CURRENT TOPICS IN GENOME ANALYSIS 2016

WEEK 5: REGULATORY AND EPIGENETIC LANDSCAPES
OF MAMMALIAN GENOMES

MARCH 16, 2016

LAURA ELNITSKI, Ph.D.
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

Outline: From blueprint to implementation

1. Genome composition
2. Functional elements
3. Histone modifications
4. DNA methylation
5. Revisiting

What’s next?
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)

Characterizing the human genome

• ~ 3.2 billion bases
• ~ 22,000 protein-coding genes
• 98% noncoding
Gene and genome content:

- 1 billion bp
- 3.2 billion bp
- 13.8 billion bp

Genetic disease:

- Mendelian or monogenic
- Complex or multifactorial

Susceptibility genes
Environmental factors
Clinical Phenotype

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

From blueprint to implementation

"nothing in the genome makes sense, except in 3D"
Central dogma - 1958

Noncoding RNA

1. Central dogma - 1958

2. Noncoding RNA

- DNA → RNA → Protein
- lincRNA-Protein Complex
- Adapter
1.2. Examining the linear sequence

EVOLUTIONARY CLUES

Evolutionary Distance

Human
Chimpanzee
Horse
Rat
Platypus
Pufferfish
Phylogenetic trees

Phylogenetic footprints

* = identical
- = gap
Phylogenetic shadowing

Conserved elements in the human genome

alignments in neutral regions

all human-mouse alignments

Selective constraint on 5% of human genome

85% id on average

[Mouse consortium, Nature 2002]
Ultraconserved elements in the human genome

- 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

http://www.mccray.com/sickle-cell/
Lactose intolerance

Regions of accelerated change

Regulatory variation

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

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

Insulators/boundaries
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

Variants in enhancers

Allele Specific activity

full#hash.ru:activity.pdf
**SHH enhancer mutations**


*Development* 2005 132: 4797-805


**Hemingway cat with six toes**

**XIST dispersion**

http://www.youtube.com/watch?v=V3sAcgHoeA4#t=150
Regulatory architecture

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

2.3 3D packaging

Inter-chromosomal networks

Lamin domains
Progeria

Progeria

Altered genome topology in cancer

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

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

b  Active enhancer

\[ \text{H3K27ac} \]
\[ \text{H3K4me1} \]

\text{Enhancer}

d  Closed or poised enhancer

\[ \text{H3K4me1} \]
\[ \text{H3K27me3} \]

Rosetta stone of chromatin?

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

Gradual shifts in methylation

http://ajp.physiology.org/content/109/1/243
5-methyl cytosine

Waddington’s epigenetic landscape


Inherited at each generation
Nuclear reprogramming


Inherited at each generation

Muscle Liver Leg Eye

Oct4/Nanog/Sox2

Behavioral effects

Methylation and cancer

(a) Repetitive sequence
(b) CpG island promoters
(c) CpG island shore

Methylated cytosine ▲ Unmethylated cytosine

Genome misfolding in cancer

PDGFRα and FIP1L1

normal ▲ tumor

PDGFRA and FIP1L1
5.1. Revisiting genome composition

ENCODE

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

<table>
<thead>
<tr>
<th>Genomic element</th>
<th>Number</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein-coding genes</td>
<td>20330</td>
<td>GENCODE V17 (Feb2013, GRCh37) Ensembl 72</td>
</tr>
<tr>
<td>Long non-coding RNAs</td>
<td>13333</td>
<td>GENCODE V17</td>
</tr>
<tr>
<td>IncRNAs</td>
<td>6020</td>
<td>GENCODE V17</td>
</tr>
<tr>
<td>Pseudogenes</td>
<td>14154</td>
<td>GENCODE V17</td>
</tr>
<tr>
<td>Short non-coding RNAs</td>
<td>9078</td>
<td>GENCODE V17</td>
</tr>
<tr>
<td>miRNAs</td>
<td>3086</td>
<td>GENCODE V17</td>
</tr>
<tr>
<td>Promoters</td>
<td>70292</td>
<td>ENCODE [3]</td>
</tr>
<tr>
<td>Enhancers</td>
<td>369124</td>
<td>ENCODE [3]</td>
</tr>
<tr>
<td>TFBS (ChIP peaks)</td>
<td>636336</td>
<td>ENCODE [3]</td>
</tr>
</tbody>
</table>

Introns, long intergenic non-coding RNAs: miRNAs, micro RNAs: TFBS, transcription factor binding sites.

Defining function

The absence of evidence is not the evidence of absence.

— Carl Sagan —
Deletion of promising regions

Deletion of Ultraconserved Elements Yields Viable Mice

Nadav Ahituv1,2,*, Yuesen Zhu1, Axel Visel4, Amy Holt1, Veena Afzal1, Len A. Pennacchio3,4, Edward M. Rubin1,2

1 Genomic Research, 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 excised the elements in a mouse transgenic assay and that are near genes that exhibit imprinted expression in the mouse and when their expression is altered due to other genetic modifications. Lines of mice lacking these ultraconserved elements were viable and did not exhibit abnormalities when assayed for a variety of phenotypes, including growth, fertility, and behavior. In addition, more targeted screens, informed by the abnormalities observed in investigated elements had been altered, also failed to reveal notable abnormalities of all the possible phenotypic impact of the deleted sequences, indicating that these elements are not necessarily reflect crucial functions required for viability.

Shadow enhancers (redundancy/resilience?)
Structural conservation in IncRNA

5.2. Seemingly nonfunctional regions
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/1535-7697(12)00131-1
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