Introduction

- Methylation arrays from blood are the most common type of epigenetic data collected, and are generally comprised of measurements from >100,000 genomic locations, called loci.
- When the methylation level at a locus is different between two sample groups, this is called differential methylation.
- Cell-type methylation levels vary, but are known to be related by the haematopoietic lineage.
- Differential methylation can be cell-type-specific, where only a subset of blood cell-types in the sample are differentially methylated.
- Objective: Identify loci with differential methylation for specific cell-types

Model

Let \( y_i \) be the blood methylation level of sample \( i \). Since \( y_i \) is constrained to the unit interval, a logit-Normal distribution was used.

\[
\pi(y_i | \mu_i, \rho) = \text{logitNormal}(y_i; \mu_i, \rho).
\]
Assumption: median of the blood methylation level is a linear combination of constituent cell-type methylation levels \( \eta_1, \ldots, \eta_K \), weighted by the cell-type proportions \( p_1, \ldots, p_K \).

\[
\text{logit}^{-1}(\mu_i) = \sum_{k=1}^{K} p_k \eta_k
\]
\( \eta_k \) is parameterised in terms of a baseline \( \theta_k \) and a shift \( \phi_k \) for each cell-type. \( \delta \in \{0, 1\} \) represents the binary covariate of interest (e.g. control = 0, case = 1).

Method: Inference and Case Study

- Model fitted using numerical optimisation procedure in STAN [1].
- Obtained Maximum A Posteriori (MAP) estimates for \( \phi \) parameters and a Hessian matrix estimate.
- Laplace approximations to the posterior calculated for each for \( \phi_k \).
- Predicted differential methylation for cell-type \( k \) if

\[
\text{Pr}(|\phi_k - \phi_k^{\text{MAP}}| > \alpha | \text{Data}) < \alpha
\]

Case study: find differentially methylated loci associated with sex.
- Data-set contained 5 females and 9 males.
- Cell-sorted data contained ground-truth for comparison with predictions.

Results

- CSMA method outperformed other methods for CD8^T and CD19^B.
- OE-LR method outperformed CSMA for the Monocyte and Neutrophil cell-types.
- CSMA tended to detect more differentially methylated loci specific to the given cell-type.
- OE-LR tended to detect differentially methylated loci where all cell-types were differentially methylated.
- PSEA, Agg-LR, and LASSO methods were sub-optimal.

Conclusions

- CSMA and OE-LR are both useful for finding differentially methylated loci.
- Best method may be to use an ensemble approach for finding both cell-type specific and unspecific differentially methylated loci.

Current Work:
- Extending CSMA model to include multiple covariates and different data distributions.
- Reducing potential bias from inaccurate proportion estimates.
- Developing empirical Bayes approach for optimal value of \( \lambda_b \).

Acknowledgments

References