NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Current Topics in Genome Analysis 2012

THE REGULATORY AND EPIGENETIC LANDSCAPES OF MAMMALIAN GENOMES

Laura Elnitski, Ph.D.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

JOHNS HOPKINS MEDICINE
CONTINUING MEDICAL EDUCATION

Current Topics in Genome Analysis 2012

Laura Elnitski, Ph.D.

No Relevant Financial Relationships with Commercial Interests

2

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

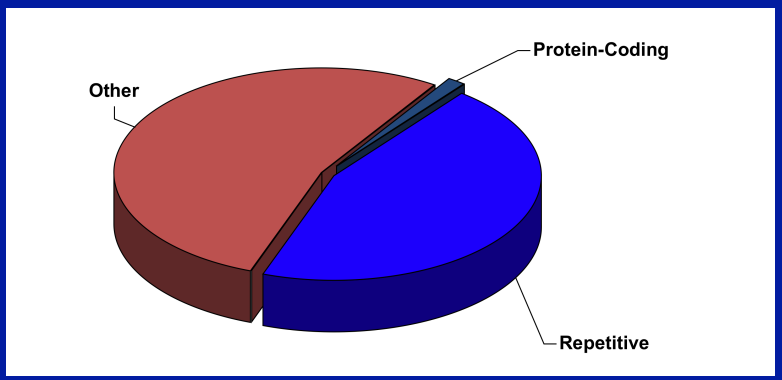
"the most important, most wondrous map ever produced by human kind"

Bill Clinton

3

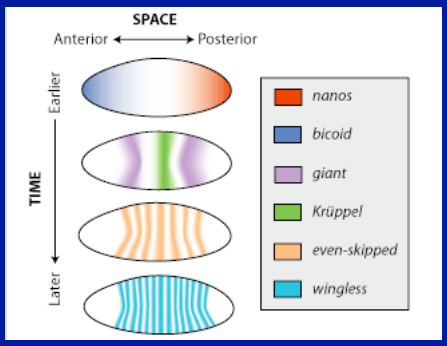
Characterizing the human genome

- ~ 3 billion bases
- 2% protein-coding regions
- 20,000-25,000 protein-coding genes



4

Regulatory elements

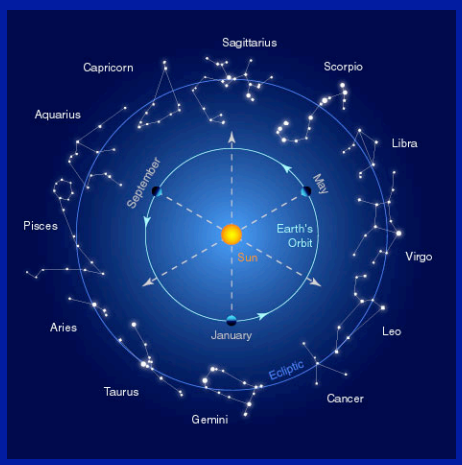


Mutations → Altered processes → Disease

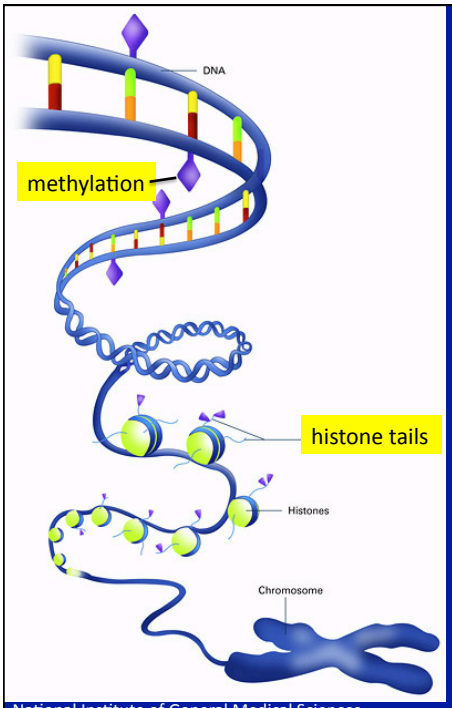
Mutation discovery → Mechanistic understanding → Therapy

5

Regulatory code



6



National Institute of General Medical Sciences

Epigenetic patterns

- Identify cell type
- Distinguish functional elements
- Indicate gene expression levels

7

Mendelian diseases	# of Genes
molecular basis known	3,398
molecular basis unknown	1,790
suspected mendelian basis	1,924

Gut 2011. 60:1739-1753

8

Have we really learned "nothing but probabilities" from the genome ?

9

4 important outcomes of the human genome project

I. Comparative genomics

II. Mapping functional elements

III. Interpretation of disease processes

IV. Reading the epigenetic code

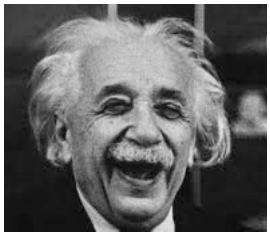
10

At 3 gigabases the genome is equivalent to how many Mozilla browsers?

- 28
- 2.8
- 0.5

11

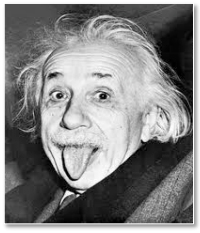
<



<

17,000 23,000 genes 30,000

<



<

1 billion bp 3.4 billion bp 132.8 billion bp

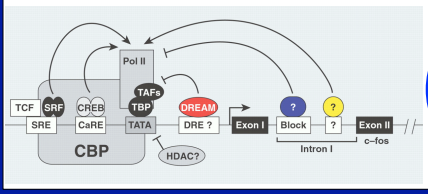
12

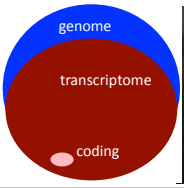
Genomic diversity

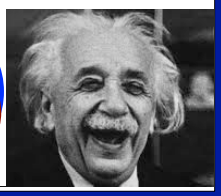
- multiple noncoding elements used in combination
- alternative processing and alternative promoters
- noncoding RNA

13

regulation | function | speciation







Mutational consequences are varied

Original Sentence
TIME TO DREAM

Single Letter Change
T I/A ME TO DREAM → TAME TO DREAM

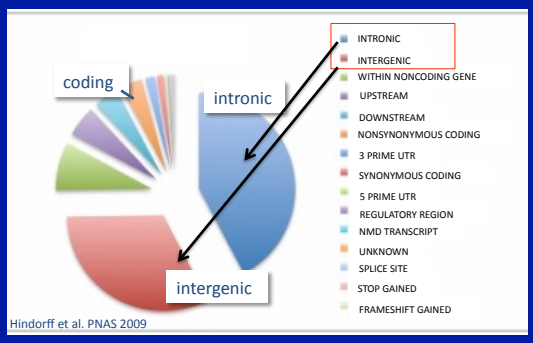
Reverse Order
I M/T E TO DREAM → EMIT TO DREAM

Deletion
T IM E TO DREAM → TETO DR EAM

Insertion
TI ME TO DREAM → TIIM ET ODREAM

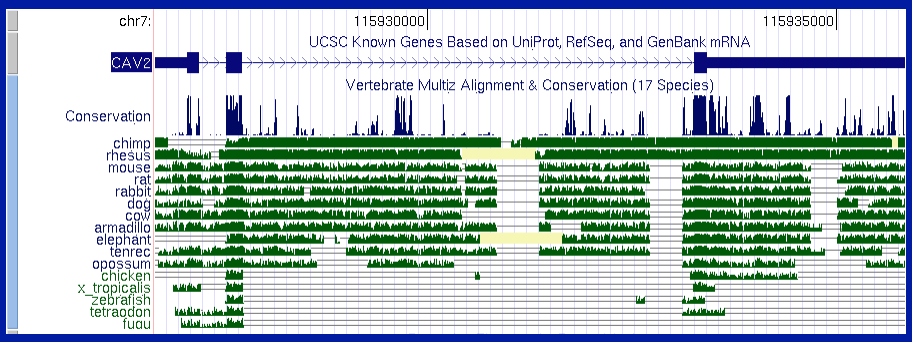
14

Genome wide association data



15

Comparative genomics



Mutations in functional DNA are less likely to be tolerated

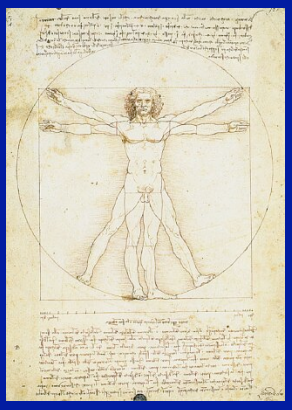
fastest divergence : immune, reproduction and olfaction

slowest divergence : developmental genes

16

How much overlap is there between the human and mouse genomes?

- ***20%?***
- ***40%?***
- ***10%?***



17

Evolutionary Distance

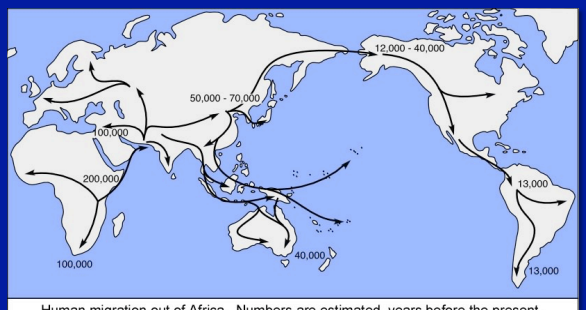

Human


Chimpanzee


Horse


Rat


Platypus

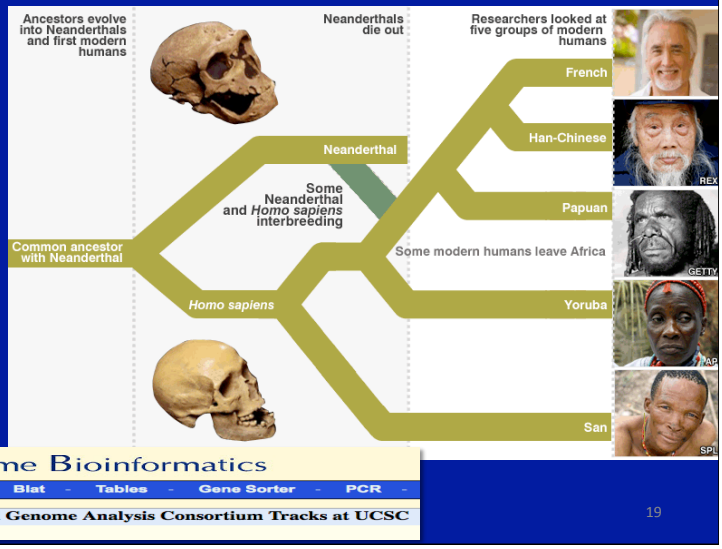


Human migration out of Africa. Numbers are estimated years before the present.

bonner@indiana.edu

What percent of genomic DNA was inherited from Neanderthals in people from Eurasia and the southwestern Pacific?

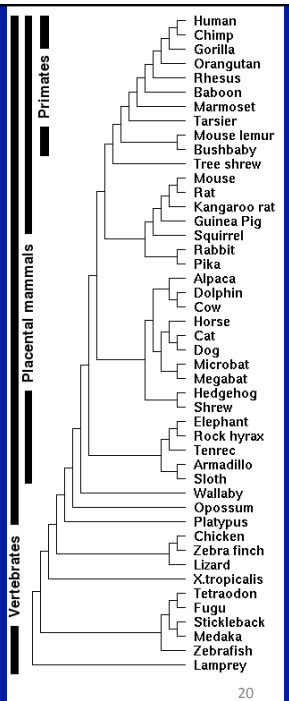
- 1-4%
- 10-12%
- 0.1-0.4%



19

Functional elements

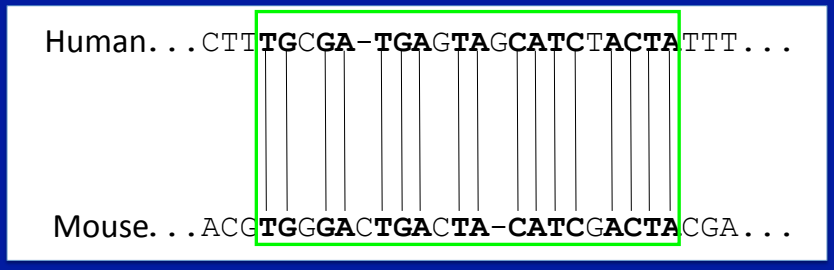
- protein-coding genes
- non-coding RNAs
- regulatory sequences



20

Comparative genomics of related species highlights functional regions

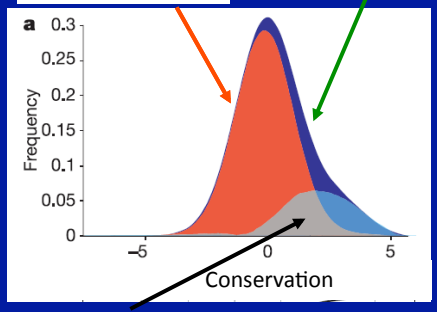
purifying selection



Human-mouse comparisons

neutral regions

40% aligning



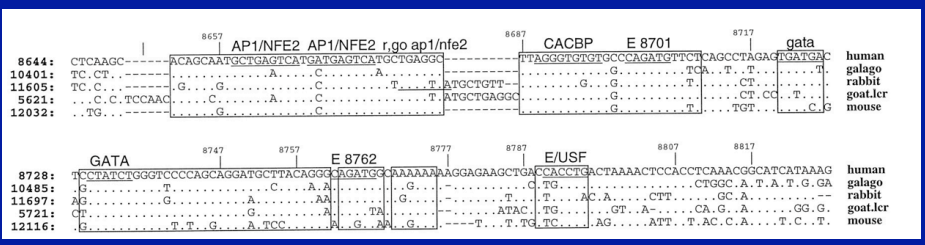
Selectively constrained:

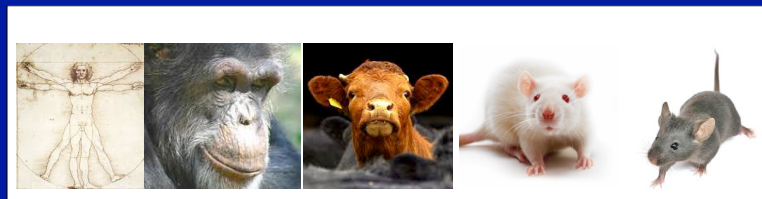
5%

[Mouse consortium, *Nature* 2002]

22

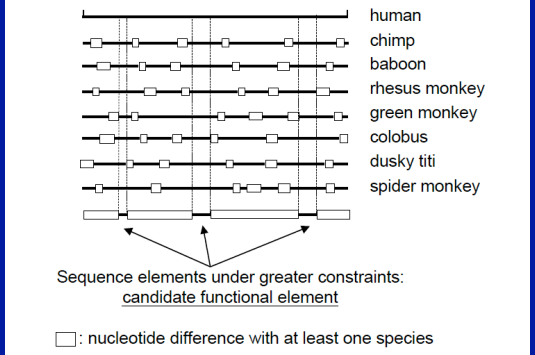
Phylogenetic footprints





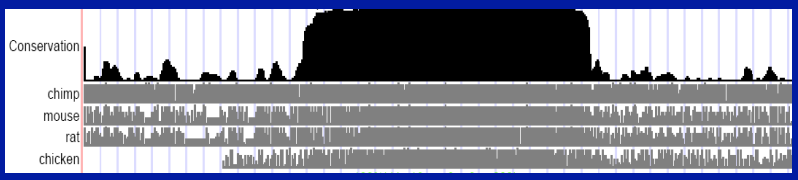
23

Phylogenetic shadowing

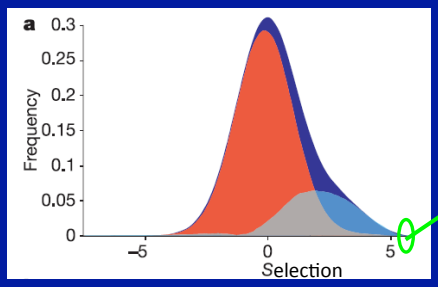


24

Ultraconserved elements

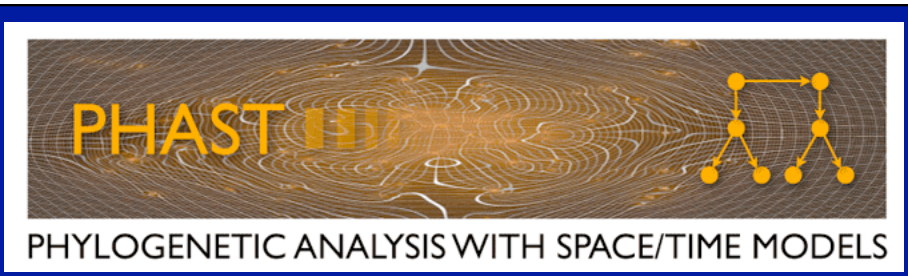


UCSC browser conservation track

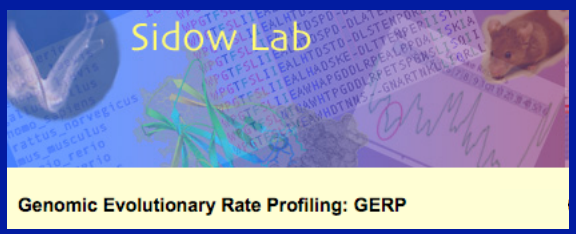


Ultraconservation

25



generative model-based approach



bottom-up approach

26

Unconstrained functional regions

I. lack biological assays

II. chromatin accessibility was more important than sequence composition

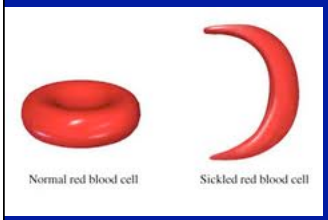
III. lineage-specific

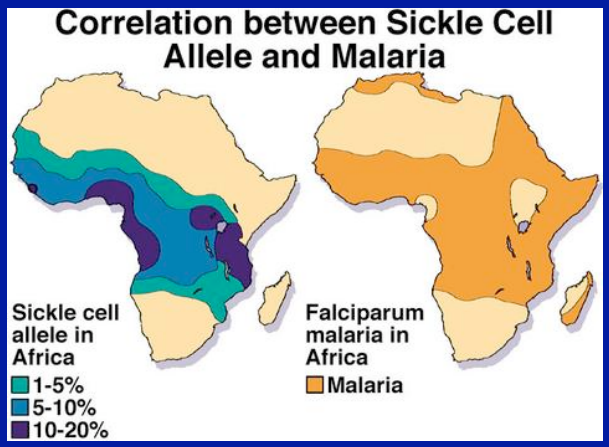
IV. functionally conserved but non-orthologous

V. did not confer a selective advantage or disadvantage

27

Positive selection

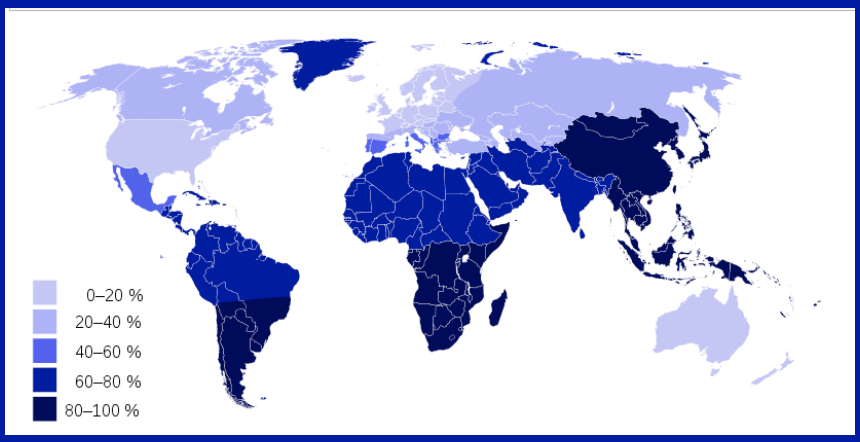




The McGraw-Hill Companies

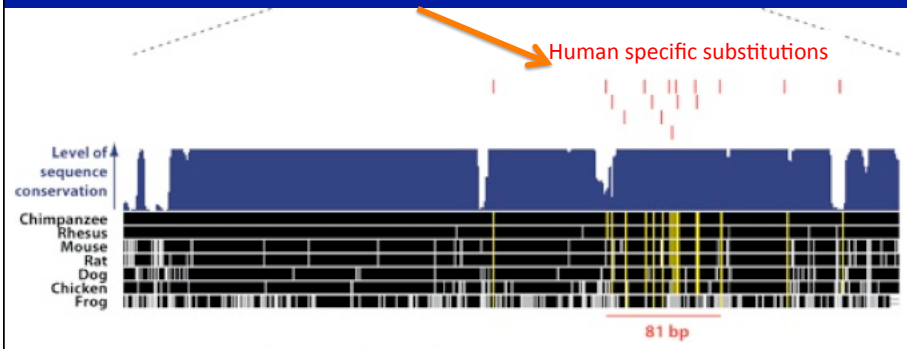
28

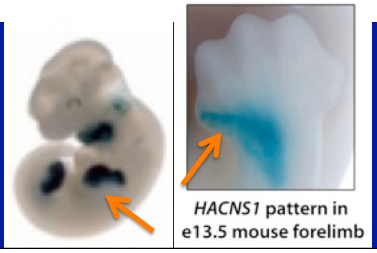
Lactose intolerance



29

Human accelerated enhancer: HACNS1





HACNS1 pattern in e13.5 mouse forelimb

Processes in accelerated change

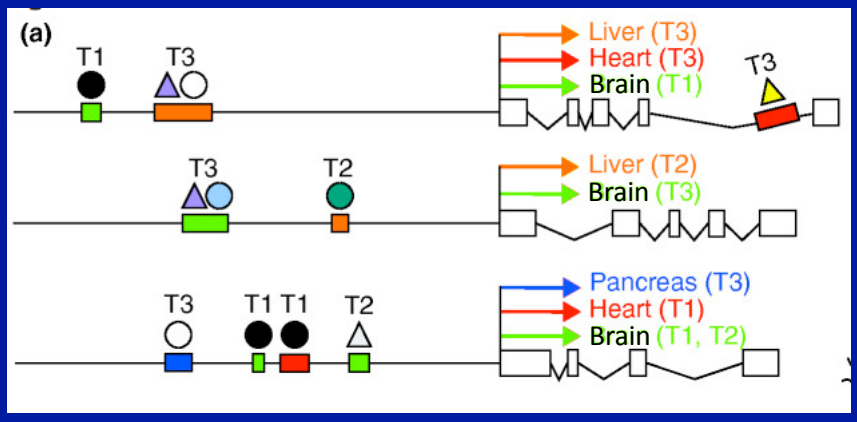
- positive selection
- biased gene conversion
- relaxation of negative constraint

30

II. Mapping functional elements

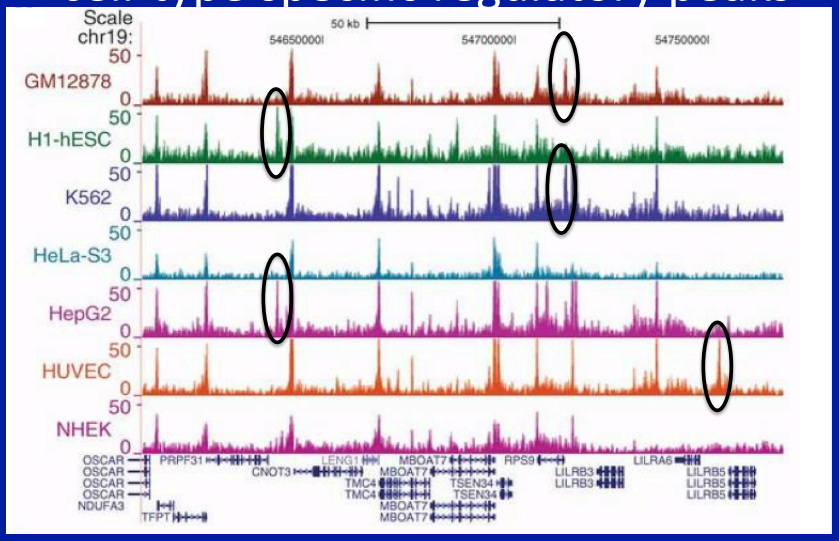
31

Tissue-specific regulatory differences….



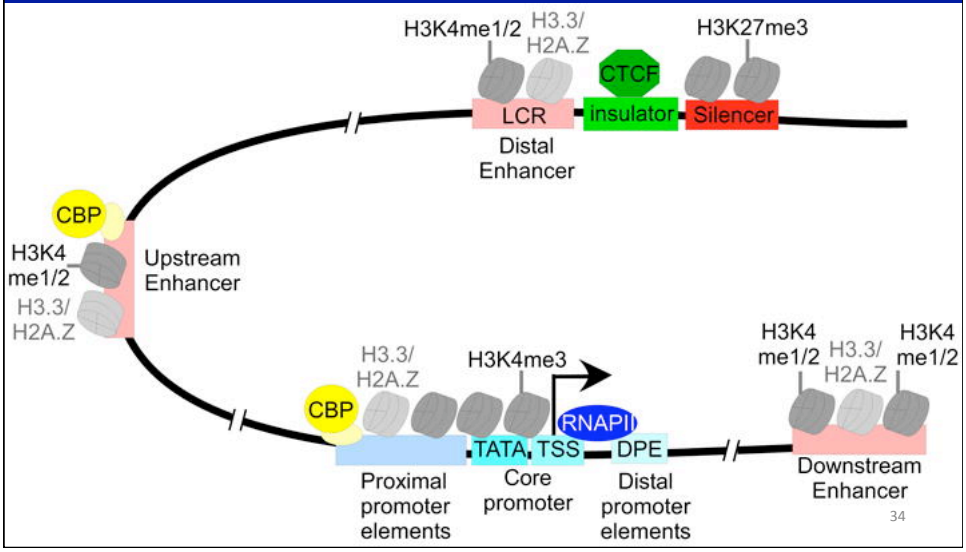
32

…are revealed in ChIP-seq data as cell-type specific regulatory peaks



33

Cis-regulation



34

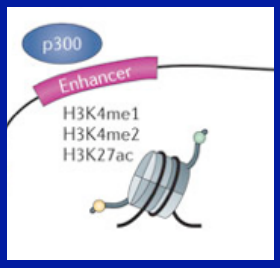
Enhancers

Sequence features

- Clusters of TFs
- Sequence conservation

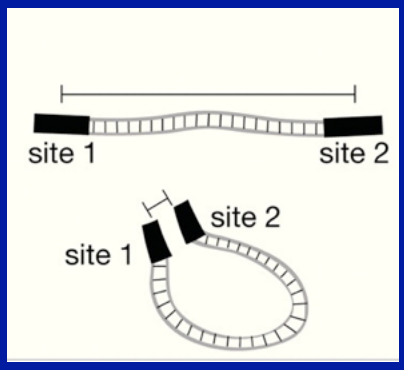
Epigenetic features

- DNAse hypersensitivity
- H3K4me1, me2, H3K27ac
- p300

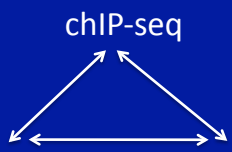


35

Long distance interactions



ChIA-pet
3C, 4C, 5C
HiC



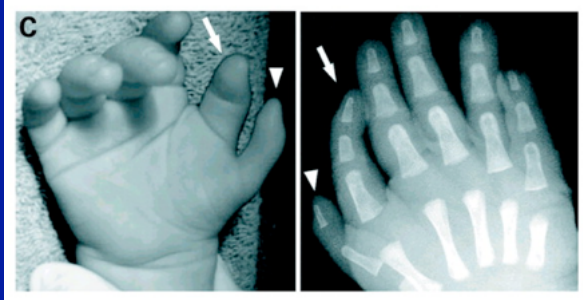
36

III. Interpretation of disease processes

37

SHH enhancer mutations

- Gain of function



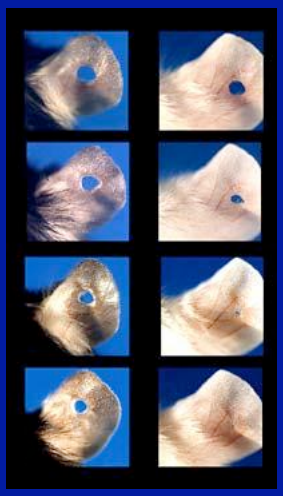
Hum. Mol. Genet. 2003 12 : 1725-1735

- Loss of function

Development 2005 132 : 4 797-803

38

Regenerating tissue



39

Common disease, common variant

enhancer mutations?

type II diabetes
colorectal cancer
breast cancer
pancreatic cancer
coronary artery disease

CELL-TYPE SPECIFIC DATA FROM ENCODE

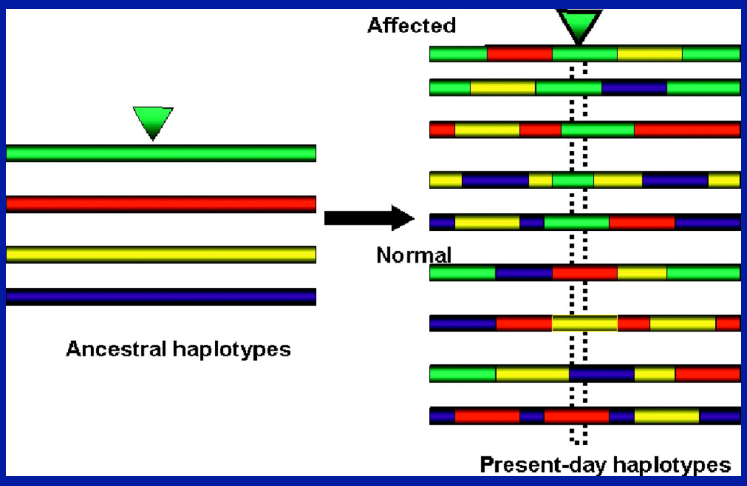
40

How do we know if a variant disrupts a functional element or is neutral?

- sequence conservation and phylogenetic footprints provide evidence
- histone modifications and DNAse hypersensitivity indicate function
- p300 binding and looping interactions show activity
- Look to ENCODE data for evidence

41

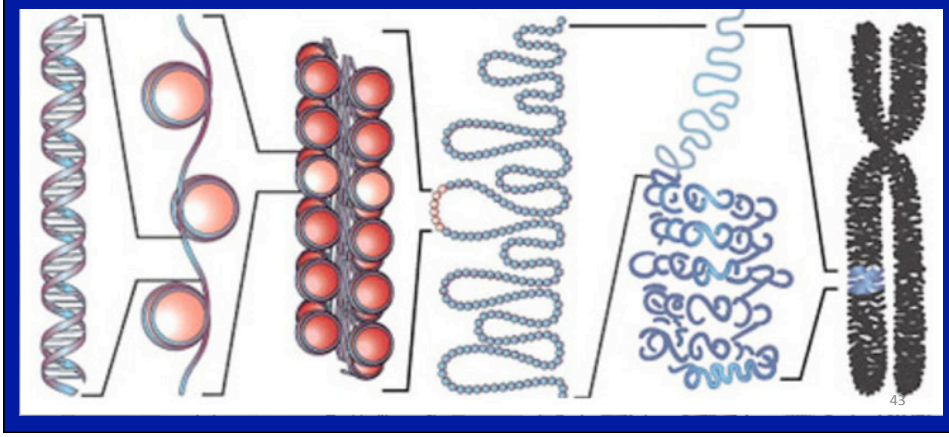
GWAS variant might not be the causal variant



Bioinformatics 2005. 21:4384-4393

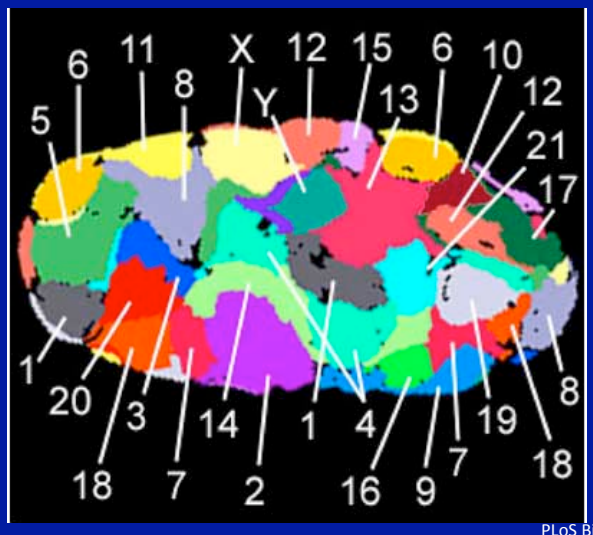
42

Regulatory connections through chromatin



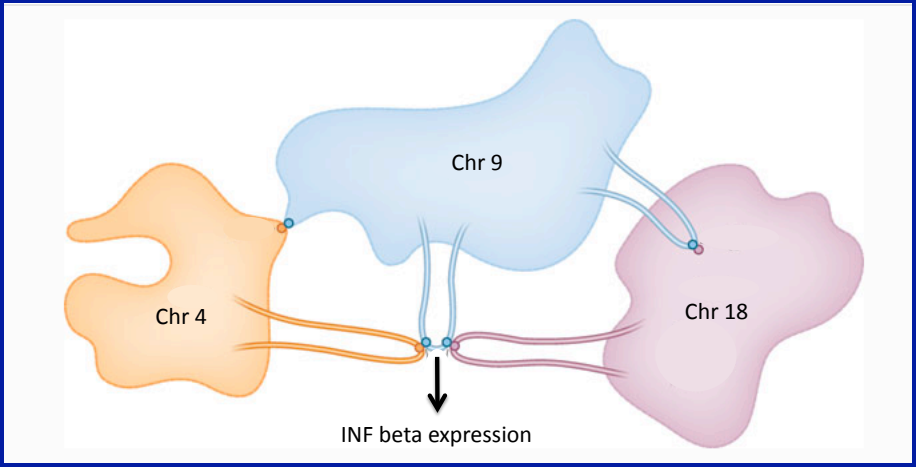
43

Chromosome territories



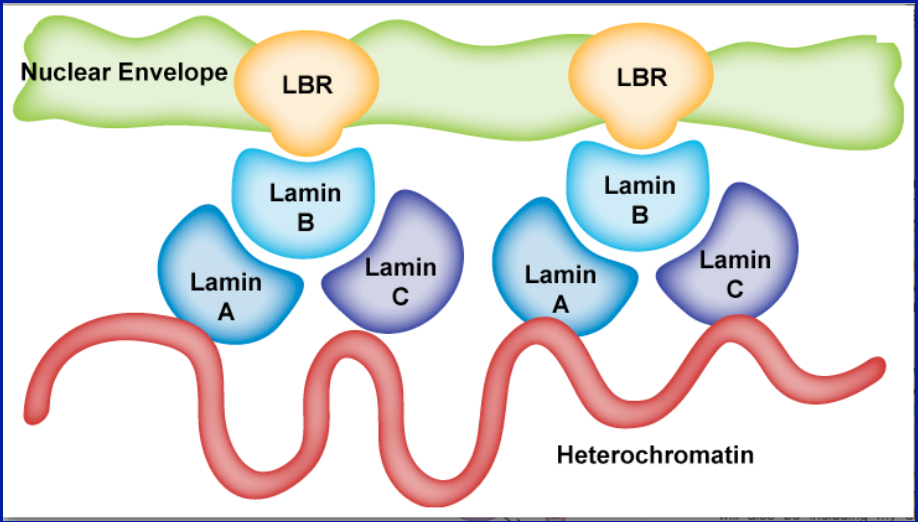
PLoS Biol. 2005 3:e157

Interchromosomal interactions



45

Lamina



poorhungrydrowsy.blogspot.com

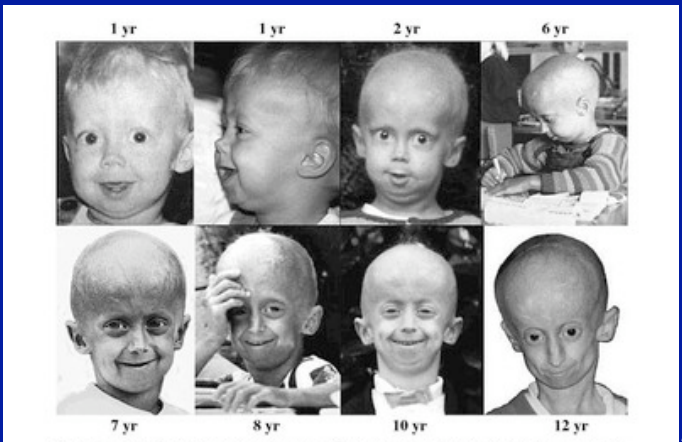
46

How do we confirm that laminar interactions are important?

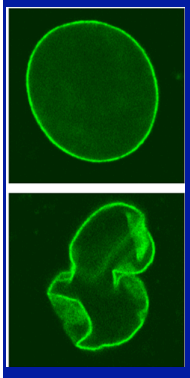
- *Find laminar mutations that cause disease*
- *Presence of sequence conservation in lamina-interacting domains*
- *Deletion of lamina structures*

47

Progeria



Normal nucleus



Affected nucleus

Am J. Med Genet. 23:2603-24

PLoS Biology Vol. 3/11/2005, e395

48

"Age is not a particularly interesting subject.

Anyone can get old.

All you have to do is live long enough."

(Groucho Marx)

49

Changes to Chromatin With Aging

General heterochromatinization
DNA repair decrease
Chromatin aberrations
Telomere shortening
Loss of histone ADP-ribosylation
Enrichment of tri-methylated histone H4 K20
Appearance of rDNA circles in yeast
Loss of 5-methylcytosine
Changes in H1 distribution

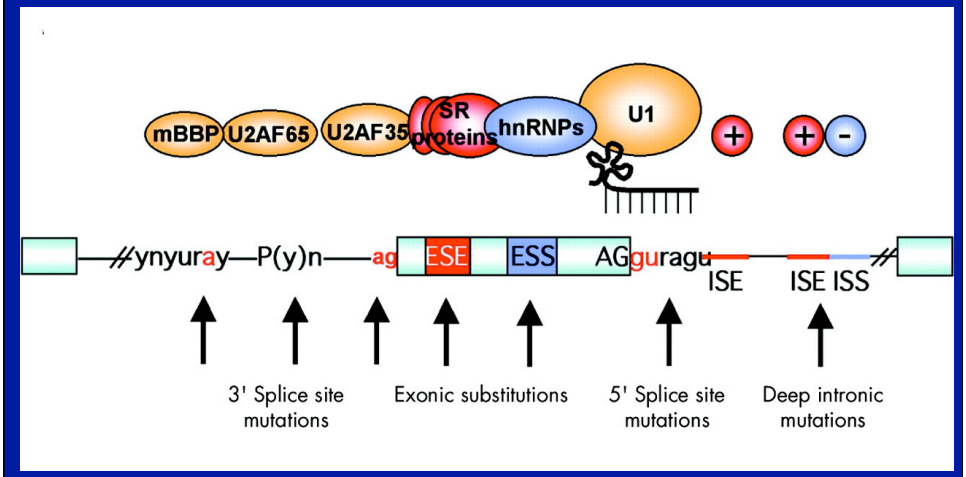
Aging

50

III. Interpretation of disease processes

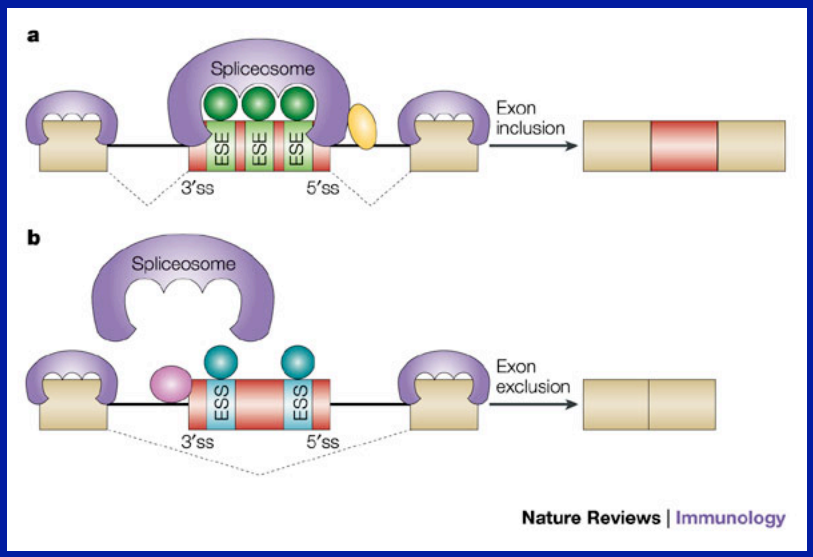
51

Splicing



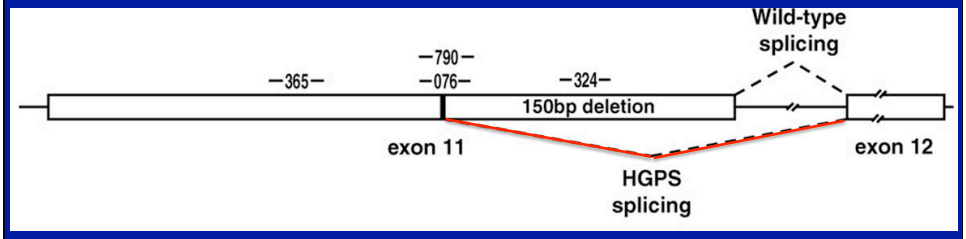
52

J Med Genet 2005;42:737-748



53

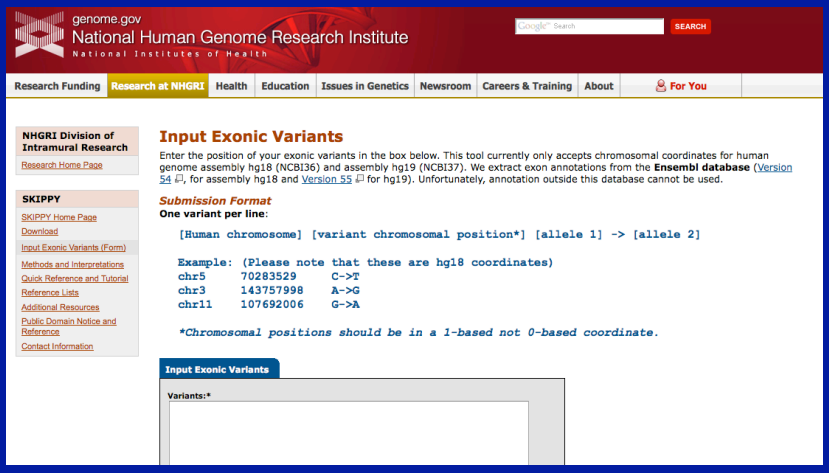
Lamin A splicing mutation



Challenges:
mapping elements
consequences of mutations
extensive noncoding space

54

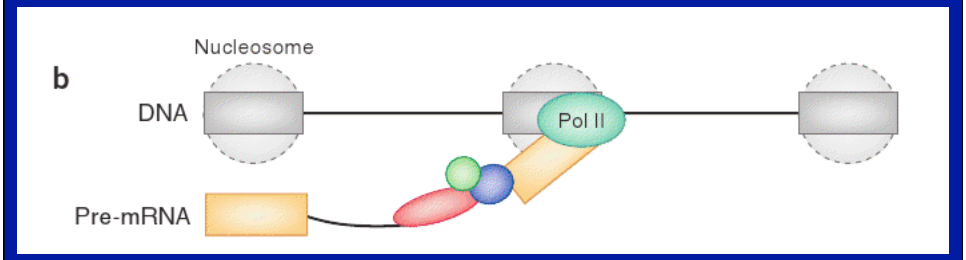
Predicting variants that cause exon skipping



* Woolfe A, Mullikin J, **Elnitski L** (2010)

55

Speed Bumps



56

nature structural & molecular biology 2009 16: 9

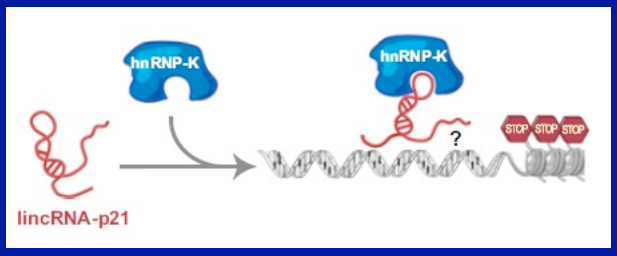
What's the prevailing connection between functional elements in the cell?

- Aging cells sabotage each other
- Noncoding RNA orchestrate many events
- Conserved elements underlie all important features

57

Noncoding RNA

DNA ⟶ RNA ⟶ ~~Protein~~

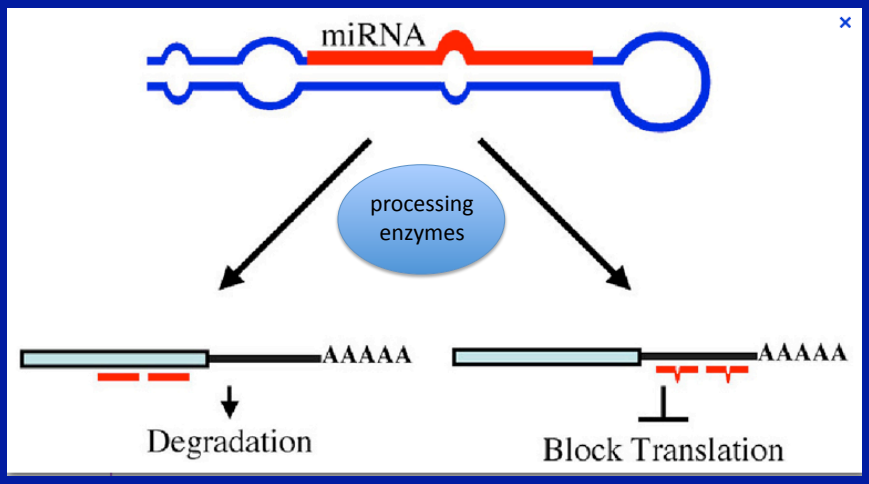


I. guides
II. signals
III. decoys
IV. scaffolds

Oncogenic & tumor-suppressor lncRNAs

58

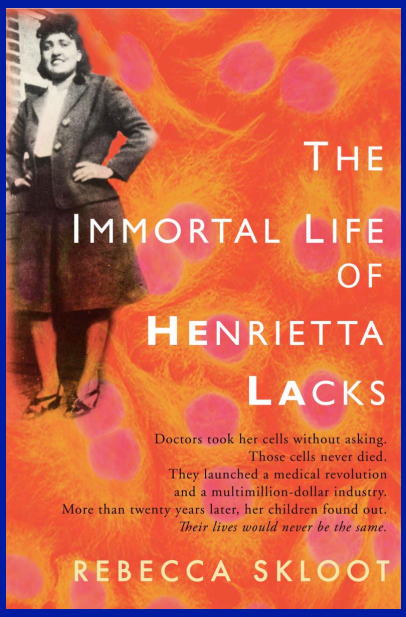
MicroRNA



59

IV. How to read the epigenetic code

60



"Good science

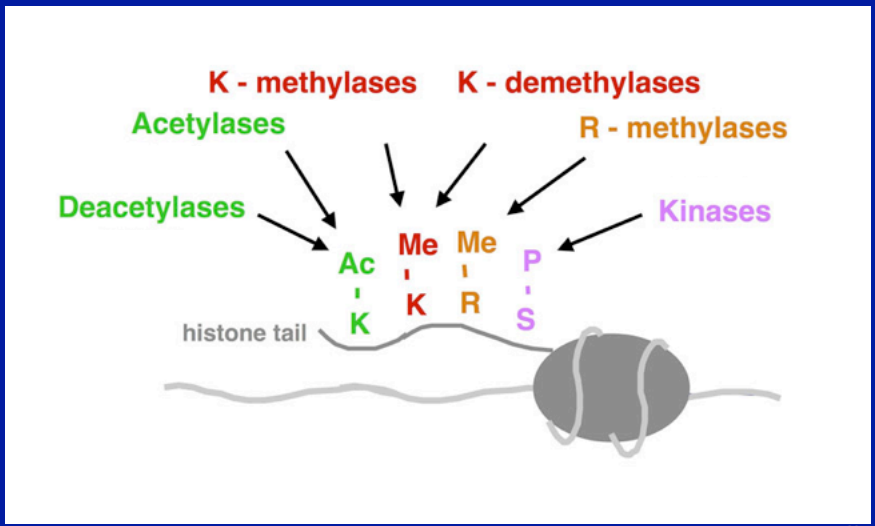
is all about following the data as it shows up

letting yourself be proven wrong,

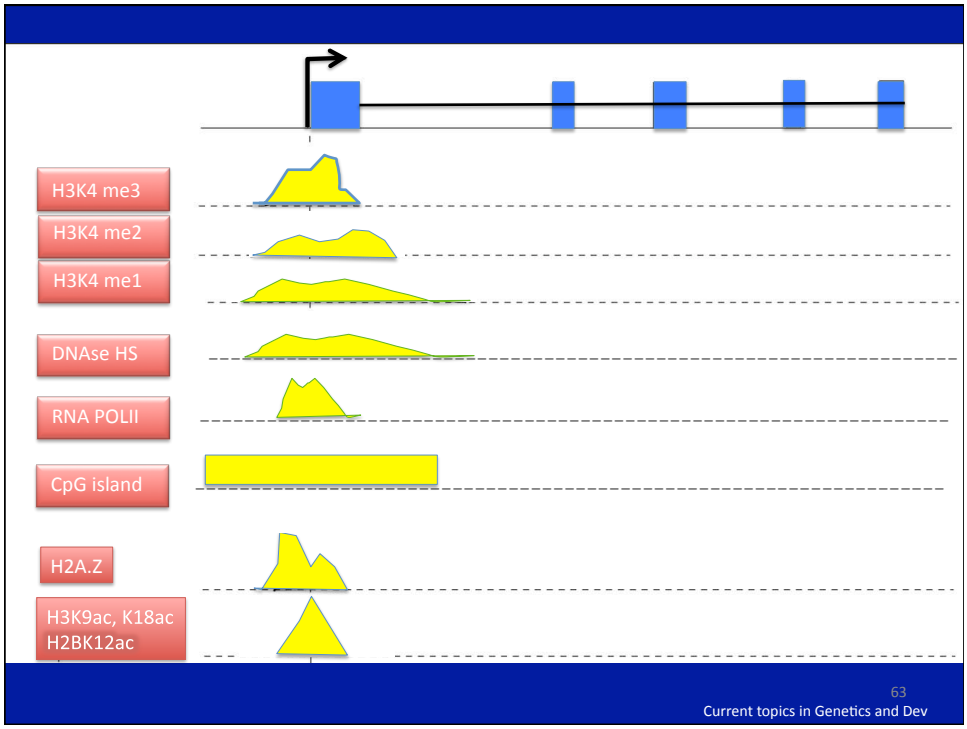
and letting everything change while you're working on it "

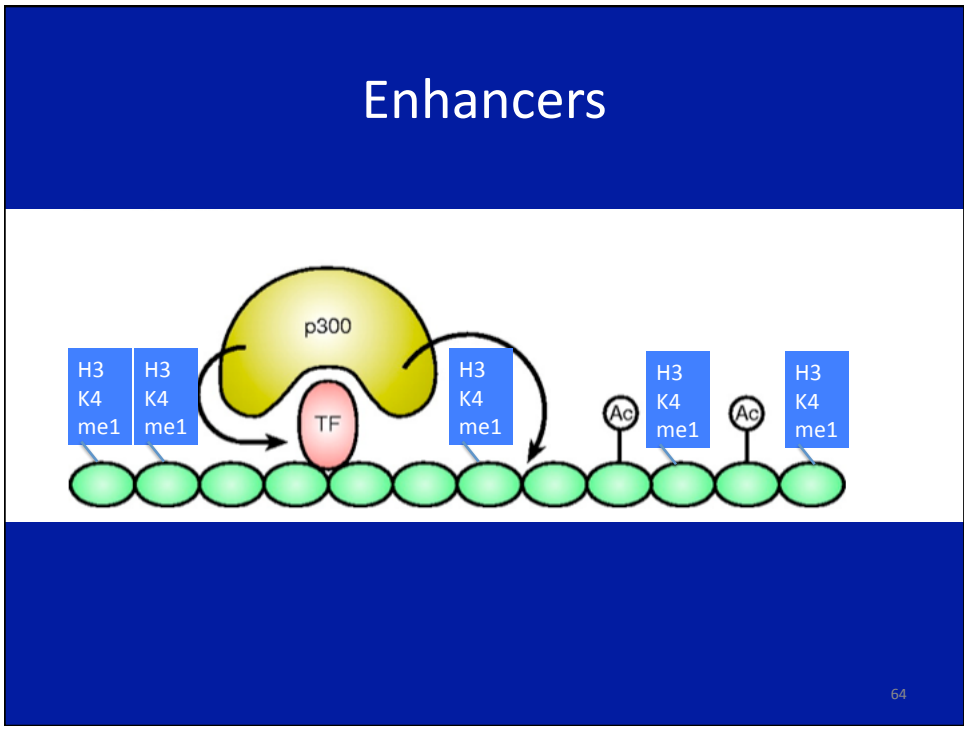
61

Diversity of histone modifications

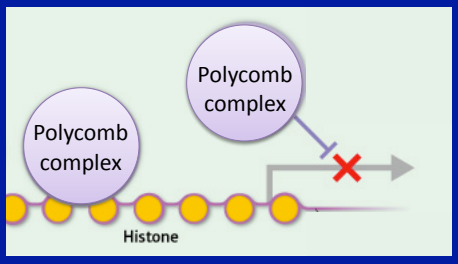


62


H3K4 me3
H3K4 me2
H3K4 me1
DNAse HS
RNA POLII
CpG island
H2A.Z
H3K9ac, K18ac
H2BK12ac
63
Current topics in Genetics and Dev


Enhancers
p300
TF
H3 K4 me1
H3 K4 me1
H3 K4 me1
Ac
H3 K4 me1
Ac
H3 K4 me1
64

Repressed promoters



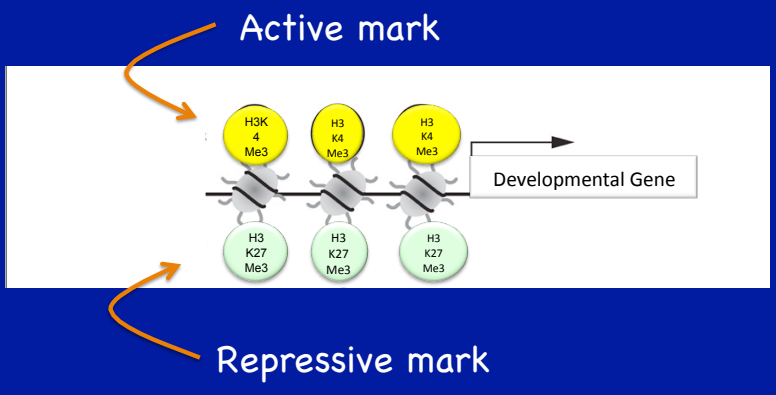
H3K27me3
H3K9me3
DNA methylation

Annu. Rev Genomics Hum. Genet. 8:299-325

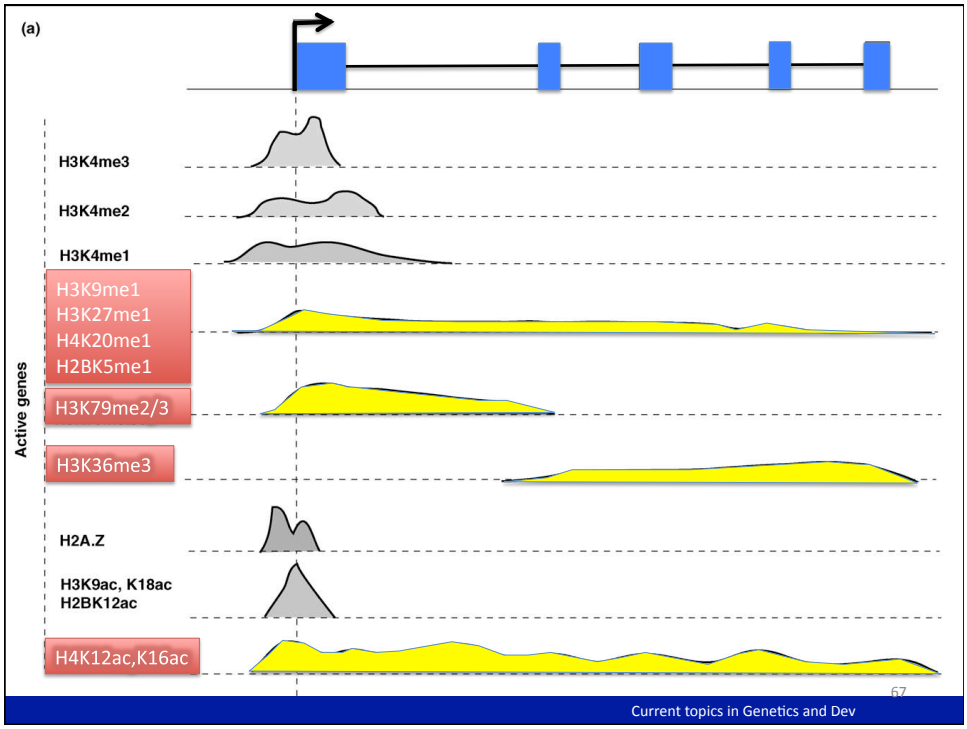
www.rikenresearch.riken.jp/eng/frontline/5514

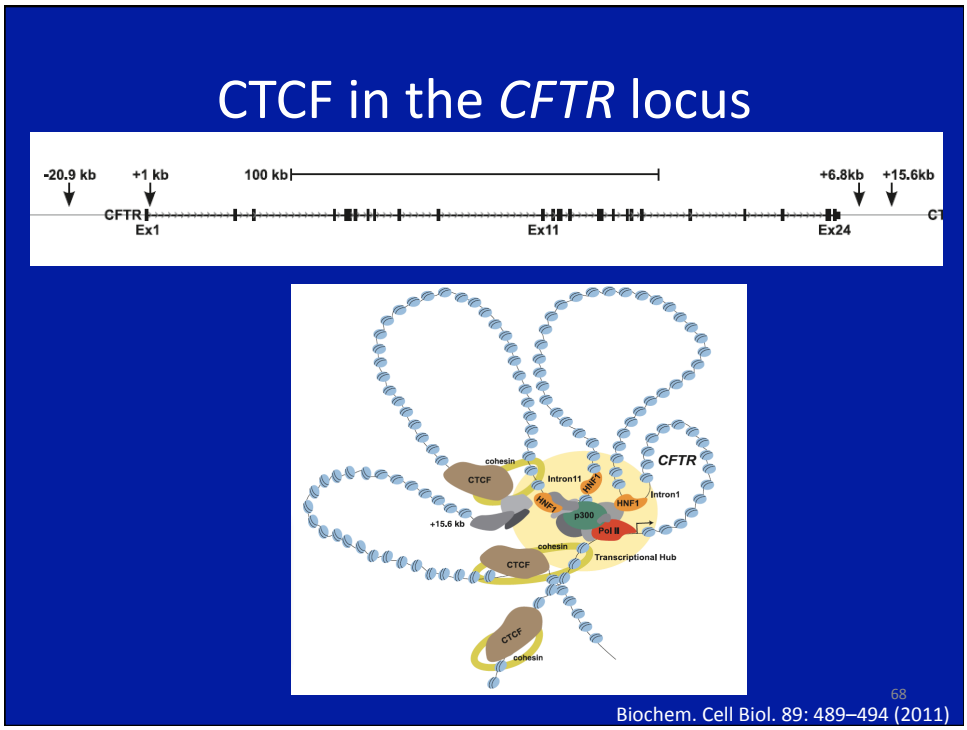
65

Bivalent marks in ES cells



Kondo Yonsei Med J. 2009 Aug;50(4):455-463.


(a)
H3K4me3
H3K4me2
H3K4me1
H3K9me1
H3K27me1
H4K20me1
H2BK5me1
H3K79me2/3
H3K36me3
H2A.Z
H3K9ac, K18ac
H2BK12ac
H4K12ac,K16ac
Active genes
67
Current topics in Genetics and Dev

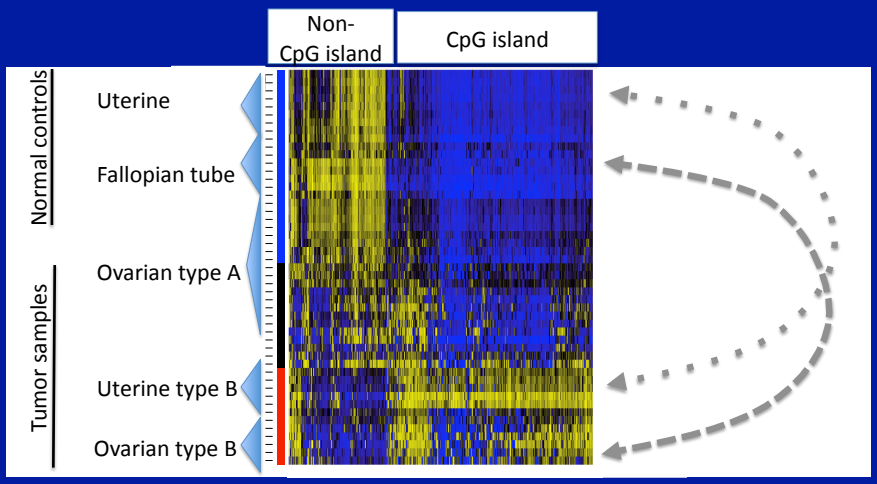

CTCF in the CFTR locus
-20.9 kb
+1 kb
100 kb
+6.8kb
+15.6kb
CFTR
Ex1
Ex11
Ex24
cohesin
CTCF
Intron11
HNF1
CFTR
Intron1
p300
Pol II
+15.6 kb
Transcriptional Hub
68
Biochem. Cell Biol. 89: 489–494 (2011)

DNA methylation

- specific for a tissue type
- stably alters gene expression patterns
- suppresses the expression of viral genes and prevents genomic rearrangements
- plays a crucial role in the development many types of cancer

69

Ovarian tumor methylation



Kolbe et al. Elnitski, in press

70

Contributions of the human genome project

I. Understanding evolutionary diversity

II. Genome function and regulatory elements

III. Variants that disrupt function and explain

71

The end

72