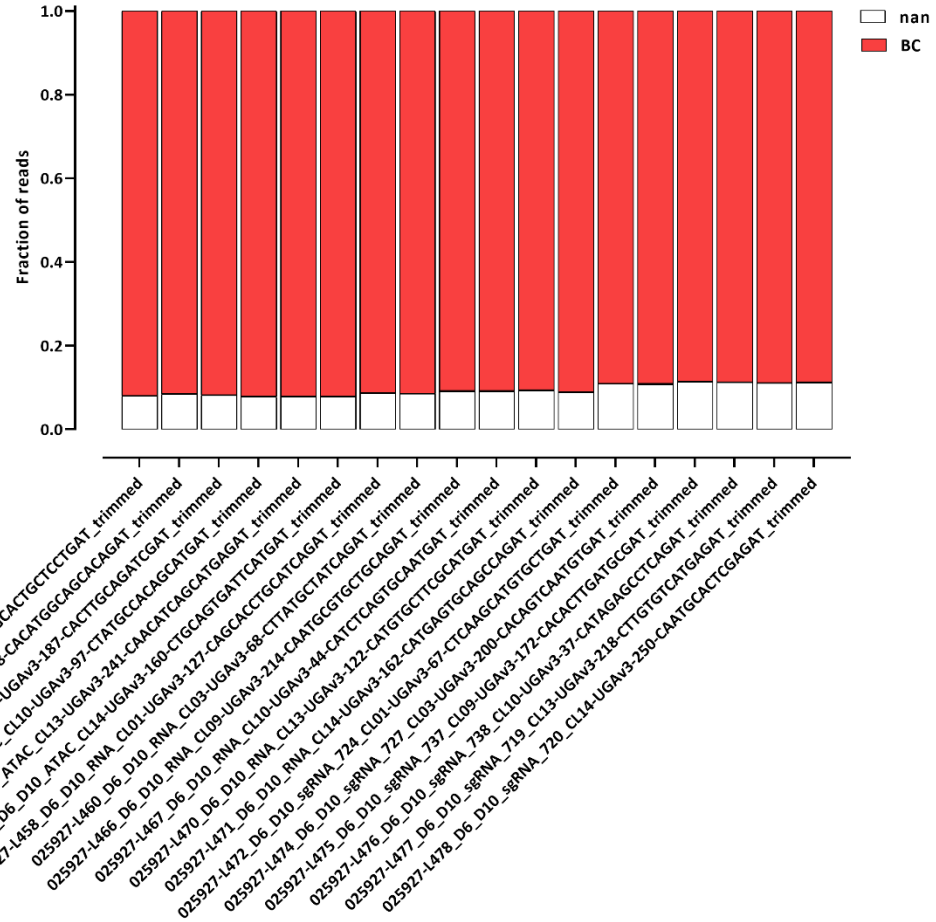
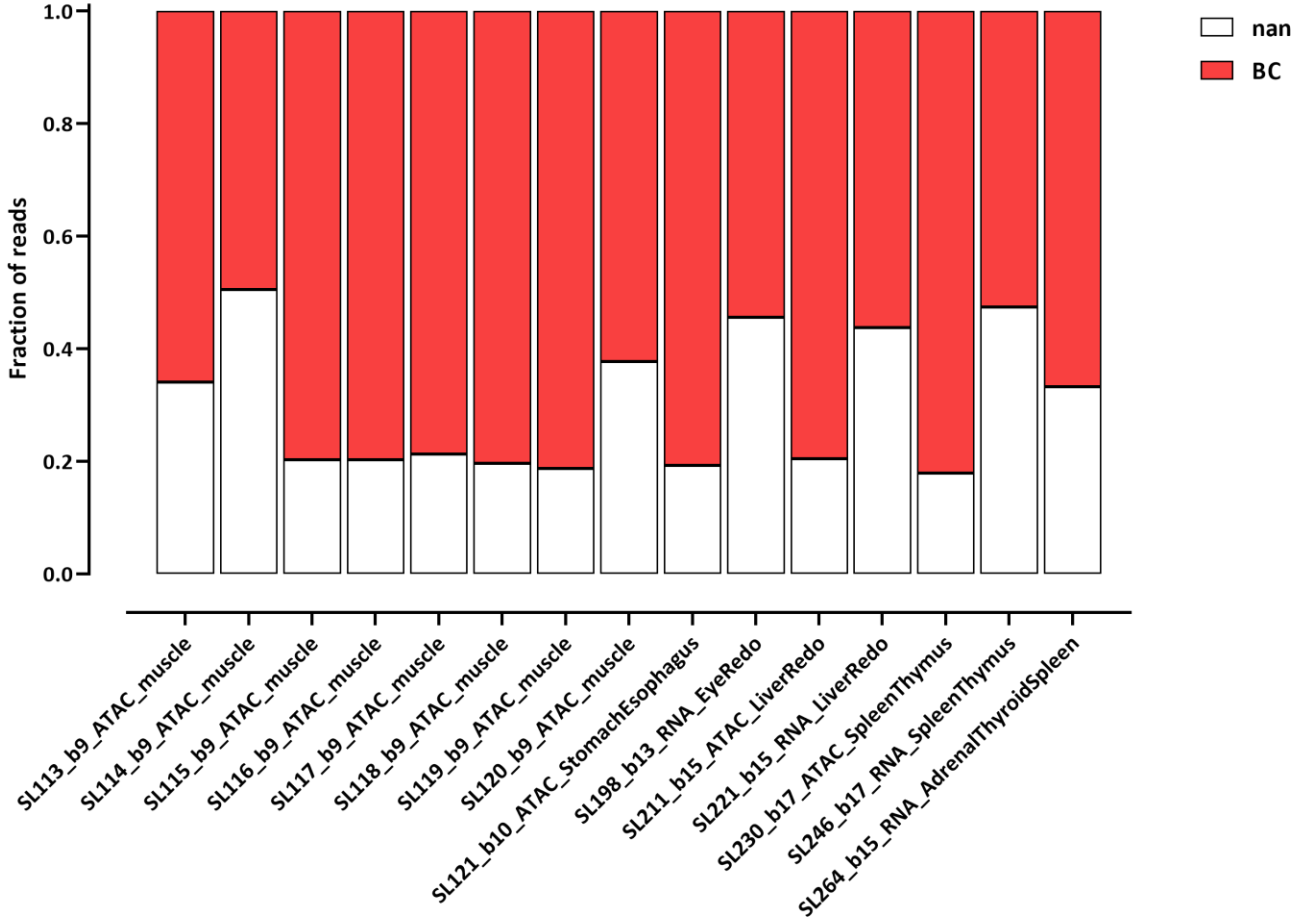
ULTIMA PERTURB-SHARE FIRST PRODUCTION RUN

Barcode assignment



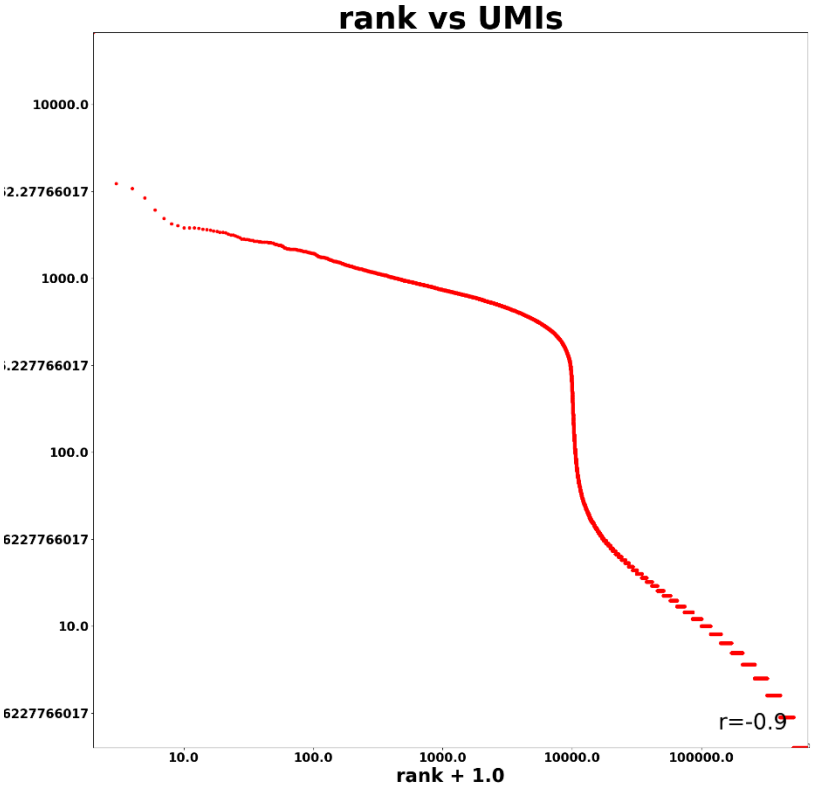
Typical Illumina for comparison:

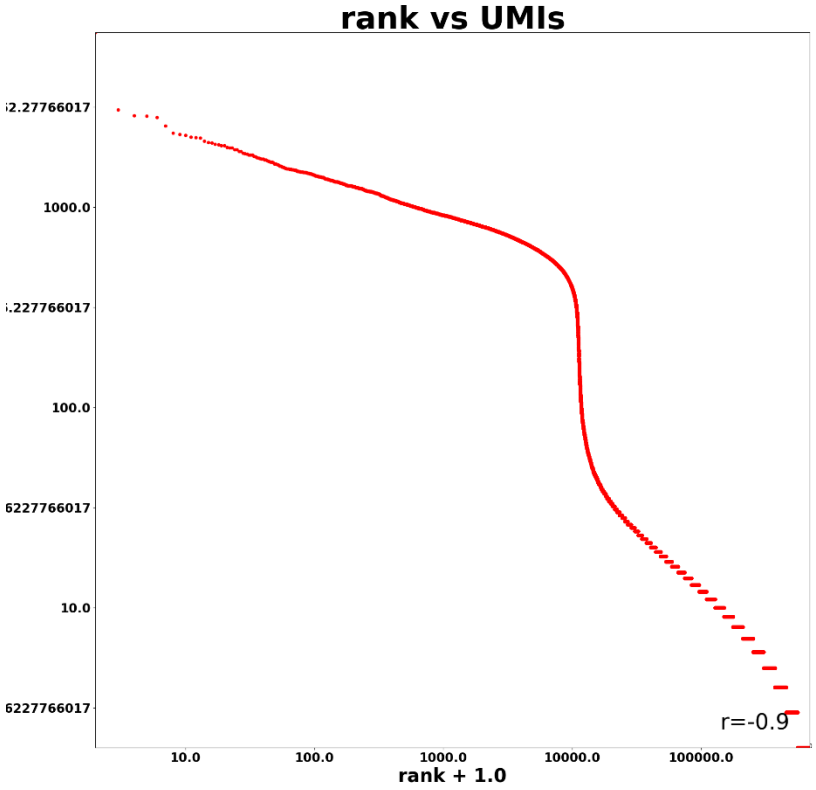


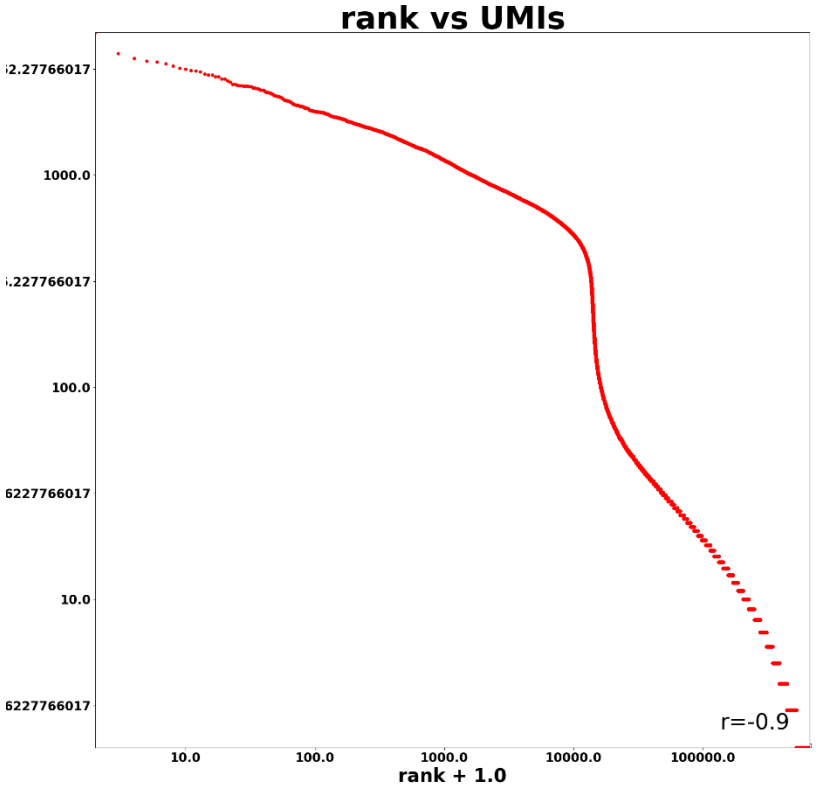
Mapping stats RNA

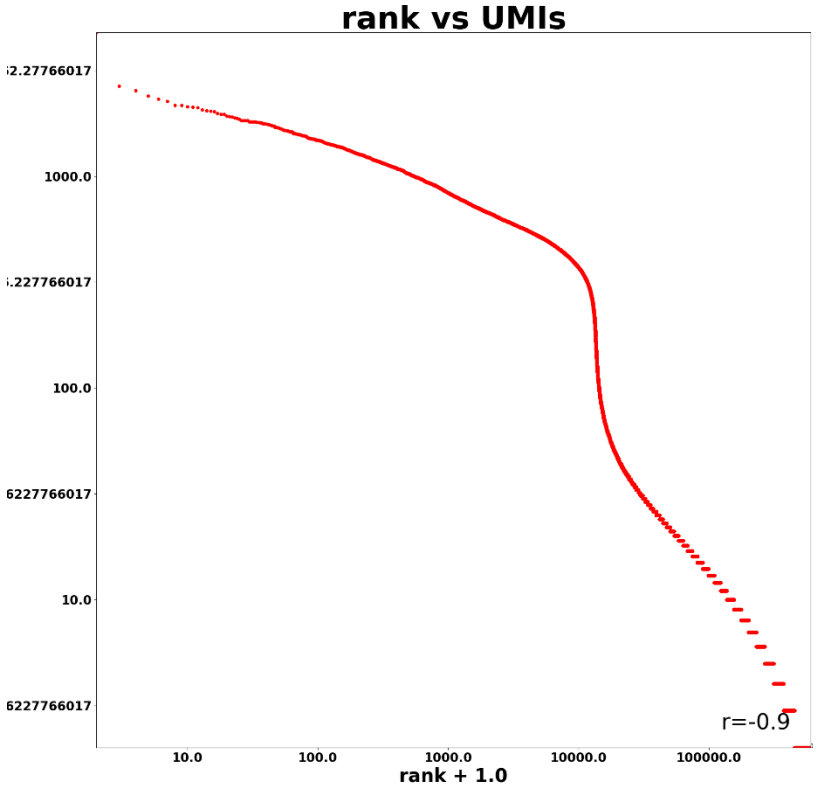
Library	Read Length	Mapping	Species	Raw fragments	Complexity	Unique	Unique Splices	Multi	Multi Splices	Fraction mapped	Cells >500 UMIs
025927-L458_D6_D10_RNA_CL01-UGAv3-127-CAGCACCTGCATCAGAT_trimmed	UG	STAR-2.5.3a	hg38	274,422,379	0.86	75,005,250	11,751,724	11,382,064	8,301,374	0.39	6,856
025927-L460_D6_D10_RNA_CL03-UGAv3-68-CTTATGCTATCAGAT_trimmed	UG	STAR-2.5.3a	hg38	329,406,009	0.86	88,443,553	14,174,062	13,018,514	9,408,943	0.38	7,996
025927-L466_D6_D10_RNA_CL09-UGAv3-214-CAATGCGTGCTGCAGAT_trimmed	UG	STAR-2.5.3a	hg38	430,357,449	0.85	117,852,488	18,217,556	16,751,508	12,205,983	0.38	10,605
025927-L467_D6_D10_RNA_CL10-UGAv3-44-CATCTCAGTGCAATGAT_trimmed	UG	STAR-2.5.3a	hg38	276,404,811	0.86	79,013,809	11,757,490	11,512,173	7,643,773	0.40	5,578
025927-L470_D6_D10_RNA_CL13-UGAv3-122-CATGTGCTTCGCATGAT_trimmed	UG	STAR-2.5.3a	hg38	261,390,428	0.87	75,179,869	11,110,998	11,180,641	7,706,249	0.40	4,884
025927-L471_D6_D10_RNA_CL14-UGAv3-162-CATGAGTGCAGCCAGAT_trimmed	UG	STAR-2.5.3a	hg38	372,899,440	0.85	99,175,479	15,296,687	14,135,832	10,044,644	0.37	8,864

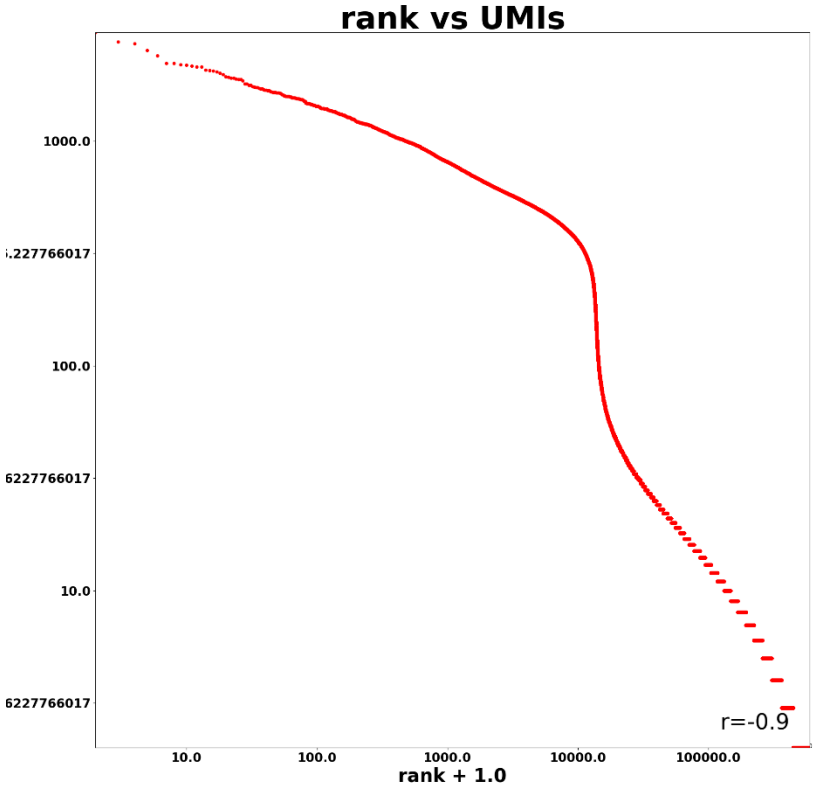
#	Exonic:	Intergenic:	Intronic:
025927-L458_D6_D10_RNA_CL01-UGAv3-127-CAGCACCTGCATCAGAT_trimmed	0.38	0.10	0.52
025927-L460_D6_D10_RNA_CL03-UGAv3-68-CTTATGCTATCAGAT_trimmed	0.38	0.09	0.52
025927-L466_D6_D10_RNA_CL09-UGAv3-214-CAATGCGTGCTGCAGAT_trimmed	0.37	0.10	0.54
025927-L467_D6_D10_RNA_CL10-UGAv3-44-CATCTCAGTGCAATGAT_trimmed	0.37	0.10	0.53
025927-L470_D6_D10_RNA_CL13-UGAv3-122-CATGTGCTTCGCATGAT_trimmed	0.37	0.10	0.54
025927-L471_D6_D10_RNA_CL14-UGAv3-162-CATGAGTGCAGCCAGAT_trimmed	0.37	0.10	0.53

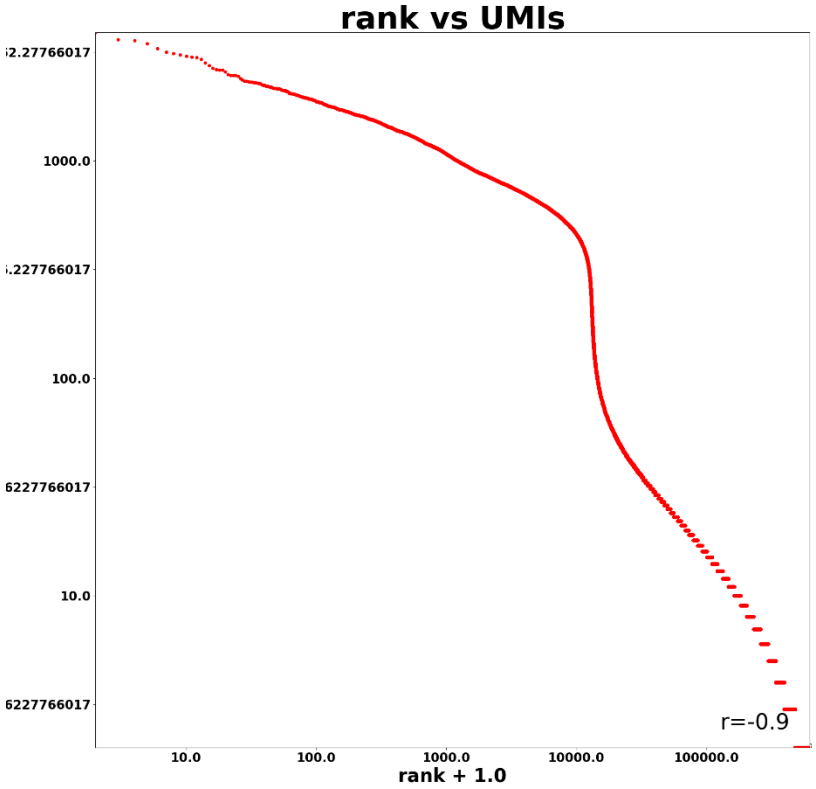

rank vs UMIs
10000.0
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0


rank vs UMIs
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0


rank vs UMIs
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0


rank vs UMIs
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

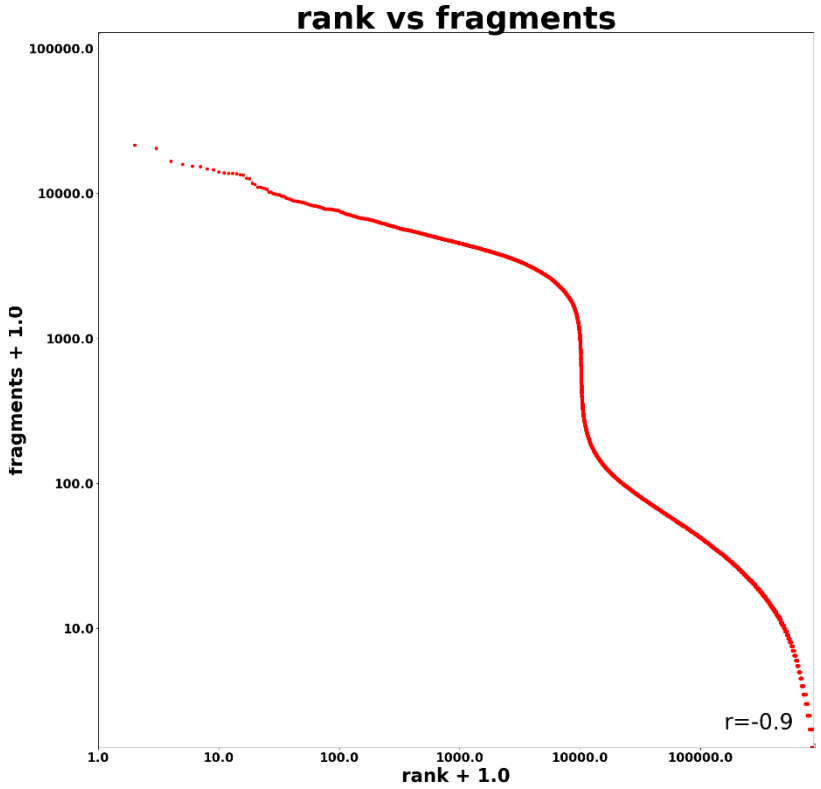

rank vs UMIs
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

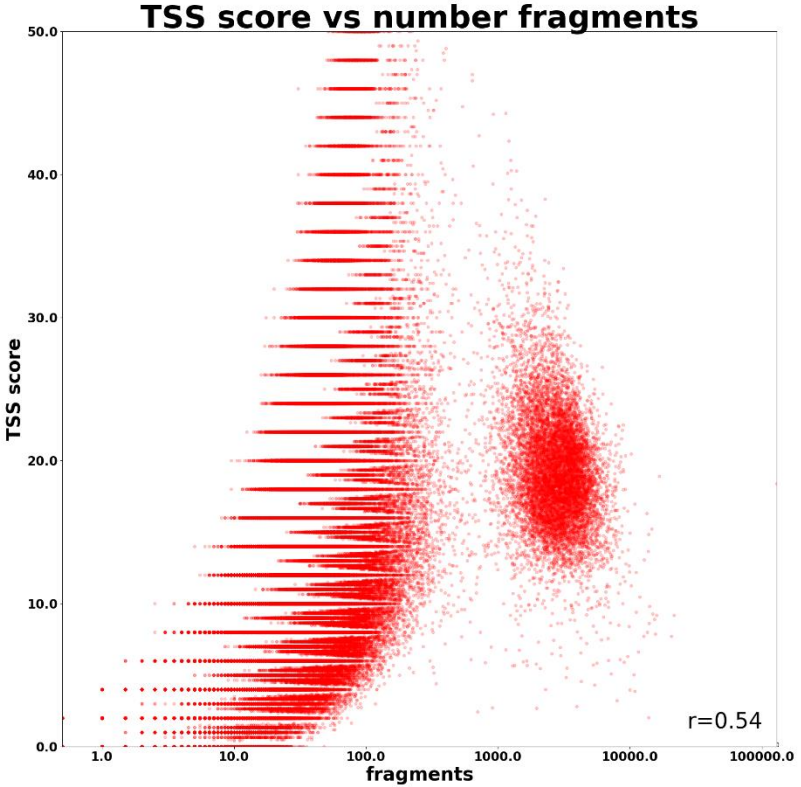

rank vs UMIs
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

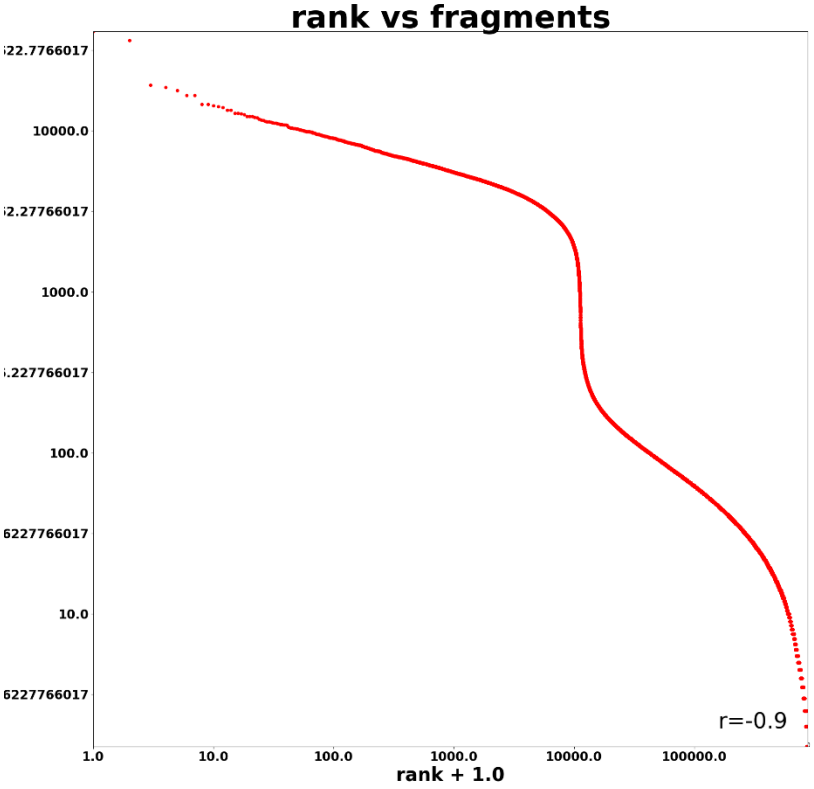
Mapping stats ATAC

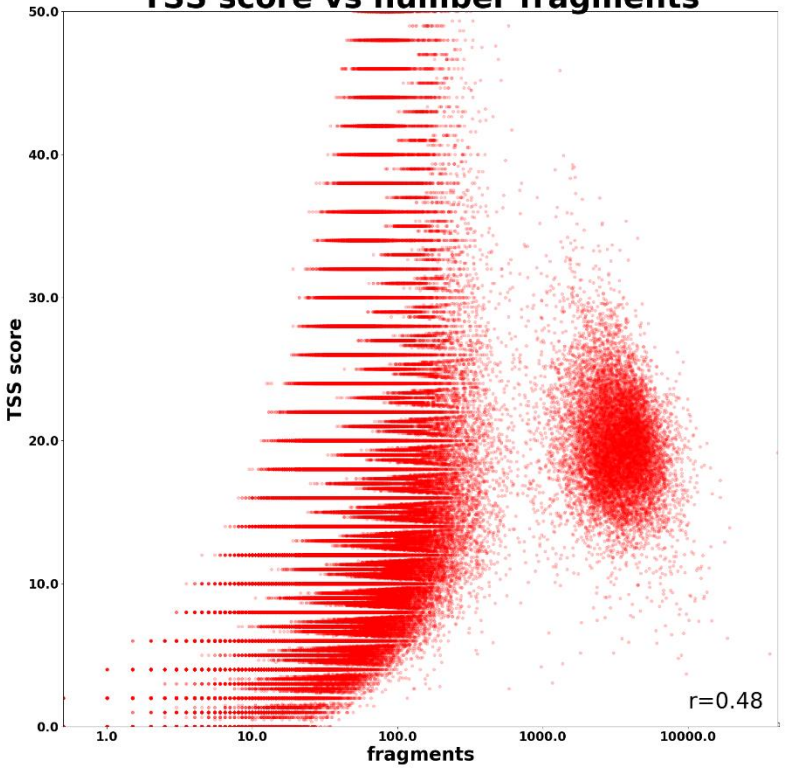
Species	Read Length	Library	Raw fragments	Unique non-chrM reads	Alignment fraction	Complexity	chrM reads	chrM fraction	scATAC dedup reads	scATAC dedup Complexity	TSS ratio	Cells >500 fragments
hg38	UG, BT	025927-L418_D6_D10_ATAC_CL01-UGAv3-174-CGCACTGCTCCTGAT_trimmed	156,234,486	55,912,131	0.41	0.82	8,241,312	0.13	48,152,191	0.91	17.98	9,998
hg38	UG, BT	025927-L420_D6_D10_ATAC_CL03-UGAv3-178-CACATGGCAGCACAGAT_trimmed	216,357,427	76,625,741	0.41	0.77	11,848,238	0.13	63,981,968	0.87	18.50	11,174
hg38	UG, BT	025927-L430_D6_D10_ATAC_CL09-UGAv3-187-CACTTGCAGATCGAT_trimmed	260,714,998	93,393,799	0.41	0.78	14,273,670	0.13	81,195,478	0.86	18.59	14,052
hg38	UG, BT	025927-L431_D6_D10_ATAC_CL10-UGAv3-97-CTATGCCACAGCATGAT_trimmed	254,435,715	92,487,050	0.42	0.77	14,335,734	0.13	79,184,568	0.85	19.00	13,554
hg38	UG, BT	025927-L434_D6_D10_ATAC_CL13-UGAv3-241-CAACATCAGCATGAGAT_trimmed	259,873,812	94,589,137	0.42	0.77	14,398,000	0.13	81,261,057	0.85	18.88	13,755
hg38	UG, BT	025927-L435_D6_D10_ATAC_CL14-UGAv3-160-CTGCAGTGATTCATGAT_trimmed	236,457,489	82,420,812	0.40	0.78	12,245,198	0.13	70,613,312	0.87	18.83	13,009
hg38	UG, BWA	025927-L418_D6_D10_ATAC_CL01-UGAv3-174-CGCACTGCTCCTGAT_trimmed	156,234,486	112,462,210	0.72	0.91			92,153,895	0.95	17.30	10,243
hg38	UG, BWA	025927-L420_D6_D10_ATAC_CL03-UGAv3-178-CACATGGCAGCACAGAT_trimmed	216,357,427	161,548,474	0.75	0.88			126,597,646	0.94	18.35	11,462
hg38	UG, BWA	025927-L430_D6_D10_ATAC_CL09-UGAv3-187-CACTTGCAGATCGAT_trimmed	260,714,998	194,476,327	0.75	0.88			160,195,606	0.93	18.44	15,496
hg38	UG, BWA	025927-L431_D6_D10_ATAC_CL10-UGAv3-97-CTATGCCACAGCATGAT_trimmed	254,435,715	189,648,508	0.75	0.87			153,775,132	0.92	18.73	14,829
hg38	UG, BWA	025927-L434_D6_D10_ATAC_CL13-UGAv3-241-CAACATCAGCATGAGAT_trimmed	259,873,812	194,494,602	0.75	0.88			158,224,902	0.92	18.72	15,092
hg38	UG, BWA	025927-L435_D6_D10_ATAC_CL14-UGAv3-160-CTGCAGTGATTCATGAT_trimmed	236,457,489	177,884,292	0.75	0.89			143,746,881	0.93	18.51	14,004

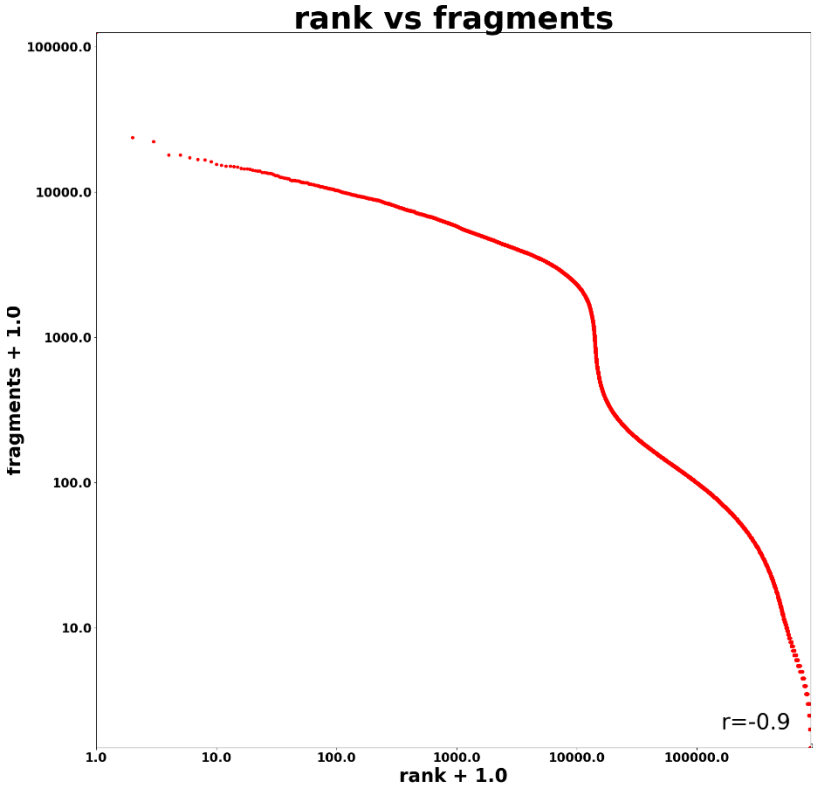
Note: cutoff should be higher than 500 in this case, but doing it on the RNA side will do the job fine

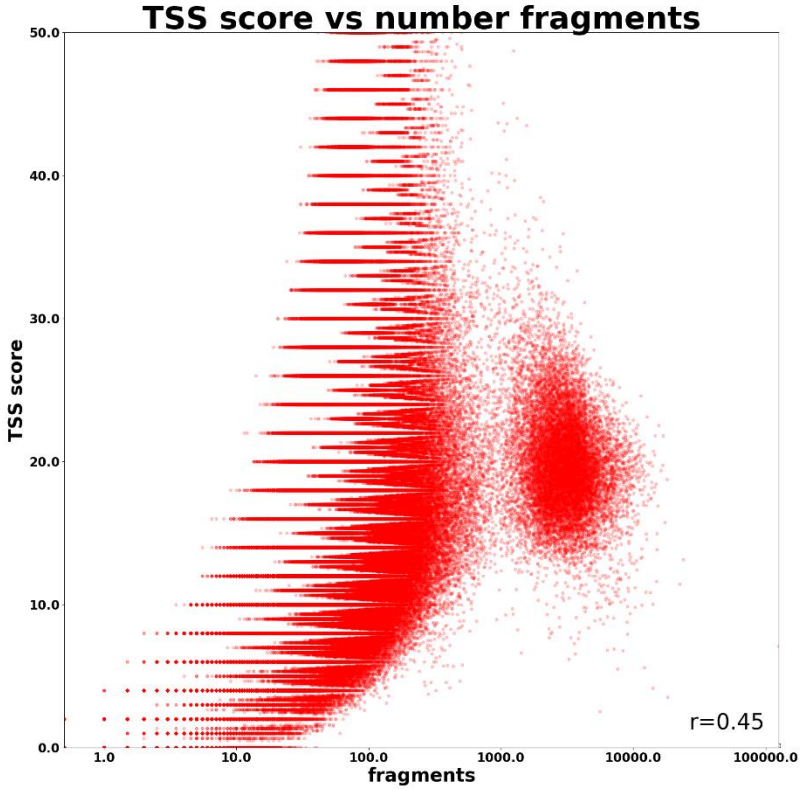

rank vs fragments
fragments + 1.0
rank + 1.0
100000.0
10000.0
1000.0
100.0
10.0
1.0
10.0
100.0
1000.0
10000.0
100000.0
r=-0.9

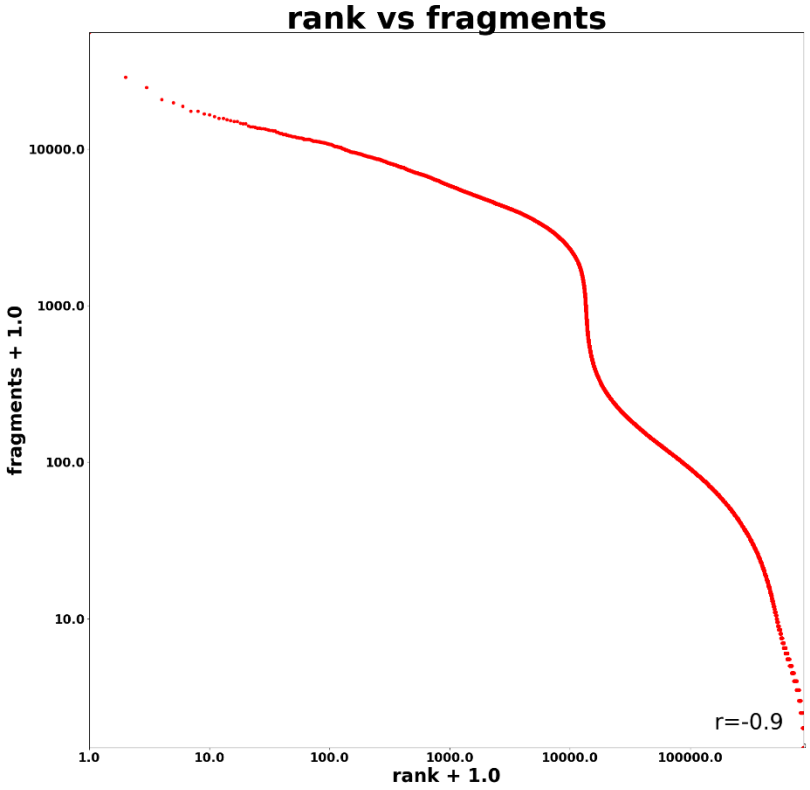

TSS score vs number fragments
TSS score
fragments
50.0
40.0
30.0
20.0
10.0
0.0
1.0
10.0
100.0
1000.0
10000.0
100000.0
r=0.54

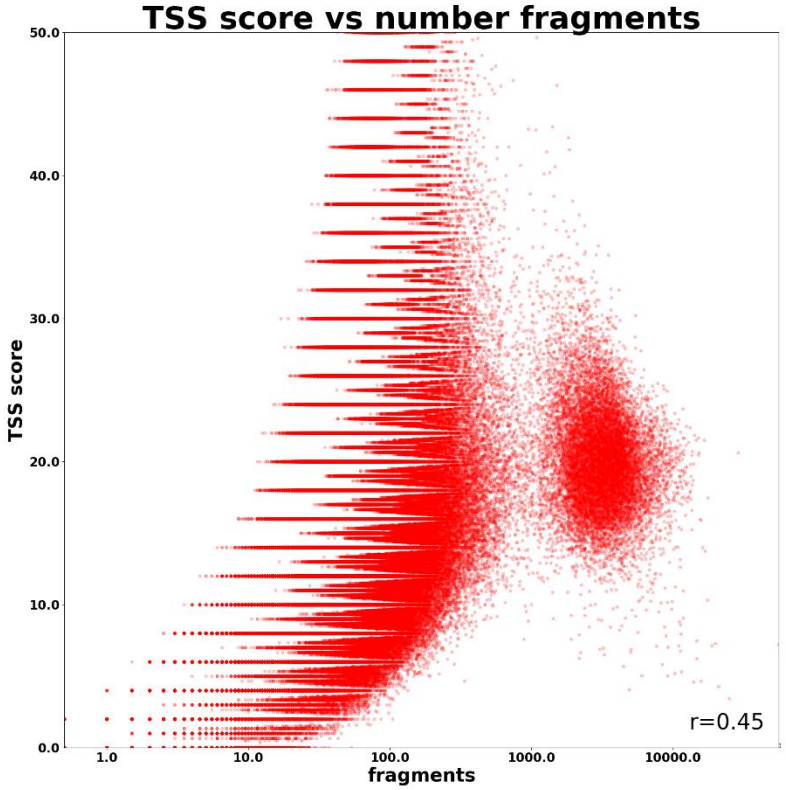

rank vs fragments
22.7766017
10000.0
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
1.0
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

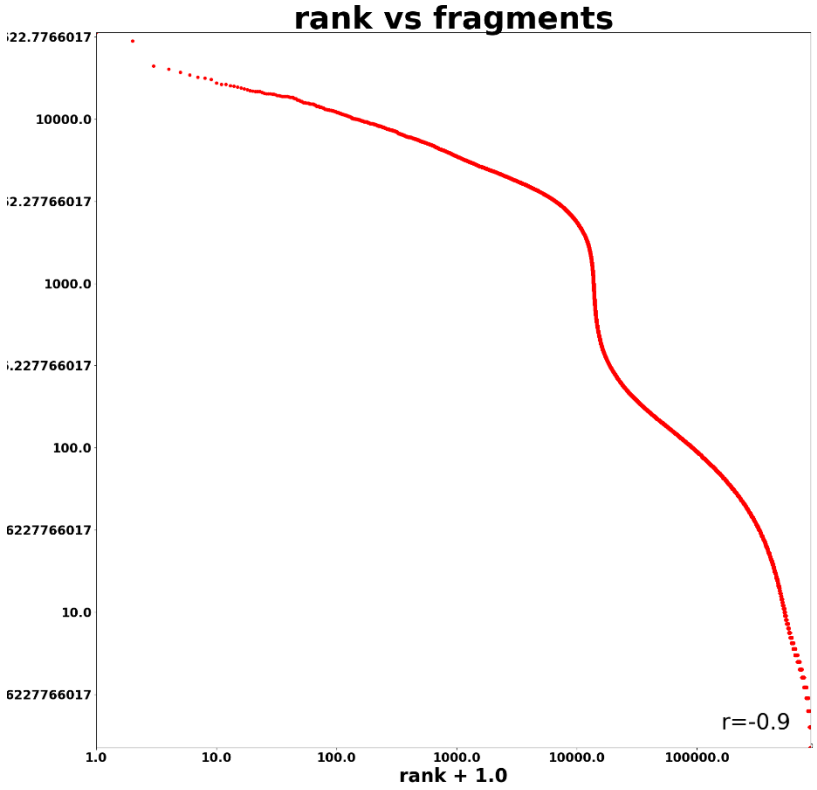

TSS score vs number fragments
50.0
40.0
30.0
20.0
10.0
0.0
TSS score
r=0.48
1.0
10.0
100.0
1000.0
10000.0
fragments

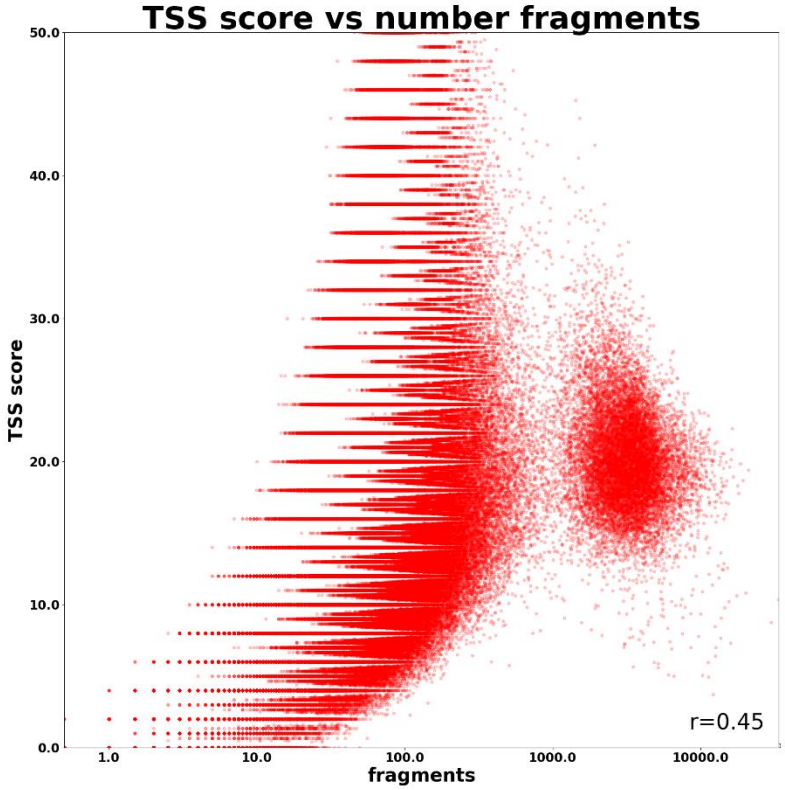

rank vs fragments
100000.0
10000.0
1000.0
100.0
10.0
fragments + 1.0
r=-0.9
1.0
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

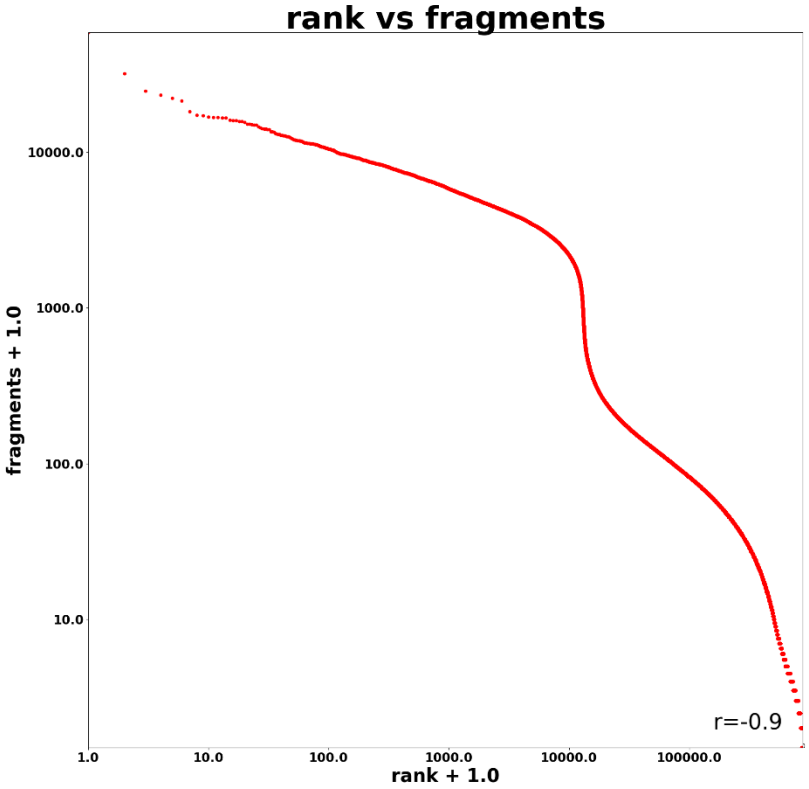

TSS score vs number fragments
50.0
40.0
30.0
20.0
10.0
0.0
TSS score
r=0.45
1.0
10.0
100.0
1000.0
10000.0
100000.0
fragments

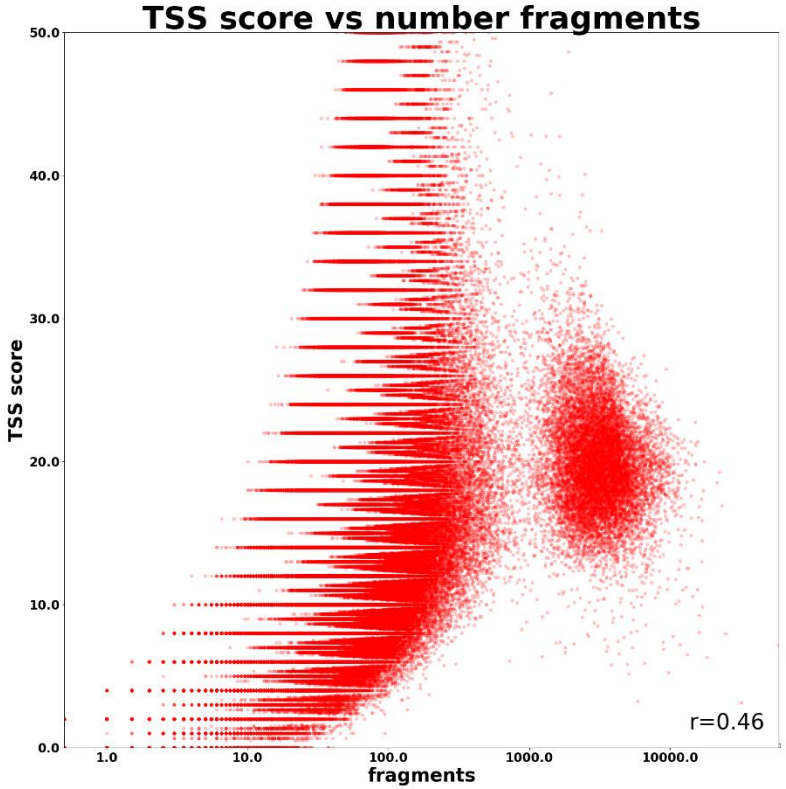

rank vs fragments
fragments + 1.0
10000.0
1000.0
100.0
10.0
1.0
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0
r=-0.9


TSS score vs number fragments
TSS score
50.0
40.0
30.0
20.0
10.0
0.0
1.0
10.0
100.0
1000.0
10000.0
fragments
r=0.45


rank vs fragments
22.7766017
10000.0
2.27766017
1000.0
.227766017
100.0
6227766017
10.0
6227766017
r=-0.9
1.0
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0

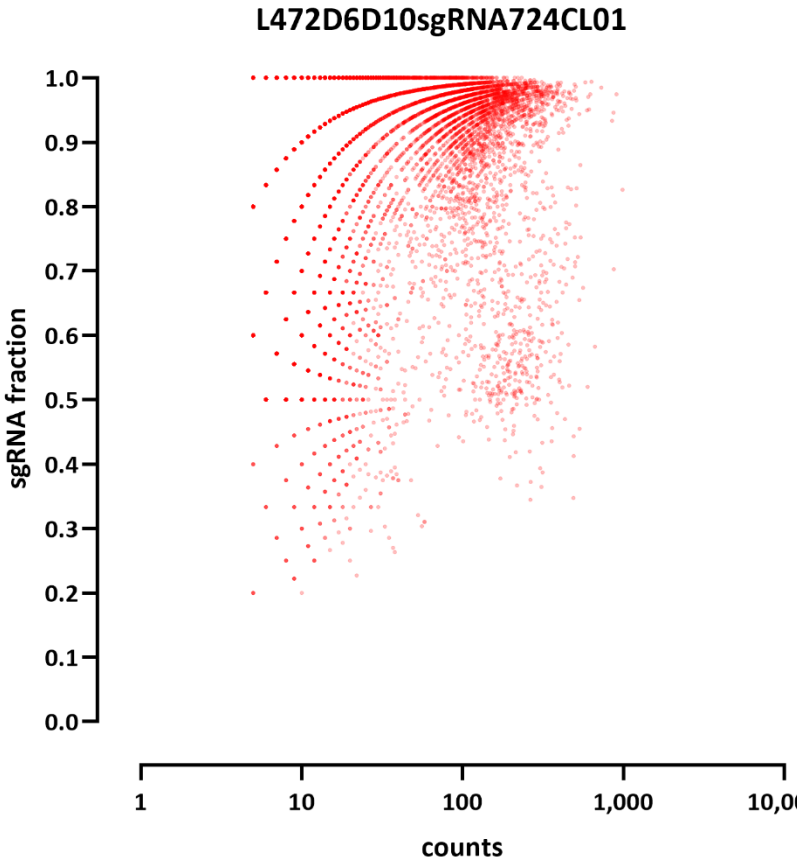

TSS score vs number fragments
50.0
40.0
30.0
20.0
10.0
0.0
TSS score
r=0.45
1.0
10.0
100.0
1000.0
10000.0
fragments

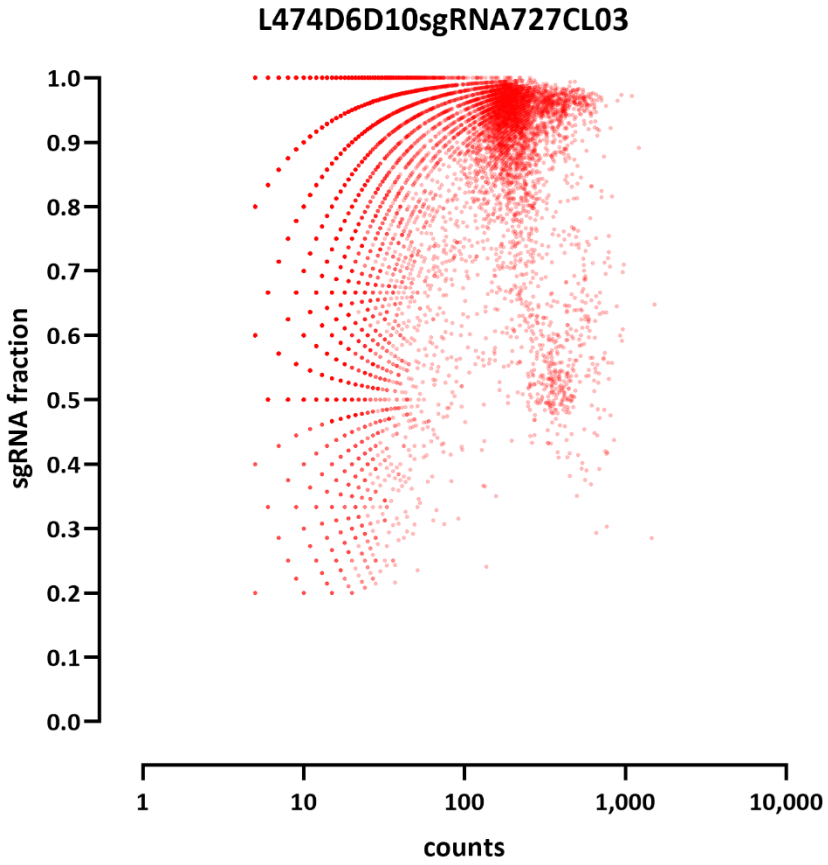

rank vs fragments
fragments + 1.0
10000.0
1000.0
100.0
10.0
1.0
10.0
100.0
1000.0
10000.0
100000.0
rank + 1.0
r=-0.9


TSS score vs number fragments
TSS score
50.0
40.0
30.0
20.0
10.0
0.0
1.0
10.0
100.0
1000.0
10000.0
fragments
r=0.46

sgRNA reads

#library	raw reads	reads assigned to sgRNA	fraction	barcodes >10 reads	barcodes >5 reads
025927-L472_D6_D10_sgRNA_724_CL01-UGAv3-67-CTCAAGCATGTGCTGAT_trimmed	49,244,850	26,476,962	0.54	51,601	56,521
025927-L474_D6_D10_sgRNA_727_CL03-UGAv3-200-CACAGTCAATGTGAT_trimmed	29,309,339	16,607,662	0.57	70,577	74,241
025927-L475_D6_D10_sgRNA_737_CL09-UGAv3-172-CACACTTGATGCGAT_trimmed	25,862,530	13,758,687	0.53	118,619	140,456
025927-L476_D6_D10_sgRNA_738_CL10-UGAv3-37-CATAGAGCCTCAGAT_trimmed	21,395,424	11,637,254	0.54	95,563	109,678
025927-L477_D6_D10_sgRNA_719_CL13-UGAv3-218-CTTGTGTCATGAGAT_trimmed	37,991,352	20,917,201	0.55	123,819	135,451
025927-L478_D6_D10_sgRNA_720_CL14-UGAv3-250-CAATGCACTCGAGAT_trimmed	18,307,722	9,858,059	0.54	72,458	124,585


L472D6D10sgRNA724CL01
sgRNA fraction
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
10
100
1,000
counts


L474D6D10sgRNA727CL03
sgRNA fraction
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
10
100
1,000
10,000
counts

sgRNA assignment intersect with ATAC

Asking for:

- >= 10 sgRNA reads
- >= 75% of reads assigned to a single sgRNA

3,937 cells in CL01
4,074 cells in CL03
4,994 cells in CL09
4,785 cells in CL10
4,778 cells in CL13
4,179 cells in CL14

sgRNAs assigned to at least one cell:

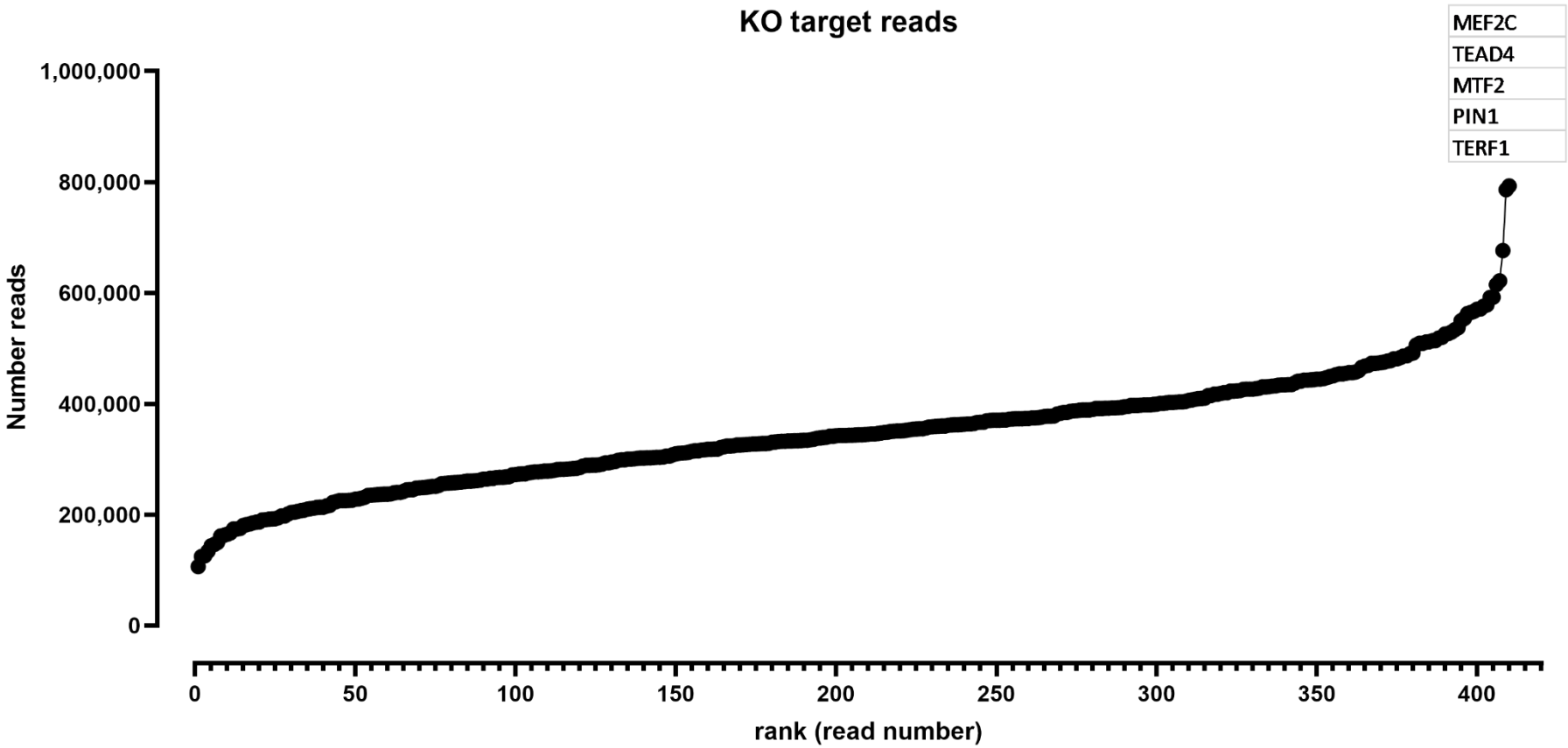
CL01	836
CL03	848
CL09	843
CL10	846
CL13	841
CL14	836

Note:

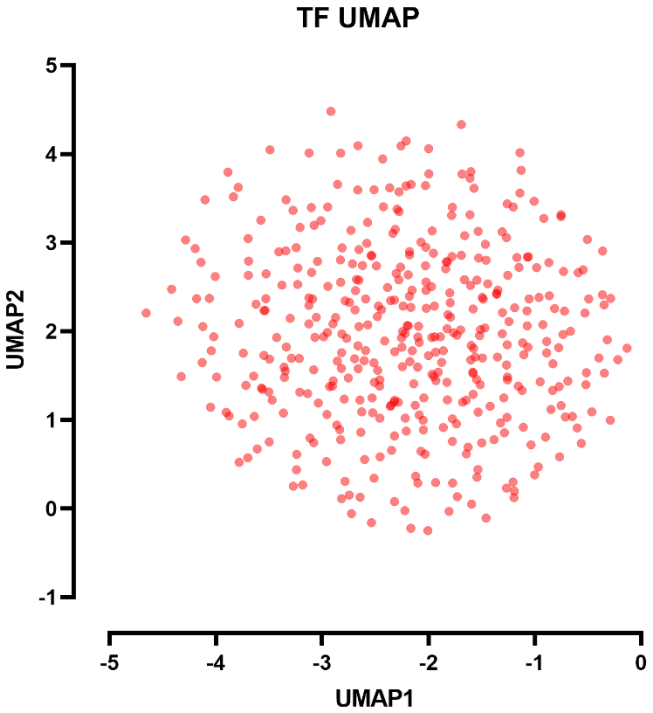
I am ignoring time points for this analysis, but these libraries were a mix of two time points in fact

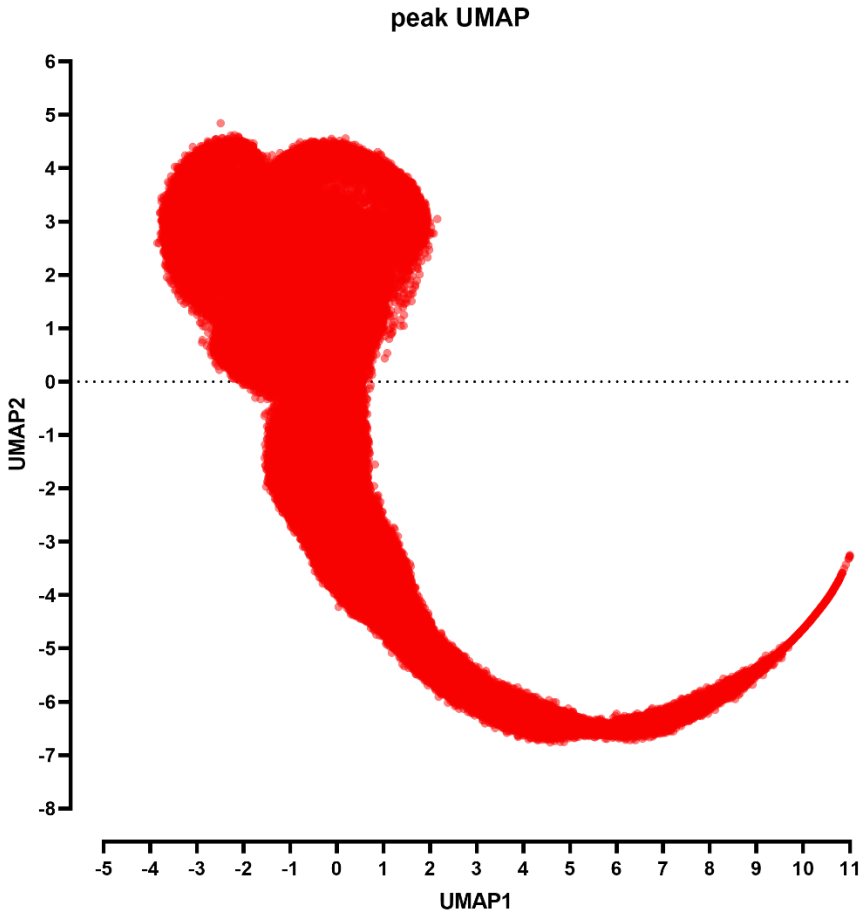
From these we get ~25 cells per sgRNA across both time points

ATAC

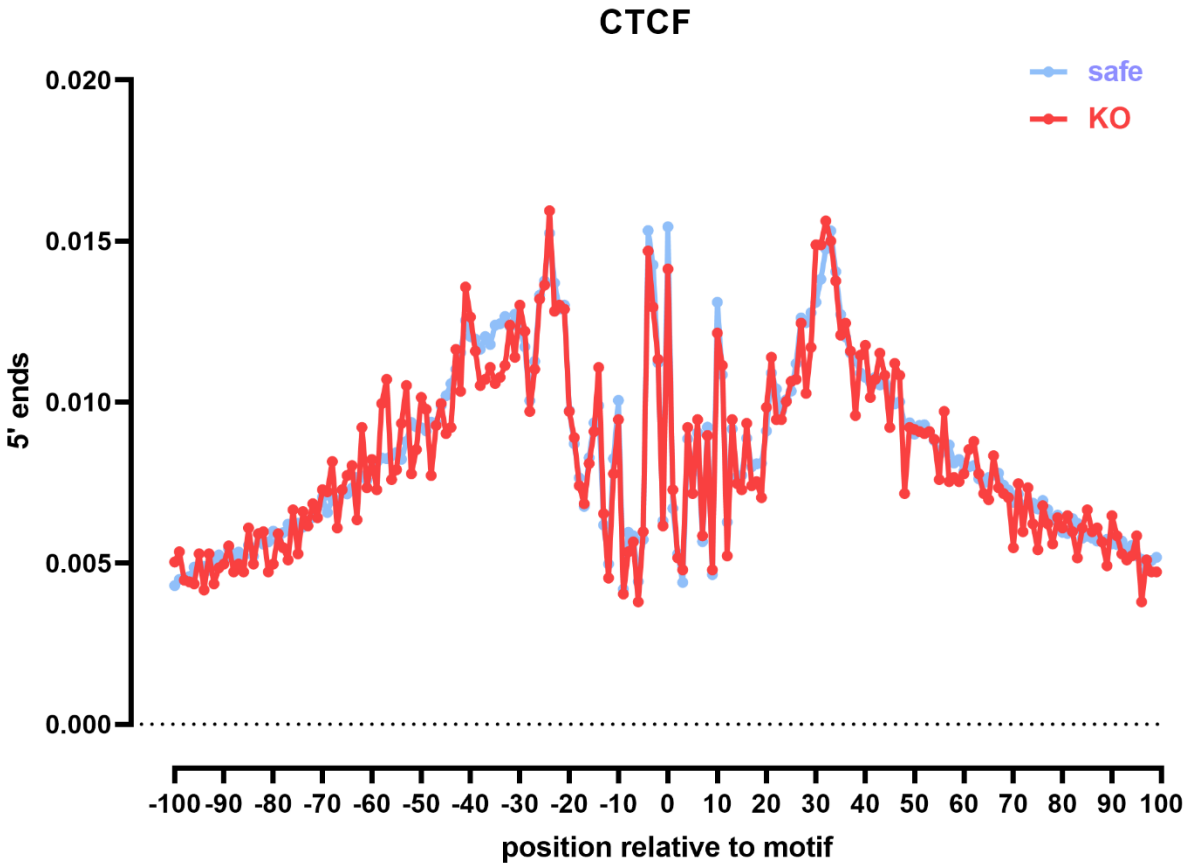

KO target reads
MEF2C
TEAD4
MTF2
PIN1
TERF1
1,000,000
800,000
600,000
400,000
200,000
0
Number reads
0
50
100
150
200
250
300
350
400
rank (read number)

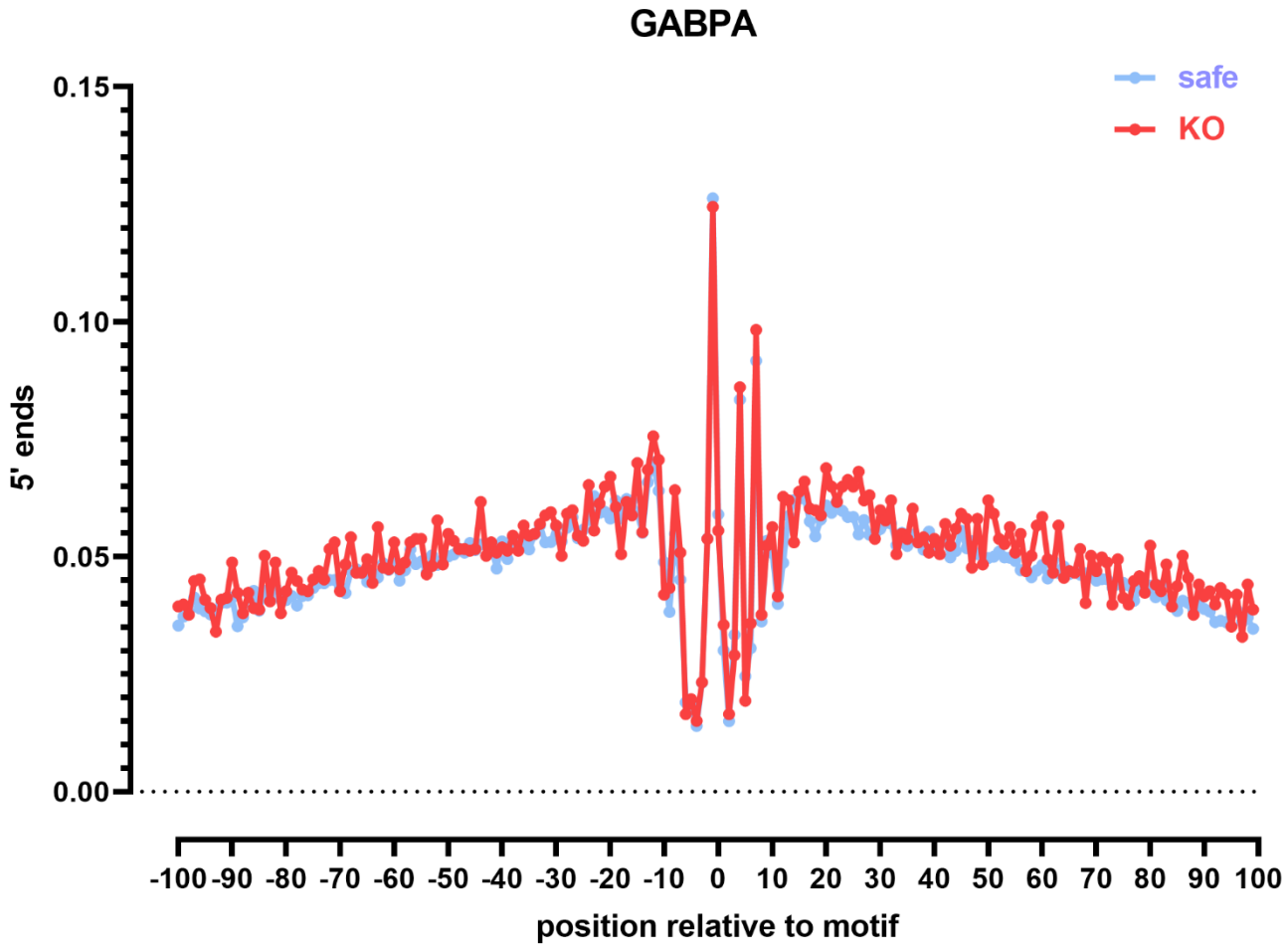
ATAC

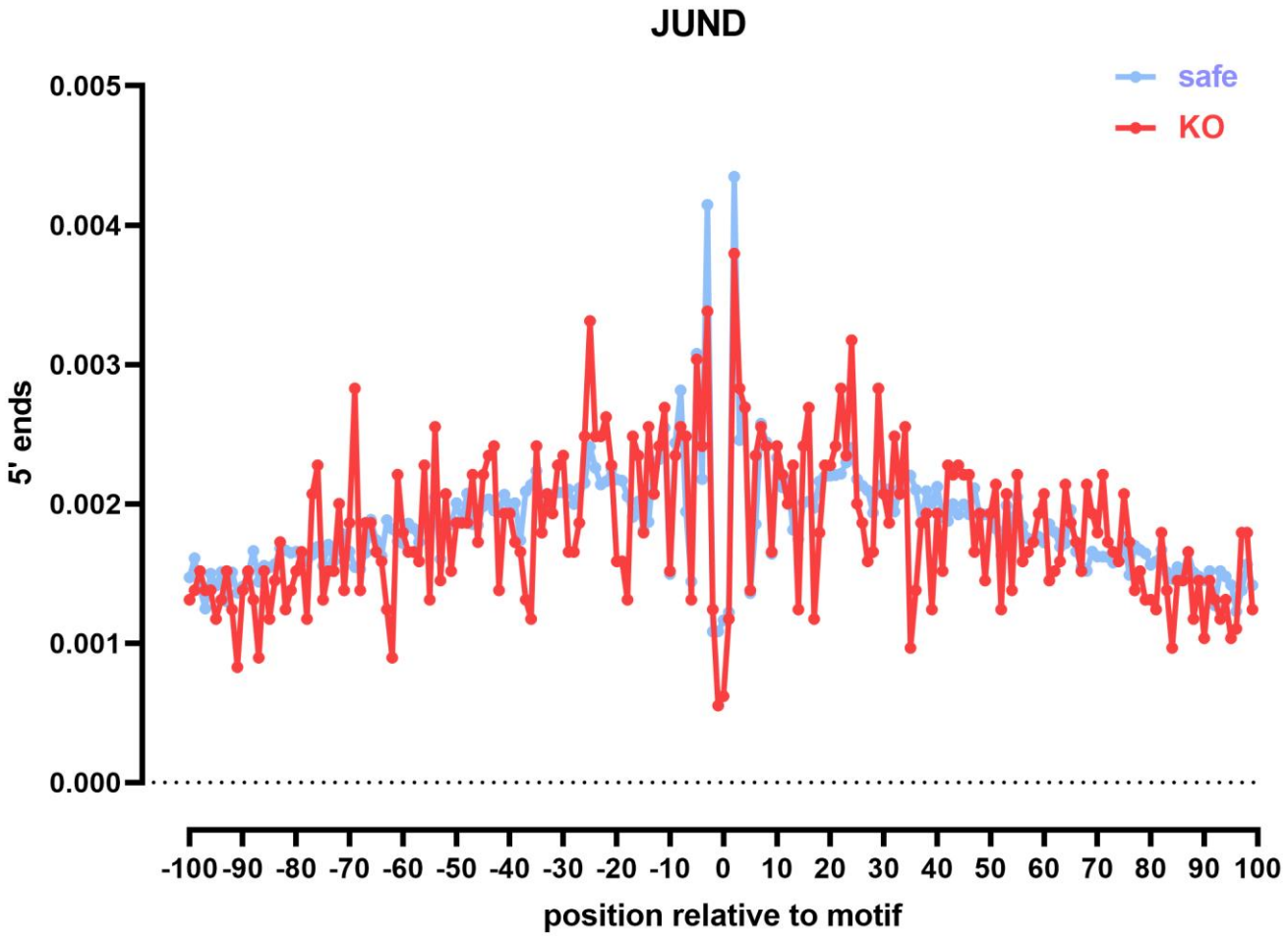

TF UMAP
UMAP2
UMAP1
5
4
3
2
1
0
-1
-5
-4
-3
-2
-1
0

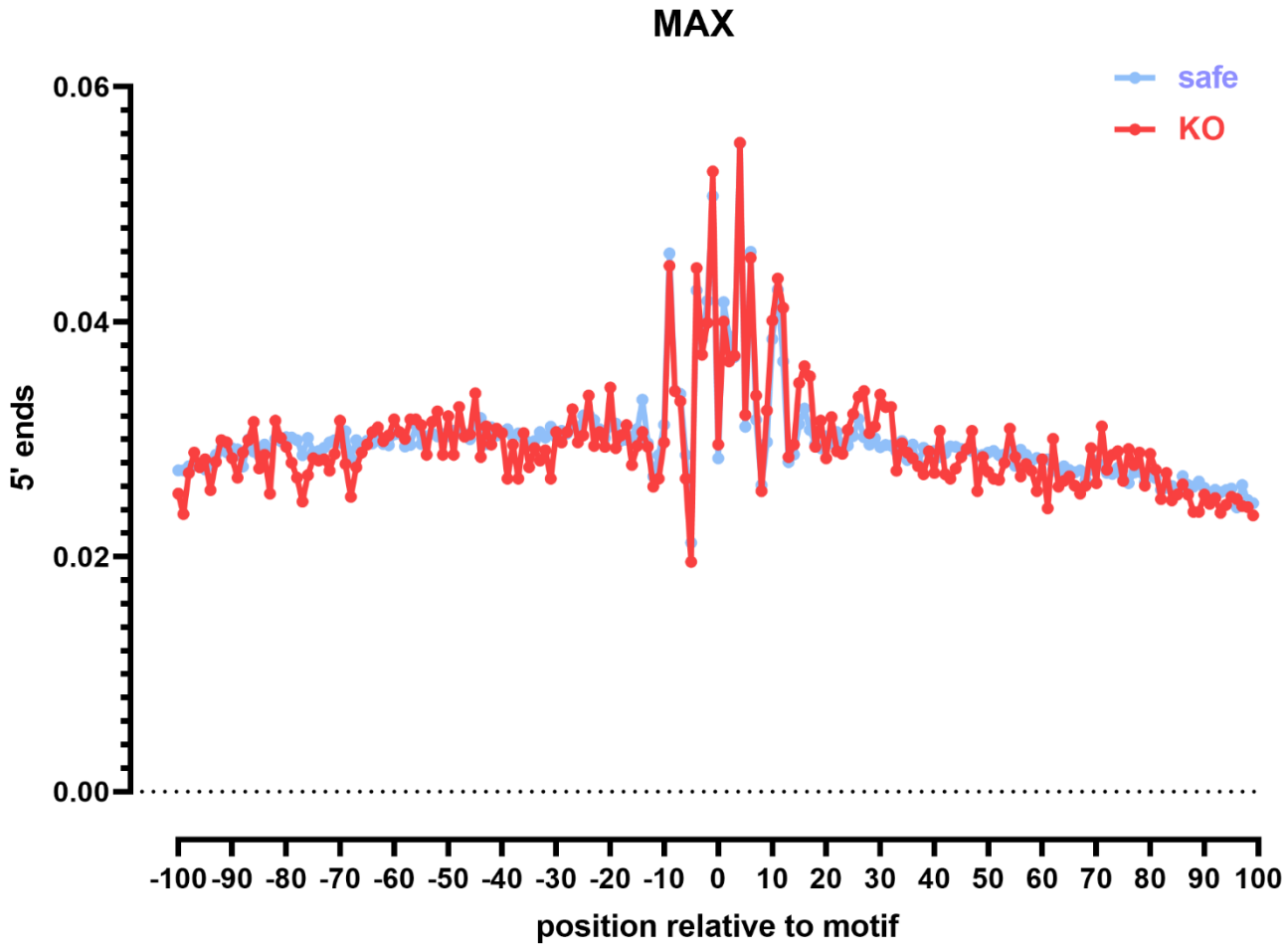

peak UMAP
UMAP2
UMAP1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11

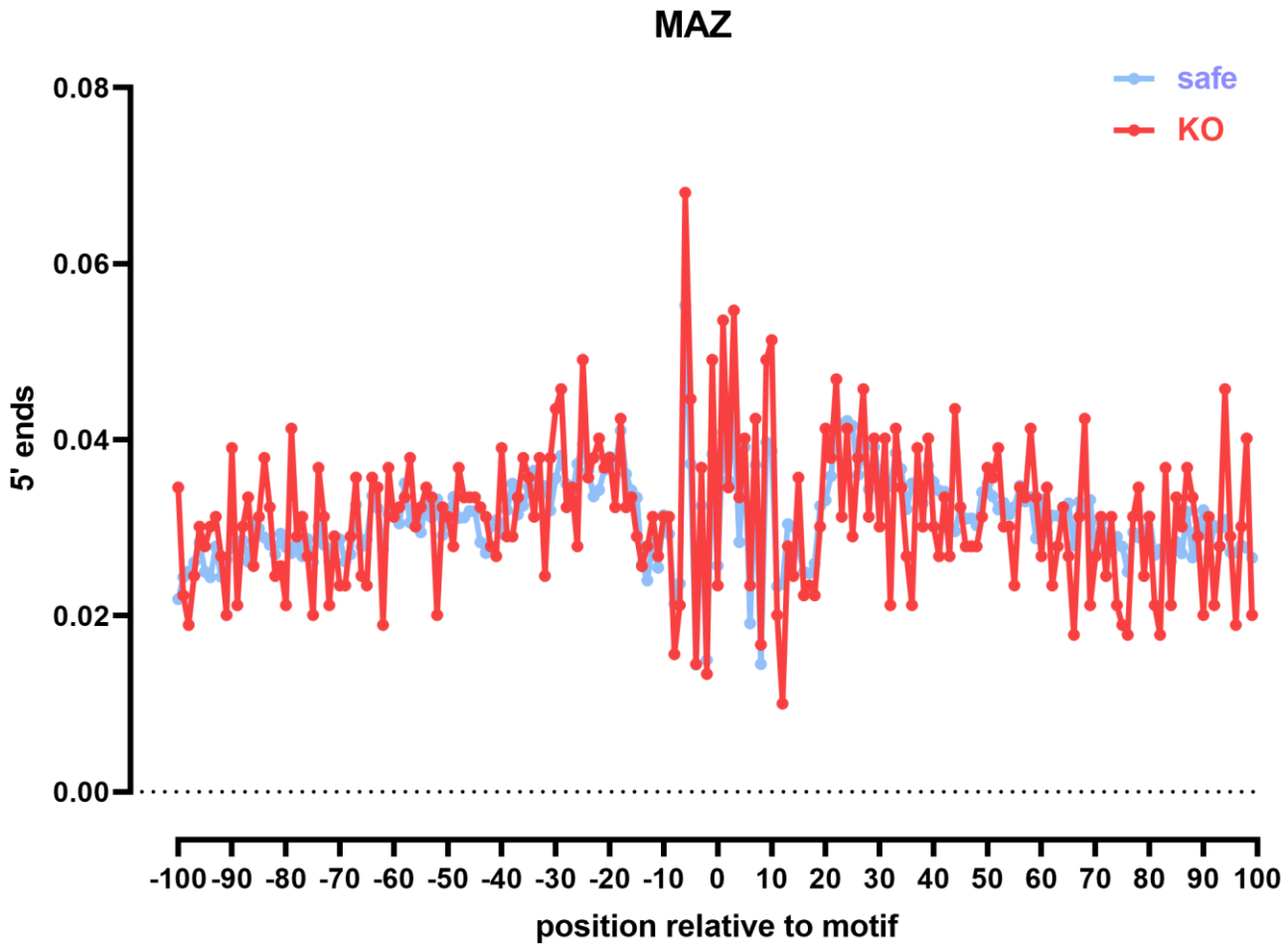
MOTIF ACCESSIBILITY

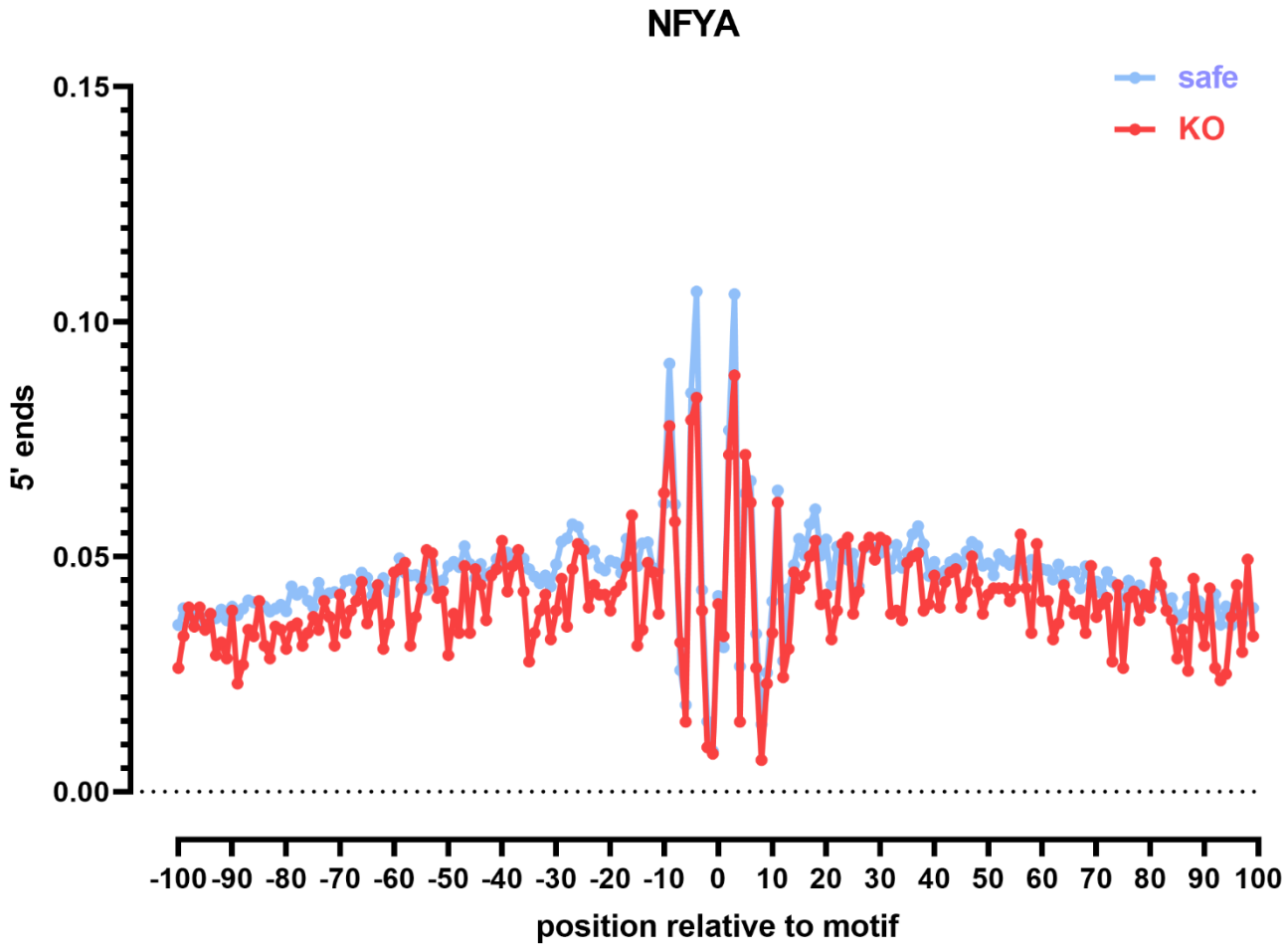


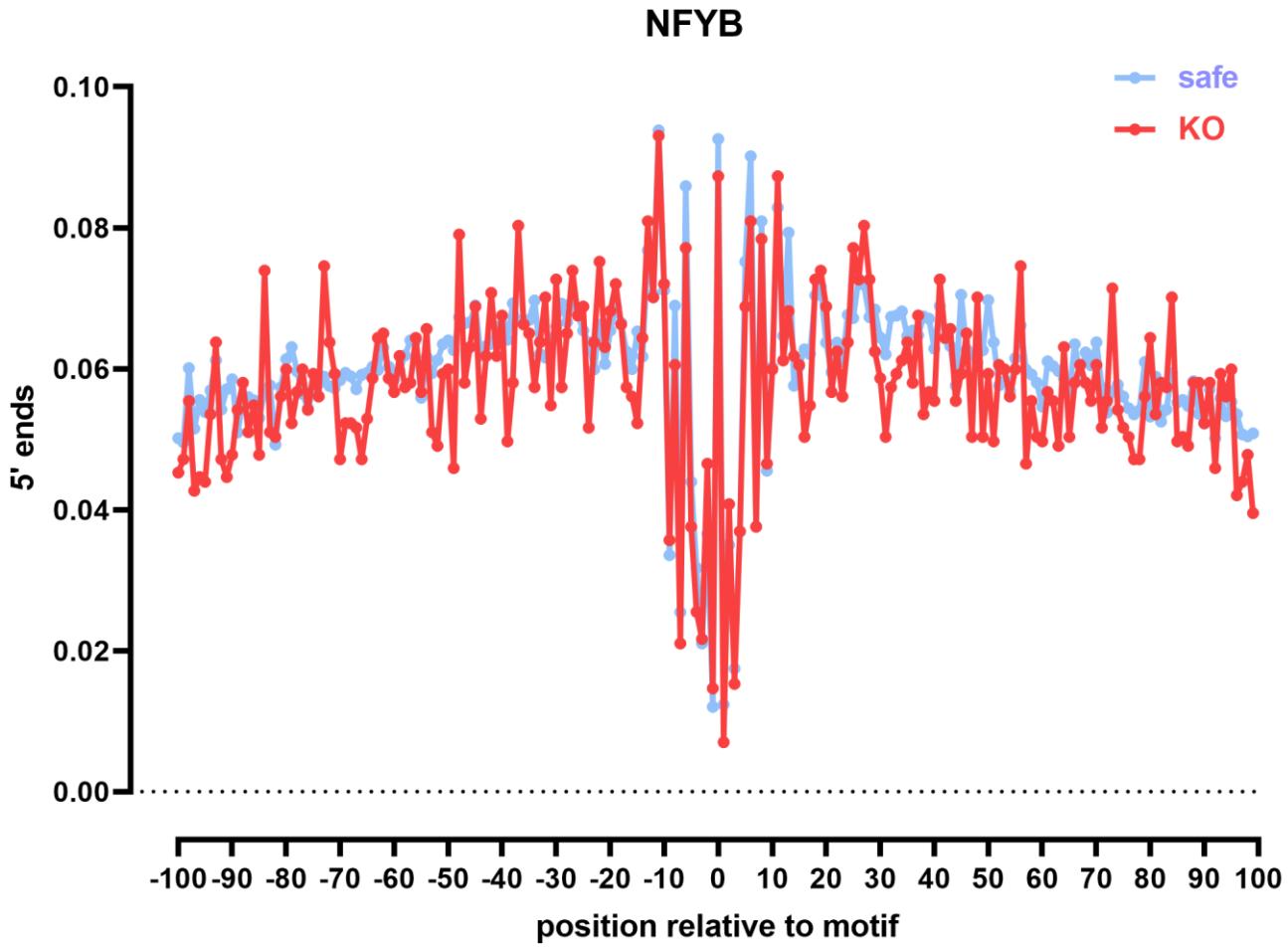

GABPA
safe
KO
0.15
0.10
0.05
0.00
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif


JUND
safe
KO
0.005
0.004
0.003
0.002
0.001
0.000
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif

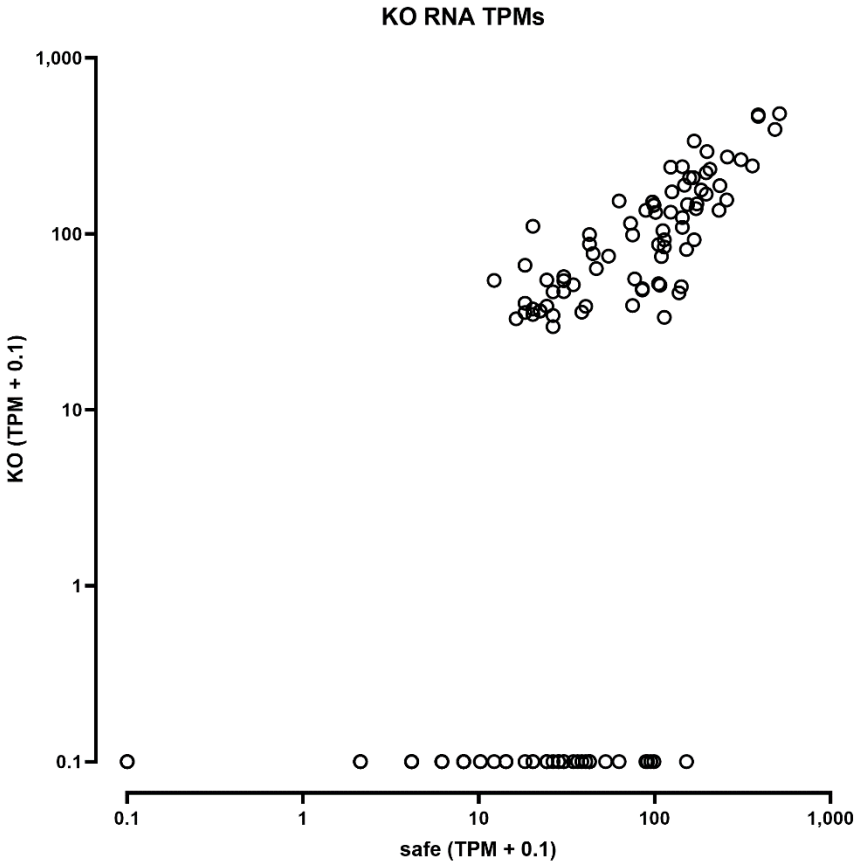

MAX
safe
KO
0.06
0.04
0.02
0.00
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif


MAZ
safe
KO
0.08
0.06
0.04
0.02
0.00
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif


NFYA
safe
KO
0.15
0.10
0.05
0.00
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif


NFYB
safe
KO
0.10
0.08
0.06
0.04
0.02
0.00
5' ends
-100 -90 -80 -70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100
position relative to motif

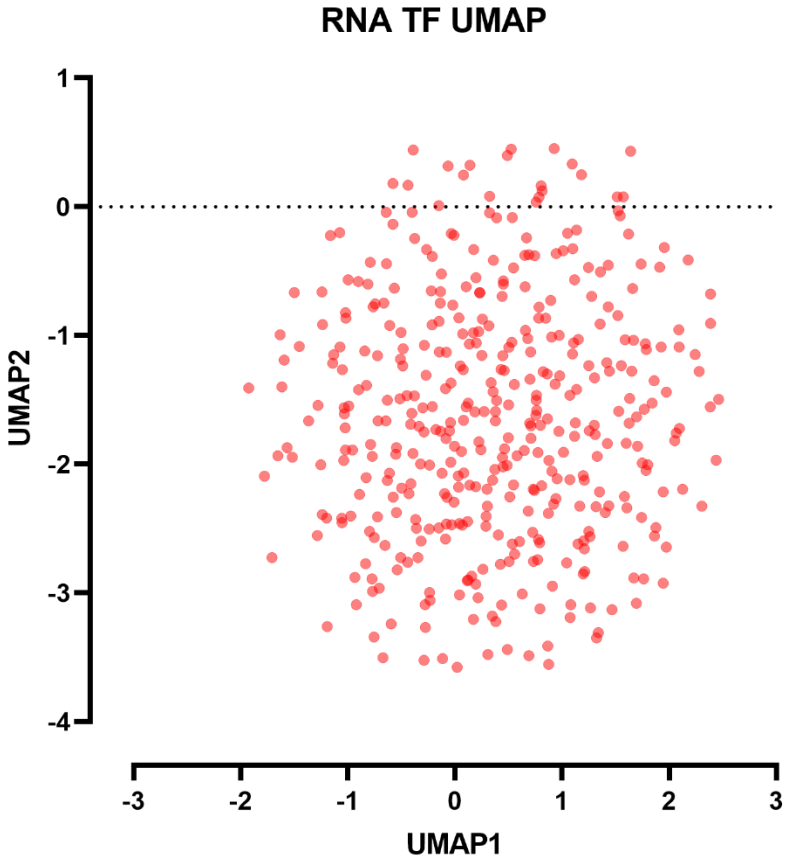
RNA



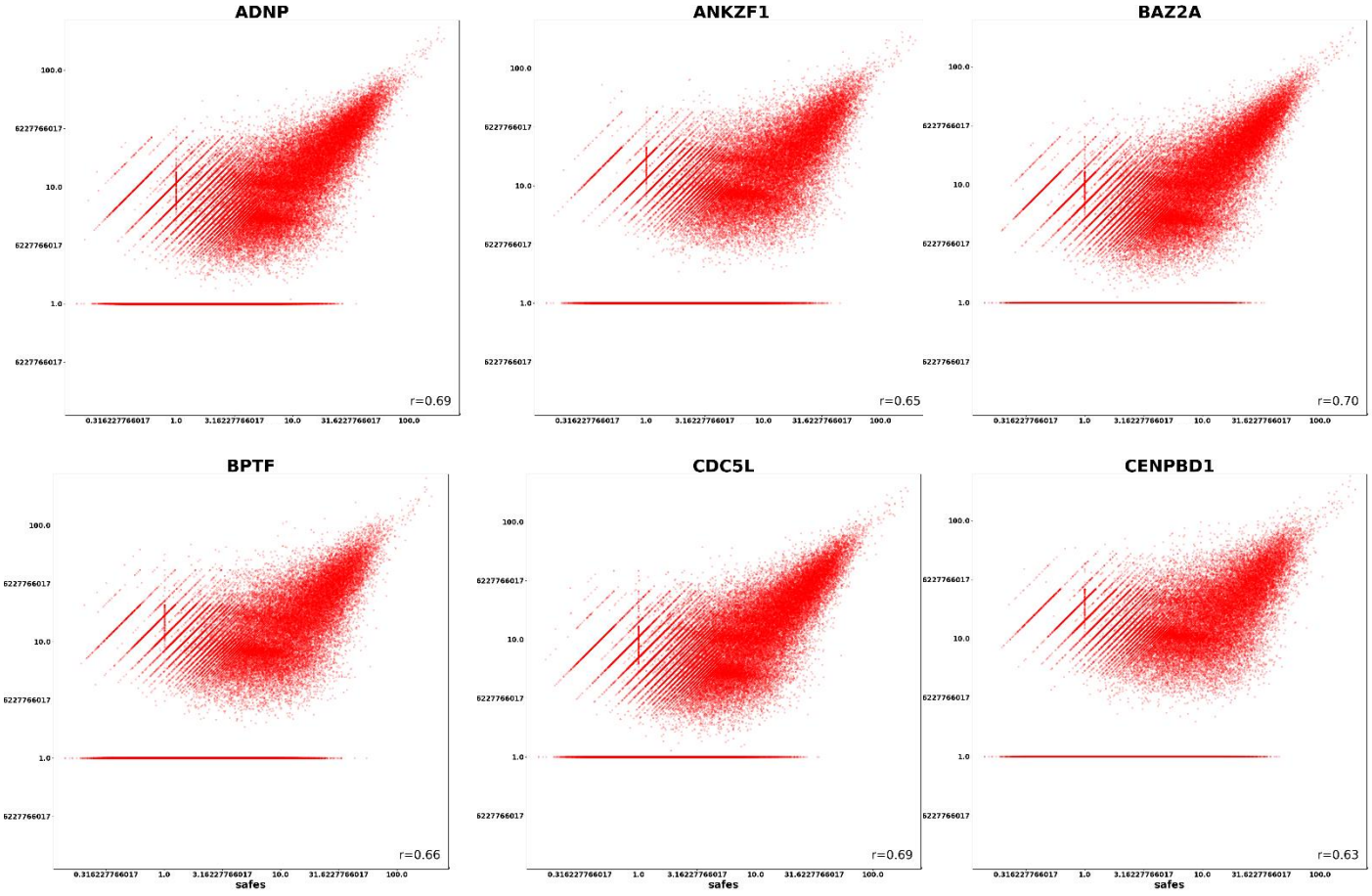
Note that:

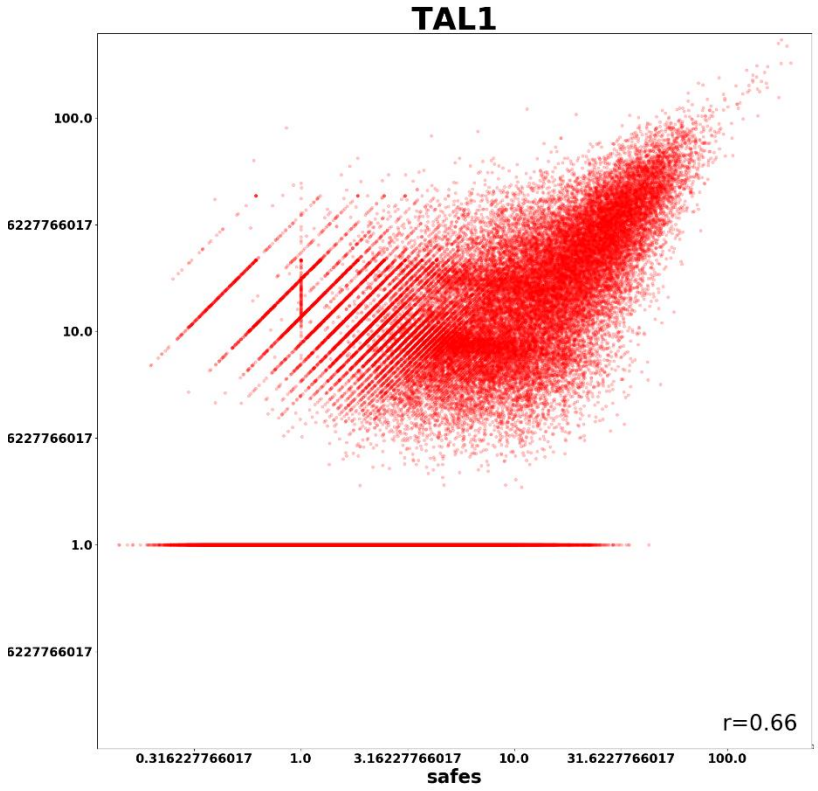
- The "safes" have higher coverage because there are 40 safe sgRNA vs. 2 KO sgRNA per gene
- We are still at only a few cells per sgRNA
- These libraries are a mixture of D2 and D7, and at D2 we do not expect a full KO yet

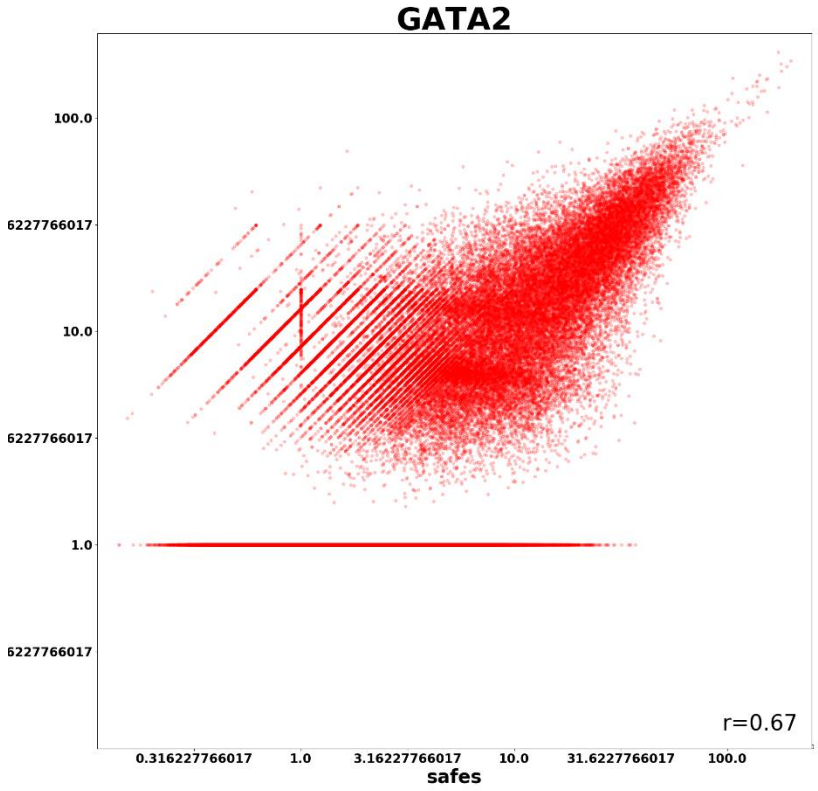
RNA

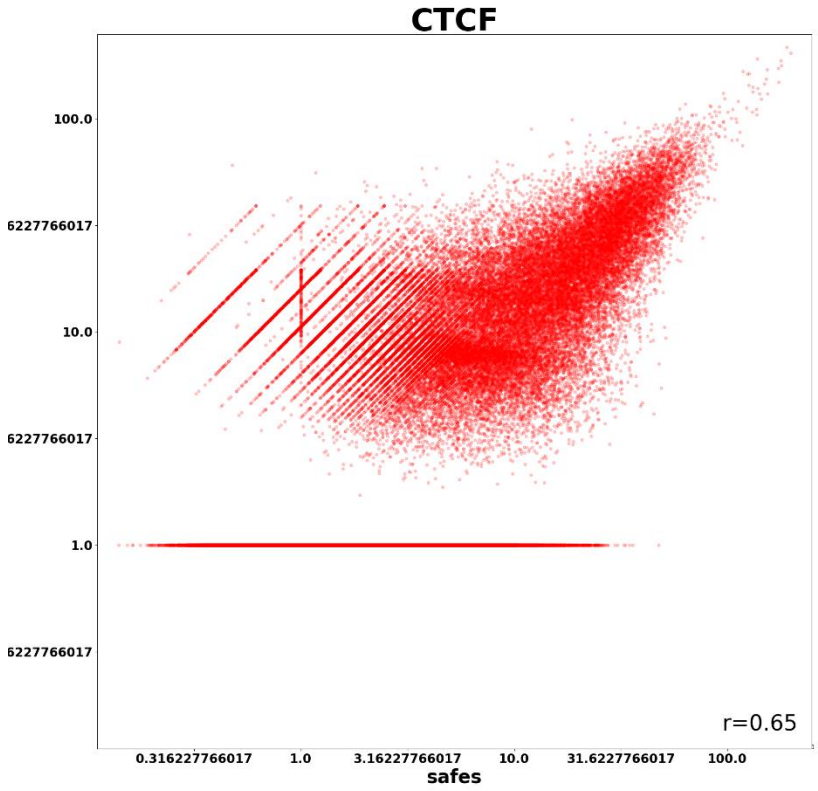


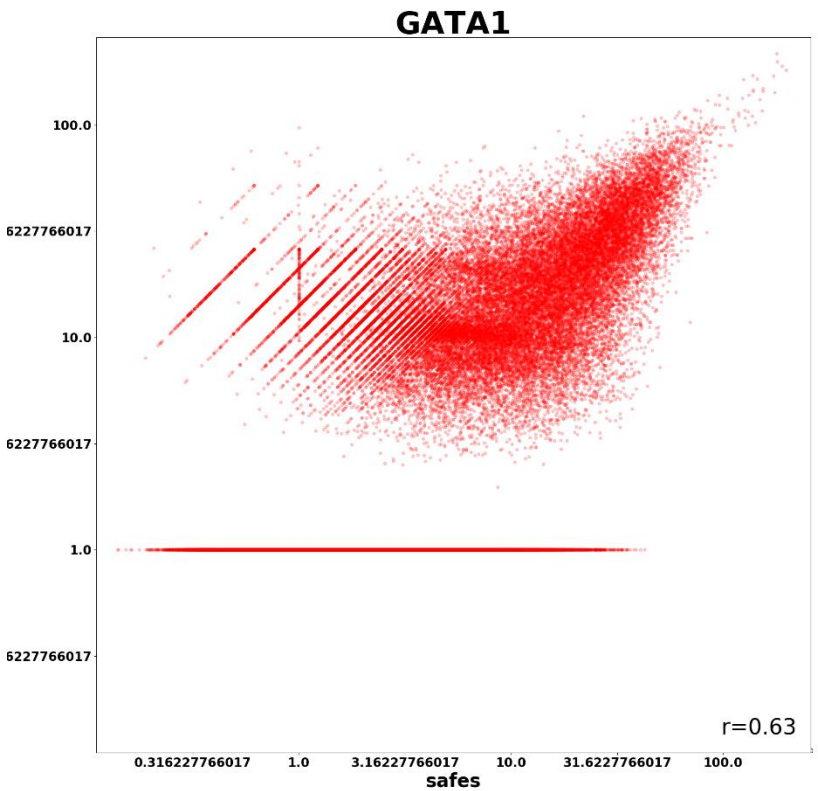
ATAC INDIVIDUAL PETRUBATIONS

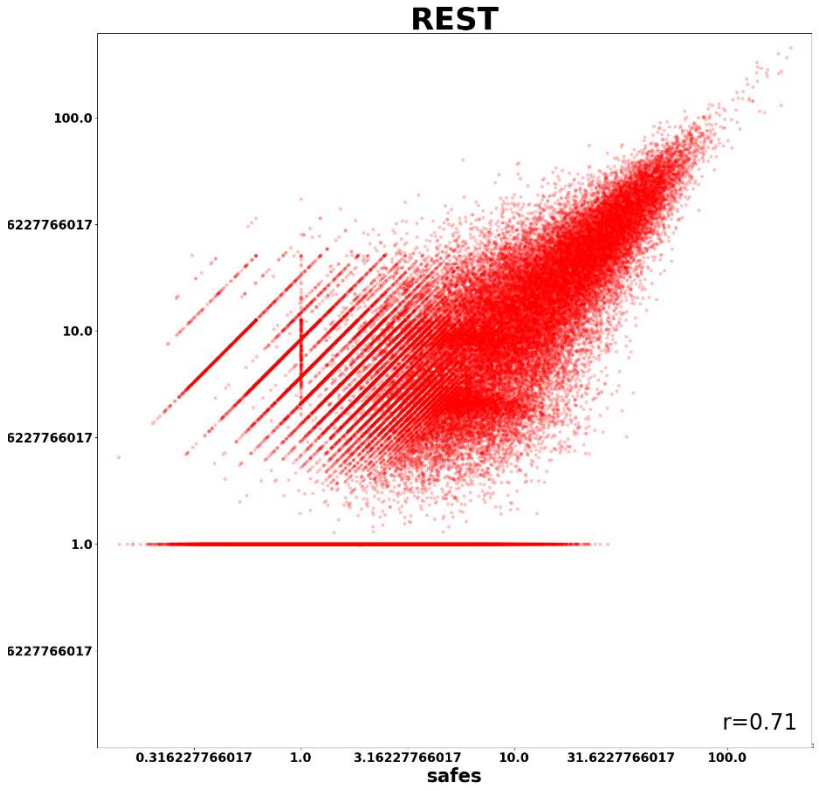

ADNP
ANKZF1
BAZ2A
BPTF
CDC5L
CENPBD1
r=0.69
r=0.65
r=0.70
r=0.66
r=0.69
r=0.63
safes
safes

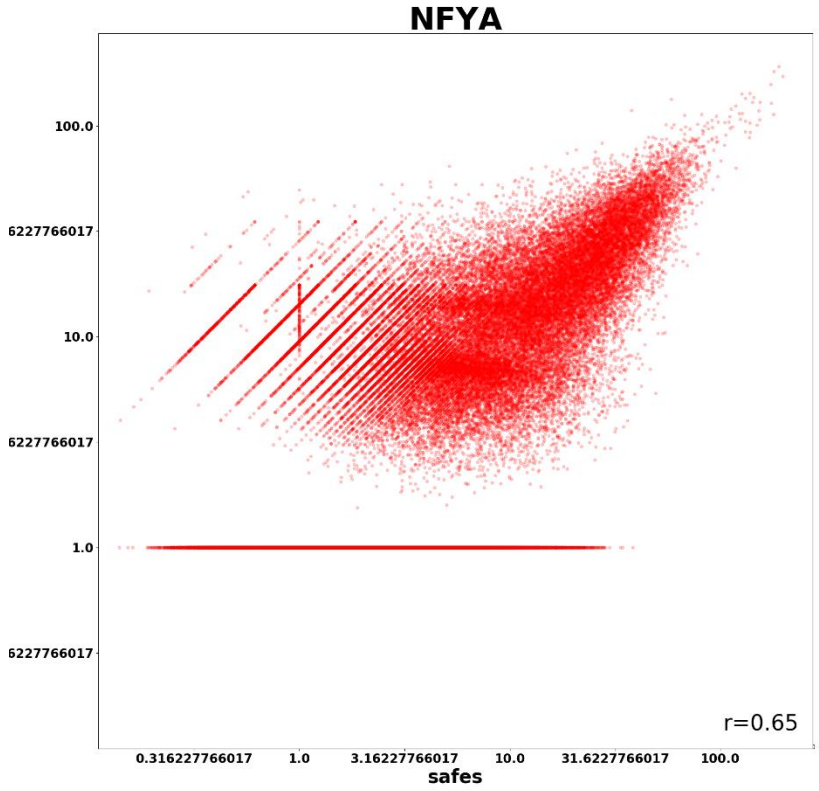

TAL1
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.66
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes

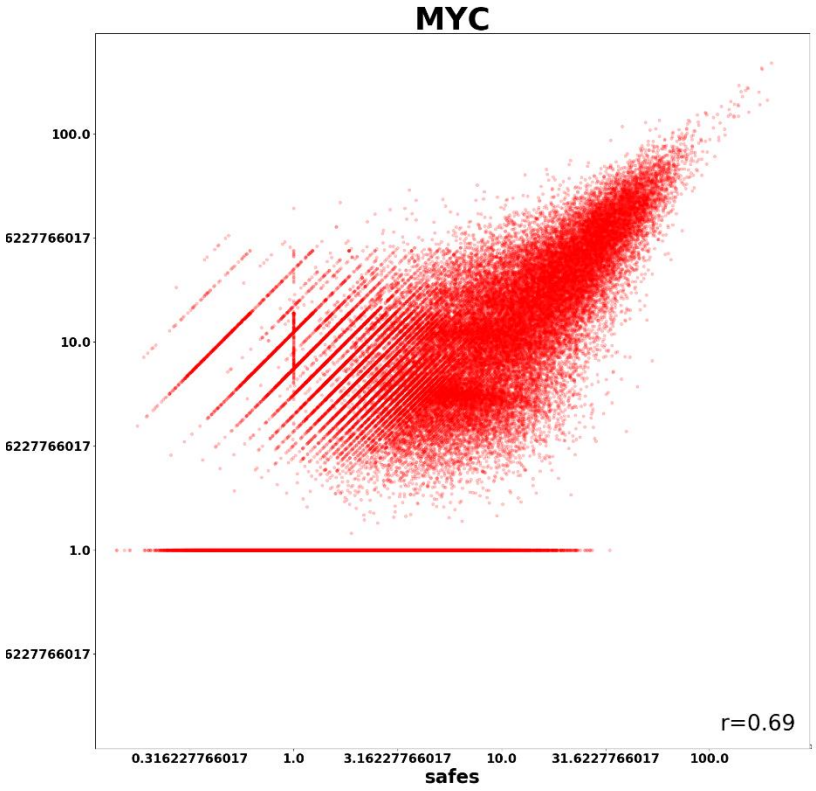

GATA2
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.67
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes

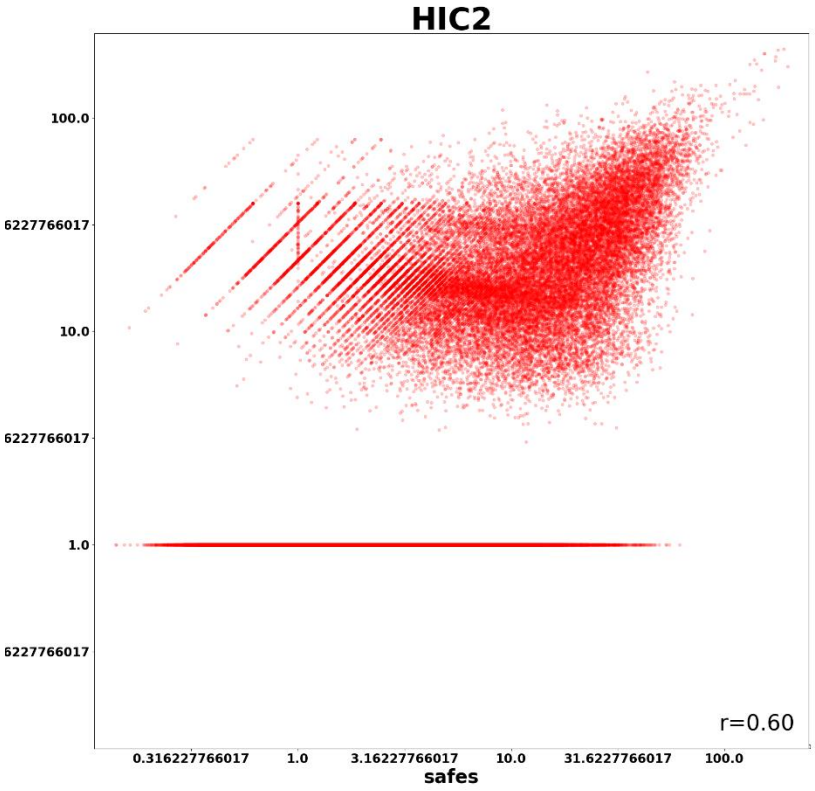

CTCF
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.65
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes


GATA1
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.63
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes


REST
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.71
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes


NFYA
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.65
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes


MYC
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.69
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes


HIC2
100.0
6227766017
10.0
6227766017
1.0
6227766017
r=0.60
0.316227766017
1.0
3.16227766017
10.0
31.6227766017
100.0
safes

Full set:

https://mitra.stanford.edu/kundaje/marinovg/oak/ENCODE4/single-cell/2023-07-11-UG-data/subbam_MACS_RPM-png/