



# GATA2 ChIP-Seq on K562

---

Paola Maragno  
Carlo Manenti

# GATA2

«GATA2 plays an essential role in regulating transcription of genes involved in the development and proliferation of hematopoietic and endocrine cell lineages» [GENECARDS](#)





# Quality Controls

% of mapped reads	Replicate 1	Replicate 2	Input
	93.1 %	96.2 %	89.1%

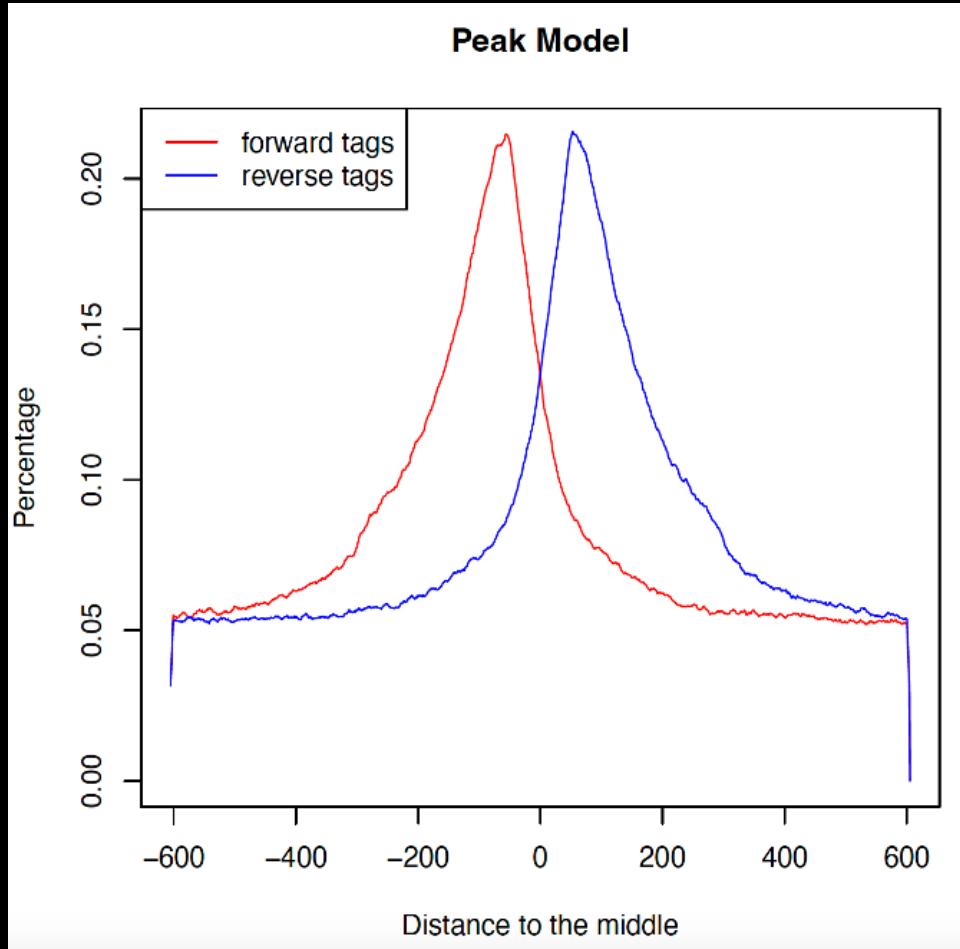
Redundant Rate	Replicate 1	Replicate 2	Pooled Replicates
	12 %	11 %	13 %

Fragment Size	Replicate 1	Replicate 2	Pooled Replicates
	149 bps	164 bps	159 bps

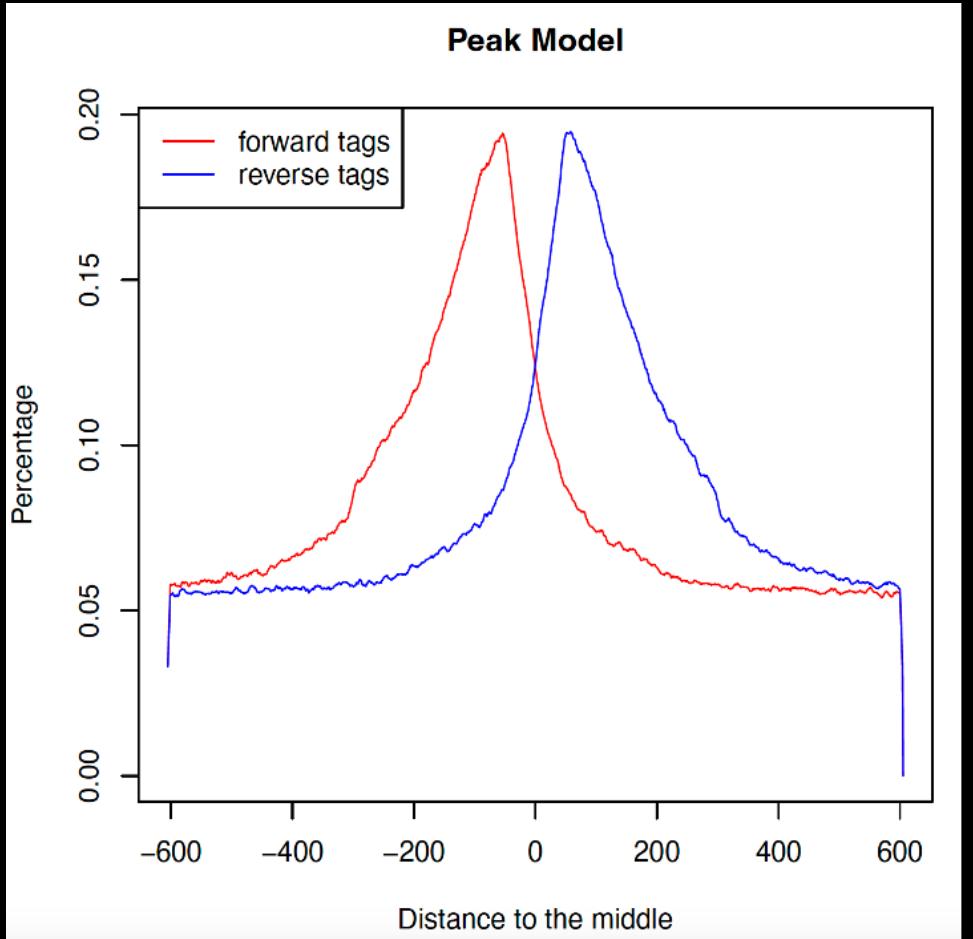


# Fragment Size Estimation Plot

Replicate 1



Replicate 2





# ENCODE Blacklisted Regions

PEAKs	Replicate 1	Replicate 2	Pooled
Before Filtering	9744	7532	7404
After Filtering	9739	7529	7404

SUMMITs	Replicate 1	Replicate 2	Pooled
Before Filtering	9744	7532	7404
After Filtering	9739	7529	7404



# % of peaks in common with ENCODE

	Replicate 1	Replicate 2	Peaks in common	Pooled peaks
Any kind of overlap	82.5 %	87.3 %	90.8 %	89.6 %
Summits within 100bp from one another	80.9 %	85.8 %	90.4 %	88.7 %

# Distribution of q-value for Any kind of overlap

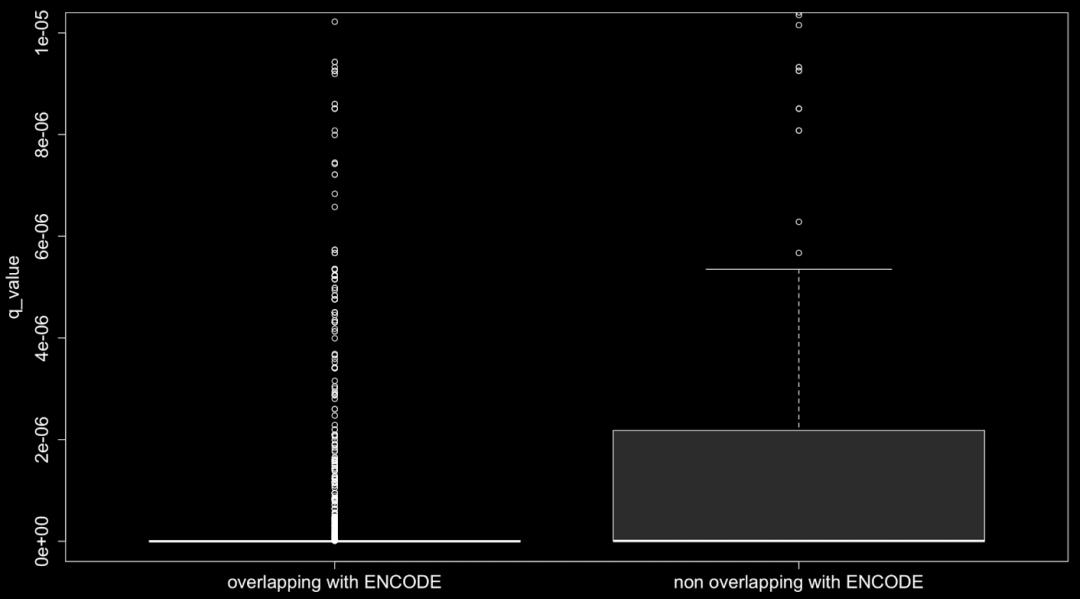
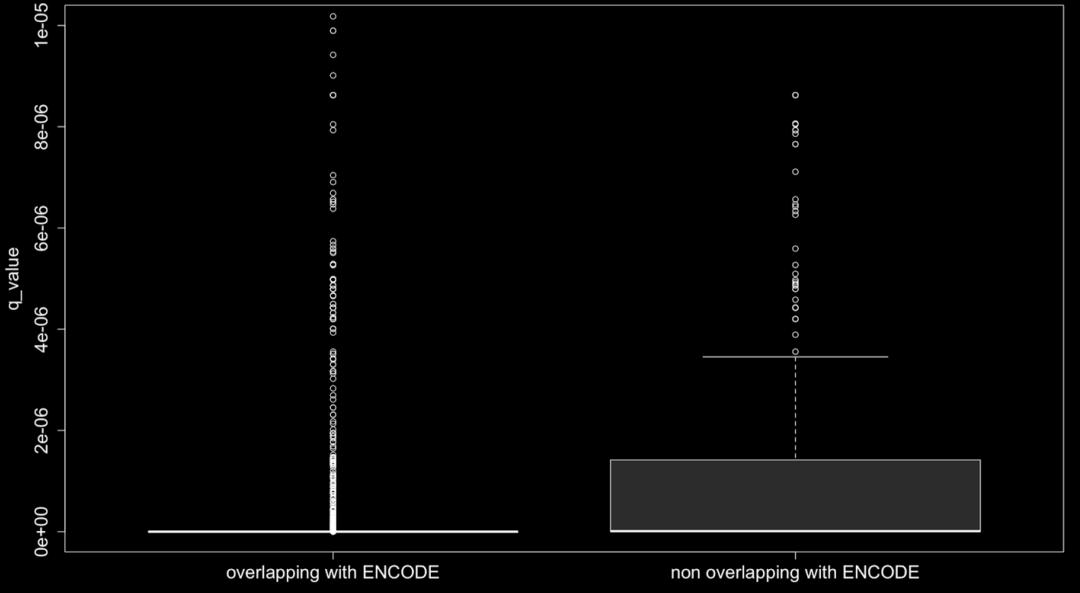
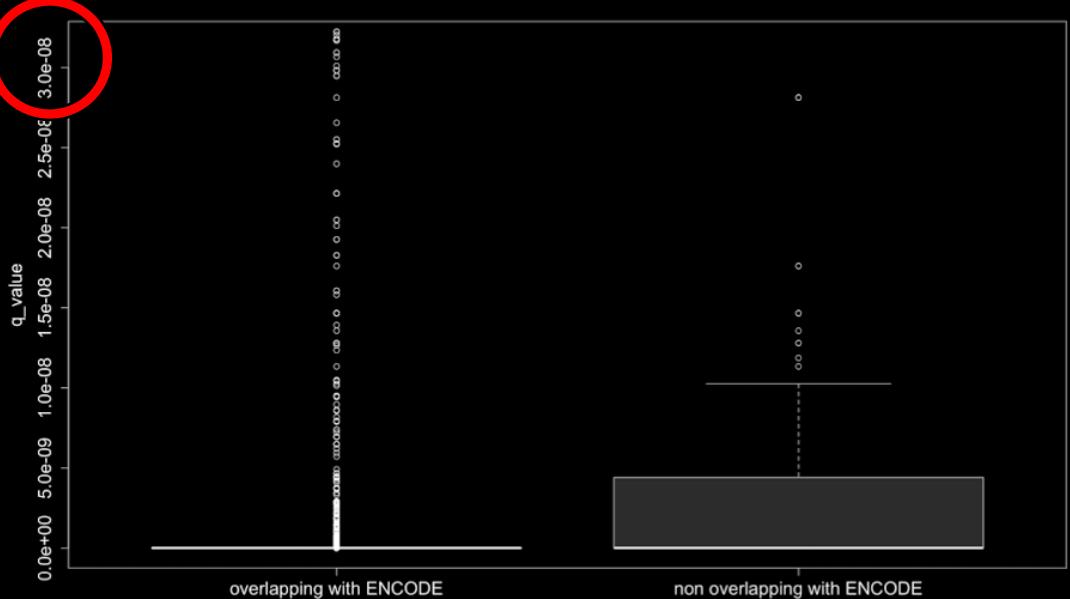
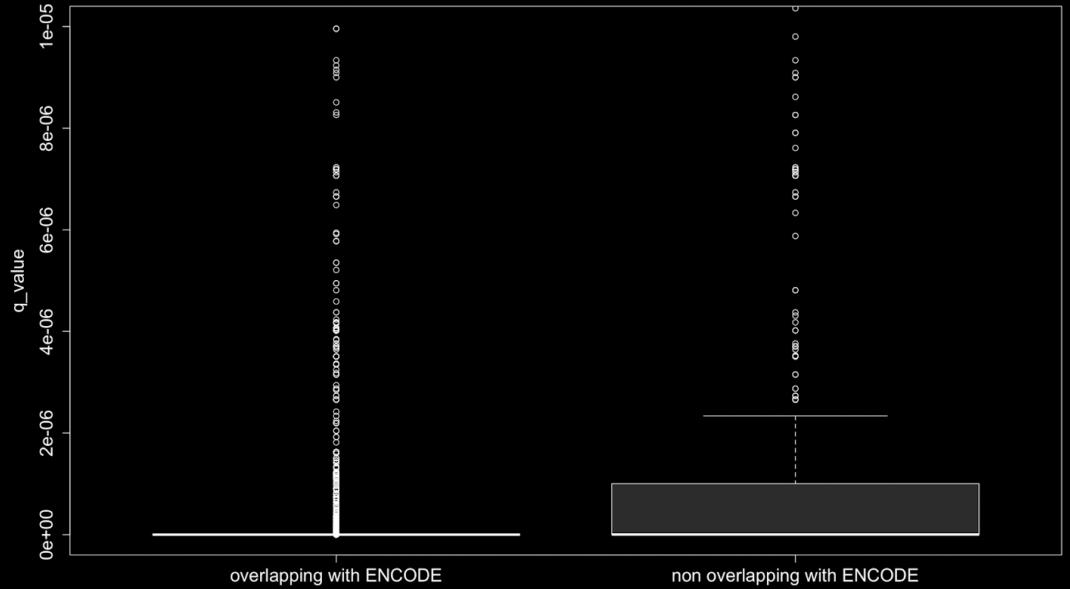
Replicate 2

Pooled

6

Replicate 1

Common peaks





# Distribution of q-value for summit proximity

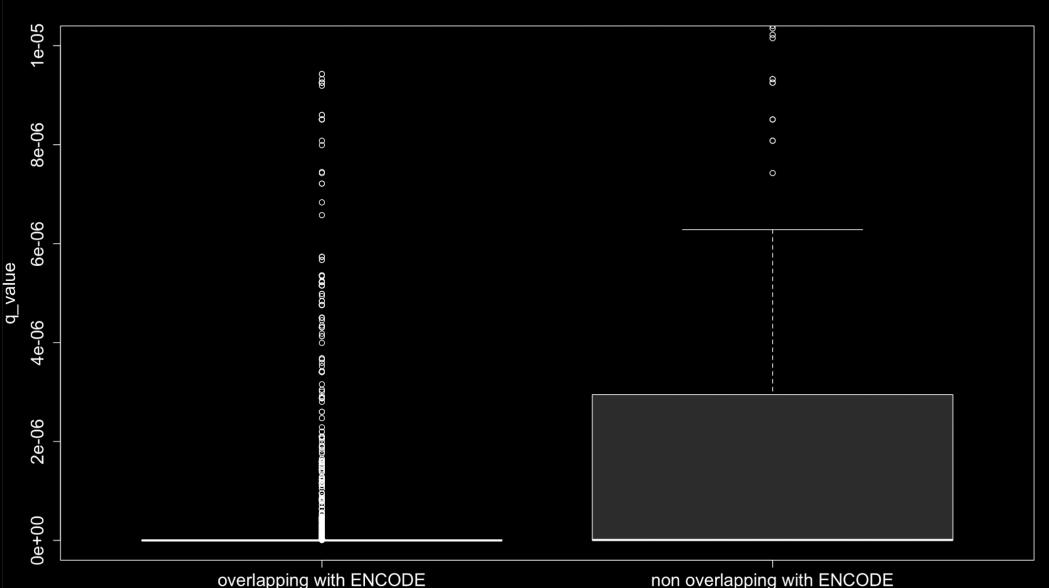
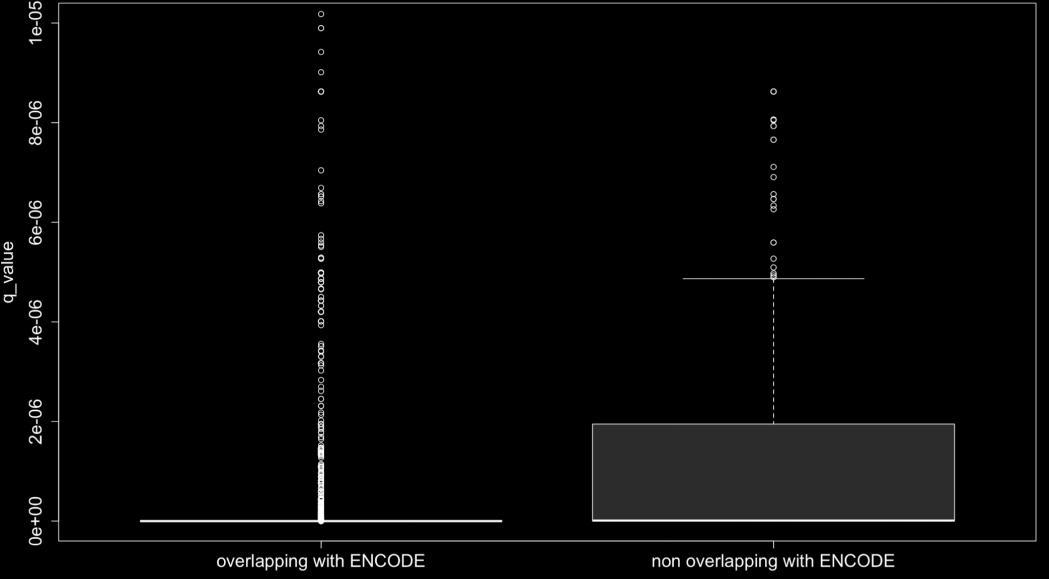
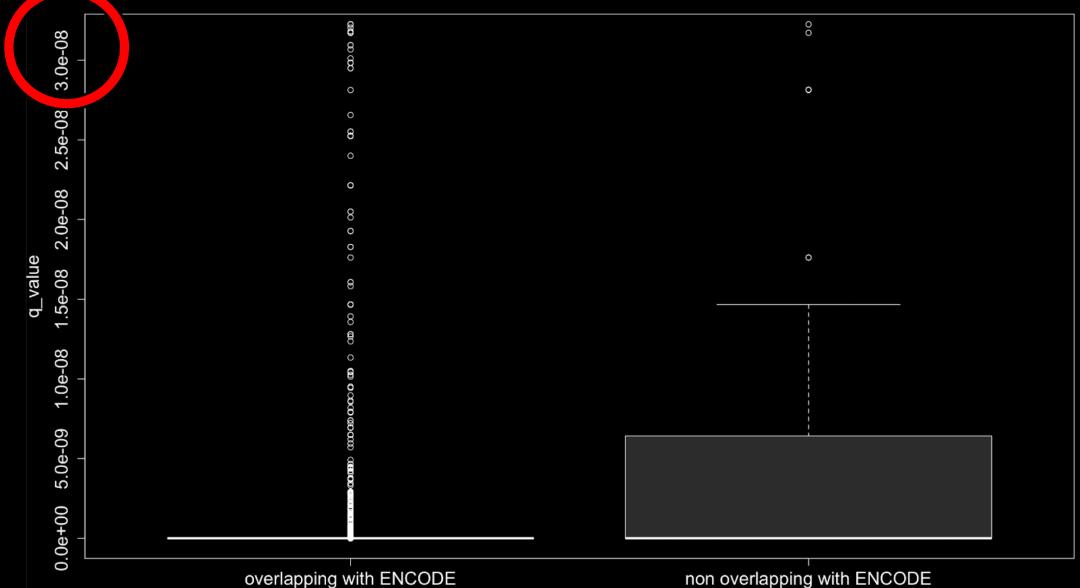
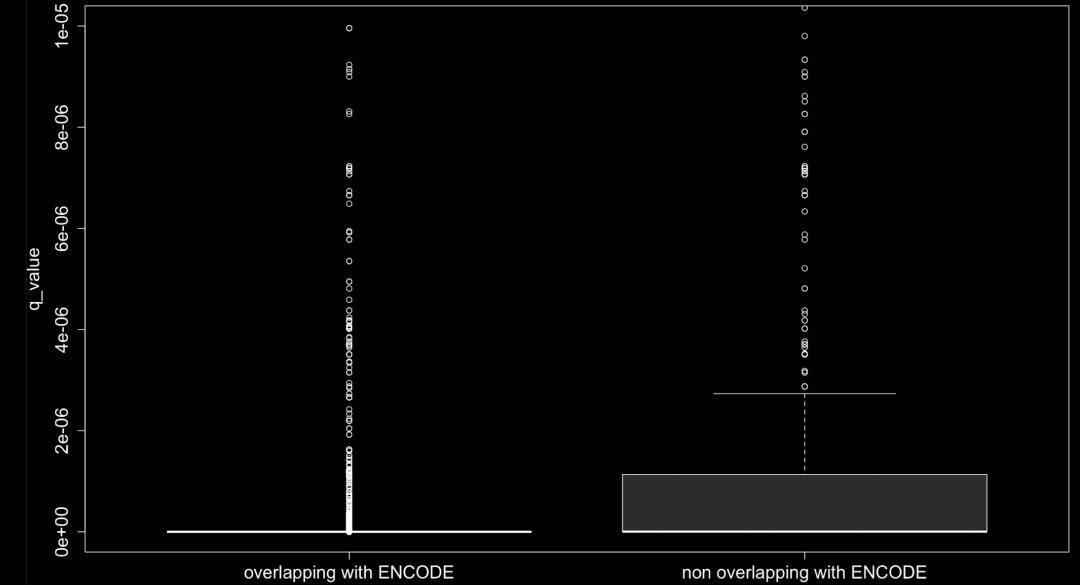
Replicate 2

Pooled

7

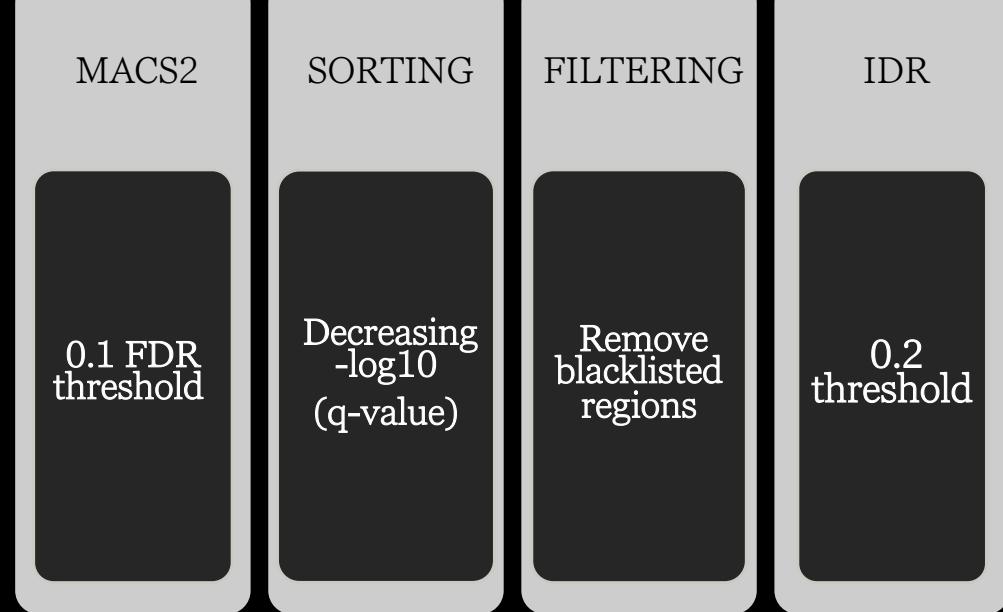
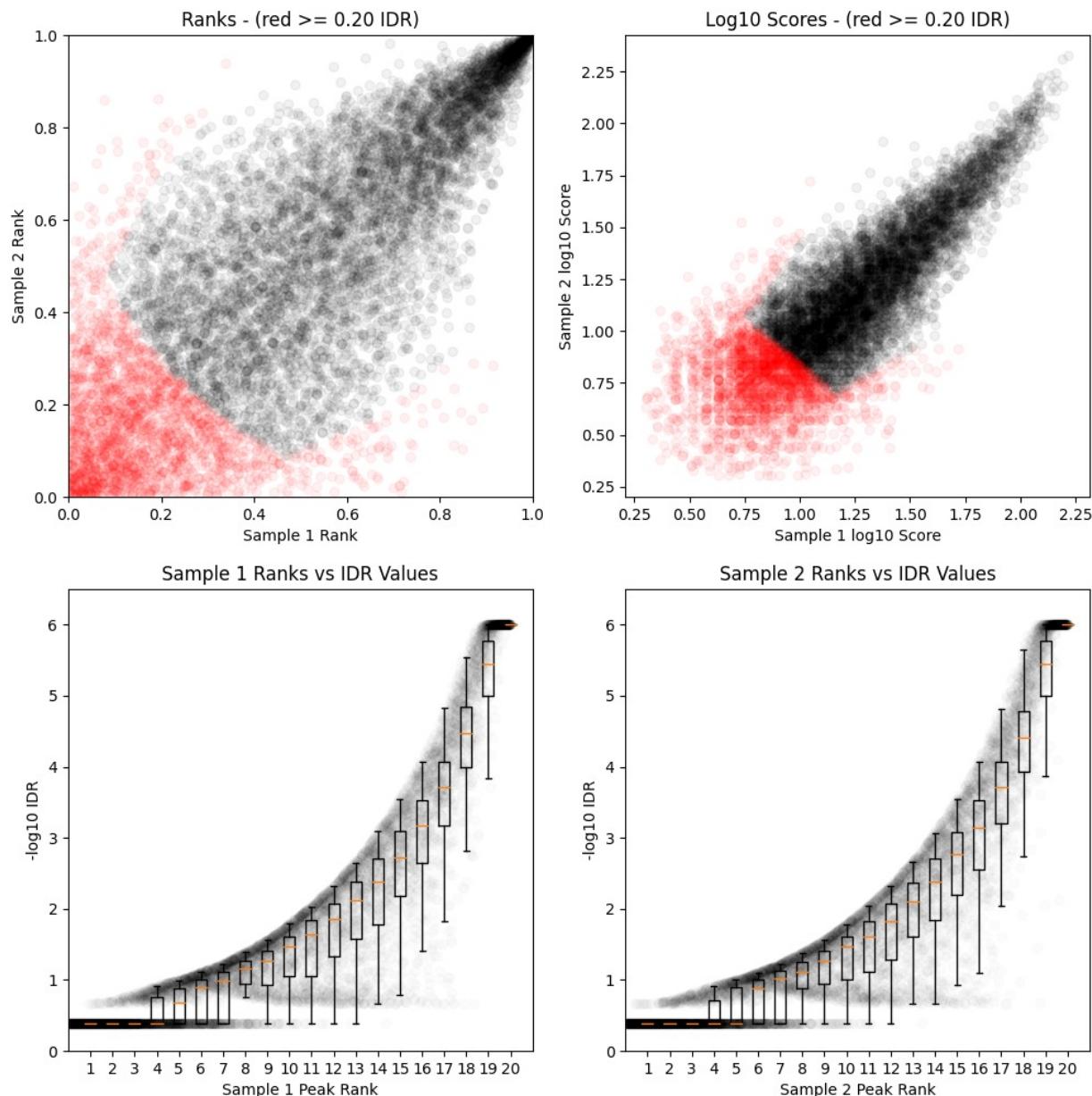
Replicate 1

Common peaks





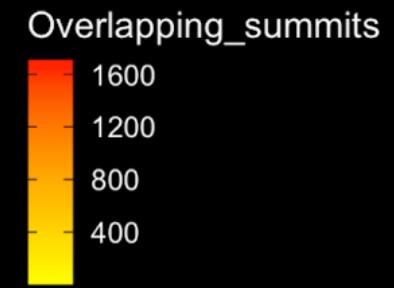
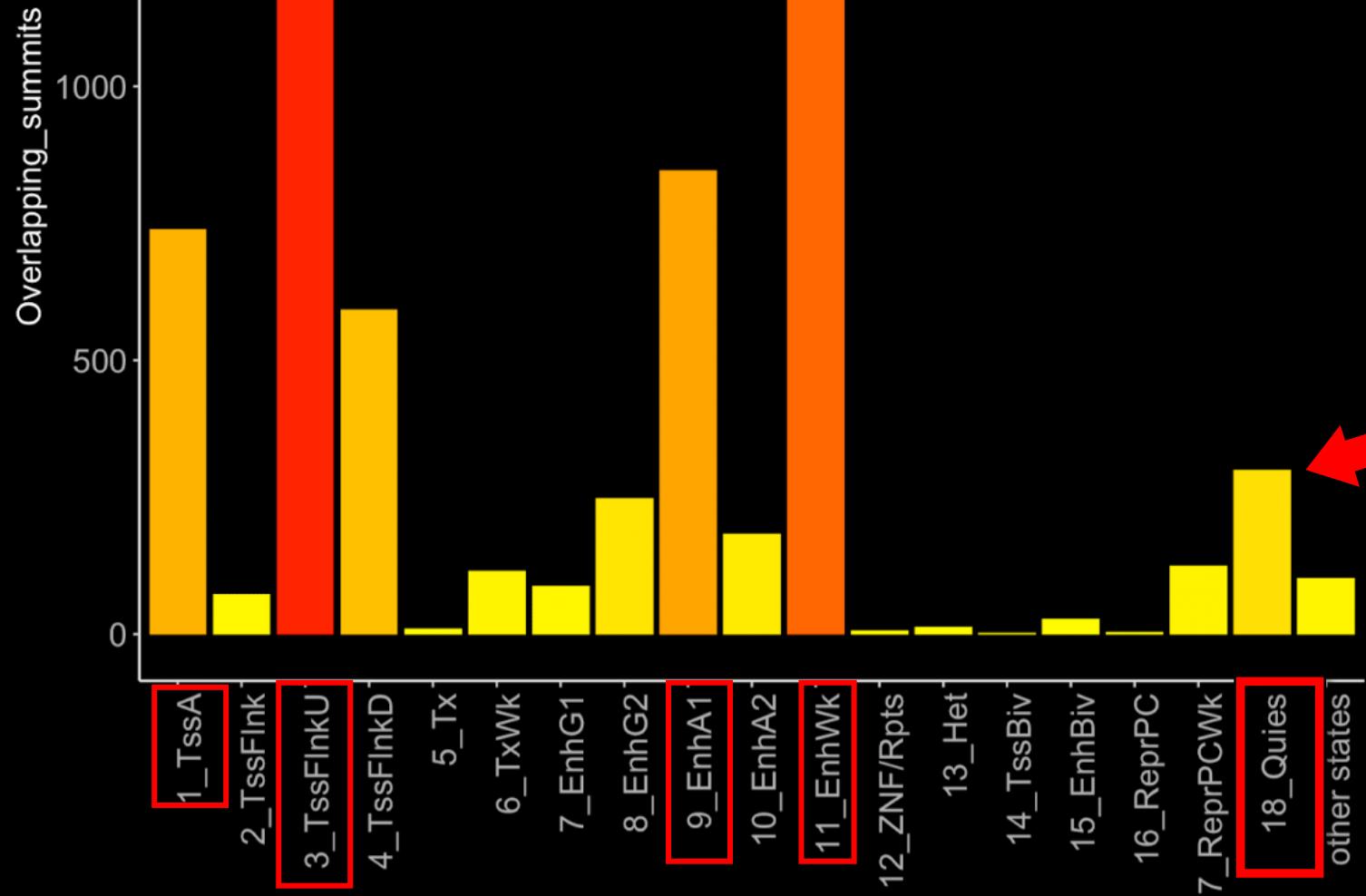
# IDR



6519 peaks  
91.2 % of overlap for any kind of overlap  
90.7 % of overlap for summit proximity



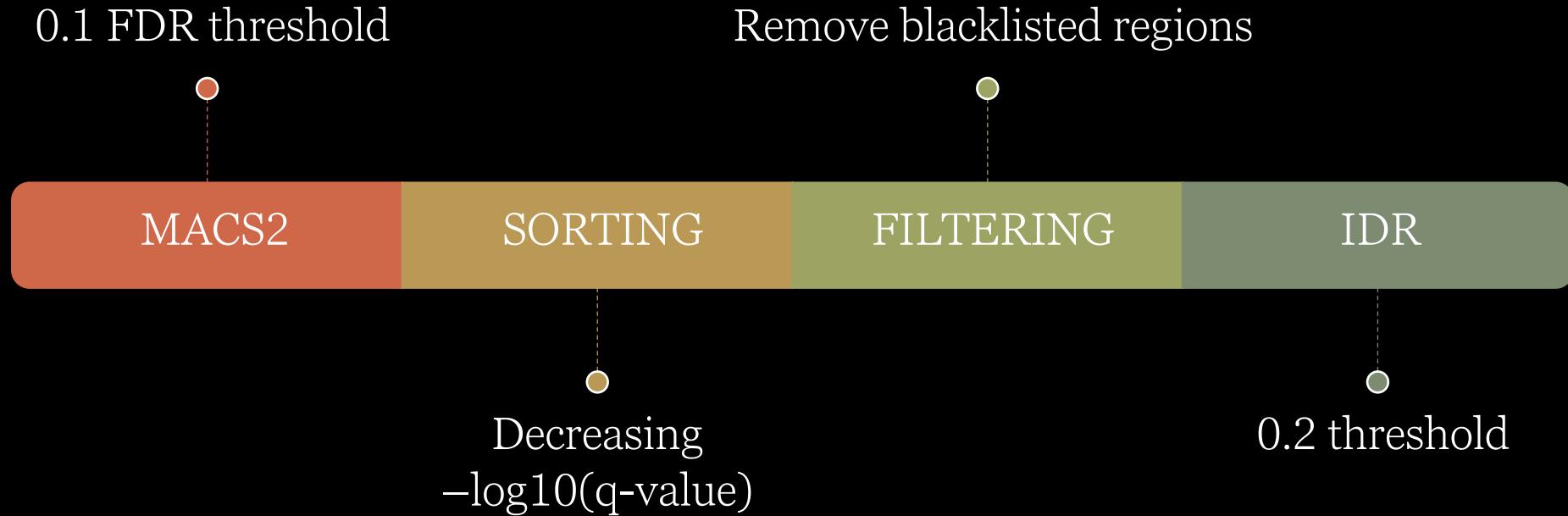
# Chromatin States



300 summits



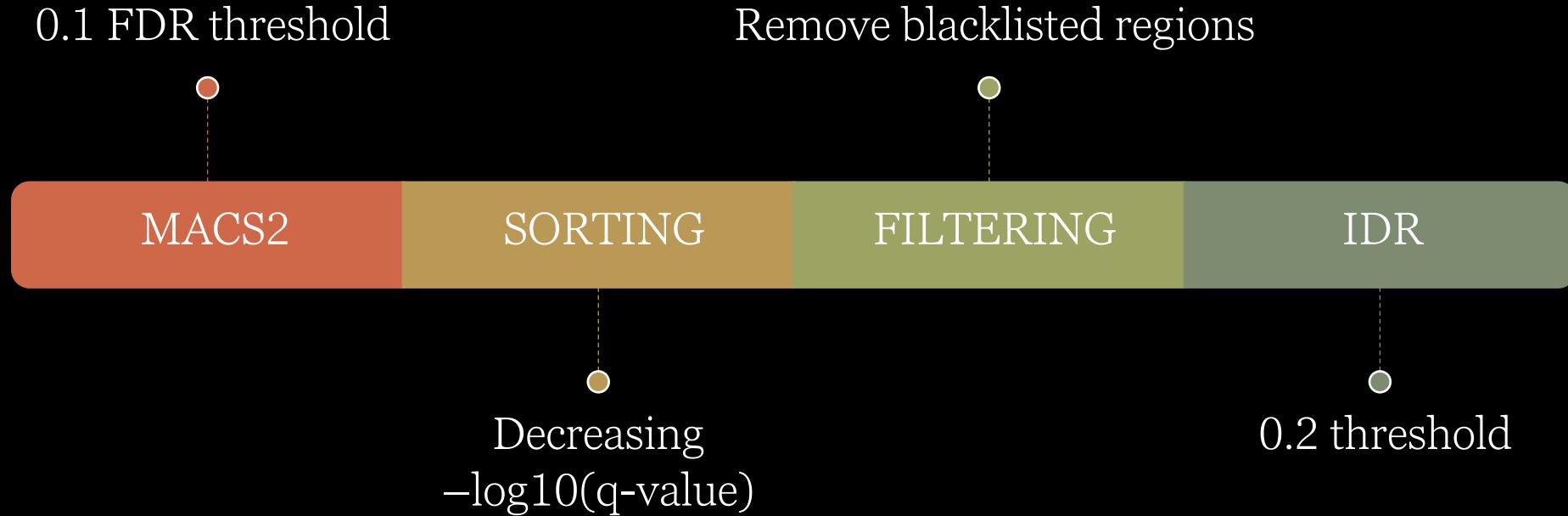
# ...starting from filtered bam files



	Replicate 1	Replicate 2
Redundant Rate	0 %	0 %
Fragment size	170 bps	186 bps

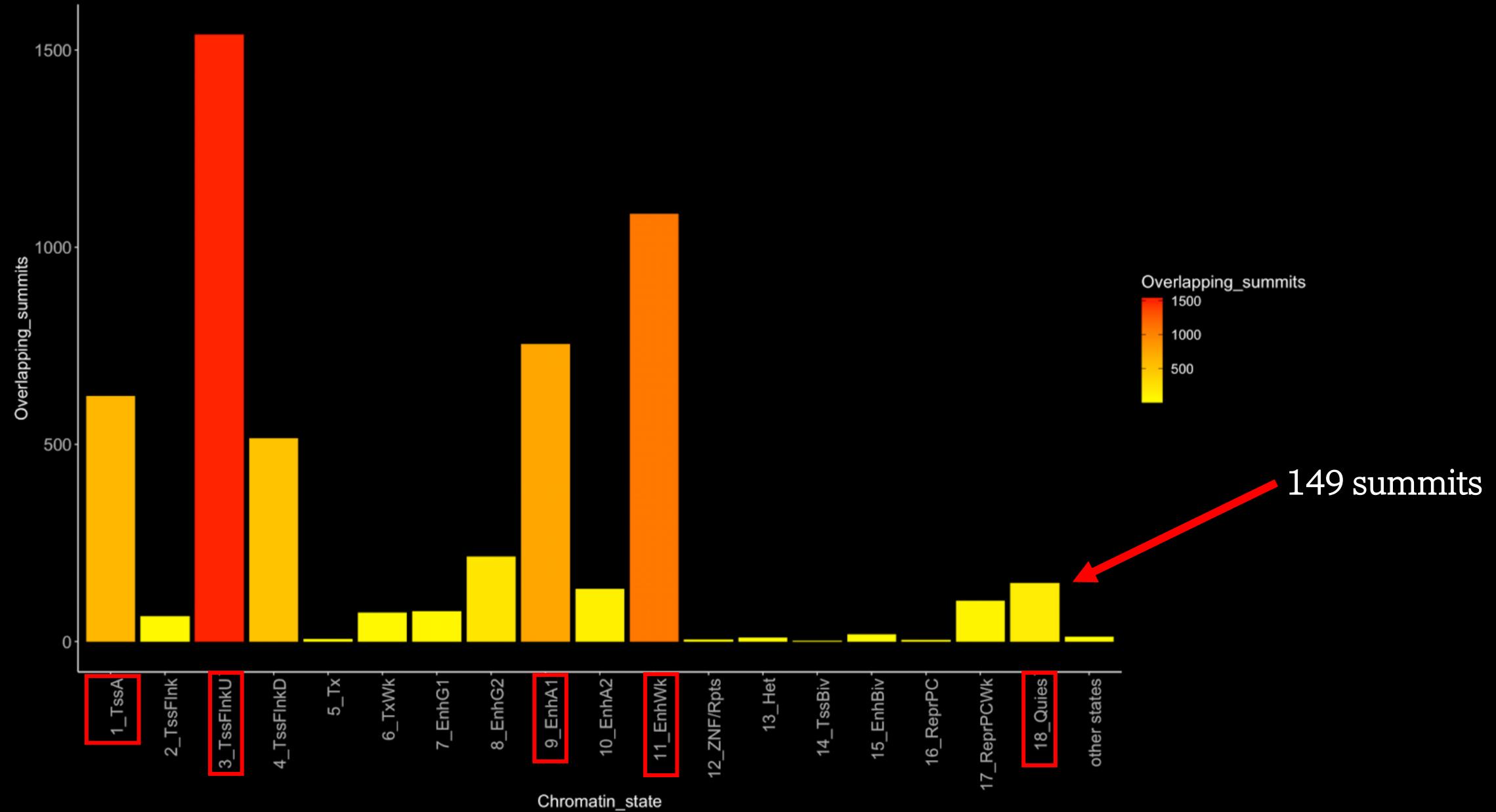


# ...starting from filtered bam files

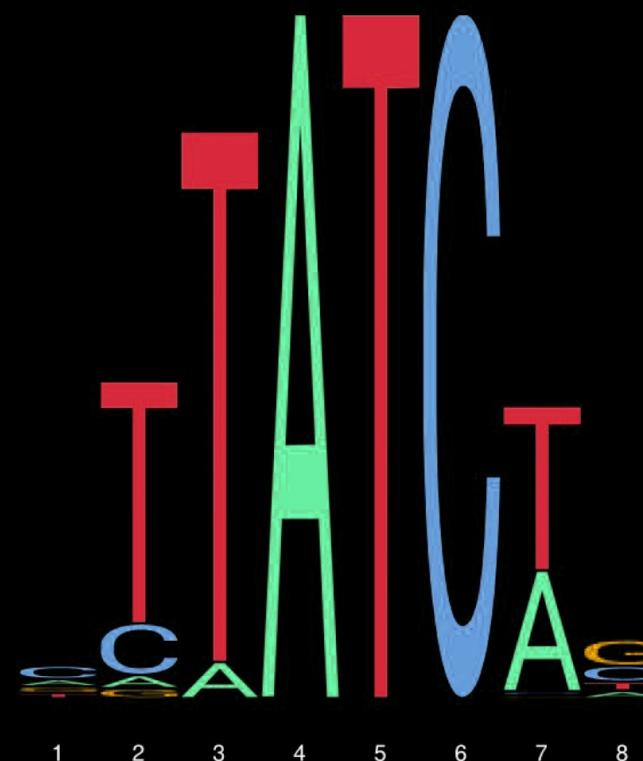




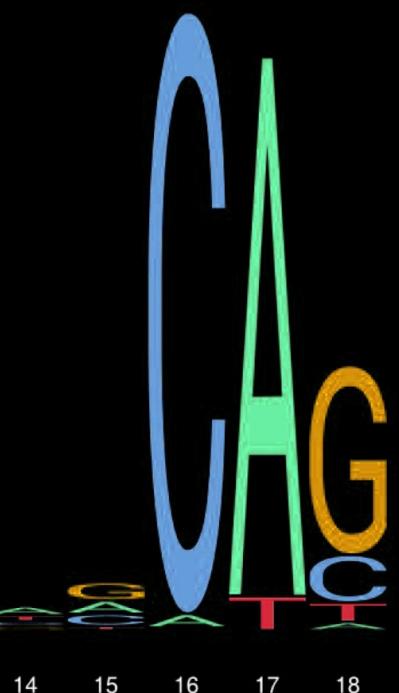
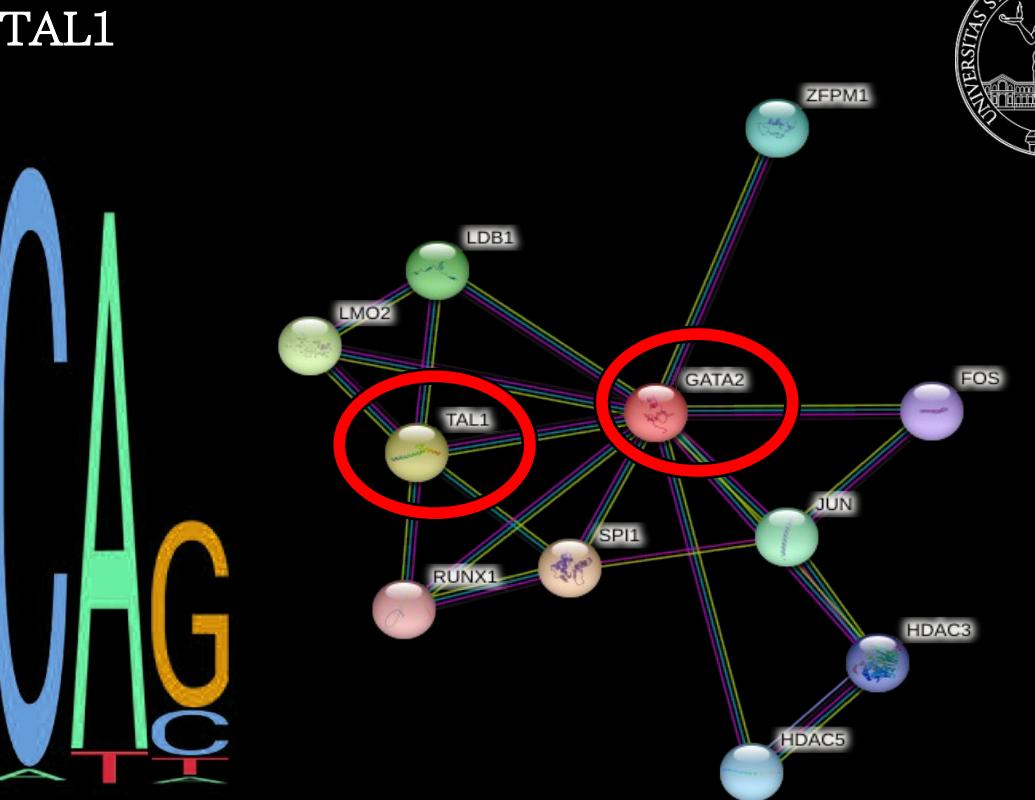
# Filtered bam files: Chromatin States



GATA1

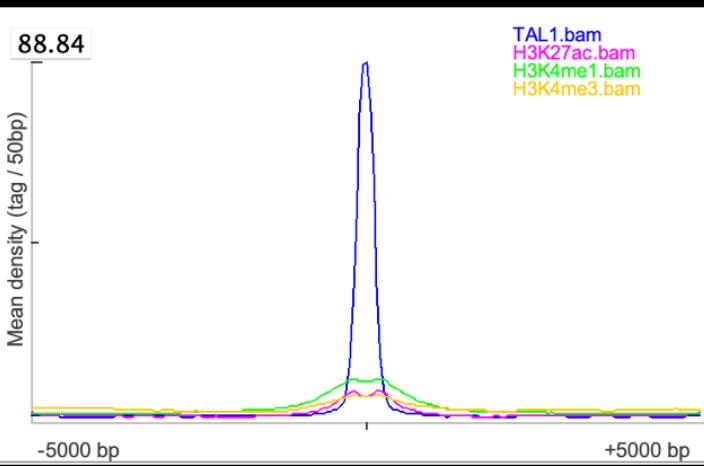
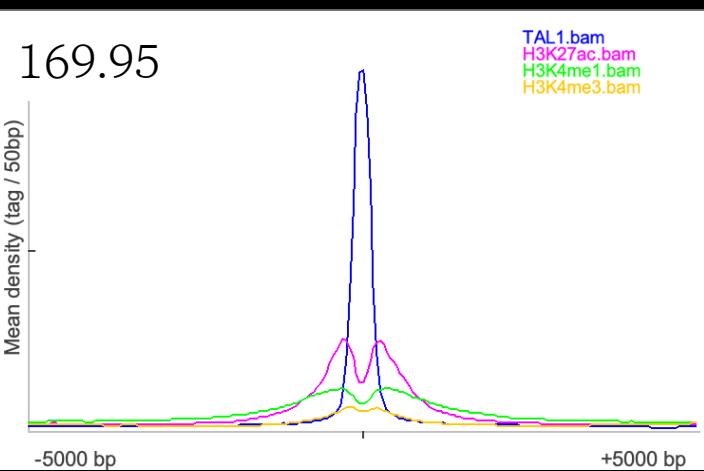
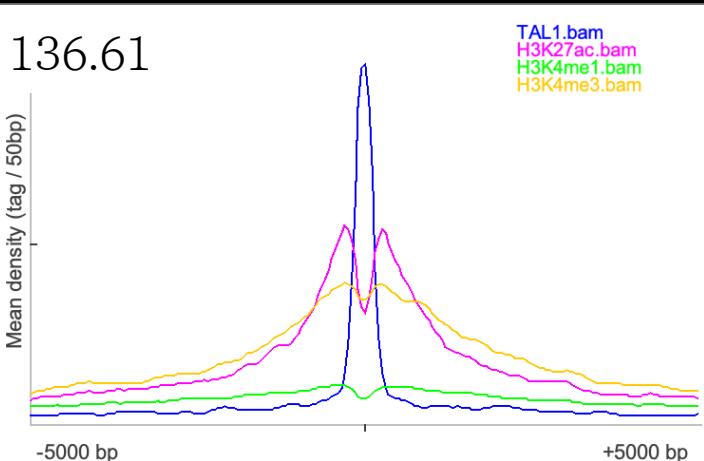
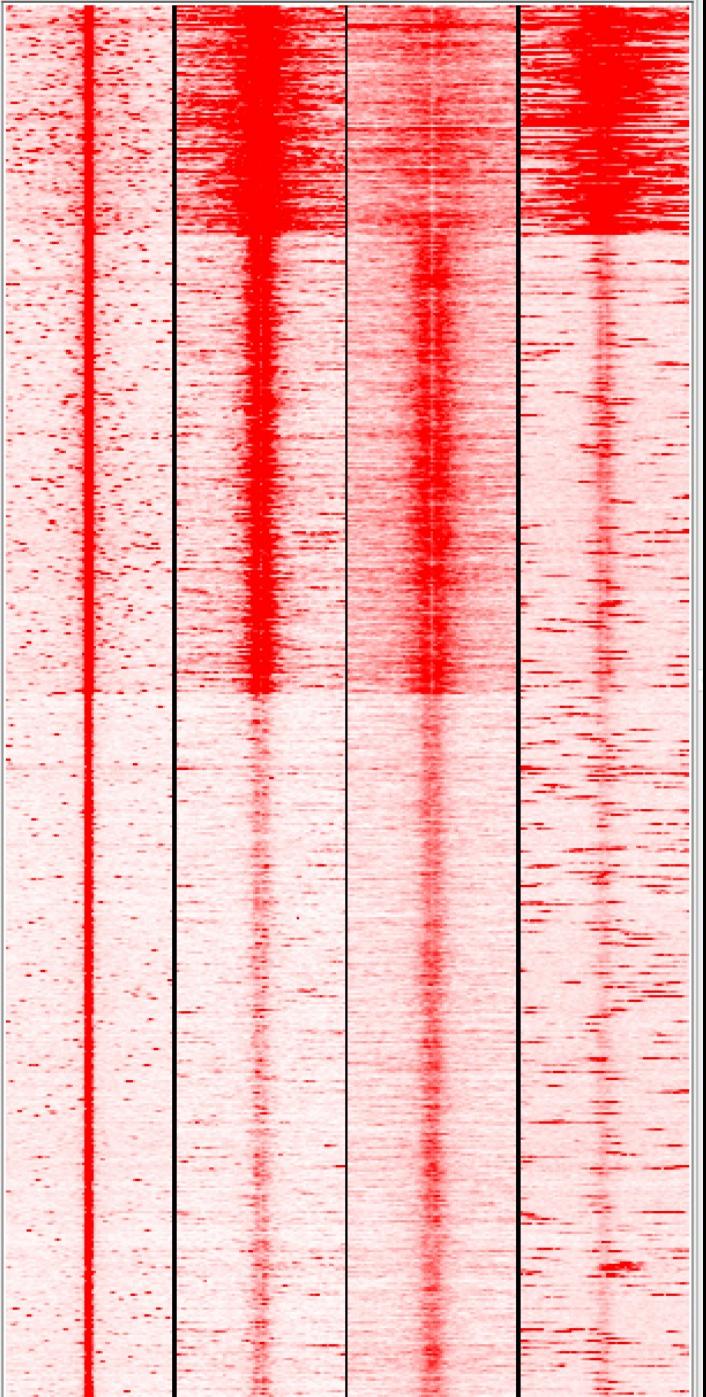


TAL1

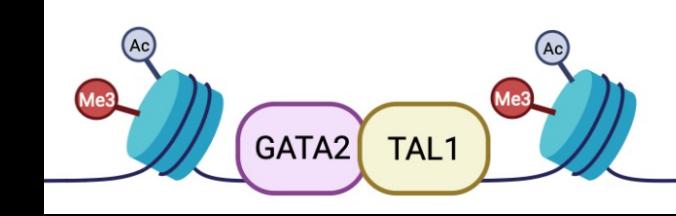
JASPARSTRING

«Recruitment by GATA transcription factors appears to be a stronger determinant of TAL1 binding to chromatin than the canonical E-box binding site motif. [...] TAL1 regulates expression after being directed to a distinct subset of genomic binding sites in each cell type via its association with different complexes containing master regulators such as GATA2, ERG, and RUNX1»

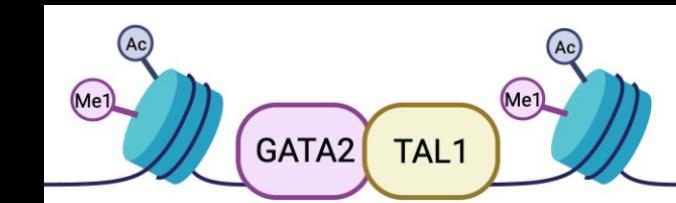
PubMed



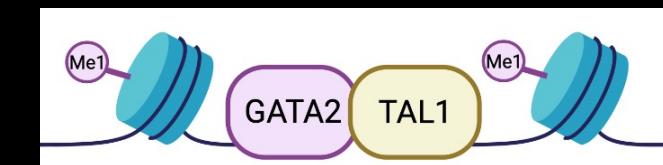
# seqMINER



TRANSCRIPTION STARTING SITE



ENHANCER



WEAK ENHANCER / OTHER STATES



# GREAT

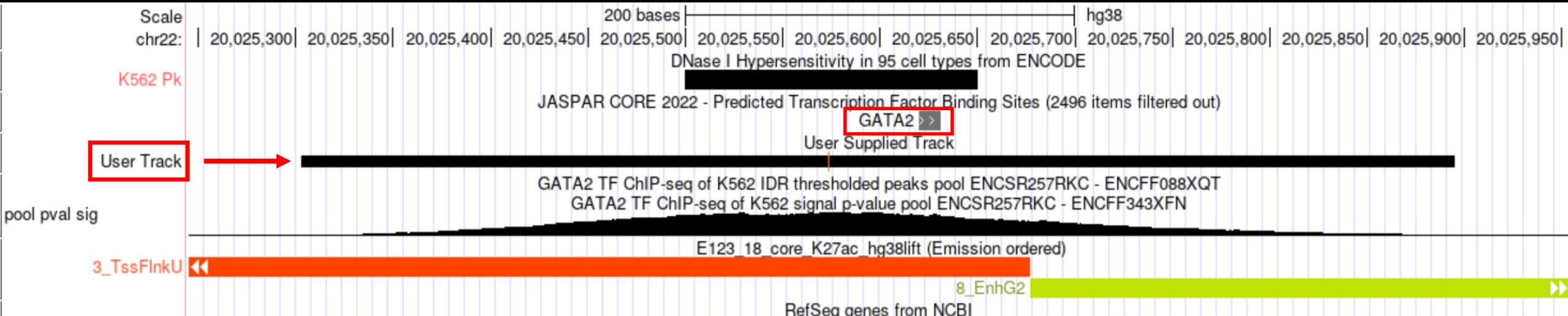
Total peaks	Regulated genes	Regulated promoters	Regulated enhancers
5392 peaks	2827	409	2499

Promoter: +/- 1 kb around the TSS

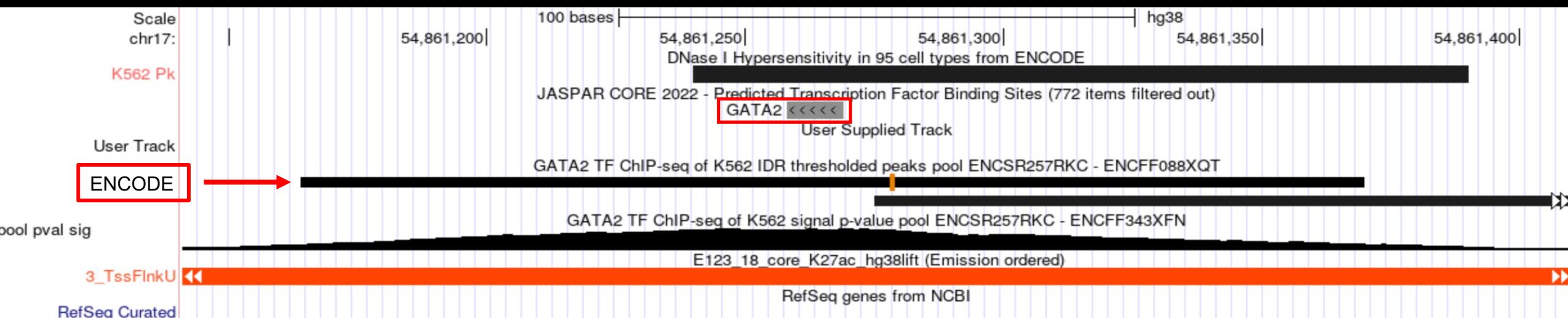
Distal regulatory regions: up to 30 kb from the TSS

GATA2 regulates itself

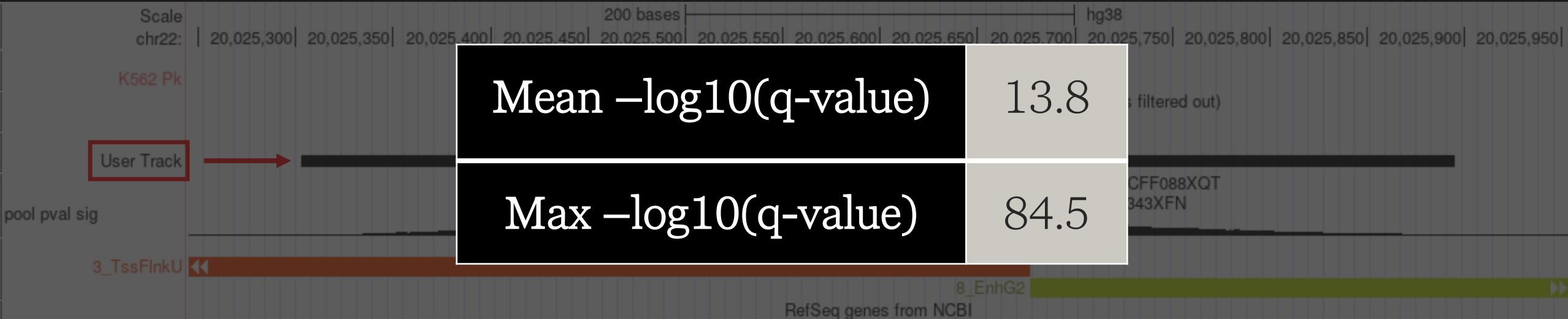
# Our most enriched peak



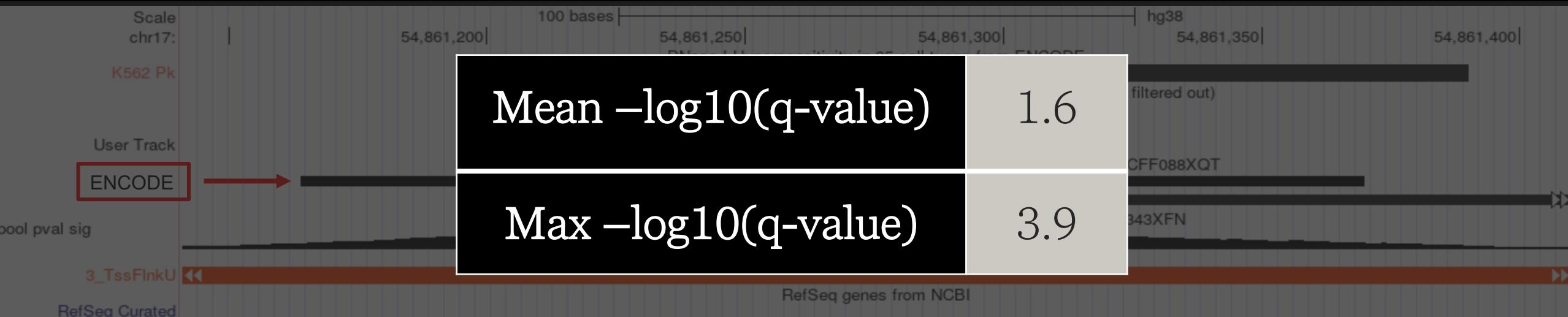
# ENCODE most enriched peak



# Our most enriched peak

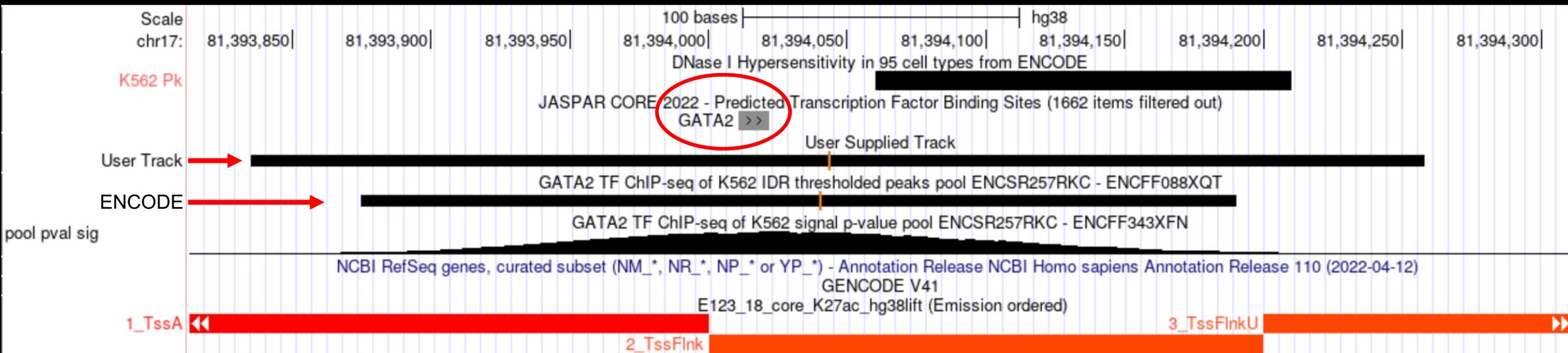


# ENCODE most enriched peak

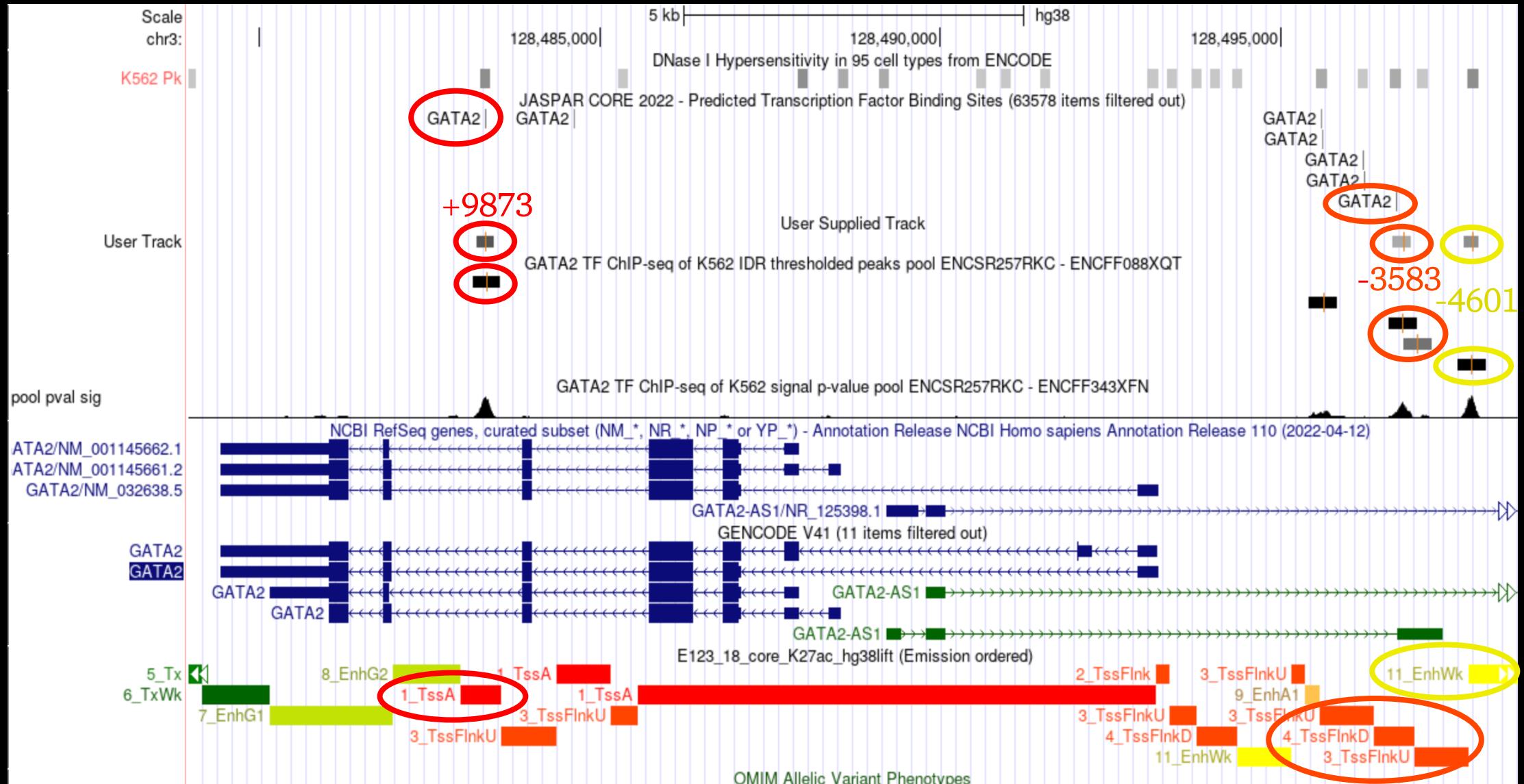




# Most enriched peak in common with ENCODE

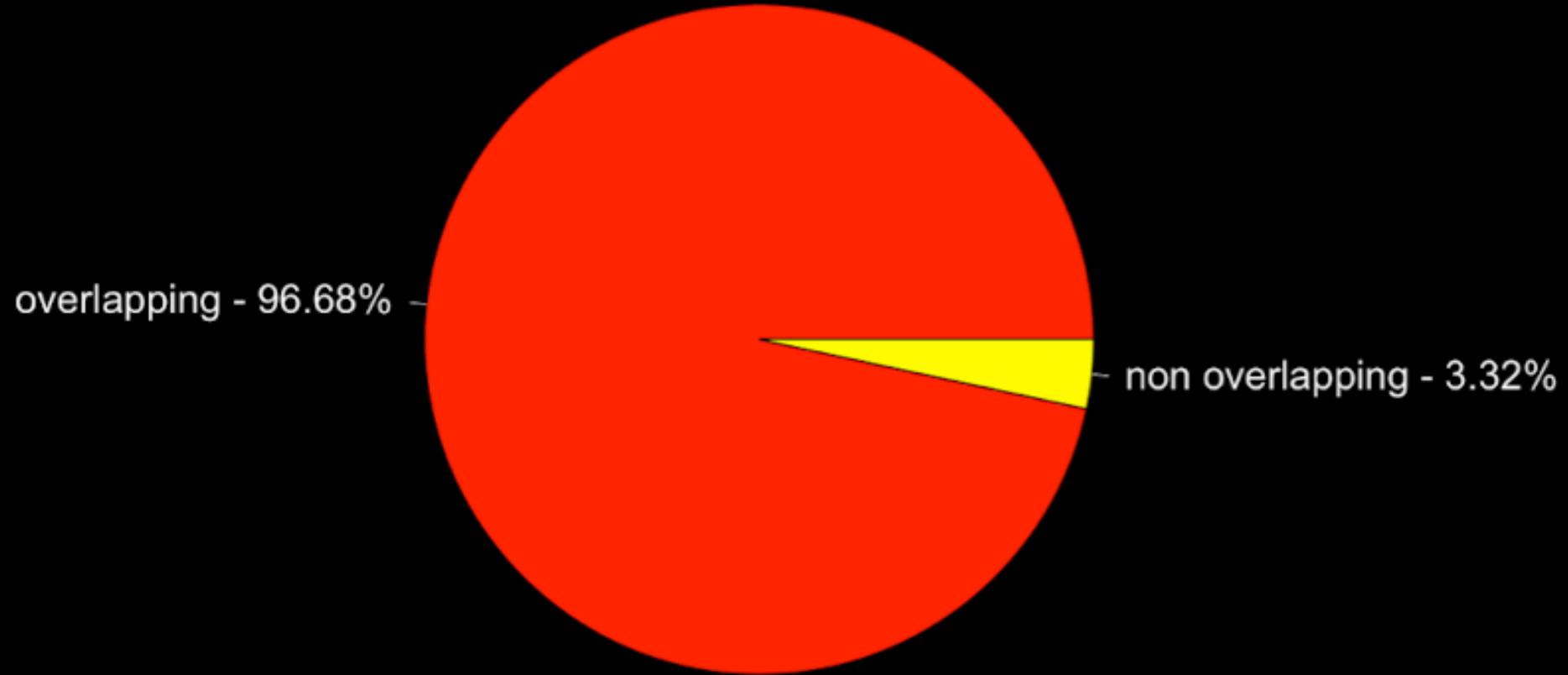


# GATA2



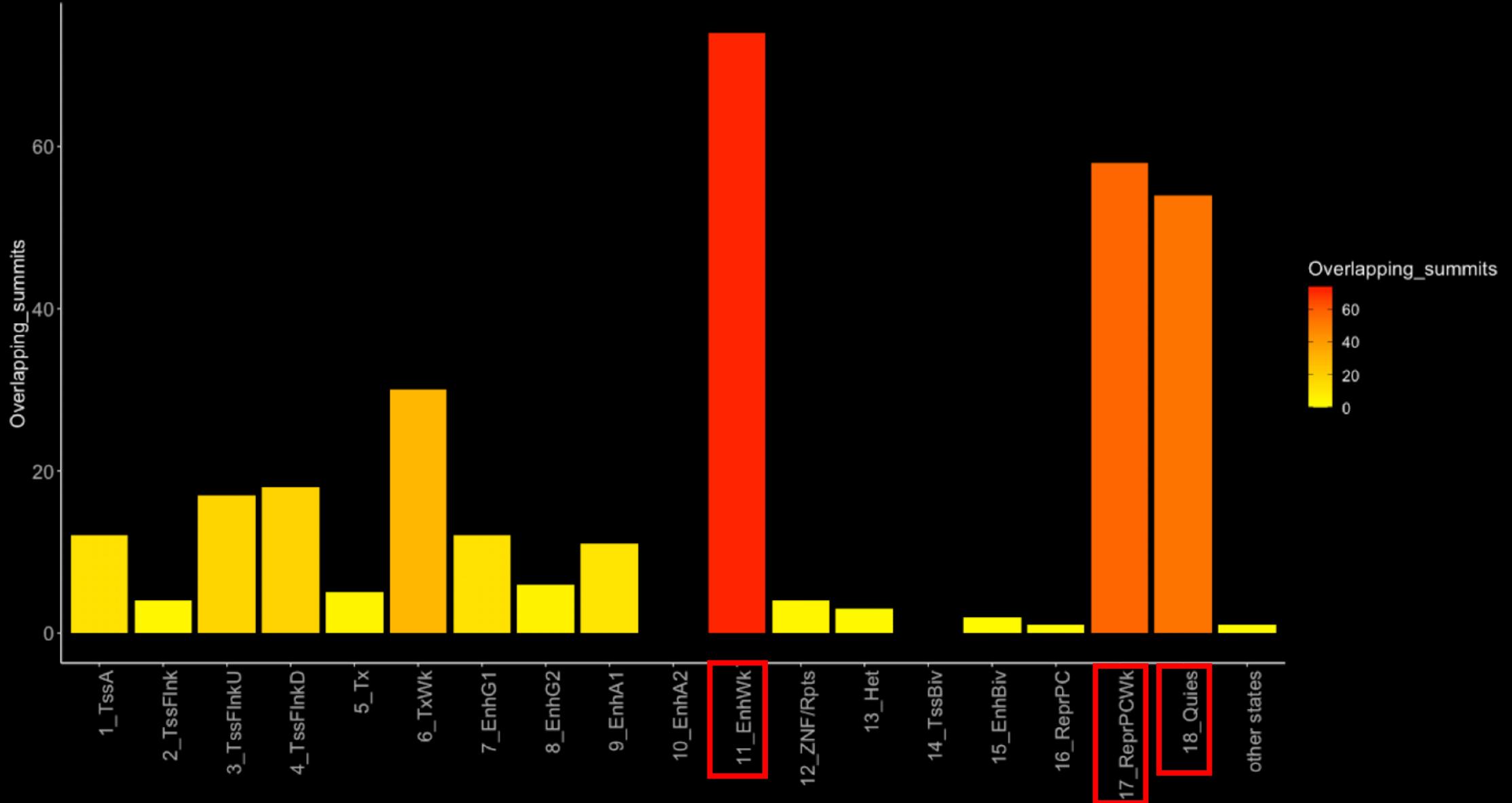


# % of GATA2 summits overlapping with DNase-Seq peaks





# Chromatin states of non-overlapping summits

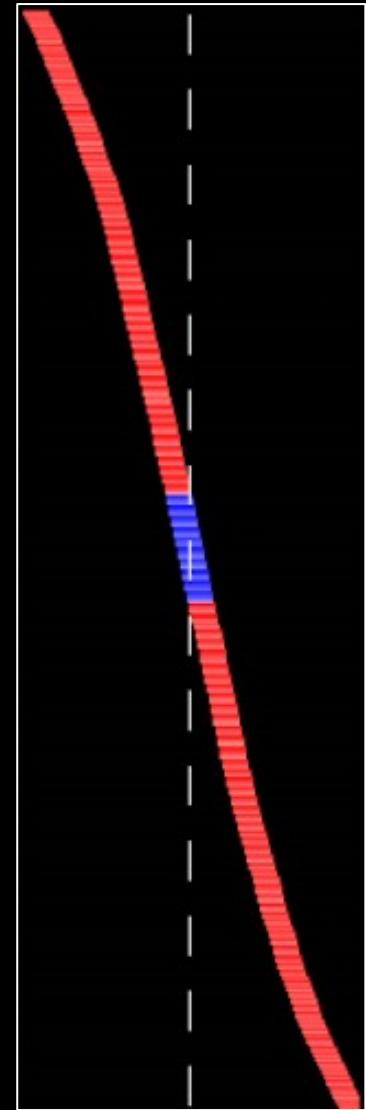
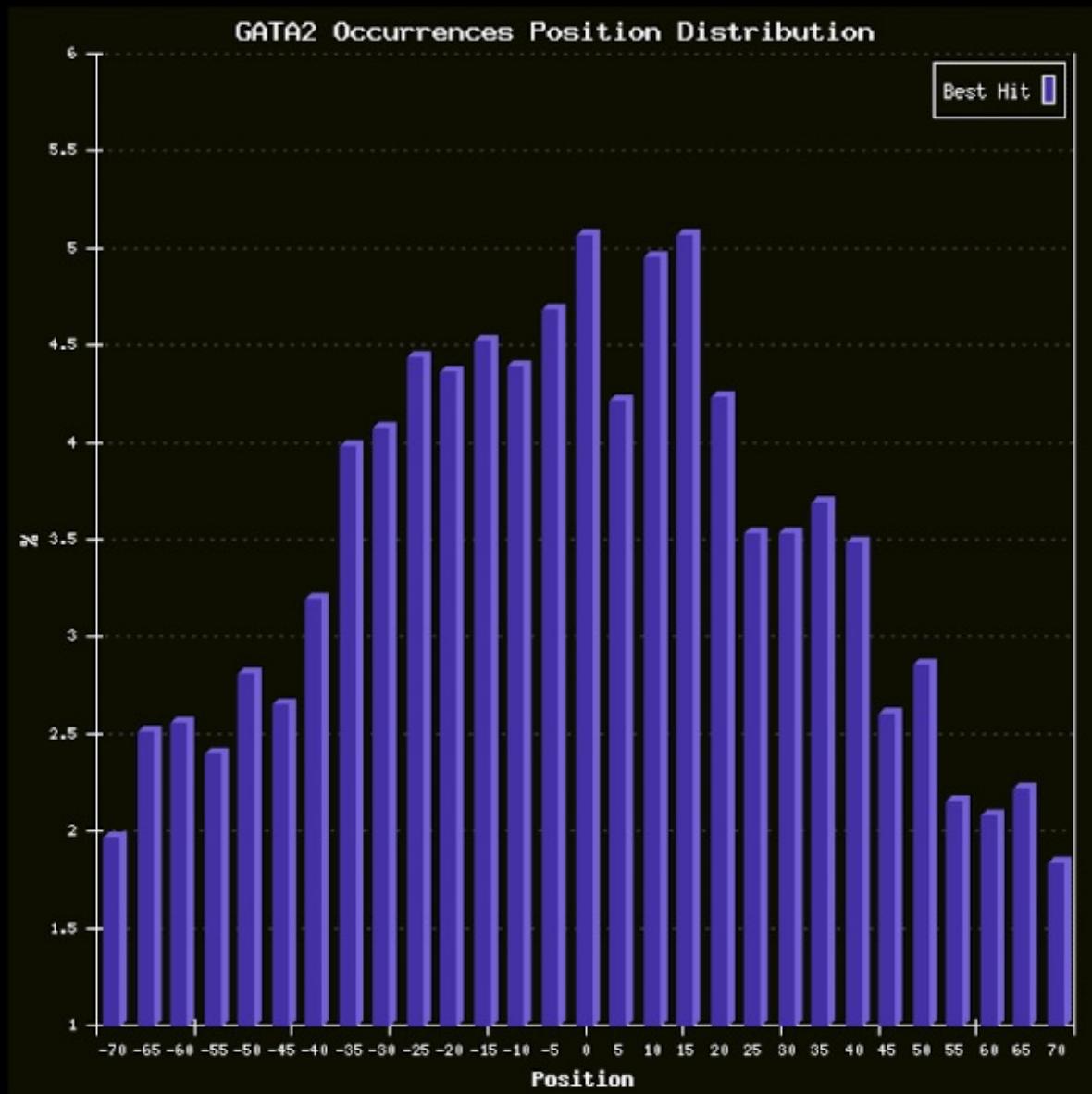


# Pscan-ChIP

NEW: Ver. 1.3 (Last update: 24 Jan 2018)



Name	ID	L.PV	L.O/U	G.PV	G.O/U	SP.COR	P.POS	P.POS.PV
<a href="#">Gata1</a>	MA0035.3	0	↑	0	↑	0.0742	[-5,5]	1.0000
<a href="#">GATA1::TAL1</a>	MA0140.2	0	↑	0	↑	0.1085	[4,14]	9.3E-7
<a href="#">Gata4</a>	MA0482.1	0	↑	0	↑	0.073	[-5,5]	0.4781
<a href="#">GATA5</a>	MA0766.1	0	↑	0	↑	0.0671	[9,19]	0.3811
<a href="#">GATA2</a>	MA0036.3	0	↑	0	↑	0.0728	[-5,5]	0.6807
<a href="#">GATA6</a>	MA1104.1	0	↑	0	↑	0.0513	[-4,6]	0.2761
<a href="#">GATA3</a>	MA0037.3	0	↑	0	↑	0.0711	[19,29]	1.0000
<a href="#">Mecom</a>	MA0029.1	1.7E-102	↑	3.9E-157	↑	-0.0136	[7,17]	0.0400
<a href="#">NEUROD1</a>	MA1109.1	5.9E-59	↑	2.0E-71	↑	0.0234	[1,11]	1.0000
<a href="#">TWIST1</a>	MA1123.1	6.6E-47	↑	3.2E-53	↑	0.0127	[-19,-9]	1.0000
<a href="#">Bhlha15</a>	MA0607.1	5.1E-46	↑	1.3E-70	↑	0.0396	[-26,-16]	1.0000



Distribution of the localization of the most likely instance of the binding motif around the summits of GATA2



## GATA FAMILY

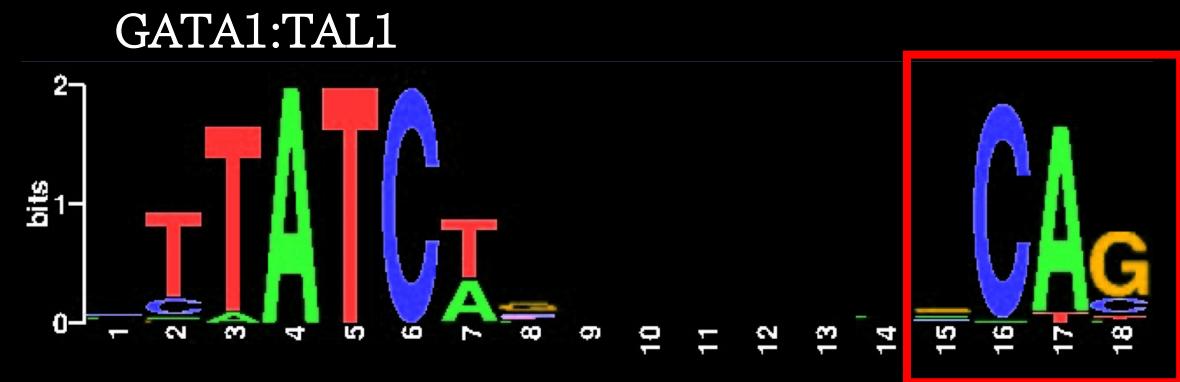
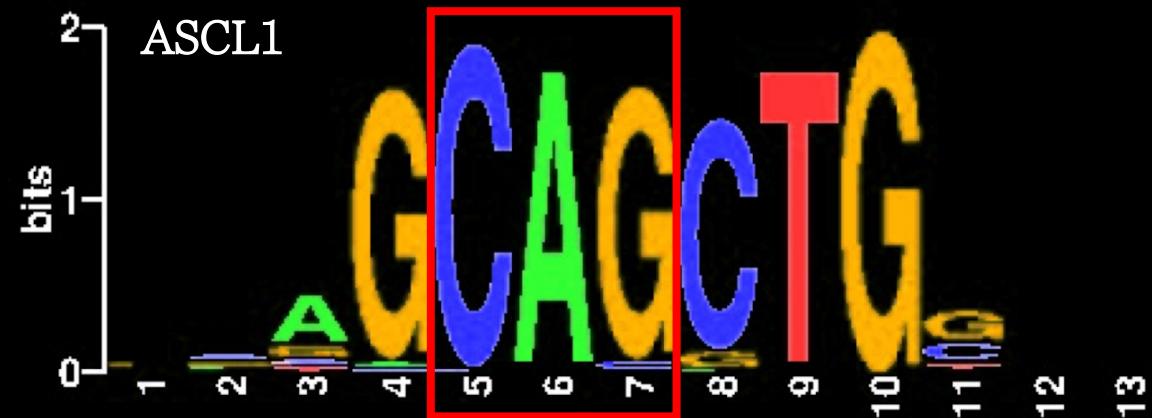
Name	ID	L.PV	L.O/U	G.PV	G.O/U	SP.COR	P.POS	P.POS.PV
<a href="#">Gata1</a>	MA0035.3	0		0		0.0742	[-5,5]	1.0000
<a href="#">GATA1::TAL1</a>	MA0140.2	0		0		0.1085	[4,14]	9.3E-7
<a href="#">Gata4</a>	MA0482.1	0		0		0.073	[-5,5]	0.4781
<a href="#">GATA5</a>	MA0766.1	0		0		0.0671	[9,19]	0.3811
<a href="#">GATA2</a>	MA0036.3	0		0		0.0728	[-5,5]	0.6807
<a href="#">GATA6</a>	MA1104.1	0		0		0.0513	[-4,6]	0.2761
<a href="#">GATA3</a>	MA0037.3	0		0		0.0711	[19,29]	1.0000

## BHLH FAMILY

<a href="#">Mecom</a>	MA0029.1	1.7E-102		3.9E-157		-0.0136	[7,17]	0.0400
<a href="#">NEUROD1</a>	MA1109.1	5.9E-59		2.0E-71		0.0234	[1,11]	1.0000
<a href="#">TWIST1</a>	MA1123.1	6.6E-47		3.2E-53		0.0127	[-19,-9]	1.0000
<a href="#">Bhlha15</a>	MA0607.1	5.1E-46		1.3E-70		0.0396	[-26,-16]	1.0000
<a href="#">GSC2</a>	MA0891.1	2.5E-42		3.6E-107		0.0092	[15,25]	1.0000
<a href="#">ASCL1</a>	MA1100.1	2.4E-40		2.8E-60		0.0638	[-4,6]	1.0000
<a href="#">NHLH1</a>	MA0048.2	9.9E-39		2.0E-61		0.0565	[-5,5]	1.0000
<a href="#">Ascl2</a>	MA0816.1	2.9E-37		3.3E-61		0.0669	[-20,-10]	1.0000
<a href="#">Tcf12</a>	MA0521.1	6.1E-37		3.0E-55		0.0595	[-5,5]	1.0000
<a href="#">TFAP4</a>	MA0691.1	4.6E-36		4.4E-46		0.0358	[-5,5]	1.0000
<a href="#">Myog</a>	MA0500.1	2.0E-34		2.7E-52		0.0597	[-5,5]	1.0000

# ASCL1

ASCL1 MA1100.1 2.4E-40 ↑ 2.8E-60 ↑ 0.0638 [-4,6] 1.0000



"Plays a key role in neuronal differentiation: acts as a pioneer transcription factor and activates neural pathways. The combination of three transcription factors, ASCL1, POU3F2/BRN2 and MYT1L, is sufficient to reprogram fibroblasts and other somatic cells into induced neuronal (iN) cells in vitro."

THPA



SO LONG,  
AND THANKS FOR ALL THE FISH



# Encode Project

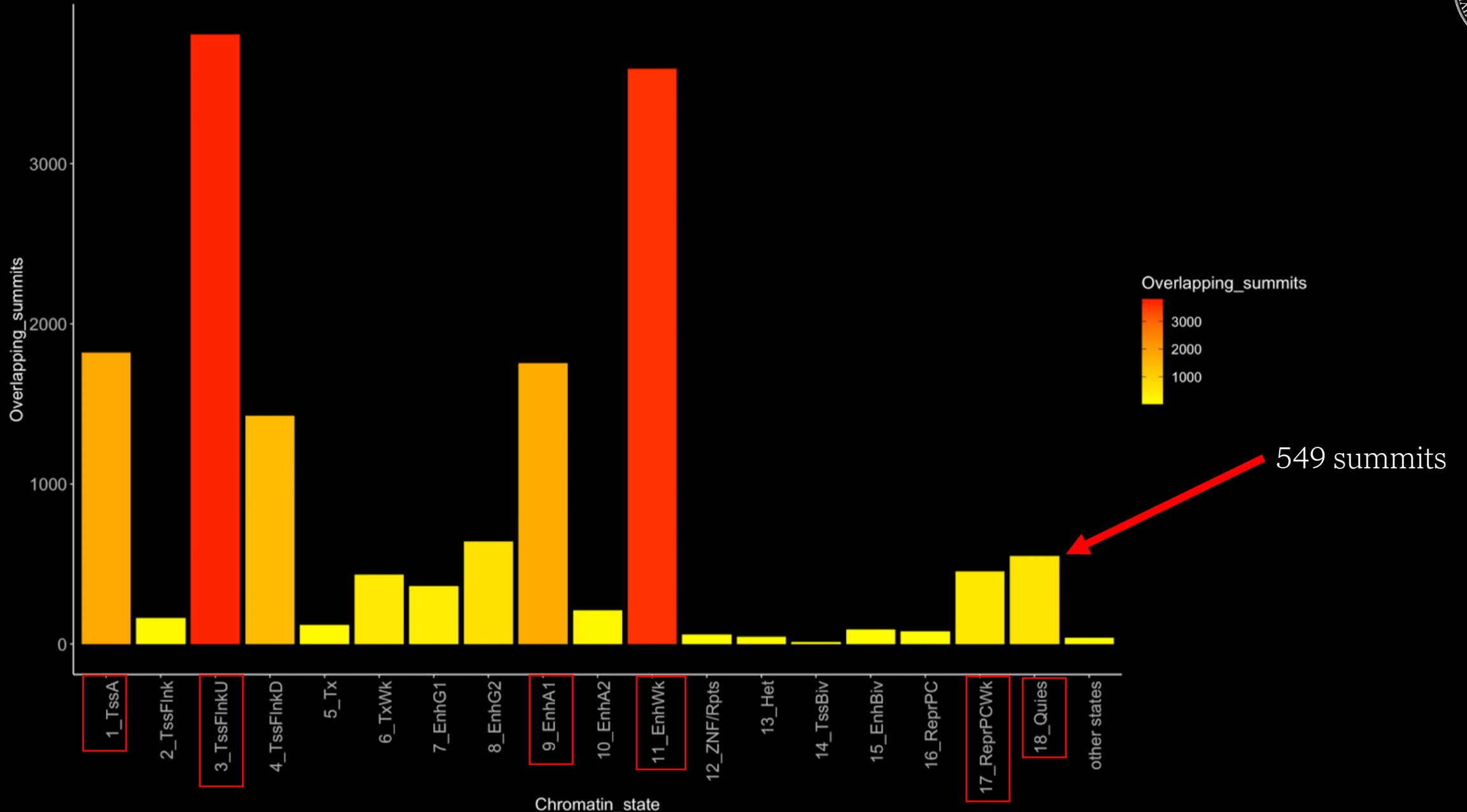
- ENCODE project provides information about the regulatory landscape of mammalian cells – K562, HeLa, GM12868 and HepG2
- It includes ChIP-Seq experiments performed on hundreds of human transcription factors
- Roadmap Epigenomics Project data of histone modifications and DNA-methylation are included in ENCODE portal
- All the data are processed with exactly the same computational pipeline, same parameters and same thresholds for quality checks

## Our Data

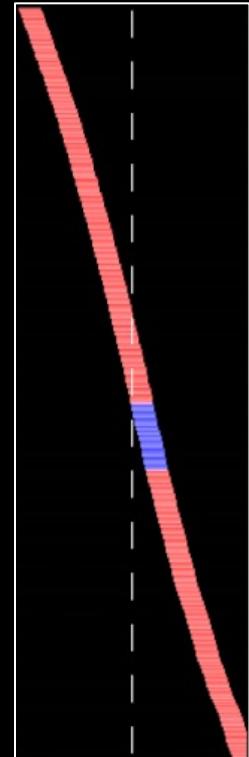
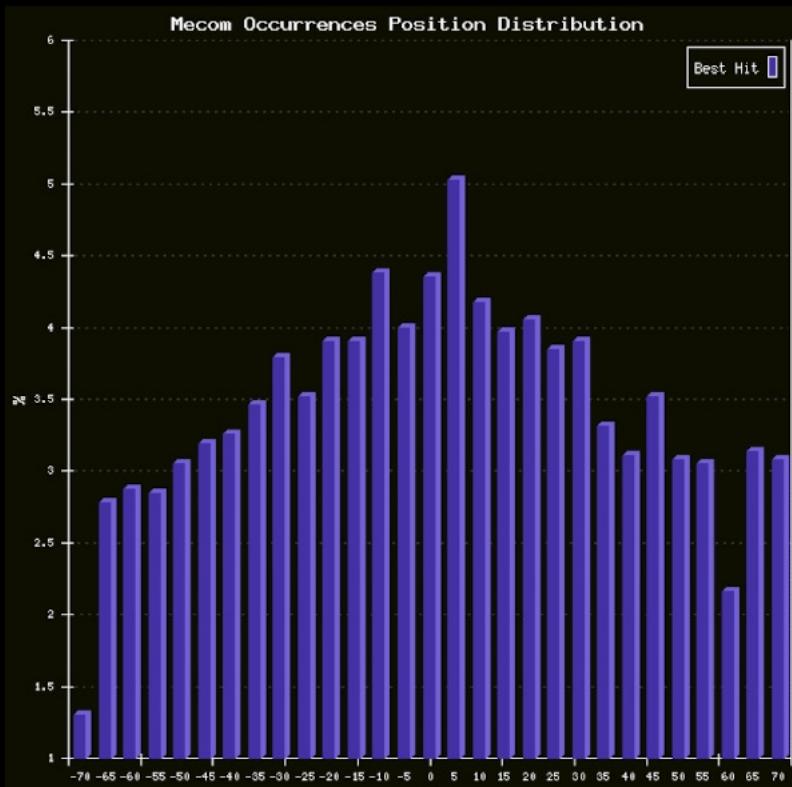
- GATA2 Transcription Factor
- In K562 cell line
- Antibody against a protein «flagged» with CRISPR editing
- ENCODE4 v1.4.0 GRCh38



# ENCODE: Chromatin States



# Mecom TF



«Transcriptional regulator binding to the promoter region of target genes and regulating positively or negatively their expression. Oncogene which plays a role in development, cell proliferation and differentiation. May also play a role in apoptosis through regulation of the JNK and TGF-beta signaling. Involved in hematopoiesis.»

# Pscan-ChIP

NEW: Ver. 1.3 (Last update: 24 Jan 2018)

