Supplementary Material for: “Genomewide significance testing of variation from single case exomes”

Amy B. Wilfert¹, Katherine R. Chao², Madhurima Kaushal³, Sanjay Jain³,⁵, Sebastian Zollner⁴, David R. Adams², Donald F. Conrad¹,⁵

¹Department of Genetics, ²Department of Internal Medicine, ³Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA
²NIH Undiagnosed Diseases Program, NIH, Bethesda, Maryland, USA
⁴Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

Corresponding Author:
Dr. Donald F. Conrad
Department of Genetics
Washington University School of Medicine
Campus Box 8232
St. Louis MO, 63110, USA
E-mail: dconrad@genetics.wustl.edu
Supplementary Note

In this supplemental note, and the display items that follow, we present a number of analyses that are meant to gauge the sensitivity of our results to choices in modeling and database usage. As an alternative to CADD, we developed our own disease mutation classifier based on a logistic regression model; we also make use of a different classifier known as rSVM published elsewhere\(^1\). We demonstrate that the use of ClinVar or HGMD as the source of training data for a disease mutation classifier does not have a substantial impact on the accuracy of the resulting classifier. We provide further background details on the UDP cohort including composition and the results of diagnosis by the clinical team at the NIH. Finally, we provide more context for the PSAP approach by comparison with other analysis methods, and outline known limitations and workarounds.

Blacklists

Poorly calibrated genes were identified empirically by looking for genes with an excess of low PSAPs in unrelated healthy individuals. Genes with more than two variants with PSAPs < 10\(^{-5}\) in a single individual were considered to be poorly calibrated and were added to the blacklist for removal in future analyses.

Logistic regression model

Four logistic regression models were trained on data from HGMD and ESP6500, ClinVar and ESP6500, and HGMD and 1000GP. For all models, the training set consisted of 60,000 randomly selected benign variants and 5,000 randomly selected disease variants. All models regressed predictions from SIFT, PolyPhen2-HDIV, PolyPhen2-HVAR, LRT, MutationTaster, and GERP++ on the classification of the variant as benign or disease-causing. The performance of all the models was assessed using receiver operator characteristic (ROC) curves and area under the curves (AUC) by using a set of 5,000 randomly selected benign and disease variants each that was excluded from the training set. The variants included in the training set from ESP6500 were the same in the HGMD and ESP6500, and ClinVar and ESP6500 models when the models were compared. The variants included in the training set from HGMD were the same in the HGMD and ESP6500, and HGMD and 1000GP models when the models were compared. A set of 5,000 randomly selected benign and disease variants each that was excluded from the training set and had no missing data across all the prediction models was used to compare the HGMD and ESP6500 model to pCADD, and the individual covariates.

Comparison of ClinVar and HGMD based analyses

In light of the release of the ClinVar database, a stringently curated set of verified disease mutations, the data collected in HGMD has been highly criticized for its reliability. We have found that there are no differences in performance between the logistic regression model trained on HGMD data and the logistic regression model trained on ClinVar. If the models differed, we would have expected them to group according to the data in which they were trained, but instead we found that the models grouped according to the dataset on which they were tested (Supplementary Figure 9). The HGMD and ClinVar had similar areas under the curve (AUCs) for both test sets. Interestingly, both models had larger AUCs when tested on a subset of the ClinVar data (Supplementary Figure 9), which does suggest the data has fewer false positives. Since only 6.7% of all HGMD variants are also in ClinVar and the percent of overlap between HGMD and ClinVar variants included in the training and test sets was even lower at 4.3%, it is unlikely these differences are caused by an artifact in the training or test data. Therefore, we conclude that the composition of the training data does not alter the strength of the signal between disease and benign mutations detected by the model.

Modeling Singleton Rate

As we developed the PSAP modeling framework, we set out to make intentionally conservative modeling choices as a first pass. The rationale was, if bias was to be introduced, we wished to bias the resulting p-values towards the large end (e.g. insignificant end, towards 1). For the analyses presented in the main text,
we use a single gene-specific singleton rate for PSAP calculations on each gene. This rate estimate is an average of the rate of singleton variants observed at highly functional sites (non-synonymous sites and splice sites) as well as less functional sites (synonymous sites). This has the impact of making rare non-synonymous variation look less interesting (i.e. more likely), assuming that the true singleton rate at non-synonymous sites is lower than at synonymous sites. Later in the project, we performed PSAP calculations using a two parameter singleton rate model – one for synonymous sites, and one for non-synonymous sites. The PSAP p-values for the AR model are unaffected. For the AD model, 90% of genes have lower average PSAP p-values, underscoring that our current method for calculating p-values is conservative. However, the median difference in p-value is only 10% (mean 20%). Similarly, 83% of genes have lower average PSAP p-values under the CHET model, with a median difference of 5% (mean 9%). It is difficult to extrapolate from first principles how PSAP p-values will be affected if use more complex singleton rate models (e.g. derived from triplet-based mutation models). However, our own empirical analysis using a 5mer mutation model on genes from chromosome 22 indicate that the position specific singleton rates will always fall within +/- 1 order of magnitude (Supplementary Figure 16).

Overview of UDP Cohort

The Undiagnosed Diseases Program is a clinical genetics project based out of the NIH Clinical Center, with a network of affiliated clinical groups spread throughout the United States. The goals and organization of the UDP have been described in detail elsewhere. Here we provide a brief characterization of the families included in the present study. All families were recruited at the NIH Clinical Center. A variety of pedigrees were sequenced, all of which were two-generation, ranging from parent-offspring trios with one affected to 6 member families with multiple affecteds. A detailed family history was obtained for each family, including medical history and geographic ancestry going back 2 generations from the proband. The self-reported ancestry revealed that 94% of the families were Caucasian, 4% were Asian, 2% were African-American or Asian-American, and 1% of unknown ancestry. Of all 135 families, 112 (83%) were sporadic singleton cases, 22 (16%) were suspected recessive patterns of inheritance based on multiple affecteds, and 1 (1%) suspected dominant inheritance based on observation of an affected parent. Full details of the diagnostic genes are provided in Supplementary Table 2. Of the 30 families with diagnoses, 3 were segregating causal variants at multiple loci. Of the remaining 27, two were single variant AR, 3 were X-linked recessive, 11 were compound het recessive, and 11 were AD. The vast majority of the 30 diagnoses described here expand the phenotype associated with a known disease gene, as opposed to identify a novel disease gene or confirm a well-accepted set of clinical symptoms associated with a known disease gene. The functional consequences of the causal mutation in most of the 30 diagnosed cases were not easily recognizable as a deleterious based on molecular consequences alone: Of the 27 diagnosed cases with single gene phenotypes:

2 de novo loss of function,

2 compound hets of splice site + missense

2 compound hets of frameshift indel + missense

21 missense only

UDP Ancestry Analysis

As demonstrated in Figure S5, population structure does appear to inflate the distribution of PSAP broadly across the genome, especially in moderate ranges of the PSAP distribution (0.01-10^-4). Importantly, we rarely observe variants in what we consider to be the range of clinically interesting variation <10^-4. We report an enrichment of low of PSAP values (<10^-5) in the UDP unaffecteds in the main text, but it is important to
consider population structure as an explanation for this observation. We performed principal components analysis on the exome genotype calls from the full set of UDP individuals and 6 reference samples from diverse populations, sequenced by the 1000 genomes project. These reference individuals consisted of 2 African Americans, 2 East Asians and 2 Caucasians. Results of the PCA are consistent with UDP participants’ self-reported ancestry: the majority of samples cluster into a single Caucasian cluster (Figure S11). We next examined the correlation of low PSAP values with principal components loadings for all UDP individuals. Notably, the majority of individuals in the UDP do not have PSAP values < $10^{-5}$ for each disease model. For instance, ~87% of samples have no PSAP < $10^{-5}$ under the CHET model, and only 3 individuals have > 1 PSAP < $10^{-5}$. For each of the three disease models considered, we found no statistical association between the PC1 loading and the number of PSAP values < $10^{-5}$ (Supplementary Figure 12).

Comparison with VAAST

Other groups have described probabilistic methods for identifying disease mutations that, like our approach, allow incorporation of functional annotation as weights or priors for each test \(^{20,21}\). One of the most commonly used approaches, VAAST, uses a very general likelihood ratio framework to test for association between variation and a phenotype. VAAST resembles our method in that variation can be grouped for testing within a feature of arbitrary size and location, and that variation can be weighted by user-supplied functional annotation. VAAST explicitly estimates an alternative model from the case population and uses a likelihood ratio test to reject the null hypothesis of no difference between cases and controls. In contrast, the approach that we describe here forgoes explicit estimation of an alternative model. We report sampling probabilities from a well-calibrated null model, which should have all of the properties of a conventional p-value. The properties of the likelihood ratio framework mean that a likelihood test will be highly sensitive to estimation errors when obtaining parametric p-values with small sample sizes (such as n=1). This will result in an inflated type I error rate, making it more difficult to unambiguously identify causal genotype from a background of genomewide data. Indeed, in the original VAAST paper, the authors evaluate VAAST as a genotype prioritization tool for n-of-1 cases involving an autosomal recessive locus in 100 different OMIM genes, reporting an average rank of 373 for a disease-causing genotype. For the same set of genotypes and inheritance model, PSAP reports an average rank of 8.23 (median 1) in a European American genetic background, and 21.38 (median 1) in an African American background.

PSAP limitations and workarounds

During the course of this work, we identified specific limitations of the PSAP approach that should be kept in mind during a rare disease analysis. We observed that PSAP p-values are sensitive to population structure. In principle, a correction factor could be estimated and applied in this case, much in the way that genomic control is estimated from the distribution of p-values obtained in a conventional genomewide association study \(^{17}\). Second, sequencing error produces unlikely base calls that we expect will be enriched with low PSAP values. We have found that by assessing a large number of samples sequenced and analyzed at the same center, one can create “blacklists” of genes with CADD distributions that are poorly modeled (Methods, Supplementary Figure 16). Third, we have investigated the genotypes where CADD vastly outperforms PSAP as a prioritization tool; this almost always occurs in genes where there are high-frequency variants that have CADD scores as large or larger than a known HGMD variant within that gene (Fig. 3C). We suggest that such situations can be best addressed by (a) applying an allele frequency filter along with PSAP or (b) using the combined PSAP and CADD scores when investigating an exome. Fourth, the calculation of PSAP p-values that we present for the CHET model does not explicitly account for the phase of the variants. When large, phased control sequencing datasets become available, phase can be explicitly incorporated into the CHET calculation, but in the short term investigators will want to consider the observed phase of the heterozygous sites in the proband when interpreting an interesting CHET PSAP p-value.
Supplementary figure 1: Impact of database size on the percent of variants observed in an individual that are reported in the database as described by Mitchell et al. 

- **Left panel:** Probability all variants in an individual are observed in the database.
- **Right panel:** Probability one or more variants in an individual are not observed in the database.
Supplementary Figure 2: PSAP performs well using a variant annotation method other than CADD. Here we show the results of the spike-in analysis in described in the main text, but in addition to using CADD as our annotation metric, we show results using a different classifier, rSVM$^1$. Results are presented for the homozygous recessive disease model (n=20,701 HGMD variants) The y-axis indicates the log rank of the disease variant within the simulated exome, and the x-axis indicates the rank of variant based on its predicted pathogenicity (e.g. the C-score, rSVM score, PSAP p-value). Compare further to Figure S10.
Supplementary Figure 3: Comparison of gene-specific CADD score distributions simulated from the PSAP null model (in red) to those observed in 100 healthy individuals (in black). Four random genes were selected for this figure.
Supplementary Figure 4: Histogram of the average difference between simulated and observed cumulative distributions of summary scores across a gene for the single variant recessive (right) and compound heterozygous disease models. Red line indicates the sample mean.

To ensure our gene-specific null models were well calibrated, we compared our fitted models to empirical data from 434 healthy exomes. We found that our null models are conservative but well calibrated representations of the healthy population (Figure 2A). On average, we observed a Kullback-Leibler divergence of 0.01 or smaller between our models and the empirical data, when averaged across all genes, and found the recessive models better fit the empirical data than the dominant model. The most common error we made was overestimating the frequency of low CADD scores. Low CADD scores are typically associated with common variants, which are more likely to be found in regions of high LD.
Supplementary Figure 5. Estimation of an LD correction term for the CHET model. When we initially inspected the calibration of our CHET p-values calculated on control datasets, we observed a modest but consistent bias of our p-values to be 30-40% lower than expected. This is likely due to the fact that our CHET probability calculations require the assumption of independence (linkage equilibrium) between variation within the same gene. Any violation of this assumption will bias our p-values. To improve the calibration of our CHET p-values, we fit a linear model relating the observed CHET p-value distribution from the GTEx cohort (blue points) to the p-value distribution expected under the null. We then apply this fitted linear model to recalibrate our CHET p-values (green points).
Supplementary Figure 6: We performed the same variant prioritization analysis as shown in Figure 4A of the main text, this time simulating cases of Mendelian disease using genomes of both European American (EA, n=189) and African American (AA, n=189) descent. We found that demographic background has a measurable but modest impact, with Mendelian disease variants ranking higher in AA genomes than in EA genomes. This difference is most noticeable in the autosomal dominant disease model, with a median difference in rank between the AA and EA background of 25, but the trend is present across all models (median differences of 2 and 0 in homozygous and compound heterozygous recessive disease models, respectively). Considering the fact that African Americans have an increased amount of heterozygosity and rare variation across their genome, this result is unsurprising.
Supplementary Figure 7. A single gene example demonstrating differences in the strength of signal between benign genotypes and known disease genotypes using gene-specific null models for the single variant recessive and dominant disease models. The grey lines in each panel indicate the average CADD score for a substitution at each nucleotide in the PTEN gene.
Supplementary Figure 8. The theoretical distribution of possible CADD scores (or the “state space” of CADD) varies among genes in the genome, as illustrated here with four different genes. For each gene we have plotted the complete distribution of CADD for all 3k substitutions possible away from the reference genome sequence (k = number of coding nucleotides in the gene). While nearly 50% of possible substitutions in SH3BGRL3 have phred-scaled CADD scores greater than 15, it is not possible to observe a CADD score of 15 in AACS and DSPP due to the calibration of the CADD scoring method. Thus it is difficult to compare the significance of CADD scores directly among genes without additional standardization.
Supplementary Figure 9: Comparison of true positive and false positive rates for logistic regression models trained on HGMD (n=5,000 variants for test and training sets) and ESP6500 (n=60,000 for test and training sets) or ClinVar (n=5,000 variants for test and training sets) and ESP6500 data (n=60,000 for test and training sets).
Supplementary Figure 10. Benchmarking our ability to identify the causal gene in simulated n-of-one cases. N-of-one exomes were simulated using SNVs reported to cause disease in HGMD (AD=30,222; AR=27,110; CHET=27,110) and genotype calls from the exomes of 189 healthy individuals (Methods). For each simulated disease exome, genes were rank ordered according to (i) PSAP $p$-value (CHET model), (ii) a gene-based statistic using CADD, or (iii) the combination of CHET PSAP $p$-value and CADD. As exome analysis is often performed with an allele frequency filter, we also plot the performance of each statistic when applied to the subset of exome variants with minor allele frequency $<1\%$ in each simulated case (“+ AF”). The rank of the disease gene within the simulated case is shown on the y-axis, while the simulated disease genes are ranked against each other based on the CADD scores of their disease mutations (x-axis).
Supplementary Figure 11: Gene-specific probabilities of observing a protein-truncating variant (x-axis) correlate with an established metric for assessing haploinsufficiency (y-axis). The blue line indicates the loess regression fit. The Spearman correlation between the two sets of values is 0.37.
Supplementary Figure 12. We calculated the average CADD score for protein truncating SNVs for each gene in the genome, and used these gene-specific averages as gene-specific estimates of the functional impact of frameshift indels. We then used this scheme to obtain PSAP values for indels and evaluated our ability to prioritize indels from HGMD using the same simulation framework as in Figure 4A; however, this time, we examine the rank of each disease-causing indel within the background of all genetic variants (indels and SNVs) of each exome. Results are plotted separately for each of the three disease models that we evaluated (AD=autosomal dominant, AR= single variant recessive, CHET = compound heterozygous). Each line indicates the rank of the known disease variant within the simulated disease exome relative to the predicted severity of the disease variant.
Supplementary Figure 13. Relationship between population structure and low PSAP scores in the UDP. Top left - PCA loadings for all UDP subjects, as well as demographic control individuals sequenced by the 1000 genomes project. Each individual is colored by self-reported ancestry. Remaining three panels - for each disease model, we plot the distribution of PC1 loadings for individuals with 0, 1, 2, 3, or 8 genes with PSAP < 1e-5 ('low scoring genes'). No significant differences in ancestry were detected between individuals with 0 and more than 0 low scoring genes.
Supplementary Figure 14. Enrichment of low PSAP values in CAKUT n-of-one cases. We evaluated 120 n-of-one cases affected with congenital anomalies of the kidney and urinary tract (CAKUT) recruited in the Division of Nephrology at Washington University. We used 418 ethnicity-matched control exomes generated by the GTEx consortium for comparison. For each disease model we identified PSAP p-value thresholds corresponding to an approximately 5% false discovery rate (for AD and AR, this was p<1e-5; for CHET, p<2.5e-5). These bar plots depict the percent of affected and unaffected individuals with at least one gene with a PSAP less than 1e-5 (AD and AR) or 2.5e-5 (CHET), compared to the expected number. Results are stratified by disease model and gene set evaluated. Fisher exact tests were used to identify cohorts that were significantly enriched for individuals with low PSAP values compared to expectation. * = p-value < 0.05, *** = p-value < 0.01.
Supplementary Figure 15: Plots of CADD scores corresponding to a PSAP of $1^{-5}$ versus three unique constraint metrics. The top row looks compares the CADD scores corresponding to a PSAP of $1^{-5}$ to the probability a gene is loss-of-function intolerant (Samocha et al.) for the heterozygous dominant, homozygous recessive, and compound heterozygous recessive models, respectively. The middle row compares CADD scores corresponding to a PSAP of $1^{-5}$ to RVIS (Petrovski et al.), and the bottom row compares the CADD scores corresponding to a PSAP of $1^{-5}$ to the probability a gene is haploinsufficient (Huang et al.). The blue lines in each plot represent the line of best fit, as fitted by linear regression. All plots show that PSAP captures information that these constraint metrics do not. The correlation coefficients comparing the various constraint metrics with 3 different PSAP thresholds are shown in Supplementary Tables 4–6.
Supplementary Figure 16. In the PSAP calculations presented in the main text, we conservatively assume a uniform singleton rate for all bases in a gene. Our use of a single value for each gene-specific singleton rate is equivalent to setting the singleton rate for all bases in a gene equal to the mean singleton rate in that gene (i.e. there is not rate variation). We wish to know if this treatment of ignoring mutation (singleton) rate variation will lead to seriously biased PSAP values. In order to the impact of this decision, we applied a context-dependent 5-mer-based mutation model. For each gene on chromosome 22, we estimated base-specific mutation rates of every coding basepair position. For all genes that we analyzed, all base-specific mutation rates in a given gene are within one order of magnitude of the mean for that gene (max =9.6-fold greater, min=0.125 fold less). The histogram shows the distribution of base-specific rate/whole-gene average rate for 150kbp spanning 28 genes on chromosome 22. While PSAP p-values using this mutation rate model could still vary by greater than 1 order of magnitude from PSAP p-values derived from a single rate singleton model, this is difficult to predict from first principles and will need to be explored empirically in future work.
Supplementary Figure 17: Notable properties of the blacklisted genes. Top left figure shows distribution of the number of genes in a set of 162 randomly sampled genes that had an average coverage less than 10, repeated 10,000 times. The red line indicates the number of low coverage genes observed in the blacklist. Top right figure shows the distribution of the number of genes in a set of 162 genes where the standard deviation of coverage was greater than the mean coverage, repeated 10,000 times. The red line indicates the number of genes from the blacklist that had standard deviations of coverage greater than the mean coverage. The bottom left histogram shows the distribution of mean nucleotide coverage across a gene in ExAC. The white bars describe all genes and the red bars describe blacklisted genes. The bottom right histogram
describes the distribution of the ration of mean NT coverage across a gene to the standard deviation in that coverage. The white bars describe all genes and the red bars describe only blacklisted genes.

Supplementary Table 1. Empirical distribution of PSAP-MIN for Simulated Disease Exomes. As described in the main text, we simulated over 4 million Mendelian disease exomes for each of 3 different disease models. For each disease model, we tabulated the fraction of individuals with a PSAP-MIN p-values less than one of four benchmark values (1e-4, 1e-5, 5e-6, <=1e-6). This table represents an estimate of sensitivity for using PSAP-min as a test statistic in Mendelian disease genome identification. For the AD model 28,493 HGMD variants were used in the simulation; for CHET and AR models, 25,567 variants were used. AD = autosomal dominant disease model, AR = autosomal recessive disease model, and CHET = compound heterozygous disease model.

| PSAP-MIN threshold | AD 5.3M simulated cases | CHET 4.8M simulated cases | AR 4.8M simulated cases |
|--------------------|--------------------------|----------------------------|-------------------------|
| p-value            |                          |                            |                         |
| 1.00E-04           | 0.110                    | 0.585                      | 0.735                   |
| 1.00E-05           | 0.016                    | 0.380                      | 0.422                   |
| 5.00E-06           | 0.006                    | 0.327                      | 0.295                   |
| ≤1.00E-06          | 0.002                    | 0.000                      | 0.129                   |

Supplementary Table 2. Empirical distribution of PSAP-MIN for Three Population Control Cohorts. Here, we tabulate the proportion of individuals with PSAP-MIN < 10^-4, 10^-5, or 10^-6 for three disease models using the three controls cohorts presented in Figure 1 of the main text. Nsamps = number of individuals in the cohort. Unlike Table 2 of the main text, the (n) under each disease model heading represents the number of genes observed with at least one relevant genotype in each cohort (e.g. a heterozygous genotype in AD, a homozygous genotype in AR). AR=autosomal recessive disease model; CHET= compound heterozygote disease model; AD=autosomal dominant disease model.

| WHI EA (nsamps = 199) | AD (n =15,509 ) | AR (n =10,466 ) | CHET (n = 9,558) |
|------------------------|-----------------|-----------------|------------------|
| 1.00e-4                | 0.502           | 0.326           | 0.085            |
| 1.00e-5                | 0.057           | 0.065           | 0.015            |
| ≤ 1.00e-6              | 0.005           | 0.005           | 0.005            |

| GTEX (nsamps = 418) | AD (n = 15,614) | AR (n =11,048) | CHET (n = 10,159)|
|---------------------|-----------------|----------------|------------------|
| 1.00e-4              | 0.591           | 0.481          | 0.141            |
| 1.00e-5              | 0.124           | 0.050          | 0.019            |
| ≤ 1.00e-6            | 0.012           | 0.000          | 0.000            |

| SWED (nsamps = 2,189) | AD (n = 16,170) | AR (n =11,619) | CHET (n = 11,643)|
|-----------------------|-----------------|----------------|------------------|
| 1.00e-4               | 0.405           | 0.307          | 0.121            |
| 1.00e-5               | 0.065           | 0.055          | 0.023            |
| ≤ 1.00e-6             | 0.005           | 0.002          | 0.000            |
| Gene          | Location | Mutation Type     | ExAC Frequency | CADD | PSAP     | Inheritance Model |
|--------------|----------|-------------------|----------------|------|----------|-------------------|
| SLC52A2      | Exonic   | Nonsynonymous SNV | 6.51E-05       | 14.62| 4.63E-05 | CHET             |
| SLC52A2      | Exonic   | Nonsynonymous SNV | 3.25E-05       | 22.2 | 4.63E-05 | CHET             |
| ASXL3        | Exonic   | Stopgain          | NA             | 39   | 8.70E-05 | AD               |
| MEGF10       | Exonic   | Frameshift Indel  | NA             | 45   | 1.00E-06 | AR               |
| STIM1        | Exonic   | Nonsynonymous SNV | NA             | 34   | 2.40E-04 | AD               |
| TUBB2A       | Exonic   | Nonsynonymous SNV | NA             | 14.7 | 0.07     | AD               |
| NAGLU        | Exonic   | Nonsynonymous SNV | NA             | 23.7 | 3.86E-06 | CHET             |
| NAGLU        | Exonic   | Nonsynonymous SNV | NA             | 19.8 | 3.86E-06 | CHET             |
| CAD          | Exonic   | Nonsynonymous SNV | 8.13E-06       | 36   | 0.000394 | AD               |
| AARS2        | Exonic   | Nonsynonymous SNV | 3.00E-02       | 17.17| 3.86E-06 | CHET             |
| AARS2        | Exonic   | Nonsynonymous SNV | 1.22E-04       | 20.2 | 3.86E-06 | CHET             |
| CLN6         | Exonic   | Nonsynonymous SNV | NA             | 25.6 | 3.86E-06 | CHET             |
| CLN6         | Exonic   | Frameshift Indel  | NA             | 37   | 3.86E-06 | CHET             |
| KCNA1        | Exonic   | Stopgain          | NA             | 46   | 2.70E-05 | AD               |
| ABCD1        | Exonic   | Nonsynonymous SNV | NA             | 22   | 1.00E-04 | XL               |
| SPAST        | Exonic   | Nonsynonymous SNV | NA             | 27.1 | 0.000159 | AD               |
| AARS         | Exonic   | Nonsynonymous SNV | 4.88E-05       | 20.5 | 1.23E-03 | CHET             |
| AARS         | Exonic   | Nonsynonymous SNV | NA             | 29.4 | 1.23E-03 | CHET             |
| SMARCB1      | Exonic   | Nonsynonymous SNV | NA             | 24.4 | 0.00023  | AD               |
| PLA2G6       | Exonic   | Nonsynonymous SNV | NA             | 14.8 | 0.027671 | AD               |
| SLC13A5      | Exonic   | Nonsynonymous SNV | 0.0002033      | 19.74| 1.93E-05 | CHET             |
| GARS         | Exonic   | Nonsynonymous SNV | NA             | 35   | 3.86E-06 | CHET             |
| GARS         | Exonic   | Frameshift Indel  | NA             | 40   | 3.86E-06 | CHET             |
| RNAEH2B      | Exonic   | Nonsynonymous SNV | 0.001301       | 14.94| 1.90E-05 | AR               |
| ERCC6        | Exonic   | Nonsynonymous SNV | NA             | 35   | 2.61E-04 | CHET             |
| ERCC6        | Exonic   | Splicesite        | NA             | 13.8 | 2.61E-04 | CHET             |
| SMC3         | Exonic   | Nonsynonymous SNV | NA             | 20.1 | 0.000672 | AD               |
| DARS         | Exonic   | Nonsynonymous SNV | 8.13E-06       | 18.75| 3.34E-05 | CHET             |
| ROBO3        | Exonic   | Nonsynonymous SNV | 0.001113       | 15.04| 7.62E-03 | CHET             |
| SPTAN1       | Exonic   | Splicing          | NA             | 21.4 | 0.01     | AD               |
| GRB10        | Exonic   | Nonsynonymous SNV | NA             | 22.9 | 2.00E-03 | AD               |
| DYSR1A       | Exonic   | Frameshift Indel  | NA             | 49   | 4.00E-05 | AD               |
| SLC52A2      | Exonic   | Nonsynonymous SNV | NA             | 14.14| 1.20E-05 | AR               |
| PLEC         | Exonic   | Nonsynonymous SNV | 5.73E-05       | 12.9 | 0.00799  | AR               |
| ANGPTL3      | Intronic | Splicesite        | NA             | 3E-05| 0.000072 | AR               |
| ATP6V0A4     | Exonic   | Frameshift Indel  | NA             | 4E-05| 0.000072 | AR               |
| COL9A2       | Exonic   | Nonsynonymous SNV | 0.002392       | 26.5 | 1.10E-05 | AR               |
| ABCA4        | Exonic   | Nonsynonymous SNV | 1.63E-05       | 28.6 | 2.10E-05 | AR               |
| PARK7        | Exonic   | Frameshift Indel  | NA             | 30   | 4.00E-05 | AR               |
| AVPR2        | Exonic   | Nonsynonymous SNV | NA             | 7.597| 0.0014   | XL               |
| UDP802 | NPC1 | Exonic | Nonsynonymous SNV | 0.0001138 | 27.8 | 3.86E-06 | CHET |
|--------|------|--------|------------------|-----------|-------|----------|------|
| UDP802 | NPC1 | Exonic | Nonsynonymous SNV | NA        | 33    | 3.86E-06 | CHET |
| UDP930 | SMS  | Exonic | Nonsynonymous SNV | NA        | 16.6  | 9.20E-05 | XL   |

### Supplementary Table 4: UDP diagnostic variant detection with and without gene expression information

| Expression summaries used | Cross-tissue ranking of expression metrics | Within-tissue ranking of expression metrics |
|---------------------------|------------------------------------------|------------------------------------------|
|                           | Mean | Median | Max | Mean | Median | Max |
| Baseline (PSAP only)      | 2.4  | 1      | 14  | 2.4  | 1      | 14  |
| Mean                      | 2.3  | 1      | 11  | 2.3  | 1      | 9   |
| CV                        | 2.5  | 1      | 14  | 2.6  | 1      | 14  |
| SI                        | 2.9  | 1      | 15  | 3.0  | 2      | 9   |
| CV + SI                   | 2.8  | 2      | 13  | 2.8  | 1      | 13  |
| Mean + CV + SI            | 2.3  | 1      | 7   | 1.9  | 1      | 5   |

### Supplementary Table 5: Correlation of several constraint metrics to CADD scores corresponding to a PSAP of 0.05

| PSAP model | pLoF Intolerant (Samocha) | RVIS (Petrovski) | pHaploinsufficient (Huang) |
|------------|----------------------------|------------------|----------------------------|
| HET        | -0.103                     | 0.081            | -0.095                     |
| HOM        | -0.057                     | 0.058            | -0.059                     |
| CHET       | -0.069                     | 0.080            | -0.091                     |

### Supplementary Table 6: Correlation of several constraint metrics to CADD scores corresponding to a PSAP of 0.5

| PSAP model | pLoF Intolerant (Samocha) | RVIS (Petrovski) | pHaploinsufficient (Huang) |
|------------|----------------------------|------------------|----------------------------|
| HET        | -0.055                     | 0.068            | -0.072                     |
| HOM        | -0.015                     | 0.012            | -0.014                     |
| CHET       | -0.039                     | 0.115            | -0.038                     |

### Supplementary Table 7: Correlation of several constraint metrics to CADD scores corresponding to a PSAP of $1^{-5}$

| PSAP model | pLoF Intolerant (Samocha) | RVIS (Petrovski) | pHaploinsufficient (Huang) |
|------------|----------------------------|------------------|----------------------------|
| HET        | 0.184                      | -0.29            | 0.255                      |
| HOM        | 0.102                      | -0.078           | 0.104                      |
| CHET       | -0.091                     | -0.023           | -0.018                     |
1 Dong, C. et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Human molecular genetics 24, 2125-2137 (2015).
2 Lage, K. et al. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proceedings of the National Academy of Sciences of the United States of America 105, 20870-20875 (2008).
3 Koboldt, D. C. et al. Exome-based mapping and variant prioritization for inherited Mendelian disorders. American journal of human genetics 94, 373-384 (2014).
4 Mitchell, A. A., Chakravarti, A. & Cutler, D. J. On the probability that a novel variant is a disease-causing mutation. Genome research 15, 960-966 (2005).
5 Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and Predicting Haploinsufficiency in the Human Genome. PLoS genetics 6, e1001154 (2010).