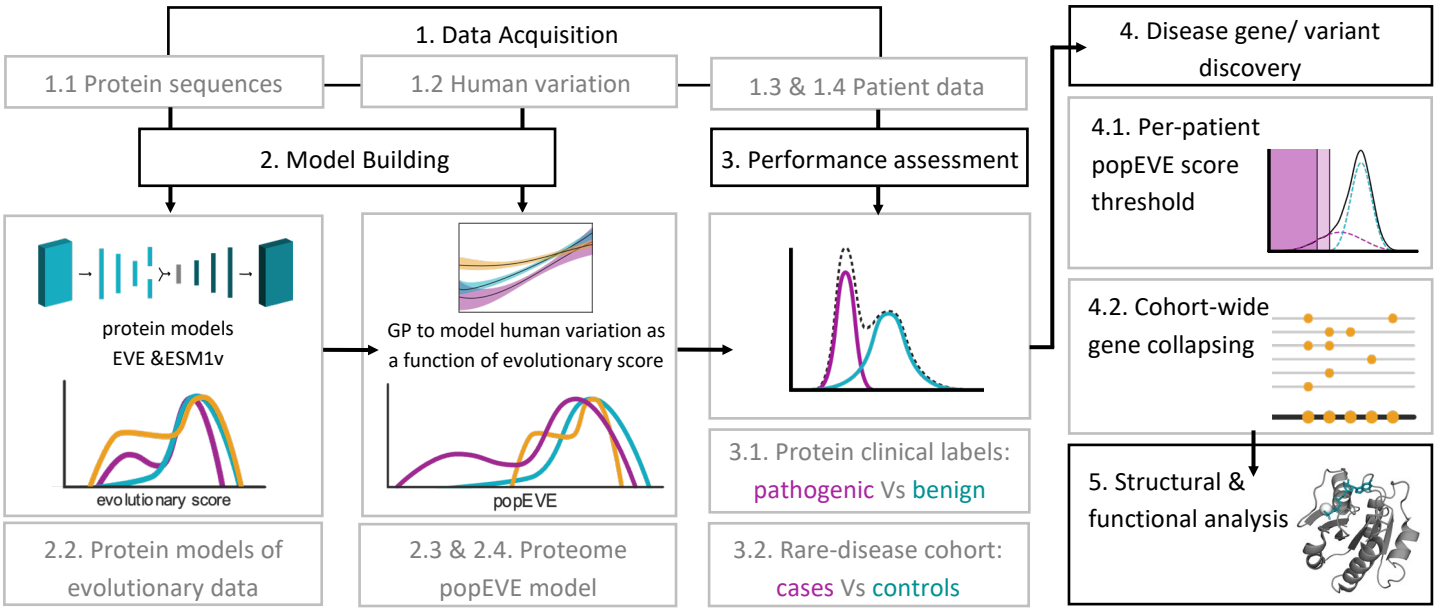
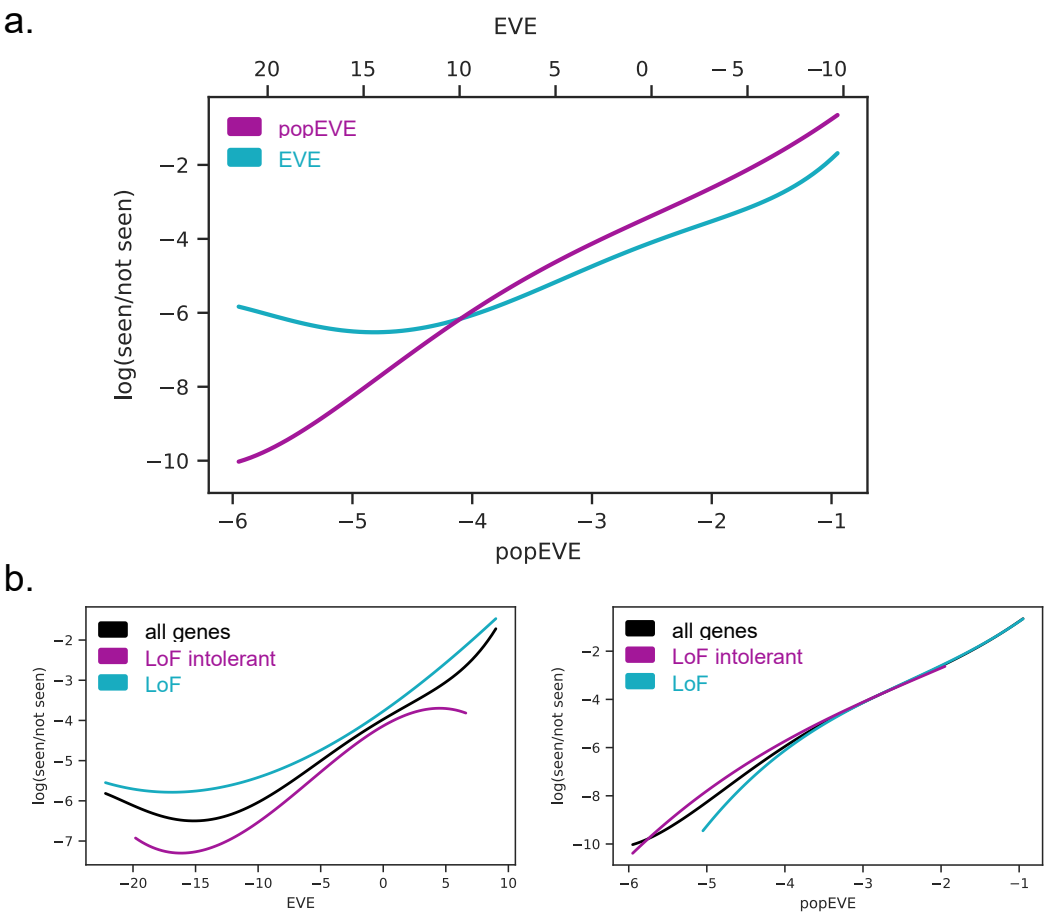


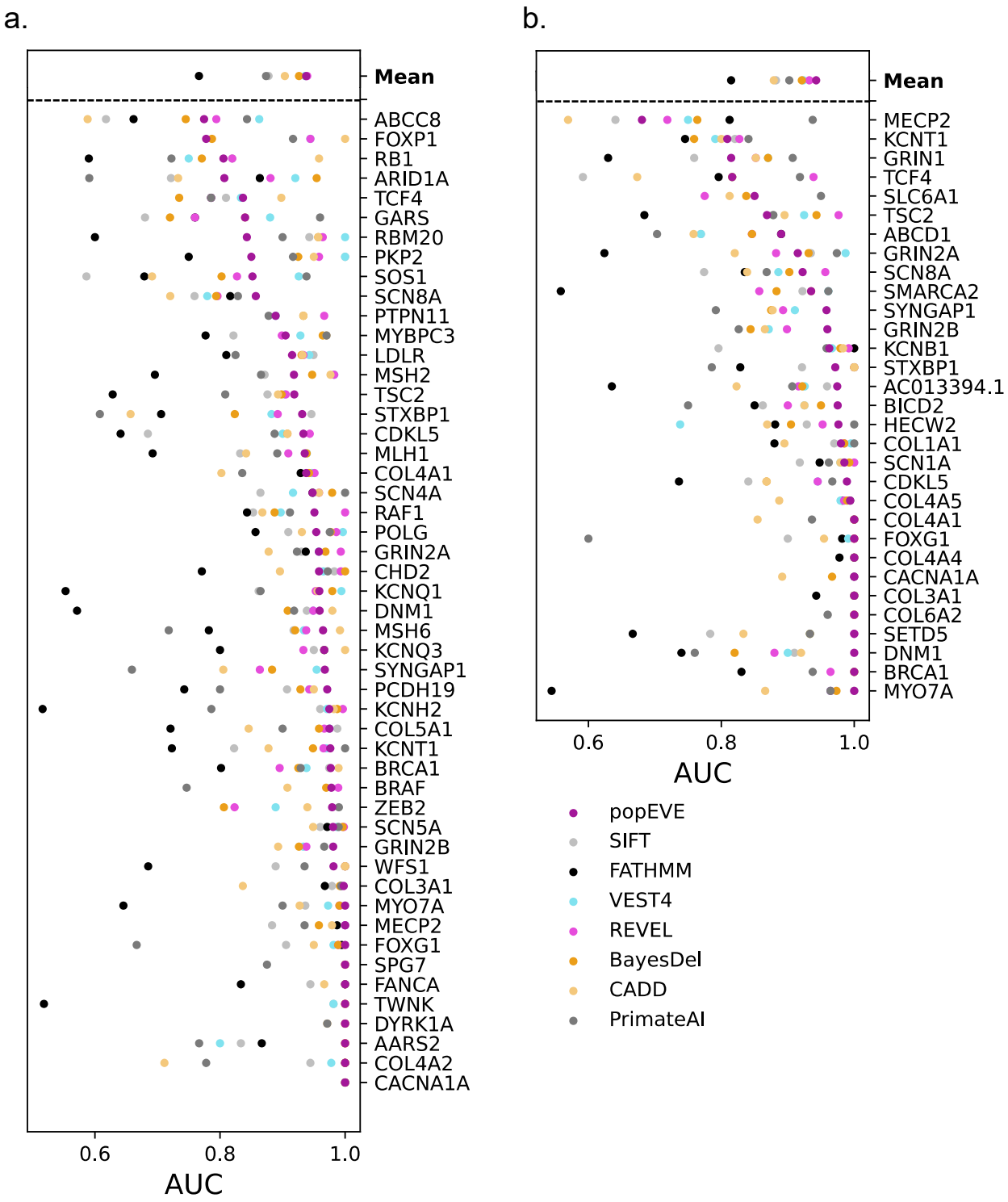
Supplementary Figure 1. Structure of workflow and methods section.

Our workflow for this project involved 1. data acquisition, 2. model building, 3. validation and performance assessment, 4. analysis of patient data towards discovery of disease variants and genes, and 5. structural and functional analysis of candidate disease variants. Model building has two main components: modeling of protein sequences across the entire tree of life from which we obtain a protein level fitness score and predicting the presence of a variant in the UKbiobank given its fitness score using Gaussian Processes, from which we obtain a score for the spectrum of pathogenicity across the human proteome, the popEVE score. We assess the performance of popEVE at predicting pathogenicity within proteins – by predicting benign or pathogenic variants from the ClinVar database and correlation with deep mutational scanning assays – and at predicting a spectrum of pathogenicity across the proteome – by identifying patients with severe developmental disorders amongst controls based on de novo variants. A detailed analysis of severe developmental disorders cohorts was performed with the objective of identifying putative disease-causing variants or genes. Our strategy was two pronged. On the one hand, identify pathogenic variants based on a popEVE score threshold; this strategy can be applied to any patient without the need for a disease-cohort. And on the other hand, take advantage of the cohort structure and identify disease causing genes by performing a burden test of pathogenic variants at the gene level across the cohort. The results of both approaches were compared to state-of-art approaches and putative variants were explored from a structural and functional perspective.



Supplementary Figure 2. Relation between evolutionary score and presence in UKbiobank.

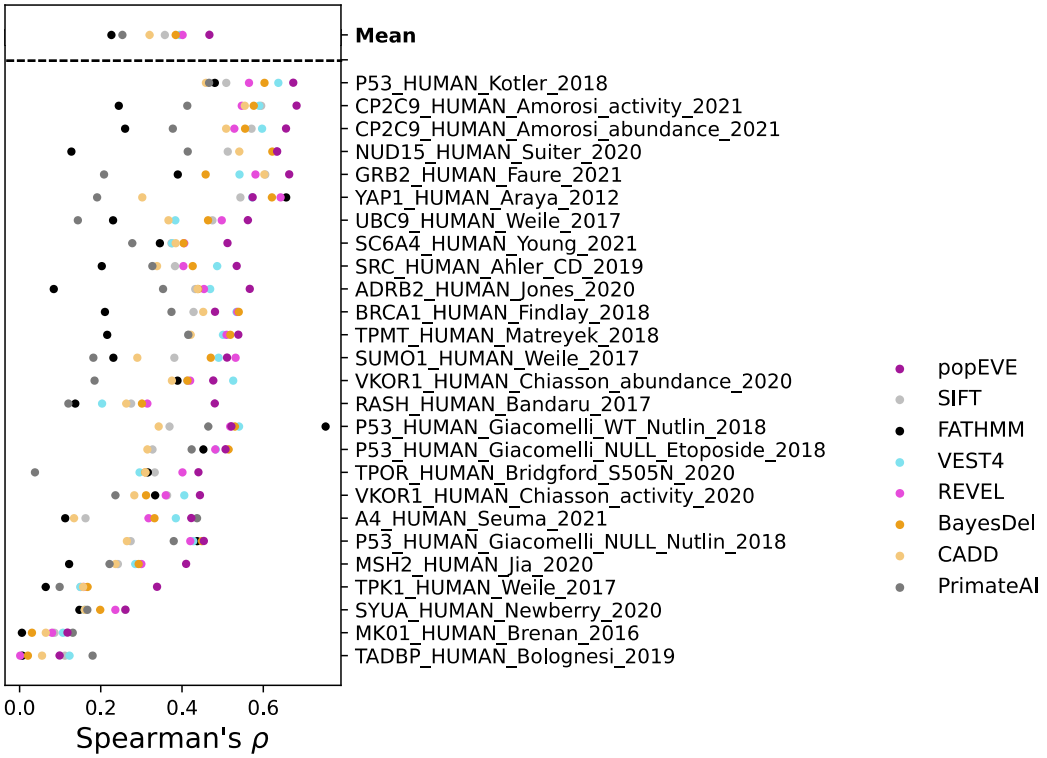
a. both binned EVE and popEVE variant scores have a linear relationship with the log odds that a variant is seen in the Ukbiobank population – however, EVE shows non-monotonicity at the less fit end of the spectrum. b. Before population adjustment, the lowest fitness variant scores for LoF tolerant genes are more deleterious than those of LoF intolerant genes, not unexpected as these models are not necessarily comparable across proteins (left). After adjustment, low fitness variants in LoF intolerant genes are now more deleterious than those in LoF tolerant genes, in line with expectation (right).



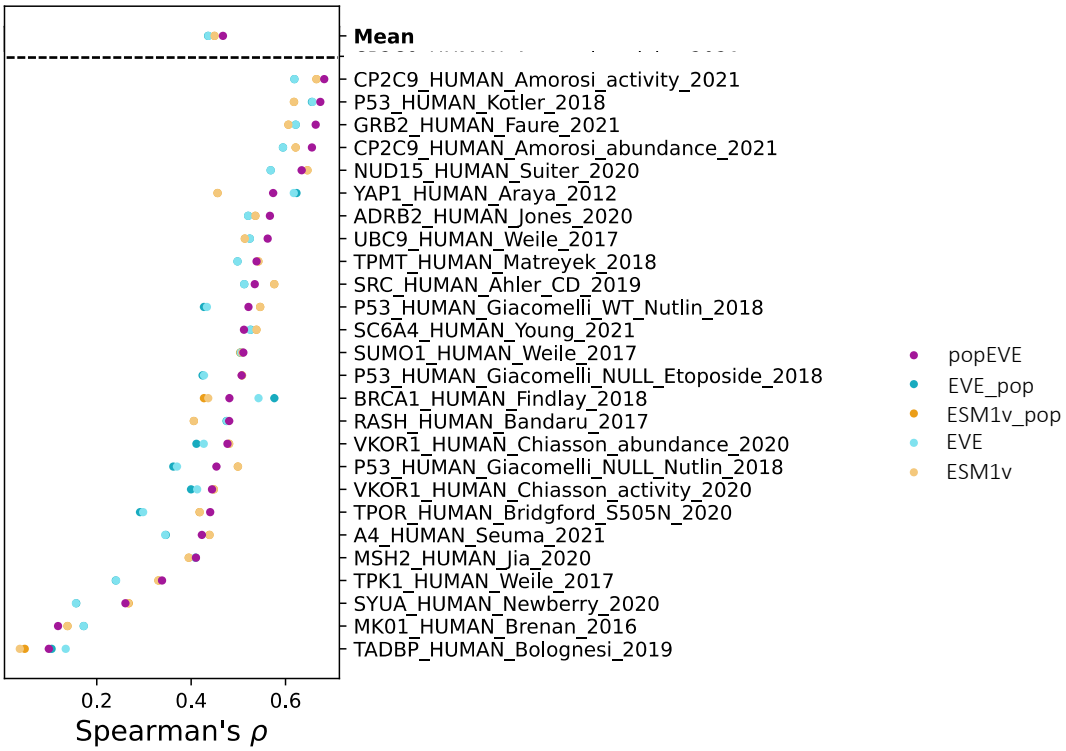
Supplementary Figure 3. Performance summary for separating Benign/Likely Benign from Pathogenic/Likely Pathogenic ClinVar labels.

Assessing the performance of popEVE and popular supervised and unsupervised variant effect prediction models on individual genes that have at least 5 benign and 5 pathogenic variants from the ClinGen curation of a. ClinVar 2019 and b. ClinVar 2020. The ClinGen dataset attempts to address data leakage in the estimation of performance of supervised methods by removing ClinVar variants used in training.

a.

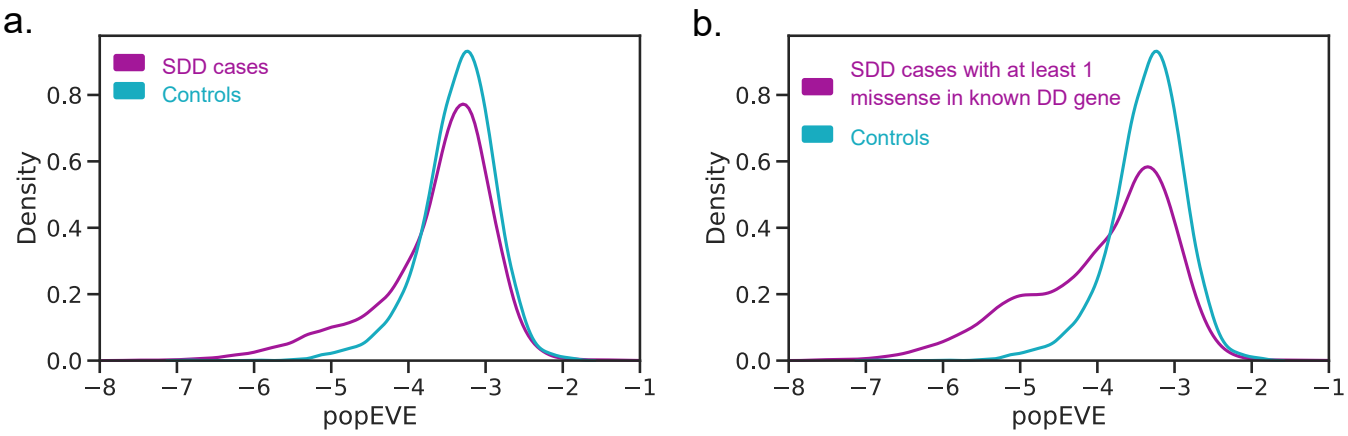


b.



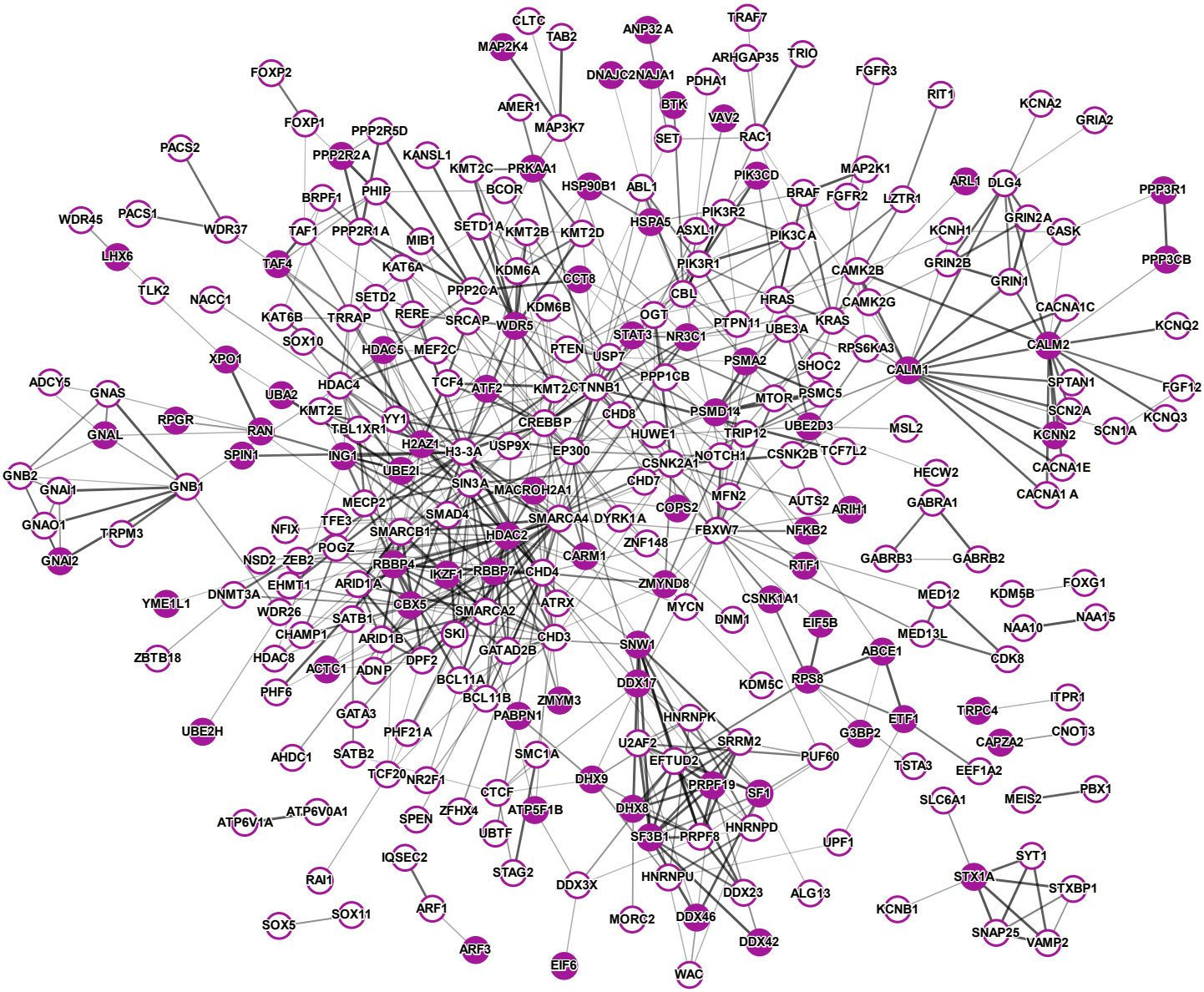
Supplementary Figure 4. Correlation of computational variant effect predicting models with high-throughput experimental assays.

a. Assessing the performance of popEVE compared to popular supervised and unsupervised variant effect prediction models when compared to high-throughput functional assays. On average popEVE outperforms other models. b. Assessing the performance of popEVE and its constituent models when compared to high-throughput functional assays. The suffix "_pop" refers to the model after rescaling via Gaussian Process to model UKbiobank variation.



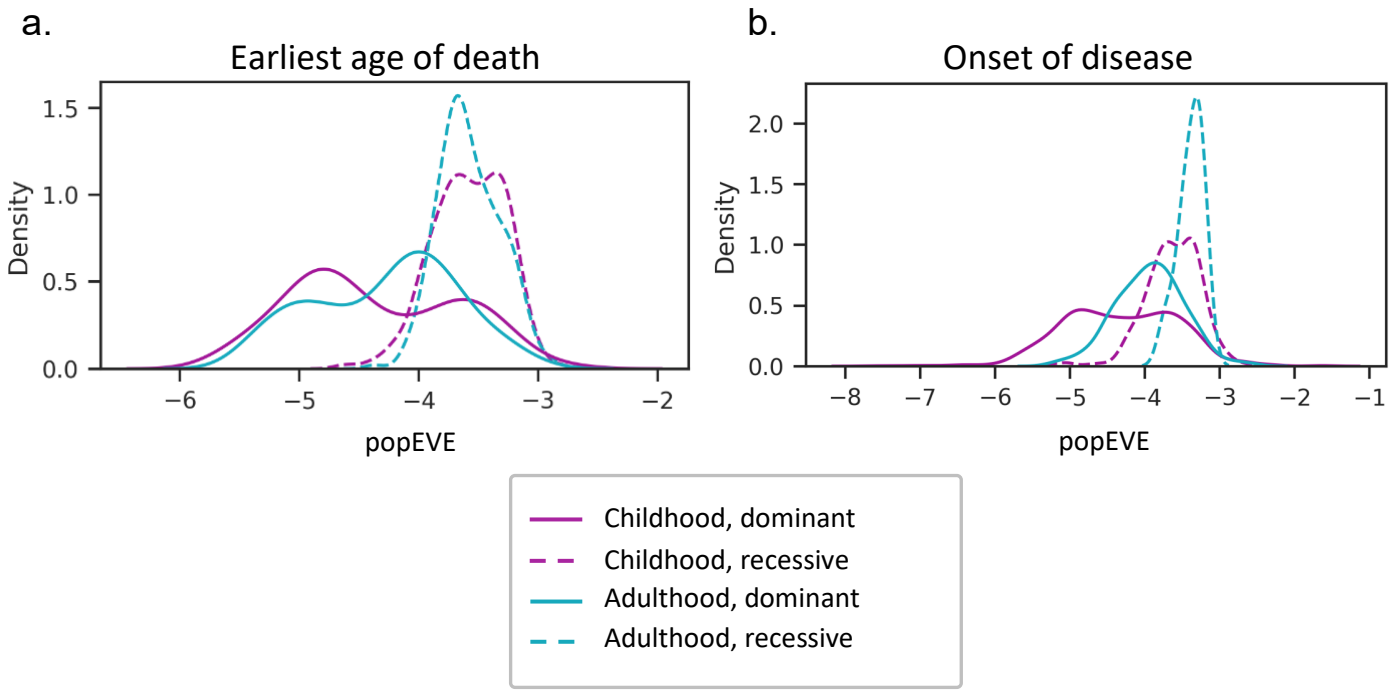
Supplementary Figure 5. The tail end of de novo missense variant popEVE scores in cases is more pathogenic than in controls.

a. The distribution of popEVE scores for de novo missense variants in severe developmental disorder cases and in unaffected controls. b. Cases were filtered for individuals with at least one de novo missense variant in a known developmental disorder-associated gene to construct a test set of cases with likely causal variants for our precision-recall analysis.



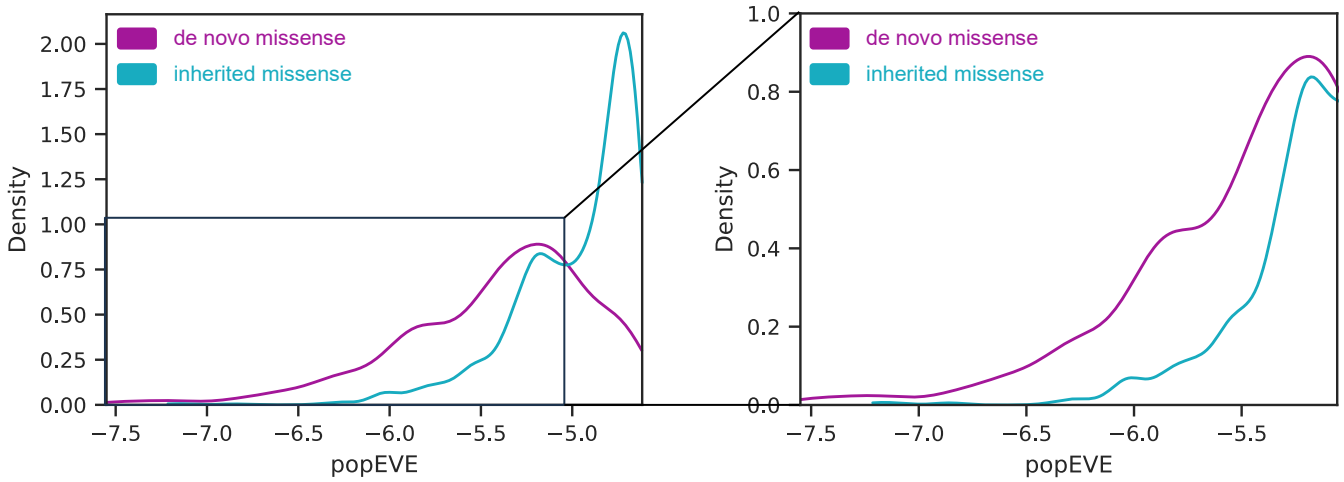
Supplementary Figure 6. Functional network of previously known genes and newly discovered genes (extension of Fig. 3a top).

Novel popEVE discovered genes are embedded into the network of genes previously discovered in the severe developmental disorder cohorts. Taking the set of 99.99 confidence threshold popEVE genes, we built a network using STRINGdb ('experiments' and 'coexpression' at a medium 0.4 score threshold). Colored nodes are novel discoveries and white nodes are known developmental disorder-associated genes.



Supplementary Figure 7. Scores of ClinVar pathogenic variants in genes associated with premature onset of disease and death, for both dominant and recessive inheritance.

Each gene with pathogenic variants in ClinVar was associated with an OMIM phenotype and corresponding mode of inheritance from OrphaNet. Scores of pathogenic ClinVar variants are plotted for genes associated with premature childhood or adulthood death (a) and onset of disease (b). Solid lines represent dominant phenotypes whereas dotted represent recessive; magenta refers to childhood and teal to adulthood. ClinVar pathogenicity spans a range of popEVE scores and severities. There is an enrichment of variants with low popEVE scores in genes associated with dominant childhood death and age of onset of disease, compared to the adulthood and recessive counterparts. Furthermore, the model expects that not all pathogenic variants in genes associated with dominant premature death and age of disease onset lead to equally severe phenotypes – indeed one would expect some variants classified pathogenic to not lead to the most severe outcome, as is suggested by the model.



Supplementary Figure 8. De novo missense variants in the DDD cohort are generally more pathogenic than inherited missense variants according to popEVE.

popEVE scores of de novo missense variants seen in the DDD patients are on average more pathogenic than missense variants inherited from their parents. The bulk of popEVE pathogenic inherited variants fall in the moderately pathogenic category (left) while the de novo variants tend to fall in the severely pathogenic range. Additionally, the de novo variants show a long tail of pathogenicity (right).