E-Risk Study Concept Paper template

Provisional Paper Title: Identifying cell-type interacting associations between DNA methylation and atopic dermatitis
Proposing Author: Josine Min
Author's Email: josine.min@bristol.ac.uk
Academic supervisor: Click or tap here to enter text. (if the proposing author is a student)
E-Risk Sponsor: Chloe Wong (if the proposing author is not an E-Risk co-investigator)
Today's Date: 7/17/2023
Please indicate if you will require an E-Risk independent reproducibility check: ☐

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

Background & objective of the study:

Epigenome-wide association studies (EWAS) of Atopic Dermatitis (AD) in whole blood haven't yielded reproducible DNA methylation (DNAm) changes yet. This could be due to low sample sizes, poor AD definitions, gene-environmental interactions or tissue/cell type specific effects. Given the role of the immune response and inflammation in AD, it is likely that AD-associated DNAm changes in blood may exist but may occur in specific immune-cell subtypes only. As an alternative to an EWAS in specific cell types, for which little data is available, or whole blood EWAS, which may not capture associations in non-abundant cell types, we propose to apply computational algorithms to estimate cell type interacting DNAm changes from whole blood DNAm data.

In addition to this, previous EWAS have focused solely on differences in mean levels of DNAm between cases and controls. However, variability of human regulatory factors, such as DNAm, may influence ability to adapt to exposures and/or genotype (1). Testing for whether variability of DNAm sites differs between AD cases and controls may therefore also help with biological understanding of the disease.

In this analysis we aim to run 3 types of EWAS of AD using whole blood data in teenage individuals:

1. Cell-type interacting EWAS to obtain mean effect estimates of the association between DNAm and AD within individual cell types. As validation of these methods in real cell type data are missing, we plan to use three methods to assess this: CellDMC (2) and TCA (3) for discovery and omicWAS (4) for replication.
2. Conventional EWAS adjusted and unadjusted for cell counts to obtain mean effect estimates of the association between DNAm and AD averaged across all blood cell types.
3. Variance EWAS to identify variability difference estimates for each DNAm site between cases and controls.

Note, this proposal comes from a consortium called BIOMAP (BIOmarkers of Atopic dermatitis and Psoriasis) who are interested in identifying biomarkers for eczema. They are seeking cohorts to contribute in a "federated" model, whereby the analyses will be conducted by each participating

cohort (using provided scripts) and the results brought together centrally for meta-analysis. Meta-analysis EWAS summary statistics will be made publicly available.

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

Up to 20% of children and 15% of adults worldwide experience some form AD (5,6), yet the mechanisms underlying disease development are poorly understood. Identifying cell type interacting or variance DNAm changes in relation to AD, may improve our understanding of how molecular changes within specific cells impact disease development. It could also help provide information on the relevant cell types pertinent to AD development and whether DNAm-environment interactions influence AD.

Data analysis methods:

Overview

- Quality Control
- Cell counts for 12 blood cells will be derived using either the meffil R package (this may require access to the DNAm IDAT files) or the EpiDISH R package, instructions will follow and will be within the scripts provided.
- Estimate and adjust for genetic relatedness using GRAMMAR method
- Estimation of surrogate variables using SmartSVA implemented in meffil
- Conventional EWAS will be performed for each cohort using linear regression (DNA methylation will be the outcome) using the ewaff R package.
- Cell type interacting EWAS on whole blood will be performed using the cell-type interacting R packages (see below).
- Variance EWAS will be run using the jlst R package

QC and normalisation

- Meffil provides tutorials on sample QC and normalization. We prefer that you would use functional normalization. Probes will be excluded if they have a detection P value < 0.01 across over 10% of samples. If you prefer to use another normalisation method then please ensure that you use **normalized betas** for your DNAm levels.
- You should keep nonspecific binding probes, probes with SNPs in their sequence, multimapping probes. We will do post-hoc filtering on the meta-analysis results.
- Beta values should be used to measure methylation levels.

Conventional EWAS

Analyses will be run to estimate associations between DNAm at individual CpG sites and AD case/ control status. These will be run using linear regression with the models below:

Conventional EWAS models

1. DNAm ~ AD + sex + surrogate variables + age + (selection factors) + (ancestry)
2. DNAm ~ AD + sex + surrogate variables + age + cell counts + (selection factors) + (ancestry)

Covariates in brackets are the optional covariates as mentioned above.

Cell-type interacting analyses

We will run analyses to assess whether there are any cell-type interacting associations between DNAm and AD for each of the 12 cell types. The covariates for these analyses will be those in **Model 1**. We will use 3 methods to estimate these effects. Firstly, we will run CellDMC (2) and TCA

(3) across all CpG sites. Any associations at FDR < 0.05 (number of tests = number of DNAm sites tested multiplied by the number of cell types) identified within one method that replicate, with the same direction of effect, when using the other method at FDR < 0.05 (number of tests = number of sites to replicate) will be taken forward. After meta-analysis, we will then run omicWAS (4), limiting our analyses to only replicated sites from the previous analyses*. Sites that replicate in the same direction of effect when using omicWAS at FDR < 0.05, will be reported as having good evidence for cell-type interacting differential DNAm between eczema subtype cases and controls.

* We do not intend to run omicWAS on all sites as that would be too computationally intensive

Cell-type interacting EWAS steps

1. Remove outliers using Tukey method
2. Fit age, sex, genetic principal components as fixed effects, and family relatedness as a random effect, against each methylation probe and keep residuals.
3. Estimate SVs and discard any cell counts related SVs
4. Run cell type interacting model:

DNAm ~ AD + sex + surrogate variables + age + (selection factors) + (ancestry)

Covariates in brackets are the optional covariates as mentioned above.

Variance EWAS

We will run variance EWAS to assess whether variability in DNAm at any CpG site differs between AD cases and controls. We will use the jlst R package to run this and run several variance methods. Variance differences between cases and controls may capture interaction effects of some variables such as age, sex and cell type proportions. Therefore, we will run a completely unadjusted model and will compare this with two adjusted models (the same as those in the conventional EWAS) – see models below.

Variance EWAS steps

1. Remove outliers using Tukey method
2. Fit genetic relatedness as a random effect using kinship matrix, against each methylation probe and keep residuals.
3. Estimate SVs
4. Run variance EWAS models:

    1. DNAm ~ AD
    2. DNAm ~ AD + sex + surrogate variables + age + (selection factors) + (ancestry)
    3. DNAm ~ AD + sex + surrogate variables + age + cell counts + (selection factors) + (ancestry)

Covariates in brackets are the optional covariates

Scripts will be provided to run the following analyses:

- Estimation of the 12 blood cell counts
- Surrogate variable generation
- Conventional EWAS

- Cell-type interacting EWAS using CellDMC, TCA, omicWAS
- Variance EWAS

Variables needed and at which ages:

DNAm measured by the Illumina Infinium HM450 BeadChip or the EPIC BeadChip, with the EPIC BeadChip being preferred. Any batch variables related to the DNAm data will be needed.
Eczema/Atopic dermatitis: Individuals will be defined as AD cases if they have recorded a doctor diagnosis or self-report of AD between 13 and 20 years old. All individuals without a self-report or doctor diagnosis of AD will be assigned as controls.

- FAMILYID Unique family identifier
- ATWINID Twin A ID (ex chkdig)
- BTWINID Twin B ID (ex chkdig)
- ZYGOSITY Zygosity
- SAMPSEX Sex of Twins: In sample
- TAGEE18 (Age at Interview - P18 – Elder (and Younger))
- TAGEGE18 (Age at Interview (Grouped) - P18 – Elder (and Younger))
- LIFHEA3 Since you were 12 have you been told by a doctor that you have eczema? (age 18 variable; elder and younger needed)
- LIFHEA3b Do you have a problem with eczema now? (age 18 variable; elder and younger needed)
- Illumina 450K DNA methylation data from peripheral blood at age-18 + related variables (probes, batch number, methylation array control probe principal components, chipID etc) for both elder and younger twin.
- Cell count variation for samples used to generate DNA methylation profiles at age 18
- Genetic principal components (PCs) calculated on the subset of twins with DNA methylation data at age 18
- Kinship matrix

References cited:

1. Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A*. 2010;107 Suppl 1(Suppl 1):1757-1764. doi:10.1073/pnas.0906183107
2. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods*. 2018;15(12):1059-1066. doi:10.1038/s41592-018-0213-x
3. Rahmani, E., Schweiger, R., Rhead, B. *et al.* Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat Commun* **10**, 3417 (2019). https://doi.org/10.1038/s41467-019-11052-9
4. Takeuchi F, Kato N. Nonlinear ridge regression improves cell-type-specific differential expression analysis. *BMC Bioinformatics*. 2021;22(1):141. Published 2021 Mar 22. doi:10.1186/s12859-021-03982-3
5. Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *Lancet* [Internet]. 2020;**396**(10247):345–360. Available from: https://www.sciencedirect.com/science/article/pii/S0140673620312861
6. Nutten S. Atopic Dermatitis: Global Epidemiology and Risk Factors. *Ann Nutr Metab* [Internet]. Karger Publishers; 2015 May 6 [cited 2022 Mar 25];**66**(Suppl. 1):8–16. Available from: https://www.karger.com/Article/FullText/370220