

Caregiver contact time and sex effects on methylation status in infants

Sumaiya Islam^{1,2}, Nicole Gladish^{1,2}, Jonathan Baik¹, Olivia de Goede^{1,3}, Shannon Erdelyi¹, Jessica Pilsworth^{1,4}

¹University of British Columbia; ²Centre for Molecular Medicine and Therapeutics; ³Child and Family Research Institute; ⁴Genome Sciences Centre, BC Cancer Agency

BACKGROUND

Caregiving in early life has a profound influence on the long-term health and well-being of offspring. Differences in caregiving in various animal models have been shown to result in differential DNA methylation profiles in offspring¹, but this phenomenon has yet to be demonstrated in humans.

Table 1. Sample sex and contact

| | Contact High | Contact Low |
|--------|--------------|-------------|
| Female | 20 | 19 |
| Male | 36 | 19 |

Study Objective

To test if caregiver-infant contact influences the DNA methylation profile in human infants.

Study Design

From buccal samples of n = 94 infants (5 weeks of age), genome-wide DNA methylation was measured at ~450,000 CpG sites using the Illumina Infinium 450K BeadChip (450K array). Site-specific methylation was compared between contact and sex groups (Table 1).

PRE-PROCESSING

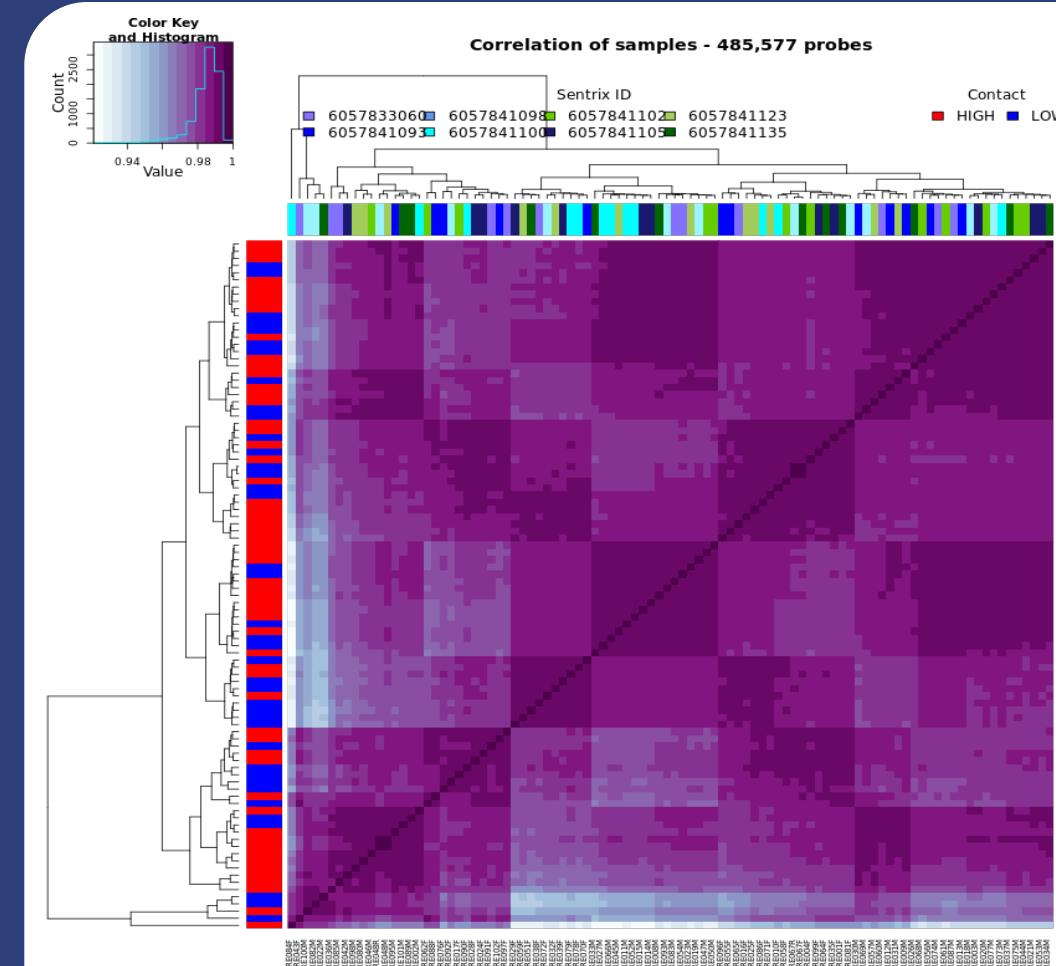


Figure 1. Sample correlation heatmap of pre-processed data using unsupervised cluster analysis.

Table 2. Summary of probes filtered.

| Probes | Samples | Filtering |
|--------|---------|--------------------------|
| I | 485,577 | 96 |
| II | 485,512 | SNP Probes |
| III | 473,864 | XY Probes |
| IV | 472,503 | Bad Probes |
| V | 452,222 | Polymorphic Probes |
| VI | 441,566 | XY Multiple Binders |
| VII | 441,566 | 94 Removal of Replicates |

- There is no apparent clustering of samples based on contact (Figure 1).
- There seems to be no batch effect, as shown by the lack of sample clustering by Sentrix ID (Figure 1).

Probes were filtered out of the dataset based on the following criteria (Table 2):

1. SNP Probes: probes that assay sites of known common DNA sequence polymorphisms
2. XY Probes: probes on the sex chromosomes
3. Bad Probes: probes with a detection P-value < 0.01 or with missing beta values in >5% of samples
4. Polymorphic Probes: probes that have SNP sites located at the CpG site being measured
5. XY Multiple Binders: probes that bind non-specifically to sex chromosomes

NORMALIZATION

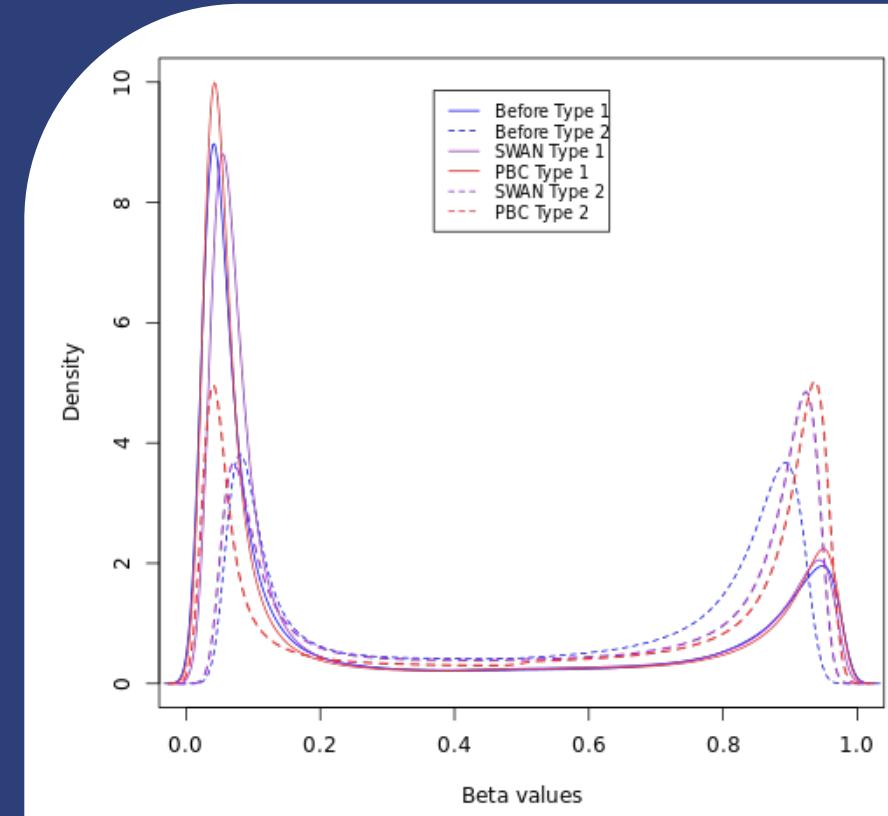
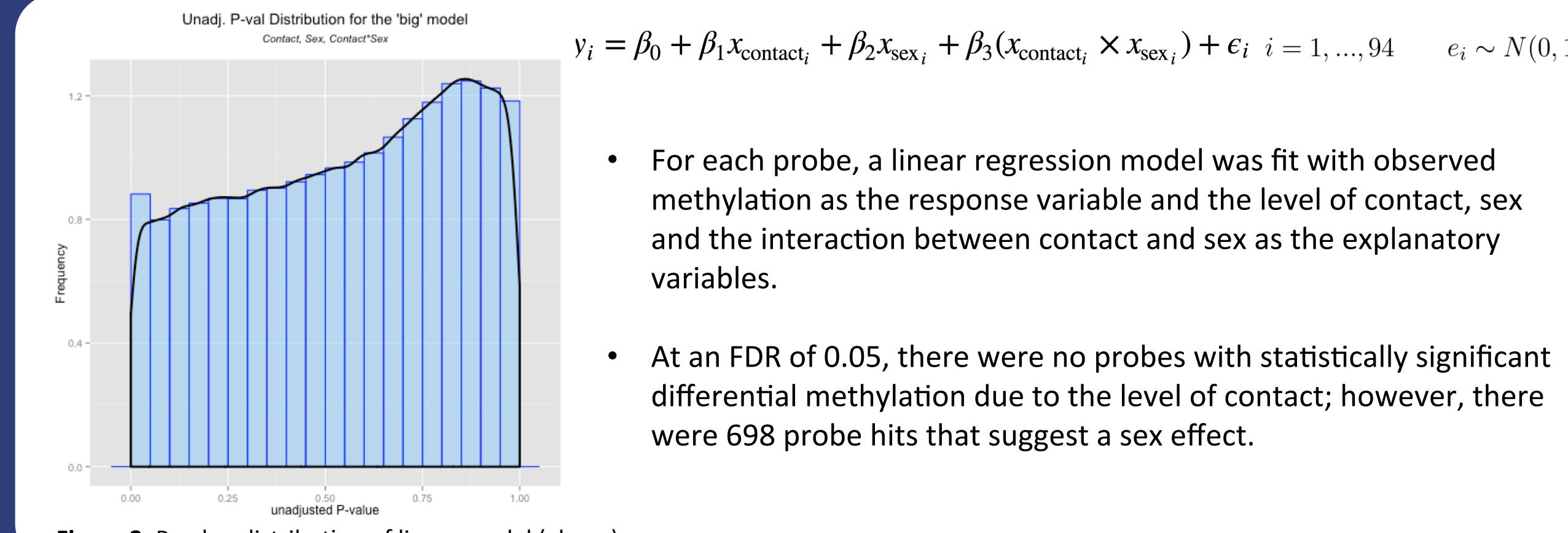


Figure 2. Beta value density plot for all samples from raw (blue), PBC-normalized (red) and SWAN-normalized (purple) data

| | RAW | COLOR CORRECTION | PBC | SWAN |
|----------|--------|------------------|--------|--------|
| CARE04M | 0.9979 | 0.9983 | 0.9984 | 0.9983 |
| CARE067F | 0.9985 | 0.9987 | 0.9988 | 0.9987 |
| AVG | 0.9982 | 0.9985 | 0.9986 | 0.9985 |

- Data must be colour corrected to account for differences between the two fluorescent labels (red or green).
- The beta value distributions differ due to the different chemistries between Type I and II probes (Figure 2), therefore, normalization is required before both probe types can be compared.
- We tried two commonly used normalization methods:
 - Peak-based correction (PBC)² rescales the M-values³ of Type II to match the range of Type I probes.
 - Subset-quantile within array normalization (SWAN)⁴ subsets the probes based on their underlying CpG content prior to normalization.
- Based on correlation of technical replicates (Table 3) and similarity of beta-value distributions between the two probe types (Figure 2), we chose the PBC-normalized data for analysis.

LIMMA



- $y_i = \beta_0 + \beta_1 x_{\text{contact}_i} + \beta_2 x_{\text{sex}_i} + \beta_3 (x_{\text{contact}_i} \times x_{\text{sex}_i}) + \epsilon_i \quad i = 1, \dots, 94 \quad \epsilon_i \sim N(0, 1)$
- For each probe, a linear regression model was fit with observed methylation as the response variable and the level of contact, sex and the interaction between contact and sex as the explanatory variables.
 - At an FDR of 0.05, there were no probes with statistically significant differential methylation due to the level of contact; however, there were 698 probe hits that suggest a sex effect.

SVA & CLUSTER ANALYSIS

Surrogate Variable Analysis (SVA)

$$y_i = \beta_0 + \beta_1 x_{\text{contact}_i} + \beta_2 x_{\text{sex}_i} + \beta_3 (x_{\text{contact}_i} \times x_{\text{sex}_i}) + \beta_4 x_{\text{sv1}_i} + \beta_5 x_{\text{sv2}_i} + \epsilon_i \quad i = 1, \dots, 94 \quad \epsilon_i \sim N(0, 1)$$

- There appeared to be an unmeasured source of variation unexplained by contact or sex (Figure 3).
- Surrogate variable analysis⁵ (SVA) was used to identify hidden variables underlying this heterogeneity.
- The linear regression model above was fit to each probe.
- At an FDR of 0.05, there were no probes with significant differences in methylation due to contact, but there were 990 probes with significant differences in methylation due to sex.

Cluster Analysis

$$y_i = \beta_0 + \beta_1 x_{\text{contact}_i} + \beta_2 x_{\text{sex}_i} + \beta_3 (x_{\text{contact}_i} \times x_{\text{sex}_i}) + \sum_{c=4}^7 \beta_c x_{\text{cluster}_c} + \epsilon_i \quad i = 1, \dots, 94 \quad \epsilon_i \sim N(0, 1)$$

- The Euclidean distance between samples was computed based on scaled methylation data, and unsupervised hierarchical clustering using ward linkage was used to identify sample clusters.
- The linear regression model above was fit to each probe.
- At an FDR of 0.05, there were no probes with significant differences in methylation due to contact, but there were 1006 probes with significant differences in methylation due to sex.

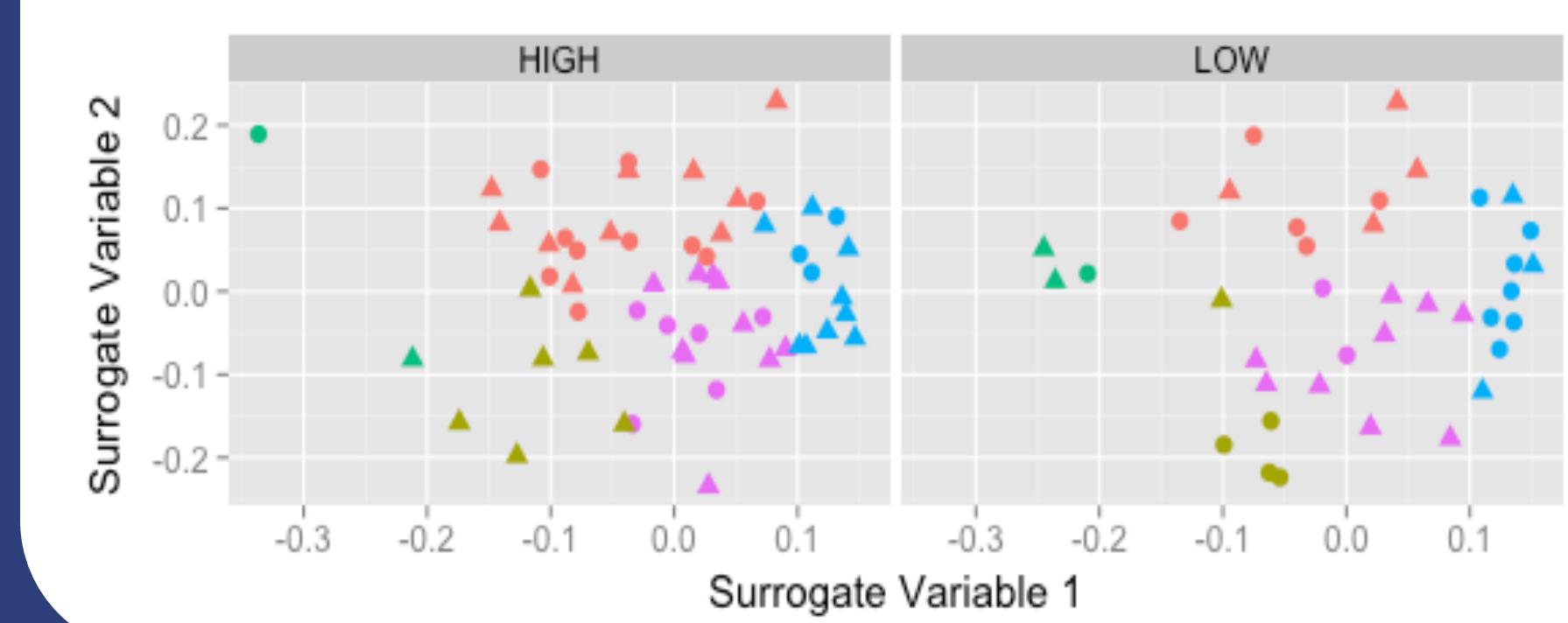


Figure 3. The relationship between cluster membership and the identified surrogate variables indicate that both methods yield similar groupings that are unrelated to sex and contact.

PCA

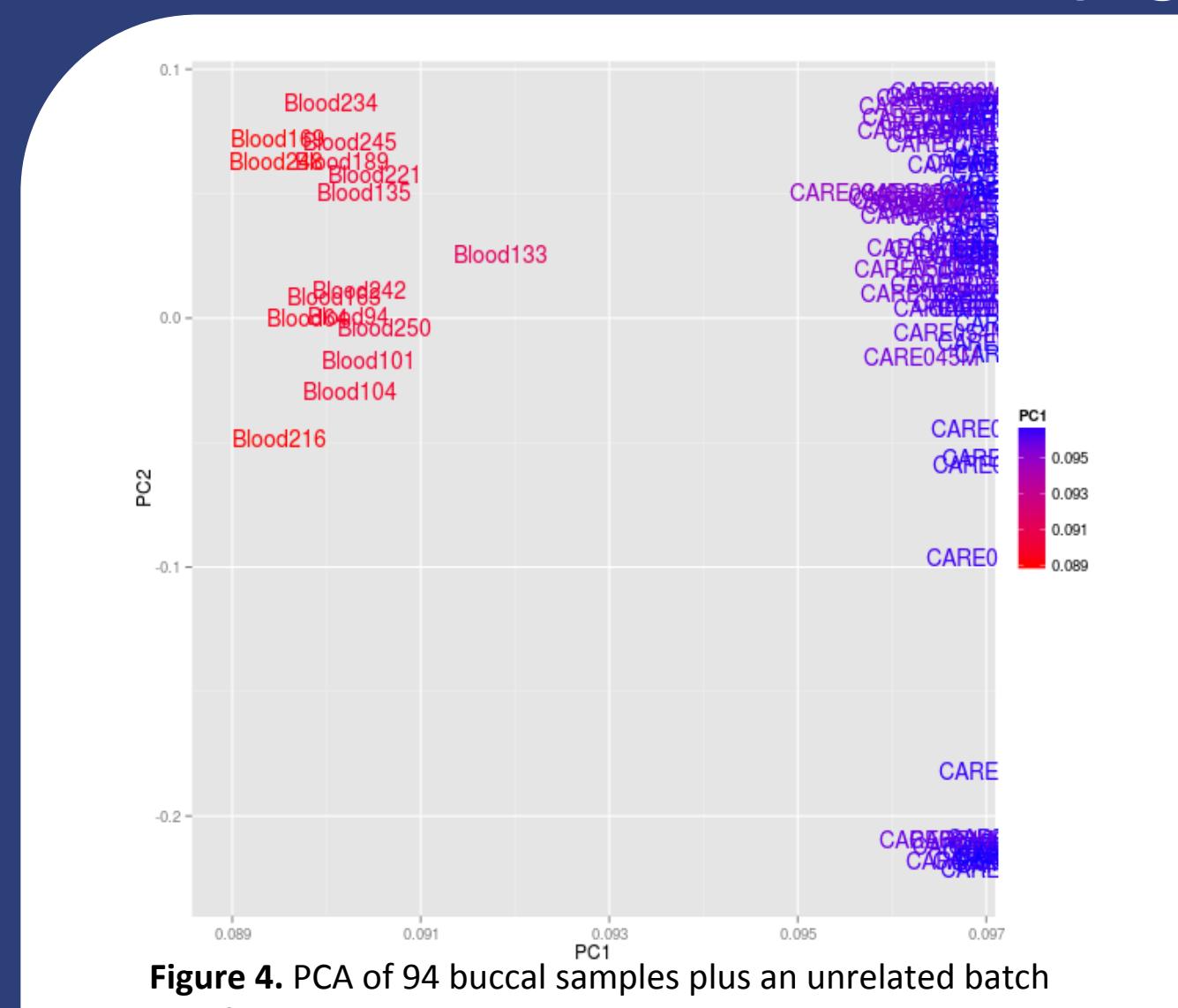


Figure 4. PCA of 94 buccal samples plus an unrelated batch of 16 whole blood samples to assess potential blood contamination.

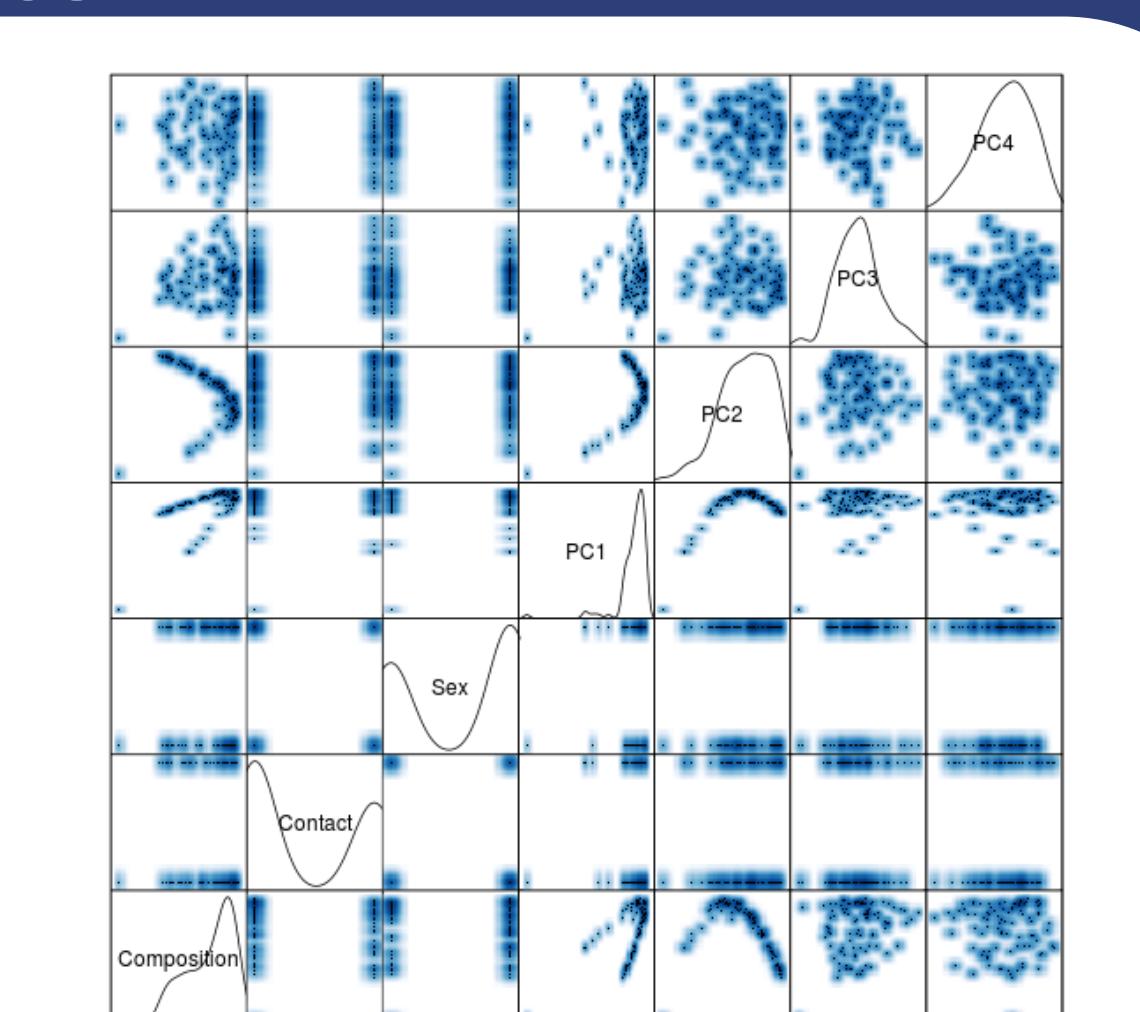


Figure 5. A scatterplot matrix of covariates and principal components indicating strong correlation between PC1 (and to some extent PC2) and tissue composition.

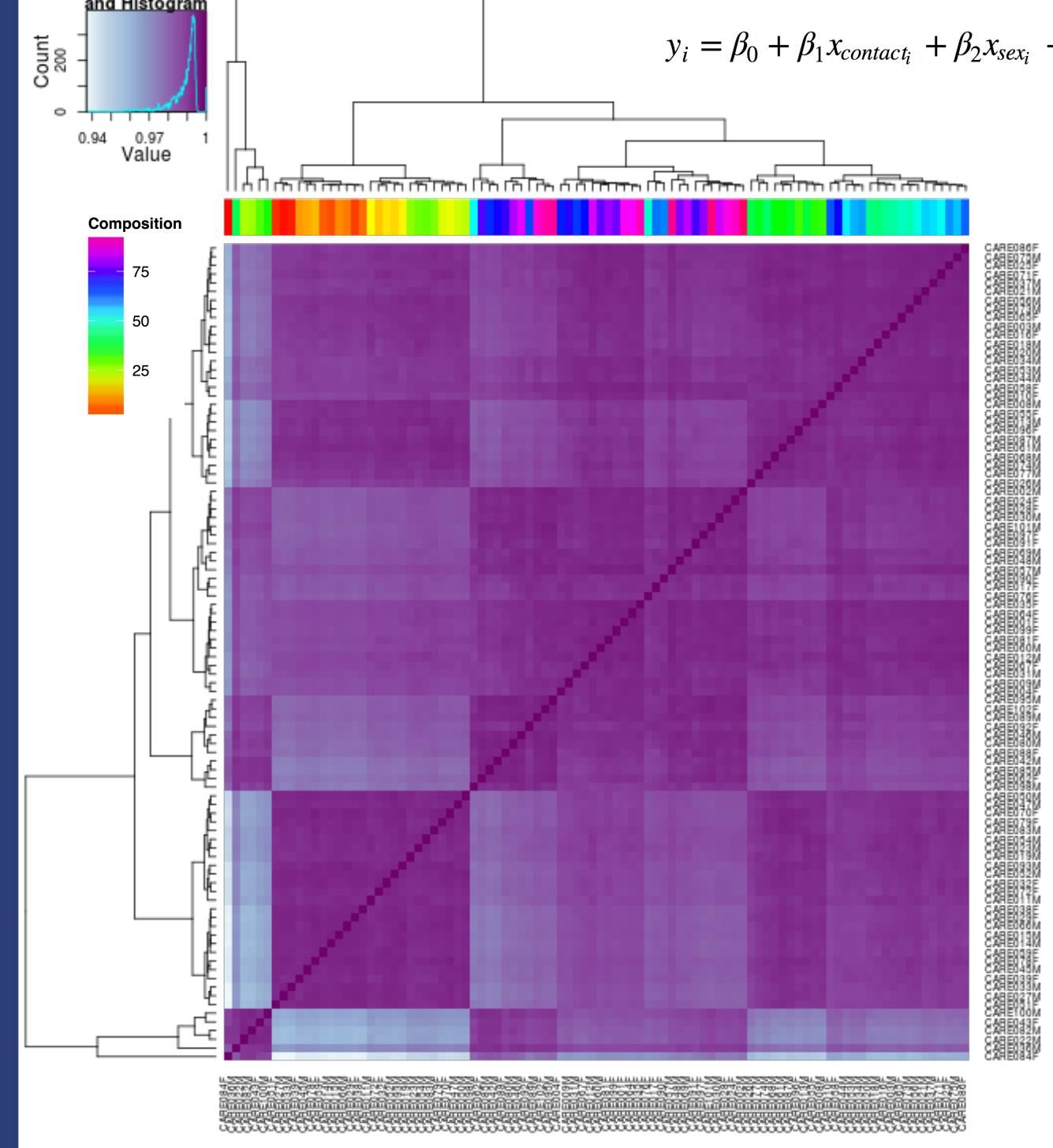


Figure 6. Sample correlation heatmap suggestive of a tissue composition effect. Tissue composition is indicated by the top bar colour, and evaluated based on ranked PC1 score.

HIT COMPARISON



Figure 7. Box plots of top sex hits among randomly-selected probe "hits" from various analysis methods.

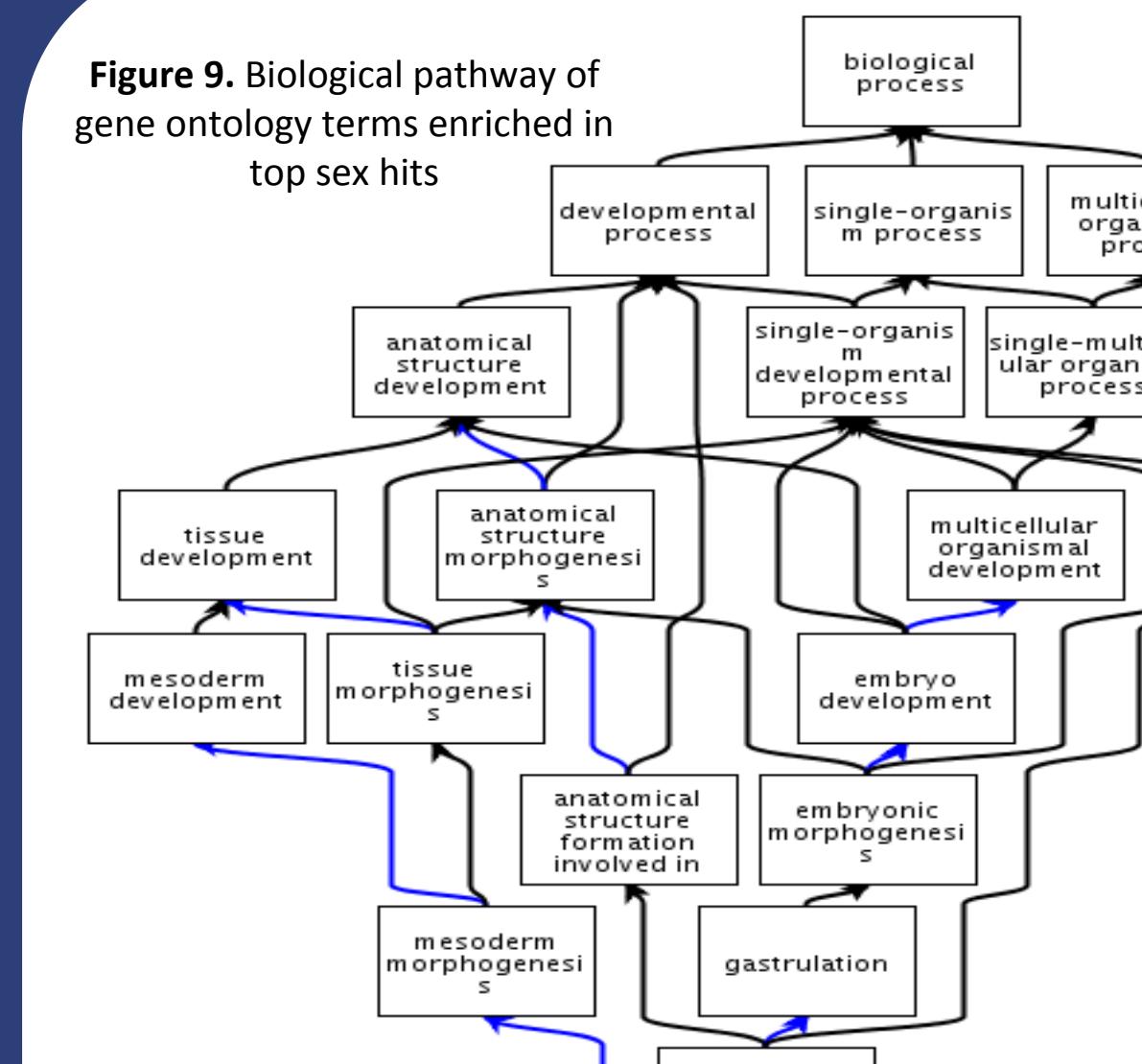
- The four analysis methods showed differences in probe hits for sex effects (Figure 7).
- 378 hits were common between all four methods (Figure 8).
- The common sex effect hits were used in pathway analysis.

Figure 8. Venn diagram showing overlapping hits from the various analysis methods.



PATHWAY ANALYSIS

Figure 9. Biological pathway of gene ontology terms enriched in top sex hits



- None of the gene ontology (GO) term hits achieved statistical significance (Table 4).

- Top GO hits were associated with mesoderm formation & embryonic development (Figure 9).

- Specific genes within these GO terms are involved with the TGF-β signaling pathway (Table 5), which is a critical regulator of cell growth, differentiation, and development⁷.

- TGF-β is involved in production of IgA, the major humoral effector of mucosal immunity⁷.

- Early in lactation, breast milk contains an abundant amount of TGF-β; therefore, it is expected that infants will be enriched for gene products involved in this pathway⁷.

| GENE | Description of genes associated with GO:0001707 mesoderm formation |
|-------|--|
| WNT3 | Provides instructions for making proteins that participate in chemical signaling pathways in the body |
| SMAD3 | Provides instructions for making a protein involved in transmitting chemical signals from the cell surface to the nucleus in the TGF-β signaling pathway |
| BMP7 | Encodes a member of the bone morphogenic protein family - member of the TGF-β superfamily |
| BMP4 | Plays an important role in the onset of endochondral bone formation in humans - member of the TGF-β superfamily |
| NODAL | May be essential for organization of axial structures in early embryonic development - member of the TGF-β superfamily |

Table 5: Genes associated with gene ontology terms

SUMMARY

Contact does not seem to be a significant source of variability for DNA methylation in this study, but there may be an effect due to the sex of the infant. Limitations of this study include:

- Buccal swabs used as samples are known to contain heterogeneous cell populations, and differential tissue composition is one of the largest contributors to variation in DNA methylation.
- The amount of caregiver-infant contact was supplied as a discrete variable, reducing the statistical power in the analyses. Having a continuous contact variable may have been more appropriate for our analyses.

References

- Weaver, I. et al. Epigenetic programming by maternal behavior. *Nature Neuroscience* 7, 847–854 (2004).
- Dedeurwaerdher, S. et al. Evaluation of the Infinium Methylation 450K technology. *Epidemiology* 3, 771–784 (2011).
- Du, P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11, 587–587 (2010).
- Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* 13, R44 (2012).
- Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735 (2007).
- Jones, M. et al. Distinct DNA methylation patterns of cognitive impairment and trisomy 21 in down syndrome. *BMC Med Genomics* 6, 58 (2013).
- Ogawa, J., Sashara, A., Yoshida, T. et al. Role of transforming growth factor-β in breast milk for initiation of IgA production in newborn infants. *Early Hum Dev.* 2004;77:67–75