Aenne Solvejg Thormählen
Dr. med.

The contribution of rare genetic variants to the heritability of myocardial infarction: Systematic functional characterization of LDL-Receptor missense alleles identified through exome sequencing of large clinical cohorts

Fach/Einrichtung: Humangenetik
Doktorvater: Priv.-Doz. Dr. med. Heiko Runz

Exome sequencing has proven powerful to identify protein-coding genetic variation across the human genome and unravel the basis of monogenic diseases. It has recently been shown that exome sequencing is further suited to discover rare alleles that confer risk for complex disease. Nevertheless, successful application of exome sequencing to complex phenotypes faces two key challenges: First, most alleles identified in a population are extremely rare. And second, most alleles are neutral on protein activities. Consequently, association tests that rely on enumerating rare alleles in cases and controls (termed rare variant association studies, RVAS) are typically underpowered, as the many neutral alleles dampen association signals that arise from the few alleles that disrupt protein functions. It is believed that RVAS would profit considerably if only clearly disruptive variants were considered for association testing. However, strategies to securely discriminate disruptive from neutral variants are immature, particularly for missense variants whose function cannot be unambiguously delineated from sequence.

This study shows that the statistical power of RVAS improves dramatically if variants are stratified according to their *in vitro* ascertained functions. It establishes novel technology to profile the biological effects of missense variants objectively and in a scalable manner. At the example of the low-density lipoprotein receptor (*LDLR*) – a Mendelian disease gene and one of yet few genes in which rare variants associate with complex disease - odds ratios of rare *LDLR* allele carriers for high plasma LDL- cholesterol (LDL-C) are refined from 4.5 to 25.3, and from 2.1 to 20.0 for early-onset myocardial infarction (MI). Importantly, these results provide proof-of-concept that combining sequencing with biological studies may make it possible to reduce RVAS sample sizes by more than 2-fold.

Overall, exomes of 1,716 cases with MI and 1,519 MI-free controls were leveraged. Furthermore, exome chip datasets from close to 40,000 individuals characterized for LDL-C were analysed. Biological functions of the *LDLR* missense alleles identified in these cohorts were experimentally characterized in an unbiased and quantitative manner through systematic overexpression and complementation experiments with mutated GFP-labelled cDNAs in cells. Interestingly, carriers of *LDLR* alleles, which experiments identified as “disruptive-missense”, had higher plasma LDL-C. Additionally, considering *in vitro* data made the association of *LDLR* with LDL-C and MI considerably more robust.

This study addresses a fundamental challenge to modern genetics (to distinguish disruptive from neutral alleles) by applying novel technology (systematic cell-based functional characterization of rare protein-coding variants) to a significant unmet health need (hypercholesterolemia, MI). This innovative approach to extract function from sequence is scalable and contributes to overcome crucial barriers in the interpretation of protein-coding genetic variation. This combined sequencing-biological study is one of the first to address this fundamental challenge of current genetics and will be a lead example for any geneticist dealing with rare variants.