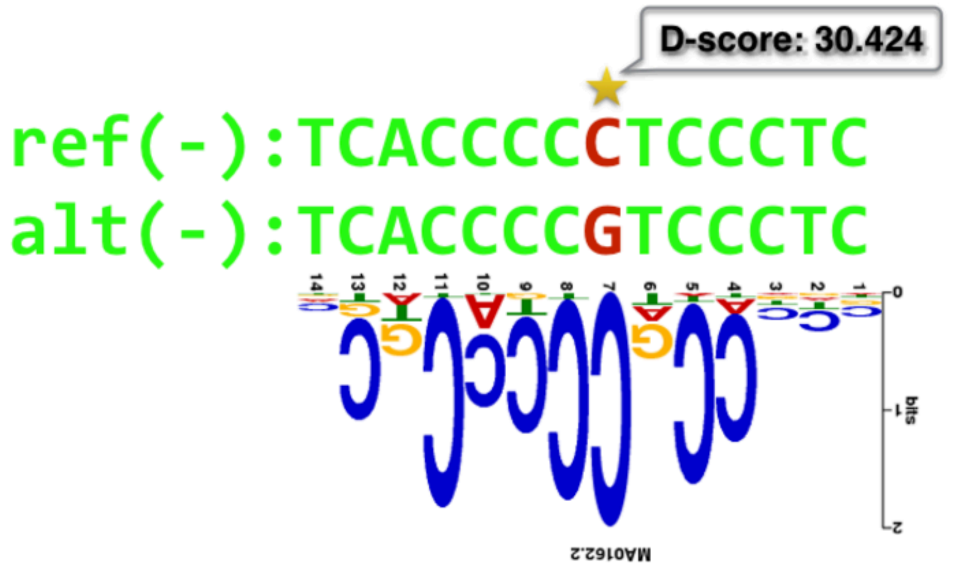
TIP+Enhancer Net

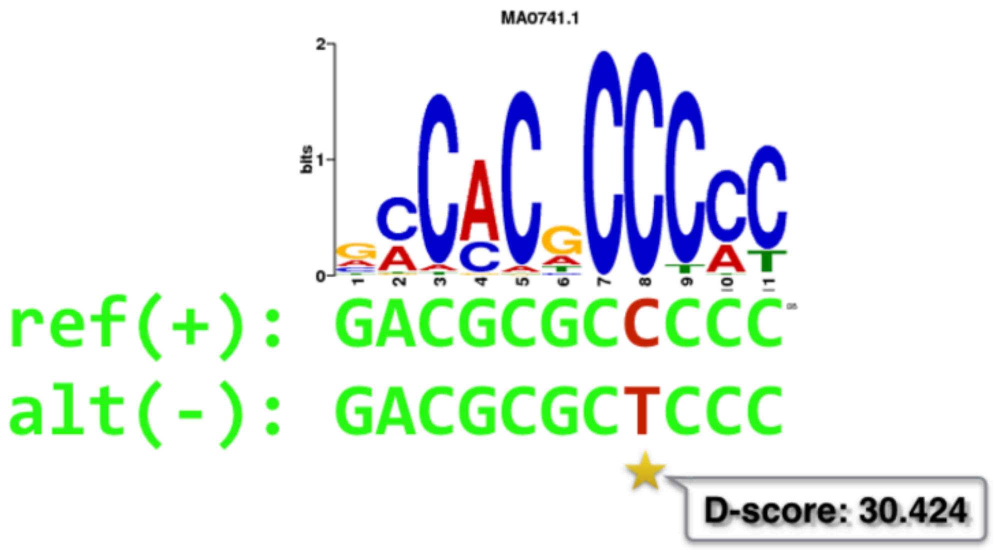
(a)
Color Key and Histogram
Count
0 1 2 3 4 5
0 0.2 0.4 0.6 0.8 1
Value
ZBTB40
SPI1
BHLHE40
RAD21
SRF
MYC
ELF1
EZH2
SMAD5
GABPA
MAFK
ATF3
REST
CEBPB
E2F4
NFYA
UBTF
RCOR1
ETV6
ZBTB33
SP1
ETS1
ELK1
CHD1
ZNF274
MAX
PML
SMC3
NFE2
JUND
NR2C2
NRF1
EGR1
MAZ
IKZF1
MEF2A
BCLAF1
TBL1XR1
TARDBP
NBN
MXI1
ZBED1
STAT5A
JUNB
HDGF
MLLT1
MTA2
CBX5
YBX1
CHD2
CTCF
RFX5
ZNF143
SUZ12
NFYB
YY1
FOS
USF1
USF2
SIX5
CEBPZ
Dynamic
Enhanced
Suppressed
Static
K
com
G
(b)
ZNF274
JUND
IKZF1
ZNF143
(c)
Dynamic
Enhanced
Static
Suppressed
0.0 0.2 0.4 0.6
0.00 0.25 0.50 0.75 1.00
Gain
Loss
Common
(d)
TPM log10 fold change
DNase log10 fold change
H3K27me3 log10 fold change
H3K4me3 log10 fold change
5mC log10 fold change
Histogram
Dynamic
Enhanced
Static
Suppressed
0.00 0.04 0.08 0.12
0.0 0.1 0.2 0.3 0.4
0.0 0.2 0.4 0.6 0.8 1.0
-4 -2 0 2 4

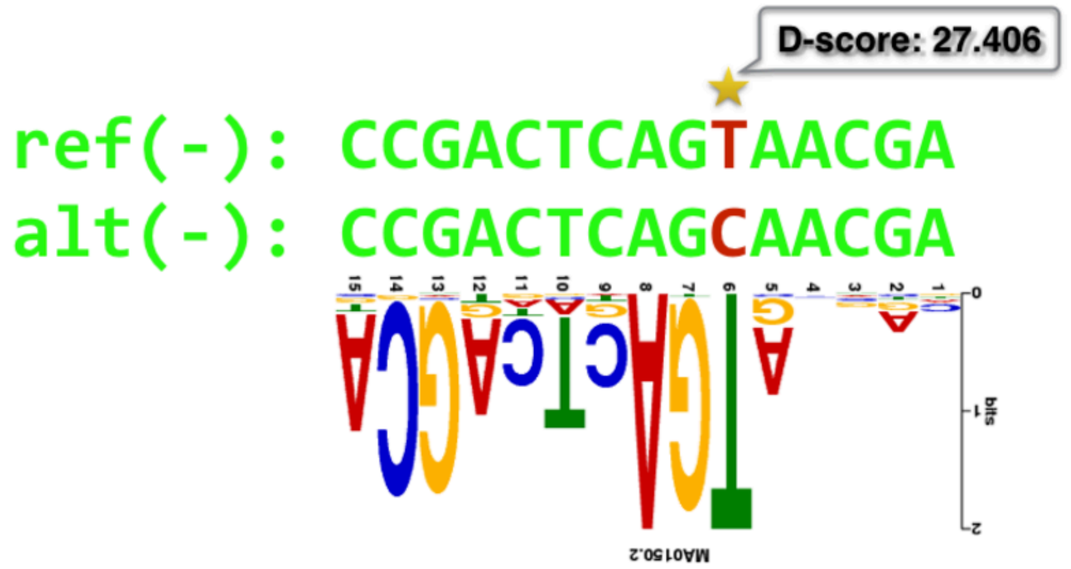
MotifTools

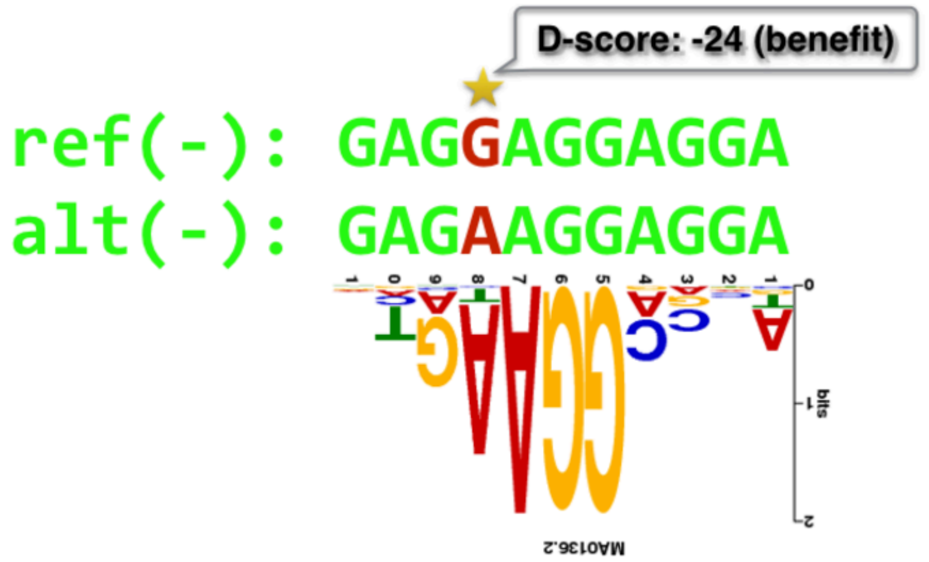
A collection of python tools to evaluate a variant on TF motif

- D-score: motif-breaking or motif-gaining power
- B-score: burden score









How to interpret D-score

D-score is a motif "[D]isruptive score" of a variant. It is calculated by difference between P-value between reference genome and alternate genome.

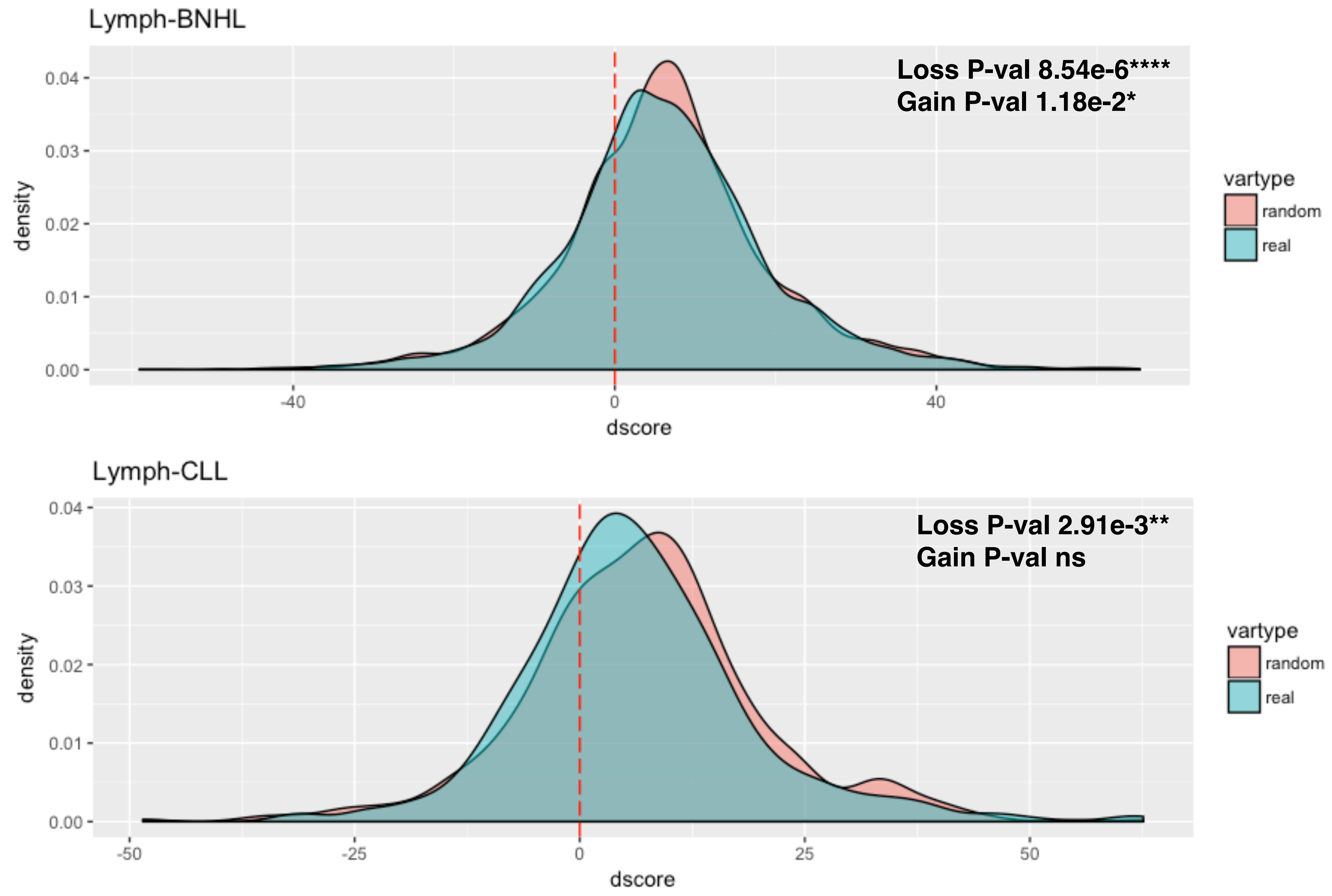
```
D-score = [-10 * log(P-val_Ref)] - [-10 * log(P-val_Alt)]
D-score = -10 * log(P-val_Ref/P-val_Alt)
```

- *Positive D-score denotes a variant is decreasing the likelihood of TF to bind the motif (motif-break)*
- *Negative D-score denotes a variant is increasing the likelihood of TF to bind the motif (motif-gain)*

Table X. Comparison between real and simulated variants effect on TFSS motif

		Lymph-BNHL		Lymph-CLL	
		Real (PCAWG)	Random (Sanger Neutral)	Real (PCAWG)	Random (Sanger Neutral)
GM12878 (n=50 TFSS)	# var within peak	111,423	79,211	13,930	11,995
	# var w/ motif loss	1,627	1,447	212	250
	# var w/ motif gain	732	571	99	109
K562 (n=79 TFSS)	# var within peak	156,679	134,397	20,829	22,459
	# var w/ motif loss	2,921	2,822	414	544
	# var w/ motif gain	1,138	1,087	171	202

Fig X. Distribution of motif disruption scores of real and simulated variants in K562



Fisher's Exact Test
ns P > 0.05, * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001, **** P ≤ 0.0001

Fig X. TF ChIP-seq peaks with motif variants in K562

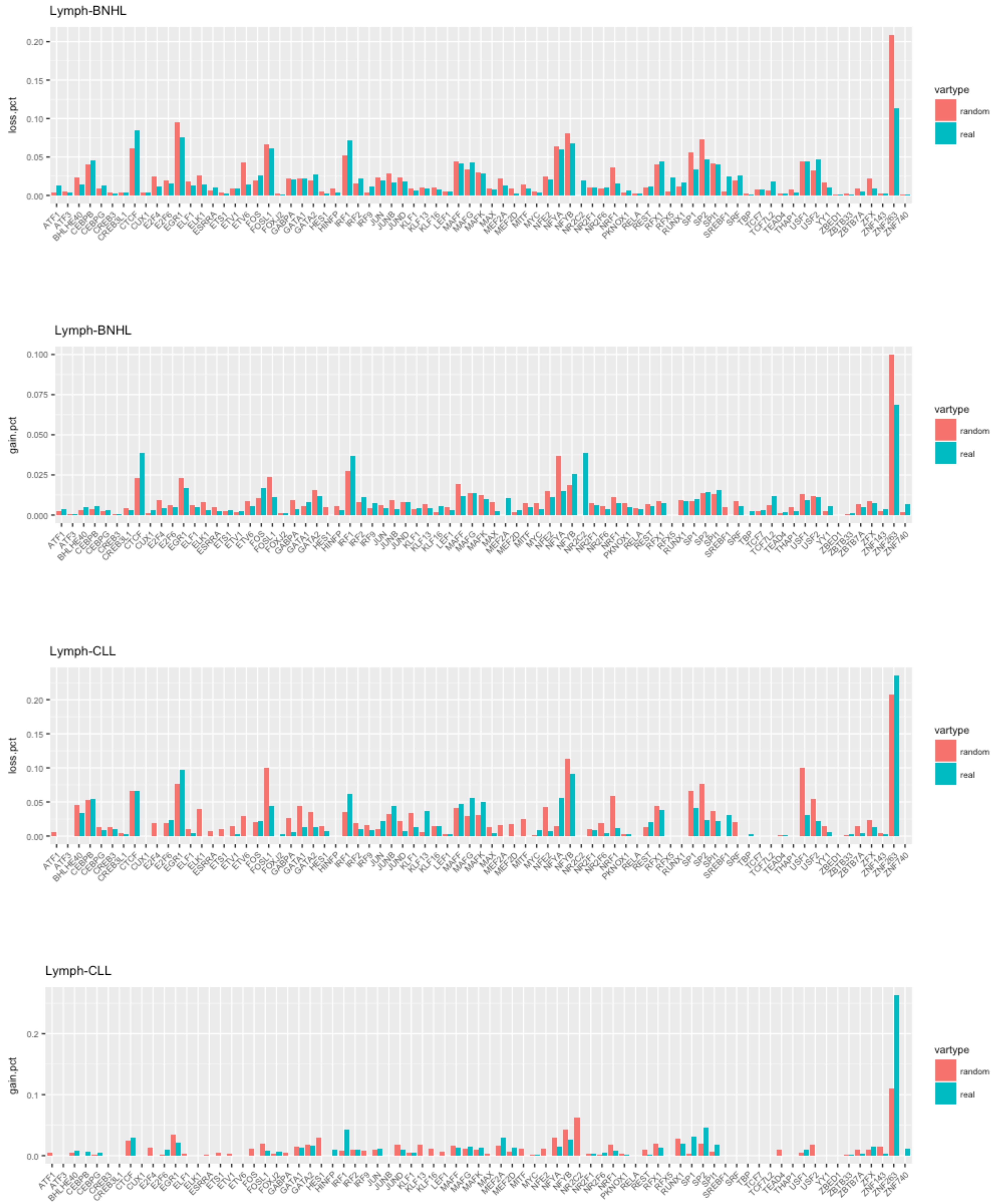
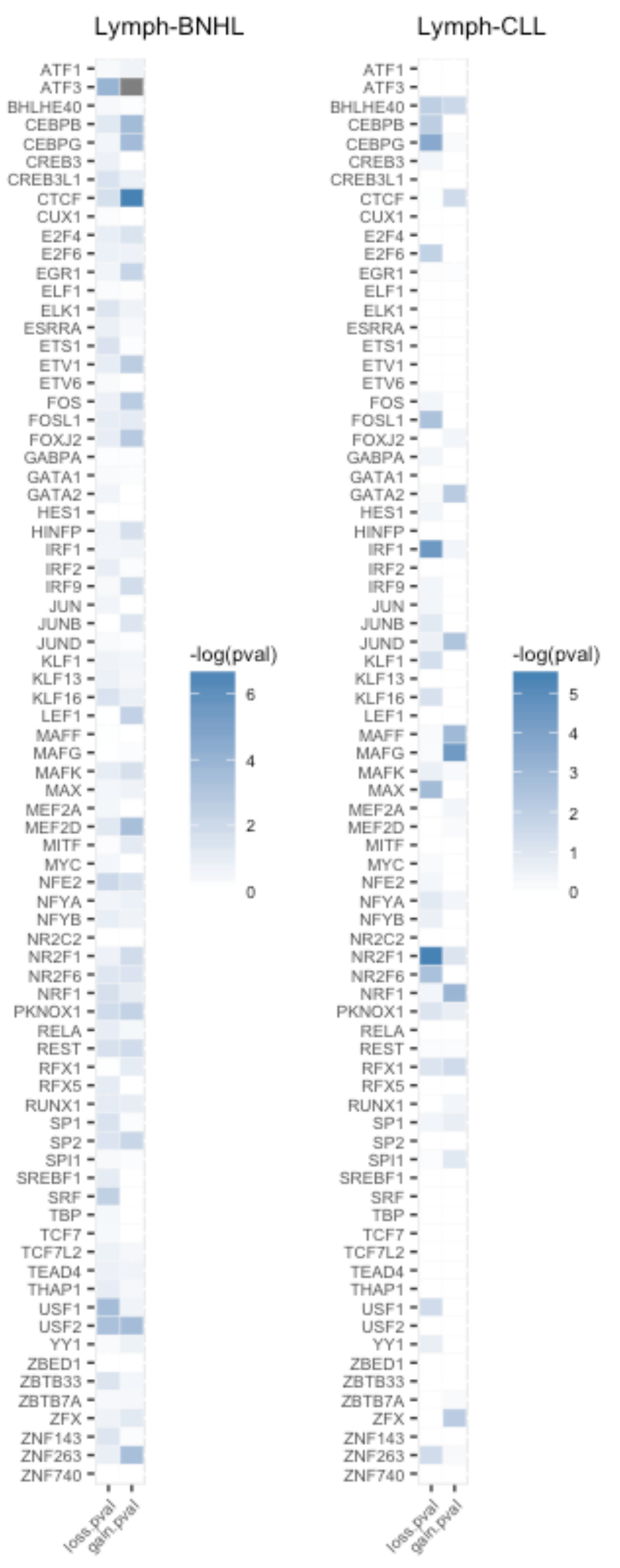


Fig X. Comparison between real and simulated motif variants in K562



Wilcox Rank Sum Test

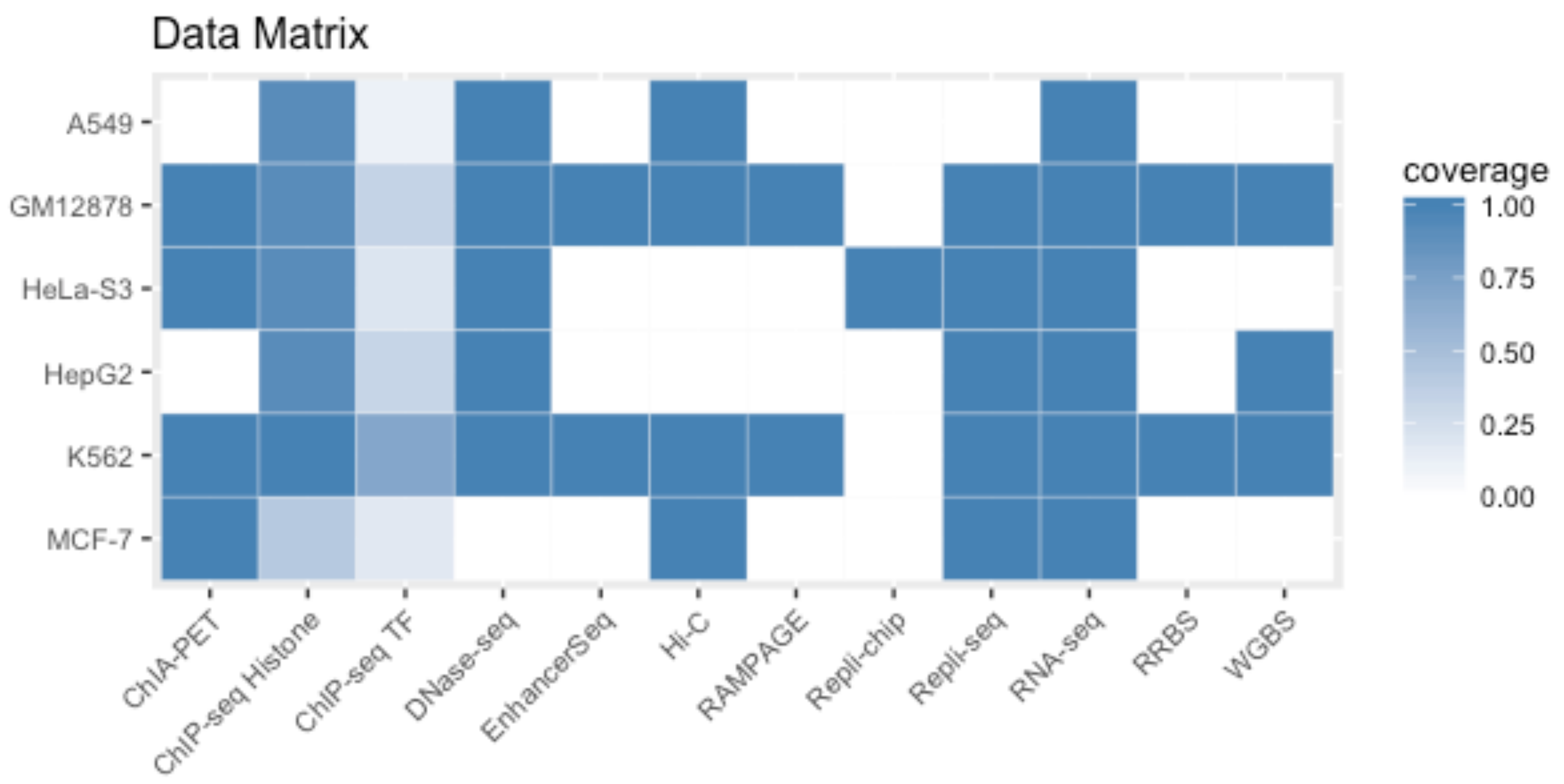
Table X. TFSS network edges with motif gain/loss in K562

			Lymph-BNHL			Lymph-CLL	
	# TSS edges	w/ variants	w/ motif loss	w/ motif gain	w/ variants	w/ motif loss	w/ motif gain
ARID3A	2249	352	0	0	33	0	0
ARNT	3332	773	0	0	48	0	0
ATF1	5153	409	5	3	50	0	0
ATF3	6459	339	1	1	46	0	0
BHLHE40	8211	884	9	3	84	0	2
CEBPB	7105	599	12	0	86	3	0
CEBPG	4548	755	5	0	107	1	1
CREB3	2835	596	2	0	45	0	0
CREB3L1	8408	1719	6	4	139	0	0
CTCF	9820	795	35	24	70	5	1
CUX1	1199	404	1	0	24	0	0
E2F4	6942	791	10	2	63	0	0
E2F6	11089	1505	17	7	121	4	0
EGR1	11825	1096	85	8	103	11	3
ELF1	7094	1126	19	8	95	0	0
ELK1	2722	370	8	1	26	0	0
ESRRA	8732	1476	7	2	124	0	0
ETS1	7354	1432	5	8	88	0	0
ETV1	7566	993	13	5	88	1	0
ETV6	1236	212	4	2	19	0	0
FOS	3694	190	1	0	25	0	0
FOSL1	2887	193	2	0	21	2	1
FOXJ2	3765	1069	1	0	79	0	1
GABPA	8108	807	26	3	66	0	0
GATA1	1199	274	4	3	21	0	0
GATA2	3503	417	8	3	42	0	2
HES1	2696	547	2	0	31	0	0
HINFP	2287	233	1	1	27	0	1
IRF1	2568	771	39	37	62	5	1
IRF2	3295	448	2	1	41	1	0
IRF9	1741	173	2	2	30	1	0
JUN	3044	138	3	1	16	1	0
JUNB	1241	199	1	0	20	0	0
JUND	9641	973	0	2	110	1	3
KLF1	4865	1132	8	4	115	0	0
KLF13	5415	673	5	0	51	1	0
KLF16	4395	618	4	2	48	3	0
LEF1	4524	895	4	2	62	0	0
MAFF	3769	272	8	3	43	2	0
MAFG	4877	689	32	14	61	4	2
MAFK	4549	455	7	1	53	2	1
MAX	11372	1725	10	5	151	0	0
MEF2A	1787	195	6	3	14	0	0
MEF2D	3541	503	3	2	56	0	0
MITF	2730	348	3	2	29	0	0
MYC	10159	1713	2	4	159	0	0
NFE2	4953	370	3	3	31	0	0
NFYA	3070	210	9	2	36	2	0
NFYB	4436	275	9	6	34	3	0
NR2C2	578	51	2	3	4	0	0
NR2F1	6825	1038	6	8	105	1	0
NR2F6	4965	830	6	2	91	0	0
NRF1	11691	2992	35	23	259	2	4
PKNOX1	9766	1360	3	2	139	0	0
RELA	1912	362	0	0	31	0	0
REST	9488	1545	11	11	145	4	0
RFX1	4797	578	17	3	60	4	0
RFX5	1290	185	4	0	18	0	0
RUNX1	2240	151	5	2	17	0	0
SP1	8349	1039	32	11	111	6	4
SP2	2928	176	6	2	24	1	1
SPI1	7103	454	15	6	50	0	2
SREBF1	1465	111	1	0	22	1	0
SRF	2961	188	1	0	23	0	0
TBP	9695	1117	1	2	105	0	0
TCF7	1898	389	3	2	38	0	0
TCF7L2	1246	110	2	1	17	0	0
TEAD4	6803	1119	3	2	116	0	0
THAP1	3252	279	1	1	30	0	0
USF1	6107	439	10	3	42	0	0
USF2	1732	296	6	1	18	0	0
YY1	8490	807	8	6	74	0	0
ZBED1	1975	869	1	0	34	0	0
ZBTB33	8049	1245	7	7	123	0	0
ZBTB7A	9733	1485	12	9	122	0	0
ZFX	10378	3457	29	25	240	5	5
ZNF143	8322	780	1	4	71	1	1
ZNF263	1426	241	15	12	16	0	5
ZNF740	1045	311	0	0	36	0	0

Table X. Transcription factor classification

TF	MAJOR CLASS	TF FAMILY	TF DOMAIN
ATF3	TFSS	bZIP	
BCLAF1	TFSS	bZIP	
BHLHE40	TFSS	HLH	
CBX5	chromatin		
CEBPB	TFSS	bZIP	
CEBPZ	TFSS	bZIP	
CHD1	chromatin	Homeodomain	
CHD2	chromatin	Homeodomain	
CTCF	TFSS	ZNF	ZNF-C2H2
E2F4	TFSS	wHTH	TDP | wHTH
EGR1	TFSS	ZNF	ZNF-C2H2
ELF1	TFSS	ETS	ETS | wHTH
ELK1	TFSS	ETS	ETS | wHTH
EP300	general		
ETS1	TFSS	ETS	ETS | wHTH
ETV6	TFSS	ETS	ETS | wHTH
EZH2	chromatin		
FOS	TFSS	bZIP	
GABPA	TFSS	ETS	ETS | wHTH
HDGF	TFSS		
IKZF1	TFSS	ZF-C2H2	
JUNB	TFSS	bZIP	
JUND	TFSS	bZIP	
MAFK	TFSS	bZIP	
MAX	TFSS	HLH	
MAZ	TFSS	HLH	
MEF2A	TFSS	MADs-box	
MLLT1	TFSS		
MTA2	TFSS	ZF-GATA	
MXI1	TFSS	HLH	
MYC	TFSS	HLH	
NBN	TFSS		
NFE2	TFSS	bZIP	
NFYA	TFSS	CBF-NFY	
NFYB	TFSS	CBF-NFY	
NR2C2	TFSS	NR	
NRF1	TFSS	bZIP	
PML	cofactor		
POLR2A	general		
POLR2AphosphoS2	general		
POLR2AphosphoS5	general		
POLR3G	general		
RAD21	chromatin		
RCOR1	TFSS	MYB	
REST	TFSS	ZNF	ZNF-C2H2
RFX5	TFSS	wHTH	RFX | wHTH
SIN3A	general		
SIX5	TFSS	Homeodomain	
SMAD5	TFSS	MH1	
SMC3	chromatin		
SP1	TFSS	ZNF	ZNF-C2H2
SPI1	TFSS	ETS	ETS | wHTH
SRF	TFSS	MADs-box	
STAT5A	TFSS	STAT	p53 | STAT
SUZ12	chromatin	ZNF	ZNF-C2H2
TAF1	general		
TARDBP	TFSS		
TBL1XR1	cofactor		
TBP	general		
UBTF	TFSS	HMG	
USF1	TFSS	HLH	
USF2	TFSS	HLH	
YBX1	TFSS	CSD	
YY1	TFSS	ZNF	ZNF-C2H2
ZBED1	TFSS	ZNF	ZNF-C2H2
ZBTB33	TFSS	ZNF	ZNF-C2H2
ZBTB40	TFSS	ZNF	ZNF-C2H2
ZNF143	TFSS	ZNF	ZNF-C2H2
ZNF274	TFSS	ZNF	ZNF-C2H2

Fig X. ENCODE experiment data matrix by biosample



coverage =
target / # available unique target

Repli-seq
WBGS
RAMPAGE
(needs review)

Fig X. TF ChIP-seq experiments by cell line

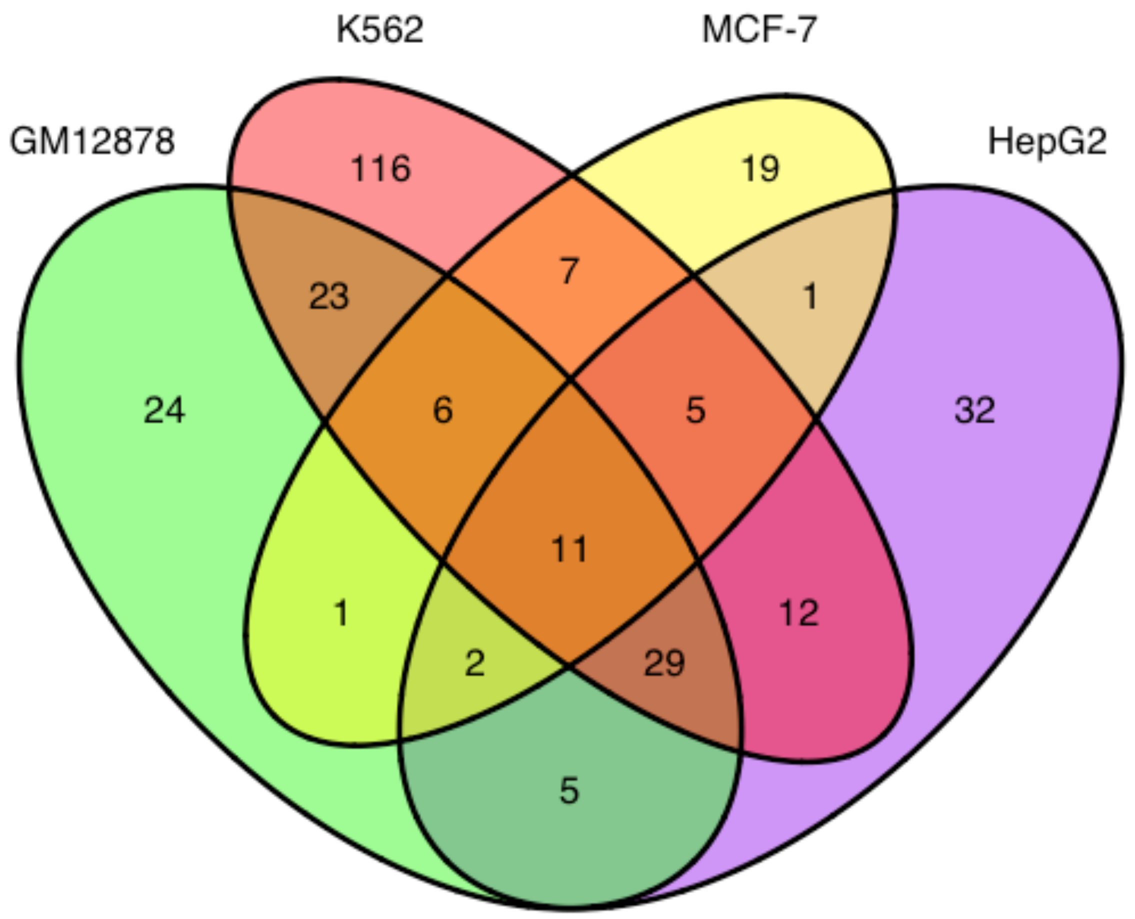
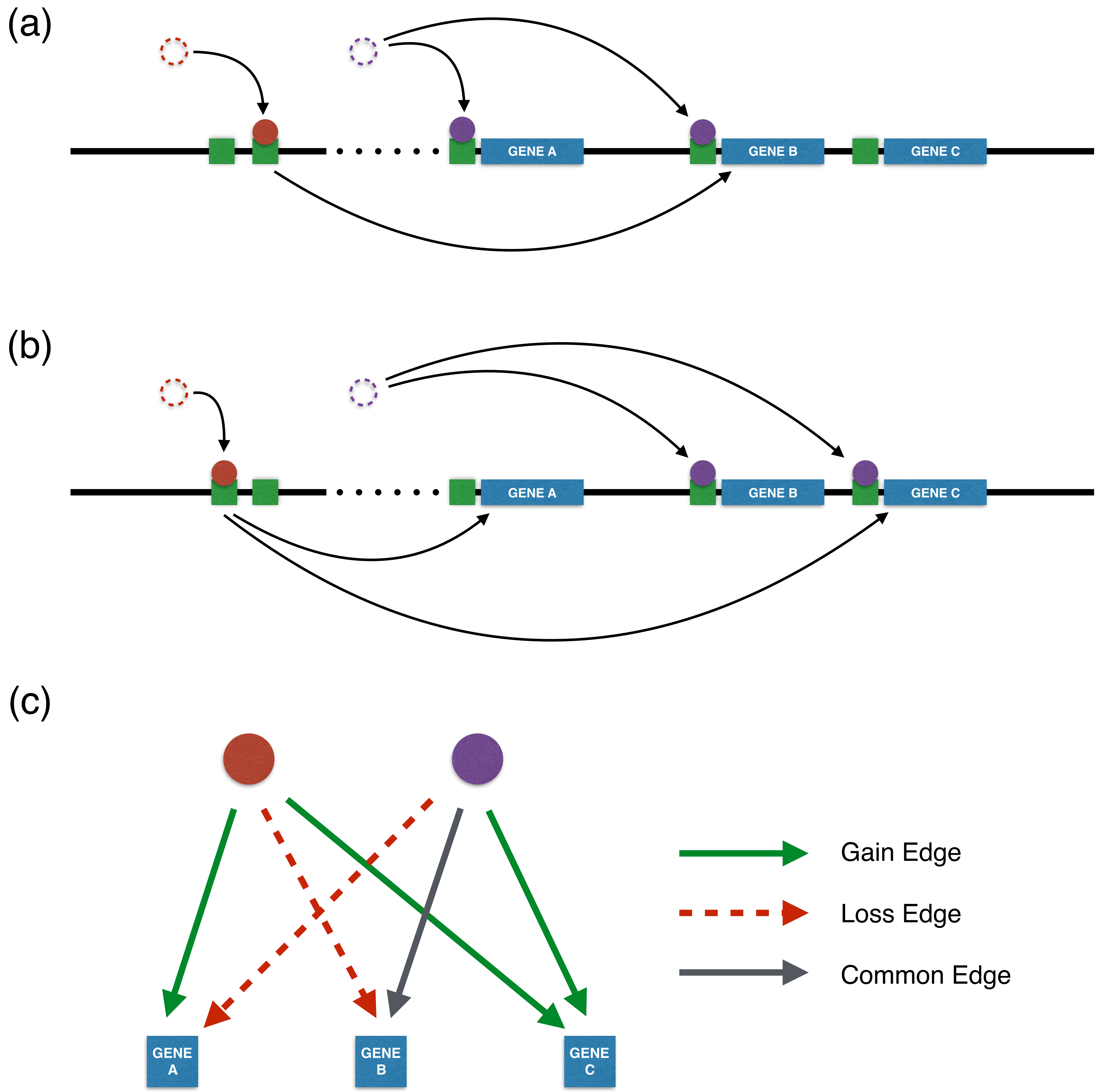


Table X. Proximal and distal regulatory elements of TFSS

TF	Peaks	Proximal	Distal	% Proximal
ATF3	16,011	4,604	11,407	28.8%
BCLAF1	4,444	1,934	2,510	43.5%
BHLHE40	22,497	6,726	15,771	29.9%
CEBPB	38,715	6,088	32,627	15.7%
CHD1	9,350	5,327	4,023	57.0%
CHD2	7,797	3,634	4,163	46.6%
CTCF	54,387	9,483	44,904	17.4%
E2F4	8,181	4,558	3,623	55.7%
EGR1	36,997	12,852	24,145	34.7%
ELK1	2,961	1,662	1,299	56.1%
ETS1	10,726	5,471	5,255	51.0%
EZH2	1,685	347	1,338	20.6%
FOS	7,646	2,330	5,316	30.5%
GABPA	14,393	5,802	8,591	40.3%
JUND	40,052	8,668	31,384	21.6%
MAFK	26,965	3,589	23,376	13.3%
MAX	31,436	11,211	20,225	35.7%
MAZ	33,323	11,915	21,408	35.8%
MEF2A	5,631	1,186	4,445	21.1%
MXI1	6,711	3,381	3,330	50.4%
MYC	24,153	9,000	15,153	37.3%
NFYA	4,286	1,866	2,420	43.5%
NFYB	10,096	2,806	7,290	27.8%
NR2C2	587	309	278	52.6%
PML	15,895	5,975	9,920	37.6%
RAD21	34,725	6,045	28,680	17.4%
RCOR1	35,741	7,829	27,912	21.9%
RFX5	2,201	807	1,394	36.7%
SIX5	4,194	2,251	1,943	53.7%
SMC3	23,598	4,068	19,530	17.2%
SPI1	28,677	5,748	22,929	20.0%
SRF	4,717	1,776	2,941	37.7%
STAT5A	9,811	1,743	8,068	17.8%
TBL1XR1	8,505	2,092	6,413	24.6%
UBTF	6,002	2,972	3,030	49.5%
USF1	18,521	4,501	14,020	24.3%
USF2	3,083	1,044	2,039	33.9%
YY1	12,677	6,094	6,583	48.1%
ZNF143	29,069	6,842	22,227	23.5%
ZNF274	1,997	427	1,570	21.4%
CBX5	4,868	884	3,984	18.2%
CEBPZ	1,012	492	520	48.6%
ELF1	13,928	5,638	8,290	40.5%
ETV6	2,625	840	1,785	32.0%
HDGF	6,405	2,307	4,098	36.0%
IKZF1	49,278	12,142	37,136	24.6%
JUNB	3,933	850	3,083	21.6%
MLLT1	9,234	3,607	5,627	39.1%
MTA2	13,804	2,356	11,448	17.1%
NBN	13,928	4,406	9,522	31.6%
NFE2	26,075	3,770	22,305	14.5%
NRF1	28,662	11,371	17,291	39.7%
REST	43,207	8,713	34,494	20.2%
SMAD5	17,763	9,366	8,397	52.7%
SP1	12,101	6,506	5,595	53.8%
SUZ12	2,360	1,103	1,257	46.7%
TARDBP	8,702	3,706	4,996	42.6%
YBX1	773	129	644	16.7%
ZBED1	3,652	1,568	2,084	42.9%
ZBTB33	48,989	9,569	39,420	19.5%
ZBTB40	22,988	7,398	15,590	32.2%

Fig X. Schematic of TF-Gene network rewiring



(a)
GENE A
GENE B
GENE C
(b)
GENE A
GENE B
GENE C
(c)
GENE A
GENE B
GENE C
Gain Edge
Loss Edge
Common Edge