GROWUPGROWOLD
HUMAN DEVELOPMENT, BIRTH TO DEATH

Moffitt & Caspi
Genes Environment Health Behavior

Concept Paper Form

Provisional Paper Title: **The contribution of specific blood cell types to whole blood DNA methylation profiles: implications for interpreting findings in epigenetic epidemiology**
Proposing Author: Eilis Hannon
Author's Email: E.J.Hannon@exeter.ac.uk
P.I. Sponsor: Jonathan Mill (J.Mill@exeter.ac.uk) (if the proposing author is a student or colleague of an original PI)
Today's Date: 1/23/2020

Please describe your proposal in 2-3 pages with sufficient detail for helpful review.

Objective of the study:

There is increasing interest in the role of epigenetic variation in health and disease, with the primary focus of epigenetic epidemiology being DNA methylation (DNAm)[1]. A critical issue for epigenome-wide association study (EWAS) analyses is the fact that, unlike germline genetic variation, DNAm signatures are tissue- and cell type-specific[2] and the selection of tissue type for epigenetic profiling has implications for any conclusions made from these studies. For practical reasons, most EWAS have been performed using DNA isolated from easily-accessible peripheral tissues (e.g. whole blood or buccal epithelial tissue); although these may not be the primary tissue-/cell type relevant to the phenotype under study. Therefore, where associations have been reported from analyses of whole blood, it is unclear which specific blood cell type is affected. Our understanding of how variation within individual blood cell types contributes to variability in whole blood DNAm profiles is limited by the lack of data available on purified cell types from multiple individuals. To our knowledge, no group has assessed purified blood cell-types using the Illumina EPIC array, which is now the gold-standard platform for EWAS.

In this study use genome-wide DNAm data profiled in DNA isolated from buccal epithelial, nasal epithelial, whole blood and five major blood cell types (monocytes, granulocytes, CD4 T-cells, CD8 T-cells and B-cells) from 30 individuals (15 twin pairs). With these data we will characterize patterns of covariation between peripheral tissues and major blood cell types to identify sites where variation identified in whole blood is attributed to variation within a single blood cell type.

Data analysis methods:

After performing our standard, stringent quality control pipeline on the raw DNA methylation data we will perform a series of analysis to characterize the differences between and co-variation between peripheral tissues and major blood cell types.

1) To identify DNA methylation sites with significant different DNA methylation means and/or variances between tissues and cell types we will use an ANOVA and Levine's test at each site.
2) To quantify the level of covariation between each sample type and whole blood we will calculate Pearson's correlation coefficients across matched samples; the values will be squared and multiplied by 100 to obtain the percentage of variance explained for each site.
3) To quantify the proportion of variance in whole blood explained by the five major blood cell types, we will fit a linear model with whole blood DNAm as the outcome predicted by the DNAm level of each of the five cell types included as covariates extracting the r^2 value of the full model.
4) To identify which cell types are influencing the variation observed in whole we will calculate characteristic scores. First, DNAm values at all autosomal probes will be adjusted for differences in mean level of DNAm between blood cell types, by taking the residuals from a linear model where DNAm was regressed against cell type. Then, characteristic scores for each DNAm site and cell type will be calculated by fitting a one-sided Levene's test comparing the variation of a single cell type against the variation across all samples from the other four cell types specifically testing for a larger variance in that cell type.

Variables needed at which ages:

No phenotypic variable required – for this study all we require is knowing which individual each sample type originates from.

DNAm data has been generated for purified cell-types obtained from the 30 samples recruited for the American Asthma Foundation E-Risk sub-project run by Jonathan Mill

Significance of the Study (for theory, research methods or clinical practice):

The results of this study will provide important insights for the interpretation of EWAS performed in whole blood, including both existing and future studies. It will clarify how cell-specific DNA methylation profiles aggregate into whole profiles, and determine for each DNA methylation site on the EPIC array which cell-types are driving the variation.

References cited:

1. Murphy TM, Mill J. Epigenetics in health and disease: heralding the EWAS era. *Lancet* 2014; **383**: 1952-4.
2. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014; **15**: R31.

Data Security Agreement

Provisional Paper Title: The contribution of specific blood cell types to whole blood DNA methylation profiles: implications for interpreting findings in epigenetic epidemiology
Proposing Author: Eilis Hannon
Today's Date: 1/23/2020

☒	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
☒	My project is covered by the Duke ethics committee OR I have /will obtain ethical approval from my home institution.
☒	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
☒	I will not "sync" the data to a mobile device.
☒	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Moffitt or Caspi.
☒	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
☒	I will not post data online or submit the data file to a journal for them to post. *Some journals are now requesting the data file as part of the manuscript submission process. Study participants have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to Temi or Avshalom for strategies for achieving compliance with data-sharing policies of journals.*
☒	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. This data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.

Signature: