TRAIN MALTA NETWORKING EVENT Day 2

Cambridge 2nd June 2017

Genes of interest

LRRK2

PINK1

GNRHR

GNRH1

FGFR2

WDR11

PKD1

SOD1

CXCL3

HIST1H1C

LHX2

ALDH1L2

IL1R2

TLR8

MPEG1

CLEC4D

VNN2

MGAM

FPR1

CXCL11

IL1A

IFIT1

Pu.1

CEMP

POU5F1

Nanog

SOST

SLFN14

KALRN

CLIC1

SPINT2

BRD3

SMIM1

TUBB1

ACTN1

DIAPH1

SRC

PAR4

TSPAN9

COMT

Haemoglobin genes on Chr 11

Genes involved in switching between CD4 and CD8 T-cells e.g. thPOK, runx3

SOD genes in various organisms

BED file formats

BED (Browser Extensible Data) format provides a flexible way to define the data lines that are displayed in an annotation track. BED lines have three required fields and nine additional optional fields. The first three required BED fields are:

- chrom - The name of the chromosome (e.g. chr3, chrY, chr2_random) or scaffold (e.g. scaffold10671).
- chromStart - The starting position of the feature in the chromosome or scaffold.
- chromEnd - The ending position of the feature in the chromosome or scaffold.

The 9 additional optional BED fields are:

- name - Defines the name of the BED line to be displayed.
- score - A score between 0 and 1000.
- strand - Defines the strand - either '+' or '-'.
- thickStart - The starting position at which the feature is drawn thickly (for example, the start codon in gene displays).

- thickEnd - The ending position at which the feature is drawn thickly (for example, the stop codon in gene displays).
- itemRgb - An RGB value of the form R,G,B (e.g. 255,0,0). If the track line itemRgb attribute is set to "On", this RBG value will determine the display color of the data contained in this BED line.
- blockCount - The number of blocks (exons) in the BED line.
- blockSizes - A comma-separated list of the block sizes. The number of items in this list should correspond to blockCount.
- blockStarts - A comma-separated list of block starts. All of the blockStart positions should be calculated relative to chromStart. The number of items in this list should correspond to blockCount.

Loading files on IGV

The Integrative Genomics Viewer (IGV) is a high-performance visualisation 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. Because NGS datasets are very large, it is often impossible or inefficient to read them entirely into a computer's memory. In order to retrieve quickly the data we are interested in, most programs have a way of treating these data files as databases. Database indexes enable one to rapidly pull specific subsets of the data from them.

Resources used

Integrative Genomics Viewer (IGV) : http://software.broadinstitute.org/software/igv/home

- Launch IGV and switch to the Human hg19 genome.
- From the main window of IGV, click on Genomes; then "Load Genome from URL" and specify the url http://igv.broadinstitute.org/genomes/hg19.genome .

There are a lot of things you can do in IGV. Here are a few:

- Zoom in using the slider in the upper right.
- Navigate by clicking and dragging in the window. This is how you move left and right along the genome.
- Navigate more quickly. Use page-up page-down, home, end.

- Jump right to a gene. Type its name into the search box. Try "TAL1".
- Change the appearance of genes. Right click on the gene track and try "expanded".
- You can load one or multiple bed files specifying them in the menu "File"; "Load from File".

Available bed files

MK_IDEAS_genome_segmentation.bed

EB_IDEAS_genome_segmentation.bed

2016-01-08-unstitched_EBs_strong_enhancers_IDEAS_minusP1000_collapsed_2_Gateway_SuperEnhancers.bed

2016-01-08-unstitched_MKs_strong_enhancers_IDEAS_minusP1000_collapsed_2_Gateway_SuperEnhancers.bed