Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure: implications for variant effect prediction

Lukas Gerasimavicius, Benjamin J Livesey and Joseph A. Marsh*

MRC Human Genetics Unit, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh, UK

**joseph.marsh@ed.ac.uk*

Abstract

Most known pathogenic mutations occur in protein-coding regions of DNA and change the way proteins are made. Taking protein structure into account has therefore provided great insight into the molecular mechanisms underlying human genetic disease. While there has been much focus on how mutations can disrupt protein structure and thus cause a loss of function (LOF), alternative mechanisms, specifically dominant-negative (DN) and gain-of-function (GOF) effects, are less understood. Here, we have investigated the protein-level effects of pathogenic missense mutations associated with different molecular mechanisms. We observe striking differences between recessive *vs* dominant, and LOF *vs* non-LOF mutations, with dominant, non-LOF disease mutations having much milder effects on protein structure, and DN mutations being highly enriched at protein interfaces. We also find that nearly all computational variant effect predictors underperform on non-LOF mutations, even those based solely on sequence conservation. However, we do find that non-LOF mutations could potentially be identified by their tendency to cluster in space. Overall, our work suggests that many pathogenic mutations that act via DN and GOF mutations are likely being missed by current variant prioritisation strategies, but that there is considerable scope to improve computational predictions through consideration of molecular disease mechanisms.

Introduction

Missense single nucleotide variants, which result in the substitution of a single amino acid residue at the protein level, are responsible for a large fraction of all currently known human genetic disorders[1,2]. However, missense variants that are known to be pathogenic represent only a tiny fraction of those that have been observed in the human population[3,4]. Thus, identifying missense changes that are likely to be clinically relevant remains a major challenge for diagnosis, and ultimately treatment, of human genetic disease.

A large number of computational methods have been developed for the prediction of variant effects, and are widely used in clinical sequencing pipelines for the prioritisation of potentially damaging variants. Such variant effect predictors (VEPs) are diverse in their methodologies and implementation, leveraging protein sequences and/or structures, as well as various contextual annotations at the gene, protein or residue level[5,6]. While their current utility in a clinical setting is still limited by their insufficient accuracy[7], the improvement and application of VEPs is expected to remain a major avenue to variant prioritisation[8].

VEPs can provide useful information on the likelihood of mutations being pathogenic, but most tell us nothing about the molecular mechanisms underlying disease mutations. For this, consideration of the protein structural context of mutations can be very informative. In particular, protein stability predictors, which directly evaluate the change in Gibbs free energy of folding (ΔΔG) upon mutation, represent an alternate computational strategy for understanding the effects of missense mutations. Most stability predictors directly utilise protein structures to model the change in stability between the wild-type and variant proteins through a scoring function of pairwise atomic or coarse-grained interactions[9]. While these methods were not specifically designed for the identification of pathogenic variants, they are routinely used when evaluating candidate mutations[10–13] in order to identify those that are damaging to protein structure and will thus cause a loss of function (LOF)[14–16]. Alternatively, increased protein stabilisation can also be associated with disease, and it has been shown that using the absolute ΔΔG values results in higher accuracy when identifying disease mutations, although this may also be due to predictor inability to correctly distinguish the direction of the effect[17,18]. Interestingly, a recent study found that stability predictors performed much better in the identification of pathogenic missense mutations in genes associated with haploinsufficiency[19], supporting the utility of stability predictors for identifying LOF mutations.

Although many pathogenic missense mutations cause a simple LOF, a large number are known to operate via alternate molecular mechanisms. For example, with the dominant-negative (DN) effect, the expression of a mutant protein interferes with the activity of a wild-type protein[20]. This is most commonly observed for proteins that form homomeric complexes, in which the mutant subunits can be considered as poisoning the assembly[21]. Such mutations should not be highly destabilising, as the DN effect is reliant on the mutant protein being stable enough to co-assemble into a complex with the wild type. In fact, it has previously been observed, for a limited subset of transmembrane channel proteins, that DN missense mutations tend to have low predicted ΔΔG values[13]. Similarly, we can hypothesise that gain-of-function (GOF) mutations, which can occur through various mechanisms, such as constitutive activation, shift of substrate or binding target specificity, or protein aggregation[22], should also tend to be mild at a protein structural level.

In this study, we have investigated the effects of pathogenic missense mutations associated with different molecular disease mechanisms on protein structure. We find clear differences between LOF *vs* non-LOF mutations in terms of their structural context, their predicted effects on protein stability, and their clustering in three-dimensional space. Most importantly, we find that nearly all

the VEPs we tested perform worse on DN and GOF mutations, which shows that there are systematic limitations in the ability of current computational predictors to identify disease mutations associated with non-LOF mechanisms.

Results

Consideration of full protein complex structures improves the identification of disease mutations

As the basis for this study, we first compiled a dataset of human missense mutations mapped to three-dimensional protein structures from the Protein Data Bank[23] of 1,261 Mendelian disease genes. This included 13,050 annotated pathogenic and likely pathogenic missense mutations from ClinVar[3], and 211,266 missense variants observed across >140,000 people from gnomAD v2.1[24]. While we recognise that the gnomAD dataset will contain some damaging variants, *e.g.* those that are associated with late-onset disease, population-specific penetrance or are pathogenic under homozygous conditions. The vast majority of the gnomAD variants are classified as rare according to clinical genetics standards (<0.1% allele frequency), and performing allele frequency-based filtering of the gnomAD dataset would drastically diminish the available data and statistical power of our analyses, despite most rare variants observed in the general population being benign or of sub-clinical significance[25]. Thus, we have chosen to include all the gnomAD variants (excluding those annotated as pathogenic in ClinVar) as our 'putatively benign' dataset.

Next, we predicted the effects of all missense changes using FoldX v5[26], a structure-based tool we previously showed to have the best performance in distinguishing between ClinVar pathogenic and gnomAD variants[17]. In agreement with our previous analysis, we found that strongest effects were observed using absolute values, |ΔΔG|. This could reflect two things. First, some pathogenic mutations might increase protein stability, and so the magnitude of the stability perturbation is most useful for identifying these stabilising disease mutations. Second, FoldX, as well as other stability predictors, may be better at predicting the magnitude than the sign of the stability perturbation. This was supported by the fact that different stability predictors are often discordant in whether or not a mutation is stabilising or destabilising[17]. For most analyses, we found it makes very little difference whether we use raw or absolute ΔΔG, as the large majority are destabilising. We will further address the issue of stabilising *vs* destabilising mutations later in the manuscript.

Our previous study on the identification of pathogenic mutations with protein stability predictors considered only monomeric proteins, in particular because many of the predictors we tested work only for single polypeptide chains[17]. As the dataset used in this study was derived from both monomeric and complex structures, we first investigated the impact of using full structures. For each mutation, we calculated ΔΔG for both the protein monomer alone, and for the full biological assembly containing all molecules. Monomer |ΔΔG| represents the effect on the stability of a monomeric protein or isolated subunit, while full |ΔΔG| includes both intra- and intermolecular interactions. For the structures of monomeric proteins, and for mutations not close to intermolecular interfaces, the full and monomer |ΔΔG| values will be identical or nearly identical.

Figure 1A shows that using the full structures in the FoldX calculations increased the average extent of observed stability perturbations of pathogenic mutations by almost 15%, from 3.39 to 3.89 kcal/mol, compared to only using the monomeric subunits. To see how this affects the identification of disease mutations, we performed receiver operating characteristic (ROC) analysis to assess discrimination between the ClinVar and gnomAD mutations. The area under the curve (AUC), which corresponds to the probability of a randomly chosen disease mutation being assigned a higher-ranking |ΔΔG| value than a random gnomAD variant[27], was used as a quantitative classification

performance metric. As is evident from **Figure 1B**, the curve, derived from full |ΔΔG| values, resides higher than the performance curve of monomer prediction values over the entire threshold range. Using full |ΔΔG| values results in an AUC of 0.677, which is significantly higher ($P = 4.3 \times 10^{-71}$) than the AUC of 0.652 when only using monomeric structures. While this is unsurprising, given the common role of intermolecular interactions in human genetic disease[28,29], it emphasises the importance of considering full protein complex structures, when available.

Recessive mutations are more structurally perturbing than dominant mutations

Next, we investigated mode of inheritance, which is closely related to molecular disease mechanism. Autosomal recessive (AR) disorders are overwhelmingly associated with LOF, whereas autosomal dominant (AD) disorders can have different underlying molecular mechanisms[22]. While some dominant mutations will cause disease via LOF (*i.e.* haploinsufficiency), many will be DN or GOF. Thus, we expect that differences in the structural effects of recessive *vs* dominant missense mutations should be reflective of the differences between LOF and non-LOF mechanisms.

Disease inheritance annotations for genes were obtained from OMIM[30]. To allow a simplified analysis at the gene level, we only investigated genes with either autosomal recessive (726 genes) or autosomal dominant (535 genes) inheritance, excluding those with mixed inheritance. **Figure 2A** compares the monomer and full |ΔΔG| values for AD and AR disease mutations, and the putatively benign gnomAD variants. While all groups exhibit a high degree of heterogeneity, AR mutations are significantly more perturbing, with a mean difference of 1.1 kcal/mol for monomer |ΔΔG| and 0.9 kcal/mol for full |ΔΔG| compared to AD mutations. Interestingly, AD mutations are approximately intermediate compared to gnomAD and AR variants.

The differences in perturbation magnitude across the different mutation groups can be partially explained by their enrichment in different spatial locations (**Figure 2B**). It is well known that pathogenic mutations are common in protein interiors and at interfaces, where they can act via protein destabilisation or disruption of interactions, while they are underrepresented on protein surfaces[28,29,31–33]. We find that AR mutations are most enriched in protein interiors (58%), while only 15% occur at the protein surface. In contrast, 43% of the gnomAD variants occur on the surface and 38% in the interior. AD mutations appear intermediate between AR and gnomAD, with 43% in the interior and 26% on the surface. Interestingly, however, the AD group is the most highly enriched at protein interfaces (31%), compared to 27% for AR and 20% for gnomAD, hinting at the importance of intermolecular interactions for understanding alternate molecular disease mechanisms. Importantly, in **Figure S1** we show that the differences in perturbation magnitudes are still observed when controlling for interior, interface and surface locations.

The ability to identify pathogenic mutations also depends upon the properties of the benign variants from which they must be discriminated, so we also compared the |ΔΔG| values of gnomAD variants from AD *vs* AR disease genes. The properties of gnomAD variants from AR genes are expected to be quite different compared to those from AD genes, as AR genes will, by definition, tend to be more tolerant of heterozygous damaging mutations. Therefore, we also selected a subset of variants that have been observed in a homozygous state at least once in gnomAD, with the expectation that these should be more reflective of truly benign variants (although at the cost of being a much smaller dataset). Interestingly, we find in **Figure S2** that, while the gnomAD variants in AR genes are collectively much more damaging than those from AD genes, when we consider the homozygous subset of variants from AR genes, they are very similar to AD mutations.

Finally, we assessed whether the inheritance mode of variants affects the performance of disease variant identification using predicted stability effects. We observe a remarkable improvement in the

identification of pathogenic missense mutations from AR genes compared to AD genes as measured by ROC AUC, which is even greater when the homozygous set of putatively benign gnomAD variants are used (**Figure 2C**). This suggests that predicted changes in protein stability are far more useful for the identification of recessive mutations, due to the fact that recessive disorders are much more likely to be associated with LOF.

Gain-of-function and dominant-negative mutations have much milder effects on protein structure than loss-of-function mutations

To compare the effects of mutations associated with different molecular disease mechanisms, we attempted to classify AD disease genes into those associated with haploinsufficiency (HI), DN effects, or GOF, using a combination of keyword searches, manual curation of OMIM[30] entries, and ClinGen[34] annotations (see Methods). While this approach is necessarily imperfect, in that it assumes that all disease mutations from the same gene will be associated with the same mechanism, it represents the most feasible strategy we currently have available for investigating the general properties of these different types of mutations on a larger scale.

We first explored predicted stability effects using monomer |ΔΔG| values (**Figure 3A, left**). Interestingly, a clear difference emerges between the loss-of-function (HI) and non-LOF (DN and GOF) mutations, with the HI mutations being far more disruptive to protein structure. The HI mutations show a nearly identical distribution to the AR mutations, with |ΔΔG| means of 4.18 and 4.02 kcal/mol, respectively, and no statistically significant difference. In contrast, the DN and GOF mutations are much milder, with |ΔΔG| means of 2.66 and 2.42 kcal/mol, respectively, and also showing no statistically significant differences from each other. In fact, although the DN and GOF mutations are considerably more perturbative than the putatively benign gnomAD variants, they are closer to those in magnitude than they are to the LOF mutations.

When intermolecular interactions from full protein structures are also taken into consideration, using full |ΔΔG| values (**Figure 3A, right**), a similar overall pattern is observed, with LOF mutations being the most damaging, and DN and GOF mutations being intermediate between LOF and gnomAD. Interestingly, however, DN mutations show a large rise in predicted |ΔΔG|, increasing from 2.66 to 3.32 kcal/mol, while GOF mutations show only a small increase, from 2.42 to 2.65 kcal/mol, and the overall difference between DN and GOF mutations becomes highly significant.

Figure 3B demonstrates that the DN mutations are far more likely to affect intermolecular interactions, with 45% of mutations occurring at the interface. In contrast, GOF mutations show little interface enrichment, thus explaining the significant difference between DN and GOF mutations when considering full |ΔΔG| values. Of course, this result is probably influenced by the fact that DN mutations are far more likely to be found in protein complexes[21].

Using ROC analysis, we tested the ability of the predicted stability perturbations to distinguish between pathogenic and putatively benign gnomAD variants across all the different molecular disease mechanism groups (**Figure 3C)**. Most strikingly, we see that LOF mutations (HI and AR) are predicted far better than DN and GOF mutations, with AUCs 0.1-0.15 higher for LOF *vs* non-LOF mutations. We also note a 0.05 AUC improvement in the prediction of DN mutations when considering intermolecular effects full |ΔΔG| values, consistent with our observations above, although this is still well below the performance observed for LOF mutations.

We previously demonstrated that optimal performance of stability predictors for the identification of pathogenic mutations is subject to significant gene and predictor-specific heterogeneity[20]. While FoldX |ΔΔG| values showed an optimal threshold for distinguishing between pathogenic and benign

close to 1.5 kcal/mol, a value that has been previously utilised in several studies on variant stability perturbation[35–37], we observed that on a per-gene level the optimal thresholds varied considerably. Therefore, we explored whether knowledge of the underlying molecular disease mechanism could inform us of a more optimal threshold to use for variant prioritisation. **Table S1** lists the optimal FoldX |ΔΔG| thresholds for gene groups based on our molecular mechanism annotations. We observe considerable differences in optimal thresholds, ranging from 1.16 kcal/mol for GOF mutations to 1.59 kcal/mol for HI mutations when using full |ΔΔG|. Interestingly, the two LOF group thresholds correspond quite closely to the previously used 1.5 kcal/mol. However, our findings signify that using this single threshold in a practical setting would lead to considerable underprediction of non-LOF disease variants. It is currently unlikely that practical prediction performance can be dramatically improved by mechanism-specific threshold choices, but awareness of the unique protein-and variant-level features underlying non-LOF variants could reveal future avenues for protein-specific prediction approaches.

Stabilising mutations may be important for gain of function

We also hypothesised that, in addition to be milder in magnitude, pathogenic GOF mutations might have a tendency to increase protein stability, *e.g.* by stabilising activated states. Similarly, DN mutations could conceivably be associated with increased rigidification, allowing a mutant protein to 'poison' a complex containing a mixture of mutant and wild-type subunits. Indeed, a number of pathogenic mutations that increase protein stability exist in the literature[38–40]. To address this, we need to consider the actual ΔΔG values, including sign, and not the absolute values used above. In **Figure S3A**, we plot the distribution for full ΔΔG values different types of pathogenic mutations. From this, we can see that, while the large majority of mutations in each class are predicted to have a net destabilising effect (ΔΔG > 0), there are still many stabilising mutations. **Figure S3B** shows the fraction of stabilising mutations (ΔΔG < 0) for each group. This shows that predicted stabilising effects are most common in GOF mutations (21%), followed by DN (17%). In contrast, LOF mutations have a much lower tendency to be stabilising (12% for HI and 10% for AR).

While the above analysis shows that non-LOF mutations are more likely to have predicted stabilising effects, it is possible that this is simply due to the fact that they tend to be of lower magnitude. If one assumes a typical error of ~1 kcal/mol for a FoldX ΔΔG prediction[41], then we can expect that many of the low magnitude ΔΔG values might appear to be stabilising due to this margin of error. We attempted to control for this by considering predicted stabilising *vs* destabilising effects in two groups: those with |ΔΔG| < 1 kcal/mol, and those with |ΔΔG| from 1-4 kcal/mol. We then calculated Fisher log odds enrichment of stabilising mutations within each class from among all disease mutations. As illustrated in **Figure S3C**, the tendency for predicted stabilisation to be most enriched among GOF mutations is observed in both groups ($P = 1.73 \times 10^{-4}$ and 0.015 for |ΔΔG| < 1 kcal/mol and 1-4 kcal/mol groups, respectively). In contrast, destabilisation is most enriched among AR mutations ($P = 4.36 \times 10^{-5}$ and 7.75×10^{-4}). Interestingly, stabilisation appears to be more common in HI than AR mutations, which is difficult to explain, considering that we would expect both to act via similar LOF mechanisms. One possible explanation relates to the common occurrence of haploinsufficiency among transcription factors: increased stability might occasionally be a mechanism that inhibits DNA binding via a loss of flexibility.

Nearly all variant effect predictors underperform on non-loss-of-function mutations

Given that DN and GOF mutations tend to be mild at a protein structural level, and are thus poorly identified using FoldX |ΔΔG| values, we wondered whether this might also have similar implications for the performance of other computational VEPs. Initially, our hypothesis was that, while predictors that incorporate protein structural information might perform worse on non-LOF mutations, those

based primarily on evolutionary conservation should be relatively insensitive to differences in molecular disease mechanism. Therefore, we tested a diverse set of 41 VEPs for their performance on different types of disease mutations.

Table S2 outlines the different VEPs used, which we classified into several groups. First, we included those based purely on amino acid sequence conservation. Multi-feature methods are those that rely on multiple different features, although they all also include sequence conservation. Metapredictors are those VEPs that derive predictions based on the outputs of two or more other predictors. We also considered several simple amino acid substitution matrices. Finally, we considered some methods based purely on DNA-level features, primarily nucleotide conservation. It is important to note that a majority of the VEPs in our multi-feature and metapredictor groups were derived in a supervised fashion, which makes them susceptible to circularity due to dataset overlaps with training data[6,42]. While this prevents us from being able to make fair judgements on absolute performance of different predictors, as some are likely to have been trained on some of the mutations in our dataset, we are more interested in the relative performance within predictors against different molecular disease mechanism groups.

Figure 4 shows the ROC AUC values for discriminating between pathogenic ClinVar and putatively benign gnomAD missense variants across all predictors. Remarkably, we observe nearly universal underperformance on non-LOF mutations. For the large majority of protein-level predictors, the LOF mutations (HI and AR) show higher AUC values than the non-LOF (DN and GOF) mutations.

How can we explain this consistent underperformance against non-LOF mutations? One clue comes from examination of the substitution matrices. These are very simple models, derived from physicochemical properties of amino acids, or patterns of evolutionary substitutions, that will always give the same value for the same type of amino acid substitution. Effectively, they show how different the mutant amino acid residue is from the wild type. Across all substitution matrices, HI mutations are predicted better than DN and GOF mutations, which simply tells us that the non-LOF mutations tend to involve more similar amino acid changes than LOF mutations. Thus, despite the fact that these models do not use protein structural information, they are still reflecting the milder nature of the pathogenic non-LOF substitutions.

In principle, the substitution matrix results could also explain the performance of some other predictors based on amino acid sequence conservation thaty also incorporate a substitution matrix (*e.g.* SIFT[43]). Notably, however, the underperformance of non-LOF mutations is also observed for DeepSequence[44], which is based purely on multiple sequence alignments of the protein of interest and does not explicitly use any external substitution scores for variant evaluation. Moreover, the LOF mutations also tend to be better predicted than non-LOF mutations by those methods based on nucleotide-level sequence conservation, which utilise no information about amino acid substitution type. Thus, it does appear that amino acid residues associated with DN or GOF mutations genuinely tend to show weaker evolutionary conservation than those associated with LOF mutations, thus providing another factor that can explain their poor identification by VEPs.

While the tendency for LOF mutations to be better predicted than non-LOF mutations is observed across all of the top-performing predictors, there are a few exceptions. Notably, MPC shows very similar performance for HI, DN and GOF mutations, and worse performance for AR mutations. This can likely be explained by the fact that it was developed specifically using haploinsufficient and dominant disease genes[45]. It is also interesting to note the performance of the nucleotide-level prediction methods. Consistent with our previous observation, they tend to be worse than protein-level predictors for identifying pathogenic mutations[6]. Interestingly, however, all of these methods perform better on GOF *vs* DN mutations, and 6/7 perform better on GOF *vs* HI mutations.

Overall, the tendency for LOF mutations to be predicted better than non-LOF mutations is very clear. However, the differences are relatively small between DN and GOF mutations across most predictors. Interestingly, the DN mutations are better predicted across all substitution matrices, suggesting that DN mutations tend to involve more perturbative amino acid substitutions. In contrast, GOF mutations are better predicted by all methods based purely on sequence conservation (the nucleotide-level methods and DeepSequence). It appears that these two competing factors - DN mutations being more perturbative, and GOF mutations occurring at more conserved positions - effectively cancel each other out across most VEPs that utilise both sources of information, resulting in very similar predictions for the two groups. However, it does suggest that comparison of conservation and amino acid substitutions may hold some value in discriminating between DN and GOF mutations.

Gain-of-function and dominant-negative mutations tend to cluster in three-dimensional space

Finally, since non-LOF mutations tend to be poorly predicted by existing computational approaches, we wondered whether there is other information that could be used to better predict them. We hypothesised that DN and GOF missense mutations should tend to cluster within specific regions, while LOF missense mutations should be more evenly spread throughout a protein. The reason for this is simple: destabilising mutations should be able to occur throughout a protein sequence, at least within its folded domains, while the more specific effects associated with DN and GOF mutations may be more localised to particular regions. Previously, there has been some evidence for this. For example, a recent study found that genes with more spatially clustered disease variants tend to be associated with mechanisms other than haploinsufficiency[46]. Other studies have shown dominant GOF mutations exhibit more focal, shorter distance clustering at sites of functional potential, compared to recessive LOF variants, which occur more uniformly throughout structures[47,48]. Another study of cancer-associated variation found that GOF mutations in oncogenes tend to be more clustered compared to LOF mutations in tumour suppressors and other targets[49].

To address this in more detail, we defined a simple metric for disease mutation clustering, based on the proximity of each protein residue to a known disease mutation at another residue (see Methods). The final clustering metric is presented as a ratio for each protein, whereby a value of one indicates that disease mutations tend to be randomly spaced throughout the protein, while a higher value indicates that the disease mutations tend to cluster together.

In **Figure 5A**, we compare the distributions of clustering values for AR compared to AD disease genes, under the assumption that AR genes should be primarily associated with LOF mechanisms, while AD genes can be associated with LOF or non-LOF mechanisms. Remarkably, we observe a very strong, highly significant tendency for disease mutations in AD genes to be more clustered than those in AR genes. In **Figure 5B**, we compare the distributions of clustering values for AD genes associated with different molecular disease mechanisms. Consistent with our previous observations, we find that DN and GOF mutations tend to be significantly more clustered than HI mutations. Together, these results strongly support the idea that non-LOF mutations tend to cluster together more than LOF mutations. However, there is no significant difference between clustering of DN and GOF mutations. Thus, the utilisation of information about mutation clustering in three-dimensional space shows promis for improving predictions of pathogenic non-LOF mutations.

Discussion

Our first observation in this study was the importance of considering full protein complex structures, if available, when using stability predictors to investigate potential mutation pathogenicity. This is

unsurprising, as many pathogenic mutations are known to occur at protein interfaces[33], and the impact of these will be missed when considering only monomeric structures. Thus, the effects of a mutation on a protein complex structure should be tested if possible. While there will be many cases where no structure of a biologically relevant complex is available, advances in protein modelling and intermolecular docking may be helpful in the future.

Next, we found a striking association between predicted changes in protein stability, inheritance and molecular disease mechanisms. LOF mutations associated with AR and HI genetic disorders can be predicted quite well using changes in protein stability, but dominant GOF and DN mutations tend to have mild effects at a protein structural level. While there has been limited evidence of this is the past for DN mutations in transmembrane proteins[13], this is the first time this has been assessed on such a large scale for both DN and GOF mutations.

Given the above observations, it is essential to emphasise that the primary benefit of stability predictors with respect to disease mutations is in understanding molecular mechanisms. When a mutation is predicted to destabilise a protein or disrupt an interaction, this gives us a plausible molecular disease mechanism. However, that a mutation is structurally mild should not be taken as strong evidence that it is likely to be benign, unless all other known pathogenic mutations are known to be destabilising. There is a tendency in the literature when reporting new disease mutations to only include predicted changes in stability if they are high. We suggest that even mild changes in predicted or observed stability should be reported, as this can provide clues as to potential molecular mechanisms.

A major limitation of our study arises from our strategy for annotating HI *vs* DN *vs* GOF mutations. We are reliant on the descriptions of mutations that have been compiled into the OMIM database, which are somewhat qualitative. In fact, it is often very difficult to distinguish between DN and GOF mutations experimentally, and thus there is a large set of mutations we excluded from our analyses because OMIM described them as potentially being either DN or GOF, but could not distinguish between the two. Moreover, for simplicity, we assume that all pathogenic missense mutations in the same gene are associated with the same molecular mechanism, which is clearly not true for all genes. Ideally, there would be some large-scale, unbiased method for classifying the molecular mechanisms underlying different mutations from different genes, but this is unlikely in the near term. However, the significant trends that we observe with our approach strongly suggests that we are accurately capturing systematic differences between different types of mutations. It is highly likely that, if we had more accurate, mutation-level classifications of molecular mechanisms of disease mutations, the trends we have observed here would be even more pronounced.

While we have shown that DN and GOF mutations tend to have mild effects on protein stability compared to LOF mutations, the actual specific mechanistic details underlying most non-LOF mutations remain indeterminate. We know that the DN effect tends to be associated with the assembly of proteins into complexes[21,50,51], and this is supported by our observation of an enrichment of DN mutations at interface positions. However, our finding that including intermolecular interactions and using the full $|\Delta\Delta G|$ values significantly improve the identification of pathogenic DN mutations is in a way paradoxical, as it implies that DN mutations tend to perturb interactions. A mutation that completely disrupts assembly of a protein complex would not be compatible with a DN effect, as there would be no way for the mutant protein to affect the wild-type protein. However, it is possible that DN mutations often cause milder perturbations, not entirely blocking protein complex assembly, but sufficient to affect the function of a full complex. Alternatively, by 'loosening' interactions in the complex, DN mutations may cause all subunits, including wild type, to be mislocalised or degraded, as in the example of ALDH2[21,52]. As for GOF

mutations, they are likely to be much more heterogeneous in their specific molecular effects, *e.g.* excessive protein stabilisation, toxic aggregation, modulation of protein activity or interaction selectivity[39,40,53,54].

Previously, there has been very little consideration of how variant effect predictors perform for pathogenic mutations associated with different molecular mechanisms. One previous study found that SIFT and PolyPhen appeared to underperform on GOF mutations, but this involved a very small set of known disease mutations[55]. Our observation that nearly all tested VEPs perform worse on non-LOF mutations was very surprising, and has major implications for the identification of pathogenic variants. Computational predictors are very widely used for the prioritisation of potentially causal variants, and if they are systematically underperforming on DN and GOF mutations, this could translate into large numbers of true pathogenic variants being missed by current sequencing studies. Given our results, there are no specific predictors we can particularly recommend if DN or GOF mutations are suspected, although we continue to, in general, recommend DeepSequence[44] or other related unsupervised methods based upon our previous benchmarking study[6]. At the moment, we suggest that, if a DN or GOF mechanism is suspected, one should be very careful about filtering out variants due to lower VEP scores.

The clustering of DN and GOF mutations in three-dimensional space presents one possible strategy for improving their prediction. However, this is limited to those genes where multiple disease mutations are already known, which greatly limits the applicability of clustering-based approaches. In addition, even though we see a strong enrichment of clustering in DN and GOF mutations, clustering of pathogenic missense mutations can still be associated with a LOF mechanism. For example, if a protein has multiple folded domains, separated by flexible linkers, and destabilising LOF mutations are most likely to occur in the folded domains, this will result in clustering of LOF mutations. The strategy we have used here may be relatively immune to this, as we only considered structured regions of proteins, but if one uses full-length protein models that include flexible/disordered regions (*e.g.* AlphaFold models[56]), or a sequence-only clustering approach, then more significant clustering of LOF mutations might be observed.

Another way of improving the identification of non-LOF mutations is with high-throughput experimental approaches, *e.g.* deep mutational scanning (DMS). These experiments have shown tremendous power for the direct identification of pathogenic variants[57–61], and if an appropriate experimental phenotype is chosen, it should be possible to measure an appropriate variant effect for DN and GOF mutations. However, not all DMS strategies are appropriate for different types of mutations. A recent study using fluorescent-coupled PTEN to measure the abundance of different variants was able to identify destabilising disease mutations, but not known DN mutations[62]. Future work will be required to determine the best ways to systematically identify pathogenic mutations associated with different molecular mechanisms using a DMS-like strategy.

Methods

The pathogenic missense mutations used in this study were downloaded from ClinVar on 2021.01.28, selecting those labelled as ‘pathogenic’ and ‘likely pathogenic’. Putatively benign missense variants were taken from gnomAD v2.1.1. Any variants present in the ClinVar dataset were excluded from the gnomAD set. No filtering for allele frequency was performed in the gnomAD variants, as this would dramatically reduce the size of our dataset. For certain analyses involving recessive mutations, we used only those gnomAD variants that had been observed at least once in a homozygous state.

Protein structures were downloaded from the Protein Data Bank on 2020.08.17. Mutations were mapped to protein structures in the same manner as previously described[33], considering those polypeptide chains with >90% sequence identity to a human protein over a region of at least 50 amino acid residues. While in some cases this includes non-human structures, allowing us to substantially increase the size of our dataset, mutations were only mapped to structures where the residue of interest, as well as its adjacent neighbours, were the same as the human wild-type sequence.

FoldX 5.0 calculations were performed using all default parameters, essentially as previously described[17]. The procedure was modified to include structures of protein complexes, and the FoldX calculations were set to also take into account intermolecular interactionss. Other VEP values were obtained using our previously described pipeline[6]. Where available, other VEP predictions were obtained using the dbNSFP database version 4.0[63]. Further predictor scores were download from online sources (PonP2), obtained from predictor web-interfaces (SNAP2, SuSPect, NetDiseaseSNP) or run locally (SIFT, DeepSequence). While for most VEPs, we have predictions for the vast majority of mutations considered in this study, due to the high computational requirements of DeepSequence, we were only able to run it for 168 proteins. The full list of variant effect predictors employed in this work together with the sample sizes for ClinVar and gnomAD variant predictions can be found in **Table S2**.

Structural locations were classified as interior, surface or interface according to a previously classification[64]. Interface residues show a solvent accessible surface area difference between the free subunit and full protein structure. Other residues with less than 25% relative solvent accessible surface area in the full structure were classified as interior, while the remainder was designated as surface.

Molecular disease mechanism annotations for genes were derived based upon information available in the OMIM[30] and ClinGen databases[34]. First, only genes that were annotated with an inheritance of 'AD' in OMIM were considered. Those genes annotated as 'Sufficient evidence for dosage pathogenicity' in ClinGen were classified as HI. Next, we searched all OMIM entries for the keywords 'dominant negative', 'gain of function' and 'activating mutation'. Then, we manually read the OMIM entries for all genes identified in this search. If there was evidence in the OMIM entry that a pathogenic missense mutation was due to one of these mechanisms, then it was assigned as DN or GOF. Importantly, we tried to keep this process as unbiased as possible, so we only considered information available in the OMIM entries. While for many genes, there is further evidence available in the literature regarding molecular disease mechanisms available, this could lead to bias on our part if we spend more time investigating certain genes than others.

For the disease mutation clustering analysis, we defined a metric based on the proximity of each protein residue to a known disease mutation at another residue. Essentially, for each residue in a protein structure, we calculated the Cα:Cα distance to all other residues with known disease mutation, and the closest distance was selected. The average distance to another disease mutation residue was calculated for the sites of all disease mutations, and for all other residues. The final clustering metric is presented as a ratio of the two values for each protein, whereby a value of one indicates that disease mutations tend to be randomly distributed throughout the protein, while a higher value indicates that the disease mutations tend to cluster together.

All mutations, variant and gene annotations, corresponding structure identifiers, predictor and inheritance or mechanisms-based group gene clustering values are provided in the **Supplemental Dataset**.

Pairwise statistical comparisons between ΔΔG value groups were carried out using Dunn's test implementation in the R 'ggstatsplot'[65] package, with the p-values for comparisons involving more than two groups being adjusted through Holm's multiple comparison correction. Statistical differences for variant spatial location ratios were assessed using the pairwise Chi-square test implemented in 'rmngb'[66] R package, fraction error bars were derived using the 'binom.confint' function from the 'binom'[67] package. Fisher odds ratios for the stabilising variant analysis were derived using the base R 'fisher.test' function.

ROC analysis was performed in R using the 'pROC' package[68], with AUC curve differences being statistically assessed through 2,000 bootstraps using the 'roc.test' function. Optimal classification thresholds were calculated at default settings using the 'closest.topleft' argument for all proteins with at least 20 gnomAD and pathogenic or likely pathogenic ClinVar variants per-gene, and determined through the closest ROC point distance to the top-left corner of the plot[69]. In the VEP ROC analysis case-control direction was adjusted individually for each predictor to produce positive predictiveness values above 0.5 AUC.

Acknowledgements

This project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001169), and the Medical Research Council (MR/M02122X/1). JM is a Lister Institute Research Fellow. BL is supported by the Medical Research Council Precision Medicine Doctoral Training Programme.

Data Availability

The Supplemental Dataset is available at https://doi.org/10.7488/ds/3153

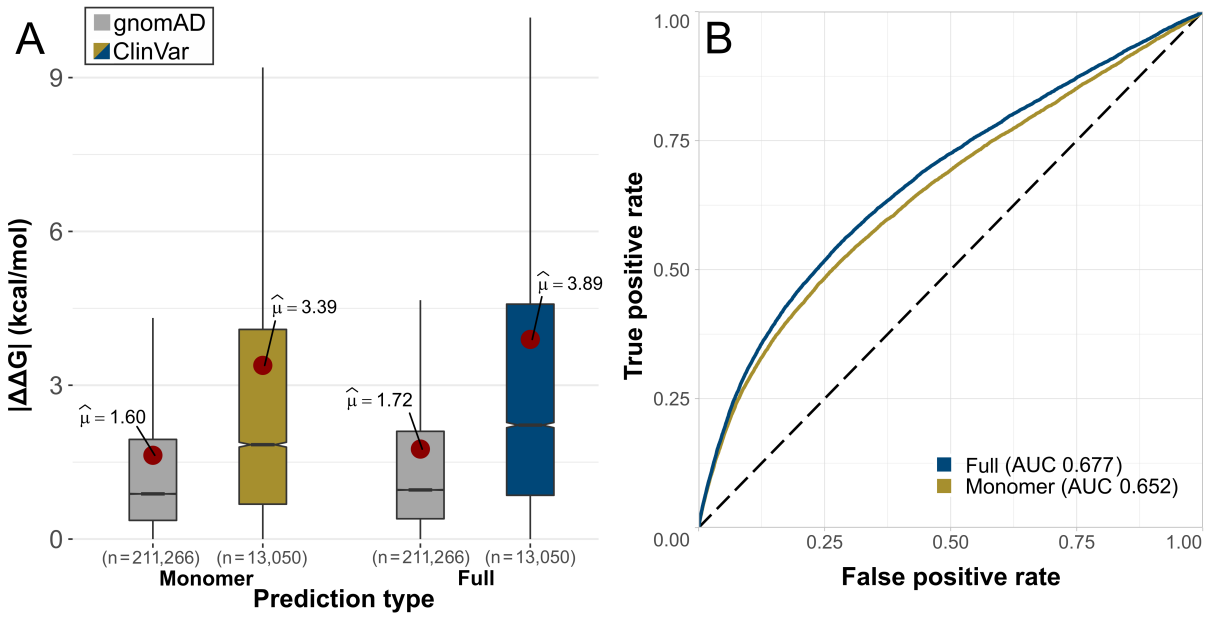

Figure 1. Using full protein complex structures improves discrimination between pathogenic and putatively benign missense variants. (A) Comparison of predicted |ΔΔG| values for putatively benign gnomAD and pathogenic ClinVar mutations when calculated using monomer structures only, or by also employing full biological assemblies and therefore also considering intermolecular interactions. All pairwise group comparisons are significant ($P < 2.5 \times 10^{-28}$, Holm-corrected Dunn's test). **(B)** Receiver operating characteristic (ROC) curves and area under the curve (AUC) for discrimination between ClinVar and gnomAD mutations using monomer and full |ΔΔG| values.

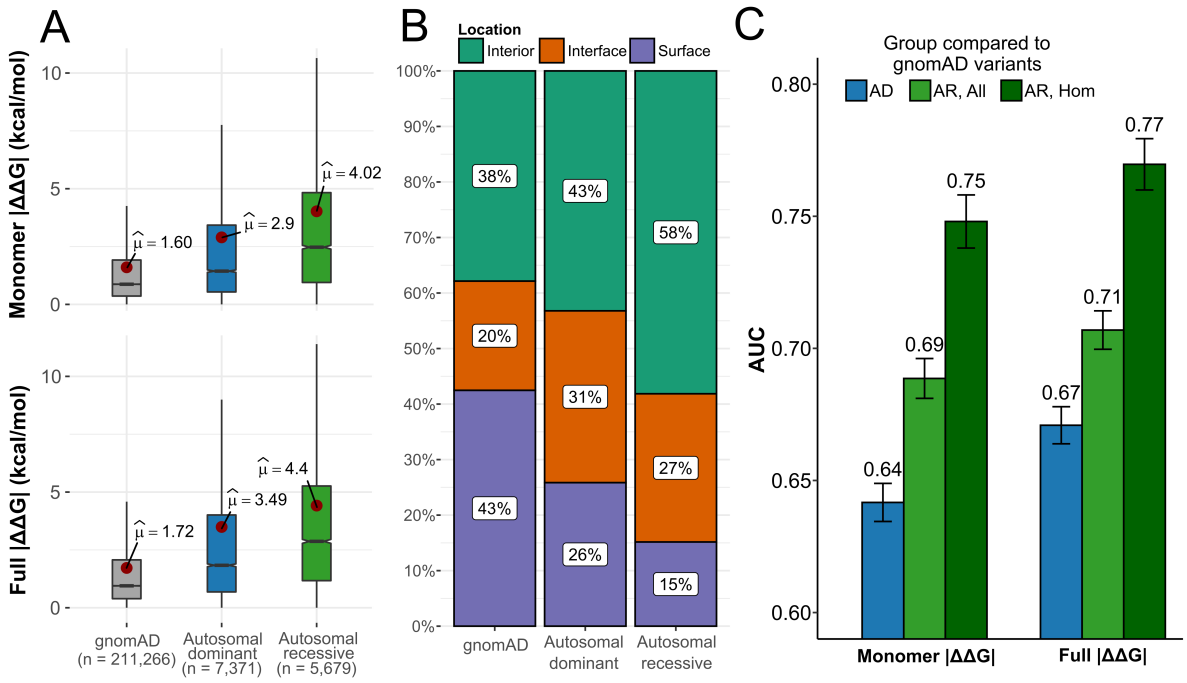

Figure 2. Autosomal dominant mutations are milder than autosomal recessive mutations. (A) Comparison of predicted |ΔΔG| values for gnomAD variants and pathogenic ClinVar mutations from autosomal dominant (AD) and autosomal recessive (AR) genes. All pairwise comparisons between groups are significant ($P < 6.4 \times 10^{-72}$, Holm-corrected Dunn's test). **(B)** Proportions of gnomAD, AD and AR mutations occurring in protein interiors, surfaces and interfaces. **(C)** AUC values calculated from ROC curves for discriminating between pathogenic ClinVar mutations from AD or AR genes, and putatively benign gnomAD variants. 'AR, All' considers all gnomAD variants from the AR genes, while 'AR, Hom' only includes those variants that have been observed in a homozygous state in gnomAD at least once. Error bars denote 95% confidence intervals, which were computed with 2000 stratified bootstrap replicates.

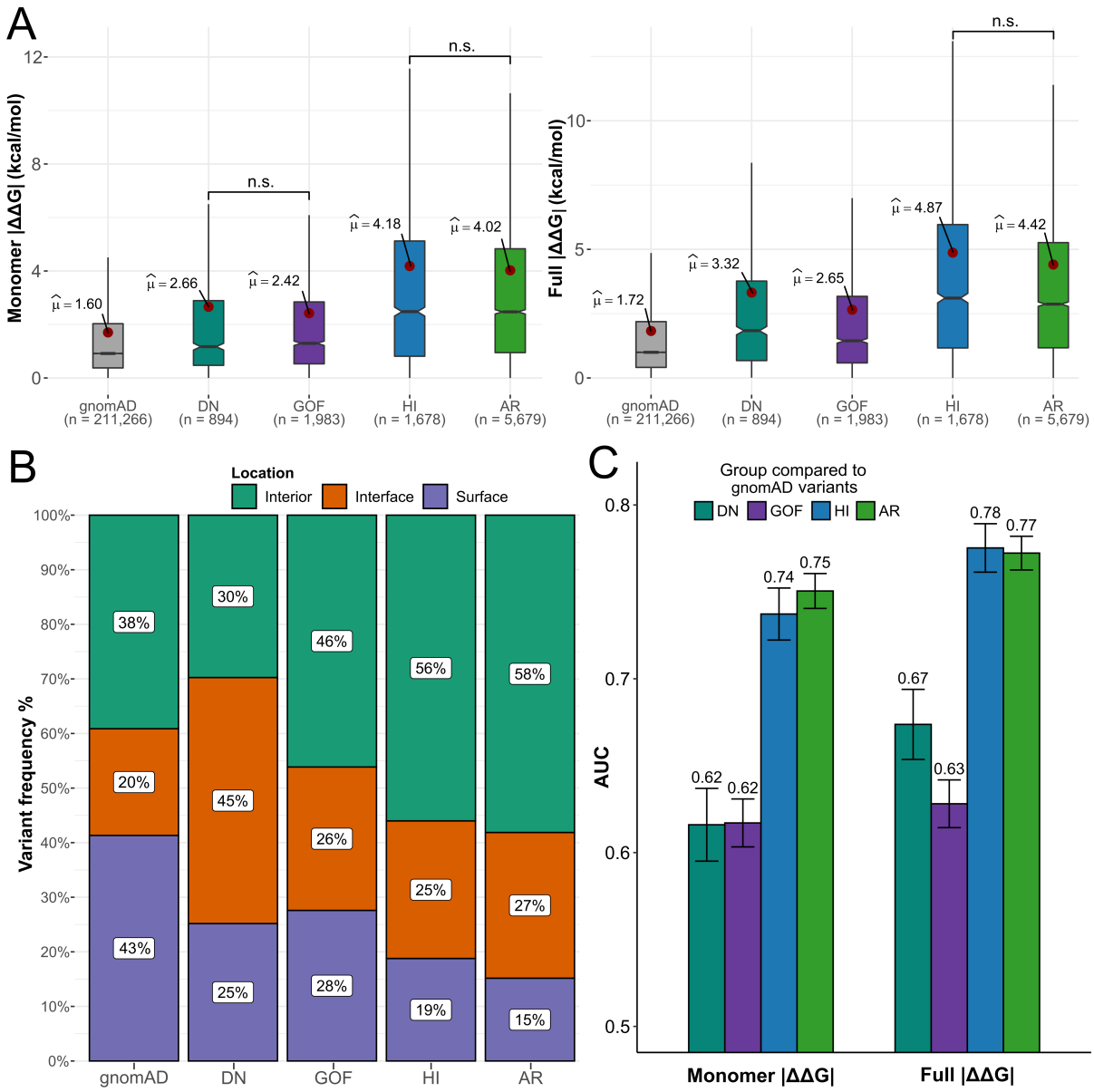

Figure 3. Gain-of-function and dominant-negative mutations are milder than loss-of-function mutations. (A) Comparison of predicted |ΔΔG| values for gnomAD variants and pathogenic ClinVar mutations from AD genes classified as dominant negative (DN), gain of function (GOF) and haploinsufficient (HI), and pathogenic ClinVar mutations from AR genes. Pairwise group comparisons are significant ($P < 0.001$, Holm-corrected Dunn's test) unless specified (n.s.). **(B)** Proportions of different types of mutations occurring in protein interiors, surfaces and interfaces. **(C)** AUC values calculated from ROC curves for discriminating between different types of pathogenic ClinVar mutations, and putatively benign gnomAD variants. Only homozygous gnomAD variants were included for the AR analysis. Error bars denote 95% confidence intervals, which were computed with 2000 stratified bootstrap replicates.

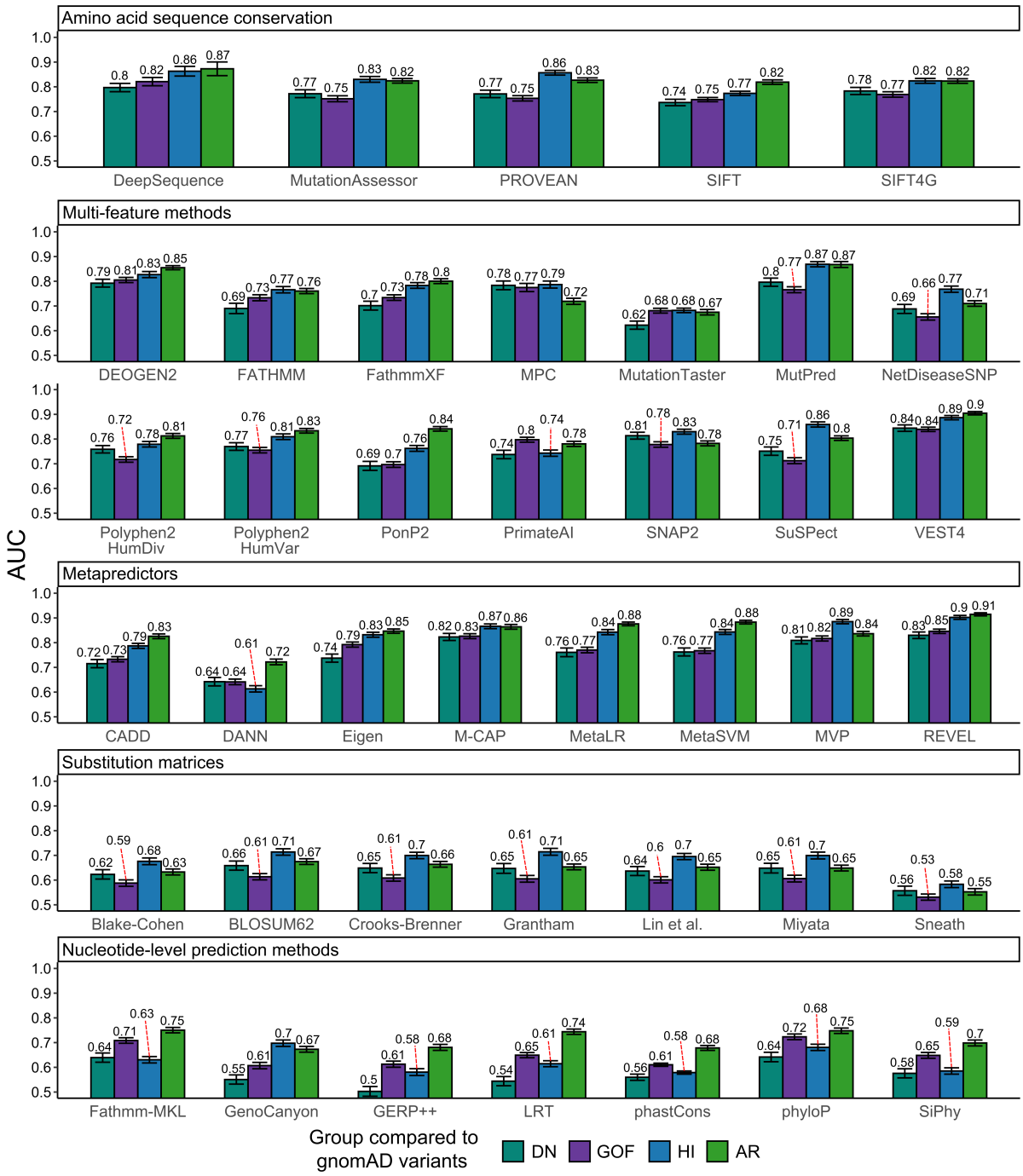

Figure 4. Dominant-negative and gain-of-function mutations are less well identified by nearly all computational variant effect predictors than loss-of-function mutations. AUC values calculated from ROC curves for discriminating between different types of pathogenic ClinVar mutations and putatively benign gnomAD variants, using the outputs of different computational variant effect predictors. Only homozygous gnomAD variants were included for the AR analysis. Error bars denote 95% confidence intervals, which were computed with 2000 stratified bootstrap replicates.

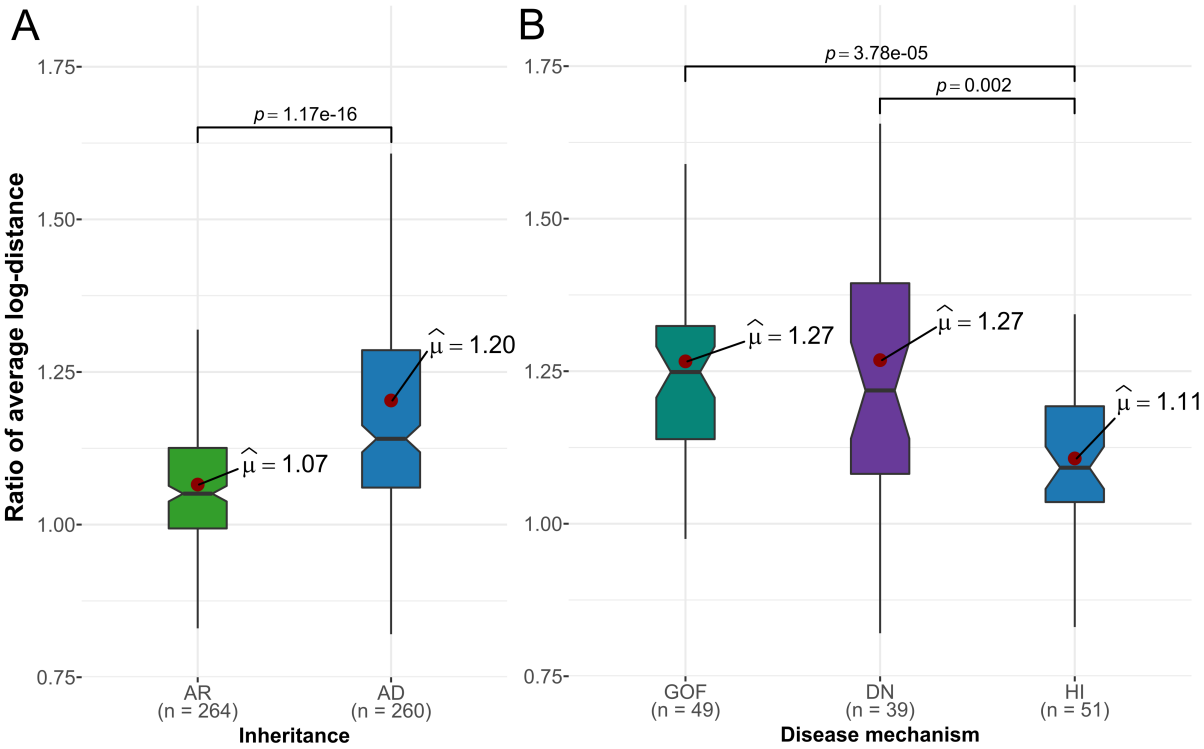

Figure 5. Pathogenic mutations in genes associated with dominant-negative and gain-of-function mechanisms are significantly more clustered in space than those associated with loss-of-function mechanisms. (A) Distribution of clustering values for autosomal recessive (AR) compared to autosomal dominant (AD) disease genes. **(B)** Distribution of clustering values for AD disease genes associated with gain-of-function (GOF), dominant-negative (DN) and haploinsufficient (HI) mechanisms. P-values are calculated with the Wilcoxon rank-sum test. The metric for disease mutation clustering was derived based on the Euclidian distance of each protein residue to a known disease mutation position. The final clustering metric is presented as a ratio for each protein, with higher values indicating that the disease mutations tend to cluster closer together in space.

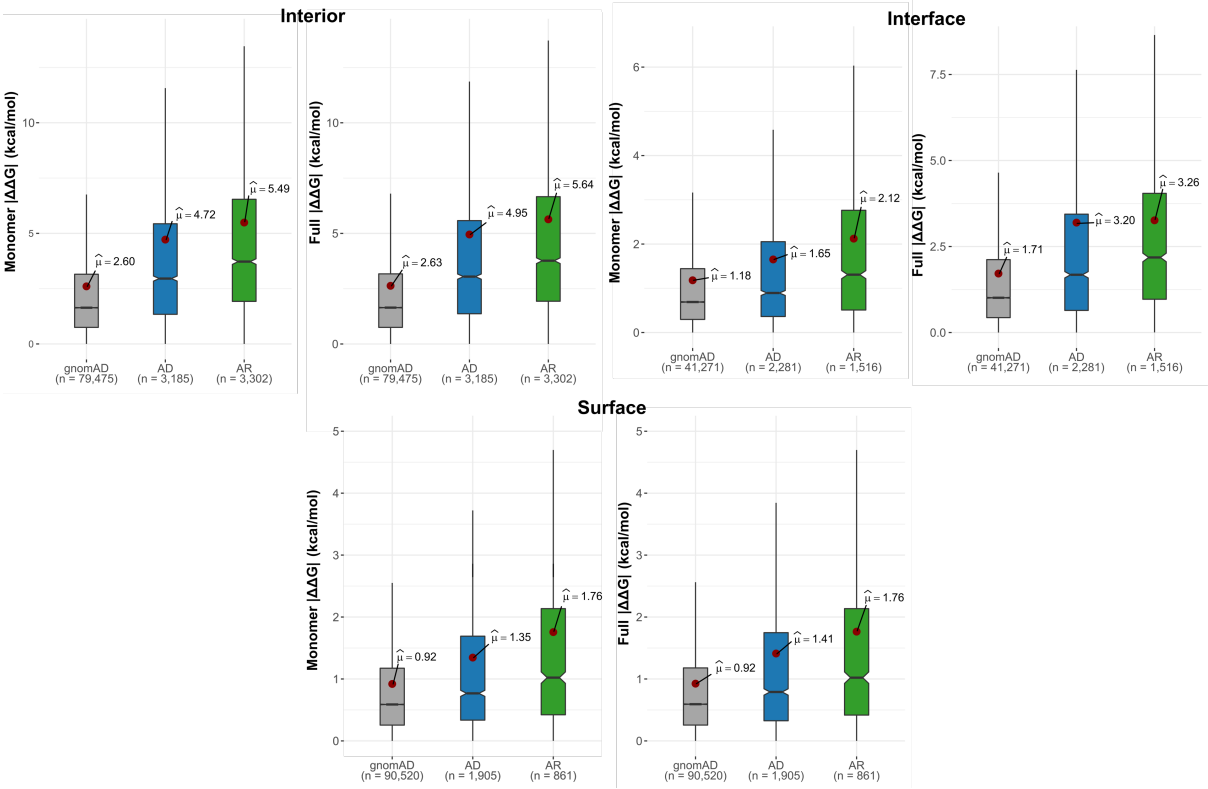

Figure S1. Differences in |ΔΔG| values between gnomAD, autosomal dominant and autosomal recessive mutations are observed across interior, interface and surface locations. All pairwise group comparisons showed significant differences ($P < 5.0 \times 10^{-6}$, Holm-corrected Dunn's test).

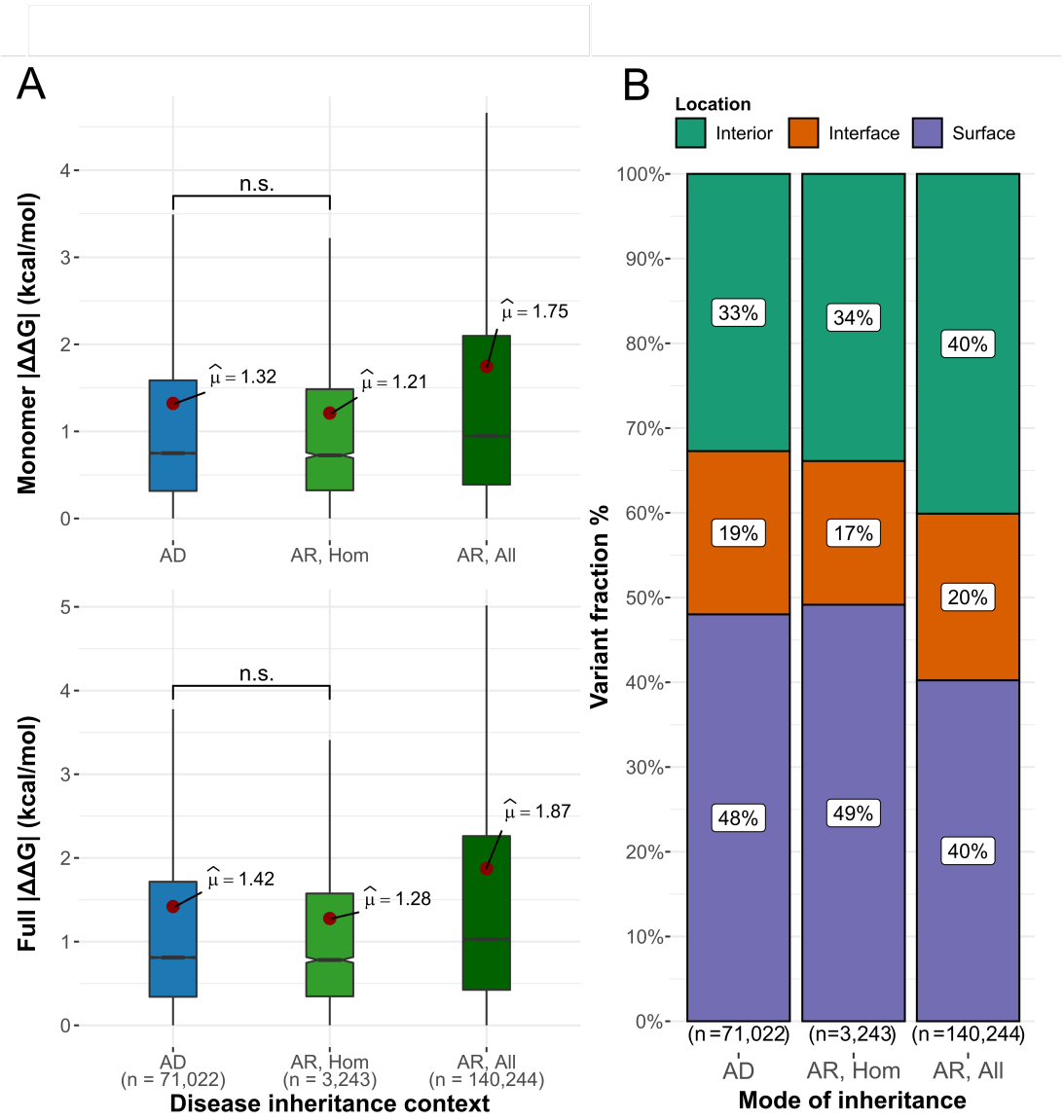

Figure S2. Accounting for variant zygosity reveals highly similar |ΔΔG| values and location distributions for gnomAD variants in autosomal dominant and autosomal recessive disease genes. **(A)** Stability perturbation differences between gnomAD variants of different inheritance contexts and zygosity. 'AR, All' considers all gnomAD variants from the AR genes, while 'AR, Hom' only includes those variants that have been observed in a homozygous state in gnomAD at least once. Pairwise group comparisons are significant unless specified ($P < 1.4 \times 10^{-34}$, Holm-corrected Dunn's test). **(B)** Proportions of gnomAD variants throughout spatial structure locations for gene variants characterised by different inheritance context groups.

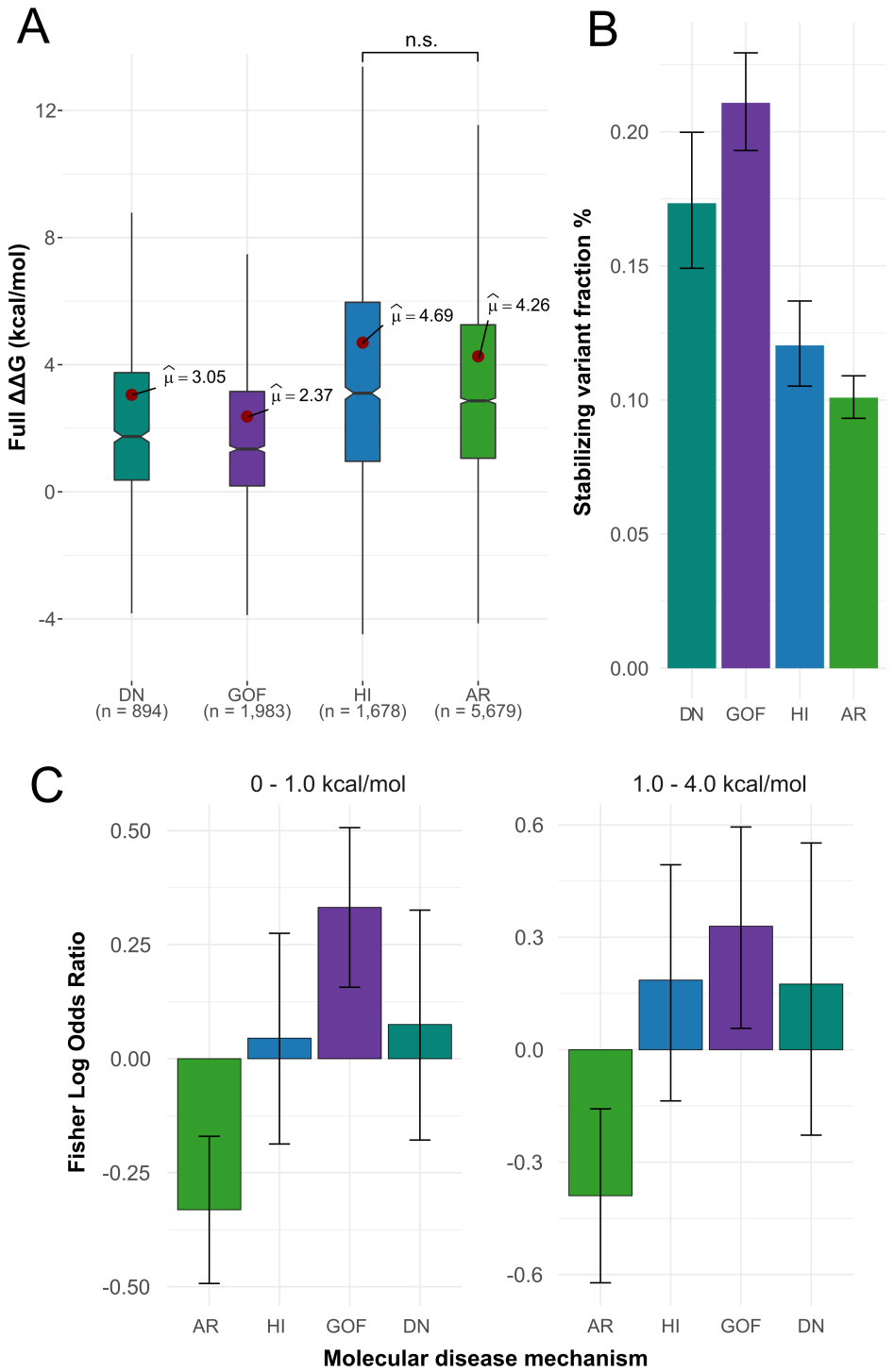

Figure S3. Stabilising mutations may be important for gain-of-function disease. (A) ΔΔG distributions (not absolute values) for pathogenic mutations associated with different molecular mechanisms. **(B)** Fraction of pathogenic mutations associated with different molecular mechanisms with negative (stabilising) ΔΔG values **(C)** Fisher log odds enrichments of stabilising mutations within each class from among all disease mutations. Error bars represent 95% confidence intervals.

Table S1. Optimal FoldX |ΔΔG| thresholds to distinguish between pathogenic and putatively benign mutations for genes associated with different molecular disease mechanisms.

Molecular disease mechanism group	Predictor	Optimal threshold (kcal/mol)	Specificity (low est.)	Specificity (median)	Specificity (high est.)	Sensitivity (low est.)	Sensitivity (median)	Sensitivity (high est.)
DN	FoldX Monomer |ΔΔG|	1.03	52.01	55.15	58.39	60.46	61.78	63.06
GOF	FoldX Monomer |ΔΔG|	0.98	56.68	58.80	60.97	57.88	58.72	59.58
HI	FoldX Monomer |ΔΔG|	1.38	62.22	64.66	66.92	72.76	73.63	74.54
AR (homozygous gnomAD)	FoldX Monomer |ΔΔG|	1.41	64.61	65.87	67.14	72.12	73.64	75.21
DN	FoldX Full |ΔΔG|	1.28	57.05	60.18	63.53	64.70	65.95	67.29
GOF	FoldX Full |ΔΔG|	1.16	55.87	57.89	60.01	61.29	62.06	62.84
HI	FoldX Full |ΔΔG|	1.59	66.03	68.30	70.50	74.81	75.65	76.52
AR (homozygous gnomAD)	FoldX Full |ΔΔG|	1.55	67.11	68.30	69.50	73.27	74.68	76.19

Table S2. Variant effect predictors used in this study.

Predictor	Feature-based classification	Disease dataset size	gnomAD dataset size	Source access	Reference
Blake-Cohen	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:BLAJ010101	(Blake & Cohen, 2001)[70]
BLOSUM62	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:HENS920102	(Henikoff & Henikoff, 1992)[71]
CADD	Metapredictor	12804	208495	https://cadd.gs.washington.edu/snv	(Kircher et al., 2014)[72]
Crooks-Brenner	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:CROG050101	(Crooks & Brenner, 2005)[73]
DANN	Metapredictor	12799	208111	dbNSFP database	(Quang et al., 2015)
DeepSequence	Amino acid sequence conservation	4572	18285	https://github.com/debbiemarkslab/DeepSequence	(Riesselman et al., 2018)[74]
DEOGEN2	Multi-feature	12093	202004	https://deogen2.mutaframe.com/	(Raimondi et al., 2017)[75]
Eigen	Metapredictor	12770	207393	dbNSFP database	(Ionita-Laza et al., 2016)[76]
FATHMM	Multi-feature	12078	196601	http://fathmm.biocompute.org.uk/inherited.html	(Shihab et al., 2013)[77]
Fathmm-MKL	Multi-feature	12799	208111	dbNSFP database	(Shihab et al., 2015)[78]
FathmmXF	Nucleotide-level prediction method	12710	207634	http://fathmm.biocompute.org.uk/fathmm-xf/	(Rogers et al., 2018)[79]
GenoCanyon	Nucleotide-level prediction method	12799	208111	dbNSFP database	(Lu et al., 2015)[80]
GERP++	Nucleotide-level prediction method	12799	208102	dbNSFP database	(Davydov et al., 2010)[81]
Grantham	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:GRAR740104	(Grantham, 1974)[82]
Lin et al.	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:LINK010101	(Lin et al., 2001)[83]
LRT	Nucleotide-level prediction method	12544	201244	dbNSFP database	(Chun & Fay, 2009)[84]
M-CAP	Metapredictor	12744	206051	http://bejerano.stanford.edu/mcap/	(Jagadeesh et al., 2016)[85]
MetaLR	Metapredictor	12770	207393	dbNSFP database	(Dong et al., 2015)[86]
MetaSVM	Metapredictor	12770	207393	dbNSFP database	(Dong et al., 2015)[86]
Miyata	Substitution matrix	13048	211134	https://www.genome.jp/entry/aaindex:MIYT790101	(Miyata et al., 1979)[87]
MPC	Multi-feature	8816	160998	dbNSFP database	(Samocha et al., 2017)[45]
MutationAssessor	Amino acid sequence conservation	11871	195872	dbNSFP database	(Reva et al., 2011)[88]

MutationTaster	Multi-feature	12758	205957	dbNSFP database	(Schwarz et al., 2014)[89]
MutPred	Multi-feature	11552	170552	dbNSFP database	(Pejaver et al., 2017)[90]
MVP	Metapredictor	12526	205824	dbNSFP database	(Qi et al., 2018)[91]
NetDiseaseSNP	Multi-feature	12789	205379	http://www.cbs.dtu.dk/services/NetDiseaseSNP/	(Johansen et al., 2013)[92]
phastCons	Nucleotide-level prediction method	12804	208495	dbNSFP database	(Siepel et al., 2005)[93]
phyloP	Nucleotide-level prediction method	12804	208495	http://papi.unipv.it/	(Pollard et al., 2010)[94]
PolyPhen2 HumDiv	Multi-feature	11769	198823	http://genetics.bwh.harvard.edu/pph2/	(Adzhubei et al., 2010)[95]
PolyPhen2 HumVar	Multi-feature	11769	198823	http://genetics.bwh.harvard.edu/pph2/	(Adzhubei et al., 2010)[95]
PonP2	Multi-feature	12030	195087	http://structure.bmc.lu.se/PON-P2/	(Niroula et al., 2015)
PrimateAI	Multi-feature	12636	206450	dbNSFP database	(Sundaram et al., 2018)[96]
PROVEAN	Amino acid sequence conservation	12009	197164	http://provean.jcvi.org/index.php	(Choi et al., 2012)[97]
REVEL	Metapredictor	12770	207393	https://sites.google.com/site/revelgenomics/	(Ioannidis et al., 2016)[98]
SIFT	Amino acid sequence conservation	12851	208985	https://sift.bii.a-star.edu.sg/www/code.html	(Sim et al., 2012)[99]
SIFT4G	Amino acid sequence conservation	12416	203352	dbNSFP database	(Vaser et al., 2016)[100]
SiPhy	Nucleotide-level prediction method	12797	208090	dbNSFP	(Garber et al., 2009)[101]
SNAP2	Multi-feature	13009	210582	https://www.rostlab.org/services/snap/	(Hecht et al., 2015)[102]
Sneath	Substitution matrix	13048	211134	(Sneath, 1966)	(Sneath, 1966)[103]
SuSPect	Multi-feature	12997	210252	http://www.sbg.bio.ic.ac.uk/suspect/about.html	(Yates et al., 2014)[104]
VEST4	Multi-feature	12631	206618	https://www.cravat.us/CRAVAT/	(Carter et al., 2013)[105]

References

1. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
2. Katsonis, P. *et al.* Single nucleotide variations : Biological impact and theoretical interpretation. **23**, 1650–1666 (2014).
3. Landrum, M. J. *et al.* ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, 980–985 (2014).
4. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
5. Miosge, L. A. *et al.* Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5189–E5198 (2015).
6. Livesey, B. J. & Marsh, J. A. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Mol. Syst. Biol.* **16**, 1–12 (2020).
7. Richards, S. *et al.* 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).
8. Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J. & Fowler, D. M. Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. *Cell Syst.* **6**, 116-124.e3 (2018).
9. Sanavia, T. *et al.* Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. *Comput. Struct. Biotechnol. J.* **18**, 1968–1979 (2020).
10. Heyn, P. *et al.* Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* **51**, 96–105 (2019).
11. Holt, R. J. *et al.* De Novo Missense Variants in FBXW11 Cause Diverse Developmental Phenotypes Including Brain, Eye, and Digit Anomalies. *Am. J. Hum. Genet.* **105**, 640–657 (2019).
12. Williamson, K. A. *et al.* Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA–protein interaction. *Genet. Med.* **22**, 598–609 (2020).
13. McEntagart, M. *et al.* A Restricted Repertoire of de Novo Mutations in ITPR1 Cause Gillespie Syndrome with Evidence for Dominant-Negative Effect. *Am. J. Hum. Genet.* **98**, 981–992 (2016).
14. Yue, P., Li, Z. & Moult, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* **353**, 459–473 (2005).
15. Glusman, G. *et al.* Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: A proposed framework. *Genome Med.* **9**, 1–10 (2017).
16. Bhattacharya, R., Rose, P. W., Burley, S. K. & Prlić, A. Impact of genetic variation on three dimensional structure and function of proteins. *PLoS ONE* **12**, 1–22 (2017).
17. Gerasimavicius, L., Liu, X. & Marsh, J. A. Identification of pathogenic missense mutations using protein stability predictors. *Sci. Rep.* **10**, 1–10 (2020).
18. Casadio, R., Vassura, M., Tiwari, S., Fariselli, P. & Luigi Martelli, P. Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. *Hum. Mutat.* **32**, 1161–1170 (2011).
19. Birolo, G. *et al.* Protein Stability Perturbation Contributes to the Loss of Function in Haploinsufficient Genes. *Front. Mol. Biosci.* **8**, 620793 (2021).
20. Herskowitz, I. dominant negative mutations CS-G-CS. *Nature* **329**, 1–4 (1987).
21. Bergendahl, L. T. *et al.* The role of protein complexes in human genetic disease. *Protein Sci.* **28**, 1400–1411 (2019).
22. Veitia, R. A., Caburet, S. & Birchler, J. A. Mechanisms of Mendelian dominance. *Clin. Genet.* **93**, 419–428 (2018).

23. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **10**, 980–980 (2003).
24. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019) doi:10.1101/531210.
25. Gudmundsson, S. *et al.* Variant interpretation using population databases: lessons from gnomAD. 56.
26. Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168–4169 (2019).
27. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
28. Sahni, N. *et al.* Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **161**, 647–660 (2015).
29. Jubb, H. C. *et al.* Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* **128**, 3–13 (2017).
30. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019).
31. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol.* **369**, 1318–1332 (2007).
32. Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J. Mol. Biol.* **315**, 771–786 (2002).
33. Livesey, B. J. & Marsh, J. A. The properties of human disease mutations at protein interfaces. (2021) doi:10.1101/2021.08.20.457107.
34. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
35. Seifi, M. & Walter, M. A. Accurate prediction of functional, structural, and stability changes in PITX2 mutations using in silico bioinformatics algorithms. *PLoS ONE* **13**, 1–23 (2018).
36. Potapov, V., Cohen, M. & Schreiber, G. Assessing computational methods for predicting protein stability upon mutation: Good on average but not in the details. *Protein Eng. Des. Sel.* **22**, 553–560 (2009).
37. Bromberg, Y. & Rost, B. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics* **10**, 1–9 (2009).
38. Witham, S., Takano, K., Schwartz, C. & Alexov, E. A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins Struct. Funct. Bioinforma.* **79**, 2444–2454 (2011).
39. Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R. & Alexov, E. Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* **425**, 3919–3936 (2013).
40. Nishi, H. *et al.* Cancer Missense Mutations Alter Binding Properties of Proteins and Their Interaction Networks. *PLoS ONE* **8**, (2013).
41. Stein, A., Fowler, D. M., Hartmann-Petersen, R. & Lindorff-Larsen, K. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem. Sci.* **44**, 575–588 (2019).
42. Grimm, D. G. *et al.* 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).
43. Ng, P. C. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
44. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018).
45. Samocha, K. E. *et al.* Regional missense constraint improves variant deleteriousness prediction. *bioRxiv* (2017) doi:10.1101/148353.

46. Lelieveld, S. H. *et al.* Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *Am. J. Hum. Genet.* **101**, 478–484 (2017).
47. Sivley, R. M., Dou, X., Meiler, J., Bush, W. S. & Capra, J. A. Comprehensive Analysis of Constraint on the Spatial Distribution of Missense Variants in Human Protein Structures. *Am. J. Hum. Genet.* **102**, 415–426 (2018).
48. Turner, T. N. *et al.* Proteins linked to autosomal dominant and autosomal recessive disorders harbor characteristic rare missense mutation distribution patterns. *Hum. Mol. Genet.* **24**, 5995–6002 (2015).
49. Stehr, H. *et al.* The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. *Mol. Cancer* **10**, 54 (2011).
50. Hijikata, A., Tsuji, T., Shionyu, M. & Shirai, T. Decoding disease-causing mechanisms of missense mutations from supramolecular structures. *Sci. Rep.* **7**, 1–8 (2017).
51. Veitia, R. A. Exploring the Molecular Etiology of Dominant-Negative Mutations. *Plant Cell Online* **19**, 3843–3851 (2007).
52. Chen, C.-H. *et al.* Novel and prevalent non-East Asian ALDH2 variants; Implications for global susceptibility to aldehydes' toxicity. *EBioMedicine* **55**, 102753 (2020).
53. De Baets, G., Van Doorn, L., Rousseau, F. & Schymkowitz, J. Increased Aggregation Is More Frequently Associated to Human Disease-Associated Mutations Than to Neutral Polymorphisms. *PLOS Comput. Biol.* **11**, e1004374 (2015).
54. Li, X. H. & Babu, M. M. Human Diseases from Gain-of-Function Mutations in Disordered Protein Regions. *Cell* **175**, 40–42 (2018).
55. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genet. Test. Mol. Biomark.* **14**, 533–537 (2010).
56. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
57. Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
58. Majithia, A. R. *et al.* Prospective functional classification of all possible missense variants in PPARG. *Nat. Genet.* **48**, 1570–1575 (2016).
59. Starita, L. M. *et al.* Variant Interpretation: Functional Assays to the Rescue. *Am. J. Hum. Genet.* **101**, 315–325 (2017).
60. Nielsen, S. V., Hartmann-Petersen, R., Stein, A. & Lindorff-Larsen, K. Multiplexed assays reveal effects of missense variants in MSH2 and cancer predisposition. *PLOS Genet.* **17**, e1009496 (2021).
61. Cagiada, M. *et al.* Understanding the Origins of Loss of Protein Function by Analyzing the Effects of Thousands of Variants on Activity and Abundance. *Mol. Biol. Evol.* **38**, 3235–3246 (2021).
62. Matreyek, K. A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
63. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **12**, 103 (2020).
64. Levy, E. D. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. *J. Mol. Biol.* **403**, 660–670 (2010).
65. Patil, I. Visualizations with statistical details: The 'ggstatsplot' approach. *J. Open Source Softw.* **6**, 3167 (2021).
66. Pierucci, A. F. CRAN - Package rmngb. http://cran.nexr.com/web/packages/rmngb/index.html.
67. Dorai-Raj, S. *binom: Binomial Confidence Intervals For Several Parameterizations*. (2014).
68. Turck, N. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **8**, 12–77 (2011).
69. Greiner, M., Pfeiffer, D. & Smith, R. D. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.* **45**, 23–41 (2000).

70. Blake, J. D. & Cohen, F. E. Pairwise sequence alignment below the twilight zone. *J. Mol. Biol.* **307**, 721–735 (2001).
71. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* **89**, (1992).
72. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
73. Crooks, G. E. & Brenner, S. E. An alternative model of amino acid replacement. *Bioinformatics* **21**, 975–980 (2005).
74. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **15**, 816–822 (2018).
75. Raimondi, D. *et al.* DEOGEN2: Prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. *Nucleic Acids Res.* **45**, W201–W206 (2017).
76. Ionita-Laza, I., Mccallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).
77. Shihab, H. A. *et al.* Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum. Mutat.* **34**, 57–65 (2013).
78. Shihab, H. A. *et al.* An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543 (2015).
79. Rogers, M. F. *et al.* FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511–513 (2018).
80. Lu, Q. *et al.* A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Sci. Rep.* **5**, 1–13 (2015).
81. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, (2010).
82. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
83. Lin, K., May, A. C. W. & Taylor, W. R. Amino acid substitution matrices from an artificial neural network model. *J. Comput. Biol.* **8**, 471–481 (2001).
84. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
85. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
86. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
87. Miyata, T., Miyazawa, S. & Yasunaga, T. Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12**, 219–236 (1979).
88. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* **39**, 37–43 (2011).
89. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
90. Pejaver, V. *et al.* Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **11**, (2020).
91. Qi, H. *et al.* MVP predicts the pathogenicity of missense variants by deep learning. *Nat. Commun.* **12**, (2021).
92. Johansen, M. B., Izarzugaza, J. M. G., Brunak, S., Petersen, T. N. & Gupta, R. Prediction of Disease Causing Non-Synonymous SNPs by the Artificial Neural Network Predictor NetDiseaseSNP. *PLoS ONE* **8**, (2013).
93. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
94. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

95. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
96. Sundaram, L. *et al.* Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* **50**, 1161–1170 (2018).
97. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **7**, (2012).
98. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
99. Sim, N. L. *et al.* SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, 452–457 (2012).
100. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
101. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, 54–62 (2009).
102. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genomics* **16**, (2015).
103. Sneath, P. H. A. Relations between chemical structure and biological activity in peptides. *J. Theor. Biol.* **12**, 157–195 (1966).
104. Yates, C. M., Filippis, I., Kelley, L. A. & Sternberg, M. J. E. SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J. Mol. Biol.* **426**, 2692–2701 (2014).
105. Carter, H., Douville, C., Stenson, P. D., Cooper, D. N. & Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14 Suppl 3**, (2013).