

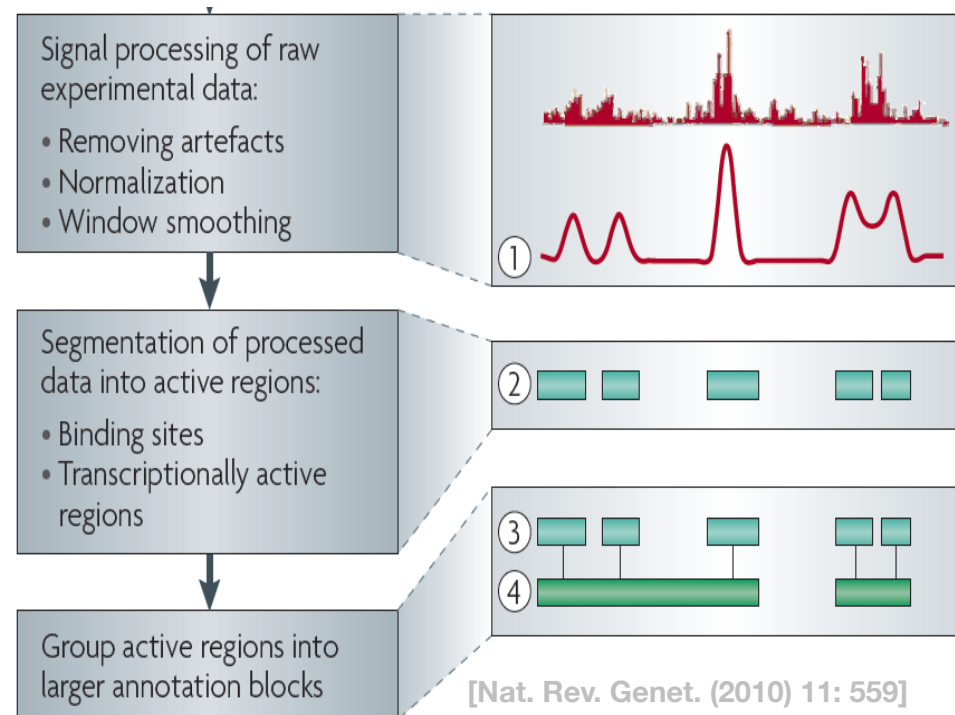
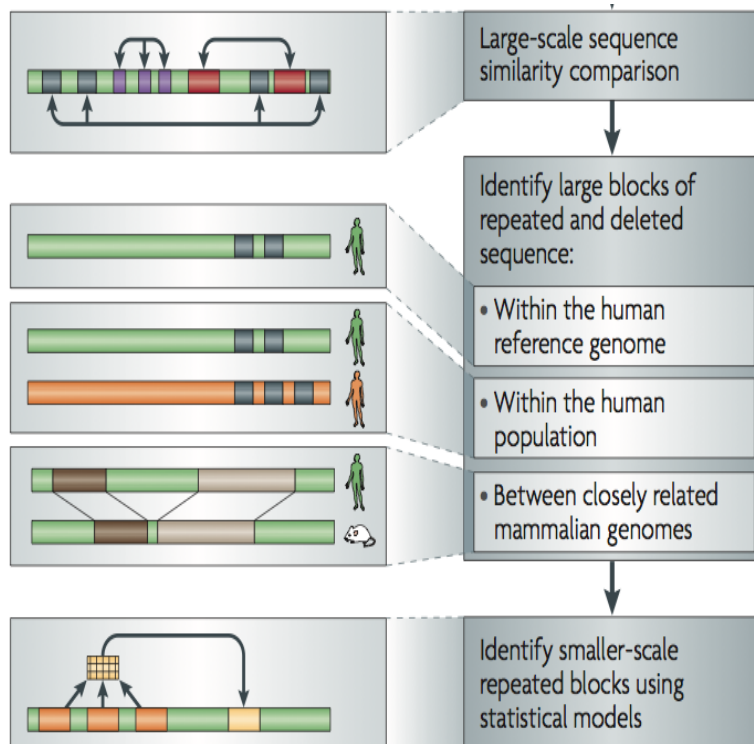
# Non-coding Annotations: Overview

There are several collections of information "tracks" related to non-coding features

Sequence features, incl. Conservation

Functional Genomics

ChIP-seq (Epigenome & seq. specific TF)  
and ncRNA & un-annotated transcription



# Functional Genomics Annotations

## A) PEAKS

1. DNase peaks at the UCSC genome browser {on many cell lines}
2. The regulation track at the UCSC genome browser, with compilation of TF ChIP-seq peaks from uniform processing (individual peaks are annotated with TF and cell line)
3. Blacklist Regions

## B) PROMOTERS

Annotated GENCODE TSSes (also, TSSes with FANTOM CAGE support)

## C) ENHANCERS (Supervised)

## D) UNSUPERVISED SEGMENTATIONS, INCLUDING ENHANCERS

ChromHMM, SegWay, HiHMM....

## E) HOT/LOT REGIONS

## F) CONNECTIVITY

1. Enhancer-target gene connection
2. TF-target network connectivity
3. TADs: Topologically Associated Domain

## G) MOTIFS

for TF binding

## H) RNA

1. A matrix of expression data of known genes (or exons) for protein-coding genes & known ncRNAs {on many cell lines}
2. Novel RNA contigs track, i.e., possible novel transcripts (ie Transcriptionally Active Regions or TARs)
3. Novel junctions

## I) OTHER

1. List of Allelic SNPs & Regions
2. Models

# Genomic annotations

## Introduction

The ENCODE Project provides a set of candidate genomic regions that can serve as predictions for further investigation. This page provides links to download a set of candidate genomic regions as well as a list of publications that contain additional data.

## Candidate genomic regions

- Gene expression matrix over ENCODE2 cell lines (~60 cell lines in total) in GENCODE 19 [Download data | Download methods]
- GENCODE v19 TSS list stratified by Fantom5 CAGE data [View README]
  - Strict CAGE clusters [Download]
  - Robust CAGE clusters [Download]
  - Permissive CAGE clusters [Download]
- Candidate enhancers based on DNase hypersensitivity and H3K27ac and annotated with TF-ChIP peaks as well as candidate promoters annotated with TF-ChIP peaks. [Visualize data | Download methods]
  - Distal DNase peaks [Download]
  - Proximal DNase peaks [Download]
  - H3K27ac annotations [Download]
  - Distal TF binding sites [Download]
  - Proximal TF binding sites [Download]

## Additional annotations

Papers previously published by the ENCODE Consortium contain data files that include additional genomic annotations. Search for all publications with ENCODE element data

### Peaks

Peaks are enriched regions of the genome corresponding to either sites of transcription factor binding or DNase hypersensitivity identified during various functional genomic assays. In this section, we provide a list of peaks in various cell lines using both DNase-Seq and ChIP-Seq assays. [View publications.](#)

### RNAs

RNA represents the direct readout of the genetic information encoded by genomes and a significant proportion of a cell's regulatory capabilities are focused on its synthesis, processing, transport, modification and translation. A catalogue of the RNA species made inside the cell and the amount of RNA from each of these loci across various cell lines is provided in this section. [View publications.](#)

### Promoters

The promoter is the region proximal to the transcription start site of a gene that regulates its transcription using transcription factor binding sites. These transcription factors recruit RNA polymerase after binding to the promoter and initiate transcription of the gene. [View publications.](#)

**Simplified**

**Comprehensive**  
(published annotation, mostly  
in '12 & '14 rollouts)

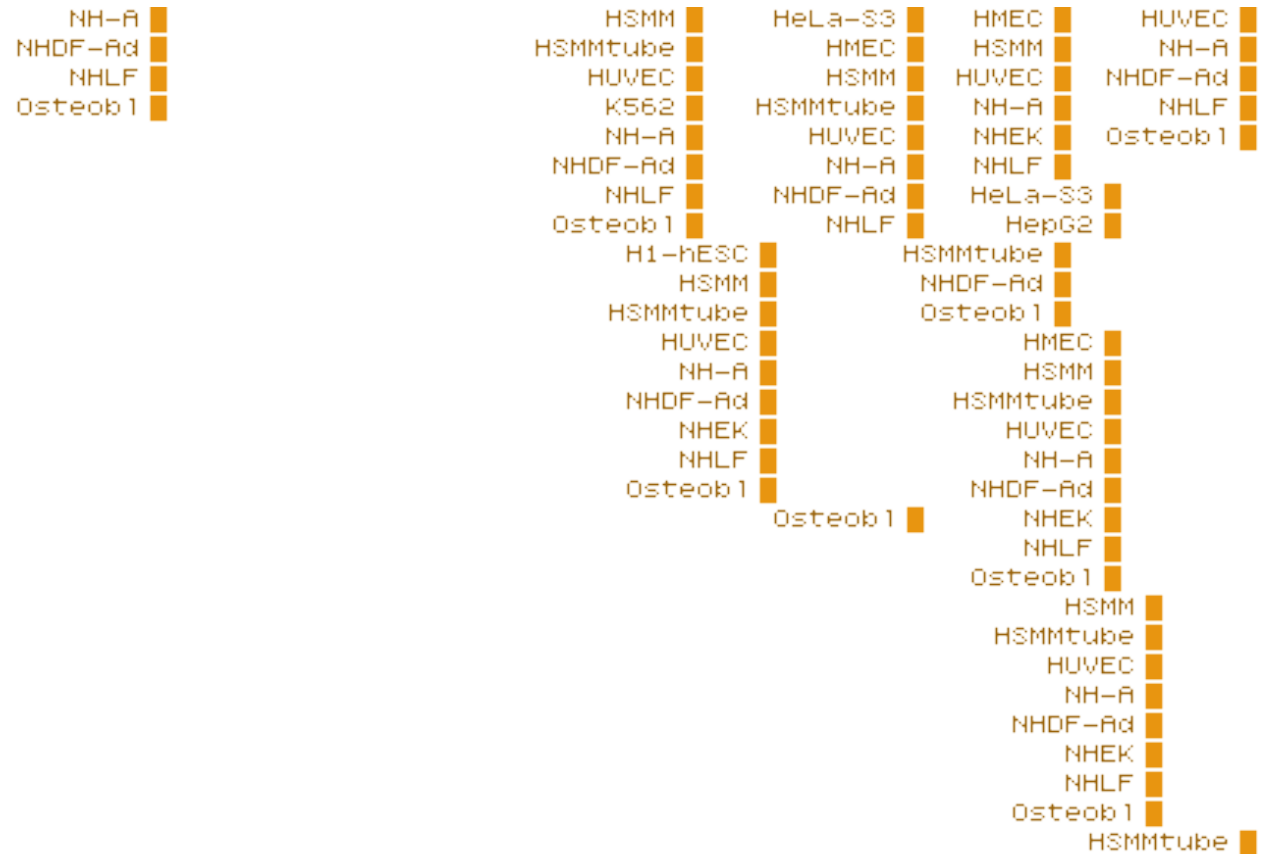
## "Simplified" Annotation

- "Slice" through the ENCODE, providing close-to-data subset of the annotations
- Gene expression matrix
  - over ENCODE2 cell lines (~60 cell lines in total) in GENCODE 19
- TSS list
  - GENCODE v19
- "Tissue type" facet for the cell lines (DCC)



# Simplified regulatory sites

- **Candidate enhancers:** The master list of TSS-distal DHS peaks annotated with
  - **H3K27ac enrichment** (percentile over background) in a cell-type-specific manner.
  - TF ChIP-seq peaks across cell-types
- **Candidate promoters:** The master list of TSS-proximal DHS peaks annotated with TF ChIP-seq peaks across cell types.



Access candidate genomic annotations via [encodeproject.org](http://encodeproject.org) on the "Data" menu bar

The screenshot shows the ENCODE website interface. The top navigation bar includes 'ENCODE', 'Data', 'Methods', 'About ENCODE', and 'Help'. A search bar and a 'Sign in' link are also present. The 'Data' menu is open, showing options for 'Assays', 'Biosamples', 'Antibodies', 'Annotations', and 'Release policy'. The main content area features the title 'ENCODE Encyclopedia of DNA Elements' and a diagram illustrating the relationship between various genomic annotations and their functional elements. The diagram shows a DNA strand with several regions highlighted: 'Long-range regulatory elements (enhancers, repressors/silencers, insulators)', 'Promoters', and 'Transcripts'. Above the DNA, various assays are shown in boxes, with arrows pointing to their corresponding genomic features: SC ChIA-PET, DNase-seq FAIRE-seq, ChIP-seq, WGBS RRBS methyl450k, Computational predictions and RT-PCR, RNA-seq, and CLIP-seq RIP-seq. The diagram also depicts 'Hypersensitive Sites', 'CH<sub>3</sub>' and 'CH<sub>3</sub>CO' methyl groups, and 'RNA polymerase'.

The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

*Image credits: Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)*

[encodeproject.org/data/annotations](http://encodeproject.org/data/annotations)