Learning about proteins from sequencing studies

S. Sunyaev
Harvard University

Sequencing studies of large population cohorts discover myriads of amino acid variants segregating in the human population. This generates a need in functional and structural interpretation of these variants. At the same time, analysis of natural selection and phenotypic effects of human alleles has a potential to generate new knowledge about proteins. It allows mapping structural properties of amino acid variants onto their genetic effects. At the next level, sequencing studies may inform us about genetic interactions between missense variants. We found that an appreciable fraction of apparently fully penetrant disease-causing human alleles are present in the genomes of other mammals, suggesting a role for genomic context. We developed a model of genetic interactions that predicts most of these to be simple pairwise compensations. Functional testing of this model on two known human disease genes revealed discrete cis amino acid residues that, although benign on their own, could rescue the human mutations in vivo. This approach was also applied to ab initio gene discovery to support the identification of a de novo disease driver in BTG2 that is subject to protectivecis-modification in >50 species.