Pathogenic Variants of a Visual Protein Constrain Mutational Pathways in Genotype Space

Steven Chen[1], BSc, Jing Liu[1], BSc, Alexander Van Nynatten[2], PhD, Belinda Chang[3], PhD,

[1]Department of Cell and Systems Biology, University of Toronto
[2]Department of Biological Sciences, University of Toronto, Scarborough
[3]Department of Cell and Systems Biology and Department of Ecology and Evolutionary Biology, University of Toronto

Introduction: A significant effort in biomedical research has been to identify mutations responsible for diseases from leukemia to visual degeneration. Yet, even characterizing all possible single mutations underlying a disease is insufficient for understanding its progression. How does pathogenicity arise as mutations accumulate in the soma throughout life or in the germline over generations? This question has remained elusive, even for studies focused on single genes, given the immense space of possible mutational combinations and trajectories that need to be tested. Previous attempts at mapping genotype-phenotype-fitness landscapes have focused primarily on adaptive mutational trajectories inferred from genome sequence data. However, disease-causing mutations are difficult to detect and are not amenable to such approaches, as they are often non-adaptive, transient, and rarely establish at high frequencies in the population.

Methods: To investigate how disease phenotypes may arise from vast mutational space, we developed a high-throughput platform in yeast capable of assaying two crucial aspects of protein function for the visual protein rhodopsin: light-dependent activation and stability. We achieved this by engineering a fluorescence-based transcriptional reporter for measuring light-dependent rhodopsin activation, and a separate fluorescent protein fusion with distinct absorbance and emission spectra for measuring rhodopsin stability.

Results: Using this platform, together with fluorescence-activated cell sorting and ultradeep sequencing, we screened a combinatorial mutation library encoding all possible combinations of amino acid substitutions at four sites in rhodopsin representing 160,000 protein variants and discovered hundreds that neither compromise rhodopsin stability nor light activatability. With such a large volume of data, we applied systems-level analyses to generate a map of protein variant effects, revealing that ~4.6 percent of variants are functional with the remaining variants compromising at least one aspect of function.

Conclusion: This result suggests that although genotype space is dominated by non-functional pathogenic variants, a small but substantial proportion of variants are functionally permissive and likely constitute viable mutational pathways. Together our platform enables the high-throughput discovery of protein variant effects for visual proteins and is demonstrated to be capable of identifying non-functional, and potentially pathogenic, variants that constrain viable mutational pathways in genotype space.

.