




NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research



Current Topics in Genome Analysis
Spring 2016

GENOMIC REGULATORY AND EPIGENETIC LANDSCAPES
Laura Elnitski, Ph.D.

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




Current Topics in Genome Analysis 2016


Laura Elnitski

*No Relevant Financial Relationships with
Commercial Interests*

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



Human genome project

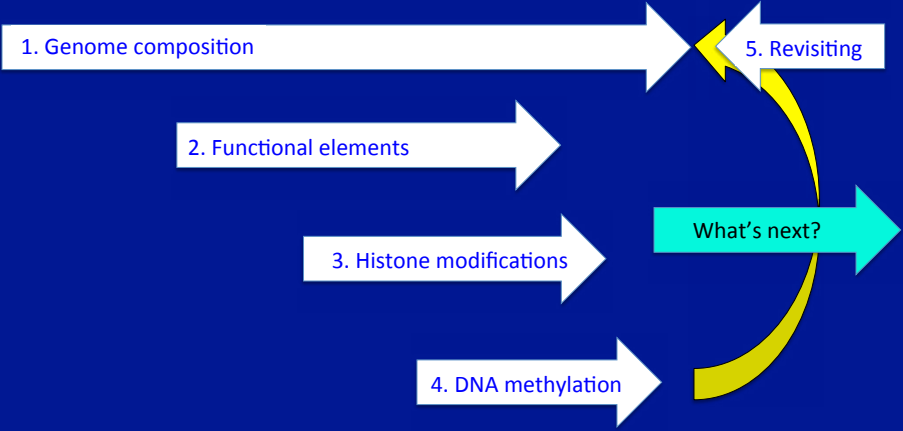


“The nitrogen in our DNA, the calcium in our teeth, the iron in our blood, the carbon in our apple pies were made in the interiors of collapsing stars. We are made of starstuff.”

— Carl Sagan, *Cosmos*

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Outline: From blueprint to implementation



1. Genome composition

2. Functional elements

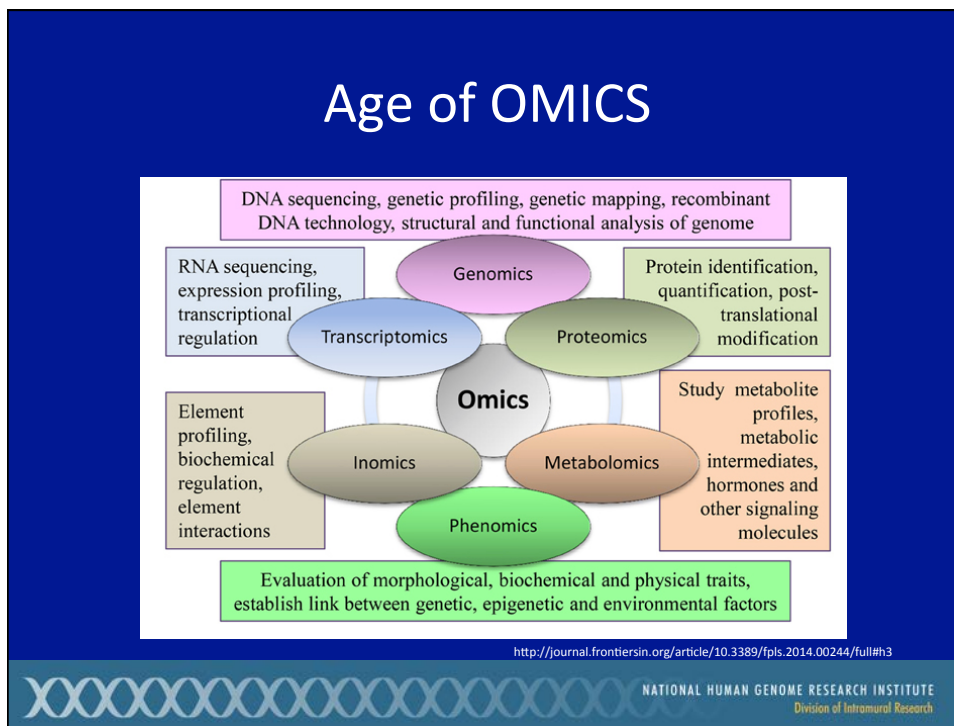
3. Histone modifications

4. DNA methylation

5. Revisiting

What's next?

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



1.1. Genome composition

Human genome project



**"NOW THIS IS
NOT THE END.
IT IS NOT EVEN
THE BEGINNING
OF THE END.
BUT IT IS,
PERHAPS,
THE END OF THE
BEGINNING."**

SIR WINSTON CHURCHILL (1874-1965)

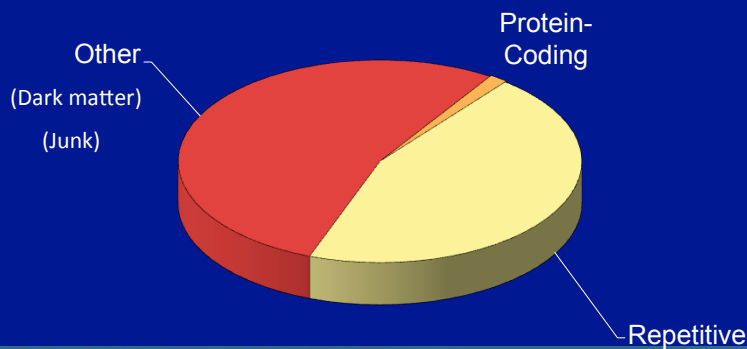
© UNFINISHED POLIS



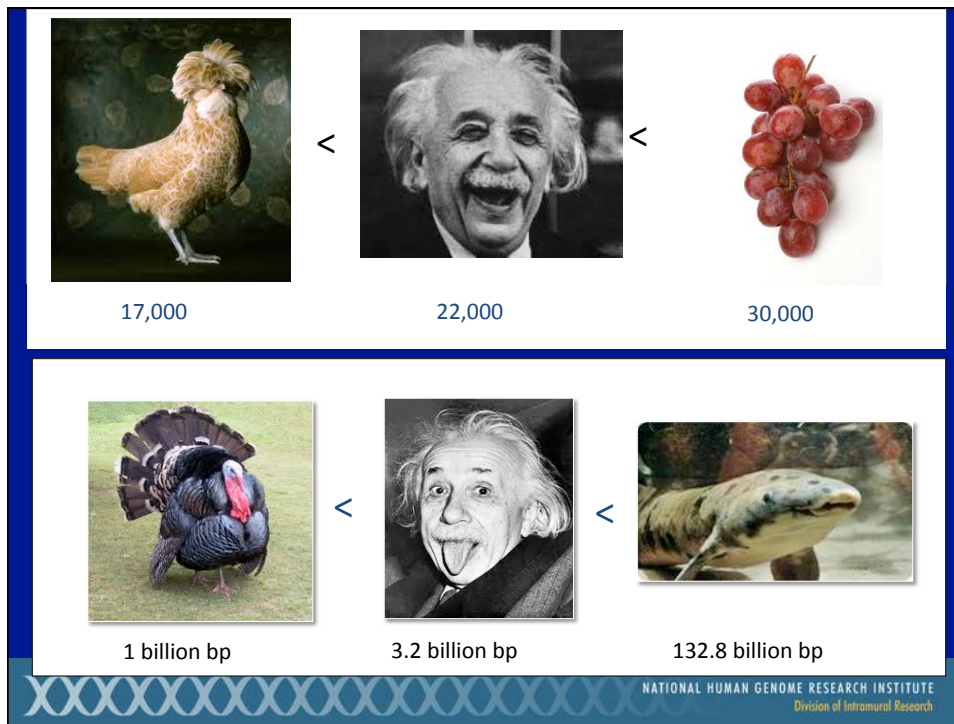
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Characterizing the human genome

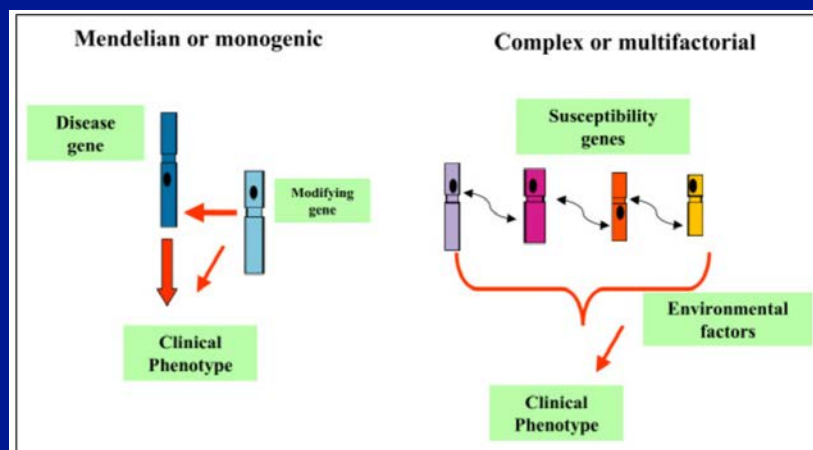
- ~ 3.2 billion bases
- ~ 22,000 protein-coding genes
- 98% noncoding



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



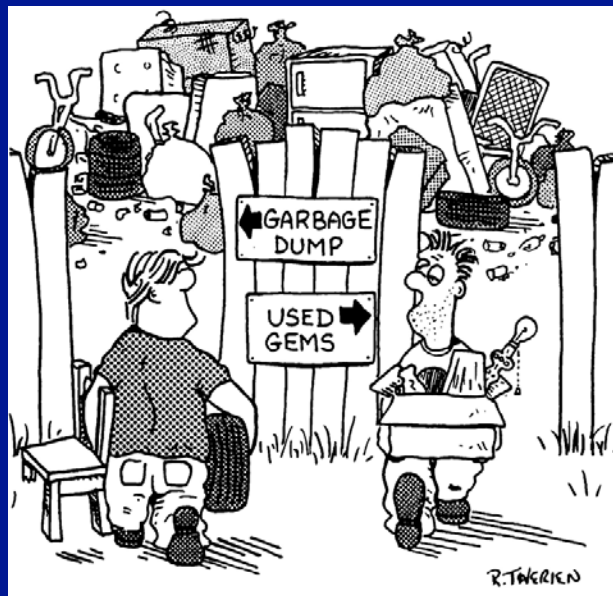
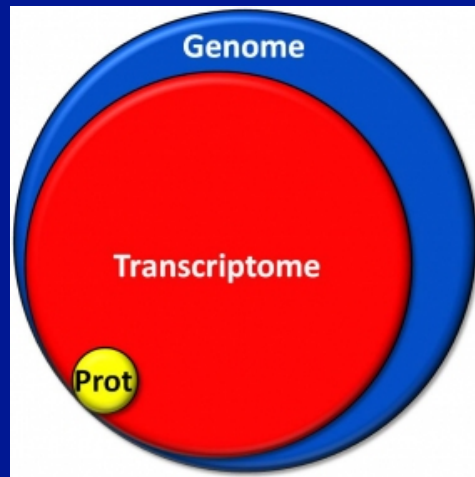
Genetic disease



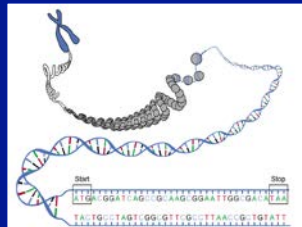
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2630295/>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Pervasive transcription

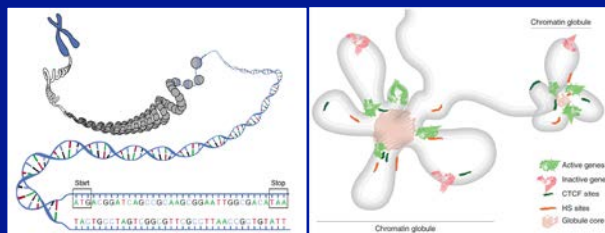


Envisioning the genome



A. Linear sequence

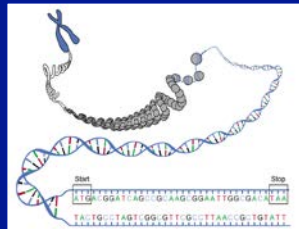
Envisioning the genome



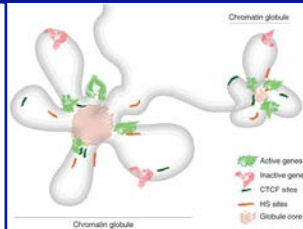
A. Linear sequence

B. Looping interactions

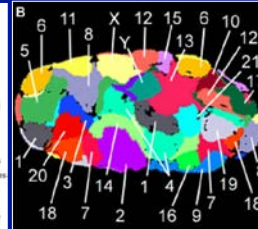
Envisioning the genome



A. Linear sequence



B. Looping interactions



C. 3D packaging



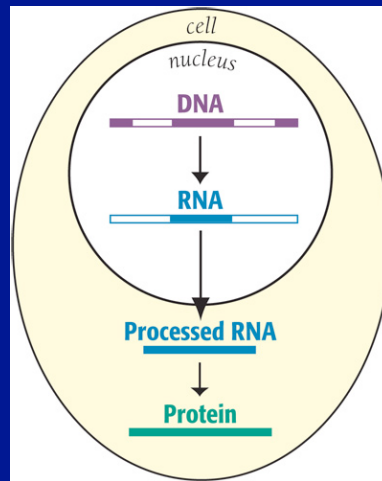
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

From blueprint to implementation

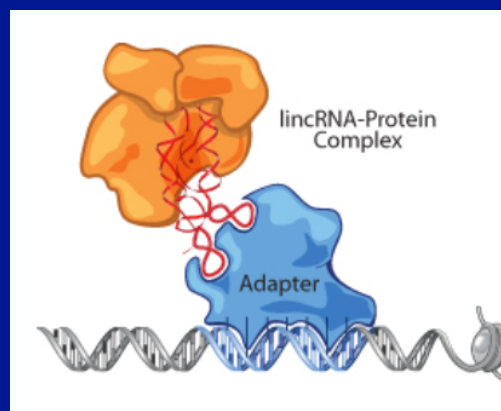


"nothing in the genome makes sense, except in 3D" NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Central dogma - 1958



Noncoding RNA



1.2. Examining the linear sequence

EVOLUTIONARY CLUES

Evolutionary Distance



Human



Chimpanzee



Horse



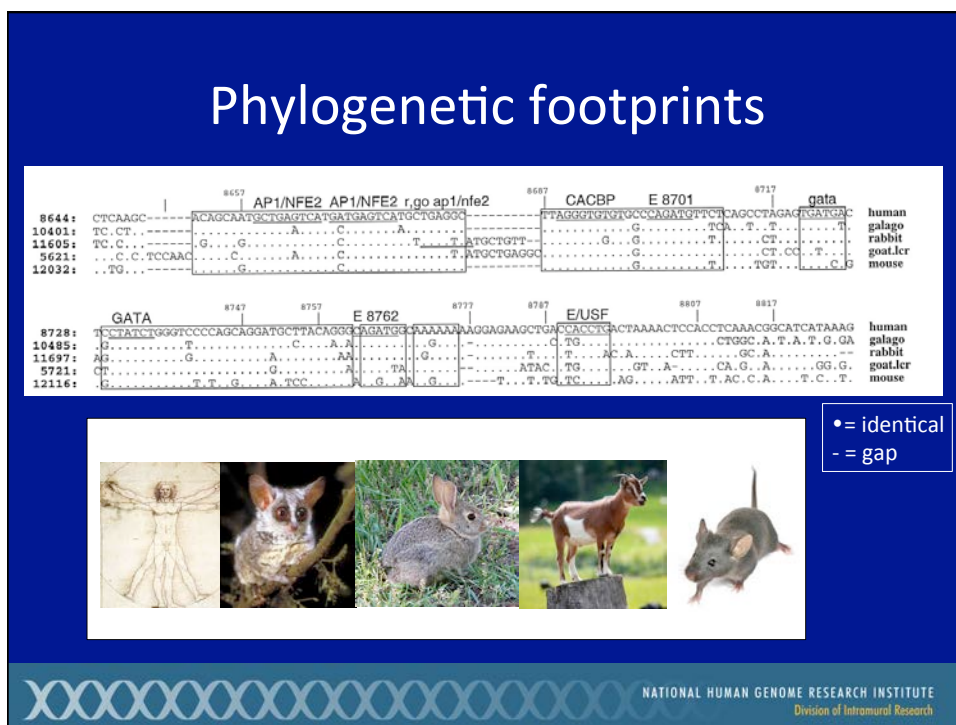
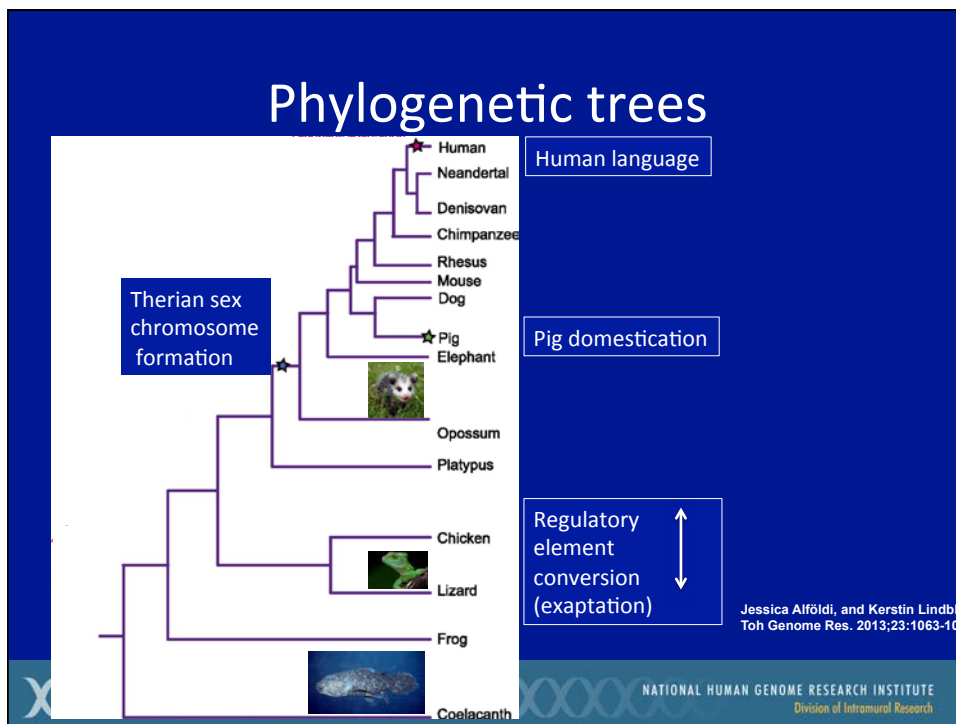
Rat



Platypus



Pufferfish



Phylogenetic shadowing

Sequence elements under greater constraints:
candidate functional element

□: nucleotide difference with at least one species

THE RESEARCH INSTITUTE
Division of Intramural Research

Conserved elements in the human genome

alignments in neutral regions

all human-mouse alignments

a

Sequence level view

```

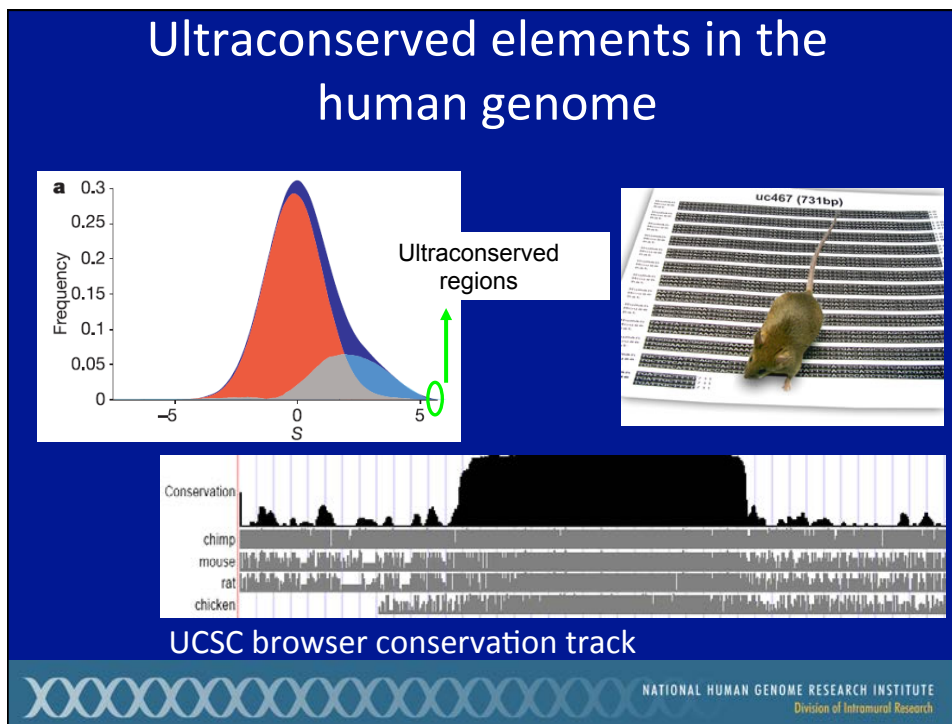
            ...CTTTGCGA-TGAGTAGCATCTACTATT...
            ...ACCTGGGACTGACTA-CATCGACTACGA...
        
```

Selective constraint
on
5%
of
human genome

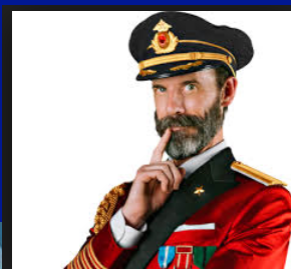
85% id on average

[Mouse consortium, *Nature* 2002]

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

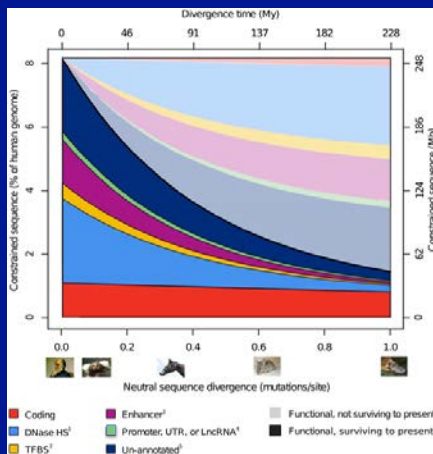


- There are *known knowns*; these are things we know we know.
- We also know there are *known unknowns*; that is to say we know there are some things we do not know.
- There are also *unknown unknowns* – the ones we don't know we don't know.
- It is the latter category that tend to be the *difficult* ones



New estimates of selective constraint

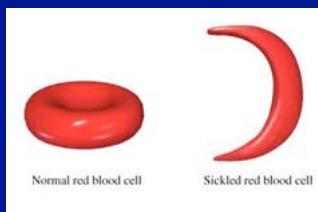
fraction of
 constrained
 sequence that has
 been retained
 (saturated colors)



fraction of
 constrained
 sequence that has
 been turned over
 (pastel colors)

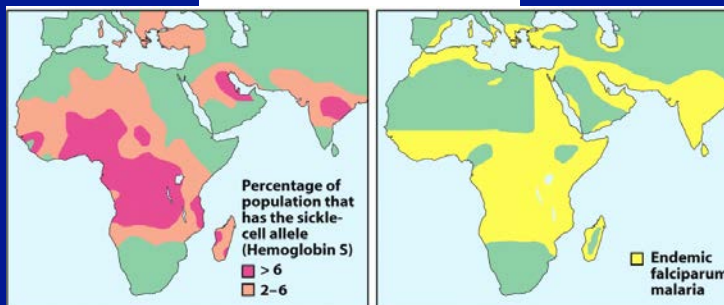
<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004525>

Sickle cell anemia

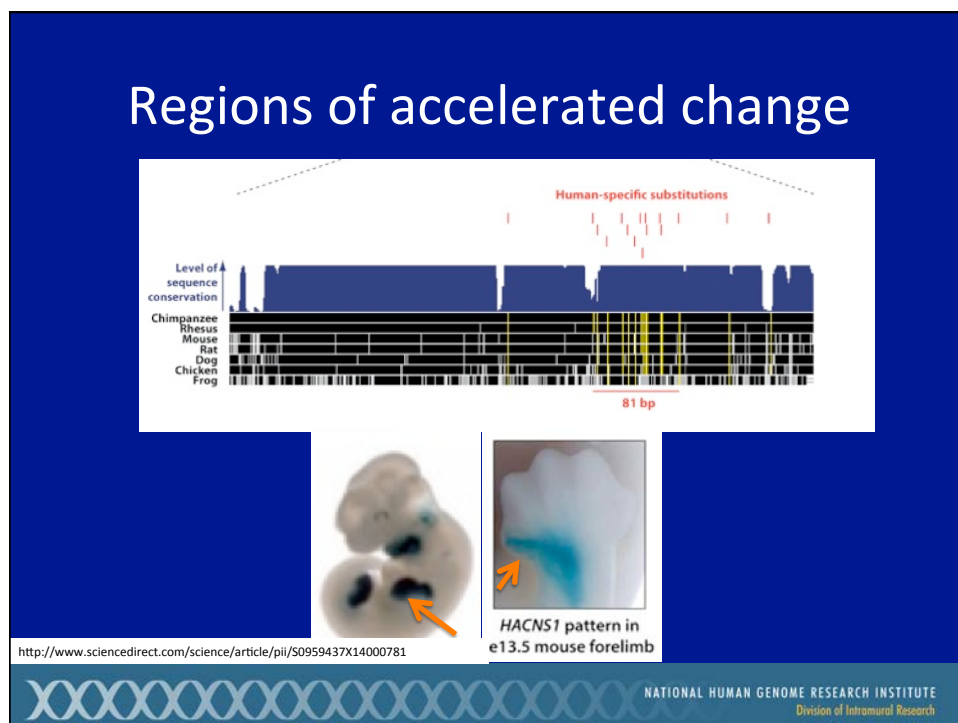


Normal red blood cell

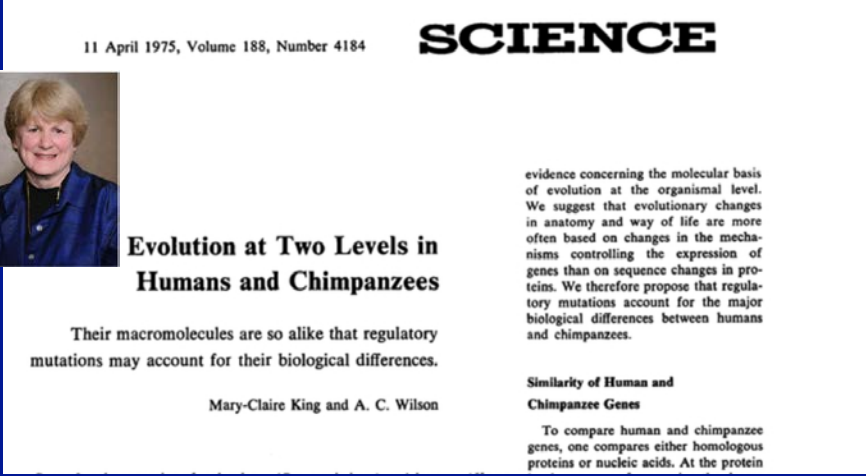
Sickled red blood cell



The McGraw-Hill
 Companies




Regulatory variation



11 April 1975, Volume 188, Number 4184

SCIENCE



Evolution at Two Levels in Humans and Chimpanzees

Their macromolecules are so alike that regulatory mutations may account for their biological differences.


Mary-Claire King and A. C. Wilson

evidence concerning the molecular basis of evolution at the organismal level. We suggest that evolutionary changes in anatomy and way of life are more often based on changes in the mechanisms controlling the expression of genes than on sequence changes in proteins. We therefore propose that regulatory mutations account for the major biological differences between humans and chimpanzees.

Similarity of Human and Chimpanzee Genes

To compare human and chimpanzee genes, one compares either homologous proteins or nucleic acids. At the protein

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research



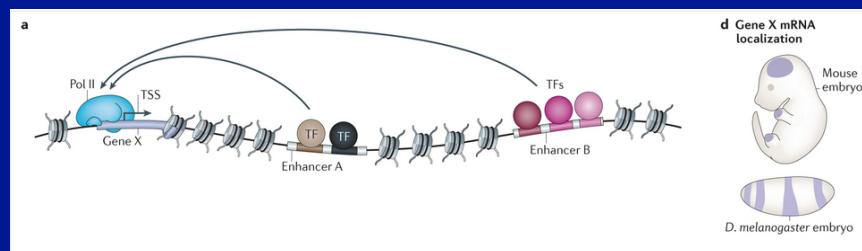
Out of intense complexities
intense simplicities emerge.

Winston Churchill

2.1 Functional elements in linear sequence



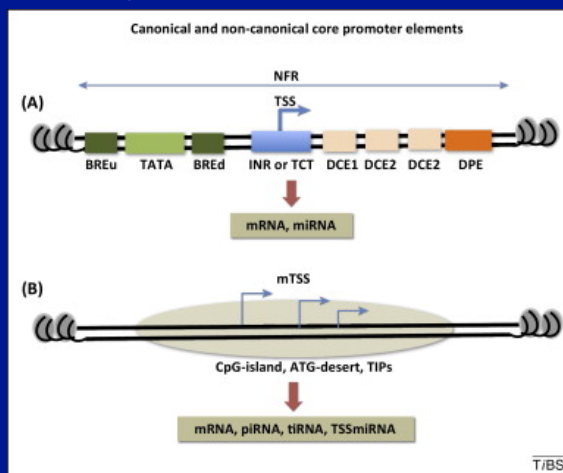
Enhancers



http://www.nature.com/nrg/journal/v15/n4/fig_tab/nrg3682_F1.html



Core promoter elements

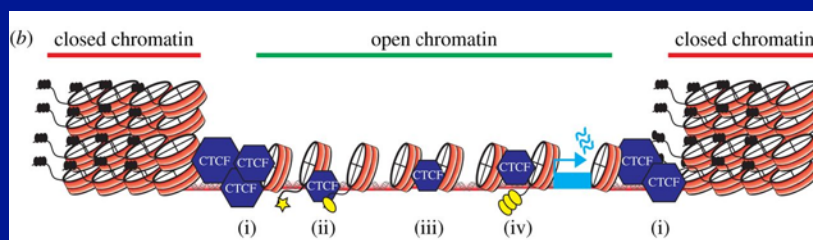


Trends in Biochemical Sciences 2015 40, 165-171 DOI: (10.1016/j.tibs.2015.01.007)



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Insulators/boundaries

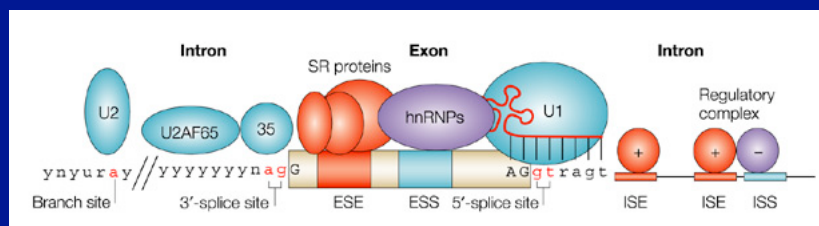


<http://rstb.royalsocietypublishing.org/content/368/1620/20120369>



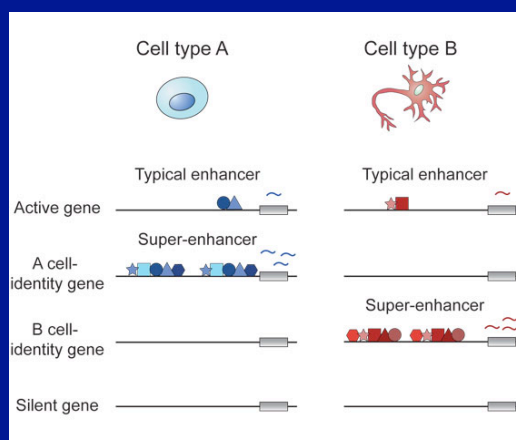
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Splicing elements

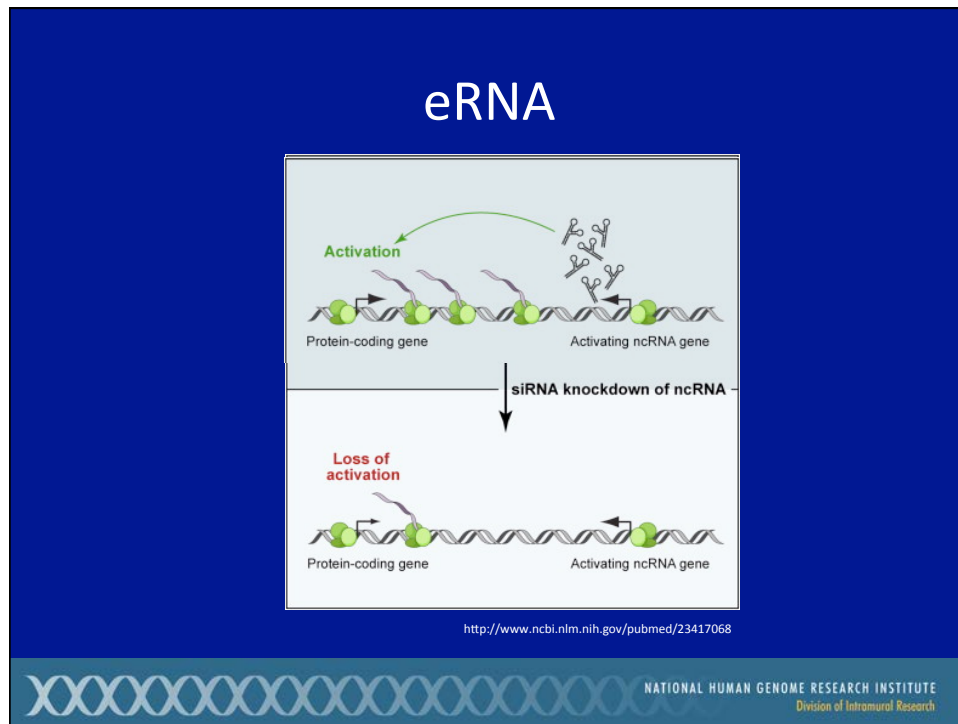


Nature Reviews Genetics 2004 5: 389-396

Super & stretch enhancers



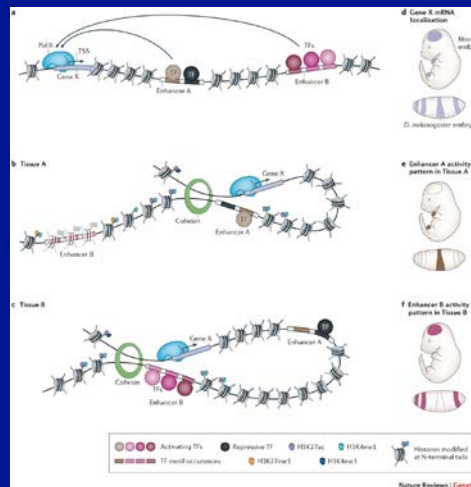
<http://www.pnas.org/content/110/44/17921.abstract>



2.2. Looping interactions



Looping interactions

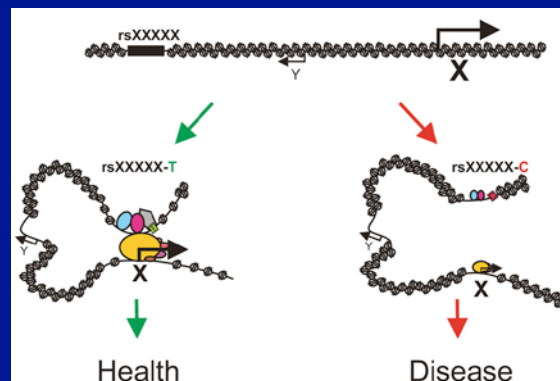


<http://www.nature.com/nrg/journal/v15/n4/full/nrg3682.html>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Variants in enhancers

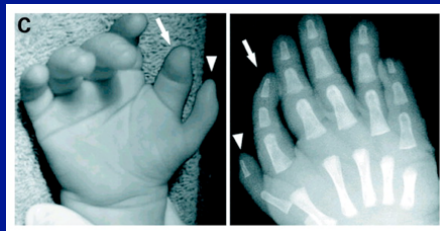
Allele Specific activity



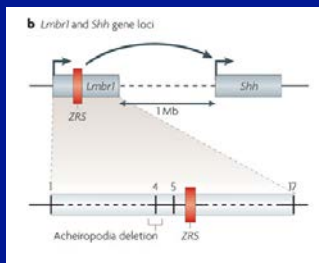
<http://journal.frontiersin.org/Journal/10.3389/fgene.2012.00195/full#sthash.nUct1Orj.dpuf>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

SHH enhancer mutations

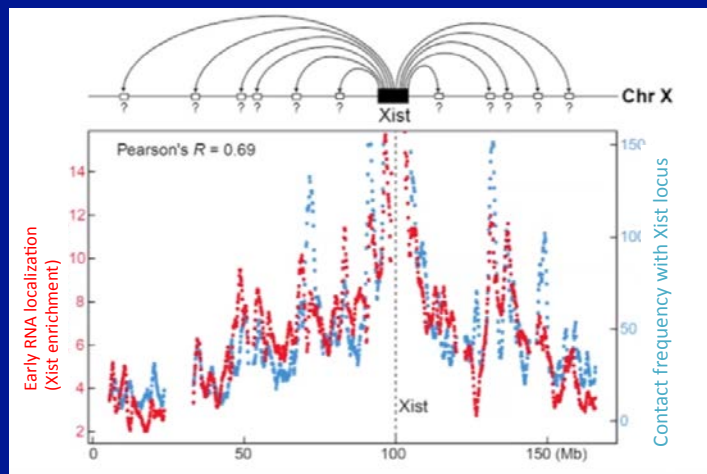


Hemingway cat with six toes



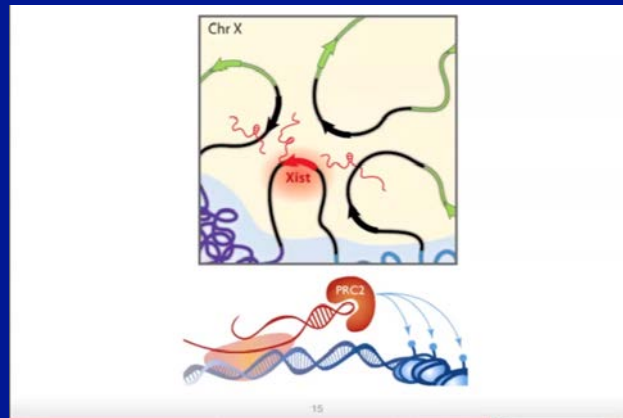
Hum. Mol. Genet. 2003 12 : 1725-1735
Development 2005 132 : 4 797-803
Hum. Mol. Genet. (2008) 17 (7): 978-985

XIST dispersion



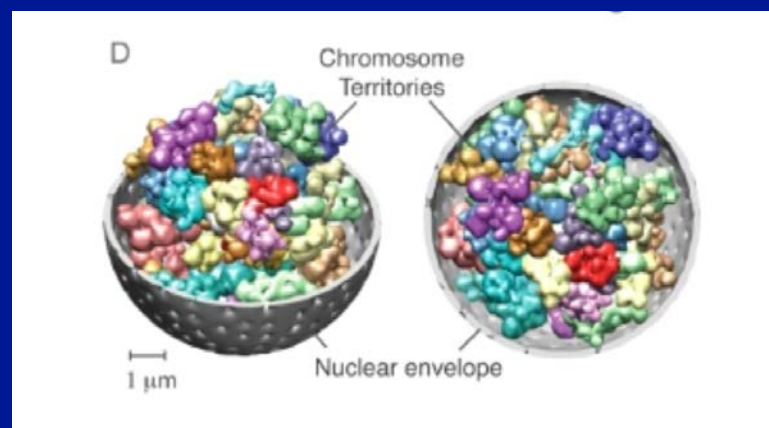
<http://www.youtube.com/watch?v=P3X4ujzRxc4#t=150>

Regulatory architecture



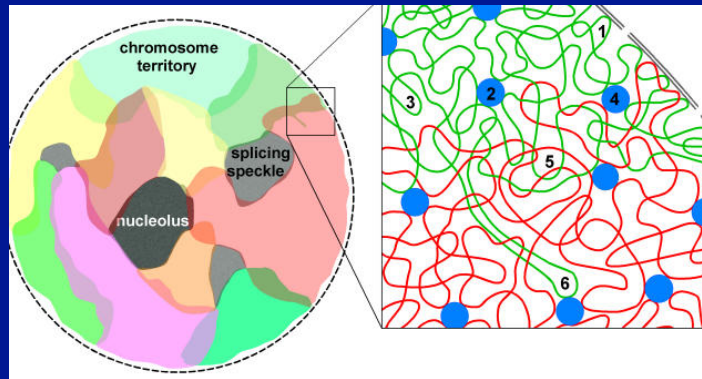
<http://www.youtube.com/watch?v=P3X4ujzRxc4#t=150>

2.3 3D packaging



PLoS Biol. 2006 May; 4(5): e138.

Inter-chromosomal networks

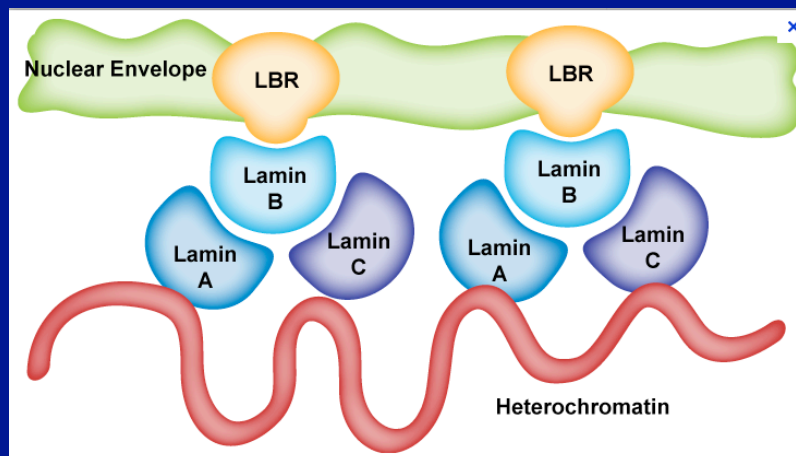


PLoS Biol. 2006 May; 4(5): e138.



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Lamin domains

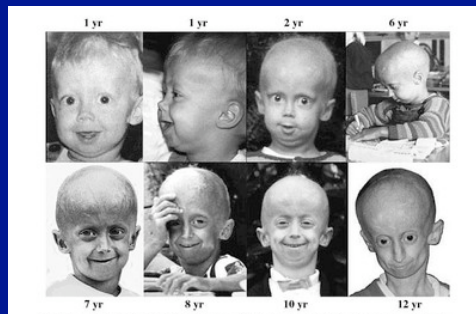


PLoS Biol. 2005 3:e157

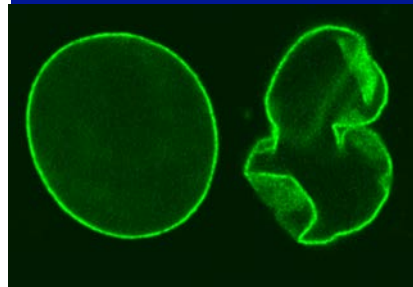


NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Progeria



Am J. Med Genet. 23:2603-24

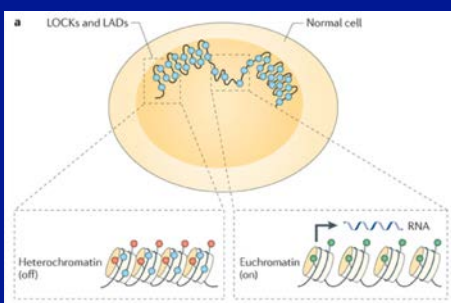


<http://www.pnas.org/content/109/41/16666>

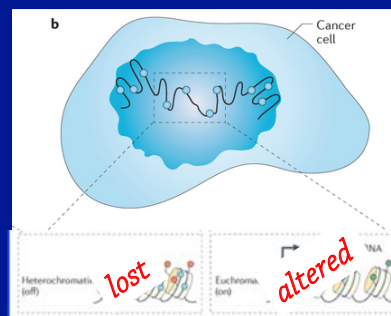


NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Altered genome topology in cancer



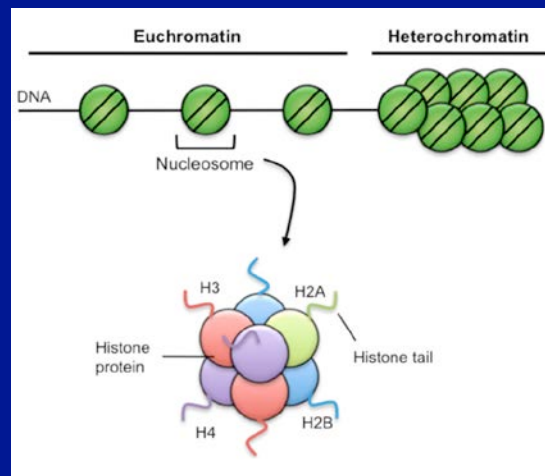
<http://www.nature.com/nrc/journal/v13/n7/full/nrc3486.html>



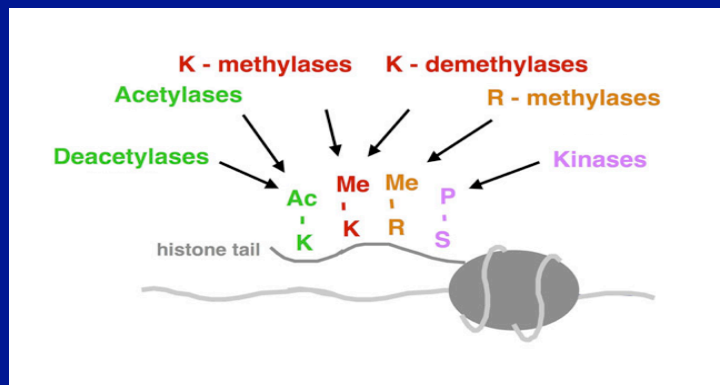
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

3. Histone modifications

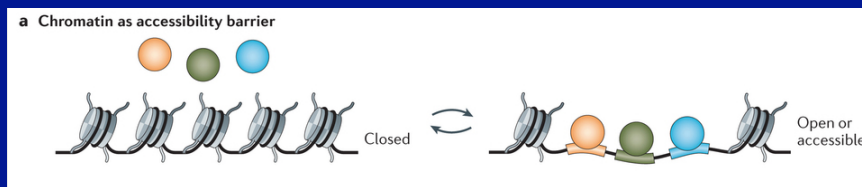
Active and inactive chromatin



Types of histone modifications



Chromatin as a regulator



http://www.nature.com/nrg/journal/v15/n4/fig_tab/nrg3682_F2.html



Enhancer dynamics

b Active enhancer

Enhancer

H3K27ac

H3K4me1

d Closed or poised enhancer

H3K4me1

H3K27me3

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Rosetta stone of chromatin?

ENCODE ChromHMM Chromatin State

- Active Promoter
- Weak Promoter
- Inactive/poised Promoter
- Strong enhancer
- Weak/poised enhancer
- Insulator/CTCF
- Transcribed
- Weak transcribed
- Polycomb repressed
- Heterochromatin; low signal

← P53 ChIPseq peak

p53RE TSS Transcription

Gene exon

ChromHMM

ChromHMM state at p53 ChIP-seq peak maximum

ChromHMM states at TSS of gene nearby p53 peak

<http://www.nature.com/nmeth/journal/v9/n3/full/nmeth.1906.html>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
 Division of Intramural Research

Kabuki Syndrome


(a) Repressive complex: SUZ12, RBBP7/4, EED, EZH2. Histone tails with Me and K 27, A R T K 4 sites.

(b) Active complex: WDR5, RbBP5, NCOA6, ASH2L, PTIP, PA1, KMT2D, DPY30, KDM6A. RNA Pol II. transcription.

<http://onlinelibrary.wiley.com/doi/10.1111/cge.12051/full#fig-0003>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

4. DNA methylation



“Age is not a particularly interesting subject.

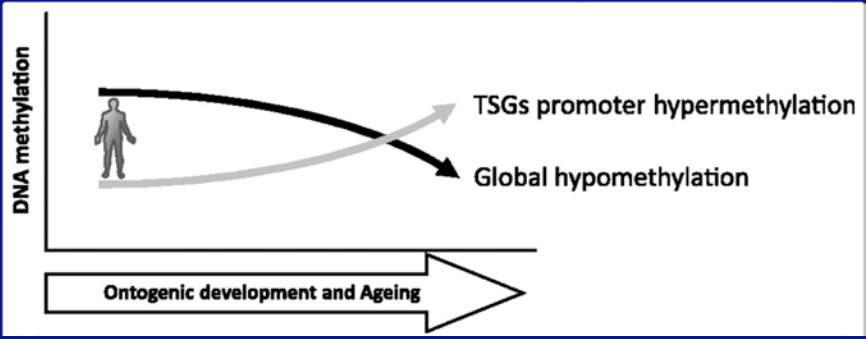
Anyone can get old.

All you have to do is live long enough.”

(Groucho Marx)

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Gradual shifts in methylation



DNA methylation

TSGs promoter hypermethylation

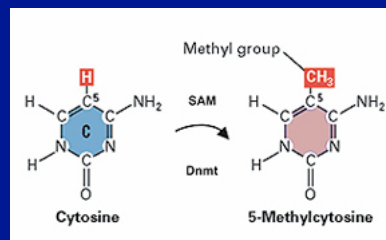
Global hypomethylation

Ontogenic development and Ageing

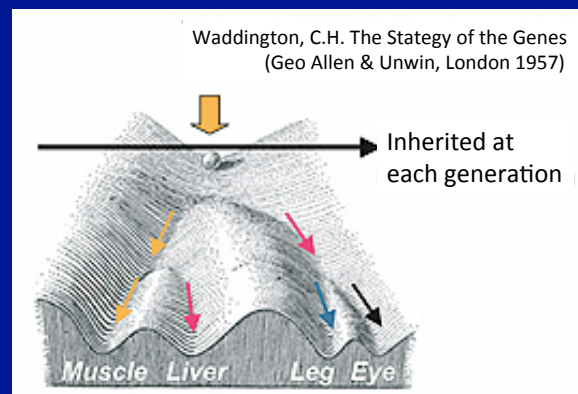
<http://jap.physiology.org/content/109/1/243>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

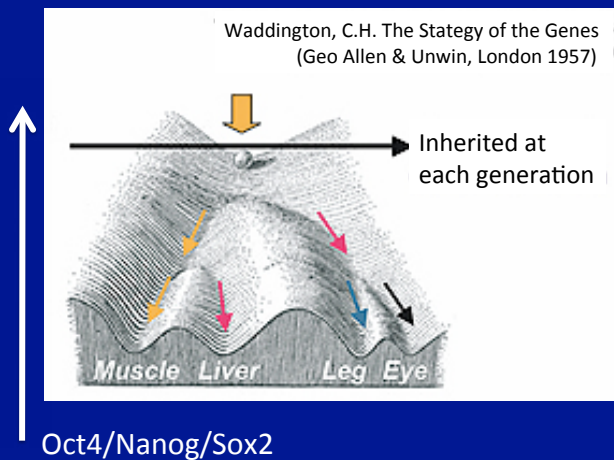
5-methyl cytosine



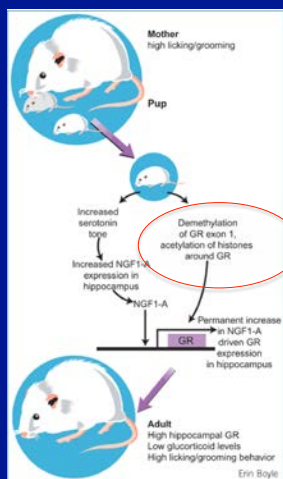
Waddington's epigenetic landscape



Nuclear reprogramming



Behavioral effects



Nature Neuroscience 7, 791 - 792 (2004)



Methylation and cancer

(a) Repetitive sequence

(b) CpG island promoters

(c) CpG island shore

Legend: | Methylated cytosine | Unmethylated cytosine

Alcohol Research: Current Reviews, Volume 35, Issue Number 1

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Genome misfolding in cancer

A B

normal tumor

PDGFRA and FIP1L1

<https://www.broadinstitute.org/news/7706>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

5.1. Revisiting genome composition



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

ENCODE

a surprisingly large amount of the human genome, 80.4%, is covered by at least one ENCODE-identified element

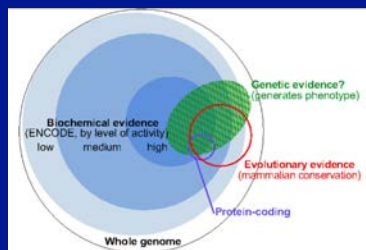
ARTICLE doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

<http://www.nature.com/nature/journal/v489/n7414/full/nature11247.html>



<http://www.pnas.org/content/111/17/6131.abstract>

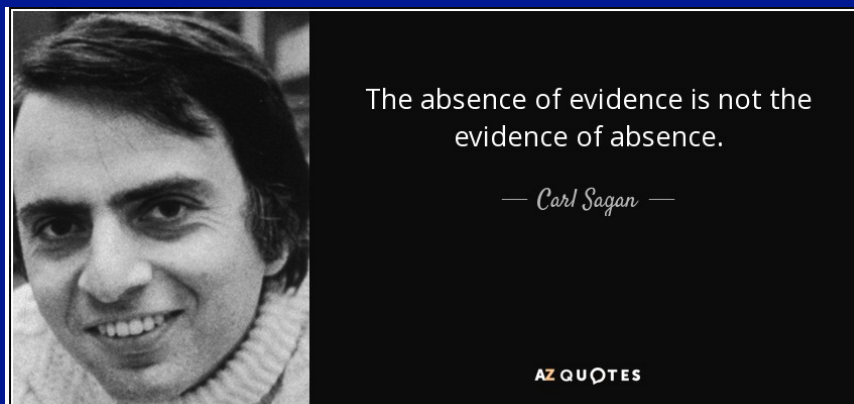
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Tallying genomic features

Genomic element	Number	Source
Protein-coding genes	20330	GENCODE V17 (Feb2013, GRCh37) Ensembl 72
Long non-coding RNAs	13333	GENCODE V17
lincRNAs	6020	GENCODE V17
Pseudogenes	14154	GENCODE V17
Short non-coding RNAs	9078	GENCODE V17
miRNAs	3086	GENCODE V17
Promoters	70292	ENCODE [3]
Enhancers	399124	ENCODE [3]
TFBS (ChIP peaks)	636336	ENCODE [3]

lincRNAs, long intergenic non-coding RNAs; miRNAs, micro RNAs; TFBS, transcription factor binding sites.

Defining function



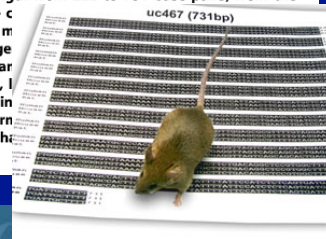
Deletion of promising regions

Deletion of Ultraconserved Elements Yields Viable Mice

Nadav Ahituv^{1,2*}, Yiwen Zhu¹, Axel Visel¹, Amy Holt¹, Veena Afzal¹, Len A. Pennacchio^{1,2}, Edward M. Rubin^{1,2*}

¹ Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America, ² United States Department of Energy Joint Genome Institute, Walnut Creek, California, United States of America

Ultraconserved elements have been suggested to retain extended perfect sequence identity between the human, mouse, and rat genomes due to essential functional properties. To investigate the necessities of these elements *in vivo*, we removed four noncoding ultraconserved elements (ranging in length from 222 to 731 base pairs) from the mouse genome. To maximize the likelihood of observing a phenotype, we *inactivated* enhancers in a mouse transgenic assay and that are near genes that exhibit *in vivo* inactivation in the mouse and when their expression is altered due to other genetic elements. The resulting lines of mice lacking these ultraconserved elements were viable and showed no abnormalities when assayed for a variety of phenotypes including growth, development, and behavior. In addition, more targeted screens, informed by the abnormalities observed in other studies, investigated elements had been altered, also failed to reveal notable abnormalities. These results indicate that all of the possible phenotypic impact of the deleted sequences, indicate that these elements do not necessarily reflect crucial functions required for viability.



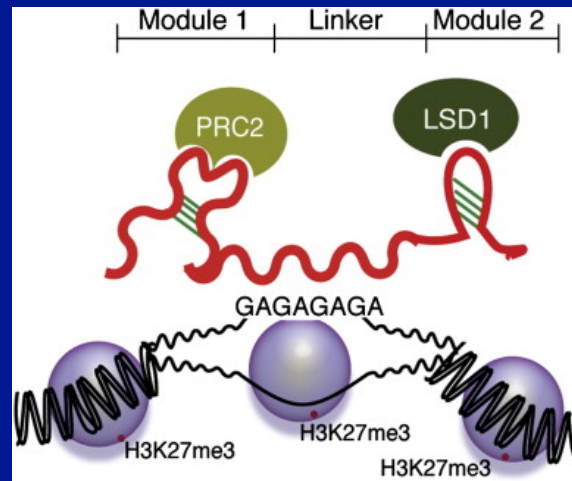
Division of Intramural Research

Shadow enhancers (redundancy/resilience?)



NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

Structural conservation in lncRNA



<http://www.sciencedirect.com/science/article/pii/S0304416513004753>



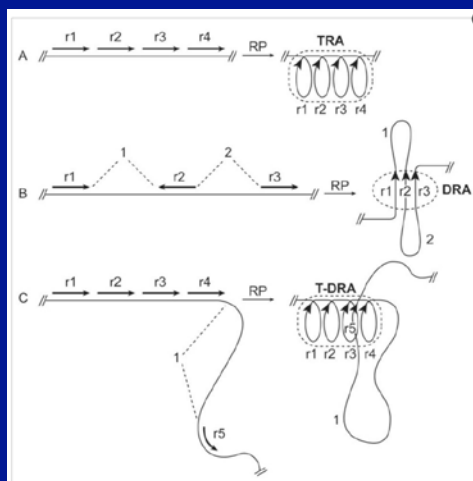
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

5.2. Seemingly nonfunctional regions



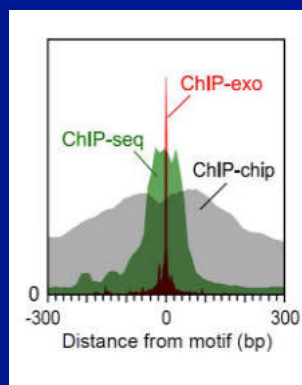
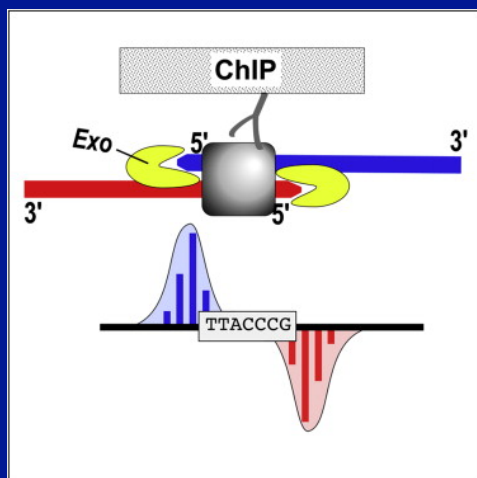
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
Division of Intramural Research

TE repeat pairing



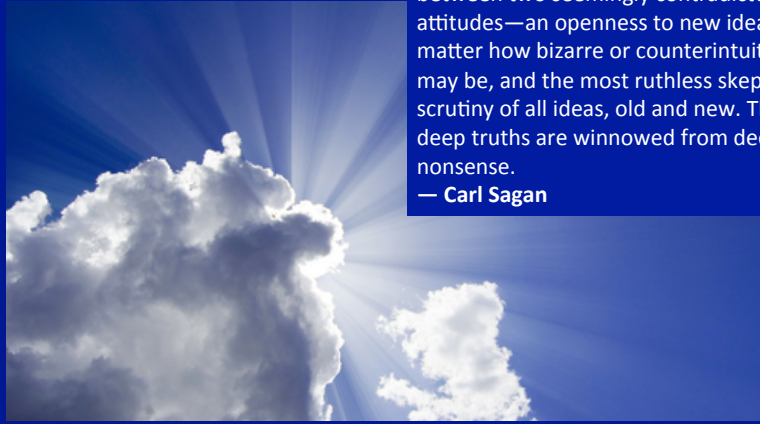
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3927610/figure/f1-genes-02-00502/>

Wide-spread regulated RNA Pol II initiation



[http://www.cell.com/abstract/S0092-8674\(11\)01351-1](http://www.cell.com/abstract/S0092-8674(11)01351-1)

Illuminating the dark matter



At the heart of science is an essential balance between two seemingly contradictory attitudes—an openness to new ideas, no matter how bizarre or counterintuitive they may be, and the most ruthless skeptical scrutiny of all ideas, old and new. This is how deep truths are winnowed from deep nonsense.

— Carl Sagan