Annotate Peaks

ChIP-Seq and ATAC-Seq identify enriched regions or peaks in genome. Depending on the assay, the biological meaning of enrichment changes; in ChIP-Seq, enrichment indicates protein binding, while in ATAC-Seq, enrichment indicates open chromatin. To understand the importance of enriched regions in regulating gene expression, we can add information about overlapping or nearby genomic features.

What is Annotate peaks?

Annotate peaks takes an input set of regions and checks for overlap between those regions and a gene/feature annotation. This gives regulatory context for enriched regions.

Running Annotate peaks

The input for Annotate peaks is a Peaks type data node.

- Click a Peaks data node
- Click the Peak analysis section in the toolbox
- Click Annotate peaks
- Set the *Genomic overlaps* parameter

The Genomics overlaps parameter lets you choose one of two options (Figure 1).

- Report one gene region per peak (precedence applies) chooses one gene section for each peak using the precedence order to settle cases where more than one gene section overlaps a peak. The order of precedence is TSS, TTS, CDS Exon, 5' UTR Exon, 3' UTR Exon, Intron, Intergenic.
- Report all gene regions per peak creates a row for each gene section that overlaps a peak in the task report.

Genomic overlaps 🧃	 Report one gene region per peak (precedence applies) Report all gene regions per peak
TSS upstream limit (1000
TSS downstream limit (100
TTS upstream limit (100
TTS downstream limit 🏼 🚺	1000
Select Annotation file	
Assembly	Homo sapiens (human) - hg38
Annotation model	Ensembl Transcripts release 104
Back Finish	

Figure 1. Annotate peaks dialog

User should define the transcription start site (TSS) and transcription termination site (TTS) limit in the unit of bp

- · Choose a gene/feature annotation from the drop-down menu
- · Click Finish to run

Annotate peaks output

Annotate peaks produces an Annotated peaks data node. The Annotated peaks task report adds a Gene section breakdown pie chart and adds columns with information about the Gene IDs, Transcript IDs, Gene section, Distance to TSS, and Distance to TTS of each peak to the standard Peaks report (Figure 2). If run with the option to report all gene sections selected, each peak will have a row for each gene section it overlaps. If run with the option to report one gene section selected, each peak will have one row with the gene section it overlaps chosen using the order of precedence.

4 B				• TSS	TTS C	Gene sector	breakdown	non • Intergenic				
ptional c	column iew	S Sample name Search	Chromosome Search	≎ Start Search	≎ End Search	✓ -log10(pvalue) Search	Fold enrichment Search	Gene IDs Search	Transcript IDs Search	Gene section Search	Distance to TSS Search	Distance to TTS Search
1 5	-	ChIP_vs_IGG	22	50638871	50640269	44.89988	20.31955	SELENOO	NM_031454	TSS	0	-15777
2 4	2	ChIP_vs_IGG	22	36783579	36783989	40.07011	18.47232	MYH9	NM_002473	5' UTR Exon	124	-10625
3 4	2	ChIP_vs_IGG	22	31884921	31886689	37.34150	10.97828	EIF4ENIF1	NM_001164501	TSS	0	-4957
4 -5	2	ChIP_vs_IGG	22	50699516	50700287	33.48045	15.16395	MAPK12	NM_001303252	TSS	0	-818
5 4	-	ChIP_vs_IGG	22	46932456	46933530	31.01766	14.21620	CELSR1	NM_014246	TSS	0	-17572
6 4	-	ChIP_vs_IGG	22	39151508	39152036	30.08033	12.65124	SUN2	NM_001199579	TSS	0	-2078
7 4	2	ChIP_vs_IGG	22	41487854	41488676	30.08033	12.65124	EP300	NM_001429	TSS	0	-8740
8 4	2	ChIP_vs_IGG	22	46972671	46973346	30.08033	12.65124			Intergenic		
9 4	2	ChIP_vs_IGG	22	21896929	21897472	29.89290	8.98223			Intergenic		
10 🔥	-	ChIP_vs_IGG	22	37914280	37915657	28.07369	10.85044	CARD10	NM_014550	TSS	0	-2788
11 🔸	-	ChIP_vs_IGG	22	50912699	50913483	26.52464	13.51025	SBF1	NM_002972	TSS	18	-2926
12 🔥	2	ChIP_vs_IGG	22	51158429	51159522	26.23506	7.98420	SHANK3	NM_033517	CDS Exon 21	45359	-1211
13 🔥	-	ChIP_vs_IGG	22	46371257	46373008	26.17806	12.32071	WNT7B	NM_058238	TSS	1	-5500
14 🔸	2	ChIP_vs_IGG	22	22221169	22221902	26.12878	12.93062	MAPK1	NM_138957	TSS	69	-9785
15 🔥	2	ChIP_vs_IGG	22	28051662	28052640	24.76785	10.70490			Intergenic		
16 🔥	2	ChIP_vs_IGG	22	30783008	30783338	23.80540	11.37296	RNF215	NM_001017981	TSS	0	-820
17 🔸	-	ChIP_vs_IGG	22	26825245	26825782	23.04523	12.30407	ASPHD2	NM_020437	TSS	0	-1519
18 🔥	2	ChIP_vs_IGG	22	46466116	46468094	23.04523	12.30407	LOC642648	NR_148973	TSS	0	-558
19 🔸	-	ChIP_vs_IGG	22	38794149	38794884	22.62836	6.98618	LOC400927	NR_002821	TSS	48	-5347
20 🔥	-	ChIP_vs_IGG	22	36234318	36234813	21.70336	11.08339	RBFOX2	NM_001349997	Intron 1	1705	-9953
21 🔥	2	ChIP_vs_IGG	22	42509520	42509805	21.70336	11.08339	NDUFA6-AS1	NR_034118	5' UTR Exon	22583	-1155
22 - \$	2	ChIP_vs_IGG	22	20118786	20120043	19.60703	8.75855	ZDHHC8	NM_013373	TSS	0	-1548
aa 1	-	ChIP_vs_IGG	22	30116029	30116336	19.54145	10.15978	CABP7	NM_182527	TSS	-8	-1148
23 1		ChIP_vs_IGG	22	38301996	38302752	19.16733	9.47747	MICALL1	NM_033386	TSS	0	-3571
23 -) 24 - /	e											
23 1 24 1 25 1	-	ChIP_vs_IGG	22	42915365	42915809	19.16733	9.47747	RRP7A	NM_015703	TSS	21	-1102

Figure 2. Gene section breakdown pie chart

The table can be sorted by any of its columns (Figure 3). Click on the *Optional columns* on the upper-left corner of the table to add more information on each region

P	Peaks 20388												
5.7°	3			Gene se	rss est	edown							
		TSS	• TTS • CDS	S Exon • 5' U	TR Exon 🌘	3' UTR Exon 🌘 Intro	on 🔵 intergenic						
Option	nal columns	• TSS	• TTS • CDS	S Exon • 5' U	TR Exon	3' UTR Exon 🌘 Intro	on 🌒 Intergenic	1		1	1		
Option	nal columns Sample name ≎	TSS Chromosome +	● TTS ● CDS Start ≎	S Exon ● 5' U End ≎	TR Exon ● Length ≎	3' UTR Exon ● Intro -log10(pvalue) ≎	Fold enrichment \$	Gene IDs ≎	Transcript IDs ≎	Gene section ≎	Distance to TSS \$	Distance to TTS \$	gene_name ≎
Option 1-	nal columns Sample name ≎ SRR1944628	TSS Chromosome chr1	● TTS ● CDS Start ≎ 10.011	S Exon ● 5' U End ≎ 10,456	TR Exon • Length \$ 445	3' UTR Exon ● Intro -log10(pvalue) ≎ 31.195	Fold enrichment \$	Gene IDs ≎	Transcript IDs ≎	Gene section ≎ Intergenic	Distance to TSS \$	Distance to TTS \$	gene_name ≎
Optior &	nal columns Sample name ≎ SRR1944628 SRR1944628	TSS Chromosome chr1 chr1 chr1	• TTS • CDS Start \$ 10,011 180,745	S Exon ● 5' U End ≎ 10,456 180,948	TR Exon • Length \$ 445 203	3' UTR Exon ● Intro -log10(pvalue) ♥ 31.195 31.200	Fold enrichment 7.816 7.653	Gene IDs ≎	Transcript IDs \$	Gene section ≎ Intergenic Intergenic	Distance to TSS \$	Distance to TTS \$	gene_name \$
Option + + +	al columns Sample name ≎ SRR1944628 SRR1944628 SRR1944628	TSS Chromosome chr1 chr1 chr1 chr1	● TTS ● CDS Start ≎ 10,011 180,745 397,710	S Exon • 5' U End ¢ 10,456 180,948 398,092	TR Exon • Length \$ 445 203 382	3' UTR Exon ● Intro -log10(pvalue) ♥ 31.195 31.200 20.581	Intergenic Fold enrichment 7.816 7.653 7.179	Gene IDs ≎	Transcript IDs ¢	Gene section ≎ Intergenic Intergenic Intron 1	Distance to TSS \$	Distance to TTS ¢	gene_name \$
Optior + + + + + + +	al columns Sample name ≎ SRR1944628 SRR1944628 SRR1944628 SRR1944628	TSS Chromosome chr1 chr1 chr1 chr1 chr1	● TTS ● CDS Start ≎ 10,011 180,745 397,710 718,524	S Exon • 5' U End ¢ 10,456 180,948 398,092 718,837	TR Exon ● Length ≎ 445 203 382 313	3 UTR Exon ● Intro -log10(pvalue) \$ 31.195 31.200 20.581 20.013	Intergenic Fold enrichment 7.816 7.653 7.179 8.888	Gene IDs \$ 	Transcript IDs ¢	Gene section ≎ Intergenic Intergenic Intron 1 Intron 2	Distance to TSS \$	Distance to TTS ♀ 	gene_name \$
Option + + + + + + + + +	al columns Sample name \$ SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628	• TSS Chromosome • chr1 chr1 chr1 chr1 chr1	• TTS • CDS Start ¢ 10,011 180,745 397,710 718,524 2,135,498	S Exon 5' U End ¢ 10,456 180,948 398,092 718,837 2,135,772	TR Exon ● Length ≎ 445 203 382 313 274	3' UTR Exon ● Infro -log10(pvalue) ≎ 31.195 31.200 20.581 20.013 23.872	■ Intergenic Fold enrichment ≎ 7.816 7.653 7.179 8.888 6.109	Gene IDs ≎ 	Transcript IDs ¢ ENST00000455207 ENST00000414688 ENST00000470511	Gene section ≎ Intergenic Intergenic Intron 1 Intron 2 5' UTR Exon	Distance to TSS \$	Distance to TTS \$	gene_name \$
Option + + + + + + + + + + + +	al columns Sample name ♀ SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628	• TSS Chromosome • chr1 chr1 chr1 chr1 chr1 chr1	■ TTS ● CDS Start ≎ 10,011 180,745 397,710 718,524 2,135,498 2,189,546	S Exon 5' U End \$ 10,456 180,948 398,092 718,837 2,135,772 2,190,021	TR Exon ● Length ≎ 445 203 382 313 274 475	3' UTR Exon ● Intro -log10(pvalue) ≎ 31.195 31.200 20.581 20.013 23.872 20.832	n ● Intergenic Fold enrichment ◆ 7.816 7.653 7.179 8.888 6.109 7.287	Gene IDs ¢ ENSG00000237094 ENSG00000230021 ENSG00000067506 ENSG00000162585	Transcript IDs ≎ ENST00000455207 ENST00000414688 ENST00000470511 ENST00000378546	Gene section ≎ Intergenic Intregenic Intron 1 Intron 2 5' UTR Exon TTS	Distance to TSS \$	Distance to TTS ♀ 	gene_name
Option + + + + + + + + + +	al columns Sample name ≎ SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628	● TSS Chromosome → chr1 chr1 chr1 chr1 chr1 chr1 chr1	• TTS • CDS Start ¢ 10,011 180,745 397,710 718,524 2,135,498 2,189,546 4,650,392	S Exon ● 5' U End ≎ 10,456 180,948 398,092 718,837 2,135,772 2,190,021 4,650,829	TR Exon ● Length ≎ 445 203 382 313 274 475 437	3 UTR Exon Intro log10(pvalue) © 31,195 31,200 20,581 20,013 23,872 20,832 35,766	Fold enrichment 0 7.816 7.653 7.179 8.888 6.109 7.287 7.385 7.385	Gene IDs 0 ENSG0000237094 ENSG0000023021 ENSG0000057606 ENSG0000016255	Transcript IDs ¢ ENST00000455207 ENST00000414688 ENST00000470511 ENST00000378546	Gene section Intergenic Intergenic Intron 1 Intron 2 5' UTR Exon TTS Intergenic	Distance to TSS ♦ 	Distance to TTS ¢ 24,528 6,657 15,126	gene_name
Option + + + + + + + + + + + + + + + + + + +	al columns Sample name ¢ SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628	• TSS Chromosome - chr1 chr	■ TTS ■ CDS Start ≎ 10,011 180,745 397,710 718,524 2,189,546 4,650,392 4,702,915	End \$ Exam \$ 5'U End \$ 10,456 180,948 398,092 718,837 2,135,772 2,190,021 4,650,829 4,703,412	TR Exon ● Length ◆ 445 203 382 313 274 475 437 497	3 UTR Exon ● Intro -log10(pvalue) ≎ 31.195 31.200 20.581 20.013 23.872 20.832 35.766 23.184	■ Intergenic Fold enrichment 7 816 7 653 7 179 8 888 6 109 7 287 7 385 7 879	Gene IDs © ENSG0000237094 ENSG0000037094 ENSG0000057606 ENSG00000162585 ENSG00000166581	Transcript IDs ¢ ENST000045207 ENST0000414688 ENST0000470511 ENST0000378546 ENST0000378191	Gene section 0 Intergenic Intergenic Intron 1 S UTR Exon TTS Intergenic Intergenic	Distance to TSS © 87,117 13,376 19,293 4,752 48,306	Distance to TTS © -24,528 -6,657 -15,126 0 -89,123	gene_name O
Option + + + + + + + + + + + + + + + + + + +	al columns Sample name © SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628 SRR1944628	TSS Chromosome - drr1	TTS CDS Start C 10,011 180,745 397,710 718,524 2,135,5498 2,189,546 4,650,392 4,702,915 5,424,895	End C 10,456 10,456 180,948 398,092 718,837 2,135,772 2,190,021 4,650,829 4,703,412 5,425,289	TR Exon ● Length ◆ 445 203 382 313 274 475 437 497 394	3' UTR Exon • Infra -log10(pvalue) • 31.195 31.200 20.581 20.013 23.872 20.832 35.766 23.184 24.216	■ Intergenic Fold enrichment 7 816 7 653 7 179 8 888 6 6109 7 287 7 385 7 879 9 387	Gene IDa © ENSG0000237094 ENSG0000037094 ENSG0000057606 ENSG00000165265 ENSG0000196581	Transcript IDs \$ ENST00000455207 ENST0000041658 ENST00000378546 ENST00000378546	Gene section ≎ Intergenic Intergenic Intron 1 5 UTR Exon TTS Intergenic Intron 1	Distance to TSS ©	Distance to TTS ♥ 	gene_name

Figure 3. Filtering the Annotated peaks task report

Gene sections

TSS

Transcription start site (TSS) is -1000bp and +100bp (default setting) from the TSS for a transcript

TTS

Transcription termination site (TTS) is -100bp and +1000bp (default setting) from the TTS for a transcript

CDS Exon

Coding sequence (CDS) Exon is overlapping a coding exon in a transcript

5' UTR Exon

5' Untranslated Region (UTR) Exon is overlapping an exon in the 5' UTR of a transcript

3' UTR Exon

3' Untranslated Region (UTR) Exon is overlapping an exon in the 3' UTR of a transcript

Intron

Intron is overlapping an intron in a transcript

Intergenic

Intergenic is not located within 1000bp of a transcript

