GRAVI: Gene Regulatory Analysis Using Variable Inputs

Stevie Pederson[1,2], Geraldine Laven-Law[1], Richard Iggo[1,3], Amy Dwyer[1], Theresa E Hickey[1] and Wayne D Tilley[1]

[1]Dame Roma Mitchell Cancer Research Laboratories, Adelaide Medical School, University of Adelaide
[2]Black Ochre Data Labs, Telethon Kids Institute [3]Bordeaux Institute of Oncology, University of Bordeaux

GRAVI Outline

GRAVI is a `snakemake`[1] workflow for performing high-throughput, high-quality Differential Signal Analysis for ChIP-Seq data, exploiting the `extraChIPs` Bioconductor package. GRAVI standardises multiple steps to perform a *complete standalone analysis*, and also enables easier *integration across multiple complex datasets*. The workflow can be applied to *one or more* ChIP targets under *two or more* conditions. Optional RNA-Seq and HiC data further extend GRAVI to **directly map dynamic changes in ChIP signal to regulatory targets**.

Find GRAVI on GitHub

Key Aims

1. Best Practice Differential ChIP Signal Analysis
2. Accurate Mapping of Binding Sites to Regulatory Targets
3. Integration of multiple ChIP targets & treatments
4. Highly flexible input data

Input Data

Minimal Input

- 1×ChIP Target (.bam)
  −2× Conditions
- Gene Annotations (.gtf)
- Blacklisted Regions (.bed)

Optional Input

- Additional ChIP targets/treatments
- Differential Gene Expression (.tsv)
- HiC Interactions (.bedpe)
- Features of Interest (.gtf)
- External Coverage Tracks (.bw)

GRAVI Steps

Always Performed	1. Annotation Preparation 2. Peak Calling and Sample QC 3. Differential Signal Analysis
>1 Comparison Only	4. Pairwise Comparisons

Key Outputs

- Compiled Multi-Page HTML (Figures, Tables, `R` Code, etc)
- Bed Files (Peaks, Key Results)
- BigWig Files (Visualisation)
- Spreadsheets (Integrated results with mappings)
- `R` Data Objects for Custom Downstream Analysis

Motivation

Activation of the Androgen Receptor (AR) induces an *anti-proliferative phenotype* in many breast cancer models.[2] Integrating the dynamics of AR binding with the Estrogen Receptor (ERα), additional transcription factors, H3K27ac marks and changes in transcription, we were able to investigate the underlying molecular mechanisms. Comparison across multiple cancer models was then used to identify key targets and mechanisms.

Differential Signal Analysis

Always Performed

1. Sliding Windows[3] using either
   (a) Quasi-Likelihood Models,[4] or
   (b) SQN[5] with limma-trend[6]
2. Range-Based H_0[7]
3. Independent Hypothesis Weighting[8]
4. Mapped to Genes, Regulatory Regions and Optional Features
5. Tables, Figures and Bed files
6. Enrichment Analysis

If DGE Results Included

1. Direct Targets Identified
2. *Combined* Enrichment Analysis

Coming Soon

1. Fixed-Width Windows
2. More Normalisation Options
3. ATAC-Seq Methods

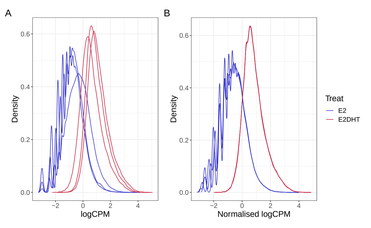

Fig. 1: AR logCPM values A) before and B) after Smooth Quantile Normalisation. AR is translocated from the cytoplasm to the nucleus when activated by DHT

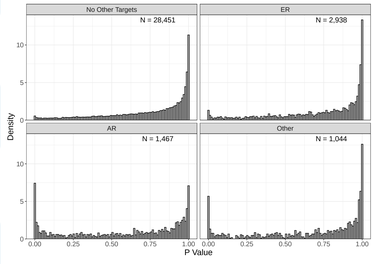

Fig. 2: Partitioned *p*-values for H3K27ac Differential Signal Analysis based on co-detection of AR and ER. Here, 21% more windows were considered significant

Pairwise Comparisons

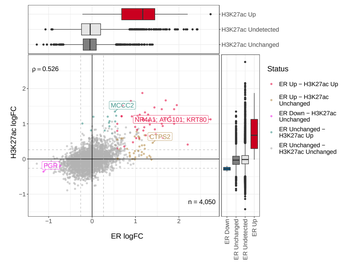

Fig. 3: Pairwise DSA results for ER and H3K27ac

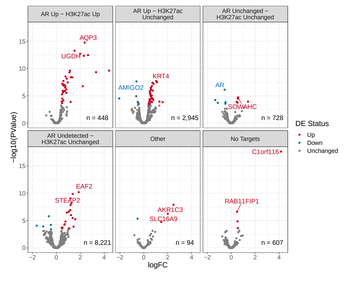

Fig. 4: DE Genes by combined AR and H3K27ac signal.

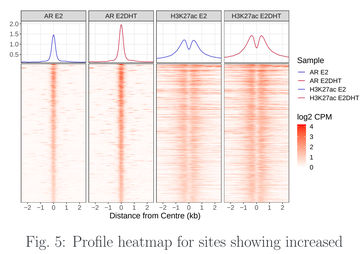

Fig. 5: Profile heatmap for sites showing increased signal for both AR and H3K27ac.

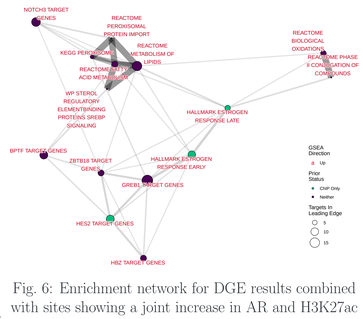

Fig. 6: Enrichment network for DGE results combined with sites showing a joint increase in AR and H3K27ac

All examples taken from an analysis of AR, ER and H3K27ac in ZR-75 cells treated with DHT

References

[1] Felix Mölder et al. "Sustainable data analysis with Snakemake". en. In: *F1000Res.* 10 (Jan. 2021), p. 33.

[2] Theresa E. Hickey et al. "The androgen receptor is a tumor suppressor in estrogen receptor–positive breast cancer". In: *Nature Medicine 2021 27:2* 27.2 (Jan. 2021), pp. 310–320. ISSN: 1546-170X.

[3] Aaron T L Lun and Gordon K Smyth. "csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows". In: *Nucleic Acids Res.* 44.5 (2016), e45.

[4] Steven P Lund et al. "Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates". en. In: *Stat. Appl. Genet. Mol. Biol.* 11.5 (Oct. 2012).

[5] Stephanie C Hicks et al. "Smooth quantile normalization". In: *Biostatistics* 19.2 (July 2017), pp. 185–198. ISSN: 1465-4644.

[6] Charity W Law et al. "voom: Precision weights unlock linear model analysis tools for RNA-seq read counts". en. In: *Genome Biol.* 15.2 (Feb. 2014), R29.

[7] Davis J McCarthy and Gordon K Smyth. "Testing significance relative to a fold-change threshold is a TREAT". en. In: *Bioinformatics* 25.6 (Mar. 2009), pp. 765–771.

[8] Nikolaos Ignatiadis et al. "Data-driven hypothesis weighting increases detection power in genome-scale multiple testing". en. In: *Nat. Methods* 13.7 (July 2016), pp. 577–580.

See A Complete Report Here