



Supplementary Materials for

Accurate proteome-wide missense variant effect prediction with AlphaMissense

Jun Cheng *et al.*

Corresponding authors: Jun Cheng, jucheng@google.com; Pushmeet Kohli, pushmeet@google.com; Žiga Avsec, avsec@google.com

Science **381**, eadg7492 (2023)
DOI: 10.1126/science.adg7492

The PDF file includes:

Supplementary Note
Materials and Methods
Figs. S1 to S10
Tables S1 to S6
References

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist
Data S1 to S9

Supplementary Note

Challenges of using current clinical labels for model training and evaluation

Variant effect prediction tools often use known clinical labels (e.g. ClinVar) for training and evaluation. It is crucial to be aware of the bias in these labels so that the trained models are generalizable and the evaluations are objective. Overinflated evaluation performance is misleading and may cause further circularities for future research. Two types of data circularities were discussed by Grimm et al (11). Here we summarize four types of data leakage and two types of bias in training and evaluation with known clinical labels.

Four types of data leakage

First, leakage from training variants in the test set. Although most methods report performance on the holdout test set, different methods use different subsets of variants for training and model evaluation. In most cases the list of training variants is not available. The community also lacks a standard way to share variants in terms of genome versions, gene identifiers, and isoforms. Models might do well on the training variants by overfitting to the training labels but fail to generalize. When comparing methods with known clinical labels, it is very difficult or impossible to create a true held out set for all methods in the benchmark. This is even more complicated in the case of ensemble methods (e.g. REVEL) since the training variants of ensemble models are a union of all training variants used by all models in the ensemble, making it even more difficult to make clean training and test separation.

Second, leakage from variants in the same amino acid position. More than one missense variant can lead to amino acid substitutions at the same protein position. These variants have highly correlated pathogenicity since they share the same structural context (examples can be seen from Fig. 3D). However, existing methods mostly split training and test by genomic variant identifier, resulting in variants from the same protein position presented in both training and test. Despite not training on curated labels, we removed all protein positions seen in our genetic benchmarks from training.

Third, existing methods often do not consider data leakage from paralogous genes or homologous protein domains. This type of data leakage is hard to avoid due to the lack of a comprehensive list of paralogous genes and homologous protein domains, but needs to be considered when interpreting or comparing performance metrics with models trained on clinical labels.

Fourth, label circularity. Certain models have been widely used in clinical genetics for a long time, therefore predictions of these models likely influenced the classification labels of newly curated variants, leading to further label circularity. This type of label circularity is impossible to correct at model evaluation stage when evaluating models on clinical labels.

To quantify the first two types of data leakage, we took VARTY as an example, since the training data is available. The auROC of VARTY on the ClinVar test set dropped from 0.949 to 0.935 after removing the first and second type of data leakage in contrast to AlphaMissense (0.949 to 0.947) (**Fig. S9A**).

Two types of label bias

First, gene label bias. The ratio of pathogenic variants versus benign variants is extremely biased in ClinVar, with many genes having predominantly pathogenic or benign variants (**Fig. S9B**). A model trained on ClinVar can simply learn this bias per gene instead of actual biology. To test the performance (bias) of simply using the ratio of pathogenic variants as a predictor, we split ClinVar test set variants randomly into train (60%) and test (40%). We use the percentage of pathogenic variants per gene as a predictor ($\mathcal{G}(g)$), variants from the same gene are predicted with the same pathogenicity). If a gene is absent from the training data, the pathogenicity for variants from that gene is predicted as 1/(number of training variants). Such a simple predictor achieved auROC=0.914 on the 40% test ClinVar variants, signifying the bias level (**Fig. S9C**). If we combine the gene-based predictor with AlphaMissense as follows:

$$\mathcal{G}(g) \cdot \mathcal{P}(v),$$

where $\mathcal{P}(v)$ is the AlphaMissense score before calibration, the auROC of AlphaMissense on the 40% test set improved from 0.950 to 0.962 (**Fig. S9C**), suggesting that even a well performing model can take advantage of the gene labeling bias.

Second, the number of variants per gene is biased in curated databases (e.g. ClinVar) with a small number of well-studied genes representing most variants. Training on these databases means that the model might perform better on well-studied genes but not on others due to the lack of training data. Around 74% of ClinVar labels with at least one review star are from 20% of genes in ClinVar (1,595 genes, which is less than 10% of total human protein coding genes) (**Fig. S9D**).

Gene-level AlphaMissense predictions mimic loss of function gene constraint

We observe that gene-level average AlphaMissense pathogenicity of missense variants is correlated with a commonly-used measure of gene-level evolutionary constraint in humans, “loss of function observed/expected upper bound fraction” (LOEUF) (3). This measure is based on the relationship between the expected and observed numbers of predicted loss of function (pLoF) variants found within a population cohort. The reliability of this estimate depends on the expected number of pLoF variants, which in turn depends on the size of the coding sequence of a gene (3). As noted by the authors (3), a large fraction of genes are too small for the LOEUF measure to be statistically powered. In fact, for 22% of genes the expected number of pLoF variants is less than or equal to 10 (**Fig 4A**). We investigated whether AlphaMissense pathogenicity is informative about which genes would likely be intolerant to functionally-altering perturbations in humans, especially for genes that would be underpowered in population cohort-based approaches. To assess this, we compared the average AlphaMissense pathogenicity (across all possible missense variants in a gene) to multiple independent biological measures of a gene’s sensitivity to disruption in humans, similar to that carried out by the authors of the LOEUF measure (3) (**Methods**). We further checked that the same relationships hold for the subset of genes underpowered in LOEUF, particularly among genes with the highest average AlphaMissense pathogenicity, as these represent genes we would predict to have the most functional importance. Since AlphaMissense is trained using MSA, it is likely to have learned properties of a gene’s conservation across species. We therefore also assessed whether a conservation-based metric, PhyloP (47), has similar properties to AlphaMissense. To be

consistent with the LOEUF-based analysis, where low values indicate high gene constraint, we binned genes into deciles using negative AlphaMissense pathogenicity (similarly for PhyloP) such that low deciles correspond to higher average pathogenicity (**Methods**, **Fig. S10A**).

Average AlphaMissense pathogenicity has similar biological correlates to LOEUF (**Fig. S7**, **Supplementary Data S2 & S3**). Considering genes in the most pathogenic decile of AlphaMissense, they are depleted of rare structural deletions (mean depletion=0.21 (CI95 0.15-0.28)), are expressed in more tissues on average (mean percent=31.4% (CI95 30.8-31.9)), and are enriched for genes of known functional importance: heterozygous lethal mouse knock-out genes (enrichment=1.95, $P_{\text{hyper}}=1.9\text{e-}8$), genes essential to cell survival in cancer cell lines (enrichment=3.78, $P_{\text{hyper}}=1.09\text{e-}85$), and genes known to cause severe developmental disorders (enrichment=1.72, $P_{\text{hyper}}=5.8\text{e-}22$). Particularly striking is the relationship with 636 genes that are essential for cell survival. These genes were identified by genome-wide CRISPR screens in a variety of human cell-lines (46), and the enrichment for AlphaMissense is even stronger than LOEUF (3.78 vs. 2.27). We discuss this feature in more detail in the main text and **Fig. 4B-C**.

Among genes underpowered for reliable LOEUF scores, the enrichment of functionally important gene sets amongst the most pathogenic decile of AlphaMissense always remains statistically significant (**Fig. S7A-D**, **Supplementary Data S2**). We note the total number of underpowered genes in the Haploinsufficient gene set is small (19), precluding strong conclusions, but the enrichment remains significant amongst the larger set (166) of genes specific to severe developmental disorders (enrichment=2.54, $P_{\text{hyper}}=1.8\text{e-}9$). For the two quantitative properties (**Fig. S7E-F**, **Supplementary Data S3**) the fraction of tissues remains high (mean=32.5 (CI95 31.4-33.5)) but the depletion of structural deletions is slightly weaker than for the well-powered genes, although the confidence intervals overlap (mean depletion=0.40 (CI95 0.26-0.56)).

PhyloP also shows similar properties to LOEUF (**Fig. S7A-F**), which is unsurprising as they are correlated (Spearman correlation = -0.55, $P<2.2\text{e-}16$). However, amongst genes underpowered for LOEUF, the most conserved decile of Phylop is no longer enriched with genes of known functional importance ($P_{\text{hyper}}>=0.0025$; Bonferroni), except for cell essential genes (enrichment=4.61, $P_{\text{hyper}}=5.8\text{e-}9$). However, the enrichment is lower than for AlphaMissense (enrichment=5.91, $P_{\text{hyper}}=5.6\text{e-}46$), and many fewer underpowered genes are in the lowest decile (147 vs. 442), limiting its recall within this category of genes.

The effectiveness of AlphaMissense amongst the small genes underpowered for LOEUF is illustrated in the 10 genes with the highest average AlphaMissense pathogenicity (**Supplementary Data S4**), which all play a role in core cellular functions. For example, HIST4H4 (Uniprot_ID: P62805, Transcript_ID: ENST00000539745; average AlphaMissense pathogenicity=0.89) encodes Histone 4, a core component of the nucleosome, comprising only 103 amino acid residues. Pathogenic *de novo* missense variants have been identified in other H4 genes, all presenting with neurodevelopmental features of intellectual disability and motor and/or gross developmental delay (59). The LOEUF (1.9) for this gene would indicate an unconstrained gene, but the estimate is highly uncertain (oe_lof_lower = 0.39, oe_lof_upper = 1.9), as there are only 1.8 pLoF variants expected under neutral selection (3). Similarly, there were no constrained coding regions detected by (31) in this gene, and it is surprisingly not conserved according to the Phylop score (1.4; 6th decile). A second example of the complex SF3b is discussed in the main text (**Fig. 4**).

Together, these results imply that AlphaMissense is informative for identifying small genes of functional significance that are difficult for other *in silico* approaches. Unlike approaches that test for gene-specific depletion of observed deleterious variation in population cohorts (1, 3), (31), AlphaMissense scores are not as limited by cohort sizes or the length of a gene's coding sequence. Approaches that exploit phylogenetic relationships across many species (such as PhyloP) also share similar biological properties to AlphaMissense, likely due to the common use of MSA. However, most of the properties of PhyloP no longer hold when considering smaller genes, indicating that such approaches are also limited by gene size. We provide gene-level average AlphaMissense pathogenicity scores as part of the AlphaMissense data resource. The gene-level data used in this analysis is provided in **Supplementary Data S4**. This includes 442 small genes that AlphaMissense predicts as highly sensitive to genetic perturbations (decile bin=0), but which LOEUF is underpowered to detect (**Fig S10B**).

Limitations

AlphaMissense does not predict differences in structural coordinates for proteins harboring missense substitutions. Rather, wild-type structural predictions are used to guide the AlphaMissense model. It also does not predict the impact of missense variants explicitly on more biophysical properties such as protein stability or protein-protein binding affinity. AlphaMissense also is not trained to explicitly consider the interaction with other proteins when predicting pathogenicity. AlphaMissense currently only supports single amino acid substitutions and does not cover insertions, deletions or epistatic interaction of multiple missense variants. Since AlphaMissense relies on input MSA for structure prediction and estimating evolutionary conservation, it might have limited performance on de novo proteins lacking evolutionary information. Calibration of AlphaMissense scores is not perfect and can deviate from the expected probability on ClinVar (Fig. 2D).

Materials and Methods

Human missense variants

For our analysis, we generated and used two distinct resources. First, we generated AlphaMissense predictions for every possible single amino acid substitution within each UniProt canonical isoform in the human proteome (see “Proteome-wide predictions”). We will refer to this resource as *proteome-wide predictions*, which contains 216M protein variants (19 times protein length per protein). Working with UniProt canonical isoforms enables different analyses on different genome versions, as every genetic missense variant can be mapped to a UniProt protein variant, regardless of the genome build. This table is released alongside this publication, and was used for the MAVE analysis, as the experiments measure the effects of protein variants, regardless of whether they are a feasible missense variant (single nucleotide substitution).

Second, we generated a table of all feasible missense variants. This required a mapping from proteome to genome coordinates which we produced by matching the amino acid sequence of the UniProt canonical isoforms (used above) to the amino acid sequence corresponding to GENCODE transcript IDs (separately for HG38 v32 and HG19 v27), allowing up to 2 single amino acid differences. Transcripts from HG19 and HG38 matched with 18,761 and 19,233 UniProt canonical isoforms respectively. For each transcript, we generated all possible single nucleotide substitutions and kept those that resulted in a single amino acid substitution in their translation compared to the reference genome, excluding the start and stop codon. We refer to this resource as the *missense proteome map*. The HG19 version of the missense proteome map was only used to map the training data to UniProt IDs since the training variants are in HG19 (detailed in the Training variants section). The HG38 version of the missense proteome map was used for the remaining analysis that focused on genetic variants observed in the human population. This table contained 71,140,163 missense variants for HG38 corresponding to 19,233 gencode transcripts with a matched UniProt sequence.

For the purposes of this manuscript, the following terminology will be used wherever the distinction is relevant: “missense variants” refers to single nucleotide substitutions that result in a single amino acid substitution, and “protein variants” refers to amino acid substitutions which may or may not be from feasible missense variants. “Canonical transcripts” refers to transcripts that could be matched to a UniProt canonical isoform using the matching procedure described above.

Evaluation data

ClinVar variants processing

ClinVar missense variants (release date 2022-09-10) were first mapped to the canonical transcript. Variants with CLNSIG labels as “Benign” and “Likely benign” are considered as benign (negative), variants labeled as “Pathogenic” and “Likely pathogenic” were considered as pathogenic (positive). Variants with the term “splice” in their description were removed (29 variants), since our missense predictor is unlikely to be able to predict variants affecting splicing. In total, 100,796 variants (54,724 benign and 46,072 pathogenic) with at least one review star were considered for this study.

Validation variants

A small subset of ClinVar (30) missense variants were used as the validation set. To select a subset of the missense ClinVar variants as the validation set, we first randomly selected 300 proteins. For each protein, we then selected the maximum possible equal number of positive and negative variants. In total, 2,526 variants (1,263 pathogenic) were selected and referred to as the validation set.

Test variants

ClinVar test set

After removing the ClinVar variants used as the validation set, the remaining 82,872 missense variants were used as the test set. In contrast to the validation set, this dataset is not balanced and contains 30,884 pathogenic and 51,988 benign variants from 7,951 proteins. We provide the list of variants and our predictions as Supplementary Data S5.

Cancer hotspot mutations

Mutations from inferred cancer hotspots and negative control variants were collected from the Supplementary Table S2 of (10). The original cancer hotspot mutations were from (60) and negative control variants are rare variants from the DiscoverEHR database (61). 868 positive and 1,734 negative variants were mapped to the canonical transcripts as missense variants and were considered in this benchmark. We provide the list of variants and our predictions as Supplementary Data S6.

De novo variants from rare disease patients

We curated the same data as (12). The de novo variants from rare disease patients were a combination of de novo variants from the Deciphering Developmental Disorders cohort (DDD) patients (54) and healthy controls from (62). 353 patient variants and 57 control variants from 215 DDD related genes were mapped to the canonical transcripts as missense variants and therefore were considered. We provide the list of variants and our predictions as Supplementary Data S7.

ProteinGym

ProteinGym is a set of MAVE datasets curated and processed by (19) to assess predictors on predicting variant effects on protein fitness. The substitution benchmark consists of ~1.5M variants from 87 MAVE experiments measured with 72 proteins. Data was accessed from https://marks.hms.harvard.edu/tranception/ProteinGym_substitutions.zip (accessed 2023-01-13). We provide the list of variants and our predictions as Supplementary Data S8.

Additional MAVE datasets

To generate an additional independent MAVE benchmark data set, we searched across MaveDB (Esposito et al., 2019), MaveRegistry (Da Kuang et al., 2021), PubMed and BioRxiv. Studies were selected for inclusion if they were (1) not originally in ProteinGym; (2) covered a human protein; (3) measured a phenotype relevant to protein abundance/trafficking, protein-protein

interaction, enzymatic activity, or function; (4) had more than 100 variant measurements. MAVE data was typically reported as a table available for download from the publication or data repository. The exception was PPARG_HUMAN, for which we queried the webserver accompanying the paper (<https://miter.broadinstitute.org/>) by searching through all possible missense mutations at prevalence 1e-5. This was performed by changing the [ID] field in this URL: [https://miter.broadinstitute.org/mitergrade/?query=p.\[ID\]&prevalence=1.0e-5](https://miter.broadinstitute.org/mitergrade/?query=p.[ID]&prevalence=1.0e-5). Data sources for these additional MAVE datasets are listed in Table S5.

Training data

Training variants

Benign variants (negative class) in the training set were derived from observed variants in human and primate species. The benign training was collected from three major sources. Variants from humans were collected from gnomAD version 2.1.1. (genome version HG19). Primate variants from the Great Ape project (63) which was mapped to HG18 were obtained from <https://eichlerlab.gs.washington.edu/greatape/data/VCFs/SNPs/> and lifted over to HG19 using ENSEMBL release 100 chain file (NCBI36_to_GRCh37.chain.gz) (64). Bonobo (*Pan paniscus*) variants were from the FigShare repository (https://figshare.com/articles/dataset/Han_et_al_Data_tsv_gz/7855850) accompanying Han 2019 (65, 66). These authors uniformly processed data from previous studies (63, 67) with 59 chimpanzees and 10 bonobo genomes. These MAF thresholds were chosen such that approximately at least 2 individuals from the primate population carry the variant. Detailed statistics of benign variants from different primate species are listed in the **Table S1**. All primate variants were lifted over to the human HG19 genome and intersected with the HG19 version of the *missense proteome map* to get the corresponding UniProt IDs and GENCODE transcript ids.

We removed all training variants at protein positions where at least one variant appeared in any of the validation or test sets, except for the MAVE data. Note that for the non-MAVE validation or test datasets, the labels are potentially influenced by allele frequency. This allows us to better evaluate and calibrate model performance on amino acids where allele frequency information is lacking.

The pathogenic labels for the training set are sampled (as described in the next section) from the remaining 65,314,044 variants, out of all the possible missense variants in the *missense proteome map*, that were not observed in any primate or human population (data sources in **Table S2**).

To account for label uncertainty for the low frequency variants, which are more likely to be pathogenic than the frequent ones, we introduce per-variant loss weight coefficients. These weights multiply the classification loss such that errors on the frequent variants contribute more strongly to the training signal. Human variants are split into four bins depending on the associated MAF with logarithmically spaced bin edges over the range [1e-5, 2e-4]. A summary of how human variants are weighted in model training is listed in **Table S2**. Primate variants are given weight 1.0.

Unknown variant sampling in training

We create a balanced training set by sampling variants from the unobserved set in equal numbers as the benign set. The model is trained to classify these sampled unobserved variants as

pathogenic and therefore we refer to them as pathogenic for brevity. The probability of sampling a variant from the unobserved set depends on its trinucleotide context and the protein it belongs to. Specifically, we ensure that the pathogenic set contains the same number of variants for each trinucleotide context as the benign set. Then, for each trinucleotide context, the probability of sampling an unobserved variant depends on its protein. The probability of sampling a protein is jointly determined by the length of the protein and the abundance of the protein in the benign set. The trinucleotide context describes the nucleotide substitution and the two flanking bases and are computed as described in (12).

In practice, this is achieved by weighted sampling. The probability of sampling a variant (V) from the unobserved set (U) is dependent on its protein (G) and trinucleotide context (T) through the sampling weight:

$$\rho(V) \propto \rho(T) \rho(G | T)$$

With the first term in the sampling weight we match the number of variants per trinucleotide context between the benign and pathogenic set. The weight coefficient is the frequency ratio of that trinucleotide context in the benign set and in the unobserved set:

$$\rho(T) = \frac{n(T, B)}{n(B)} \Big/ \frac{n(T, U)}{n(U)}$$

where $n(T, B)$ denotes the counts of variants in the training set that are benign (B) and have trinucleotide context T , and $n(B)$ denotes the total number of benign variants. The symbols $n(T, U)$ and $n(U)$ denote the analogous counts for the unobserved set.

The second term in the variant sampling weight modifies the probability of sampling a protein conditioned on the trinucleotide context. With this term we balance two objectives: to match the distribution over proteins in the benign set and to maximize the diversity of training sequences. Specifically, the probability of sampling a protein G from the unobserved set given a trinucleotide context T is the average of two probabilities: one equal to the frequency of the protein in the corresponding benign set and the other proportional to the protein's sequence length:

$$\rho(G|T) = \frac{1}{2} \left(\frac{n(G, T, B)}{n(T, B)} + p(\text{length}(G) | T) \right) \frac{n(T, U)}{n(G, T, U)}$$

Here the probability $p(\text{length}(G) | T)$ is proportional to the length of G 's sequence. It is computed by dividing the length of G by the sum of the lengths of all proteins in the unobserved set U with trinucleotide context T .

We sample separate subsets of pathogenic variants mirroring the properties of each of the four benign subsets summarized in **Table S2**. Each pathogenic subset has the same size and weight in the loss function as the corresponding benign subset.

Data self-distillation

Without data self-distillation, for each trinucleotide context, all unobserved variants from the same protein have equal probability of being sampled. With data self-distillation we leverage an initial round of trained models to lower the probability of sampling likely benign variants and increase the probability of sampling likely pathogenic variants from the unobserved set. This is

obtained by extending the formula for the variant sampling weights in the previous section with a term $C(V | G, T)$:

$$\rho(V) \propto C(V|G, T) \rho(G | T) \rho(T)$$

$C(V | G, T)$ is a coefficient computed from the pathogenicity scores predicted by models trained without dataset self-distillation. This term decreases the probability of sampling likely benign variants from the unobserved set into the pathogenic set.

We trained two models without data self-distillation and computed the pathogenicity scores for all variants in the unobserved set. Then, we grouped the variants by their context T and protein G and calculated the percentile rank of each variant within each group (i.e. a number from 0 and 1, corresponding to the variant with the lowest and highest predicted pathogenicity score). Because of this grouping, the data self-distillation sampling weights do not affect the overall counts of proteins and trinucleotide contexts contained in the pathogenic set. In other words, the coefficient $C(V | G, T)$ is the percentile rank of variant V , among all variants with the same G and T , averaged across the two models.

Data pipeline

The data pipeline of AlphaMissense takes a table of variants as inputs and produces input features for the model. The data pipeline is largely the same with AlphaFold. Below we describe in detail the data pipeline of AlphaMissense and refer to (21) whenever the process is shared:

Spatial cropping

We use maximum sequence length (crop size) $N_{\text{res}}=256$ for both training and inference. Protein sequences longer than 256 are cropped to keep 256 residual spatially closest to the sampled variant. The spatial distance is defined as the minimum Euclidean distance between any atom of the residue and any atom of variant residue. Protein structures are AlphaFold predictions from the AlphaFold Protein Structure Database. During inference, in case the structure of the target sequence is unknown, we crop to the nearest 256 residues around the variant in the sequence space.

Variant MSA row

As in the original AF architecture, the first row of the MSA corresponds to the wild-type sequence for which the structure is computed. We repeat this sequence in the second row of the MSA with N residues masked, where N differs between training and inference. In both cases we first sample a single variant (i.e. the central variant) and crop the sequence as described in the previous section. During training, we then query the training set for other variants for the same gene and within the cropped region. Of these, up to 49 additional variants with unique amino acid positions are sampled. The position of all N variants are masked in the second row of the MSA. Therefore, the model has access to the wild-type sequence but does not observe the specific amino acid variants whose pathogenicity it is tasked to predict.

MSA filtering

Genetic search was performed the same as described in Jumper et al. At the final inference run, we run each model twice, once with MSA diversity filter and once with only random sampling. HHFilter (68) was used for diversity with the following parameters:

```
--diff 1000 --cov 50 --id 95.
```

MSA masking

Masking was applied to all the variants in the second row of the MSA. During training, all other rows of the MSA are masked and perturbed with the same probabilities as in the AF paper for training the masked MSA head. During inference, only variant positions in the second MSA row are masked as described above.

Variant process and featurization

For every sampled variant, the input features required by AlphaFold are derived from the reference sequence as described in (21). Most input features are the same as AlphaFold except: 1. We are not using cluster_profile, cluster_has_deletion, cluster_deletion_value, cluster_deletion_mean, and all template related features. 2. We moved MSA masking from the model into the data pipeline, therefore masking for the masked language model task is precomputed and does not change in recycling. We summarize the features we added compared to AlphaFold in **Table S3**.

Similar to AlphaFold, input features are grouped as follow to the model:

- **target_feat** Same as AlphaFold. This is the feature of size $[L, 21]$ consisting of the “aatype” feature.
- **residue_index** Same as AlphaFold. This is the feature of size $[L]$ consisting of the “residue_index” feature.
- **msa_feat** This only contains the cluster_msa feature. Same as AlphaFold, N_{cycle} random samples of this feature were drawn and provided to the network for each recycling iteration.
- **msa_profile** Average amino acid frequencies for all positions in the alignment.

AlphaMissense model

The model predicts variant pathogenicity classification logits from features derived from the reference amino-acid sequence and MSA as outlined in Algorithm 1. Because most of the model architecture is identical to that of AF, we will detail the few differences in the following sections.

Algorithm 1 AlphaMissense Model Inference

```

def Inference  $(\{\mathbf{f}_i^{\text{aatype}}\}, \{f_i^{\text{residue\_index}}\}, \{\mathbf{f}_{s_a i}^{\text{msa}}\}, \{\mathbf{f}_i^{\text{msa\_profile}}\})$  :
    # Recycling iterations, the parameters are shared across iterations:
    1:  $\hat{\mathbf{z}}_{ij}^{\text{prev}}, \vec{\mathbf{x}}_i^{\text{prev,C}^\beta} = \mathbf{0}, \vec{\mathbf{0}}$ 
    2: for all  $c \in [1, \dots, N_{\text{cycle}}]$  do  $N_{\text{cycle}} = 4$ 
        # Sample a subset of the full MSA:
        3:  $\{\mathbf{f}_{s_i}^{\text{msa\_feat}}\}, \{\mathbf{f}_{s_e i}^{\text{extra\_msa\_feat}}\} = \text{SampleMSA}(\{\mathbf{f}_{s_a i}^{\text{msa}}\})$ 
        # Initialize the MSA and pair representations:
        4:  $\{\mathbf{m}_{s_i}\}, \{\mathbf{z}_{ij}\} = \text{InputEmbedder}(\{\mathbf{f}_i^{\text{aatype}}\}, \{f_i^{\text{residue\_index}}\}, \{\mathbf{f}_i^{\text{msa\_profile}}\}, \{\mathbf{f}_{s_i}^{\text{msa\_feat}}\})$ 
        # Inject previous outputs for recycling:
        5:  $\{\mathbf{z}_{ij}\} += \text{RecyclingEmbedder}(\{\hat{\mathbf{z}}_{ij}^{\text{prev}}\}, \{\vec{\mathbf{x}}_i^{\text{prev,C}^\beta}\})$ 
        # Embed extra MSA features:
        6:  $\{\mathbf{a}_{s_e i}\} \leftarrow \{\mathbf{f}_{s_e i}^{\text{extra\_msa\_feat}}\}$ 
        7:  $\mathbf{e}_{s_e i} = \text{Linear}(\mathbf{a}_{s_e i})$   $\mathbf{e}_{s_e i} \in \mathbb{R}^{K_{\text{extra}}}, K_{\text{extra}} = 64$ 
        8:  $\{\mathbf{z}_{ij}\} \leftarrow \text{ExtraMsaStack}(\{\mathbf{e}_{s_e i}\}, \{\mathbf{z}_{ij}\})$ 
        # Main trunk of the network:
        9:  $\{\mathbf{m}_{s_i}\}, \{\mathbf{z}_{ij}\}, \{\mathbf{s}_i\} \leftarrow \text{EvoformerStack}(\{\mathbf{m}_{s_i}\}, \{\mathbf{z}_{ij}\})$ 
        # Update structure and pair representation:
        10:  $\{\vec{\mathbf{x}}_i^{\text{prev,C}^\beta}\} \leftarrow \text{StructureModule}(\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\})$ 
        11:  $\{\hat{\mathbf{z}}_{ij}^{\text{prev}}\} \leftarrow \{\mathbf{z}_{ij}\}$ 
    12: end for
    # Variant effect prediction at position  $i$  variant  $a$ :
    13:  $\{\mathbf{p}_i\} = \text{SoftMax}(\text{Linear}(\{\mathbf{m}_{2i}\}))$   $\mathbf{p}_i \in \mathbb{R}^{21}$ 
    14:  $s_i^a = \log p_i^{\text{ref}} - \log p_i^a$ 
    15: return  $s_i^a$ 

```

Algorithm 2 MSA sampling

```

def SampleMSA( $\{\mathbf{f}_{s_a i}^{\text{msa}}\}$ ) :
    # Uniform random shuffle of the full MSA. During fine-tuning the first and second rows, which contain
    # respectively the reference and masked variant amino acid sequences, are kept in their place.
    1:  $\{\mathbf{f}_{s_a i}^{\text{msa}}\} \leftarrow \text{RandomShuffle}(\{\mathbf{f}_{s_a i}^{\text{msa}}\})$ 
    2:  $\{\mathbf{f}_{s_i}^{\text{msa\_feat}}\} \leftarrow \{\mathbf{f}_{s_i}^{\text{msa}}\}, s' \in [1, \dots, N_{\text{msa}}]$ 
    3:  $\{\mathbf{f}_{s_e i}^{\text{extra\_msa\_feat}}\} \leftarrow \{\mathbf{f}_{s_e i}^{\text{msa}}\}, s' \in [N_{\text{msa}} + 1, \dots, N_{\text{msa}} + N_{\text{extra}}]$ 
    4: return  $\{\mathbf{f}_{s_i}^{\text{msa\_feat}}\}, \{\mathbf{f}_{s_e i}^{\text{extra\_msa\_feat}}\}$ 

```

MSA sampling

Unlike in AF, the MSA masks are computed only once (in the data pipeline, not shown here) rather than anew for every recycling iteration. This was done to prevent the model from identifying the masked residues from the recycled embeddings. Algorithm 2 describes sampling the N_{msa} sequences used as MSA features and the $N_{\text{extra}}=1152$ extra MSA sequences from the input MSA, which consists of up to $N_{\text{all}}=2048$ sequences. N_{msa} is set to 128 during pre-training and 256 during fine-tuning. Moreover, during fine-tuning the first and second row of the MSA, which contain the reference sequence and the masked variants, are never shuffled and therefore always included in the MSA features.

Input embeddings

During each round of recycling, the InputEmbedder, described in Algorithm 3 and by **Fig. S1**, initializes the MSA representation $\{\mathbf{m}_{si}\}$ and the pair representation $\{\mathbf{z}_{ij}\}$. Unlike AF, we do not add an embedding of the target sequence to the MSA representation $\{\mathbf{m}_{si}\}$. Instead, during pre-training, the MSA representation is initialized from the MSA features alone. During fine-tuning, we add a non-pre-trained embedding of the MSA profile (zero-initialized). Likewise, only during fine-tuning, we add an embedding of the MSA profile to the pair representation.

Algorithm 3 Embeddings for initial representations

```
def InputEmbedder( $\{\mathbf{f}_i^{\text{target\_feat}}\}, \{f_i^{\text{residue\_index}}\}, \{\mathbf{f}_{si}^{\text{msa\_feat}}\} \{\mathbf{f}_i^{\text{msa\_profile}}\}$ ) :
    1:  $\mathbf{a}_i, \mathbf{b}_i = \text{Linear}(\mathbf{f}_i^{\text{target\_feat}})$   $\mathbf{a}_i, \mathbf{b}_i \in \mathbb{R}^{K_{\text{pair}}}, K_{\text{pair}} = 128$ 
    2:  $\mathbf{c}_i, \mathbf{d}_i = \text{Linear}(\mathbf{f}_i^{\text{msa\_profile}})$   $\mathbf{c}_i, \mathbf{d}_i \in \mathbb{R}^{K_{\text{pair}}}, K_{\text{pair}} = 128$ 
    3:  $\mathbf{z}_{ij} = \mathbf{a}_i + \mathbf{b}_j + \mathbf{c}_j + \mathbf{d}_j$   $\mathbf{z}_{ij} \in \mathbb{R}^{K_{\text{pair}}}, K_{\text{pair}} = 128$ 
    4:  $\{\mathbf{z}_{ij}\} += \text{relpos}(\{f_i^{\text{residue\_index}}\})$ 
    5:  $\mathbf{m}_{si} = \text{Linear}(\mathbf{f}_{si}^{\text{msa\_feat}}) + \text{Linear}(\mathbf{f}_i^{\text{msa\_profile}})$   $\mathbf{m}_{si} \in \mathbb{R}^{K_{\text{msa}}}, K_{\text{msa}} = 256$ 
    6: return  $\{\mathbf{m}_{si}\}, \{\mathbf{z}_{ij}\}$ 
```

Recycled embeddings

The embeddings are zero-initialized and iteratively updated over $N_{\text{cycle}}=4$ recycling iterations. In contrast to AF, we recycle only the pair representation $\{\mathbf{z}_{ij}\}$ and atom coordinates but not the MSA representation $\{\mathbf{m}_{si}\}$.

Evoformer stack

We do not use templates neither during structure pre-training nor during variant fine-tuning. Moreover, not all of the models use the extra MSA stack. The single representation $\{\mathbf{s}_i\}$, which is used for the structure module, is taken from the diagonal of the pair representation $\{\mathbf{z}_{ii}\}$, instead of the first row of the MSA representation.

Variant pathogenicity prediction

During structure training, masked MSA reconstruction was trained as an auxiliary task. During fine-tuning, we warm-start the pathogenicity prediction output layer parameters from those learned for the masked MSA task. Using the same notation from the AF paper, the pathogenicity logit score of a variant in position i along the sequence from reference amino acid (ref) to an alternative amino acid (a) is estimated as:

$$s_i^a = \log p_i^{\text{ref}} - \log p_i^a$$

The pathogenicity score of variant combinations is estimated as the sum of scores for each variant computed independently.

During training, multiple masked variants are sampled and included in the variant embedding. Each variant is associated with a label y_i (0 for benign and 1 for the variants sampled from the unobserved set), and a weight w_i . The training loss function for the pathogenicity logit scores is:

$$\mathcal{L}_{\text{variant}} = \frac{1}{N} \sum_i w_i (y_i \log[\sigma(s_i|_{>-1})] + (1 - y_i) \log[1 - \sigma(s_i|_{>0})])$$

Here, $x|_{>y}$ denotes $\max(x, y)$ and σ is the sigmoid function.

Model training details

We independently trained three AlphaFold models and fine-tuned them independently on variants. We followed the training procedure described in (21), (only the “Initial training” stage) without structure violation loss and “Experimentally resolved” loss. We increased the weight of the masked MSA task to 10.0 with other loss weights unmodified during AF training. The loss weights at the fine-tuning stage are as follow:

$$\mathcal{L} = 0.1\mathcal{L}_{\text{FAPE}} + 0.1\mathcal{L}_{\text{aux}} + 0.1\mathcal{L}_{\text{dist}} + 10.0\mathcal{L}_{\text{variant}}$$

We use Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e-5$ for both for AF training and fine-tuning, with gradient clipped to global norm at 0.1, and exponential weight decay. The parameters used for evaluation, and checkpointed for inference and later use, are the exponential moving average of the trained parameters with weight coefficient 0.999.

AF training is carried out for about 7e6 steps on single-chain structures (**Table S4**). 75% of the training structures are self-distillation data sampled from MGnify (69) and UniRef90 (70) with structures predicted by AF. Fine-tuning is carried out until auROC of the evaluation set converges (about 350k samples, each training sample contains maximum 50 variants). We first train the parameters in the output heads for 1000 steps with all other parameters frozen. Moreover, we apply weight decay toward the pre-trained AF weights with coefficient 1e-2. All parameters associated with input embeddings or output heads are exempted from this weight decay.

Calibrating predictions

The training procedure thus described produces a classifier that is accurate (i.e. the logits of benign and pathogenic variants are well separated) but not calibrated (i.e. the pathogenicity

probability out of the box does not approximate the actual label probability). We obtained calibrated predictions by training a linear logistic rescaling function that maps the output logits to a probability score which we refer to as AlphaMissense (pathogenicity) score throughout the manuscript:

$$\tilde{s} = \sigma(c_1 s + c_0)$$

where σ is the sigmoid function and c_1 and c_0 are the two scalar parameters. The model is trained by logistic regression on the ClinVar validation set (described above). Because the rescaling is monotonic it does not affect the accuracy of the classifier which is measured by scale-invariant metrics like auROC and Spearman correlation. The calibration is evaluated on the class-balanced test set and shown in **Fig. 2D**.

Finally, we label variants by their AlphaMissense score as likely benign, likely pathogenic, or ambiguous. The corresponding classification thresholds were derived from the precision curves (**Fig. S4A**) on the (non-balanced) ClinVar test set. We label variants as likely benign if their AlphaMissense score is below the threshold that obtains 90% negative-class (benign) precision. Similarly, a variant is considered likely pathogenic if its AlphaMissense score is above the threshold that obtains 90% positive-class precision.

Proteome-wide predictions

We made predictions for all 216M protein variants described above in the “Human missense variants” section (*proteome-wide predictions*). Specifically, for each protein variant, we extracted a 256 amino acid sequence context around the variant, matching the method for model inference described above. MSA queries were generated based on the whole protein sequence for proteins shorter than 2700 amino acids. For proteins longer than 2700 aa, we fragmented the protein into a maximum length of 1400 amino acid fragments with 200 aa overlaps. MSA were queried based on the fragments. Variant with residue index i is mapped to fragment f_i , where $f_i = \max(0, (i-500)/200) + 1$. We used the same MSA query as described in (21). For each variant, we obtained predictions from all 3 models in the ensemble. Each model was run twice with and without HHFilter (-diff 1000, -cov 50, -id 95). The final predictions were averages of all 6 predictions on a logit scale.

Using the same strategy, we made predictions for 400M variants from 60k non-canonical isoforms.

To provide predictions for all feasible missense variants, the *proteome-wide predictions* were merged with the *missense proteome map* (see “Human missense variants” section) by UniProt ID and protein variant id.

With the *missense proteome map*, we computed the mean AlphaMissense pathogenicity per gene (transcript id) by taking the average predicted pathogenicity for all possible missense variants.

Evaluation on the clinical benchmarks

In order to compare performance of AlphaMissense to other published models, we accessed scores from dbNSFP (71) (v4.3; via <https://sites.google.com/site/jpopgen/dbNSFP>) for the following models: SIFT (72), Polyphen2 (HDIV and HVAR) (7), PrimateAI (12), CADD (73),

Eigen (74), and REVEL (8). VARTY_R_LOO (9) scores were accessed via <http://varity.varianteffect.org/>. Predictions of gMVP (10) were accessed via the link provided at <https://github.com/ShenLab/gMVP>. Predictions of EVE (17) for 3,219 proteins were downloaded from evemodel.org. Proteome-wide predictions of ESM1b (28) were accessed via https://huggingface.co/spaces/ntranoslabs/esm_variants/tree/main (75).

Predictions of ESM1v (18) were generated with the public available code and model weights from <https://github.com/facebookresearch/esm>. The final proteome-wide predictions are average predictions of 5 models made with the masked marginals method implemented by the authors. For input sequences longer than 1022 aa, we followed a weighted slicing window approach described in (75) under the section “Handling long sequences”.

Scores from all methods were added to the *missense proteome map* by UniProt ID, protein variant position and alternative amino acid (*protein variant id*). In most cases (SIFT, Polyphen2, CADD, Eigen, REVEL, VARTY_R_LOO, ESM1b, ESM1v, EVE, AlphaMissense), the UniProt IDs for which scores were provided matched the canonical transcript, and we were able to rely on the protein variant id provided. In the case of gMVP, there was insufficient proteome coverage when using canonical transcripts only. In this case, we mapped the scores to the canonical UniProt transcript by genome coordinates before integrating into the *missense proteome map*. For all methods, in the case that multiple scores matched to the same *protein variant id* (e.g. variants from the same codon created the same alternative amino acid) the most pathogenic prediction was used. The *protein variant ids* were used to match scores with the benchmarking variants.

Significance levels and confidence intervals when comparing the performance metrics of two models were estimated using bootstrap resampling. Let t_1 be the performance metric (auROC) AlphaMissense and t_2 be the performance metrics of a competing model, the test statistics to be considered is $d=t_1-t_2$. We sampled with replacement the test data (e.g. ClinVar variants) $B=999$ times and each time (i) computed the bootstrapped metric difference d_i . The one-sided P value was approximated as:

$$P = \frac{1 + \#\{d_i \leq 0; i = 1 \dots B\}}{B + 1}$$

The confidence intervals of performance metrics were estimated with a percentile bootstrap approach. We generated 1000 bootstrap datasets of the same size by sampling test data with replacement. For a given performance metric t , the reported 95% confidence interval is $(t_{0.025}, t_{0.975})$, where $t_{0.025}$ and $t_{0.975}$ are the 2.5 and the 97.5 percentiles of the bootstrap resamples, respectively.

Evaluation on the MAVE benchmarks

ProteinGym

AlphaMissense predictions were generated with the reference amino acid sequences used in the assays (accessed from <https://github.com/OATML-MarksLab/Tranception>). Note that these reference sequences may differ from the wild-type sequences from UniProt. When scoring for variant combinations in the ProteinGym benchmark, we score each variant separately assuming independence between variants and sum up the predicted raw scores to represent the effect of variant combinations. This approach is applied for all methods unless the predictions are already

provided by ProteinGym. The absolute value of the Spearman correlation between predicted and observed assay scores were calculated per MAVE experiment then averaged by UniProt ID. These performance metrics were reported alongside the performance metrics reported by ProteinGym for all other models.

We note that the absolute level of performance for MAVE (0.514 Spearman correlation on ProteinGym) is much lower compared to clinical benchmarks such as ClinVar (0.95 auROC) because (i) the intrinsic noise of MAVE assays (the correlation between two different TP53 assays in ProteinGym is around 0.6 to 0.75) and (ii) the task is more challenging since the whole spectrum of variant pathogenicity is profiled, not just the clear pathogenic or benign cases. Spearman correlation puts a lot of emphasis on ranking the intermediate predictions correctly.

Additional MAVE benchmark

Evaluations on the additional MAVE benchmark (**Table S5**) were calculated as described above for ProteinGym. The absolute value of the Spearman correlation between predicted and observed assay scores were calculated per MAVE experiment then averaged by UniProt ID. For the RASK_HUMAN_binding data, we used the average variant effects over eight binding screens reported by the authors (76).

For the supplemental evaluations (**Fig. S5**), we aggregated results by either (i) averaging all results grouped by amino acid position or (ii) averaging all results grouped by the alternate amino acid generated by the mutation. We then calculated the Spearman correlation between the aggregated results.

Additional resources

The ACMG geneset was derived from <https://www.gimjournal.org/cms/10.1016/j.gim.2022.04.006/attachment/2f5b52ca-8ac2-49f8-b4a5-3bb434efad2c/mmc1.xlsx>. The MAVE prioritized gene list was derived from https://github.com/rothlab/mave-gene-prioritization/blob/master/ranked_clinvar_genes.csv.

The SHOC2 amino acid annotations were taken from https://github.com/jkwonbio/Structure-function-analysis-of-the-SHOC2-MRAS-PP1C-holophosphatase-complex/blob/main/data/contact_v_nocontact.csv. The SHOC2 domain annotations were taken from https://github.com/jkwonbio/Structure-function-analysis-of-the-SHOC2-MRAS-PP1C-holophosphatase-complex/blob/main/data/Correct_LRRnum_CIRCpos.csv

The GCK in vitro activities were taken from Table 2, column ‘The Relative Activity Index’ of (56). The GCK clinical labels were taken from (45).

Ablation studies

We estimate the relative contribution of key model architecture and dataset preparation choices to the final model performance with an ablation study. For each ablation we average the scores of six inference runs (produced with three models, each with and without HHFilter), generated with the same training hyper-parameters and ensembling strategy as the reference model. Unless otherwise stated in the description of the ablation, we warm-start the ablated models from the three pre-trained parameter sets used for the reference models. We evaluate the models on

ClinVar and ProteinGym. We compare the reference model, AlphaMissense, to the following ablations:

- **No fine-tuning on missense variants.** These models most closely approximate the performance of AF on the benchmarks.
- Three ensemble models where protein structure information is either not available during fine-tuning, or the model is not required to preserve it (marked by purple bars in **Fig. S6**):
 - **No AF pre-training.** We do not warm-start the models with parameters pre-trained for structure prediction and masked-MSA reconstruction.
 - **No structure loss during AF pre-training.** We warm-start the models with parameters pre-trained only for masked-MSA reconstruction. The structure loss is also not used during fine-tuning.
 - **Finetune without structure loss.** We disable the structure loss only during fine-tuning. This ablation expectedly produces a minor change in performance because structural context is available at the start of fine-tuning.
- Five ensemble models where we modify data pre-processing steps (marked by teal bars in **Fig. S6**):
 - **Train with 1 variant in MSA at the time.** During fine-tuning, we apply only $N = 1$ variant to the reference sequence in the second row of the MSA.
 - **$p(G | T) = 1$ when sampling the pathogenic set.** When sampling unobserved variants to form the pathogenic set, we do not introduce per-protein importance weights, see the section “Unknown variant sampling in training”.
 - **No MAF-based loss function weights.** The loss function computed for variants with $MAF > 1e-3$ and those for $MAF < 1e-3$ have the same weight.
 - **No dataset distillation.** When sampling unobserved variants to form the pathogenic set, we do not lower the probability of sampling variants that are classified as likely benign by previously trained models, i.e. we set $C(V | G, T_i) = 1$, see section “Data self-distillation”.
 - **Uniform sampling the pathogenic set.** The variants in the pathogenic set are sampled without importance weights and without self-distillation.
- Two ensemble models where we reduce the sources of training data (marked by orange bars in **Fig. S6**):
 - **No benign variants from primates.** We don’t include benign variants derived from primate genome sequences in the training data.
 - **No low-MAF variants.** We don’t include human variants with $MAF < 1e-3$ as benign samples in the training set.

Characterizing predictions

For characterization of predictions, we used the proteome-wide predictions of all possible missense variants (‘missense proteome map’).

To analyze the depth of non-redundant entries in the MSA we computed the number of effective sequences (N_{eff}) (77) with a similarity threshold of 0.8. For per-residue estimates we only considered rows where that residue position is non-gap in the sequence (21).

To get the per-residue disorder predictions we used the AlphaFold “experimentally resolved head”, trained to determine whether a residue would be resolved in an experimental structure. It has shown to perform as a competitive disorder predictor on the CAID benchmark dataset (25).

We obtained the allele frequency (AF) data from gnomAD version 3.1.1. For MAF we used the global allele frequency provided in the variant sites vcf files (url: <https://storage.googleapis.com/gcp-public-data--gnomad/release/3.1.1/vcf/genomes/gnomad.genomes.v3.1.1.sites.chr1.vcf.bgz>) and used 1-AF where AF>0.5. We linked gnomAD variants to our *missense proteome map* using genome variant ID.

For gene constraint we used the “loss-of-function observed/expected upper bound fraction” (LOEUF) metric available in gnomAD v2.1.1. Specifically, we used the ‘oe_lof_upper’ column in the table https://storage.googleapis.com/gcp-public-data--gnomad/release/2.1.1/constraint/gnomad.v2.1.1.of_metrics.by_transcript.txt.bgz. We matched this to a corresponding protein in our *missense proteome map* using transcript ID (16,975 unique transcripts/Uniprot IDs), and plotted the average pathogenicity score per transcript.

Comparison with properties of LOEUF

To establish whether the average AlphaMissense pathogenicity score per gene captures similar properties to a widely-used measure of gene constraint (LOEUF), we adapted the code provided by the authors of (3) (https://github.com/broadinstitute/gnomad_lof/tree, accessed March 2023)). Specifically, using the missense proteome map we computed the average AlphaMissense pathogenicity score for 19,233 unique transcripts and binned these into deciles (**Fig. S10A**). Using set of 16,975 transcripts that could be matched to the gnomAD transcript-level constraint data file (see above; matching on transcript ID), and the predefined decile bins for LOEUF (‘oe_lof_upper_bin’), we analysed the same gene lists and computed the quantitative features as reported in Figures 3 and 4b of (3), with one additional gene set (**Table S6**). One of these lists - known haploinsufficient disease-causing genes - is relatively small (only 19 genes in the ‘underpowered’ category), so we additionally analyzed a set of genes with strong clinical evidence that variation in these genes cause severe developmental disorders. Specifically, the developmental disorder gene panel (DDG2P) curated by the Gene2Phenotype project, as of 28th April 2023 (78). We further subset to the genes where the ‘confidence category’ is either ‘definitive’ or ‘strong’ (see <https://www.ebi.ac.uk/gene2phenotype/terminology> for details), and which could be matched with the 16,975 genes used in our analysis.

We defined the set of genes as ‘underpowered for LOEUF’ as those with ‘exp_lof’ <=10. This threshold was defined by the authors of (3) as genes with insufficient statistical power to be classified in the most constrained decile in a cohort of 125K individuals. As the distribution of genes in the pre-defined decile bins are not necessarily uniform after subsetting to different genes, we also computed the fraction of all genes in each decile bin, shown as gray horizontal bars in bar plots in **Fig. 4C** and **Fig. S7**. To quantify the enrichment of functional gene sets amongst a given decile we computed the proportion of the given gene set within the decile and the proportion of all genes within the decile. We defined enrichment as the ratio of these two values and computed p-values using the hypergeometric test.

Spearman correlations between LOEUF, average AlphaMissense pathogenicity, and average PhyloP were computed using the set of transcripts overlapping with the GnomAD resource as described above, and p-values calculated using a two-sided test for association.

Evaluating performance on gene essentiality prediction

To assess the performance of gene-level scores at classifying cell essential / non-essential genes (**Fig. 4B**) we used a list of common essential genes sourced from [DepMap.org](#). DepMap resources were queried from the DepMap Public 23Q2 version of the dataset (79) using their downloads page (<https://depmap.org/portal/download/all/>). For the classification of cell essential genes, the negative set was taken from the AchillesNonessentialControls.csv file, while the positive set was taken from the AchillesCommonEssentialControls.csv file, containing a union of (80) and (81). We mapped the above genes to the GnomAD resource using the DepMap_ID column. This ID is formatted as {hugo_symbol} ({entrez_id}), where {hugo_symbol} and {entrez_id} are columns from the HUGO coding gene table, downloaded in May 2023 from their URL

(https://ftp.ebi.ac.uk/pub/databases/genenames/hgnc/tsv/locus_types/gene_with_protein_product.txt). We partitioned genes into well-powered (exp_lof>10) and underpowered (exp_lof<=10) groups before calculating auROC (sklearn.metrics.roc_auc_score). Bootstrapping was performed as described above with 999 resamples.

Cell essentiality of individual SF3b complex proteins was queried using the DepMap Public 23Q2 portal, which designates them all as “Common essential”.

Complex traits analysis

To assess the relationship between AlphaMissense predictions and rare variants associated with complex traits, we used genotype/phenotype association statistics from a recent analysis of UK Biobank exomes (52). Specifically, for **Fig. 5C** we used the variant-level association statistics data (gs://ukbb-exome-public/500k/qc/variant_qc_metrics_ukb_exomes_500k.mt).

To this data we applied the following filters to generate a subset of reliable association statistics:

1. Keep only rows (variants) with all of:
 - a. ‘Keep_var_expected_ac’=True
 - b. ‘Keep_var_annt’=True
 - c. MAF < 0.01 where MAF was derived from ‘call_stats.AF’.
2. Keep only entries (variants/phenotype pairs) with ‘keep_entry_expected_ac’

This resulted in association statistics for 210,998 variants. We further partitioned variants into groups using the annotations defined by (52) (see STAR*Methods “Annotations”): ‘missense|LC’ (129,344), ‘pLoF’ (5,633) and ‘synonymous’ (76,011). For missense variants we further partitioned them into three groups defined by AlphaMissense predictions using the calibrated thresholds as defined above. Predictions were linked to the UK Biobank association data by merging on our *missense proteome map* using genetic variant IDs (117,709 unique variants). Of these genetic variants, 97 correspond to more than one protein variant, and for these we used the average AlphaMissense pathogenicity score. These data are provided as **Supplementary Data S9**. For each class of variants (e.g. pLoF) we computed the percentage of variants with at least one significantly associated phenotype. Error bars on plots show 95%

confidence intervals, which were approximated using the binomial distribution (`scipy.stats.binomtest.proportion_ci`). In **Fig. 5C** we use $p\text{-value} = 1\text{e-}5$, but the result is robust to the p -value threshold used (**Fig. S8A**). Minor allele frequency ranges (MAF) shown in **Fig. S8A** are derived from $\min(\text{AF}, 1\text{-AF})$ where AF is the ‘`call_stats.AF`’ column provided in the association statistics data described above.

For the analysis shown in **Fig. S8B** we used the same data but only applied filter 1c., and considered only pLoF (493,104 variants) and ‘missense|LC’ variants classified as ‘likely pathogenic’ by AlphaMissense (1,081,581 variants). Following (52), we considered genes with a cohort-wide cumulative minor allele count (MAC) > 50 as likely having sufficient numbers of carriers for producing reliable statistics in an association test. However, we note that for a binary case/control trait this will also depend on the cumulative allele count among *cases* within the cohort. Such traits would therefore likely require a larger cohort-wide cumulative MAC to meet the statistical requirement (e.g. 500 for a disease of 10% prevalence). In **Fig S7B** we show the boost in the number of testable genes for different cohort-wide cumulative MAC thresholds.

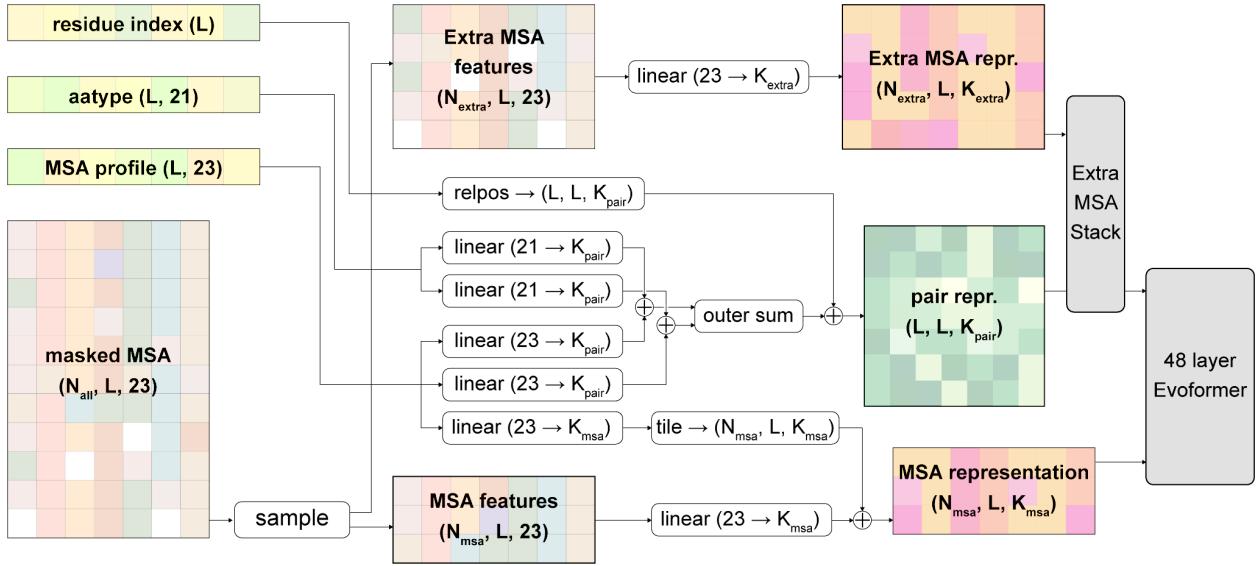


Fig. S1.

Partial diagram of the input feature embeddings of AlphaMissense. Recycling the pair representation and positions is omitted). We note that not all models employ the extra MSA stack, as summarized in **Table S4**.

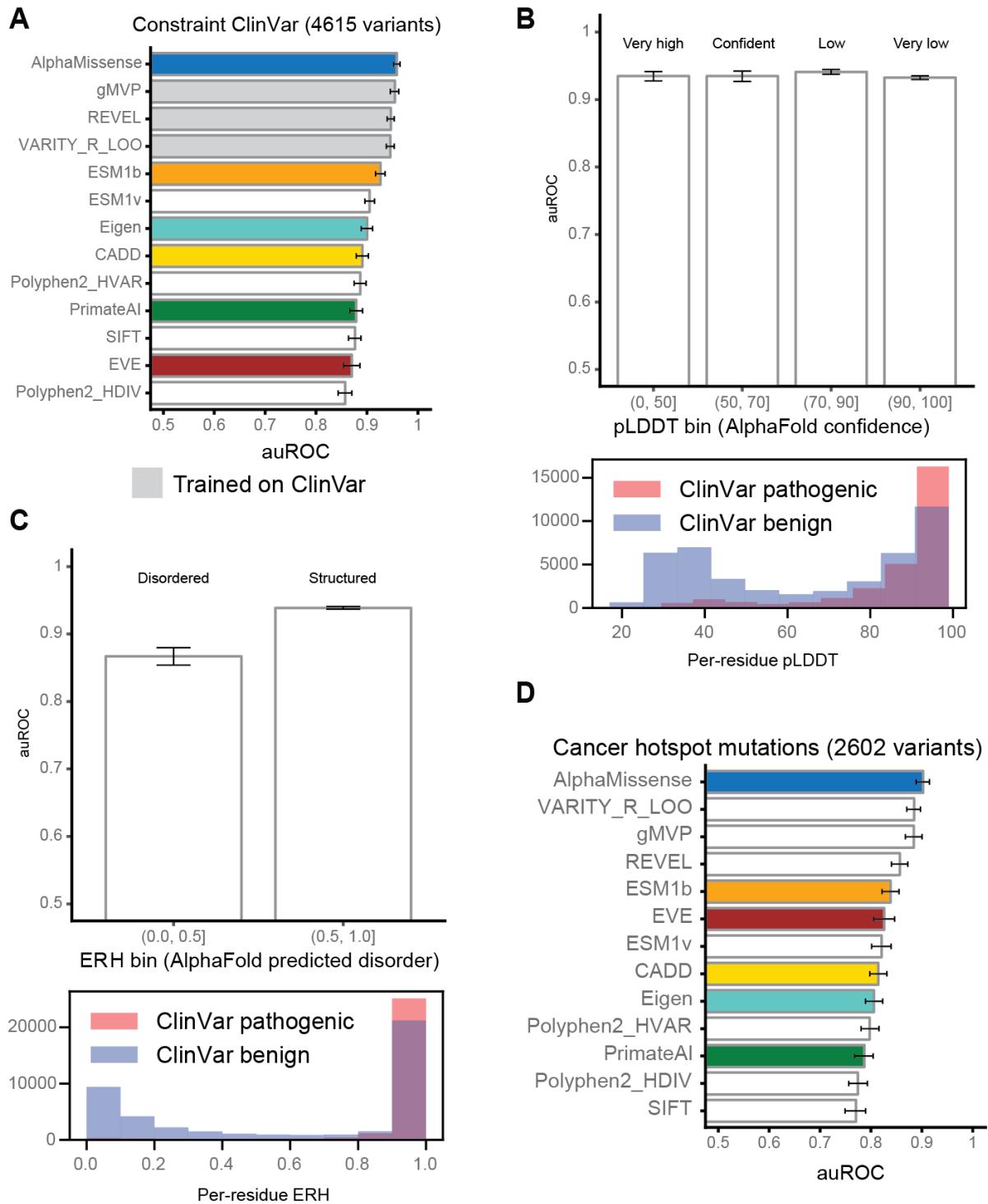


Fig. S2.

Performance of AlphaMissense on ClinVar variants compared to other methods. Methods that used ClinVar in their training data are highlighted in gray. As their performance is inflated by data leakage and label circularity, they are not considered in the performance comparison.

- A. Performance on classification of ClinVar (3,430 pathogenic and 1,185 benign) variants on regions of high evolutionary constraint (>95th percentile of constrained coding regions as

defined by (31)). Error bars show the 95% confidence interval of 1000 bootstrap resamples (Methods).

- B. Barplot of AlphaMissense auROC for variants at residues with different AlphaFold confidence levels (upper). Error bars show the 95% confidence interval of 1000 bootstrap resamples. Histogram of pathogenic (red) and benign (blue) residue pLDDT of ClinVar test variants (lower).
- C. Barplot of AlphaMissense auROC for variants at AlphaFold predicted disordered region and structured region (upper). AlphaFold experimental resolved head (ERH) predictions are used as a protein disorder predictor, ERH prediction larger than 0.5 is predicted to be structured and smaller than 0.5 is predicted to be disordered (25). Error bars show the 95% confidence interval of 1000 bootstrap resamples. Histogram of pathogenic (red) and benign (blue) residue ERH of ClinVar test variants (lower).
- D. Comparison of AlphaMissense and other predictors on 868 missense variants in hotspots from 202 cancer driver genes (60) versus 1,734 randomly selected negative variants from the DiscovEHR data (61) sourced from (10).

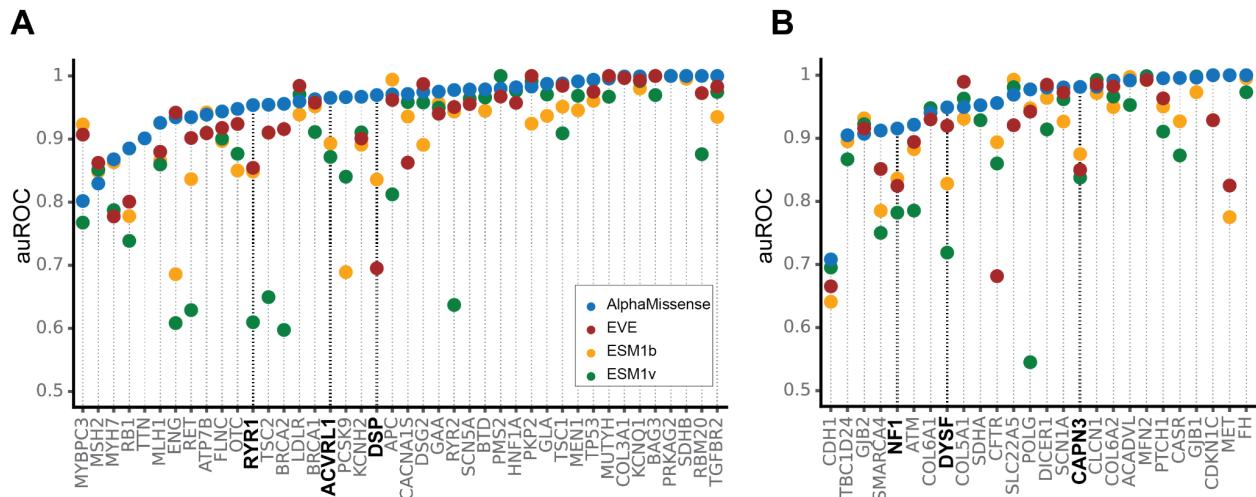


Fig. S3.

Comparing AlphaMissense and other methods on classifying ClinVar test variants per gene. Only proteins with at least 5 benign and pathogenic ClinVar test variants are considered. Methods that used ClinVar in their training data are highlighted in gray. As their performance is inflated by data leakage and label circularity, they are not considered in this comparison.

- A. Performance on ClinVar variants for ACMG clinically actionable genes (32). Genes with at least 5 benign and 5 pathogenic ClinVar labels are shown. Genes shown in Fig. 2F are indicated in bold.
- B. Performance on ClinVar variants for genes prioritized for MAVE (33), which were not already included in A. Genes with at least 5 benign and 5 pathogenic ClinVar labels are shown. Genes shown in Fig. 2G are indicated in bold.

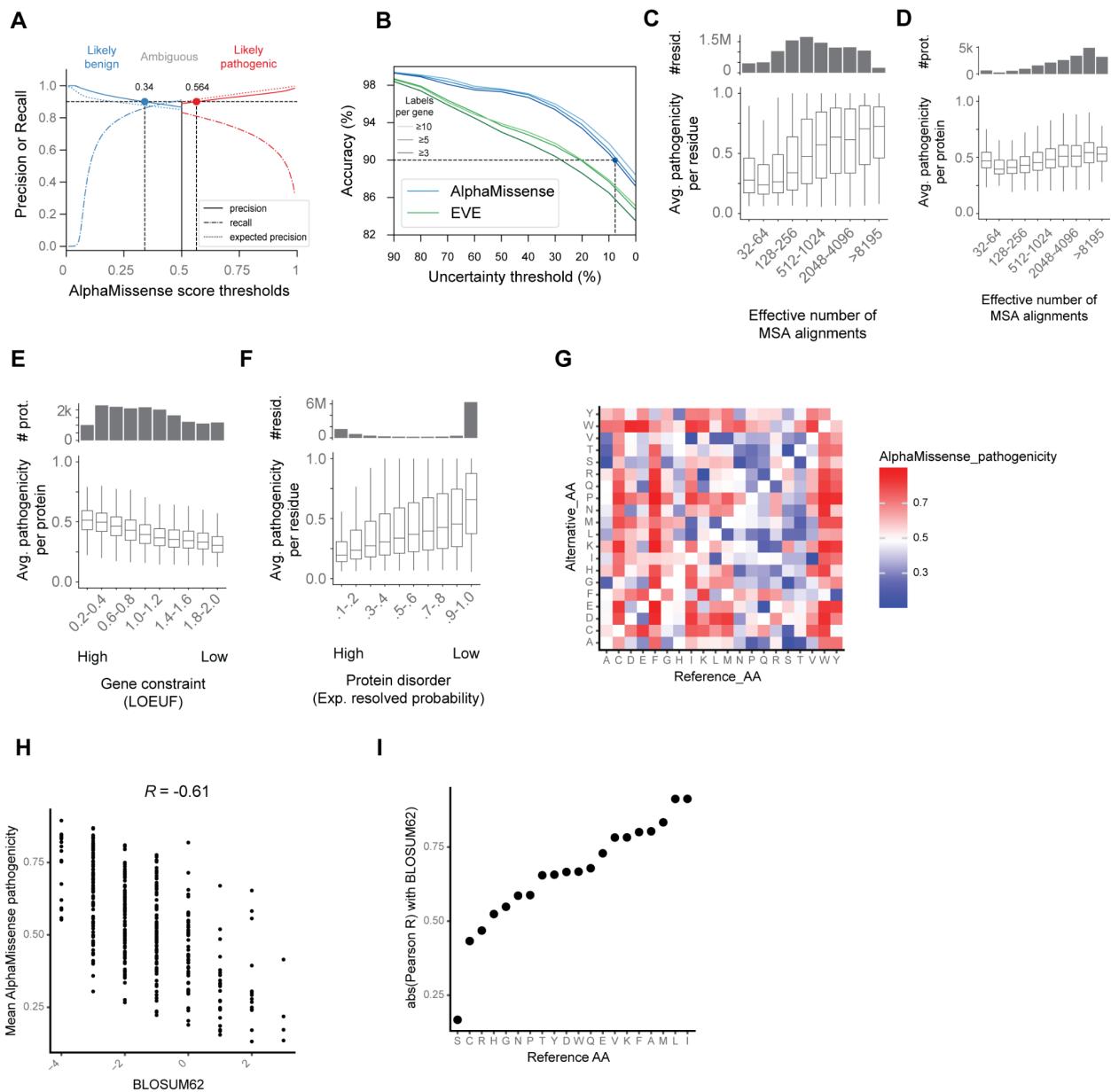


Fig. S4.

Additional calibration results and global prediction properties.

- Precision and recall of AlphaMissense predictions on ClinVar test set. The blue curves are computed on variants labeled benign (negative) and shown for AlphaMissense score in [0, 0.5]. The red curves are computed on variants labeled as pathogenic (positive) and shown for AlphaMissense score in [0.5, 1]. Classification thresholds were chosen at 90% precision for both classes. Ambiguous variants were defined as those not satisfying the 90% precision criterion. Dotted line shows the expected precision computed as the average predicted probability of the variant being pathogenic or benign.
- Compared with EVE, the trade-off between prediction accuracy in ClinVar versus uncertainty threshold, where uncertainty is measured as the fraction of variants falling in

the ‘ambiguous’ class. Accuracy is defined as the fraction of correct predictions in both pathogenic and benign classes with unambiguous predictions. Accuracy was computed over all labels for proteins with at least three (five or ten) pathogenic and benign variants.

- C. Average AlphaMissense pathogenicity score across all human proteins stratified by residues with different number of effective MSA alignments (**Methods**).
- D. Average AlphaMissense pathogenicity score stratified by proteins with different numbers of effective MSA alignments.
- E. Average AlphaMissense pathogenicity score stratified by gene constraint bins as measured by loss-of-function observed/expected upper bound fraction (LOEUF) (3).
- F. Average AlphaMissense pathogenicity score stratified by residues binned by protein disorder as predicted by AlphaFold.
- G. Heatmap of AlphaMissense predicted substitution matrix. Each color square represents the mean predicted pathogenicity across all 216 million single amino acid changes proteome-wide, stratified by reference amino acid (*Reference_AA*) and alternative amino acid type (*Alternative_AA*).
- H. Scatter plot of BLOSUM62 (x-axis) versus mean AlphaMissense pathogenicity scores across all missense variants (y-axis). Each dot represents one substitution type.
- I. Pearson correlation (absolute) between mean AlphaMissense pathogenicity per amino acid substitution type (y-axis) and BLOSUM62 evaluated per reference amino acid (x-axis).

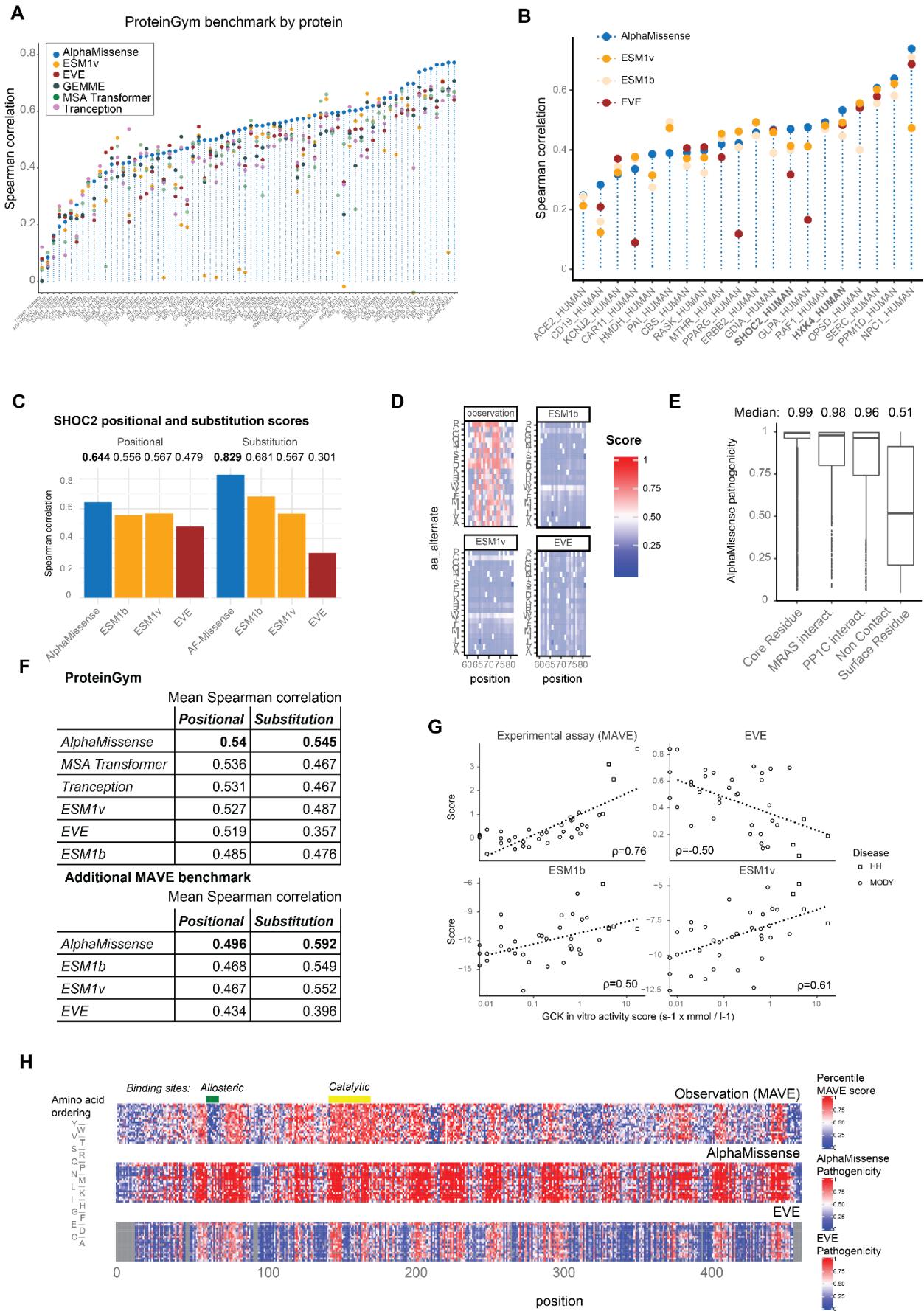


Fig. S5.

Supporting information for multiplexed assays of variant effect benchmarks.

- A. Per-protein performance on ProteinGym benchmark. Each dot shows model performance as measured by Spearman correlation (y-axis) between predictions and MAVE data, for each combination of protein in ProteinGym (x-axis) and model (color/shape). Proteins are ordered by AlphaMissense performance.
- B. Per-protein performance on the custom MAVE benchmark dataset. Each dot shows model performance as measured by Spearman correlation (y-axis) between predictions and MAVE data, for each combination of protein (x-axis) and model (color/shape). Proteins are ordered by AlphaMissense performance. Bolded proteins (SHOC2 and GCK) were highlighted for further investigation in **Fig. 3**.
- C. Correlations between MAVE data and model predictions for SHOC2, where Spearman correlation is calculated in two ways. In the ‘Positional’ benchmark, the average prediction at each amino acid position (resulting in N data points, where N is the length of the protein) is correlated against the average MAVE score at each amino acid position. This captures the broad agreement over the length of the protein. In the ‘Substitution’ benchmark, the average predicted effect of mutating to a particular amino acid is calculated (resulting in 20 data points, one for each possible amino acid). This is correlated against the average observed effect of mutating to a particular amino acid in the MAVE data. This captures the broad agreement over amino-acid specific properties that may drive the variant effect results.
- D. Heatmap of amino-acid substitutions in positions 60-83 of SHOC2 as scored by (i) MAVE data of cell growth in cancer cells dependent on SHOC2 (43), (ii) ESM1b, (iii) ESM1v or (iv) EVE. Each score was independently percentile normalized (ranging from 0 to 1) using scores across the whole protein. Variants with scores closer to zero (blue) retain SHOC2 function, whereas scores closer to one (red) lose SHOC2 function.
- E. Distribution of AlphaMissense pathogenicity scores for four different residue groups in SHOC2 defined by (43): Core residues, MRAS and PP1C interacting residues, and Non contact surface residues. The median score is shown above.
- F. Same as C, but reporting the mean Spearman correlation over all proteins in either the ProteinGym or the additional MAVE benchmark.
- G. Comparison of relative activity index for glucokinase (GCK; UniProt: P35557) mutants (56). On the log x-axis, a score of one indicates in vitro activity equivalent to wild type, a score of zero indicates no activity, and a score above one indicates hyper-activity. On the y axis, scores are shown for MAVE (high-through experimental determination of GCK activity), ESM1b, ESM1v and EVE predictions. Each score has been individually percentile normalized. Each dot represents a unique protein variant, and Spearman correlation for each score is shown.
- H. Heatmaps of observed and predicted effects of amino acid substitutions for GCK. Top heatmap: Observed variant effects (MAVE) of cell growth in yeast cells dependent on human GCK (HXK4_HUMAN) (45). Observations are percentile normalized. Variants with scores closer to one (red) show decreased GCK activity, while scores closer to zero (blue) reflect increased GCK activity. Lower heatmaps: EVE and AlphaMissense

pathogenicity scores. Both scores range from zero to one, with higher scores corresponding to increased pathogenicity. Variants with no prediction are colored in gray (see EVE heatmap). Top: Annotations of important binding regions.

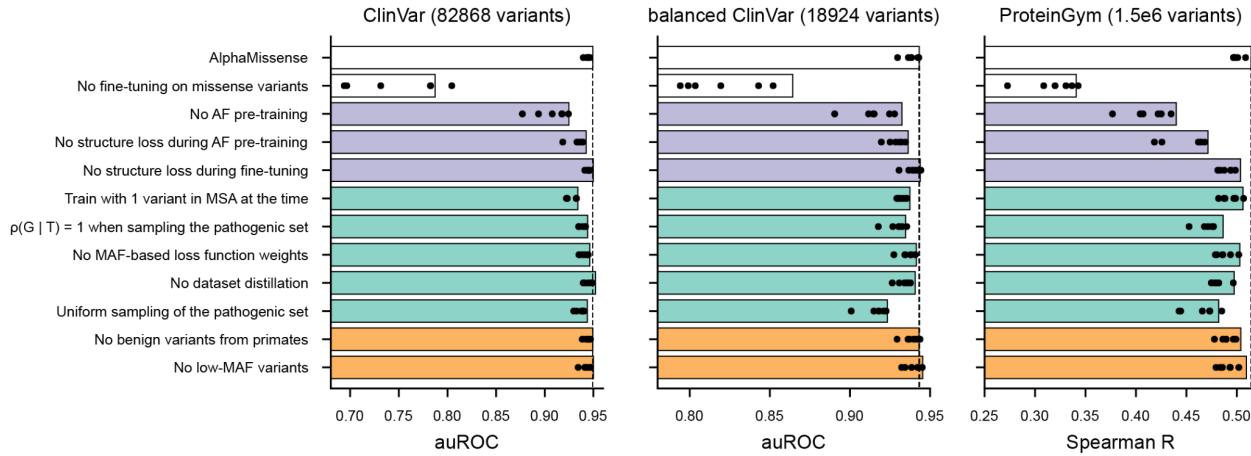


Fig. S6.

AlphaMissense ablation study.

Accuracy of the ablation models compared to the baseline AlphaMissense. Purple bars refer to ablations where structural context is either not provided during pre-training or not enforced during fine-tuning. Teal bars represent ablations that alter the variant sampling process. Orange bars represent ablations that exclude sources of data. Left: Performance on ClinVar test variants measured by auROC. The performance of ensembles of six models (**Methods**) are shown as bars and each single model is shown as dots. Middle: The same as Left, but on the class-balanced ClinVar set. Right: Spearman correlations between predictions and ProteinGym MAVE data for each ensemble (bars) and the corresponding single models (dots).

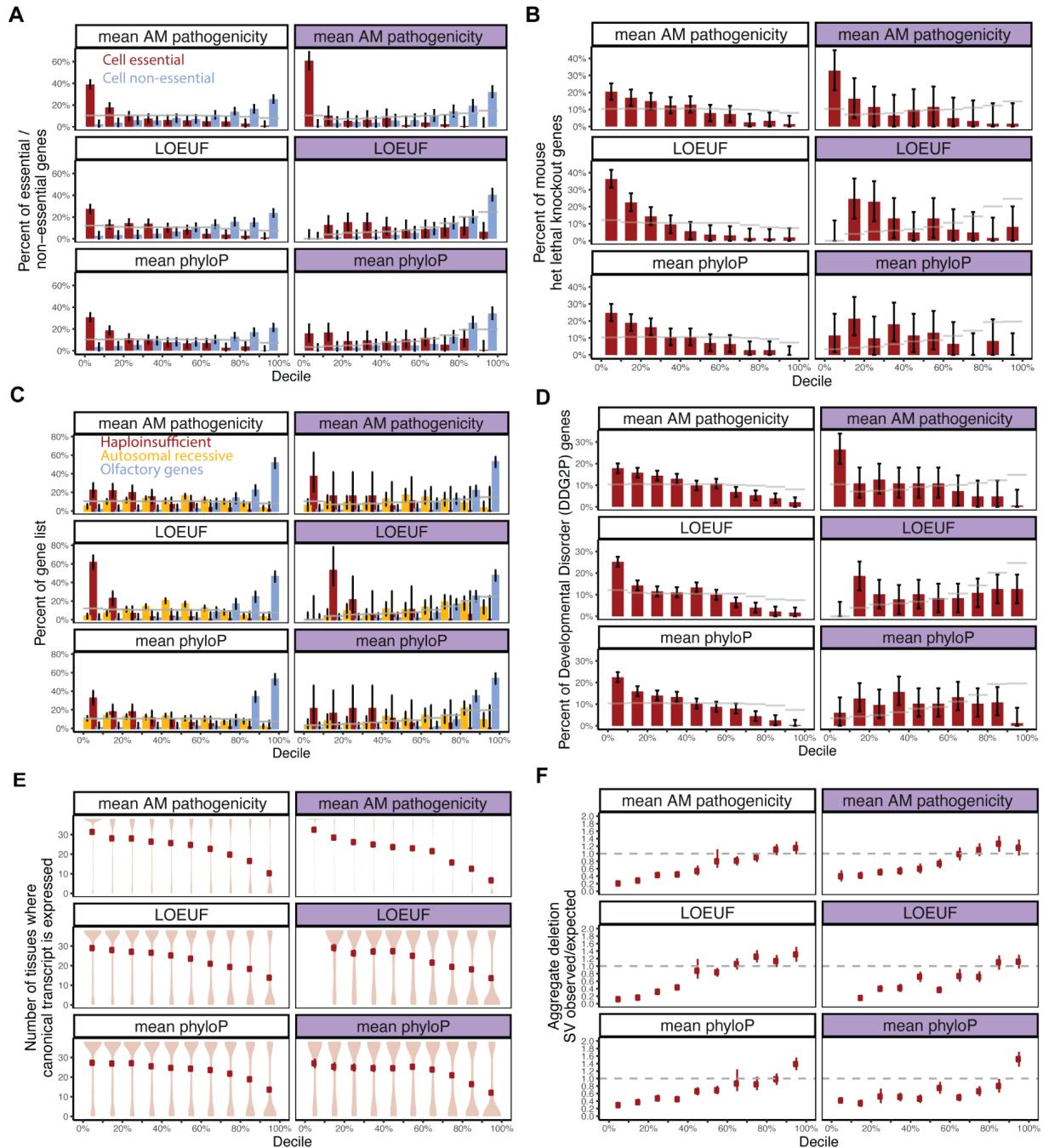


Fig S7.

Relationship between gene-level scores and biological correlates of gene constraint.

A. Distribution of cell essential and non-essential genes (46) (**Table S6**) among deciles defined by three different gene-level scores: average AlphaMissense pathogenicity, LOEUF (3), average PhyloP (47). Plots under white headers (left) show results for all genes available for each analysis; plots under purple headers (right) show results for genes underpowered for LOEUF ('exp_lof' <= 10) (**Methods**). For scores based on AlphaMissense and PhyloP, deciles were defined using negative scores to align with the direction of LOEUF, where low values correspond to more constrained genes. Error bars

show the 95% confidence interval of multinomial proportions. Gray lines indicate the fraction of all genes in each decile bin. This can vary as deciles were computed based on all genes. “AM” = AlphaMissense.

- B. As in A, but for heterozygous lethal knockout genes (3) (**Table S6**).
- C. As in A, but for haploinsufficient, autosomal recessive, and olfactory genes (1, 3) (**Table S6**).
- D. As in A, but for genes underlying developmental disorders (DDG2P) (78) (**Table S6**).
- E. Variation in the number of tissues where the canonical transcript is expressed. Points show the average within each decile, and error bars show 95% confidence intervals estimated using $1.96 \times$ standard error of the mean.
- F. Variation in the depletion of structural deletions observed in a gene, measured as the ratio of observed/expected in the gnomAD cohort, as estimated by (3). Points show the average depletion within each decile, and error bars show 95% confidence intervals estimated using 1,000 bootstrap resamples, as implemented by (3).

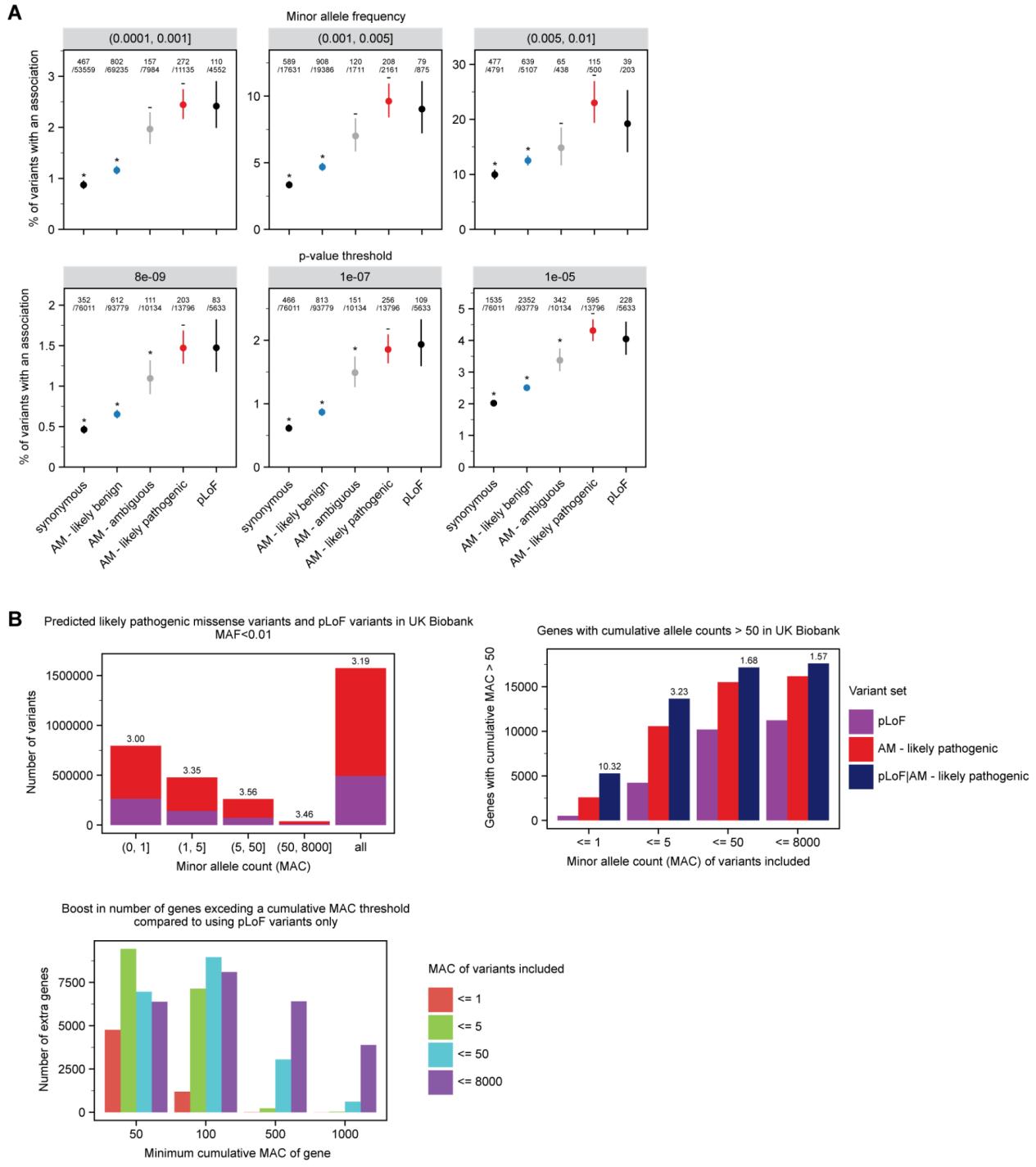


Fig. S8.

Supporting information for properties of proteome-wide predictions.

- A. Similarity between AlphaMissense ‘likely pathogenic’ variants and pLoF variants is robust to choice of allele frequency and *p*-value. Both plots are exactly as Fig. 5C, but for different subsets of variants. Top: Variants within different MAF ranges. Bottom: Variants

subsetted by the *p*-value threshold used to define a variant/phenotype association as statistically significant.

- B. Boost in the number of likely deleterious variants available for complex traits association analyses such as burden testing. All counts are based on variants with MAF<0.01 observed in 450K UK Biobank exomes, as published by (52) (**Methods**). ‘AM likely pathogenic’ refers to variants we classify as ‘likely pathogenic’ using AlphaMissense (**Methods**), and ‘pLoF|AM - likely pathogenic’ refers to variants in either category. Top left: Counts of variants in different minor allele count (MAC) ranges. Top right: Counts of genes with a cumulative MAC of at least 50, for different subsets of variants. Cumulative MAC is the sum of the variant-level MAC across the included variants found in a gene. This approximates the count of individual carriers of putative deleterious variants under the assumption that no one individual carries more than one rare variant in a given gene. Numbers above bars show the fold increase in the number of genes when including both pLoF and ‘likely pathogenic’ missense variants (blue) compared to pLoF variants only (purple). The largest increase is driven by very rare variants with MAC \leq 5. Bottom: Number of additional genes when including variants in ‘pLoF|AMe - likely pathogenic’ compared to only pLoF, for different thresholds of cumulative MAC.

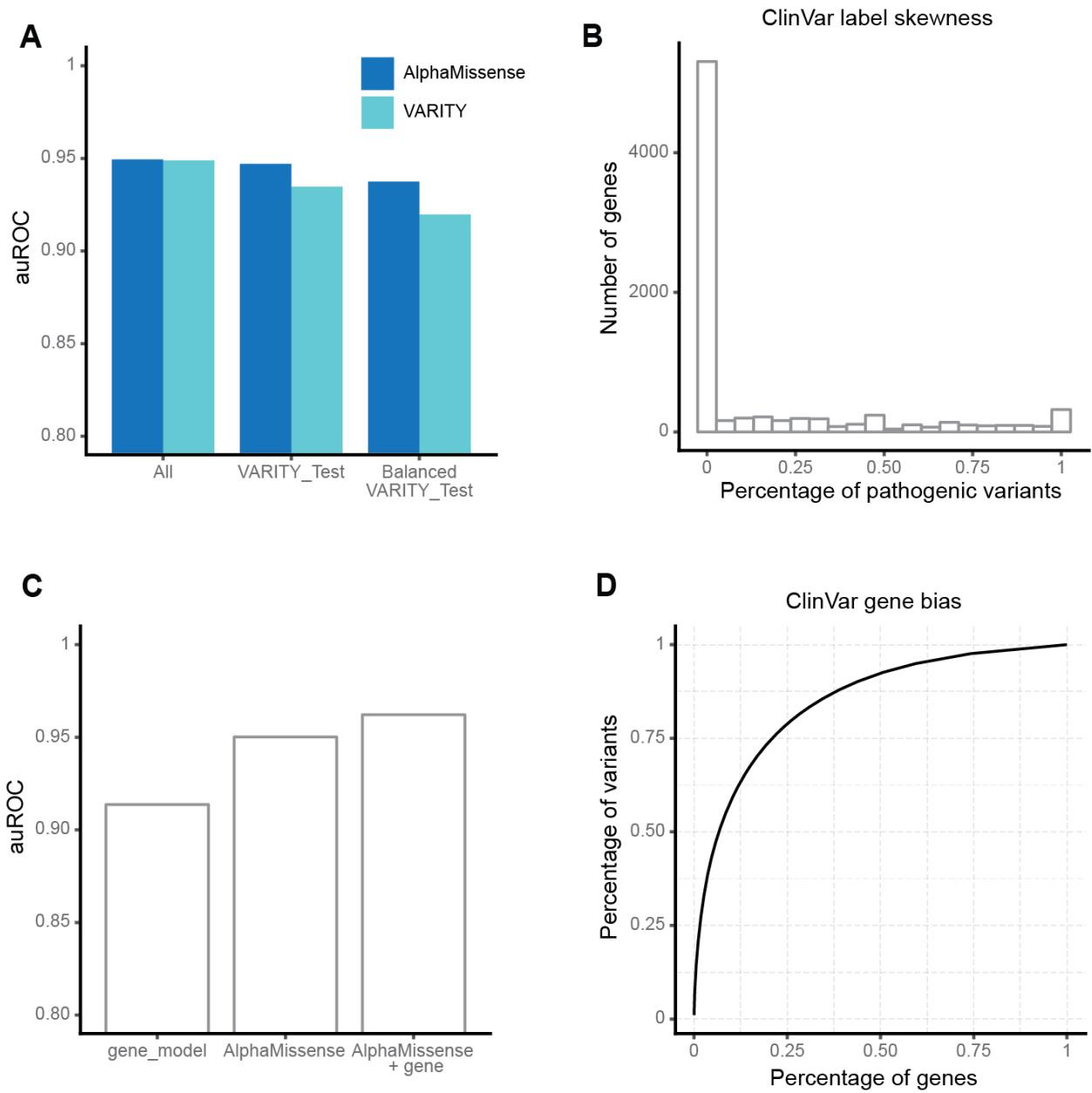


Fig. S9.
Label bias in ClinVar.

- Example of data leakage in benchmarking. Barplot of auROC on all ClinVar test variants (*ALL*), subset of amino acid positions not seen in VARIETY training (*VARIETY_Test*), and subset of variants balancing the number of positive and negative variants per gene in VARIETY_Test (*Balanced VARIETY_Test*).
- Histogram of the percentage of pathogenic variants per gene in ClinVar.
- Gene bias in ClinVar. Barplot of auROC with only the fraction of pathogenic variants per gene as a predictor (*gene_model*), AlphaMissense and AlphaMissense combined with the fraction of pathogenic variants per gene (*AlphaMissense + gene*).
- Percentage of observed genes in ClinVar versus percentage of observed ClinVar variants.

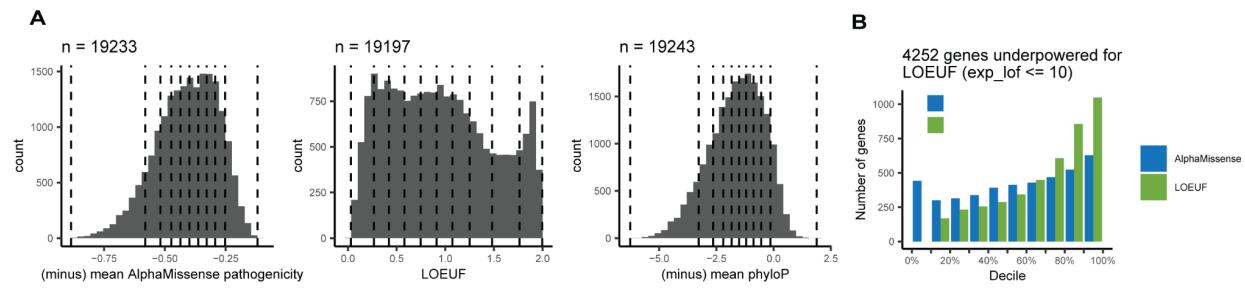


Fig. S10.

Distributions of three gene-level scores in analysis of cell essentiality and gene constraint.

- Distribution of three different gene-level scores and their decile boundaries (dashed lines). LOEUF is a measure of gene constraint under loss of function mutations in humans, and PhyloP (47) is a measure of evolutionary conservation across multiple species. For both AlphaMissense pathogenicity and PhyloP, we compute the average score per gene and plot the negative of these values to align with LOEUF, where low values indicate high gene constraint. Counts are the total number of genes with non-missing values (**Methods**).
- Counts of underpowered genes (**Methods**, **Fig. 4A**) in the deciles defined by gene-level average AlphaMissense pathogenicity or LOEUF.

Species	Number of variants treated as benign in training	Benign cutoff	Data source
Human	1,248,533	MAF > 1e-5	gnomAD (3)
Gorilla	32,089	MAF > 0.03, AN > 50	(63)
Chimpanzee	27,537	MAF > 0.04, AN > 40	(63)
Orangutan, Sumatran subspecies	21,918	MAF > 0.1, AN > 9	(63)
Orangutan, Bornean subspecies	13,130	MAF > 0.1, AN > 9	(63)
Bonobo	2,398	MAF > 0.2	(65) using data from (63) and (67)

Table S1.
Number of benign training variants from different primate species.

The minor allele frequency as (MAF) is defined as $\min(\text{AF}, 1-\text{AF})$ from the global allele frequency (AF). Benign variants were defined as those satisfying the allele frequency and the number of alleles (AN) requirements in at least one organism.

MAF range	Number of variants	Weight in loss
MAF > 2e-4	120,088	1.0
7.37e-5 < MAF < 2e-4	129,631	0.8
2.71e-5 < MAF < 7.37e-5	330,443	0.4
1e-5 < MAF < 2.71e-5	668,371	0.2

Table S2.

Human benign variant weights based on MAF in training.

Feature & Shape	Description
aatype [L, 21]	One-hot representation of the reference amino acid sequence (20 amino acids + unknown).
msa [N _{msa} , L, 23]	One-hot representation of the MSA sequences with 15% of the residues masked or altered by uniform sampling (20 amino acids + unknown + gap + masked msa token) except for the second row, which is the alternative sequence with variants. All positions with variants are masked.
msa_profile [L, 23]	The distribution across amino acid types for each residue in the (unmasked) MSA. The last column represents, for each position along the sequence, the number of deleted residues averaged over all aligned sequences in the MSA (20 amino acids + unknown + gap + deletion mean).
residue_index [L]	The index into the original amino acid sequence.

Table S3.
Input features to AlphaMissense.

Model	AF Pre-training (stage 1)			Variant fine-tuning (stage 2)		
	1	2	3	1.1	2.1	3.1
Parameter initialized from	Random	Random	Random	Model 1	Model 2	Model 3
Sequence crop size L	256	256	256	256	256	256
Number of sequences N_{msa}	128	128	128	256	256	256
Extra MSA	No	No	Yes	No	No	Yes
Training samples (1e6)	7.8	7.5	5.85	-	-	-
Initial learning rate	1e-3	1e-3	1e-3	5*1e-4	5*1e-4	5*1e-4
Number of recycle N_{cycle}	4	4	4	4	4	4

Table S4.
AlphaMissense training details.

In the first stage, we train three AlphaFold models after the specified number of training samples in the table. In the second stage, each of the three models were used to fine-tune on variant data until model performance (auROC) converges on the validation set.

Screen name	UniProt name	UniProt accession	Assay Type	Citation	Original data table (and link if available)
SHOC2_HUMAN	SHOC2_HUMAN	Q9UQ13	Fitness	(43)	https://github.com/jkwonbio/Structure-function-analysis-of-the-SHOC2-MRAS-PP1C-holophophatase-complex/blob/main/data/2022.3.16.Extended%20Data%20Table%204.csv
CAR11_HUMAN	CAR11_HUMAN	Q9BXL7	Fitness	(82)	https://www.cell.com/ajhg/fulltext/S0002-9297(20)30373-6#supplementaryMaterial
CAR11_HUMAN_lof	CAR11_HUMAN	Q9BXL7	Fitness	(82)	https://www.cell.com/ajhg/fulltext/S0002-9297(20)30373-6#supplementaryMaterial
KCNJ2_HUMAN_function	KCNJ2_HUMAN	P63252	Capacity	(83)	https://raw.githubusercontent.com/schmidt-lab/KirDMS/main/Manuscript_SourceData/input_data/function_mod.csv
KCNJ2_HUMAN_surface	KCNJ2_HUMAN	P63252	Abundance	(83)	https://raw.githubusercontent.com/schmidt-lab/KirDMS/main/Manuscript_SourceData/input_data/surface_mod.csv
NPC1_HUMAN	NPC1_HUMAN	O15118	Capacity	(84)	https://static-content.springer.com/esm/art%3A10.1038%2Fs41587-021-01201-1/MediaObjects/41587_2021_1201_MOESM3_ESM.xlsx ; Supplementary Table 1. Saturation

					prime editing function scores for assays performed in NPC1 haploidized HEK293T cells
CBS_HUMAN	CBS_HUMAN	P35520	Fitness	(85)	MaveDB (urn:mavedb:00000005-a-5/)
MTHR_HUMAN	MTHR_HUMAN	P42898	Fitness	(86)	MaveDB (urn:mavedb:00000049-a-4)
RAF1_HUMAN	RAF1_HUMAN	P04049	Binding	(87)	MaveDB (urn:mavedb:00000061-a-1)
OPSD_HUMAN	OPSD_HUMAN	P08100	Abundance	(88)	MaveDB (urn:mavedb:00000099-a-1)
HXK4_HUMAN	HXK4_HUMAN	P35557	Fitness	(45)	MaveDB(urn:mavedb:00000096-a-1/)
SERC_HUMAN	SERC_HUMAN	Q9Y617	Fitness	(89)	Table S7 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7444316/bin/NIHMS1612744-supplement-Supp_TableS7.txt)
GDIA_HUMAN	GDIA_HUMAN	P31150	Fitness	(90)	Table S1 (https://www.biorxiv.org/content/biorxiv/early/2022/10/11/2021.10.06.463360/DC1/embed/media-1.xlsx?download=true)
PPARG_HUMAN	PPARG_HUMAN	P37231	Capacity	(91)	https://miter.broadinstitute.org/mitergraduate/
CD19_HUMAN	CD19_HUMAN	P15391	Binding	(92)	https://figshare.com/articles/dataset/SI_DeepSeqData_for_Retargeting_CD19_CAR_T_cells_via_engineered_CD19-fu

					sion_proteins/82917 59
GLPA_HUMAN	GLPA_HUMAN	P02724	Abundance	(93)	MaveDB (urn:mavedb:00000 051-c/)
HMDH_HUMAN	HMDH_HUMAN	P04035	Fitness	(94)	MaveDB (urn:mavedb:00000 035-a)
ERBB2_HUMAN	ERBB2_HUMAN	P04626	Abundance	(93)	MaveDB (urn:mavedb:00000 051-b-1/)
ACE2_HUMAN	ACE2_HUMAN	Q9BYF1	Binding	(95)	MaveDB (urn:mavedb:00000 069-a-1/)
PPM1D_HUMA N	PPM1D_HUMAN	O15297	Capacity	(96)	https://www.nature.com/articles/s41467-022-30463-9#Sec30
PAI_HUMAN	PAI_HUMAN	P05121	Abundance	(97)	https://www.biorxiv.org/content/biorxiv/early/2022/09/25/2022.07.19.500671/D_C1/embed/media-1.xlsx?download=true; first table
RASK_HUMAN_ binding	RASK_HUMAN	P01116	Binding	(76)	https://www.biorxiv.org/content/biorxiv/early/2022/12/08/2022.12.06.519122/D_C5/embed/media-5.xlsx?download=true
RASK_HUMAN_ folding	RASK_HUMAN	P01116	Abundance	(76)	https://www.biorxiv.org/content/biorxiv/early/2022/12/08/2022.12.06.519122/D_C5/embed/media-5.xlsx?download=true

Table S5.

MAVE screens used for the additional benchmark.

We manually classified each screen into the following assay types: Abundance (protein abundance), Binding (protein-protein interaction), Capacity (protein enzymatic or other

function), Fitness (organismal fitness). This classification is inspired by (98).

Gene feature	Figure	Source		Number of genes analyzed	
		Reference	Data file	All	Underpowered (exp_lof <= 10)
Cell essential genes	S7A	Genes identified by (46) as essential to cell viability across multiple cultured cell lines.	https://github.com/broadinstitute/gnomad_lof/blob/master/R/ko_gene_lists/list_CEGv2.tsv	636	125
Cell non-essential genes	S7A	Genes identified by (46) as non-essential to cell viability across multiple cultured cell lines.	https://github.com/broadinstitute/gnomad_lof/blob/master/R/ko_gene_lists/list_NEGv1.tsv	695	346
Mouse het lethal knockout genes	S7B	Genes identified as lethal to heterozygous knock-out mice. Data originally from MGI (http://www.informatics.jax.org) but curated by (3) - see their Supplementary Materials for details.	https://github.com/broadinstitute/gnomad_lof/blob/master/R/ko_gene_lists/list_mouse_het_lethal_genes.tsv	356	61
Haploinsufficient genes	S7C	Genes known to be haploinsufficient in humans as curated and described by (1, 3).	https://github.com/macarthur-lab/gene_lists/blob/master/lists/haploinsufficiency_*_curated_2016.tsv	155	19
Autosomal recessive genes	S7C	Genes known to underlie disease in humans in an autosomal recessive manner, as curated and described by (1, 3), originally from (99, 100)	https://github.com/macarthur-lab/gene_lists/blob/master/lists/all_ar.tsv	1058	144

Olfactory genes	S7C	Identified by gene names beginning with “OR” (3)	See https://github.com/broadinstitute/gnomad_lof/blob/e09d9fdc13d43c8ac996c93735ed206ba41ed473/R/fig3_spectrum.R#L8	308	300
Developmental disorder genes (DDG2P)	S7D	Gene2Phenotype developmental disorder gene panel with definitive causal evidence (78).	https://ftp.ebi.ac.uk/pub/databases/gen2phenotype/28_04_2023/2023-04-28DDG2P.csv.gz	1608	166
Number of tissues canonical transcript is expressed	S7E	Data originally from Genotype-Tissue Expression Project (GTEx v7), processed and summarized by (3)	https://storage.googleapis.com/gcp-public-data--gnomad/papers/2019-flagship-lof/v1.0/misc/files/GTEx.v7.median_expression_per_tx_per_tissue.021018.tsv.bgz	16836	4251
Aggregate deletion structural variants (observed/expected)	S7F	The ratio of observed vs. expected number of rare coding deletion structural variants (SV) per gene as described in (3).	https://storage.googleapis.com/gcp-public-data--gnomad/papers/2019-flagship-lof/v1.1/misc/files/gnomAD-SV_v2_rev1_release_forFlagship.rare_biallelic_LoF_deletions_per_gene.obs_exp.txt	15432	3341

Table S6.

Details of biological features of gene constraint as shown in Fig. S7. The analysis of these features is discussed in Supplementary Note: *Gene-level AlphaMissense predictions mimic loss of function gene constraint*, and results are available in **Supplementary Data S2 & S3**.

Supplementary Data S1: Gene-level scores and related data for the SF3b complex

Supplementary Data S2: Gene set enrichment statistics among the most constrained decile as defined by AlphaMissense and other scores.

Supplementary Data S3: Quantitative properties of deciles defined by AlphaMissense and other scores.

Supplementary Data S4: Gene-level AlphaMissense scores for all genes used in comparison with LOEUF and PhyloP.

Supplementary Data S5: List of variants and AlphaMissense predictions for the ClinVar benchmark.

Supplementary Data S6: List of variants and AlphaMissense predictions for the Cancer hotspot mutations benchmark.

Supplementary Data S7: List of variants and AlphaMissense predictions for the Deciphering Developmental Disorders benchmark.

Supplementary Data S8: List of variants and AlphaMissense predictions for the ProteinGym benchmark.

Supplementary Data S9: Association statistics and AlphaMissense scores for genetic variants used in the analysis of complex traits.

References and Notes

1. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardiissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016). [doi:10.1038/nature19057](https://doi.org/10.1038/nature19057) [Medline](#)
2. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018). [doi:10.1038/s41586-018-0579-z](https://doi.org/10.1038/s41586-018-0579-z) [Medline](#)
3. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, B. M. Neale, M. J. Daly, D. G. MacArthur; Genome Aggregation Database Consortium, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020). [doi:10.1038/s41586-020-2308-7](https://doi.org/10.1038/s41586-020-2308-7) [Medline](#)
4. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014). [doi:10.1038/nmeth.3027](https://doi.org/10.1038/nmeth.3027) [Medline](#)
5. G. M. Findlay, R. M. Daza, B. Martin, M. D. Zhang, A. P. Leith, M. Gasperini, J. D. Janizek, X. Huang, L. M. Starita, J. Shendure, Accurate classification of *BRCA1* variants with saturation genome editing. *Nature* **562**, 217–222 (2018). [doi:10.1038/s41586-018-0461-z](https://doi.org/10.1038/s41586-018-0461-z) [Medline](#)
6. AVE Alliance Founding Members, The Atlas of Variant Effects (AVE) Alliance: understanding genetic variation at nucleotide resolution, version 4a, Zenodo (2021); <https://doi.org/10.5281/zenodo.7508716>.

7. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1–7.20.41 (2013).
[doi:10.1002/0471142905.hg0720s76](https://doi.org/10.1002/0471142905.hg0720s76) [Medline](#)
8. N. M. Ioannidis, J. H. Rothstein, V. Pejaver, S. Middha, S. K. McDonnell, S. Baheti, A. Musolf, Q. Li, E. Holzinger, D. Karyadi, L. A. Cannon-Albright, C. C. Teerlink, J. L. Stanford, W. B. Isaacs, J. Xu, K. A. Cooney, E. M. Lange, J. Schleutker, J. D. Carpten, I. J. Powell, O. Cussenot, G. Cancel-Tassin, G. G. Giles, R. J. MacInnis, C. Maier, C.-L. Hsieh, F. Wiklund, W. J. Catalona, W. D. Foulkes, D. Mandal, R. A. Eeles, Z. Kote-Jarai, C. D. Bustamante, D. J. Schaid, T. Hastie, E. A. Ostrander, J. E. Bailey-Wilson, P. Radivojac, S. N. Thibodeau, A. S. Whittemore, W. Sieh, REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016). [doi:10.1016/j.ajhg.2016.08.016](https://doi.org/10.1016/j.ajhg.2016.08.016) [Medline](#)
9. Y. Wu, H. Liu, R. Li, S. Sun, J. Weile, F. P. Roth, Improved pathogenicity prediction for rare human missense variants. *Am. J. Hum. Genet.* **108**, 1891–1906 (2021).
[doi:10.1016/j.ajhg.2021.08.012](https://doi.org/10.1016/j.ajhg.2021.08.012) [Medline](#)
10. H. Zhang, M. S. Xu, X. Fan, W. K. Chung, Y. Shen, Predicting functional effect of missense variants using graph attention neural networks. *Nat. Mach. Intell.* **4**, 1017–1028 (2022).
[doi:10.1038/s42256-022-00561-w](https://doi.org/10.1038/s42256-022-00561-w) [Medline](#)
11. D. G. Grimm, C. A. Azencott, F. Aicheler, U. Gieraths, D. G. MacArthur, K. E. Samocha, D. N. Cooper, P. D. Stenson, M. J. Daly, J. W. Smoller, L. E. Duncan, K. M. Borgwardt, The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Hum. Mutat.* **36**, 513–523 (2015). [doi:10.1002/humu.22768](https://doi.org/10.1002/humu.22768) [Medline](#)
12. L. Sundaram, H. Gao, S. R. Padigepati, J. F. McRae, Y. Li, J. A. Kosmicki, N. Fritzilas, J. Hakenberg, A. Dutta, J. Shon, J. Xu, S. Batzoglou, X. Li, K. K.-H. Farh, Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* **50**, 1161–1170 (2018). [doi:10.1038/s41588-018-0167-z](https://doi.org/10.1038/s41588-018-0167-z) [Medline](#)
13. M. Kircher, D. M. Witten, P. Jain, B. J. O’Roak, G. M. Cooper, J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014). [doi:10.1038/ng.2892](https://doi.org/10.1038/ng.2892) [Medline](#)
14. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003). [doi:10.1093/nar/gkg509](https://doi.org/10.1093/nar/gkg509) [Medline](#)
15. T. A. Hopf, J. B. Ingraham, F. J. Poelwijk, C. P. I. Schärfe, M. Springer, C. Sander, D. S. Marks, Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017). [doi:10.1038/nbt.3769](https://doi.org/10.1038/nbt.3769) [Medline](#)
16. E. Laine, Y. Karami, A. Carbone, GEMME: A simple and fast global epistatic model predicting mutational effects. *Mol. Biol. Evol.* **36**, 2604–2619 (2019).
[doi:10.1093/molbev/msz179](https://doi.org/10.1093/molbev/msz179) [Medline](#)
17. J. Frazer, P. Notin, M. Dias, A. Gomez, J. K. Min, K. Brock, Y. Gal, D. S. Marks, Disease variant prediction with deep generative models of evolutionary data. *Nature* **599**, 91–95 (2021). [doi:10.1038/s41586-021-04043-8](https://doi.org/10.1038/s41586-021-04043-8) [Medline](#)

18. J. Meier, R. Rao, R. Verkuil, J. Liu, T. Sercu, A. Rives, Language models enable zero-shot prediction of the effects of mutations on protein function. bioRxiv 2021.07.09.450648 [Preprint] (2021); <https://doi.org/10.1101/2021.07.09.450648>.
19. P. Notin, M. Dias, J. Frazer, J. Marchena Hurtado, A. N. Gomez, D. Marks, Y. Gal, Tranception: Protein fitness prediction with autoregressive transformers and inference-time retrieval. *Proc. Mach. Learn. Res.* **162**, 16990–17017 (2022).
20. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023). [doi:10.1126/science.adc2574](https://doi.org/10.1126/science.adc2574) Medline
21. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). [doi:10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2) Medline
22. S. Ittisoponpisan, S. A. Islam, T. Khanna, E. Alhuzimi, A. David, M. J. E. Sternberg, Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J. Mol. Biol.* **431**, 2197–2212 (2019). [doi:10.1016/j.jmb.2019.04.009](https://doi.org/10.1016/j.jmb.2019.04.009) Medline
23. A. Schmidt, S. Röner, K. Mai, H. Klinkhammer, M. Kircher, K. U. Ludwig, Predicting the pathogenicity of missense variants using features derived from AlphaFold2. *Bioinformatics* **39**, btad280 (2023). [doi:10.1093/bioinformatics/btad280](https://doi.org/10.1093/bioinformatics/btad280) Medline
24. B. Li, D. M. Roden, J. A. Capra, The 3D mutational constraint on amino acid sites in the human proteome. *Nat. Commun.* **13**, 3273 (2022). [doi:10.1038/s41467-022-30936-x](https://doi.org/10.1038/s41467-022-30936-x) Medline
25. K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywelt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021). [doi:10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1) Medline
26. G. R. Buel, K. J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* **29**, 1–2 (2022). [doi:10.1038/s41594-021-00714-2](https://doi.org/10.1038/s41594-021-00714-2) Medline
27. R. M. Rao, J. Liu, R. Verkuil, J. Meier, J. Canny, P. Abbeel, T. Sercu, A. Rives, MSA Transformer. *Proc. Mach. Learn. Res.* **139**, 8844–8856 (2021).
28. A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, R. Fergus, Biological structure and function emerge from scaling unsupervised learning to

- 250 million protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016239118 (2021). [doi:10.1073/pnas.2016239118](https://doi.org/10.1073/pnas.2016239118) [Medline](#)
29. B. J. Livesey, J. A. Marsh, Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Mol. Syst. Biol.* **16**, e9380 (2020). [doi:10.15252/msb.20199380](https://doi.org/10.15252/msb.20199380) [Medline](#)
30. M. J. Landrum, J. M. Lee, M. Benson, G. R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J. B. Holmes, B. L. Kattman, D. R. Maglott, ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018). [doi:10.1093/nar/gkx1153](https://doi.org/10.1093/nar/gkx1153) [Medline](#)
31. J. M. Havrilla, B. S. Pedersen, R. M. Layer, A. R. Quinlan, A map of constrained coding regions in the human genome. *Nat. Genet.* **51**, 88–95 (2019). [doi:10.1038/s41588-018-0294-6](https://doi.org/10.1038/s41588-018-0294-6) [Medline](#)
32. D. T. Miller, K. Lee, N. S. Abul-Husn, L. M. Amendola, K. Brothers, W. K. Chung, M. H. Gollob, A. S. Gordon, S. M. Harrison, R. E. Hershberger, T. E. Klein, C. S. Richards, D. R. Stewart, C. L. Martin; ACMG Secondary Findings Working Group, ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **24**, 1407–1414 (2022). [doi:10.1016/j.gim.2022.04.006](https://doi.org/10.1016/j.gim.2022.04.006) [Medline](#)
33. D. Kuang, R. Truty, J. Weile, B. Johnson, K. Nykamp, C. Araya, R. L. Nussbaum, F. P. Roth, Prioritizing genes for systematic variant effect mapping. *Bioinformatics* **36**, 5448–5455 (2021). [doi:10.1093/bioinformatics/btaa1008](https://doi.org/10.1093/bioinformatics/btaa1008) [Medline](#)
34. S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm; ACMG Laboratory Quality Assurance Committee, Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015). [doi:10.1038/gim.2015.30](https://doi.org/10.1038/gim.2015.30) [Medline](#)
35. M. H. Høie, M. Cagiada, A. H. Beck Frederiksen, A. Stein, K. Lindorff-Larsen, Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. *Cell Rep.* **38**, 110207 (2022). [doi:10.1016/j.celrep.2021.110207](https://doi.org/10.1016/j.celrep.2021.110207) [Medline](#)
36. K. A. Matreyek, L. M. Starita, J. J. Stephany, B. Martin, M. A. Chiasson, V. E. Gray, M. Kircher, A. Khechaduri, J. N. Dines, R. J. Hause, S. Bhatia, W. E. Evans, M. V. Relling, W. Yang, J. Shendure, D. M. Fowler, Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018). [doi:10.1038/s41588-018-0122-z](https://doi.org/10.1038/s41588-018-0122-z) [Medline](#)
37. A. Laddach, J. C. F. Ng, F. Fraternali, Pathogenic missense protein variants affect different functional pathways and proteomic features than healthy population variants. *PLOS Biol.* **19**, e3001207 (2021). [doi:10.1371/journal.pbio.3001207](https://doi.org/10.1371/journal.pbio.3001207) [Medline](#)
38. D. Munro, M. Singh, DeMaSk: A deep mutational scanning substitution matrix and its use for variant impact prediction. *Bioinformatics* **36**, 5322–5329 (2021). [doi:10.1093/bioinformatics/btaa1030](https://doi.org/10.1093/bioinformatics/btaa1030) [Medline](#)

39. S. Henikoff, J. G. Henikoff, Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915–10919 (1992). [doi:10.1073/pnas.89.22.10915](https://doi.org/10.1073/pnas.89.22.10915) Medline
40. S. Fayer, C. Horton, J. N. Dines, A. F. Rubin, M. E. Richardson, K. McGoldrick, F. Hernandez, T. Pesaran, R. Karam, B. H. Shirts, D. M. Fowler, L. M. Starita, Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in *BRCA1*, *TP53*, and *PTEN*. *Am. J. Hum. Genet.* **108**, 2248–2258 (2021). [doi:10.1016/j.ajhg.2021.11.001](https://doi.org/10.1016/j.ajhg.2021.11.001) Medline
41. B. J. Livesey, J. A. Marsh, Updated benchmarking of variant effect predictors using deep mutational scanning. *Mol. Syst. Biol.* **19**, e11474 (2023). [doi:10.15252/msb.202211474](https://doi.org/10.15252/msb.202211474) Medline
42. J. J. Kwon, W. C. Hahn, A leucine-rich repeat protein provides a SHOC2 the RAS circuit: a structure-function perspective. *Mol. Cell. Biol.* **41**, e00627-20 (2021). [doi:10.1128/MCB.00627-20](https://doi.org/10.1128/MCB.00627-20) Medline
43. J. J. Kwon, B. Hajian, Y. Bian, L. C. Young, A. J. Amor, J. R. Fuller, C. V. Fraley, A. M. Sykes, J. So, J. Pan, L. Baker, S. J. Lee, D. B. Wheeler, D. L. Mayhew, N. S. Persky, X. Yang, D. E. Root, A. M. Barsotti, A. W. Stamford, C. K. Perry, A. Burgin, F. McCormick, C. T. Lemke, W. C. Hahn, A. J. Aguirre, Structure-function analysis of the SHOC2-MRAS-PP1C holophosphatase complex. *Nature* **609**, 408–415 (2022). [doi:10.1038/s41586-022-04928-2](https://doi.org/10.1038/s41586-022-04928-2) Medline
44. S. M. Sternisha, B. G. Miller, Molecular and cellular regulation of human glucokinase. *Arch. Biochem. Biophys.* **663**, 199–213 (2019). [doi:10.1016/j.abb.2019.01.011](https://doi.org/10.1016/j.abb.2019.01.011) Medline
45. S. Gersing, M. Cagiada, M. Gebbia, A. P. Gjesing, A. G. Coté, G. Seesankar, R. Li, D. Tabet, J. Weile, A. Stein, A. L. Gloyn, T. Hansen, F. P. Roth, K. Lindorff-Larsen, R. Hartmann-Petersen, A comprehensive map of human glucokinase variant activity. *Genome Biol.* **24**, 97 (2023). [doi:10.1186/s13059-023-02935-8](https://doi.org/10.1186/s13059-023-02935-8) Medline
46. T. Hart, A. H. Y. Tong, K. Chan, J. Van Leeuwen, A. Seetharaman, M. Aregger, M. Chandrashekhar, N. Hustedt, S. Seth, A. Noonan, A. Habsid, O. Sizova, L. Nedyalkova, R. Climie, L. Tworzyanski, K. Lawson, M. A. Sartori, S. Alibeh, D. Tieu, S. Masud, P. Mero, A. Weiss, K. R. Brown, M. Usaj, M. Billmann, M. Rahman, M. Costanzo, C. L. Myers, B. J. Andrews, C. Boone, D. Durocher, J. Moffat, Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3* **7**, 2719–2727 (2017). [doi:10.1534/g3.117.041277](https://doi.org/10.1534/g3.117.041277) Medline
47. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010). [doi:10.1101/gr.097857.109](https://doi.org/10.1101/gr.097857.109) Medline
48. J. M. Dempster, J. Rossen, M. Kazachkova, J. Pan, G. Kugener, D. E. Root, A. Tsherniak, Extracting biological insights from the Project Achilles genome-scale CRISPR screens in cancer cell lines. bioRxiv 720243 [Preprint] (2019); <https://doi.org/10.1101/720243>.
49. C. Sun, The SF3b complex: Splicing and beyond. *Cell. Mol. Life Sci.* **77**, 3583–3595 (2020). [doi:10.1007/s00018-020-03493-z](https://doi.org/10.1007/s00018-020-03493-z) Medline
50. L. Bomba, K. Walter, N. Soranzo, The impact of rare and low-frequency genetic variants in common disease. *Genome Biol.* **18**, 77 (2017). [doi:10.1186/s13059-017-1212-4](https://doi.org/10.1186/s13059-017-1212-4) Medline

51. S. Lee, G. R. Abecasis, M. Boehnke, X. Lin, Rare-variant association analysis: Study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014). [doi:10.1016/j.ajhg.2014.06.009](https://doi.org/10.1016/j.ajhg.2014.06.009) [Medline](#)
52. K. J. Karczewski, M. Solomonson, K. R. Chao, J. K. Goodrich, G. Tiao, W. Lu, B. M. Riley-Gillis, E. A. Tsai, H. I. Kim, X. Zheng, F. Rahimov, S. Esmaeeli, A. J. Grundstad, M. Reppell, J. Waring, H. Jacob, D. Sexton, P. G. Bronson, X. Chen, X. Hu, J. I. Goldstein, D. King, C. Vittal, T. Poterba, D. S. Palmer, C. Churchhouse, D. P. Howrigan, W. Zhou, N. A. Watts, K. Nguyen, H. Nguyen, C. Mason, C. Farnham, C. Tolonen, L. D. Gauthier, N. Gupta, D. G. MacArthur, H. L. Rehm, C. Seed, A. A. Philippakis, M. J. Daly, J. W. Davis, H. Runz, M. R. Miller, B. M. Neale, Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* **2**, 100168 (2022). [doi:10.1016/j.xgen.2022.100168](https://doi.org/10.1016/j.xgen.2022.100168) [Medline](#)
53. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022). [doi:10.1093/nar/gkab1061](https://doi.org/10.1093/nar/gkab1061) [Medline](#)
54. Deciphering Developmental Disorders Study, Prevalence and architecture of *de novo* mutations in developmental disorders. *Nature* **542**, 433–438 (2017). [doi:10.1038/nature21062](https://doi.org/10.1038/nature21062) [Medline](#)
55. P. Petit, M. Antoine, G. Ferry, J. A. Boutin, A. Lagarde, L. Gluais, R. Vincentelli, L. Vuillard, The active conformation of human glucokinase is not altered by allosteric activators. *Acta Crystallogr. D* **67**, 929–935 (2011). [doi:10.1107/S0907444911036729](https://doi.org/10.1107/S0907444911036729) [Medline](#)
56. A. L. Gloyn, S. Odili, C. Buettger, P. R. Njølstad, C. Shiota, M. A. Magnuson, F. M. Matschinsky, “Glucokinase and the regulation of blood sugar: A mathematical model predicts the threshold for glucose stimulated insulin release for GCK gene mutations that cause hyper- and hypoglycemia” in *Glucokinase and Glycemic Disease: From Basics to Novel Therapeutics*, F. M. Matschinsky, M. A. Magnuson, Eds., vol. 16 of *Frontiers in Diabetes* (S. Karger AG, 2004), pp. 92–109.
57. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritchel, L. H. Wong, M. Zielinski, T. Sargeant, R. G. Schneider, A. W. Senior, J. Jumper, D. Hassabis, P. Kohli, Ž. Avsec, Source code for AlphaMissense, version 1.0.0, Zenodo (2023); <https://doi.org/10.5281/zenodo.8208697>.
58. J. Cheng, G. Novati, J. Pan, C. Bycroft, A. Žemgulytė, T. Applebaum, A. Pritchel, L. H. Wong, M. Zielinski, T. Sargeant, R. G. Schneider, A. W. Senior, J. Jumper, D. Hassabis, P. Kohli, Ž. Avsec, Predictions of AlphaMissense, version 1.0.0, Zenodo (2023); <https://doi.org/10.5281/zenodo.8208688>.
59. F. Tessadori, K. Duran, K. Knapp, M. Fellner, S. Smithson, A. Beleza Meireles, M. W. Elting, Q. Waisfisz, A. O'Donnell-Luria, C. Nowak, J. Douglas, A. Ronan, T. Brunet, U. Kotzaeridou, S. Svhovec, M. S. Saenz, I. Thiffault, F. Del Viso, P. Devine, S. Rego, J.

- Tenney, A. van Haeringen, C. A. L. Ruivenkamp, S. Koene, S. P. Robertson, C. Deshpande, R. Pfundt, N. Verbeek, J. M. van de Kamp, J. M. M. Weiss, A. Ruiz, E. Gabau, E. Banne, A. Pepler, A. Bottani, S. Laurent, M. Guipponi, E. Bijlsma, A.-L. Bruel, A. Sorlin, M. Willis, Z. Powis, T. Smol, C. Vincent-Delorme, D. Baralle, E. Colin, N. Revencu, E. Calpena, A. O. M. Wilkie, M. Chopra, V. Cormier-Daire, B. Keren, A. Afenjar, M. Niceta, A. Terracciano, N. Specchio, M. Tartaglia, M. Rio, G. Barcia, S. Rondeau, C. Colson, J. Bakkers, P. D. Mace, L. S. Bicknell, G. van Haaften; Deciphering Developmental Disorders Study, Recurrent *de novo* missense variants across multiple histone H4 genes underlie a neurodevelopmental syndrome. *Am. J. Hum. Genet.* **109**, 750–758 (2022). [doi:10.1016/j.ajhg.2022.02.003](https://doi.org/10.1016/j.ajhg.2022.02.003) Medline
60. M. T. Chang, T. S. Bhattacharai, A. M. Schram, C. M. Bielski, M. T. A. Donoghue, P. Jonsson, D. Chakravarty, S. Phillips, C. Kandoth, A. Penson, A. Gorelick, T. Shamu, S. Patel, C. Harris, J. Gao, S. O. Sumer, R. Kundra, P. Razavi, B. T. Li, D. N. Reales, N. D. Socci, G. Jayakumaran, A. Zehir, R. Benayed, M. E. Arcila, S. Chandarlapaty, M. Ladanyi, N. Schultz, J. Baselga, M. F. Berger, N. Rosen, D. B. Solit, D. M. Hyman, B. S. Taylor, Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* **8**, 174–183 (2018). [doi:10.1158/2159-8290.CD-17-0321](https://doi.org/10.1158/2159-8290.CD-17-0321) Medline
61. F. E. Dewey, M. F. Murray, J. D. Overton, L. Habegger, J. B. Leader, S. N. Fetterolf, C. O'Dushlaine, C. V. Van Hout, J. Staples, C. Gonzaga-Jauregui, R. Metpally, S. A. Pendergrass, M. A. Giovanni, H. L. Kirchner, S. Balasubramanian, N. S. Abul-Husn, D. N. Hartzel, D. R. Lavage, K. A. Kost, J. S. Packer, A. E. Lopez, J. Penn, S. Mukherjee, N. Gosalia, M. Kanagaraj, A. H. Li, L. J. Mitnaul, L. J. Adams, T. N. Person, K. Praveen, A. Marcketta, M. S. Lebo, C. A. Austin-Tse, H. M. Mason-Suarez, S. Bruse, S. Mellis, R. Phillips, N. Stahl, A. Murphy, A. Economides, K. A. Skelding, C. D. Still, J. R. Elmore, I. B. Borecki, G. D. Yancopoulos, F. D. Davis, W. A. Faucett, O. Gottesman, M. D. Ritchie, A. R. Shuldiner, J. G. Reid, D. H. Ledbetter, A. Baras, D. J. Carey, Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, aaf6814 (2016). [doi:10.1126/science.aaf6814](https://doi.org/10.1126/science.aaf6814) Medline
62. I. Iossifov, B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paeper, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y.-H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State, M. Wigler, The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014). [doi:10.1038/nature13908](https://doi.org/10.1038/nature13908) Medline
63. J. Prado-Martinez, P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley, B. Lorente-Galdos, K. R. Veeramah, A. E. Woerner, T. D. O'Connor, G. Santpere, A. Cagan, C. Theunert, F. Casals, H. Laayouni, K. Munch, A. Hobolth, A. E. Halager, M. Malig, J. Hernandez-Rodriguez, I. Hernando-Herraez, K. Prüfer, M. Pybus, L. Johnstone, M. Lachmann, C. Alkan, D. Twigg, N. Petit, C. Baker, F. Hormozdiari, M. Fernandez-Callejo, M. Dabad, M. L. Wilson, L. Stevison, C. Camprubí, T. Carvalho, A. Ruiz-Herrera, L. Vives, M. Mele, T. Abello, I. Kondova, R. E. Bontrop, A. Pusey, F. Lankester, J. A. Kiyang, R. A. Bergl, E. Lonsdorf, S. Myers, M. Ventura, P. Gagneux, D. Comas, H. Siegismund, J.

Blanc, L. Agueda-Calpena, M. Gut, L. Fulton, S. A. Tishkoff, J. C. Mullikin, R. K. Wilson, I. G. Gut, M. K. Gonder, O. A. Ryder, B. H. Hahn, A. Navarro, J. M. Akey, J. Bertranpetti, D. Reich, T. Mailund, M. H. Schierup, C. Hvilsom, A. M. Andrés, J. D. Wall, C. D. Bustamante, M. F. Hammer, E. E. Eichler, T. Marques-Bonet, Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
[doi:10.1038/nature12228](https://doi.org/10.1038/nature12228) [Medline](#)

64. F. Cunningham, J. E. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, J. G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D. N. Ogeh, A. Parker, A. Parton, M. Perry, I. Piližota, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, J. G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S. E. Hunt, G. R. Iisley, J. E. Loveland, F. J. Martin, B. Moore, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, S. Dyer, P. W. Harrison, K. L. Howe, A. D. Yates, D. R. Zerbino, P. Flliceck, Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022). [doi:10.1093/nar/gkab1049](https://doi.org/10.1093/nar/gkab1049) [Medline](#)
65. S. Han, A. M. Andrés, T. Marques-Bonet, M. Kuhlwilm, Genetic variation in *Pan* species is shaped by demographic history and harbors lineage-specific functions. *Genome Biol. Evol.* **11**, 1178–1191 (2019). [doi:10.1093/gbe/evz047](https://doi.org/10.1093/gbe/evz047) [Medline](#)
66. M. Kuhlwilm, Han_et.al_Data.tsv.gz, version 1, figshare (2019);
<https://doi.org/10.6084/m9.figshare.7855850.v1>.
67. M. de Manuel, M. Kuhlwilm, P. Frandsen, V. C. Sousa, T. Desai, J. Prado-Martinez, J. Hernandez-Rodriguez, I. Dupanloup, O. Lao, P. Hallast, J. M. Schmidt, J. M. Heredia-Genestar, A. Benazzo, G. Barbujani, B. M. Peter, L. F. Kuderna, F. Casals, S. Angedakin, M. Arandjelovic, C. Boesch, H. Kühl, L. Vigilant, K. Langergraber, J. Novembre, M. Gut, I. Gut, A. Navarro, F. Carlsen, A. M. Andrés, H. R. Siegismund, A. Scally, L. Excoffier, C. Tyler-Smith, S. Castellano, Y. Xue, C. Hvilsom, T. Marques-Bonet, Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**, 477–481 (2016). [doi:10.1126/science.aag2602](https://doi.org/10.1126/science.aag2602) [Medline](#)
68. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019). [doi:10.1186/s12859-019-3019-7](https://doi.org/10.1186/s12859-019-3019-7) [Medline](#)
69. A. L. Mitchell, A. Almeida, M. Beracochea, M. Boland, J. Burgin, G. Cochrane, M. R. Crusoe, V. Kale, S. C. Potter, L. J. Richardson, E. Sakharova, M. Scheremetjew, A. Korobeynikov, A. Shlemov, O. Kunyavskaya, A. Lapidus, R. D. Finn, Mgnify: The microbiome analysis resource in 2020. *Nucleic Acids Res.* **48**, D570–D578 (2020). [Medline](#)

70. B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu; UniProt Consortium, UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015). [doi:10.1093/bioinformatics/btu739](https://doi.org/10.1093/bioinformatics/btu739) [Medline](#)
71. X. Liu, C. Li, C. Mou, Y. Dong, Y. Tu, dbNSFP v4: A comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **12**, 103 (2020). [doi:10.1186/s13073-020-00803-9](https://doi.org/10.1186/s13073-020-00803-9) [Medline](#)
72. P. C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001). [doi:10.1101/gr.176601](https://doi.org/10.1101/gr.176601) [Medline](#)
73. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019). [doi:10.1093/nar/gky1016](https://doi.org/10.1093/nar/gky1016) [Medline](#)
74. I. Ionita-Laza, K. McCallum, B. Xu, J. D. Buxbaum, A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016). [doi:10.1038/ng.3477](https://doi.org/10.1038/ng.3477) [Medline](#)
75. N. Brandes, G. Goldman, C. H. Wang, C. J. Ye, V. Ntranos, Genome-wide prediction of disease variant effects with a deep protein language model. *Nat. Genet.* **55**, 1512–1522 (2023). [doi:10.1038/s41588-023-01465-0](https://doi.org/10.1038/s41588-023-01465-0) [Medline](#)
76. C. Weng, A. J. Faure, B. Lehner, The energetic and allosteric landscape for KRAS inhibition. bioRxiv 2022.12.06.519122 [Preprint] (2022); <https://doi.org/10.1101/2022.12.06.519122>.
77. T. Wu, J. Hou, B. Adhikari, J. Cheng, Analysis of several key factors influencing deep learning-based inter-residue contact prediction. *Bioinformatics* **36**, 1091–1098 (2020). [doi:10.1093/bioinformatics/btz679](https://doi.org/10.1093/bioinformatics/btz679) [Medline](#)
78. A. Thormann, M. Halachev, W. McLaren, D. J. Moore, V. Svinti, A. Campbell, S. M. Kerr, M. Tischkowitz, S. E. Hunt, M. G. Dunlop, M. E. Hurles, C. F. Wright, H. V. Firth, F. Cunningham, D. R. FitzPatrick, Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat. Commun.* **10**, 2373 (2019). [doi:10.1038/s41467-019-10016-3](https://doi.org/10.1038/s41467-019-10016-3) [Medline](#)
79. Broad DepMap, DepMap 23Q2 Public, version 3, figshare (2023); <https://doi.org/10.6084/m9.figshare.22765112.v3>.
80. V. A. Blomen, P. Májek, L. T. Jae, J. W. Bigenzahn, J. Nieuwenhuis, J. Staring, R. Sacco, F. R. van Diemen, N. Olk, A. Stukalov, C. Marceau, H. Janssen, J. E. Carette, K. L. Bennett, J. Colinge, G. Superti-Furga, T. R. Brummelkamp, Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015). [doi:10.1126/science.aac7557](https://doi.org/10.1126/science.aac7557) [Medline](#)
81. T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K. R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F. P. Roth, O. S. Rissland, D. Durocher, S. Angers, J. Moffat, High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015). [doi:10.1016/j.cell.2015.11.015](https://doi.org/10.1016/j.cell.2015.11.015) [Medline](#)

82. I. Meitlis, E. J. Allenspach, B. M. Bauman, I. Q. Phan, G. Dabbah, E. G. Schmitt, N. D. Camp, T. R. Torgerson, D. A. Nickerson, M. J. Bamshad, D. Hagin, C. R. Luthers, J. R. Stinson, J. Gray, I. Lundgren, J. A. Church, M. J. Butte, M. B. Jordan, S. S. Aceves, D. M. Schwartz, J. D. Milner, S. Schuval, S. Skoda-Smith, M. A. Cooper, L. M. Starita, D. J. Rawlings, A. L. Snow, R. G. James, Multiplexed functional assessment of genetic variants in *CARD11*. *Am. J. Hum. Genet.* **107**, 1029–1043 (2020). [doi:10.1016/j.ajhg.2020.10.015](https://doi.org/10.1016/j.ajhg.2020.10.015) Medline
83. W. Coyote-Maestas, D. Nedrud, Y. He, D. Schmidt, Determinants of trafficking, conduction, and disease within a K⁺ channel revealed through multiparametric deep mutational scanning. *eLife* **11**, e76903 (2022). [doi:10.7554/eLife.76903](https://doi.org/10.7554/eLife.76903) Medline
84. S. Erwood, T. M. I. Bily, J. Lequyer, J. Yan, N. Gulati, R. A. Brewer, L. Zhou, L. Pelletier, E. A. Ivakine, R. D. Cohn, Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022). [doi:10.1038/s41587-021-01201-1](https://doi.org/10.1038/s41587-021-01201-1) Medline
85. S. Sun, J. Weile, M. Verby, Y. Wu, Y. Wang, A. G. Cote, I. Fotiadou, J. Kitaygorodsky, M. Vidal, J. Rine, P. Ješina, V. Kožich, F. P. Roth, A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. *Genome Med.* **12**, 13 (2020). [doi:10.1186/s13073-020-0711-1](https://doi.org/10.1186/s13073-020-0711-1) Medline
86. J. Weile, N. Kishore, S. Sun, R. Maaieh, M. Verby, R. Li, I. Fotiadou, J. Kitaygorodsky, Y. Wu, A. Holenstein, C. Bürer, L. Blomgren, S. Yang, R. Nussbaum, R. Rozen, D. Watkins, M. Gebbia, V. Kožich, M. Garton, D. S. Froese, F. P. Roth, Shifting landscapes of human MTHFR missense-variant effects. *Am. J. Hum. Genet.* **108**, 1283–1300 (2021). [doi:10.1016/j.ajhg.2021.05.009](https://doi.org/10.1016/j.ajhg.2021.05.009) Medline
87. J. Zinkus-Boltz, C. DeValk, B. C. Dickinson, A phage-assisted continuous selection approach for deep mutational scanning of protein-protein interactions. *ACS Chem. Biol.* **14**, 2757–2767 (2019). [doi:10.1021/acscchembio.9b00669](https://doi.org/10.1021/acscchembio.9b00669) Medline
88. A. Wan, E. Place, E. A. Pierce, J. Comander, Characterizing variants of unknown significance in rhodopsin: A functional genomics approach. *Hum. Mutat.* **40**, 1127–1144 (2019). [doi:10.1002/humu.23762](https://doi.org/10.1002/humu.23762) Medline
89. A. Sirr, R. S. Lo, G. A. Cromie, A. C. Scott, J. Ashmead, M. Heyesus, A. M. Dudley, A yeast-based complementation assay elucidates the functional impact of 200 missense variants in human PSAT1. *J. Inherit. Metab. Dis.* **43**, 758–769 (2020). [doi:10.1002/jimd.12227](https://doi.org/10.1002/jimd.12227) Medline
90. R. A. Silverstein, S. Sun, M. Verby, J. Weile, Y. Wu, M. Gebbia, I. Fotiadou, J. Kitaygorodsky, F. P. Roth, A systematic genotype-phenotype map for missense variants in the human intellectual disability-associated gene *GDI1*. bioRxiv 2021.10.06.463360 [Preprint] (2022); <https://doi.org/10.1101/2021.10.06.463360>.
91. A. R. Majithia, B. Tsuda, M. Agostini, K. Gnanapradeepan, R. Rice, G. Peloso, K. A. Patel, X. Zhang, M. F. Broekema, N. Patterson, M. Duby, T. Sharpe, E. Kalkhoven, E. D. Rosen, I. Barroso, S. Ellard, S. Kathiresan, S. O’Rahilly, K. Chatterjee, J. C. Florez, T. Mikkelsen, D. B. Savage, D. Altshuler; UK Monogenic Diabetes Consortium; Myocardial Infarction Genetics Consortium; UK Congenital Lipodystrophy Consortium, Prospective functional classification of all possible missense variants in *PPARG*. *Nat. Genet.* **48**, 1570–1575 (2016). [doi:10.1038/ng.3700](https://doi.org/10.1038/ng.3700) Medline

92. J. R. Klesmith, L. Su, L. Wu, I. A. Schrack, F. J. Dufort, A. Birt, C. Ambrose, B. J. Hackel, R. R. Lobb, P. D. Rennert, Retargeting CD19 chimeric antigen receptor T cells via engineered CD19-fusion proteins. *Mol. Pharm.* **16**, 3544–3558 (2019). [doi:10.1021/acs.molpharmaceut.9b00418](https://doi.org/10.1021/acs.molpharmaceut.9b00418) [Medline](#)
93. A. Elazar, J. Weinstein, I. Biran, Y. Fridman, E. Bibi, S. J. Fleishman, Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. *eLife* **5**, e12125 (2016). [doi:10.7554/eLife.12125](https://doi.org/10.7554/eLife.12125) [Medline](#)
94. R. J. Jiang, “Exhaustive mapping of missense variation in coronary heart disease-related genes,” thesis, University of Toronto (2019); <http://hdl.handle.net/1807/98076>.
95. K. K. Chan, D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert, E. Prock, Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261–1265 (2020). [doi:10.1126/science.abc0870](https://doi.org/10.1126/science.abc0870) [Medline](#)
96. P. G. Miller, M. Sathappa, J. A. Moroco, W. Jiang, Y. Qian, S. Iqbal, Q. Guo, A. O. Giacomelli, S. Shaw, C. Vernier, B. Bajrami, X. Yang, C. Raffier, A. S. Sperling, C. J. Gibson, J. Kahn, C. Jin, M. Ranaghan, A. Caliman, M. Brousseau, E. S. Fischer, R. Lintner, F. Piccioni, A. J. Campbell, D. E. Root, C. W. Garvie, B. L. Ebert, Allosteric inhibition of PPM1D serine/threonine phosphatase via an altered conformational state. *Nat. Commun.* **13**, 3778 (2022). [doi:10.1038/s41467-022-30463-9](https://doi.org/10.1038/s41467-022-30463-9) [Medline](#)
97. L. M. Haynes, Z. M. Hutterer, A. Yee, C. A. Kretz, D. R. Siemieniak, D. A. Lawrence, D. Ginsburg, Deep mutational scanning and massively parallel kinetics of plasminogen activator inhibitor-1 functional stability to probe its latency transition. *J. Biol. Chem.* **298**, 102608 (2022). [doi:10.1016/j.jbc.2022.102608](https://doi.org/10.1016/j.jbc.2022.102608) [Medline](#)
98. D. M. Keeling, P. Garza, C. M. Nartey, A.-R. Carvunis, The meanings of ‘function’ in biology and the problematic case of de novo gene emergence. *eLife* **8**, e47014 (2019). [doi:10.7554/eLife.47014](https://doi.org/10.7554/eLife.47014) [Medline](#)
99. J. S. Berg, M. Adams, N. Nassar, C. Bizon, K. Lee, C. P. Schmitt, K. C. Wilhelmsen, J. P. Evans, An informatics approach to analyzing the incidentalome. *Genet. Med.* **15**, 36–44 (2013). [doi:10.1038/gim.2012.112](https://doi.org/10.1038/gim.2012.112) [Medline](#)
100. R. Blekhman, O. Man, L. Herrmann, A. R. Boyko, A. Indap, C. Kosiol, C. D. Bustamante, K. M. Teshima, M. Przeworski, Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18**, 883–889 (2008). [doi:10.1016/j.cub.2008.04.074](https://doi.org/10.1016/j.cub.2008.04.074) [Medline](#)