PERSPECTIVE

# Towards Increasing the Clinical Relevance of In Silico Methods to Predict Pathogenic Missense Variants

**David L. Masica[1], Rachel Karchin[1,2]** *

1 Department of Biomedical Engineering and The Institute for Computational Medicine, The Johns Hopkins University, Baltimore, Maryland, United States of America, 2 Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America

* karchin@jhu.edu

## Overview

As genetic sequencing throughput continues to accelerate, so does the accumulation of variants of unknown clinical significance. The great majority of these variants cause amino acid substitutions (cSNVs) in protein sequence. The need to interpret these variants continues to motivate development of better in silico bioinformatic methods. Despite the development of dozens of such methods over the past 15 years, clinically relevant prediction accuracy remains elusive. Here, we present some recent progress and shortcomings in the development of bioinformatics missense variant classifiers, and we argue for the increased use of endophenotypes. Endophenotypes are quantitative measurements that are correlated with phenotypes via shared genetic causes (e.g., enzyme catalytic activity, serum cholesterol or glucose level, volumetric lung capacity). In many cases, endophenotypes are more directly influenced by genetic variation, increasing their power to detect genotype-endophenotype associations relative to genotype-phenotype associations. The data required to train and benchmark bioinformatic methods to predict endophenotype from cSNVs and other variant types is increasingly available and could be made widely available by concerted community effort to enhance locus-specific and disease variant databases. We highlight some currently available data and present results from published bioinformatics studies that use endophenotypes.

## Introduction

The 21st century has thus far been marked by a heroic effort in genomic science and technology. If not yet upon us, the age of personalized genomic medicine appears imminent. Deriving medically relevant insight from these advances requires the ability to interpret the genetic variation and similarity observed in the population. One approach for interpreting genetic variation is the use of bioinformatics methods; simply put, these approaches are unified by a reliance on molecular/biological information (DNA, RNA, and protein sequence and annotation, protein structure, etc.) [1–4]. Many bioinformatics classifiers, primarily focused on amino acid substitution variants (cSNVs), have been and continue to be developed, typically achieving classification accuracies much better than random and thus supporting the use of molecular information [5–9]. These methods appear, however, to have reached a performance bottleneck; currently realized limits in performance all but forbid the consultation of bioinformatics cSNV classifiers in clinical settings. This bottleneck might be, in part, the result of simplifying

assumptions inherent to the prediction of qualitative, dichotomous, or categorical phenotypes from individual missense variants.

A complementary and beneficial strategy could be greater use of endophenotypes, quantitative measurements that are related to phenotypes via shared underlying genetics [10,11]. Endophenotypes can include phenomena at diverse biological scales; some examples include protein catalytic rate or melting temperature (stability), cell growth rate, and blood pressure. In this perspective we make a case for the increased use of endophenotypes, beginning with a brief overview of in silico bioinformatics methods for assessing phenotypic impact of cSNVs and the performance reported in recently published comparative studies. We compare the utility of endophenotypes and phenotypes in the context of *CFTR* cSNVs in cystic fibrosis, and of *LDLR* cSNV impact on cardiovascular-related diseases such as familial hypercholesterolemia. We also provide examples of bioinformatics methods that have been used to predict endophenotypes from cSNVs.

## Bioinformatics Classifiers of Phenotype: Background and Performance

Bioinformatics methods to predict the pathogenicity of cSNVs typically utilize gene/protein sequence, protein structure, annotation, or some combination of the three [1–4]. Table 1 shows 13 methods tested in up to five independent large-scale studies, one from each of the past five years [5–9]. Each of the five independent studies used large sets of putatively neutral and pathogenic variants, and three of these studies considered two such datasets; Olatubosun et al. used the Protein Mutant Database (PMD) twice, the second time using only a subset of variants defined as being reliably predicted by their own PON-P method. Criteria for inclusion in this table was that an independent group directly compared the method with other methods; for the fairest comparison, if an author included an assessment of their own method, that assessment is not shown in Table 1. Each of the five studies included the statistical sensitivity, specificity, and accuracy, or included the relevant contingency table; therefore, we present these three performance metrics for each of the tested method-dataset combinations (Table 1). While these performance metrics alone cannot provide a truly comprehensive estimate of classifier performance, they do facilitate a reasonable comparison of methods and a general assessment of method utility.

SIFT [12] and PolyPhen-2 [13] are two of the most commonly cited and used methods to predict disease phenotype from cSNVs. Across the five independent tests in Table 1, accuracies achieved by both methods typically ranged from the mid-60s to mid-70s. Another important result for these two methods is that the trade off between sensitivity and specificity varies significantly among benchmarks. For instance, SIFT and PolyPhen-2 both had high sensitivities in the Olatubosun et al. benchmark, and SIFT had disproportionately high specificity in both of the Rapakoulia et al. benchmarks. A similar variability in sensitivity and specificity across benchmarks, as well as accuracy range, was reported for SNAP [14]. PhD-SNP [15] was tested in four of the five studies, achieving accuracies comparable to most methods presented in Table 1, with a reasonable balance between sensitivity and specificity in all tests. Panther [16] was also tested in four of the studies, achieving accuracies similar to SIFT, PolyPhen-2, PhD-SNP, and SNAP, but typically realizing better gains in specificity than sensitivity. SNPs&GO [17] was also skewed toward specificity, realizing a slightly higher accuracy than SIFT, PolyPhen-2, Panther, and SNAP. MutPred [18] and Mutation Assessor [19] were tested in three and two of the studies, respectively; these methods achieved relatively high accuracies and balanced sensitivities and specificities (Table 1).

**Table 1. Five years of independent testing of cSNV variant classifiers.**

| | Thusberg et al. 2011 | Olatubosun et al. 2012 | Shihab et al. 2013 | Rapakoulia et al. 2014 | Dong et al. 2015 |
|---|---|---|---|---|---|
| | + 19,335 SwissProt & LSDBs<br>- 21,170 dbSNP | + 902 PMD<br>- 858 PMD | + ~23,000 SwissVar<br>- ~34,000 SwissVar | + 8,871 SwissProt<br>- 24,342 dbSNP | + 120 Curated by author<br>- 124 CHARGE seq. proj. |
| | + 3,594 LSDBs<br>- 3,594 dbSNP | Subset reliable at 0.95 | | + HumVar 22,196<br>- HumVar 21,119 | + 6,279 VariBench II<br>- 13,240 VariBench II |
| SIFT | 68%, 62% (65%)<br>65%, 62% (64%) | 83%, 41% (62%)<br>83%, 48% (66%) | 67%, 82% (74%) | 38%, 81% (69%)<br>37%, 90% (63%) | 68%, 75% (71%)<br>75%, 69% (71%) |
| PolyPhen-2 | 73%, 70% (71%)<br>62%, 69% (66%) | 84%, 40% (62%)<br>81%, 50% (66%) | 86%, 61% (73%) | 85%, 69% (73%)<br>87%, 72% (80%) | 92%, 59% (78%)<br>88%, 49% (63%) |
| Panther | 77%, 76% (76%)<br>52%, 75% (64%) | NA<br>NA | 62%, 75% (68%) | 31%, 90% (74%)<br>31%, 89% (59%) | 38%, 94% (66%)<br>NA |
| PhD-SNP | 63%, 79% (71%)<br>40%, 79% (60%) | 67%, 61% (64%)<br>69%, 74% (72%) | 66%, 74% (70%)<br>77%, 79% (78%) | NA<br>NA | 70%, 83%, (77%)<br>NA |
| SNAP | 88%, 56% (72%)<br>74%, 56% (65%) | 73%, 54% (63%)<br>72%, 65% (68%) | NA | NA<br>NA | 53%, 70% (61%)<br>NA |
| SNPs&GO | 71%, 92% (82%)<br>73%, 92% (82%) | NA<br>NA | 76%, 89% (82%) | NA<br>NA | 55%, 94% (75%)<br>NA |
| MutPred | 85%, 78% (81%)<br>71%, 77% (74%) | NA<br>NA | 90%, 90% (90%) | NA<br>NA | 74%, 81% (77%)<br>NA |
| Mutation Assessor | NA<br>NA | NA<br>NA | NA | 73%, 88% (84%)<br>77%, 84% (80%) | 70%, 80% (74%)<br>73%, 74% (74%) |
| nsSNAPAnalyzer | 61%, 58% (60%)<br>58%, 56% (57%) | NA<br>NA | NA | NA<br>NA | NA<br>NA |
| MutationTaster | NA<br>NA | NA<br>NA | NA | NA<br>NA | 94%, 74% (86%)<br>91%, 49% (64%) |
| FATHMM | NA<br>NA | NA<br>NA | NA | NA<br>NA | 55%, 91% (75%)<br>86%, 82% (83%) |
| CONDEL | NA<br>NA | NA<br>NA | NA | NA<br>NA | 71%, 73% (72%)<br>71%, 71% (71%) |
| CADD | NA<br>NA | NA<br>NA | NA | NA<br>NA | 79%, 74% (77%)<br>75%, 65% (69%) |

Selected studies, one from each of the past five years, comparing multiple bioinformatic variant classifiers on large datasets of putatively pathogenic and benign variants. To avoid any potential bias, tests of an author's own method, if present, were excluded. For each entry, results are given as *sensitivity*, *specificity* (*accuracy*). Variants datasets used as the pathogenic or disease-causing class are indicated by a plus sign (+) and the neutral or benign class with a negative sign (-). Entries with multiple lines indicate that two unique datasets were used for benchmarking. Both datasets used in Olatubosun et al. 2012 were the same, except that variants in the second set were filtered to meet a 95% prediction confidence using their own PON-P method. PMD is the Protein Mutant Database; NA indicates the method was not tested in the corresponding benchmark.

## Origins and Implications of Variant Classifier Performance

Results for some of the method-benchmark combinations presented in Table 1 are promising. However, for most methods—in particular, those subject to the greatest scrutiny—reported performance is inconsistent across benchmarks, including highly unbalanced sensitivity and specificity. Bioinformatics methods are not currently recommended for medical decisions that require variant interpretation [20]. We believe that these perceived methodological shortcomings might, in part, arise from assumptions inherent in the prediction of qualitative biological phenomena from individual genetic variants. Genome-wide association studies (GWAS) have thus far indicated that the vast majority of genetic variation in complex diseases likely impact gene regulation and have low effect size [21,22]. Similarly, pedigree studies have recovered relatively few high-penetrance genes/variants [23,24]. And recent whole-genome and targeted

sequencing efforts are revealing that healthy individuals, on average, harbor tens of so-called "disease alleles," including many in the homozygous state [24,25]. Thus, bioinformatics predictors are challenged with predicting disease from individual variants, yet individual variants appear to rarely carry significant disease liability. This is particularly true of cSNVs, which on average have a more subtle effect on disease than truncating variants [26]. And because "disease alleles" used for classifier training and benchmarking may be relatively common among healthy individuals, expectations about the relevance of subsequent predictions should be tempered.

Fig 1 illustrates some factors that may confound bioinformatic cSNV disease phenotype prediction, using as an example the impact of *LDLR* (low-density lipoprotein receptor) cSNVs on cardiovascular-related diseases. LDLR is a protein localized to the plasma cell membrane that enables endocytosis of low-density lipids into cells and controls plasma concentration of low-density lipid cholesterol (LDL-c). Changes in LDLR protein stability and activity, which are most closely related to cSNVs in LDLR, are likely to be easiest to predict (Fig 1A). The further downstream an effect is from LDLR function (a diagnosis of coronary artery disease, for instance), the more likely additional factors are to influence the effect. These influences can include other endogenous inputs (e.g., genetic and epigenetic factors) as well as exogenous factors such as lifestyle and environment (Fig 1B). Thus, a cSNV that perturbs LDLR stability and/or activity will not necessarily cause disease. This is of practical importance because bioinformatics predictors rely heavily on features of protein sequence, structure, and function, but are typically tasked with predicting disease. Further complicating is the fact that diagnosis of a particular disease phenotype can result from different diagnostic tests, combination of tests, and varied interpretation of test results (Fig 1B). In contrast, measured effects in cellular assays and diagnostics such as high serum LDL cholesterol are closer to the LDLR protein stability and activity required for normal function, and they are not in theory confounded by subjective factors implicit to associating cSNVs with disease diagnosis.

Fig 1C shows 12 LDLR cSNVs that were common to multiple studies with published variant-specific mean patient LDL-c level, cellular assays of LDL uptake, bioinformatics prediction with SIFT and Polyphen-2, and curation in the Human Gene Mutation Database (HGMD) [27–29]. Supporting the idea that endophenotypic measurements may be more consistent than disease phenotype categories, the clinical LDL-c concentration and experimental LDL-uptake assays from two different laboratories show the expected negative correlation (Pearson correlation = −0.69; $p$-value = 0.013). (Normal LDLR uptake of LDL lowers plasma LDL-c levels.) However, the assignment of each cSNV to a disease phenotypic category differs depending on the selected endophenotype studied and is often in disagreement with HGMD. As in the independent benchmarking studies (Table 1), the accuracy and specificity/sensitivity trade-off of the bioinformatics classifiers depends on which categorization is considered to be the gold standard.

Using clinical diagnostics, experimental assays, and biomedical literature to derive gold-standard mutation databases is reasonable and commonplace. But, given that most variants, and cSNVs in particular, are low effect and incompletely penetrant, the disagreement among these potential gold standards should be unsurprising. Therefore, it becomes unclear how and to what extent the disappointing performance of bioinformatics methods should be interpreted, given that many cSNVs could reasonably be placed in multiple phenotypic categories. In contrast, the continuous-valued cellular LDL-uptake and serum LDL-c measurements rely only on accurate determination, rather than varied and arbitrary thresholds for classification. These types of endophenotypic measurements represent a practical and useful target for prediction and help circumvent some potentially unreliable presuppositions currently associated with bioinformatics cSNV prediction.
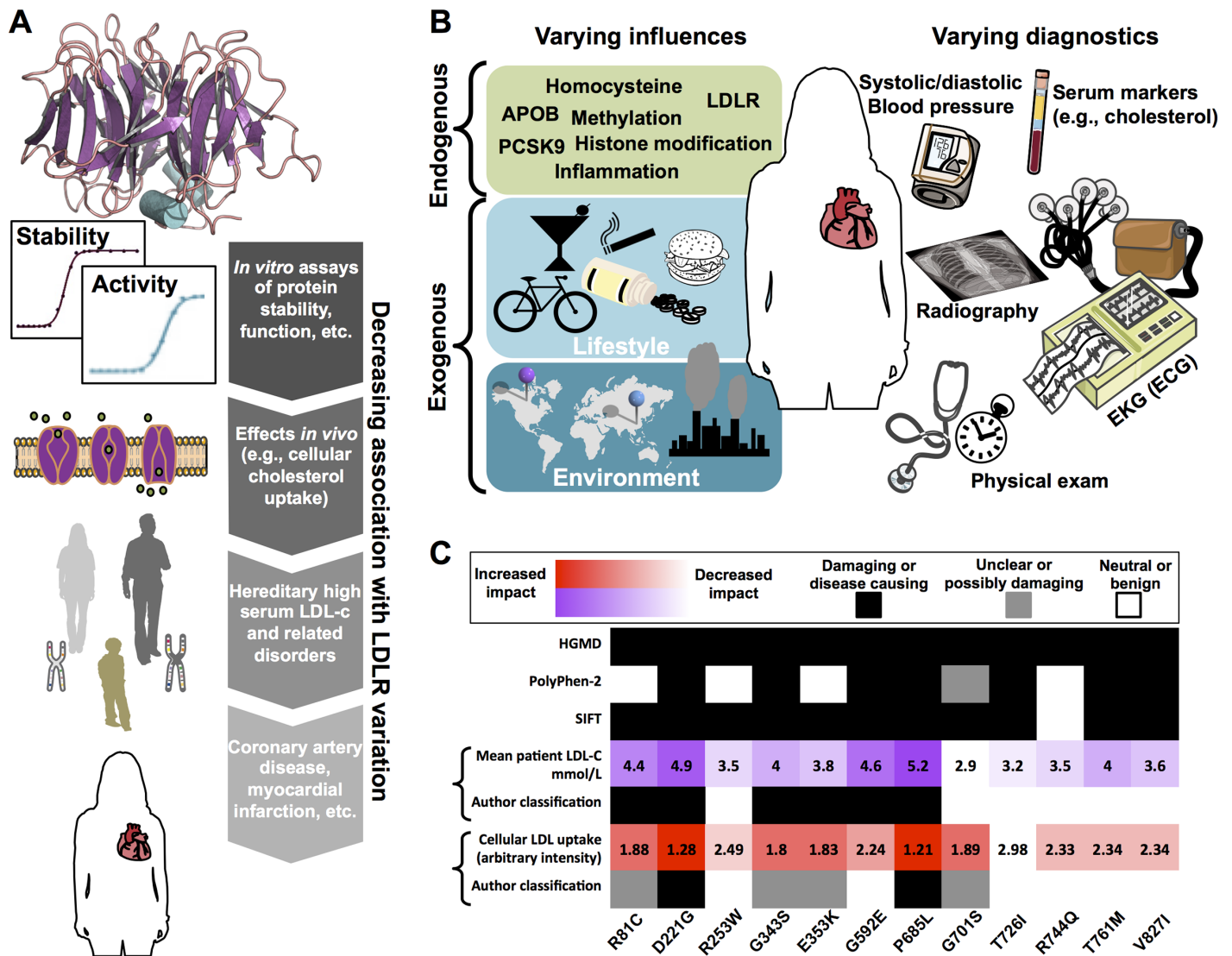
**Fig 1. How context can influence the impact and inferred impact of LDLR variation on different experimental and clinical parameters.** LDLR variation is likely to have the largest observable and reproducible impact on parameters most directly influenced by the protein (*A*). For instance, if a variant produces effects downstream from the protein, then protein structure and function are likely perturbed. Further downstream, cellular LDL uptake could be modulated, which can increase risk for familial hypercholesterolemia, which might appreciably increase risk for heart disease or heart attack. Liability for complex cardiovascular diseases is influenced by many endogenous and exogenous factors other than LDLR mutation (*B*). Furthermore, diagnoses can result from a varied combination of clinical and laboratory diagnostics, which can result in differential or conflicting diagnoses (*B*). In *C*, cellular studies, pedigree studies, a disease mutation database, and popular bioinformatics methods are used to classify LDLR variants as disease causing or benign. On the heatmap, black and white indicate a classification of disease causing and benign, respectively, for different classification methods (gray indicates an intermediate or unclear classification). Mean patient LDL-c concentration from pedigree studies (purple) and cellular LDL uptake (red) shown with darker colors indicating more severe impact (numbers indicate published values).

doi:10.1371/journal.pcbi.1004725.g001

## Endophenotypes: An Alternative and Complementary Framework

The term "endophenotype" was coined in 1966 to distinguish the observable, external phenotype (exophenotype) from internal or microscopic traits [30]. In 1972, Gottesman and Shields reintroduced the term in the context of schizophrenia to describe internal phenotypes discoverable by biochemical assays or microscopic examination [31]. Used infrequently over the next several decades, the word "endophenotype" experienced quite a renaissance after the publication

of a 2003 review article by Gottesman and Gould [11]. Endophenotypes are most often explicitly used in the context of psychiatric disorders such as schizophrenia or bipolar disorder, but the endophenotype concept has been applied in the context of many diseases, including obesity [32], diabetes [33], osteoporosis [34], heart disease [27–29,35], hypertension [36], phenylketonuria [37], and cystic fibrosis [38–40]. Requirements of heritability and co-segregation have been suggested in order for a quantitative trait to be considered a true endophenotype [10,11,41]. For this perspective, we use a broad definition of endophenotype; in short: endophenotypes are quantitative traits that are associated with qualitative traits (phenotypes) via shared genetic influences. Importantly, endophenotypes include the quantitative risk factors that are often used to diagnose and define disease (e.g., serum metabolite concentrations, blood pressure, sweat chloride), as well as molecular-scale phenomena such protein stability or catalytic rates.

By this definition, we believe that there are considerable benefits to bioinformatic approaches for predicting the genotype-endophenotype relationship, relative to that of the genotype-phenotype relationship: (1) Endophenotypes are closer to the level of gene action and protein function than the associated phenotypes, increasing the effect size and power to detect variant-endophenotype associations relative to that of the variant-phenotype associations. An example of this benefit is depicted in Fig 1A, where a cSNV in the LDLR protein is expected to have a more measureable effect on cellular LDL uptake than it would on the dichotomous prediction of having a heart attack or not. (2) By virtue of being qualitative, phenotypes rely on subjective and often arbitrary definitions. Although quantifying phenotypic descriptions is an active area of informatics research [42], the exact defining characteristics of a phenotype can change over time and be subject to disagreement among experts. Conversely, by virtue of being quantitative, endophenotypes should, in principle, rely only on accurate measurements. (3) Endophenotypes facilitate the ranking of biological states (e.g., disease severity) within the otherwise arbitrarily defined phenotypic categories. (4) The reliance on objective measurements rather than subjective definitions, along with the disposing of arbitrary thresholds for partitioning phenotypic categories, reduces data contamination and in turn benefits algorithmic training and benchmarking. (5) Endophenotypes can describe both severity and molecular mechanism with higher resolution than can phenotypes (Figs 1C and 2).

## The Benefit of Using Endophenotypes: An Illustrative Example

Endophenotypes can provide information that supplements phenotypic categories and increases their clinical utility, by pointing to specific disease severity and mechanism associated with a variant. This utility is illustrated by three cystic fibrosis transmembrane conductance regulator (*CFTR*) cSNVs shown in Fig 2, each having a distinct, clinically defined impact on cystic fibrosis disease liability (benign, indeterminate, and disease causing) [40].

The first endophenotype shown in Fig 2 is the continuous-valued clinical diagnostic of patient sweat chloride, which increases across the three phenotypic categories. The "sweat test" is considered the gold standard for diagnosing cystic fibrosis. Healthy individuals have sweat chloride concentrations of less than 30 to 40 mmol/L and a test reporting 60 mmol/L or greater most often results in cystic fibrosis diagnosis. As expected, the mean sweat chloride of patients harboring the benign cSNV S1235R is lower than that of patients harboring the indeterminate cSNV D614G, and is highest in patients with the disease-causing mutation G551D. The second endophenotype measures chloride conductance by in vivo cellular assays; transport of chloride ions through the plasma membrane of epithelial cells is a major function of CFTR. Defects in chloride conductance result in the mucus build-up associated with cystic fibrosis. Again as expected, chloride conductance negatively correlates with increasing disease severity. It is highest in patients with the benign cSNV, substantially lower in those with the indeterminate
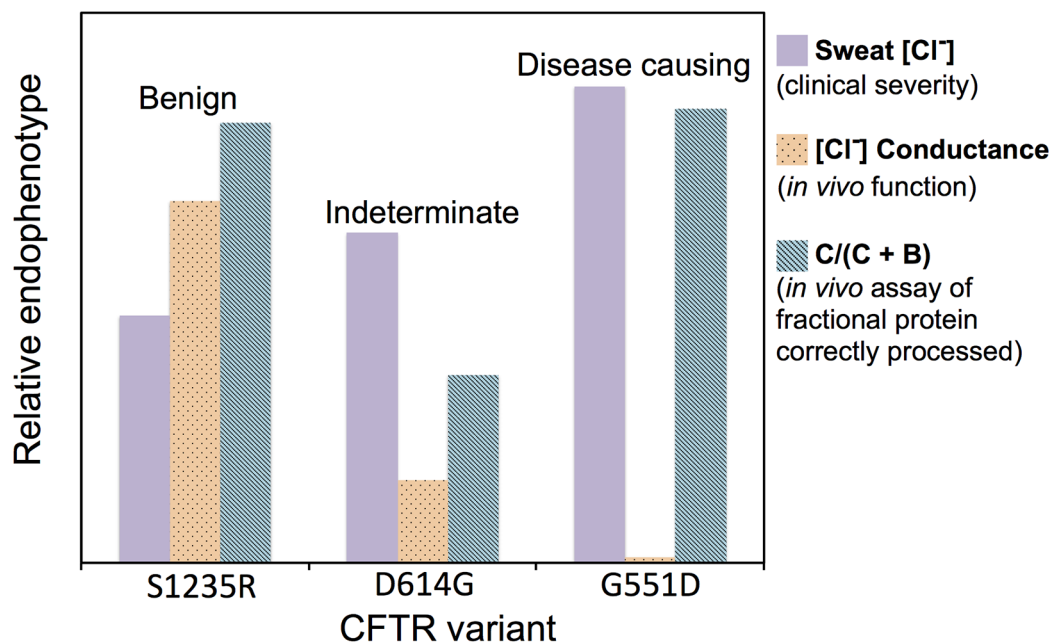
**Fig 2. Some advantages of considering endophenotypes, relative to phenotypes, illustrated using three CFTR variants.** Mean sweat chloride from individuals harboring the three variants (S1235R, D614G, and G551D), and results from two distinct in vivo experiments performed in cells expressing the variants. Increasing sweat chloride is associated with increasing disease severity, whereas in the two in vivo assays decreasing values correspond to decreasing protein function or abundance. Endophenotypes were scaled for purposes of presenting on a single chart, such that three sweat chlorides could be compared with one another, the three chloride conductance measurements could be compared with one another, etc.

cSNV, and undetectable in those with disease-causing G551D. Lastly, there is a different and important trend for the third endophenotype, in vivo measurements of CFTR C-band B-band ratio or $C/(C + B)$, which measures the fraction of CFTR protein that is fully processed (glycosylated) and trafficked to the cell surface. Correct processing and trafficking is necessary but not sufficient for normal CFTR function. As expected, CFTR is correctly processed and trafficked for the benign cSNV, and for the indeterminate cSNV the fraction of correctly processed protein decreases. But surprisingly, for the disease-causing G551D mutation, the fraction of correctly processed protein is approximately equal to that found with the benign variant; it is this differential impact of G551D—benign with respect to post-translational processing and damaging with respect to proper chloride transport function—that facilitates the efficacy of the landmark, G551D-specific, cystic fibrosis drug Ivacaftor [43]. Importantly, G551D-mutant CFTR protein is processed and trafficked to the epithelial cell surface, but once there it exhibits decreased chloride conductance. Ivacaftor potentiates cells harboring the CFTR G551D mutation, restoring chloride conductance. Following clinical trials, Ivacaftor was approved in 2014 to treat patients harboring several other CFTR mutations characterized by high C-band B-band ratio and low chloride conductance [40,44].

Previously, it has been proposed that multiple phenotypic categories, spanning the range from the most benign to most pathogenic variants, might alleviate problems with potential subjectivity and over-simplification of disease/benign or disease/indeterminate/benign classifications [45]. Indeed, the American College of Medical Genetics (ACMG) has recently published guidelines that include a five-category standard for clinical variant interpretation in genes that

cause Mendelian disorders [20]. The guidelines emphasize the limited clinical utility of the current generation of in silico bioinformatic prediction methods, in particular citing low specificity.

Bioinformatics methods designed to predict endophenotypes might be able to achieve greater accuracy and reliability than those designed to predict phenotypic categories. The assessment of such methods is not confounded by subjective choices about the number of phenotypic categories or the assignment of a variant to the correct category. In silico interpretation of a variant in terms of one or more endophenotypes may capture clinically important differences between variants that are placed in the same phenotypic category. For example, both the CFTR G551D mutation described above and CFTR N1303K are pathogenic according to ACMG standards, but the two mutations have different endophenotypic patterns. Unlike G551D, the N1303K mutation has both low C-band B-band ratio and low chloride conductance [40], indicating that the mechanism of CFTR dysfunction is different in N1303K. These differences are relevant to clinical decision-making, since Ivacaftor is indicated for G551D, while the newer drug Lumacaftor may be effective for mutations that impact post-translational processing of CFTR [44].

The ability to visualize variants in a multidimensional landscape of several endophenotypes could be valuable for clinicians. Fig 3 shows a hypothetical landscape of cystic fibrosis severity along three orthogonal coordinates: post-translational process of CFTR, in vivo chloride conductance, and sweat chloride levels. Each cSNV can be represented as a point in the coordinate system, enabling clinical assessment of the relationship between the cSNV, disease severity, and multiple measures of disease mechanism.

## Bioinformatics Prediction of Endophenotypes

The output of most bioinformatic cSNV classifiers is a raw, continuous-valued score, which is transformed into the assignment of each cSNV to one of two or more categories. For instance, SIFT returns the probability that a protein will tolerate a cSNV, while PolyPhen-2 returns the probability that a cSNV is protein damaging. Although these methods were not developed to predict endophenotypes, their continuous-valued outputs could be used as informative endophenotypic correlates. This insight was utilized by Wettstein et al. to predict phenylalanine hydroxylase (PAH) activity and three phenylketonuria (PKU)-related endophenotypes as a function of PAH cSNVs [37]. In that study, the authors scored up to 834 PAH cSNVs using the SIFT, Polyphen-2, FoldX [46], and SNPs3D [47] packages. In the case of PAH activity, the authors found statistically significant correlation between FoldX score and PAH enzymatic activity, as well as for SNPs3D and PAH activity; neither SIFT or PolyPhen-2 scores were correlated with PAH activity. Similarly, the authors compared scores from the four methods with three PKU-related disorders (PKU, mild PKU, and mild hyperphenylalaninemia), and found significant association between continuous-valued FoldX, SNPs3D, and PolyPhen-2 scores and the three disease subtypes; SIFT scores and disease subtype were not significantly associated. This work shows that existing cSNV classifiers can be repurposed for predicting endophenotypic severity, as well as recovering categorical phenotype without necessarily requiring the use of arbitrary thresholds for partitioning scores.

The above-cited work of Wennstein et al. demonstrates the potential to repurpose existing cSNV classifiers; however, these existing methods are limited because they are agnostic to the endophenotype being predicted. This limitation is important, because different variants in the same gene can affect disease via distinct mechanisms (Fig 2), or be causal of different diseases entirely (e.g., *NF1* mutation can drive cancer or cause neurofibromatosis). We hypothesize that detecting the subtle biological underpinnings that converge to influence a particular mutation-
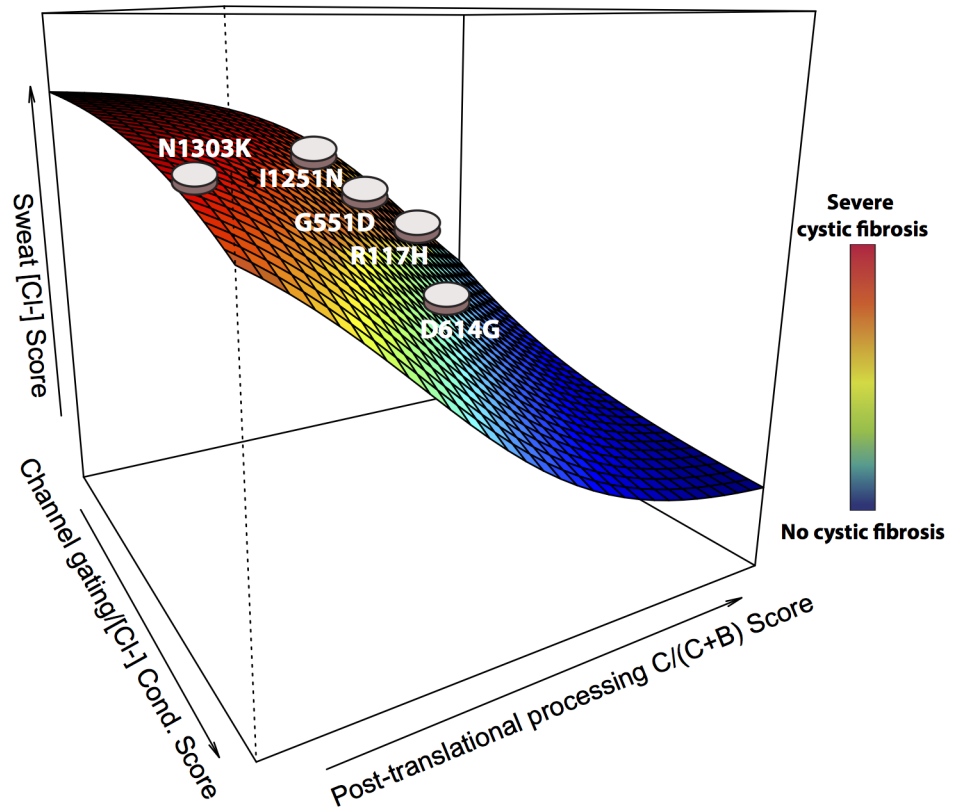
**Fig 3. Hypothetical visualization of a multidimensional endophenotypic landscape for cystic fibrosis.**
Each cSNV can be represented as a point in a three-dimensional space of three endophenotypic scores
relevant to cystic fibrosis disease: post-translational processing (glycosylation) and trafficking of the CFTR
protein to the epithelial cell plasma membrane, an in vivo cellular assay of chloride conductance that
measures channel gating, and chloride concentration in a diagnostic sweat test. Each point on the landscape
can be interpreted with respect to disease severity, shown in the color bar to the right of the landscape.

dependent endophenotype will benefit from classifiers that are trained to predict specific endo-
phenotypes, rather than classifiers that are nonspecific or agnostic.

We have recently developed an endophenotype prediction algorithm that trains endophe-
notype-specific cSNV classifiers [39]. The classifiers are, in part, a multiple-sequence alignment
(MSA), the gene composition of which is optimized by iteratively maximizing the coefficient of
determination (R-squared of regressing continuous-valued variables) between an internal
score function and the cSNV-specific endophenotypes from the training set [39,48]. The score
function considers amino acid conservation and the conservation of amino acid biophysical/
biochemical properties, derived from the MSA. The score function can optionally consider 3-D
structural data, as well. We refer to a classifier whose gene composition is optimized to predict
an endophenotype as an endoPhenotype-Optimized Sequence Ensemble (ePOSE), and hence
we call the method the ePOSE algorithm.

Fig 4 shows results from predicting three cystic fibrosis-related endophenotypes from 20
CFTR cSNVs (20 data points on each panel in Fig 4). For each of the three endophenotypes,
individually, the ePOSE algorithm trained using 19 of 20 CFTR cSNVs, and prediction was
made on the remaining cSNV; this process was repeated for each cSNV (leave-one-out cross-
validation). Predictions were typically well correlated with the endophenotype being predicted,
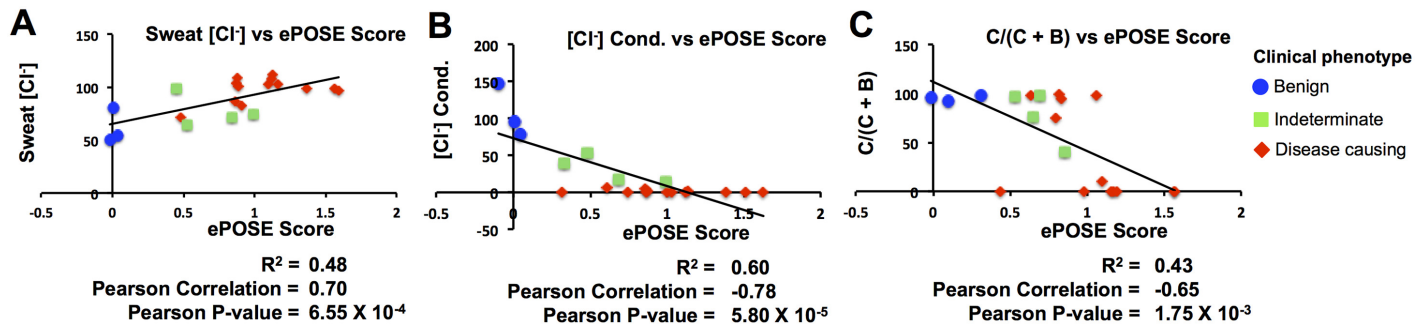
**Fig 4. Correlation of ePOSE score with three individual endophenotypes.** Measured endophenotype versus predicted impact (*ePOSE Score*) for 20 *CFTR* variants using classifiers trained with (A) sweat chloride, (B) chloride conductance, or (C) fraction of correctly processed CFTR protein. Each plot is the result of 20 leave-one-out cross-validation calculations (i.e., one data point for each of the 20 variants). Blue circles, green squares, and red diamonds denote benign, indeterminate, and disease-causing annotated phenotype, respectively, for each of the 20 variants. Note: increasing sweat chloride is associated with increasing disease severity, whereas for the two in vivo assays, decreasing values correspond to decreasing protein function or processing.

doi:10.1371/journal.pcbi.1004725.g004

including reasonable separation of three clinically defined phenotypes associated with each cSNV (denoted by color and shape in Fig 4).

In Fig 5, ePOSE predicts differential mechanisms associated with disease severity, including predictions for a validation set of three additional cSNVs, for which experimental and clinical data was collected prospectively. In contrast to Figs 3 and 5 is not hypothetical and shows actual ePOSE scores for each of three endophenotypes. The ePOSE algorithm accurately
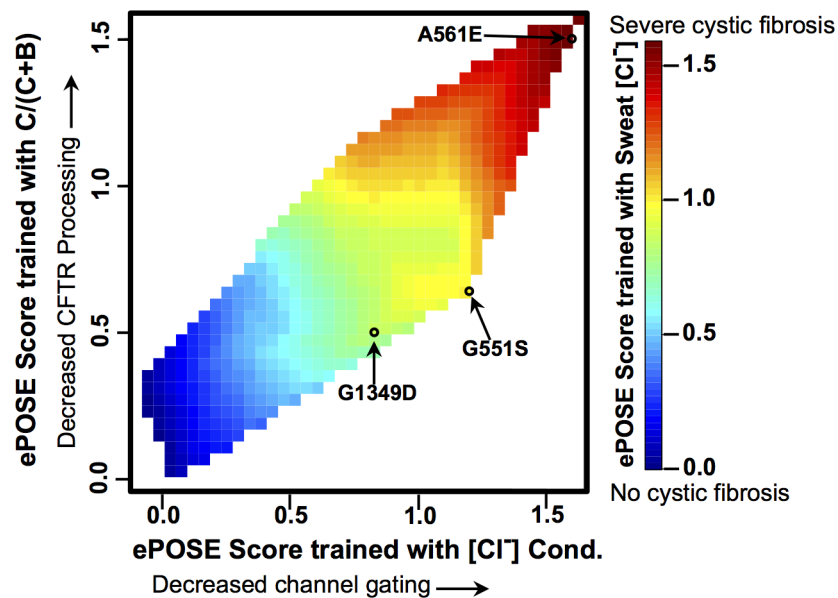


**Fig 5. Interpolation plot of predicted endophenotypes resulting from the separate leave-one-out cross-validation calculations shown in Fig 4.** ePOSE score for the 20 *CFTR* variants from Fig 4 plotted and interpolated (color shows ePOSE scores resulting from training with sweat chloride data). Using the resulting classifiers, each endophenotype was predicted for three additional variants (G551S, A561E, and G1349D) and subsequently validated. A561E was accurately predicted to affect disease via drastically reduced CFTR processing and channel gating. G551S was accurately predicted to affect cystic fibrosis primarily via channel gating.

doi:10.1371/journal.pcbi.1004725.g005

predicted that a significant fraction of the G551S cSNV would be processed and trafficked to the cell surface, but that chloride conductance would be significantly attenuated in cells expressing this cSNV. As described above, this same observation led to the development of Ivacaftor, a drug initially approved to target another G551 variant, G551D. Indeed, Ivacaftor has some efficacy for potentiating cells expressing the G551S cSNV as well [49].

Many of the existing in silico bioinformatic cSNV classifiers, originally designed to predict disease phenotypes, could be adapted to predict endophenotypes. Such efforts will require continued community-wide collection of data for algorithmic training and for independent benchmarking efforts. Locus-specific databases (LSDBs) already contain variant-specific endophenotypic information. Table 2 shows examples of training data currently available for endophenotypes associated with cystic fibrosis (CFTR), Li-Fraumeni and hereditary cancers (TP53) [50], phenylketonuria (PAH) [51], hypercholesterolemia and cardiovascular disease susceptibility (LDLR) [52], hereditary breast cancer (BRCA2) [53], and hyperhomocysteinemia (CBS) [54]. The CFTR2 database [40], which contains variants and endophenotypic data from ~40,000 cystic fibrosis patients, illustrates the potential of our suggested approach. For each patient, a reported genotype and up to six endophenotypes is provided. A mean value (and standard error) for an endophenotype of interest can be estimated, using reported values from patients with the same genotype (e.g., ~1,400 patients have one copy of the G551D allele and their mean sweat chloride is 104). Mean sweat chloride estimation is currently possible for ~250 unique variants in CFTR2, if a minimum of five patients with measured sweat chloride and sharing the identical allele is required [39].

**Table 2. Six disease-associated genes with sources of variant-specific endophenotypic data.**

| Protein or gene (number of amino acids) | Endophenotype (number of unique mutations) | Disease phenotype(s) | References and database URLs |
|---|---|---|---|
| CFTR (1,480) | Sweat [Cl-] (>250) | Cystic fibrosis | [40]; cftr2.org |
| | FEV%1 (>250) | | |
| | P.a. Infection (>250) | | |
| | Pancreas function (>250) | | |
| | C/ (C + B) (59) | | |
| | [Cl-] Conductance (59) | | |
| TP53 (393) | *WAF1* transactivation (2,314) | Li-Fraumeni and nonhereditary cancers | [50]; P53.iarc.fr |
| PAH (452) | Catalytic activity % wild type (80) | Phenylketonuria | [51]; www.pahdb.mcgill.ca |
| | PAH protein % wild type (80) | | |
| | *PAH* mRNA % wild type (80) | | |
| | Serum Phe and BH$_4$ response (>35) | | |
| LDLR (860) | Cellular LDL uptake (79) | Cardiovascular disease | [27,52]; umd.be/LDLR/ |
| | Expression (79) | Hypercholesterolemia | ucl.ac.uk/ldlr |
| | Plasma LDL-c (~79) | | |
| BRCA2 (3,418) | Homologous recombination (140) | Hereditary breast cancer | [53] |
| CBS (551) | Cell growth (204) | Hyperhomocysteinemia | [54] |

Both locus-specific databases (LSDBs) and published manuscripts contain data to enable development of new in silico bioinformatics methods to predict variant impact on endophenotypes. For each gene, available endophenotypes, number of unique mutations with endophenotypic values, disease phenotype(s) and links/references to data sources is provided.

doi:10.1371/journal.pcbi.1004725.t002

## Conclusion

As next-generation sequencing is integrated into routine patient care, in silico bioinformatic missense cSNV prediction tools have the potential to contribute to clinical practice. As of this writing, independent assessments of these tools indicate that they do not perform consistently, and there is considerable skepticism about their clinical utility. We reason that many of the apparent limitations are the result of a weakly defined paradigm. The tools are tasked with classifying cSNVs as disease causing, but most cSNVs by themselves do not have a large effect on whether an individual develops disease. The tools are also expected to assign cSNVs to phenotypic classes, although there is disagreement about how many of these classes should be considered and even which cSNVs belong in each class.

There are several potentially important, additional considerations regarding the performance of phenotypic prediction and what might be reasonably expected from endophenotypic prediction. First, many methods presented in Table 1 advertise the ability to predict variant impact on protein native state. While a connection between variant impact on protein native state and disease is often drawn, it is also acknowledged that these variables are not synonymous. Given that classifiers are often tasked with predicting impact on health or benchmarked using databases of putatively disease-causing variants, rather than assessing protein *damage*, methods are developed and challenged using disparate criteria. The Protein Mutant Database (PMD) employed in the Olatubosun et al. study [6] (Table 1) does record the impact of mutation on protein *activity*, potentially circumventing some of the above-described limitations. However, the PMD reduces continuous-valued activities (percent of wild type) to six discrete categories, and Olatubosun et al. further reduced categorization to either "functional" or "nonfunctional"; this clearly results in information loss, similar to that encountered when dichotomizing variants into discrete pathogenic and benign categories. It could be informative to compare the continuous-valued output from classifiers with the actual, non-stratified continuous-valued protein activities. This approach would be similar to that pursued in the PAH-PKU example from Wettstein et al. (above) [37]; this endophenotypic approach avoids the potentially dubious dichotomization of both algorithmic output and the experimental protein activities.

Endophenotype prediction presents new technical challenges, both in data acquisition and methods development. Although the large-scale development of gene-endophenotype databases will require community-wide effort, we see this as a tractable problem. Given that diseases are defined and diagnosed using quantitative endophenotypic risk factors, screens of genetic risk factors and association studies could, when possible, catalogue the continuous-valued endophenotypes used to partition the cases and controls. Some examples of this type of database curation are included in Table 2.

Wettstein et al. showed that some existing methods could potentially be repurposed for endophenotype prediction [37]. Even though most cSNV classifiers return dichotomous or categorical predictions, the underlying score functions calculate continuous-valued scores that could be correlated with measured endophenotypes. Although these existing classifiers suffer the limitation that they are not endophenotype specific, assessing the correlation of different endophenotypes and the continuous-valued output of existing methods could be informative. Also, the individual components of some methods' score functions might be useful to help infer mechanism. For instance, SNPeffect combines four scores that each estimate distinct protein phenomena (amyloid formation, aggregation, stability, and chaperone binding) [55]. Classifiers that are gene and endophenotype specific—such as those produced by the ePOSE algorithm—will benefit from learning which variants (or genes) contribute to specific components of disease: as an example, CFTR variants that effect processing versus chloride

conductance, or complex heart disease phenotypes that can result from varying combinations of *LDLR*-specific cholesterol plaques [27] or *LPA*-specific calcium plaques [56]. Undoubtedly, successful endophenotype prediction will benefit from diverse approaches.

Endophenotype predictors could be a useful complement for predicting complex diseases. A hallmark of complex disease is the presentation of varied combinations of traits associated with that disease, in which different traits can be influenced by different genetic risk factors. The quantitative traits are themselves endophenotypes, and predicting these objective traits, rather than a clinically defined abstraction of traits (i.e., phenotypes), could provide unique opportunities. For instance, predicting these quantitative traits facilitates the decomposition of complex disease into simpler, individual risk factors. For endophenotype predictors that are gene specific, this benefit will largely depend on a priori knowledge regarding causal genes.

In contrast to disease phenotype classes, endophenotypes are quantitative measurements having shared genetic underpinnings with disease phenotypes of interest. We suggest that in silico tools can be developed to predict the impact of cSNVs on endophenotypes, yielding improved accuracy and added value into the study of the mechanism and severity of cSNV impact on disease. The ePOSE algorithm provides a proof of concept and yields promising results in predicting three endophenotypes for a small set of cystic fibrosis cSNVs from the CFTR2 database [39]. The feasibility of such an approach will require community-wide efforts to augment the information currently available in LSDBs and other mutation databases.

# References

1.  Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. Annual Review of Genomics and Human Genetics. 2006; 7(1):61–80.

2.  Jordan DM, Ramensky VE, Sunyaev SR. Human allelic variation: perspective from protein function, structure, and evolution. Current Opinion in Structural Biology. 2010; 20(3):342–50. doi: 10.1016/j.sbi.2010.03.006 PMID: 20399638

3.  Karchin R. Next generation tools for the annotation of human SNPs. Briefings in Bioinformatics. 2009; 10(1):35–52. doi: 10.1093/bib/bbn047 PMID: 19181721

4.  Cline MS, Karchin R. Using bioinformatics to predict the functional impact of SNVs. Bioinformatics. 2011; 27(4):441–8. doi: 10.1093/bioinformatics/btq695 PMID: 21159622

5.  Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Human Molecular Genetics. 2015; 24(8):2125–37. doi: 10.1093/hmg/ddu733 PMID: 25552646

6.  Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M. PON-P: Integrated predictor for pathogenicity of missense variants. Human Mutation. 2012; 33(8):1166–74. doi: 10.1002/humu.22102 PMID: 22505138

7.  Rapakoulia T, Theofilatos K, Kleftogiannis D, Likothanasis S, Tsakalidis A, Mavroudi S. EnsembleGASVR: a novel ensemble method for classifying missense single nucleotide polymorphisms. Bioinformatics. 2014; 30(16):2324–33. doi: 10.1093/bioinformatics/btu297 PMID: 24771561

8.  Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Human Mutation. 2013; 34(1):57–65. doi: 10.1002/humu.22225 PMID: 23033316

9.  Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. Human Mutation. 2011; 32(4):358–68. doi: 10.1002/humu.21445 PMID: 21412949

10. Glahn DC, Knowles EE, McKay DR, Sprooten E, Raventos H, Blangero J, et al. Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. Am J Med Genet B Neuropsychiatr Genet. 2014; 165B(2):122–30. doi: 10.1002/ajmg.b.32221 PMID: 24464604

11. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry. 2003; 160(4):636–45. PMID: 12668349

12. Ng PC, Henikoff S. Accounting for Human Polymorphisms Predicted to Affect Protein Function. Genome Research. 2002; 12(3):436–46. doi: 10.1101/gr.212802 PMID: 11875032

13. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Meth. 2010; 7(4):248–9.

14. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Research. 2007; 35(11):3823–35. PMID: 17526529

15. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics. 2006; 22(22):2729–34. PMID: 16895930

16. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Research. 2003; 13(9):2129–41. PMID: 12952881

17. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Human Mutation. 2009; 30(8):1237–44. doi: 10.1002/humu.21047 PMID: 19514061

18. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009; 25(21):2744–50. doi: 10.1093/bioinformatics/btp528 PMID: 19734154

19. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Research. 2011; 39(17):e118. doi: 10.1093/nar/gkr407 PMID: 21727090

20. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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. 2015; 17 (5):405–24. doi: 10.1038/gim.2015.30 PMID: 25741868

21. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–53. doi: 10.1038/nature08494 PMID: 19812666

22. Visscher Peter M, Brown Matthew A, McCarthy Mark I, Yang J. Five Years of GWAS Discovery. The American Journal of Human Genetics. 2012; 90(1):7–24. doi: 10.1016/j.ajhg.2011.11.029 PMID: 22243964

23. Zlotogora J. Penetrance and expressivity in the molecular age. Genet Med. 2003; 5(5):347–52. PMID: 14501829

24. Cooper D, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013; 132(10):1077–130. doi: 10.1007/s00439-013-1331-2 PMID: 23820649

25. Consortium GP. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–73. doi: 10.1038/nature09534 PMID: 20981092

26. Waddell N, Ten Haaf A, Marsh A, Johnson J, Walker LC, kConfab I, et al. BRCA1 and BRCA2 missense variants of high and low clinical significance influence lymphoblastoid cell line post-irradiation gene expression. PLoS Genet. 2008; 4(5):e1000080. doi: 10.1371/journal.pgen.1000080 PMID: 18497862

27. Thormaehlen AS, Schuberth C, Won H-H, Blattmann P, Joggerst-Thomalla B, Theiss S, et al. Systematic Cell-Based Phenotyping of Missense Alleles Empowers Rare Variant Association Studies: A Case for LDLR and Myocardial Infarction. PLoS Genet. 2015; 11(2):e1004855. doi: 10.1371/journal.pgen.1004855 PMID: 25647241

28. Huijgen R, Kindt I, Defesche JC, Kastelein JJP. Cardiovascular risk in relation to functionality of sequence variants in the gene coding for the low-density lipoprotein receptor: a study among 29 365 individuals tested for 64 specific low-density lipoprotein-receptor sequence variants. European Heart Journal. 2012; 33(18):2325–30. doi: 10.1093/eurheartj/ehs038 PMID: 22390909

29. Huijgen R, Kindt I, Fouchier SW, Defesche JC, Hutten BA, Kastelein JJP, et al. Functionality of sequence variants in the genes coding for the low-density lipoprotein receptor and apolipoprotein B in individuals with inherited hypercholesterolemia. Human Mutation. 2010; 31(6):752–60. doi: 10.1002/humu.21258 PMID: 20506408

30. John B, Lewis KR. Chromosome variability and geographic distribution in insects. Science. 1966; 152 (3723):711–21. PMID: 17797432

31. Gottesman II, Shields J. Schizophrenia and genetics: A twin study vantage point. Oxford, England: Academic Press; 1972. xviii, 433 p.

32. Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, et al. A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. Nat Genet. 1997; 15(3):273–6. PMID: 9054940

33. Mitchell BD, Cole SA, Bauer RL, Iturria SJ, Rodriguez EA, Blangero J, et al. Genes Influencing Variation in Serum Osteocalcin Concentrations Are Linked to Markers on Chromosomes 16q and 20q. The Journal of Clinical Endocrinology & Metabolism. 2000; 85(4):1362–6. doi: 10.1210/jcem.85.4.6571

34. Kiel D, Demissie S, Dupuis J, Lunetta K, Murabito J, Karasik D. Genome-wide association with bone mass and geometry in the Framingham Heart Study. BMC Medical Genetics. 2007; 8(Suppl 1):S14. PMID: 17903296

35. Sing CF, Stengård JH, Kardia SLR. Genes, Environment, and Cardiovascular Disease. Arteriosclerosis, Thrombosis, and Vascular Biology. 2003; 23(7):1190–6. PMID: 12730090

36. Lynn K-S, Li L-L, Lin Y-J, Wang C-H, Sheng S-H, Lin J-H, et al. A neural network model for constructing endophenotypes of common complex diseases: an application to male young-onset hypertension microarray data. Bioinformatics. 2009; 25(8):981–8. doi: 10.1093/bioinformatics/btp106 PMID: 19237446

37. Wettstein S, Underhaug J, Perez B, Marsden BD, Yue WW, Martinez A, et al. Linking genotypes database with locus-specific database and genotype-phenotype correlation in phenylketonuria. Eur J Hum Genet. 2015; 23(3):302–9. doi: 10.1038/ejhg.2014.114 PMID: 24939588

38. Stanke F, Hedtfeld S, Becker T, Tummler B. An association study on contrasting cystic fibrosis endophenotypes recognizes KRT8 but not KRT18 as a modifier of cystic fibrosis disease severity and CFTR mediated residual chloride secretion. BMC Medical Genetics. 2011; 12(1):62.

39. Masica DL, Sosnay PR, Raraigh KS, Cutting GR, Karchin R. Missense variants in CFTR nucleotide-binding domains predict quantitative phenotypes associated with cystic fibrosis disease severity. Human Molecular Genetics. 2015; 24(7):1908–17. doi: 10.1093/hmg/ddu607 PMID: 25489051

40. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nature genetics. 2013; 45 (10):1160–7. doi: 10.1038/ng.2745 PMID: 23974870

41. Lenzenweger MF. Thinking clearly about the endophenotype–intermediate phenotype–biomarker distinctions in developmental psychopathology research. Development and Psychopathology. 2013; 25 (25th Anniversary Special Issue 4pt2):1347–57.

42. Kohler S, Schulz MH, Krawitz P, Bauer S, Dolken S, Ott CE, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. American journal of human genetics. 2009; 85 (4):457–64. doi: 10.1016/j.ajhg.2009.09.003 PMID: 19800049

43. Van Goor F, Hadida S, Grootenhuis PDJ, Burton B, Cao D, Neuberger T, et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. Proceedings of the National Academy of Sciences. 2009; 106(44):18825–30.

44. Brodlie M, Haq IJ, Roberts K, Elborn JS. Targeted therapies to improve CFTR function in cystic fibrosis. Genome medicine. 2015; 7(1):101. doi: 10.1186/s13073-015-0223-6 PMID: 26403534

45. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Human mutation. 2008; 29 (11):1342–54. doi: 10.1002/humu.20896 PMID: 18951461

46. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Research. 2005; 33(suppl 2):W382–W8.

47. Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. BMC bioinformatics. 2006; 7:166. doi: 10.1186/1471-2105-7-166 PMID: 16551372

48. Masica DL, Sosnay PR, Cutting GR, Karchin R. Phenotype-optimized sequence ensembles substantially improve prediction of disease-causing mutation in cystic fibrosis. Human Mutation. 2012; 33 (8):1267–74. doi: 10.1002/humu.22110 PMID: 22573477

49. Yu H, Burton B, Huang C-J, Worley J, Cao D, Johnson JP Jr, et al. Ivacaftor potentiation of multiple CFTR channels with gating mutations. Journal of Cystic Fibrosis. 2012; 11(3):237–45. doi: 10.1016/j.jcf.2011.12.005 PMID: 22293084

50. Leroy B, Anderson M, Soussi T. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum Mutat. 2014; 35(6):672–88. doi: 10.1002/humu.22552 PMID: 24665023

51. Scriver CR, Hurtubise M, Konecki D, Phommarinh M, Prevost L, Erlandsen H, et al. PAHdb 2003: what a locus-specific knowledgebase can do. Hum Mutat. 2003; 21(4):333–44. PMID: 12655543

52. Villeger L, Abifadel M, Allard D, Rabes JP, Thiart R, Kotze MJ, et al. The UMD-LDLR database: additions to the software and 490 new entries to the database. Hum Mutat. 2002; 20(2):81–7. PMID: 12124988

53. Guidugli L, Pankratz VS, Singh N, Thompson J, Erding CA, Engel C, et al. A classification model for BRCA2 DNA binding domain missense variants based on homology-directed repair activity. Cancer research. 2013; 73(1):265–75. doi: 10.1158/0008-5472.CAN-12-2081 PMID: 23108138

54.  Wei Q, Wang L, Wang Q, Kruger WD, Dunbrack RL. Testing computational prediction of missense mutation phenotypes: Functional characterization of 204 mutations of human cystathionine beta synthase. Proteins: Structure, Function, and Bioinformatics. 2010; 78(9):2058–74.

55.  De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, et al. SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. Nucleic Acids Research. 2012; 40(D1):D935–D9.

56.  Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith AV, Peloso GM, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. New England Journal of Medicine. 2013; 368 (6):503–12. doi: 10.1056/NEJMoa1109034 PMID: 23388002