

Longitudinal Genetic Modelling: Revisiting Associations of SNPs Associated With Blood Fasting Glucose in Normoglycemic Individuals

Mickaël Canouil¹, Ghislain Rocheleau¹, Loïc Yengo¹, Philippe Froguel^{1,2}

¹Univ. Lille, CNRS, CHU Lille, Institut Pasteur de Lille, UMR 8199 - EGID, F-59000 Lille, France

²Department of Genomics of Common Disease, Imperial College London, London, United Kingdom

New statistical methods need to be proposed as an alternative to the current cross-sectional design predominantly used in genome-wide association studies (GWAS). When longitudinal (repeated) measures of a trait are available, an efficient modelling of the temporal trajectories is expected to increase statistical power to detect genetic loci associated with that trait.

Using genotypes assayed with the Metabochip DNA arrays (Illumina) from 4,500 subjects recruited in the French cohort D.E.S.I.R. (Données Épidémiologiques sur le Syndrome d'Insulino-Résistance), we re-examine published GWAS findings for some confirmed loci associated with fasting plasma glucose (FPG).

We compared several approaches to test the SNP main effect, on the one hand, and to test the interaction SNP-by-time effect, on the other hand. For the former, we compared five methods: linear regression models using only baseline measures or using the average of measures across all time-points, Two-Step approach with random intercept, Generalised Estimating Equations (GEE) and Linear Mixed Model (LMM); while for the latter we compared Two-Step approach with random slope, Conditional Two-Step, GEE and LMM with interaction term.

Type I error and power were computed using permutations and resampling procedures on the full dataset for the SNP effect, and using numerical simulations for the interaction effect. Across all models tested, the type I error was not inflated. In contrast, power analysis sometimes showed an increased statistical power for the baseline approach compared to methods dealing with repeated measures of FPG. We provide mathematical conditions showing why this counterintuitive situation might happen.

In the context of large GWAS with millions of imputed SNPs and when repeated measures are available for exploration, implementing methods which approximate a full longitudinal model seems at present the most efficient and fastest way to identify genetic associations without major loss in power. More importantly, these approximate methods run much faster than the full modelling approaches like GEE or LMM and could help picking the most associated SNPs for further testing in full models.