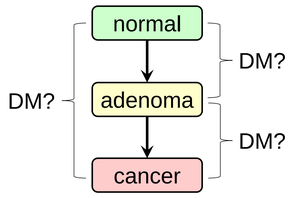
Identifying aberrant methylation patterns underlying colorectal cancer progression

Beryl Zhuang ✦ Eva Yap ✦ Ka Ming Nip ✦ Rashedul Islam ✦ Santina Lin

The University of British Columbia, Vancouver, BC, Canada

Research Question

Aberrant DNA methylation can lead to malignancy by hyper-methylation of CpG islands (CGIs) resulting in transcriptional silencing of tumor suppressor genes. CGIs are sequences with high CpG fractions (>50%) located within gene promoters and methylation at these sites promotes association of methyl-binding proteins and subsequent recruitment of transcriptional repressors[1,2].



Colorectal cancer (CRC) accounts for the second highest cancer-related mortality among men and third among women in Canada, and it progresses from precursor lesions such as adenomas[1,3]. For our project, we compared methylation patterns between normal mucosa, adenoma, and colorectal tumor. By identifying differentially methylated (DM) CGIs between these three groups, we hope to determine aberrant methylation underlying CRC progression.

Dataset

Table 1. Illumina HumanMethylation450 array dataset of 147 colon samples from the Gene Expression Omnibus GSE48684.

Group	Description	Sample Size
normal-H	normal colon from healthy individuals	17
normal-C	normal colon from CRC patients	24
adenoma	colon adenoma	42
cancer	tumor colon from CRC patients	64

Data Normalization

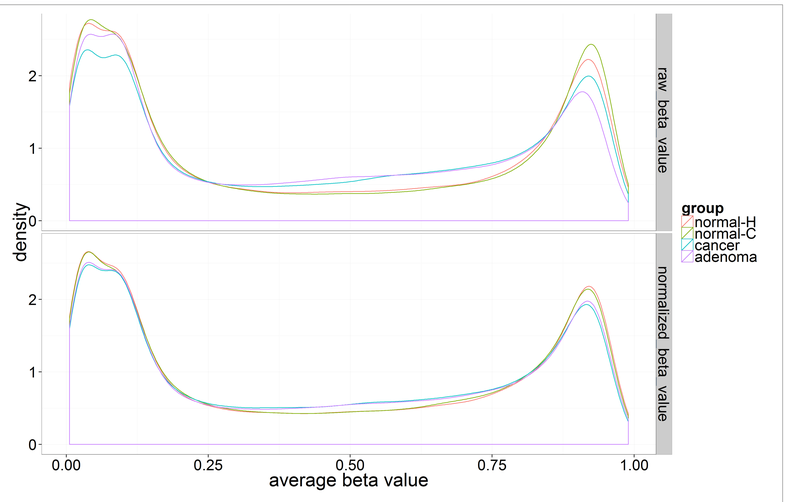


Figure 1. Density plot of average beta values for each probe across all samples before and after quantile normalization.

Acknowledgements
UBC 2015 W2 STAT 540 instructors and teaching assistants.

References
1. Li, J. et al. "Epigenetic Biomarkers: Potential Applications in Gastrointestinal Cancers." ISRN gastroenterology. Mar 6; 2014:464015.
2. Luo, Y. et al. "Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer." Gastroenterology. 2014 Aug; 147(2):418-29.e8.
3. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2014. Toronto, ON: Canadian Cancer Society; 2014
4. Yasukochi, Y. et al. "X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils." Proc Natl Acad Sci U S A. 2010 Feb; 107(8):3704-9.
5. Ostman, A. et al. "Protein-tyrosine phosphatases and cancer." Nat Rev Cancer. 2006 Apr; 6(4):307-20.
6. Derynck, R. et al. " TGF-β signaling in tumor suppression and cancer progression." Nat Genet. 2001 Oct; 29(2): 117-29.
7. Garcia-Lora, A. et al. "MHC Class I Antigens, Immune Surveillance, and Tumour Immune Escape." J Cell Physiol. 2003 Jun; 195(3): 346-55.

Exploratory Analysis

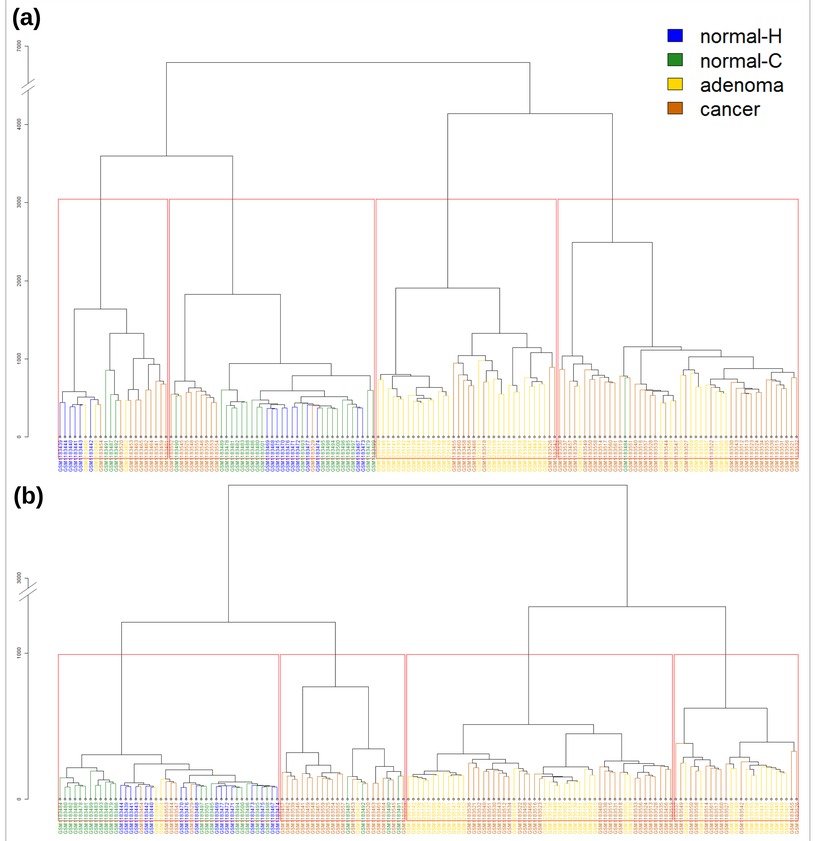


Figure 2. Unsupervised hierarchical clustering using Ward's method on **(a)** raw beta values after removing probes in chrX and those not in CGIs and **(b)** normalized beta values averaged across each CGI[4].

Functional Enrichment Analysis

Table 2. Enrichment of GO terms in differentially methylated CGIs according to Fisher's exact test and Kolmogorov-Smirnov test.

GO.ID	Term	Annotated	Significant	Expected	Rank in classic Fisher	Classic Fisher	Classic KS
GO:0003674	molecular_function	15366	11	9.33	106	0.330	< 1e-30
GO:0005001	transmembrane receptor protein tyrosine phosphatase (TR-PTP) activity	98	0	0.06	123	1.000	1.1e-11
GO:0046332	SMAD binding	128	0	0.08	124	1.000	1.8e-10
GO:0042288	MHC class I protein binding	10	0	0.01	126	1.000	0.0094

- TR-PTPs carry out tumour-suppressing activity through antagonizing RTK signaling and inactivation of TR-PTP has been implicated in CRC[5].
- SMAD proteins are transcriptional activators and their suppression leads to the progression of various cancer types including CRC[6].
- Most cancer cells express low levels of MHC class I molecules to evade immune clearance[7].
- Enrichment of GO terms like molecular function and binding (not shown) are uninformative and likely caused by over-citations.

Differential Methylation Analysis

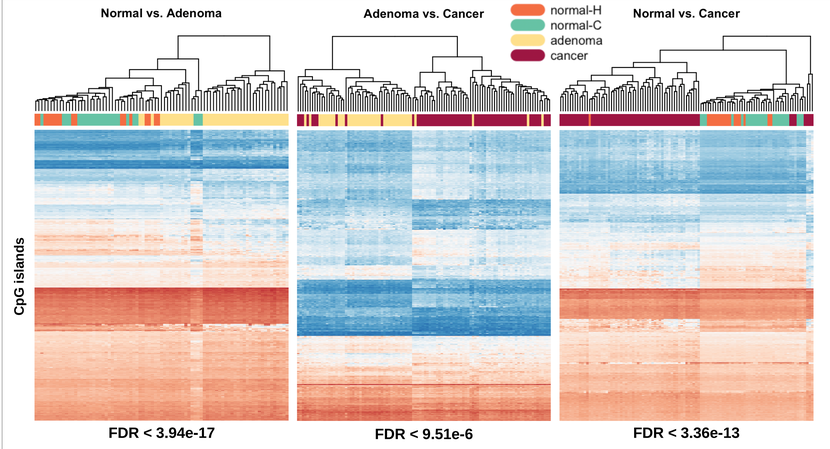


Figure 3. Heatmaps of M values of top 450 hits from Bioconductor Linear Models for Microarray Data (limma).

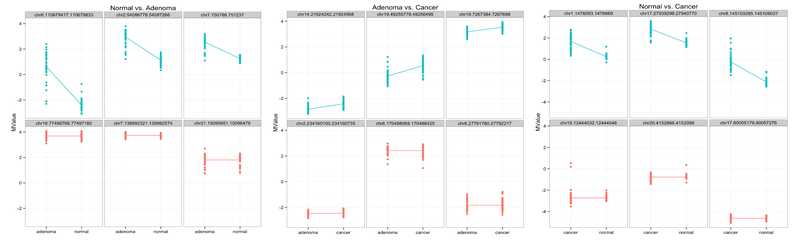


Figure 4. Top 3 differentially methylated and 3 non-differentially methylated CGIs for each pairwise group comparison.

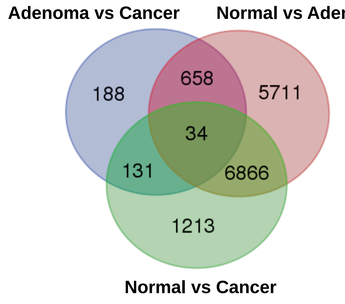


Figure 5. Venn diagram of differentially methylated CGIs at FDR < 1e-4.

$$Y = X\alpha_1 + \alpha_0 + \varepsilon$$

where:

- **Y** = methylation levels (M value)
- **X** = design matrix indicating different sample groups
- α_1 = change in mean M value
- α_0 = mean M value for reference sample group
- ε = error ~ $N(0, \sigma^2)$

Equation 1. Linear model of methylation level of each CGI using sample groups as single covariate.

Conclusion

We identified 34 DM CGIs associated with 21 genes in our pairwise comparisons between sample groups. Enriched functions are implicated in progression of various cancer types including CRC. Given that these DM CGIs are also identified in adenomas compared to normal mucosa, these aberrant methylation patterns could be markers for early detection, risk assessment, and disease monitoring of CRC.