IGV Practical Session.

Tom Carroll, MRC Clinical Sciences.
Mark Dunning, Cancer Research Uk, Cambridge Institute.

This session will give you short introduction into using IGV. We will be using externally hosted data available from Encode including ChIP-seq and RNA-seq data.
Please feel free to try other datasets once you have run through the tasks but remember IGV is hosted on your machine so less powerful computers will feel the memory load of large numbers of datasets.

ChIP-seq data

This section looks at how to load and manipulate some ChIP-seq data. The data is from a lymphoblastoma cell line and contains ChIP-seq samples for several histone types.

Loading IGV and moving around the gene of interest.

- First, load IGV and select the hg19 genome.
- Navigate to chr12:6,641,585-6,649,537.
- Navigate to Gapdh.
- Zoom out to see the surrounding genes.
- Select the "RefSeq Genes" track and expand to see all transcript isoforms.
- Scroll to the left to discover the non-coding RNA SCRNA10.
- Click on SCRNA to bring up a window with further information.
- Double click on Gapdh to recentre window on the gene.
- Add a region of interest at the Gapdh locus using the IGV menu
  Regions -> Region Navigator -> add
- Right click the "Region of interest" and edit description to show "Active gene"

Loading data

- Go to the IGV menu -> File -> load from server
- From the menu follow drop-down
  Tutorials -> UI Basics (Encode)
  GM12878 CFCF
  GM12878 H3K27ac
  GM12878 H3K27me3
  GM12878 H3K36me3
  GM12878 H3K4me3
  GM12878 H3K9ac

Controlling IGV display

- Select all tracks and then select "autoscale."
- At the Gapdh locus, investigate the enrichment of signal over gene body.
- Set all the tracks to maintain current data ranges (*deselect autoscale*)

- Once set, navigate to PIANP.
- Note the difference in enrichment of ChIP-seq.
- Add as "Region of interest" and edit description to show "Inactive gene"
- Zoom out to compare enrichment across neighbouring genes.

- Go to Regions -> Gene Lists.. -> "proneural dev genes"
- Inspect the signal across genes to determine their expression state.
- Return to main view by selecting all from chromosome dropdown.

- Colour all track by a unique colour. Good idea to make K27me3 the most distinct colour to rest.

- Autoscale all tracks.
- Select all tracks and create an overlay,
  select tracks -> Create Overlay Track
- Change track height to 100
  Select tracks -> Change Track Height
- Revisit Gapdh, PIANP and gene list. Scan across genome to identify silent and active gene expression.

RNA-seq data

Start a new session

- Clear data by going to IGV menu -> File -> New Session…

Loading the data

- Go to the IGV menu -> File -> load from server
- From the menu follow drop-down
  Tutorials -> RNA-seq (Body Map)
  Heart
  Liver
- Go to IGV menu -> Views -> Preferences -> Alignments
  Tick "show junction track"
  Set "visibility range threshold' to 15kb
- Go to SLC25A3 gene.
- Collapse reads track (named Heart/Liver)

Inspect RNA-seq data

- Select tracks -> Colour Alignments by -> Read Strand to identify strand of transcript.
- Expand "Features track" to identify alternative exons/transcripts.
- Inspect coverage tracks to discover areas of coverage unique to Heart or Liver sample.
- Inspect junction tracks to evaluate alternative splicing of transcripts between tissue
- Select junction tracks -> Expand
- Compare major (first) transcript variant in Heart and Tissue

- Click on junctions to identify start and end of spans.

Another way to inspect splicing

- Select tracks -> Sashimi plot
  Select Gene Track -> Refseq Genes
  Select Alignment Tracks –> Heart + Liver

- Save image and compare junctions across tissue.

Copy-Number Data

For this section, we will use the copy-number dataset from Metabric (The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups- Curtis et al (2012))

Loading the data

IGV menu -> File -> Load from URL

http://igvdata.broadinstitute.org/data/hg18/metabric/Metabric_CN_IGV.seg

Zoom-out to show all chromosomes. You should see a pattern reminiscent of Figure 2 of the Metabric paper

http://www.nature.com/nature/journal/v486/n7403/fig_tab/nature10983_F2.html

Selecting a region

Go to region

chr8:35,982,271-39,004,295

Define this as a new region in Regions -> Region Navigator

Right-Click on the region you have just defined and select 'Sort by Amplification'

Zoom-out to region

chr8:21,970,223-52,680,530

Verify that the the region you created defines a narrow amplicon on 8p12.

Loading sample information

Now load the sample information for the metabric cohort

IGV menu -> File -> Load from URL

http://igvdata.broadinstitute.org/data/hg18/metabric/Metabric_Clinical_IGV.txt

Sorting tracks

Try sorting the data according to different clinical parameters (e.g. IntClustMemb, ER status, Pam50 subtype).

If you hover over a particular variable name at the top of the sample information panel, you should see an option to "click to sort". Alternatively use the menus

Tracks → Sort Tracks

Which IntClust group does the 8p12 amplification appear to be most common in? What Pam50 subtype?

Also try grouping by tracks by Intclust Memb, ER Status and see if you can identify different patterns of copy-number change between different groups

Traks → Group Tracks

DNA Re-sequencing data

The data from this section can be found on the Group Drive. These files are the aligned sequencing reads from three 1000 genomes individuals, and corresponding SNV calls.

Core/bioinformatics/public_folders/Training/GenomeBrowsers/exampleData

Load the aligned reads

File -> Load from File

NA19909.chr22.bam

NA19914.chr22.bam

NA19916.chr22.bam

Note: the corresponding index file (same file name, but with .bai) must be found in the same directory in order to load the bam files.

Three new data tracks will appear, but each panel will be blank unless you are sufficiently zoomed-in.

Navigate to chr22:22,309,703-22,329,832

You should now start to see individual reads coloured in grey. Coloured vertical lines on a read indicated bases in that read that are different from the reference genome, and potentially interesting to us. Hovering over the coverage track at any position shows the total number of bases observed in a particular location, and how many are on the positive and negative strand.

These data were generated after performing an exome enrichment.

Load the exome-enrichment file

The targets for an exome enrichment are often given as a bed file. Once such file is given in the example folder

File -> Load from File

Agilent_SureSelect_Human_All_Exon_38Mb_Kit_hg19.bed

You should now see a greater coverage in the enriched regions (i.e. exons)

Load the SNV calls

File -> Load from File

chr22.flt.vcf.gz

These are the SNV and indels calls generated using samtools (see appendix)

Investigating individual variants

Go to

chr22:22,312,790

What is the reference base at this position?

In which of our three samples was a variant called in? How many reference and alternate bases were found?

Loading external annotation

Load the file

common_and_clinical-latest.vcf.gz

These data are from;

http://www.ncbi.nlm.nih.gov/variation/docs/human_variation_vcf/

and show the locations of common variants that have also been linked to disease.

Use these file to locate potentially disease-relevant variants called in the three samples.

e.g.

chr22:18,900,969-18,901,046

What is the dbSNP ID for this variant? (Hover over the **common_and_clinical_latest** track for this postion)

Appendix

commands to generate bam and vcf file

samtools view -h ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/other_exome_alignments/NA19914/exome_alignment/NA19914.mapped.illumina.mosaik.ASW.exome.20111114.bam 22 | samtools view -bS - > exampleData/NA19914.chr22.bam

samtools view -h ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/other_exome_alignments/NA19909/exome_alignment/NA19909.mapped.illumina.mosaik.ASW.exome.20111114.bam 22 | samtools view -bS - > exampleData/NA19909.chr22.bam

samtools view -h ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/other_exome_alignments/NA19916/exome_alignment/NA19916.mapped.illumina.mosaik.ASW.exome.20111114.bam 22 | samtools view -bS - > exampleData/NA19916.chr22.bam

samtools index exampleData/NA19914.chr22.bam
samtools index exampleData/NA19909.chr22.bam
samtools index exampleData/NA19916.chr22.bam

```
samtools mpileup -uf human_g1k_v37.fasta.gz exampleData/NA*.bam | bcftools view -bvcg - > chr22.raw.bcf

bcftools view chr22.raw.bcf  | vcfutils.pl varFilter  -D100 > chr22.flt.vcf

bgzip -c chr22.flt.vcf  > chr22.flt.vcf.gz

tabix -p vcf chr22.flt.vcf.gz
```