Investigating the genetic basis of meiotic recombination in large biobanks

Julia Sidorenko[1], Loic Yengo[1], Peter M. Visscher[1,2]

[1] Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia
[2] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Background. Meiotic recombination is a fundamental biological process that generates and maintain genetic variation within populations. This process is tightly regulated such that both too low and too high recombination rates (RR) can have detrimental effects on both cell fate and later-life events. In humans, RR is shown to be moderately (up to 30%) heritable (Fledel-Alon et al., 2011) and, recently, multiple RR-associated genomic loci have been mapped (Halldorsson et al., 2019). That includes well characterized *RNF212* and *MSH4* genes, which are shared across multiple species. However, the studies of the genetic control of RR associated phenotypes have been largely impeded by the availability of suitable data (e.g., genotype data of parents and multiple offspring or embryos).

Methods & Results. Here we aim to leverage the increasingly available large population level biobanks to increase the available sample size to study meiotic RR even in the absence of extensive families. We show that the average familial RR and locations of crossovers can be inferred from the patterns of identity-by-descent (IBD) sharing between full siblings. Moreover, the genotypes of the offspring capture 2/3 of the information of the genotypes of their parents and thus can inform on genetic factors influencing the parental RR. Using 19,454 full-sib pairs in UK Biobank we perform a GWAS of familial RR on mean sibling genotype and replicate four out of six genome-wide significant loci for sex-combined RR identified in the large Icelandic study of ~60,000 individuals (Halldorsson et al., 2019). Moreover, we demonstrate that such GWAS can be simply performed on observed (not average) sibling genotypes, and thus can be easily applied to large biobank data.

Conclusions. Overall, our study provides an efficient framework to study familial RR in biobank setting and replicates existing associations obtained using individual level phenotype and genotype data. We are currently extending these analyses to additional biobanks and ancestries and expect to reach a sample size of approximately 150,000 families to uncover new insights into genetic control of meiotic recombination rate.