NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

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

Elusive Genomic Attributes

Physical Traits Illnesses Behaviors

Evolution at two levels in humans and chimpanzees
King and Wilson
Science 11 April 1975: 107-116
DOI: 10.1126/science.1090005

- "the modest divergence observed in protein sequences **cannot** account for the profound phenotypic differences between humans and chimps"

~5% of the genome is under negative selection

1.5% of that represents coding sequences

How much is functional?

Discussion Points

Nuclear Architecture

Spectrum of Genomic Mutations

Regulatory Mutations

Epigenetic Modifications

DNA Methylation in Cancer

Regulatory Dynamics

Spatial Organization

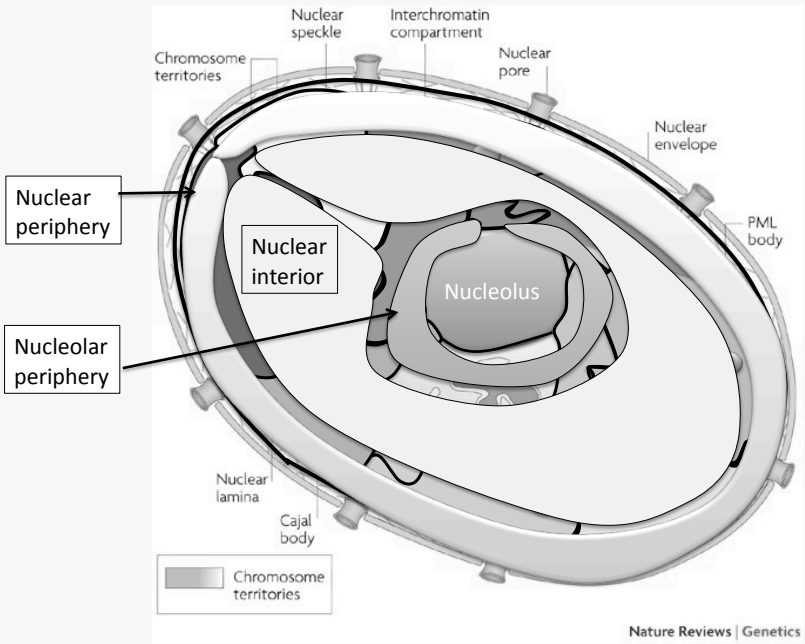
1. Individual chromosomes occupy distinct positions in the nucleus, referred to as chromosome territories

2. Different chromosome segments adopt a complex organization and topography within their chromosome territory.

Spatial Organization

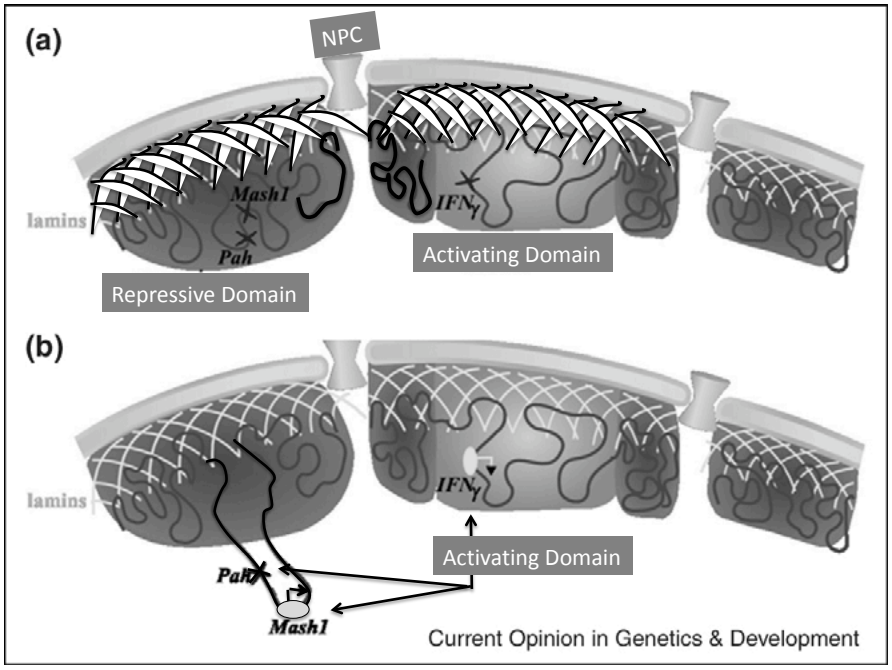
3. Gene-rich regions tend to be oriented towards the nuclear interior, whereas gene-poor regions tend to be oriented towards the periphery.

4. A polarized intranuclear distribution of gene-rich and gene-poor chromosomal segments has been shown to be an evolutionarily conserved principle of nuclear organization



Nuclear Dynamics

1. Repositioning of a gene locus is often associated with activation or silencing
2. Structural constraints impose limits on chromatin mobility
3. Understanding how the dynamic nature of the positioning of genetic material in the nuclear space and the higher-order architecture of the nucleus are integrated is essential to our overall understanding of gene regulation

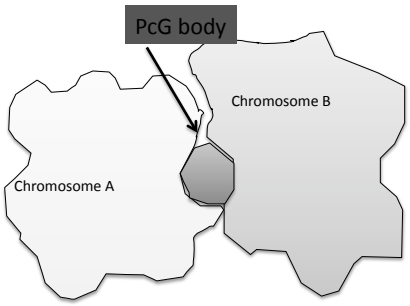


The possibility that spatial networks of genomic loci exist in the nucleus implies the presence of a previously unexplored level of gene regulation that coordinates expression across the genome.

Inter-chromosomal Interactions

Chromosome kissing

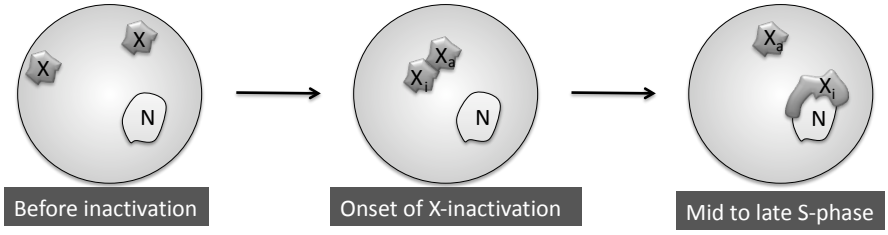
in gene silencing



Adapted from
Current opinion in Genetics and Development

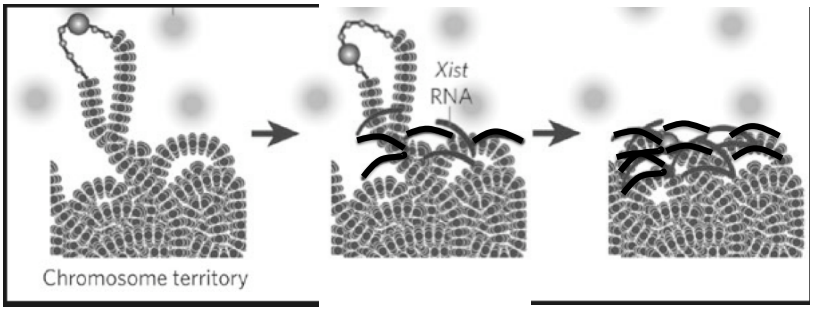
Chromosome Kissing

in X-inactivation



Adapted from Current Opinion in Genetics & Development Volume 17, Issue 5, October 2007, Pages 443-450 Differentiation and gene regulation.

Non-coding RNA



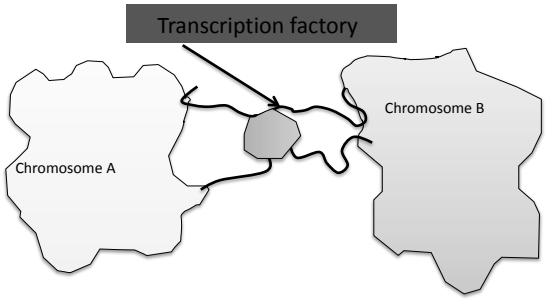
Peter Fraser

Inter-chromosomal Interactions

Chromosome kissing

in gene activation

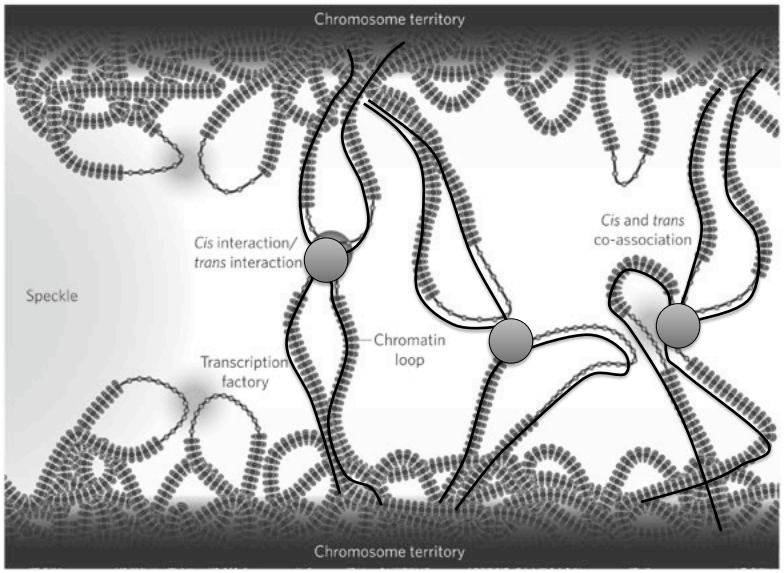
T-lymphocyte activation



also CTCF sites, splicing speckles, etc.

Adapted from Current Opinion in Genetics & Development Volume 17, Issue 5, October 2007, Pages 443-450 Differentiation and gene regulation.

Inter-chromosomal Interactions



Nature 447, 413-417 (24 May 2007)12

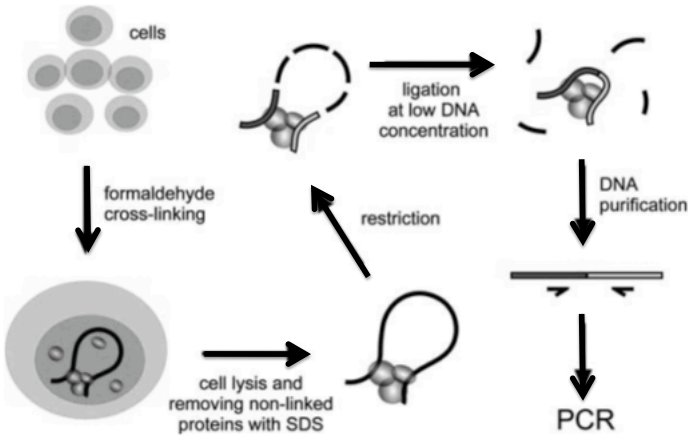
Transcription Factories

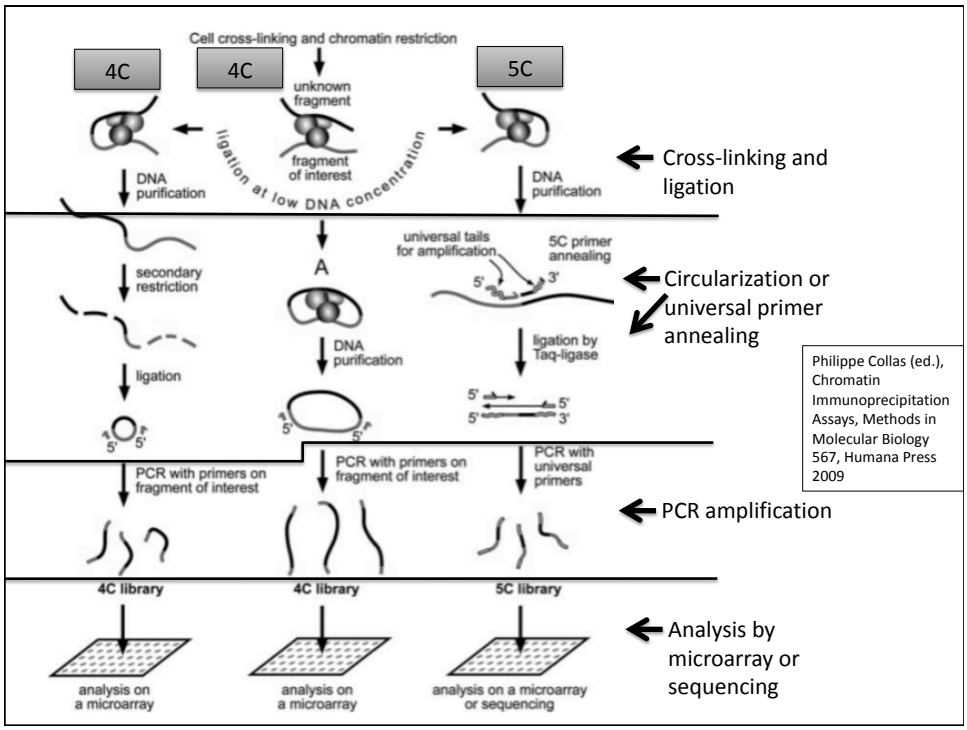
http://users.path.ox.ac.uk/%7Epcook/images/tcycle.html

The possibility that chromosome kissing events could be the origin of chromosomal rearrangements implies a way to study their derivation

Detecting long range interactions

Chromatin Conformation Capture





Philippe Collas (ed.), Chromatin Immunoprecipitation Assays, Methods in Molecular Biology 567, Humana Press 2009

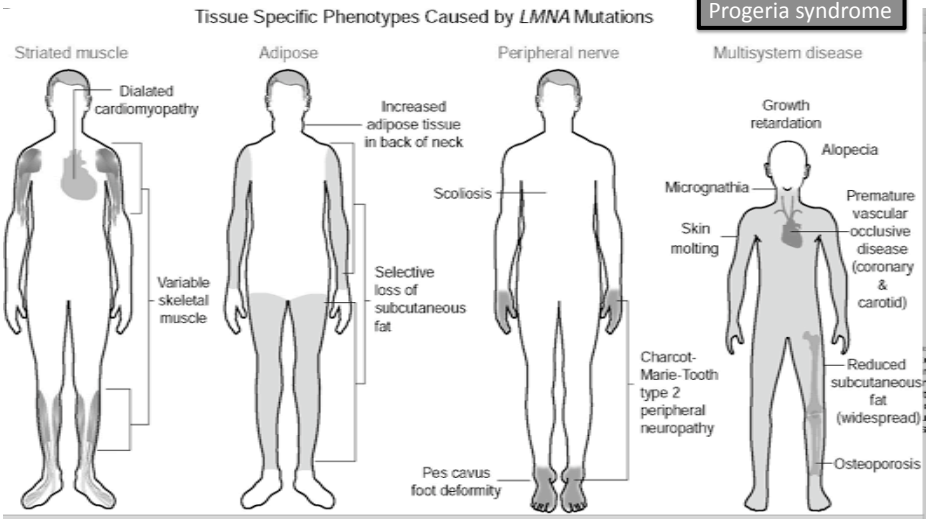
If chromosomal architecture is relevant to gene regulation, diseases stemming from mutations in these genes should be known

Disruption of the Regulatory Landscape

Mutations in genes encoding nuclear envelope proteins cause a fascinating array of diseases referred to as “nuclear envelopathies” or “laminopathies” that affect different tissues and organ systems.

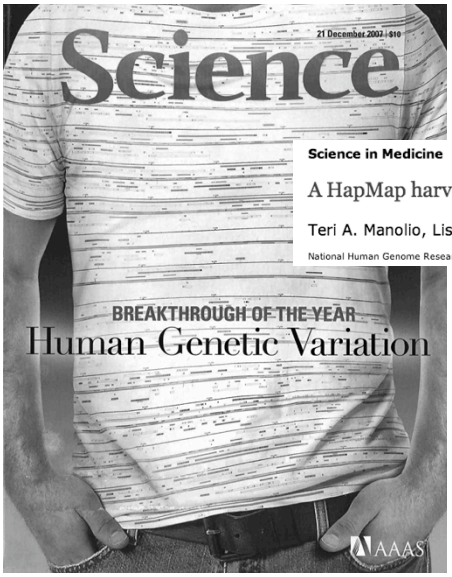
LAMINOPATHIES

Hutchinson-Gilford Progeria syndrome



Developmental Cell Volume 17, Issue 5, 17 November 2009, Pages 626-638
The Nuclear Envelope as a Signaling Node in Development and Disease
William T. Dauer and Howard J. Worman

Spectrum of Sequence Variants



Science in Medicine

A HapMap harvest of insights into the genetics of common disease

Teri A. Manolio, Lisa D. Brooks and Francis S. Collins

National Human Genome Research Institute, Bethesda, Maryland, USA.

Genome Wide Association Studies

Common Disease- Common or Rare Variant ?

SNP1 in exon 1 of Gene A
SNP1 in exon 2 of Gene A
SNP2 in exon 2 of Gene A
SNP2 in exon 2 of Gene B
SNP in reg. element of Gene A

→ Disease

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Current Opinion in Genetics & Development

Common vs. rare allele hypotheses for complex diseases

Nicholas J Schork, Sarah S Murray, Kelly A Frazer and Eric J Topol

Common Diseases Involve Multiple Variants

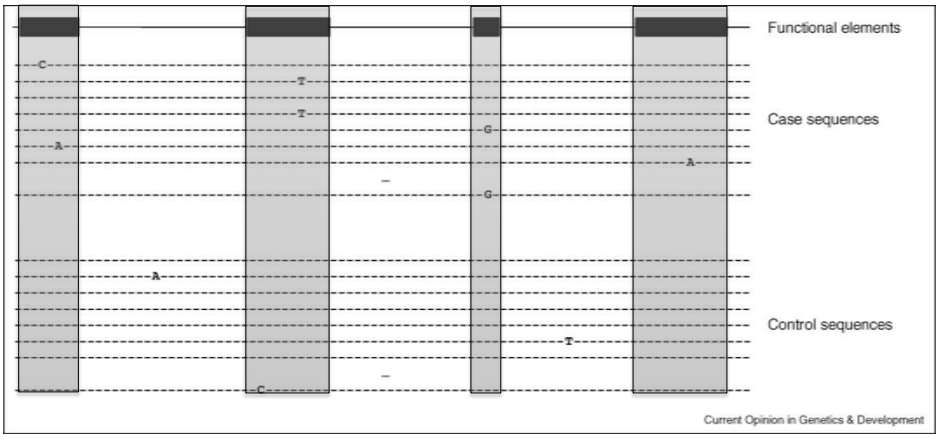
Table 1

Recent sequencing studies linking multiple rare variations to a phenotype or disease.

Reference	Gene	Phenotype	Results
[37] Nejentsev *et al.*	IFIH1	Type 1 diabetes	Multiple rare cSNPs are more frequent in T1D
[38] Marini *et al.*	MTHFR	Folate response	Multiple coding SNP effects are folate remedial
[39••] Ji *et al.*	Salt handling genes	Blood pressure	Multiple coding SNPs for individuals with low BP
[40] Azzopardi *et al.*	APC	Colorectal cancer	Multiple variations among colorectal cancer
[41] Masson *et al.*	CTRC	Pancreatitis	Multiple variations among pancreatitis patients
[42] Ma *et al.*	Toll-like receptors	Tuberculosis (TB)	Multiple coding variations influence TB
[43] Ahituv *et al.*	58 different genes	Obesity	Multiple variations among obese patients
[44] Romeo *et al.*	ANGPTL4	Elevated HDL	Multiple variations among high HDL patients
[45] Kotowski *et al.*	PCSK9	Low LDL	Frequent nonsense mutations among low LDL
[46] Cohen *et al.*2005)	PCSK9	Heart disease	Multiple sequence variations among HD patients
[47] Cohen *et al.*	NPC1L1	Low LDL	Multiple rare variants among low LDL patients
[48] Cohen *et al.*	PCSK9	Low LDL	Frequent nonsense mutations among low LDL
[49] Cohen *et al.*	ABCA1, APOA1, LCAT	Low plasma HDL	Coding SNPs differences for low HDL patients

Common vs. rare allele hypotheses for complex diseases Schork et al. Current Opinion in Genetics & Development 2009, 19:212–219

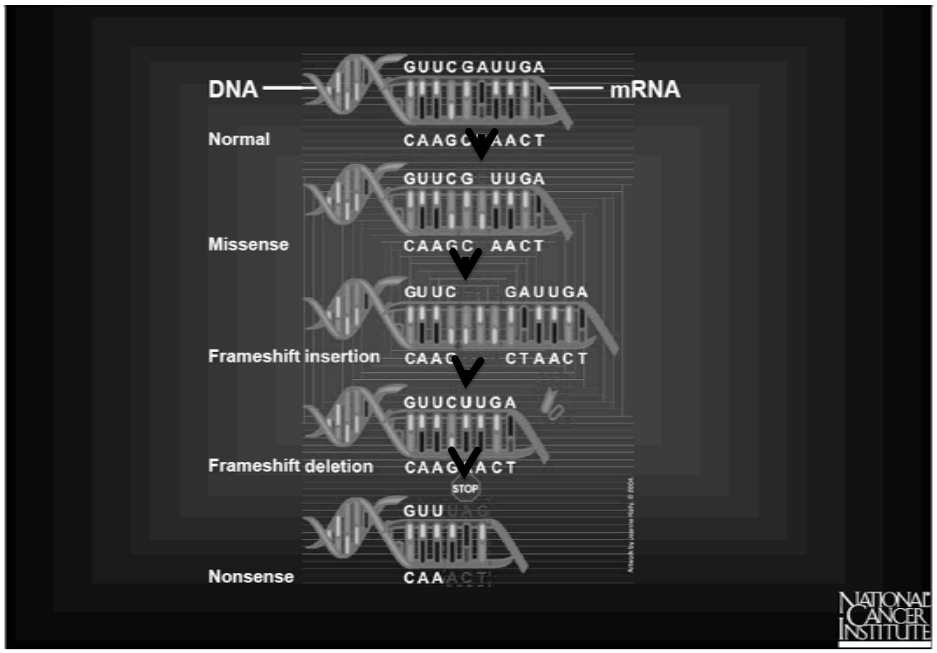
Allelic Heterogeneity



The conclusion that common diseases are multi-factorial in origin implies that many more disease-associated variants remain to be identified

Coding Mutations

Affect gene function and / or regulation of expression

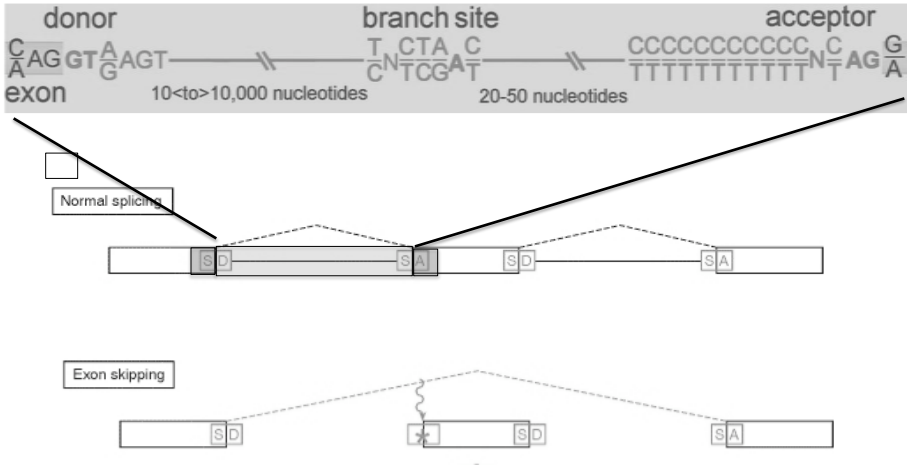


Silent Substitutions at Synonymous Positions

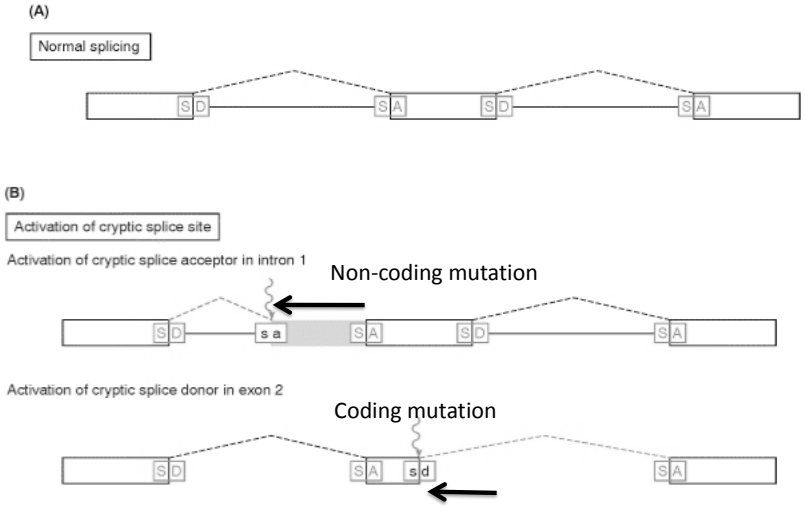
Second Letter

First Letter	T	C	A	G	*Third Letter*
T	TTT, TTC Phe; TTA, TTG Leu	TCT, TCC, TCA, TCG Ser	TAT, TAC Tyr; TAA Stop; TAG Stop	TGT, TGC Cys; TGA Stop; TGG Trp	T C A G
C	CTT, CTC, CTA, CTG Leu	CCT, CCC, CCA, CCG Pro	CAT, CAC His; CAA, CAG Gln	CGT, CGC, CGA, CGG Arg	T C A G
A	ATT, ATC, ATA Ile; ATG Met	ACT, ACC, ACA, ACG Thr	AAT, AAC Asn; AAA, AAG Lys	AGT, AGC Ser; AGA, AGG Arg	T C A G
G	GTT, GTC, GTA, GTG Val	GCT, GCC, GCA, GCG Ala	GAT, GAC Asp; GAA, GAG Glu	GGT, GGC, GGA, GGG Gly	T C A G

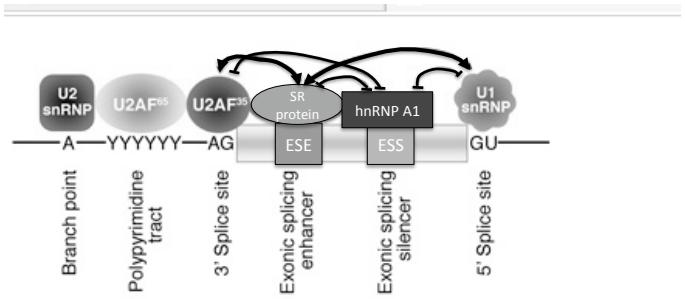
Splicing Mutations



Ectopic Splice Site Mutations



Exonic Splicing Regulators



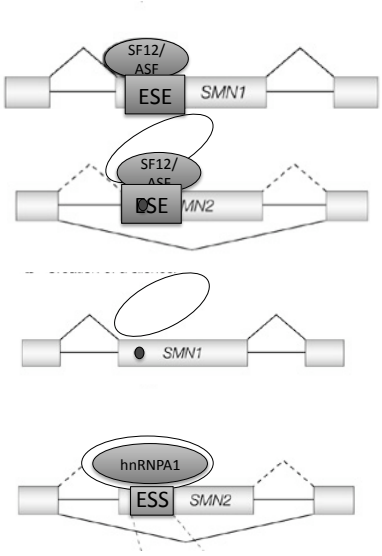
Synonymous Substitutions Affect Splicing

GGU UUC AGA CAA AAU CAA
G F R Q N Q

GGU UUU AGA CAA AAU CAA
G F R Q N Q

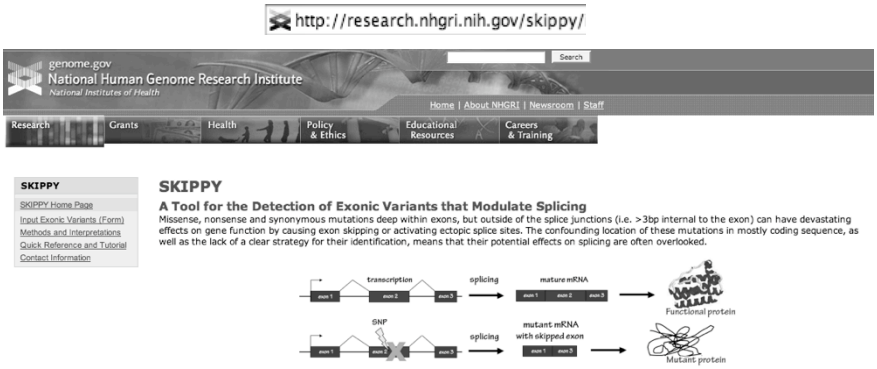
GGU UUC AGA CAA AAU CAA
G F R Q N Q

GGU UUU AGA CAA AAU CAA
G F R Q N Q



The fact that synonymous substitutions in coding sequences could interrupt regulatory processes implies that re-sequencing projects might be ignoring the most critical variants

Predicting / Evaluating splicing mutations

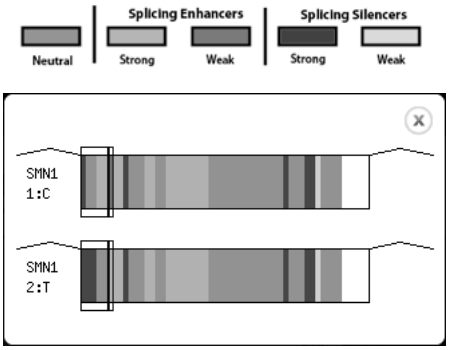


Genomic features defining exonic variants that modulate splicing

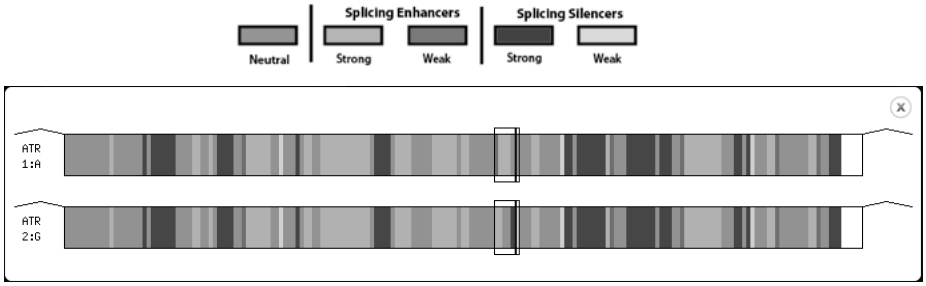
Adam Woolfe ✉, **James C Mullikin** ✉ and **Laura Elnitski** ✉

Genome Biology 2010, **11**:R20 doi:10.1186/gb-2010-11-2-r20

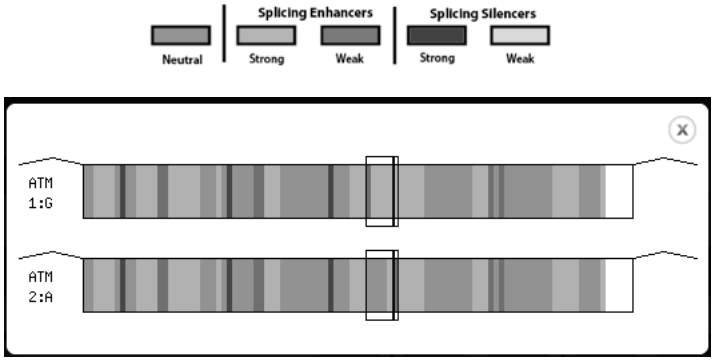
Predicting / Evaluating splicing mutations



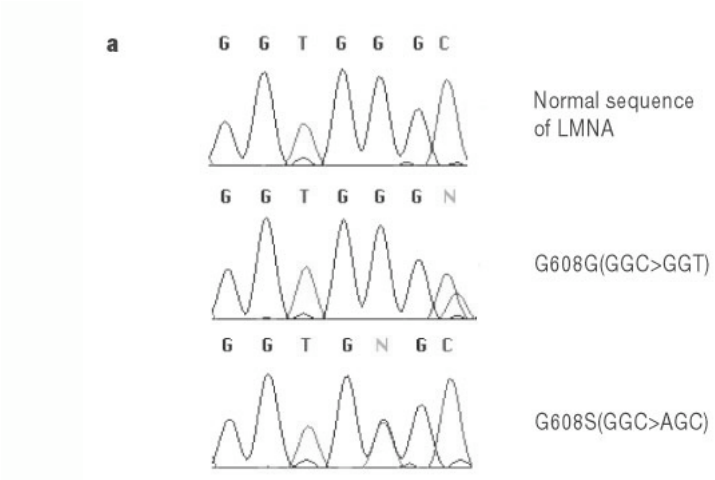
Predicting / Evaluating splicing mutations



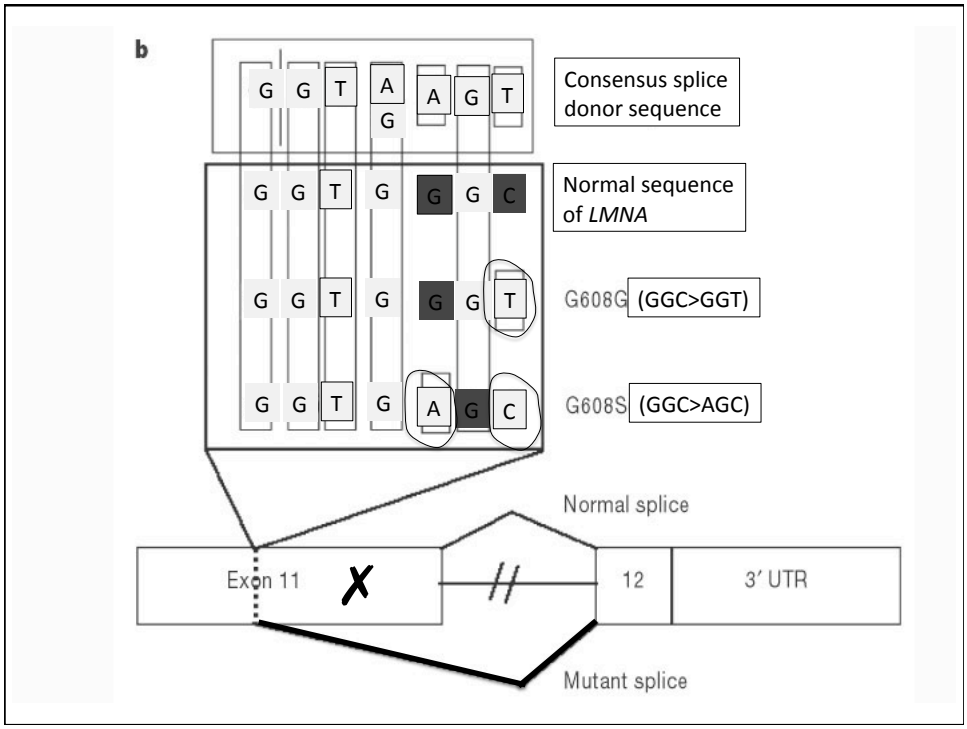
Predicting / Evaluating splicing mutations



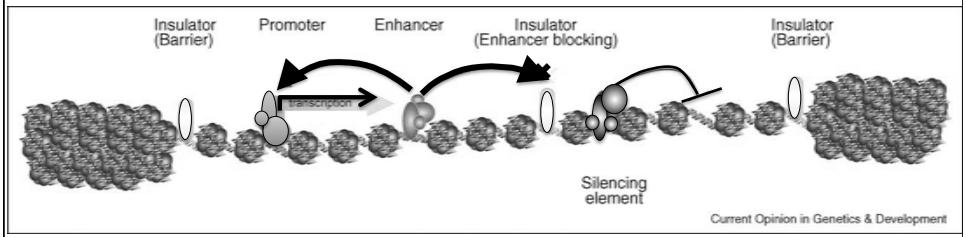
Splicing Mutations in Progeria



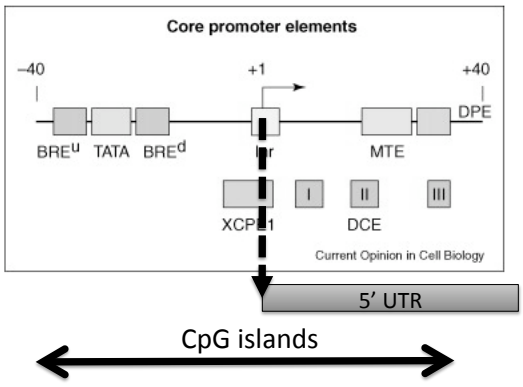
Eriksson, Brown, Gordon, Glynn, Singer, Scott, Erdos, Robbins, Moses, Berglund, Dutra, Pak, Durkin, Csoka, Boehnke, Glover and F.S. Collins Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. Nature 423, 293-298(15 May 2003)



Non-coding Regulatory Landscape



Promoter Types

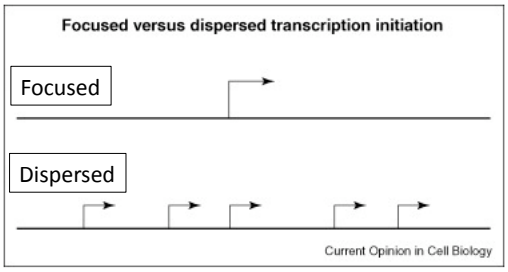


CpG islands

Promoter Characteristics

"Narrow Peak"

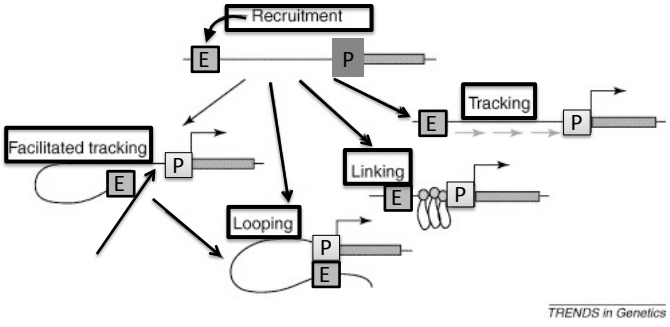
"Broad"



"TATA box"

"CpG island"

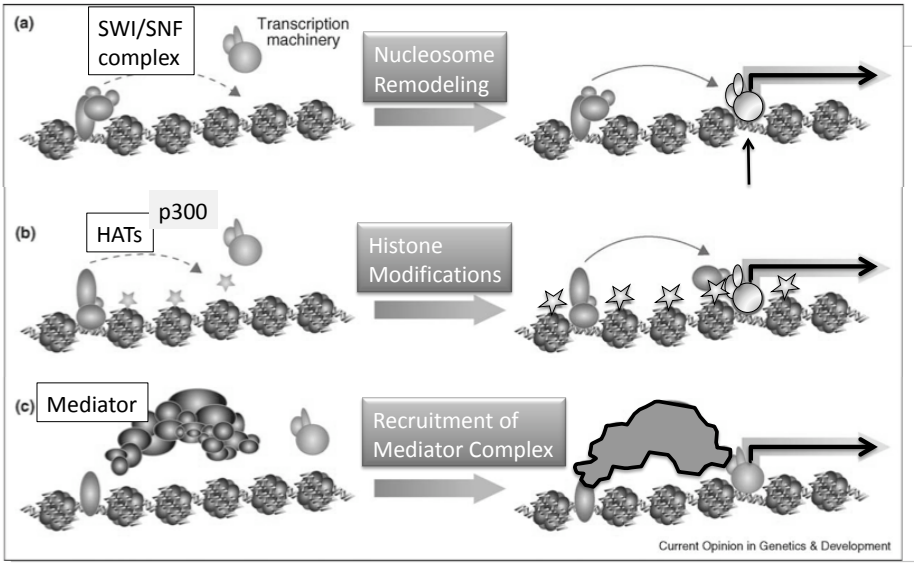
Intra-chromosomal Interactions



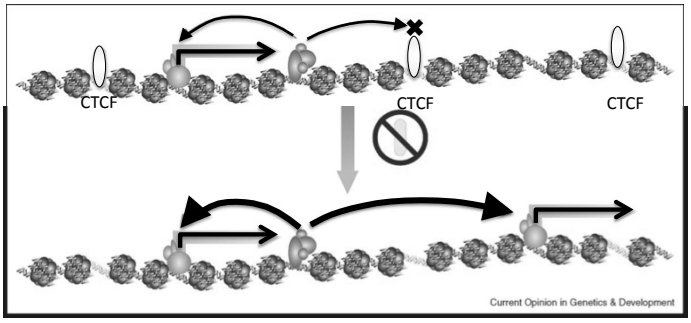
Ann Dean, 2005

-Yet another model - ratcheting a gene through an immobilized transcription factory

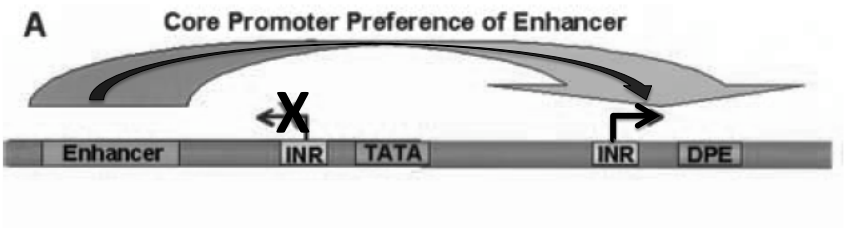
Enhancer Function



Enhancer Blocking

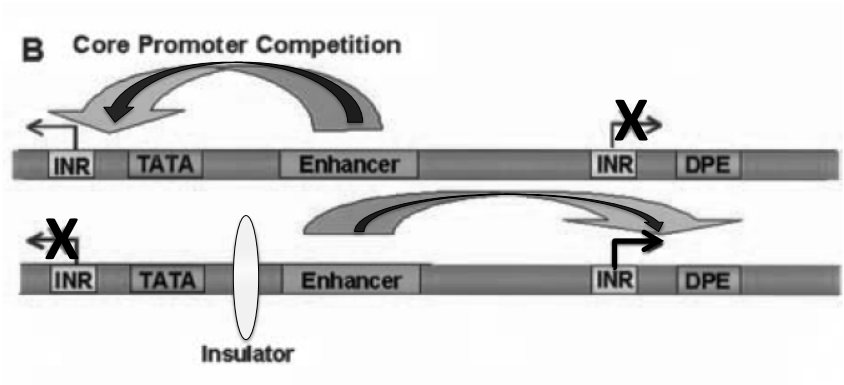


Promoters Contribute to Combinatorial Regulation



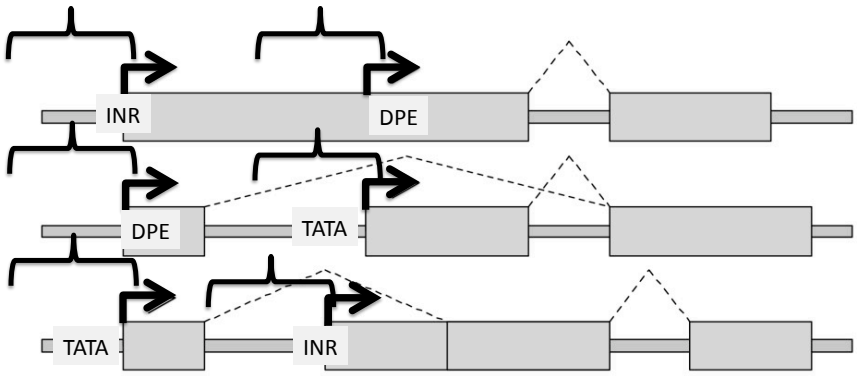
Smale S T Genes Dev. 2001;15:2503-2508

Promoters Contribute to Combinatorial Regulation



Smale S T Genes Dev. 2001;15:2503-2508

Alternative Promoters

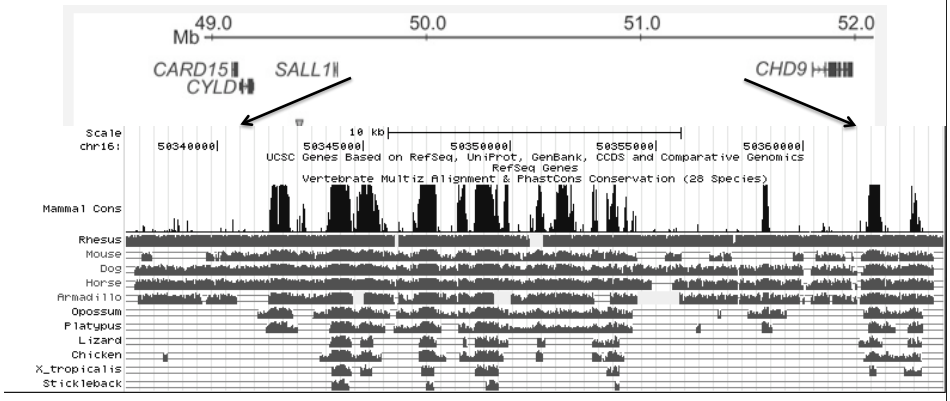


Alternative promoters use core elements to contribute specificity

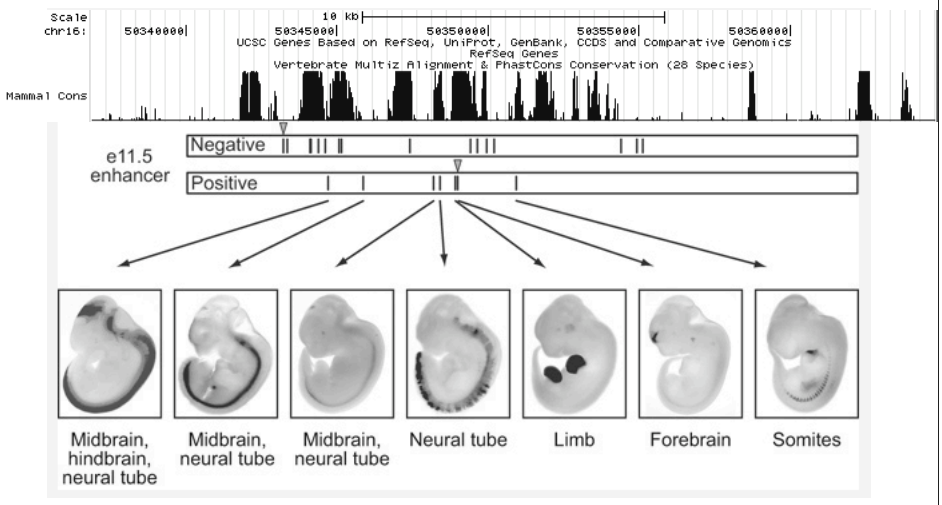
The preference of particular enhancer - promoter combinations implies inherent specificity of interactions that could be used for predictive purposes

Conservation Helps to Identify Enhancers

Conservation Identifies Enhancers

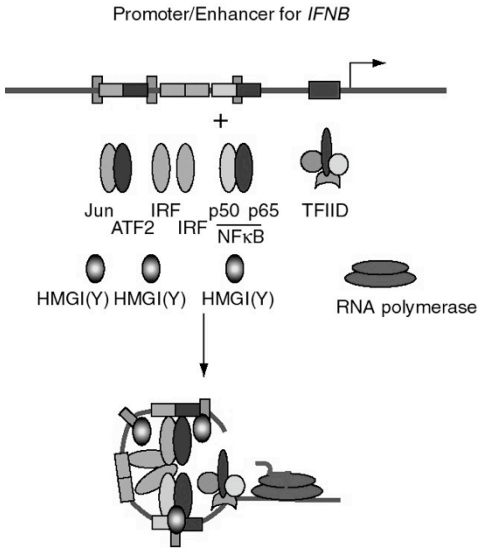


Conservation Identifies Enhancers

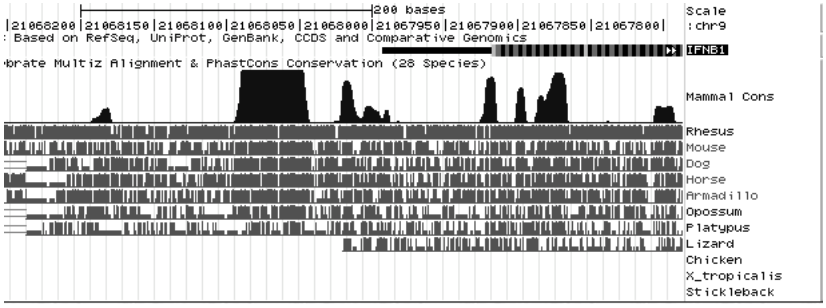


In vivo enhancer analysis of human conserved non-coding sequences. Pennacchio et al. Nature 444, 499-502.

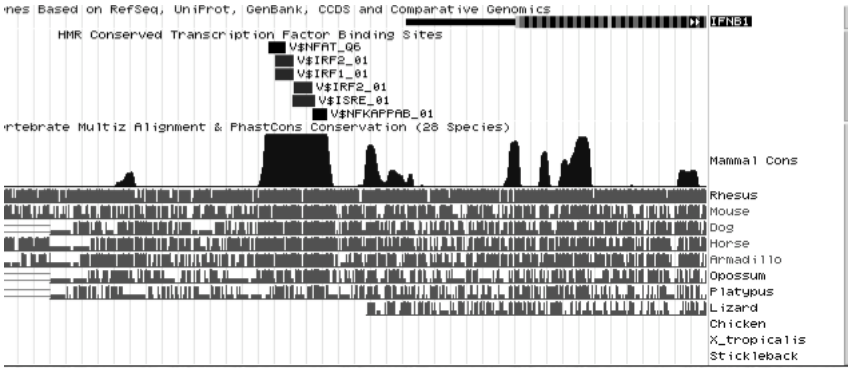
Interferon Beta Enhancer-Promoter



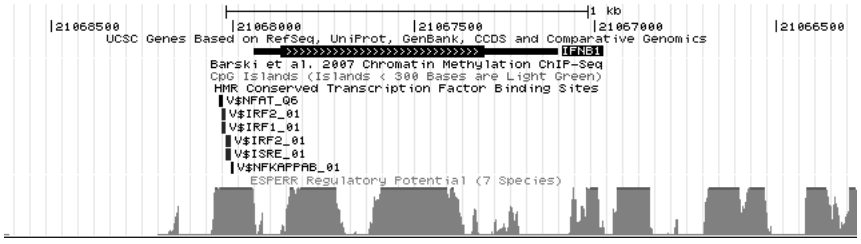
Interferon Beta Enhancer-Promoter



Interferon Beta Enhancer-Promoter



Interferon Beta Enhancer-Promoter

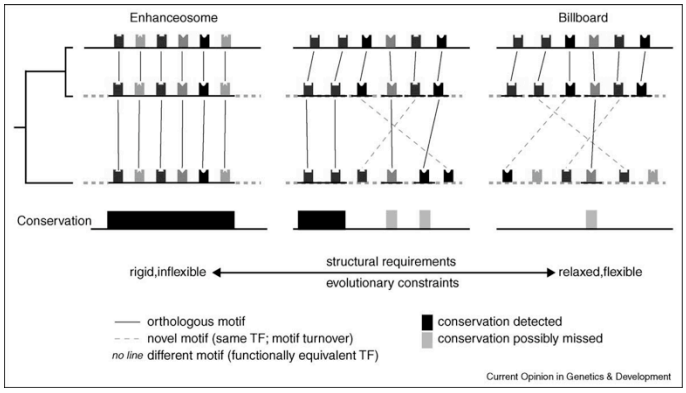


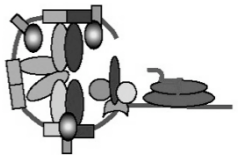
ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements.
Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F.
Genome Res. 2006 Dec;16(12):1596-604. Epub 2006 Oct 19.

Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat.
Kolbe D, Taylor J, Elnitski L, Eswara P, Li J, Miller W, Hardison R, Chiaromonte F.
Genome Res. 2004 Apr;14(4):700-7.

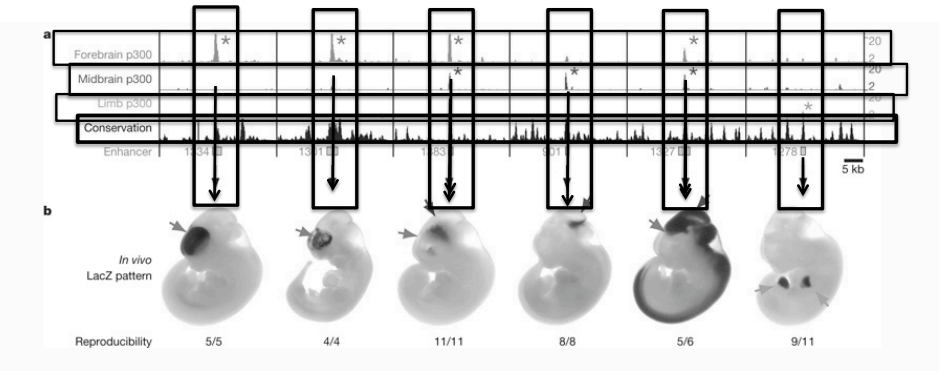
Distinguishing regulatory DNA from neutral sites.
Elnitski L, Hardison RC, Li J, Yang S, Kolbe D, Eswara P, O'Connor MJ, Schwartz S, Miller W, Chiaromonte F.
Genome Res. 2003 Jan;13(1):64-72.

Conservation & Divergence of Enhancers





Enhancer discovery



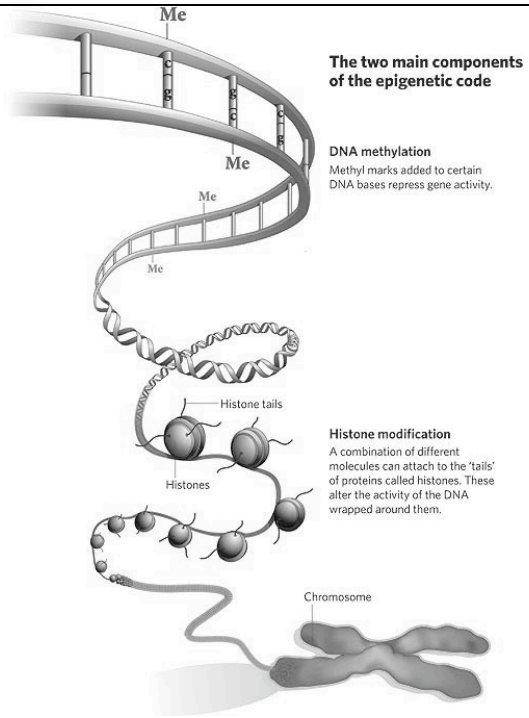
A Visel *et al. Nature* **457**, 854-858 (2009)

From genomes to epigenomes

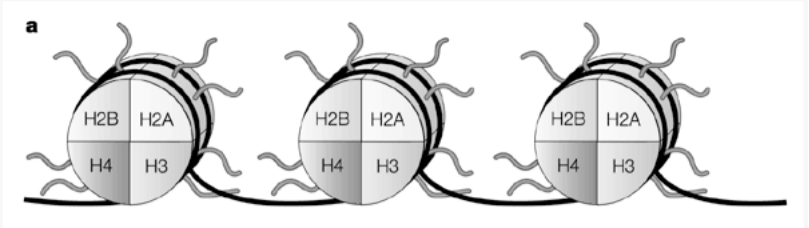
Epigenetics

Epigenetics commonly refers to the study of mitotically and/ or meiotically heritable changes in gene function that are not attributable to a change in DNA sequence.

An 'epigenome' is a representation of all epigenetic phenomena across the genome

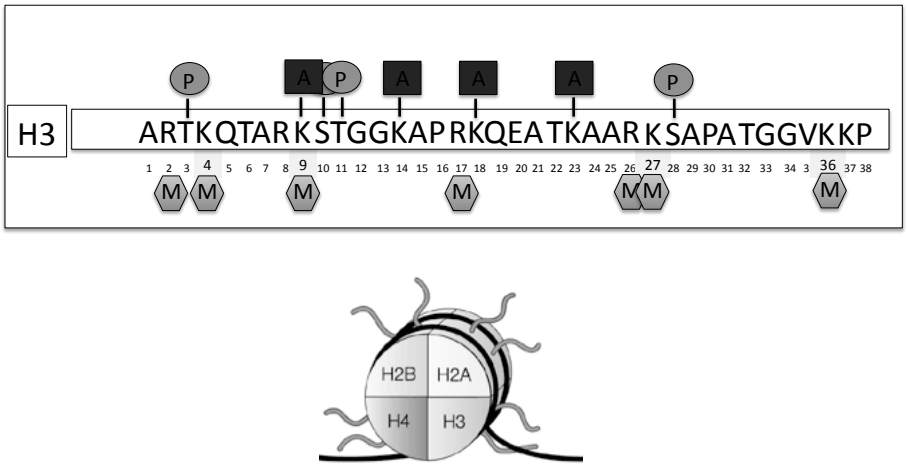


Covalent Histone Modifications

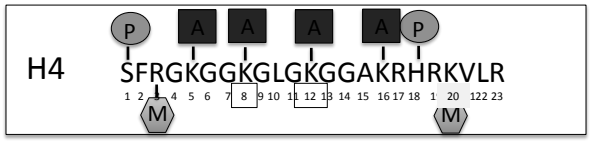


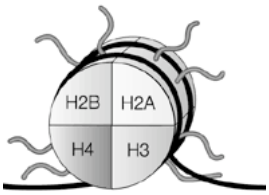
Histone deacetylases and cancer: causes and therapies. Paul A. Marks, Richard A. Rifkind, Victoria M. Richon, Ronald Breslow, Thomas Miller & William K. Kelly
Nature Reviews Cancer 1, 194-202 (December 2001)

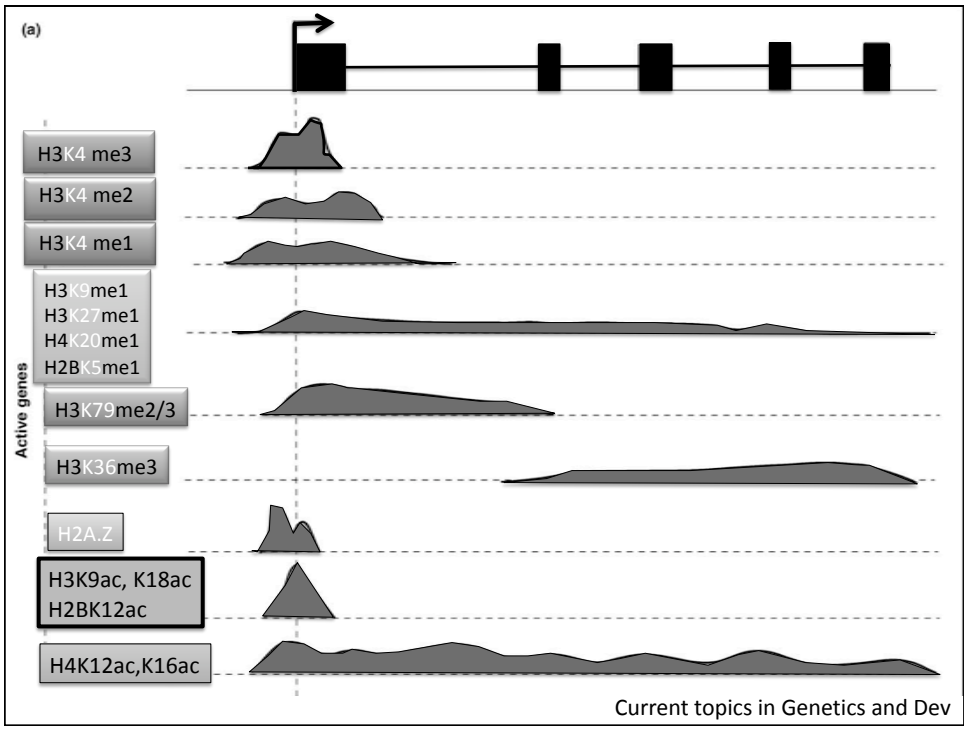
Covalent Histone Modifications



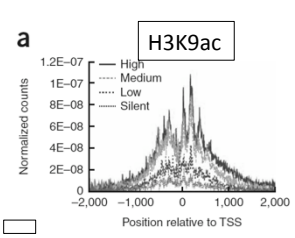
Covalent Histone Modifications

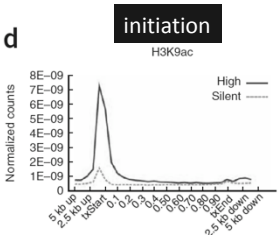






Histone Acetylation Marks

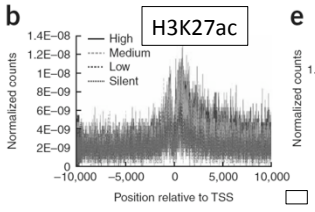


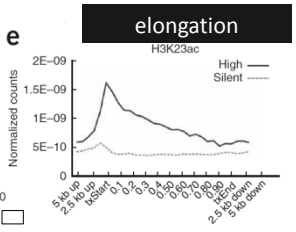


p300

Barski, et al. Cell, Volume 129, Issue 4, 823-837, 18 May 2007

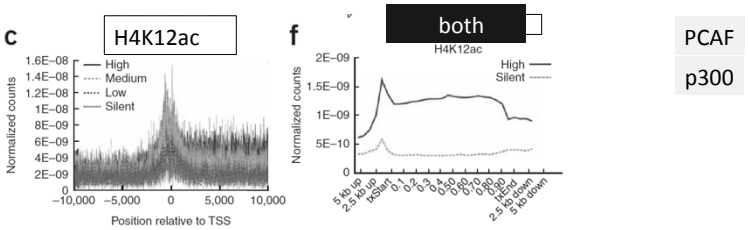
Histone Acetylation Marks





PCAF

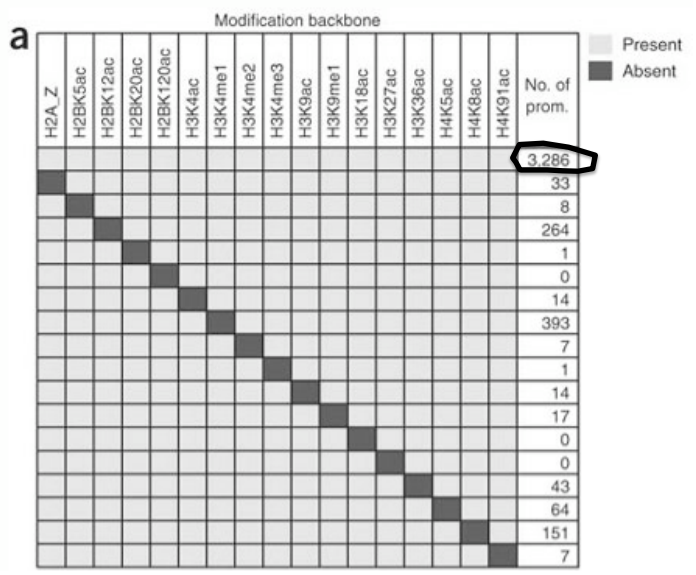
Histone Acetylation Marks



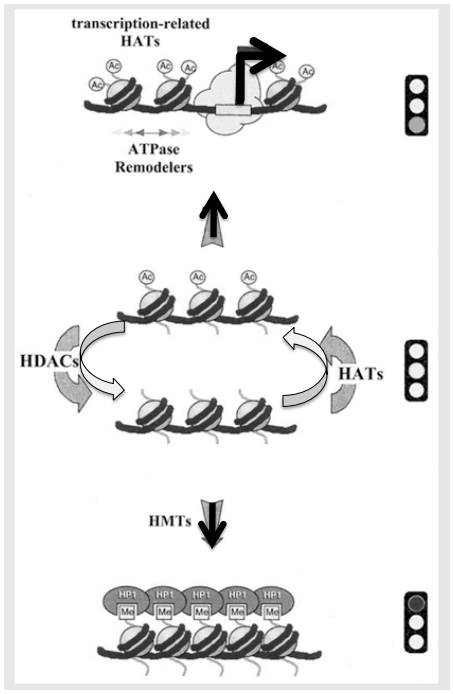
PCAF

p300

Combinatorial patterns at promoters

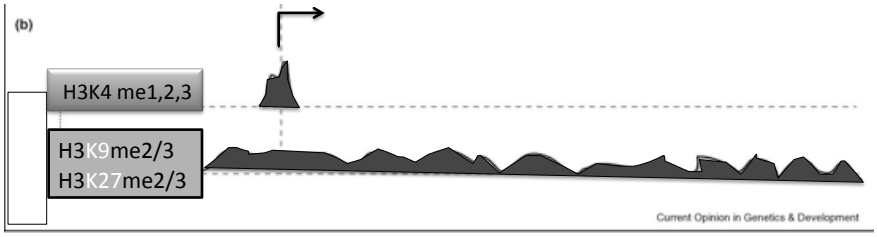


Wang et al. Nature Genetics 40, 897 - 903 (2008)

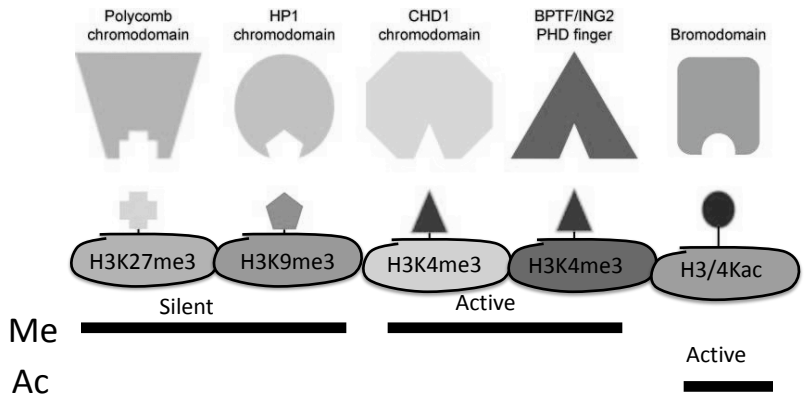


Eberharter & BeckerEMBO
3, 3, 224–229 (2002)

Inactive Genes

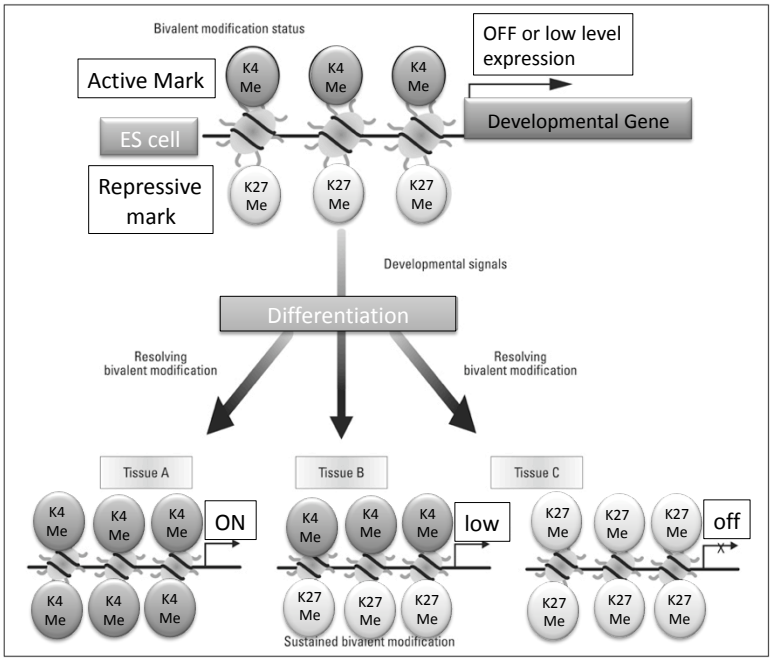


Translator Proteins



Higgs DR, et al. 2007.
Annu. Rev. Genomics Hum. Genet. 8:299–325

The fact that every cell type has a unique pattern of histone modifications attributable to the functioning of that cell implies that changes in those patterns could reveal disease processes



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

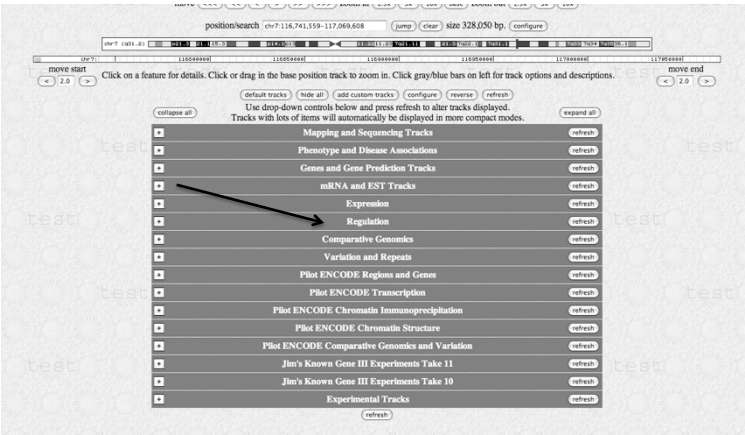
Histone Modification Summary

(1) The H3K27me3 modification appears to be dominant because all patterns containing this modification tend to be repressive;

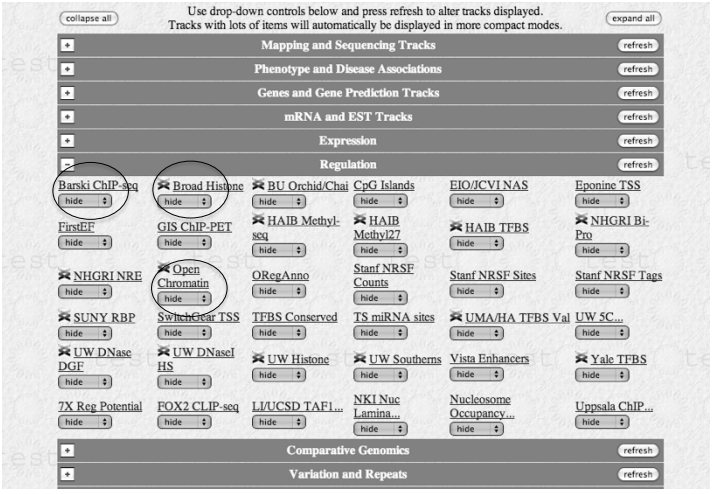
(2) The H3K4me3 modification alone is not sufficient to support active transcription because the genes associated with H3K4me3 alone tend to be silent.

(3) The histone modification pattern alone does not determine the expression level; genes associated with many patterns show an extremely broad range of expression from silent to active.

UCSC Browser Data

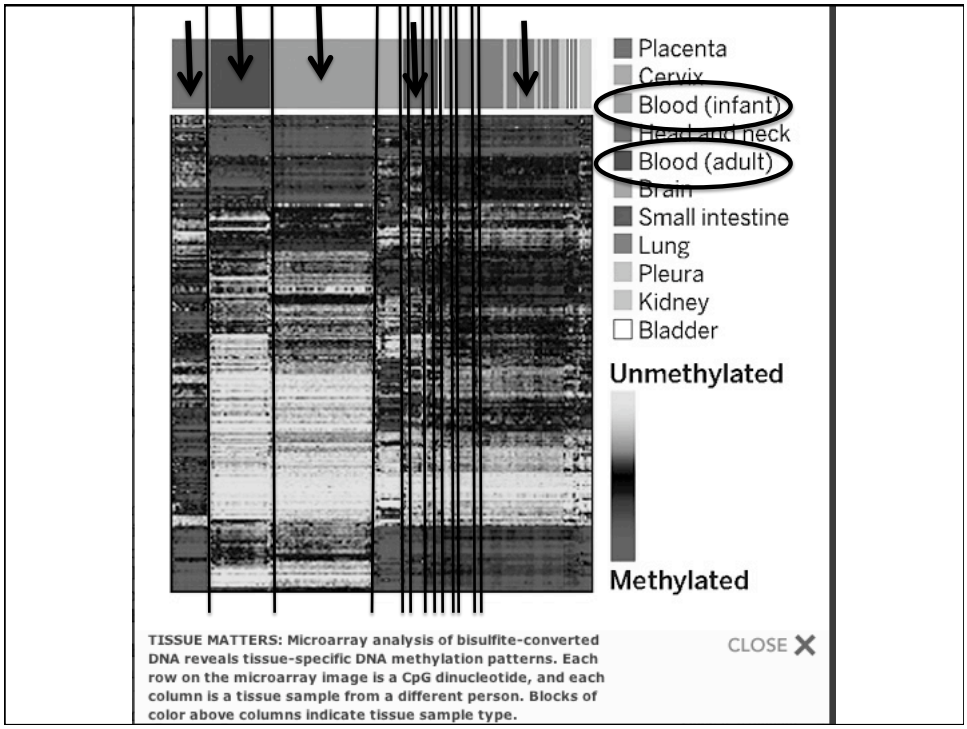


UCSC Browser Data

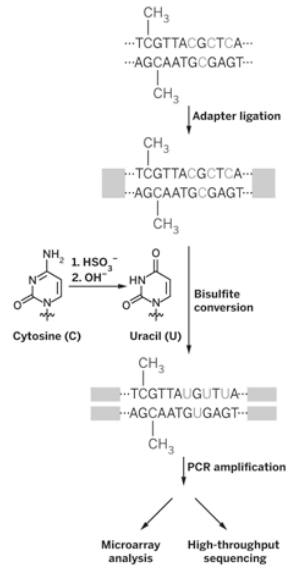


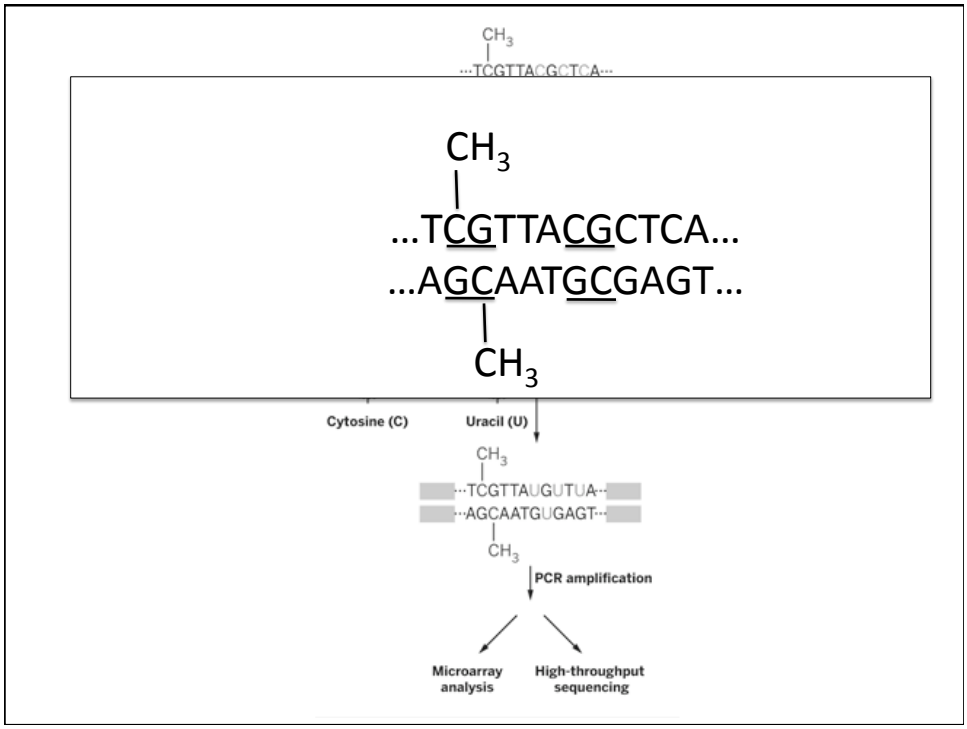
The regulatory involvement of histone modifications implies that many functional regions could be detected by these signals alone, if collected in the correct cell type.

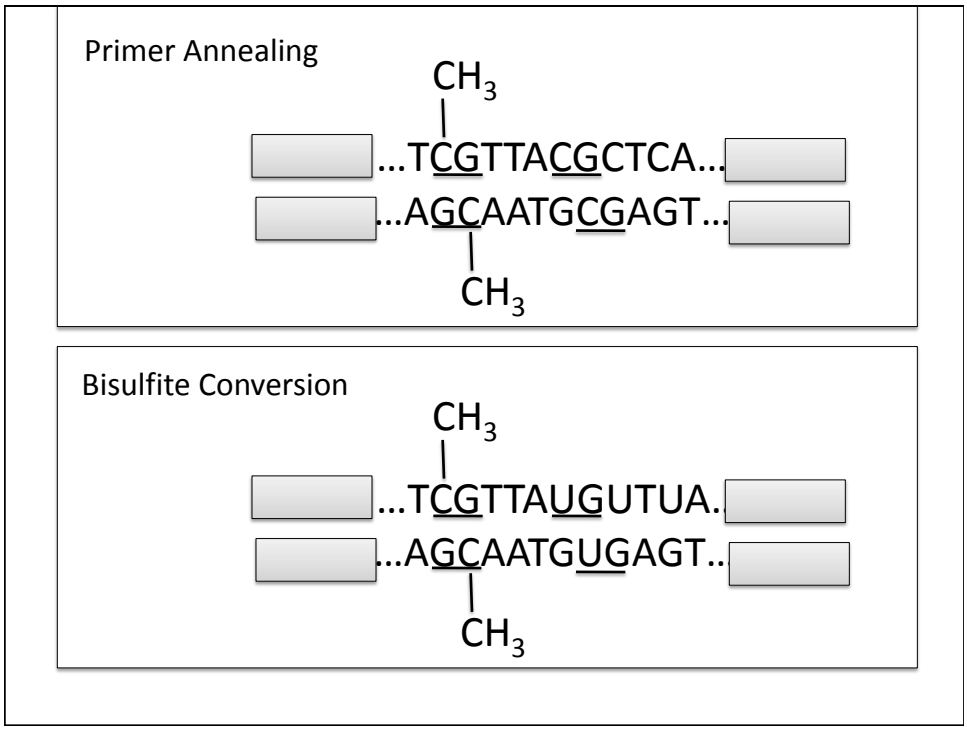
Epigenetics: DNA Methylation

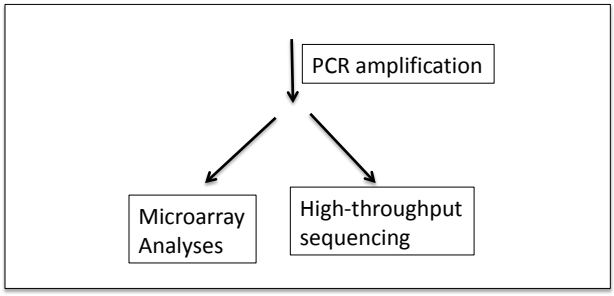


Detecting DNA methylation

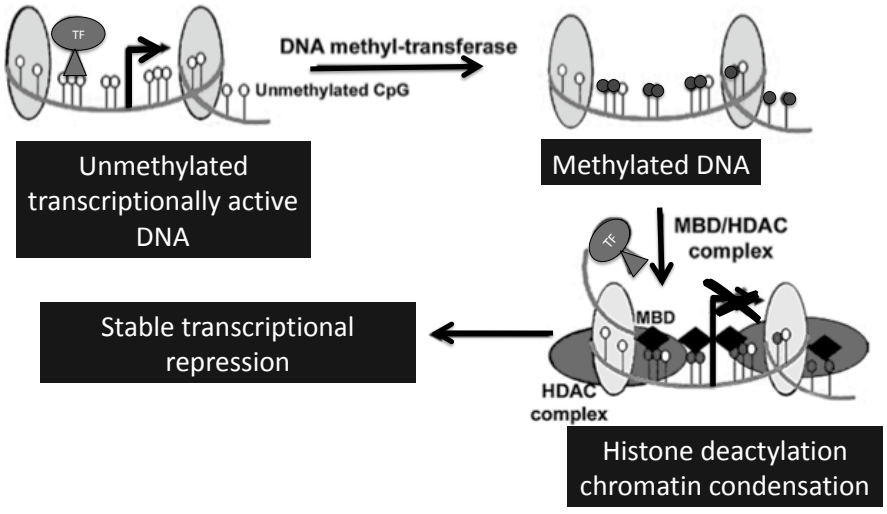



CH3
…TCGTTACGCTCA…
…AGCAATGCGAGT…
CH3
Cytosine (C)
Uracil (U)
CH3
…TCGTTAUGUTUA…
…AGCAATGUGAGT…
CH3
PCR amplification
Microarray analysis
High-throughput sequencing


Primer Annealing
CH3
…TCGTTACGCTCA…
…AGCAATGCGAGT…
CH3
Bisulfite Conversion
CH3
…TCGTTAUGUTUA.
…AGCAATGUGAGT..
CH3

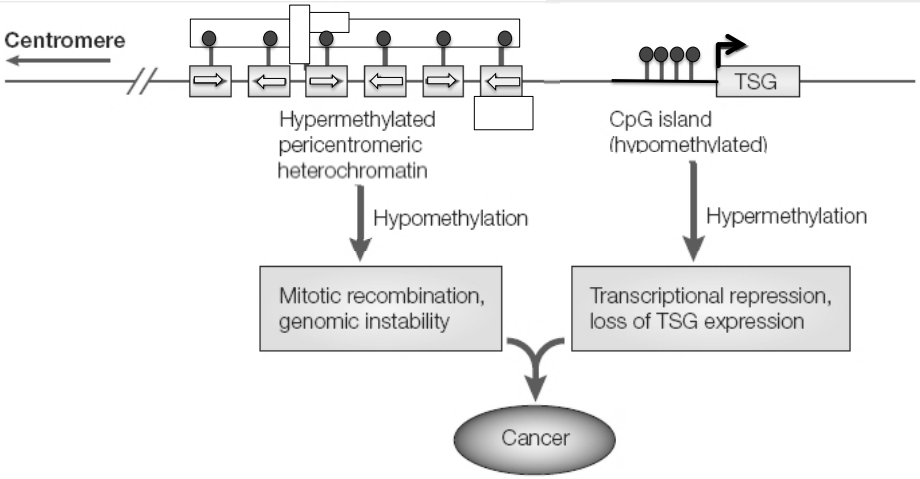


Histone modification and DNA methylation

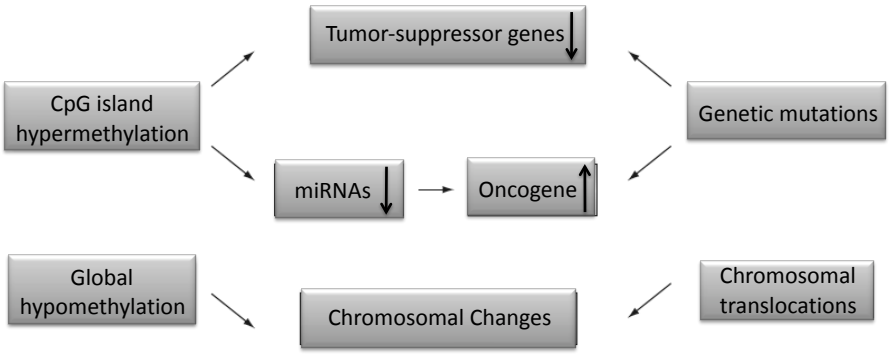


Int. braz j urol. vol.33 no.1 Rio de Janeiro Jan./Feb. 2007 Murugesan ManoharanI,II; Kavitha RamachandranII; Mark S. SolowayI,II; Rakesh SingalII,III

Methylation and Cancer



Integrated Biological Systems

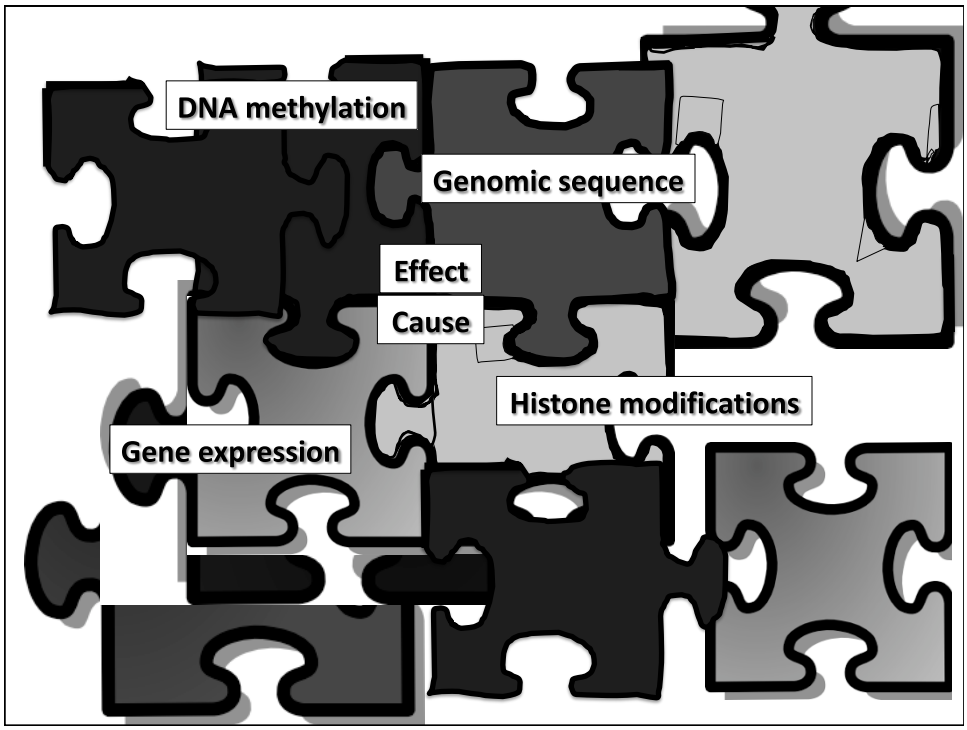


Robertson, K., DNA methylation and human disease, Nature Reviews Genetics 6, 597-561

Why do CpG islands become methylated in cancer?

Why do certain CpG islands become methylated while others do not?

Is aberrant hypermethylation a targeted or a random process?



Traveling The Pathway to Genomic Medicine

Healthcare tailored to the individual based on genomic information

The Pathway to Genomic Medicine

1000 Genomes
A Deep Catalog of Human Genetic Variation

NCGC
NIH Chemical Genomics Center

The ENCODE Project: ENCyclopedia Of DNA Elements

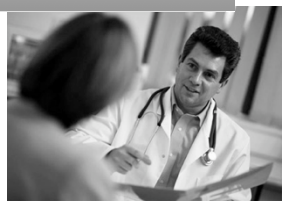
- Overview
- Consortium Membership
- Data Release Policy
- Accessing ENCODE Data

THE CANCER GENOME ATLAS

Interdisciplinary Research

HGP

Characterize
Diagnose
Treat



Realization of Genomic Medicine

Targeted re-sequencing
Whole genome re-sequencing
High-throughput genotyping

"Fulfilling the Promise"