COLUMBIA CENTER FOR CHILDREN'S ENVIRONMENTAL HEALTH
MAILMAN SCHOOL OF PUBLIC HEALTH
COLUMBIA UNIVERSITY

Confounds of Epigenetic Epidemiology using Cord Blood

Julie B. Herbstman

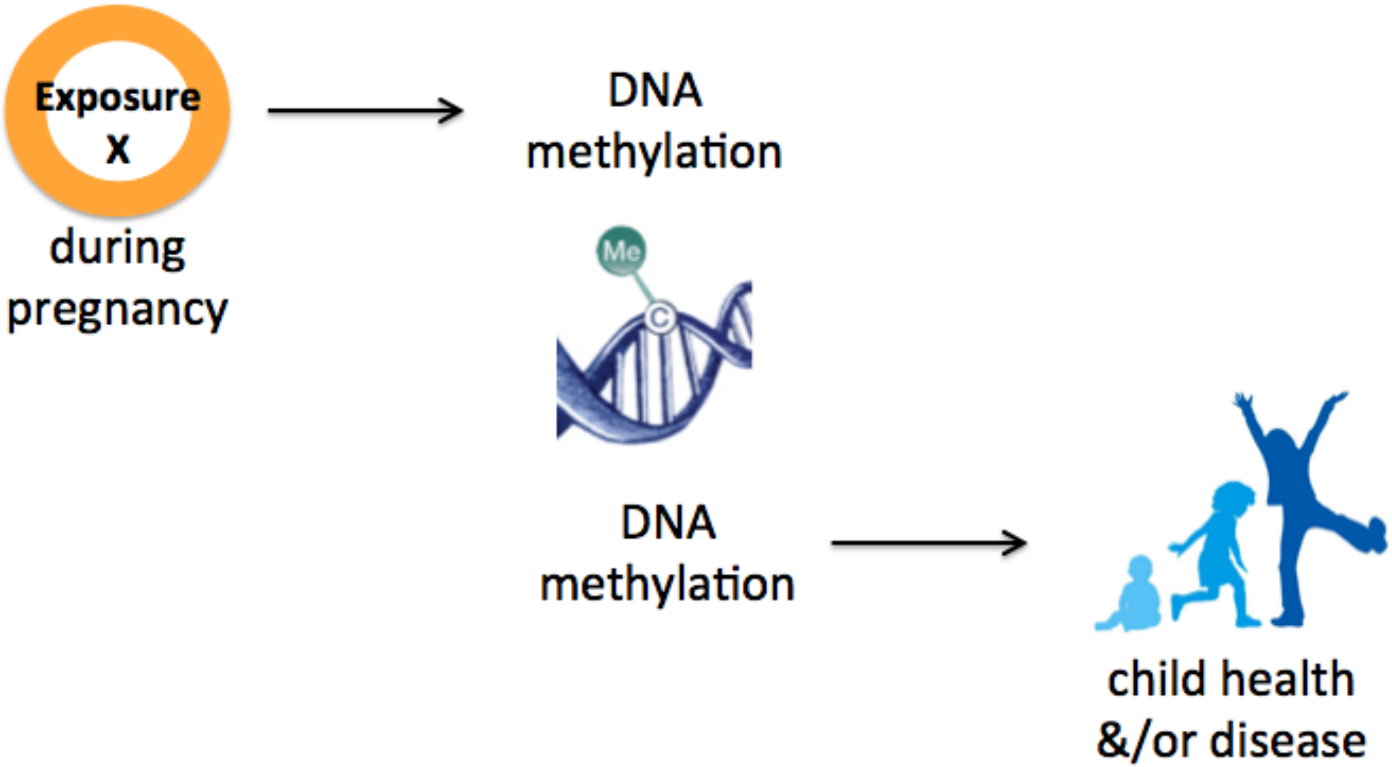
April 8, 2015

DNA methylation as a mediator

adverse outcomes in childhood (and across the lifecourse)

http:.www.niehs.nih.gov/exposurebiology/
http://www.bodyandsoulkc.com/

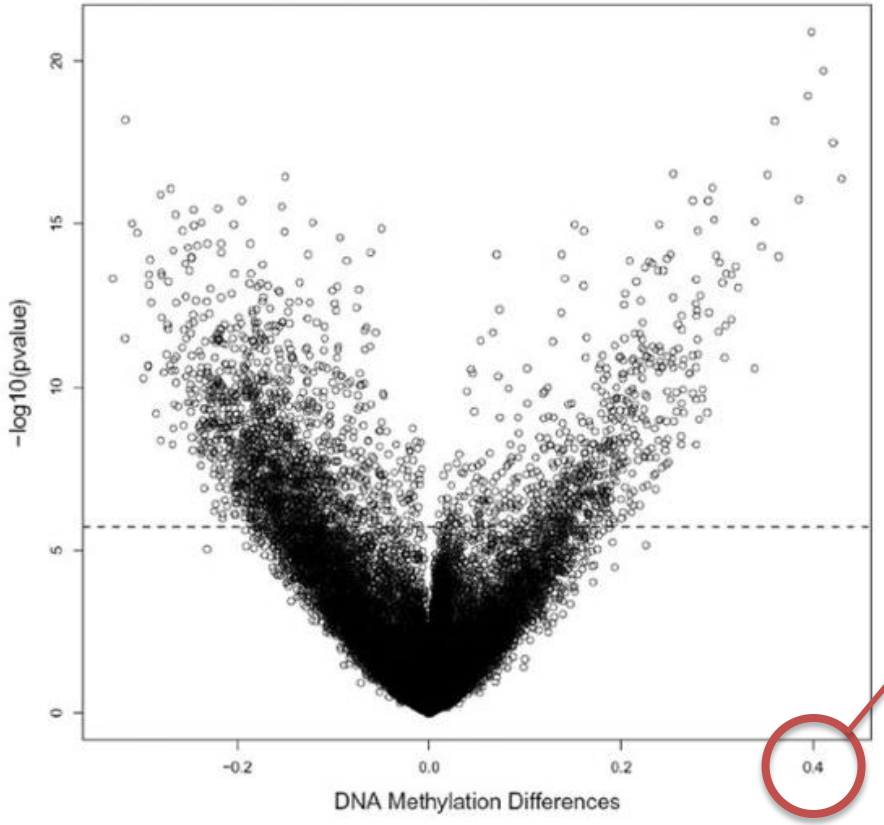
2 sets of hypotheses



- A longitudinal birth cohort is a reasonable study design
- Cord blood among the first (easily) accessible tissue
- Many groups are using the Infinium450K array to measure DNA methylation in stored cord blood

http://www.promega.com
http://www.rch.org.au/ccch/

Anticipated effect size



40% difference
in HCC tumor vs. adjacent normal

HCC tumor and adjacent normal tissue

[Shen et al. 2012]

Anticipated effect size

15% difference
in smokers
vs.
non-smokers

Cord blood methylation in maternal smokers vs. non-smokers

[Joubert et al. 2013]

Some strategies to improve detection of small effects

- Increase the sample size: consortium efforts
  - e.g., Prenatal and Childhood Epigenetics (PACE) consortium
- Improve the technical aspects of the measurements: reduce “noise”
  - e.g., normalization procedures
- Control for confounders using either statistics or design
  - e.g., twin or sibling studies (design)
  - e.g., stratification/adjustment (statistical)
    - Confounding by cell type distribution

Methylation varies between cells of different types

- Because DNA methylation is tissue and cell-type specific, methylation measured in unsorted peripheral blood may be an important source of confounding.

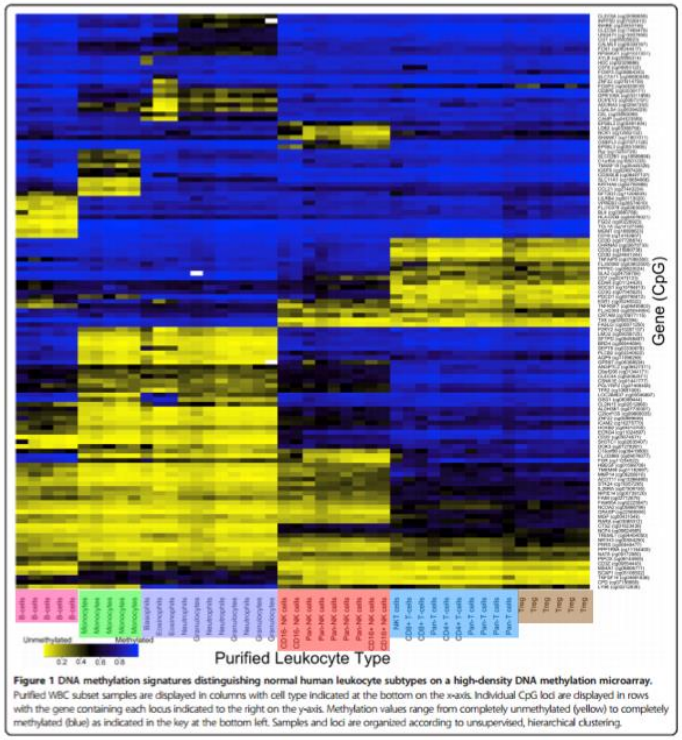


Figure 1 DNA methylation signatures distinguishing normal human leukocyte subtypes on a high-density DNA methylation microarray. Purified WBC subset samples are displayed in columns with cell type indicated at the bottom on the x-axis. Individual CpG loci are displayed in rows with the gene containing each locus indicated to the right on the y-axis. Methylation values range from completely unmethylated (yellow) to completely methylated (blue) as indicated in the key at the bottom left. Samples and loci are organized according to unsupervised, hierarchical clustering.

Unsupervised clustering of average beta values in sorted blood

[Accomando et al. Genome Biology

Methylation varies between cells of different types

- Because DNA methylation is tissue and cell-type specific, methylation measured in unsorted peripheral blood may be an important source of confounding.

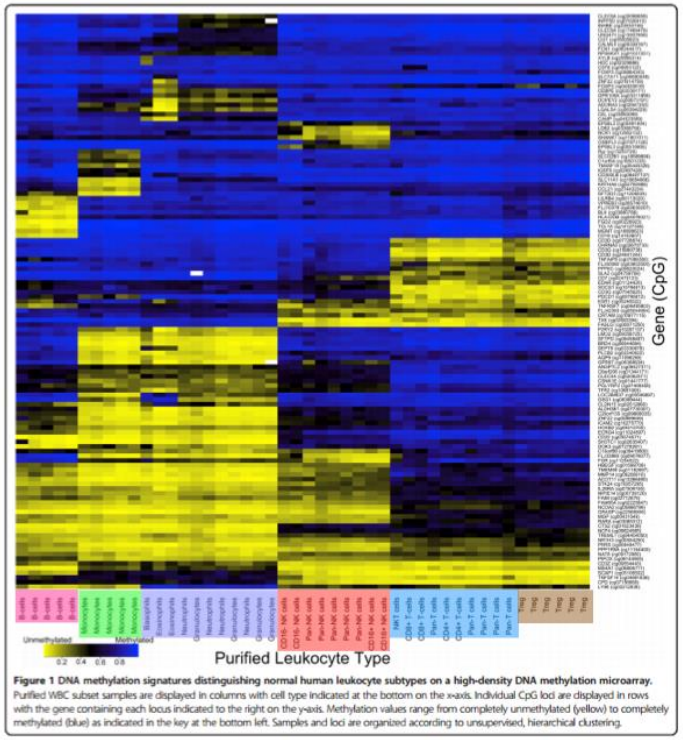


Figure 1 DNA methylation signatures distinguishing normal human leukocyte subtypes on a high-density DNA methylation microarray. Purified WBC subset samples are displayed in columns with cell type indicated at the bottom on the x-axis. Individual CpG loci are displayed in rows with the gene containing each locus indicated to the right on the y-axis. Methylation values range from completely unmethylated (yellow) to completely methylated (blue) as indicated in the key at the bottom left. Samples and loci are organized according to unsupervised, hierarchical clustering.

Unsupervised clustering of average beta values in sorted blood

[Accomando et al. Genome Biology

Prenatal Environmental Exposure → Change in DNA methylation in cord blood

Methylation varies between cells of different types

- Because DNA methylation is tissue and cell-type specific, methylation measured in unsorted peripheral blood may be an important source of confounding.

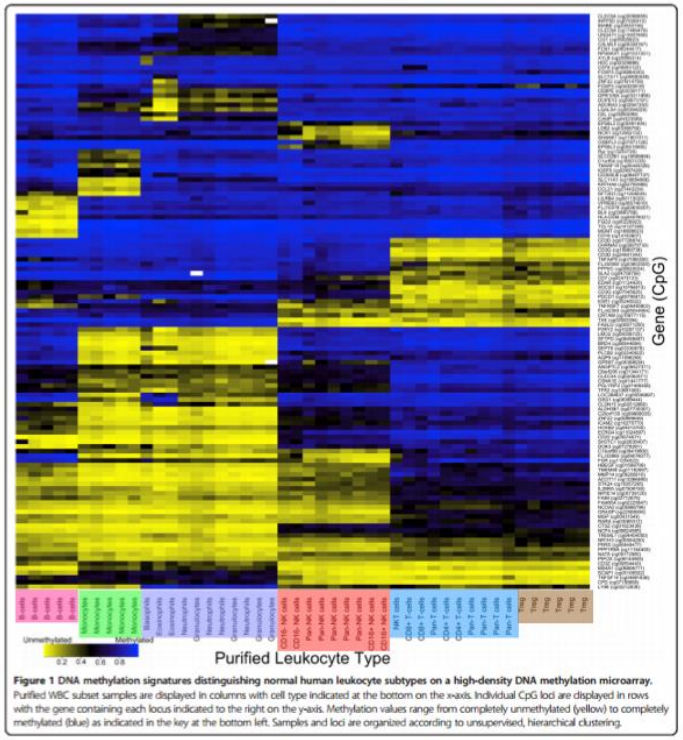
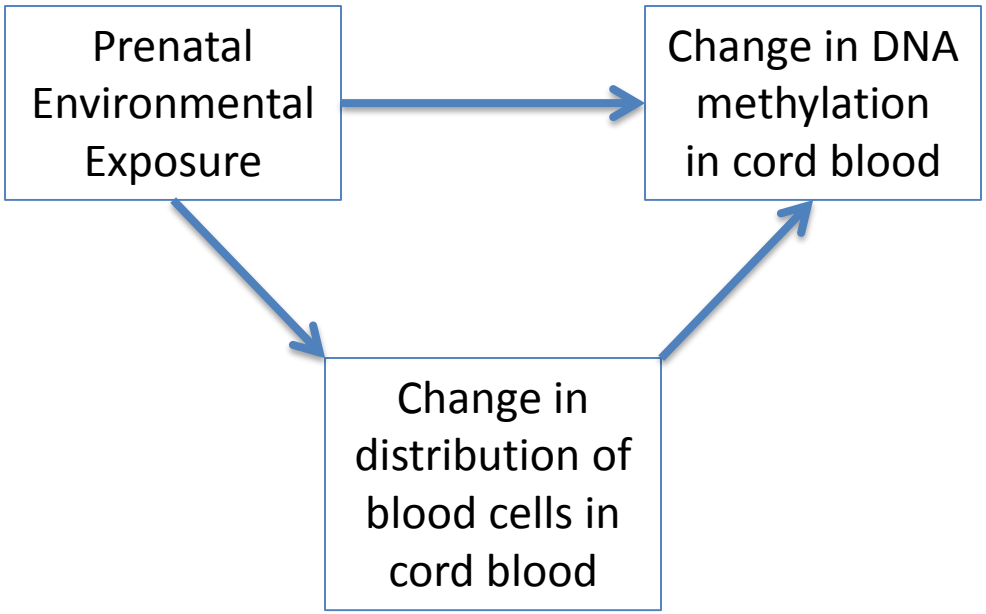


Figure 1 DNA methylation signatures distinguishing normal human leukocyte subtypes on a high-density DNA methylation microarray. Purified WBC subset samples are displayed in columns with cell type indicated at the bottom on the x-axis. Individual CpG loci are displayed in rows with the gene containing each locus indicated to the right on the y-axis. Methylation values range from completely unmethylated (yellow) to completely methylated (blue) as indicated in the key at the bottom left. Samples and loci are organized according to unsupervised, hierarchical clustering.

Unsupervised clustering of average beta values in sorted blood

[Accomando et al. Genome Biology



Problem:

- Failure to account for the cell distribution can confound hypothesized exposure-to-methylation associations, leading to spurious results or failure to detect true relationships.

Problem:

- Failure to account for the cell distribution can confound hypothesized exposure-to-methylation associations, leading to spurious results or failure to detect true relationships.

Possible solutions:

1. Restrict: only measure DNA methylation in homogenous blood samples
2. Stratify: analyze DNA methylation in cell-type-specific strata
3. Adjust for cellular composition using multivariate regression:
   - Count cellular composition
   - *Use methylation at specific CpG sites to infer cellular composition*

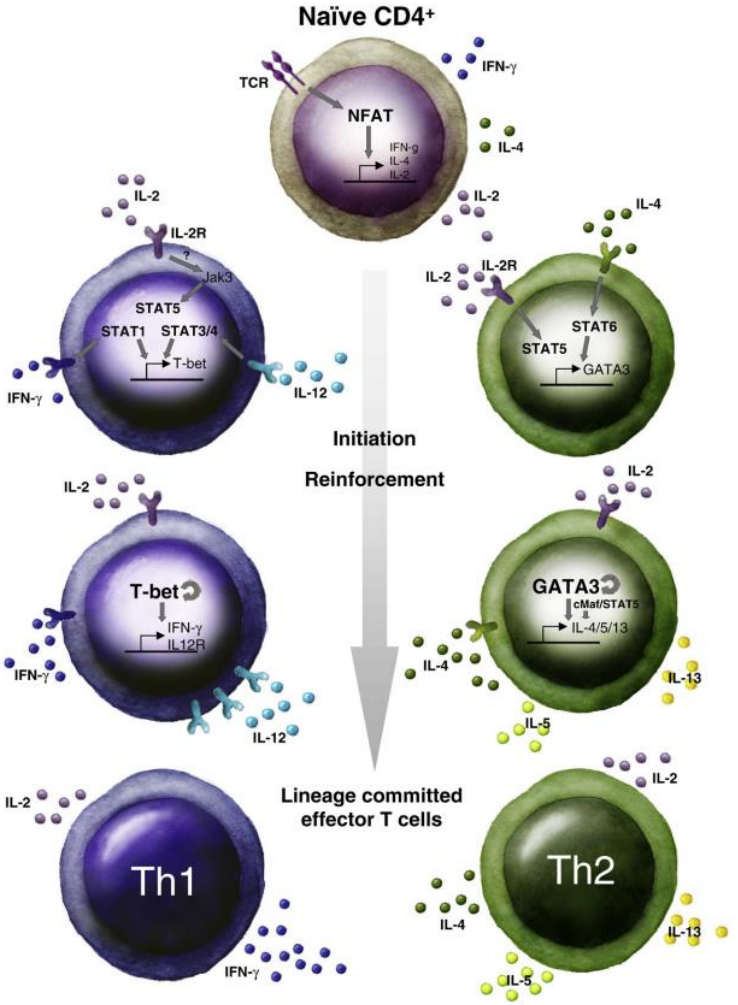
Use methylation at specific CpG sites to predict cellular composition

- **Why does this work?**

Because expression cellular surface protein markers that distinguish cell types (e.g., CD4+ T cells that become Th1 vs. Th2) are controlled epigenetically

- **What do you need to know to make this prediction?**

You need to know the methylation patterns that distinguish one cell type from another: reference set



[Janson et al. Biochimica et Biophysica Acta 2009]

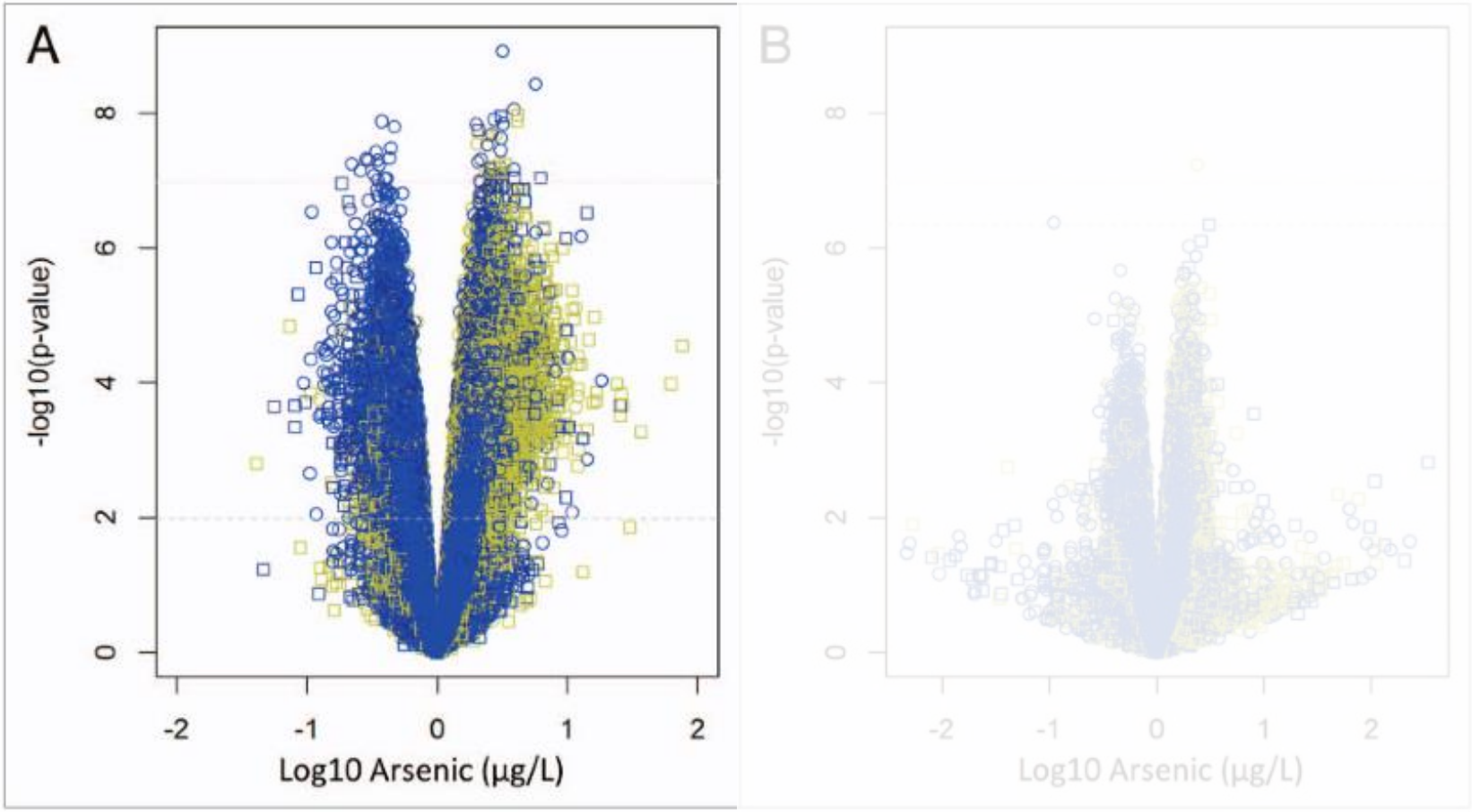
Using the Houseman method to infer underlying cell type mixture (in brief)

Step 1: Create a reference set using Infinium array in homogenous cell samples

Step 2: Fit Validation Model - using a sample where the underlying cell mixture is known (reference set), model estimates of cell counts using methylation values, save coefficients

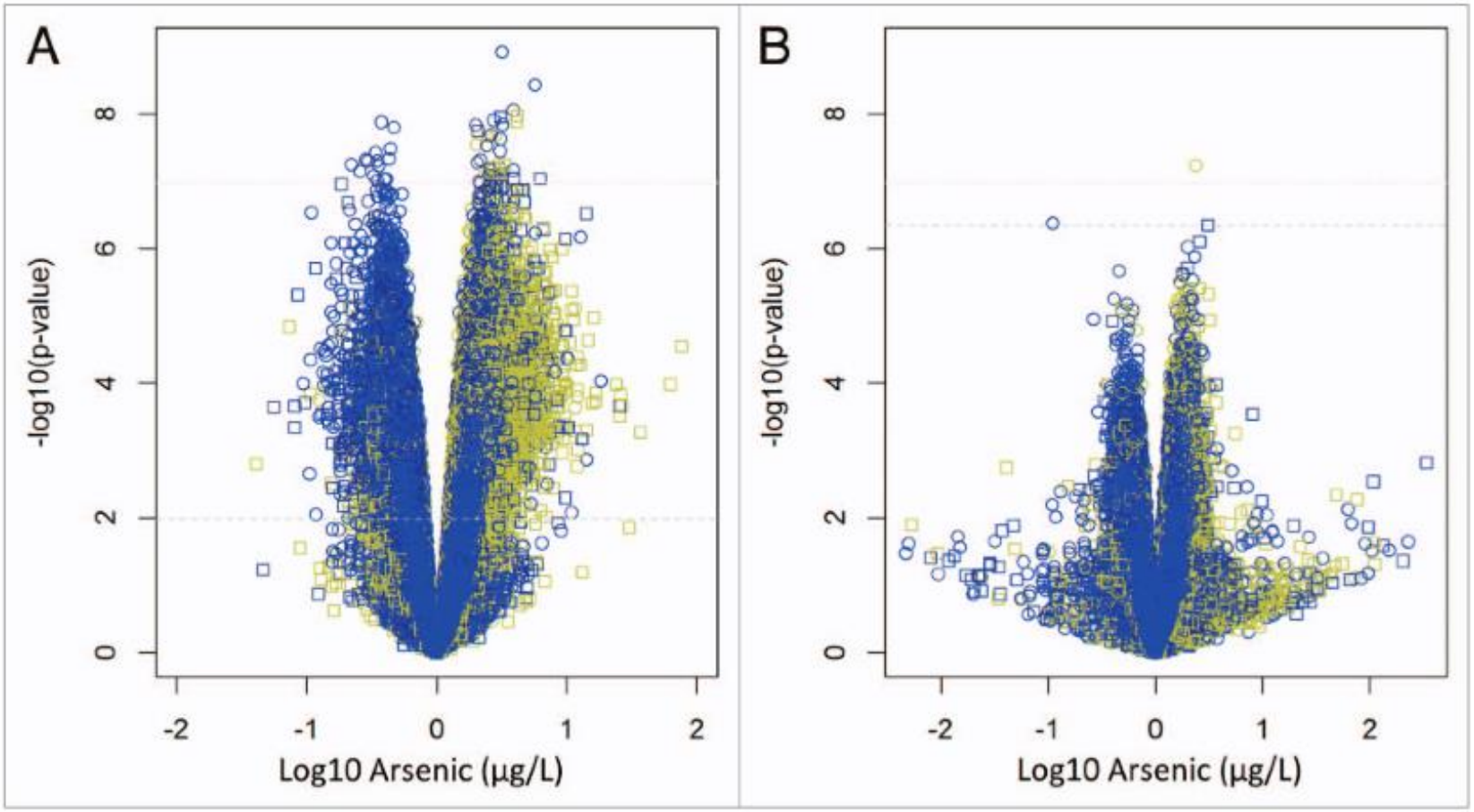
Step 3: Fit Target Model - using the most significant coefficients, estimate the effect of different covariates on the underlying cell mixture to predict the cell mixture for each individual in a target sample

An example:



[Kile et al. 2014]

An example:



[Kile et al. 2014]

The reference sets: 1) Houseman

- Blood was purchased from AllCells®, LLC (Emeryville, CA); analyzed using 27K array

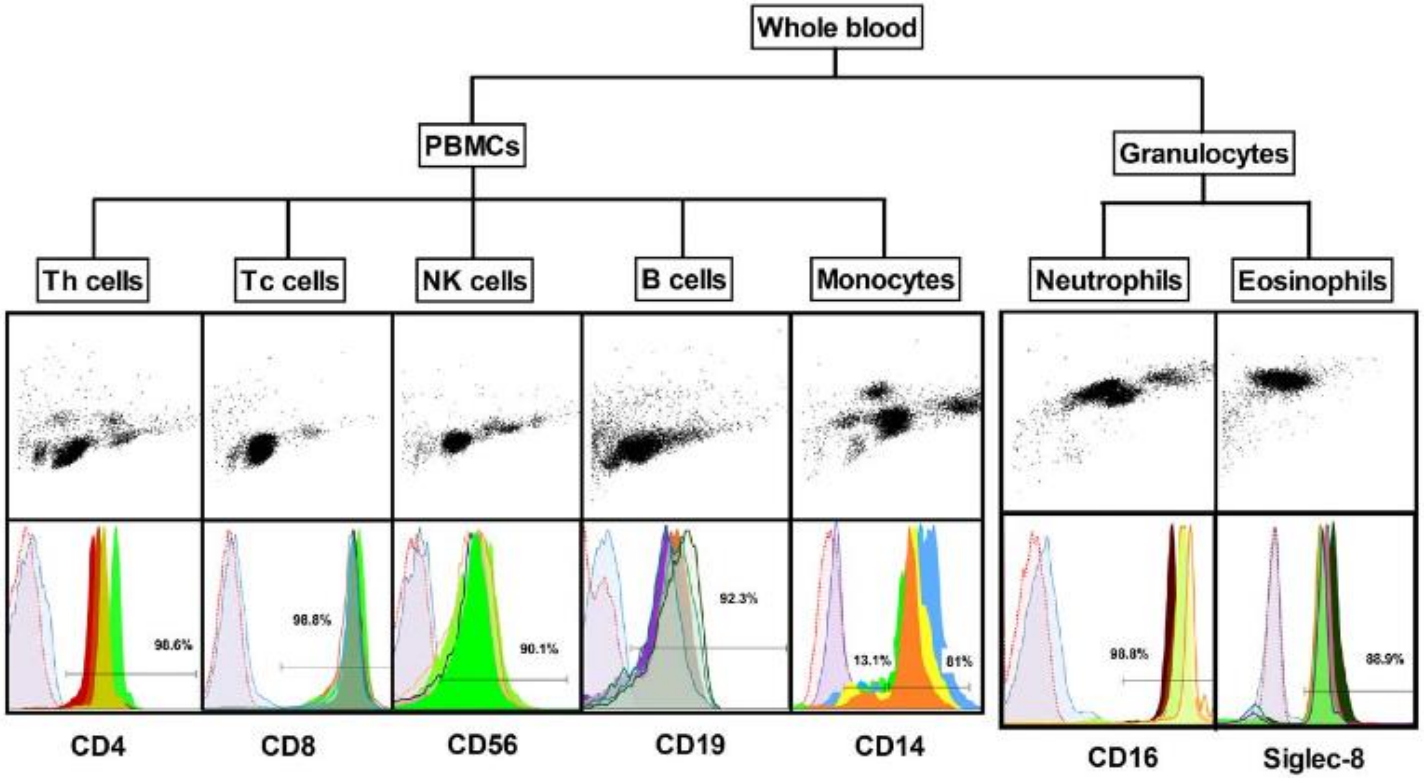
Table 1 Sorted white blood cells in S_0

Short name	Description	Number
B cells	CD19+ B-lymphocytes	6
Granulocytes	CD15+ granulocytes	8
Monocytes	CD14+ monocytes	5
NK	CD56+ Natural Killer (NK) cells	11
T cells (CD4+)[1,2]	CD3+CD4+ T-lymphocytes	8
T cells (CD8+)[1,3]	CD3+CD8+ T-lymphocytes	2
T cells (NKT)[1]	CD3+CD56+ natural killer	1
T cells (other)[1]	CD3+ T-lymphocytes	5

[Houseman et al. 2012]

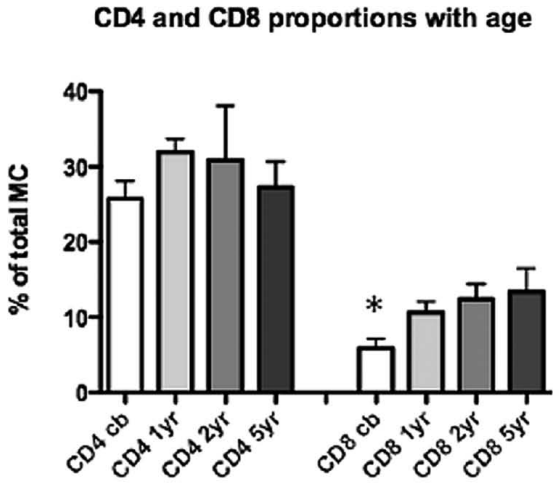
The reference sets: 2) Reinius

- Blood was collected from 6 Swedish males; analyzed using 450K array

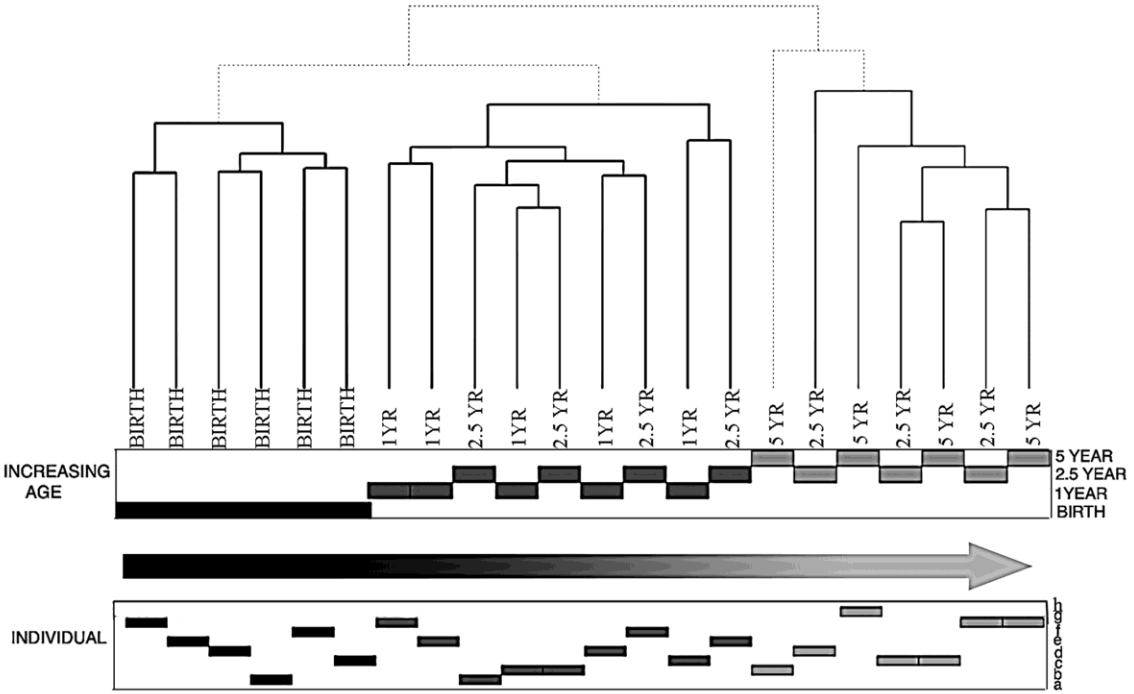


[Reinius et al. 2012]

Differences between cord and adult peripheral blood



shift in cellular distribution by age



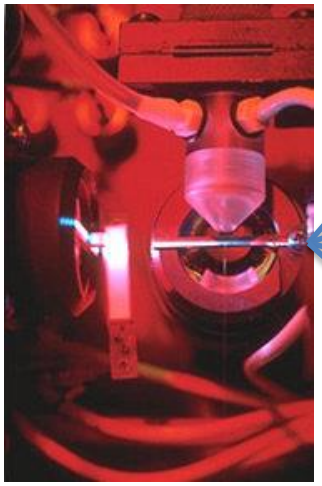
- Within a cell-type, is methylation different?

[Martino et al. Epigenetics 2011]

Generating the reference set from cord blood

cell sort using
flow cytometry

Anonymous Cord Blood



CD4 (T cells/Lymphocytes), n=5

CD14 (Monocytes), n=4

CD15 (Neutrophils), n=5

CD8 (T cells/Lymphocytes), n=5

TWBC (unsorted aliquot), n=5

Infinium
450K
array

Summary

- Many groups are using stored (whole) cord blood from birth cohorts to examine how DNA methylation might mediate prenatal exposure-to-disease relationships.
- Magnitude of the change in DNA methylation associated with exposure is likely to be small; therefore, strategies to improve detection are important
  - increase sample size, improve measurement, control confounding
- Because cell type distribution may confound associations between exposure and methylation, statistical adjustment is often necessary to improve CpG detection.
- Two reference sets necessary for adjustment exist but both are from adult blood; we created a cord-derived reference using the 450K array.
- The cord-derived CD4 cells and adult-derived CD4 cells 'look' different; the cord-derived reference set seems to predict cell distribution from cord blood better than the adult reference.

Next step: Validate the cord reference in an external population where cell distribution is known: PROGRESS cohort (in collaboration with Allan Just, Bob Wright, and Andrea Baccarelli)

Acknowledgements

Funding from:
P01 ES09600/EPA RD-83450901, R01 ES08977R03, R03 CA150140-01, R00 ES017051-04, P01 ES009089, and private foundations including John and Wendy Neu Family Foundation, the Trustees of the Blanchette Hooker Rockefeller Fund, and the New York Community Trust.

Collaborators:
Andrew Ratanatharathorn*
Allan Just
Anne Bozack
Sean T. Glenn
Catherine Tobon
Lehyla Calero,
Deliang Tang

COLUMBIA CENTER FOR CHILDREN'S ENVIRONMENTAL HEALTH

MAILMAN SCHOOL OF PUBLIC HEALTH
COLUMBIA UNIVERSITY