

EN-CODEC

v170421

DL

Σ # 1.10 (HL) Reconcile with the main ENCODE encyclopedia

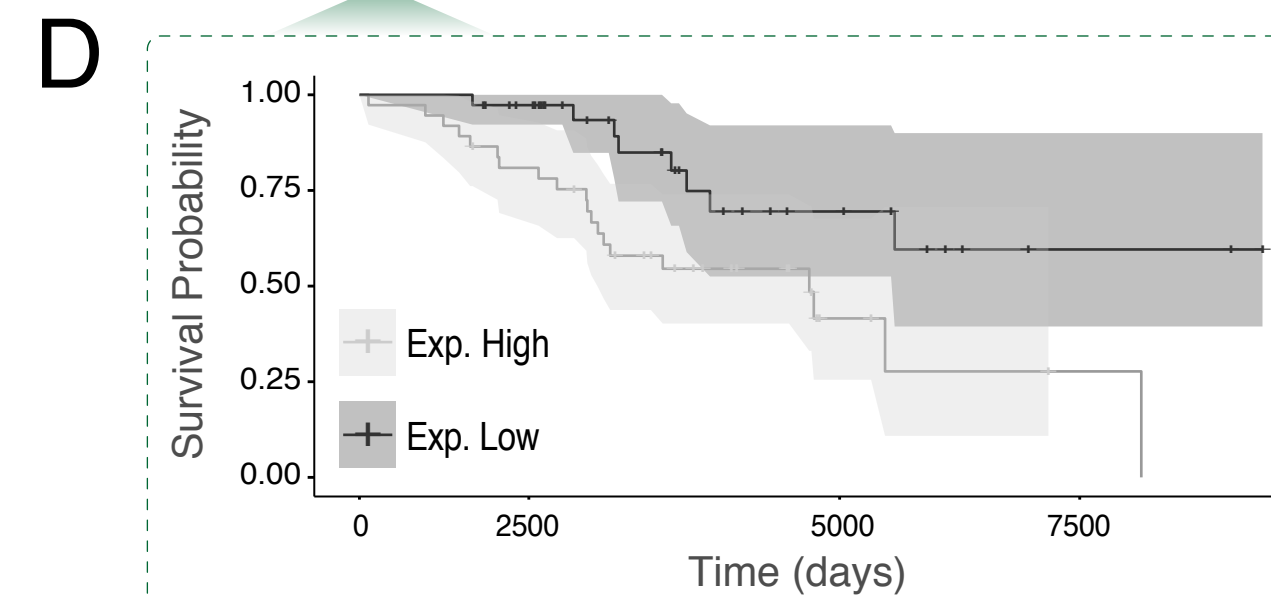
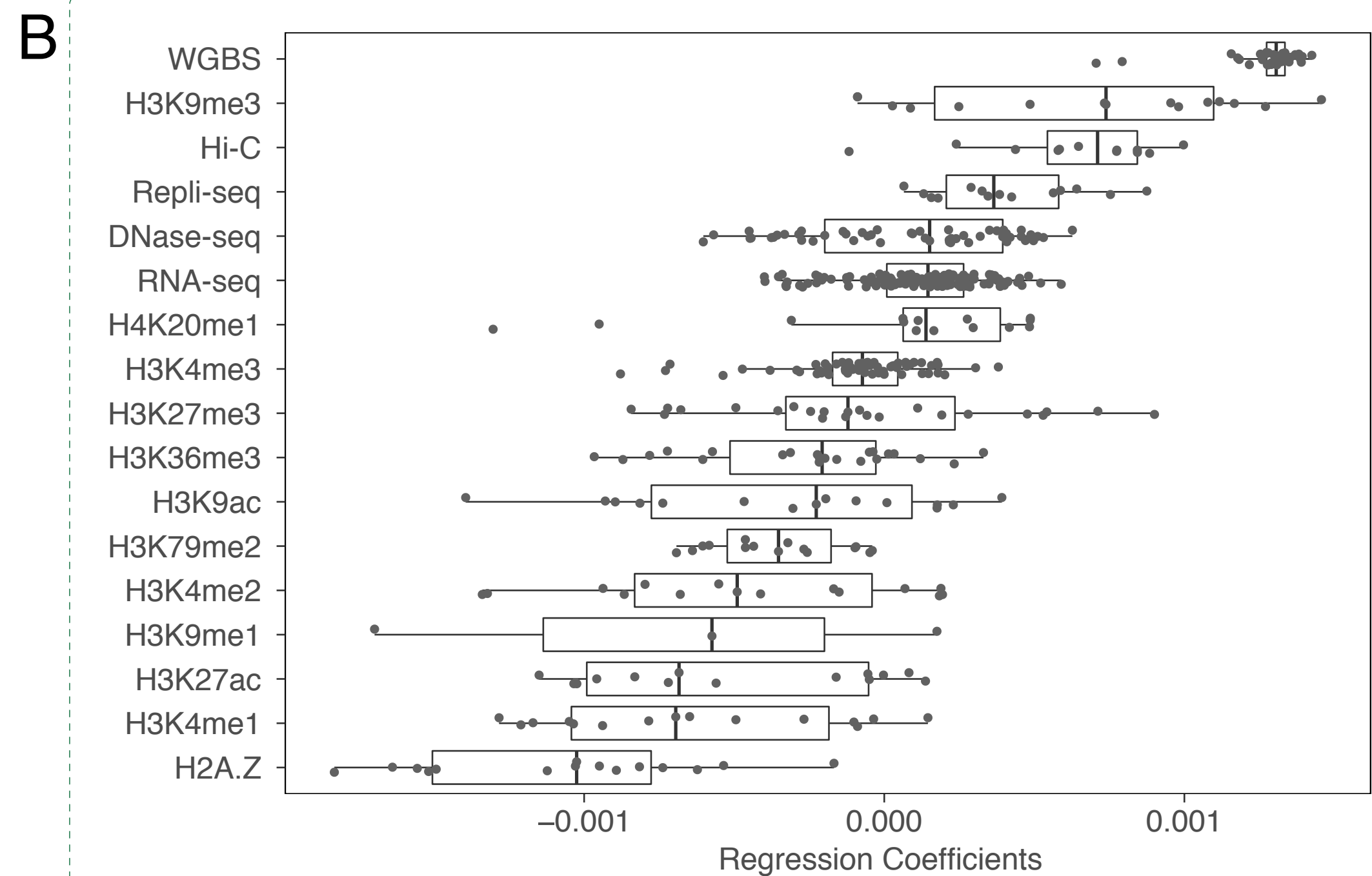
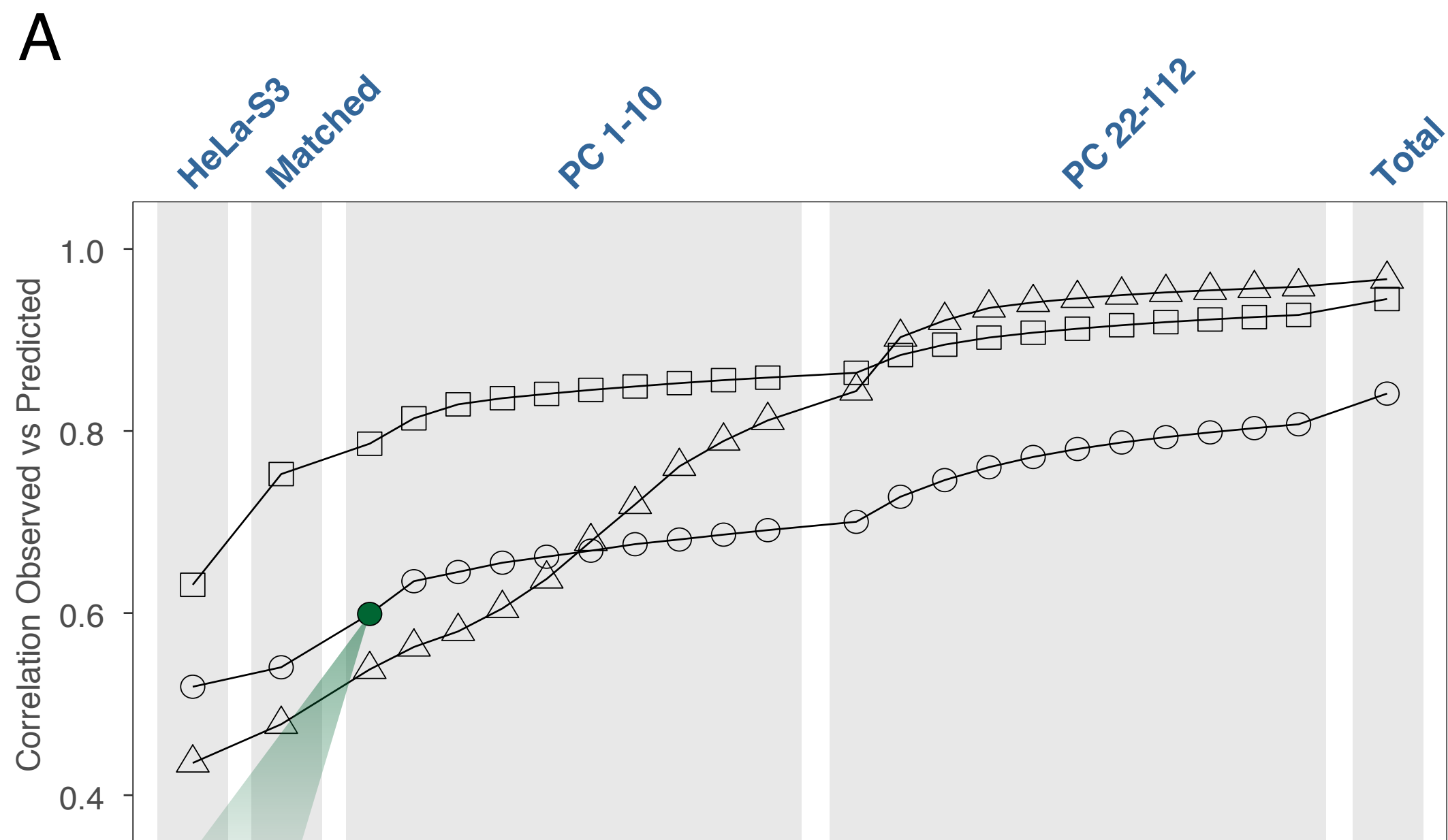
Both promoter and enhancer annotations from ENCODEC were carefully consolidated with the main ENCODE Encyclopedia resources. The ENCODE Encyclopedia comprises of three levels, two integrative levels of annotations and the ground level raw data. The ground level includes peaks and quantifications produced by uniform processing pipelines for individual data types. The integrative level contains annotations produced by integrating multiple data types. The core of the integrative level is the Registry of candidate Regulatory Elements (cREs). The registry contains approximately ~1.31M human cREs and each cRE has a cell-type non-specific accession number, which then can be browsed from SCREEN (Search Candidate Regulatory Elements by ENCODE, <http://screen.umassmed.edu/>).

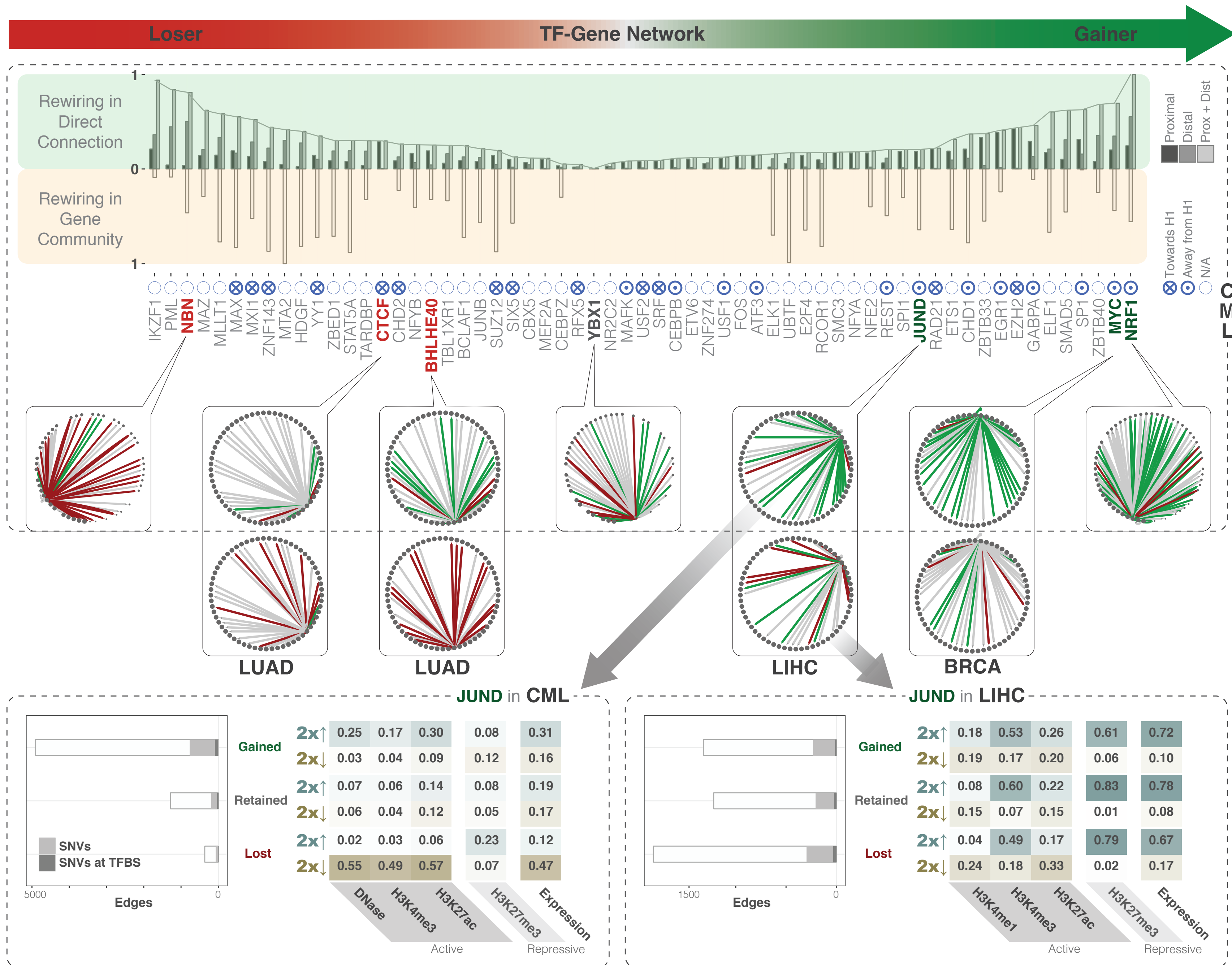
Annotations from ENCODEC were merged against the Registry of candidate Regulatory Elements (cREs). We assigned cell type non-specific cRE accession numbers to ESCAPE and CASPER integrated enhancer annotations when the region had more than 1bp overlap. When there were more than one accession numbers associated with the annotation, we assigned multiple accession numbers to the element. Overall, there was 99% overlap between integrated enhancer annotations and cREs with each element being mapped to 2.5 cRE accessions on average.

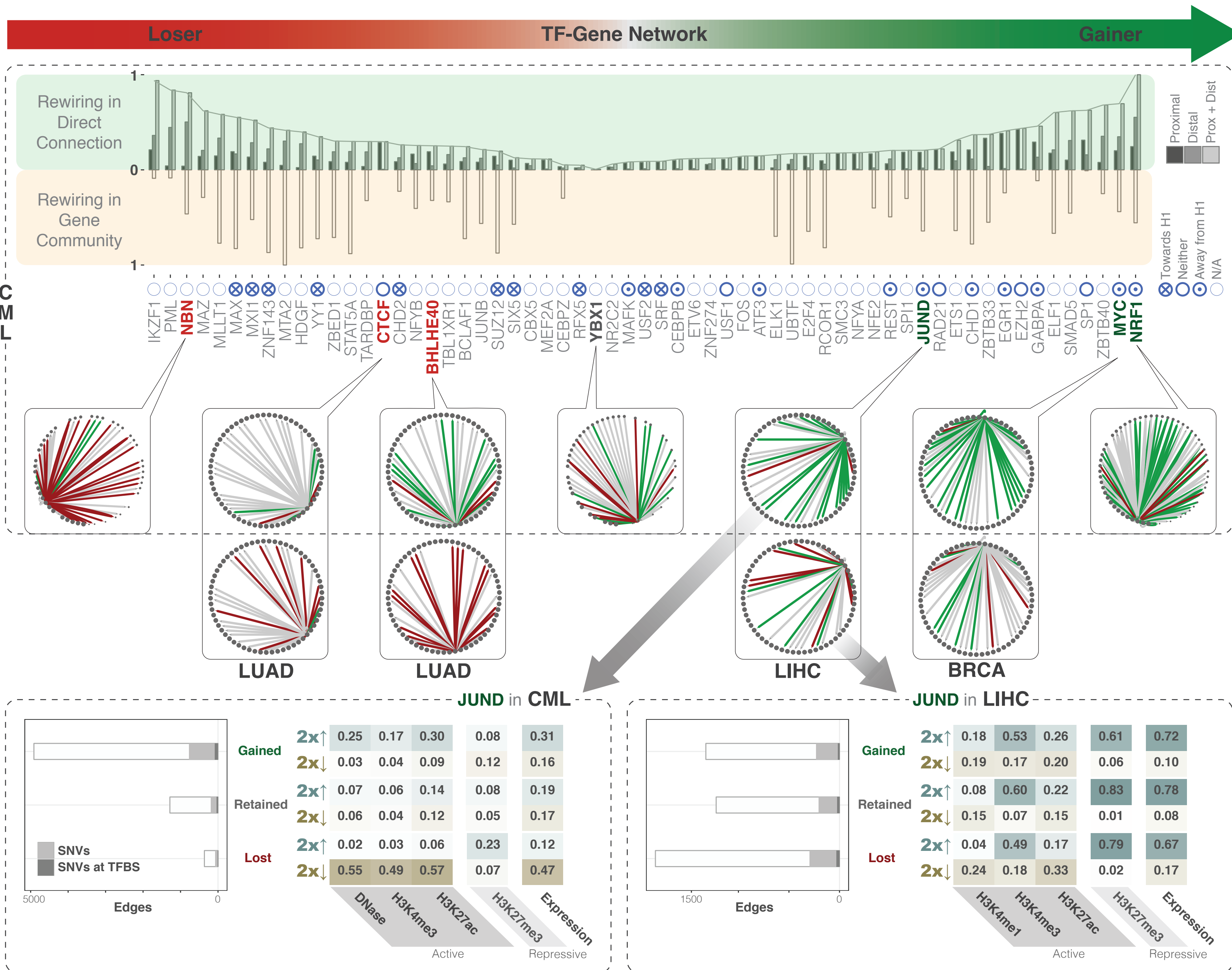
For cases without an overlap, we assigned special accession numbers EH37EXXXXXXXXX-C where XXXXXXXX are replaced with numbers starting from 0000001.

For cases without an overlap, we assigned special accession numbers EH37CXXXXXXXX where XXXXXXXX are replaced with numbers starting from 0000001.

To access the cRE using accession number, one can use the URL <http://screen.umassmed.edu/search/?q={accession}&assembly=hg19#> where {accession} is replaced with the cRE accession number. From SCREEN, one can look up H3K4me3, H3K27ac, CTCF, and DNase Z-scores and signal profiles across all available ENCODE cell







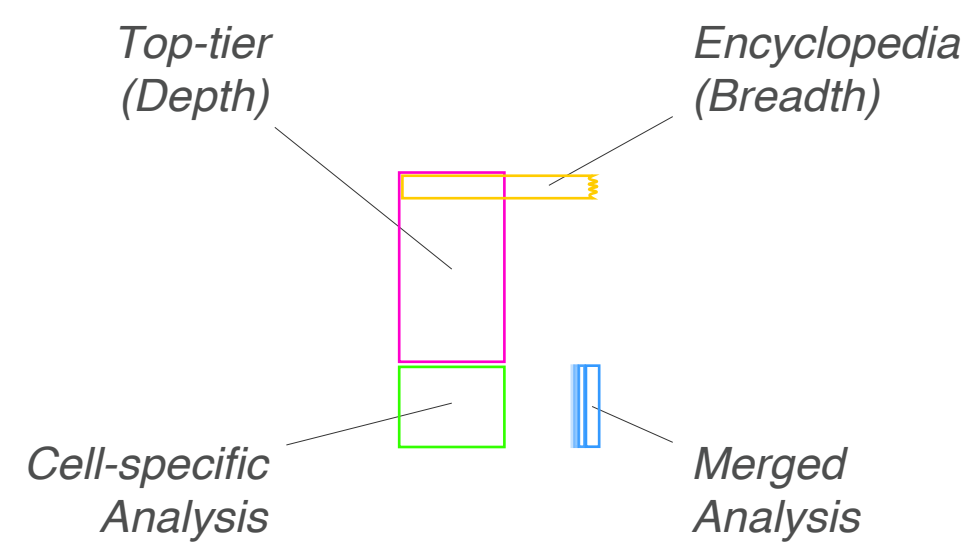
BIOSAMPLE →

← ASSAY

K562 *CD34+ PC GM12878* *HepG2* *Liver* *A549* *Lung IMR-90* *MCF-7* *HMEC MCF-10A* *HeLa-S3* *H1-hESC* *Caco-2* *HCT116* *Panc1* *LNcaP* *PC-3* *PC-9* *SK-N-MC* *DND-41* ...

		<i>T</i> Blood		<i>T</i> Liver		<i>T</i> Lung		<i>T</i> Breast		<i>T</i> Cervix	ES	COAD+READ		PAAD	PRAD	LUAD	SARC	LAML			
		<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>												
Chromatin Accessibility DS	DNase-seq	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
	Histone	12	6	11	11	7	11	7	28	5	11	3	11	11	3	12	6	1	13	13	7
Transcription TX	RNA-seq	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
	RAMPAGE	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•			
RNA-binding Proteins RP	eCLIP	89		70																	
shRNA/siRNA Knockdown KD	TF KD	85		61				2													
	RBP KD	234		225																	
3D Chromatin Structure 3D	ChIA-PET	3	3	1				4		2											
	Hi-C	▲	▲			•	▲	▲	▲		▲										
Enhancers SS	STARR-seq	•	•	▲				•													
Methylation ME	WGBS	•	▲	•	▲		▲	▲	▲		•										
	RRBS	•	▲	•	•	•	▲	•	•	•	•	•									
Replication Timing RT	Repli-seq/chip	•	•	•				•		•	•										
Transcription Factors TF	TF Total	207	99	95	7	31	1	9	52	2	4	59	49								
	TFSS	125	68	69	6	23	1	6	34	1	3	34	32								
	Chromatin Remodeller	31	11	13		3			7	1		9	9								
	Cofactor	20	9	7	1	3		2	4			6	3								
	General (GTF)	17	6	4		2		1	2		1	10	5								
	Other	14	5	2					5												
Cell Line WGS WG	SNV	▲	▲					▲		▲											
	SV	▲	▲					▲		▲											
Patient Data PD	Cohort	150			82	197		116													
	Expression	173		373		517		1100		546											

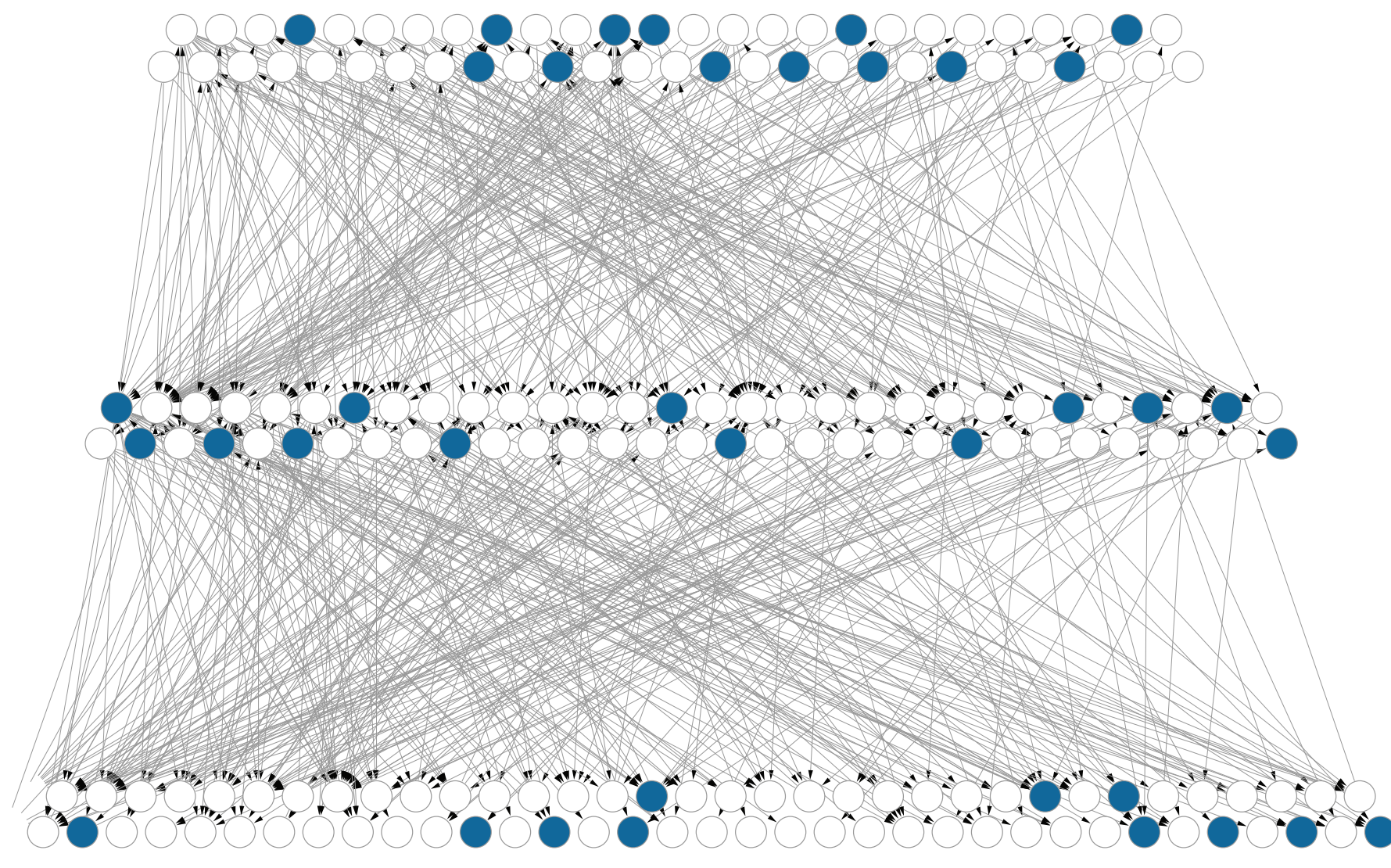
EN•CODEC



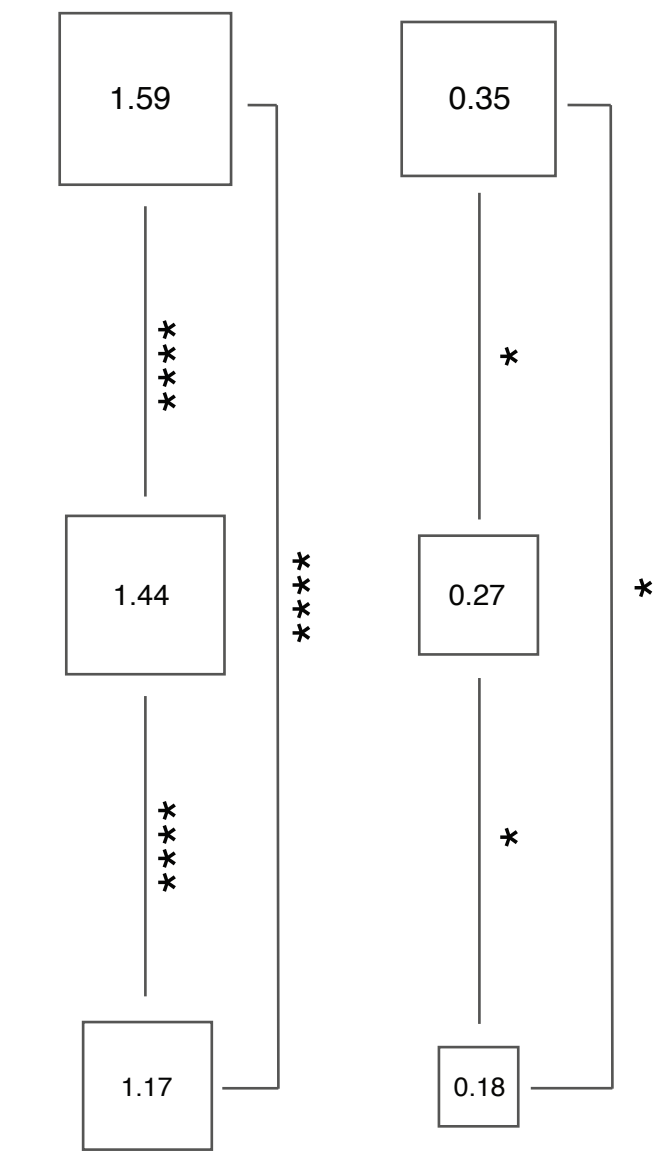
● Available in ENCODE
▲ Available as External Resource

ESCAPE (SS)	•	•						•										
CASPER (TF+DS)	•	•	•					•		•		•	•					
JEME (3D)	•	•	•					•		•		•	•					
Background Mutation Rate & Burdening																		•
TF/RBP Network	•	•	•	•	•	•	•	•	•	•	•	•	•					•
Network Hierarchy	•																	•
Network Rewiring	•		•		•		•		•									•
Expression Correlation & Network Motif																		•
TF Motif Disruption	•		•		•		•		•		•							•

A



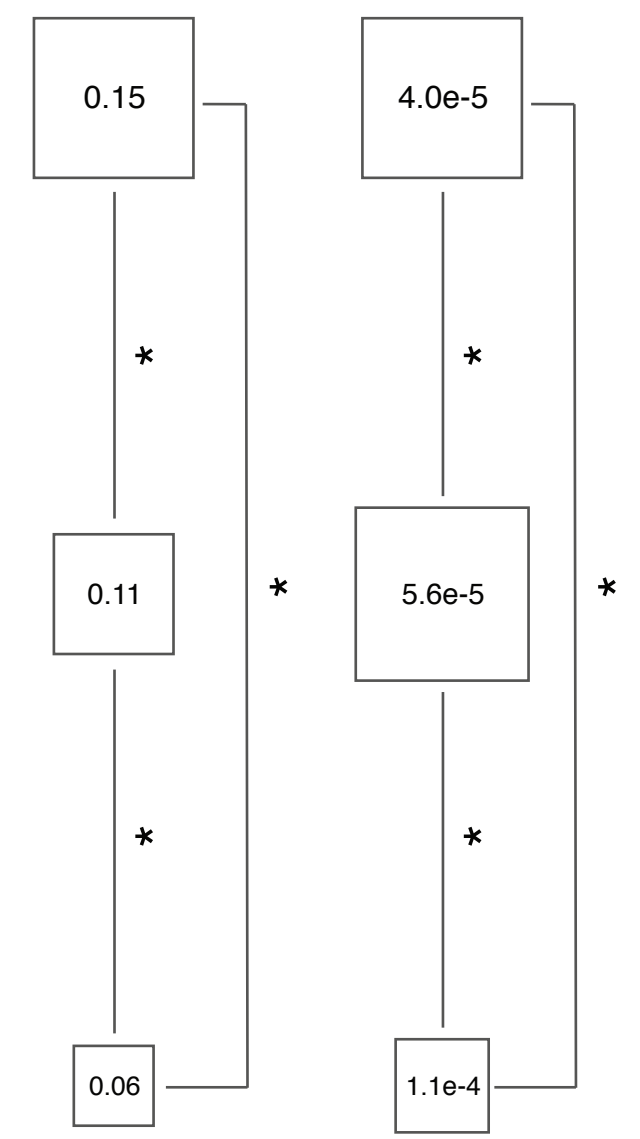
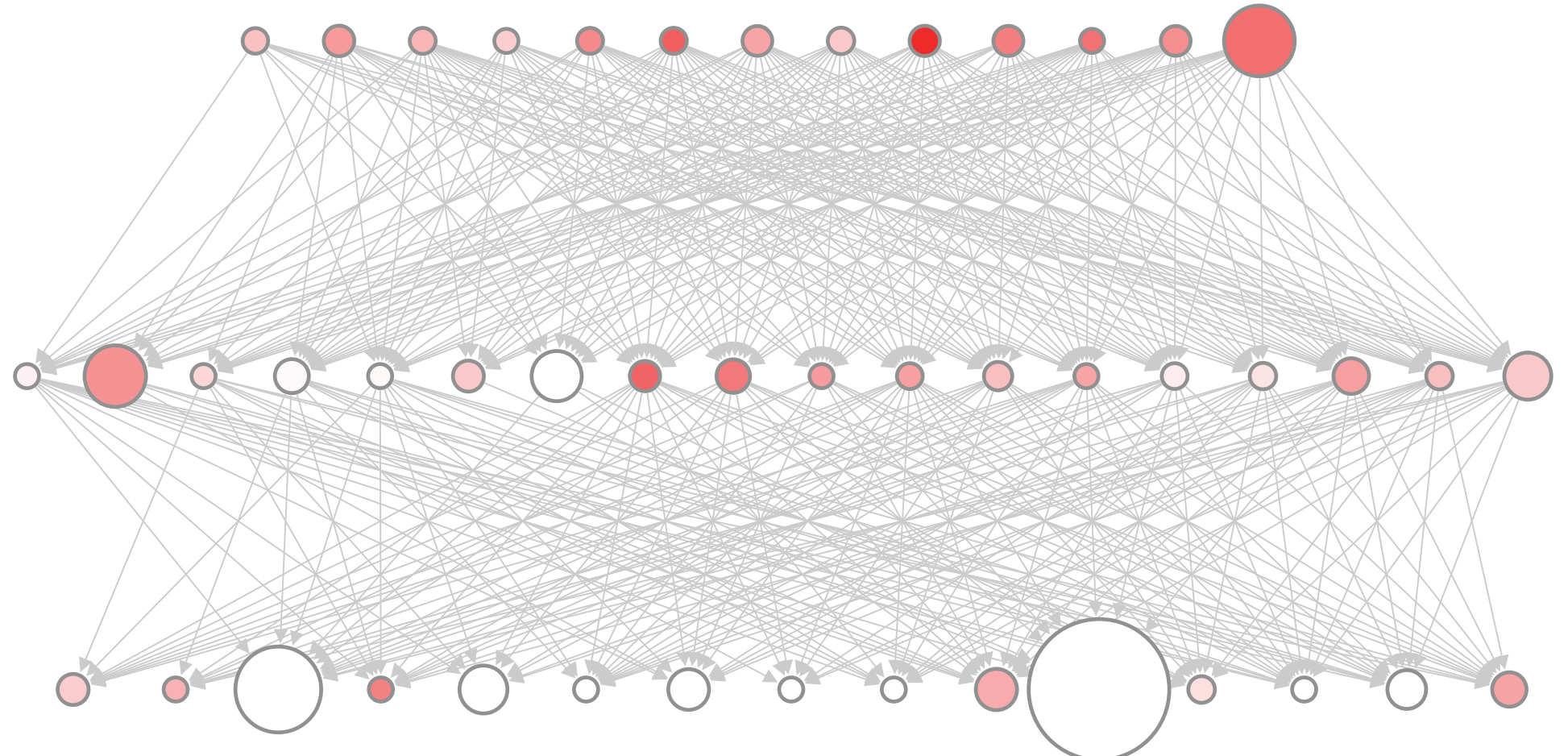
● Cancer TFSS



Target Expression Correlation

Percent Cancer TFSS

B



Expression Correlation

Percent Burdened TFBS