# Data Interpretation of RNASeq and Bisulfite sequencing Data in Cancer Research (March 2019)

## Exercise: IGV (lectured by Martina Fröhlich, March 2019)

The Integrative Genomics Viewer (IGV, http://software.broadinstitute.org/software/igv/) is a visualization tool with which you can view genomic data such as BAM files.

In the following exercise you will learn how to use IGV and see some practical examples on cancer sequencing data.

For this, we prepared 4 bam files

Chimeric RNA
RNA
Tumor DNA
Control DNA (matching control)

Reads in the DNA BAM files were aligned with either BWA align or BWA mem. The RNA sequencing reads were aligned with STAR (https://github.com/alexdobin/STAR). You will see that STAR stores the majority of reads in the RNA bam file. Some reads are in addition stored also in the Chimeric RNA BAM file. Those are reads that contain segments that are chimeric, e.g. they belong to different chromosomes, different strands or genomic regions that are far away from each other. These reads are important to identify fusion genes.

The examples and exercises given below will contain somatic SNVs, SNPs, alternative splicing and true and false positive detected fusions.

1. Looking at an alignment file in IGV

- type *igv_2.3.81.sh &* (if this does not work, simply type *igv*)
- Make sure you have the genome *Human hg19* loaded. Otherwise go to Genomes – Manage Genome List – Add From Server – *Human hg19.*
- Load the four BAM files that were described above
  Go to File – load from file

  **/bfg/data/courses/deNBI2019/commandline_IGV/*bam_files/chimericRNA_PatientX.bam***
  **/bfg/data/courses/deNBI2019commandline_IGV/*bam_files /RNA_PatientX.bam***
  **/bfg/data/courses/deNBI2019/commandline_IGV/*bam_files /tumor_DNA_PatientX.bam***
  **/bfg/data/courses/deNBI2019/commandline_IGV/*bam_files /control_DNA_PatientX.bam***

  You can choose yourself in which order you load the files. The first file loaded will appear at the top in IGV, the next one below that and so on. In practice it is helpful to use the same order for each patient you look at.

- Type into the search box (the empty field at the top) 5:112174112 and hit Go
  then the Integrative Genome Viewer jumps to window 5:112,174,092-112,174,132.
- Hover with the mouse over one of the reads. This opens the “hover element details”. While the “hover element details” are very informative, you might find it a bit annoying that they always pop up as soon as you hover over a read (default option in IGV). You can change this feature via the yellow balloon icon in the IGV tool bar to specify the behavior you want, e.g. by selecting “Show details on click”.
- The pop-up window shows the read’s properties as well as the property of the specific base in the read you clicked on. E.g. you can get information about the read name, the read’s CIGAR string, whether or not and where the read’s mate is mapped as well as the quality of the base on this exact position.
- Right-click on one of the reads. This opens the right-click menu with further options.

- How does the view change if you click “View as pairs”? (You might have to zoom out a bit to see an effect. One way of zooming is via the zoom bar in the upper right corner of IGV.)
- Click on ”blat read sequence”. Alternatively, you can click on “copy read sequence”, open blat or blast in your internet browser and align the sequence to the respective reference genome. This manual control is sometimes useful to check where else a read maps to the genome.

Another option that will be useful later is to display softclipped bases. These can be bases, where part of the read maps to a different genomic location. You see them a lot when you look at chromosomal rearrangements and gene fusions. But don’t get always overexcited when you see them. They can also be bases at the end of reads that have low base phred quality. So you always need to check with the reference.

- To display softclipped bases go to: View - Preferences - Alignments and check “Show soft-clipped bases”

2. SNV and SNPs (or germline mutations)

- Have a look at the base on position 5:112174112
- Here, you see an example for a somatic SNV. The genotype in the control is G whereas in the tumor it is G and T.
- Go to the bar above the aligned reads. How many reads align here in tumor and control? What is the frequency of the alternative allele in tumor DNA and tumor RNA?
- Now go to position 5:112176559
- What are the differences between this variant and the last one you looked at?

3. Alternative Exon Usage

- Use IGV to have a look at the exon coverage of CLDN18 (type CLDN18 into the search window).
- What can see in the RNA reads?

(In practice, you will most likely use other tools to identify alternative exon usage, e.g. DEXSeq, self made pipelines using CoverageBed etc., but knowing how alternative exon usage or exon skipping looks like in IGV can become very useful.)

4. True and false positive detected gene fusions

The RNA in this example has been aligned with the STAR aligner. On the STAR output two different in-house fusion algorithms were applied. You can find the result of both algorithms in Fusions.ods and in the Appendix of this exercise.

The EWSR1_FLI1 fusion was found by both algorithms. The AR fusion was only found by one of the algorithms

- Which fusion would you believe in and why?
- Can you find the actual breakpoints of the EWSR1-FLI1 fusion in the DNA?

Hints: For gene fusions, IGVs split view option is useful. You can type ranges or split screen locations already in the search window. In addition, the “View mate region in split screen” option from the right-click menu can be helpful.
Also, zooming in or out can help to get a better overview. In split screen mode the zoom options in the upper right corner of IGV don’t work, but you can either zoom by clicking into the window and type + or – or by right clicking into the genome viewer window and selecting “zoom in” or “zoom out”.

## Appendix:

Algorithm 1

| #gene1 | gene2 | strand1 | strand2 | breakpoint1 | breakpoint2 | site1 | site2 | split_reads1 | split_reads2 | discordant_mates | e-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 |

Algorithm 2

| #chrom | start | end | gene | fusion_nr | breakpoint | softclip_count | exonnumber | exonstart | exonend | distance | strand | reads | discordant_ma |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr22 | 29684775 | 29692263 | EWSR1 | F1 | 29688476 | 82 | 9 | 29684594 | 29684775 | 0 | + | 110 | 23 |
| chr11 | 128642841 | 128680513 | FLI1 | F1 | 128651919 | 61 | 7 | 128642676 | 128642880 | 0 | + | 110 | 23 |
| chrX | 66950025 | 66950026 | AR | F2 | 66950025 | 27 | 9 | 66943527 | 66950461 | 0 | + | 3 | 0 |
| chr1 | 192679504 | 192679505 | RGS13 | F2 | 192679505 | 74 | 7 | 192628467 | 192629440 | 50064 | + | 3 | 0 |