B!GRe
Bioinformatique des
Génomes et Réseaux

Regulatory Sequence Analysis Tools (*RSAT*)
Tutorial: command-line utilization of the tools

Jacques VAN HELDEN & the *RSAT* team

June 18, 2013

Contents

1 Regulatory variations (rSNPs and insertion/deletions)

This chapter presents a series of tools to analyze the impact of genetic polymorphism (inter-individual variations) on the binding of transcription factors.

It combines a series of tools to extract information on variations from the **EnsEMBL** database.

1.1 Installing genomes and variations from EnsEMBL

1.1.1 *download-ensembl-genome*

The program ***download-ensembl-genome*** downloads a genome from the Ensembl Web site, and installs it on the local ***RSAT*** site. This program relies on ***rsync***, which must be installed beforehand on your computer.

As usually, the complete help message can be obtained with the option `-help`.

```
## Get the description of the program + all options
download-ensembl-genome -help
```

Before downloading a genome, it is generally a good idea to know which genomes are available. For this, use the option `-available_species`.

```
## Retrieve the list of supported species on EnsEMBL
download-ensembl-genome  -available_species
```

2 References

1. van Helden, J., Andre, B. & Collado-Vides, J. (1998). Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J Mol Biol 281(5), 827-42.

2. van Helden, J., André, B. & Collado-Vides, J. (2000). A web site for the computational analysis of yeast regulatory sequences. Yeast 16(2), 177-187.

3. van Helden, J., Olmo, M. & Perez-Ortin, J. E. (2000). Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res 28(4), 1000-1010.

4. van Helden, J., Rios, A. F. & Collado-Vides, J. (2000). Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. Nucleic Acids Res. 28(8):1808-18.

5. van Helden, J., Gilbert, D., Wernisch, L., Schroeder, M. & Wodak, S. (2001). Applications of regulatory sequence analysis and metabolic network analysis to the interpretation of gene expression data. Lecture Notes in Computer Sciences 2066: 155-172.

6. van Helden, J. 2003. Prediction of transcriptional regulation by analysis of the non-coding genome. Current Genomics 4: 217-224.

7. van Helden, J. 2003. Regulatory sequence analysis tools. Nucleic Acids Res 31: 3593-3596.

8. van Helden, J. 2004. Metrics for comparing regulatory sequences on the basis of pattern counts. Bioinformatics 20: 399-406.

9. Simonis, N., J. van Helden, G.N. Cohen, and S.J. Wodak. 2004. Transcriptional regulation of protein complexes in yeast. Genome Biol 5: R33.

10. Simonis, N., S.J. Wodak, G.N. Cohen, and J. van Helden. 2004. Combining pattern discovery and discriminant analysis to predict gene co-regulation. Bioinformatics.