Navigating the ENCODE Encyclopedia:

Exploring Candidate Regulatory Elements, Linked Genes, and Genetic Variation with SCREEN

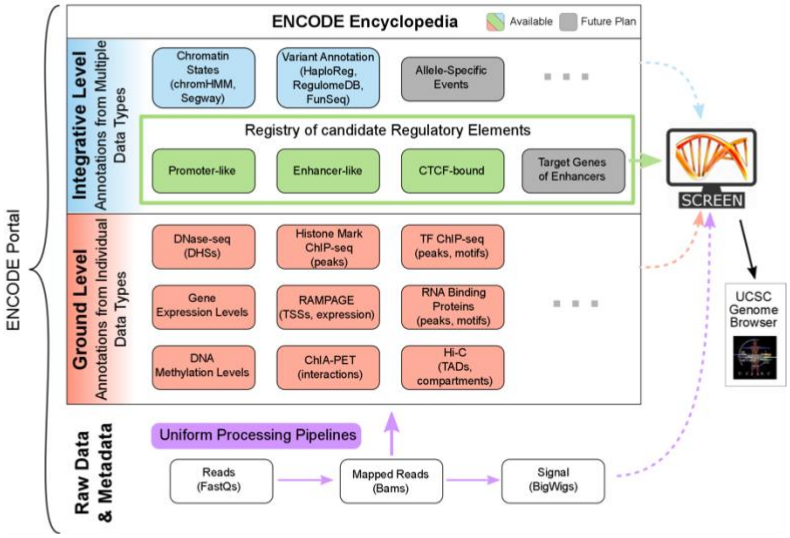
Important Links:

SCREEN Homepage: http://screen.encodeproject.org

Factorbook Homepage: http://www.factorbook.org/

ENCODE Project Portal (to access annotated and data): https://www.encodeproject.org/

Brief Overview of ENCODE Encyclopedia:

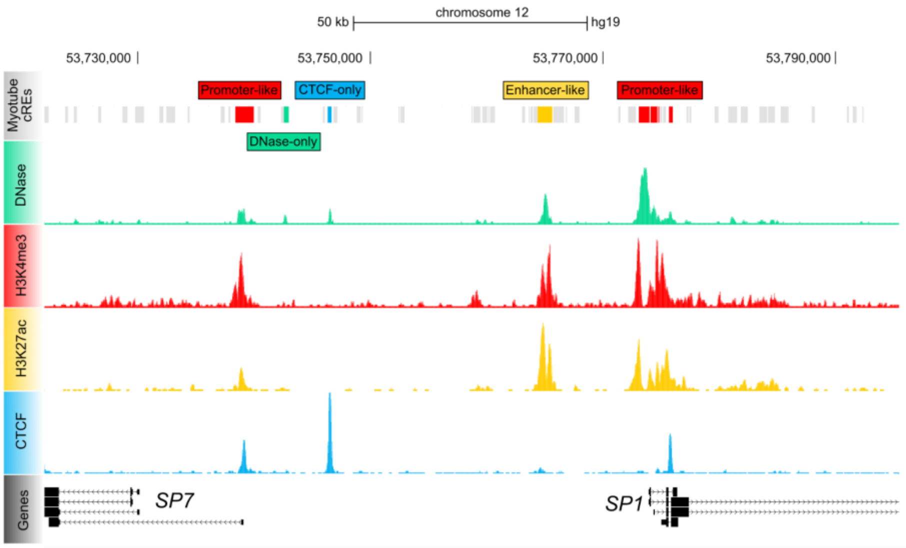


The ENCODE Consortium not only produces high-quality data, but also analyzes the data in an integrative fashion. The ENCODE Encyclopedia organizes the most salient analysis products into annotations, and provides tools to search and visualize them. The Encyclopedia has two levels of annotations:

- Integrative-level annotations integrate multiple types of experimental data and ground level annotations.
- Ground-level annotations are derived directly from the experimental data, typically produced by uniform processing pipelines.

The Registry of Candidate Regulatory Elements

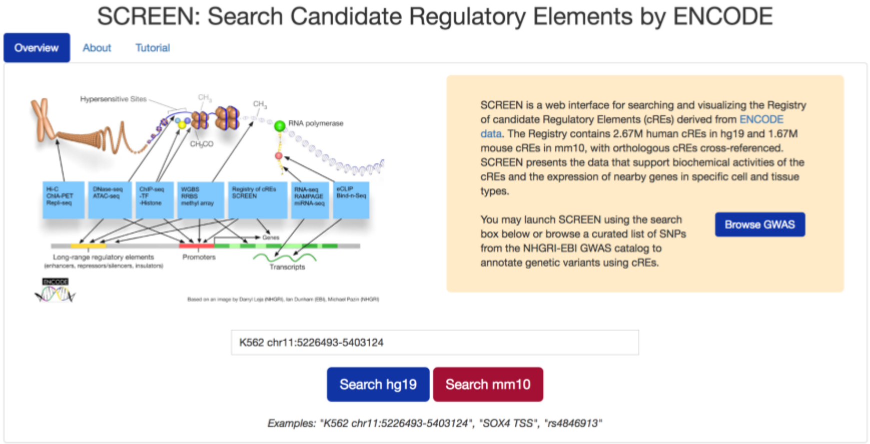
The core of the integrative level of the ENCODE Encyclopedia is the Registry of candidate Regulatory Elements (cREs), which integrates all high-quality DNase-seq and H3K4me3, H3K27ac, and CTCF ChIP-seq data produced by the ENCODE and Roadmap Epigenomics Consortia. The cREs in the Registry are the subset of representative DNase hypersensitivity sites (rDHSs) that are supported by these two histone modifications and CTCF-binding data. Currently the Registry (version 1) contains 1,310,152 human cREs and 527,001 mouse cREs.



Using H3K4me3, H3K27ac, and CTCF signals across all cell types, we classified cREs into promoter-like, enhancer-like, CTCF-bound insulator-like groups in a cell-type agnostic manner. For each specific cell type, we also classified cREs into these groups using DNase, H3K4me3, H3K27ac, and CTCF data specific for that cell type. Currently 21 human (11 mouse) cell types have complete cell-type-specific cRE classifications and 598 human (117 mouse) cell types have partial cRE classifications.

SCREEN

SCREEN is a web-based search and visualization engine specifically designed for the Registry of cREs. SCREEN allows users to explore cREs and investigate how they connect with other annotations in the Encyclopedia in a cell-type-specific manner, as well as the underlying raw ENCODE data whenever informative. SCREEN also presents the results of using cREs to interpret the variants uncovered by Genome-wide Association Studies (GWAS).


SCREEN: Search Candidate Regulatory Elements by ENCODE

SCREEN is a web interface for searching and visualizing the Registry of candidate Regulatory Elements (cREs) derived from ENCODE data. The Registry contains 2.67M human cREs in hg19 and 1.67M mouse cREs in mm10, with orthologous cREs cross-referenced. SCREEN presents the data that support biochemical activities of the cREs and the expression of nearby genes in specific cell and tissue types.

You may launch SCREEN using the search box below or browse a curated list of SNPs from the NHGRI-EBI GWAS catalog to annotate genetic variants using cREs.

K562 chr11:5226493-5403124

Examples: "K562 chr11:5226493-5403124", "SOX4 TSS", "rs4846913"

SCREEN Exercises

SPI1 (also known as PU.1) is a hematopoietic transcription factor involved with myeloid and B-lymphoid cell development. Using SCREEN, try to answer the following questions?

1. In which human tissues and cell types is *SPI1* expressed?
2. How many transcripts does the *SPI1* gene have? In which tissues are they expressed?
3. Find all of the cREs which overlap the gene body of SPI1. How many of them are active in adult B cells (hint: select "*B cell adult*")?
4. In B cells, how many of these cREs have an H3K27ac Z-score greater than 2? Use the slider bars to select for these cREs.
5. Select PLS cRE EH37E1093510. In which cell types does this cRE have high H3K4me3 and DNase Z-scores? Are these cell types consistent with the proposed biological role for SPI1?
6. For which type of cells does EH37E1093510 overlap POL2 ChIP-seq peaks?
7. Does EH37E1093510 have an orthologous mouse cRE?

In 2014 the Schizophrenia Group of the Psychiatric Genomics Consortium published a landmark study in which they identified 108 loci associated with Schizophrenia. Using SCREEN, try to answer the following questions?

1. Use the SCREEN GWAS app, select the schizophrenia (SCZ) study (Ripke *et al.* 2014). For which tissues are SCZ SNPs enriched in active cREs?
2. Select the top enriched tissue. Sort the resulting cREs by the name of the closest gene. Select the top listed ELS cRE (EH37E0354127) whose closest gene is *AL109767.1.*
3. For which cell types does this cRE have high H3K27ac and DNase signal?
4. Select the "linked genes" tab. Which gene is this cRE linked to? By which type of data?
5. Does EH37E0354127 have an orthologous mouse cRE? In which tissues is this cRE active? In which tissues is the nearest gene expressed?
6. Using the mouse differential gene expression app, compare the expression of this gene across mouse brain development. What is the general pattern of gene expression and cRE activity?