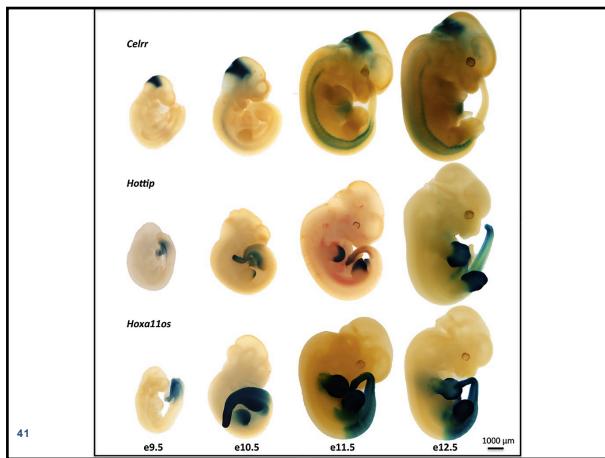


EMBL-EBI

e!Ensembl

UNDERSTANDING GENE REGULATION – PART II

Myrto Kostadima Ensembl Regulation Project Leader, EMBL-EBI

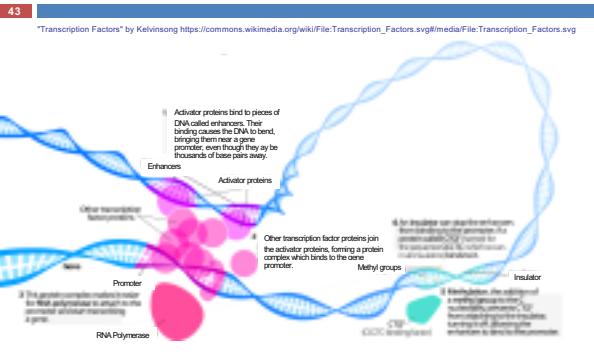


42

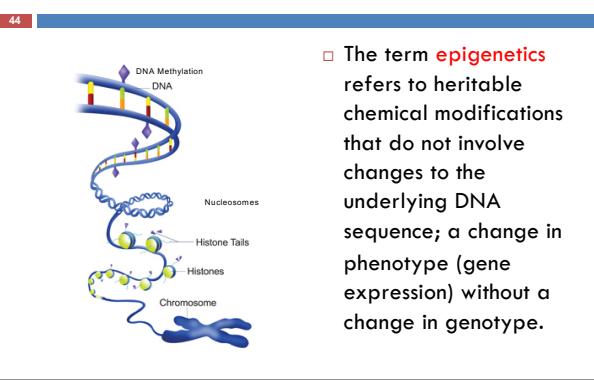
Elements of gene regulation

- Transcription factors
 - Promoters, enhancers, insulators..
- RNA Pol II machinery
- Epigenetics
 - DNA methylation
 - Nucleosome positioning
 - Histone modification
- 3D Chromatin organisation
- Long non-coding RNAs (lncRNAs)

TFs and regulatory regions

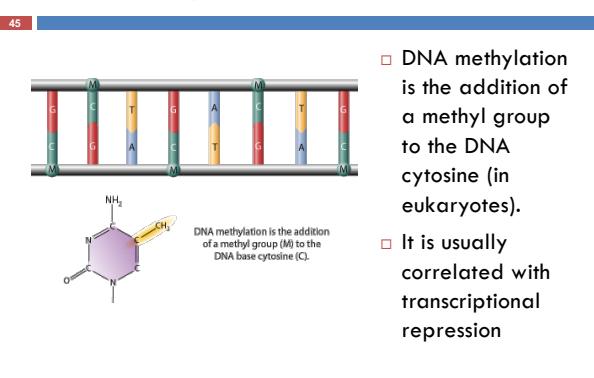


Epigenetics



- The term **epigenetics** refers to heritable chemical modifications that do not involve changes to the underlying DNA sequence; a change in phenotype (gene expression) without a change in genotype.

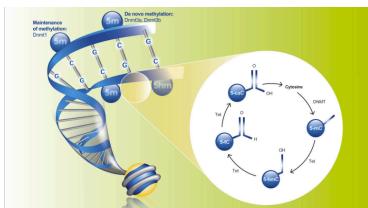
DNA methylation



- DNA methylation is the addition of a methyl group to the DNA cytosine (in eukaryotes).
- It is usually correlated with transcriptional repression

DNA methylation

46



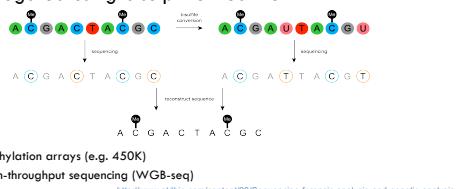
- DNA methyltransferases – DNMT1, DNMT3a/b
- Tet family of enzymes

<http://www.abcam.com/epigenetics/dna-methylation-a-guide>

DNA methylation - detection

47

- ChIP-based approach using antibodies against methylated regions (MeDIP)
- Alternatively, first use enzymatic treatment to enrich for CpG-rich regions followed by sequencing (RRBS-seq)
- Interrogated using bisulphite treatment:

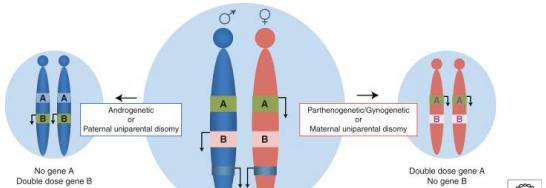


- Methylation arrays (e.g. 450K)
- High-throughput sequencing (WGB-seq)

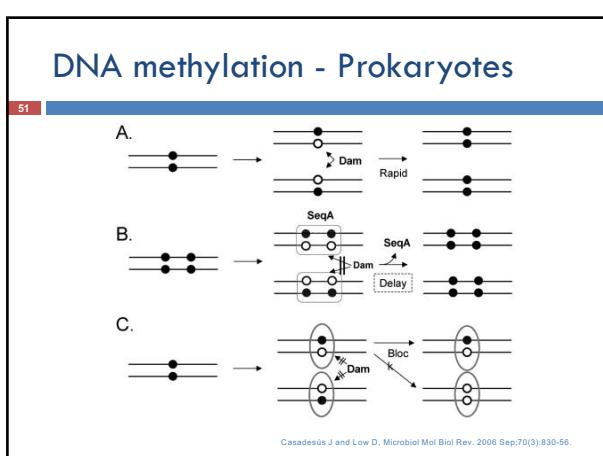
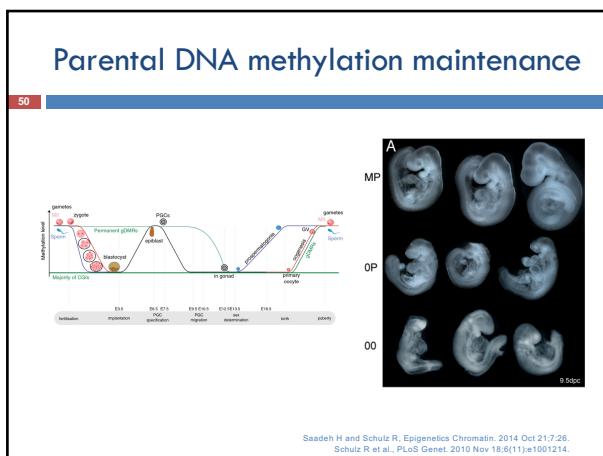
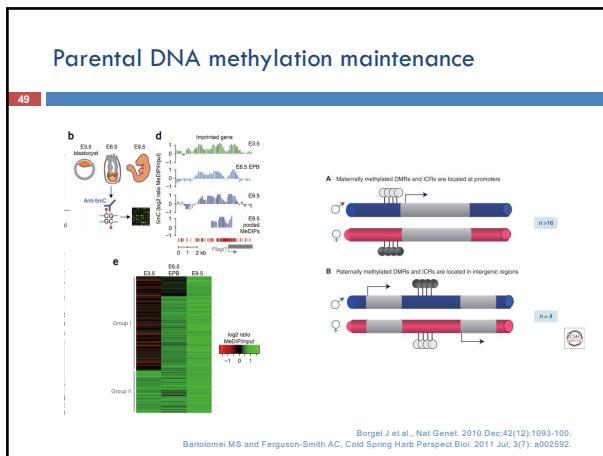
<http://www.altdbio.com/content/20/Sequencing-forensic-analysis-and-genetic-analysis>

Imprinted genes

48

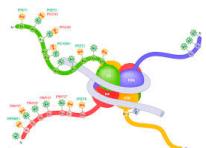


Bartolomei MS and Ferguson-Smith AC, Cold Spring Harb Perspect Biol. 2011 Jul; 3(7): a002592.



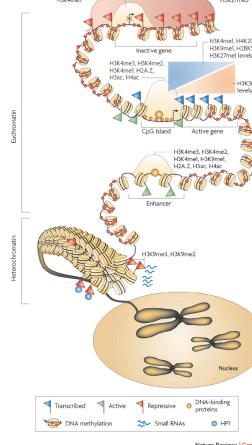
Histone modifications

- Histone proteins have protruding ‘tails’.
- Histone tails can undergo a host of modifications.
 - These are associated with regulatory function.



Schones DE and Zhao K. Nature Bey Genet. March 2008; 9: 179-191.

Histone modifications



Histone modifications

54

Insilico modification or feature	Signal characteristic	Potative functions
H3K4me1	Peak	Histone protein variant (H3K27) associated with regulatory elements with enhancers.
H3K4me1	Peak/Region	Mark of regulatory elements associated with enhancers and other distal elements, associated with transcription starts.
H3K4me2	Peak	Mark of regulatory elements associated with transcription starts.
H3K4me3	Peak	Mark of regulatory elements primarily associated with transcription starts.
TFIDB	Peak	Mark all regulatory elements with transcription starts.
H3K19me1	Region	Loosely associated with transcription, with preference for 5' end of genes.
H3K36me1	Peak/Region	Repetitive DNA sequences, centromeres, heterochromatin, repetitive elements, and transcription domains.
H3K27ac	Peak	Mark of active regulatory elements, mark of transcription start sites, and promoters from their motif occurrences.
H3K27me3	Region	Repressive mark positioned by polycomb complexes at transcription start sites and silent developmental genes.
H3K36me3	Region	Associated with genes, with preference for 5' end of genes.
H3K79me2	Region	Transcription-associated mark, with preference for 5' end of genes.
H3K9me1	Region	Loosely associated with transcription, with preference for 5' end of genes.

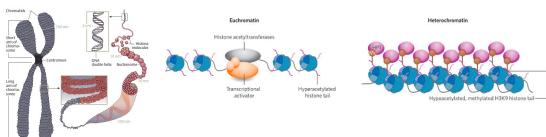
ENCODE Chaitin

Courtesy of Dr Steven Wilder

Euchromatin/Heterochromatin

55

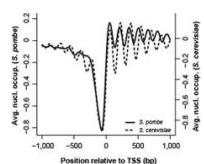
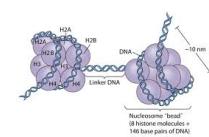
- On a more global level, there are accessible and very dense and mainly inaccessible regions.



- While some regions of the genome are stable eu/heterochromatin, others are actively compacted and decompactored

Nucleosome positioning

56



- DNA 'wrapped' around histone cores
•MNase-seq or ATAC-seq to interrogate their position

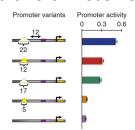
Nucleosome positioning

57

- What determines nucleosome position?

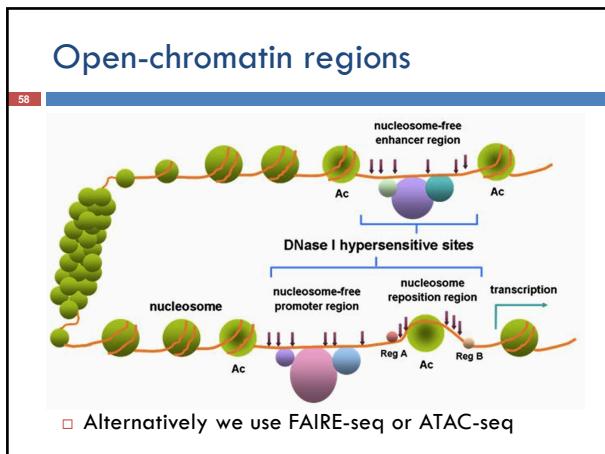
- Underlying sequence

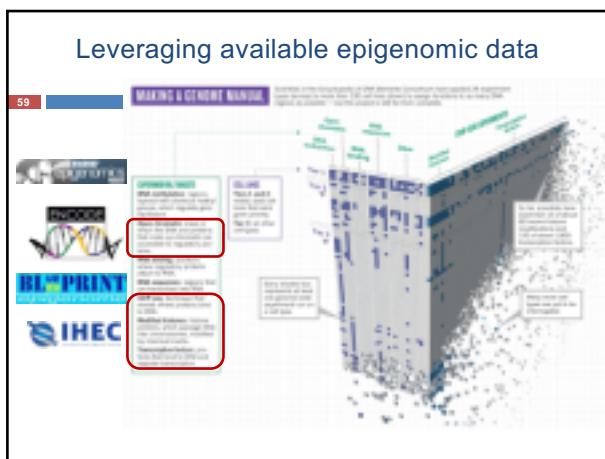
- AT-rich sequence does not 'bend' easily
- Nucleosomes avoid those regions

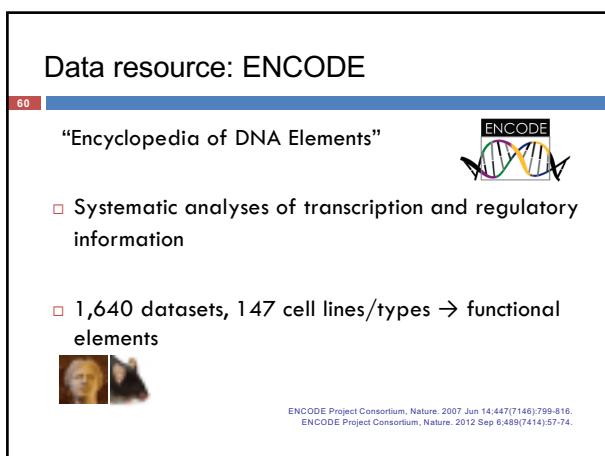


- Nucleosome remodelers (Isw2, RSC)

Raveh-Sadka T et al., Nat Genet. 2012 May 27;44(7):743-50.







Data resource: Roadmap Epigenomics

61

Public resource of normal epigenomes

- DNA methylation
- histone marks
- open chromatin
- small RNA



Roadmap Epigenomics logo

<http://www.roadmapepigenomics.org/data>
<http://www.roadmapepigenomics.org/publications/>

Data resource: Blueprint

62



Generate 100 reference epigenomes

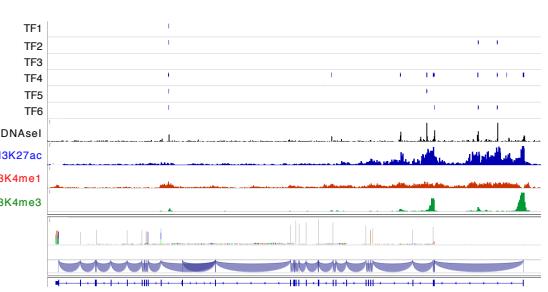
- Blood cells
- healthy individuals and
- malignant leukaemic counterparts



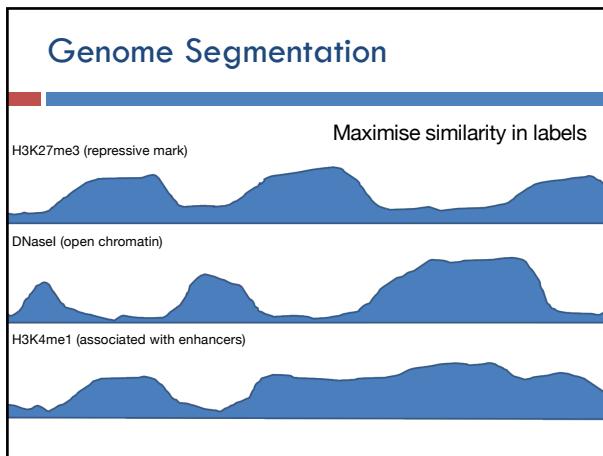
<http://www.blueprint-epigenome.eu/>
<http://dcc.blueprint-epigenome.eu/home>
<http://www.cell.com/consortium/ihec>

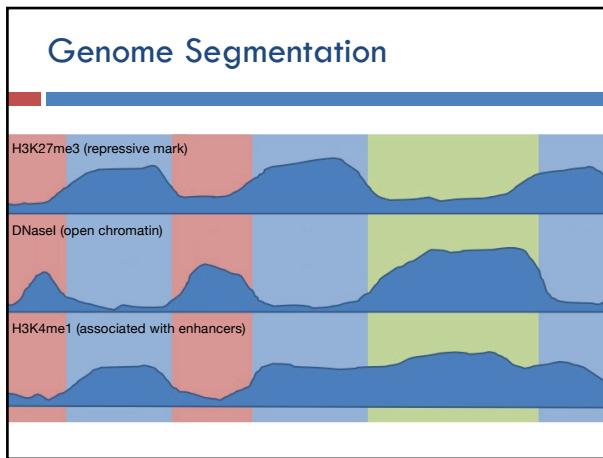
GeneX epigenetic landscape

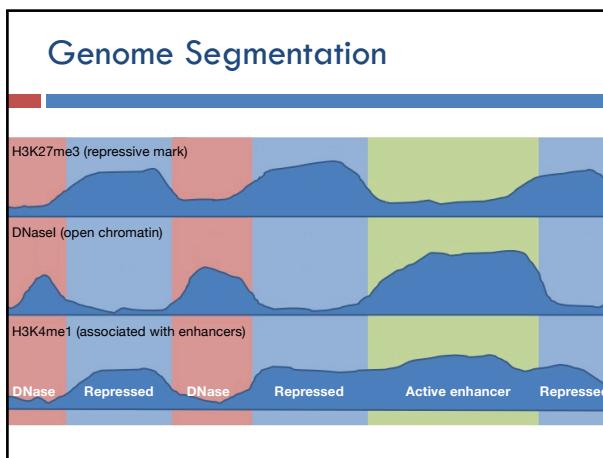
63



GeneX







Genome Segmentation – Data Input

67

- For each of the ENCODE cell lines:
 - GM12878, K562, H1-hESC, HeLa-S3, HepG2, HUVEC
 - Open Chromatin
 - DNase1 hypersensitivity
 - Transcription factor
 - CTCF (Insulator)
 - Histone modifications
 - H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac (Transcriptional Activation)
 - H3K36me3 (Transcriptional Elongation)
 - H3K27me3, H4K20me1 (Transcriptional Repression)
 - ChIP input (control)

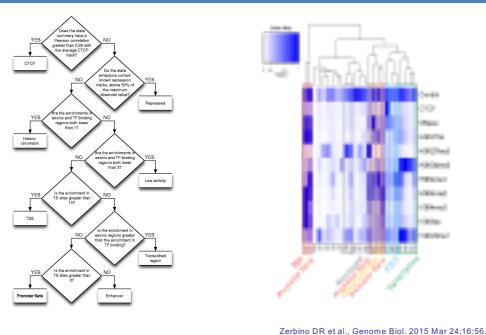
Genome segmentation algorithms

68

	ChromHMM	Segway
Modeling framework	Hidden Markov model	Dynamic Bayesian network
Number of states	25	25
Genomic resolution	200 bp	1 bp
Data resolution	Boolean	Real value
Handling missing data	Interpolation	Marginalization
Emission modeling	Bernoulli distribution	Gaussian distribution
Length modeling	Geometric distribution	Geometric plus hard and soft constraints
Training set	Entire genome	ENCODE regions (1%)
Decoding algorithm	Posterior decoding	Viterbi

Assigning function to genomic segments

69



Genome segmentation examples

70

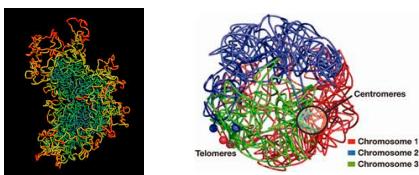
Cell types



3D organisation in the nucleus

71

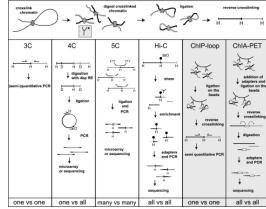
- The position of chromosomes in the nucleus is not random and it matters functionally



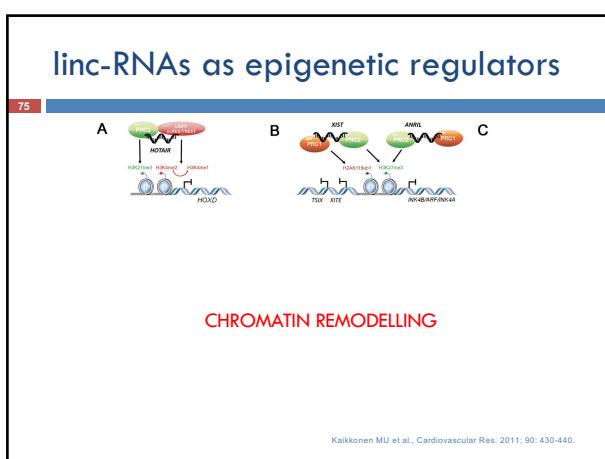
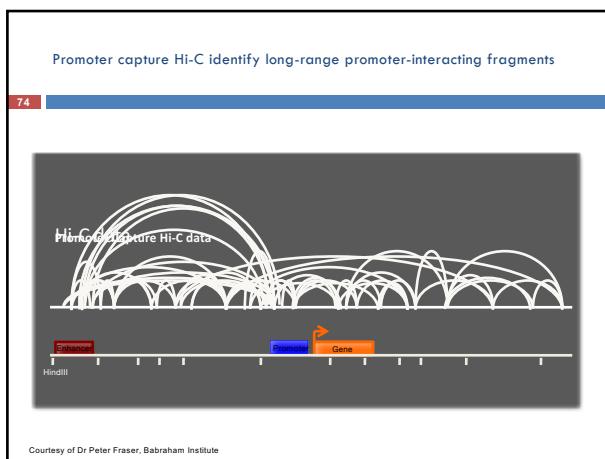
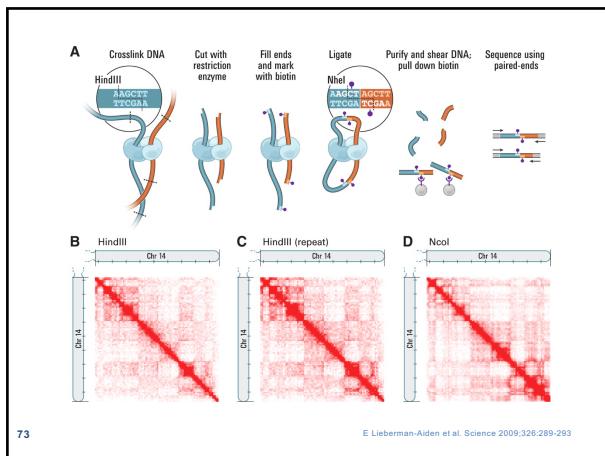
3D organisation in the nucleus

72

- Interrogate the 3D organisation by Chromosome Conformation Capture (3C, 4C, HiC)

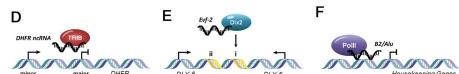


- Capture long-range interactions within chromosome
- Describe physical proximity between regions of different chromosomes



linc-RNAs as epigenetic regulators

76



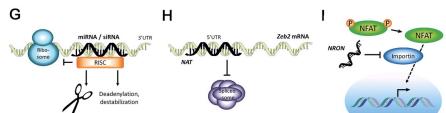
TRANSCRIPTIONAL REGULATION

Kaikkonen MU et al., Cardiovascular Res. 2011; 90: 430-440.

linc-RNAs as epigenetic regulators

77

POST-TRANSCRIPTIONAL CONTROL



Kaikkonen MU et al., Cardiovascular Res. 2011; 90: 430-440.

"Common variants associated with diverse human diseases and phenotypic traits are concentrated in regulatory DNA [...], implicating regulatory variation as an important mediator of quantitative human phenotypes." *

78

*Maurano MT et al., Nature Genet. 2015 Oct 26.

Which disease/trait is it associated with and what is the effect?

Are there any HTS data sets available in a related cell type/line?

Is there RNA-seq evidence for non-coding transcription?

What is the epigenetic signature of the locus?

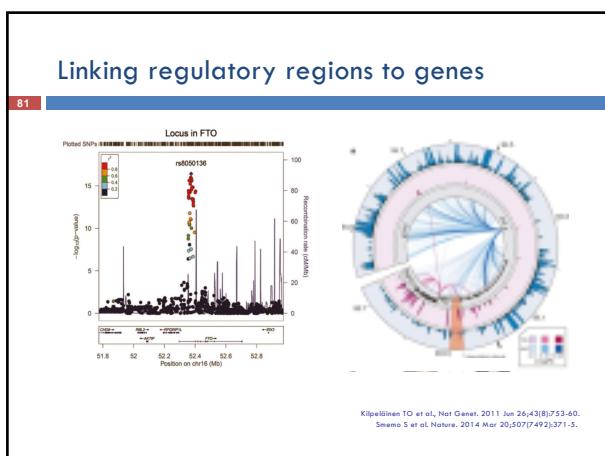
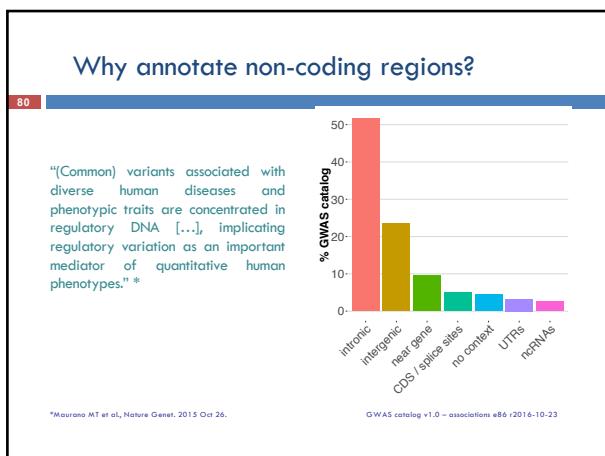
If the SNP is associated with a gene, which is the most abundant transcript?

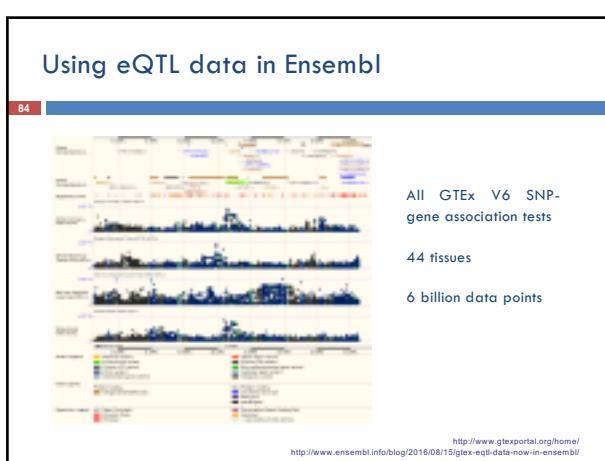
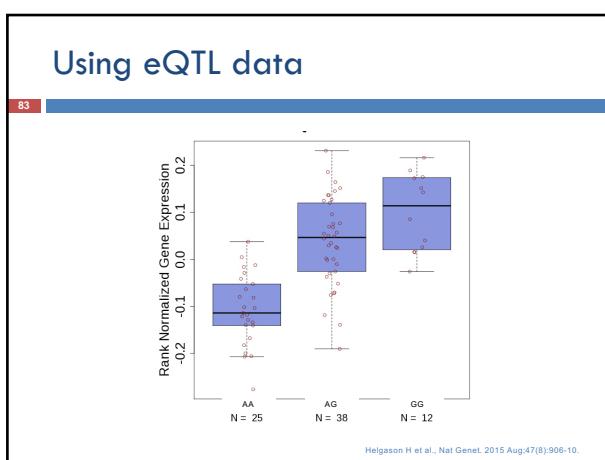
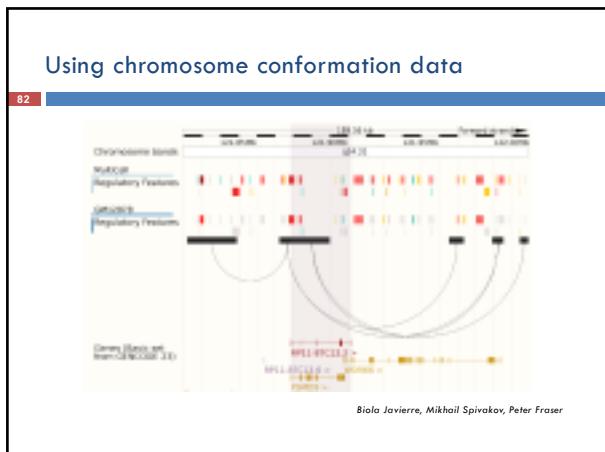
If there is no gene association, can we link it to a gene through chromatin conformation data?

Are there any transcription factors binding to the locus?

If yes, which ones?
Does the SNP affect the binding motif? How much?

79 Zhang X et al., Trends in Genet. 2014 Apr 14; 30(4):140-149.





Epigenomic and Transcriptomic Assays

