NGS read mapping : answers to questions

Céline Keime keime@igbmc.fr

Exercise 1 1. Log file

Proportion of uniquely mapped reads :

Started job on		
Started mapping on	Mar 06 10:22:06	
Finished on	Mar 06 10:22:39	searc
Mapping speed, Million of reads per hour	109.09	John
Number of input reads	1000000	NGS da
Average input read length UNIQUE READS:	50	24 show
Uniquely mapped reads number	002800	7.47 G
Uniquely mapped reads %	85.28%	
Average mapped length	49.83	14: RN
Number of splices: Total	137420	<u>4: log</u>
Number of splices: Annotated (sjdb)	136195	
Number of splices: GT/AG	136013	33 line
Number of splices: GC/AG	1157	format
Number of splices: AT/AC	111	Torrita
Number of splices: Non-canonical	139	Mar 0
Mismatch rate per base, %	0.15%	
Deletion rate per base	0.01%	run
Deletion average length	1.60	Mar 0
Insertion rate per base	0.00%	genor
Insertion average length	1.29	Mar 0
MULTI-MAPPING READS:		
Number of reads mapped to multiple loci	133764	mapp
% of reads mapped to multiple loci	13.38%	Mar 0
Number of reads mapped to too many loci	3843	sortin
% of reads mapped to too many loci	0.38%	Mar 0
UNMAPPED READS:		
% of reads unmapped: too many mismatches	0.00%	succe
<pre>% of reads unmapped: too short</pre>	0.73%	
<pre>% of reads unmapped: other</pre>	0.22%	80
CHIMERIC READS:		
Number of chimeric reads	0	Starte
<pre>% of chimeric reads</pre>	0.00%	Starte
		Finish

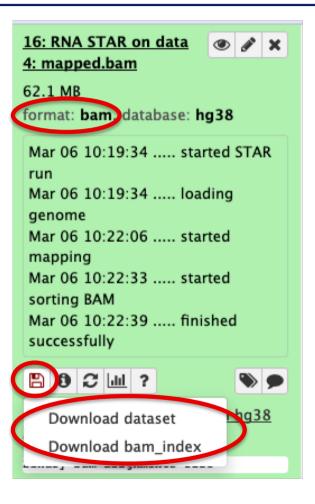


Exercise 1 2. Alignment file

Galaxy

3.1. STAR provides an alignment in BAM format

3.2. Download this file together with the corresponding index (in the same directory)





• File \rightarrow Load from file and choose the downloaded BAM file

Exercise 1 2. Splice junction

Human (hg38)	Chr1 Chr1:7,977,369-7,985,519 Go Go ★
	p36.22 p35.3 p34.1 p32.1 p31.1 p22.1 p13.3 q11 q21.1 q23.2 q25.1 q31.2 q32.2 q42.12 q43
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam Coverage	[0 - 27]
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam Junctions	
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam	chr1:7977738-7984893 Strand: Depth 21 Flanking Widths: (43,44)
Gene	→ ▲ → → → → → → → → → → → → → → → → → → →

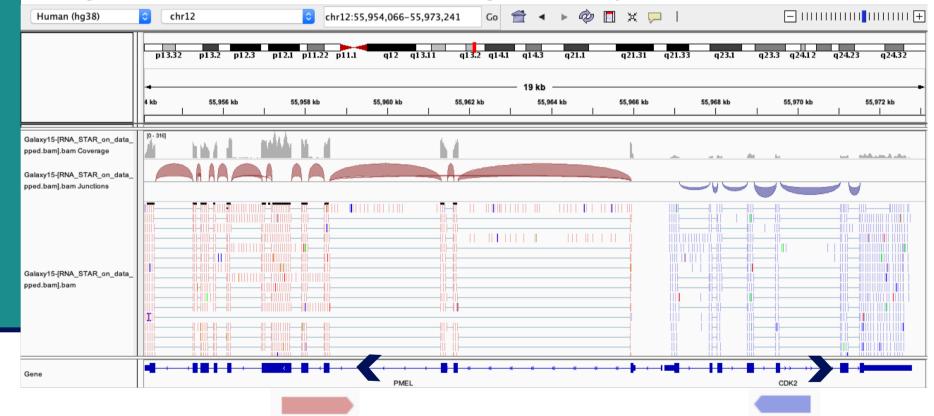
 \rightarrow 21 alignments span the junction that joins the last 2 exons of *Park7* gene

Exercise 1 2. Splice junction

Human (hg38)	
	p36.22 p35.3 p34.1 p32.1 p31.1 p22.1 p13.3 q11 q21.1 q23.2 q25.1 q31.2 q32.2 q42.12 q43
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam Coverage	
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam Junctions	
Galaxy15-[RNA_STAR_on_data_ pped.bam].bam	Hap name: null Dist: 0 Read name = HWI-ST1136:225:H5140:8:1206:5174:59018 Read length = 50bp Mapping = Primary @ MAPQ 255 Reference span = chr1:7,977,696-7,984,900 (-) = 7,205bp Cigar = 43M7155N7M Clipping = None CIGAR : 43M7155N7M
Gene	$\begin{array}{c} & \\ & \\ & \\ & \\ HI = 1 \\ & \\ nM = 0 \\ & \\ AS = 49 \end{array}$, Intron length : 7984893 - 7977738 = 7155
5 tracks loaded	hr1:7,97 Base = A @ QV 41 655M of 1,105M

Exercise 12. Strand specificity

Right click on BAM file \rightarrow Color alignments by \rightarrow read strand

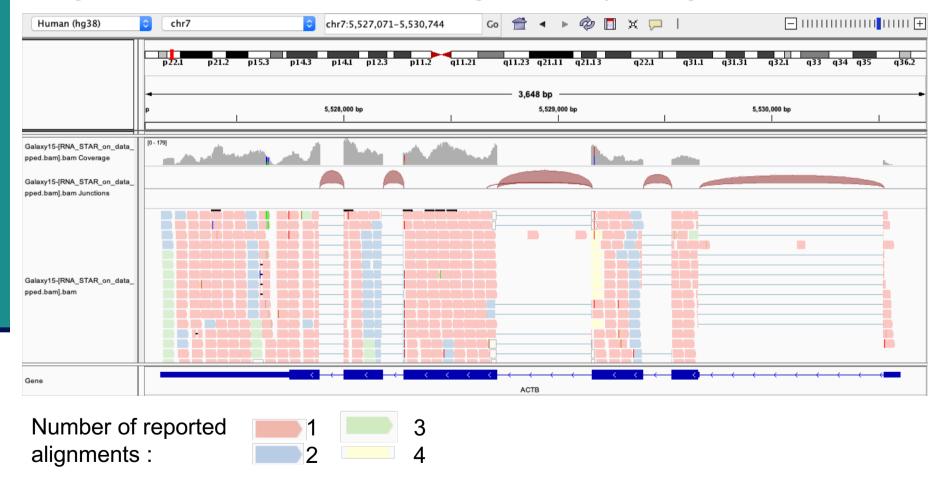


The library has been prepared with a directional mRNAseq protocol which retains strand information :

all reads are in the opposite direction as the transcribed strand

Exercise 12. Multiple mapped reads

Right click on BAM file \rightarrow Color alignments by \rightarrow tag \rightarrow NH



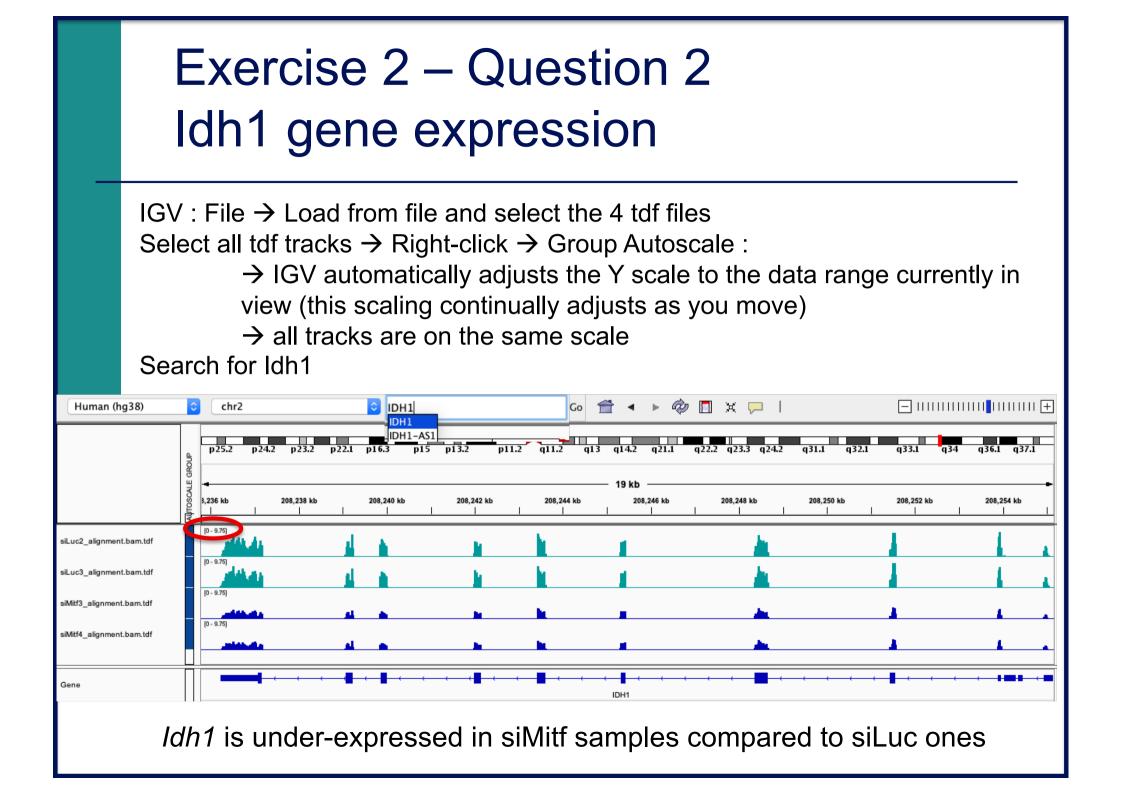
There are multiple aligned reads on this gene

Exercise 2 - Question 1 Proportion of uniquely mapped reads

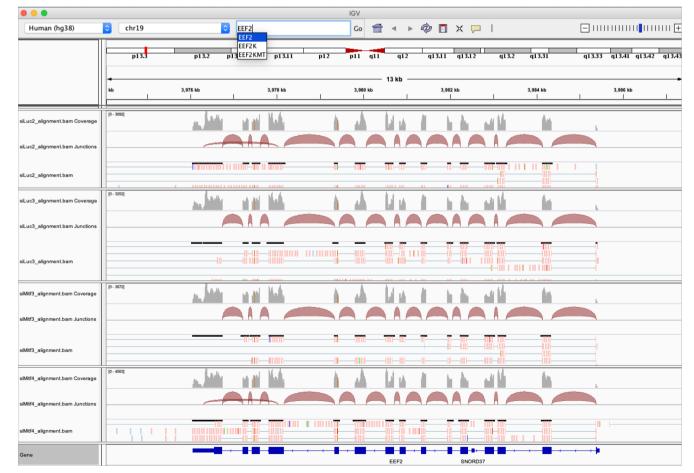
Galaxy : Shared Data \rightarrow Data Libraries \rightarrow NGS data analysis training RNAseq \rightarrow alignment \rightarrow log files :

st	Started job on tarted mapping on	Mar 05 11 Mar 05 11			History	2 🌣 🗆
Mapping speed, Million of	Finished on	Mar 05 11 123.41	:53:07		search datasets	8
	er of input reads input read length UNIQUE READS:	43672265 50			NGS data analysis tra 24 shown, 5 <u>deleted</u>	aining - RNAseq
Unique	pped reads number ly mapped reads %	85.30%			7.47 GB	S D
Number o	age mapped length of splices: Total	6001725			8: STAR on siLuc2: I	
Number of splices: Number of	Annotated (sjdb) of splices: GT/AG	5948001 5938121			33 lines	
	of splices: GC/AG	51849				
	of splices: AT/AC	6383			format: txt , database	a: hg38
	es: Non-canonical rate per base, %	5372 0.15%			Mar 05 11:30:25	started STAP
	ion rate per base	0.15%			mai 05 11.50.25	. Starteu STAR
	on average length	1.58				loading
	ion rate per base	0.00%			Mar 05 11:30:25	. loading
	on average length JLTI-MAPPING READS:	1.29			genome	
Number of reads mapped		5836055			Mar 05 11:31:53	. started
% of reads mapped		13.36%			mapping	
Number of reads mapped		167816			Mar 05 11:50:18	. started
% of reads mapped	to too many loci UNMAPPED READS:	0.38%			sorting BAM	
% of reads unmapped: too		0.00%			Mar 05 11:53:07	. finished
	mapped: too short	0.73%			successfully	
	s unmapped: other CHIMERIC READS:	0.22%			802?	۰ ا
	of chimeric reads	0 0.00%				
5 (of chimeric reads	0.00%			Started job on I	Mar 05 11:3
STAR on siLuc2:	Uniquely	mapped	reads	8	8	5.30%
STAR on siLuc3:	Uniquely	mapped	reads	8	8	5.72%
STAR on siMitf3:	Uniquely	mapped	reads	8	8	5.41%

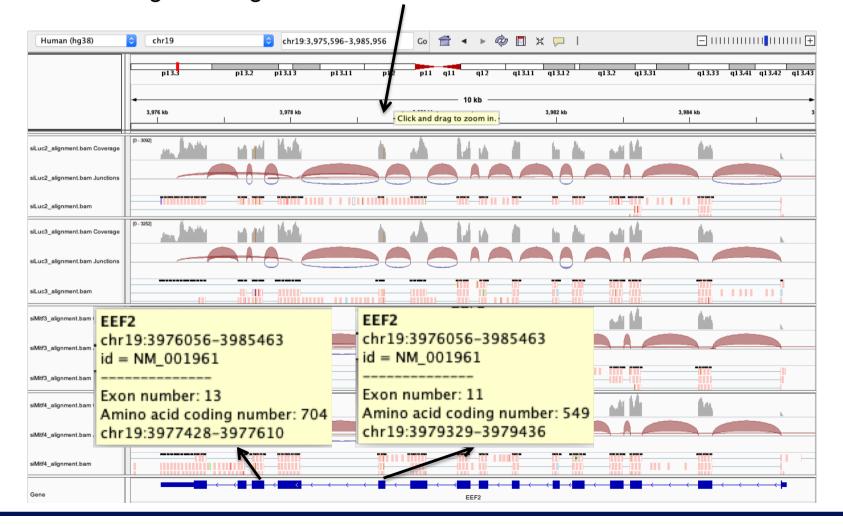
 \rightarrow This proportion is consistent across samples



- File \rightarrow new session
- File \rightarrow load from files and load the 4 BAM files
- Search for EEF2



Exon numbers are provided on annotation track Click and drag on a region to zoom in



■ *Eef*2 exon 11

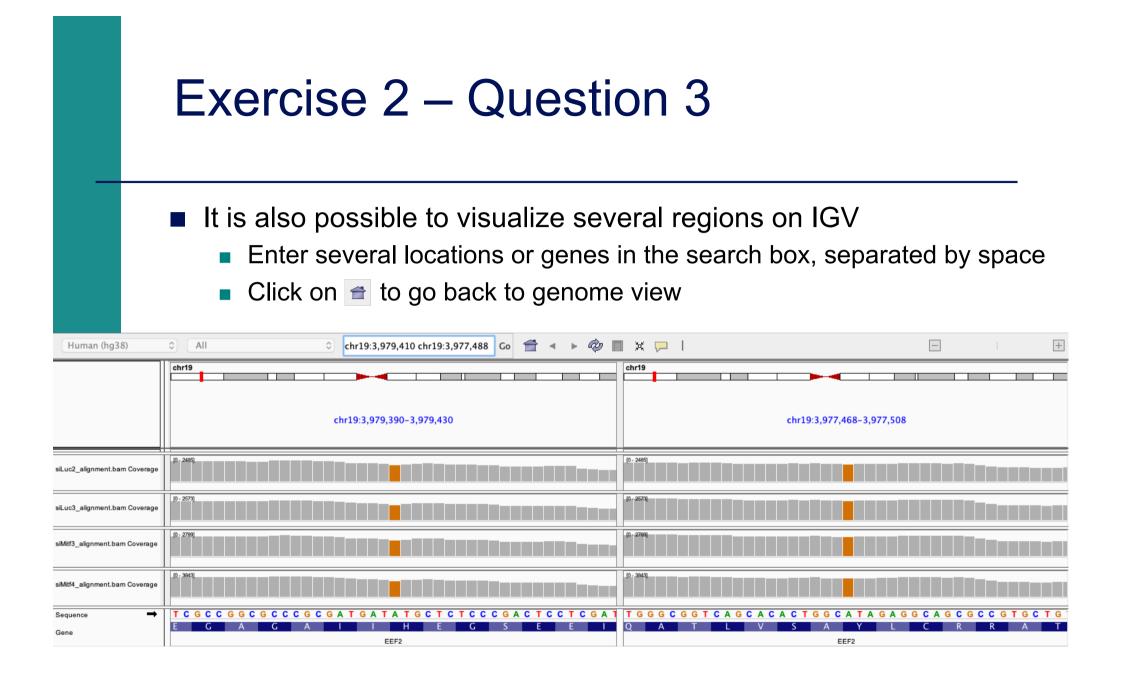
chr19:3,979,410 : G in ~100% of the reads, A in the genome

Human (hg38)	chr19 chr19:3,979,325-3,979,446 Go
	pi3.3 pi3.2 pi3.13 pi3.11 pi2 pi1 q11 q12 q13.11 q13.12 q13.2 q13.31 q13.33 q13.41 q13.42 q13.43
	- 122 bp
siLuc2_alignment.bam Coverage	
siLuc2_alignment.bam Junctions	chr19:3,979,410 Total count: 1521
siLuc2_alignment.bam	A : 3 (0%, 3+, 0-) C:0 G: 1518 (100%, 1516+, 2-) T:0
siLuc3_alignment.bam Coverage	
siLuc3_alignment.bam Junctions	
siLuc3_alignment.bam	
siMitf3_alignment.bam Coverage	
siMitf3_alignment.bam Junctions	
siMitf3_alignment.bam	G
siMitf4_alignment.bam Coverage	
siMitf4_alignment.bam Junctions	
siMitf4_alignment.bam	
Sequence →	TCACCTTGATGGGGATGCAGGCGTGGTCCTCCCAGGTCCTCCAGGCCCGGCGCGCGC

■ *Eef2* exon 13

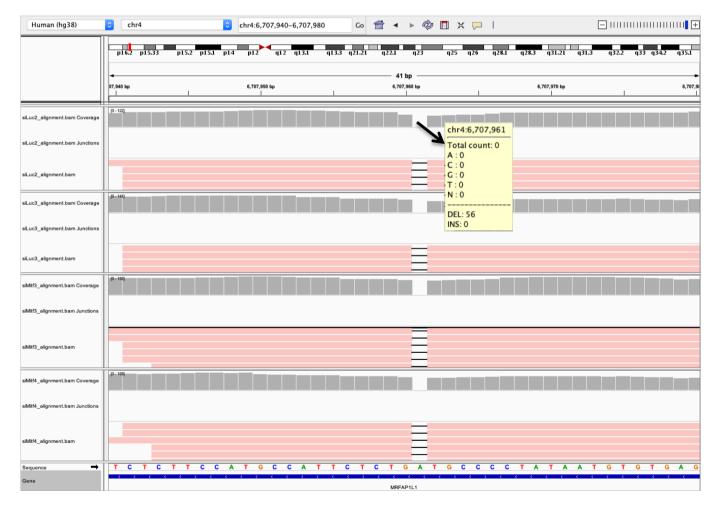
chr19:3,977,488 : G in ~100% of the reads, A in the genome





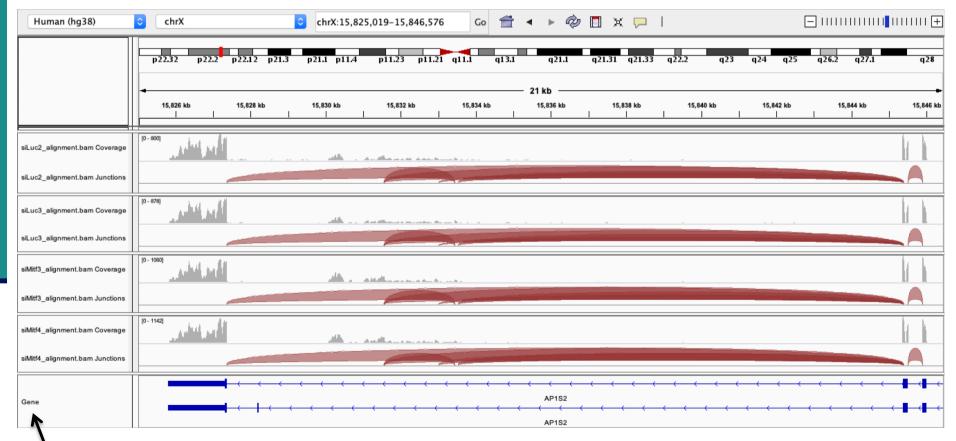
Position chr4:6707960-6707961 :

Deletion vs reference genome



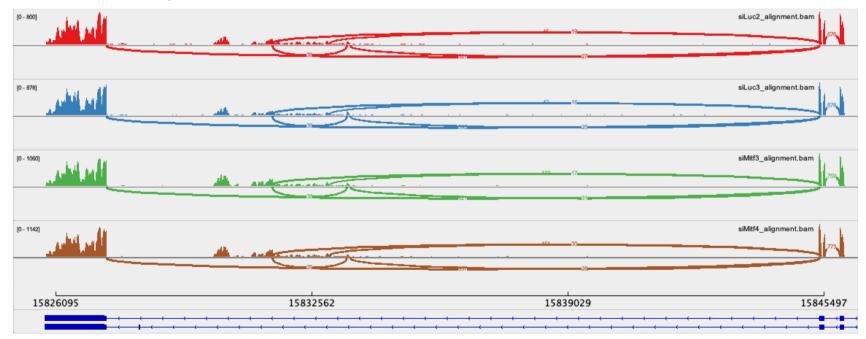
Region chrX:15,825,019-15,846,576 :

Junctions corresponding to exons not annotated in Refseq



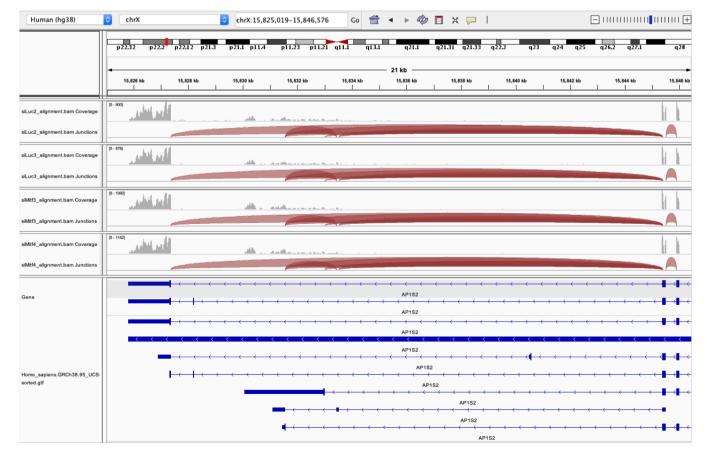
Right click on the annotation track and select Expanded to visualize all isoforms

- Region chrX:15,825,019-15,846,576 :
 - Junctions corresponding to exons not annotated in Refseq
 - Sashimi-plot :
 - Right-click on a BAM track → Sashimi plot → Select Alignment Tracks : all alignments



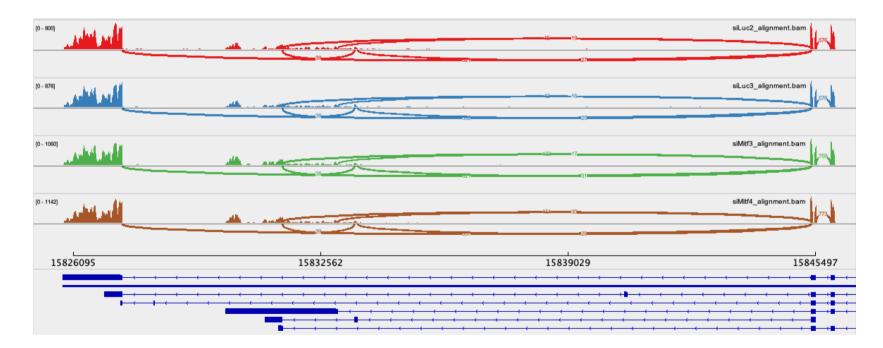
Region chrX:15,825,019-15,846,576 :

- Ensembl annotations : more exons annotated in this region
 - File → load from file → Homo_sapiens.GRCh38.95_UCSC_chr.sorted.gtf
 - Right-click on the annotation track and select Expanded

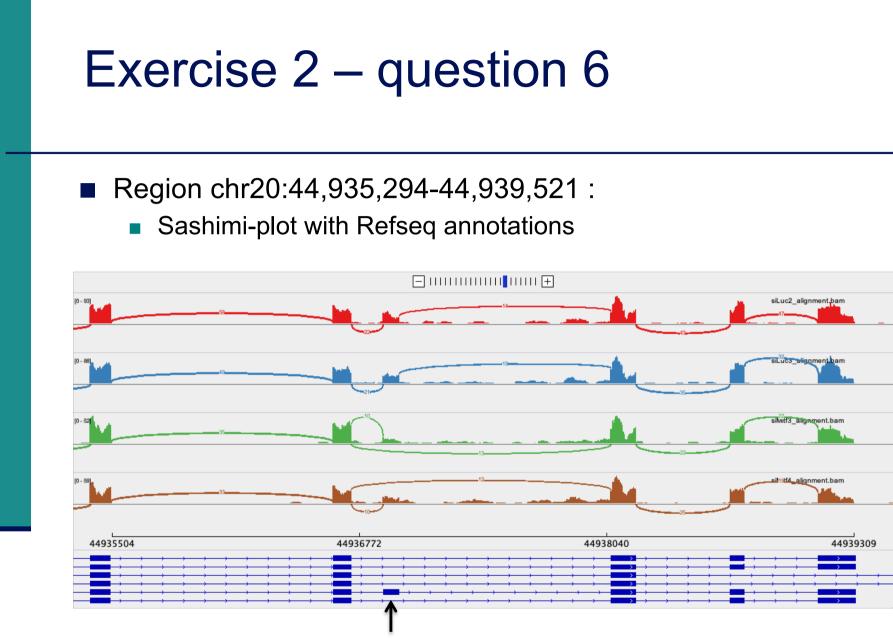


Region chrX:15,825,019-15,846,576 :

- Ensembl annotations :
 - Sashimi-plot : Right-click on a BAM track → Sashimi plot → Select Gene Track : Ensembl annotations → Select Alignment Tracks : all alignments



→ Very useful to quickly visualize splicing events along genomic regions of interest
 → More accurate with paired-end data



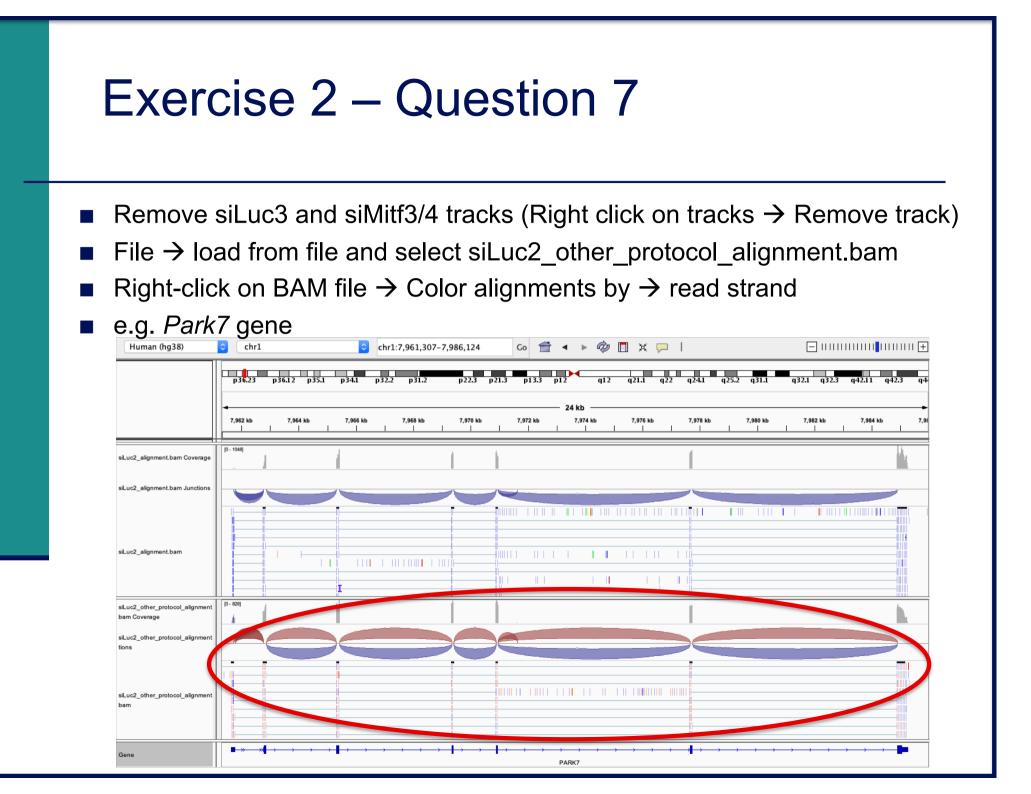
We detect an isoform without this exon in siMitf samples but not in siLuc ones

IGV is only a visualization tool

In-depth analysis using paired-end data with more coverage is needed

	Exercise 2 – question 6
	 If you want to save this region : Click on define a region of interest
	● ● ● Human (hg38)
	 Click on a track to define the start and end position of your region of interest. Interest -> a red bar appears
	 Give a name to this region (Right-click on the bar → edit description) Go to Regions → Region Navigator to display again this region
Human (hg38)	chr20 chr20:44,935,294-44,939,521 Go
	p13 p12.3 p12.2 p12.1 p11.23 p11.21 p11.1 q11.1 q11.22 q11.23 q12 q13.11 q13.12 q13.13 q13.2 q13.31 q13.33
	44,936,000 bp 44,937,000 bp 44,937,000 bp 44,939,000 bp
siLuc2_alignment.bam Coverag	e D-ssi Edit description
siLuc2_alignment.bam Junction siLuc2_alignment.bam	Copy sequence Blat sequence Delete

- You can save your IGV session
 - To save the current state of your IGV session to a named session file
 - File \rightarrow Save Session
 - Data files must stay at the same location
- Use File \rightarrow Open session to restore a saved session



	p36.23 p36.12 p35.1 p34.1 p32.2 p31.2 p22.3 p21.3 p13.3 p12 q12 q21.1 q22 q24.1 q25.2 q31.1 q32.1 q32.3 q42.11 q42.3 q4
siLuc2_alignment.bam Coverage	
siLuc2_alignment.bam Junctions	
siLuc2_alignment.bam	
siLuc2_other_protocol_alignment bam Coverage	0 - 824
siLuc2_other_protocol_alignment tions	
siLuc2_other_protocol_alignment bam	
Sequence →	$\begin{array}{c} \\ \hline \\ $

 \rightarrow This protocol is not directional (it does not preserve strand information)

You can display alignments grouped by read strand (right-click on BAM track \rightarrow Group alignments by \rightarrow read strand)