BSI Imaging Genomics Group Discussion

Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology

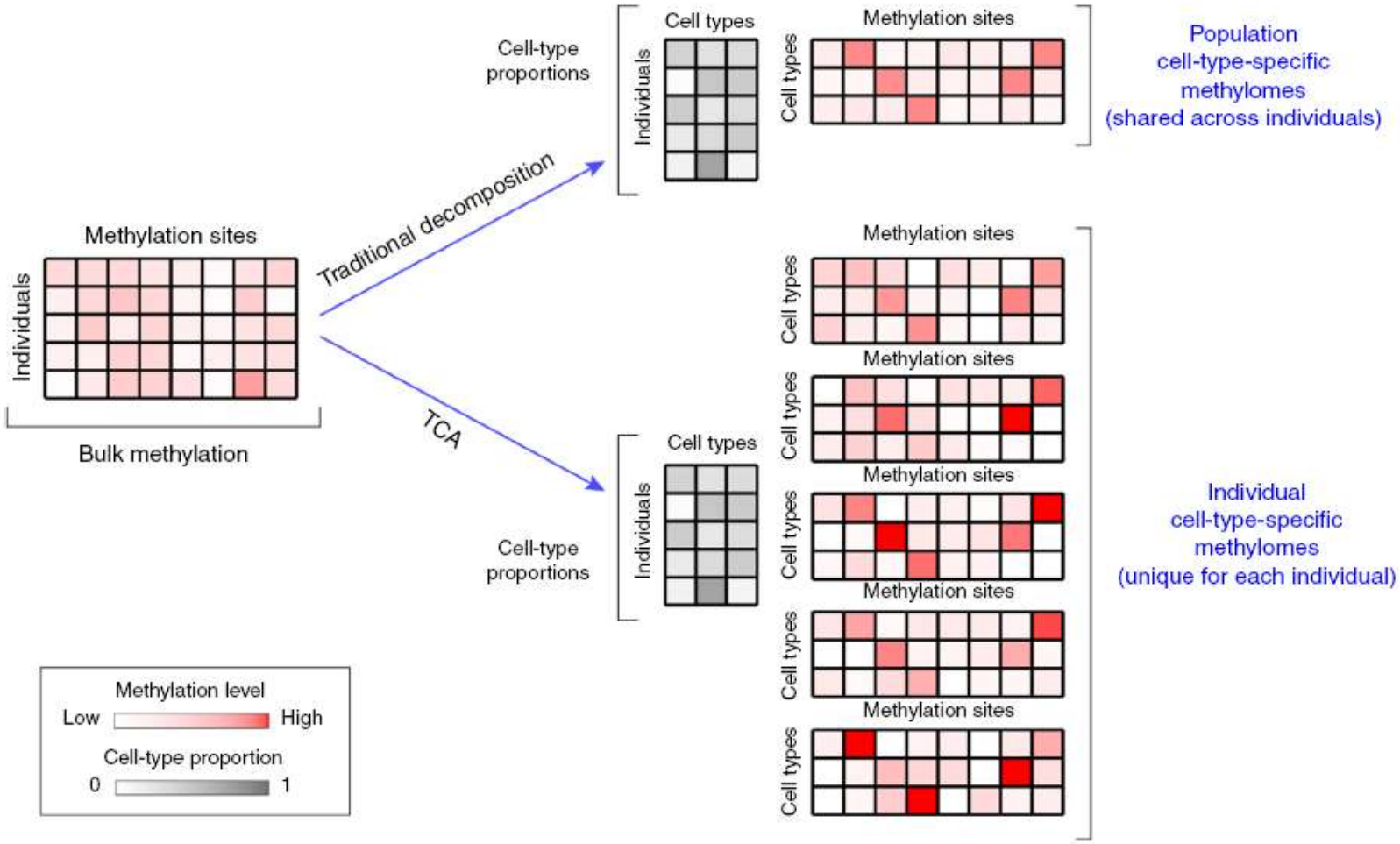
Feb 25, 2021

Jiayu Chen

Outline

- Motivation
  - Cell-sorted or single-cell DNAm profiling still limited in coverage and throughput
  - Make the best use of existing bulk tissue DNAm data (100,000+ samples in GEO)
- Method: tensor composition analysis (TCA)
- Application: apply TCA to a previous large methylation study with rheumatoid arthritis (RA) (CpG associations with case-control status)
- Validation: independent cell-sorted methylation data

TCA


Methylation sites
Individuals
Bulk methylation
Traditional decomposition
TCA
Cell-type proportions
Cell types
Individuals
Methylation sites
Cell types
Population cell-type-specific methylomes (shared across individuals)
Cell-type proportions
Cell types
Individuals
Methylation sites
Cell types
Individual cell-type-specific methylomes (unique for each individual)
Methylation level
Low
High
Cell-type proportion
0
1

TCA

- A tensor of samples by methylation sites by cell-types vs. a typical two-dimensional bulk data samples by methylation sites (**http://github.com/cozygene/TCA**)

$$X_{ij} = \sum_{h=1}^{K} w_{hi} Z_{hj}^{i} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \tau^2)$$

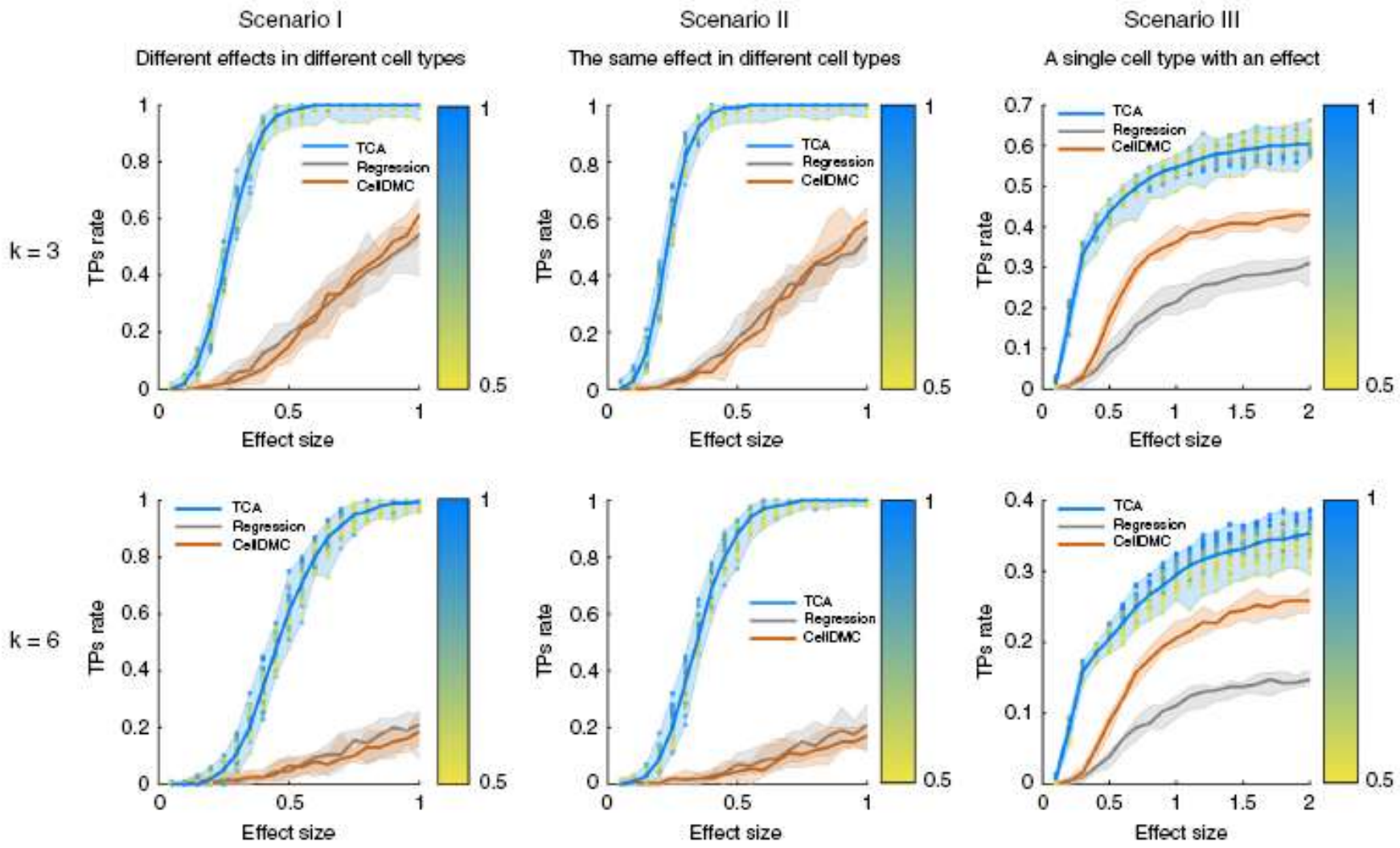
$$Z_{hj}^{i} | \mu_{hj}, \sigma_{hj} \sim N(\mu_{hj}, \sigma_{hj}^2)$$

$$Pr(Z_j^i = z_j^i | X_{ij} = x_{ij}, w_i, \mu_j, \sigma_j, \tau) \qquad Z_j^i = \left(Z_{1j}^i, \ldots, Z_{kj}^i\right)^T$$

$$\hat{z}_j^i = a_{ij} = \left(\frac{w_i w_i^T}{\tau^2} + \Sigma_j^{-1}\right)^{-1} \left(\frac{x_{ij}}{\tau^2} w_i + \Sigma_j^{-1} \mu_j\right) \qquad \Sigma_j = diag(\sigma_{1j}^2, \ldots, \sigma_{kj}^2)$$

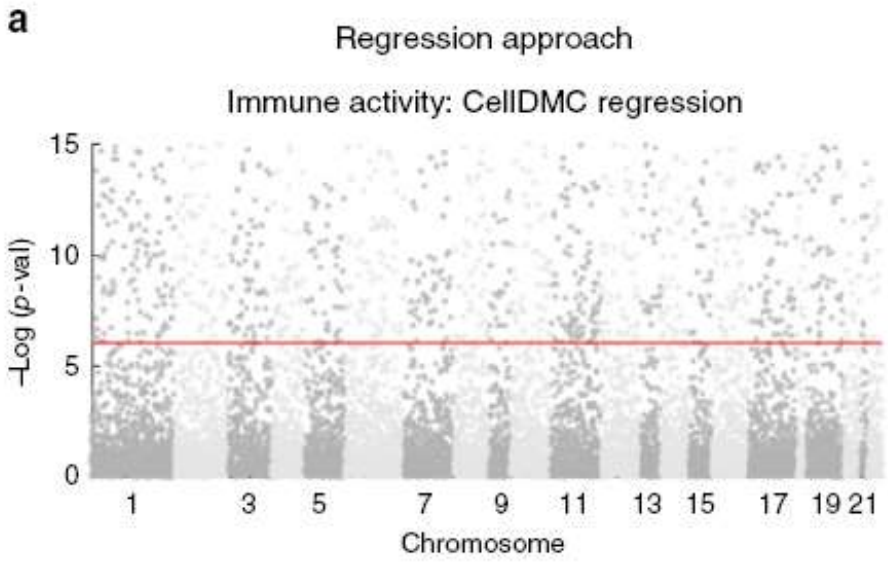
$$Y_i = Z_{lj}^i \beta_{lj} + e_i, e_i \sim N(0, \phi^2)$$

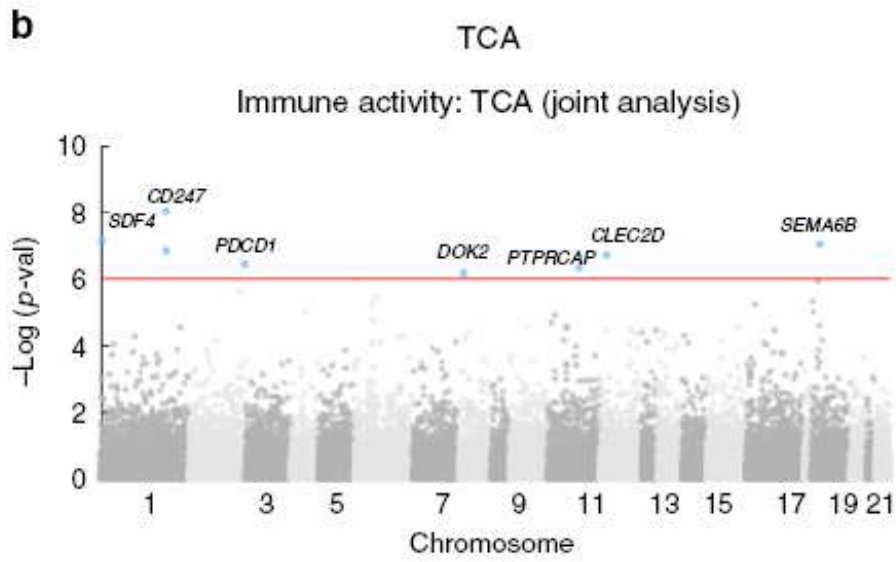
Simulation results



Cell-type-specific differential methylation in immune activity

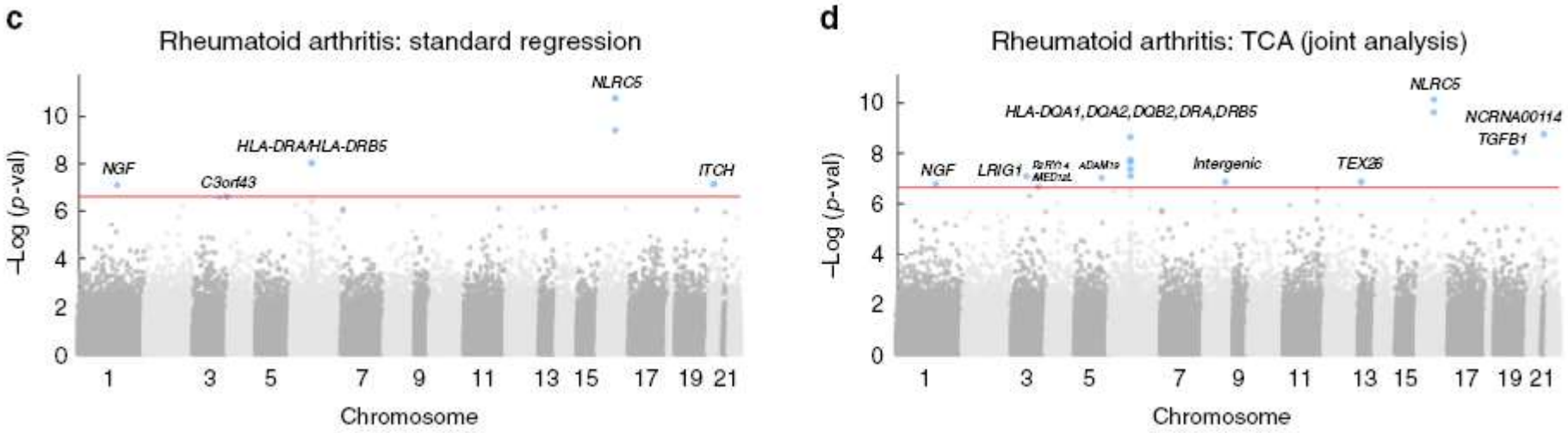
- An extreme case where the phenotype is the celltype composition
  - Defined the level of immune activity of an individual as its total lymphocyte proportion in whole-blood





Cell-type-specific differential methylation in RA

- Regression vs. TCA (joint effects of all cell types)
- 6 vs. 15 epigenome-wide findings
- Validation with cell-sorted data
  - 11 of the 15 CpGs reported by TCA (and 4 of the 6 CpGs reported by a standard regression) had a significant p-value at level 0.05 in at least one of the cell types



Cell-type-specific differential methylation in RA

- Regression vs. TCA (marginal effects of individual cell types)
- 15 cell-type-specific associations with 11 CpGs: 6 associations in CD4+, 8 in CD14+, and one in CD19+ cells
- Validation with cell-sorted data
  - Data 1: 4 of the 6 associations in CD4+ and 4 of the 8 associations in CD14+ had a significant p-value at level 0.05
  - Data 2: of the 4 CD4+ associations verified in Data 1, three associations further replicated in Data 2