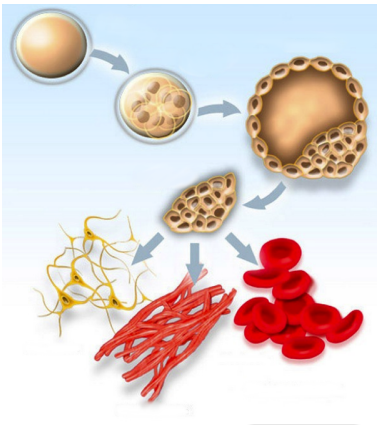
UNDERSTANDING GENE REGULATION – PART I

Myrto Kostadima

Ensembl Regulation Project Leader, EMBL-EBI

One Genome – Many cell types



- Essentially all cells of an organism share
  - the same genome
  - the same genes
- Hundreds of different cell types with a clearly distinct phenotype
- Difference?
  - Gene expression profiles

Gene regulation (almost sole) source of complexity

3

- Bacteria: 1,500-7,500 genes
- Drosophila: 14,000 genes
- C. elegans: 19,000
- Human: ?

2003

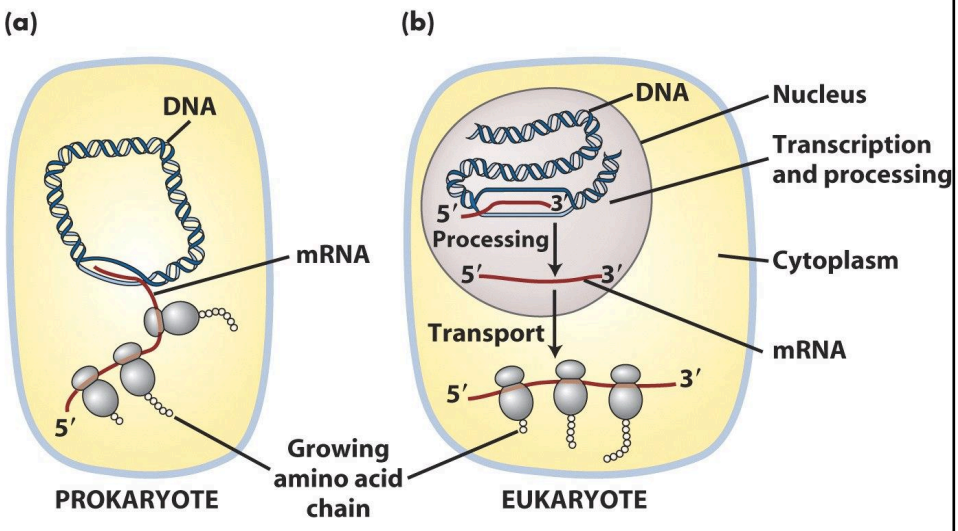
Completion of the Human Genome Project

Human genes: ~35,000

Human genes: > 100,000

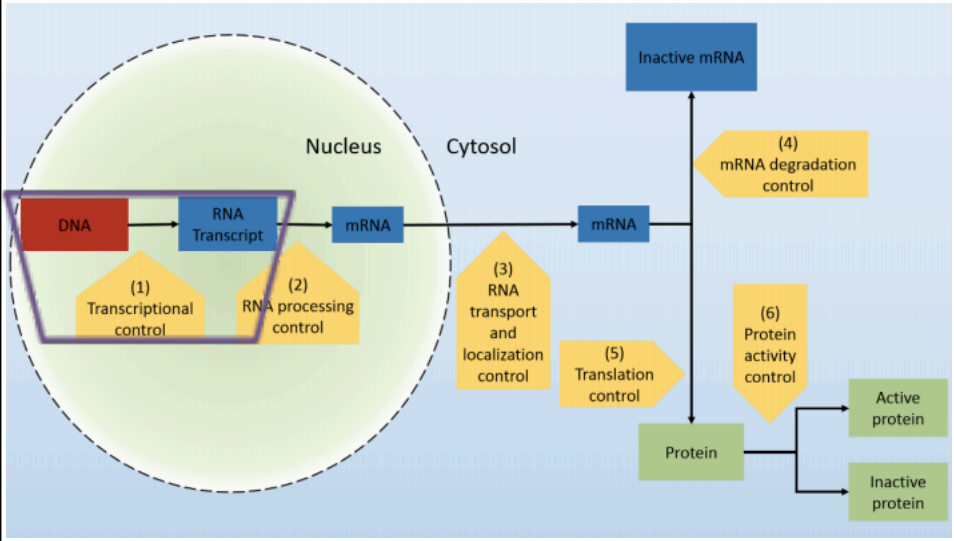
Human genes: ~21,000

4



5

http://barleyworld.org/sites/default/files/figure-08-01_0.jpg

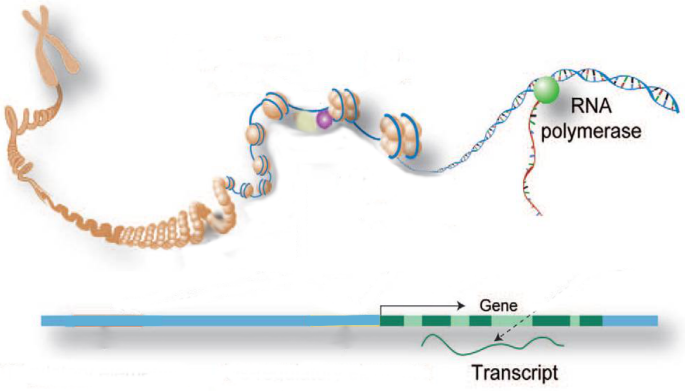


6

Alberts B et al., "Molecular Biology of the Cell", Sixth edition (2014)

From protein-coding DNA..

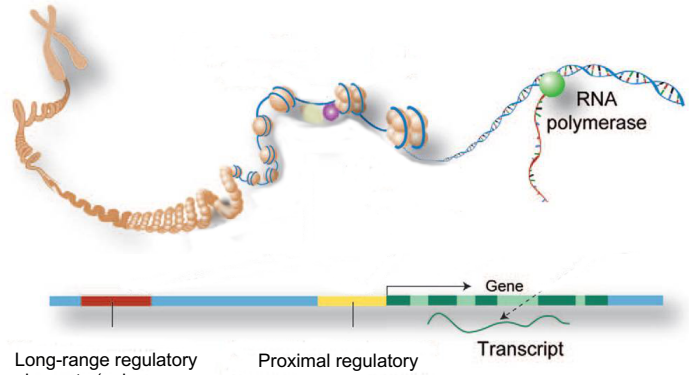
7



Adapted from "A user's guide to the encyclopedia of DNA elements (ENCODE).": PLoS Biol. 2011 Apr;9(4):e1001046

.. to regulatory DNA

8



Long-range regulatory elements (enhancers, repressors/silencers, insulators)

Proximal regulatory elements (promoters, transcription factor binding sites)

Adapted from "A user's guide to the encyclopedia of DNA elements (ENCODE).": PLoS Biol. 2011 Apr;9(4):e1001046

Elements of gene regulation

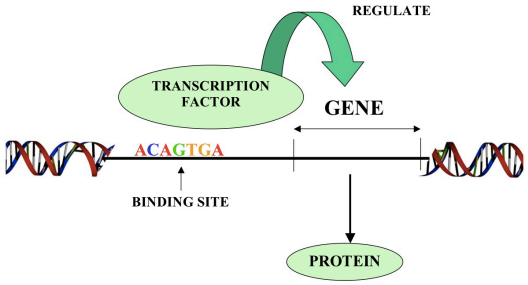
9

- Transcription factors
  - Promoters, enhancers, insulators..
- RNA Pol II machinery
- Epigenetics
  - DNA methylation
  - Nucleosome positioning/Open chromatin
  - Histone modification
  - Euchromatin/heterochromatin
- 3D Chromatin organisation

Transcription factors (TFs)

10

- TFs are a subset of genes that encode proteins that bind to sequences to regulate transcription



  - ~1, 000 human genes are TFs (depending on definition)

Vaquerizas JM et al., Nat Rev Genet. 2009 Apr;10(4):252-63.

Prokaryotes vs Eukaryotes

11

- simple process
- genes organised in operons
- one TF per operon

- complex process
- many TFs per gene
- many regulatory elements per gene

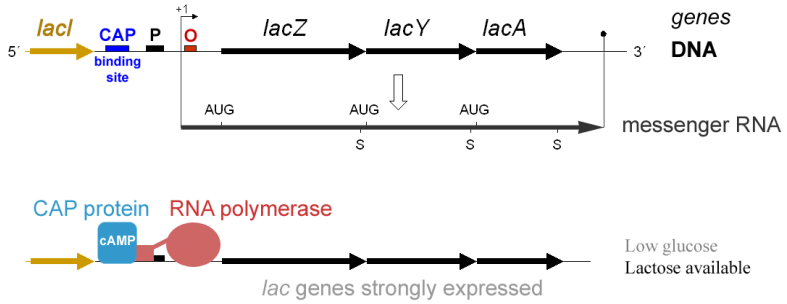
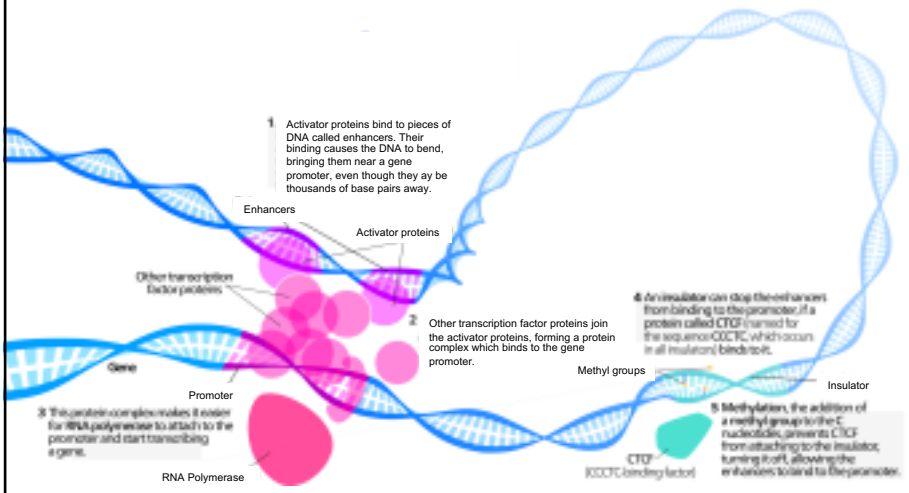


Image adapted from: https://commons.wikimedia.org/wiki/File:Lac_operon-2010-21-01.png#/media/File:Lac_operon-2010-21-01.png

TFs and regulatory regions

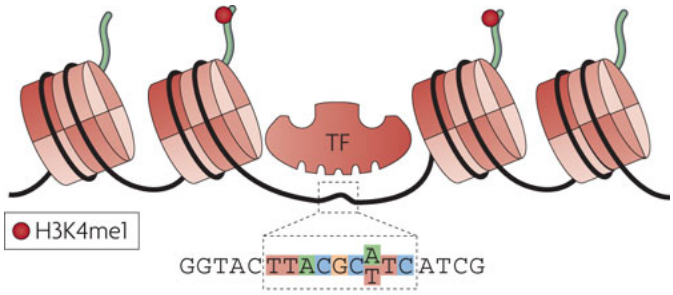
12

"Transcription Factors" by Kelvinsong https://commons.wikimedia.org/wiki/File:Transcription_Factors.svg#/media/File:Transcription_Factors.svg



TF binding site and motif

13



Nature Reviews | Genetics

Position Weight Matrices

14

GAGGTAAAC
TCCGTAAGT
CAGGTTGGA
ACAGTCAGT
TAGGTCATT
TAGGTACTG
ATGGTAACT
CAGGTATAC
TGTGTGAGT
AAGGTAAGT

Consensus sequence TAGGTAAGT

- A PWM is a commonly used representation of motifs in biological sequences.
- PWMs are often derived from a set of aligned sequences that are thought to be functionally related.
- They have become an important part of many software tools for computational motif discovery.

Position Weight Matrices

15

```
POS 1|2|3|4|5|6|7|8|9
 1. G|A|G|G|T|A|A|A|C
 2. T|C|C|G|T|A|A|G|T
 3. C|A|G|G|T|T|G|G|A
 4. A|C|A|G|T|C|A|G|T
 5. T|A|G|G|T|C|A|T|T
 6. T|A|G|G|T|A|C|T|G
 7. A|T|G|G|T|A|A|C|T
 8. C|A|G|G|T|A|T|A|C
 9. T|G|T|G|T|G|A|G|T
10. A|A|G|G|T|A|A|G|T
```

```
POS 1 2 3  4  5 6 7 8 9
A [ -  - -  -  - -  - - - ]
C [ -  - -  -  - -  - - - ]
G [ -  - -  -  - -  - - - ]
T [ -  - -  -  - -  - - - ]
```

Position Weight Matrices

16

```
POS 1|2|3|4|5|6|7|8|9
 1. G|A|G|G|T|A|A|A|C
 2. T|C|C|G|T|A|A|G|T
 3. C|A|G|G|T|T|G|G|A
 4. A|C|A|G|T|C|A|G|T
 5. T|A|G|G|T|C|A|T|T
 6. T|A|G|G|T|A|C|T|G
 7. A|T|G|G|T|A|A|C|T
 8. C|A|G|G|T|A|T|A|C
 9. T|G|T|G|T|G|A|G|T
10. A|A|G|G|T|A|A|G|T
```

```
POS 1 2 3  4  5 6 7 8 9
A [ 3  6 1  0  0 6  7 2 1 ]
C [ 2  2 1  0  0 2  1 1 2 ]
G [ 1  1 7 10  0 1  1 5 1 ]
T [ 4  1 1  0 10 1  1 2 6 ]
```

Position Weight Matrices

17

```
POS 1|2|3|4|5|6|7|8|9
 1. G|A|G|G|T|A|A|A|C
 2. T|C|C|G|T|A|A|G|T
 3. C|A|G|G|T|T|G|G|A
 4. A|C|A|G|T|C|A|G|T
 5. T|A|G|G|T|C|A|T|T
 6. T|A|G|G|T|A|C|T|G
 7. A|T|G|G|T|A|A|C|T
 8. C|A|G|G|T|A|T|A|C
 9. T|G|T|G|T|G|A|G|T
10. A|A|G|G|T|A|A|G|T
```

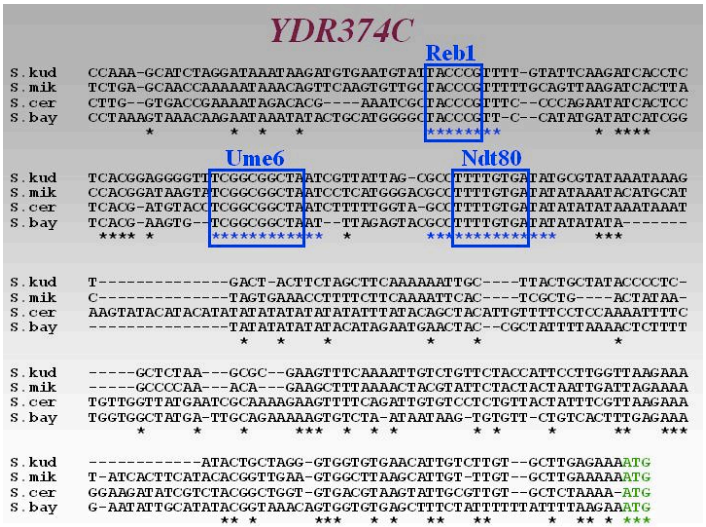
```
POS  1  2  3   4   5  6  7  8  9

A  [ 3  6  1   0   0  6   7  2  1 ]
C  [ 2  2  1   0   0  2   1  1  2 ]
G  [ 1  1  7  10   0  1   1  5  1 ]
T  [ 4  1  1   0  10  1   1  2  6 ]

A  [ .3 .6 .1   0   0 .6  .7 .2 .1 ]
C  [ .2 .2 .1   0   0 .2  .1 .1 .2 ]
G  [ .1 .1 .7   1   0 .1  .1 .5 .1 ]
T  [ .4 .1 .1   0   1 .1  .1 .2 .6 ]
```

Phylogenetic Footprinting

18



Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

19

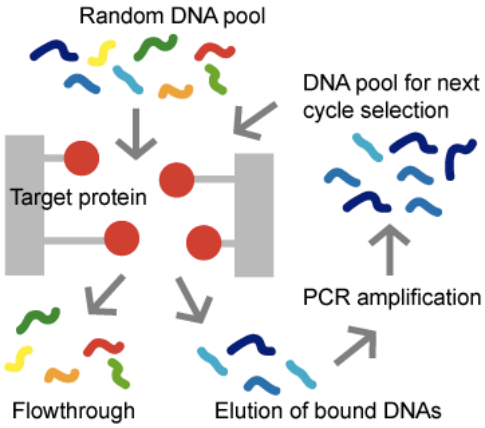
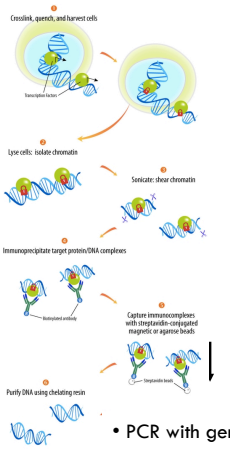


Image from: http://altair.sci.hokudai.ac.jp/g6/Projects/Selex-e.html

Chromatin ImmunoPrecipitation (ChIP)

20

- Transcription factors
- RNA Polymerase II



- PCR with gene specific primers
- Hybridisation on microarrays
- Sequencing

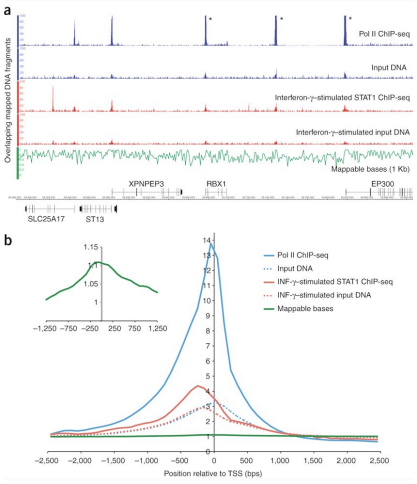
Chromatin ImmunoPrecipitation followed by sequencing (ChIP-seq)

21

- One of the early applications of high-throughput sequencing
- First studies published in 2007
  - Johnson et al (Science) - NRSF
  - Barski et al (Cell) - histone methylation
  - Robertson et al (Nature Methods) - STAT1
  - Mikkelsen et al (Nature) - histone modification
- Thousands of publications currently in PubMed

Why we need a control sample

22



- Open chromatin regions are fragmented more easily than closed regions.
- Repetitive sequences might seem to be enriched (inaccurate repeats copy number in the assembled genome).
- Uneven distribution of sequence tags across the genome.
- A ChIP-seq peak should be compared with the same region in a matched control.

Rozowsky J et al., Nat Biotechnol. 2009 Jan;27(1):66-75.

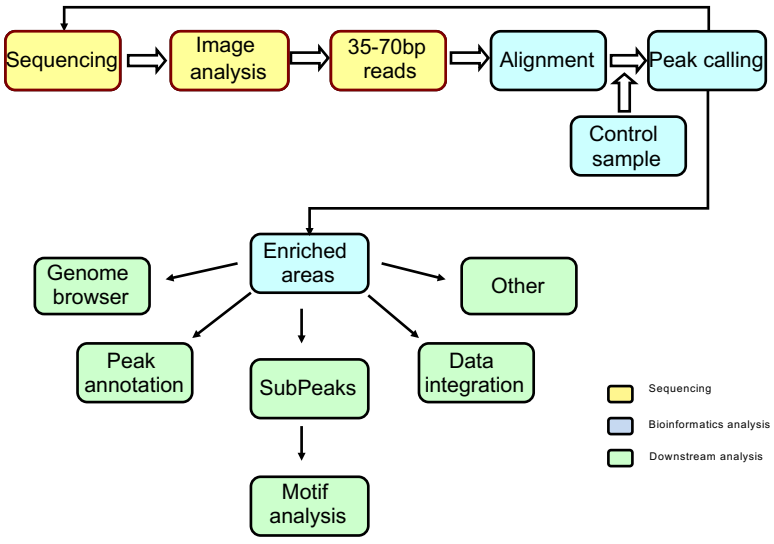
Control types in ChIP-seq

23

- Input DNA
- Mock IP - DNA obtained from IP without antibody
  - Very little material can be pulled down leading to inconsistent results of multiple mock Ips.
- Nonspecific IP - using an antibody against a protein that is not known to be involved in DNA binding
- Sequencing a control can be avoided when looking at:
  - time points
  - differential binding pattern between conditions

Analysis - Overview

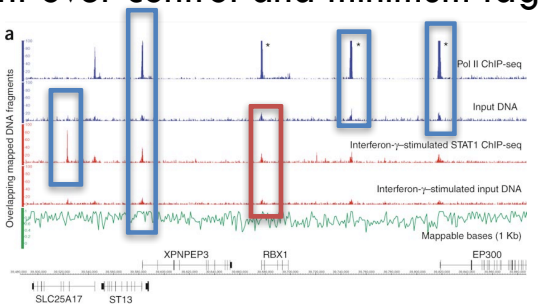
24



Peak Calling

25

Regions are scored by the number of tags in a window of a given size. Then assessed by enrichment over control and minimum tag density.



Rozowsky J et al., Nat Biotechnol. 2009 Jan;27(1):66-75.

Peak Calling - Challenges

26

- Adjust for sequence mappability - regions that contain repetitive elements have different expected tag count
- Different ChIP-seq applications produce different type of peaks. Most current tools have been designed to detect sharp peaks (TF binding)
- Alternative tools exist for broader peaks (histone modifications that mark domains - transcribed or repressed), e.g. SICER

MACS - Peak detection

27

- Remove duplicate tags (in excess of what can be expected by chance)
- Slide window across the genome to find candidate peaks with a significant tag enrichment (Poisson distribution, global background, default p-value 10e-5)
- Merge overlapping peaks, and extend each tag d bases from its center
- Also looks at local background levels and eliminates peaks that are not significant with respect to local background
- Uses the control sample to eliminates peaks that are also present there

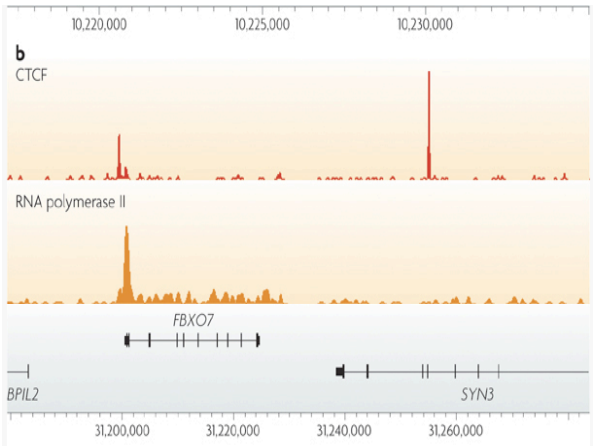
Analysis downstream to peak calling

28

- Visualisation - genome browser: Ensembl, UCSC, IGV
- Peak Annotation - finding interesting features surrounding peak regions
- Correlation with expression data
- Discovery of binding sequence motifs
- Gene Ontology analysis on genes bound by the same transcription factor
- Correlation with SNP data to find allele-specific binding

Visualisation in a genome browser

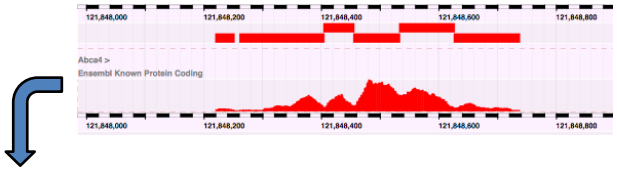
29



Park J, Nature Reviews Genetics, 2009

Motif Analysis

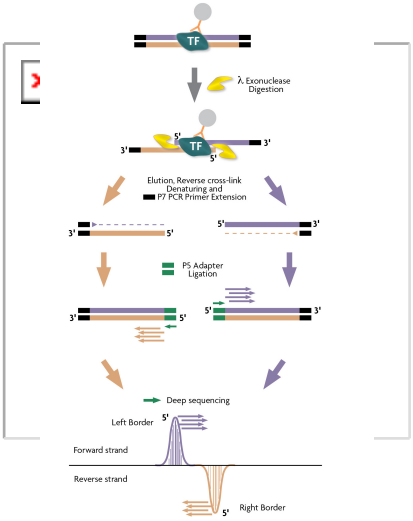
30



GAATCCCACA TTTGCATAACAAAAG ACTCCTGGTG
CAGCTGCTCT TCTGCATAACAAAGG GTGGCCCTGC
CCGGTTTTTC TTTGCATAACAATAA GATCTGGCTA
TTATTCTCAC TTTGCATAGGAATGG GGCAGTTAGA
CACAGCCACA TTTGCATAACAGAAG CCGAGCCCGC
CTTGGGTGAA TTTGCAAGACAAAGG ACAATGATCA

ChIP-exo

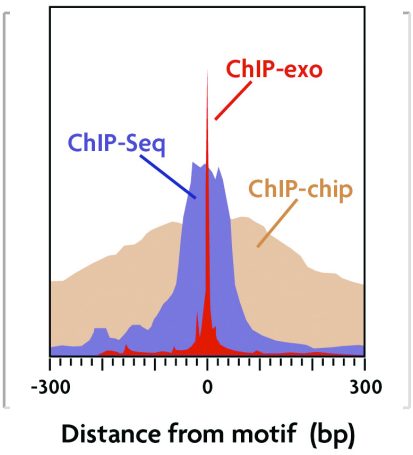
31



- ChIP-exo is a modification of the ChIP-seq protocol, improving the resolution of binding sites from hundreds of base pairs to almost one base pair.
- It employs the use of exonucleases to degrade strands of the protein-bound DNA in the 5'-3' direction to within a small number of nucleotides of the protein binding site.

ChIP-exo

32

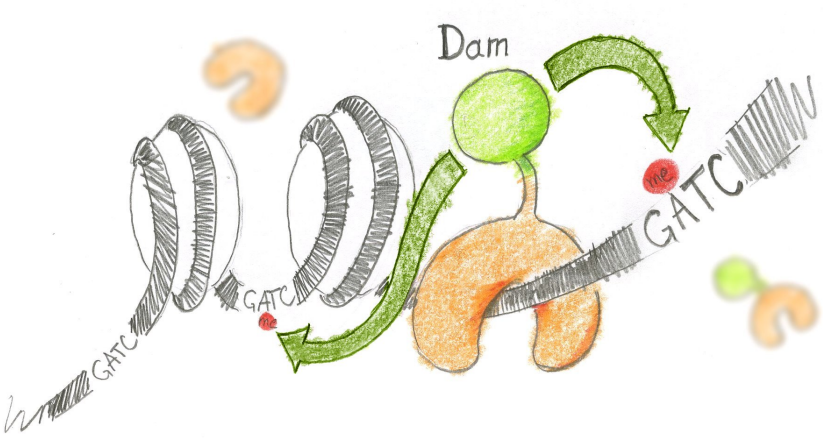


- ChIP-exo is a modification of the ChIP-seq protocol, improving the resolution of binding sites from hundreds of base pairs to almost one base pair.
- It employs the use of exonucleases to degrade strands of the protein-bound DNA in the 5'-3' direction to within a small number of nucleotides of the protein binding site.

Image adapted from: https://www.activemotif.com/images/products/ChIP-exo_Peak_image.jpg

DNA adenine methyltransferase identification

33

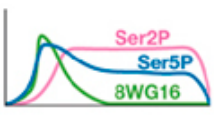


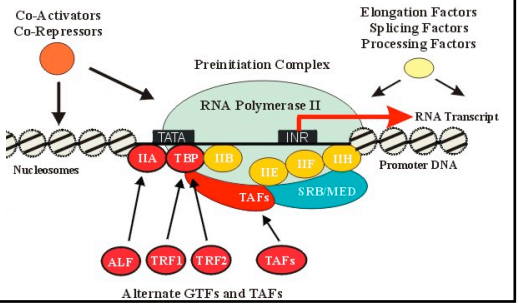
Van Steensel B et al., Nature Biotechnology 18, 424 - 428 (2000)
"DamID concept" by Guillaume Filion - Own work. Licensed under Public Domain via Commons
https://commons.wikimedia.org/wiki/File:DamID_concept.jpg#/media/File:DamID_concept.jpg

RNA Pol II and the transcriptional machinery

34

- Pol II does not act alone
- Needs a Pre-initiation complex (TFs, mediator proteins)
- Pol II is chemically modified in different phases of transcription

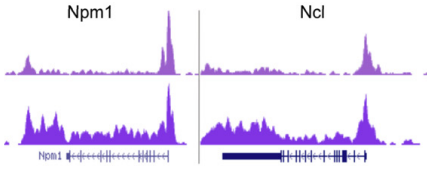




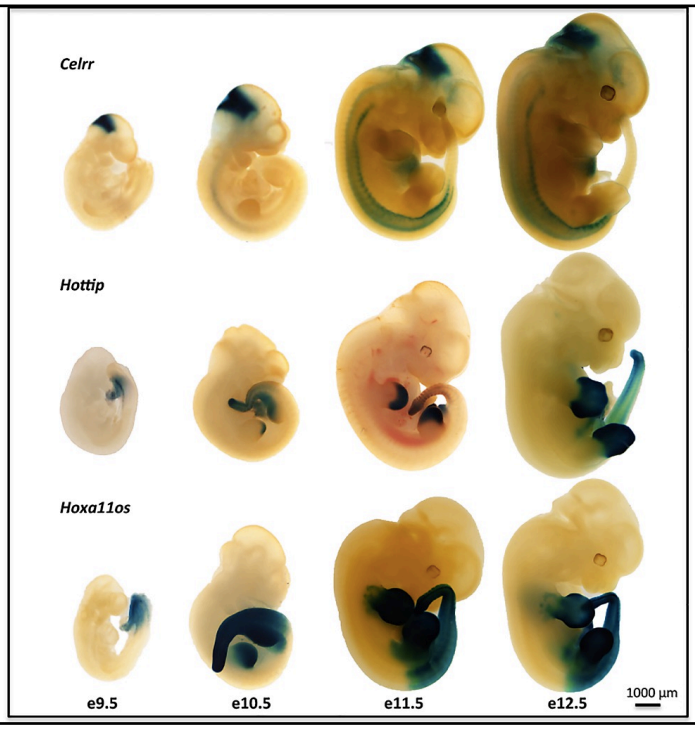
RNA Pol II pausing

35

- RNA Pol II modification & elongation factors needed for full transcripts
- RNA Pol II Pre-initiation complex can be present, but no (or low) transcription: *RNA Pol II pausing*



- RNA Pol II ChIP: Compute Pausing Index/Travelling ratio (signal at TSS vs gene body)



36

Further reading

37

REPORT

Suboptimization of developmental enhancers

Emma K. Farley[1,2,*], Katrina M. Olson[1,2], Wei Zhang[3], Alexander J. Brandt[4], Daniel S. Rokhsar[1], Michael S. Levine[1,2,*]

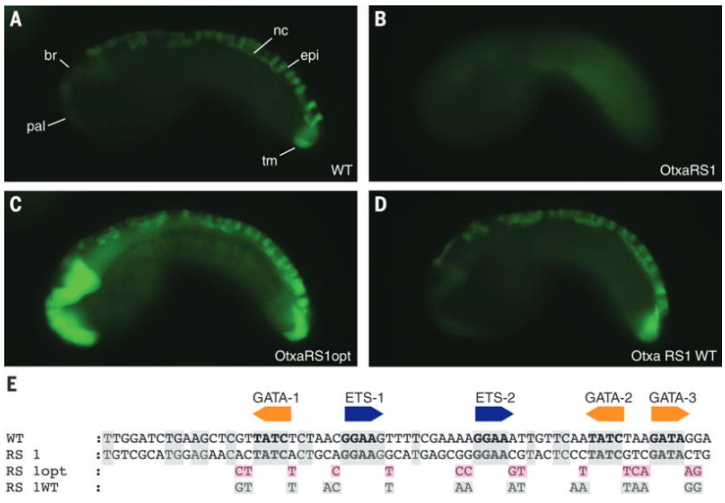
+ Author Affiliations

↵*Corresponding author. E-mail: msl2@princeton.edu (M.S.L.); ekfarley@princeton.edu (E.K.F.)

Science 16 Oct 2015:
Vol. 350, Issue 6258, pp. 325-328
DOI: 10.1126/science.aac6948

Further reading

38



39

Gene Regulation Practical

Prerequisites for next week's practical:

1. Install R packages:
    1. rtracklayer
    2. ggplo2
2. Add the following to your UNIX $PATH:
    1. /local/data/genome_informatics/programs/bedtools2

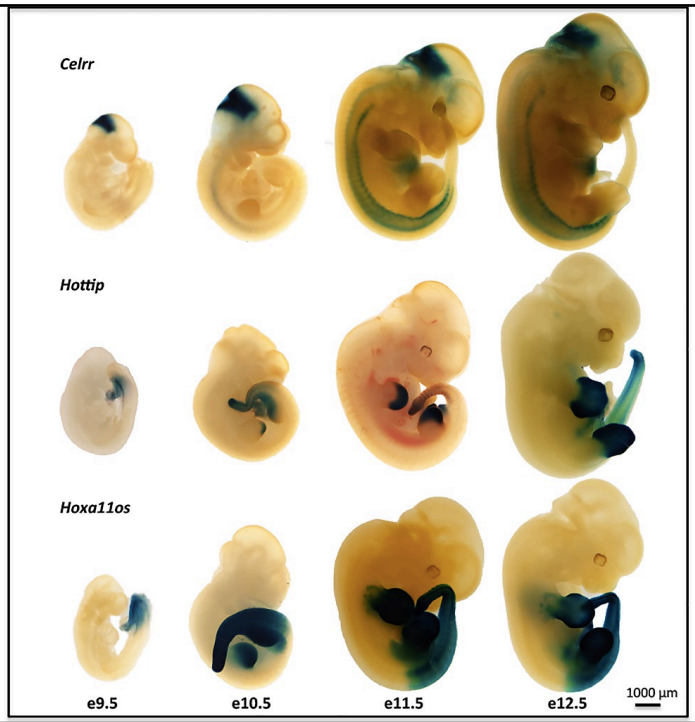
EMBL-EBI

e!Ensembl

UNDERSTANDING GENE REGULATION – PART II

Myrto Kostadima

Ensembl Regulation Project Leader, EMBL-EBI



41

Elements of gene regulation

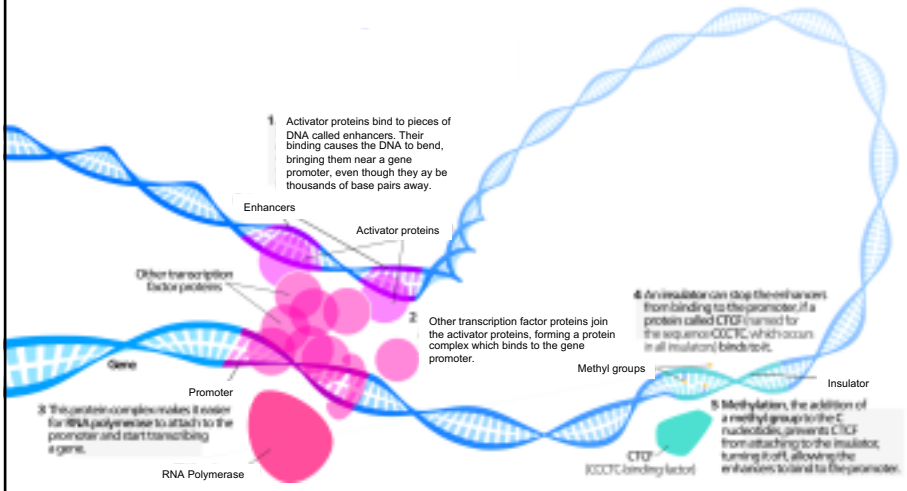
42

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

TFs and regulatory regions

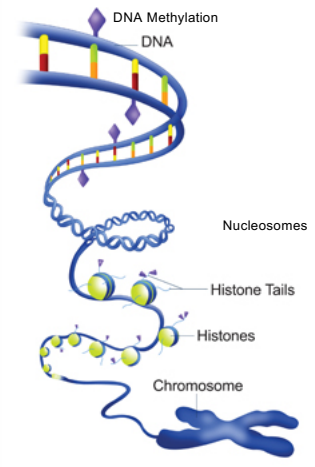
43

"Transcription Factors" by Kelvinsong https://commons.wikimedia.org/wiki/File:Transcription_Factors.svg#/media/File:Transcription_Factors.svg



Epigenetics

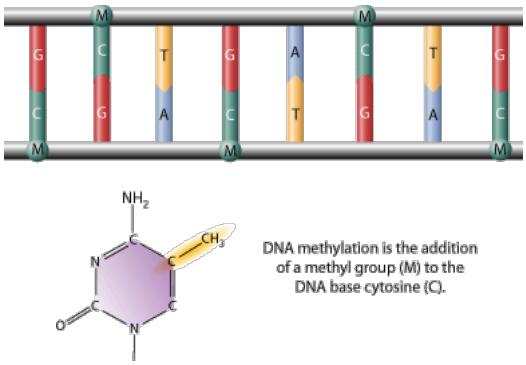
44



- 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

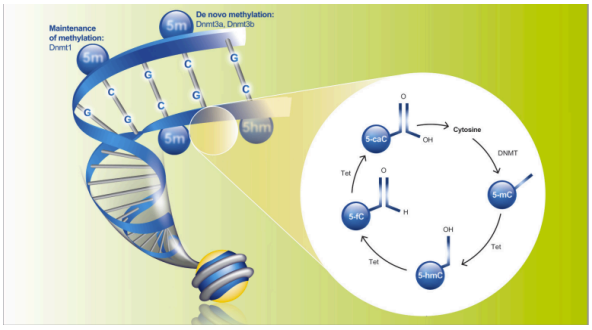
45



- 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



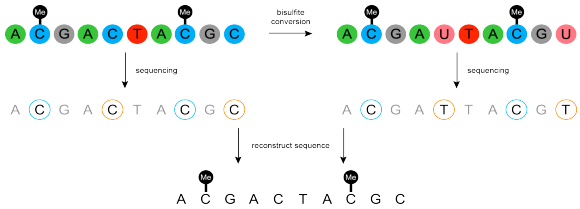
- DNA methyltransferases – DNMT1, DNMT3a/b
- Tet family of enzymes

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:

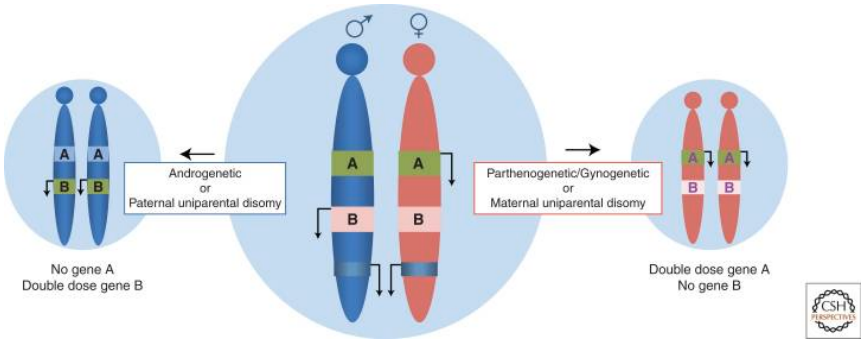


  - Methylation arrays (e.g. 450K)
  - High-throughput sequencing (WGB-seq)

http://www.atdbio.com/content/20/Sequencing-forensic-analysis-and-genetic-analysis

Imprinted genes

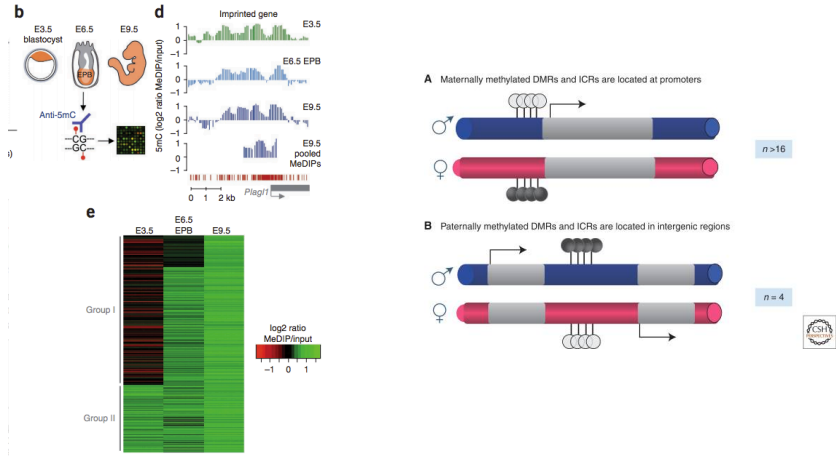
48



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

Parental DNA methylation maintenance

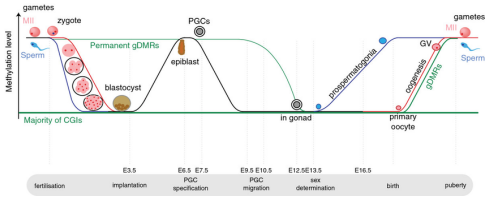
49

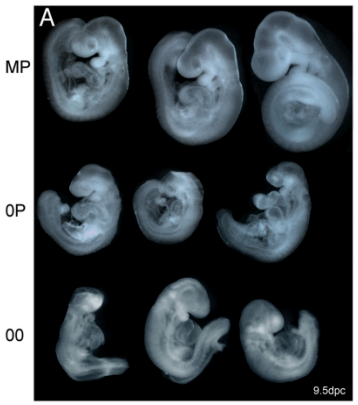


Borgel J et al., Nat Genet. 2010 Dec;42(12):1093-100.
Bartolomei MS and Ferguson-Smith AC, Cold Spring Harb Perspect Biol. 2011 Jul; 3(7): a002592.

Parental DNA methylation maintenance

50

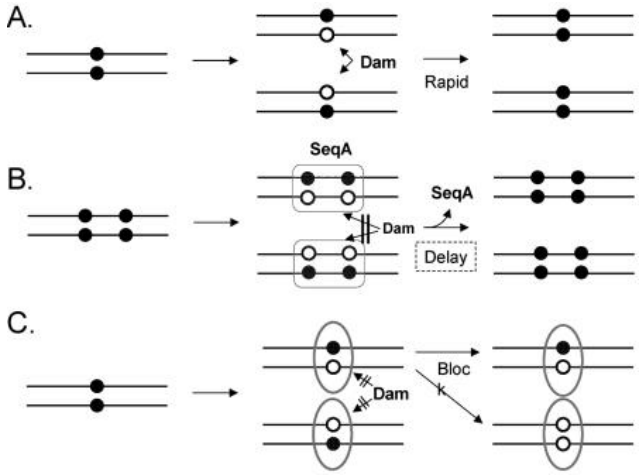




Saadeh H and Schulz R, Epigenetics Chromatin. 2014 Oct 21;7:26.
Schulz R et al., PLoS Genet. 2010 Nov 18;6(11):e1001214.

DNA methylation - Prokaryotes

51

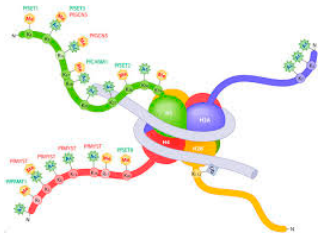


Casadesús J and Low D, Microbiol Mol Biol Rev. 2006 Sep;70(3):830-56.

Histone modifications

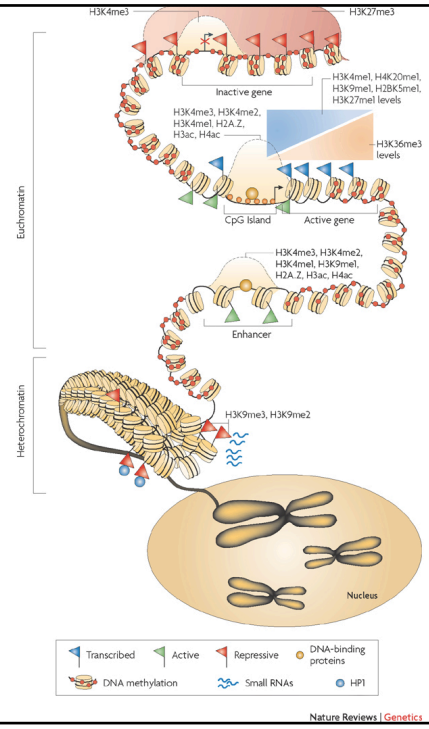
52

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



53

Histone modifications



Schones DE and Zhao K, Nature Rev Genet., March 2008: 9, 179-191.

54

Histone modifications

Histone modification or variant	Signal characteristics	Putative functions
H2A.Z	Peak	Histone protein variant (H2A.Z) associated with regulatory elements with dynamic chromatin
H3K4me1	Peak/Region	Mark of regulatory elements associated with enhancers and other distal elements, but also enriched downstream of transcription starts
H3K4me2	Peak	Mark of regulatory elements associated with promoters and enhancers
H3K4me3	Peak	Mark of regulatory elements primarily associated with promoters/transcription starts
H3K9ac	Peak	Mark of activate regulatory elements with preference for promoters
H3K9me1	Region	Loosely associated with transcription, with preference for 5' end of genes
H3K9me3	Peak/Region	Repressive mark associated with constitutive heterochromatin, repetitive elements and certain broad repressive domains
H3K27ac	Peak	Mark of active regulatory elements; may distinguish active enhancers and promoters from their inactive counterparts
H3K27me3	Region	Repressive mark established by polycomb complex activity associated with repressive domains and silent developmental genes
H3K36me3	Region	Elongation mark associated with transcribed portions of genes, with preference for 3' regions after intron 1
H3K79me2	Region	Transcription-associated mark, with preference for 5' end of genes
H4K20me1	Region	Loosely associated with transcription, with preference for 5' end of genes

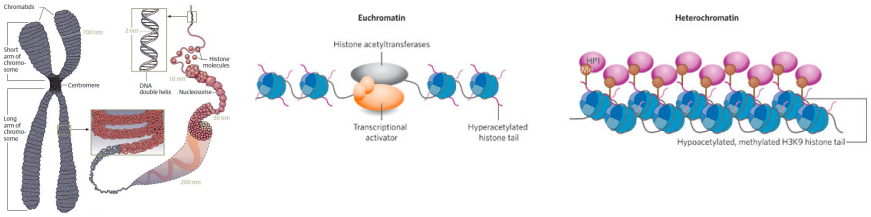
ENCODE
Cheat sheet

Courtesy of Dr Steven Wilder

Euchromatin/Heterochromatin

55

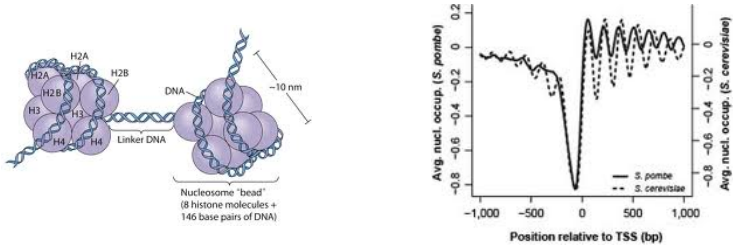
- On a more global level, there are accessible and very dense and mainly inaccessible regions.



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

Nucleosome positioning

56



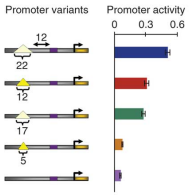
DNA 'wrapped' around histone cores

•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

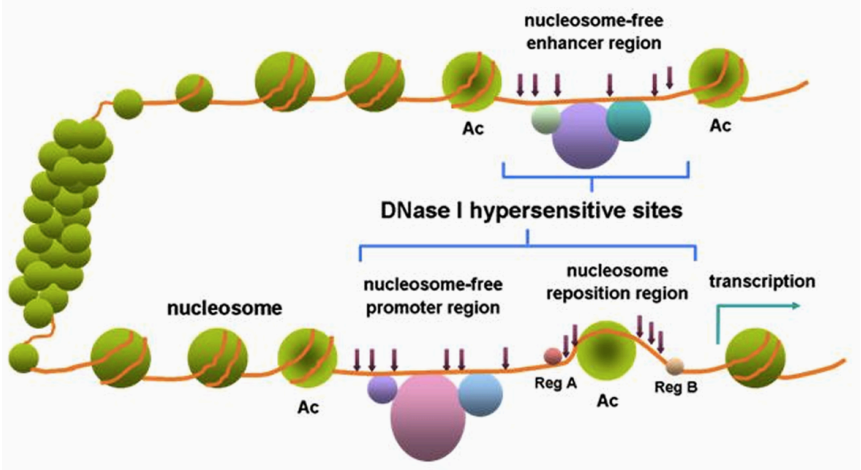


- Nucleosome remodelers (Isw2, RSC)

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

Open-chromatin regions

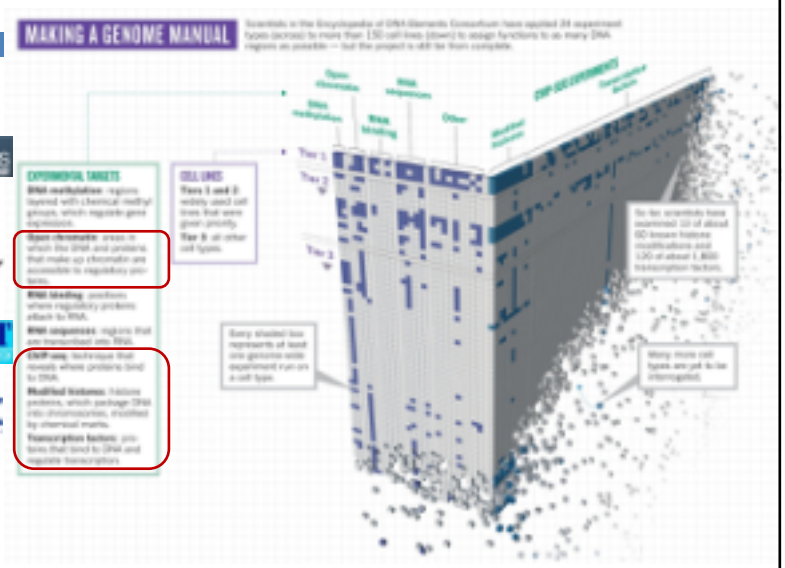
58



- Alternatively we use FAIRE-seq or ATAC-seq

Leveraging available epigenomic data

59

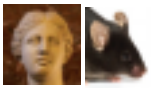


Data resource: ENCODE

60

"Encyclopedia of DNA Elements"

- Systematic analyses of transcription and regulatory information
- 1,640 datasets, 147 cell lines/types → functional elements



ENCODE Project Consortium, Nature. 2007 Jun 14;447(7146):799-816.
ENCODE Project Consortium, Nature. 2012 Sep 6;489(7414):57-74.

Data resource: Roadmap Epigenomics

61

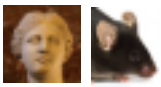
- Public resource of normal epigenomes
  - DNA methylation
  - histone marks
  - open chromatin
  - small RNA

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

Data resource: Blueprint

62

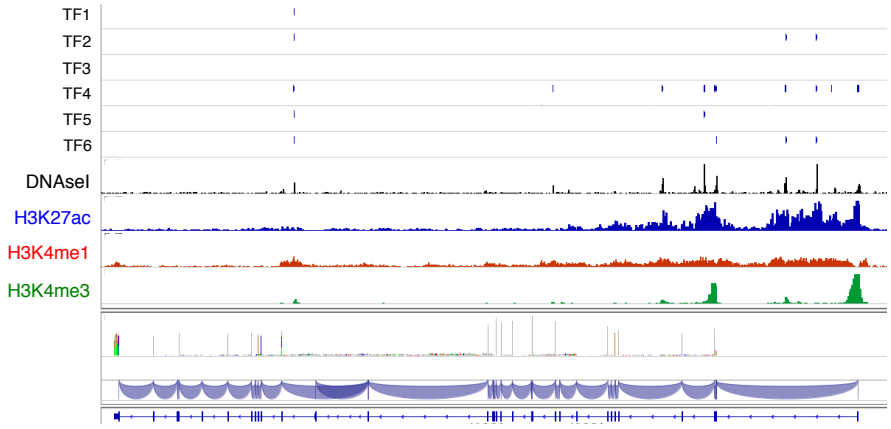
- Generate 100 reference epigenomes
  - Blood cells
  - healthy individuals and
  - malignant leukaemic counterparts



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

GeneX epigenetic landscape

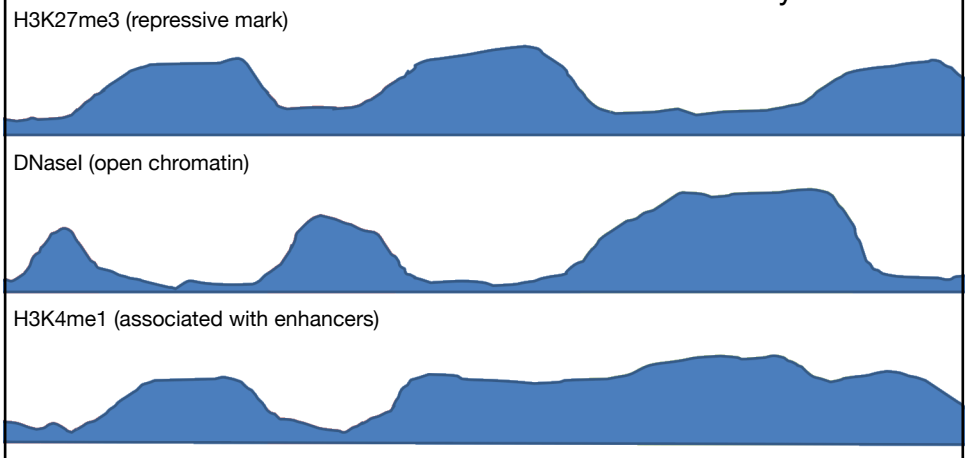
63



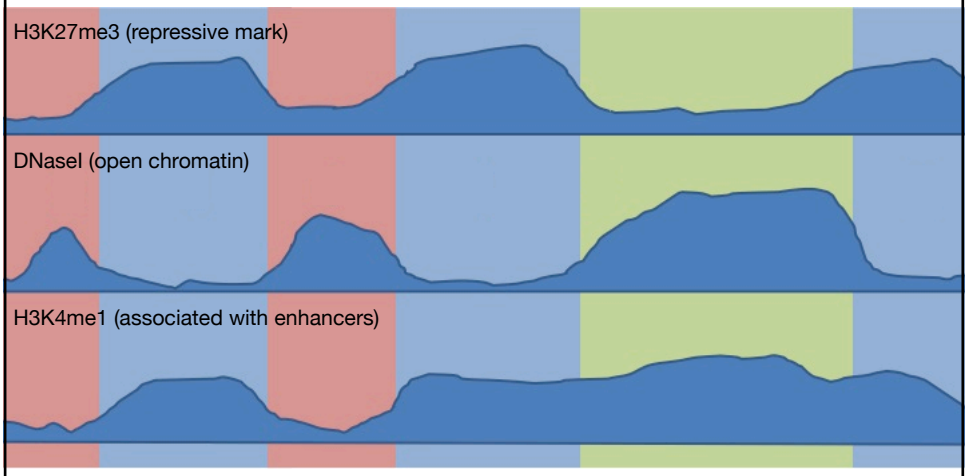
GeneX

Genome Segmentation

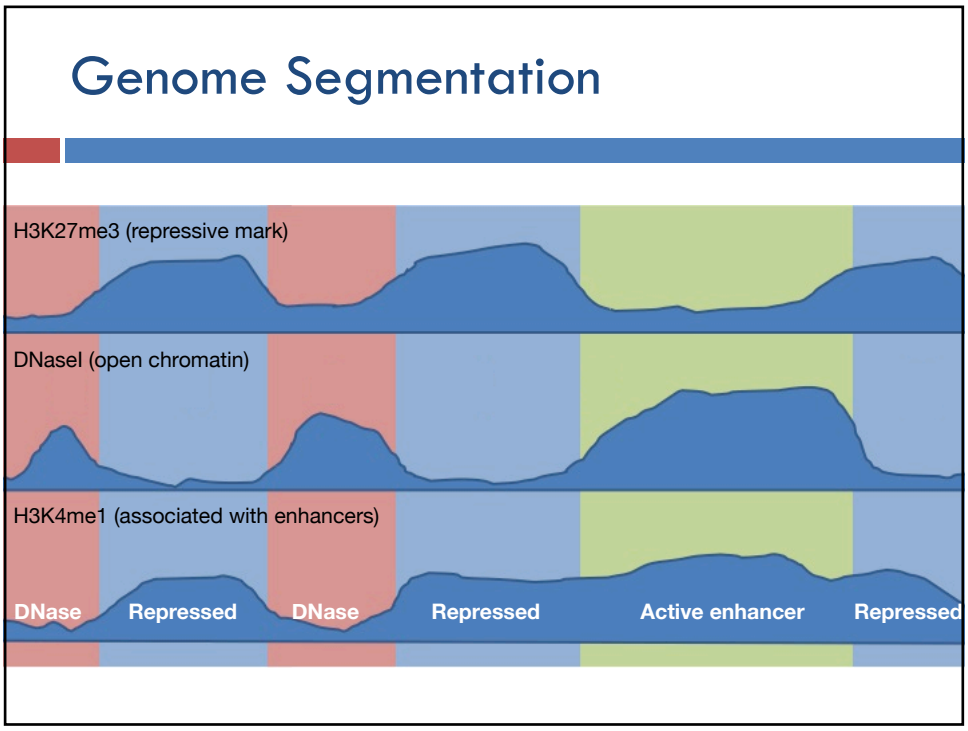
Maximise similarity in labels



Genome Segmentation



Genome Segmentation



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)

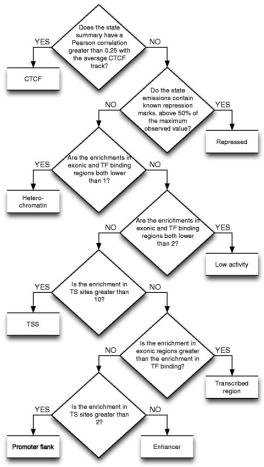
Genome segmentation algorithms

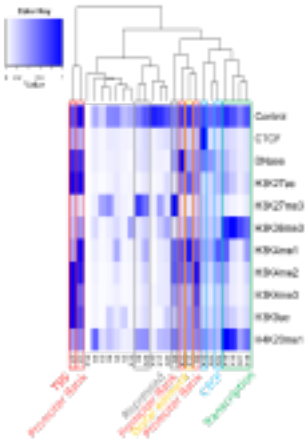
68

	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

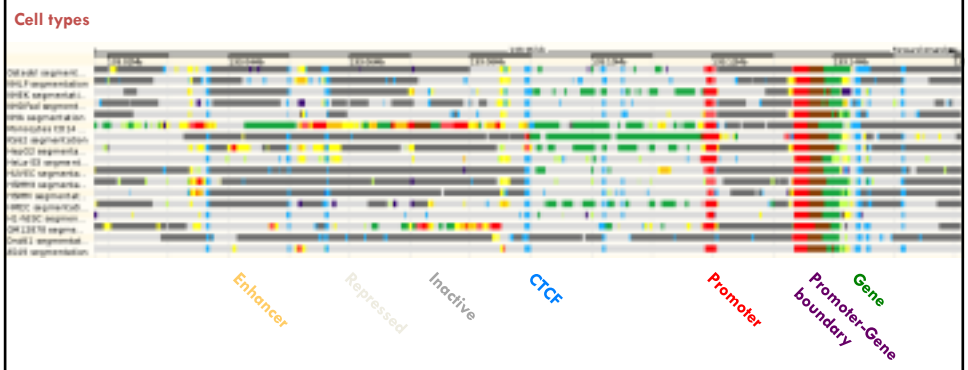




Zerbino DR et al., Genome Biol. 2015 Mar 24;16:56.

Genome segmentation examples

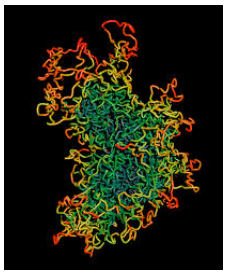
70

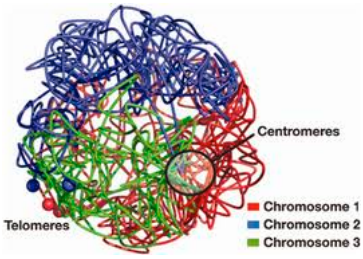


3D organisation in the nucleus

71

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

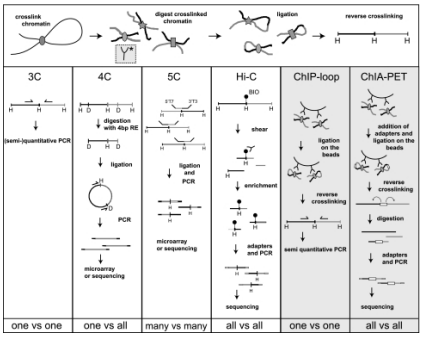




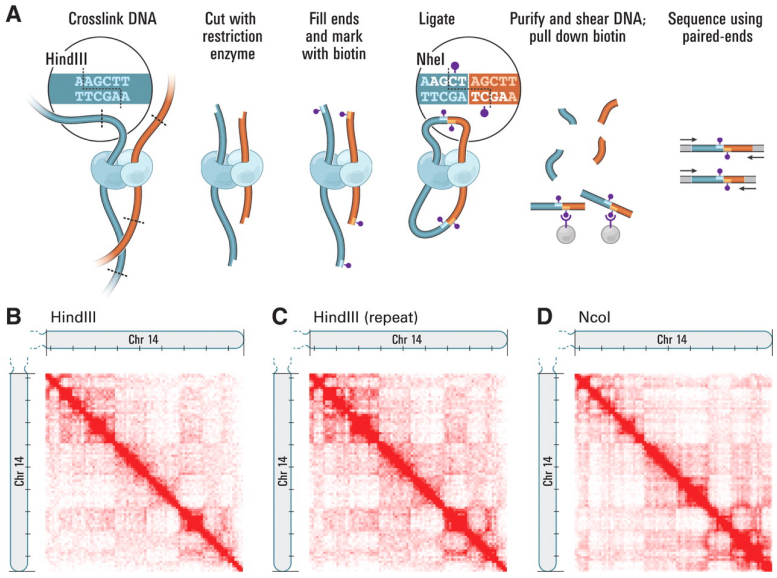
3D organisation in the nucleus

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

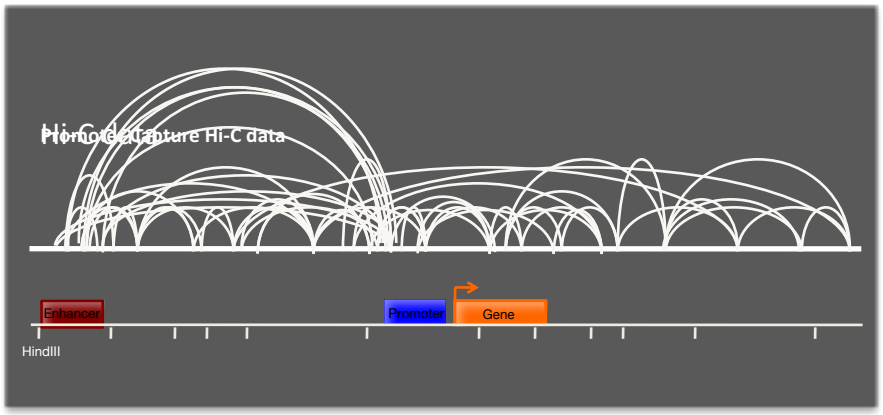


73

E Lieberman-Aiden et al. Science 2009;326:289-293

Promoter capture Hi-C identify long-range promoter-interacting fragments

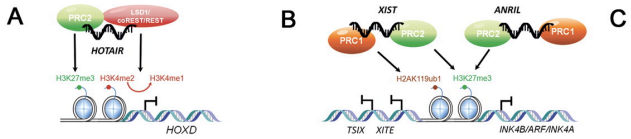
74



Courtesy of Dr Peter Fraser, Babraham Institute

linc-RNAs as epigenetic regulators

75

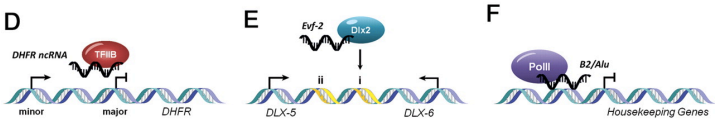


CHROMATIN REMODELLING

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

linc-RNAs as epigenetic regulators

76



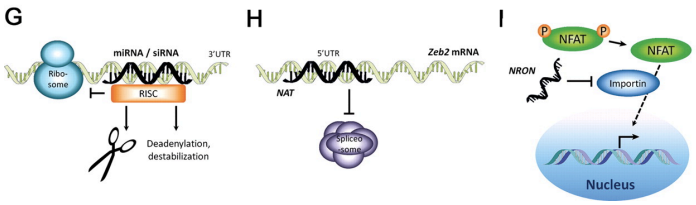
TRANSCRIPTIONAL REGULATION

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

linc-RNAs as epigenetic regulators

77

POST-TRANSCRIPTIONAL CONTROL

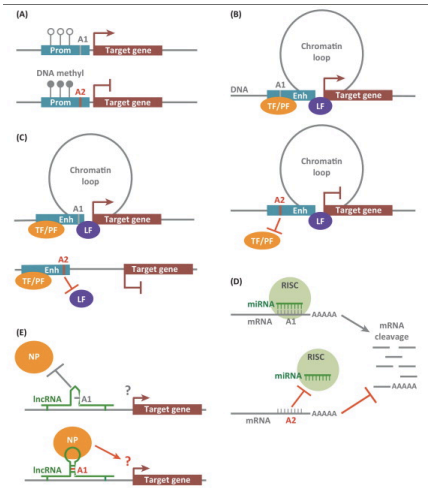


Kaikkonen MU et al., Cardiovascular Res. 2011; 90: 430-440.

"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?

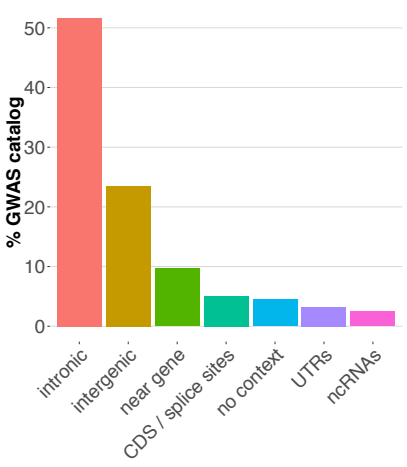
79

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

Why annotate non-coding regions?

80

"(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." *

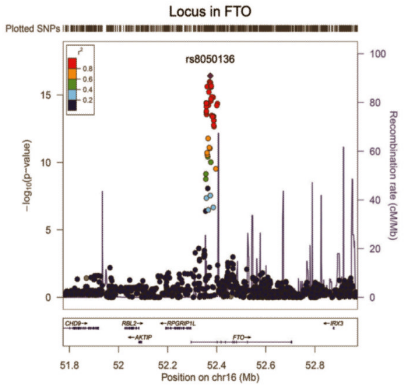


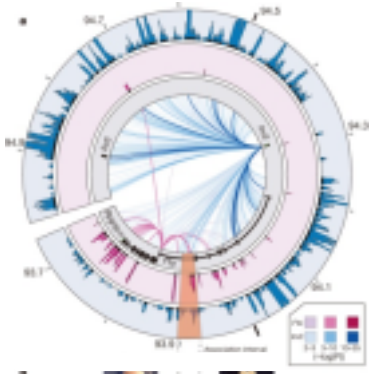
*Maurano MT et al., Nature Genet. 2015 Oct 26.

GWAS catalog v1.0 – associations e86 r2016-10-23

Linking regulatory regions to genes

81

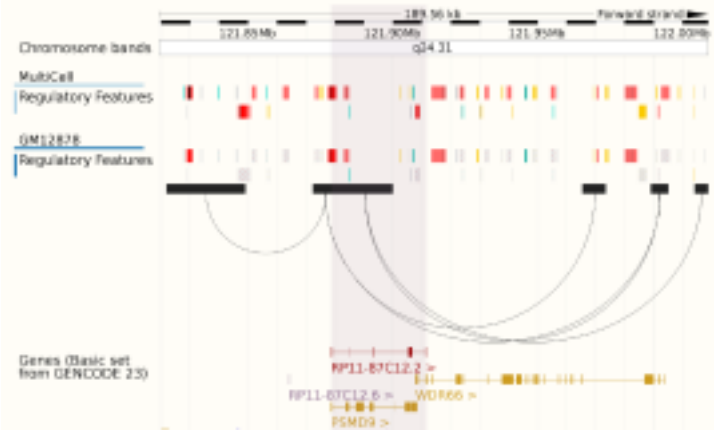




Kilpeläinen TO et al., Nat Genet. 2011 Jun 26;43(8):753-60.
Smemo S et al. Nature. 2014 Mar 20;507(7492):371-5.

Using chromosome conformation data

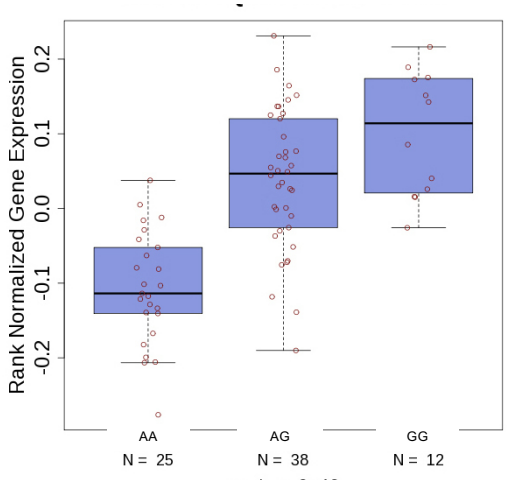
82



Biola Javierre, Mikhail Spivakov, Peter Fraser

Using eQTL data

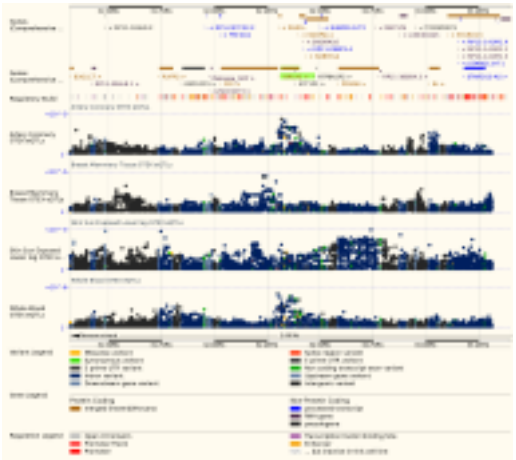
83



Helgason H et al., Nat Genet. 2015 Aug;47(8):906-10.

Using eQTL data in Ensembl

84



All GTEx V6 SNP-gene association tests

44 tissues

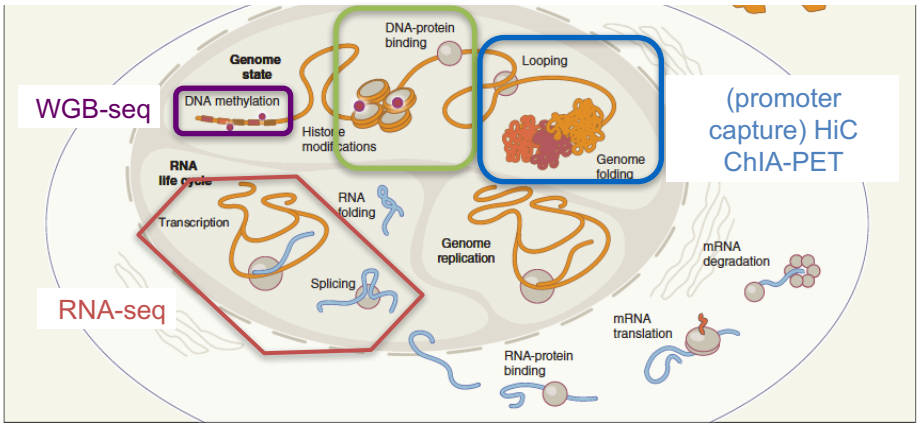
6 billion data points

http://www.gtexportal.org/home/
http://www.ensembl.info/blog/2016/08/15/gtex-eqtl-data-now-in-ensembl/

Epigenomic and Transcriptomic Assays

85

ChIP-seq



WGB-seq

(promoter capture) HiC ChIA-PET

RNA-seq

Shendure J. et al., Nat Biotechnol. 2012 Nov;30(11):1084-94

Acknowledgements

The Entire Ensembl Team

Andrew Yates[1], Wasiu Akanni[1], M. Ridwan Amode[1], Daniel Barrell[1,2], Konstantinos Billis[1], Denise Carvalho-Silva[1], Carla Cummins[1], Peter Clapham[2], Stephen Fitzgerald[1], Laurent Gil[1], Carlos García Girón[1], Leo Gordon[1], Thibaut Hourlier[1], Sarah E. Hunt[1], Sophie H. Janacek[1], Nathan Johnson[1], Thomas Juettemann[1], Stephen Keenan[1], Ilias Lavidas[1], Fergal J. Martin[1], Thomas Maurel[1], William McLaren[1], Daniel N. Murphy[1], Rishi Nag[1], Michael Nuhn[1], Anne Parker[1], Mateus Patricio[1], Miguel Pignatelli[1], Matthew Rahtz[2], Harpreet Singh Riat[1], Daniel Sheppard[1], Kieron Taylor[1], Anja Thormann[1], Alessandro Vullo[1], Steven P. Wilder[1], Amonida Zadissa[1], Ewan Birney[1], Jennifer Harrow[2], Matthieu Muffato[1], Emily Perry[1], Magali Ruffier[1], Giulietta Spudich[1], Stephen J. Trevanion[1], Fiona Cunningham[1], Bronwen L. Aken[1], Daniel R. Zerbino[1] and Paul Flicek[1,2,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and [2]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Funding

wellcome EMBL National Human Genome Research Institute Open Targets BLUEPRINT epigenome

BBSRC bioscience for the future TRANSFORMING GENETIC MEDICINE INITIATIVE MedBioinformatics

Co-funded by the European Union