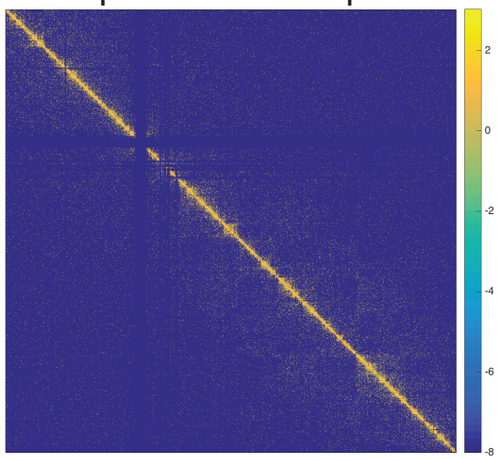
MrTADFinder

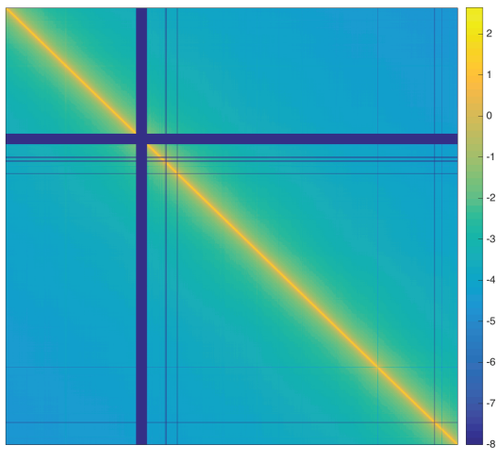
KKY, July 2016

A.

input: contact map W



null model E



$$E_{ij} = c_i^* c_j^* f(|i-j|)$$

Choose a particular resolution γ
Optimize Q over all possible partitions

$$Q = \frac{1}{2N} \sum_{ij} (W_{ij} - \gamma E_{ij}) \delta_{\sigma_i \sigma_j}$$

γ: resolution parameter

Multiple runs to define boundary scores for all adjacient bins

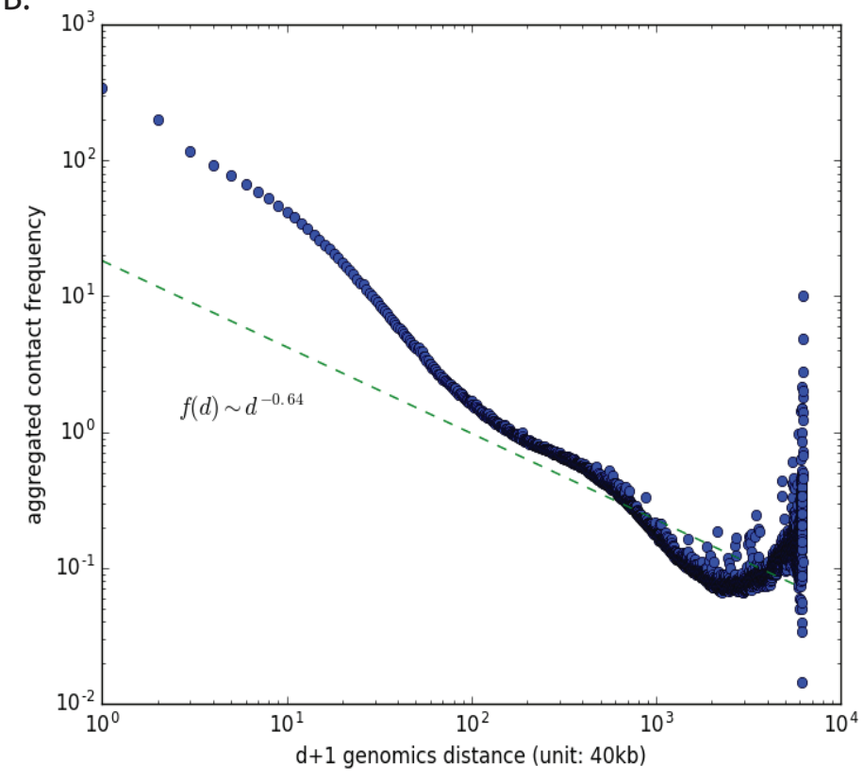
consensus boundaries based on the boundary scores

consensus domains

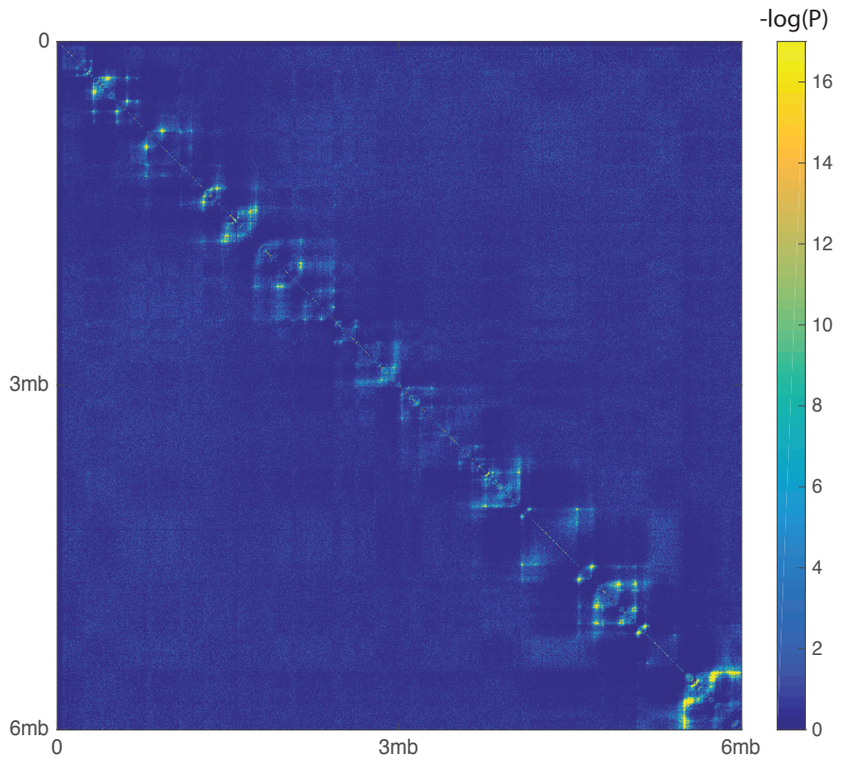
output: TADs

Row	chr	domain_st	domain_ed
1	"chr10"	40001	1120000
2	"chr10"	1120001	3240000
3	"chr10"	3240001	4840000
4	"chr10"	4840001	5680000
5	"chr10"	5680001	5760000
6	"chr10"	5760001	5920000
7	"chr10"	5920001	6000000
8	"chr10"	6000001	7560000
9	"chr10"	7560001	9360000
10	"chr10"	9360001	11520000

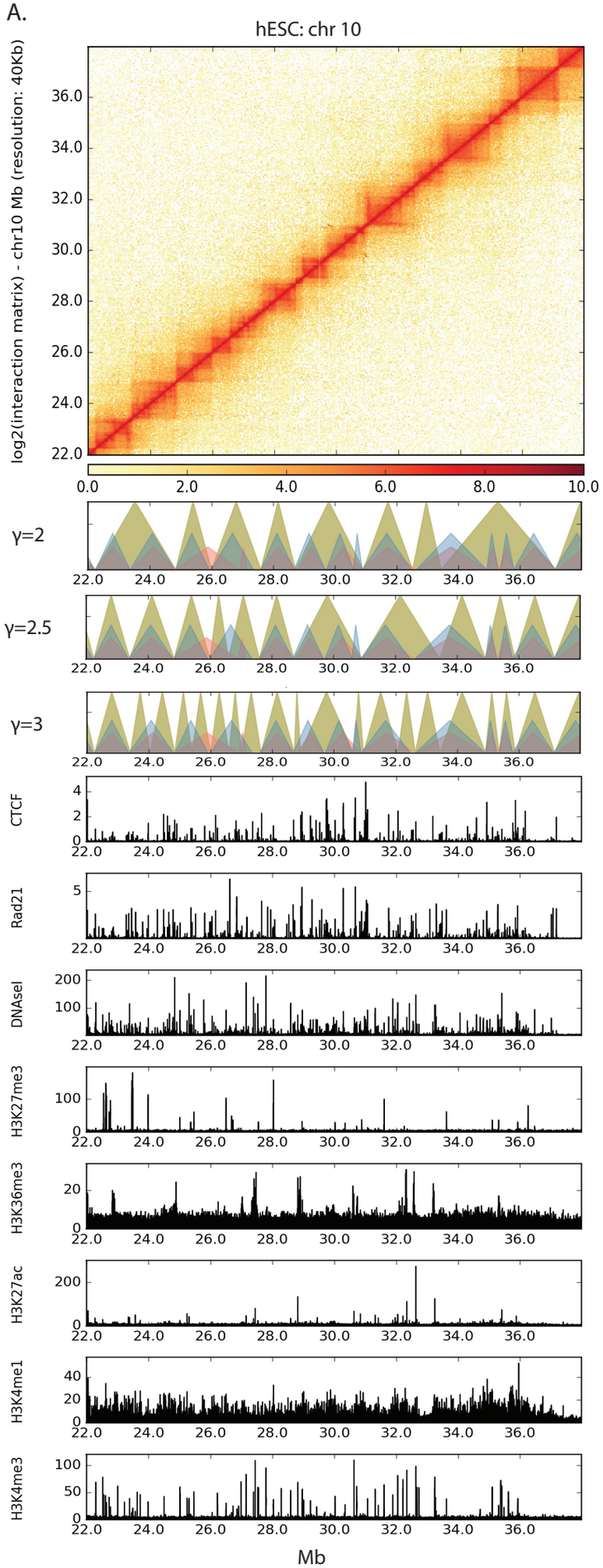
B.



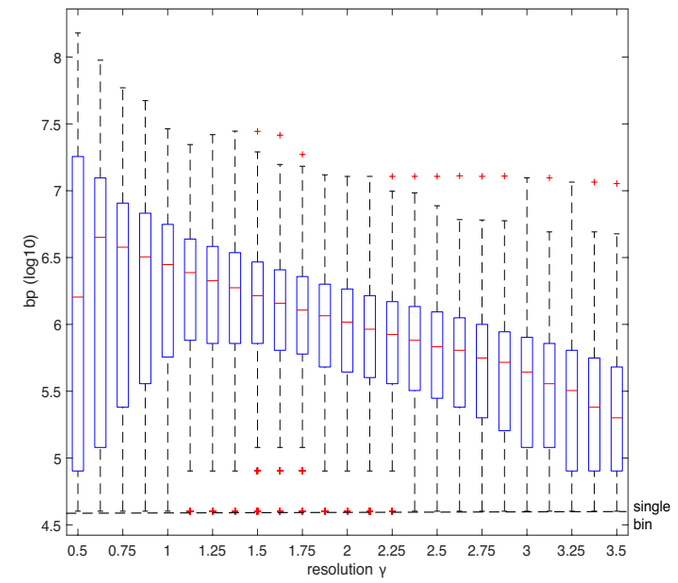
C.



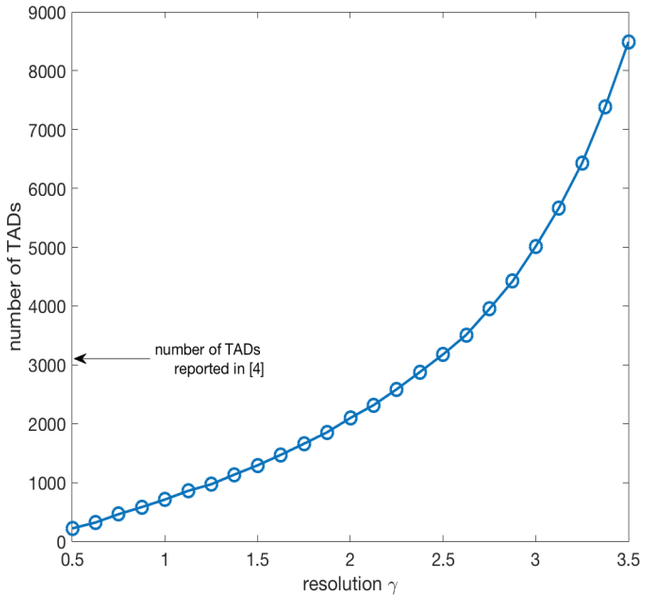
A.


hESC: chr 10
log2(interaction matrix) - chr10 Mb (resolution: 40Kb)
γ=2
γ=2.5
γ=3
CTCF
Rad21
DNAseI
H3K27me3
H3K36me3
H3K27ac
H3K4me1
H3K4me3
Mb

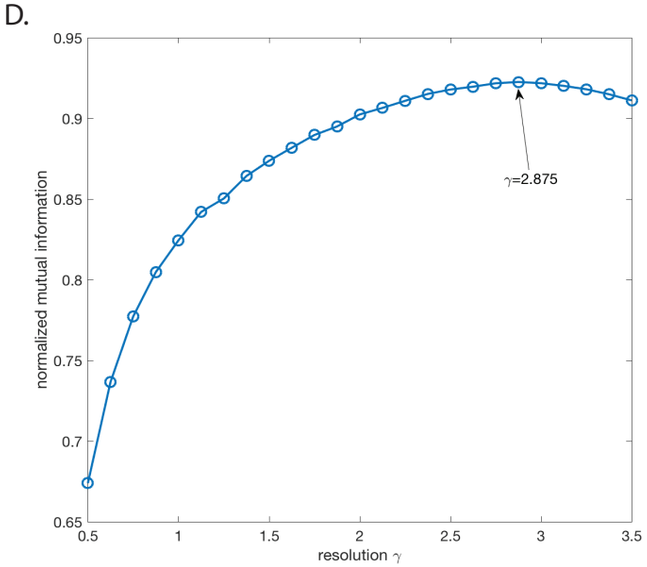
B.


bp (log10)
resolution γ
single bin

C.

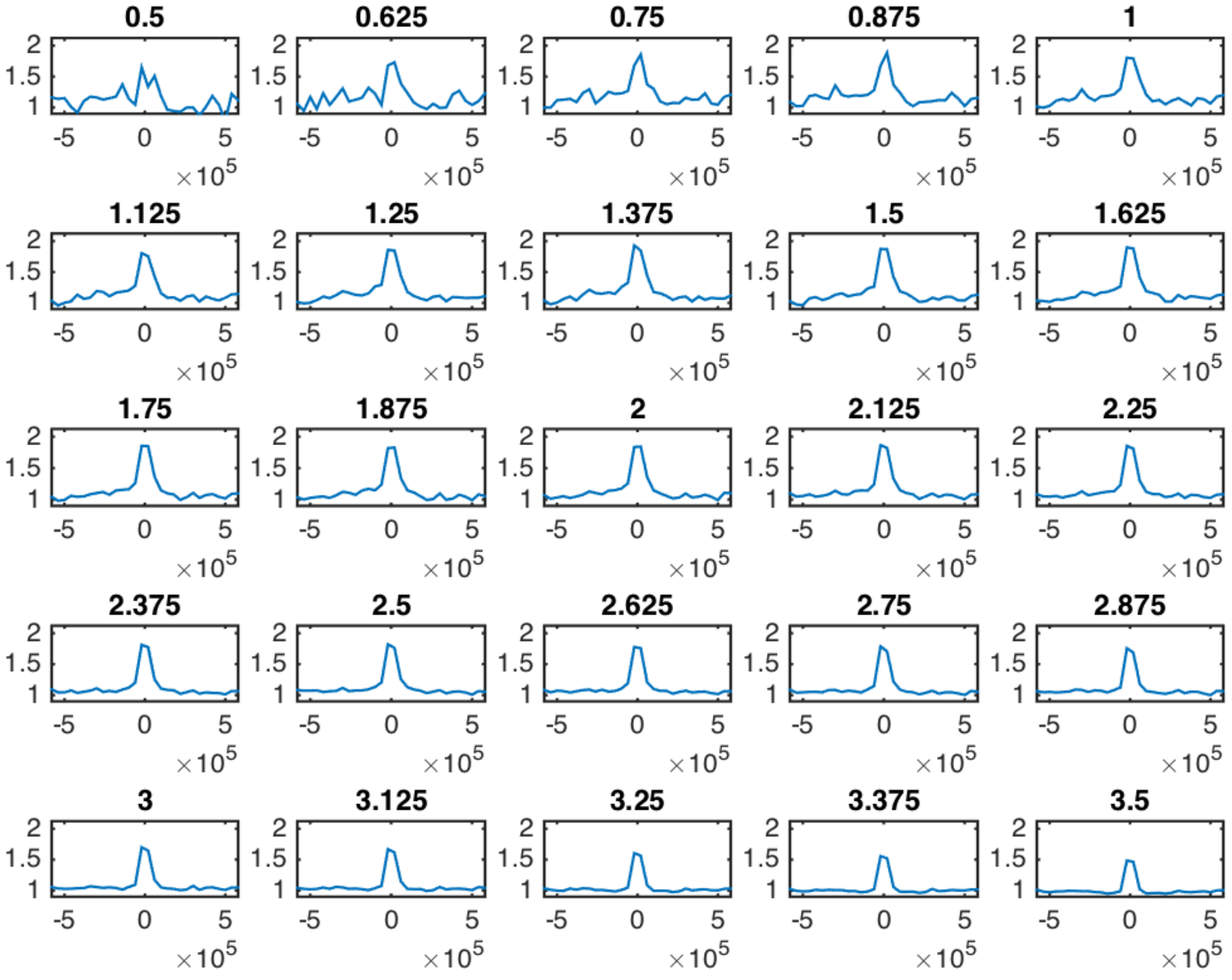

number of TADs
number of TADs reported in [4]
resolution γ

D.


normalized mutual information
γ=2.875
resolution γ

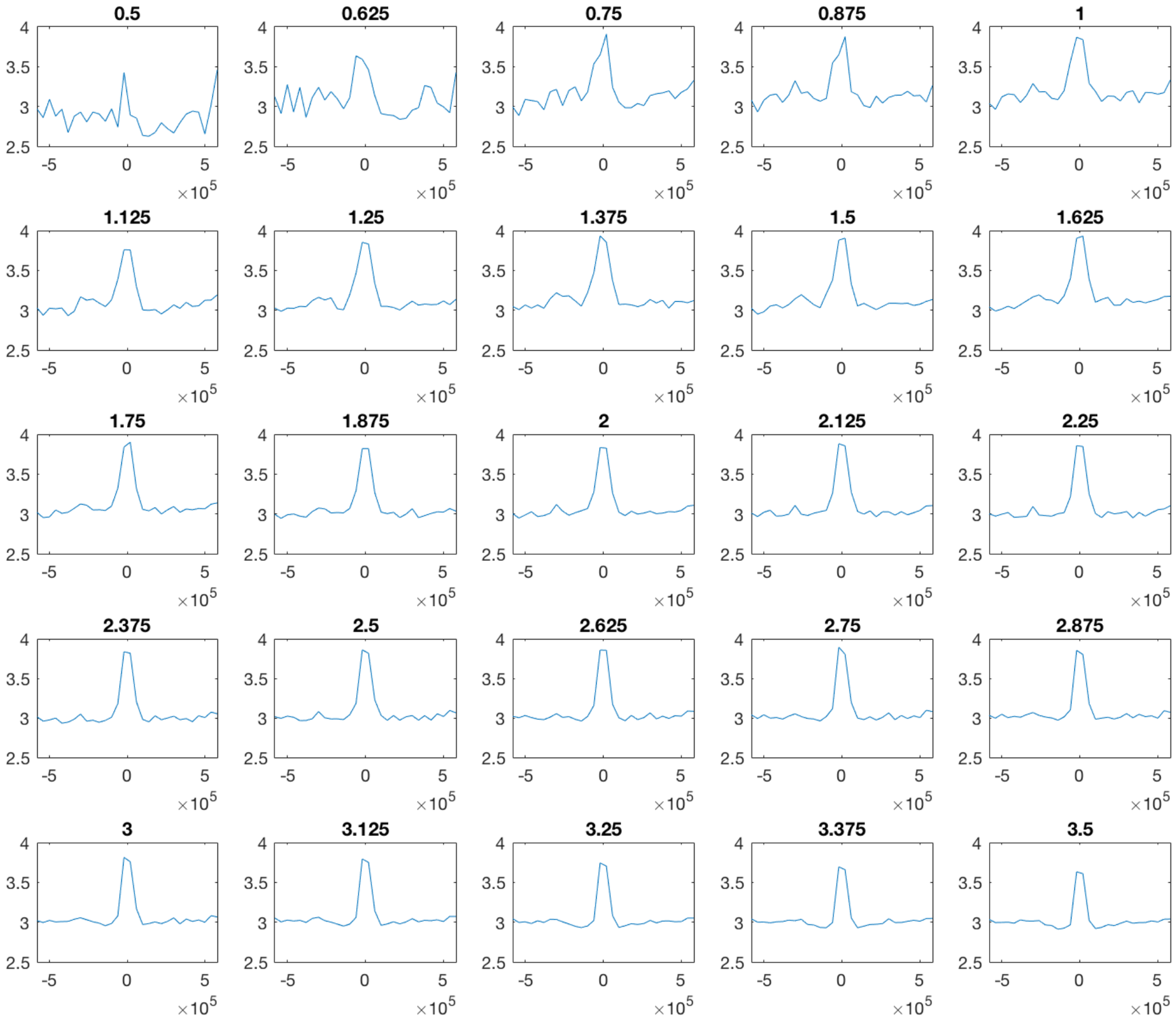
TFs near TAD boundaries at various resolutions (peak density)

Rad21



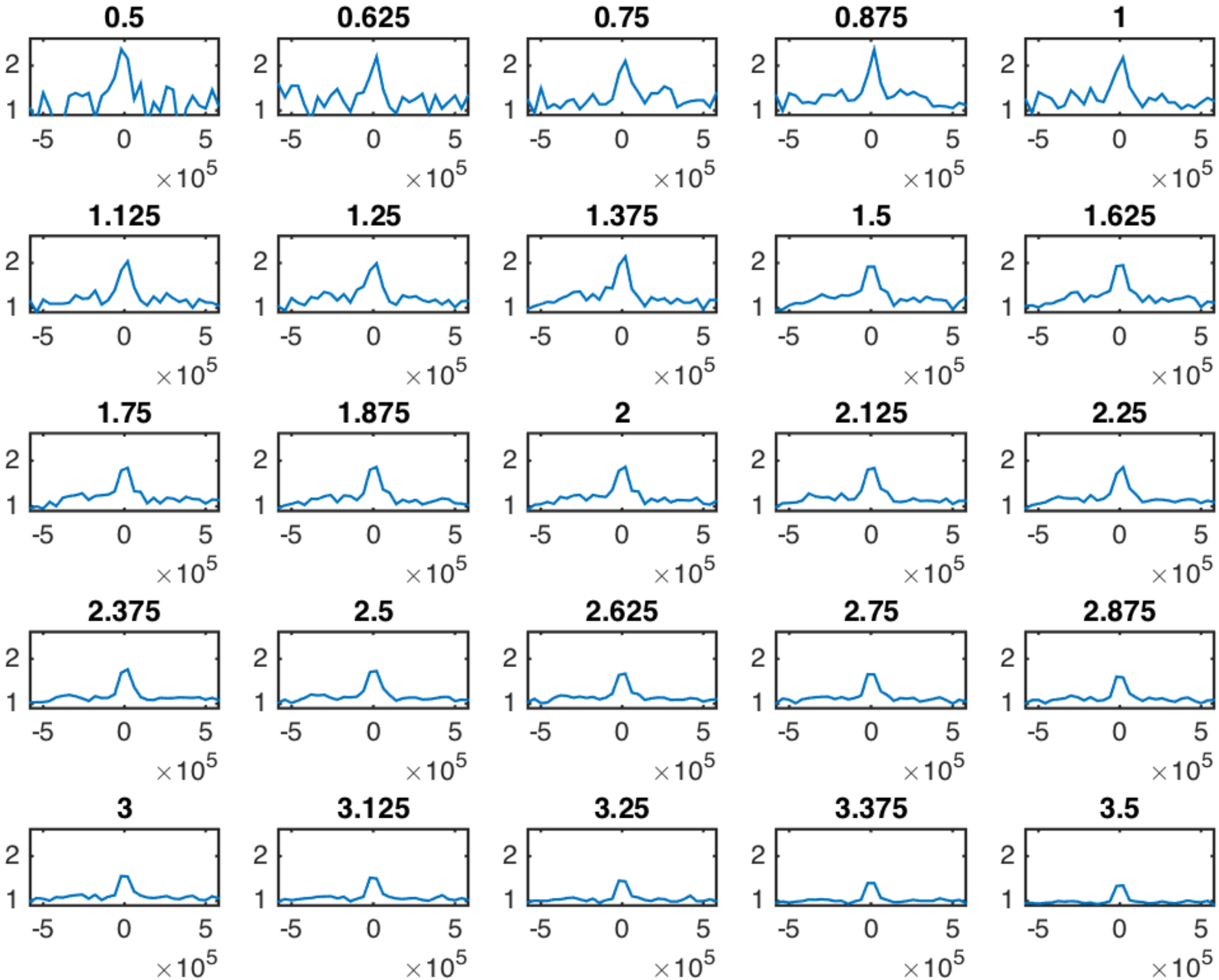
TFs near TAD boundaries at various resolutions (signal)

Rad21



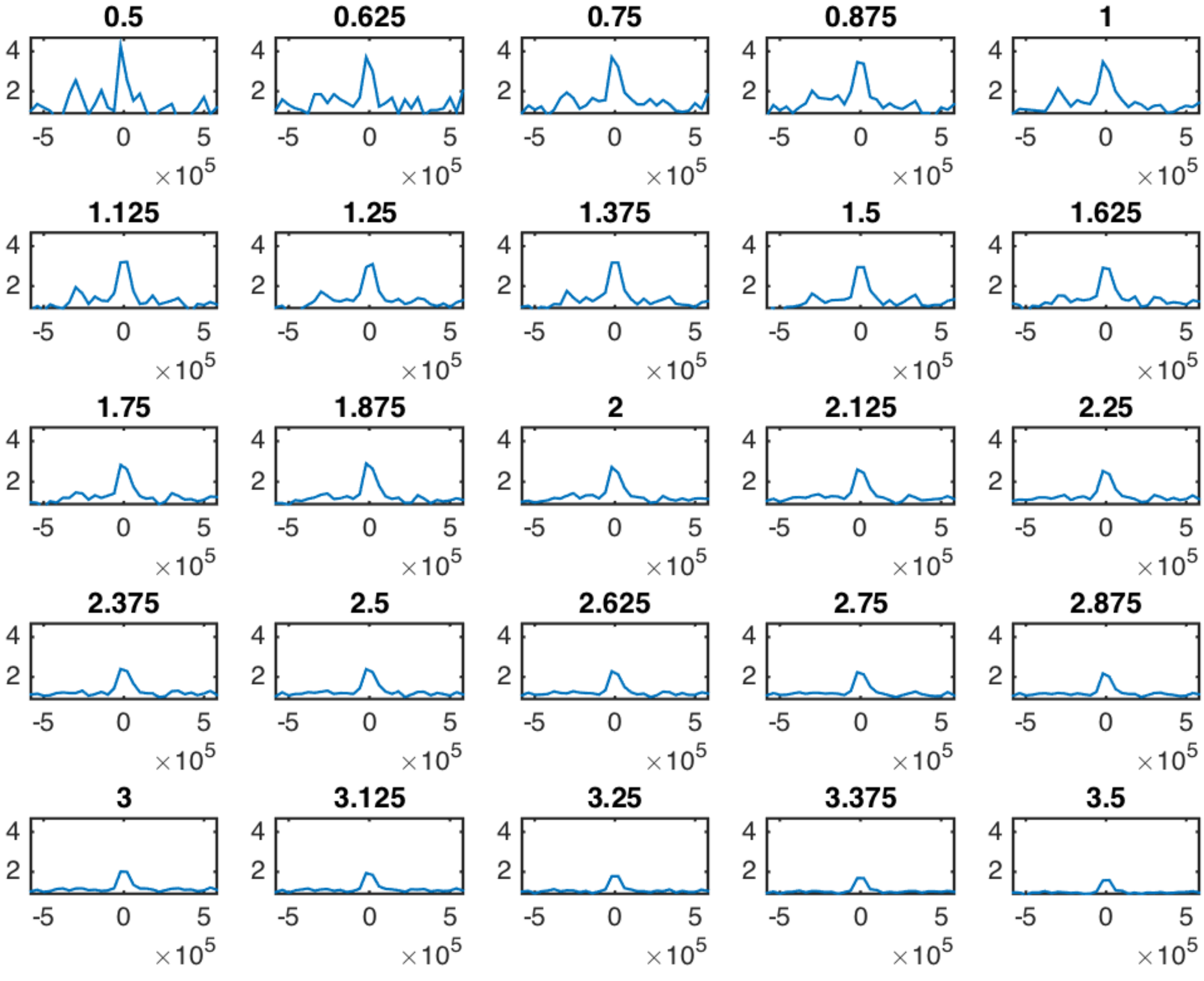
TFs near TAD boundaries at various resolutions (peak density)

MAX



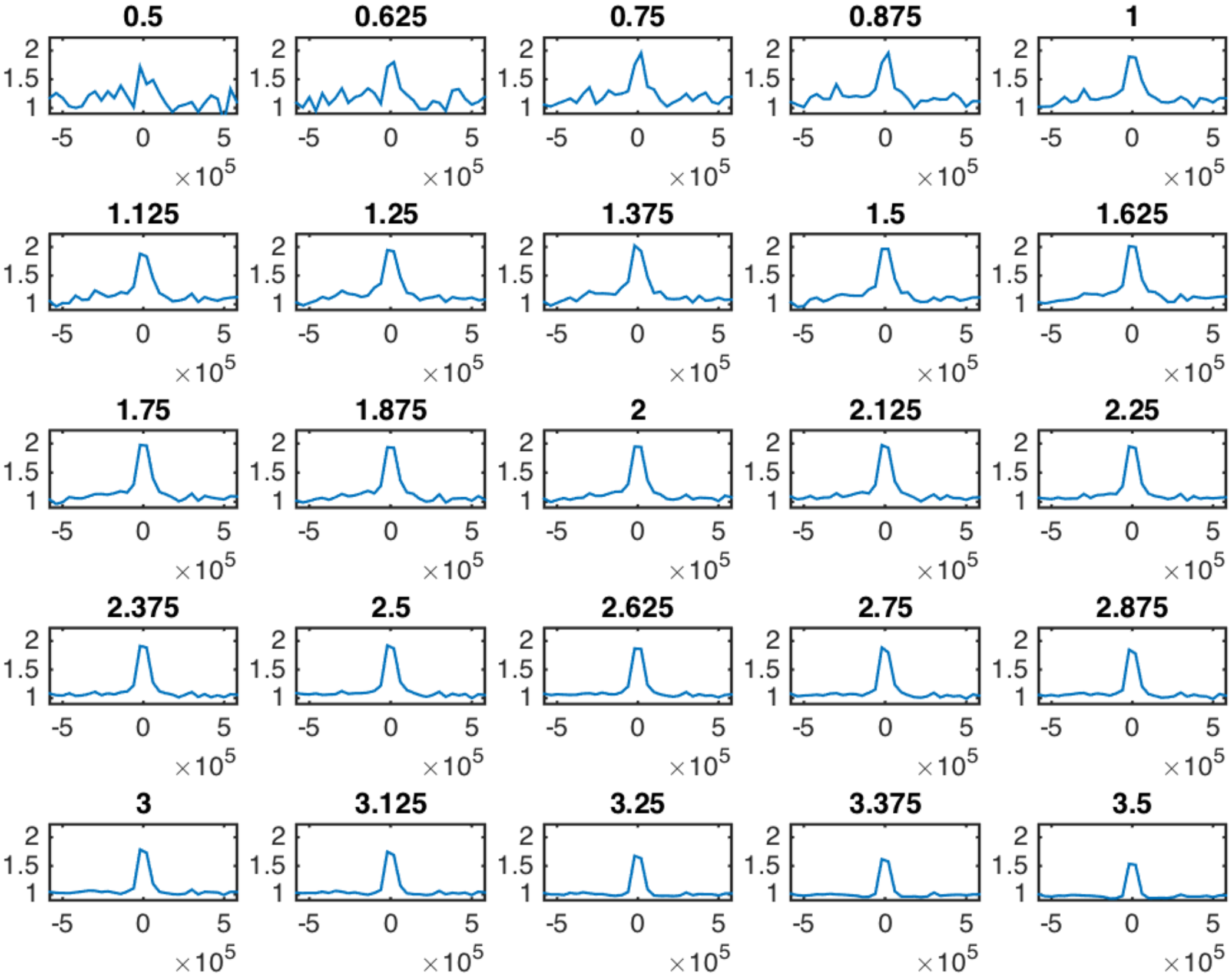
TFs near TAD boundaries at various resolutions (peak density)

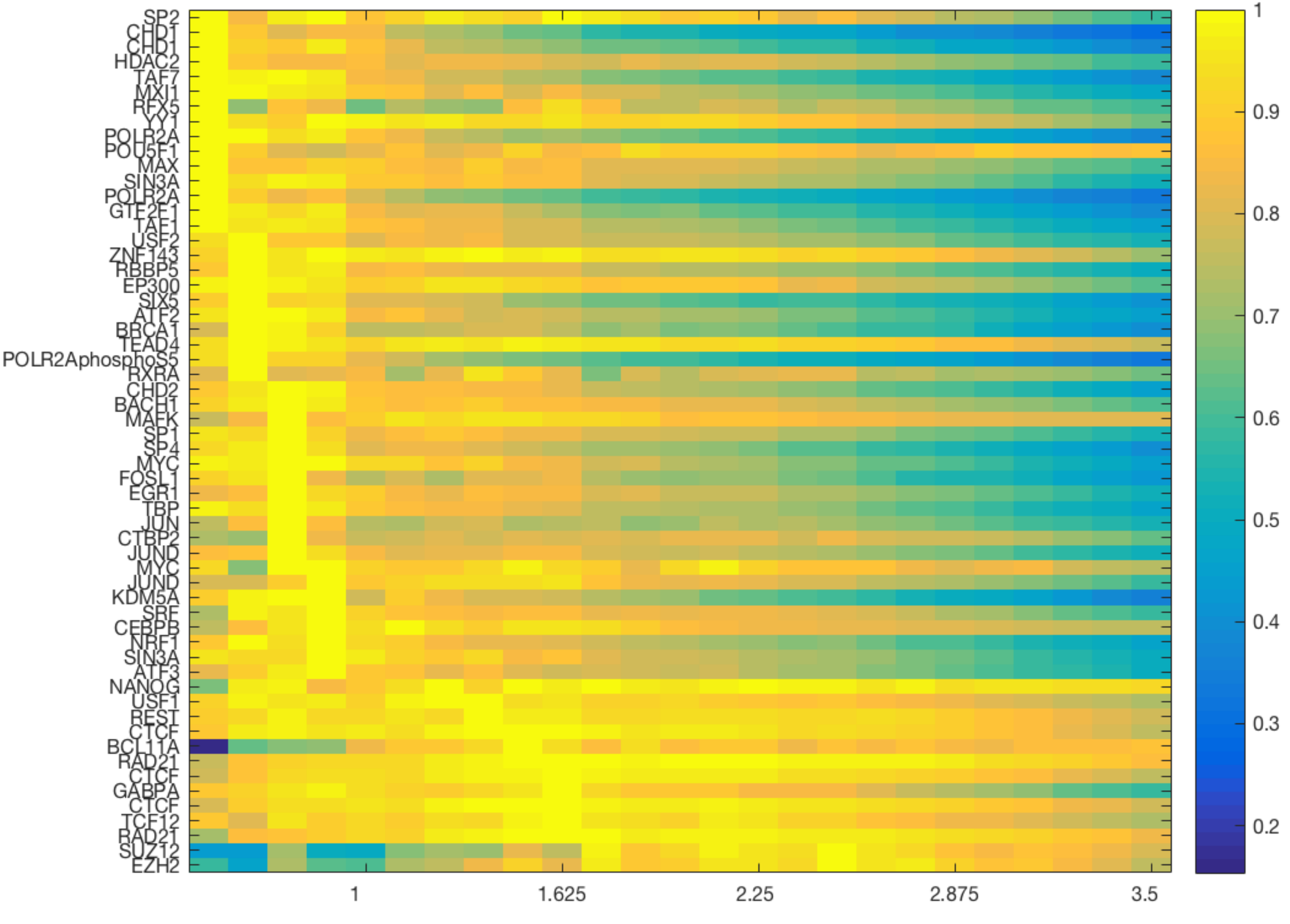
MYC



TFs near TAD boundaries at various resolutions (peak density)

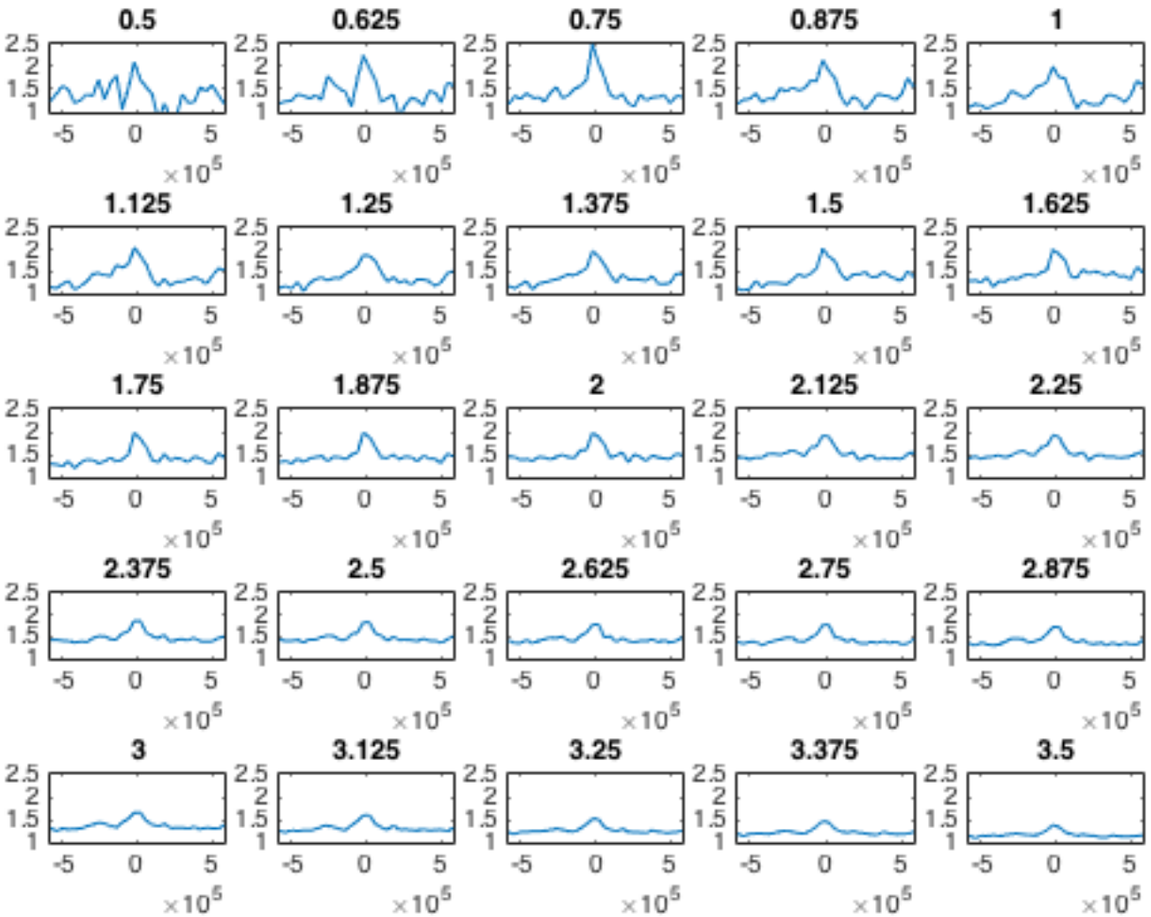
CTCF

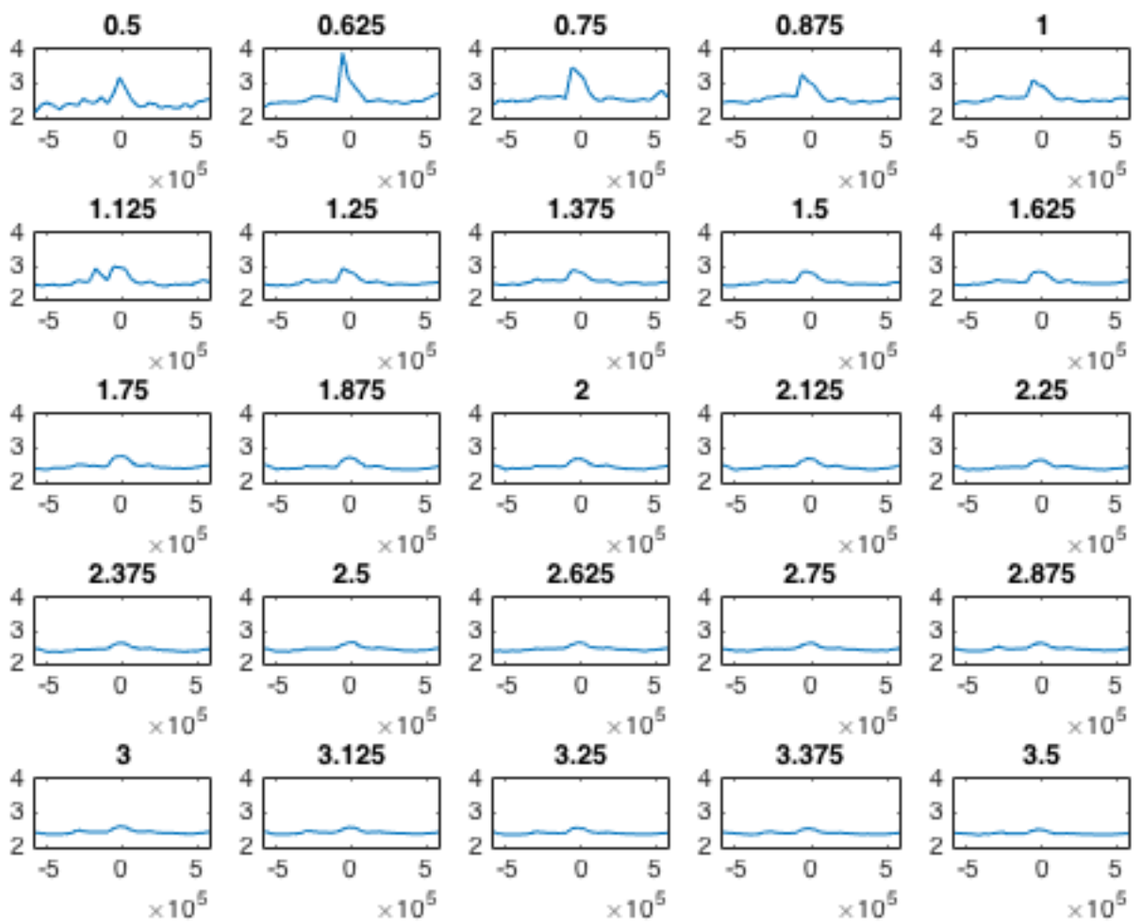



SP2
CHD1
CHD1
HDAC2
TAF7
MXI1
RFX5
YY1
YY1
POLR2A
POU5F1
MAX
SIN3A
POLR2A
GTF2F1
TAF1
USF2
ZNF143
RBBP5
EP300
SIX5
ATF2
BRCA1
TEAD4
POLR2AphosphoS5
RXRA
CHD2
BACH1
MAFK
SP1
SP4
MYC
FOSL1
EGR1
TBP
JUN
CTBP2
JUND
MYC
JUND
KDM5A
SRF
CEBPB
NRF1
SIN3A
ATF3
NANOG
USF1
RESI
CTCF
BCL11A
RAD21
CTCF
GABPA
CTCF
TCF12
RAD21
SUZ12
EZH2
1
1.625
2.25
2.875
3.5
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

Chromatin features near TAD boundaries at various resolution

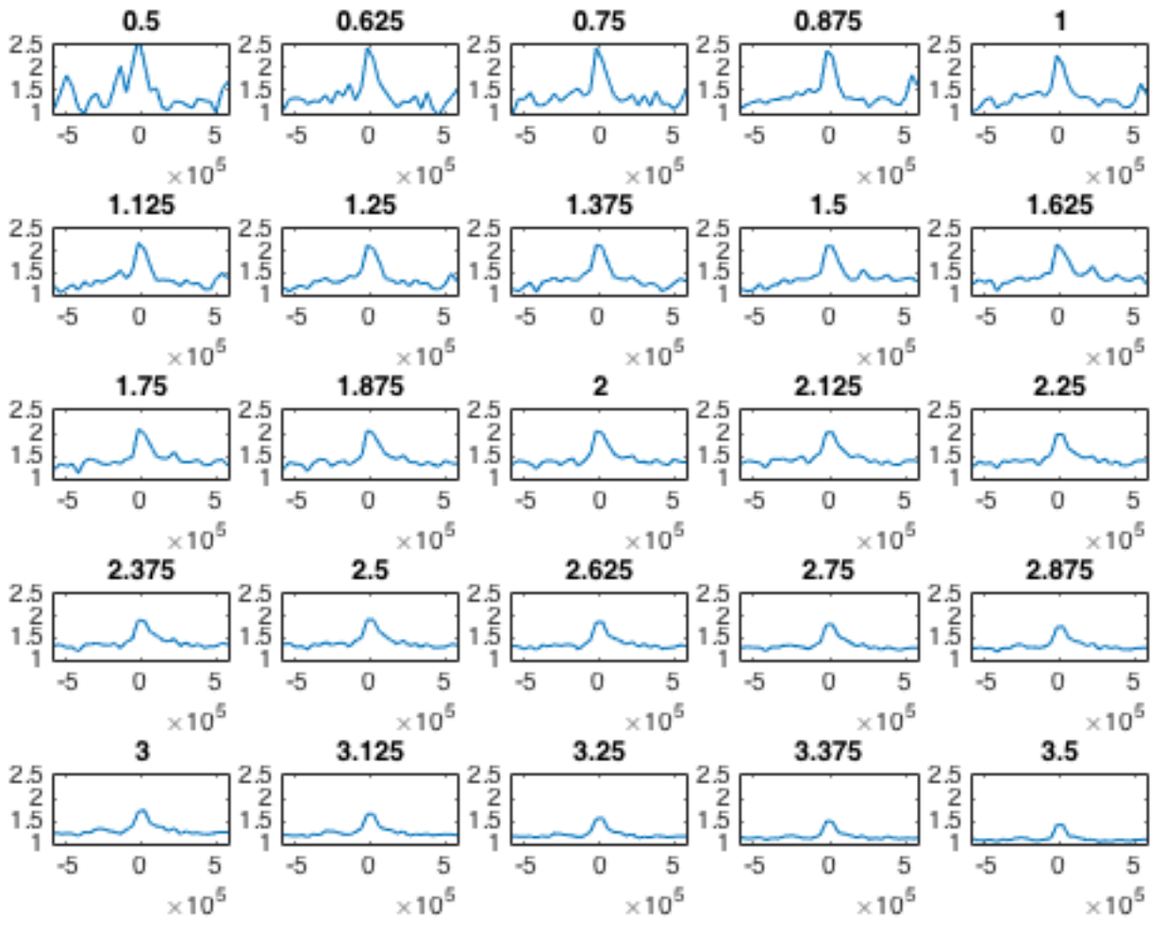
H3K20me1

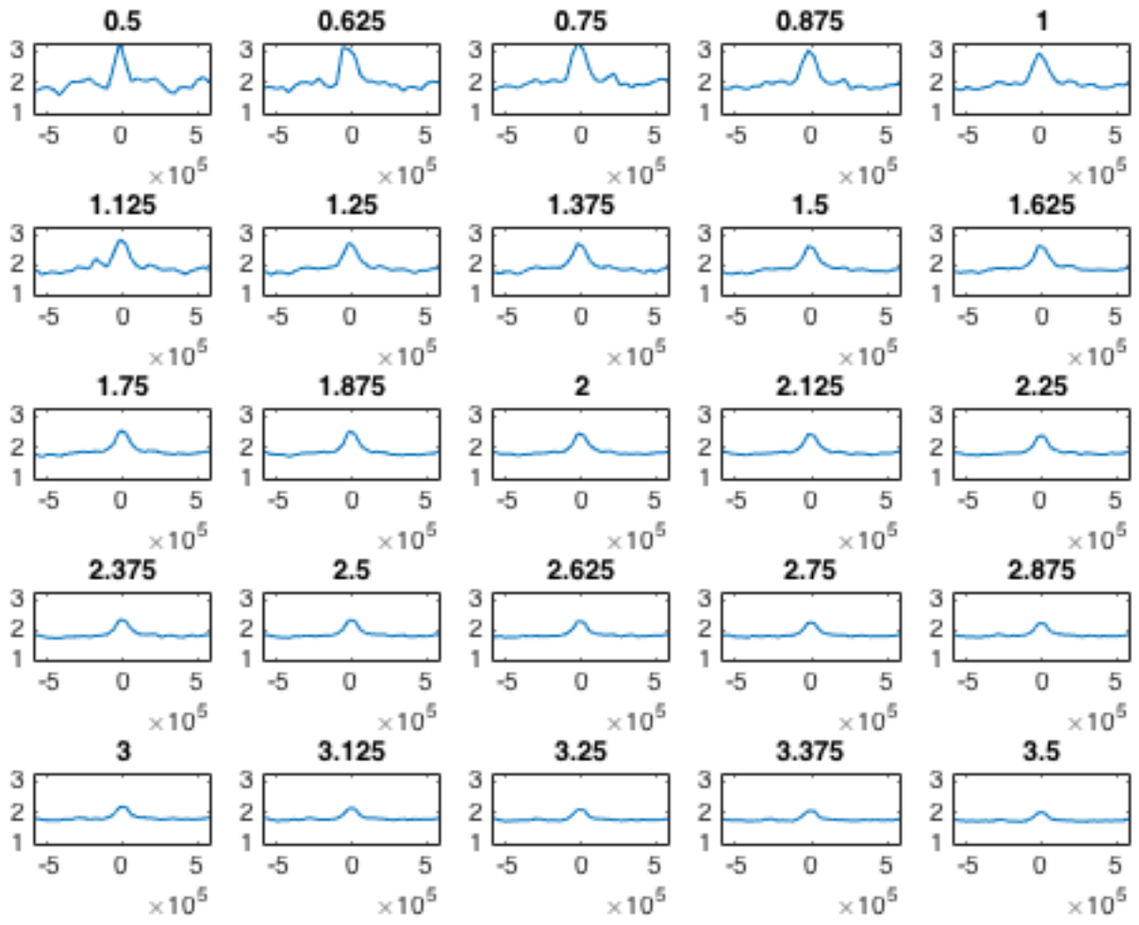




Chromatin features near TAD boundaries at various resolution

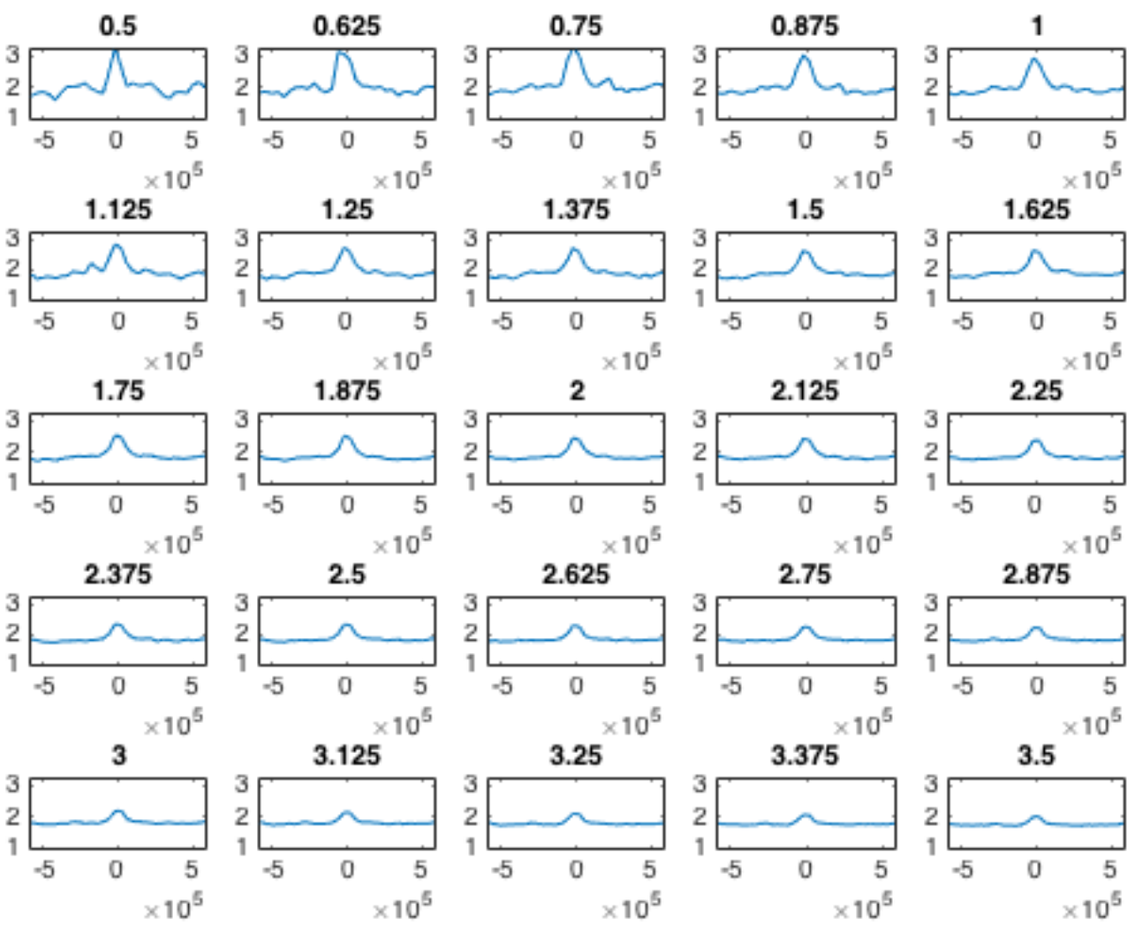
H3K4me3

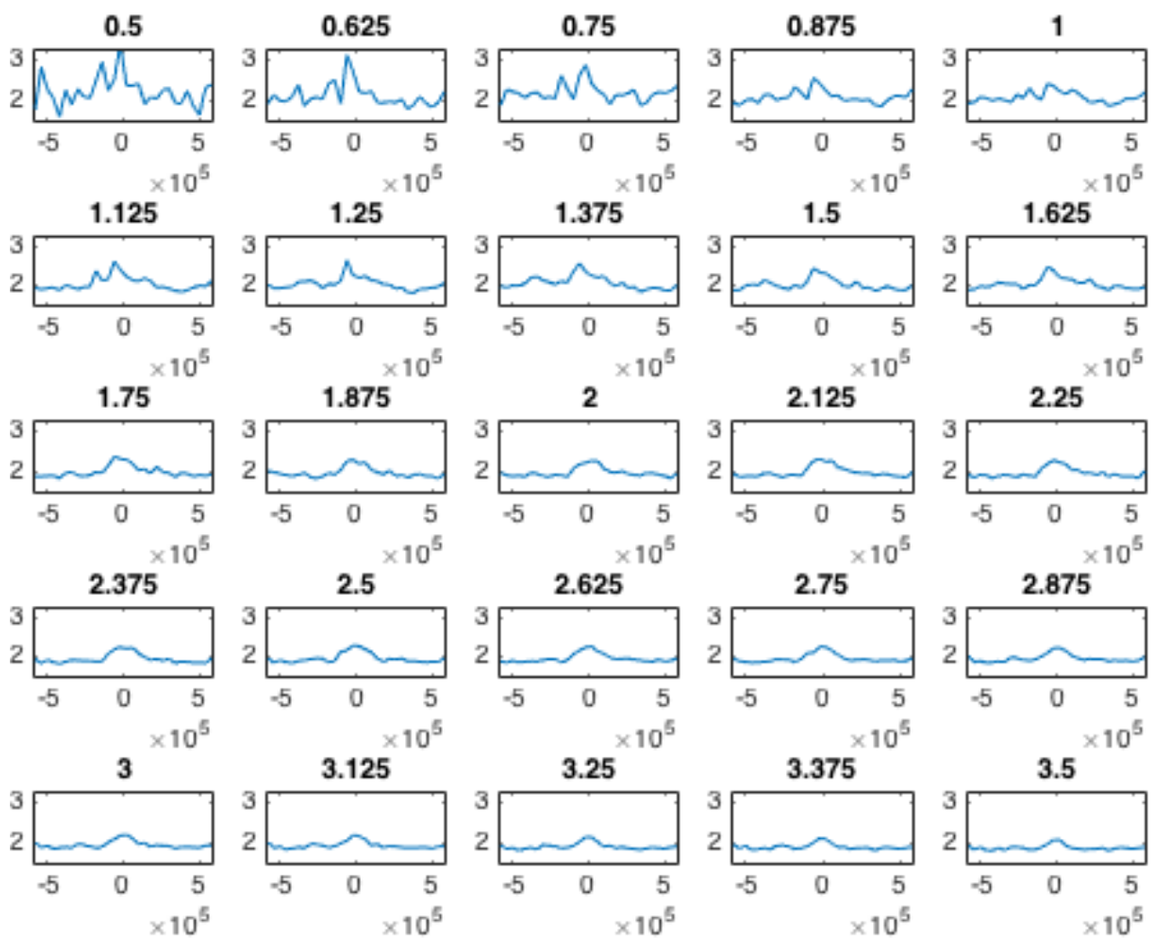




Chromatin features near TAD boundaries at various resolution

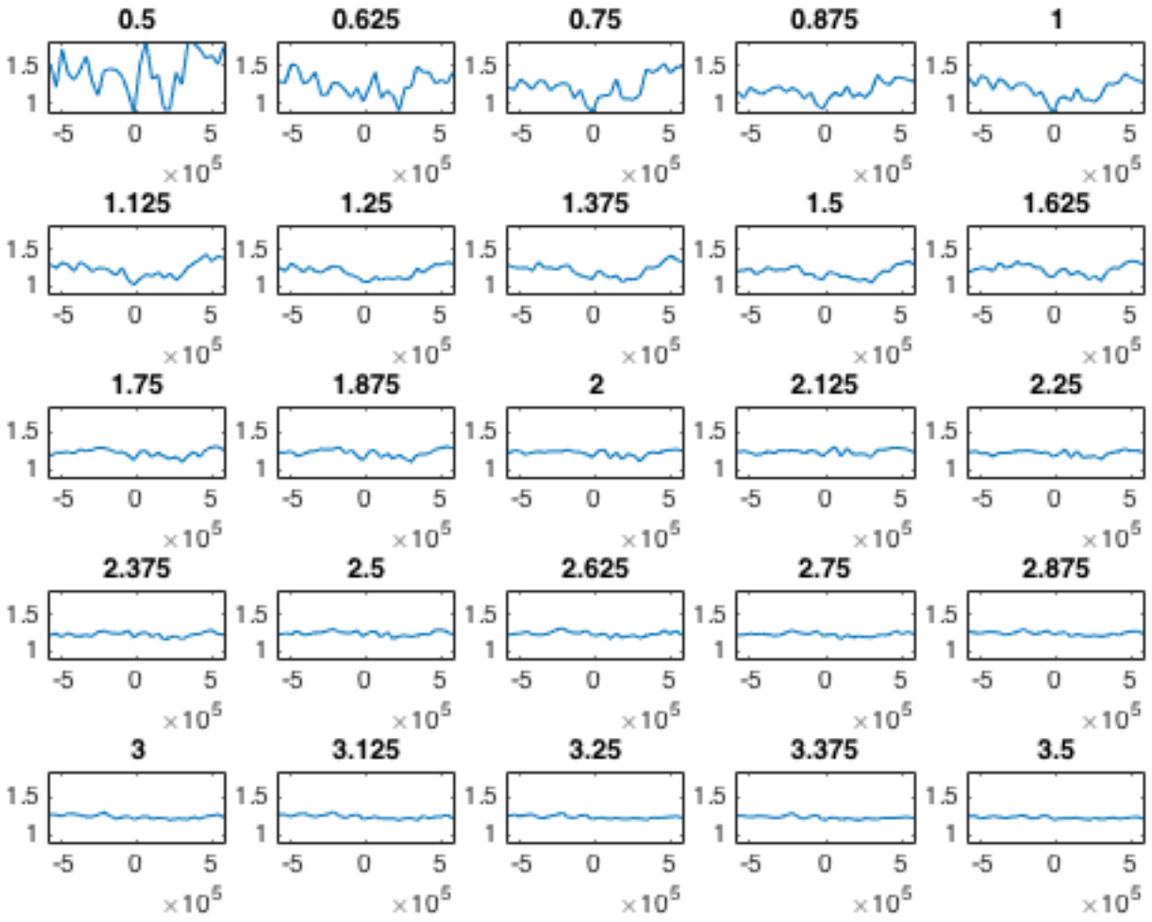
H3K79me2

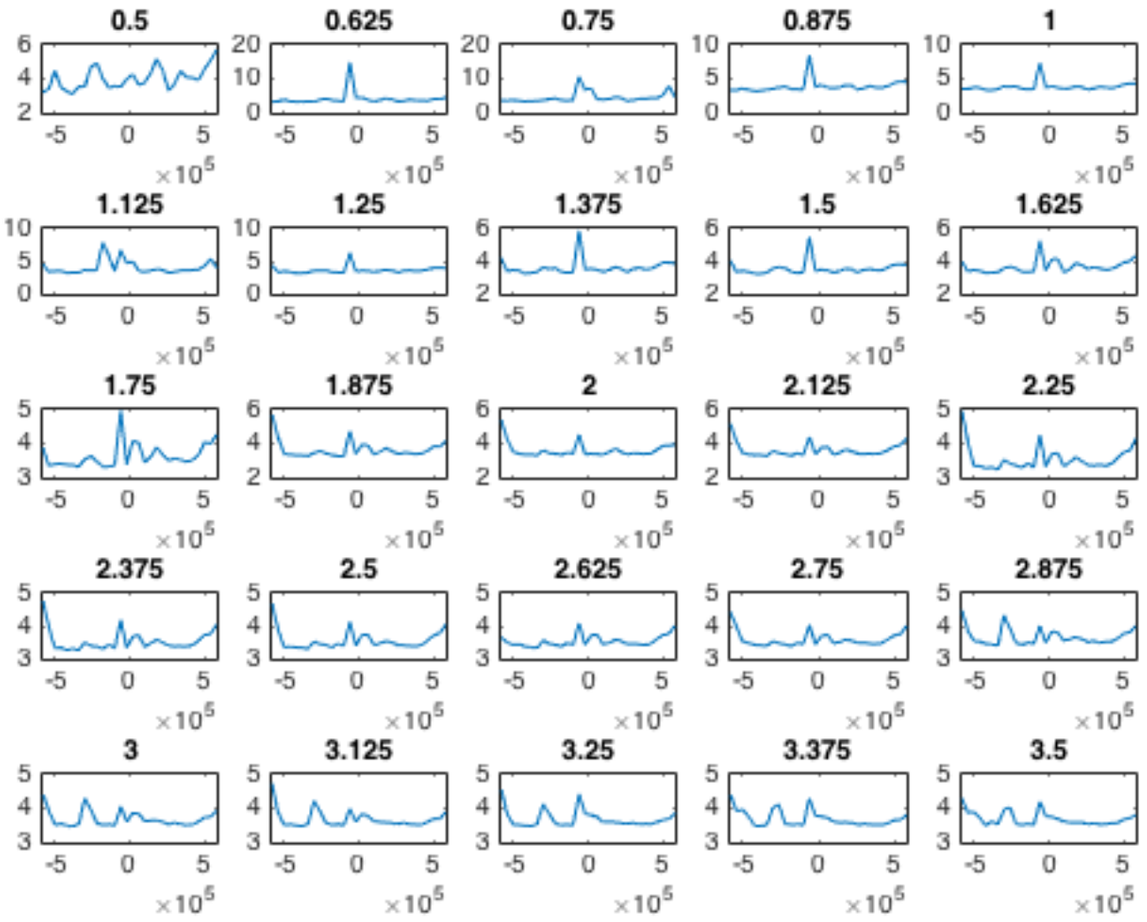




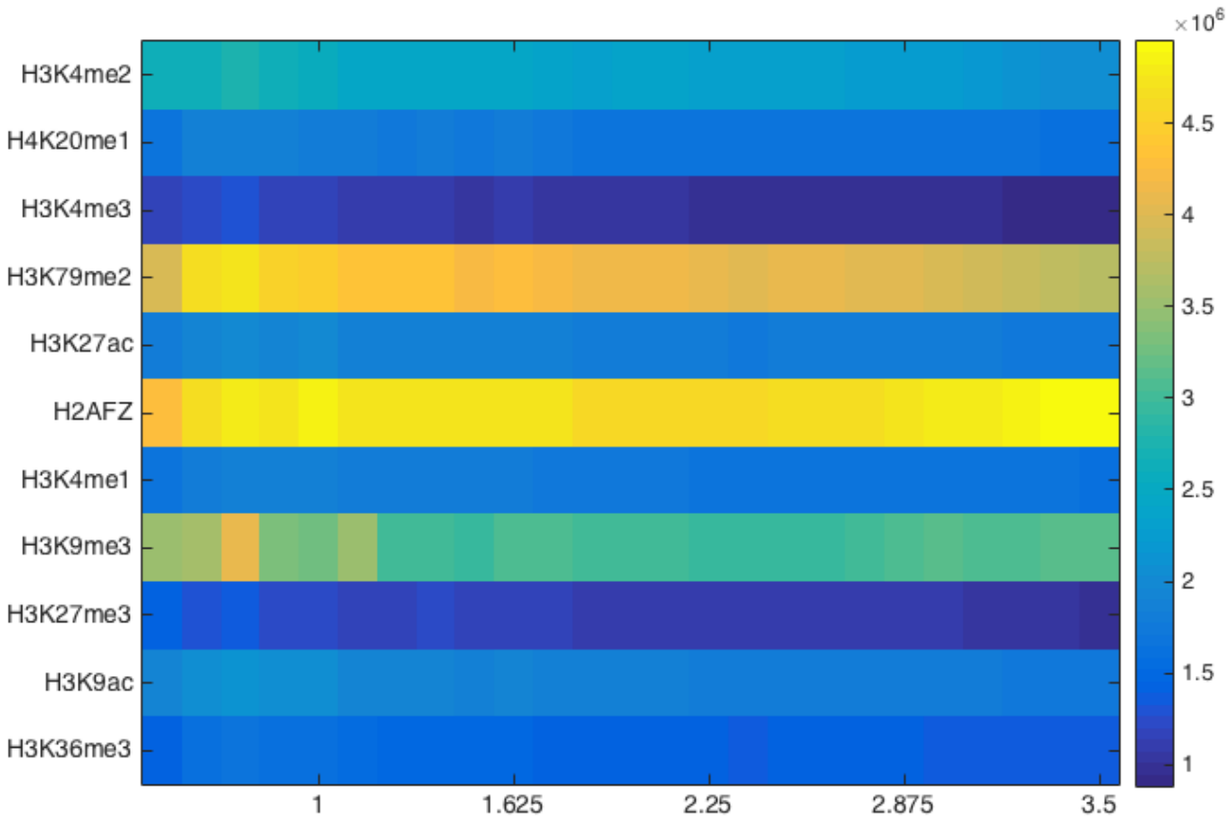
Chromatin features near TAD boundaries at various resolution

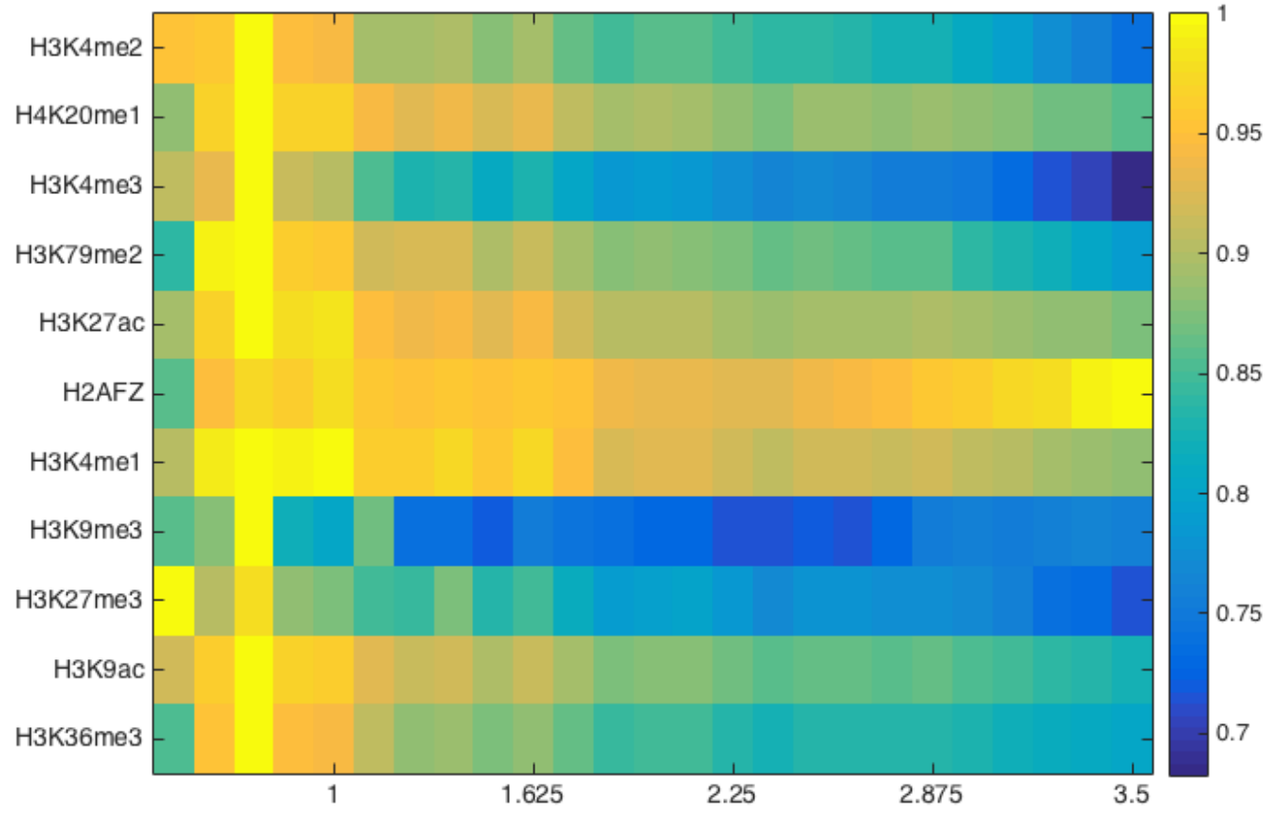
H3K9me3



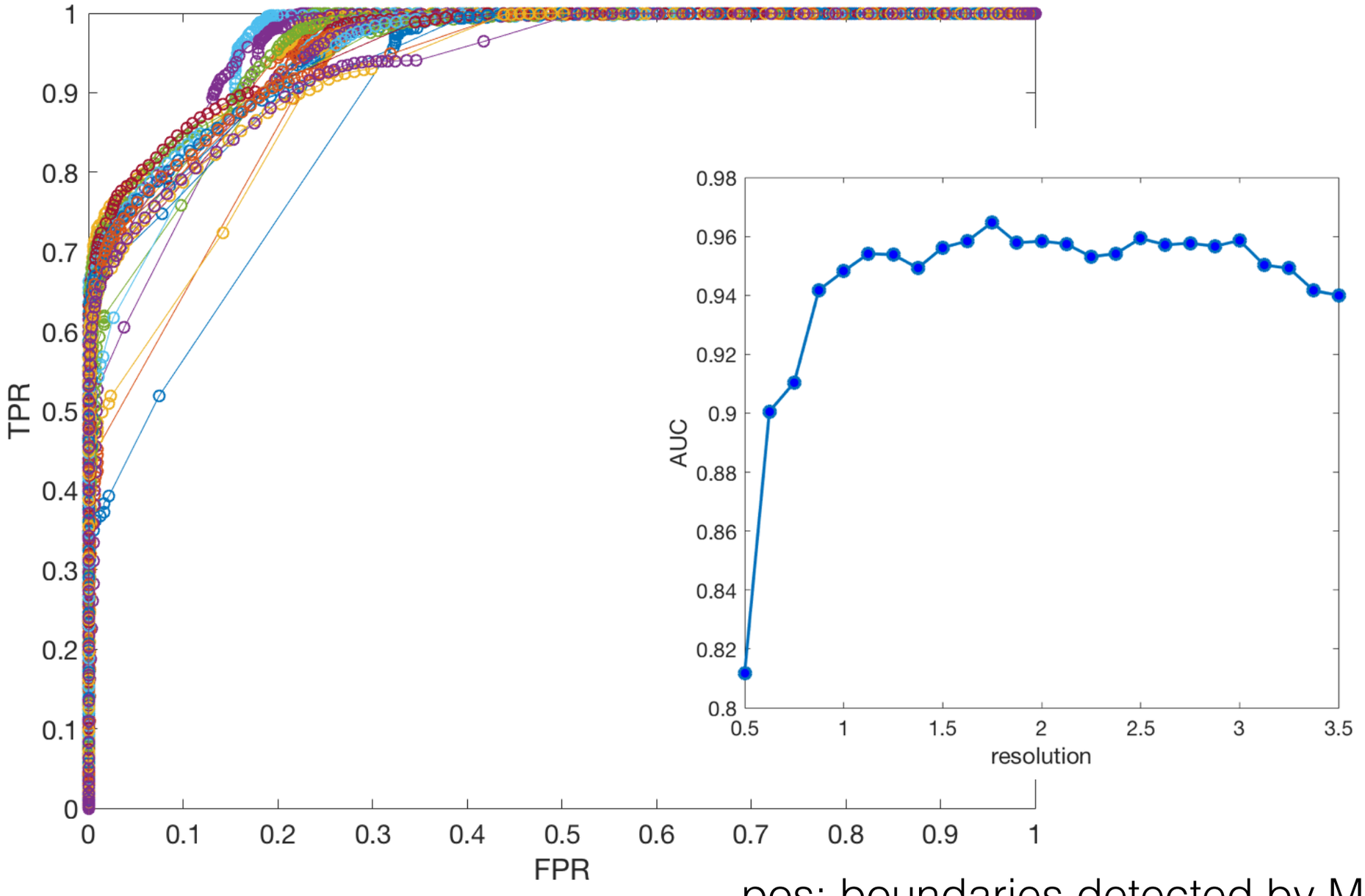


Chromatin features near TAD boundaries at various resolution





Using chromatin features to predict boundaries at various resolutions



pos: boundaries detected by MrTADFinder
neg: random boundaries of shuffled TADs

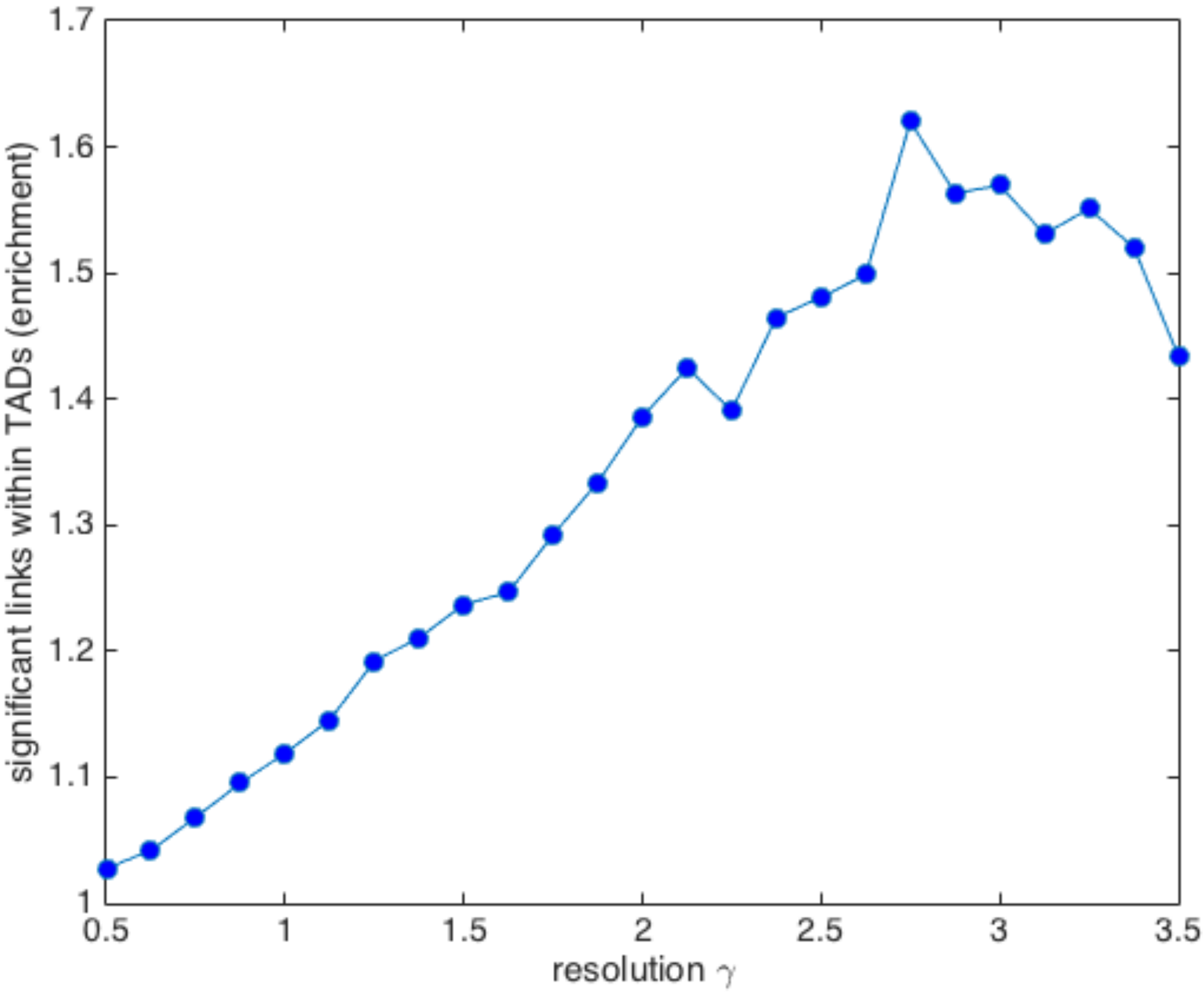
Chromatin features affect the formation of TADs in various resolution. How?

- Combining histone marks do well. It's hard to interpret the importance scores reported by random forest.

- Same issue for TFs. There are a few well known components of cohesin. Anything else? the importance of each TF, and their combinatorial effects.

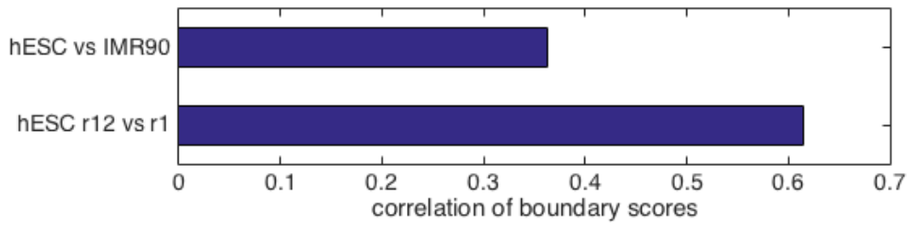
Significant contacts

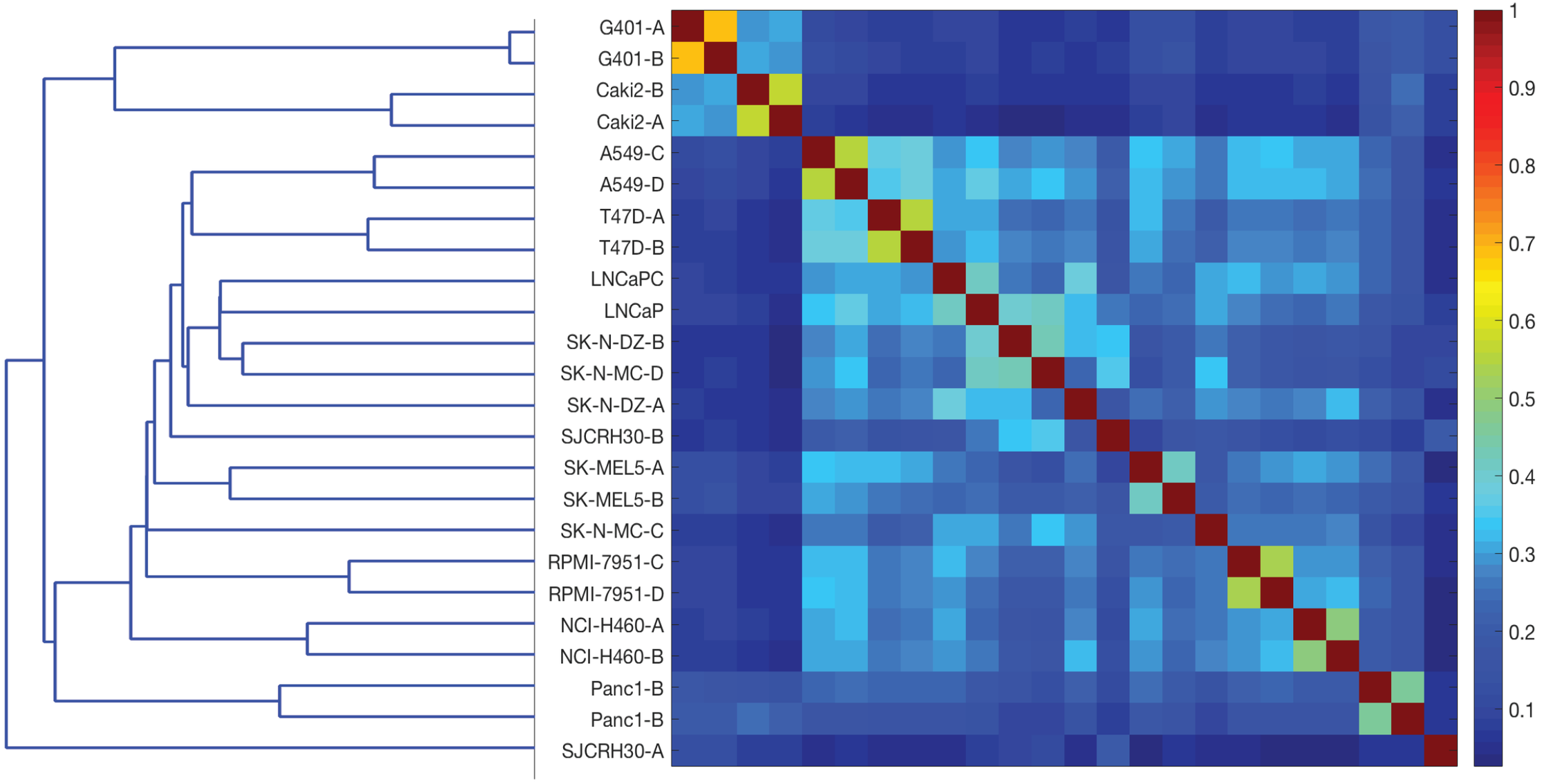
Fit-HiC



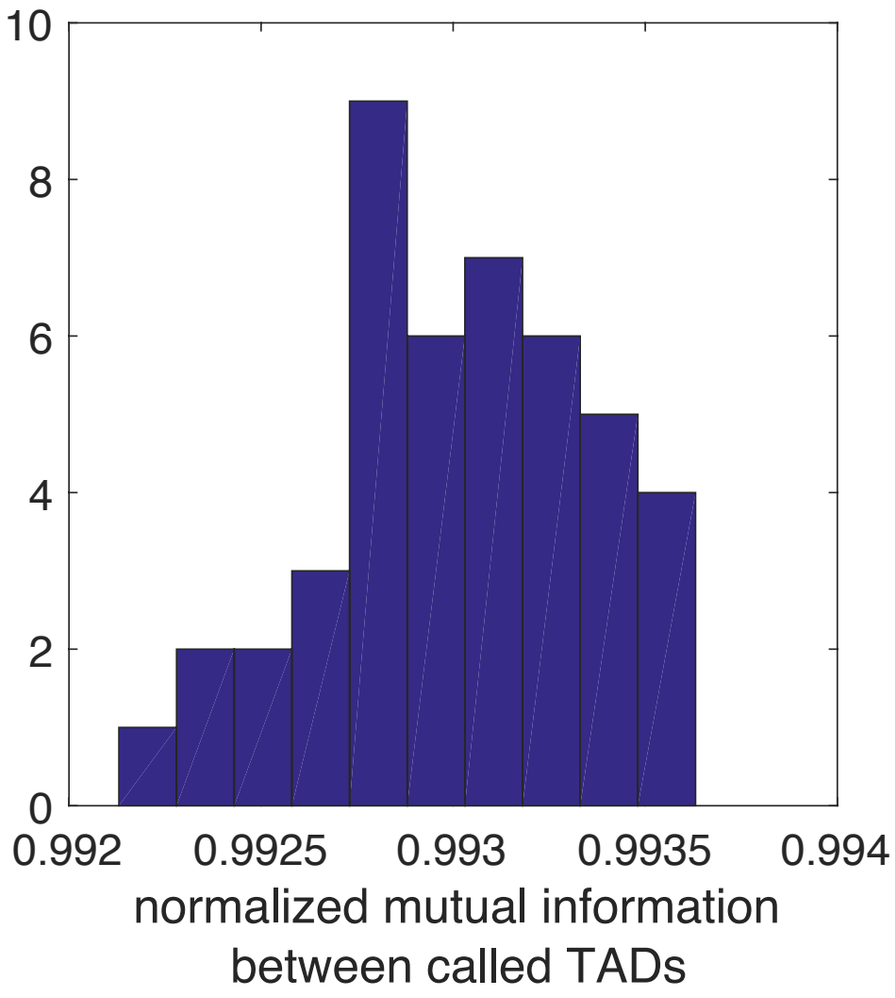
ENCODE 3 data

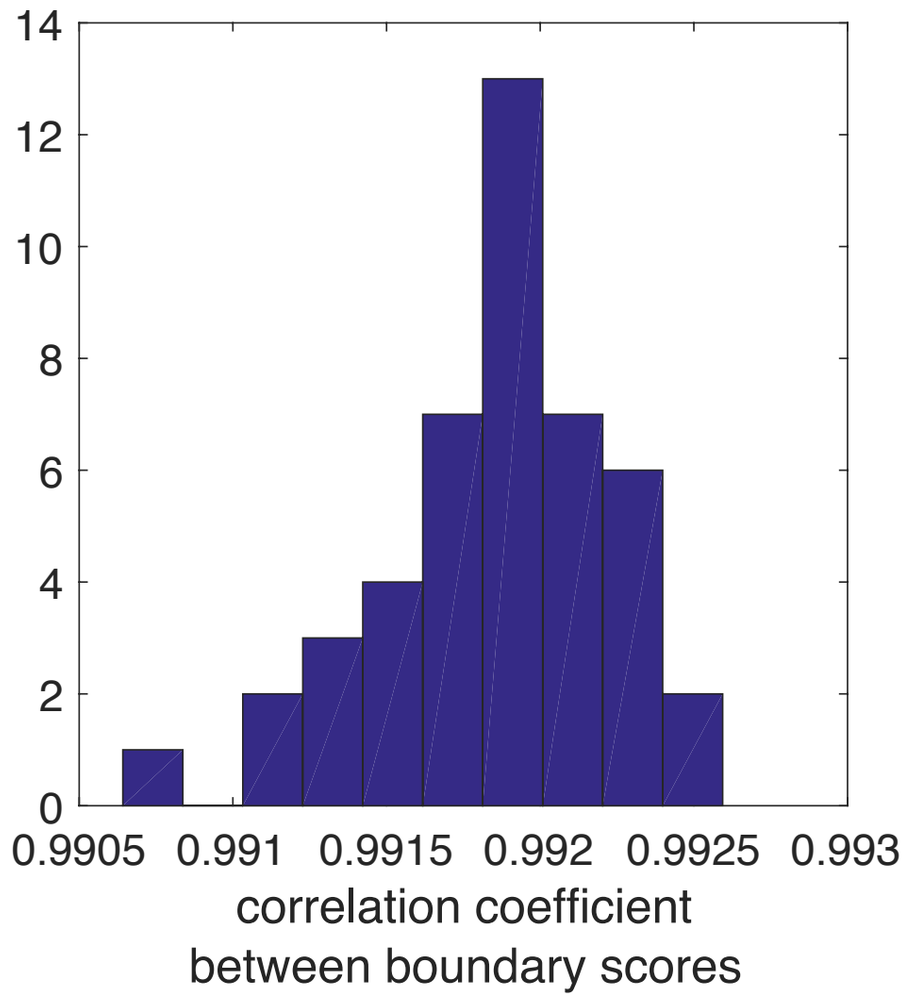
Correlation of boundary scores (resolution=1) between 24 ENCODE Hi-C experiments (include replicates).





Robustness


10
8
6
4
2
0
0.992
0.9925
0.993
0.9935
0.994
normalized mutual information
between called TADs


14
12
10
8
6
4
2
0
0.9905
0.991
0.9915
0.992
0.9925
0.993
correlation coefficient
between boundary scores