ITX Integrative Transcriptomics
Dr. K. Nieselt, Faculty of Science

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

October 5 -7, 2015

Workshop: Visual Analytics of large-scale biological data

Practical Session:

Computing and visualizing RNA-Seq transcription start site data

During this practical session you will learn how you can use our *SuperGenome* approach in conjunction with our tool called *TSSpredator* to compute transcription start sites in several strains of a bacteria in parallel, and then view them using the genome browser IGB.

The data for this practical session is stored in the "Material4PracticalSessions/SuperGenome-TSSpredator" folder. This data is from a large study in *Campylobacter jejuni* [1].

TSSpredator applied to several different strains from a bacteria (organism in general) needs a multiple whole genome alignment of the respective strains' genomes. I have already precalculated that for you.

1. **Transcription start site prediction using *TSSpredator***

   (a) Create a subfolder in your folder for this analysis, which you will need to set as the output folder in TSSpredator.

   (b) In order to use TSSpredator, there is file called 'HowTo.txt'. In this case I have already preconfigured a config file, called 'Only1replicate_config.txt', that you can load via the load button. The config file restricts the analysis to just one of the 2 replicates and calculates transcription start sites only for one replicate of each of the 4 strains.

   (c) Run TSSpredator.

   (d) Look into the results folder that you specified in (1a). Check our first the overall results in 'TSSstatistics.tsv'. How would you now visualize these?

   (e) Next inspect the results from the overall 'MasterTable.tsv' and think about an appropriate and informative visualisation.

2. **View your results using *IGB***

   (a) Load the results into IGB: start off first with the genome file 'SuperConsensus.fa'.

   (b) Then first load all genomes in the SuperGenome coordinate system called '*_super.fa' together with their respective '*_TSS.gff'. Inspect! Discuss what would be interesting features to check out and visualize????

   (c) Next load also the RNA-seq graphs '*_superFivePrime*.gr' for the enriched data in the SuperGenome coordinate system together with their non-enriched data called '*_superNormal*.gr'.

   (d) Inspect! Discuss what would now be interesting features to check out and visualize????

References

[1] Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, Sharma CM. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. PLoS Genet 2013, 9(5):e1003495 (doi:10.1371/journal.pgen.1003495).