Research Proposal Summary (3500 words including spaces)

Early determinants of cognitive behavioural development in the Canadian Healthy Infant Longitudinal Development (CHILD) study

Rationale: Genome wide association studies (GWAS) have identified numerous loci associated with complex diseases in the last 10 years. The vast majority of GWAS associated loci explain the genetic heritability of traits in Caucasian populations[1]. However, these are often not replicated in more diverse populations due to confounding factors or pleotropic effects. Hence, it is important to understand the genetic correlation of quantitative phenotypes to explain the complex behavioural traits in multiethnic cohort.

Hypothesis: Early genetic determinants of cognitive behavioural development in children as young as 5 years, correlates with psychiatric outcomes later in adulthood.

Aim 1: To perform GWAS of behavioural traits in children as early as 5 years. **Aim 2**: To investigate the burden of variants in the genes associated with cognitive behaviour in children at 5 years. **Aim 3:** To determine the genetic correlation of behavioural development with marginal effects of breast-feeding during infancy.

Materials and Methods: Our study will leverage existing genome-wide polymorphism data from 3346 subjects of diverse ethnicities from the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort. A total of 2820 CHILD subjects have available cognitive behaviour assessments (i.e. Child Behaviour Checklist (CBCL) scores for internalizing and externalizing behaviours, which are associated with anxiety and aggression, respectively). In addition, there is a wealth of pre- and post-natal environmental exposures collected from home visits, questionnaires, and nutritional assessments of the children from birth to 5 years of age[2]. **Aim 1:** We will utilize the genomics data to identify genetic factors associated with CBCL scores using a linear mixed model implemented in *BOLT-LMM*[3] tool, which will account for population structure and relatedness. **Aim 2:** We will conduct a gene-based association analysis as implemented in *MAGMA*[4] tool. SNPs will be assigned to genes based on their position according to the hg19 build of the human reference genome. *MAGMA* incorporates rare variants, pathway annotations and gene-environment interactions analysis. **Aim 3:** We will investigate the interaction effects of genes with environmental exposures such as breast-feeding up to 3, 6 and 12 months, medication use by the mothers and infants, nutrition of the mothers, etc. using *MAGMA*.

Importance: Given the availability of longitudinal data in addition to genomics and environmental data in the CHILD study, we have a unique opportunity to investigate the effects of gene-environment interactions on health outcomes Canadian children throughout early childhood. This study will improve our understanding of the underlying mechanisms of multifactorial traits and diseases in a mixed population and identify modifiable risk factors of diseases.

Feasibility: My background in bioinformatics and genomics makes me qualified to lead this study as I already have an experience in developing data analysis pipelines for identification of

genetic variants correlated with multifactorial traits. Moreover, I have been involved in the cleaning of genomics data, imputations and preliminary analysis of several clinical outcomes. Hence, I have a working knowledge of the CHILD study datasets.

References:

1. Ripke, S. *et al*. *Nature* **511** (2014)
2. Subbarao, P. *et al. Thorax* **70** (2015)
3. Loh, P. R. *et al. Nature Genetics* **47** (2015)
4. de Leeuw, C.A. *et al. PLOS Comput. Biol.* **11** (2015)