Theory: SAM/BAM files

deNBI Course
March 2019

(lectured by Martina Fröhlich, slides based on lecture by Barbara Hutter)

Alignment Output Format

- Storing alignments of fragments (reads) against reference sequences
- Current standard: **SAM** (Sequence Alignment/Map) format
  - tab-delimited text format
  - information about:
    - order and lengths of the reference sequences
    - **the alignment of each read** (fragment):
      - Start position
      - Orientation
      - Alignment quality

- quality scores preserved
- for paired end reads, information on the mate included
  - location
  - insert size or "observed template length"
    - Length of the genomic segment covered by the complete fragment
    - Number of bases from the leftmost mapped base to the rightmost mapped base
- **BAM** = binary version of SAM

Header of a SAM file

```
@HD    VN:1.0  GO:none SO:coordinate
@SQ    SN:chr1 LN:249250621
@SQ    SN:chr10        LN:135534747
@SQ    SN:chr11        LN:135006516
@SQ    SN:chr12        LN:133851895
@SQ    SN:chr13        LN:115169878
@SQ    SN:chr14        LN:107349540
@SQ    SN:chr15        LN:102531392
@SQ    SN:chr16        LN:90354753
@SQ    SN:chr17        LN:81195210
@SQ    SN:chr18        LN:78077248
@SQ    SN:chr19        LN:59128983
@SQ    SN:chr2 LN:243199373
@SQ    SN:chr20        LN:63025520
@SQ    SN:chr21        LN:48129895
@SQ    SN:chr22        LN:51304566
@SQ    SN:chr3 LN:198022430
@SQ    SN:chr4 LN:191154276
@SQ    SN:chr5 LN:180915260
@SQ    SN:chr6 LN:171115067
@SQ    SN:chr7 LN:159138663
@SQ    SN:chr8 LN:146364022
@SQ    SN:chr9 LN:141213431
@SQ    SN:chrM LN:16571
@SQ    SN:chrX LN:155270560
@SQ    SN:chrY LN:59373566
@RG    ID:111116_SN952_0061_AD0C8YACXX_L007    PL:ILLUMINA    LB:MB51_MBBL51
SM:sample_MB51_MBBL51
@RG    ID:111125_SN841_0105_BD0FVHACXX_L005    PL:ILLUMINA    LB:MB51_MBBL51
SM:sample_MB51_MBBL51
@RG    ID:111125_SN841_0105_BD0FVHACXX_L006    PL:ILLUMINA    LB:MB51_MBBL51
SM:sample_MB51_MBBL51
@RG    ID:120111_SN509_0136_AD0C9JACXX_L005    PL:ILLUMINA    LB:MB51_MBBL51
SM:sample_MB51_MBBL51
@PG    ID:bwa  PN:bwa  VN:0.5.9-r16
```

@HD: general information
SO = sort order
@SQ: reference genome information
chromosomes, their length
@RG: read group information
@PG: program used for alignment
(and further processing)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

coordinates of “spot”: identifier:**lane**:tile:x:y

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

alignment coordinates (here: chr1, start at pos. 9996)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

http://picard.sourceforge.net/explain-flags.html

SAM format flag field

Bit in HexCode	Description	numeric
0x0001	paired in sequencing	1
0x0002	mapped in a proper pair (normally Inferred during alignment form the insert size distribution)	2
0x0004	read (query sequence) itself is unmapped	4
0x0008	mate is unmapped	8
0x0010	strand of the query (1 for reverse)	16
0x0020	strand of the mate (1 for reverse)	32
0x0040	read is the first read in a pair	64
0x0080	read is the second read in a pair	128
0x0100	alignment is not primary	256
0x0200	QC failure	512
0x0400	optical or PCR duplicate	1024

- A flag is set if the respective property applies
- The values are combined (i.e. numbers added up) for the FLAG field

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

flag (here 153 = 128 + 16 + 8 + 1 = second read of pair, maps to reverse strand, mate unmapped
=> BWA assigns unmapped reads the mate's Coordinates), paired in sequencing

http://picard.sourceforge.net/explain-flags.html

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

mapping quality

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

Extended CIGAR String

- M: alignment match (can be a *sequence* mismatch! Mismatches see NM tag)
- I: insertion with respect to reference
- D: deletion with respect to reference
- N: spliced alignment

  representing intronic sequences, “reference skip” instead of deletions (RNA-Seq)
- Clipped alignment
  - **S** softclipped

  subsequences at ends that are not aligned
    - bases with low quality
    - part of the read maps somewhere else (but location of its mate indicates that it should map nearby)
  - **H** hardclipped
    - originally multi-part (chimeric) alignment
    - Shortening the read in order to map part of it if full read does not map
    - Hardclipped bases are not present in the alignment record any more
    - Supplementary alignment

Example: 6M1I94M

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

CIGAR string (here: 6 matches, 1 insertion, 94 matches)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

insert size (here: 0 because mate unmapped)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

sequence (here: original has been reverse complemented)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

base qualities (here: reversed)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

read group (here: run_lane)

Interpretation of fields and flags

Example for a BWA mapped read

HWI-ST841:105:D0FVHACXX:**6**:1102:19098:45810 153 chr1 9996 37 6M1I94M = 9996 0
TCCGATCTACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAGCCCTAACCCTAACCCTAACC BBACCC?9+?98+(,A9<<?5A<=5;5?;>A>.BB??
A=FDGIHDC>=FCFJHGGB8@@AD90GF??:?HGCC:1AEEA<+HIGD?CC;A:+2DDD=D=?: RG:Z:
111125_SN841_0105_BD0FVHACXX_L006 XT:A:U NM:i:3 XN:i:5 SM:i:37 AM:i:0 X0:i:1
X1:i:0 XM:i:2 XO:i:1 XG:i:1 MD:Z:6A72A20

MD tag (here: bases at position 7 and 73 are A in the reference genome, but different in the read)

IGV

deNBI Course
March 2019

Lectured by Martina Fröhlich,
some slides are based on lectures by Barbara Hutter and Benedikt Brors

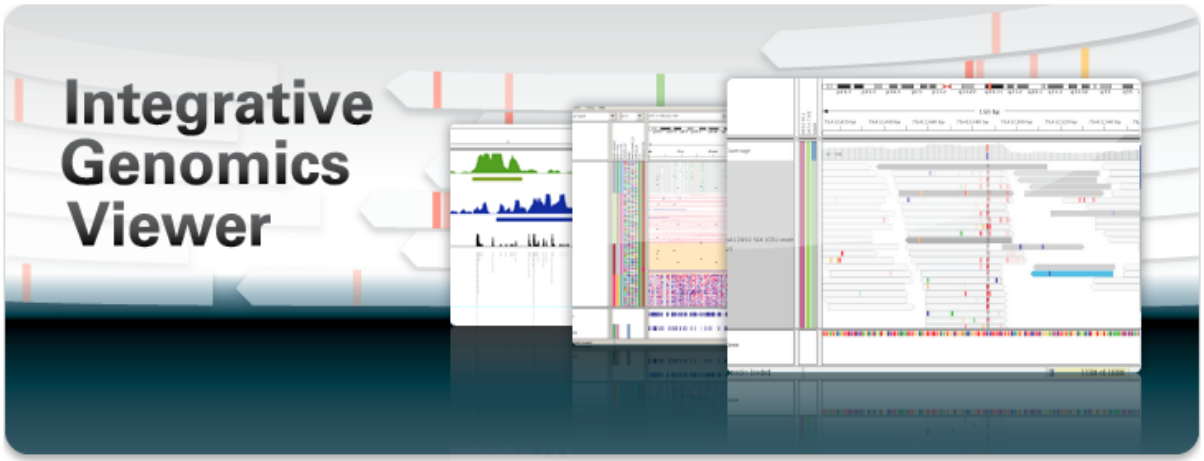
Integrative Genomics Viewer

Home
Downloads
Documents

Hosted Genomes
FAQ
IGV User Guide
File Formats
Release Notes
Credits
Contact

Search website
search

Home



Overview

The **Integrative Genomics Viewer (IGV)** is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.

Funding

Development of IGV has been supported by funding from the National Cancer Institute (NCI) of the National Institutes of Health, the Informatics Technology for Cancer Reserarch (ITCR) of the NCI, and the Starr Cancer Consortium.

IGV participates in the GenomeSpace initiative, which is funded by the National Human Genome Research Institute.

Downloads

Download the IGV desktop application and igvtools.

Citing IGV

To cite your use of IGV in your publication:

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. **Integrative Genomics Viewer**. Nature Biotechnology 29, 24–26 (2011)

Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. Briefings in Bioinformatics 14, 178-192 (2013).

Example dataset

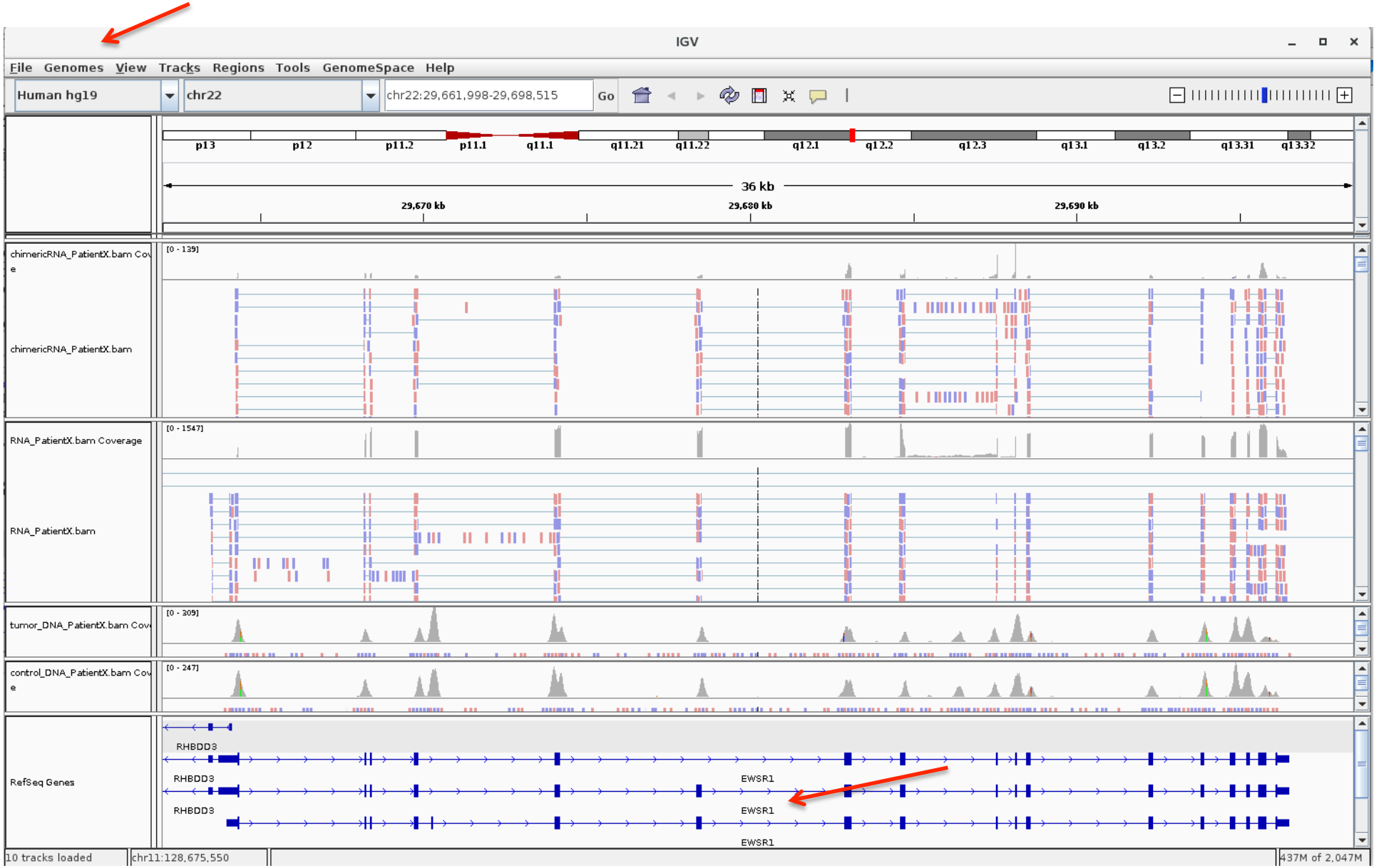
Chimeric RNA
RNA
Tumor DNA
Control DNA (matching control)

Used for aligning the reads:

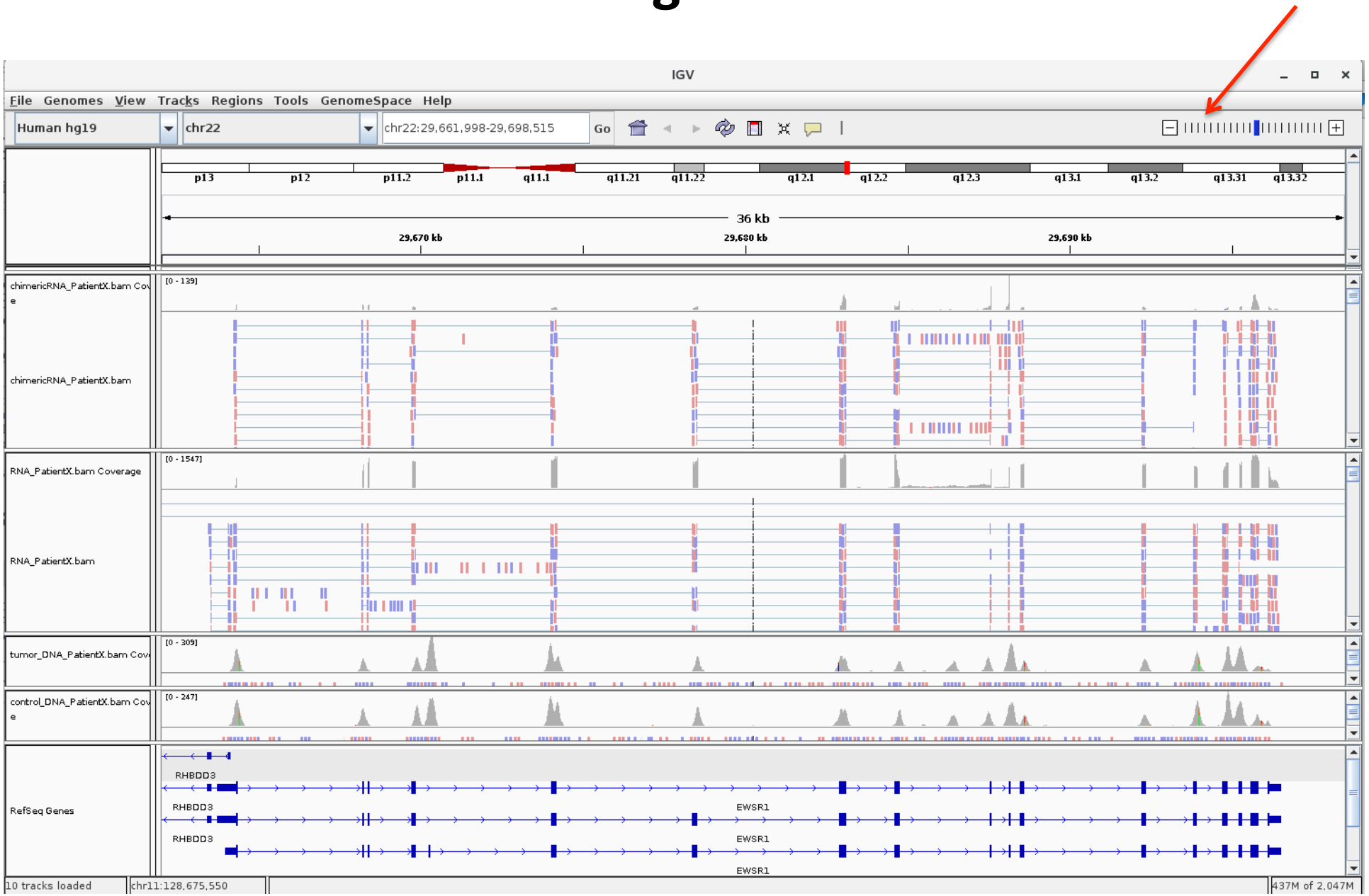
DNA: BWA align or BWA mem
RNA: STAR (https://github.com/alexdobin/STAR)

Getting started

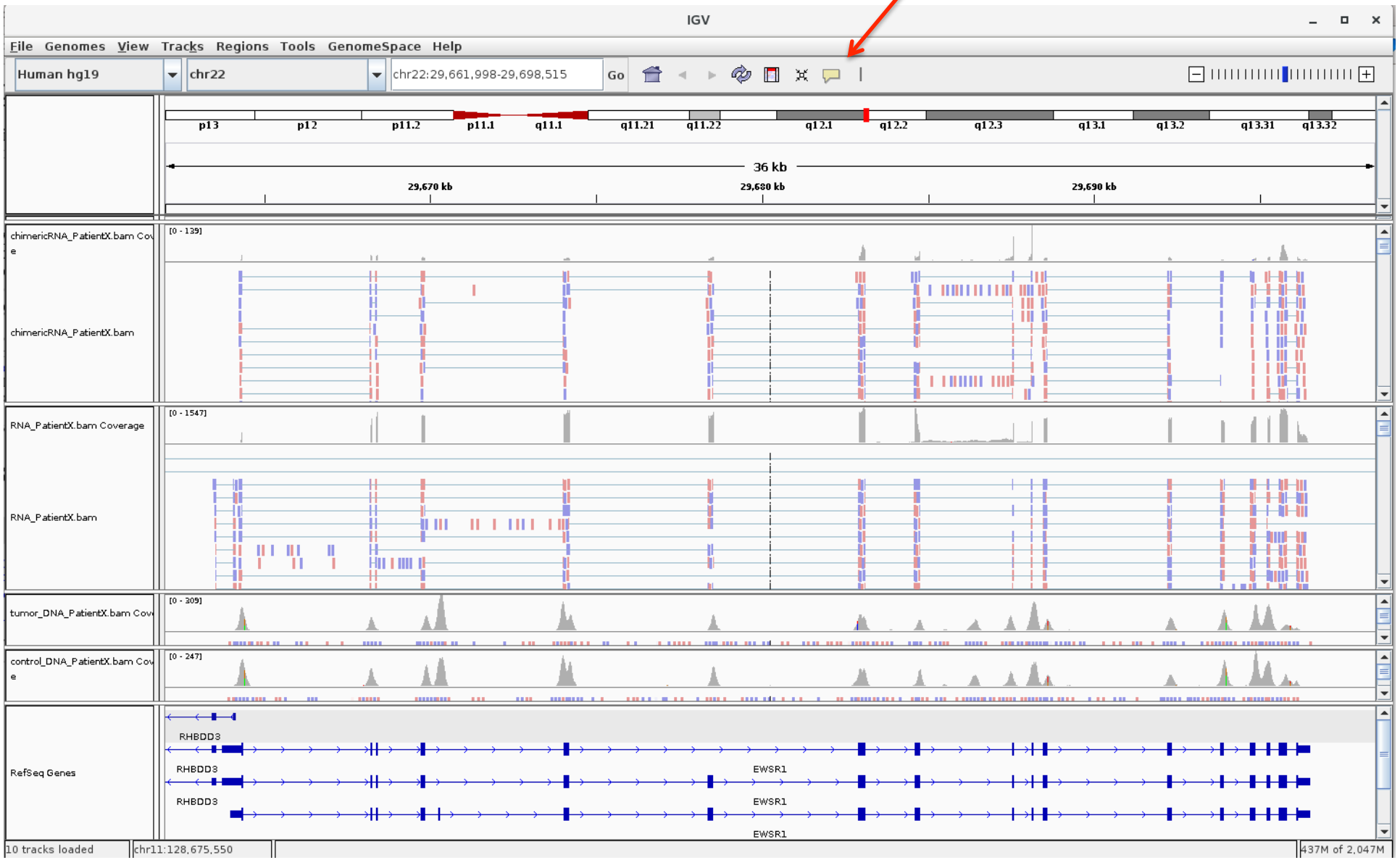
Getting started



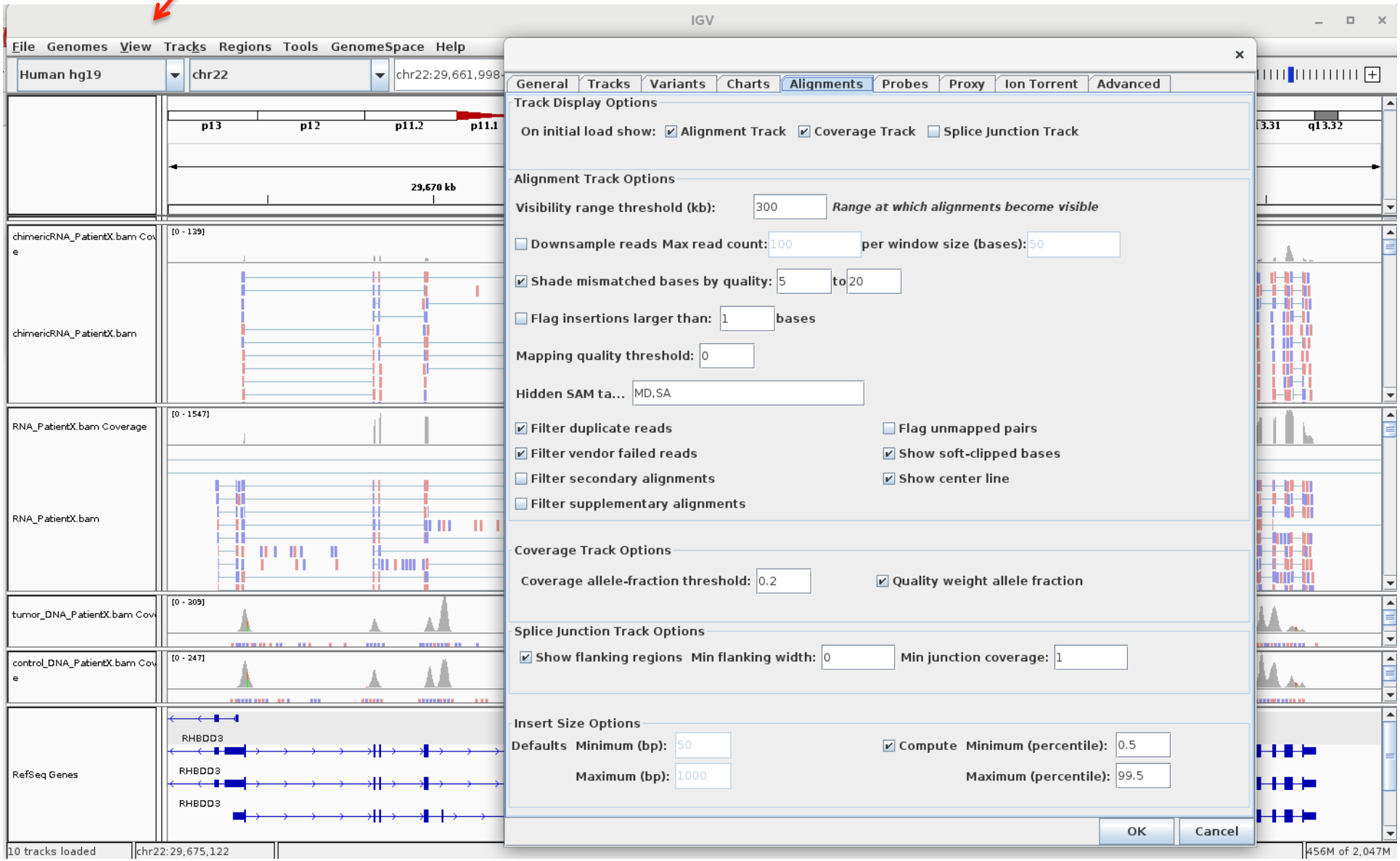
Getting started



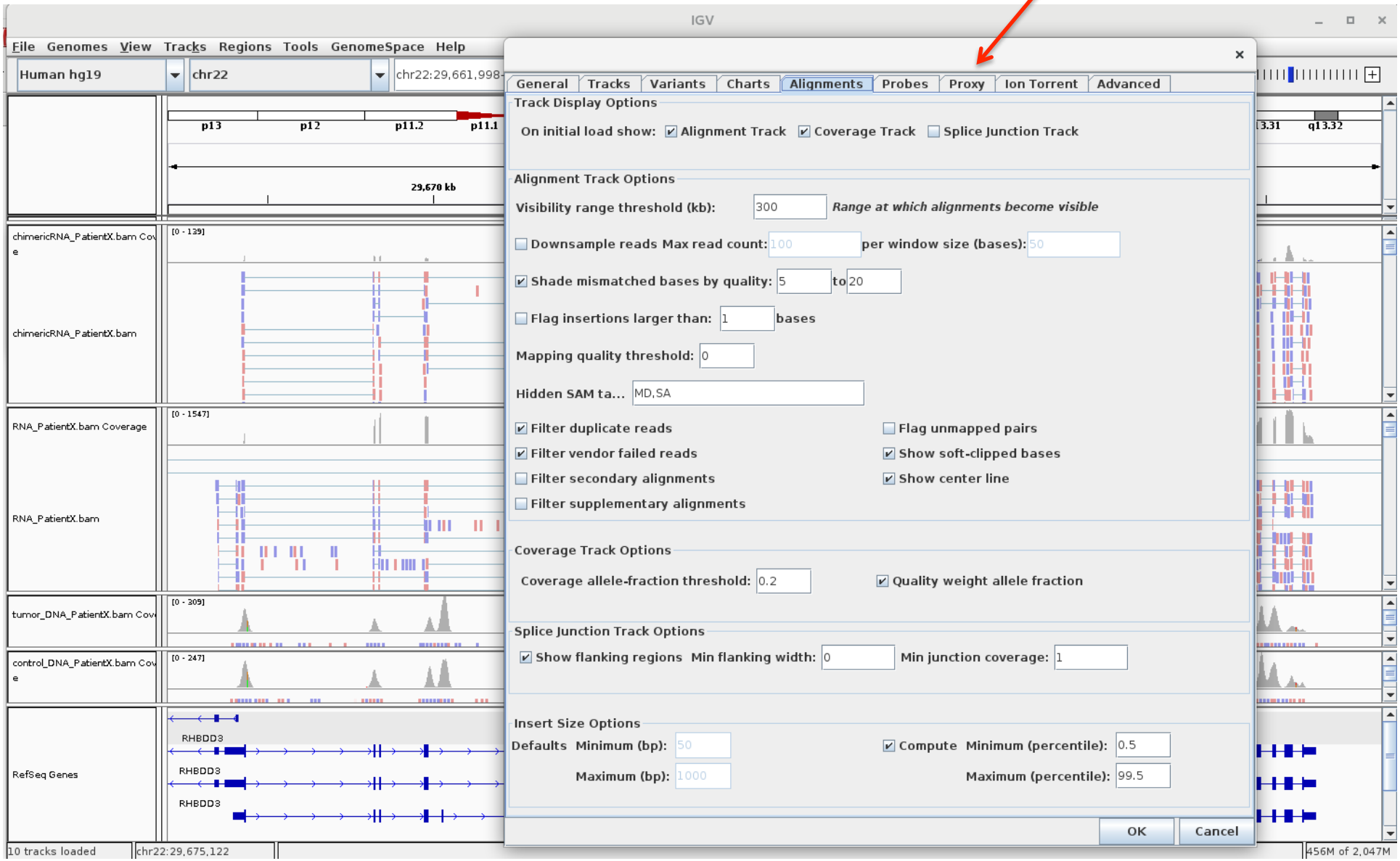
Getting started



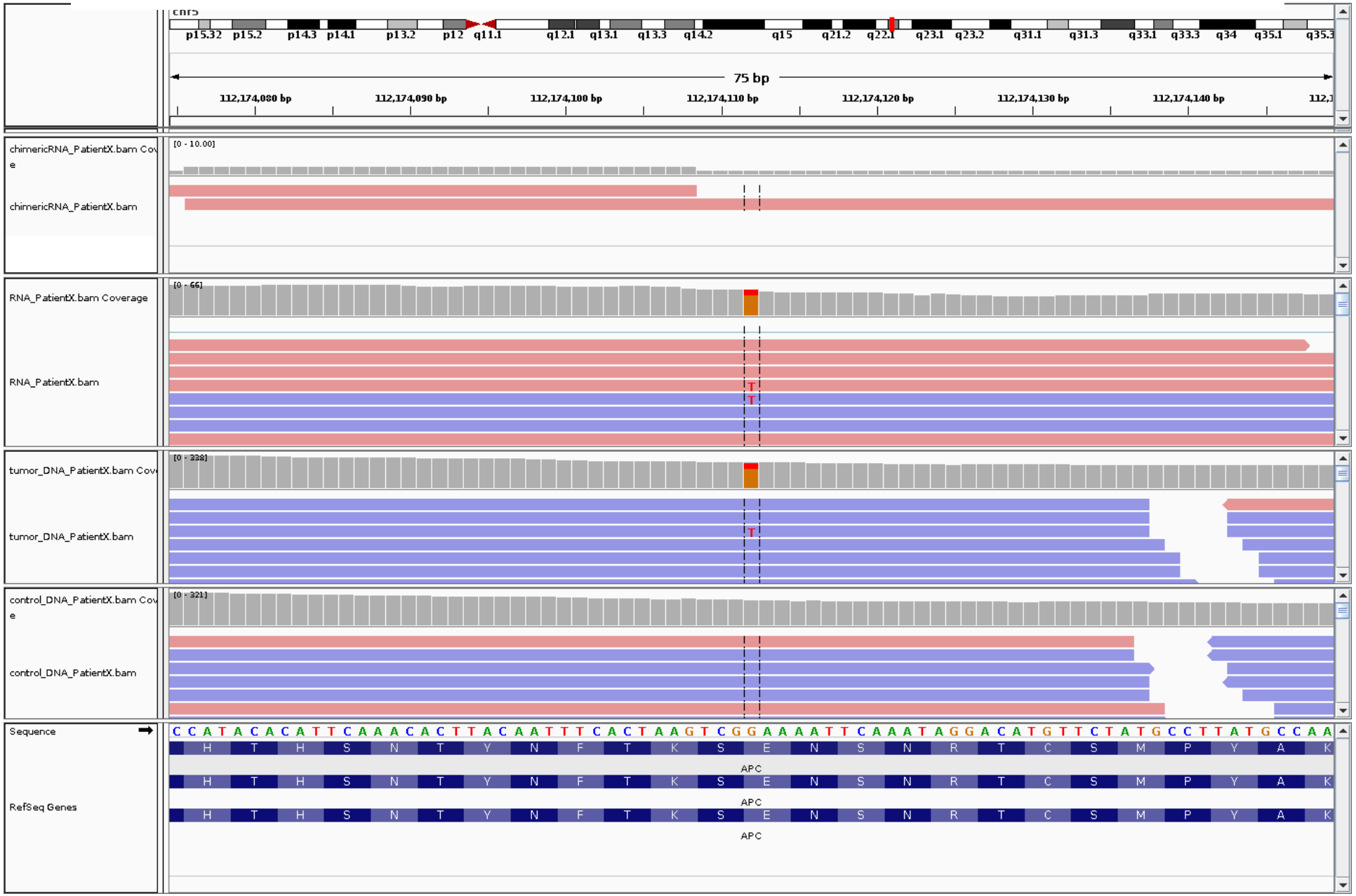
Getting started



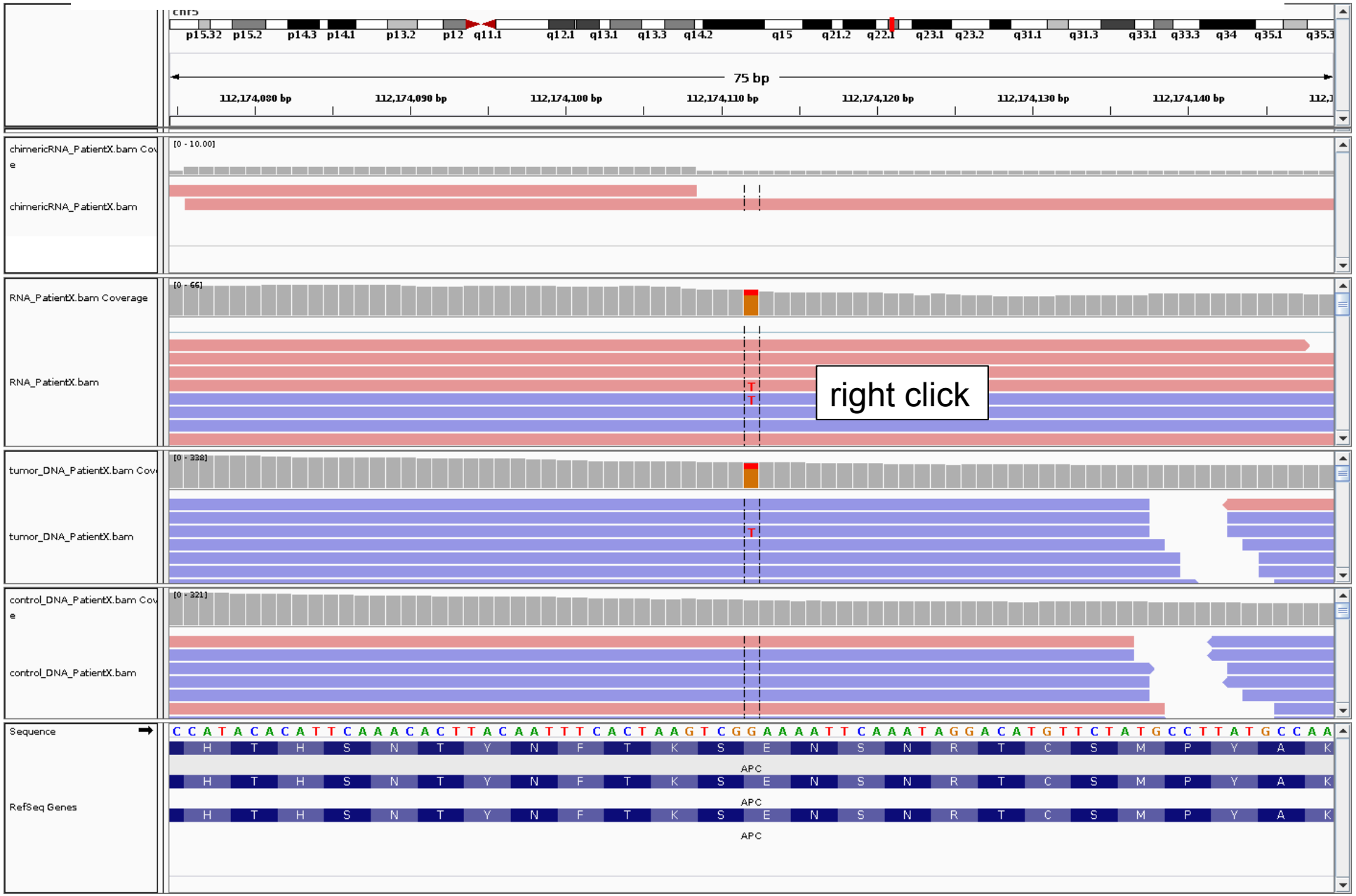
Getting started



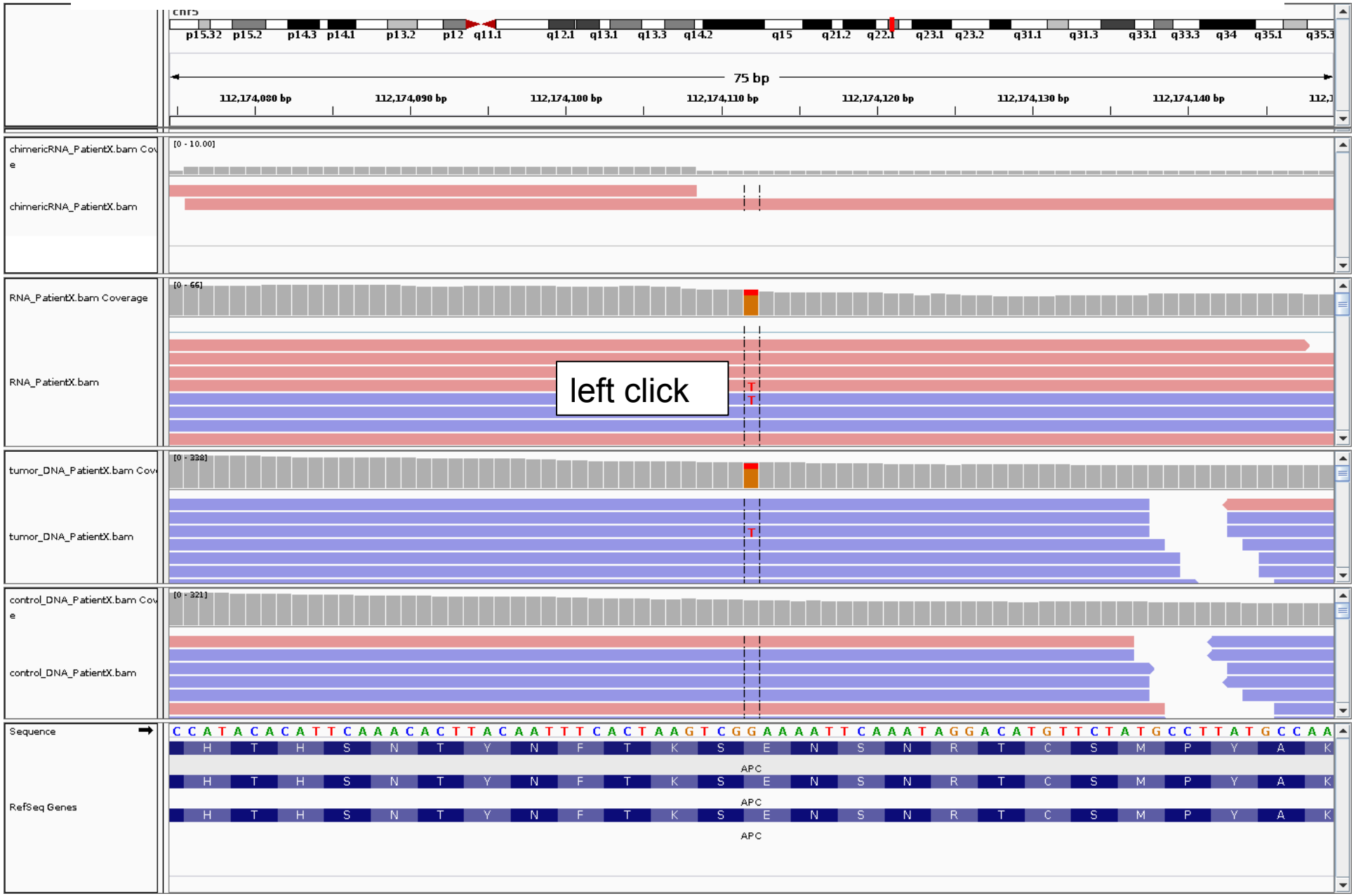
Somatic SNV


chr5
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q12.1 q13.1 q13.3 q14.2 q15 q21.2 q22.1 q23.1 q23.2 q31.1 q31.3 q33.1 q33.3 q34 q35.1 q35.3
75 bp
112,174,080 bp
112,174,090 bp
112,174,100 bp
112,174,110 bp
112,174,120 bp
112,174,130 bp
112,174,140 bp
chimericRNA_PatientX.bam Coverage
[0 - 10.00]
chimericRNA_PatientX.bam
RNA_PatientX.bam Coverage
[0 - 66]
RNA_PatientX.bam
T
T
tumor_DNA_PatientX.bam Coverage
[0 - 338]
tumor_DNA_PatientX.bam
T
control_DNA_PatientX.bam Coverage
[0 - 321]
control_DNA_PatientX.bam
Sequence
C C A T A C A C A T T C A A A C A C T T A C A A T T T C A C T A A G T C G G A A A A T T C A A A T A G G A C A T G T T C T A T G C C T T A T G C C A A
H T H S N T Y N F T K S E N S N R T C S M P Y A K
APC
RefSeq Genes
APC
APC

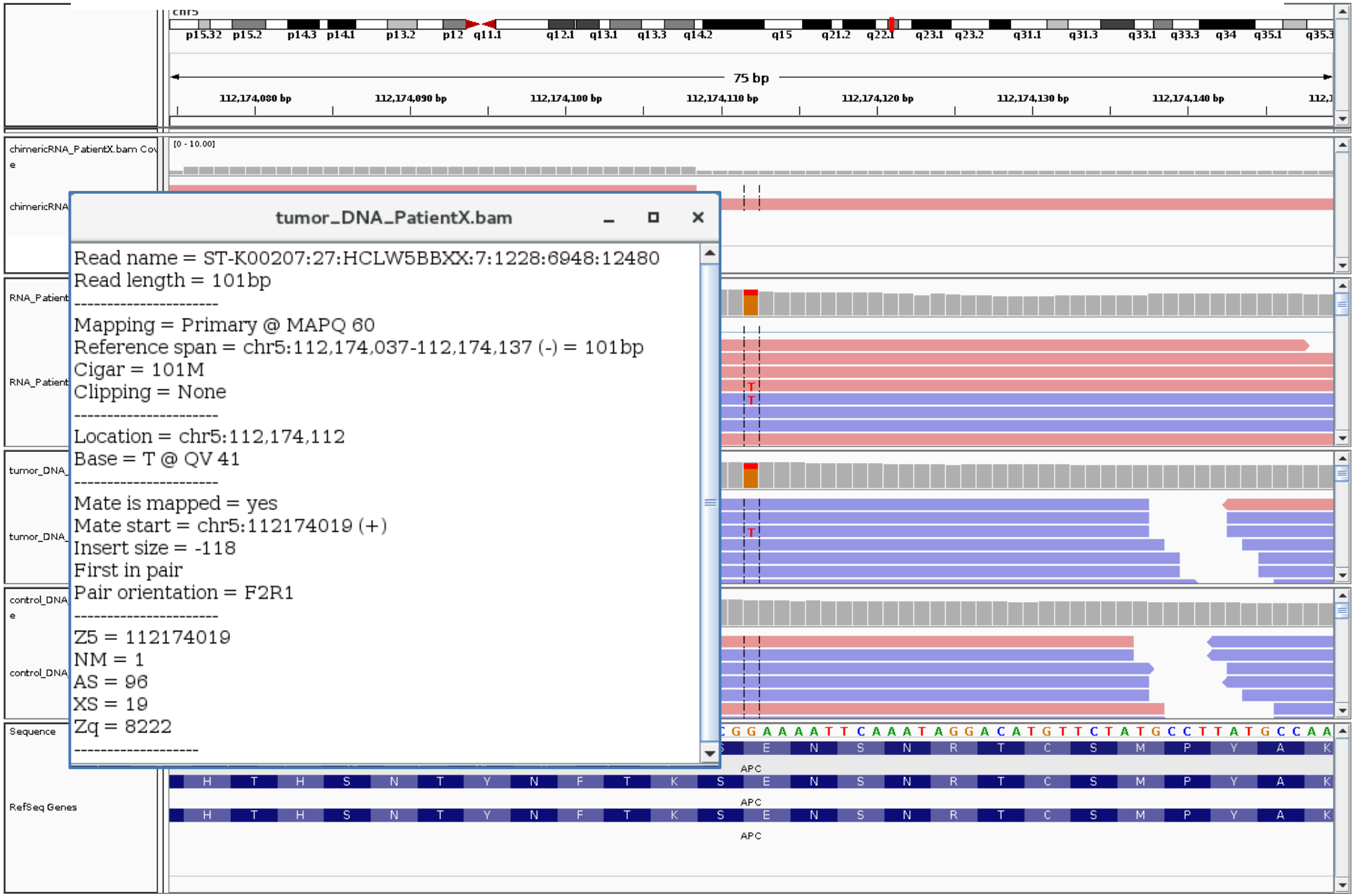
Somatic SNV


chr5
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q12.1 q13.1 q13.3 q14.2 q15 q21.2 q22.1 q23.1 q23.2 q31.1 q31.3 q33.1 q33.3 q34 q35.1 q35.3
75 bp
112,174,080 bp
112,174,090 bp
112,174,100 bp
112,174,110 bp
112,174,120 bp
112,174,130 bp
112,174,140 bp
chimericRNA_PatientX.bam Coverage
[0 - 10.00]
chimericRNA_PatientX.bam
RNA_PatientX.bam Coverage
[0 - 66]
RNA_PatientX.bam
T
T
right click
tumor_DNA_PatientX.bam Coverage
[0 - 338]
tumor_DNA_PatientX.bam
T
control_DNA_PatientX.bam Coverage
[0 - 321]
control_DNA_PatientX.bam
Sequence
C C A T A C A C A T T C A A A C A C T T A C A A T T T C A C T A A G T C G G A A A A T T C A A A T A G G A C A T G T T C T A T G C C T T A T G C C A A
H T H S N T Y N F T K S E N S N R T C S M P Y A K
APC
RefSeq Genes
APC
APC

Somatic SNV


chr5
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q12.1 q13.1 q13.3 q14.2 q15 q21.2 q22.1 q23.1 q23.2 q31.1 q31.3 q33.1 q33.3 q34 q35.1 q35.3
75 bp
112,174,080 bp
112,174,090 bp
112,174,100 bp
112,174,110 bp
112,174,120 bp
112,174,130 bp
112,174,140 bp
chimericRNA_PatientX.bam Coverage
[0 - 10.00]
chimericRNA_PatientX.bam
RNA_PatientX.bam Coverage
[0 - 66]
RNA_PatientX.bam
left click
tumor_DNA_PatientX.bam Coverage
[0 - 338]
tumor_DNA_PatientX.bam
control_DNA_PatientX.bam Coverage
[0 - 321]
control_DNA_PatientX.bam
Sequence
RefSeq Genes
APC
APC
APC

Somatic SNV


chr5
p15.32 p15.2 p14.3 p14.1 p13.2 p12 q11.1 q12.1 q13.1 q13.3 q14.2 q15 q21.2 q22.1 q23.1 q23.2 q31.1 q31.3 q33.1 q33.3 q34 q35.1 q35.3
75 bp
112,174,080 bp
112,174,090 bp
112,174,100 bp
112,174,110 bp
112,174,120 bp
112,174,130 bp
112,174,140 bp
chimericRNA_PatientX.bam Cov
e
[0 - 10.00]
chimericRNA
RNA_Patient
RNA_Patient
tumor_DNA_
tumor_DNA_
control_DNA
e
control_DNA
Sequence
RefSeq Genes
APC
APC
APC
tumor_DNA_PatientX.bam
Read name = ST-K00207:27:HCLW5BBXX:7:1228:6948:12480
Read length = 101bp

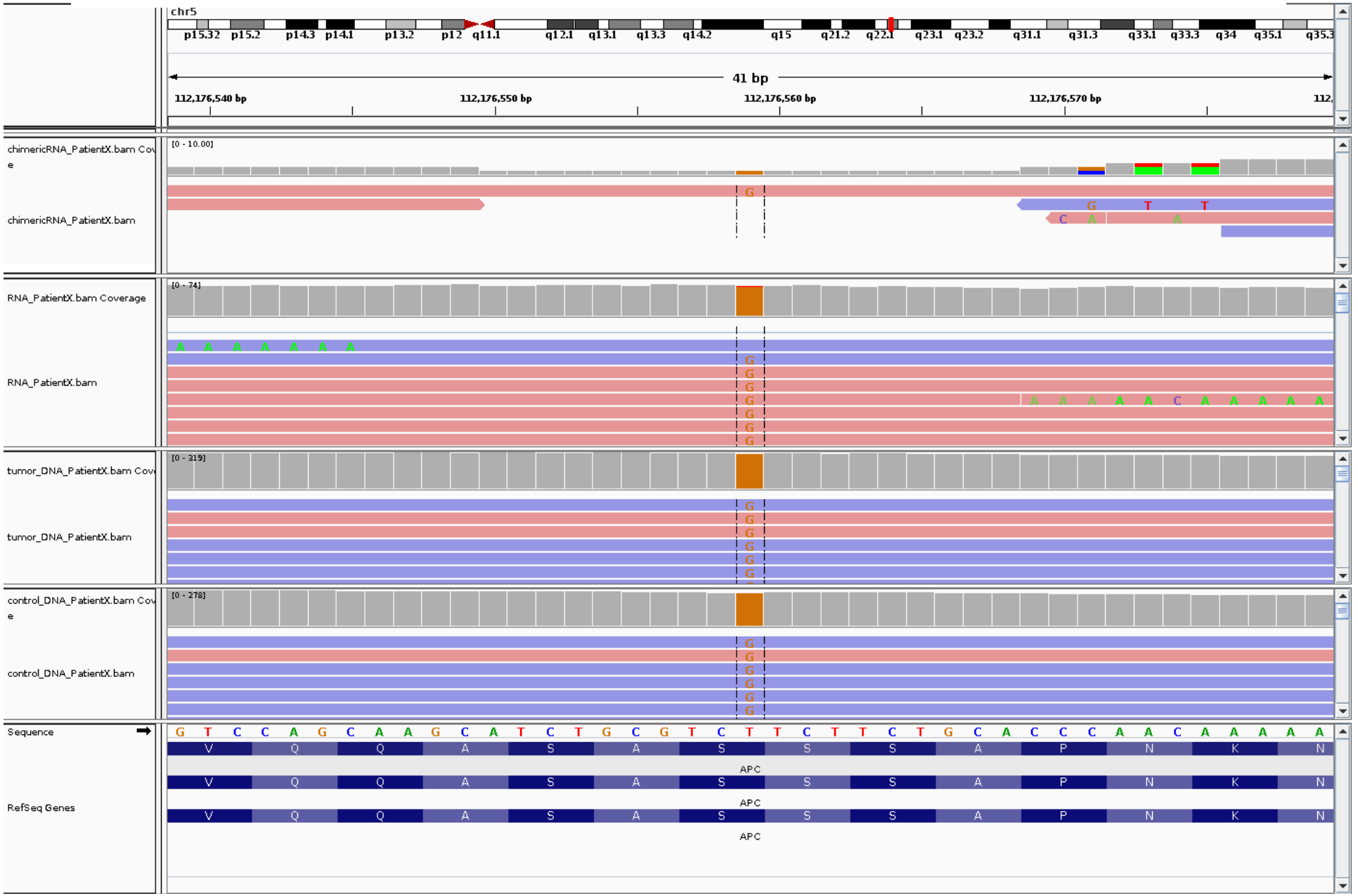
Mapping = Primary @ MAPQ 60
Reference span = chr5:112,174,037-112,174,137 (-) = 101bp
Cigar = 101M
Clipping = None

Location = chr5:112,174,112
Base = T @ QV 41

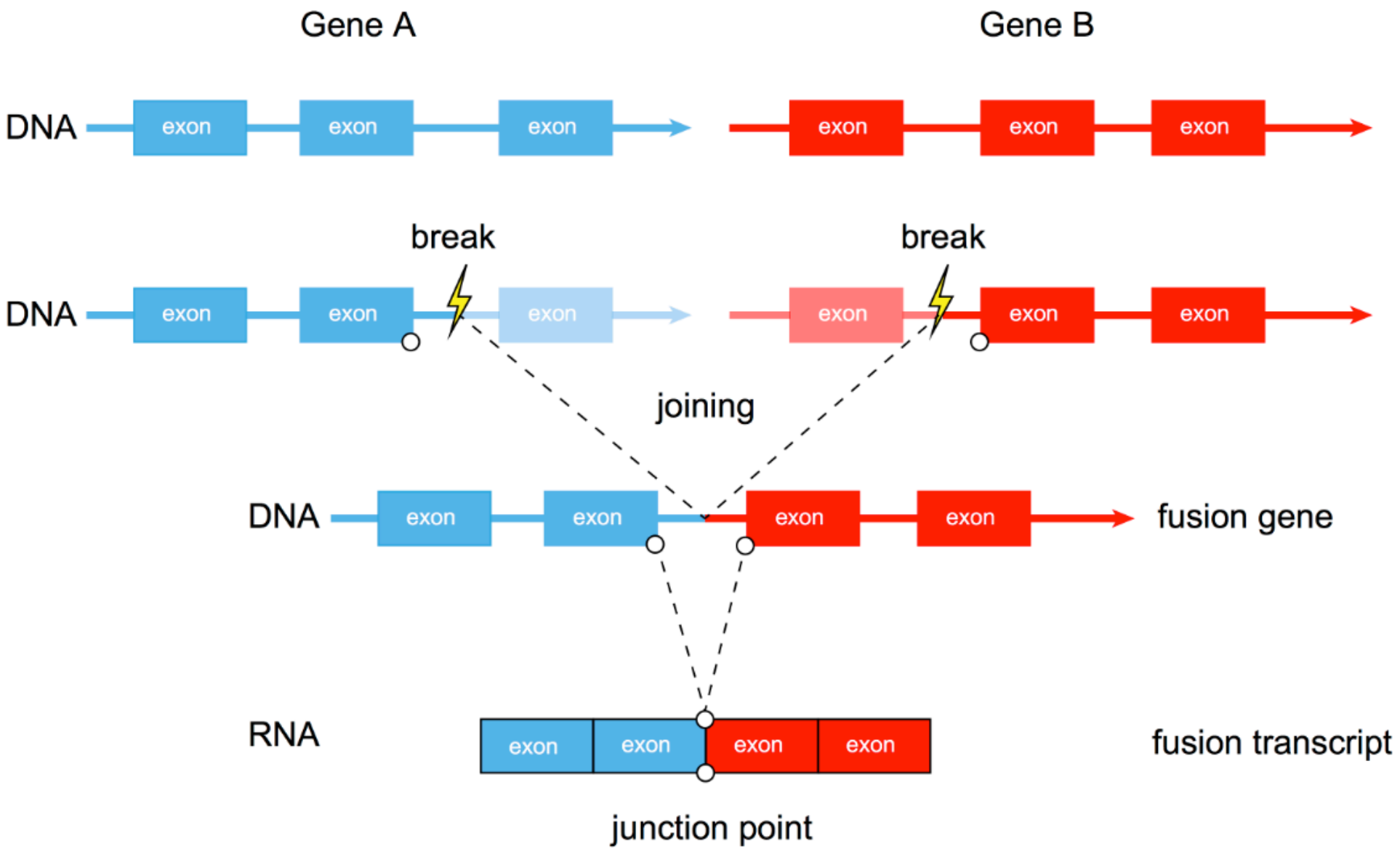
Mate is mapped = yes
Mate start = chr5:112174019 (+)
Insert size = -118
First in pair
Pair orientation = F2R1

Z5 = 112174019
NM = 1
AS = 96
XS = 19
Zq = 8222

Germline SNV



Gene fusion



TCGA Fusion Gene Data Portal

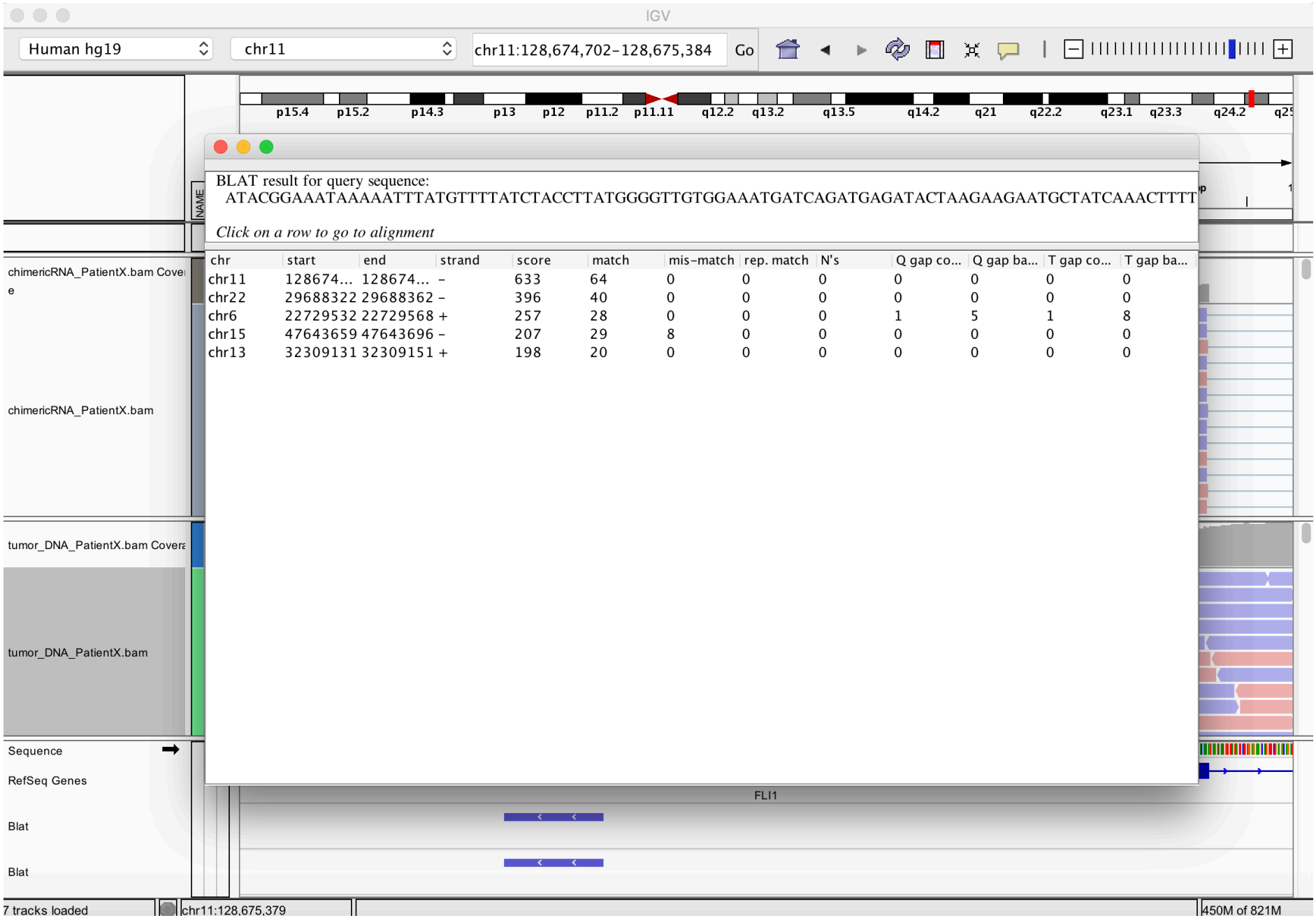
Gene fusion

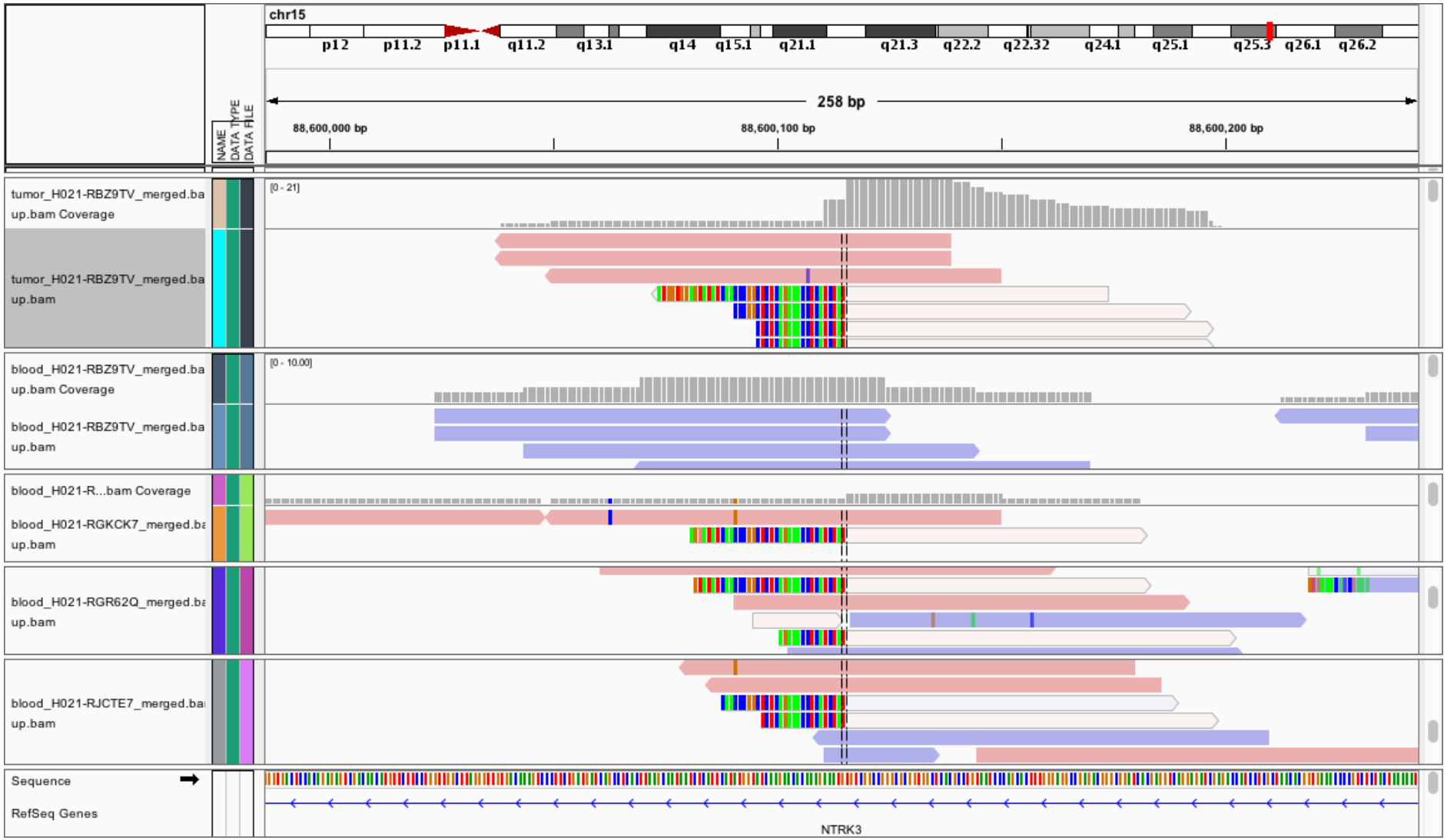
#gene1	gene2	strand1	strand2	breakpoint1	breakpoint2	site1	site2	split_reads1	split_reads2	discordant_mae	value
EWSR1	FLI1	+	+	22:29688159	11:128675261	splice-site	splice-site	42	28	39	3.57331e-43
EWSR1	FLI1	+	+	22:29688360	11:128674873	intronic	exonic	1	0	44	1.0489e-17
FLI1	EWSR1	+	+	11:128651919	22:29688477	splice-site	splice-site	0	4	4	0.00555282

Fusion detection: https://github.com/suhrig/arriba

Actual break point in DNA

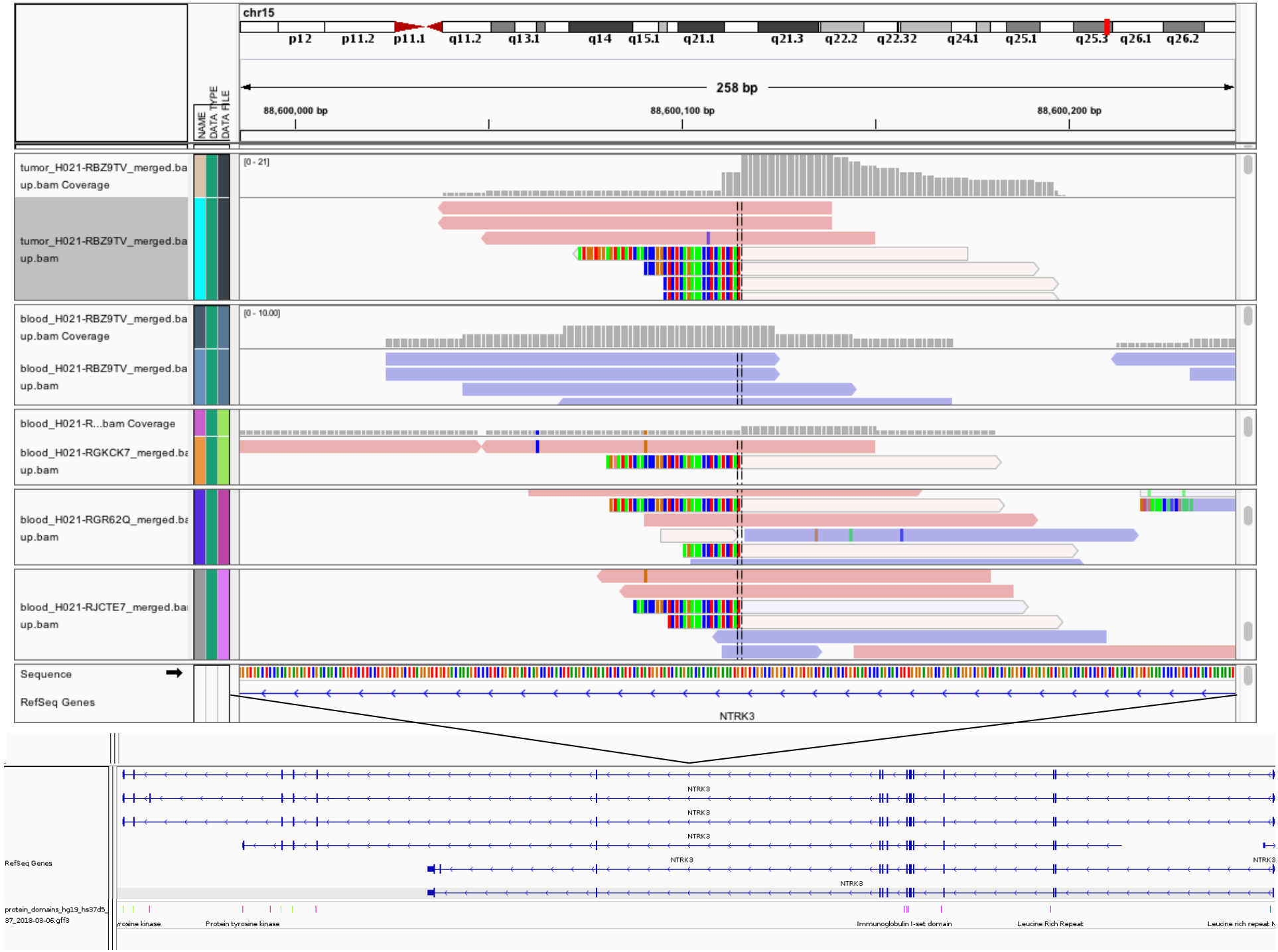
Use Blat


IGV
Human hg19
chr11
chr11:128,674,702-128,675,384
Go
p15.4 p15.2 p14.3 p13 p12 p11.2 p11.11 q12.2 q13.2 q13.5 q14.2 q21 q22.2 q23.1 q23.3 q24.2
BLAT result for query sequence:
ATACGGAAATAAAAATTTATGTTTTATCTACCTTATGGGGTTGTGGAAATGATCAGATGAGATACTAAGAAGAATGCTATCAAACTTTT
Click on a row to go to alignment
chr start end strand score match mis-match rep. match N's Q gap co... Q gap ba... T gap co... T gap ba...
chr11 128674... 128674... - 633 64 0 0 0 0 0 0 0
chr22 29688322 29688362 - 396 40 0 0 0 0 0 0 0
chr6 22729532 22729568 + 257 28 0 0 0 1 5 1 8
chr15 47643659 47643696 - 207 29 8 0 0 0 0 0 0
chr13 32309131 32309151 + 198 20 0 0 0 0 0 0 0
NAME
chimericRNA_PatientX.bam Cover
e
chimericRNA_PatientX.bam
tumor_DNA_PatientX.bam Covera
tumor_DNA_PatientX.bam
Sequence
RefSeq Genes
Blat
Blat
FLI1
7 tracks loaded
chr11:128,675,379
450M of 821M

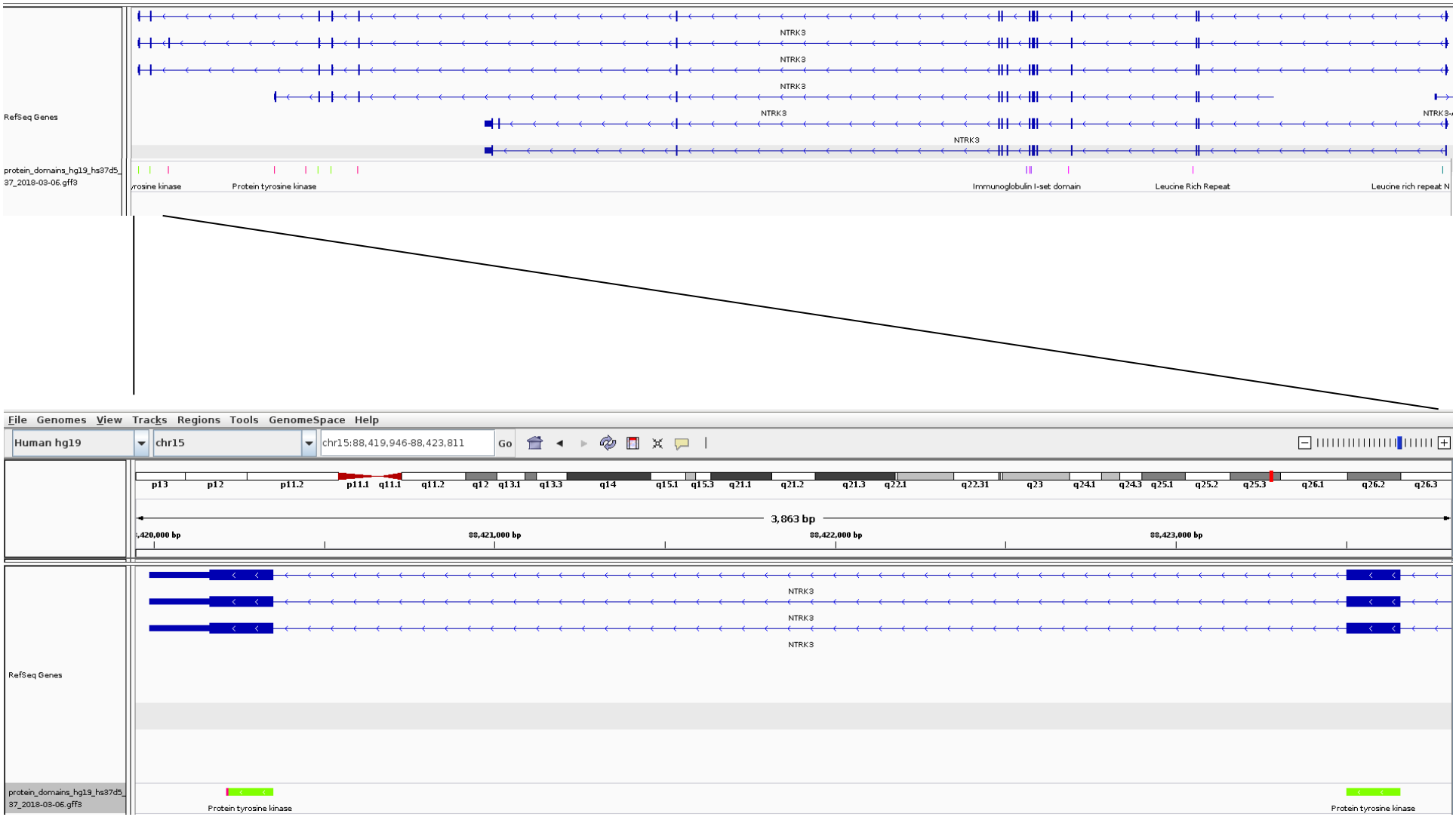


CREST: CTX: PLEKHA6_NTRK3

- Translocation/fusion would be highly clinically relevant
- Reads map also elsewhere in the genome
- Same breakpoint can be found in many randomly chosen samples (in tumor and/or blood)
- Even if true, kinase domain would be cut off

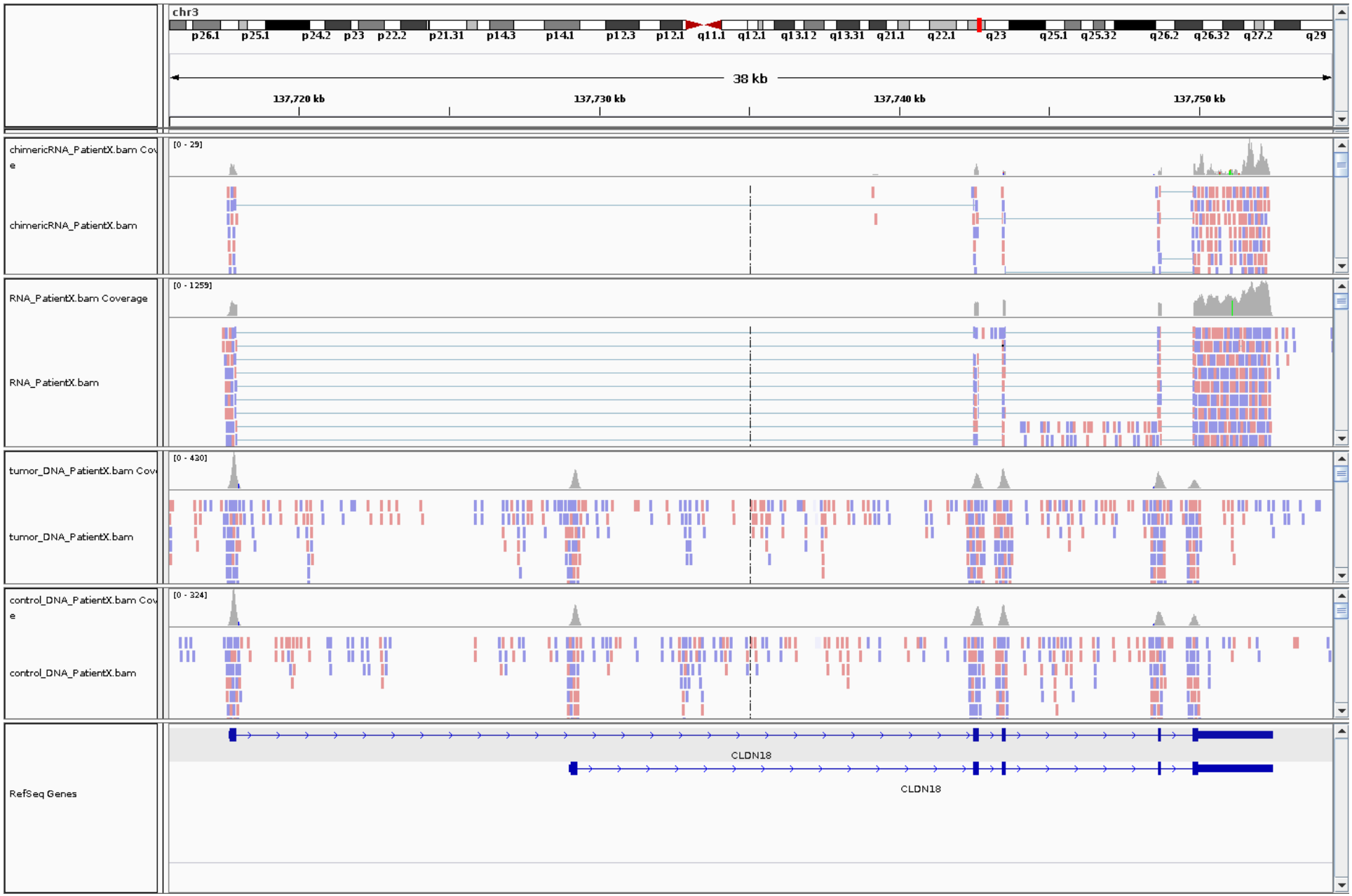

chr15
p12
p11.2
p11.1
q11.2
q13.1
q14
q15.1
q21.1
q21.3
q22.2
q22.32
q24.1
q25.1
q25.3
q26.1
q26.2
258 bp
88,600,000 bp
88,600,100 bp
88,600,200 bp
NAME
DATA TYPE
DATA FILE
tumor_H021-RBZ9TV_merged.bam Coverage
tumor_H021-RBZ9TV_merged.bam
[0 - 21]
blood_H021-RBZ9TV_merged.bam Coverage
blood_H021-RBZ9TV_merged.bam
[0 - 10.00]
blood_H021-R...bam Coverage
blood_H021-RGKCK7_merged.bam
blood_H021-RGR62Q_merged.bam
blood_H021-RJCTE7_merged.bam
Sequence
RefSeq Genes
NTRK3
NTRK3
NTRK3
NTRK3
NTRK3
NTRK3
NTRK3
RefSeq Genes
protein_domains_hg19_hs37d5_37_2018-03-06.gff3
yrosine kinase
Protein tyrosine kinase
Immunoglobulin I-set domain
Leucine Rich Repeat
Leucine rich repeat N

Protein domain track

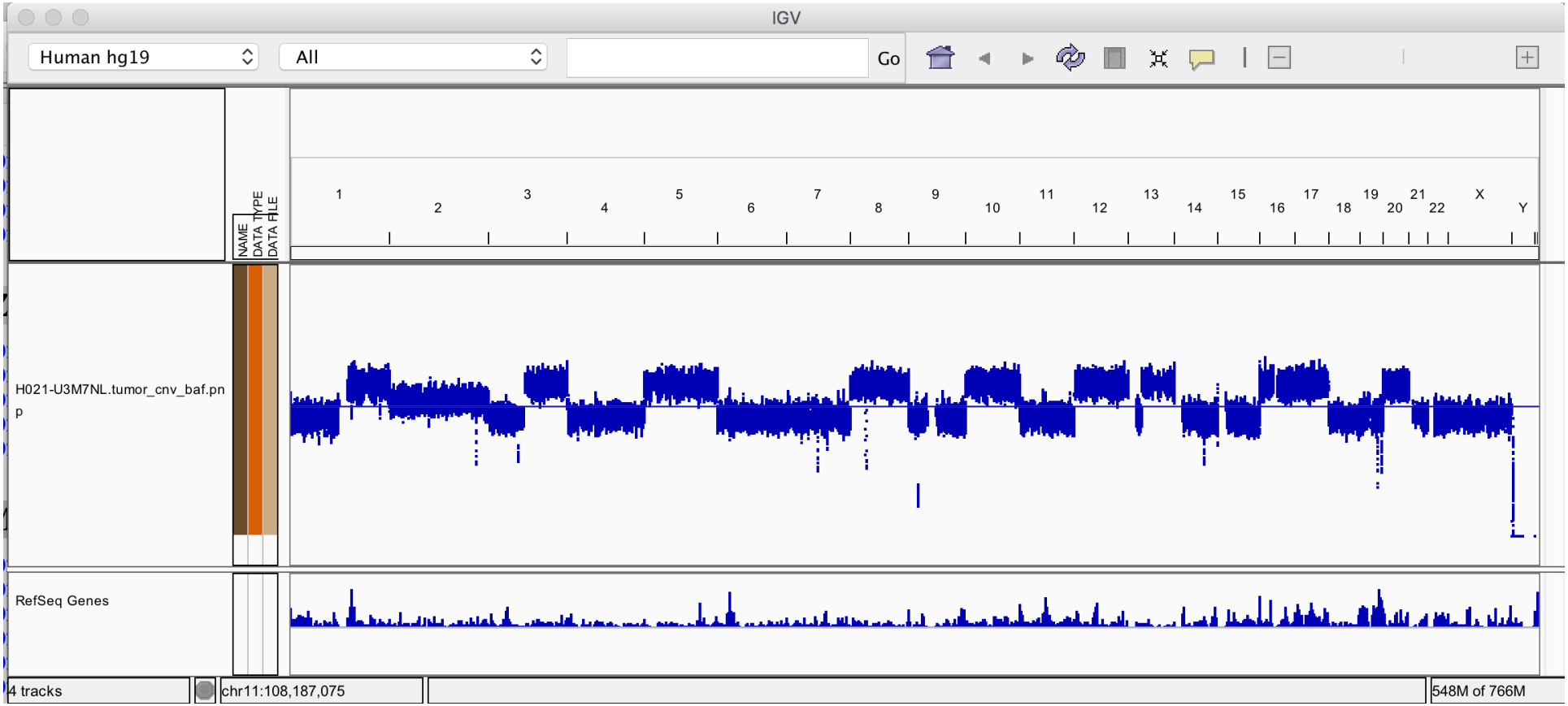


by Sebastian Uhrig

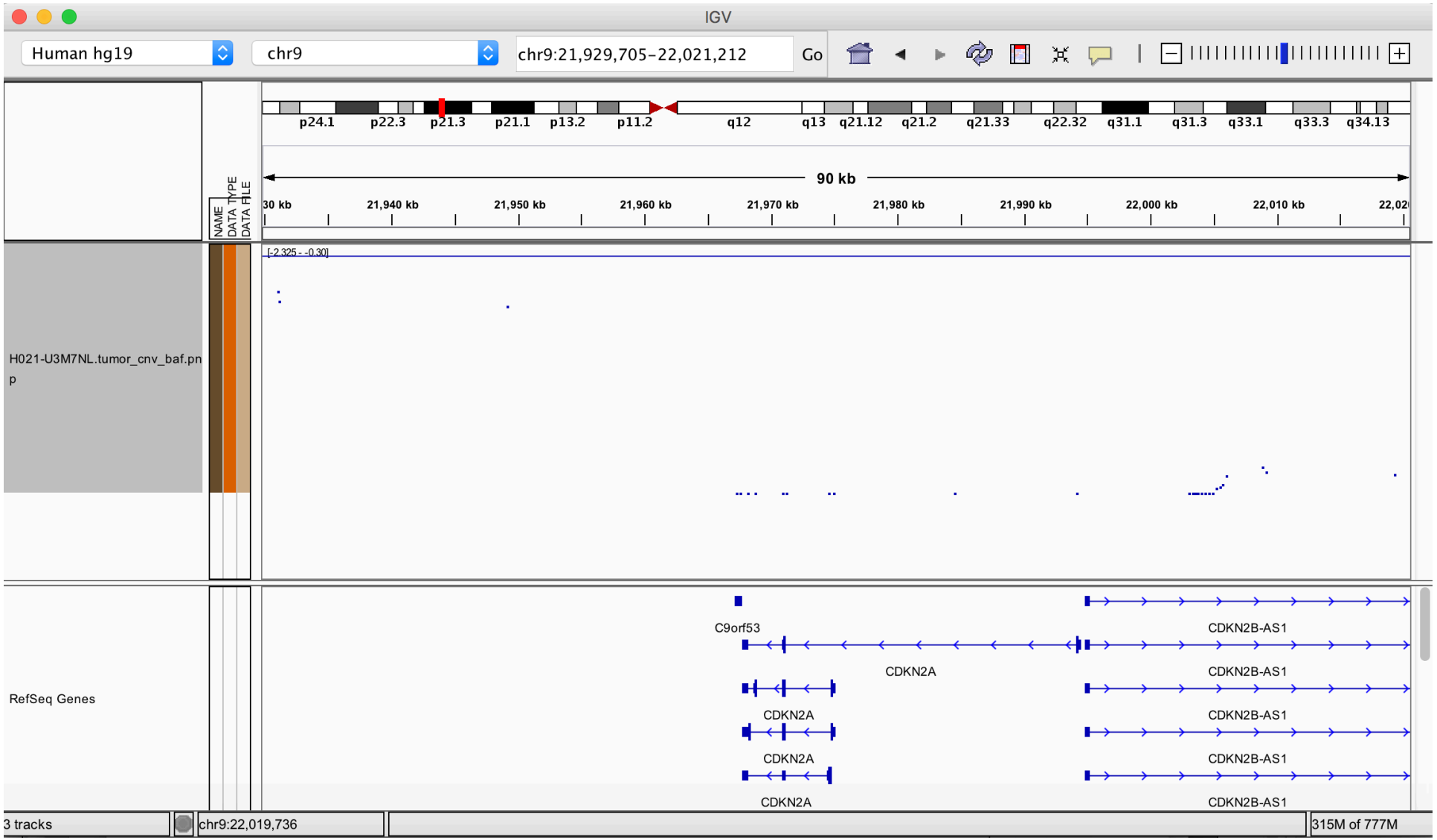
Differential exon coverage



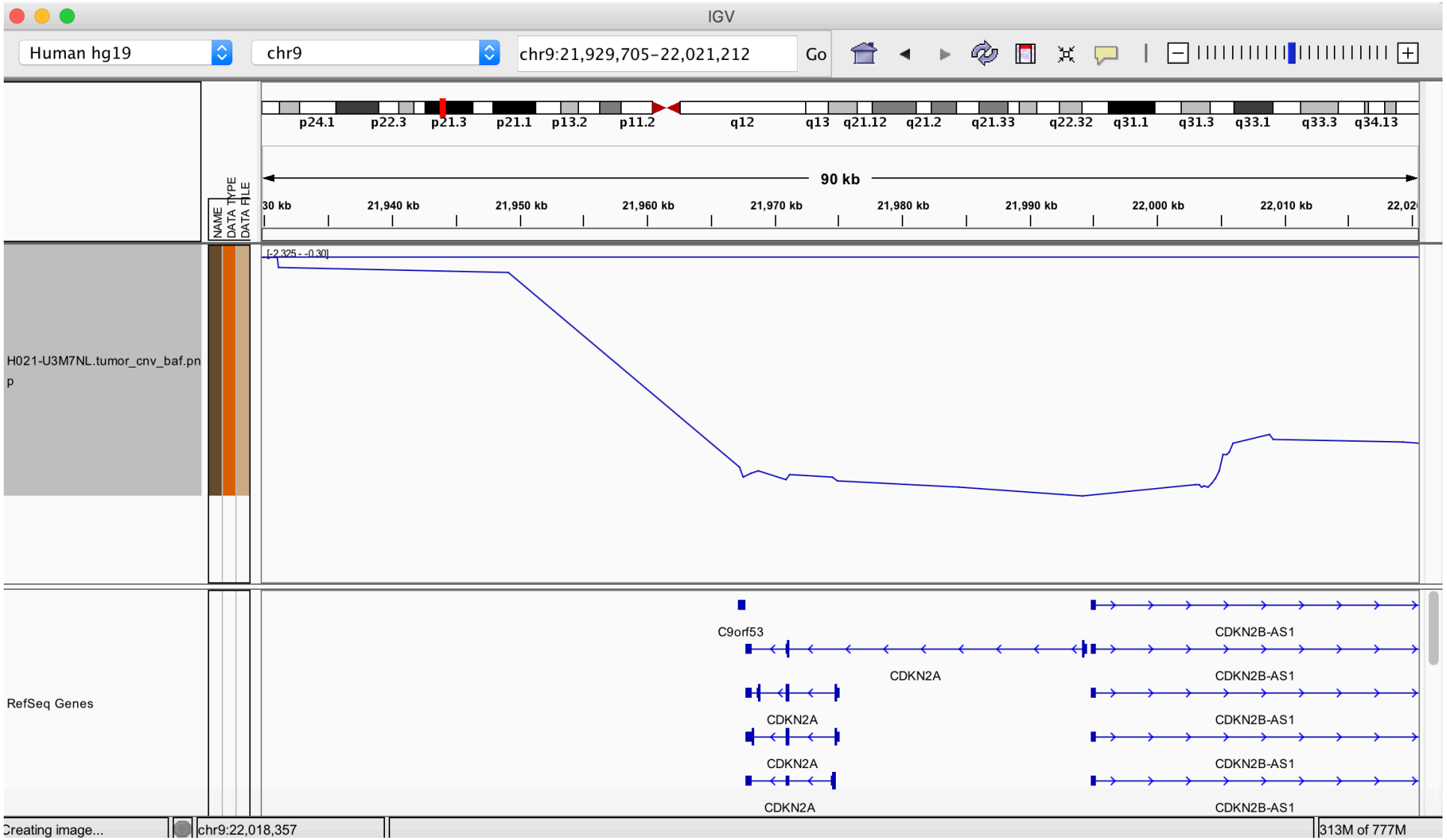
CNV Plot


IGV
Human hg19
All
Go
NAME
DATA TYPE
DATA FILE
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
H021-U3M7NL.tumor_cnv_baf.pn
p
RefSeq Genes
4 tracks
chr11:108,187,075
548M of 766M

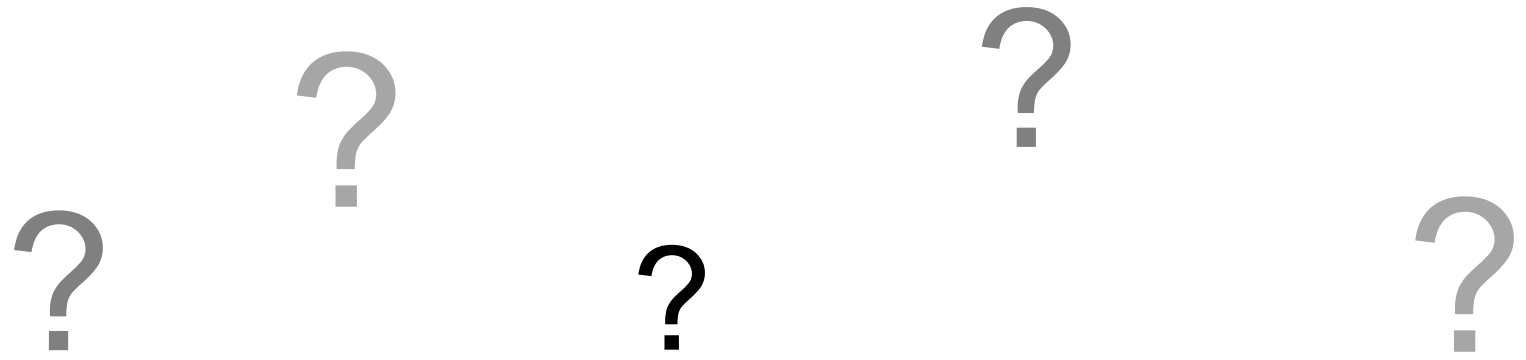
CNV Plot


IGV
Human hg19
chr9
chr9:21,929,705-22,021,212
Go
p24.1
p22.3
p21.3
p21.1
p13.2
p11.2
q12
q13
q21.12
q21.2
q21.33
q22.32
q31.1
q31.3
q33.1
q33.3
q34.13
90 kb
30 kb
21,940 kb
21,950 kb
21,960 kb
21,970 kb
21,980 kb
21,990 kb
22,000 kb
22,010 kb
NAME
DATA TYPE
DATA FILE
[-2.325 - -0.30]
H021-U3M7NL.tumor_cnv_baf.pn
p
RefSeq Genes
C9orf53
CDKN2A
CDKN2B-AS1
3 tracks
chr9:22,019,736
315M of 777M

CNV Plot


IGV
Human hg19
chr9
chr9:21,929,705-22,021,212
Go
p24.1
p22.3
p21.3
p21.1
p13.2
p11.2
q12
q13
q21.12
q21.2
q21.33
q22.32
q31.1
q31.3
q33.1
q33.3
q34.13
90 kb
30 kb
21,940 kb
21,950 kb
21,960 kb
21,970 kb
21,980 kb
21,990 kb
22,000 kb
22,010 kb
NAME
DATA TYPE
DATA FILE
H021-U3M7NL.tumor_cnv_baf.pn
p
[-2.325 - -0.30]
RefSeq Genes
C9orf53
CDKN2A
CDKN2B-AS1
Creating image...
chr9:22,018,357
313M of 777M

Time for Questions



SAM format

Column Name Description

1 QNAME - Query pair NAME if paired; or Query NAME if unpaired

2 FLAG - Bitwise FLAG (=> see next slide)

3 RNAME - Reference sequence NAME

4 POS - 1-based leftmost POSition of the clipped sequence

5 MAPQ - MAPping Quality (Phred-scaled probability of mismapping; 0 for multiple mapping possibilities)

6 CIGAR - Extended CIGAR string

7 MRNM - Mate Reference sequence NaMe; "=" if the same as RNAME (now RNEXT)

8 MPOS - 1-based leftmost Mate POSition of the clipped sequence (now PNEXT)

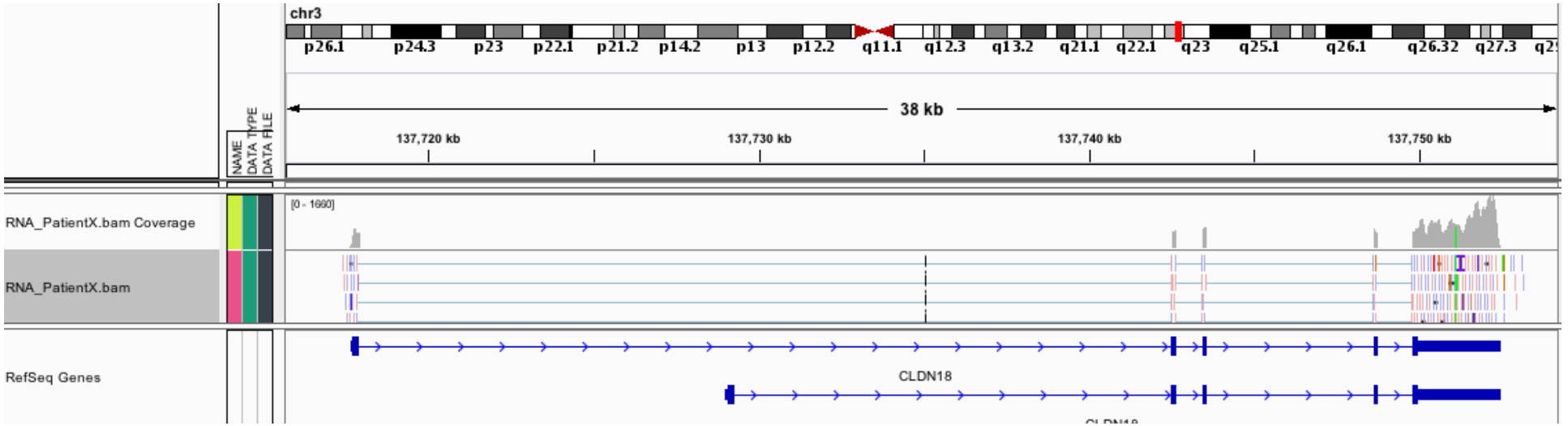
9 ISIZE - Inferred Insert SIZE (now TLEN: signed observed Template LENgth)

10 SEQ - Query SEQuence (reverse complemented if read mapped to reverse strand!)

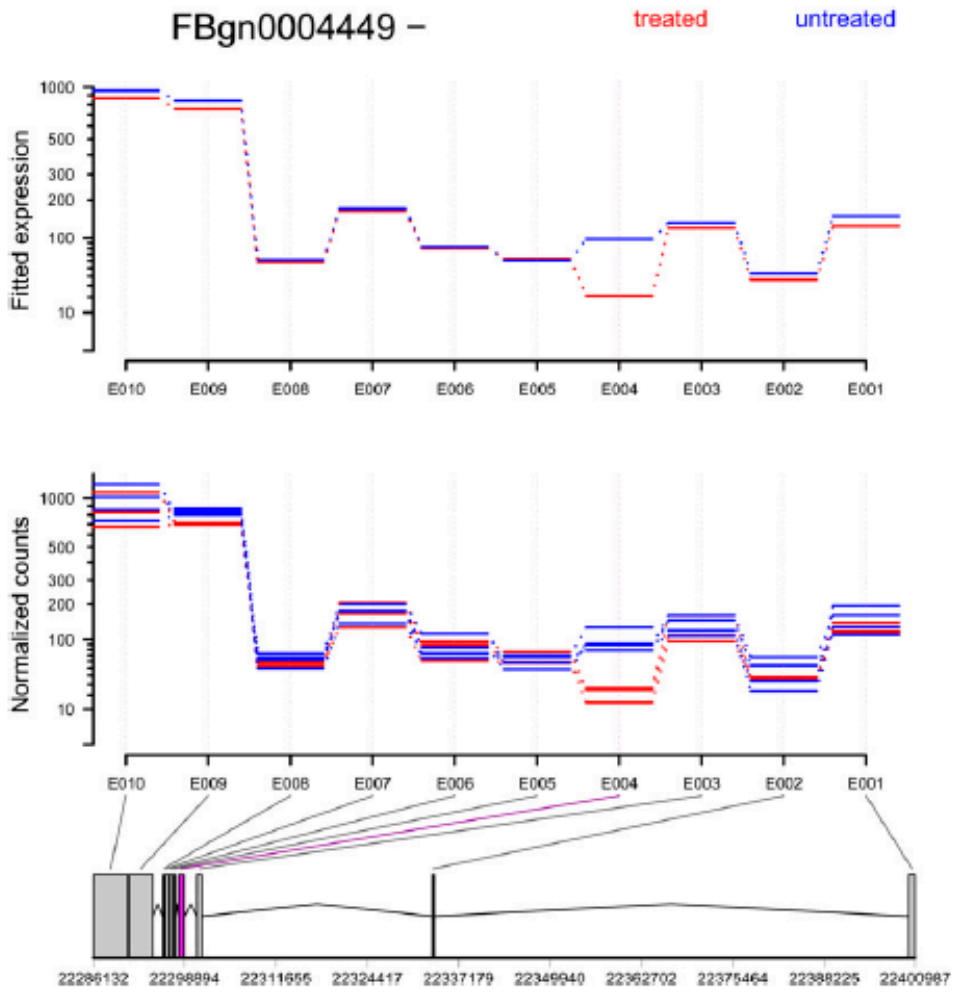
11 QUAL - Query QUALity (inverted if read mapped to reverse strand!)

Additional tags, e.g. MD: string for mismatching position (=> SNP finding)

Differential exon coverage



Differential exon coverage with DEXSeq



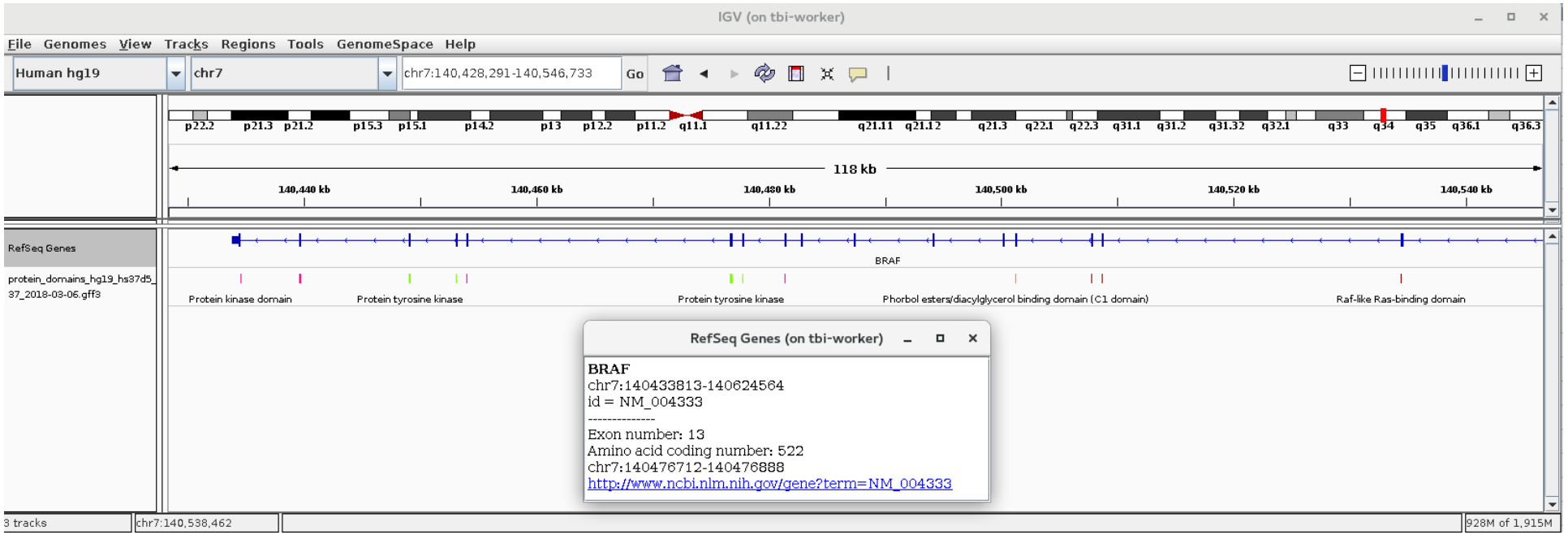
Anders et al. Genome Res 2012

Differential exon coverage

- Calculated using bedtools coverage
- Aditional scripts to get RPKM and TPM

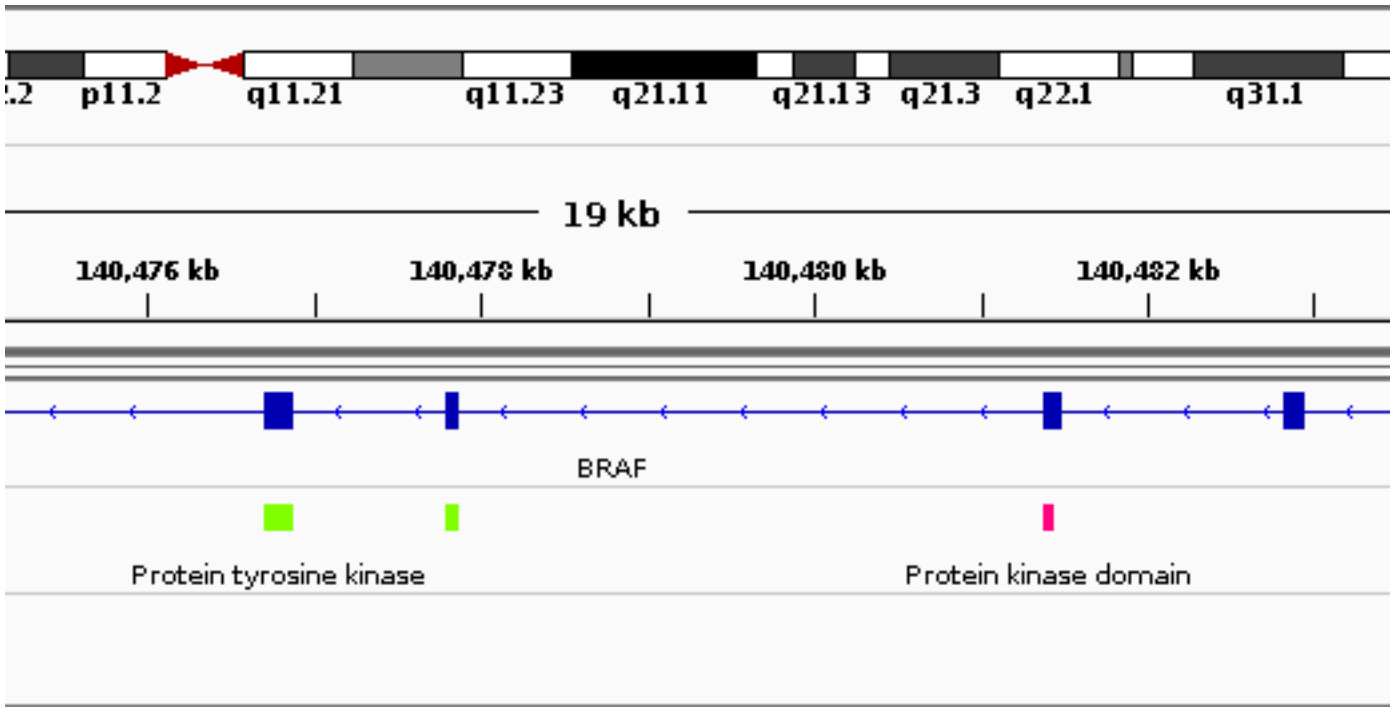
#chrom	chromStart	chromEnd	name	exonNr	strand	length	reads	bases_covered	length	coverage	RPKM	TPM
3	137717657	137717930	CLDN18	1	+	273	1292	273	273	1.0000000	46.1348	8.6
3	137729005	137729287	CLDN18	2	+	282	0	0	282	0.0000000	0.0000	0.0
3	137742499	137742664	CLDN18	3	+	165	1209	165	165	1.0000000	71.4284	13.31
3	137743448	137743566	CLDN18	4	+	118	1082	118	118	1.0000000	89.3869	16.66
3	137748638	137748749	CLDN18	5	+	111	961	111	111	1.0000000	84.3974	15.73
3	137749811	137752494	CLDN18	6	+	2683	18876	2683	2683	1.0000000	68.5832	12.78

Protein domain track



by Sebastian Uhrig

Protein domain track



by Sebastian Uhrig

Getting started

```
froehlim@tbi-worker: ~ > module avail

------------------------------------------------------------ /tbi/cluster/13.1/x86_64/modules-3.2.10/module_defs ------------------------------------------
arriba/0.10           cmake/3.7.1          htslib/1.2.1          liblmdb/0.9.21        Platypus/0.8.1.1      rnaseqc/1.1.8         SOAPnuke/1.6.0
arriba/0.12           CNVnator/0.3.3       htslib/1.3.1          libmaus2/2.0.429      PRIAM/2015            root/5.99.06          SOPHIA/34.0
arriba/0.8            CPLEX/ce_12.7.0      htslib/1.3.2          llvm/3.8.1            pypy/4.0.1            root/6.12.04          STAR/2.5.2b
bamUtil/1.0.13        fastqc/0.10.1        htslib/1.4.1          MACS2/2.1.1           python/2.7.9          rstudio/0.99-720      STAR/2.5.3a
bcftools/1.4.1        fastqc/0.11.3        igv/2.3.60            manta/1.2.2           python/3.4.3          rstudio/0.99-902      strelka/2.8.4
bedtools/2.16.2       fastqc/0.11.5        igv/2.3.72            mixcr/2.1.5           python/3.5.2          rstudio/1.1.296       subread/1.5.1
bedtools/2.23.0       GATK/3.7.0           igv/2.3.81            mutect/1.1.4          QCTools/1.0           salmon/0.8.2          subread/1.5.3
bedtools/2.24.0       gcc/5.4.x            igv/2.3.85            ncbi-blast/2.2.24     qualimap/2.2.1        sambamba/0.4.6        SYMPHONY/5.6
bioawk/20110810       gcc/6.x              igv/2.3.97            ncbi-blast/2.7.1      R/2.15.0              sambamba/0.5.4        test/1.0
biobambam/0.0.148     giggle/0.6.3         java/1.7.0_55         NetMHCIIpan/3.1a      R/3.0.0               sambamba/0.5.9        test/2.0
biobambam/0.0.191     git/2.9.4            java/1.8.0_40         NetMHCpan/3.0a        R/3.1.2               sambamba/0.6.3        tigra/0.4.0
biobambam2/2.0.81     groovy/2.4.7         Je/1.0                perl/5.20.2           R/3.2.0               sambamba/0.6.5        tigra/0.4.2
blat/34               hdf5/1.8.18          Jemultiplexer/1.0.6   PHLAT/1.0             R/3.2.2               sambamba/0.6.6        udunits/2.2.25
boost/1.62.0          HIPO2_rna/v1         joinx/1.6.17          phylowgs/d5bdea1      R/3.3.0               samtools/0.1.17       VarDict/1.4.6
bowtie2/2.2.1         HIPO2_rna/v2         joinx/42a7caf         Platypus/0.7.4        R/3.3.1               samtools/0.1.19       vcftools/0.1.10
bwa/0.7.15            HIPO2_sv/v1          kallisto/0.43.0       Platypus/0.7.4_ucs2   R/3.4.0               samtools/1.2          vcftools/0.1.12b
canvas/1.32.0.918     Homer/4.8.3          kallisto/0.43.1       Platypus/0.7.9.2      R/3.4.2               samtools/1.3.1        WebLogo/2.8.2
cdhit/4.7             htslib/0.2.5         ldc/1.2.0             Platypus/0.7.9.2_ucs2 R/3.5.0               samtools/1.5
cellranger/2.0.2      htslib/1.2           liblmdb/0.9.18        Platypus/0.8.1        razers3/3.5.7         samtools/1.6

------------------------------------------------------------------------ /software/.modules/sw ------------------------------------------------------
anaconda3/5.1.0  cellranger/2.1.1 groovy/2.4.15    jdk/8u172        picard/1.61      python/3.6.0     trimmomatic/0.38 vcflib/b1e9b31
bedops/2.4.14    crest/0.0.1      htslib/1.8       jre/8u171        pindel/0.2.5b9   sophia/35        varscan/2.3.5
caveman/1.13.1   fit-sne/d589803  igv2/2.3.97      pandoc/2.2.1     pypy/2.7-6.0.0   trimmomatic/0.30 varscan/2.3.8
froehlim@tbi-worker: ~ > module load igv/2.3.97
froehlim@tbi-worker: ~ > igv.sh
```