Normalizing for Cell Type: Does it Matter?

Implications for neuroepigenetic DNA methylation analysis and biological interpretation

Hilary Brewis, Randip Gill, Samantha Schaffner, Cassia Warren, Lisa Wei
Team: Methylhomies STAT540

Centre for Molecular Medicine and Therapeutics
BC Cancer Agency CARE + RESEARCH
UBC

BACKGROUND

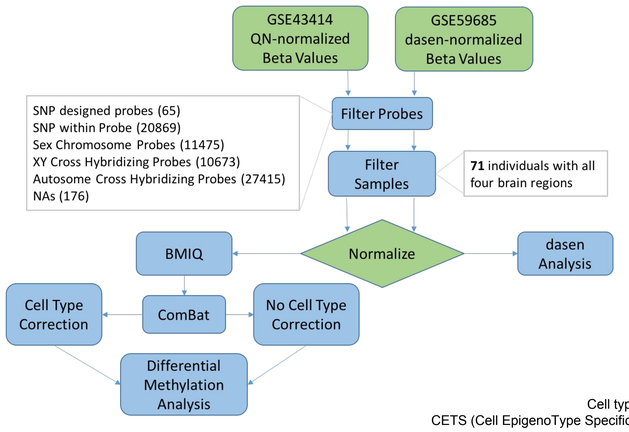
- DNA methylation (DNAm) plays a crucial role in maintaining patterns of gene expression during human development and aging.
- Understanding the human methylome is important for determining both biomarkers for - and direct pathways implicating - health and disease [1].
- Aberrant DNAm patterns have been correlated with common neurodegenerative disorders, including Alzheimer's Disease and Parkinson's Disease [2], as well as with mental disorders such as schizophrenia [3,4].
- Much of the literature on either the diseased or healthy brain methylome fails to separate DNAm data by cell type composition - a major driver of DNAm variability - or by brain region [5-8].

AIM

1. To determine whether cell type correction between brain regions is necessary in the analysis of Illumina HumanMethylation450 BeadChip array data
2. To investigate probes differentially methylated between cerebellum and cortex regions

METHODS

Dataset: Illumina HumanMethylation450 BeadChip Beta Values of 366 brain samples from 122 Alzheimer's Disease patients and controls retrieved from the Gene Expression Omnibus (GSE43414 and GSE59685). The dataset consists of four brain tissue types: entorhinal cortex, frontal cortex, superior temporal gyrus, and cerebellum.

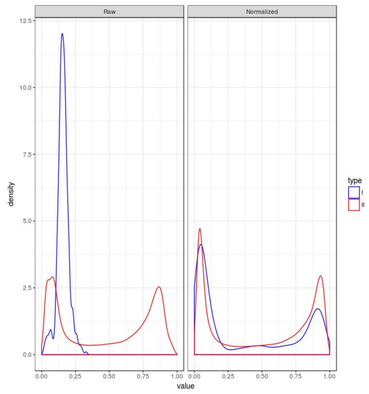


Cell type correction:
CETS (Cell EpigenoType Specific) R Package

RESULTS

1. BMIQ normalization and PCA

Distribution of 10,000 Representative Probes



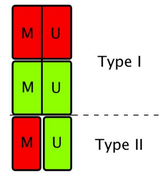
Beta-Mixture Intra-Quantile (BMIQ) Normalization

The GSE43414 dataset was already quantile normalized (QN) when we downloaded it from GEO. This is, a method which corrects for broad inter-sample methylation differences such as those due to tissue.

Wang *et al.* [9] suggest that QN followed by BMIQ to correct for probe type differences is the most stringent workflow for 450K normalization.

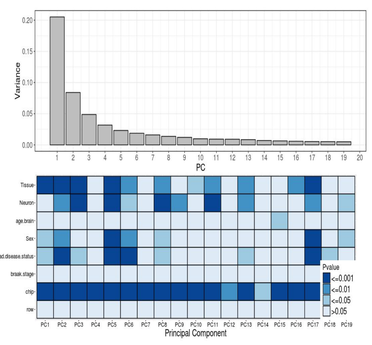
BMIQ [10] was applied to the QN-normalized data:

- Two probe types on the 450K array (I and II) with different chemistry produce different signal intensity distributions
- These must be further normalized in order to compare methylation across all probes
- BMIQ fits a beta mixture model to the type II probes, transforming their distribution into quantiles which resemble the type I distribution

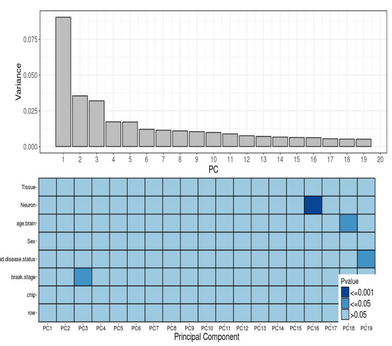


Principal Component Analysis

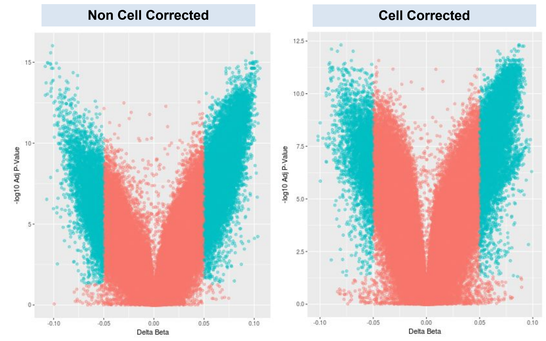
BMIQ-normalized, prior to correction



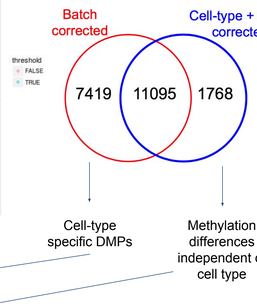
ComBat and CETS-corrected



3. Differential methylation analysis reveals DMPs that are attributed to cell type-specific variations and those that are independent of cell type



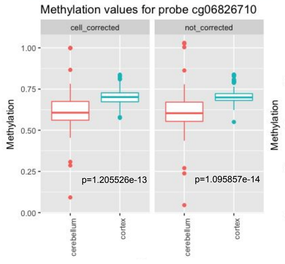
Significant differentially methylated probes between cerebellum and cortex that appeared only in analysis of batch-corrected, cell-type and batch-corrected, and those appearing in both.



Cutoffs: FDR <= 0.05, mean delta beta >= 0.05
Linear model design for DMR discovery: ~brain_region + braak.stage + Sex + Age

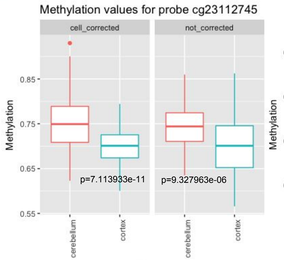
Most significant DMP in non cell-type corrected data

Methylation values for probe cg06826710



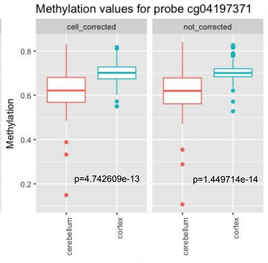
Most significant DMP in cell-type corrected data

Methylation values for probe cg23112745



Most significant DMP found in both data sets

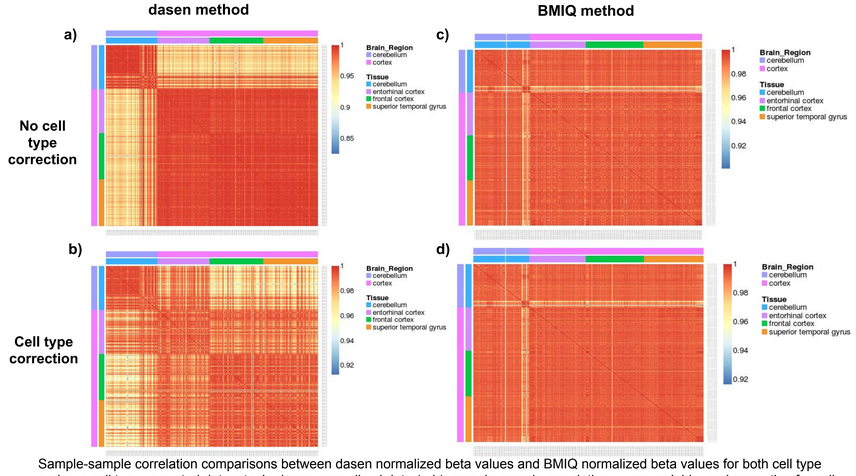
Methylation values for probe cg04197371



4. Known functional annotation of top ranked DMRs

	Unique to ComBat-corrected only			Overlapping ComBat, CETS			Unique to ComBat + CETS-corrected		
DMR Rank	Location	Gene	Function	Location	Gene	Function	Location	Gene	Function
1	chr7: 3221150-3221373	JD277612 mRNA	C9orf64: potential peroxisomal protein	chr11: 674623-674687	DEAF1	ZF TF; embryonic development and mental retardation	chr3: 42132870-42132991	TRAK1	Endosome-lysosome trafficking; hypertonia, HIV life cycle
2	chr1: 4079006-4079286	intergenic, between JD187786-JD366482 mRNAs	TONSL: NF-kappa-B mediated transcription	chr10: 131209478-131209623	intergenic, downstream of JD446374 mRNA	BRCA1: transcription, DSB repair, breast cancer	chr13: 103531040-103531058	intergenic, between ERCC5 and METTL21EP	ERCC5: DNA excision repair; xeroderma pigmentosum METTL21EP: lysine methyltransferase pseudogene
3	chr1: 21993401-21993555	RAP1GAP	Cell growth, differentiation	chr14: 101157950-101158159	JD466515 mRNA	EPB41L2: actin binding, tight junctions, synapse interaction	chr14: 72943461-72943540	RGS6	Neuronal GPCR signaling; early-onset AD, alcoholic cardiomyopathy
4	chr20: 62097336-62097681	KCNQ2	Neuronal potassium channel involved in epilepsy disorders	chr16: 87391861-87392076	FBX031	Ubiquitination and degradation, cell cycle, dendrite growth, neuronal migration	chr17: 5988176-5988249	WSCD1	Sulfotransferase; behaviour/neurological mouse phenotype *Differential expression cortex/cerebellum*

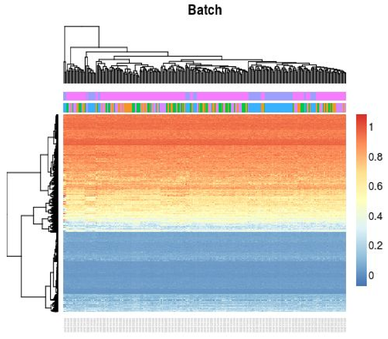
Normalization method matters: BMIQ significantly improves initial sample-sample correlations compared to dasen



Sample-sample correlation comparisons between dasen normalized beta values and BMIQ normalized beta values for both cell type and non-cell type corrected datasets. In dasen normalized data (a-b), sample-sample correlations were variable, and correcting for cell type improved these correlations. However, quantile normalization (QN) followed by beta-mixture quantile normalization (BMIQ) significantly improved the initial sample-sample correlations.

2. Batch and cell type correction reduces sample variation caused by cell type

Batch

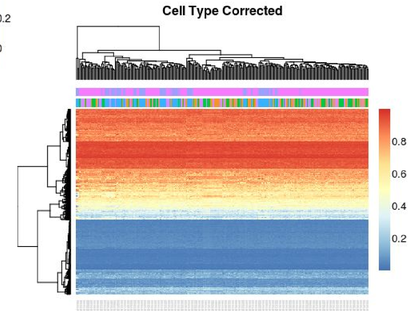


Heatmaps of Beta values for 1000 randomly selected probes

It was hypothesized that cell type correction would reduce sample clustering according to brain region and increase the amount of clustering according to tissue type.

Cell type correction did slightly decrease sample clustering according to brain region, but there was no notable increase in the clustering according to tissue type.

Cell Type Corrected



CONCLUSIONS

- BMIQ normalization is more suitable for cross-sample comparisons
- Technical and biological variability to account for are batch effects and cell type variation across brain regions to allow for comparison between tissues
- Differential methylation analysis reveals DMPs that are attributed to cell type variation and those that are independent of cell type variation
- Failing to correct for cell type may result in identification of false positive DMPs/DMRs, with implications for biological conclusions concerning clinically relevant loci

References:
[1] Bernstein *et al. Cell* (2007). [2] Sanchez-Mut *et al. Transl Pysch* (2016). [3] Huang *et al. PloS one* (2007).
[4] Huang *et al. Journal of Neuroscience* (2007). [5] Shin *et al. Nat Neurosci* (2014). [6] Jaffe *et al. Genome Biology* (2014).
[7] Guintivano *et al. Epigenetics* (2013). [8] Montano *et al. Genome Biol* (2013). [9] Wang *et al. Epigenetics* (2015). [10] Teschendorff *et al. Bioinformatics* (2013).