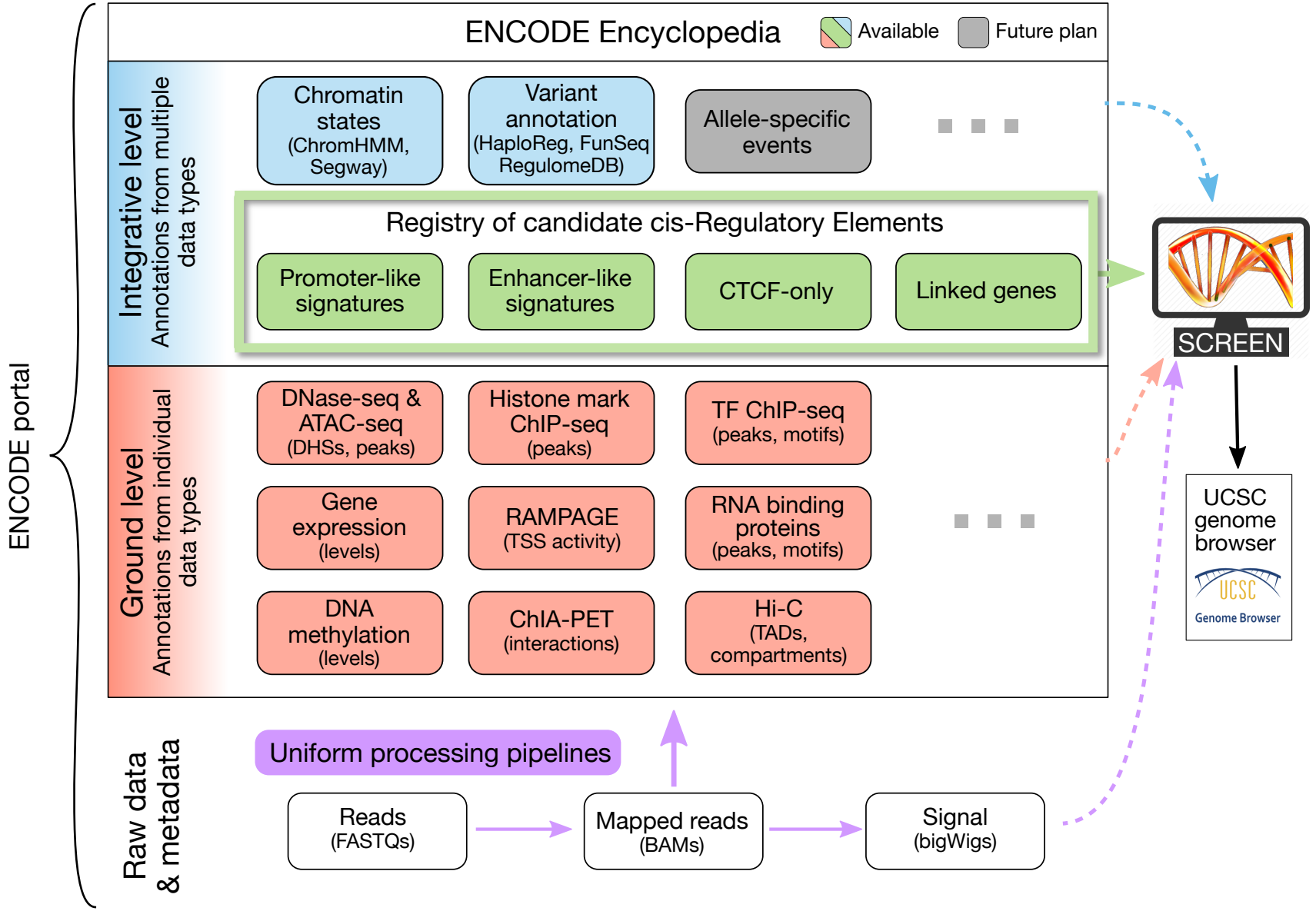
The ENCODE Registry of ccREs

and SCREEN

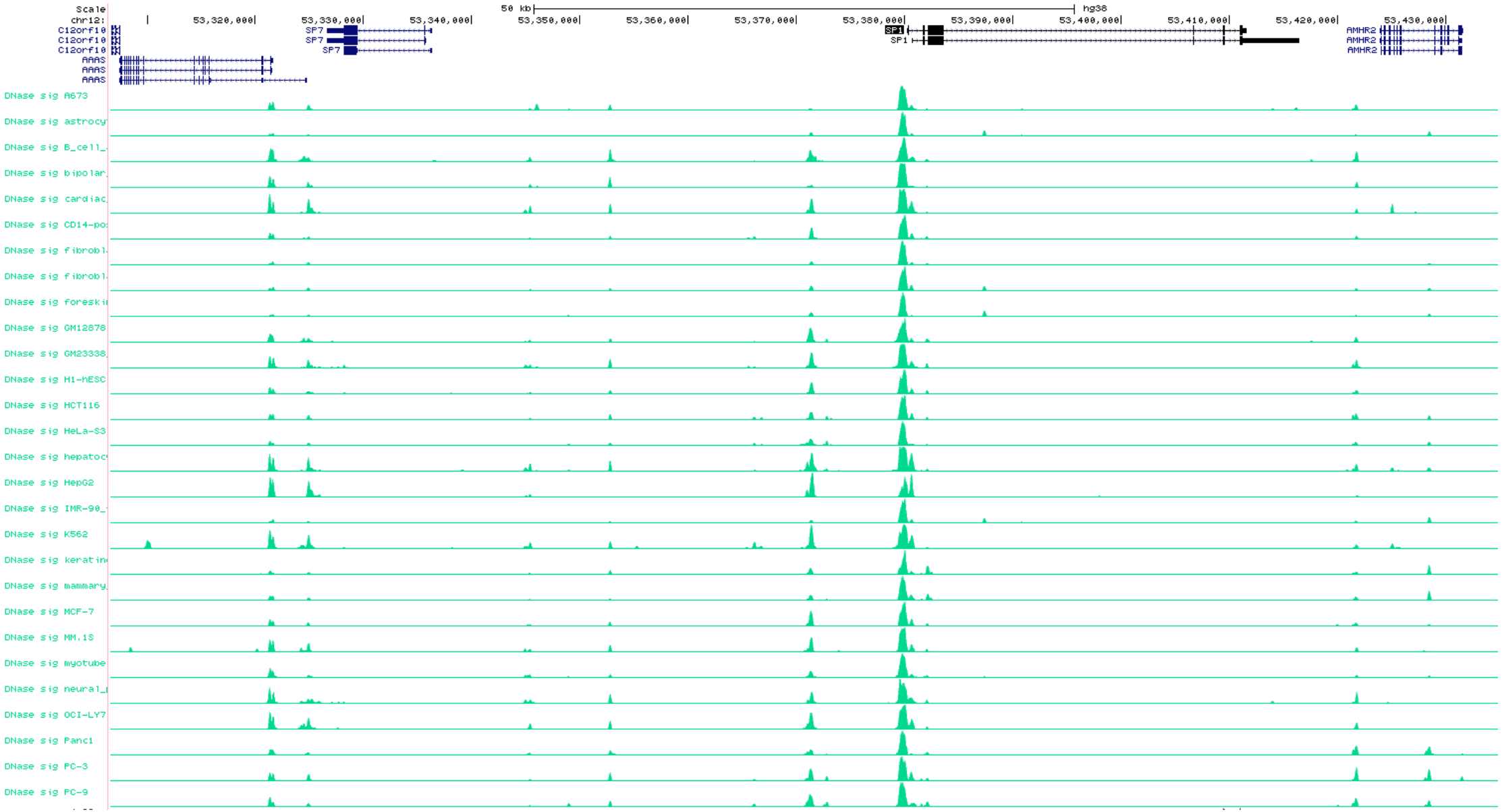
Jill Moore, PhD
Henry Pratt, MD/PhD student

Zhiping Weng Lab
University of Massachusetts Medical School

The ENCODE Encyclopedia

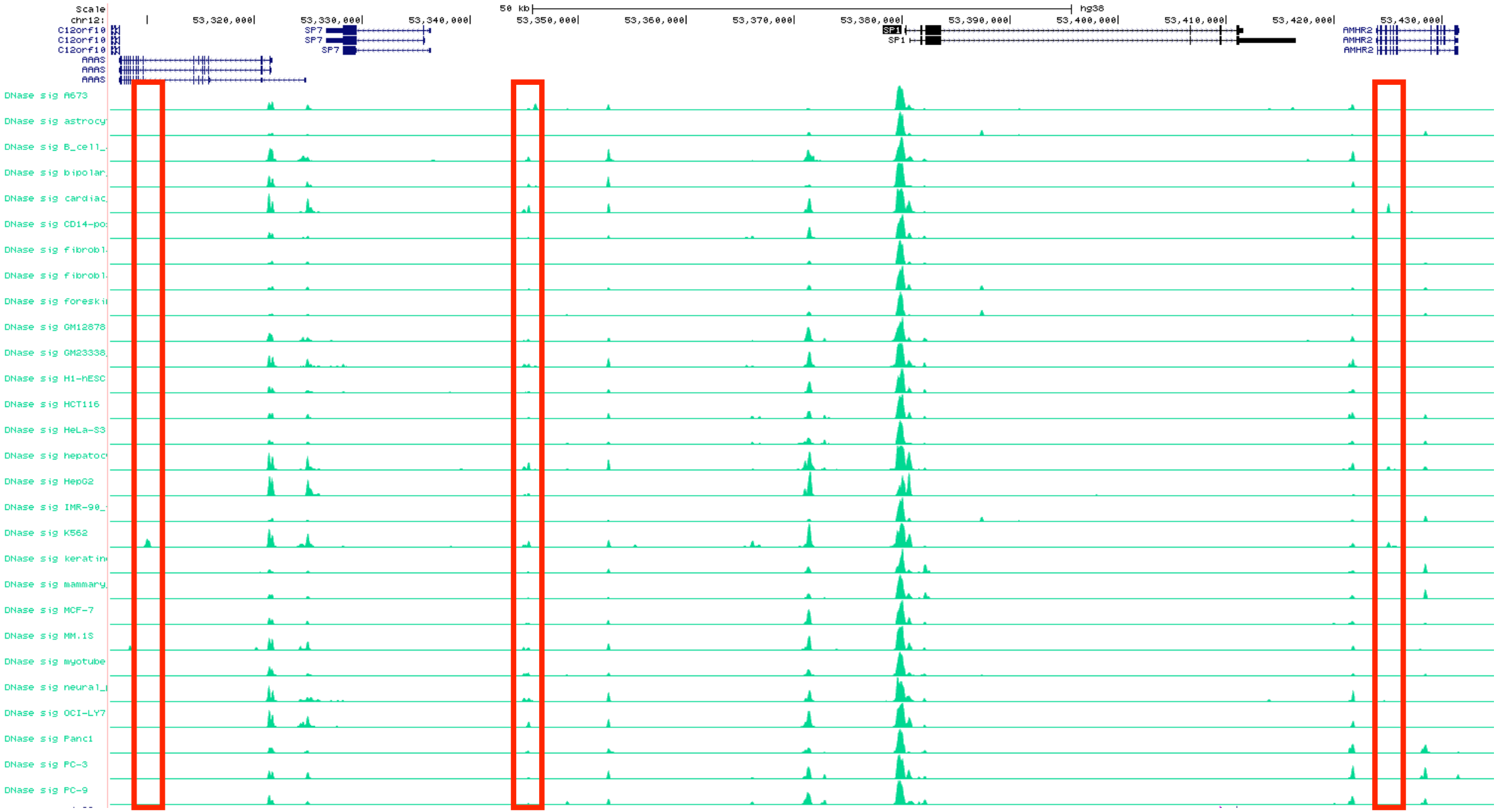


Open chromatin regions across 28 cell types

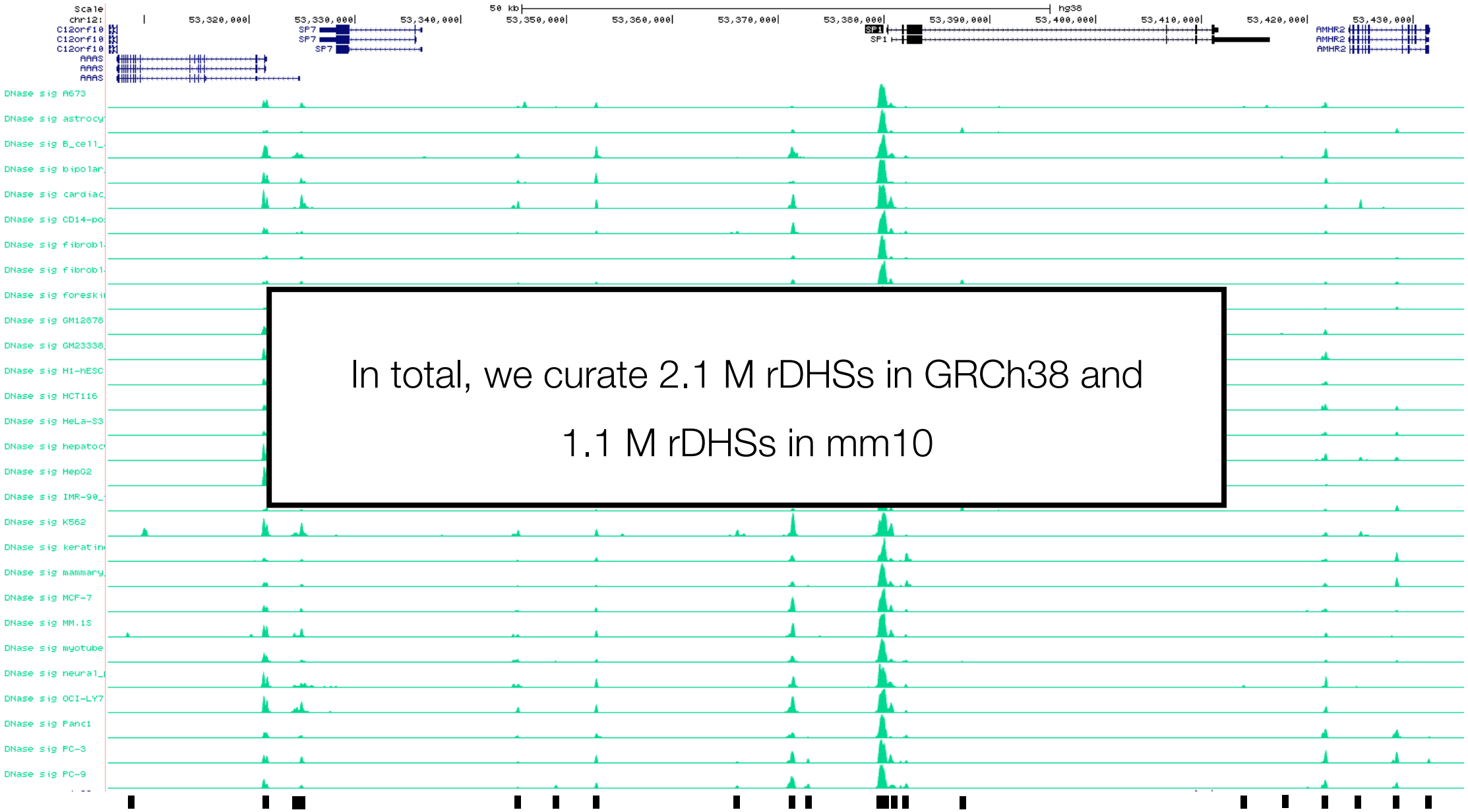


Some regions are ubiquitously open, particularly those at TSSs

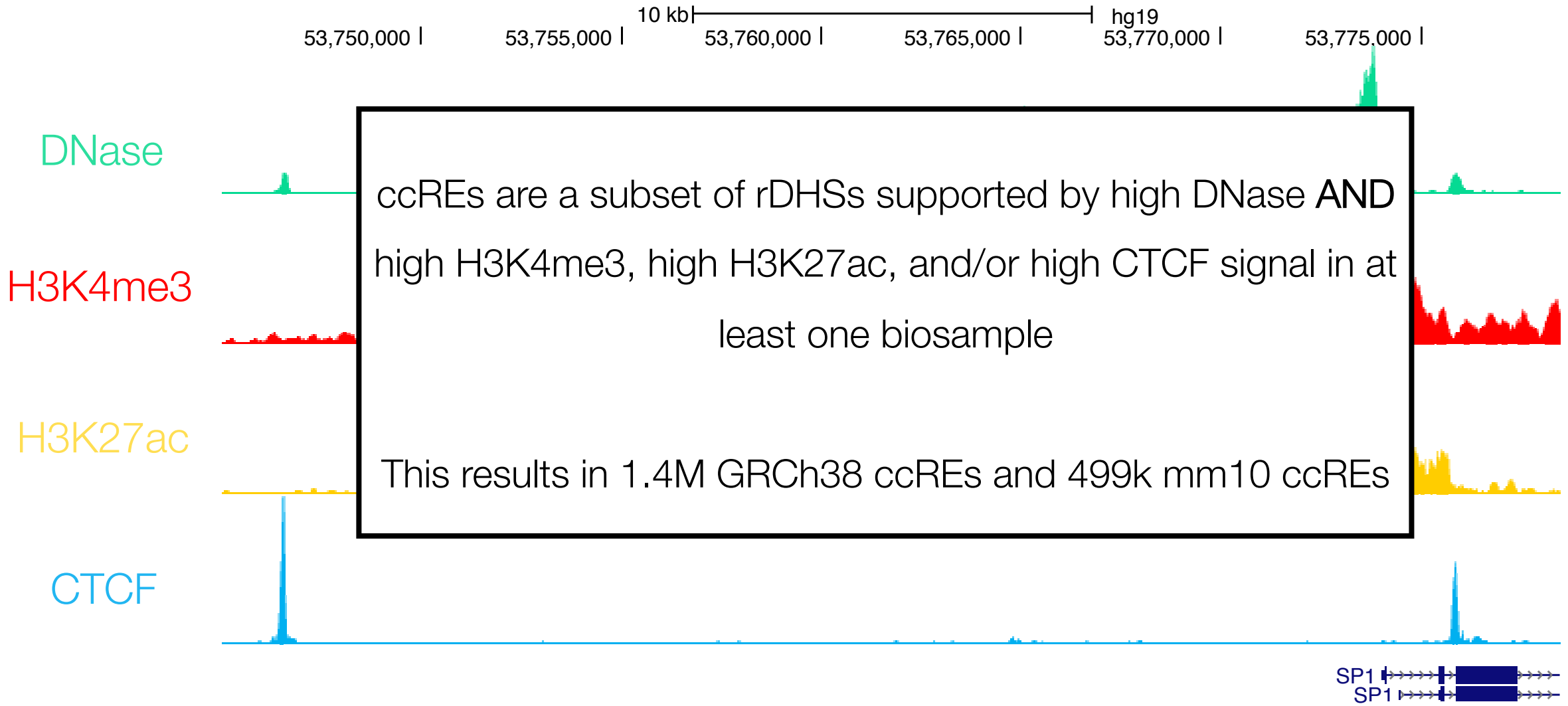
Distal regions are cell type specific but maintain same boundaries



We can represent chromatin accessibility across biosamples as a set of consensus sites called representative DNase hypersensitivity sites (rDHSs)

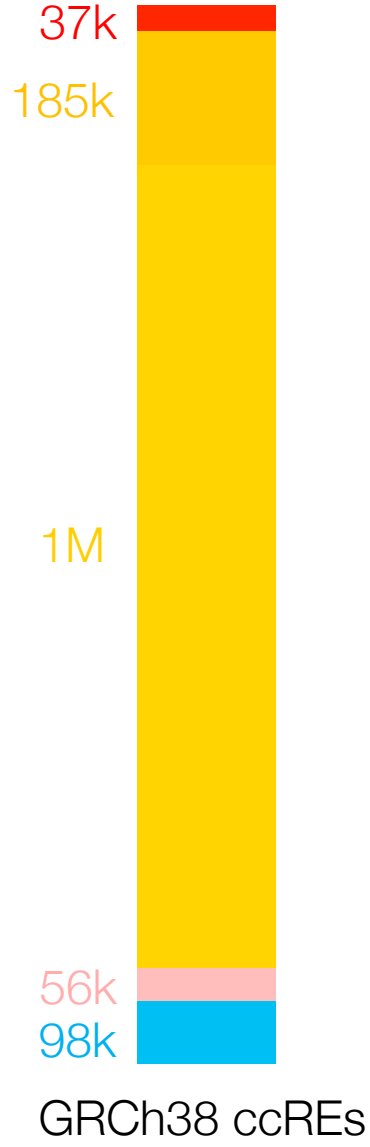


Annotating rDHSs with ChIP-seq signals to define ccREs

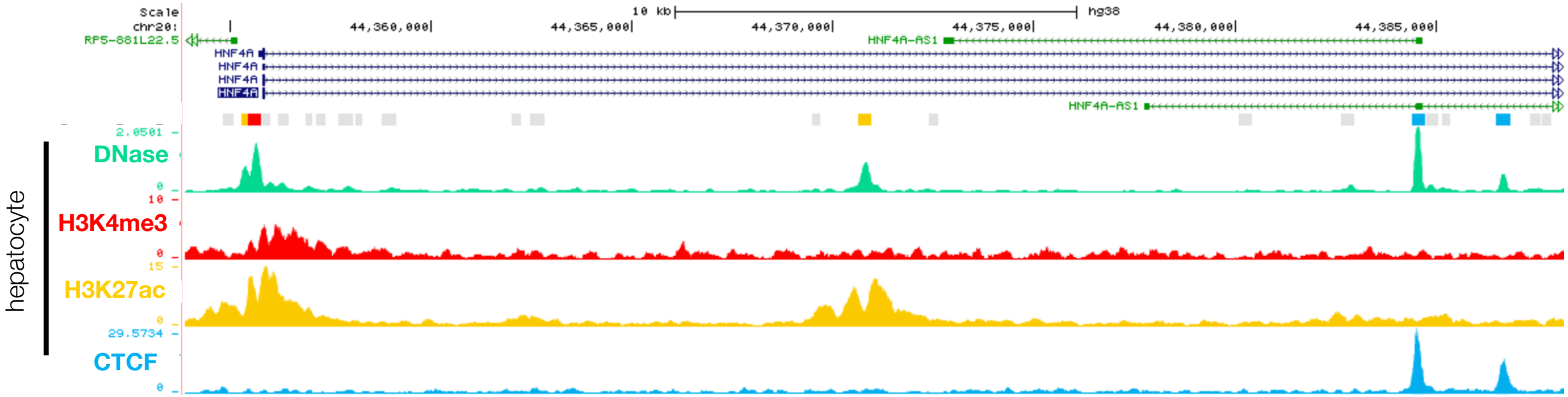


ccRE classification

- Promoters
- Enhancers
- Other regulatory elements



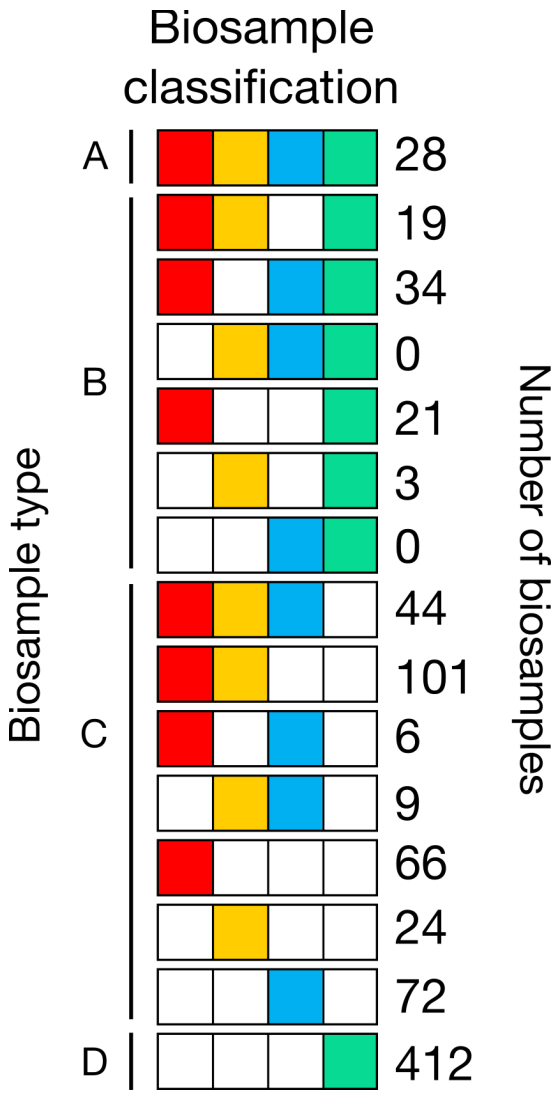
Classifying ccREs: cell type specific



hepatocyte specific ccREs (N=137k):

PLS	pELS	dELS	DNase-H3K4me3	CTCF-only	DNase-only	Low DNase
19,820	30,708	31,320	8,299	22,638	24,913	1,328,658

Classifying ccREs: partial data classification



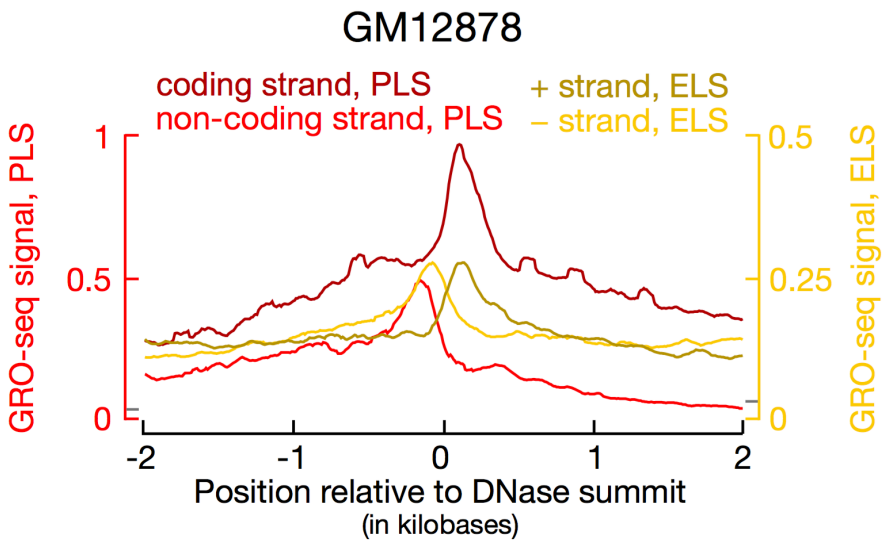
- For biosamples without all the four core marks, we implement a partial classification scheme
- Example: Spinal cord astrocyte 
  We can annotate:

  PLS
  DNase-H3K4me3
  CTCF-only
  DNase-only
  Low-DNase

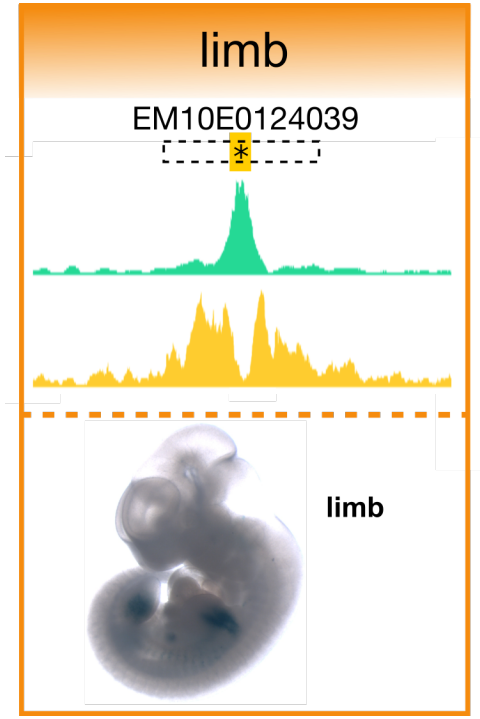
- Without DNase data we just annotate with high or low signal
- All missing data is marked in SCREEN (Henry will show in demo)

Orthogonal data support our classifications

- Transcription data:
  - RAMPAGE
  - CAGE
  - GRO-seq & PRO-seq

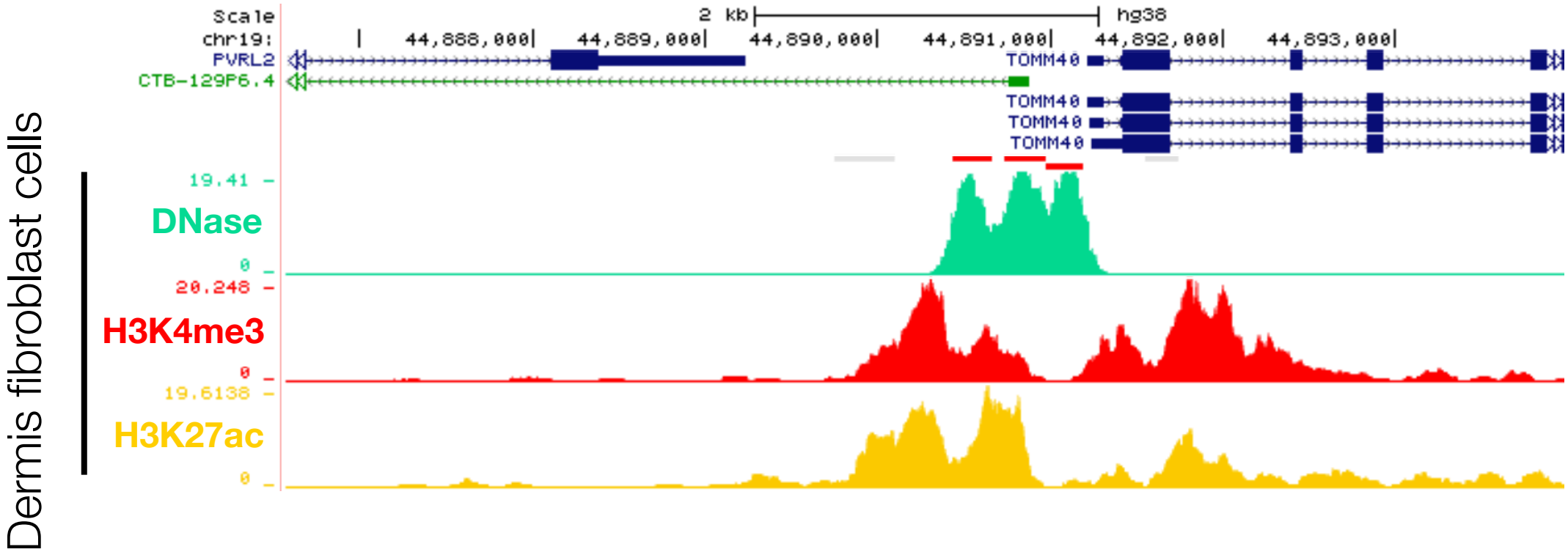


- Functional validation assays:
  - Mouse transgenic experiments
  - MPRA



Advantages to using the Registry of ccREs

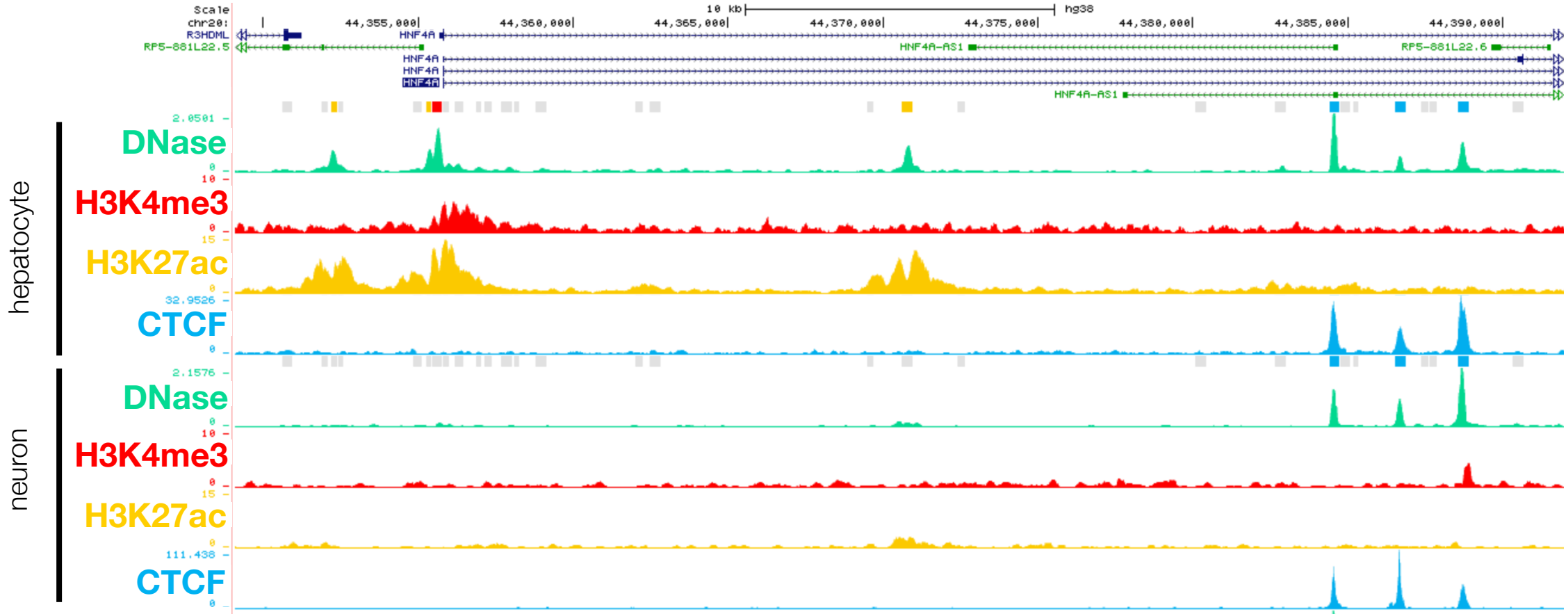
1. High resolution elements: widths between 150-350 bp



Advantages to using the Registry of ccREs

1. High resolution elements: widths between 150-350 bp

2. Boundaries of loci remain constant across hundreds of biosamples



Advantages to using the Registry of ccREs

1. High resolution elements: widths between 150-350 bp

2. Boundaries of loci remain constant across hundreds of biosamples

3. ccREs are accessioned

EH37XXXXXXX (human GRCh37/hg19 genome)

EH38XXXXXXX (human GRCh38/hg38 genome)

EM10XXXXXXX (mouse mm10 genome)

Advantages to using the Registry of ccREs

1. High resolution elements: widths between 150-350 bp

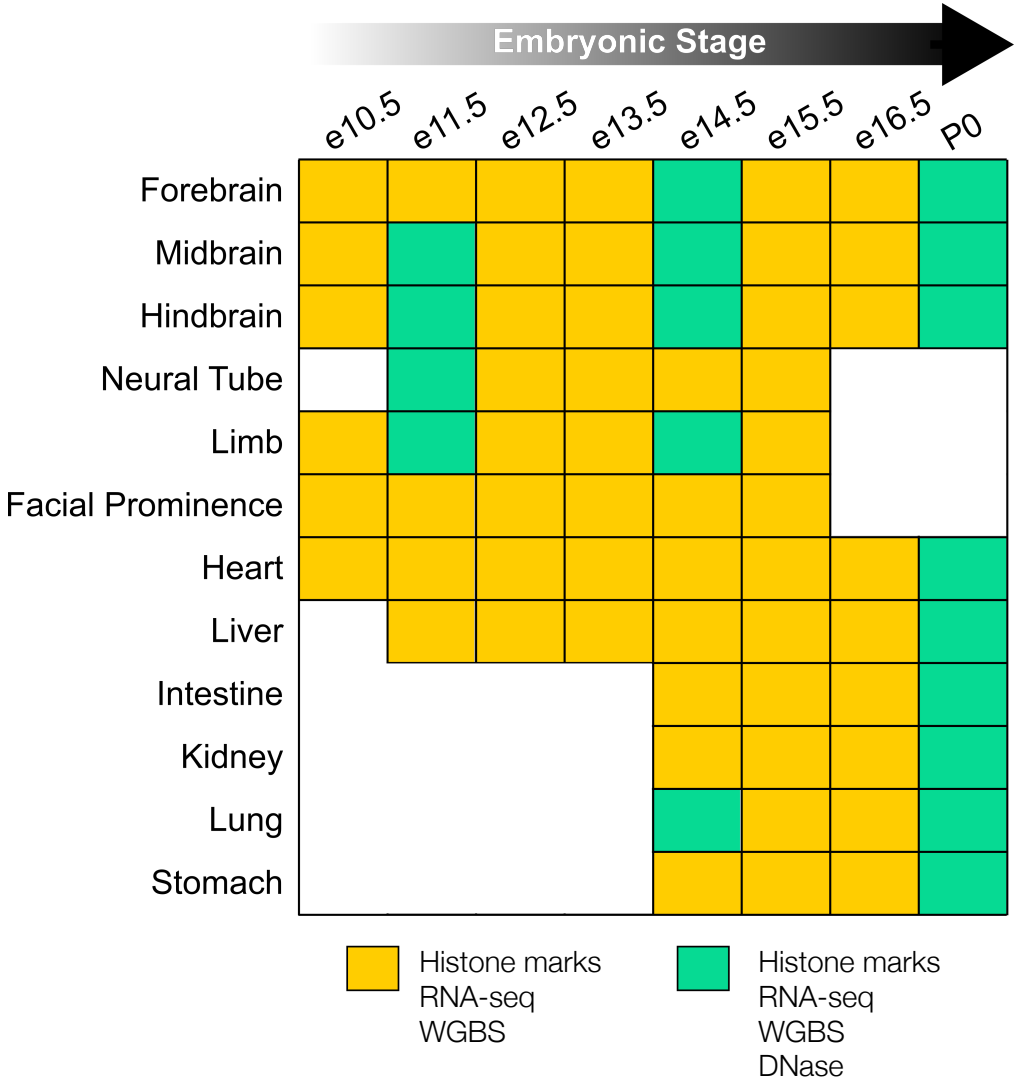
2. Boundaries of loci remain constant across hundreds of biosamples

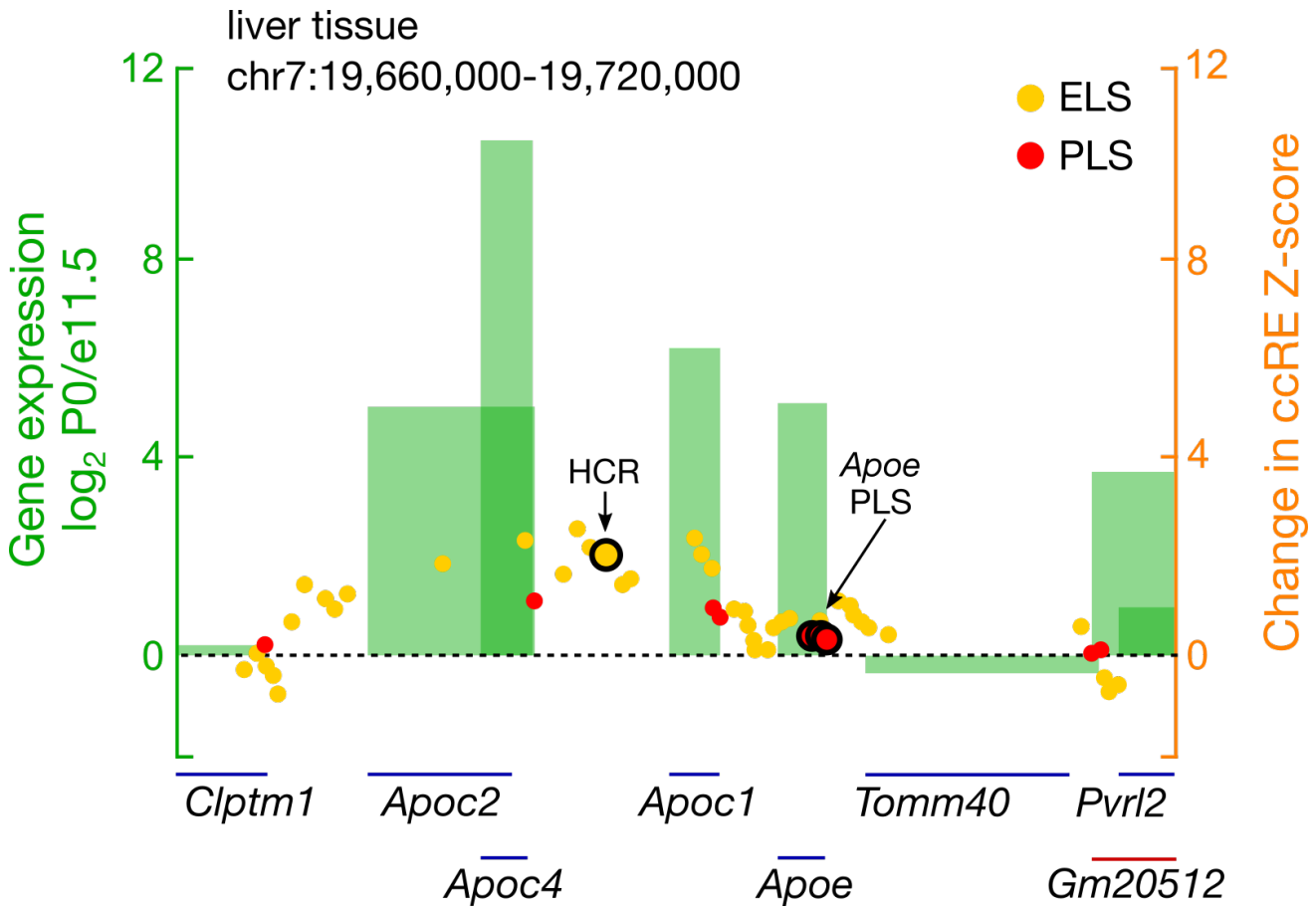
3. ccREs are accessioned

4. Easy data exploration and integration via SCREEN

Data integration examples

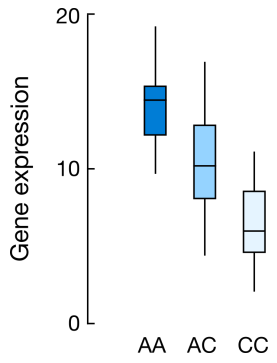
Differential gene expression and ccRE activity



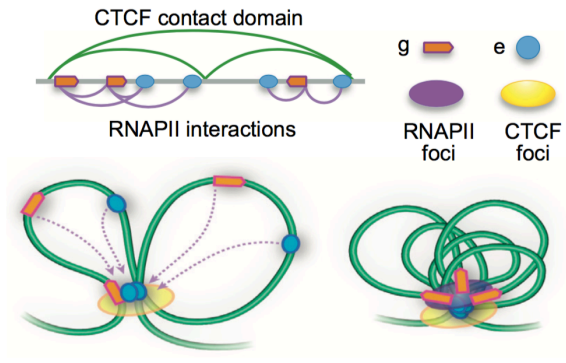


Gene-ccRE links

eQTLs



ChIA-PET



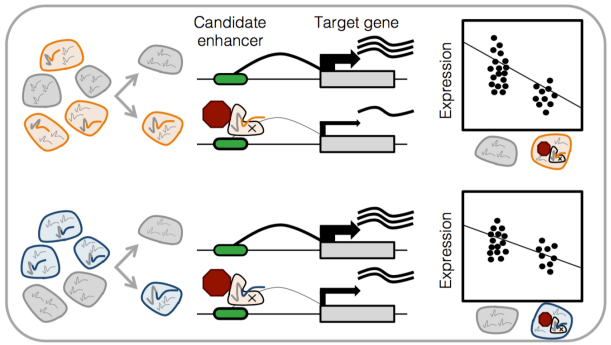
Tang...Ruan (2015) *Cell*

613k ccREs

36k genes

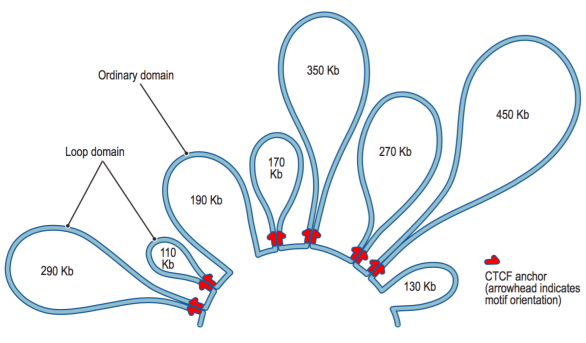
52 biosamples

crisprQTLs



Gasperini...Shendure (2019) *Cell*

Hi-C



Rao...Aiden (2014) *Cell*

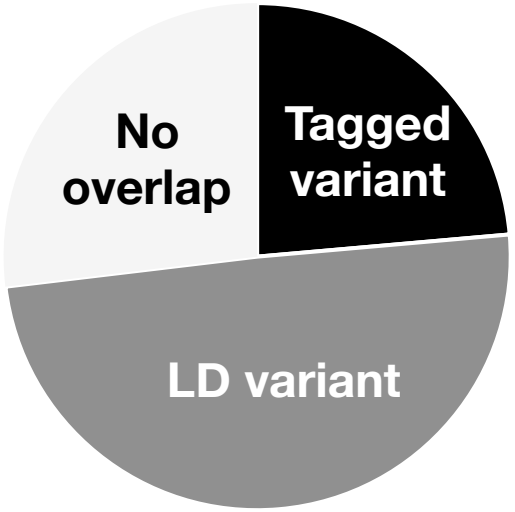
GWAS integration

NHGRI-EBI
GWAS Catalog

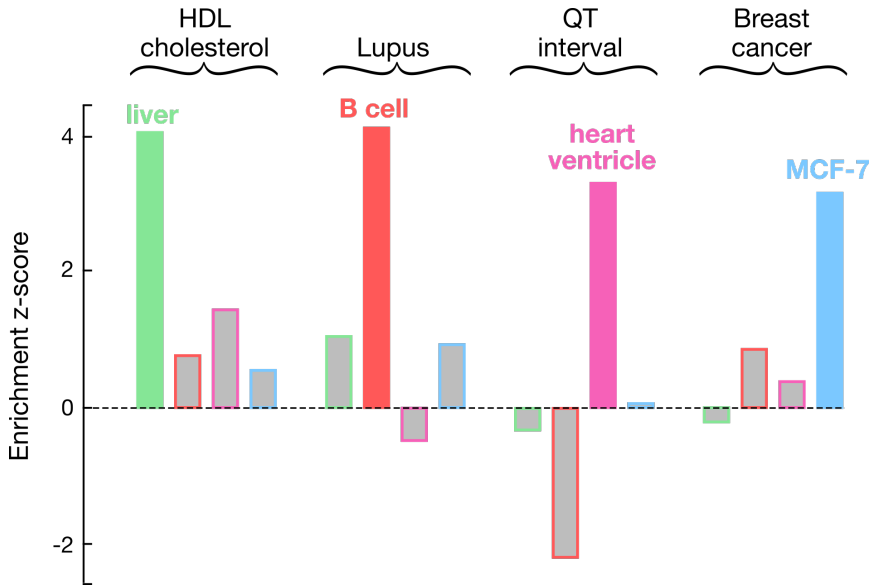
3,878 studies

397 studies with > 20 tagged SNPs

1) Predicting casual variants



2) Identifying disease relevant biosamples



Live demo