

CpG Islands: A Potential Role in Chromatin Modification

Blanca Himes
Richard Koche
Shankar Mukherji

Background - Histones

- DNA in eukaryotic genomes is organized into chromatin, which contains histones
- Histones:
 - Subject to acetylation, methylation, and other posttranslational modification
 - Modifications influence relative condensation of chromatin and thus transcriptional activation

Background - CpG Islands

- Regions that have high frequency of CpG dinucleotides compared to genome as a whole
 - Common near transcription start sites
 - Of regulatory importance in promoters
- CpGs are rare unless there is selective pressure to keep them or they are in a non-methylated area

Importance of Histone Modification

Transcriptionally Active

Transcriptionally Repressed

Figure removed due to copyright reasons.
See Figure 5 in "The Fundamental Role of Epigenetic Events in Cancer."
Nature Reviews Genetics 3, 415-428 (2002).

Motivation

- Bernstein et al:
 - Used ChIP, DNA amplification, oligo arrays
 - Maps of H3 Lys4 di- and trimethylation and H3 Lys9/14 acetylation for chr 21 and 22 generated with human hepatoma cell line
- They didn't mention CpG islands!

Hypotheses

- Because CpG islands and histone modification are both involved in epigenetic regulation, we predict that there are positional correlations between them
- CpG islands may provide a mechanism through which histone-modifying proteins find targets

First: Look for CpG Islands

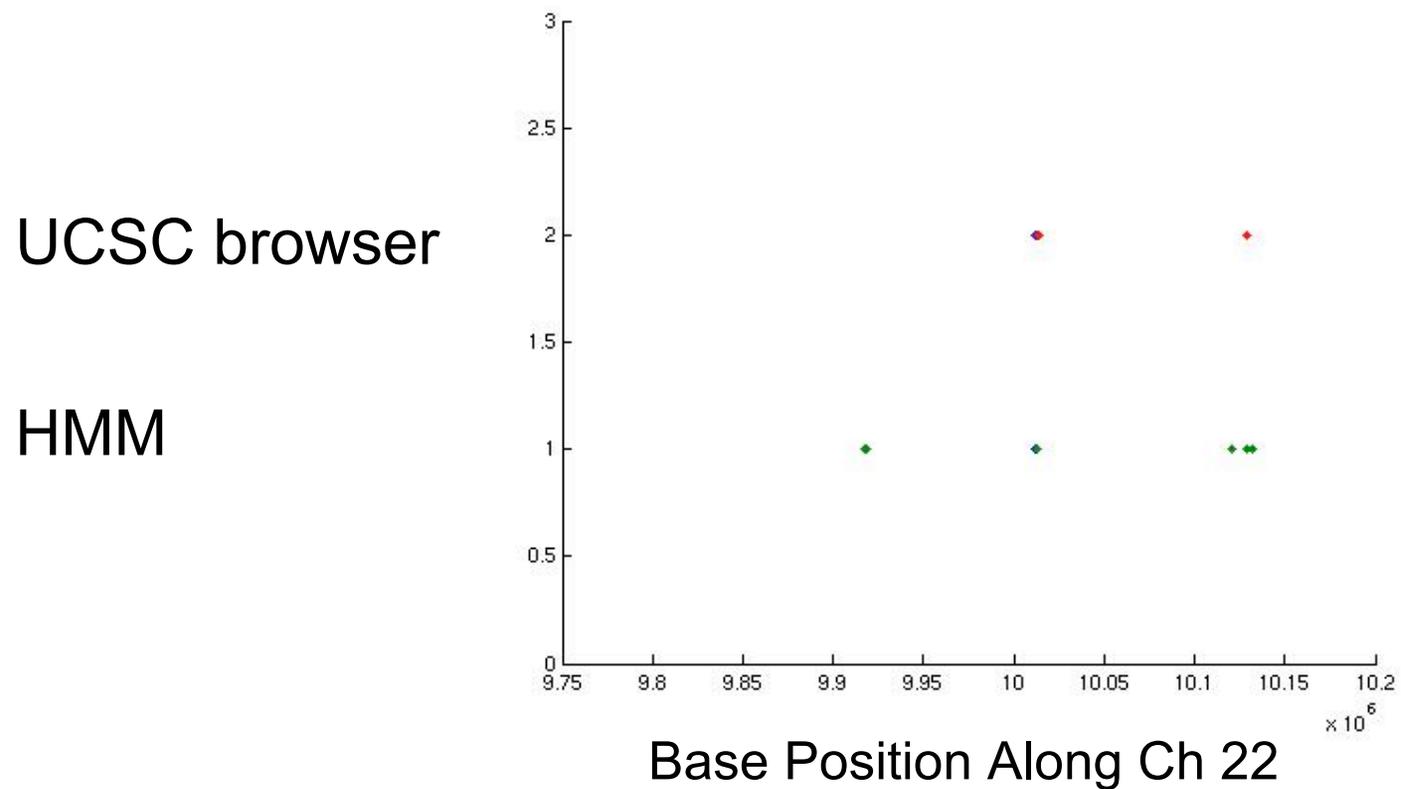
- CpG island locations can be readily obtained from the UCSC genome browser
- The method they use to identify CpG islands is:
 - Search a sequence one base at a time and score each dinucleotide (+17 for CG and -1 otherwise)
 - Identify maximally scoring segments and evaluate for:
 - GC content $\geq 50\%$
 - Length > 200 bases
 - Ratio > 0.6 of CG bases observed to those expected based on CG content of sequence

Using HMMs to Find CpGs

- As an alternative to downloading the CpG locations, we used a Hidden Markov Model to identify CpG islands.
- States are $\{A+, C+, T+, G+, A-, C-, T-, G-\}$, where the + refers to *in* a CpG island and - refers to *outside* a CpG island
- The transition probabilities within or outside an island used were (Durbin et al, Markov Chains and Hidden Markov Models pg 50):

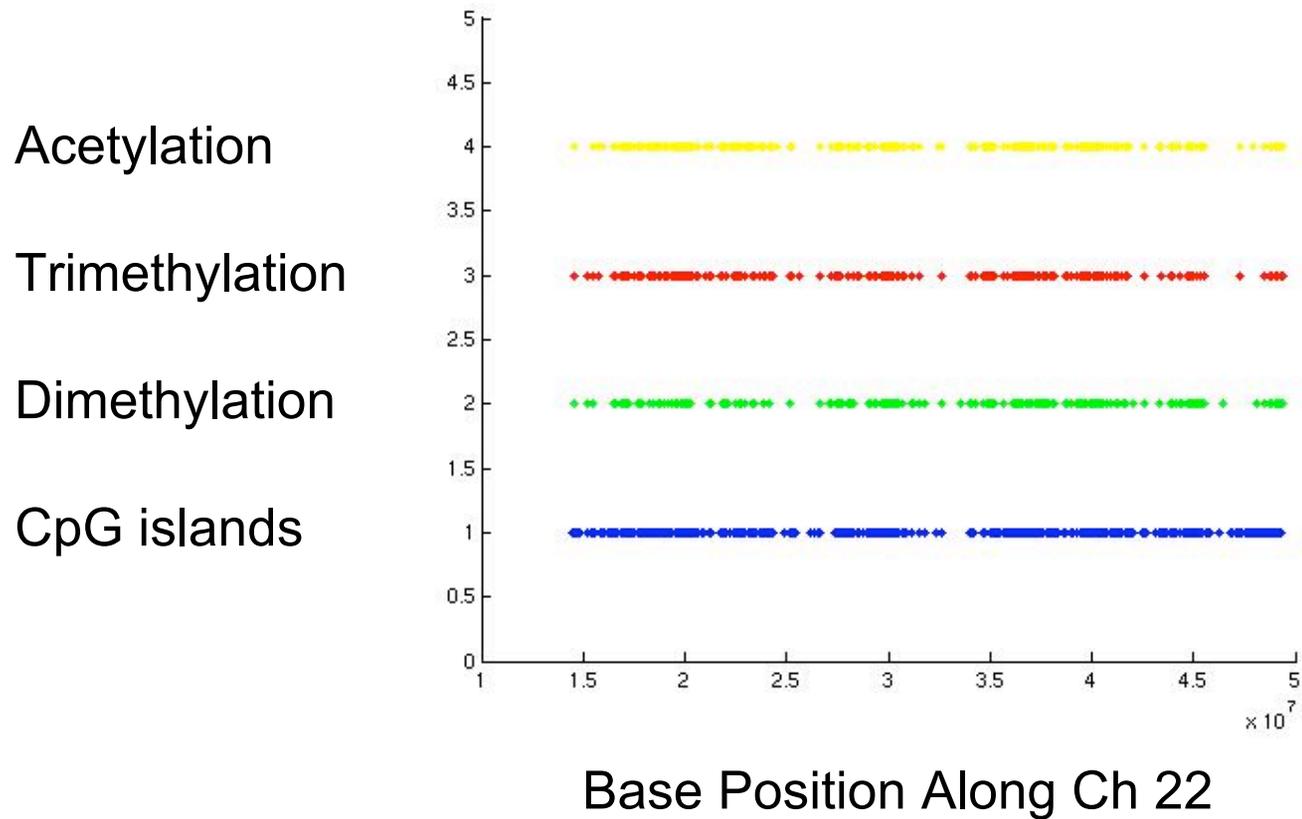
	Inside	Outside
A	[0.180, 0.274, 0.426, 0.120]	[0.300, 0.205, 0.285, 0.210]
C	[0.170, 0.368, 0.274, 0.188]	[0.322, 0.298, 0.078, 0.302]
G	[0.161, 0.339, 0.375, 0.125]	[0.248, 0.246, 0.298, 0.208]
T	[0.079, 0.355, 0.384, 0.182]	[0.177, 0.239, 0.292, 0.292]

Comparison of CpG Islands Found

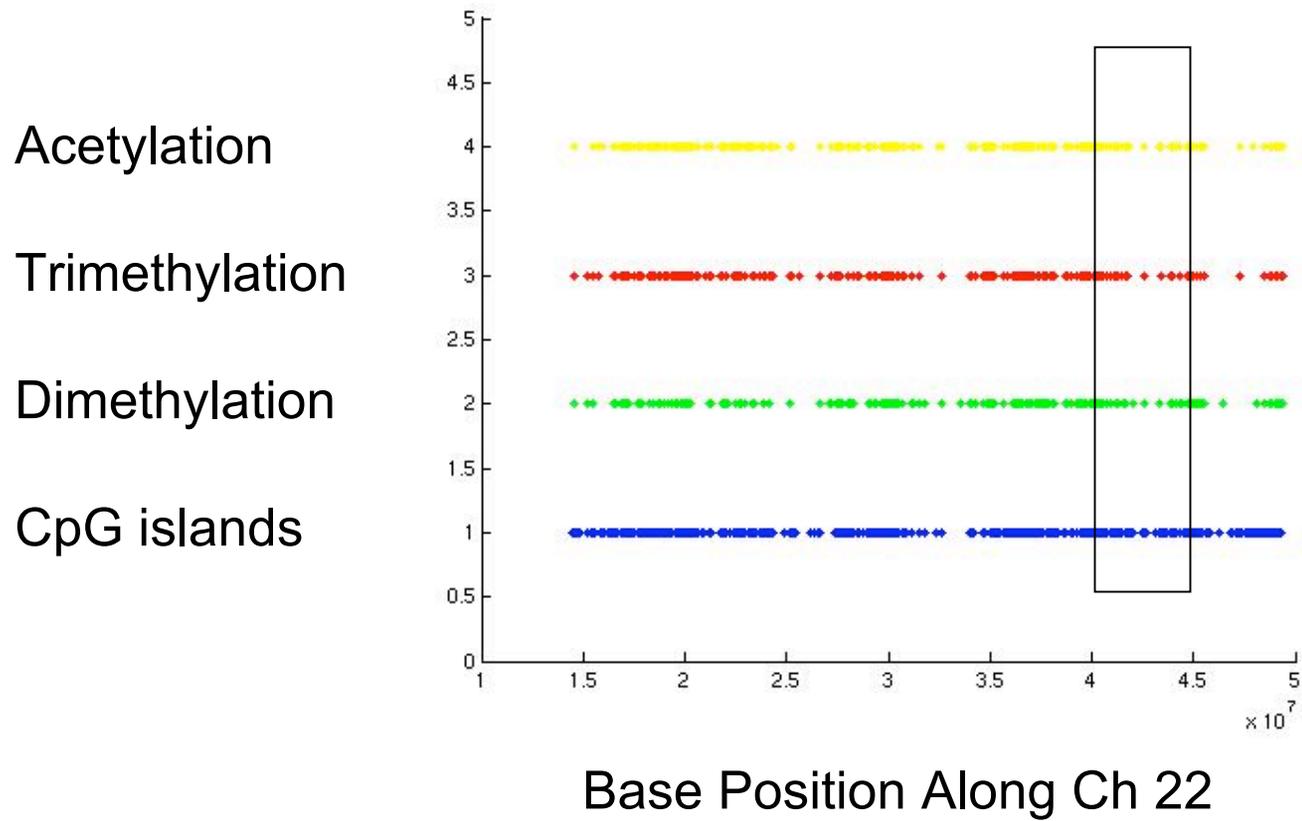


Excluded islands with size < 200 bases

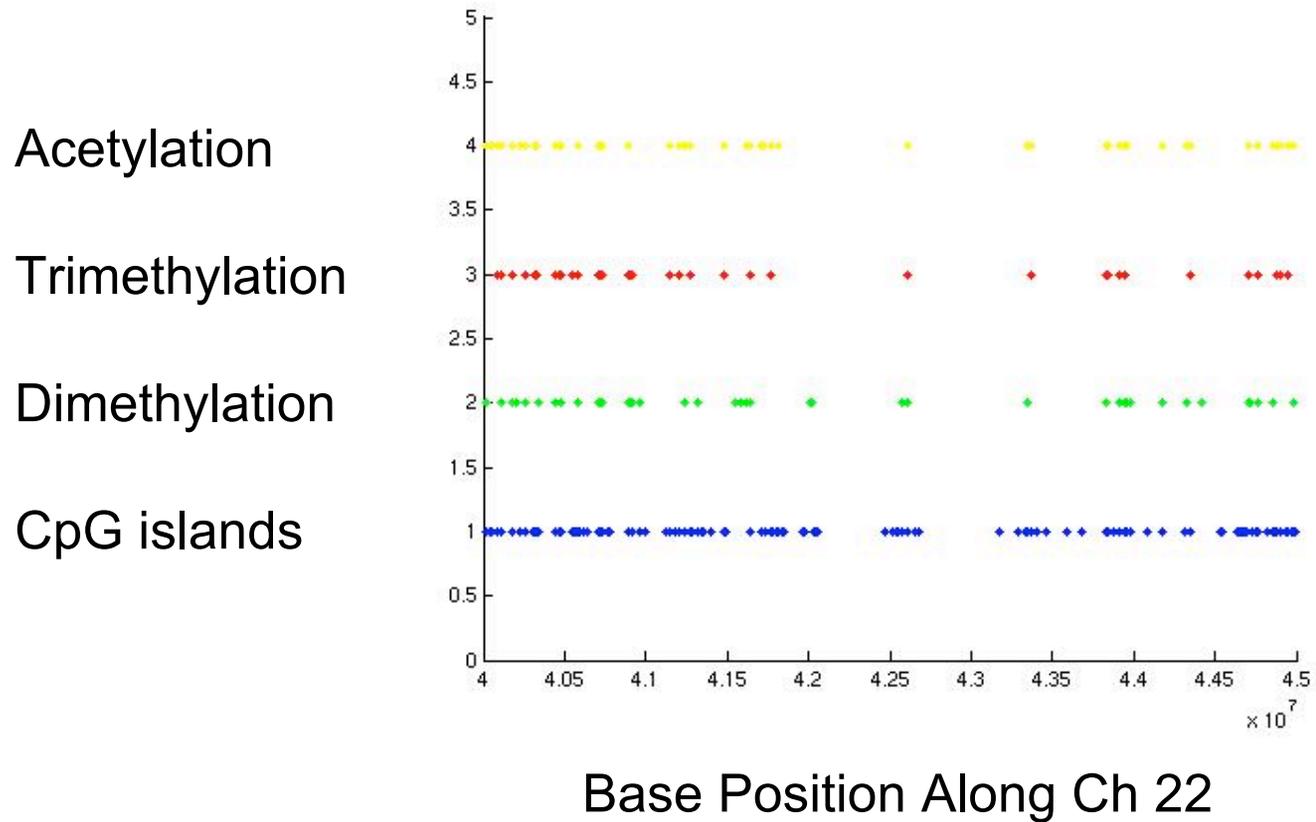
Histone Modification Data from Bernstein et al and CpG Islands



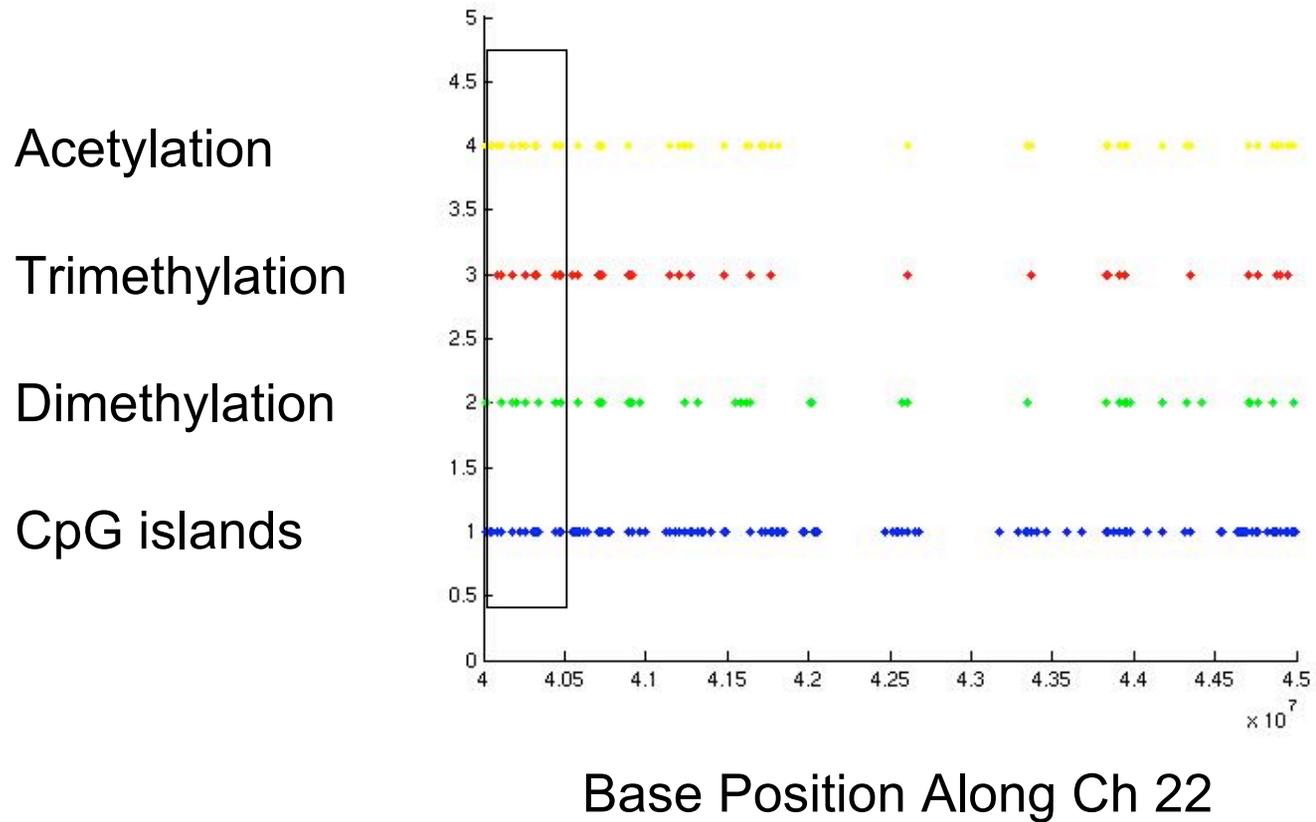
Histone Modification Data from Bernstein et al and CpG Islands



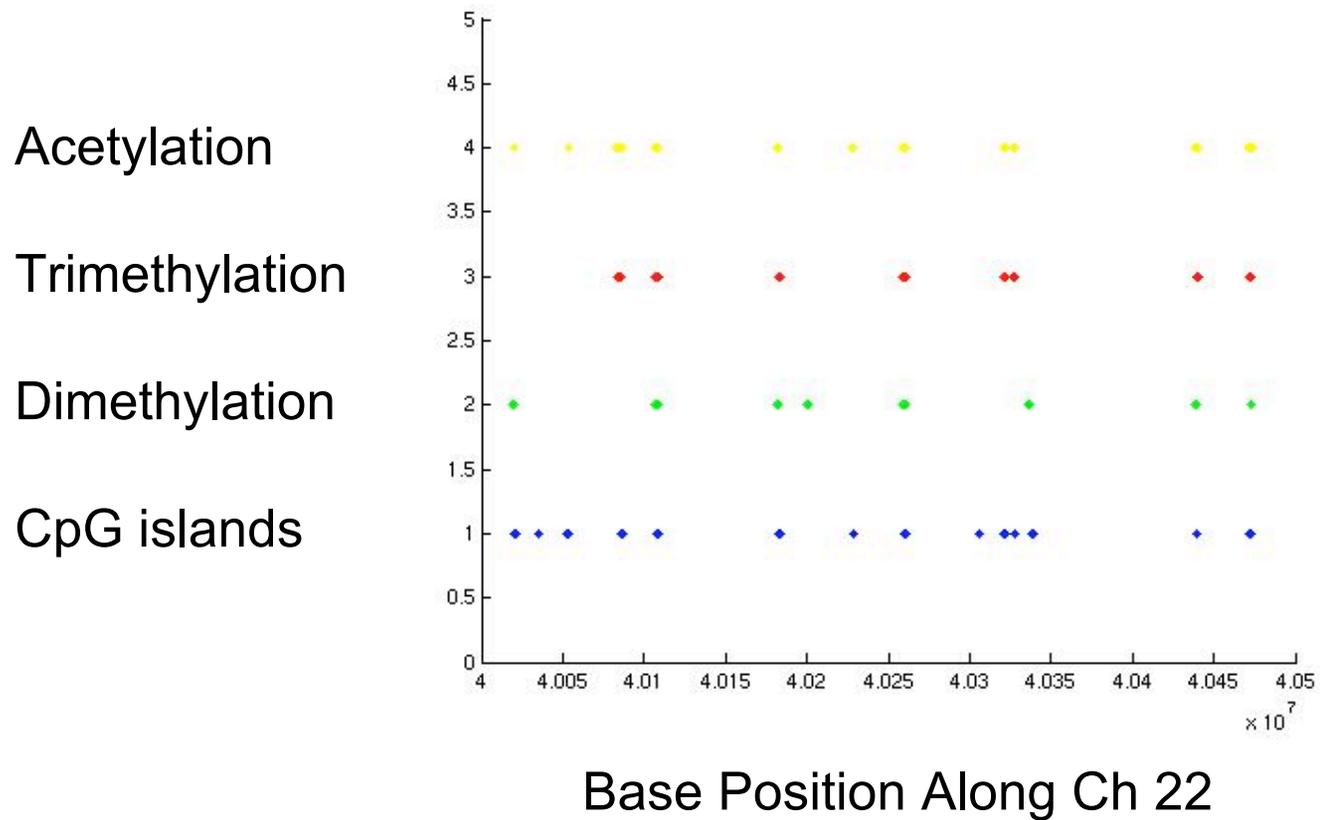
Histone Modification Data from Bernstein et al and CpG Islands



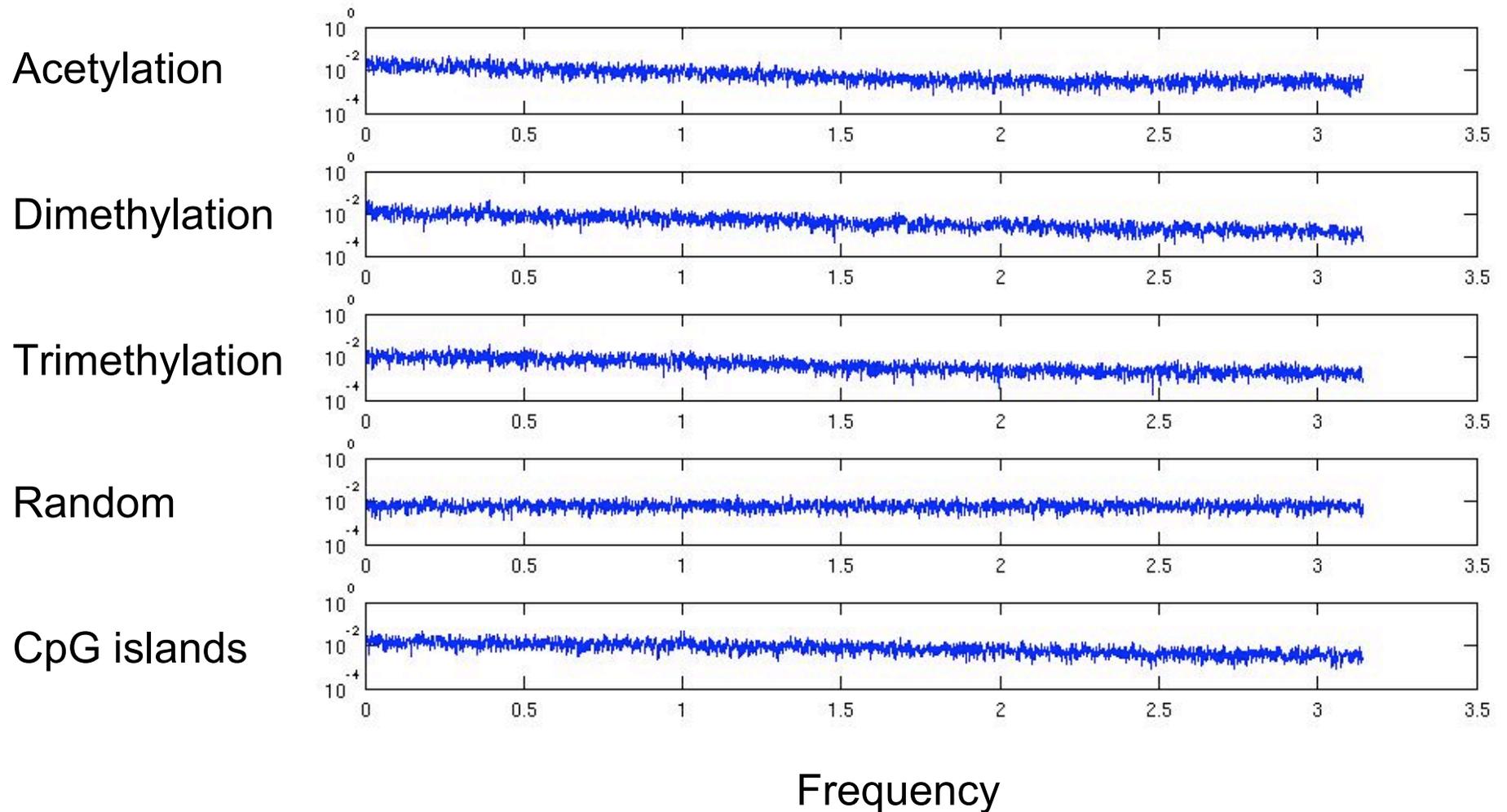
Histone Modification Data from Bernstein et al and CpG Islands



Histone Modification Data from Bernstein et al and CpG Islands



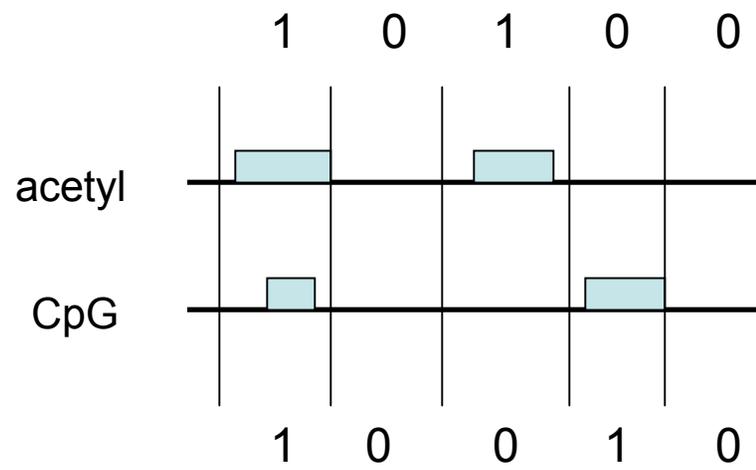
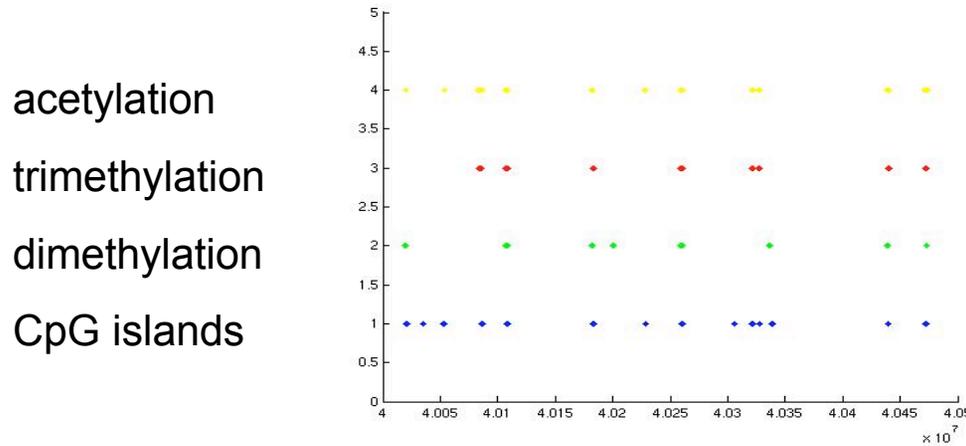
Fourier Spectra of Spatial Modification Patterns for Ch 22



Correlations of Spatial Modification Patterns for Ch 22

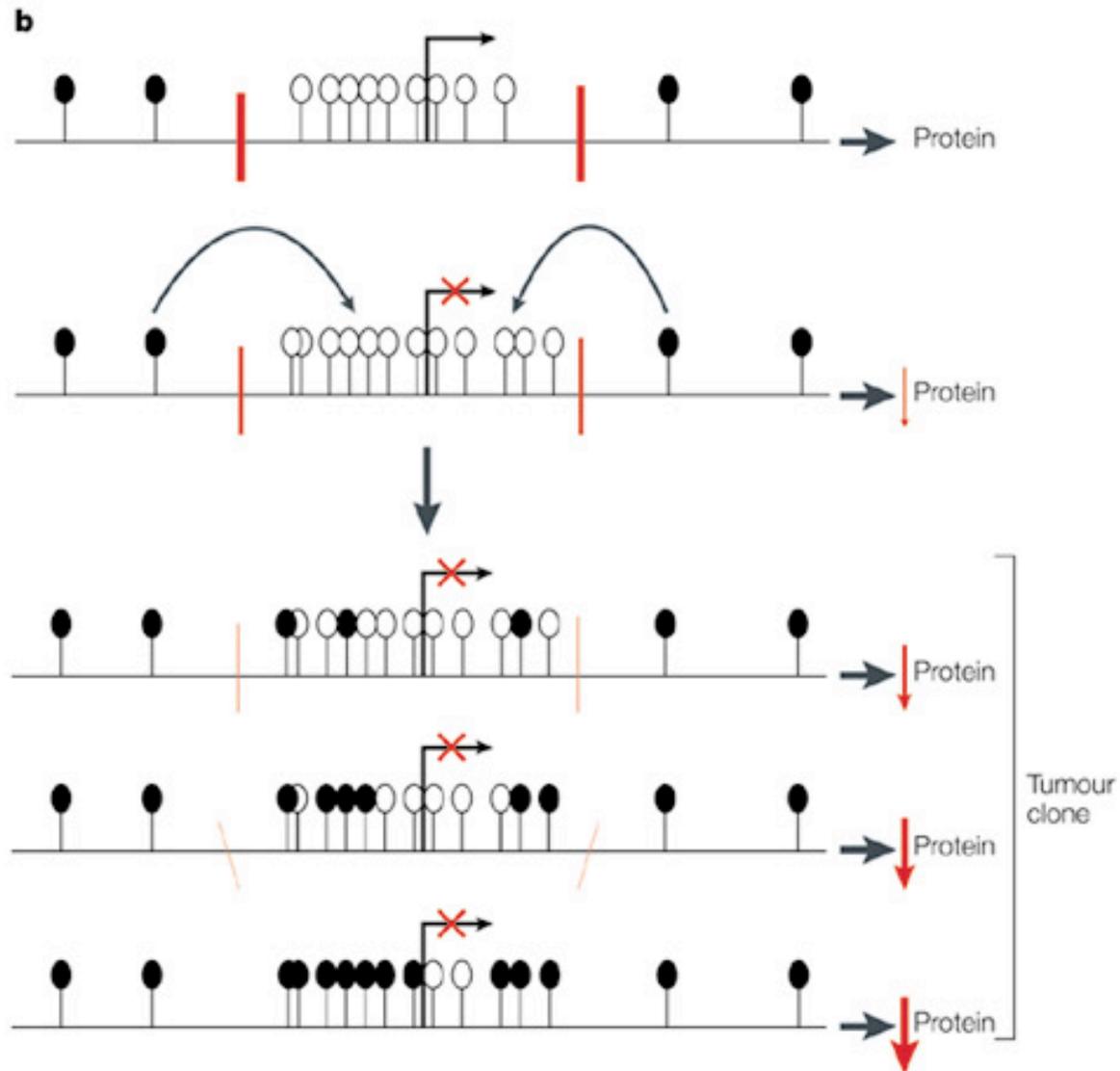
Modification Paired with CpG	Correlation Coefficient	Random Correlation
Acetylation	0.6114	0.0071
Dimethylation	0.5770	0.0035
Trimethylation	0.6221	0.0104

Pairwise CpG Positional Correlations

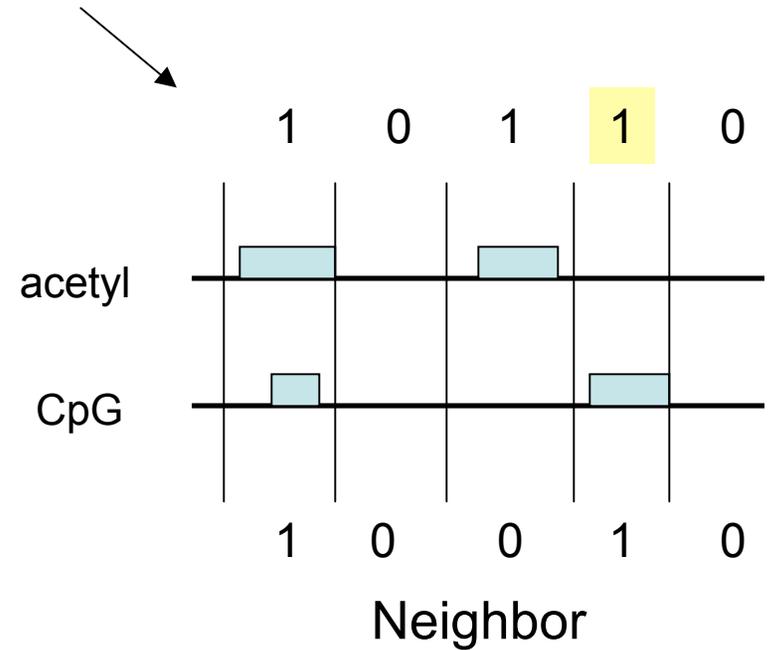
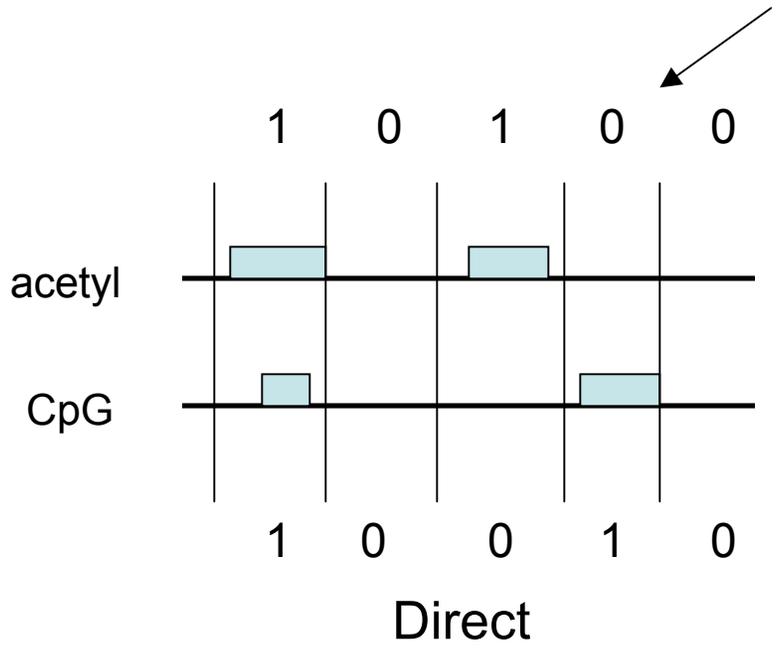
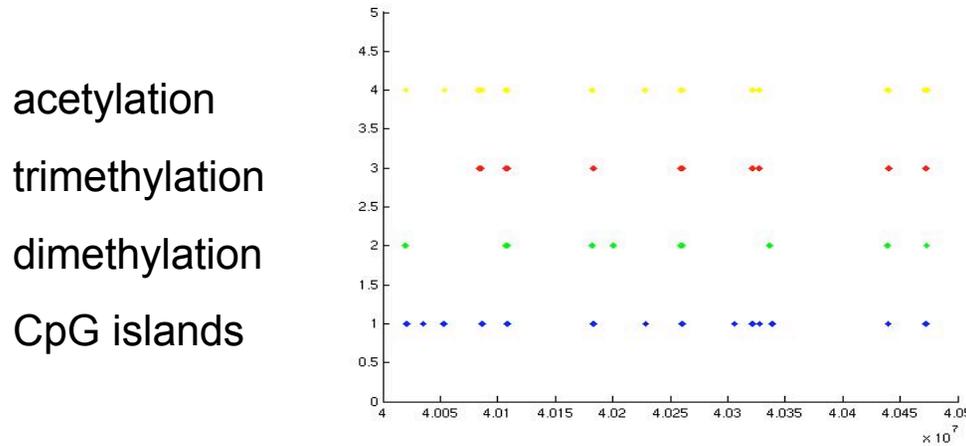


Direct

Methylation Cascade & Gene Silencing



Pairwise CpG Positional Correlations



Pairwise CpG Positional Correlations

	Acetylation		Dimethylation		Trimethylation	
	Direct	Neighbor	Direct	Neighbor	Direct	Neighbor
CpG	0.300		0.211		0.340	
Random	-0.008		0.002		-0.004	

Pairwise CpG Positional Correlations

	Acetylation		Dimethylation		Trimethylation	
	Direct	Neighbor	Direct	Neighbor	Direct	Neighbor
CpG	0.300	0.710	0.211	0.700	0.340	0.760
Random	-0.008	-0.004	0.002	-0.014	-0.004	0.006

Genes in Regions of Significant Overlap Between CpG Islands and Histone Modification States

	Acetylation-CpG		Dimethylation-CpG		Trimethylation-CpG	
	Direct	Neighbor	Direct	Neighbor	Direct	Neighbor
NF2	X	X	O	X	O	X
TIMP-3	X	X	X	X	X	X
SMARCB1	O	X	X	X	X	X
Gene Totals:	135	259	87	256	129	259

Conclusions

- Frequencies of histone modification patterns are highly correlated to frequency of CpG islands
- Significant positional overlap of histone modification patterns and CpG islands were found
- Genes known to be epigenetically regulated were identified in regions where CpG islands correlated with histone modifications