Assessing methods aiming to detect cell-specific DNAm effects

# Background

Studies analysing the association between DNAm and complex traits will often use bulk tissue samples to measure DNAm. The resulting DNAm percentage at any given CpG site is the aggregate value of DNAm across each cell type in the bulk tissue sample. If associations between a given complex trait and DNAm are cell type-specific, then aggregating DNAm across cells may cause a dilution of signal in a cell type and therefore a reduction in the ability to detect that signal. Ideally, we’d be able to separate individual cells within a tissue and measure DNAm in all of them genome-wide. Unfortunately, this is very expensive and so not viable within population-based cohorts. Further, many population-based cohorts have already assayed DNAm using bulk tissue data and do not have the resources to go back, sort the cells, and re-assay the samples.

Recently, there have been methods published that aim to assess the association between complex traits and DNAm in a cell-specific manor. These methods are relatively untested outside the papers they were published in, making it hard to determine whether they would be useful to regularly apply within an EWAS context. This project aims to conduct analysis to elucidate the usefulness of these methods.

Aims/objectives:

1. Assess the accuracy of the methods
2. Assess whether the methods identify associations that regular EWAS would fail to
3. Understand whether cell-specific associations are causal

# Methods

## Cell-specific effects methods

|  |  |  |  |
| --- | --- | --- | --- |
| Name | PMID | Paper title | implementation |
| HIRE | 31308366 | Detection of cell-type-specific risk-CpG sites in epigenome-wide association studies | R package |
| TCA | 31366909 | Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology | R package |
| NA | 31889537 | Cell Type-Specific Methylome-wide Association Studies Implicate Neurotrophin and Innate Immune Signaling in Major Depressive Disorder |  |
| NA | 33350870 | Estimating cell-type-specific DNA methylation effects in heterogeneous cellular populations |  |
| CellDMC | 30504870 | Identification of differentially methylated cell types in epigenome-wide association studies | R package |
| omicwas | 33752591 | Nonlinear ridge regression improves cell-type-specific differential expression analysis | R package |
| TOAST | 31484546 | TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis | R package |

Two papers do not provide enough detail to replicate their method (31889537, 33350870). The other methods require as input a DNAm matrix, a cell proportions matrix\*, a phenotype vector, and a covariates matrix. \*Cell proportions can be calculated by most of the methods and HIRE does not allow input of a previously calculated cell proportions matrix. Unfortunately, the cell types generated by HIRE are unreliable (previous work from Matt), and so will be excluded from the analysis leaving these four methods:

* CellDMC
* TCA
* Omicwas
* TOAST

## Simple false-positive check

To simple test of the methods will be applied: do they find associations between DNAm and randomly generated phenotype?

The methods will be run to assess associations between DNAm data from ARIES mothers and 1000 randomly generated phenotypes. The number of associations at P<1x10-7 will be checked. Given the number of tests (1000 phenotypes and 6 cell types), we’d expect around 6000 associations using each method.

EWAS using bulk tissue (conventional EWAS) will also be conducted of the random phenotypes. For each cell-specific method, a meta-analysis of cell-specific associations will be run and this will be compared to the results of the conventional EWAS.

## Identification of new signal

EWAS using each of the methods will be run using all clinical phenotypes measured at the FOM1 clinic in ARIES mothers. Age, 10 genomic PCs, and 20 SVs will be used as covariates. Conventional EWAS will also be conducted using the same covariates.

CpGs will be extracted that are associated with any complex traits at P < 1x10-7 across any of the EWAS. Associations will then be compared across methods (including conventional EWAS). A meta-analysis across cell types will be performed again to see if the results re-capitulate the conventional EWAS results.

## Assessing the most reliable method(s)

Two possibilities here:

Firstly, we could get do some more profiling of samples to get a dataset with DNAm measured in whole blood, DNAm measured in flow-sorted blood cells, and phenotype measures. Results from cell-specific methods would be compared to EWAS using flow-sorted data.

Secondly, we could use a replication framework, whereby we have two samples: one with DNAm in bulk tissue and phenotype(s) and the second with DNAm in sorted cells and the same phenotype(s). Results when applying the methods to the bulk tissue dataset would be checked for replication in the sorted cells dataset. This was how the authors of TCA + CellDMC tested their methods. Possible datasets for this: MESA, BLUEPRINT, ARIES.

The first scenario would be ideal.

## Causality

We will take the results we get from the analyses in ARIES, generate cell-specific mQTLs for the associations at P<1x10-7 and use MR to test whether they might be causally related to the trait for which they were identified in the EWAS. Given we might not be able to confidently assess the accuracy of the methods, assessment of causality allows us to assess the usefulness of the methods. Assuming individuals would mostly look to use these methods to identify effects (or proxies for effects) rather than predictive biomarkers, assessing whether they are likely picking up sites where DNAm effects the trait of interest gives a good indication of the methods’ usefulness. If there is little evidence, from the MR analyses, that DNAm at sites identified in a cell-specific manor, effect complex traits then the usefulness of the methods is greatly reduced.

Cell-specific mQTLs could be generated in MESA/BLUEPRINT.