


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***

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research



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

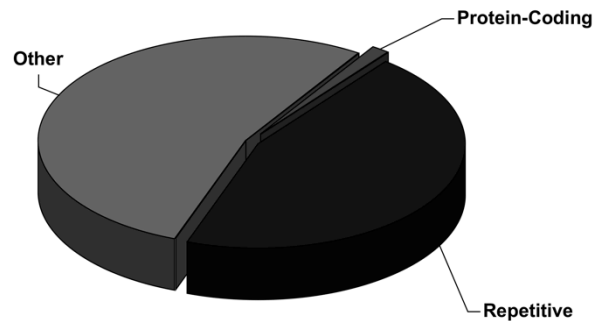
Bill Clinton



3

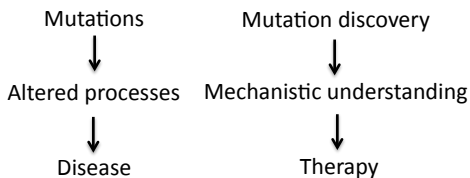
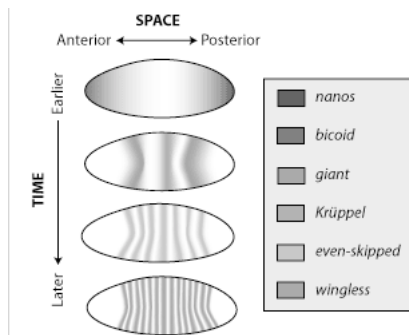
## Characterizing the human genome

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



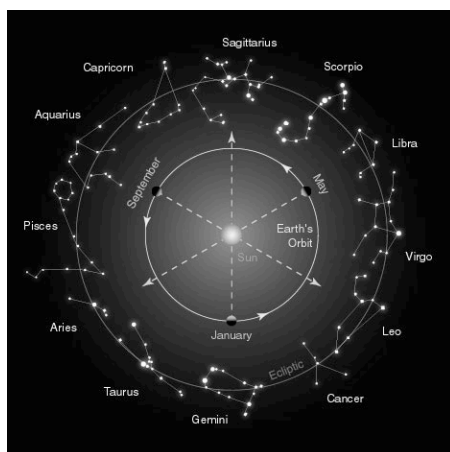
4

## Regulatory elements

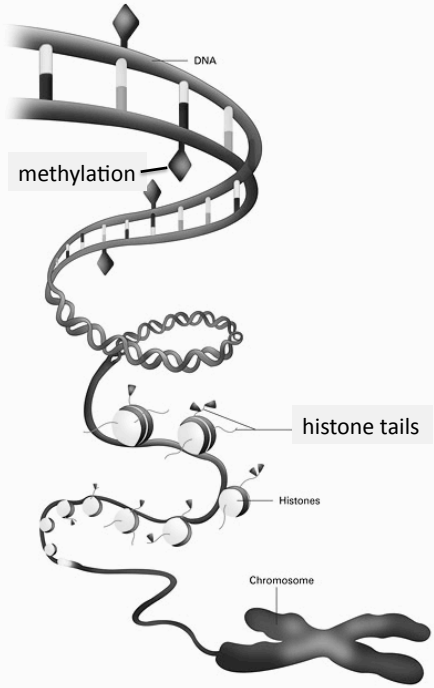


5

## Regulatory code



6



The diagram illustrates the hierarchical packaging of DNA. At the top, a DNA double helix is shown with methyl groups (represented by black diamonds) attached to the sugar-phosphate backbone, labeled 'methylation'. Below this, the DNA is wrapped around histone cores, with the protruding parts labeled 'histone tails'. The next level shows the DNA further condensed into a 'Chromosome' structure.


## Epigenetic patterns

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

7

National Institute of General Medical Sciences

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



The Science magazine cover (Vol. 296, No. 5568, 26 April 2002) features a woman's head where puzzle pieces are placed on her face and neck. The puzzle pieces are labeled with environmental and genetic factors: 'diet', 'genes', 'viruses', 'smoking', and 'pollution'. The main headline reads 'The Puzzle of Complex Diseases' and the American Association for the Advancement of Science logo is at the bottom.

8

Gut 2011. 60:1739-1753

## 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

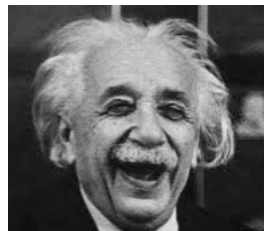


11



17,000

<



23,000 genes

<

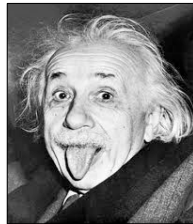


30,000



1 billion bp

<



3.4 billion bp

<



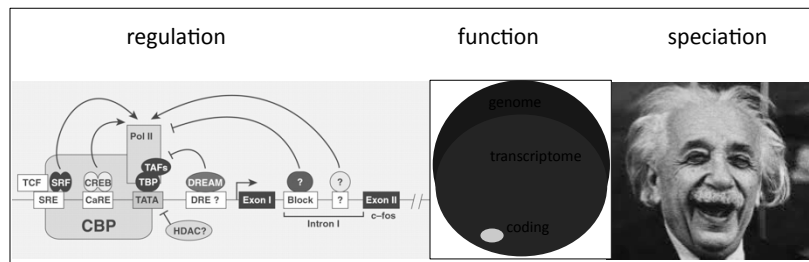
132.8 billion bp

12

## Genomic diversity

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

13

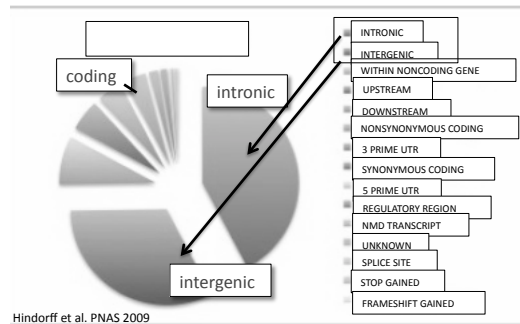


### Mutational consequences are varied

Original Sentence	<b>TIME TO DREAM</b>
Single Letter Change	<b>T<sup>I</sup>ME TO DREAM</b> → <b>TAME TO DREAM</b>
Reverse Order	<b>I<sup>M</sup>TE TO DREAM</b> → <b>EMIT TO DREAM</b>
Deletion	<b>T<sup>IM</sup>E TO DREAM</b> → <b>TETO DR EAM</b>
Insertion	<b>TI<sup>I</sup>ME TO DREAM</b> → <b>TIIM ET ODREAM</b>

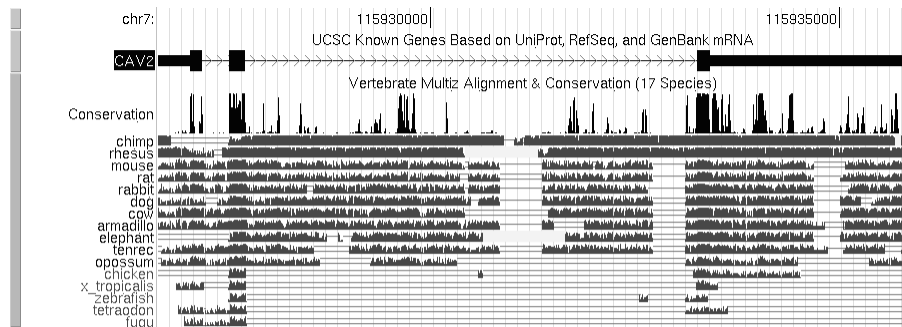
14

## Genome wide association data



15

## Comparative genomics



Mutations in functional DNA are less likely to be tolerated

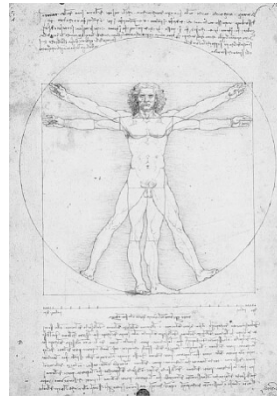
fastest divergence : immune, reproduction and olfaction  
 slowest divergence : developmental genes

16



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

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



17

## Evolutionary Distance



Human



Chimpanzee



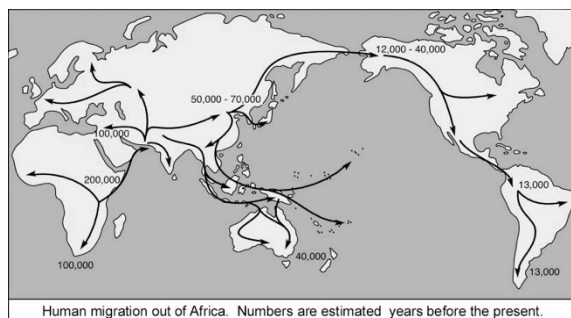
Horse



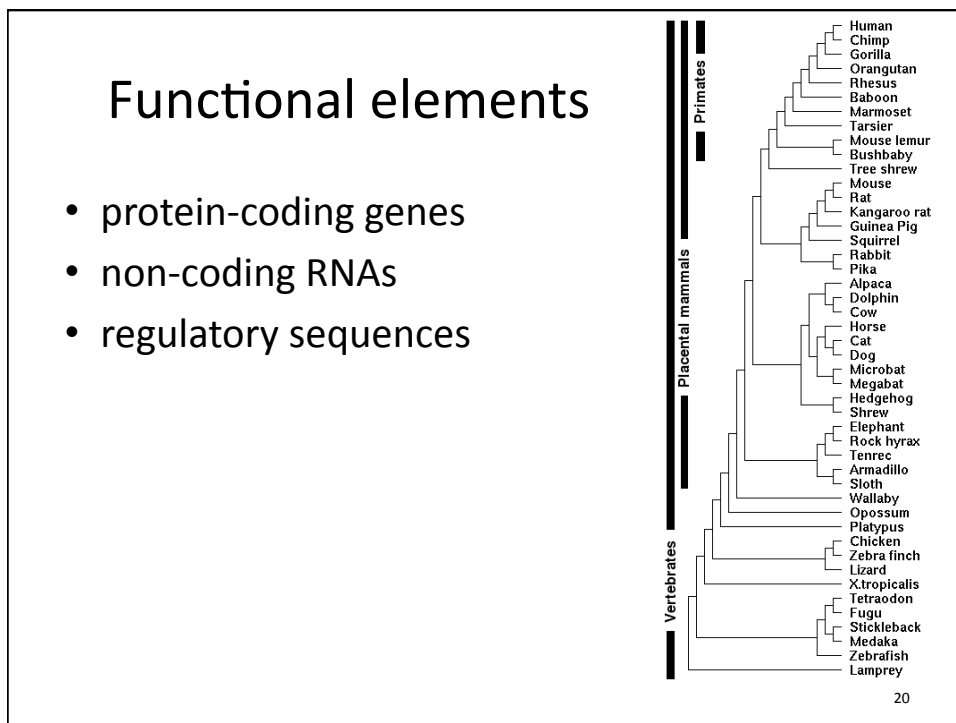
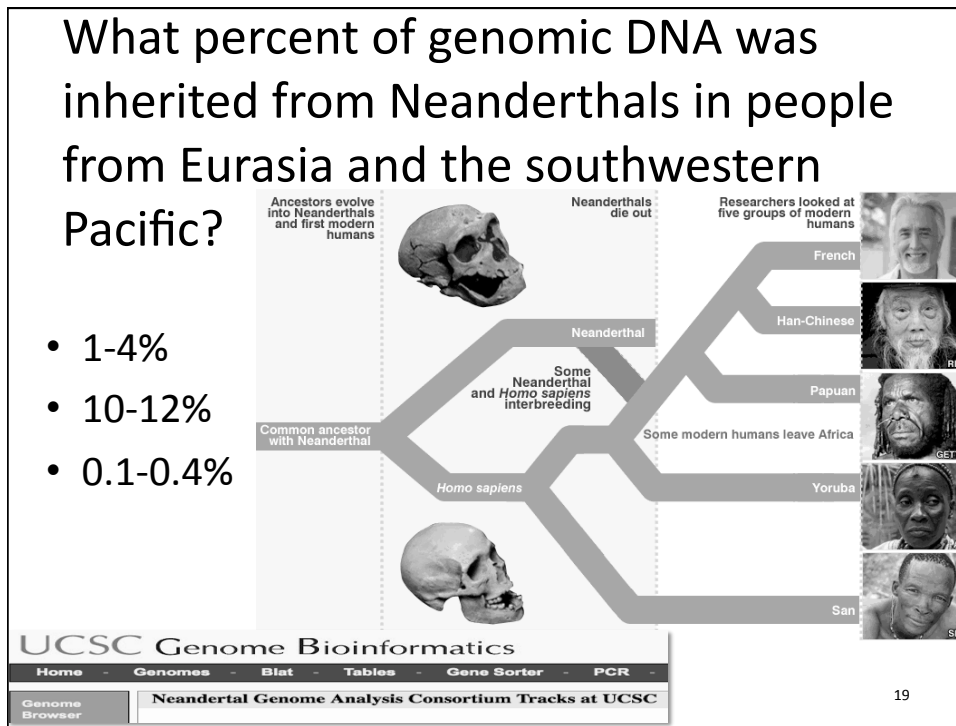
Rat



Platypus

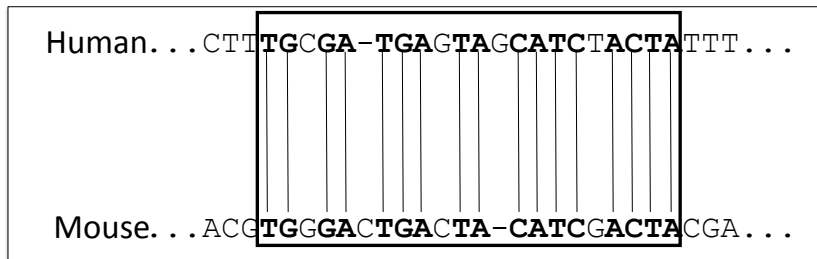


bonner@indiana.edu

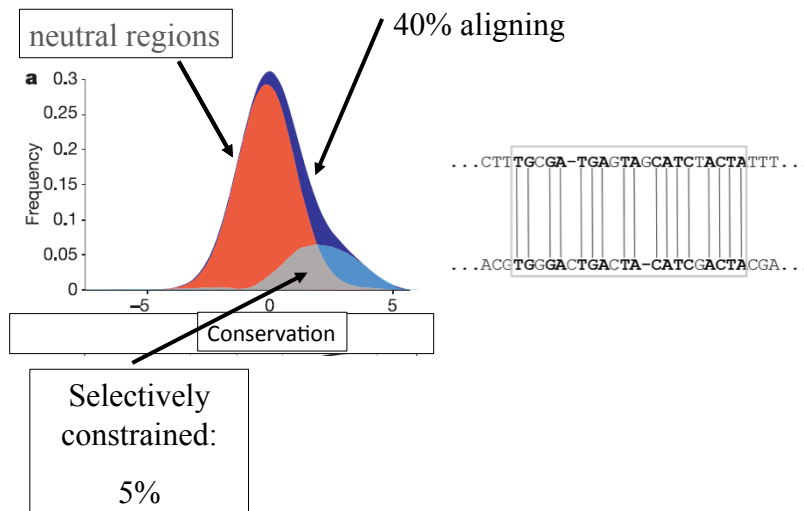


## Comparative genomics of related species highlights functional regions

purifying selection



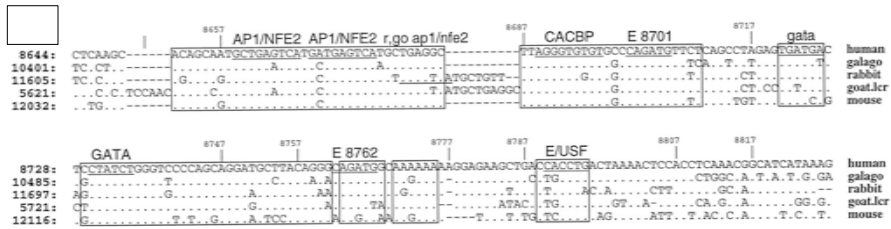
## Human-mouse comparisons



[Mouse consortium, *Nature* 2002]

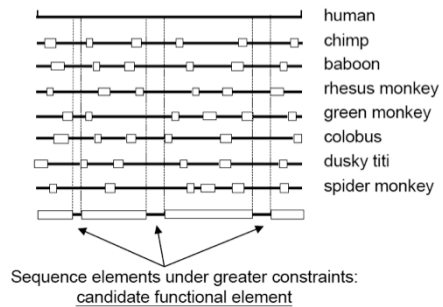
22

## Phylogenetic footprints

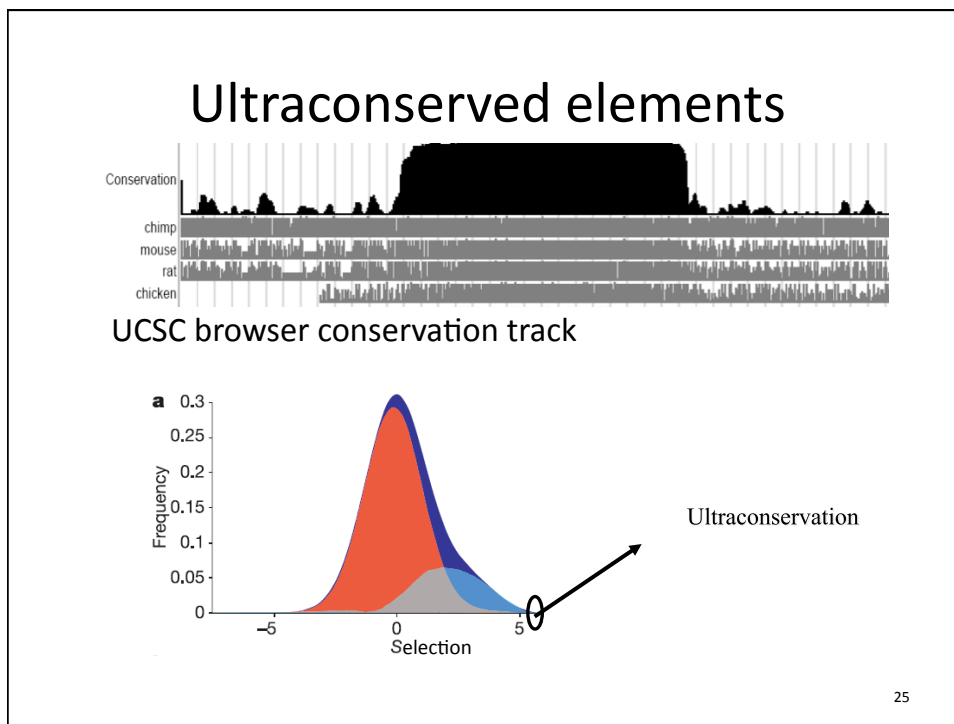


23

## Phylogenetic shadowing



24



## PHYLOGENETIC ANALYSIS WITH SPACE/TIME MODELS

generative model-based approach

**Genomic Evolutionary Rate Profiling: GERP**

bottom-up approach

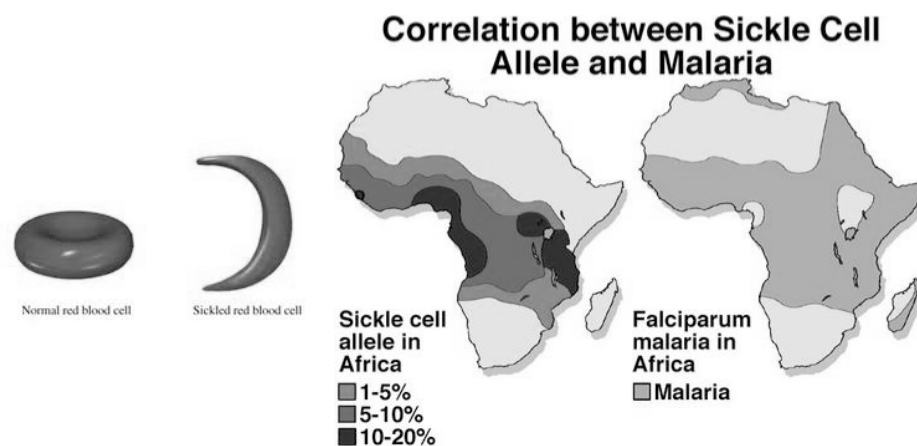
26

## 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

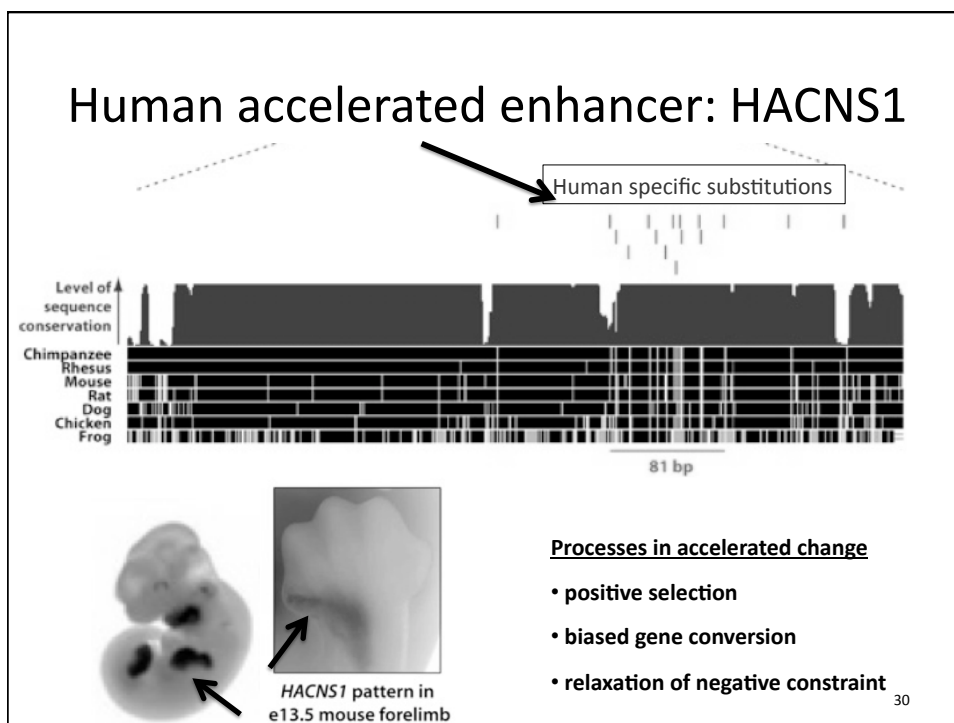
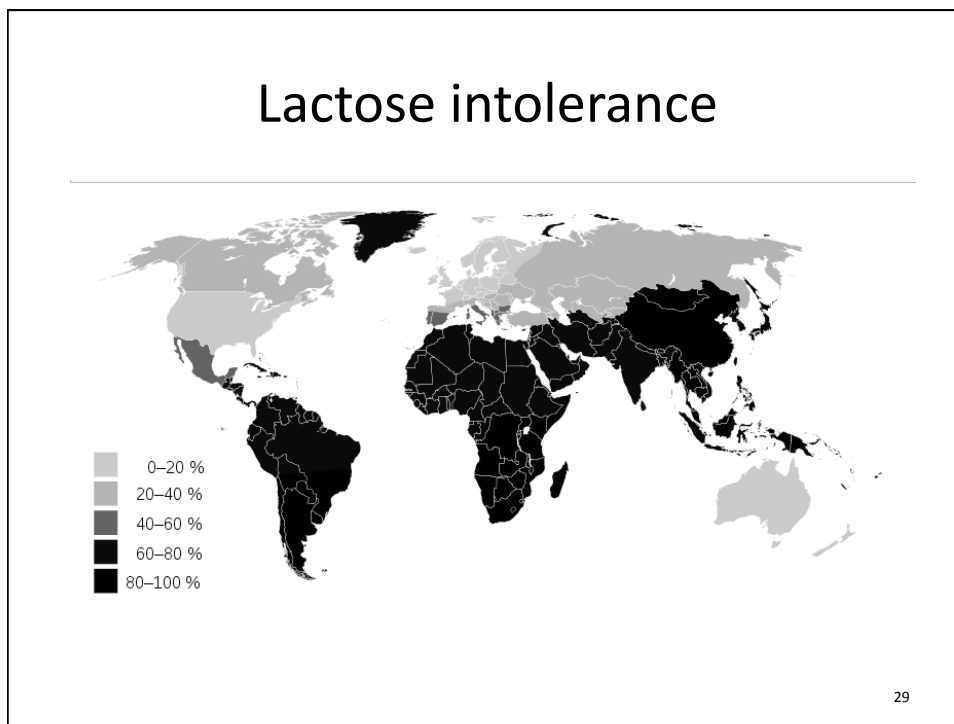
27

## Positive selection



The McGraw-Hill Companies

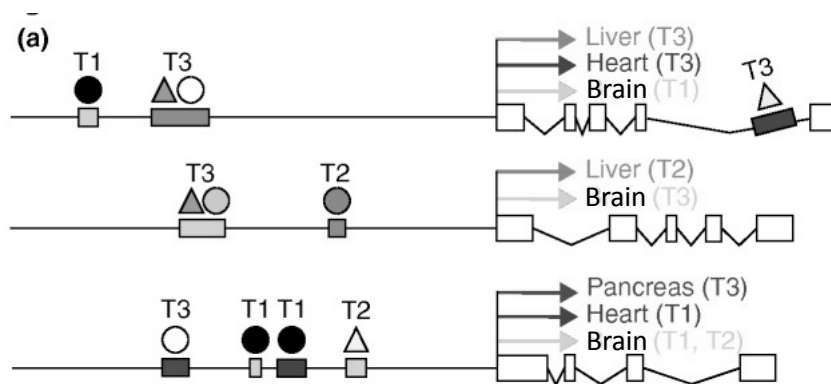
28



## II. Mapping functional elements

31

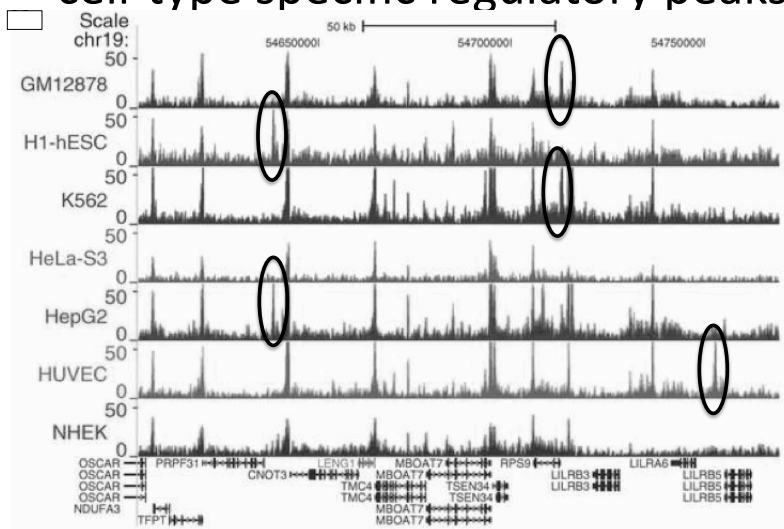
### Tissue-specific regulatory differences....



32

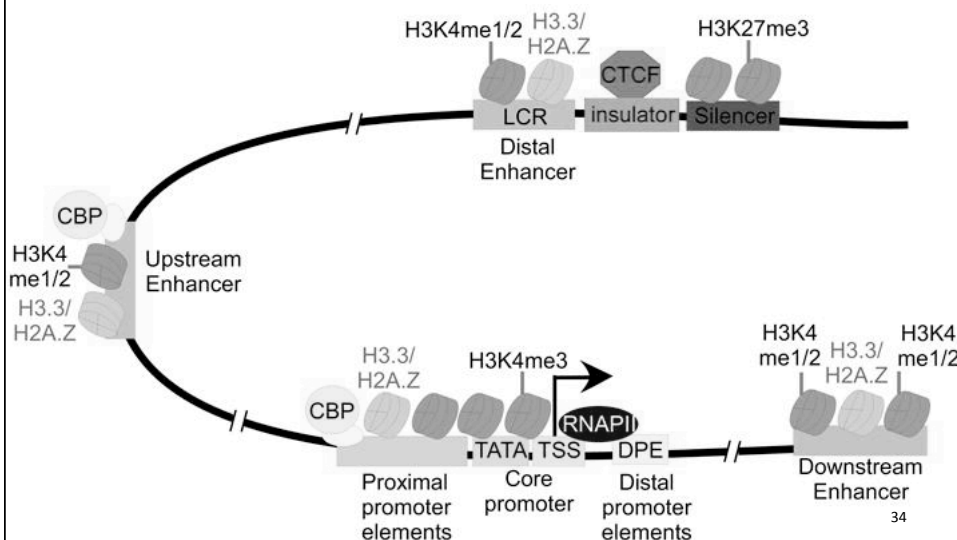


...are revealed in ChIP-seq data as cell-type specific regulatory peaks



33

### Cis-regulation



34

## Enhancers

### Sequence features

- Clusters of TFs
- Sequence conservation

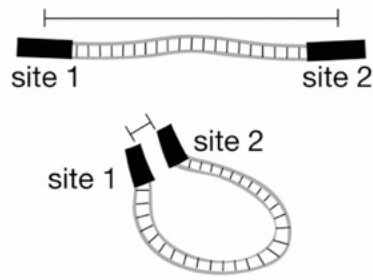
### Epigenetic features

- DNase hypersensitivity
- H3K4me1, me2, H3K27ac
- p300

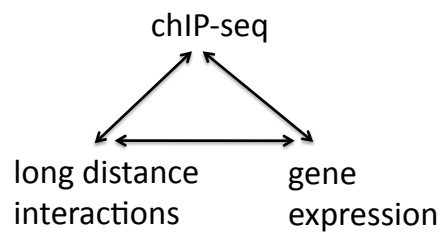


35

## Long distance interactions



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



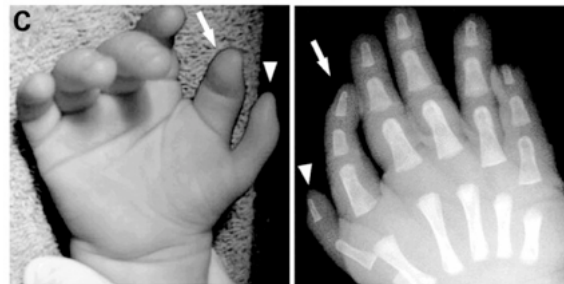
36

### III. Interpretation of disease processes

37

#### ***SHH* enhancer mutations**

- Gain of function



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

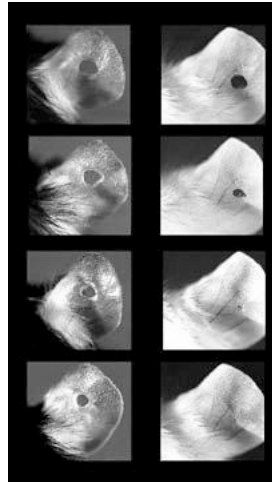
- Loss of function



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

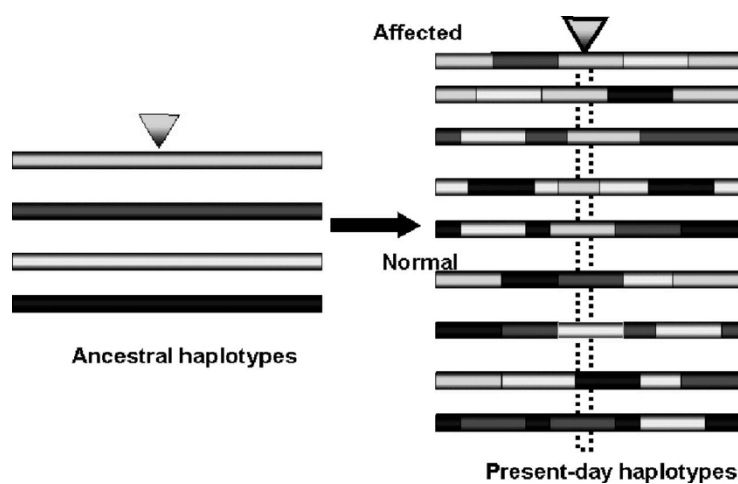
40

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

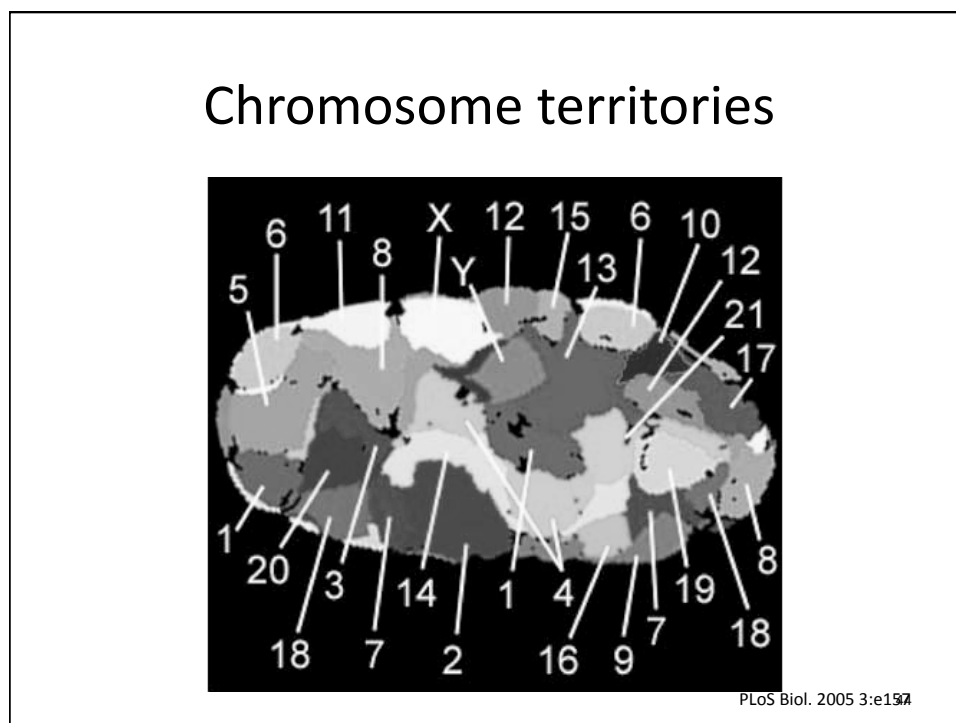
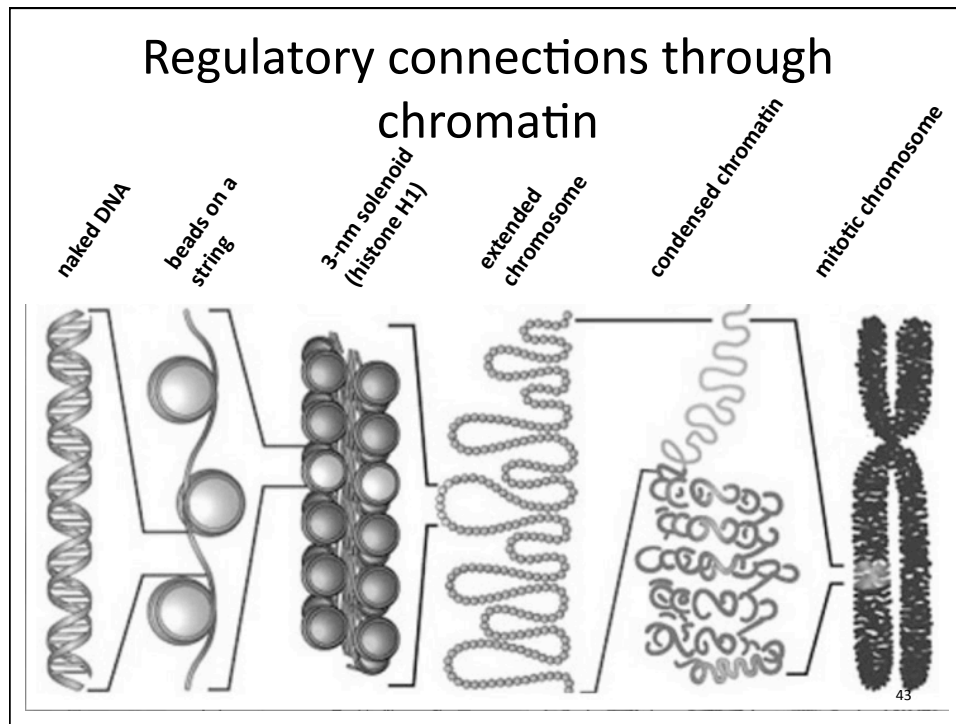
- sequence conservation and phylogenetic footprints provide evidence
- histone modifications and DNase hypersensitivity indicate function
- p300 binding and looping interactions show activity
- Look to ENCODE data for evidence

41

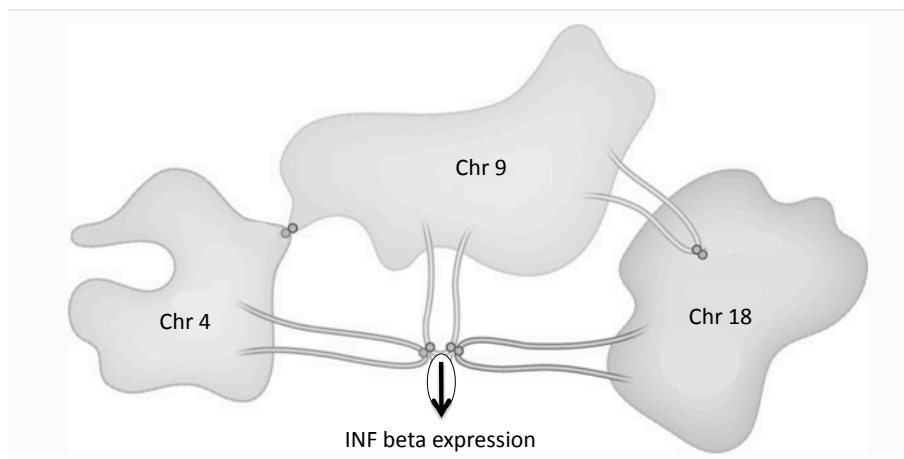
## GWAS variant might not be the causal variant



Bioinformatics 2005. 21:4384-4393<sup>42</sup>

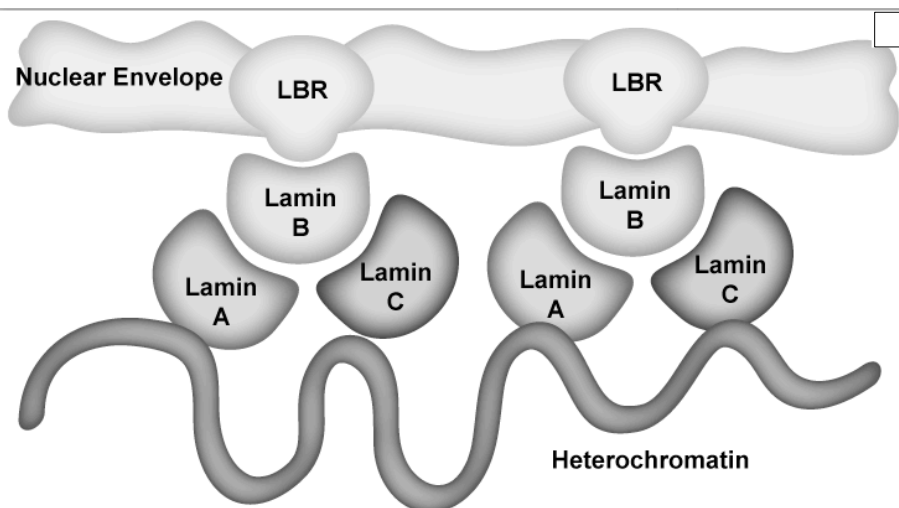


## Interchromosomal interactions



45

## Lamina



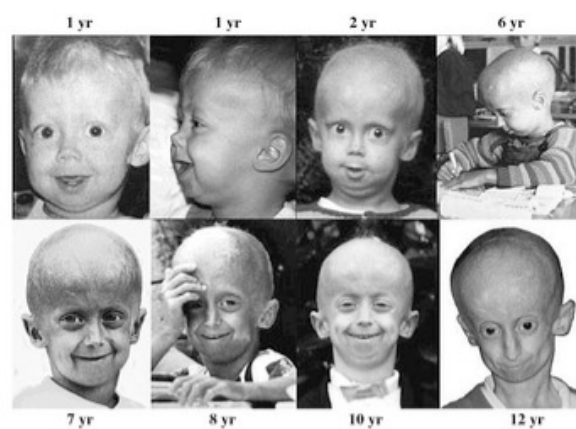
poorhungrydrowsy.blogspot.com <sup>46</sup>

## How do we confirm that laminar interactions are important?

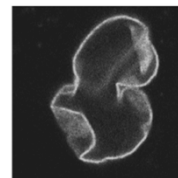
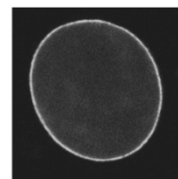
- Find laminar mutations that cause disease
- Presence of sequence conservation in lamina-interacting domains
- Deletion of lamina structures

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

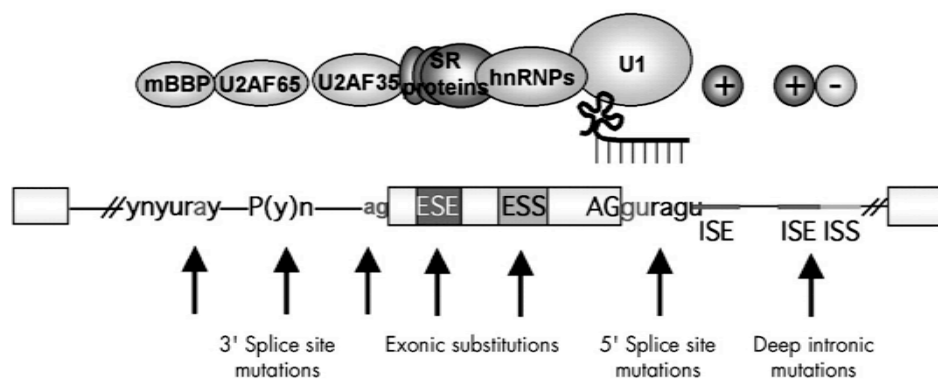
Aging

50

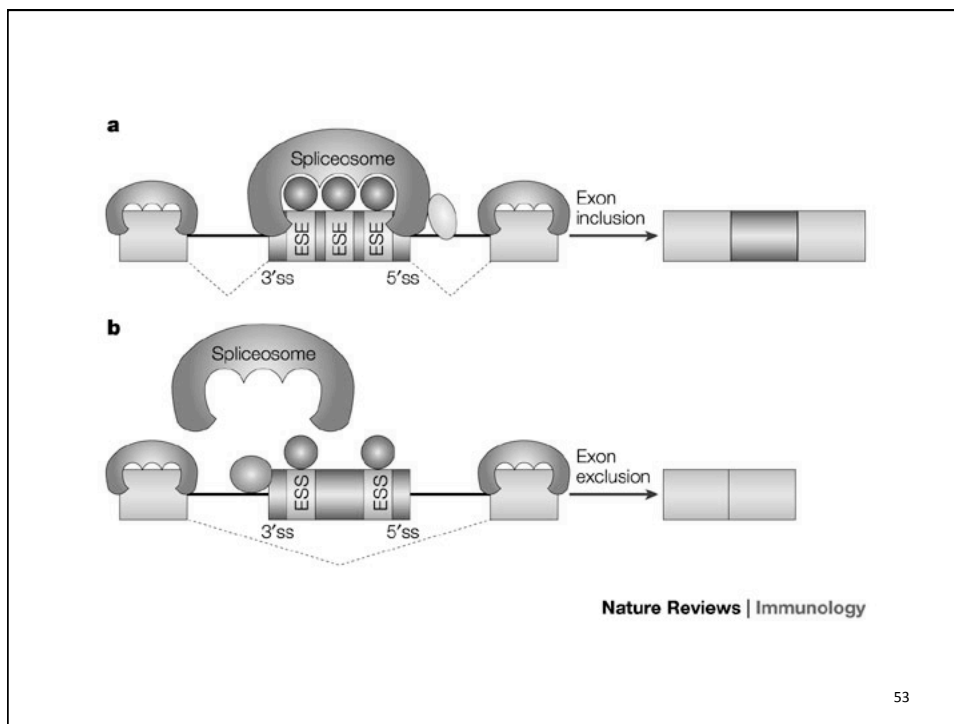
### III. Interpretation of disease processes

51

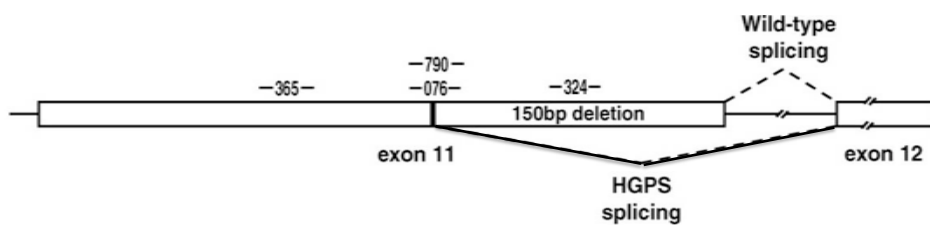
## Splicing



52  
*J Med Genet* 2005;42:737-748



## Lamin A splicing mutation



Challenges:  
mapping elements  
consequences of mutations  
extensive noncoding space

## Predicting variants that cause exon skipping

The screenshot shows the NHGRI website header with the logo and navigation menu. The main content area is titled "Input Exonic Variants" and includes a search bar, a description of the tool, a submission format, an example, and a form for inputting variants.

**genome.gov**  
National Human Genome Research Institute  
National Institutes of Health

SEARCH

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | Newsroom | Careers & Training | About | For You

**NHGRI Division of Intramural Research**  
Research Home Page

**SKIPPY**  
SKIPPY Home Page  
Download  
Input Exonic Variants (Form)  
Methods and Interpretations  
Quick Reference and Tutorial  
Reference Lists  
Additional Resources  
Public Domain Notice and Reference  
Contact Information

### Input Exonic Variants

Enter the position of your exonic variants in the box below. This tool currently only accepts chromosomal coordinates for human genome assembly hg18 (NCBI36) and assembly hg19 (NCBI37). We extract exon annotations from the **Ensembl database** (Version 54 <#>, for assembly hg18 and Version 55 <#> for hg19). Unfortunately, annotation outside this database cannot be used.

**Submission Format**  
One variant per line:  
[Human chromosome] [variant chromosomal position\*] [allele 1] -> [allele 2]

**Example: (Please note that these are hg18 coordinates)**  
chr5 70283529 C->T  
chr3 143757998 A->G  
chr11 107692006 G->A

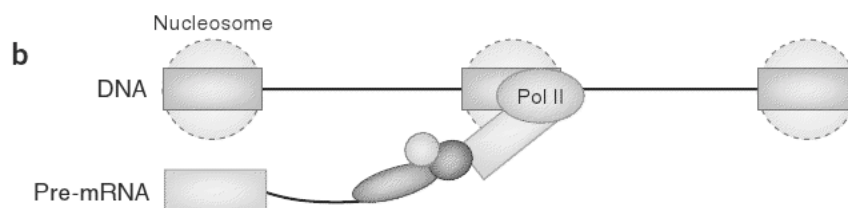
\*Chromosomal positions should be in a 1-based not 0-based coordinate.

**Input Exonic Variants**  
Variants:\*

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

55

## Speed Bumps



56  
nature structural & molecular biology 2009 16: 9

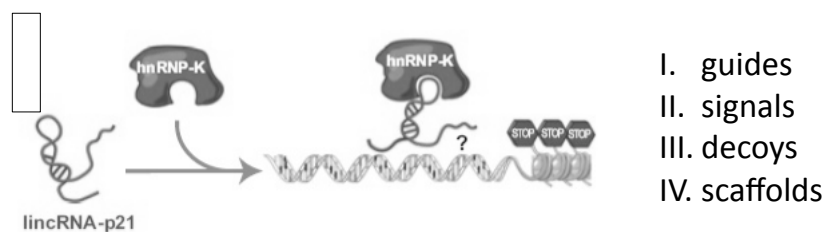
## What's the prevailing connection between functional elements in the cell?

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

57

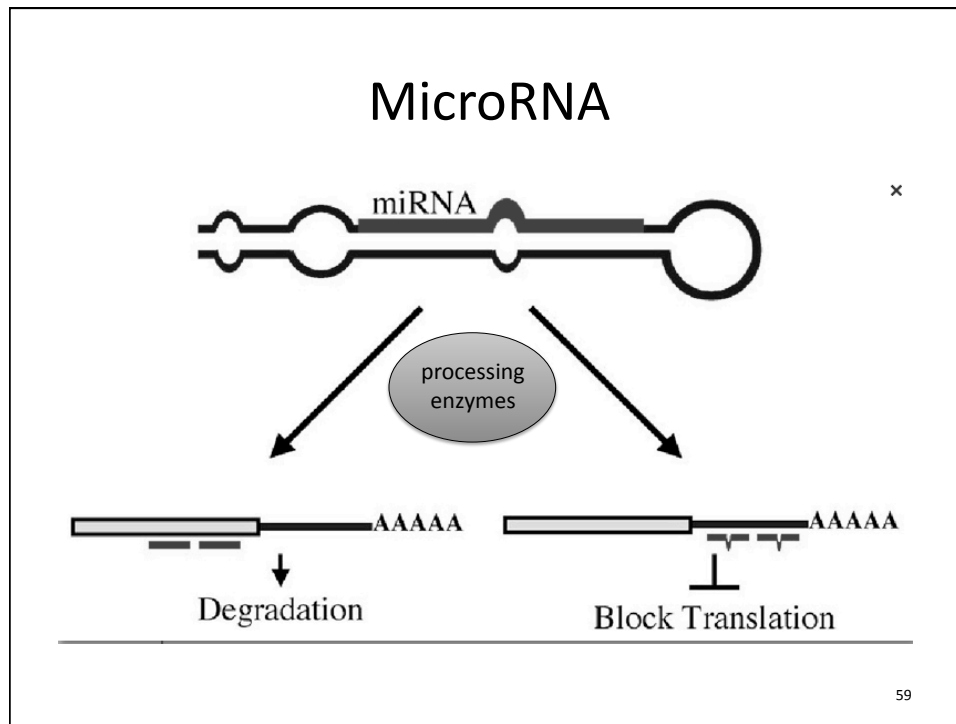
## Noncoding RNA

DNA  $\longrightarrow$  RNA  $\longrightarrow$  Protein ~~X~~

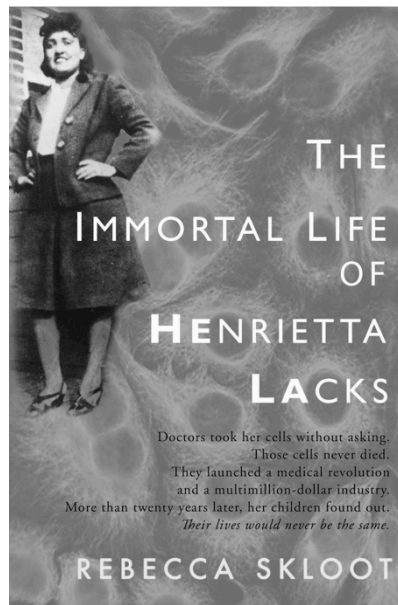


Oncogenic & tumor-suppressor lncRNAs

58



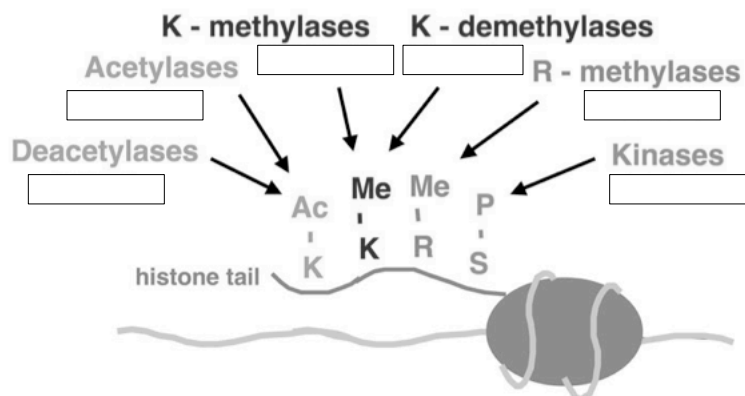
## *IV. How to read the epigenetic code*



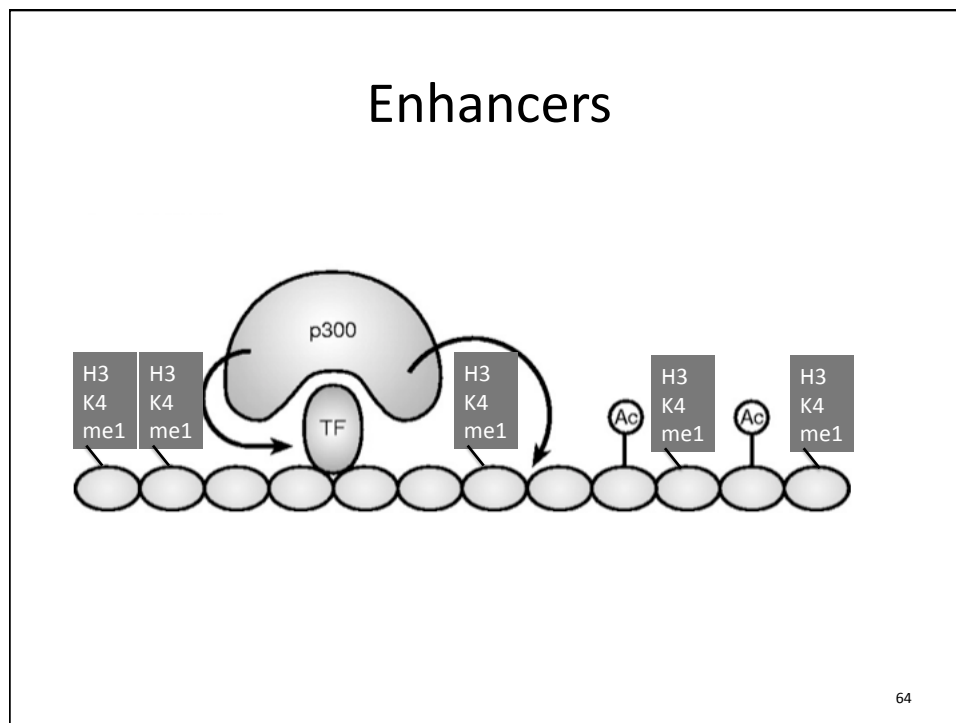
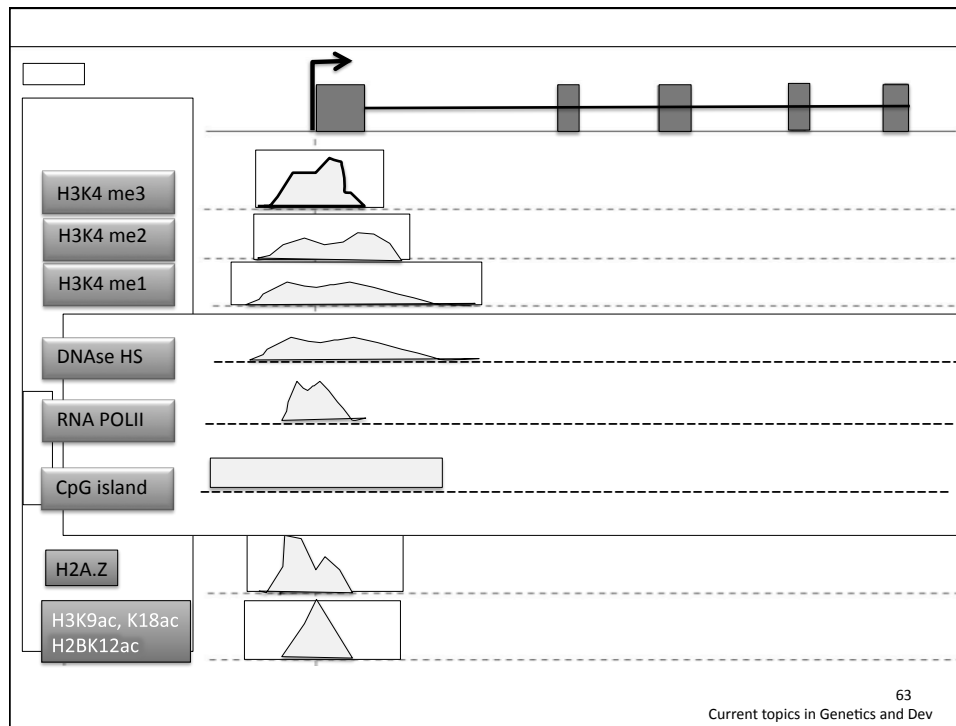
*“Good science  
is all about following the  
data as it shows up  
letting yourself be proven  
wrong,  
and letting everything  
change while you’re  
working on it “*

61

## Diversity of histone modifications

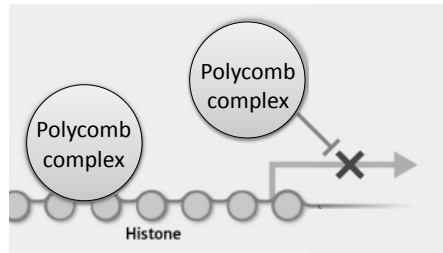
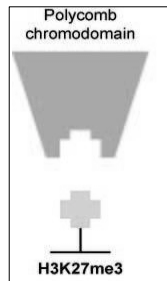


62





## Repressed promoters

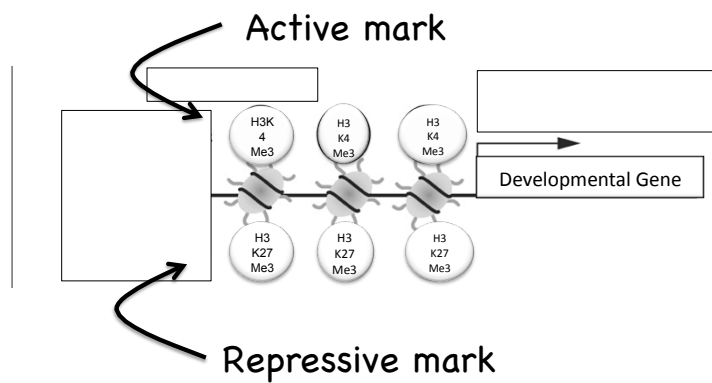


H3K27me3  
H3K9me3  
DNA methylation

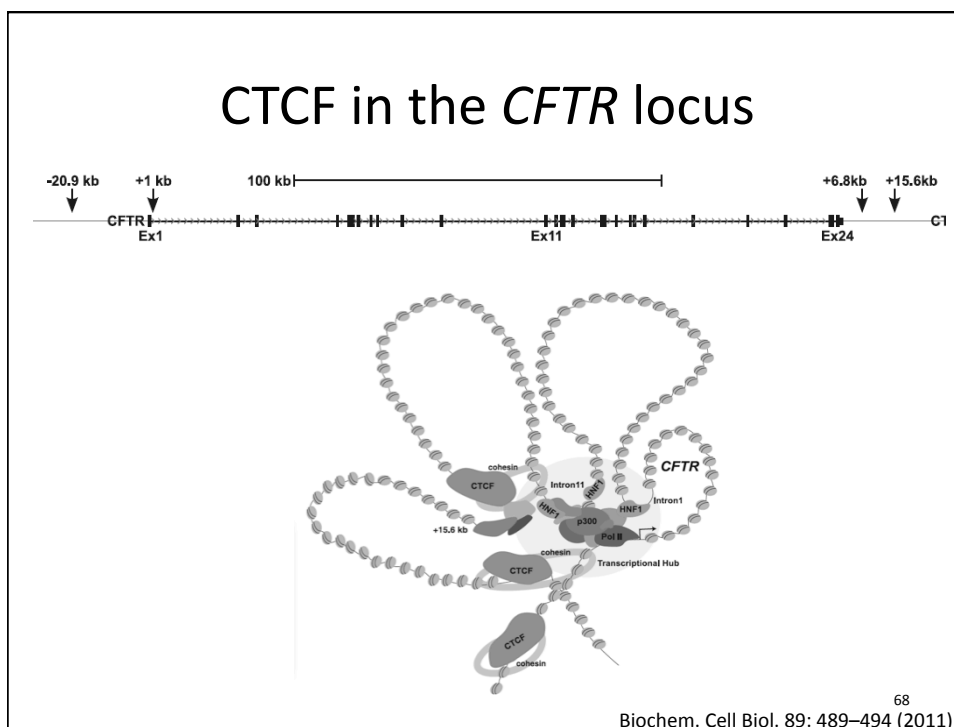
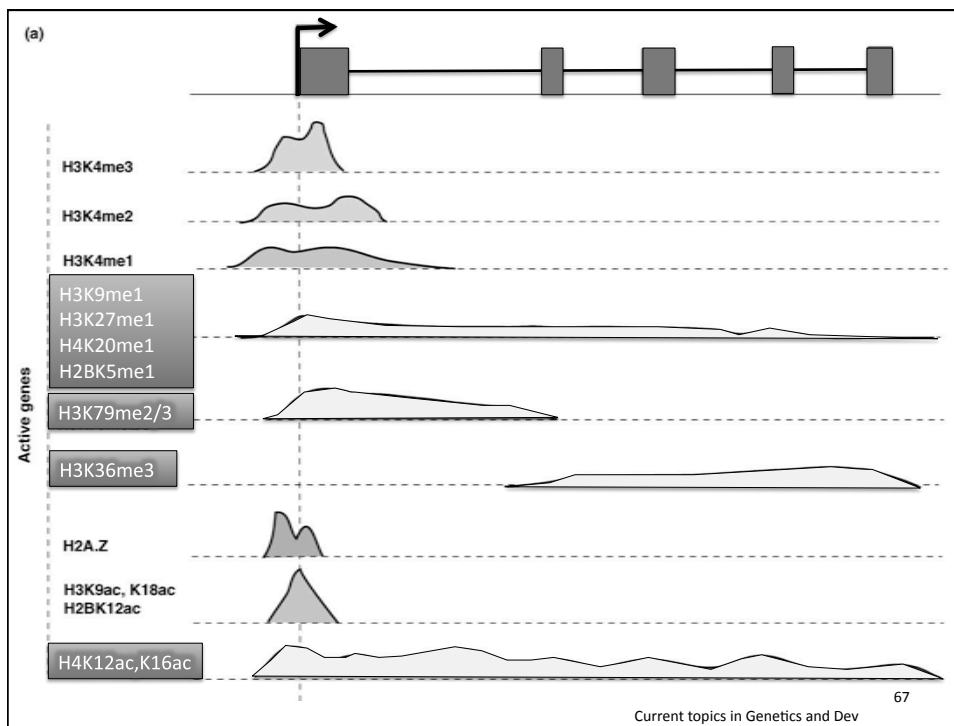
Annu. Rev Genomics Hum. Genet. 8:299-325

www.rikenresearch.riken.jp/eng/frontline/5514 65

## Bivalent marks in ES cells



Kondo Yonsei Med J. 2009 Aug;50(4):455-466.

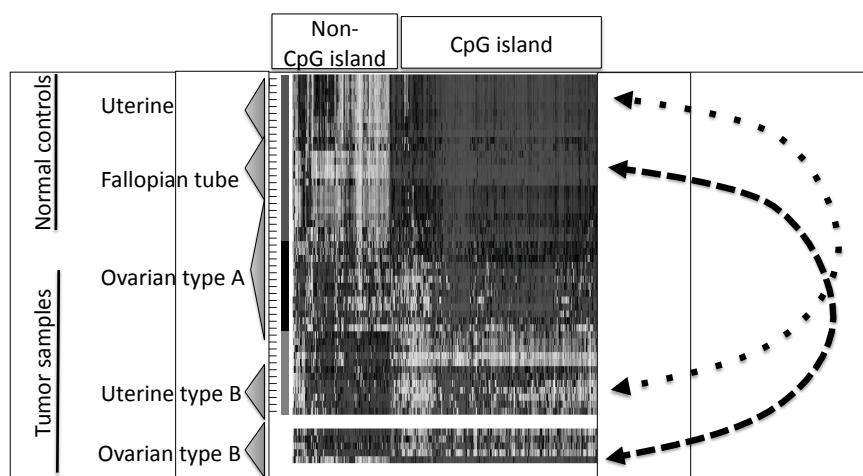


## DNA methylation

- 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<sup>70</sup>

## Contributions of the human genome project

- I. Understanding evolutionary diversity
- II. Genome function and regulatory elements
- III. Variants that disrupt function and explain



The end