

Module 4

Working with ENCODE Data

Bert Overduin, Ph.D.

Edinburgh Genomics
The University of Edinburgh
Edinburgh EH9 3FL
Scotland, United Kingdom

Overview

- ENCODE Project
- Other epigenomics projects
- Ensembl Regulatory Build

ENCODE

- Encyclopedia Of DNA Elements
- Aim: identify all functional elements in the human genome sequence
- GENCODE is part of ENCODE
- Started in 2003, still ongoing
- Tens of groups, hundreds of scientists
- Hundreds of papers

Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project

The ENCODE Project Consortium*

We report the generation and analysis of functional data from multiple, diverse experiments performed on a targeted 1% of the human genome as part of the pilot phase of the ENCODE Project. These data have been further integrated and augmented by a number of evolutionary and computational analyses. Together, our results advance the collective knowledge about human genome function in several major areas. First, our studies provide convincing evidence that the genome is pervasively transcribed, such that the majority of its bases can be found in primary transcripts, including non-protein-coding transcripts, and those that extensively overlap one another. Second, systematic examination of transcriptional regulation has yielded new understanding about transcription start sites, including their relationship to specific regulatory sequences and features of chromatin accessibility and histone modification. Third, a more sophisticated view of chromatin structure has emerged, including its inter-relationship with DNA replication and transcriptional regulation. Finally, integration of these new sources of information, in particular with respect to mammalian evolution based on inter- and intra-species sequence comparisons, has yielded new mechanistic and evolutionary insights concerning the functional landscape of the human genome. Together, these studies are defining a path for pursuit of a more comprehensive characterization of human genome function.

Nature. 2007 Jun 14;447(7146):799-816

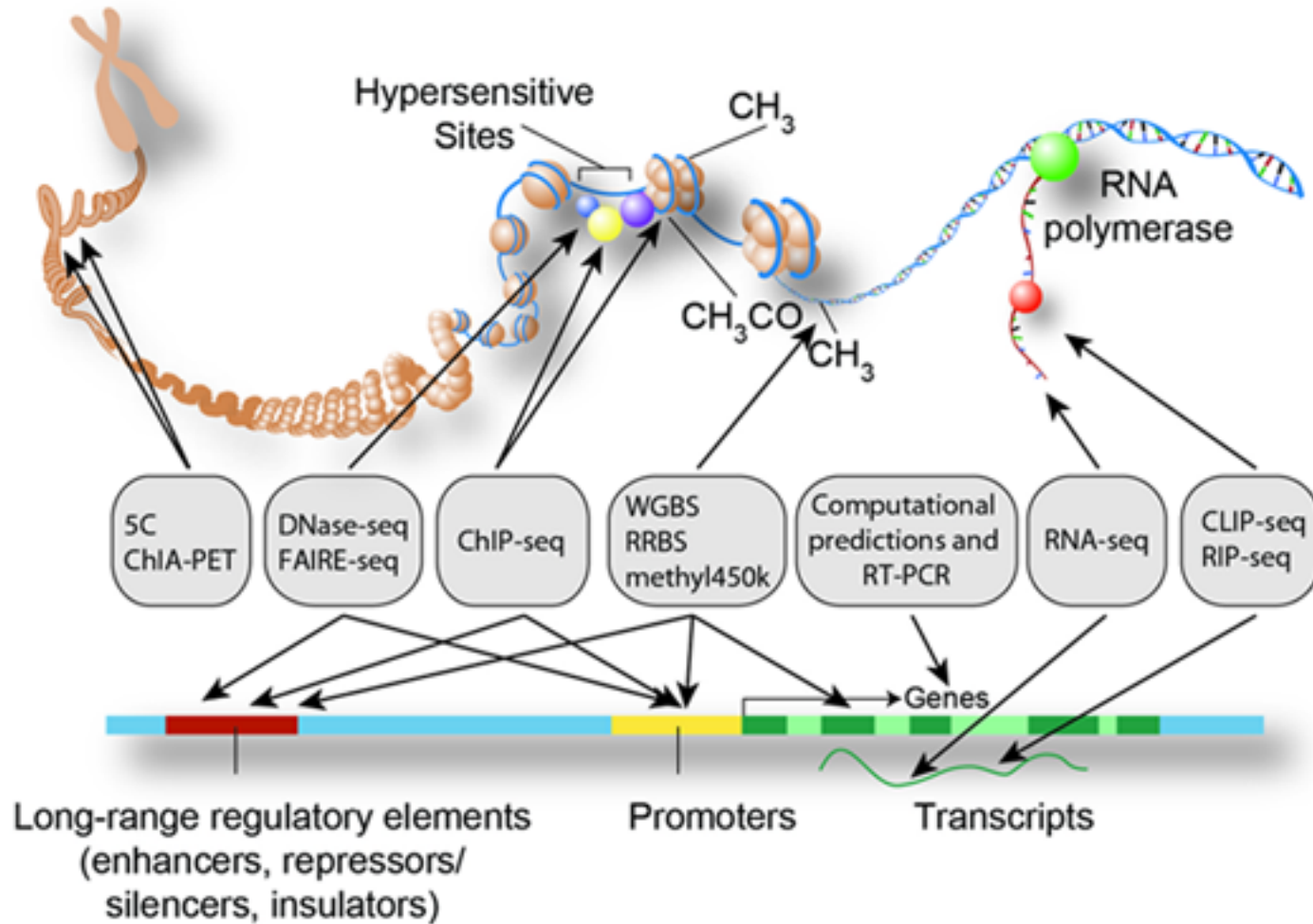
An integrated encyclopedia of DNA elements in the human genome

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.

[Nature. 2012 Sep 6;489\(7414\):57-74. doi: 10.1038/nature11247](#)

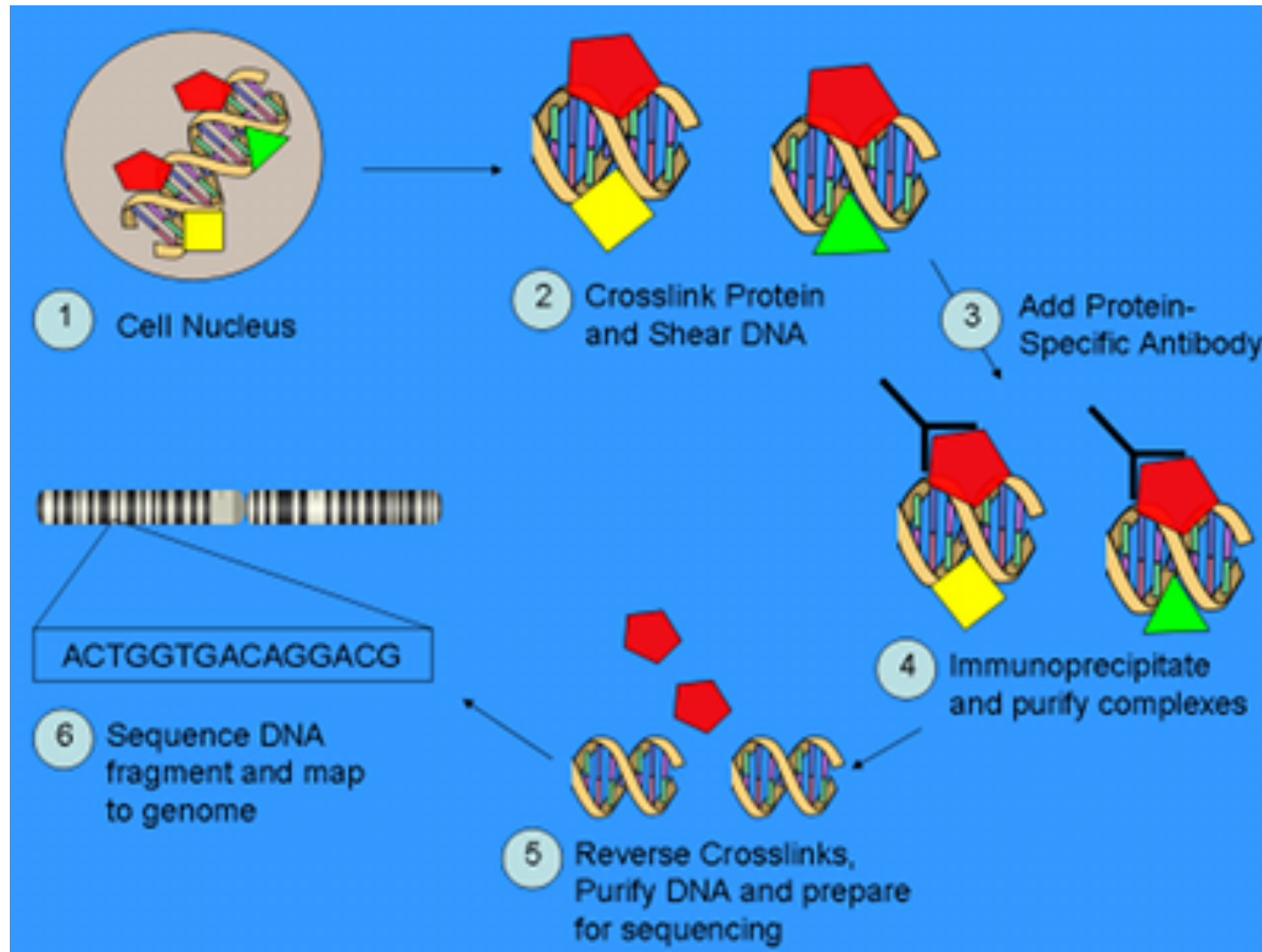
Assays



Assays

- DNase I hypersensitivity assays
- DNA methylation assays
- Immunoprecipitation of proteins that interact with DNA and RNA:
 - histone modifications
 - transcription factors
 - chromatin regulators
 - RNA-binding proteins

ChIP-seq



Other projects

- Roadmap Epigenomics Project
- <http://www.roadmapepigenomics.org>
- International Human Epigenome Consortium (IHEC)
- <http://ihec-epigenomes.org/about/>
- International Cancer Genome Consortium (ICGC)
- <https://icgc.org>
- Blueprint
- <http://www.blueprint-epigenome.eu>

Ensembl Regulatory Build

- Provides a genome-wide set of regions that are likely to be involved in gene regulation
- Based on ENCODE and Roadmap Epigenomics data
- Cell type independent and cell type specific annotation

Ensembl Regulatory Build

Classification of regions:

- Predicted promoters
- Predicted promoter flanking regions
- Predicted enhancers
- CTCF binding sites
- Unannotated transcription factor binding sites
- Unannotated open chromatin regions

Worked examples

Worked Example URLs

- ENCODE Portal: <https://www.encodeproject.org/>
- ENCODE Roadmap browser: <http://www.encode-roadmap.org>
- Roadmap Epigenomics project: <http://www.roadmapepigenomics.org>
- IHEC Data Portal: <http://epigenomesportal.ca/ihec/index.html>
- UCSC Genome Browser: <http://genome.ucsc.edu/>
- Ensembl Genome Browser: <http://www.ensembl.org>
- Ensembl GRCh37 Browser: <http://grch37.ensembl.org>