

Using vulcan, a package that combines regulatory networks and ChIP-Seq data to identify chromatin-binding cofactors

Federico M. Giorgi¹, Andrew N. Holding¹, and Florian Markowetz¹

¹Cancer Research UK, Cambridge Institute, Robinson Way, Cambridge United Kingdom

October 31, 2017

1 Overview of VULCAN

Vulcan (VirtUaL ChIP-Seq Analysis through Networks) is a pipeline that combines ChIP-Seq data and regulatory networks to obtain transcription factors that are likely affected by a specific stimulus. In order to do so, our package combines strategies from different BioConductor packages: *DESeq* for data normalization, *ChIPpeakAnno* and *DiffBind* for annotation and definition of ChIP-Seq genomic peaks, *csaw* to define optimal peak width and *viper* for applying a regulatory network over a differential binding signature. This data package contains a subset of a ChIP-Seq experiment where MCF7 cell lines were treated with estradiol and left incubating for 45 and 90 minutes. The data is available as BAM/BED/BAI files (one per sample) to illustrate how a *vulcan* pipeline could be executed on raw ChIP-Seq data. The dataset contains four replicates per time point (0, 45 and 90 minutes), including only chromosome 22 peaks for the sake of example feasibility. The *vulcandata* package contains also a dummy network of the *regulon* class as specified by the *viper* package. The package comes with a function, *vulcansheet*, which promptly generates a system-specific sample sheet for the dataset contained here, following the style of a *DiffBind* sample sheet.

2 Citation

Giorgi FM, Holding AN & Markowetz F. Network Dynamics of early ER-alpha promoter binding in response to estradiol. *Genome Biology* (Submitted).

3 Installation of *vulcandata* package

vulcandata requires the R-system (<http://www.r-project.org>) After installing R, all required components can be obtained with:

```
> source("http://bioconductor.org/biocLite.R")
> biocLite("vulcandata")
```

4 Getting started

As first step, we have to load the *vulcandata* package with:

```
> library(vulcandata)
```

5 Generate a example annotation file

```
> # Generate an annotation file from the dummy ChIP-Seq dataset
> vfile<-"deleteme.csv"
> vulcansheet(vfile)
> read.csv(vfile)
> unlink(vfile)
```

6 Access example objects

```
> # Example regulon object
> library(viper)
> load(system.file('extdata', 'network.rda', package='vulcandata', mustWork=TRUE))
> network
> # Example regulon object
> library(viper)
> load(system.file('extdata', 'network.rda', package='vulcandata', mustWork=TRUE))
> network
>
```

References

- [1] Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., & Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*.