**Prioritizing nonsense mutation with ALoFT (Pubmed Id: 28851873)**

ALoFT (annotation of loss-of-function transcripts), a method to annotate and predict the disease-causing potential of loss-of-function variants. ALoFT applies a machine learning framework where key features are associated with functional domain annotation, evolutionary conservation, and biological network annotation. We apply data from the Mendelian disease-gene discovery projects to show that ALoFT can distinguish between loss-of-function variants that are deleterious as heterozygotes and those causing disease only in the homozygous state.

**Quantifying localized frustration to prioritize missense mutations (Pubmed Id:27915290)**

We developed a new framework to evaluate the impact and prioritize missense mutations in various rare disease. In this approach localized frustration, quantifying unfavorable local interactions is employed as a metric to investigate the effect of missense mutations. This is important as previous studies have shown that local perturbations induced by missense mutations can severely impact protein functionality without strongly disrupting global stability (e.g. in relation to catalysis or allostery).

**A comprehensive resource of upstream open reading frames(uORFs) (Pubmed Id:29562350)**

Upstream open reading frames (uORFs) latent in mRNA transcripts are thought to modify translation of coding sequences by altering ribosome activity. To estimate the impact of uORFs on the regulation of translation in humans, we first circumscribed the universe of all possible uORFs based on coding gene sequence motifs and identified 1.3 million unique uORFs. To determine which of these are likely to be biologically relevant, we built a simple Bayesian classifier using 89 attributes of uORFs labeled as active in ribosome profiling experiments. This allowed us to extrapolate to a comprehensive catalog of likely functional uORFs. Our ranked list of likely active uORFs may be used to screen non-coding variants associated with rare diseases for their effect on uORFs. This information may then be used to infer the effect of non-coding variants on gene translation as mediated by uORF activity.

**Assessing the influence of missense mutations in drug binding affinity**

A key issue in drug design is understanding how rare disease variant affects drug efficacy by altering binding affinity (BA) in different individuals -- an important consideration for pharmaceutical regulators. Ideally, We take a hybrid approach using physically-based calculations to bootstrap the parameterization of a full statistical model. In particular, we do 3D-structure-based docking calculations on ~10,000 SNVs modifying known protein-drug complexes to construct a pseudo-gold-standard dataset of BAs. Then we develop a complete statistical model combining structure, ligand and sequence features and show how it can be applied to score millions of SNVs. This tools can be applied to assess the influence of Mendelian variants on drug binding affinity to determine the efficacy of potential drug targets for Mendelian diseases.