RESPONSE LETTER

Reviewer #2

-- Ref 2.0 – additional details on deltaF-thershold --

Reviewer Comment	I suggest to add some information concerning the "deltaF-threshold" which was used to discriminate deleterious from benign (as deduced from deltaF value) variants in the SIFT/Polyphen-2 complementing analyis to the main text. Is it -1.221, as explained in supplemental information? Or any other value? This should be mentioned in the main text, otherwise the ready cannot really follow what you did. There might also be an additional methods section on this analysis in the supplement.
Author Response	We would first like to thank the reviewer for providing further valuable suggestions on how we may improve this work. In the previous version of the manuscript, we provided the "deltaF-threshold" information in the method section. However, following the reviewer's suggestion, we explicitly mention this cut-off in the result section of the updated manuscript as well. Regarding additional supplementary method section for the SIFT/Polyphen-2 complementing analysis, we already provide necessary information (selection of PDB subset for the analysis and deltaF-cutoff selection method) in the method and supplement section of the current manuscript. Thus, we think additional details will be redundant here.
Excerpt From Revised Manuscript	*Excerpt from Results:* For the frustration metric, we applied ΔF threshold of -1.221 (see method for detail) to distinguish between benign and deleterious variants.

-- Ref 2.1 –Additional figure--

Reviewer Comment	I would also suggest to add supplemental figure S1 to the main article, since it gives a good overview of used data. Instead, Figure 2 and/or 6 could go to the supplement (if you have too many figures).
Author Response	We concur with reviewer's suggestion and now include Figure S1 as a main figure.

-- Ref 2.2 – Table caption and rare/common variants--

Reviewer Comment	I am a bit surprised that there are more "rare" than "common" and more conserved than variable SNVs in the 1KG and ExAC data set(s), since intuition would tell me that it should be the other way round (since these SNVs are present in healthy human populations, and as you said, 1KG and ExAC "are highly enriched in benign SNVs".). Maybe it would help to have your definitions of rare/common (MAF?) and conserved/variable (specific GERP score?) directly in the table caption.
Author Response	We update the caption of table 1 to explicitly state the MAF and GERP cutoff values distinguishing rare/common SNVs as well as conserved/variable datasets. We would also like to point out that we are evaluating the impact of only non-synonymous SNVs in the 1KG and ExAC datasets, which map to protein structure. This primarily drives the disparity in frequency of rare/common and conserved/variable SNV datasets, which reviewer is alluding to.
Excerpt From Revised Manuscript	**Table 1. Summary statistics on the number of SNVs used in comparative analyses.** This table shows variant counts for non-disease (top), HGMD (bottom-left), and pan-cancer SNVs (bottom-right). Variants were further classified as rare (MAF <= 0.5%), common (MAF > 0.5%) , conserved (GERP > 2.0) and variable (GERP <= 2.0).

-- Ref 2.3 –Schematic Figure description--

Reviewer Comment	I also still don't get Fig.1 (although I principally like it!). According to methods text and figure capture, TRP was changed to TYR: "Shown here is the result of changing residue ID 31 in plastocyanin (pdb ID 3CVD) from the wild6type residue (TRP) to a mutated residue (TYR)". These two amino acids are also highlighted / differently colored in the figure. However, the sequence context of those two highlighted amino acids is not the same. If there were only this one amino acid exchange, shouldn't the rest of the illustrated sequence be identical? Or is the illustrated sequence of amino acids not the "real" amino acid sequence but a somehow linearized spatial configuration / structural order of the amino acids, as they appear after folding to secondary and tertiary protein structure? The figure might be easier to understand if the residues were numbered (as I suggested already before).
Author Response	We thank the reviewer for bringing this point. Unfortunately, reviewer is confusing amino acids represented vertically as sequence/structural context, which is incorrect. The vertical line in the schematic should be considered more like an energy level description of protein. Each level on this energy scale corresponds to the total energy value of the protein, if the residue position (residue ID 31) was occupied by distinct amino acids. The total energy is determined by an empirical energy term, which depends on residue identity. Note that we do not perform any structural modeling for this calculation. The left vertical line represents residues based on their energy values in the native structure

	of the protein. In contrast, the right hand vertical line corresponds to the energy level based on the modeled protein structure, where wild type TRP residue was mutated to TYR using homology-modelling. We employ this modeled structure as template to further determine the energy level description of the modeled structure on the right vertical line.

-- Ref 2.4 – Neutral terms for variants --

Reviewer Comment	I would suggest to use a neutral term for variants of not further specified clinical significance, regardless whether they are rare or common. Neutral terms are "variant", "variation", "base exchange" etc. The term "mutation" should be avoided when the clinical significance of a variant is unknown or unspecified, since it is often (mis-)understood as a variant which causes disease. Example sentence, where "mutation" should be replaced by neutral term: "Furthermore, we investigated the differential influence of common and rare mutations, where SNVs with minor allele frequencies (MAF) less than or equal to 0.5% were considered to be rare mutations."
Author Response	We agree with reviewer's suggestion for using neutral term in context of variants with unknown clinical significance. We update the manuscript accordingly.
Excerpt From Revised Manuscript	

-- Ref 2.5 – Disease-associated term --

Reviewer Comment	The term "disease-associated" should be used with care in order to avoid confusion between disease-association and disease-causality. There is a dedicated method called association study, which strives to detect an association between genetic variants and a certain (mostly complex) disease, where associated variants are not necessarily causative. In contrast to this, disease-causing variants are not only statistically associated with a disease but have been shown to be causative for it, which has to be distinguished from disease-association. Therefore, some sentences should be rewritten, for example in the abstract: "disease-associated SNVs create stronger changes in localized frustration than non-disease associated variants" and in the main text "We also examined the local perturbations induced by disease-associated and benign SNVs originating in conserved and variable regions of the genome." and "[...] wherein we analyzed KF distributions for HGMD variants (disease-associated)[...]" - please check for further occurences, also in the supplemental / supporting information. There should be clarity about the

	difference between disease-association and disease-causality in your manuscript. This avoids confusion on side of your readers.
Author Response	**[[Not sure what to use instead of disease-associated.]]**
Excerpt From Revised Manuscript	

-- Ref 2.6 – GERP and DAF abbreviation --

Reviewer Comment	Should be fully spelled at least once somewhere in the manuscript. Do you mean GERP = Genome Evolutionary Rate Profiling and DAF = derived allele frequency?
Author Response	We thank the reviewer for pointing out this issue. They have now been corrected.
Excerpt From Revised Manuscript	"The distinction between conserved and variable regions were defined using genome evolutionary rate profiling(GERP) scores"

-- Ref 2.7 – definition of rare/common in table caption –

Reviewer Comment	Thresholds for your definition of rare/common (MAF?) should appear in the table caption.
Author Response	We have updated the table caption to include this definition.
Excerpt From Revised Manuscript	**Table 1. Summary statistics on the number of SNVs used in comparative analyses.** This table shows variant counts for non-disease (top), HGMD (bottom-left), and pan-cancer SNVs (bottom-right). Variants were further classified as rare (MAF <= 0.5%), common (MAF > 0.5%) , conserved (GERP > 2.0) and variable (GERP <= 2.0).