DNA methylation changes in human neural tube defects

Michelle Cui, Sarah Goodman, Alexandre Lussier, Sam Peeters, Magda Price, Gina Zhong

Background

- DNA methylation at CpG dinucleotides is involved in regulating gene expression and plays a key role in embryonic development, genomic imprinting and X chromosome inactivation
- Folate is a critical dietary methyl donor for methylation of DNA, proteins and lipids
- Incidence of neural tube defects (NTDs), including spina bifida (sb) and anencephaly (an), was reduced by 46% following fortification of grains with folic acid in Canada, suggesting that aberrant DNA methylation in fetuses contributes to the complex etiology of NTDs

OBJECTIVE: to investigate differential DNA methylation in human tissues from fetuses with spina bifida and anencephaly vs. controls

Experimental Design

- A total of 35 control, an, and sb individuals across five different tissue types (134 total samples)
- Disease groups were sex and gestationally age matched (14.5-23.7 wks) with controls

Table 1 Data Set

group	sex	br	kid	mus	sc	v	total
con	F	4	4	4	3	5	**20**
	M	6	6	6	5	6	**29**
an	F	0	7	6	5	8	**26**
	M	0	2	3	2	3	**10**
sb	F	1	3	3	3	3	**13**
	M	8	7	7	7	7	**36**

Tissue: brain(br), kidney(kid), muscle(mus), spinal cord(sc), placental villi(v)
Group: control(con), anencephaly(an), spina bifida(sb)
Sex: male(M), female(F)

Platform and Data Exploration

Table 2 Data Snippit

	Probe ID	FT13.br	FT13.kid	FT13.mus
1	cg00000029	2.807727778	0.950464501	1.269499254
2	cg00000108	4.119484331	3.560117910	3.029908421
3	cg00000109	3.669679064	3.450323026	2.363046138
4	cg00000165	-3.363771266	-2.603796907	-2.928438930
5	cg00000236	3.754887502	2.526747887	2.774121090
6	cg00000289	0.835225574	-0.194542113	0.581516696
7	cg00000292	1.476476595	-0.070439593	-0.081888208
8	cg00000321	-0.588933261	-1.768740759	-1.438950370
9	cg00000363	-3.419078119	-2.886269538	-2.069116767

- Illumina HumanMethylation450 beadchip array (485,512 probes/CpGs)
- The unit of measure reported per probe is the M value* (log2 ratio of the methylated to unmethylated probe intensities)
- M values can be converted to β values for biological interpretation, $\beta=2^M/(2^M+1)$

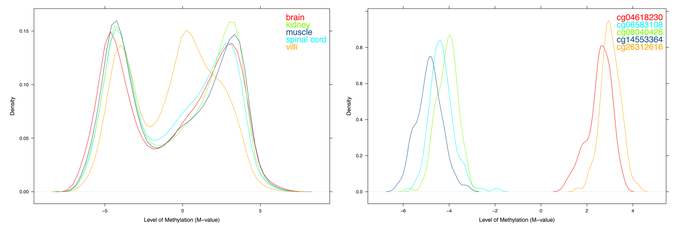


Figure 1. Bimodal distribution of M-values at all probes in each tissue (left). Distribution of M-values of five individual probes in all samples (right).

Removal of Sample Outliers

- consistently different samples were identified as outliers, based on quality control measures and clustering
- 3 samples were identified as outliers and removed

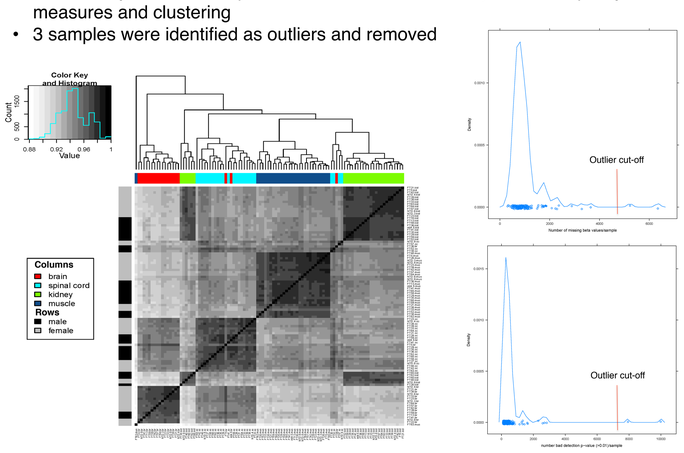


Figure 2. Identifying Sample Outliers. Correlation heatmap of samples with unsupervised clustering (left). Density plot of number of missing β-values per sample (top right). Number of bad detection p-values (quality score) per sample (bottom right).

Data Filtering

- Removal of sex chromosome probes, cross-hybridizing probes, polymorphic probes
- Removal of probes with missing β-values (in >5% of samples) and probes with bad detection p-values (in >1 samples)

485, 512 probes ⟶ 405, 010 probes

Normalization

Using functions built into the *lumi* R Bioconductor package, background normalization and colour correction across plates was performed to account for batch effects, followed by SWAN normalization

Model Selection

Does any individual have consistently high or low methylation?

- Each individual contributes multiple tissues to the data set and thus could skew group differences

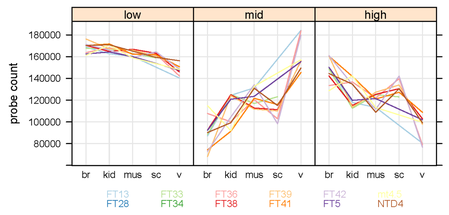


Figure 3. Methylation status of control samples. Probes were divided into three groups based on their methylation scores (<20%, 20-80%, and >80%) and plotted by tissue type for each individual.

- This is an interesting question since there have not been any studies on intra-individual differences in methylation
- Examining one individual across all tissues showed that there were no obvious 'high-methylators' or 'low- methylators' in the data set
- This was also done in spina bifida and anencephaly groups (data not shown)

Do we model all samples together?

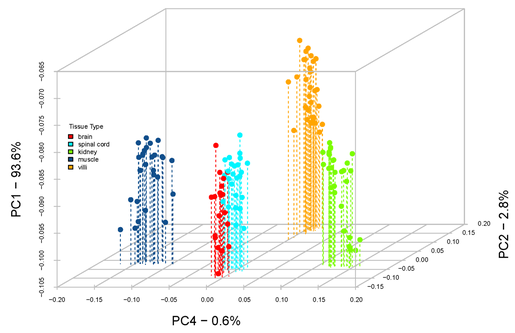


Figure 4. Unsupervised PCA. Principal components 1, 2, and 4 were plotted against each other to illustrate differences between the samples.

- > 95% of data variability is related to tissue type
- PC1 separated the samples into somatic vs extra-embryonic tissue
- PC4 further differentiated samples by germ layer origin (mesoderm – muscle, ectoderm – brain and spinal cord, endoderm – kidney, and extra-embryonic – villi)
- Due to the large proportion of variability attributed to tissue type, we will model group differences in DNA methylation within each tissue

What covariates are significant?

- Testing whether sex and gestational age had a significant effect
- A linear model was fit to each tissue/disease for several combinations of variables
- Approximately 45 unique models were applied to the data sets

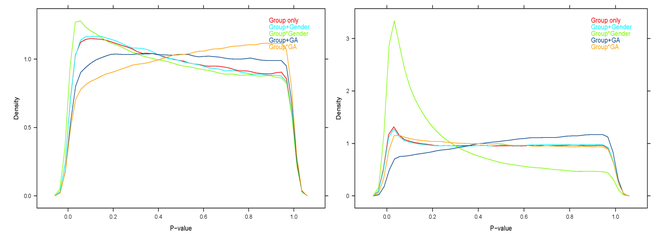


Figure 5. Examples of unadjusted p-value distribution plots. Kidney tissue models in anencephaly (left).Kidney tissue models in spina bifida (right).

- Models with the most weight around lower p-values were selected
- Note that methylation differences between anencephaly and spina bifida were not modeled

Table 3. Models by tissue/disease A total of 9 individual models were selected for spina bifida and anencephaly in each tissue type

Tissue	Anencephaly	Spina Bifida
brain	N/A	~Group*GA
spinal cord	~Group + Gender	~Group
kidney	~Group*Gender	~Group*Gender
muscle	~Group*Gender	~Group + Gender
villi	~Group*Gender +Group*GA	~Group + Gender

Differentially Methylated Probes

What were the criteria for differentially methylated probes?

- No probes were significant (p<0.05) after Benjamini-Hochberg correction
- Thus, differentially methylated probes had an unadjusted p-value<0.001
- p-values most likely did not stand up to FDR correction due the size of the data set

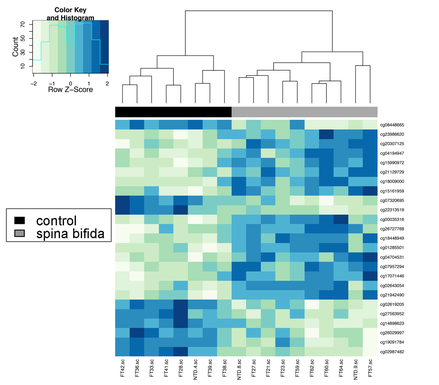


Figure 6. Heatmap displaying top 25 probes in spina bifida and controls in the spinal cord tissue.

How many probes were differentially methylated in each tissue?

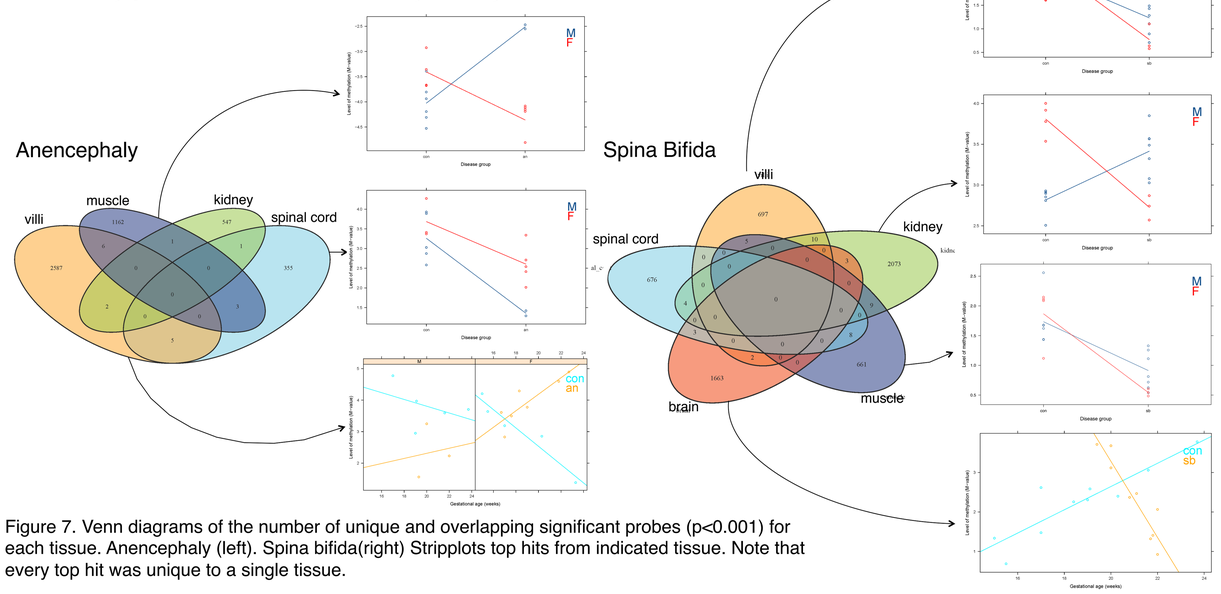


Figure 7. Venn diagrams of the number of unique and overlapping significant probes (p<0.001) for each tissue. Anencephaly (left). Spina bifida(right) Stripplots top hits from indicated tissue. Note that every top hit was unique to a single tissue.

Bioinformatic Analysis

Functional Enrichment

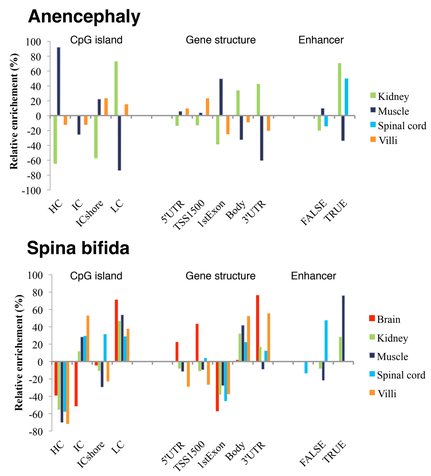
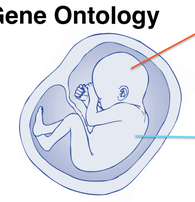


Figure 8. Relative enrichment of hit function. The percent distribution of hits in each tissue was compared to the distribution of all probes in the dataset. Significant differences were determined by Chi squared test (p<0.001) and graphed for each disease (left).

Differences in the enrichment profiles of both diseases suggests a different etiology for anencephaly and spina bifida.

Gene Ontology



BRAIN
- neuron fate commitment (1.45E-04)
- regulation of osteoblast differentiation (4.63E-04)
- negative regulation of neuron differentiation (4.94E-04)
- embryonic digestive tract development (6.93E-04)
- regulation of gliogenesis (5.93E-04)

SPINAL CORD
- embryonic skeletal system morphogenesis (2.62E-12)
- gland morphogenesis (8.38E-12)
- spinal cord development (8.46E-12)
- cell differentiation in spinal cord (8.40E-12)
- proximal/distal pattern formation (1.21E-11)

Figure 9. ErmineJ pathway enrichment. ROC analysis was performed with ordered p-values for all probes and top 5 are displayed for spina bifida. Pathways with p<0.001 were considered significant after correcting for multifunctionality (n=99 for spinal cord and n=260 for brain).

Conclusions

1. There are no large differences between NTD samples and controls after multiple correction testing
2. In general, the differences that were identified were not common between tissues or disease groups
3. Interesting candidate regions and pathways were identified in the bioinformatics analysis, which would warrant further investigation

Improvements:

1. Reducing the data's dimensionality could improve statistical significance of the results, namely by the removal of non-variable probes *a priori* or grouping probes by genomic location
2. Increased sample size would also increase the study's power