NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Current Topics in Genome Analysis
Spring 2008

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

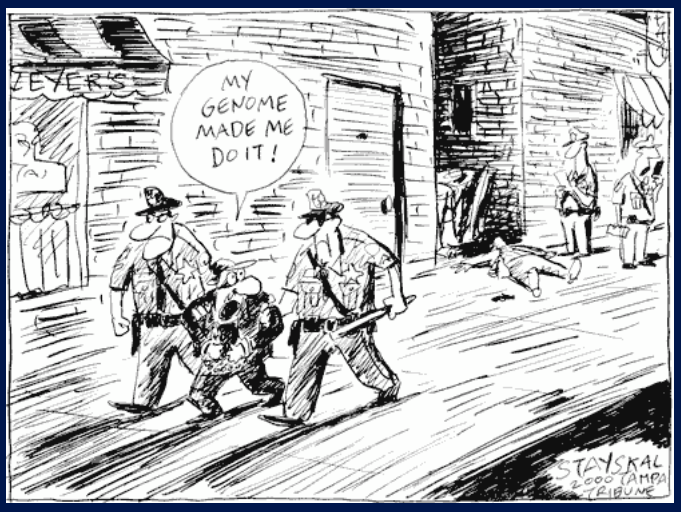
Outline

I. Global regulatory organization
II. Techniques for assessing chromosomal interactions
III. Functional elements
IV. Pattern searching in the genome
V. Epigenomics
VI. Genome methylation
VII. The landscape of regulatory mutations

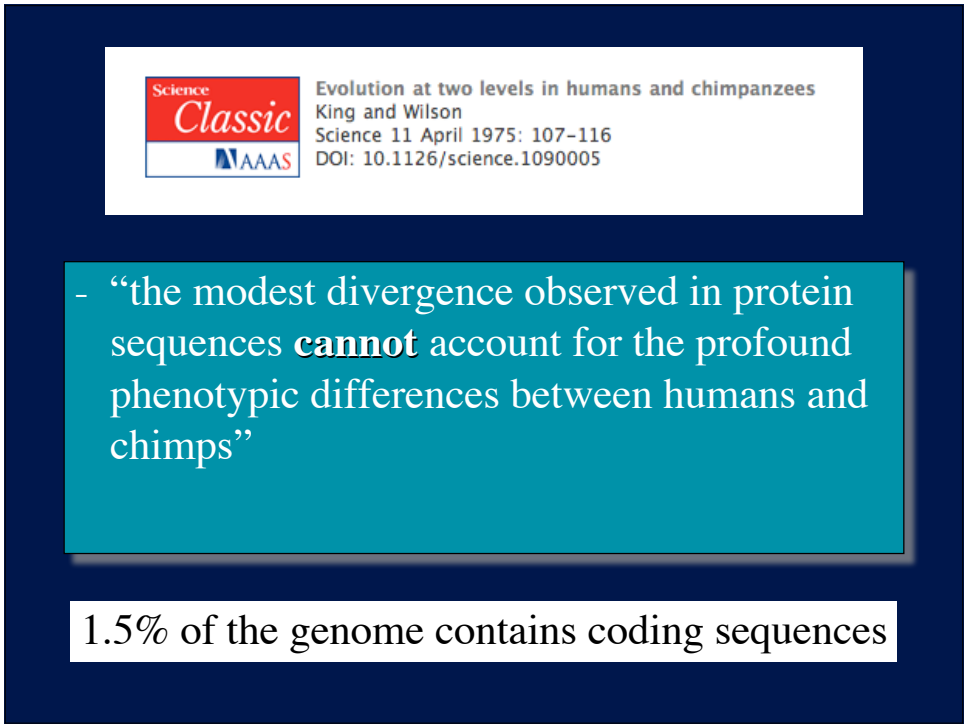
Outline

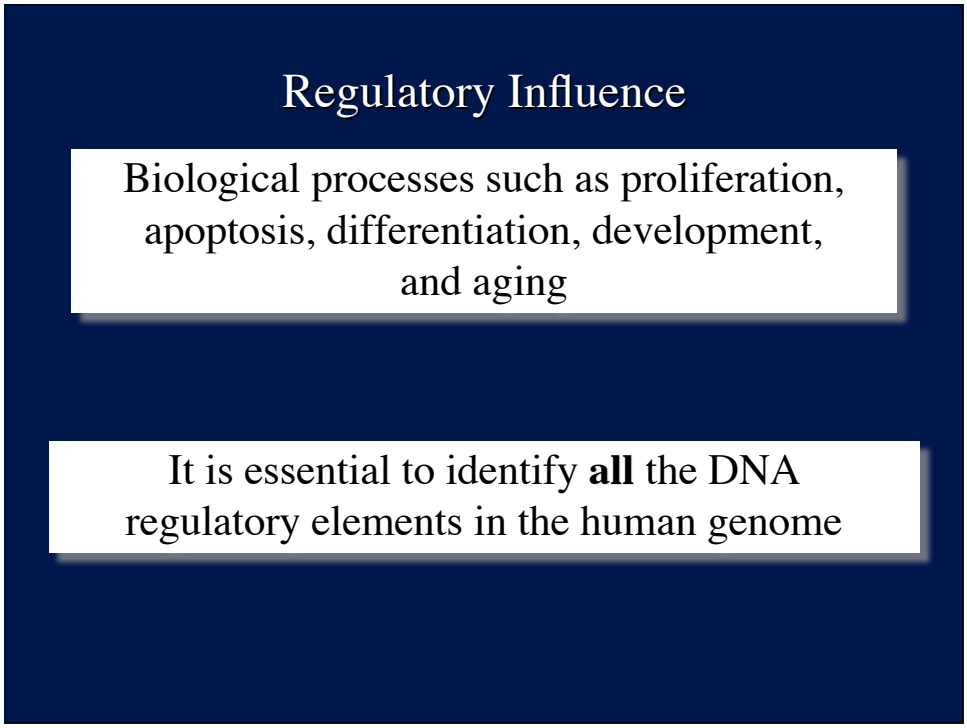
I. Global regulatory organization
II. Techniques for assessing chromosomal interactions
III. Functional elements
IV. Pattern searching in the genome
V. Epigenomics
VI. Genome methylation
VII. The landscape of regulatory mutations

Elusive Genomic Attributes

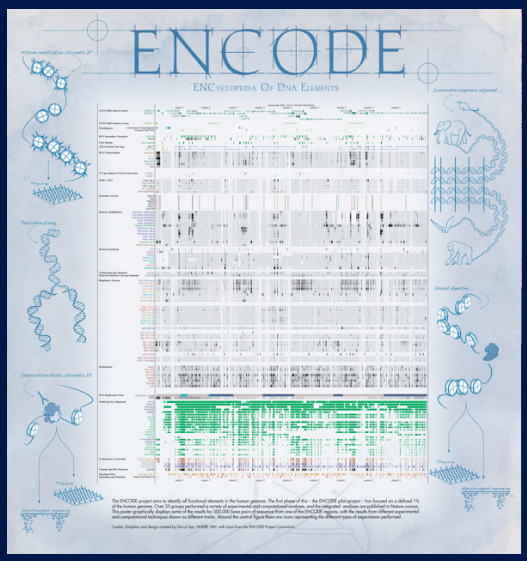


•Physical Traits •Illnesses •Behaviors


Science Classic
AAAS
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"
1.5% of the genome contains coding sequences


Regulatory Influence
Biological processes such as proliferation, apoptosis, differentiation, development, and aging
It is essential to identify **all** the DNA regulatory elements in the human genome

Deductive Reasoning



Attribute 1

Attribute 2

Attribute 3

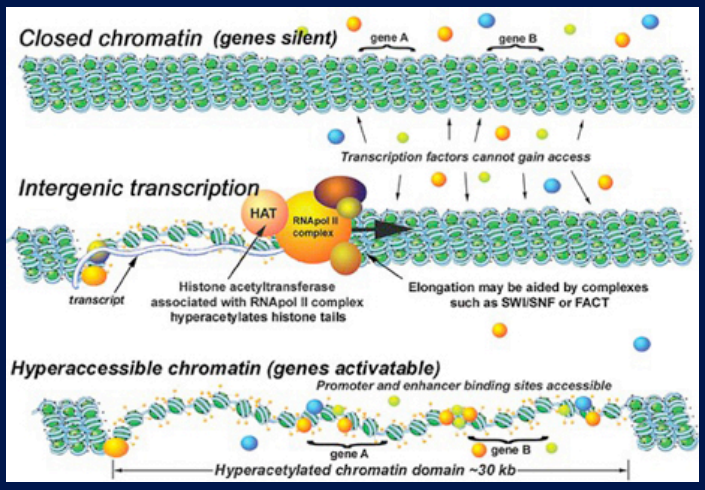
Functional sites

Often presented as static images …

… are dynamic processes within the cell

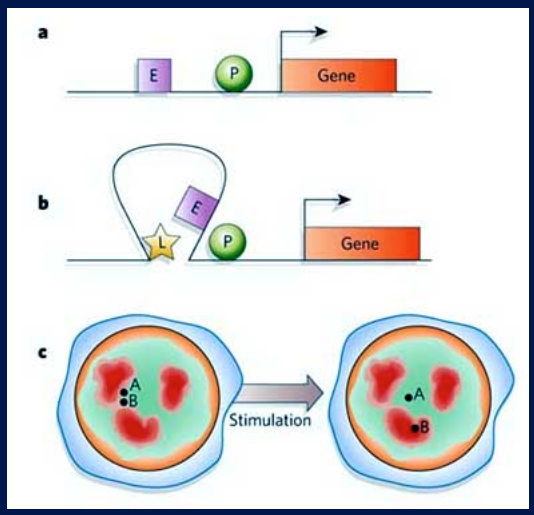
I. Global regulatory organization

Linear Concept of Regulation



Peter Fraser

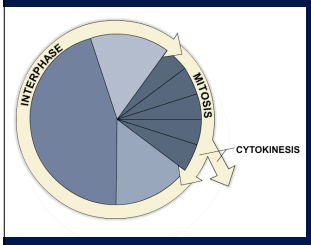
3-Dimensional Regulation

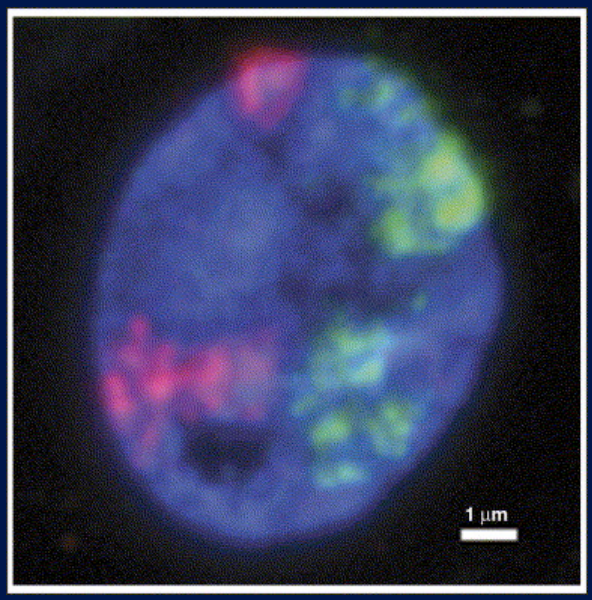


Kioussis,Nature (2005)

Chromosome Territories

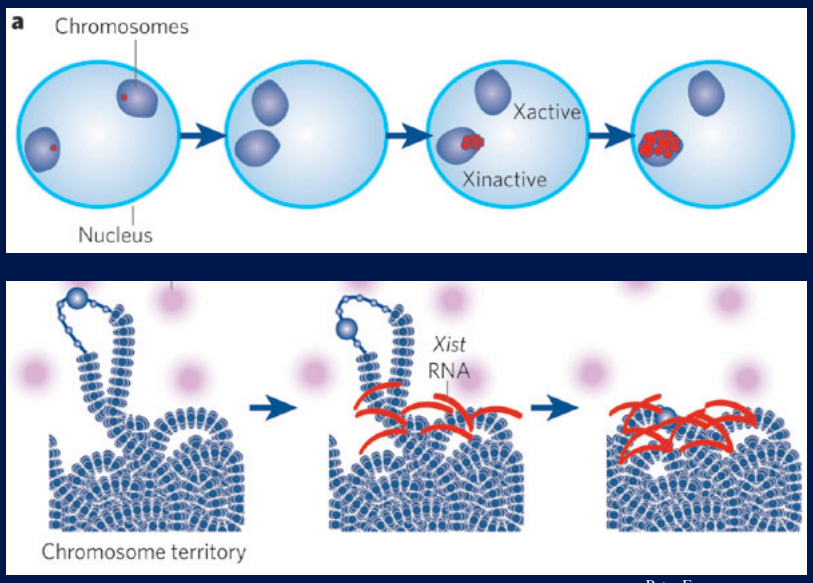
Chromosome 1 = green
Chromosome 2 = red





Branco and Pombo, 2007

Expression Fate



Peter Fraser

Transcription Factories

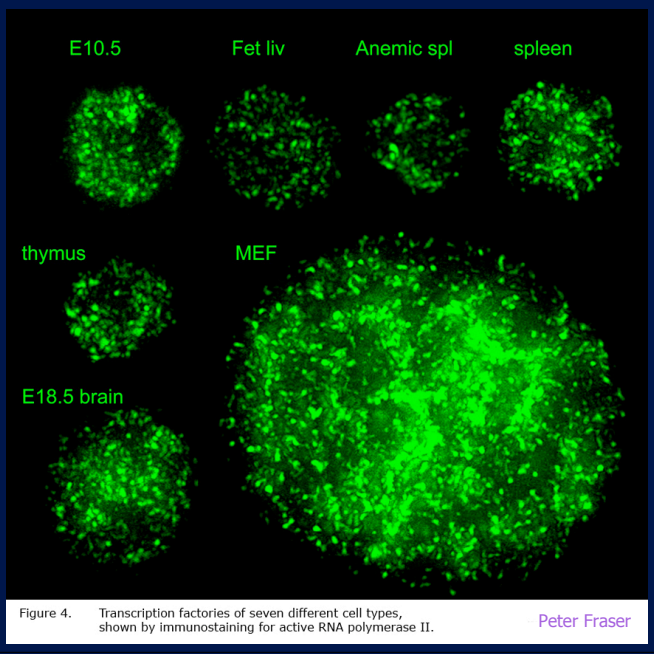
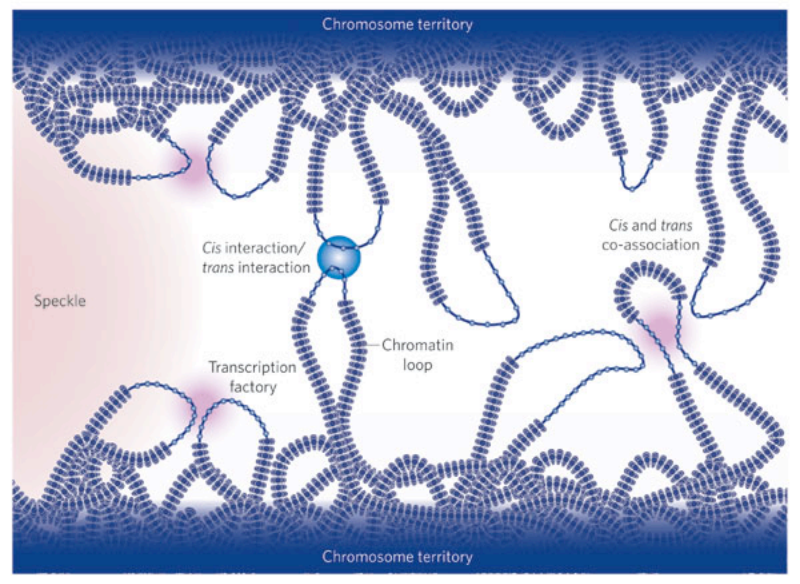


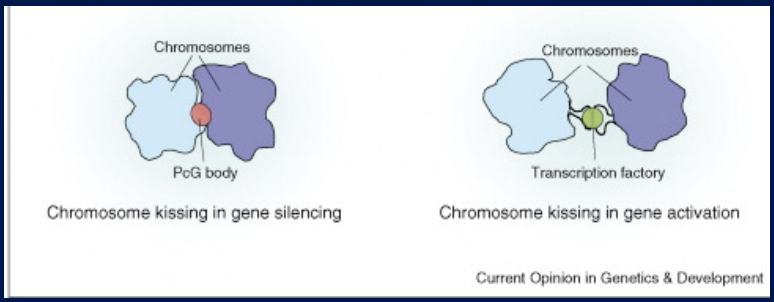
Figure 4. Transcription factories of seven different cell types, shown by immunostaining for active RNA polymerase II.

Peter Fraser

Interchromosomal Interactions



Chromosome Kissing



X-inactivation

T-lymphocyte **activation**

Types of Chromatin

Heterochromatin- a tightly packed form of DNA, aggregates at the periphery of the interphase nucleus

- **Constitutive heterochromatin**
- **Facultative heterochromatin**
- **Euchromatin**

Types of Chromatin

Constitutive heterochromatin

- stable during all stages of development and in all tissues centromeres, telomeres (and pericentromerically)
- tandemly repeated sequences
- gene-poor
- late-replicating

Types of Chromatin

Facultative heterochromatin

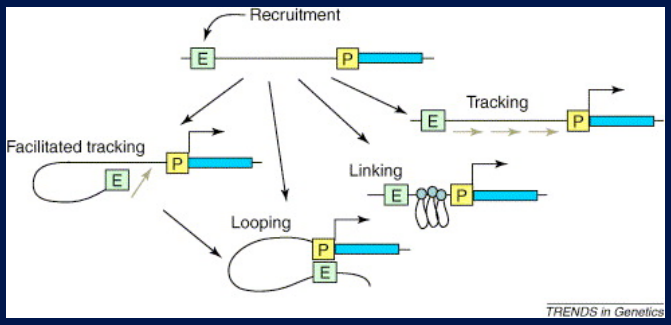
- reversible
- depends on the stage of development or cell type
- The inactive X chromosome
- relatively poor in genes
- these genes are usually not transcribed

Types of Chromatin

Euchromatin

- lightly stained appearance reflecting its less compact structure
- condensed during mitosis
- gene-rich
- often active transcribed
- early-replicating

Intrachromosomal Interactions

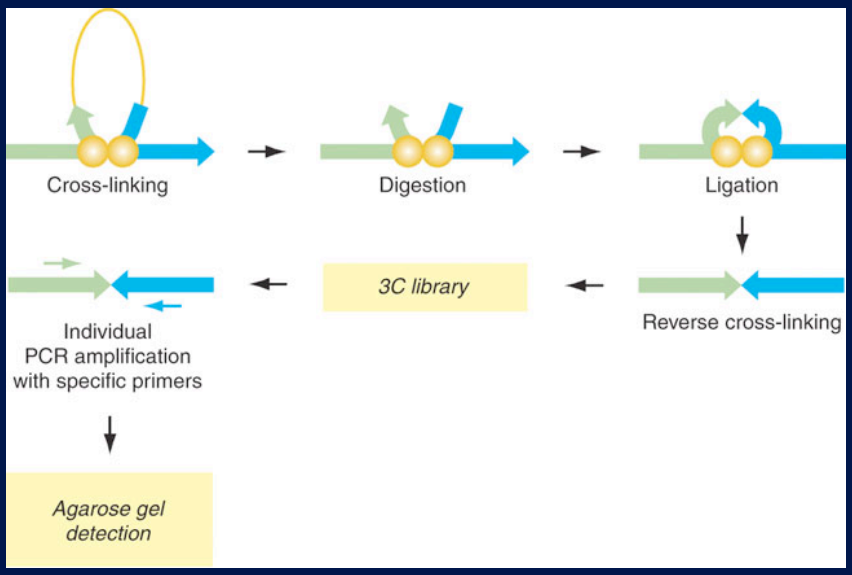


Ann Dean, 2005

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

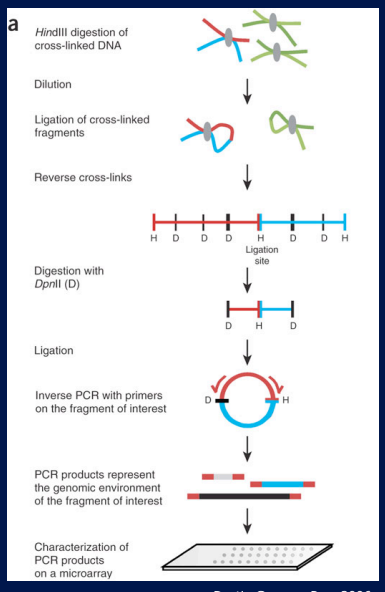
II. Techniques for assessing chromosomal interactions

Chromosome conformation capture (3C)



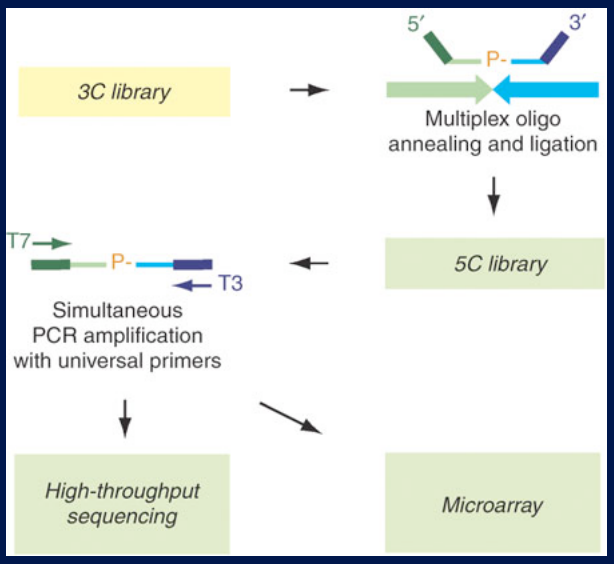
Nature Genetics - 38, 1348 - 1354 (2006)

4C Experimental strategy

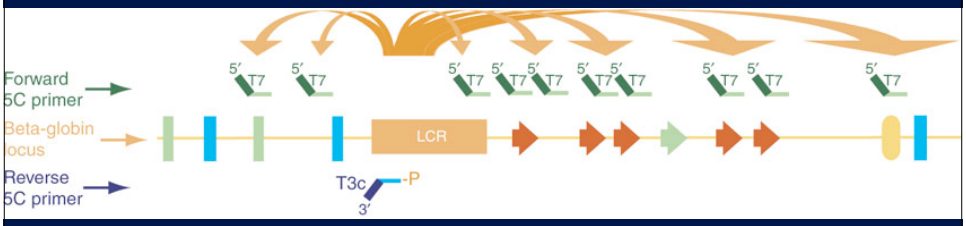


Dostie Genome Res. 2006

Chromosome conf. capture carbon copy (5C)



Globin-domain distance interactions



III. Functional Elements

Boundary Elements (I)

Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites

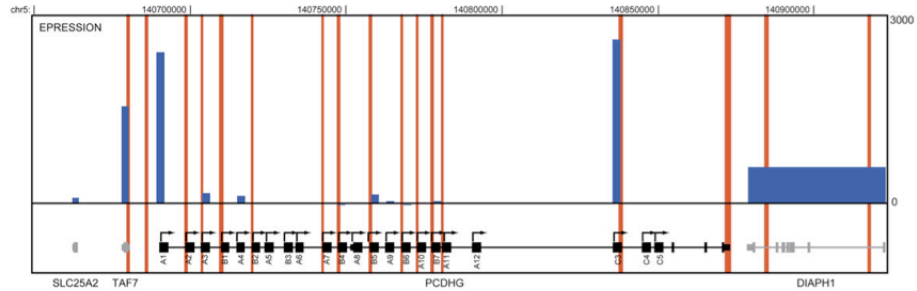
Xiaohui Xie[†], Tarjei S. Mikkelsen[†‡], Andreas Gnirke[†], Kerstin Lindblad-Toh[†], Manolis Kellis[†§], and Eric S. Lander[†¶‖††]

ID	Motif profile	No. of conserved instances
LM1		5,332
LM2		7,549

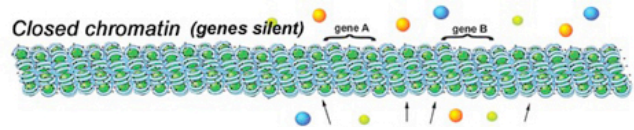
Boundary Elements (II)

Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome

Tae Hoon Kim,[1,5] Ziedulla K. Abdullaev,[2] Andrew D. Smith,[3] Keith A. Ching,[1] Dmitri I. Loukinov,[2] Roland D. Green,[4] Michael Q. Zhang,[3] Victor V. Lobanenkov,[2] and Bing Ren[1]



DNase I Hypersensitivity



Transcription factors cannot gain access

Useful for finding functional regions in a given cell type
- includes all types of functional elements
- represents removal or modification of histones

OPEN ACCESS Freely available online PLOS GENETICS

Identification and Characterization of Cell Type–Specific and Ubiquitous Chromatin Regulatory Structures in the Human Genome

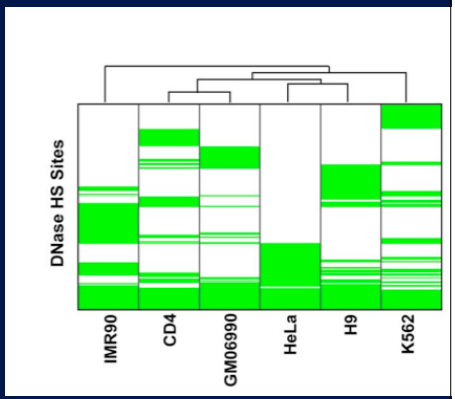
Hualin Xi[1], Hennady P. Shulha[2], Jane M. Lin[2], Teresa R. Vales[3], Yutao Fu[1], David M. Bodine[4], Ronald D. G. McKay[5], Josh G. Chenoweth[5], Paul J. Tesar[5], Terrence S. Furey[3], Bing Ren[6], Zhiping Weng[1,2*], Gregory E. Crawford[3*]

- On average for each cell type:
  - 32% are **cell type specific**
  - 46% are **common**
  - 22% are **ubiquitous**

HS sites

DNase I hypersensitive sites
- 22% are ubiquitously present
- 86% near TSS
- 10% bound by CTCF

Cell type-specific sites
- enriched for enhancer elements
- enriched for cell-type specific features & nucleosome modifications



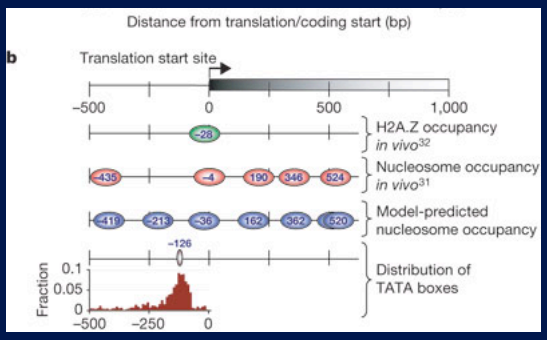
Mapping in all cell types will be important to find cell-type specific regulatory elements

Table 2. Tissue-Specific DNaseI HS Sites Are Enriched in Motifs I

Cell Type	TF Motif Group
CD4	TAL1 (T-cell acute lymphocytic leukemia)
	[25,26], E2A, E12, AP-4, or Lmo2 complex
	ETS family factors
GM06990	Lmo2 complex, Ebox, E12, or E47
	IPF1
	NF-1
H9 ES	Octamer [29] or Oct-1
	Sp-1, KROX, or VDR
	STAT1, STAT3, STAT6, or TEF-1
	SOX-9
K562	GATA [28]
	PR or GR
	GEN_INI
	Tel-2
IMR90	AP-4, Lmo2 complex, myogenin, MyoD, or LBP-1
	STAT3, STAT5A, or Ets
	AP-1
	AR
	ER
	TEF-1

These studies implicate 8% of the genome as being functional

Nucleosome Code



Nature 442, 772-778

~ 50% of *in vivo* nucleosome positioning (in yeast) is governed by an intrinsic organization encoded in the genomic

Nucleosome positioning signals in genomic DNA

Heather E. Peckham,[1,2] Robert E. Thurman,[3] Yutao Fu,[1] John A. Stamatoyannopoulos,[4] William Stafford Noble,[4,5] Kevin Struhl,[6] and Zhiping Weng[1,2,7]

[1]Bioinformatics Program, Boston University, Boston, Massachusetts 02215, USA; [2]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA; [3]Division of Medical Genetics, University of Washington, Seattle, Washington 98195, USA; [4]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA; [5]Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195, USA; [6]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Genomic Sequence Is Highly Predictive of Local Nucleosome Depletion

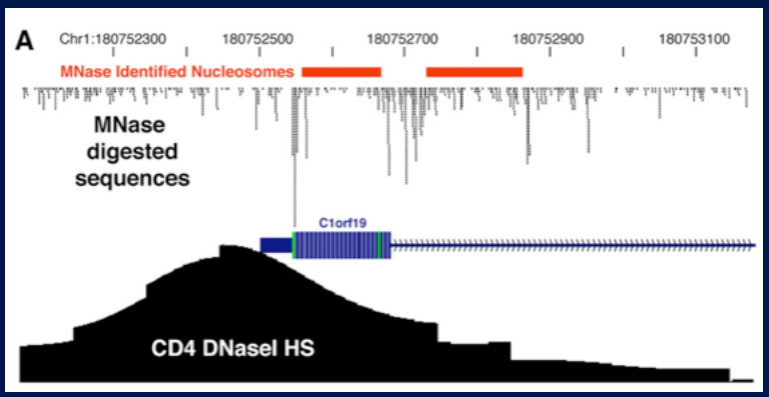
Guo-Cheng Yuan[1,2*], Jun S. Liu[1,3*]

1 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, **2** Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **3** Department of Statistics, Harvard University, Cambridge, Massachusetts, United States of America

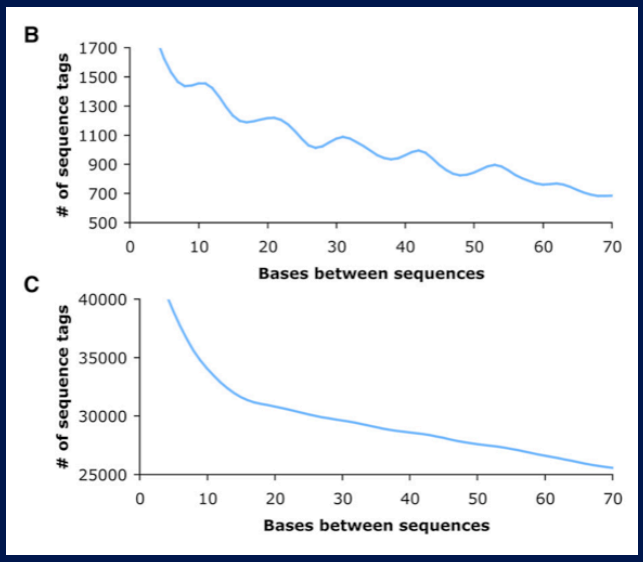
Mononucleosome Data

High-Resolution Mapping and Characterization of Open Chromatin across the Genome

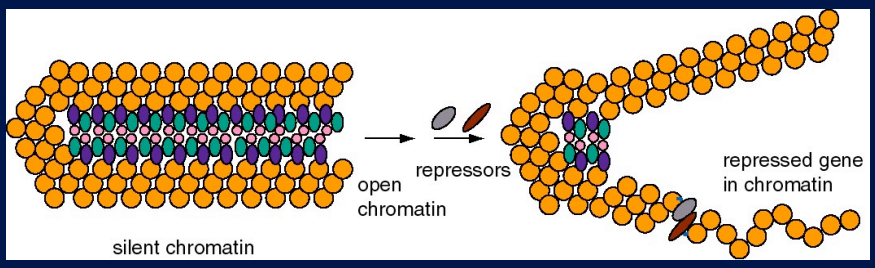
Alan P. Boyle,[1] Sean Davis,[3] Hennady P. Shulha,[2] Paul Meltzer,[3] Elliott H. Margulies,[4] Zhiping Weng,[2] Terrence S. Furey,[1,*] and Gregory E. Crawford[1,*]

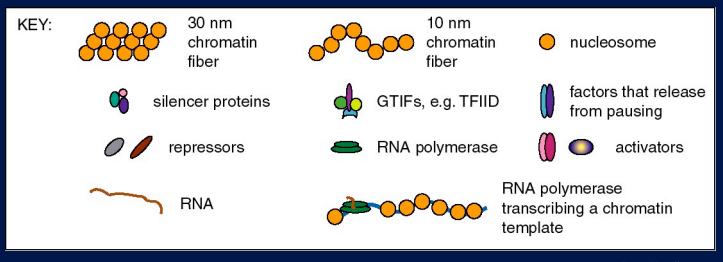


Oscillation patterns



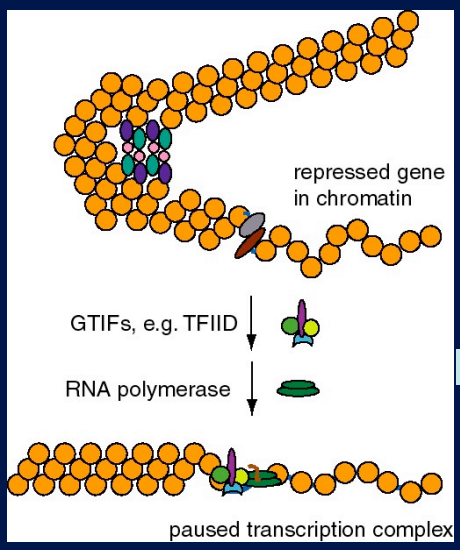
Silent and repressed chromatin





Ross Hardison

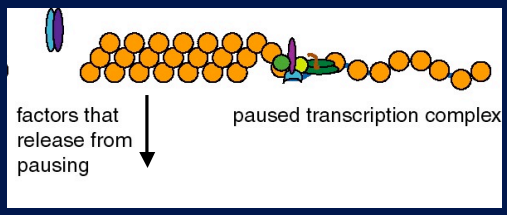
Transcription initiation and pausing

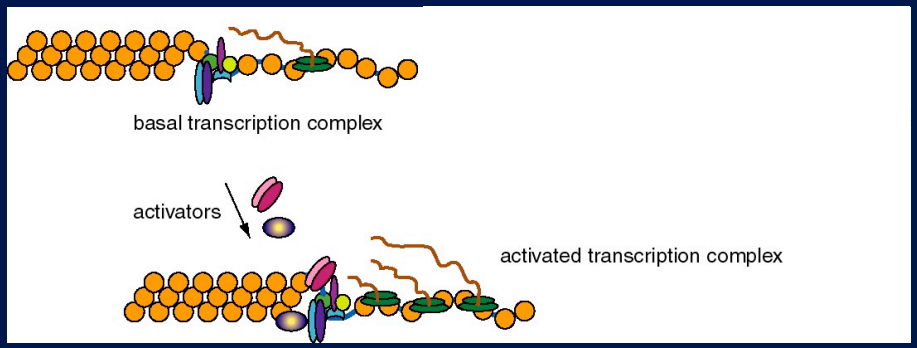


Repressors bind to negative control elements

Assemble on promoter

Basal and activated transcription



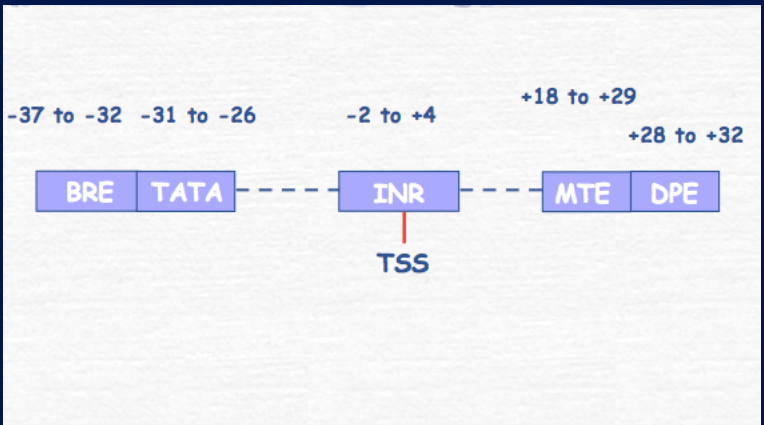


Mapping promoters

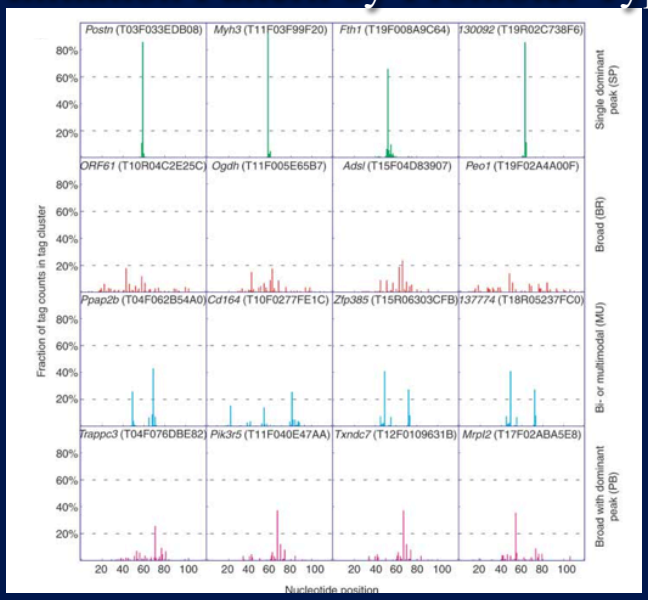
Use collections of mapped transcription start sites (TSSs)

- Categorize by motif composition
- Experimental tests of promoter mechanisms
- Computational identification of new motifs

Core Promoter Elements



Initiation Pattern by Promoter Type



SP

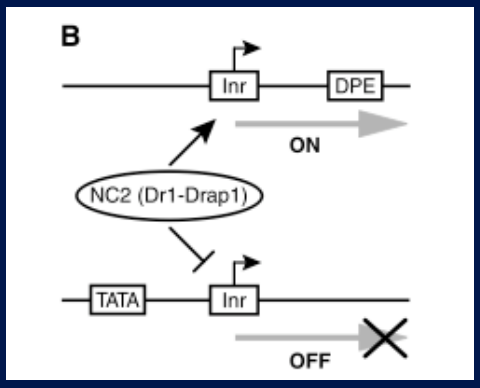
BR

MU

PB

Carninci et al. 2006

Regulatory modes of the PIC



positive

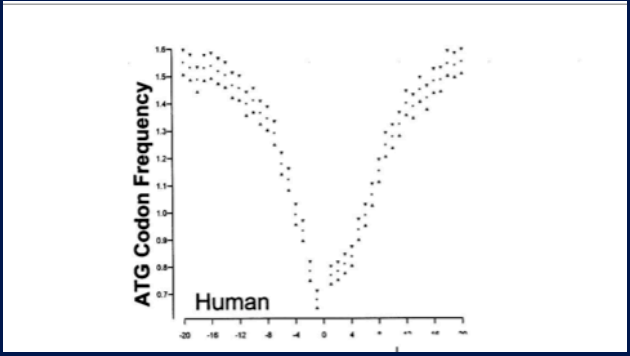
negative

Willy *et al. Science* (2000)

ATG deserts

ATG deserts define a novel core promoter subclass

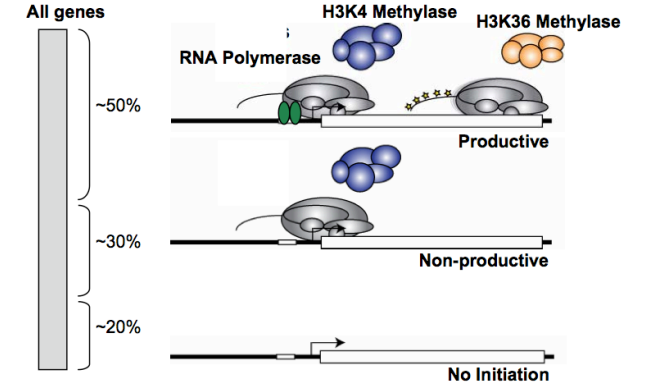
Maxwell P. Lee,[2,3] Kevin Howcroft,[3,4] Aparna Kotekar,[1] Howard H. Yang,[2] Kenneth H. Buetow,[2] Dinah S. Singer[1,5]



PIC Occupancy

A Chromatin Landmark and Transcription Initiation at Most Promoters i Human Cells

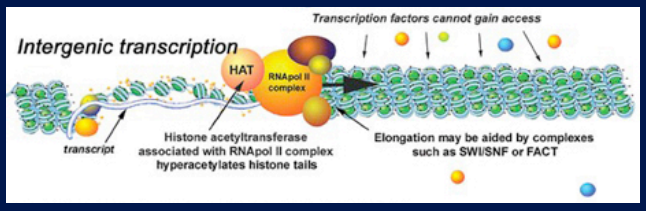
Matthew G. Guenther,[1,3] Stuart S. Levine,[1,3] Laurie A. Boyer,[1] Rudolf Jaenisch,[1] and Richard A. Young[1,2,]



Promoter Summary

Limited number of core promoter motifs
Near transcription start site
DNase hypersensitive
Occupied by PIC *in vivo*
Clusters of binding sites

Intergenic Transcription



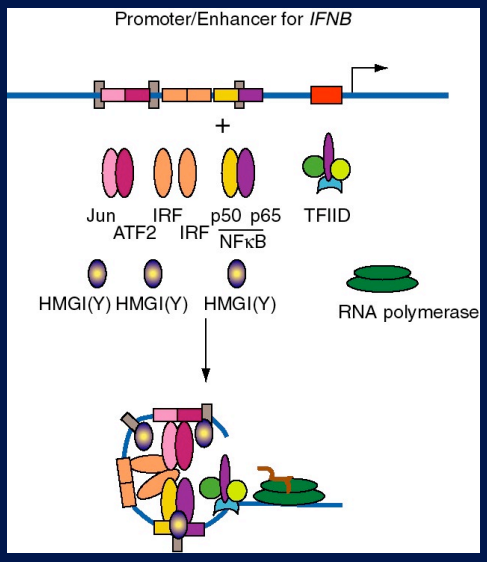
Much of the genome is transcribed
- TARS
- TUFS
- ncRNAS
  - evidence of tracking mechanism?
  - importance of ncRNA?
  - spurious events?

Enhancers

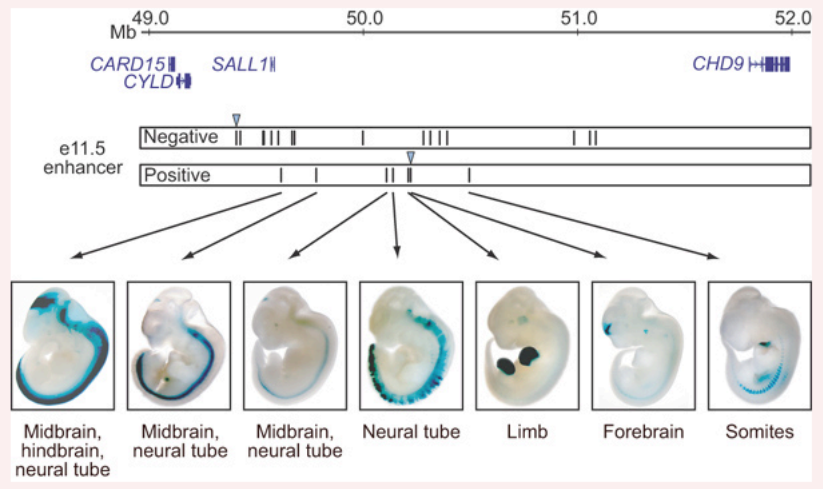
Classically defined as *cis*-acting DNA regulatory elements
stimulate transcription,
act independent of their position and orientation

- often encompass repressive sites
- usually defined by DNA sequences
- function as nucleoprotein complexes
- modify chromatin structures
- interact with components of the basal machinery

Interferon beta Enhancer-Promoter

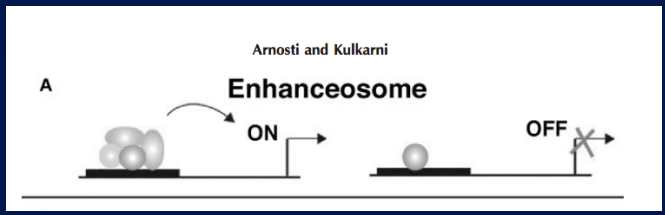


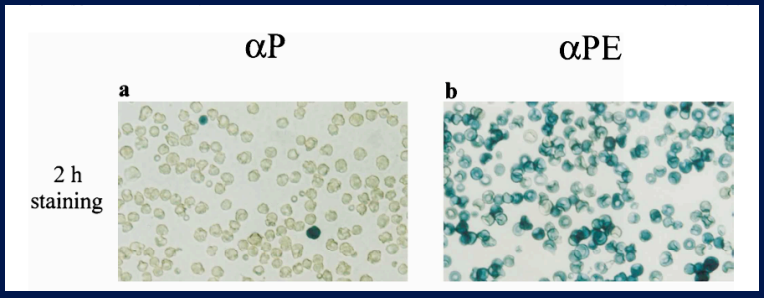
Conservation Identifies Enhancers



Binary Function

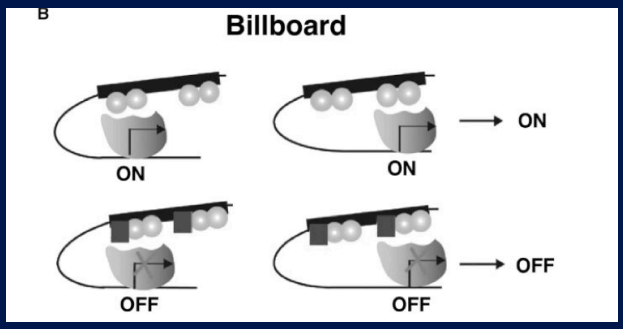
- The transcription rate of the gene is on or off.
- The enhancer shifts the balance to the active state.





Sutherland et al. MCB (1997)

Billboard Enhancers

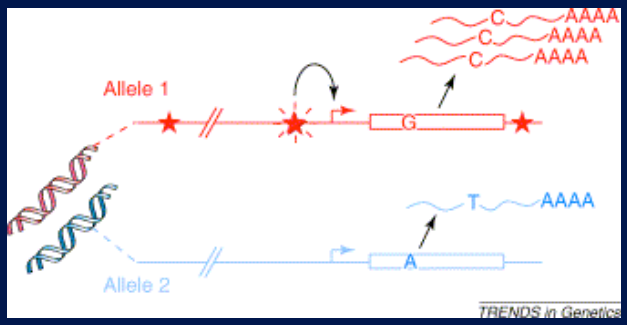


Arnosti and Kulkarni J. Cell. Biochem. (2005)

Binding sites are flexibly positioned
Ensemble of separately acting factors
Independently interact with their targets

Rheostat Function

Enhancers can quantitatively regulate transcription rates through a continuous spectrum.



Bioinformatic Implications

- Using phylogenetic analyses to identify *cis* regulatory grammar will work for enhanceosomes, but may not work for billboards.
- A lack of sequence conservation does not indicate a lack of relevance for transcriptional regulation.
- The placement of repressors relative to activators influences function.
- As the specific rules of the grammar are learned, effective bioinformatic analyses will ensue.

IV. Pattern searching in the genome

Most functional elements lends themselves to pattern mapping or discovery

- 3' UTRs are targets of microRNA
- Display conserved patterns
- Interfere with transcription or translation

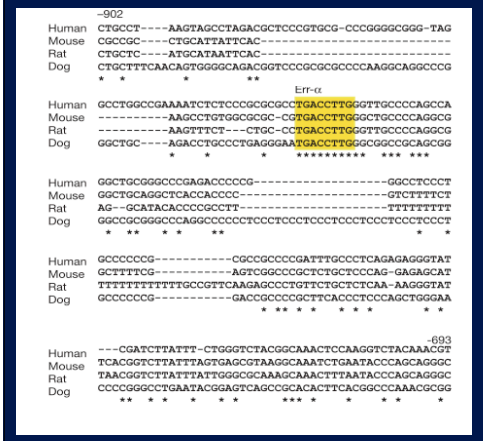
articles

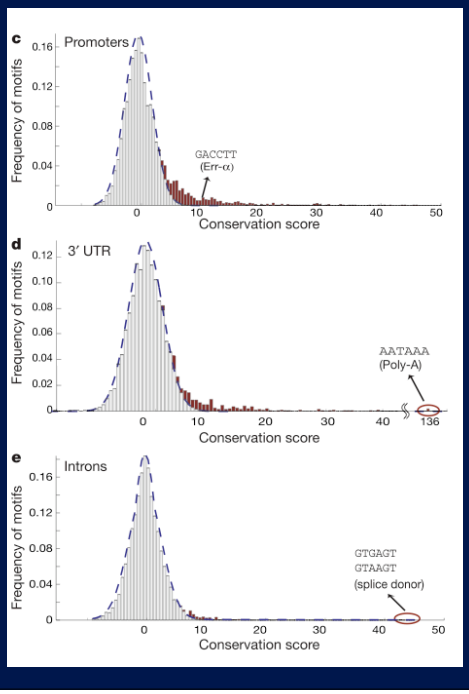
Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals

Xiaohui Xie[1], Jun Lu[1], E. J. Kulbokas[1], Todd R. Golub[1], Vamsi Mootha[1], Kerstin Lindblad-Toh[1], Eric S. Lander[1,2]* & Manolis Kellis[1,3]*

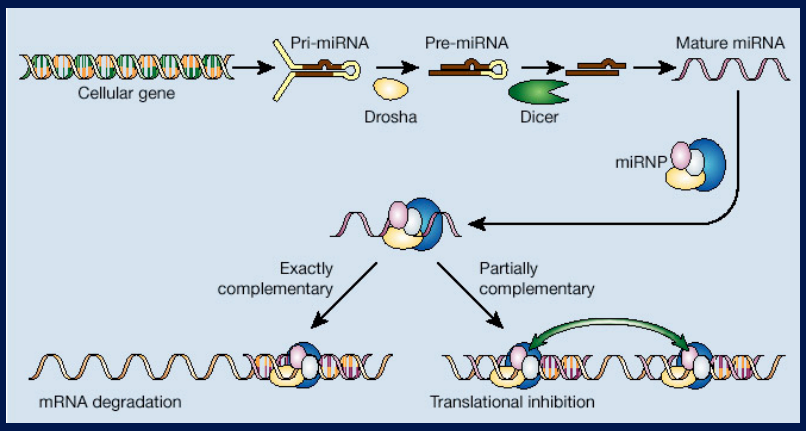
[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02139, USA
[3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* These authors contributed equally to this work

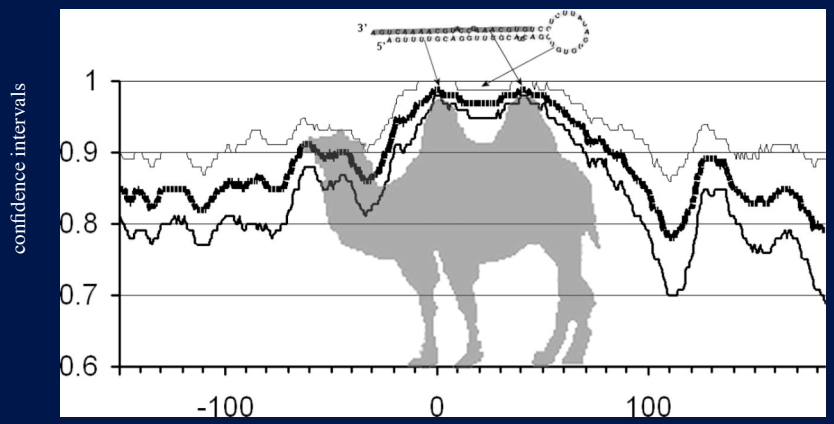




MiRNA Interference

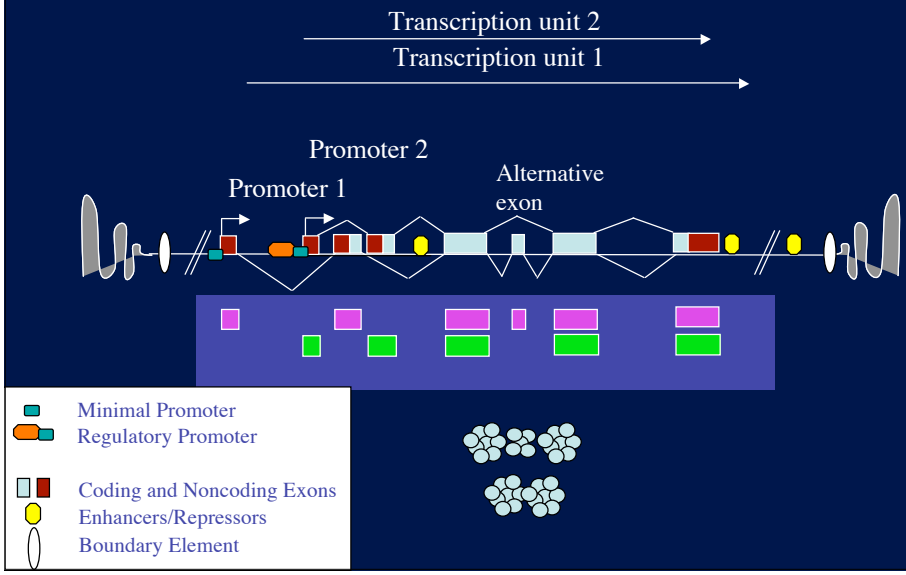


Conservation of miRNA

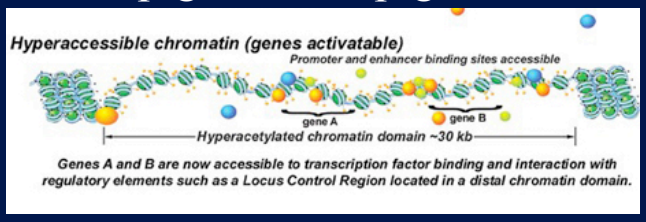


Berezikov et al. Hum. Mol. Genet. 2005

Complexity of *cis*-regulation

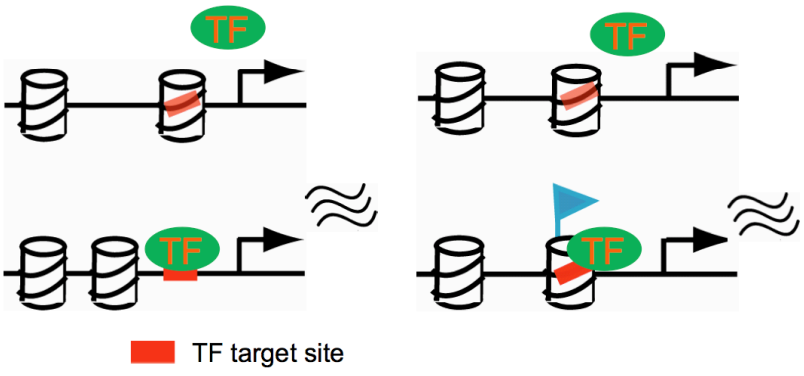


V. Epigenetics/Epigenomics

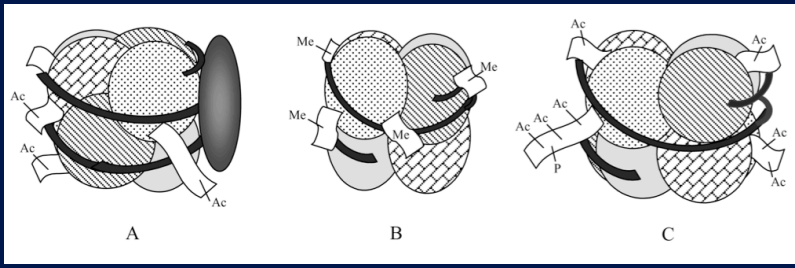


Transcriptional regulation by chromatin

- Nucleosome positioning
- Histone modification

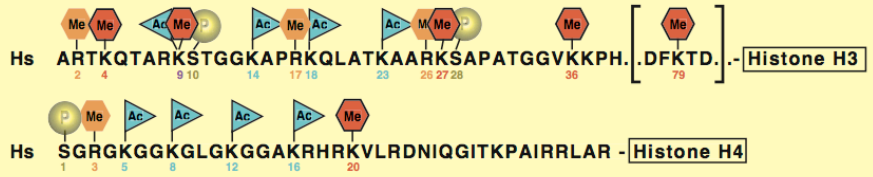


Histone Modifications

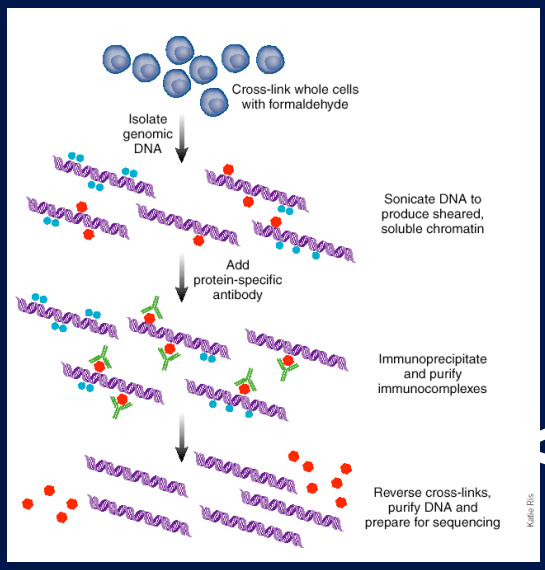


He and Lehming, 2003

Histone tail modifications



Detection Assays

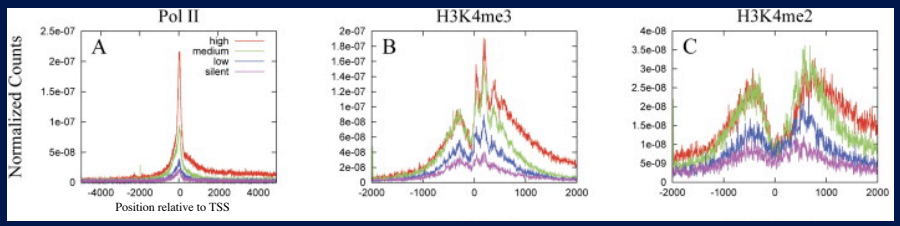


ChIP-Chip

ChIP-Seq

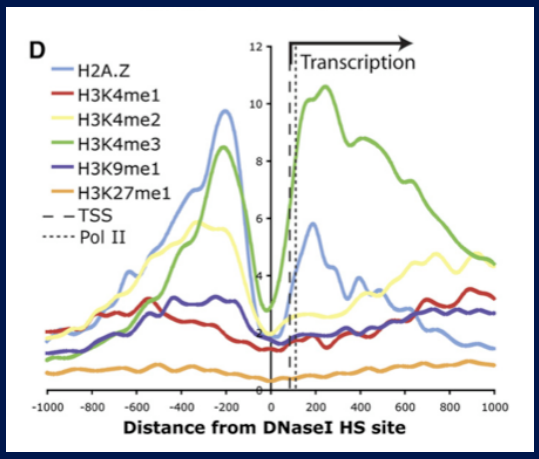
High-Resolution Profiling of Histone Methylations in the Human Genome

Artem Barski,[1,3] Suresh Cuddapah,[1,3] Kairong Cui,[1,3] Tae-Young Roh,[1,3] Dustin E. Schones,[1,3] Zhibin Wang,[1,3] Gang Wei,[1,3] Iouri Chepelev,[2] and Keji Zhao[1,*]



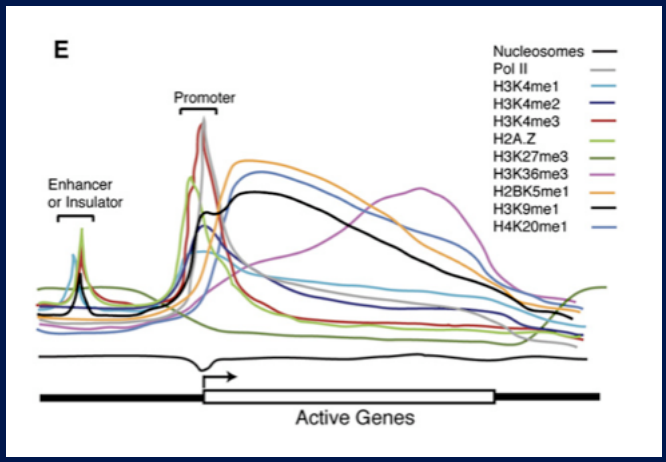
High-Resolution Mapping and Characterization of Open Chromatin across the Genome

Alan P. Boyle,[1] Sean Davis,[3] Hennady P. Shulha,[2] Paul Meltzer,[3] Elliott H. Margulies,[4] Zhiping Weng,[2] Terrence S. Furey,[1,*] and Gregory E. Crawford[1,*]



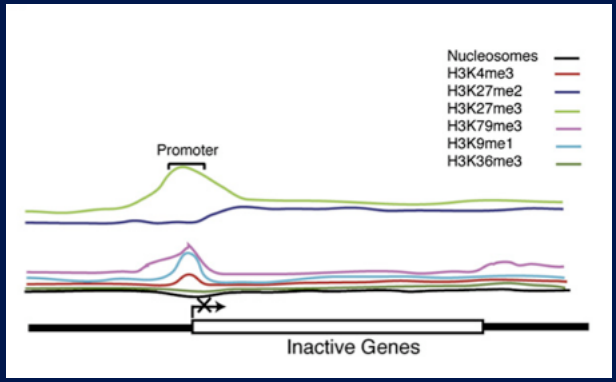
Activation Marks

High levels of H3K4me (1-3), H3K9me1, H2A.Z near the TSS
H3K36me3 and H4K20me1 in transcribed regions.

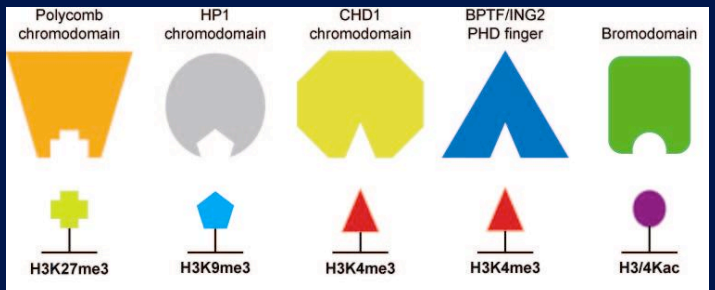


Repressive Marks

High levels of of H3K27me3 and H3K79me3



Translator Proteins

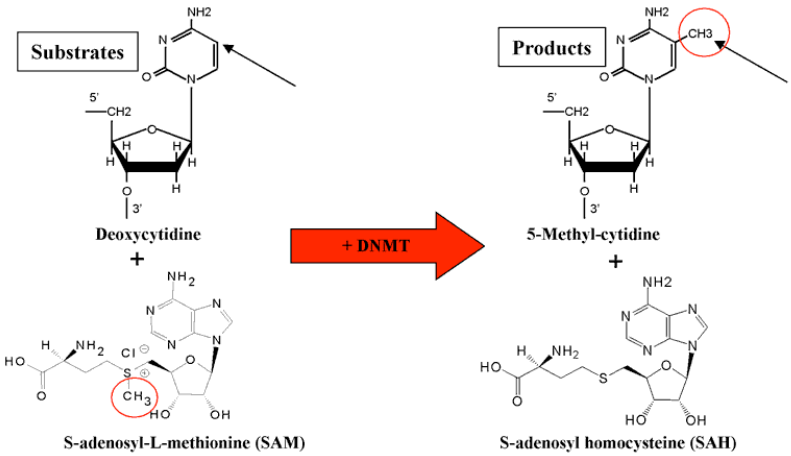


Higgs DR, et al. 2007.
Annu. Rev. Genomics Hum. Genet. 8:299–325

VI. Genome Methylation

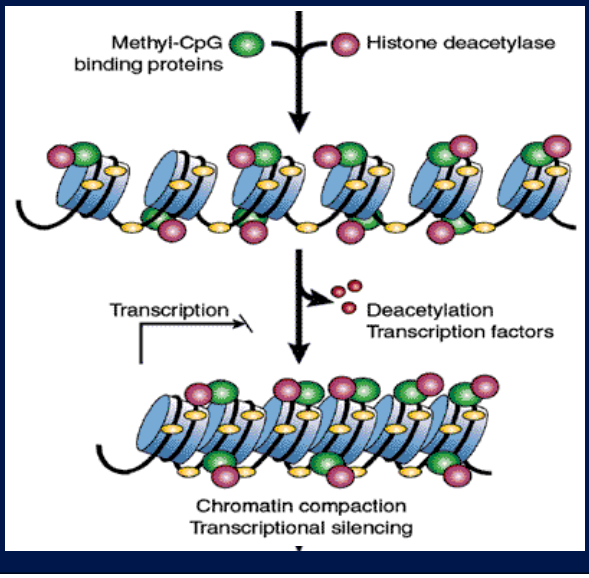
- Embryonic development
- Transcription
- Chromatin structure
- X chromosome inactivation
- Genomic imprinting
- Chromosome stability
- Human disease

Methylation of Cytosine in a CpG Dinucleotide

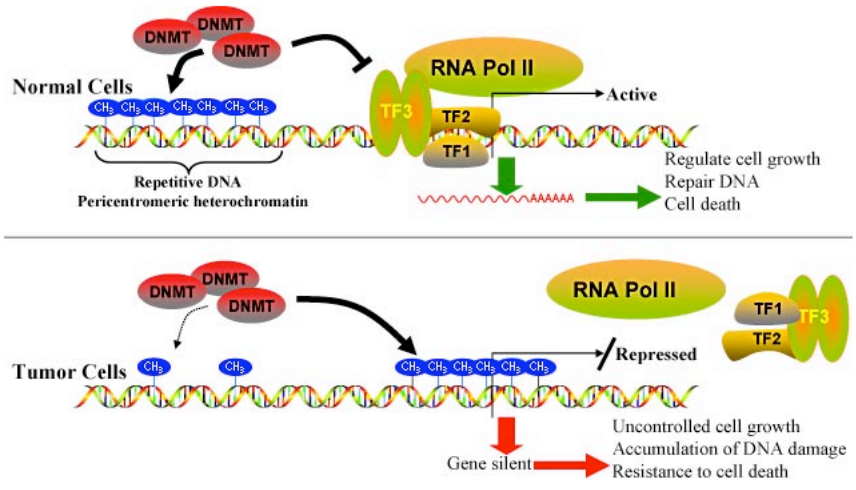


Keith D. Robertson, Ph.D.

DNA Methylation Coincides with Gene Silencing



Methylation Targets

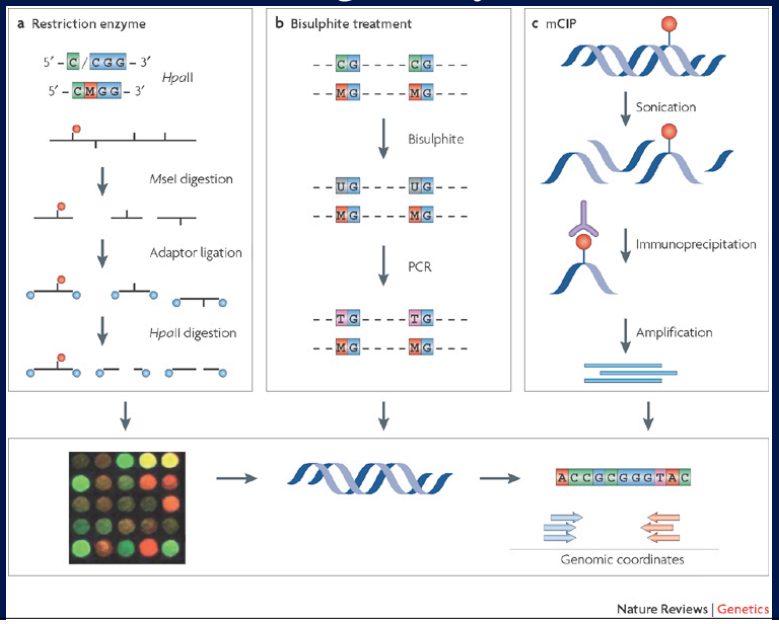


Keith D. Robertson, Ph.D.

DNA Methylation is a biomarker for cancer

- CpG island hypermethylation, have been extensively studied and are very frequent and early events
- A distinct subset of many tumor types has a CpG-island-methylator phenotype
- Detection of methylated DNA in body fluids has the potential for early cancer detection

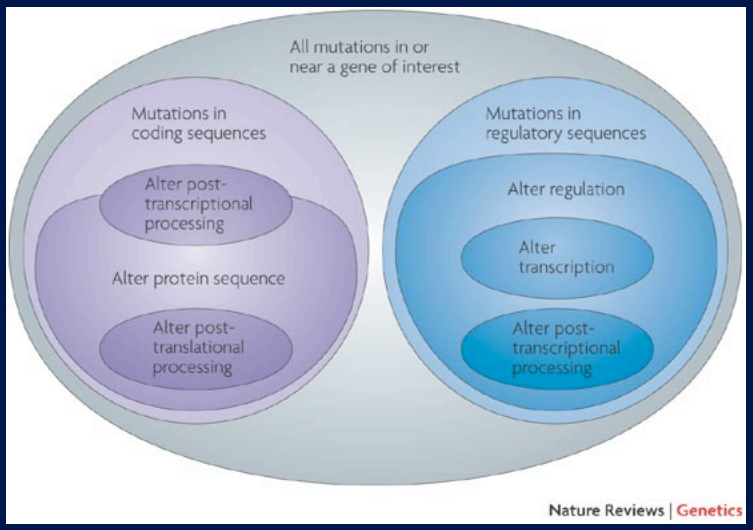
Detecting Methylation



DNA Methylation as a Chemoprevention Target

- Genes silenced by DNA methylation are intact and can be reactivated by small molecule inhibitors of the DNMTs
- Inhibitors of DNA methylation, such as 5-aza-2'-deoxycytidine (5-azadC) are capable of gene reactivation and restoration of cell growth control, apoptosis, and DNA repair capacity

VII. The landscape of regulatory mutations



Wray *Nature Reviews Genetics* (2007)

Cis-regulatory mutations

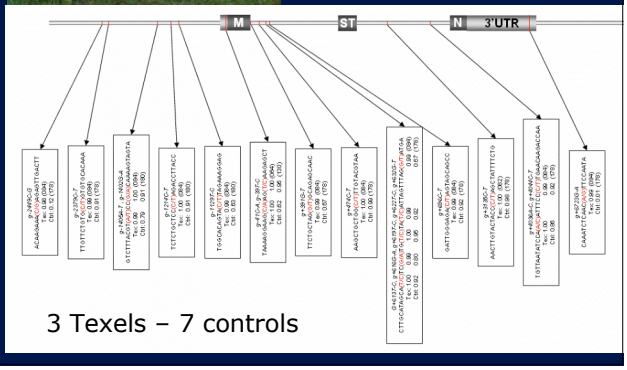
Gene	Function of product	Phenotype
AVPR1A	Vasopressin receptor	Creative dance performance
Avpr1a	Vasopressin receptor	Paternal care
Cyp6G1	P450 enzyme	Pesticide resistance
DARC	Chemokine receptor	Resistance to infection with malaria
e	Pigment synthesis	Colour pattern of abdomen
hsp70	Heat shock protein	Thermal tolerance
HTR2A	Serotonin receptor	Obsessive-compulsive behaviour
IL10	Interleukin	Outcome of infection with HIV and infection with leprosy
IL10	Interleukin	Susceptibility to schizophrenia
LCT	Digestive enzyme	Lactose persistence
LDH	Metabolic enzyme	Cardiac physiology

Polymorphic miRNA–target interactions: a novel source of phenotypic variation

Michel Georges

Hypermuscled Texel: Patrocles mutation

Resequencing the *MSTN* gene identifies 20 non-coding SNPs …

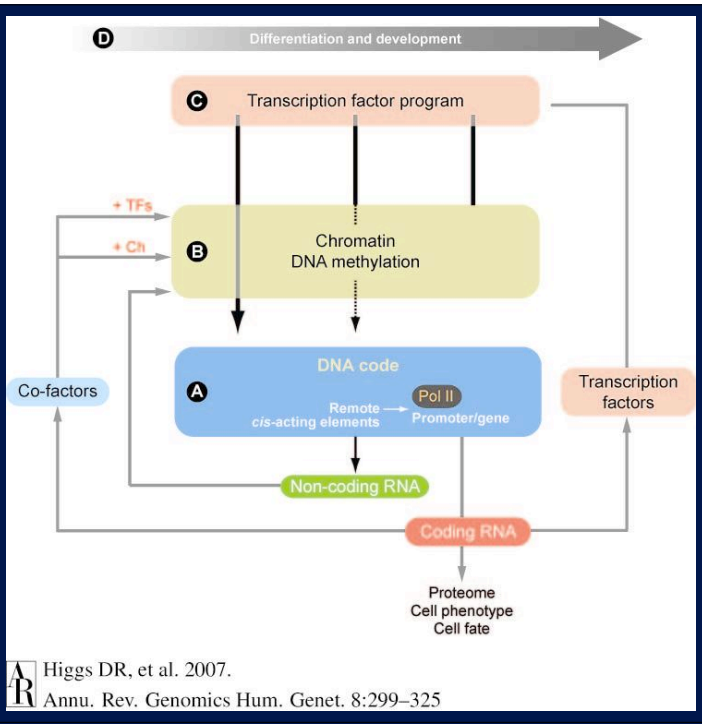


3 Texels – 7 controls

UTR (mis)Regulation

- Predicted to be a target site for *miR1*, *miR206*
- *miR1* and *miR206* are conserved in sheep and strongly expressed in skeletal muscle …
- Texel sheep have ≈ 3-fold reduction in circulating MSTN levels …
- mRNA allelic imbalance in *GA* heterozygotes …

The polymorphism created an illegitimate miRNA target site



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

Summary

1. Understand how mammalian genes are switched on and off during development and differentiation.

2. Understand and integrate the transcriptional program with the epigenetic program.

3. Apply techniques to chromosomal domains, whole chromosomes, and the entire genome by using microarray or sequencing technology.

4. Gain insights into transcriptional and epigenetic regulation to understand how they are perturbed in human genetic disease.

Current Topics in Genome Analysis

Next Lecture:

Microarray Analysis

Paul Meltzer, M.D., Ph.D.
National Cancer Institute
National Institutes of Health