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-quality ChIP-Seq analysis. GRAVI standardises multiple steps to enable a *complete standalone analysis*, as well as enabling *integration across multiple cell types*. The minimal required data is a single ChIP target under 2 conditions expanding to *any number of ChIP targets and conditions*. Optional RNA-Seq and HiC data further extend GRAVI to **directly map dynamic changes in ChIP signal to regulatory targets**.

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

Required 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. Peak Calling and Sample QC
2. Differential Signal Analysis

Performed Only When Viable

1. Pairwise Comparisons

Key Outputs

An extensive library of results for sharing with collaborators

- Compiled Multi-Page HTML (Figures, Tables, Descriptions)
- 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) has *anti-proliferative properties in many breast cancer models*[2]. Integrating the dynamics of AR binding with the Estrogen Receptor α (ERα), additional transcription factors, H3K27ac marks and changes in transcription, we are able to investigate the underlying molecular mechanisms. Comparison across cancer models can then be used to identify key targets and mechanisms.

Differential Signal Analysis

Consistent Steps

1. Sliding Windows[3] using
   (a) Quasi-Likelihood Models[4], or
   (b) SQN[5] with limma-trend[6]
2. Range-Based H_0[7]
3. Independent Hypothesis Weighting (IHW)[8]
4. Mapped to Genes, Regulatory Regions and External Features
5. Enrichment Analysis
   - Tables and Network Plots
6. Multiple tables, figures and bed files

Steps With RNA-Seq

1. Direct Targets Identified
2. *Combined* Enrichment Analysis
   - Tables and Network Plots

To Be Implemented Next

1. Alternative Fixed-Width Analysis
2. TMM/RLE Normalisation
3. ATAC-Seq Methods

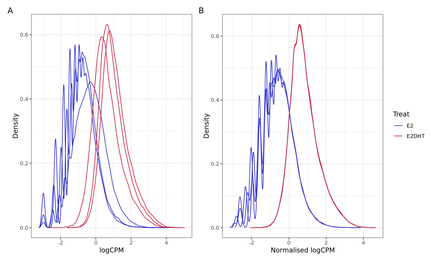

Fig. 1: AR logCPM values before and after Smooth Quantile Normalisation. AR is primarily cytoplasmic then shifts to be nuclear under E2+DHT

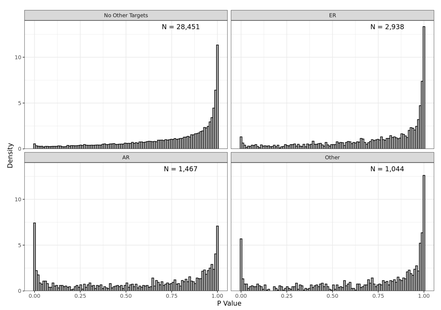

Fig. 2: Partitioned p-values for H3K27ac signal based on co-detection of AR and ER.

Pairwise Comparisons

Given prior rigour in statistical testing, comparison across pairs of factors becomes more of a *site classification problem*. When a site is found to be responsive for one transcription factor a lower p-value threshold is applied to the second factor in any overlapping sites, ensuring **unchanged sites are classified more accurately** as these are important in a pairwise analysis. Each site is classified across both ChIP targets as *Up, Down, Unchanged or Undetected* yielding a set of pairwise classifications. Comparisons of changed binding (Figure 3) are generated across all sites, and separated by annotated region or external feature. Profile heatmaps are also created (Figure 4) and enrichment testing is performed on each set of regions. Example sites for each set of regions are also plotted by default using `extraChIPs::plotHFGC()` (Figure 5). If RNA-Seq data is provided, an analysis of DE genes by pairwise changes in ChIP target binding is also performed (Figure 6).

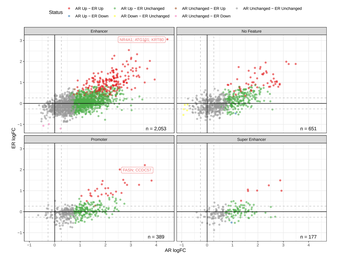

Fig. 3: Pairwise logFC values for AR and ER broken down by externally provided H3K27ac-defined features.

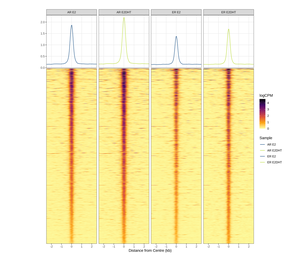

Fig. 4: Profile heatmap for sites showing increased binding for both AR and ER.

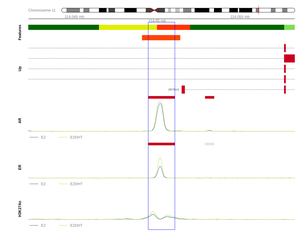

Fig. 5: *ZBTB16* is up-regulated, shows increased binding for both AR and ER, with an increase in H3K27ac signal

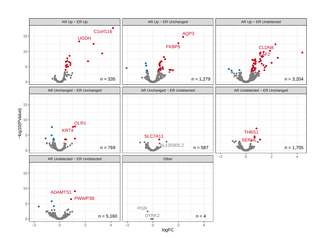

Fig. 6: DE Genes by AR and ER binding patterns. Here, the top row would be the most informative.