

Notes on processing:

- I should have asked for FASTQ, as the BAM was 2x150mers, and did not have the unaligned reads
- Thus cross-correlation is not done properly, as it needs to be run on de novo 1x36mers alignments without the second end. And it did indeed turn out to be a problem (see further below)
- Peak calling and track generation were done on the supplied BAM files

Datasets stats

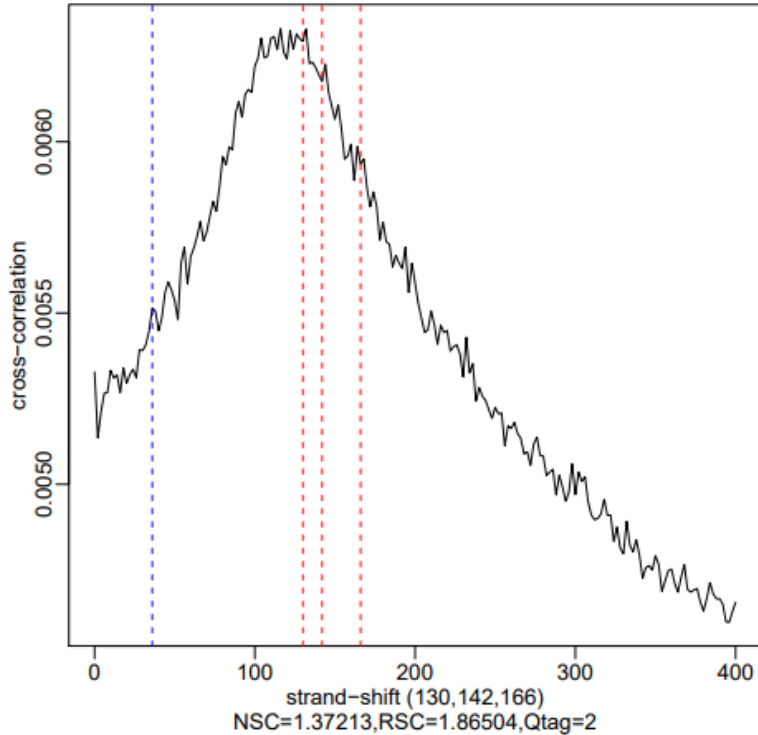
Species	#	Read Length	Raw fragments	Unique	Complexity	MACS2 NumPeaks, noBL	MACS2, noBL RPM	NSC	RSC	QC
supplied mapping	DDX50_H1_chip	2x150 BTW	N/A	29,608,873	0.79	0	0			
supplied mapping	DDX50_H1_input	2x150 BTW	N/A	22,028,070	0.78					
supplied mapping	DDX50_H2_chip	2x150 BTW	N/A	31,725,896	0.78	3	21			
supplied mapping	DDX50_H2_input	2x150 BTW	N/A	29,092,321	0.75					
supplied mapping	DDX50_L1_chip	2x150 BTW	N/A	40,354,158	0.77	1	1			
supplied mapping	DDX50_L1_input	2x150 BTW	N/A	24,699,203	0.68					
supplied mapping	DDX50_L2_chip	2x150 BTW	N/A	31,376,289	0.77	0	0			
supplied mapping	DDX50_L2_input	2x150 BTW	N/A	26,074,018	0.79					
supplied mapping	DDX50_P1_chip	2x150 BTW	N/A	36,795,954	0.76	0	0			
supplied mapping	DDX50_P1_input	2x150 BTW	N/A	30,527,213	0.79					
supplied mapping	DDX50_P2_chip	2x150 BTW	N/A	39,022,761	0.75	0	0			
supplied mapping	DDX50_P2_input	2x150 BTW	N/A	21,581,189	0.73					
supplied mapping	IFI16_D1_chip	2x150 BTW	N/A	37,197,574	0.77	17,848	37,880			
supplied mapping	IFI16_D1_input	2x150 BTW	N/A	33,707,646	0.73					
supplied mapping	IFI16_D2_chip	2x150 BTW	N/A	37,719,673	0.77	449	733			
supplied mapping	IFI16_D2_input	2x150 BTW	N/A	34,924,468	0.75					
supplied mapping	IFI16_P1_chip	2x150 BTW	N/A	37,335,783	0.77	0	0			
supplied mapping	IFI16_P1_input	2x150 BTW	N/A	18,402,083	0.66					
supplied mapping	IFI16_P2_chip	2x150 BTW	N/A	37,362,575	0.75	0	0			
supplied mapping	IFI16_P2_input	2x150 BTW	N/A	23,088,336	0.75					
hg38-male	DDX50_H1_chip	1x36						1.372	1.865	2
hg38-male	DDX50_H1_input	1x36						1.522	1.867	2
hg38-male	DDX50_H2_chip	1x36						1.321	1.819	2
hg38-male	DDX50_H2_input	1x36						1.333	1.891	2
hg38-male	DDX50_L1_chip	1x36						1.288	2.095	2
hg38-male	DDX50_L1_input	1x36						1.431	1.943	2
hg38-male	DDX50_L2_chip	1x36						1.315	2.054	2
hg38-male	DDX50_L2_input	1x36						1.318	1.813	2
hg38-male	DDX50_P1_input	1x36						1.268	2.431	2
hg38-male	DDX50_P1_nchip	1x36						1.345	1.852	2
hg38-male	DDX50_P2_chip	1x36						1.307	1.988	2
hg38-male	DDX50_P2_input	1x36						1.470	2.127	2
hg38-male	IFI16_D1_chip	1x36						1.261	2.215	2
hg38-male	IFI16_D1_input	1x36						1.359	1.961	2
hg38-male	IFI16_D2_chip	1x36						1.271	2.081	2
hg38-male	IFI16_D2_input	1x36						1.328	1.851	2
hg38-male	IFI16_P1_chip	1x36						1.283	1.831	2
hg38-male	IFI16_P1_input	1x36						1.462	2.265	2
hg38-male	IFI16_P2_chip	1x36						1.312	2.000	2
hg38-male	IFI16_P2_input	1x36						1.396	1.883	2

Note:
NSC/RSC are not meaningful because of the mapping issue

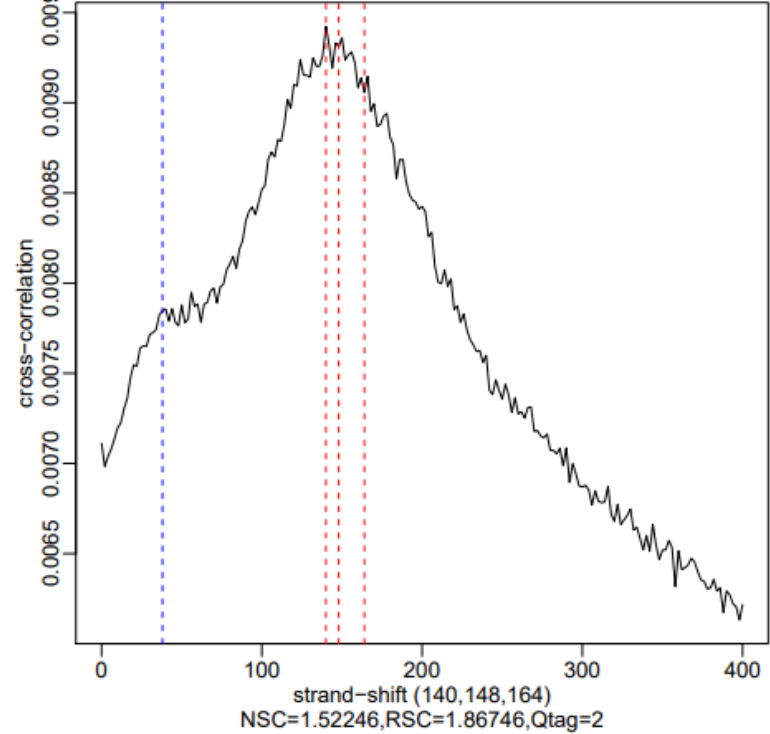
Note:
Complexity in the inputs is low

Cross-correlation





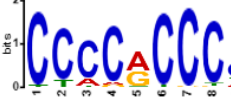
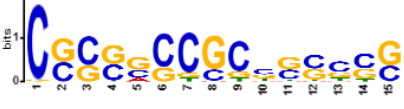
DDX50_H1_chip.mapped.end1.1x36mers.hg38-male.unique.



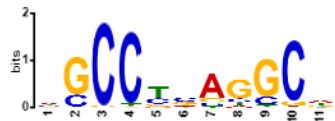
DDX50_H1_input.mapped.end1.1x36mers.hg38-male.unique.



MEME-ChIP on IFI16_D2_chip

Motif Found	Discovery/Enrichment Program	E-value	Known or Similar Motifs	Distribution	SpaMo & FIMO
 <p>bits</p> <p>Reverse Complement ↵ Show 45 More ↵ CentriMo Group ↵</p>	MEME	4.6e-056	Jun (MA0489.2) BNC2 (MA1928.1) Atf3 (MA1988.1)		<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3
 <p>bits</p> <p>Reverse Complement ↵</p>	STREME	4.2e-006	CEBPB (MA0466.3) CEBPE (MA0837.2) CEBPG (MA0838.1)	Not Centrally Enriched	<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3
 <p>bits</p> <p>Reverse Complement ↵</p>	STREME	9.6e-006	TEAD1_full_1 TEAD2 (MA1121.1) TEAD3 (MA0808.1)	Not Centrally Enriched	<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3
 <p>bits</p> <p>Reverse Complement ↵</p>	STREME	7.3e-005	KLF1 (MA0493.2) KLF5 (MA0599.1) KLF10 (MA1511.2)	Not Centrally Enriched	<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3
 <p>bits</p>	MEME	1.2e-004	ZNF610 (MA1713.1) PATZ1 (MA1961.1) Zfx (MA0146.2)	Not Centrally Enriched	<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3

Motif Found



Reverse Complement ⇄ Show 2 More J? CentriMo Group ↻?

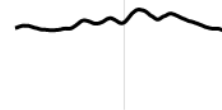
Discovery/Enrichment Program ? E-value ? Known or Similar Motifs ?

[CentriMo](#)

5.1e-004

[TFAP2C_full_3](#)

Distribution ?



SpaMo & FIMO ?

Motif Found



Reverse Complement ⇄

Discovery/Enrichment Program ? E-value ? Known or Similar Motifs ?

[STREME](#)

8.4e-004

Distribution ?

Not Centrally Enriched

SpaMo & FIMO ?

- [Motif Spacing Analysis](#)
- [Motif Sites in GFF3](#)

Motif Found



Reverse Complement ⇄ Show 9 More J? CentriMo Group ↻?

Discovery/Enrichment Program ? E-value ? Known or Similar Motifs ?

[CentriMo](#)

1.8e-003

[ASCL1 \(MA1631.1\)](#)

Distribution ?



SpaMo & FIMO ?

Motif Found



Reverse Complement ⇄

Discovery/Enrichment Program ? E-value ? Known or Similar Motifs ?

[STREME](#)

2.9e-002

[IKZF1 \(MA1508.1\)](#)
[FOXO1::ELF1 \(MA1953.1\)](#)
[ERF::FOXO1 \(MA1936.1\)](#)

Distribution ?

Not Centrally Enriched

SpaMo & FIMO ?


- [Motif Spacing Analysis](#)
- [Motif Sites in GFF3](#)

MOTIFS

The significant motifs (E-value ≤ 0.05) found by the programs MEME, STREME and CentriMo; clustered by similarity and ordered by E-value.

Expand All Clusters

Collapse All Clusters

Motif Found	Discovery/Enrichment Program ?	E-value ?	Known or Similar Motifs ?	Distribution ?	SpaMo & FIMO ?
 <p>Reverse Complement ↵</p>	MEME	1.1e-027	FOSB::JUNB (MA1135.1) FOSL2::JUNB (MA1138.1) FOS::JUN (MA0099.3)	Not Centrally Enriched	<ul style="list-style-type: none"> Motif Spacing Analysis Motif Sites in GFF3

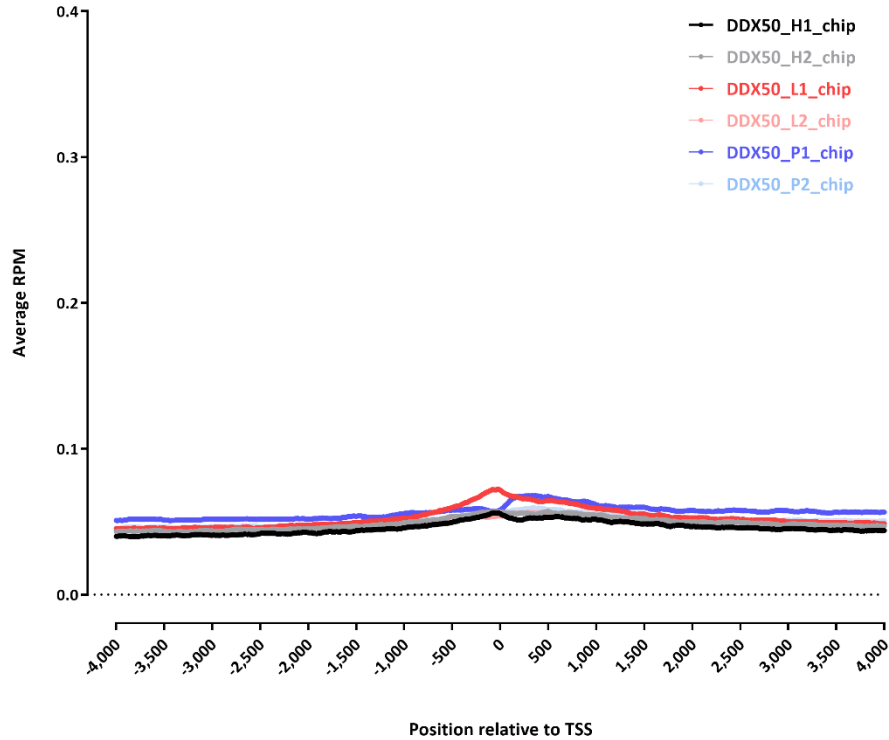
Positional Distribution [?](#)

NEW

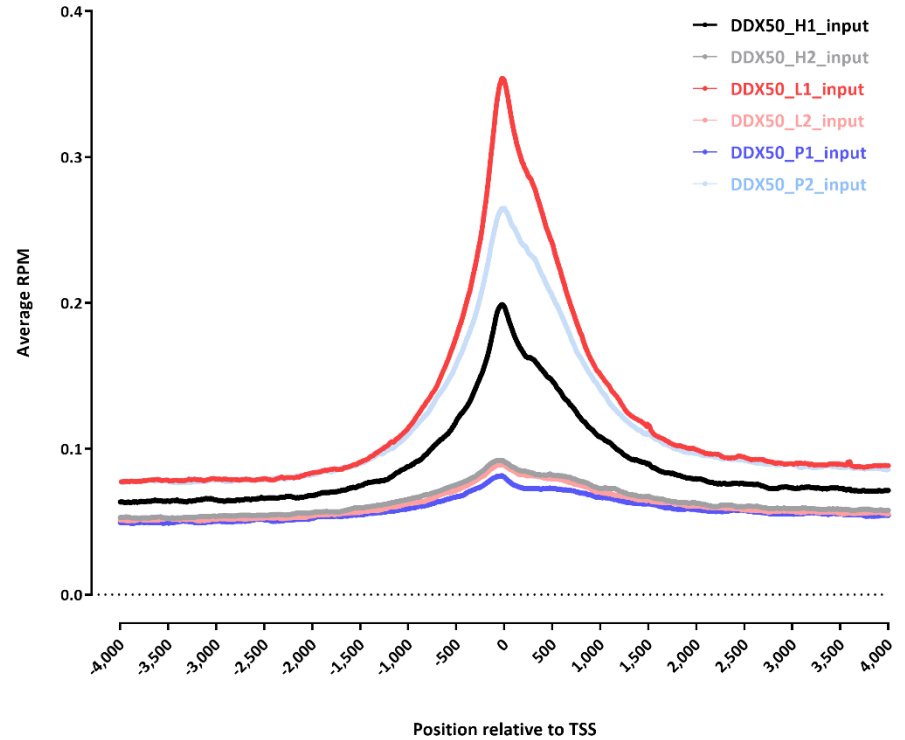


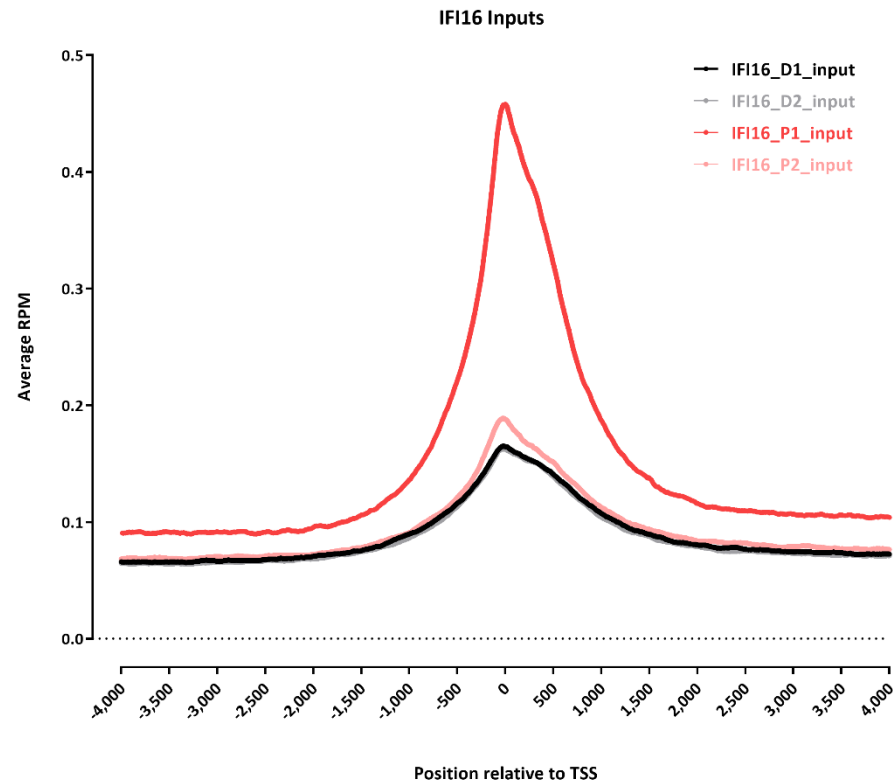
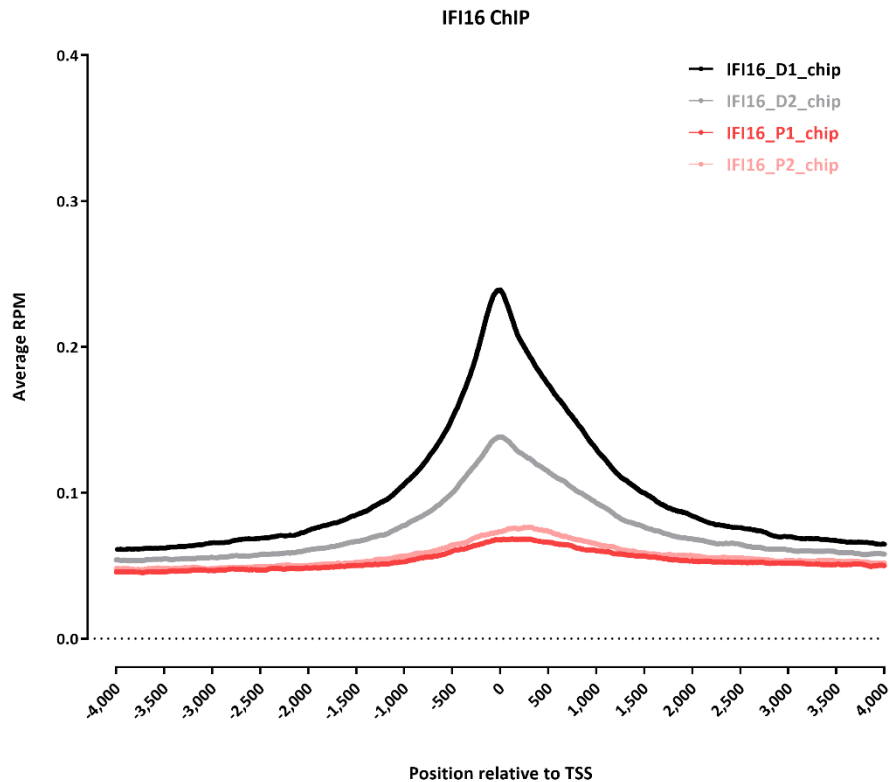
TSS profiles

DDX50 CHIP



DDX50 Inputs

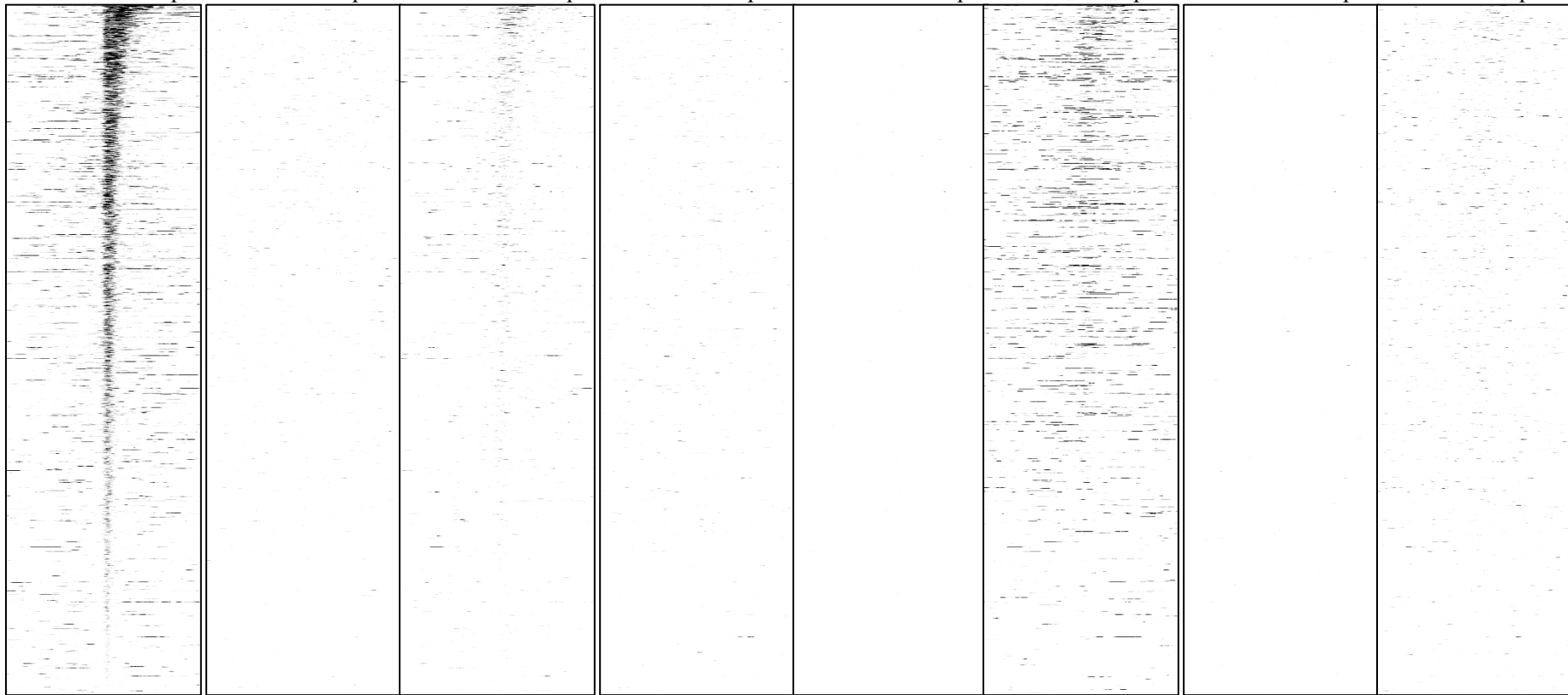




There is a rather strong TSS bias in the inputs

Signal distribution around IFI16_D1_chip peaks

IFI16_D1_chip | IFI16_D1_input | IFI16_D2_chip | IFI16_D2_input | IFI16_P1_chip | IFI16_P1_input | IFI16_P2_chip | IFI16_P2_input



-2.5kb

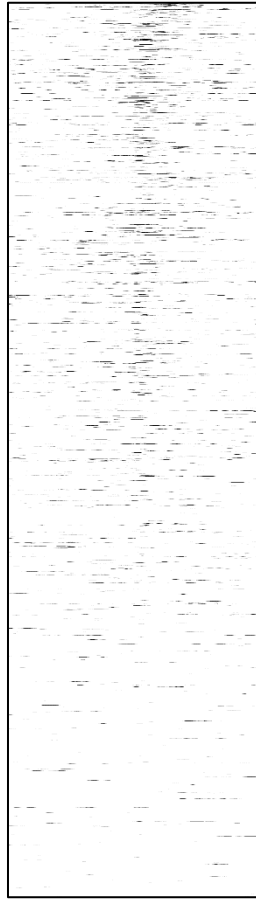
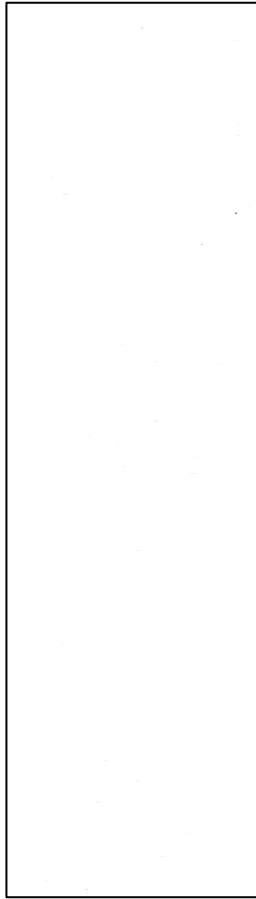
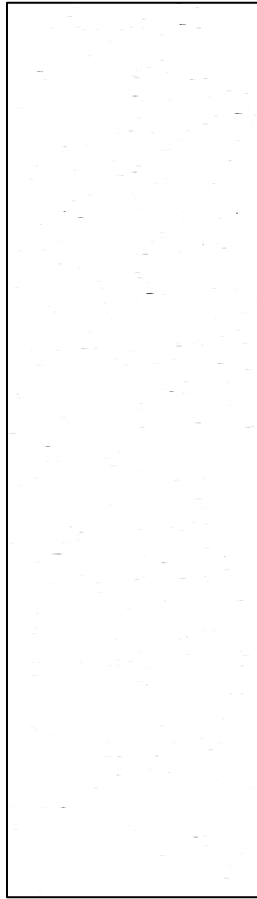
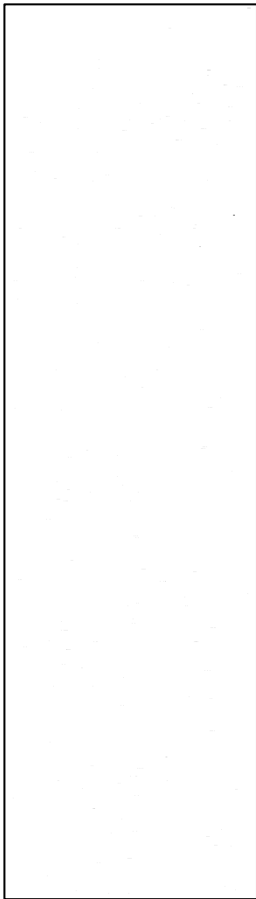
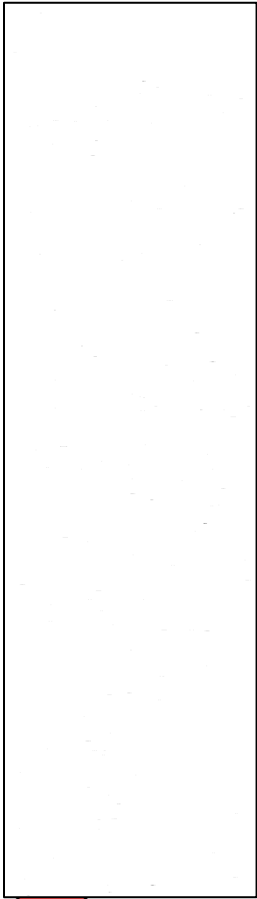
0

+2.5kb

9

The asymmetry is weird

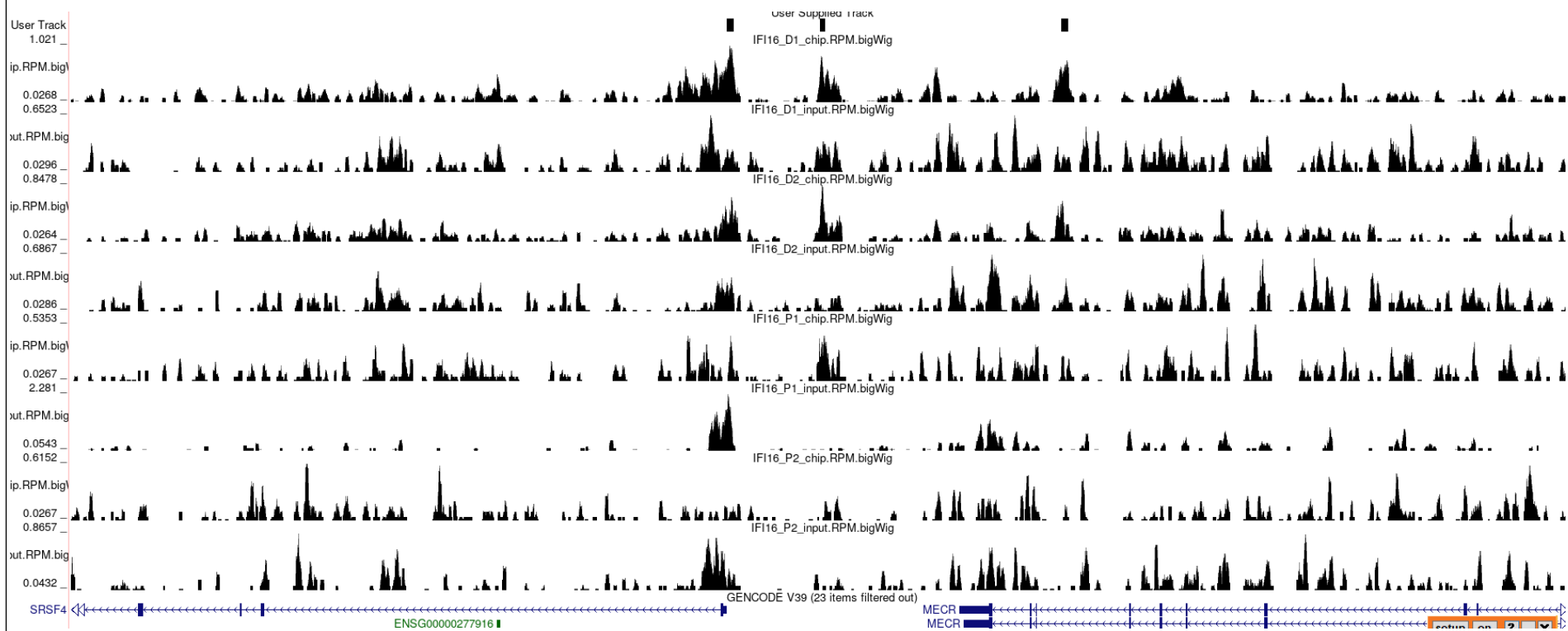
DDX50_H1_chip | DDX50_H1_input | DDX50_H2_chip | DDX50_H2_input | DDX50_L1_chip | DDX50_L1_input



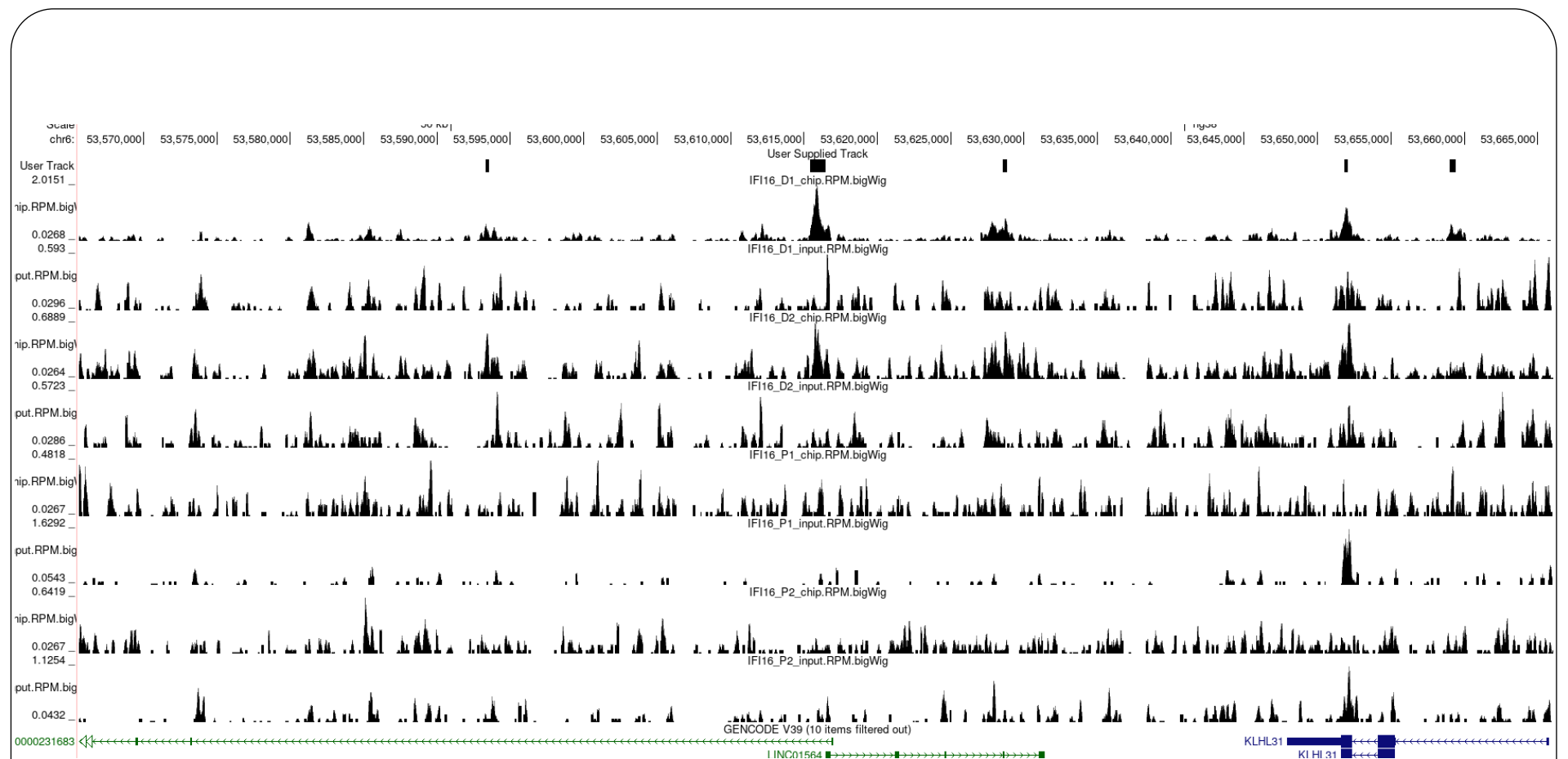
DDX50_L2_chip | DDX50_L2_input | DDX50_P1_chip | DDX50_P1_input | DDX50_P2_chip | DDX50_P2_input

--	--	--	--	--	--

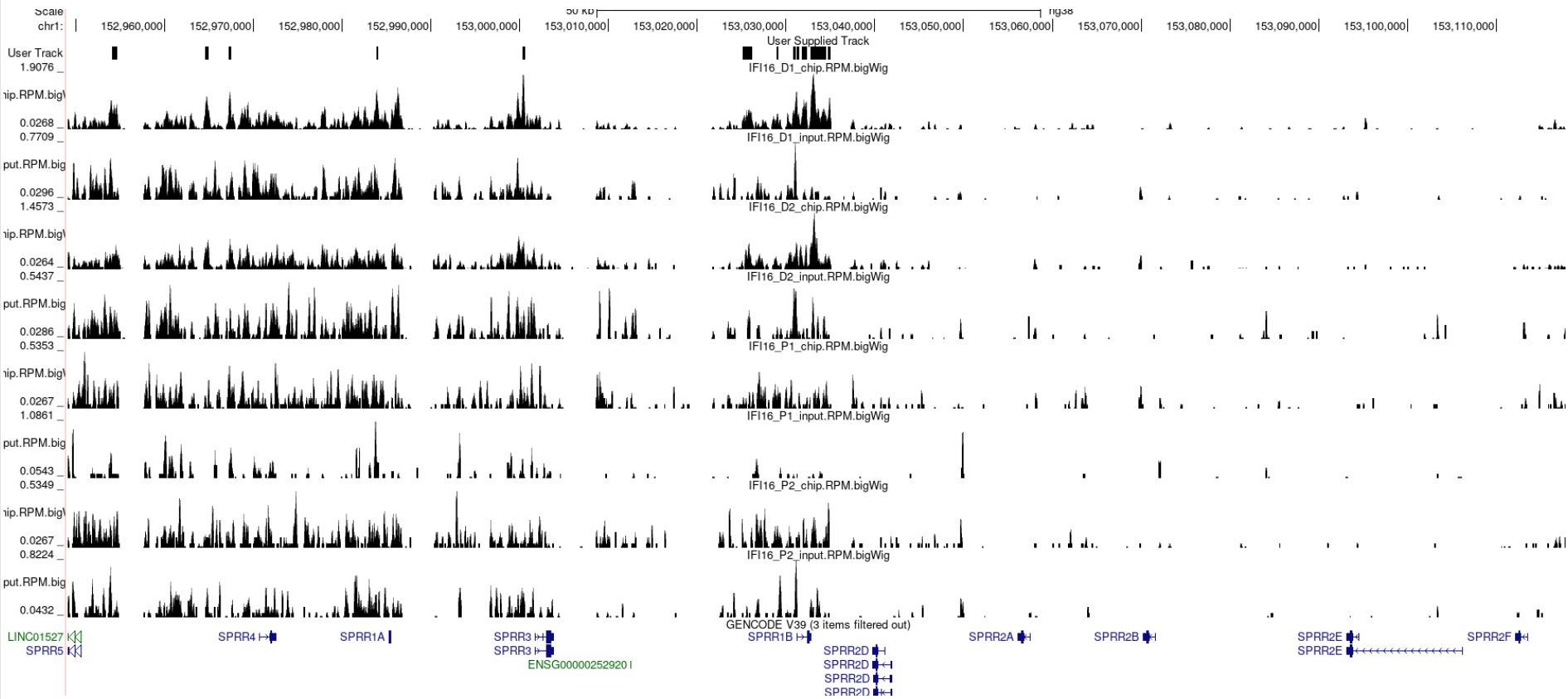
Browser snapshots



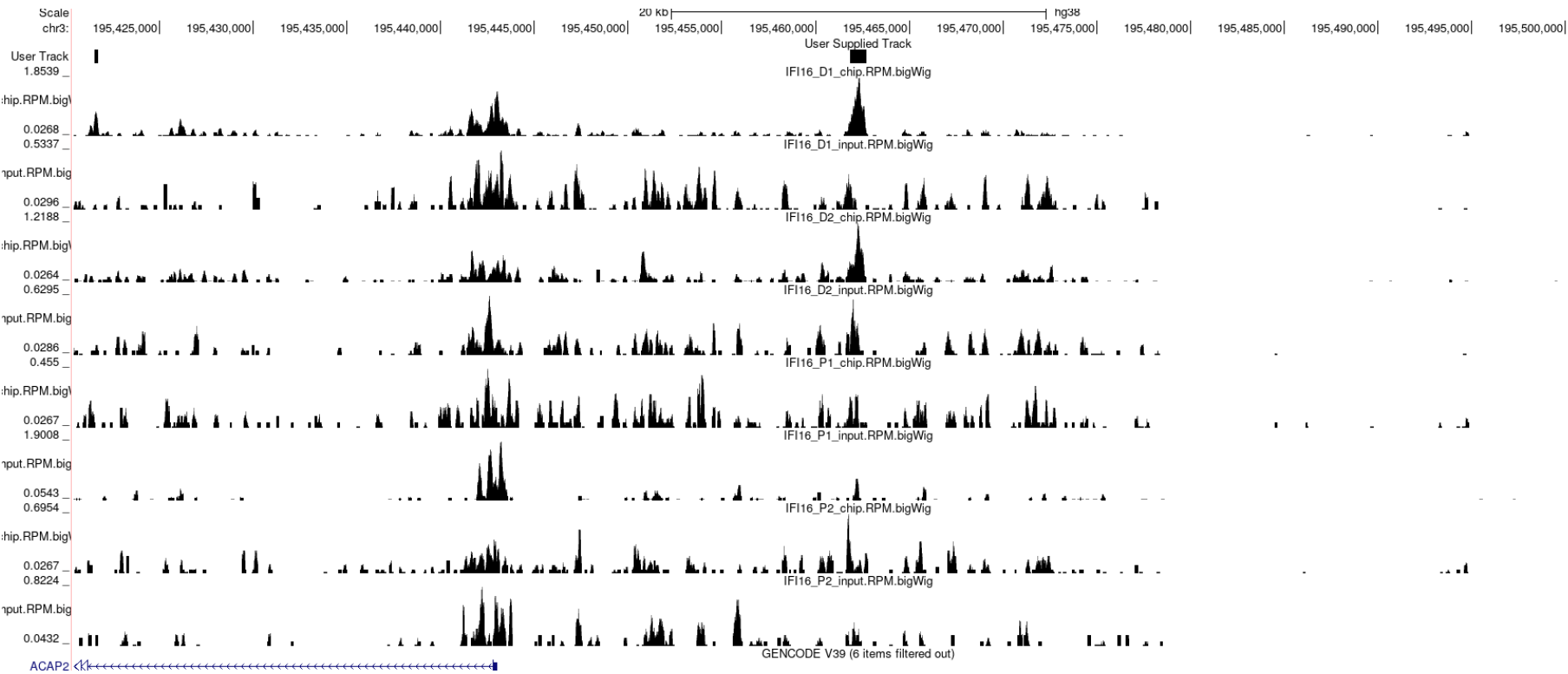
These are random peaks



This is one of the strongest peaks



This is one of the strongest peaks



This is one of the strongest peaks

Summary:

- Inputs are problematic for all datasets – low complexity and clustered enrichment
- DDX50 does not show any enrichment
- For one of the IFI16 datasets, peaks are called, another one is weak. But they are still somewhat difficult to distinguish from background
- On the other hand it is good that some motifs are returned