An Exploration of DNA Methylation in Stem vs. Somatic Cells, and How It Relates to Gene Expression

Nathaniel Lim[1], Jessica Lee[1], Analise Hofmann[1], Howard Cen[1], Scott Brown[1]

[1]Genome Science + Technology Program, University of British Columbia, Vancouver, British Columbia, Canada.

Introduction

Epigenetic factors such as DNA methylation are an essential mechanism in regulating gene expression. DNA methylation can directly prevent transcription factor binding and lead to changes in chromatin structure that restrict access of transcription factors to the gene promoter[1]. In practice, the Pearson correlation coefficient between the methylation level of the DMR and the expression of its associated gene is around -0.3[2]. DNA methylation is an important component in numerous cellular processes, including embryonic development, genomic imprinting, X-chromosome inactivation, and preservation of chromosome stability[1].

In a previous study, global analysis of 205 Human pluripotent stem cells (hPSC) and 130 somatic samples identified DNA methylation differences between somatic and pluripotent cells[3]. Tissue-specific DNA demethylation occurs during differentiation. Specific tissue types were distinguished by unique patterns of DNA hypomethylation.

Aims

- Investigate different methods of assigning methylation to a gene, and observe how this affects the correlation with gene expression
- Determine if known tissue-specific genes are identified in this dataset
- Test if there is any difference in the correlation between somatic and stem cells
- Test whether the correlation is different for genes which are differentially methylated and expressed

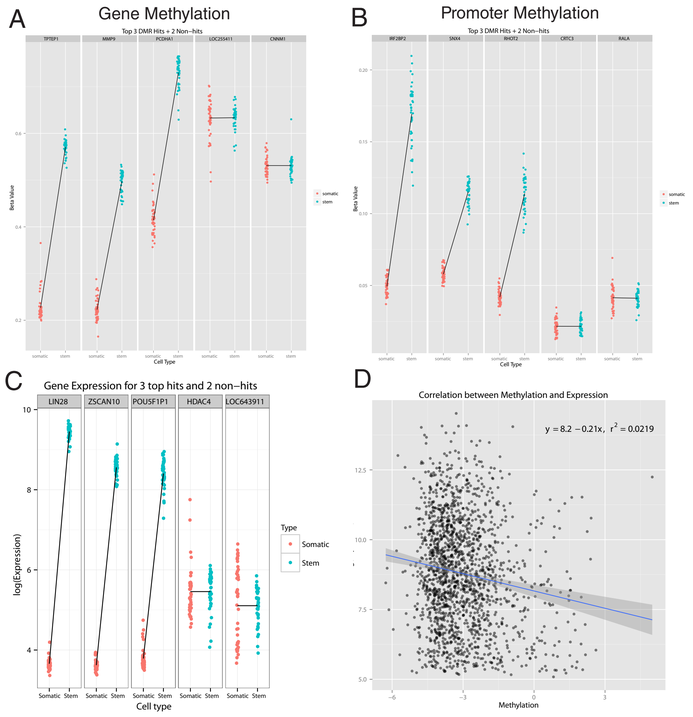

Figure 1 Examples of the data. (A) Example of differentially and non-differentially methylated genes, using average gene methylation. (B) Example of differentially and non-differentially methylated genes, using promoter methylation average. (C) Example of differentially and non-differentially expressed genes. (D) Example of the correlation between promoter methylation and expression.

Methods

1. 87 samples with both Illumina 450K Methylation data and Illumina HumanHT-12 V3.0 expression beadchip data
2. Divided these samples into two groups, somatic cells (n = 46) and stem cells (n = 41).
3. Assigned single methylation value to a gene by taking the average methylation of the entire gene, or average of the promoter region
4. Performed differential methylation and differential expression analysis (limma) between the somatic and stem cell samples using the following model:

$$Y = \mu_{STEM} + \tau_{SOMATIC} + \varepsilon$$

   Examples of hits and non-hits are shown in Figure 1
5. Investigated the correlation between expression & methylation for the differential analysis hits
6. Performed GO term enrichment using GOrilla[4]

Results

- There is a slight negative correlation between methylation and expression (Figure 1D)
- Promoter-level differential methylation shows lower methylation levels than gene-level differential methylation (Figure 2)

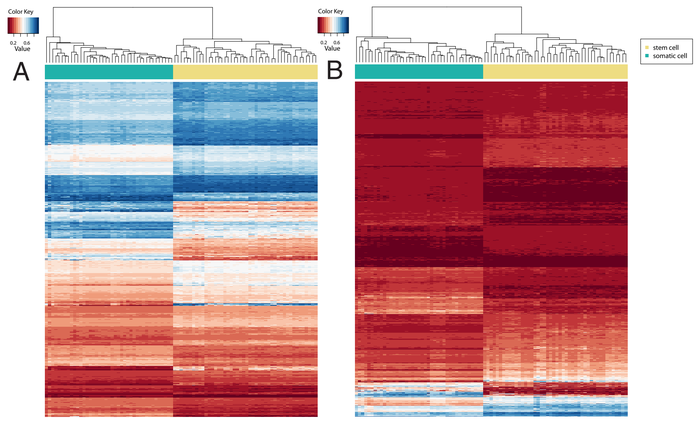

Figure 2 Heatmaps of methylation for top 1000 genes. (A) Top 1000 differentially methylated genes using average gene methylation as measure. (B) Top 1000 differentially methylated genes using average promoter methylation as measure.

- 13,436 genes had both expression and methylation data, 6,724 of these lack promoter CpG probes (Figure 3)

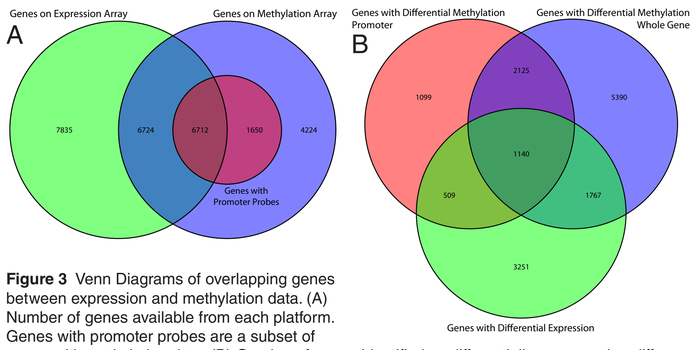

Figure 3 Venn Diagrams of overlapping genes between expression and methylation data. (A) Number of genes available from each platform. Genes with promoter probes are a subset of genes with methylation data. (B) Overlap of genes identified as differentially expressed or differentially methylated at an FDR < 1 x 10^{-5}. There are many differences between the two methods of measuring gene methylation.

- Overall, stem cells have stronger correlation between expression and methylation than somatic cells (Figure 4)
- Differentially methylated and expressed genes show decreased correlation between expression and methylation (Figure 4)

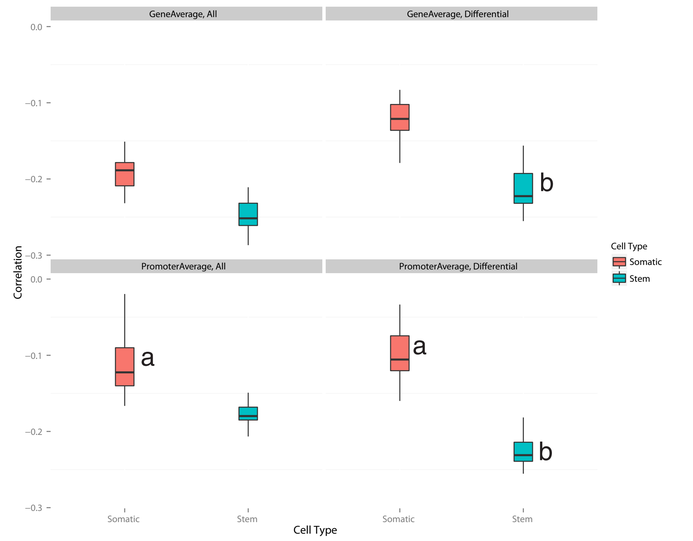

Figure 4 Boxplots for comparison of correlations between various groups. The top plots use gene-level methylation, whereas the bottom use promoter-level. Plots on the left are for all genes, plots on the right are for differentially methylated and expressed genes (FDR < 1 x 10^{-5}). All single variable comparisons are significant (P < 0.01) except pairs (a) and (b).

- Differentially expressed genes show enrichment for stem cell related GO processes, while the intersect with differential methylation does not (Table 1)

Expression			**All Methylation Intersect Expression**			**Promoter level Methylation*Expression**		
Process	p.value	q.value	Process	p.value	q.value	Process	p.value	q.value
cell fate specification	5.3 e -8	4.9 e -4	negative regulation of nitrogen compound metabolic process	1.3 e -7	1.1 e -3	negative regulation of mitosis	7.0 e-4	1.0
stem cell maintenance	5.1 e -7	2.3 e -3	negative regulation of nucleobase-containing compound metabolic process	1.4 e -7	5.5 e -4	negative regulation of nuclear division	7.0 e -4	1.0
transcription, DNA-templated	7.3 e -7	2.2 e -3	cell fate specification	1.1 e -6	3.0 e -3	pyridine nucleotide biosynthetic process	9.6 e -4	1.0
All Methylation			**Intersect of All**			**Gene Average level Methylation*Expression**		
Process	p.value	q.value	Process	p.value	q.value	Process	p.value	q.value
regulation of keratinocyte migration	3.7 e -4	1.0	response to osmotic stress	1.4 e -4	7.5 e -1	positive regulation of catenin import into nucleus	6.5 e -5	4.9 e -1
positive regulation of kertinocyte migration	3.7e -4	1.0	positive regulation of interleukin-8 production	2.8 e -4	7.3 e -1	negative regulation of cellular macromolecule biosynthetic process	1.4 e -4	5.4 e -1
negative regulation of interleukin-8 biosynthetic process	4.1 e -4	1.0	regulation of interleukin-8 production	5.3 e -4	9.3 e -1	negative regulation of macromolecule biosynthetic process	2.4 e -4	6.1 e -1

Table 1 Top three GO Process hits for different gene list intersections. Differentially expressed genes yield expected results, while subsetting this list with differentially methylated genes do not.

Conclusions

- Methylation does not add any tissue-specific information over and above expression
- This is shown by lack of relevant GO enrichment, and weaker correlation for differentially methylated and expressed genes
- With respect to gene expression, gene-level methylation appears to be a better metric to summarize the methylation state of a gene

References & Acknowledgements

1. Phillips, T. Nature Education 2008.
2. Vanderkraats et al. Nucleic Acids Research 2013.
3. Nazor et al. Cell Stem Cell 2012.
4. Eden et al. BMC Bioinformatics 2009.

We would like to thank Paul and Gloria for their assistance and advice on this project, Jenny, Gabby, and Shaun for their help in the class, and Rachel for recommending the dataset.

genome science + technology
GRADUATE PROGRAM

UBC