Analyzing DNA Methylation Patterns in Cancer

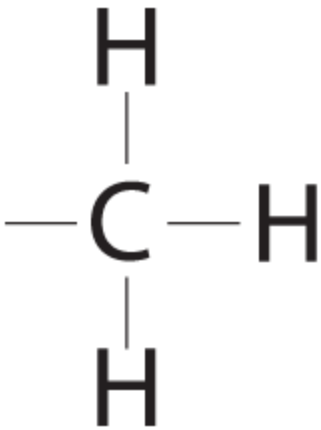
Arcadia Kratkiewicz
CSC 334

Project Overview

- Motivation
  - Methylation background
  - Intro to epigenetics
  - Methylation in cancer
- Methods
  - Obtaining data
  - F_{ST}
- Expected Results
- Future Work
  - Tree analysis
  - PCA

Motivation: Vocabulary

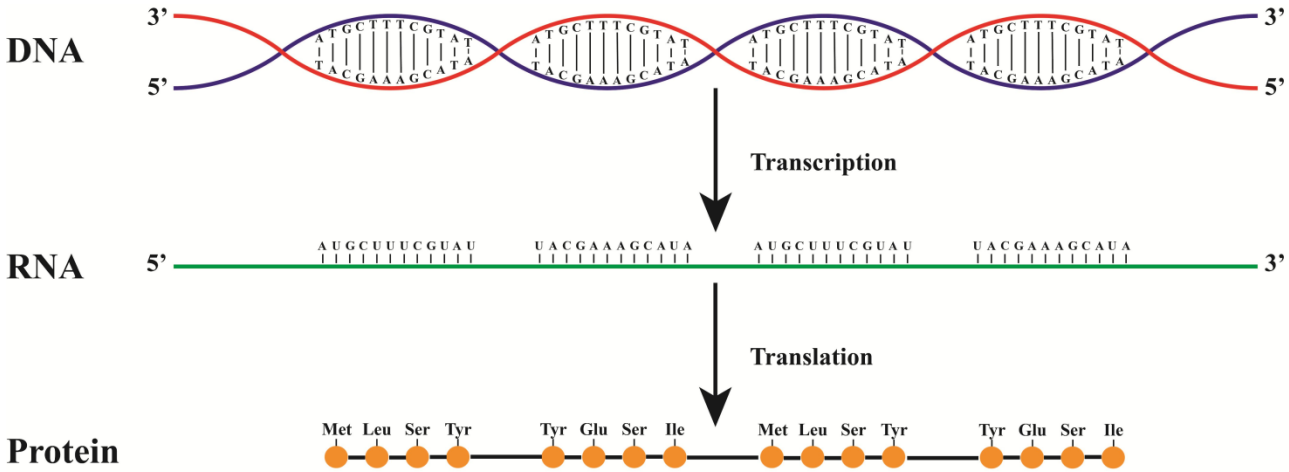
- Methylation: covalent addition of methyl (CH_3) group



- Epigenetics: genetic control by factors other than DNA sequence (change in phenotype without change in genotype)

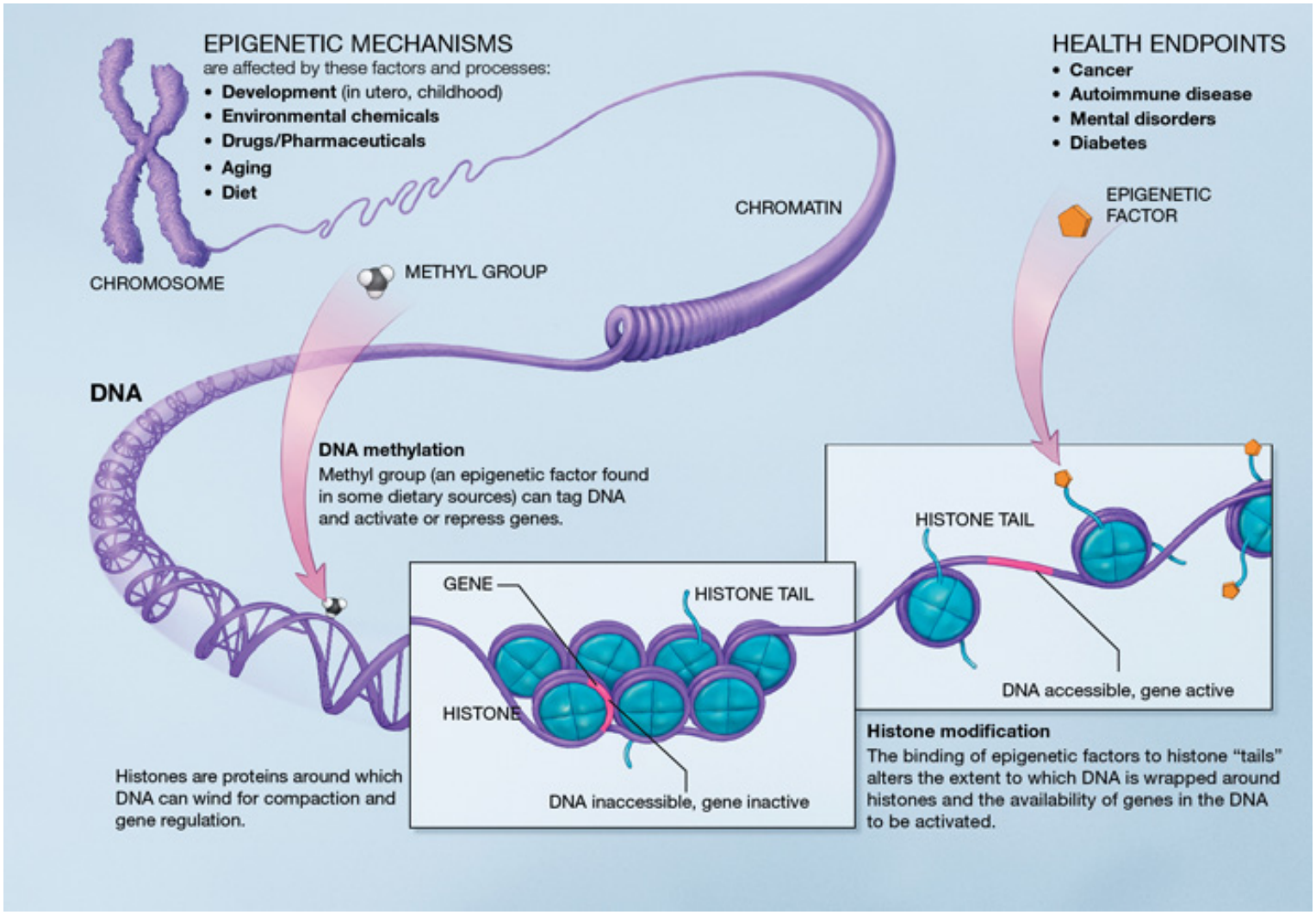
Motivation: Epigenetics

- Recall gene expression:



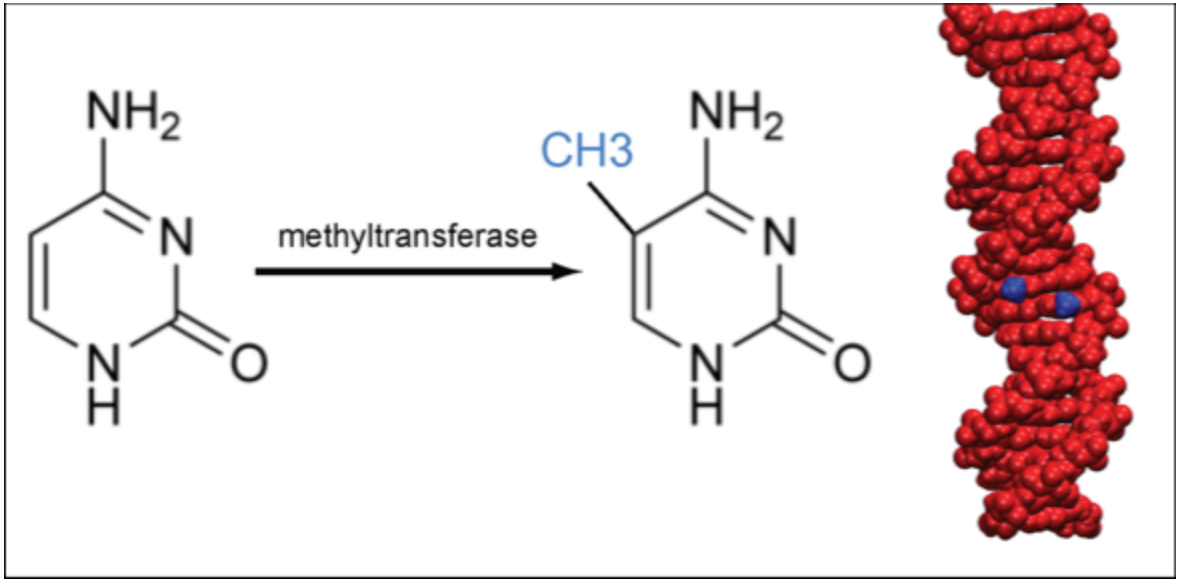
- Heritable changes in gene expression (active vs. inactive genes) without changing DNA sequence
- Factors: age, environment, disease
- Studies: identical twins

Motivation: Epigenetics

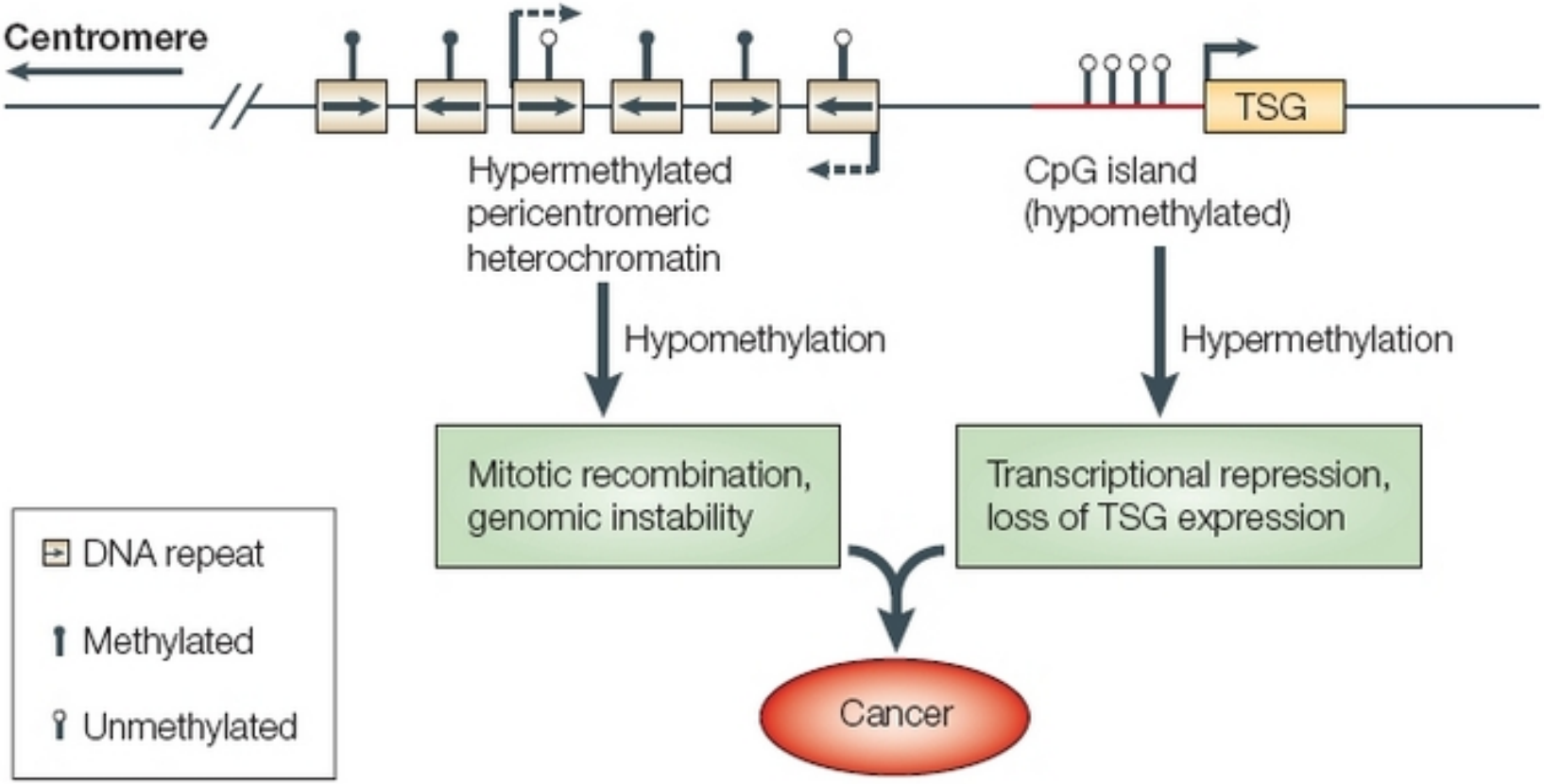


Motivation: DNA Methylation

- Epigenetic mechanism to control gene expression
- Turn genes “off”
- DNA methyltransferase adds methyl group to cytosine adjacent to guanine (CpG)



Motivation: DNA Methylation in Cancer



Methods: Plan

- Compare methylation patterns between different populations of cancerous and non-cancerous cells:

Breast cancer	Monocytes
Colorectal cancer	Liver
Ovarian cancer	Kidney
Brain cancer	Heart

Methods: Obtaining Data

DiseaseMeth

The human disease methylation database

Name/ID	Disease	Data analysis	Publication link	Experimental platform	Sample size	Download link
E-TABM-965	Neuroblastoma, Brain cancer	Only probes of Promoters (TSS[-1.5k,0.5k]) in annotation file A-MEXP-1673.adf.txt were kept and the values were mapped to [0,1] interval.	Genome-wide DNA Methylation Analysis of Neuroblastic Tumors Reveals Clinically Relevant Epigenetic Events and Large-scale Epigenomic Alterations Localized to Telomeric Regions	Nimblegen 2006-11-02_HG18_CpG_Promotor	42	
E-MTAB-144	Leukemia	Only probes of Promoters (TSS[-1.5k,0.5k]) in annotation file A-MEXP-1673.adf.txt were kept and the values were mapped to [0,1] interval.	NA	Nimblegen 2006-11-02_HG18_CpG_Promotor		
E-TABM-964	Neuroblastoma	Only probes of Promoters (TSS[-1.5k,0.5k]) in annotation file A-MEXP-1673.adf.txt were kept and the values were mapped to [0,1] interval.	Functionally Significant Epigenomic Alterations in All-trans Retinoic Acid Induced Neuroblastoma Cell Differentiation	Nimblegen 2006-11-02_HG18_CpG_Promotor		
E-GEOD-15745	NA	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain	Illumina Infinium HumanMethylation27 BeadChip	506	
E-GEOD-17448	NA	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells	Illumina Infinium HumanMethylation27 BeadChip	16	
E-GEOD-17648	Colorectal cancer	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	NA	Illumina Infinium HumanMethylation27 BeadChip	44	
E-GEOD-17769	Breast cancer	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer	Illumina Infinium HumanMethylation27 BeadChip	10	
E-GEOD-18458	Cornelia de Lange syndrome	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	Genome-wide DNA methylation analysis in cohesin mutant human cell lines	Illumina Infinium HumanMethylation27 BeadChip	72	
E-GEOD-20236	NA	Only probes of Promoters (TSS[-1.5k,0.5k]) were kept and the values were mapped to [0,1] interval.	Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains	Illumina Infinium HumanMethylation27 BeadChip	93	
		Only probes of Promoters (TSS[-1.5k,0.5k]) were kept	Human aging-associated DNA hypermethylation occurs	Illumina Infinium		

Methods: Processing Data

Reporter Identifier	VALUE
cg00000292	0.792011029
cg00002426	0.840962213
cg00003994	0.055869795
cg00005847	0.27148087
cg00006414	0.046893735
cg00007981	0.039682613
cg00008493	0.957668419
cg00008713	0.031316033
cg00009407	0.02212467
cg00010193	0.612340763
cg00011459	0.909001991
cg00012199	0.025283353
cg00012386	0.034264349
cg00012792	0.042111647
cg00013618	0.830915572
cg00014085	0.021579047
cg00014837	0.861734668

→

[[cg00000292, 1],
[cg00002426, 1],
[cg00003994, 0],
[cg00005847, 0],
[cg00006414, 0],
[cg00007981, 0],
[cg00008493, 1],
[cg00008713, 0],
[cg00009407, 0],
[cg00010193, 1],
[cg00011459, 1],
[cg00012199, 0],
[cg00012386, 0],
[cg00012792, 0],
[cg00013618, 1],
[cg00014085, 0],
[cg00014837, 1]]

→

Methods: Processing Data

Sample:	Breast cancer			Brain cancer			Liver		
Location:	A	B	C	A	B	C	A	B	C
cg00000292	1	1	1	0	0	0	0	0	0
cg00002426	1	1	0	0	0	1	1	1	1
cg00003994	1	0	1	1	1	1	1	1	1
cg00005847	0	1	0	1	0	1	0	0	0
cg00006414	0	0	1	1	1	0	0	0	0
cg00007981	1	1	1	1	1	1	1	1	1
cg00008493	0	0	0	0	0	0	0	0	0
cg00008713	1	0	1	1	1	1	1	1	1
cg00009407	1	1	1	1	1	1	1	1	1
cg00010193	1	1	1	1	1	1	0	0	0
cg00011459	0	1	0	0	0	0	0	0	0
cg00012199	0	0	0	0	0	0	0	0	0
cg00012386	1	0	1	1	1	1	0	0	0
cg00012792	1	1	1	1	1	1	1	1	1
cg00013618	1	1	0	1	1	1	0	0	0
cg00014085	1	1	1	0	0	0	1	1	1
cg00014837	1	1	1	0	0	1	0	0	0

Methods: F_{ST}

- Fixation index (F_{ST}) is a measure of population differentiation.
- Estimation of F_{ST}:

$$F\downarrow ST = \pi\downarrow between - \pi\downarrow within \,/ \pi\downarrow between$$

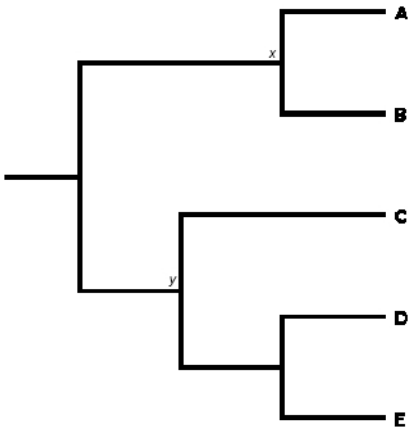
- $\pi\downarrow within$: average number of pairwise differences between individuals within the same population
- $\pi\downarrow between$: average number of pairwise differences between individuals in different populations

Expected Results

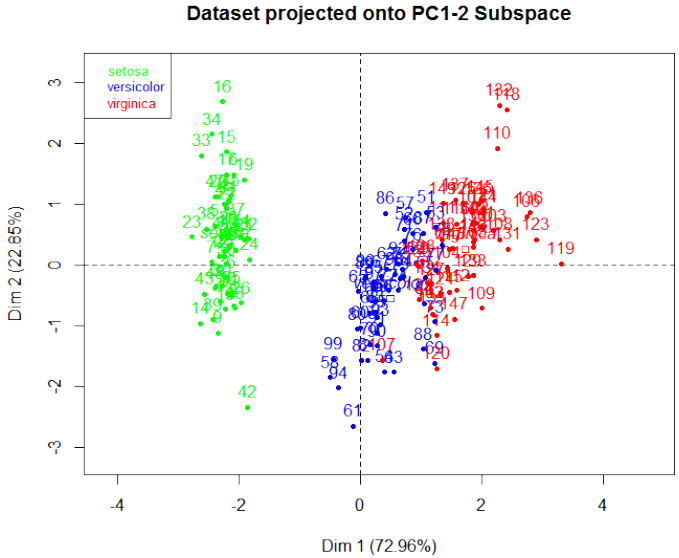
- Similar methylation patterns between cancers
- Cancer methylation patterns different from non-cancer
- Expect a higher F_{ST} for the populations than if they were randomly selected

Future Work

- Tree Analysis



- PCA



Questions?