The RNA-seq analysis of an HDAC inhibitor, ITF2357 (Givinostat), on mice with diastolic dysfunction.

By Harrison Smith

Outline

- **Background**
- **Procedure details**
- **QC of samples**
- **Differentially expressed Genes/Gene set enrichment**
- **Conclusion**

Background

- Diastolic dysfunction (DD) - Abnormal left ventricle functionality
- Studies indicate DD is attributed to cardiac myofibril relaxation.
- HDAC inhibitor, ITF2357, restores Diastolic functionality.
- RNA-seq analysis provides an observation of genetic expression

Aim

- Observe the effects of HDAC inhibition.
    - Determine differentially expressed genes in mice
    - Look at enrichment of curated gene sets
        - Infer positive and negative correlations with phenotype data with gene sets

Procedure

Summary

- Total RNA was isolated and sequenced from the whole hearts of mice.
- 20 biological replicates across two time points,
    - 4 weeks and 8 weeks, post uninephrectomy(UNX) surgery.

Negative Control	:	UNX + veh
Positive Control	:	UNX/DOCA + veh
Experimental group	:	UNX/DOCA + treatment

Procedure

Prep Protocol

- Whole heart cell tissues were lysed in the Zymo TRI Reagent
- The mRNA was poly-A selected
- Paired end reads

RNA extraction/purification	:	Zymo R2050 Direct-zol RNA Miniprep Kit
RNA library prep	:	Ultra-directional RNA library prep NEBNext Kit
NGS Sequencing	:	Illumina NovaSeq 6000

Procedure

Downstream NGS analysis

Alignment + quantification

Fastqs aligned to mm10 (mouse) reference genome from UCSC using HISAT2

Replicates were quantified and normalized to determine the overall gene expression

Cufflinks → Cuffquant → Cuffnorm tool set

Dr. Lin's R script was used for pairwise correlations, normalized expression matrix, and GSEA ready inputs.

Quality Control

Visual of normalized gene expression levels across each biological group.

Used for quickly examining outliers or contaminations that might skew further analysis

Figures:

1. Box-plot represents unnormalized expression

2. Replicate scatter plots of normalized expression

Quality Control

The log2 expression here represents the unnormalized expression levels of total genes among replicates.

The first figure to assess before viewing pairwise correlations.

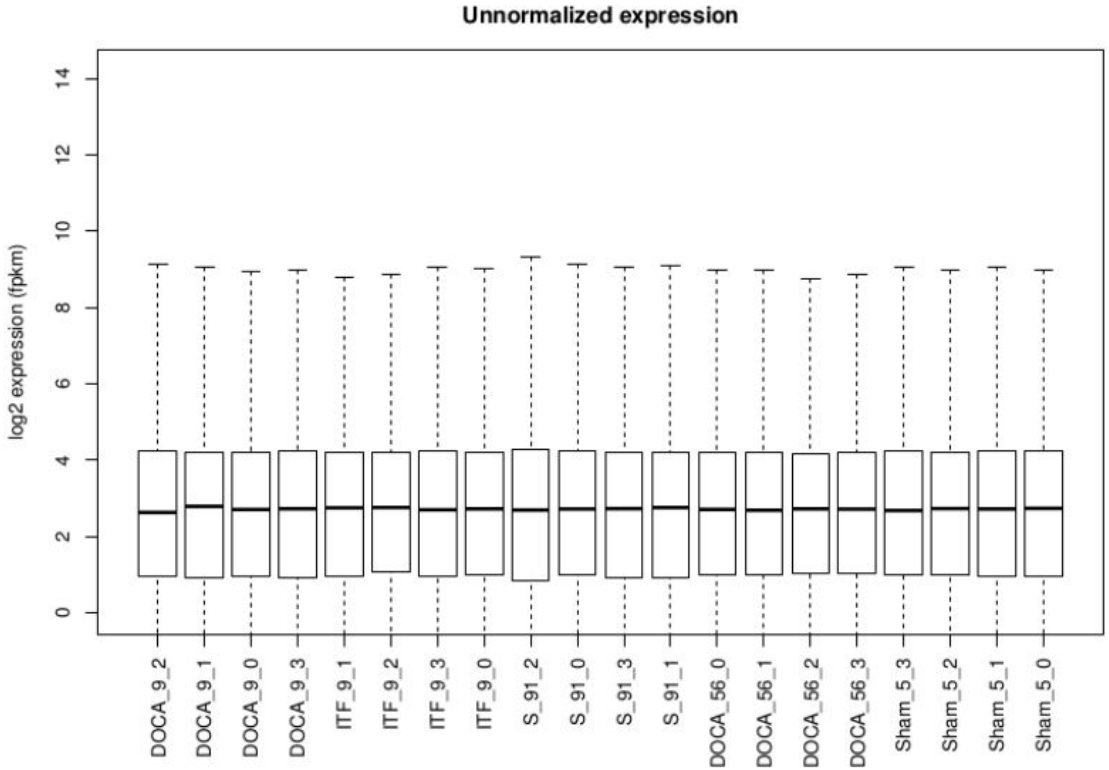
DOCA_5 : +Control mice 4 weeks

DOCA_9 : +Control mice 8 weeks

Sham_5 : -Control mice 4

S_9 : -Control 8 weeks

ITF_9 : exp + inhibitor 8 weeks



Quality Control

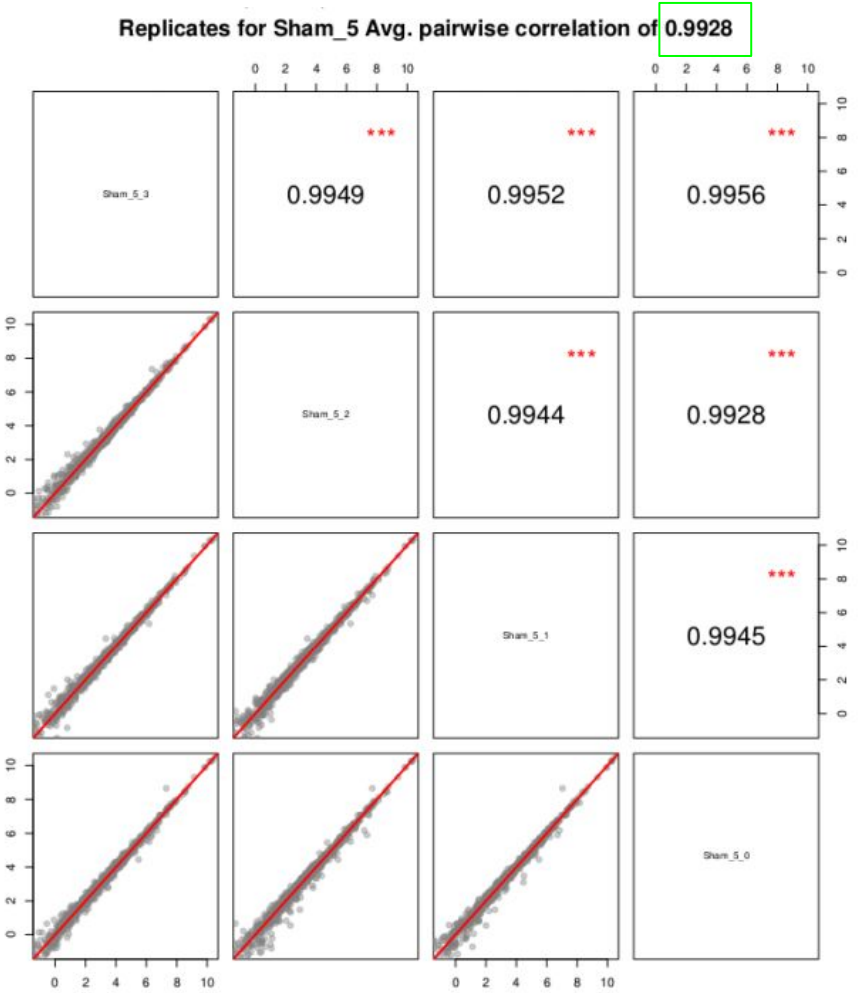
Replicate scatter plot

normalized gene expression levels which are filtered by genes.

All pairwise comparisons of replicates are plotted by group.

Helpful in quickly identifying poor quality samples. The title of each figure provides the average of all comparisons in a replicate group.

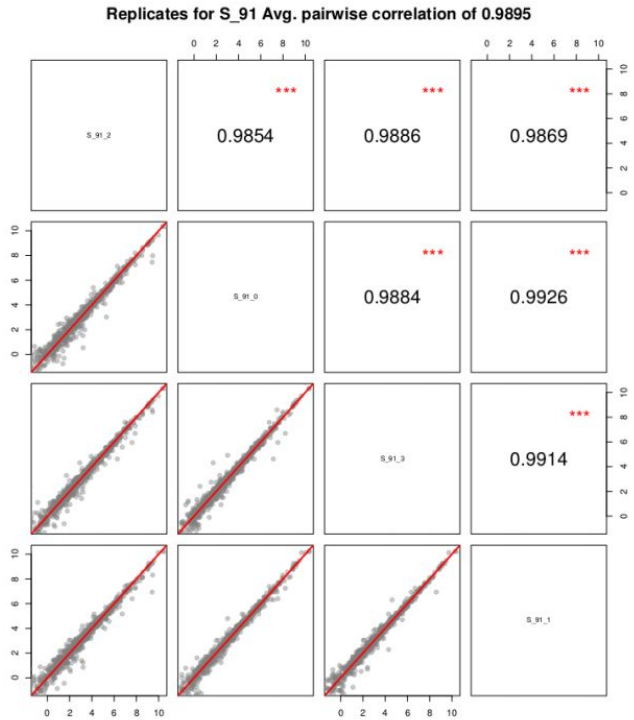
Here we have high similarity among **Sham 4 weeks** replicates at avg 0.9928



Quality Control

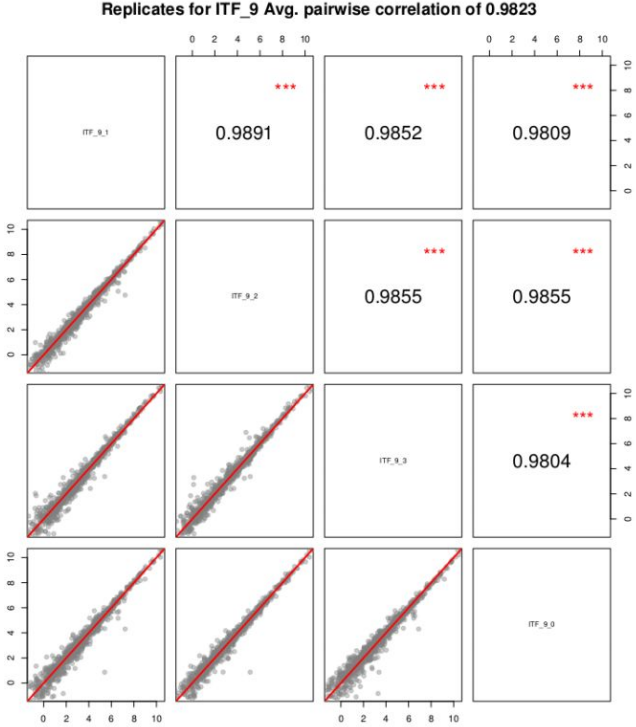
Sham 8

0.9895



ITF

0.9823

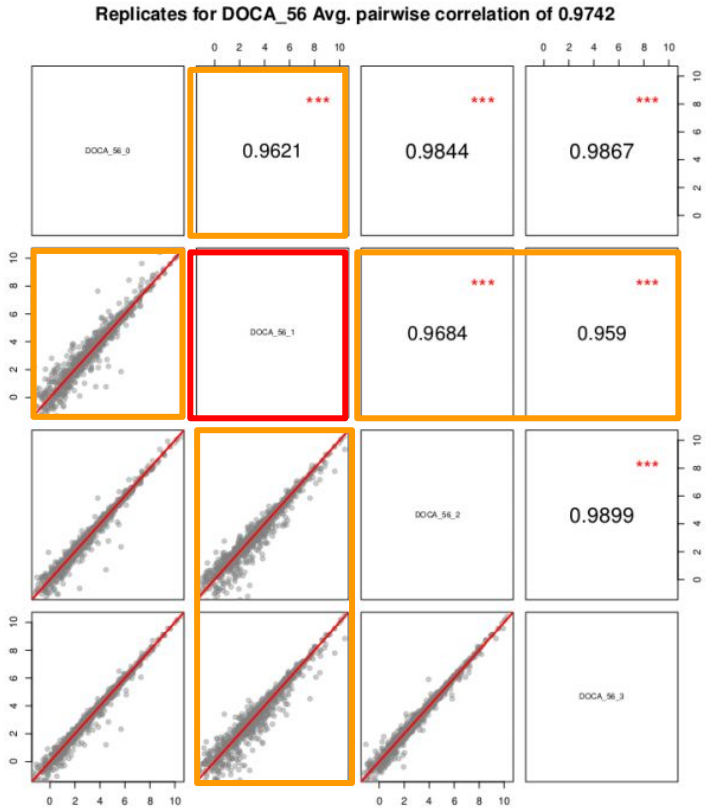


Quality Control

DOCA 4 weeks

A replicate group boxed in red appears to cause a noticeable deviation compared to the other groups.

This replicate was dropped prior to downstream analysis.

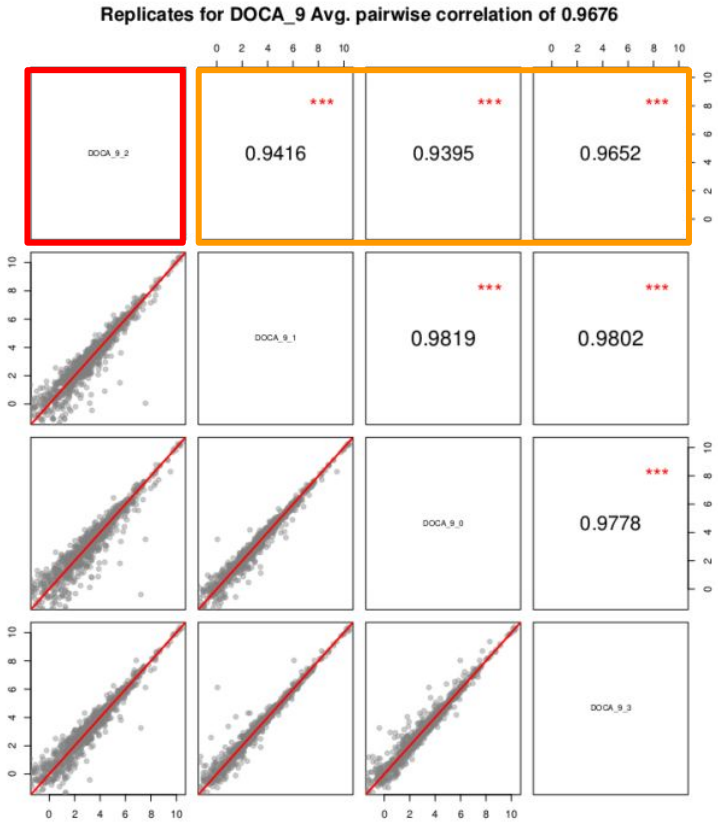


Quality Control

DOCA at 8 weeks

Replicate # 2 shows an even weaker correlation.

Will also be dropped.

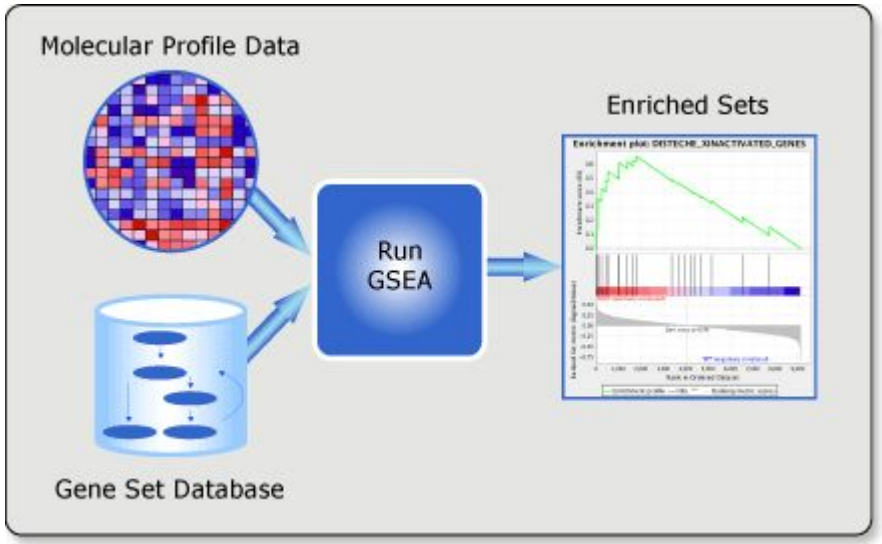


Gene set enrichment

Gene set enrichment analysis was performed with GSEA

"*gene sets*, which represent group of genes that are grouped together based on their common biological function and/or involvement in the same biological pathways"

Curated sets come from the Molecular Signatures Database (MSigDB) on the Broad Institute web page



GSEA

The peak of the graph represent the Enrichment Score

From left to right is a list of ranked genes from most upregulated to most downregulated

Positive score shows correlation with up regulated genes in the subset. (leading edge subset)

Negative score indicates downregulation correlation

Example:

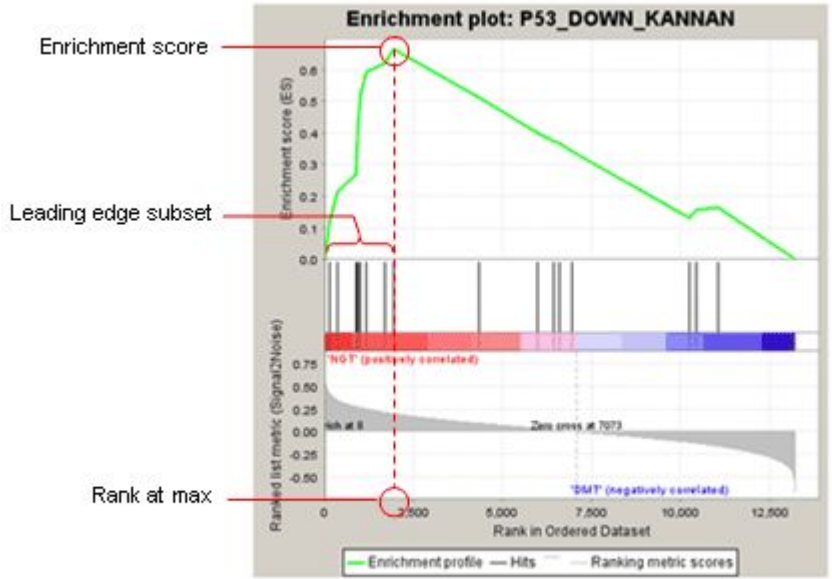


Fig 1: Enrichment plot: P53_DOWN_KANNAN
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

GSEA

The entire GSEA output documentation for each enriched pathway is provided in the project folder.

GSEA provides top 20 enrichment plots for upregulation of phenotype 1 and 20 plots for downregulation of phenotype 2 in comparison but there are typically more.

Meaning, enrichment plots for Phenotype 1 show upregulation enrichment in a pathway (gene set) compared to the phenotype 2 genes.

Enrichment plots for phenotype 2 in the same output show correlation with down regulated genes in that pathway.

Viewing individual enrichment plots and scrolling down will provide the leading edge subset genes (significant enrichment of DE genes) that make up the enriched pathway.

GSEA (example)

Here we see enrichment for downregulated genes of this pathway for Sham 8

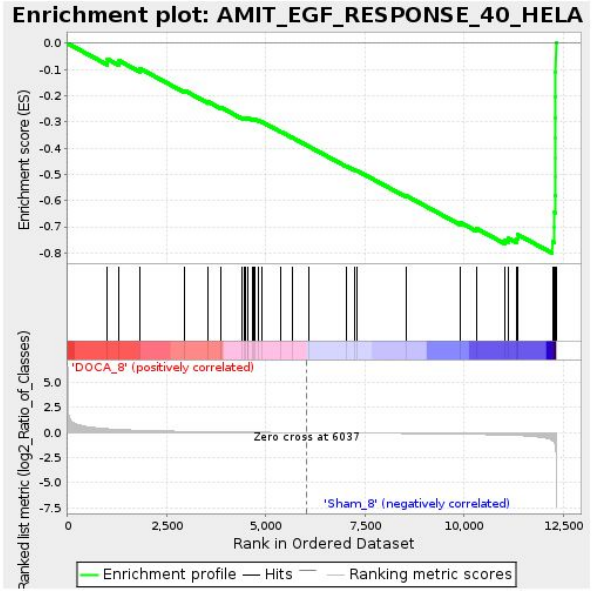


Fig 1: Enrichment plot: AMIT_EGF_RESPONSE_40_HELA
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

This plot shows enrichment for upregulated genes in this pathway

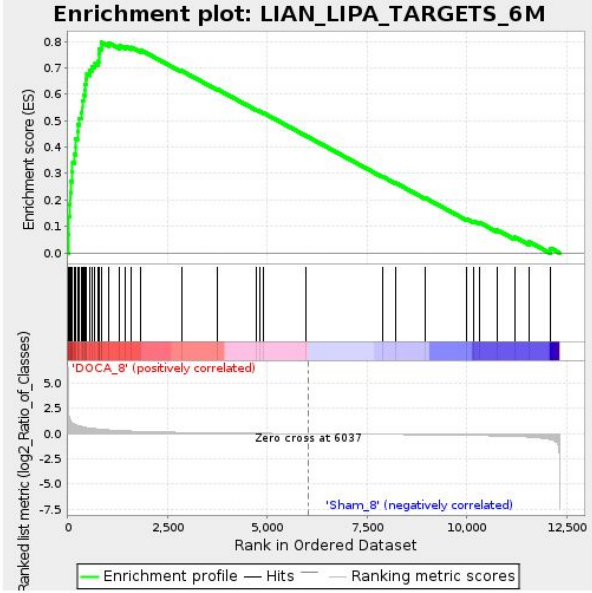


Fig 1: Enrichment plot: LIAN_LIPA_TARGETS_6M
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

GSEA

D8 vs S8

- Top 5 enriched pathways for **up-regulation** of genes in DOCA 8
- The rest can easily be viewed in the gsea project fold.
- Just open the index html and click on snapshot of results

1.Lian-LIPA_targets

2. Verhaak glioblastoma mesenchymal

3. Invasive Breast cancer

4. Mclachlan dental caries

5. Kim GLI52_targets_up

GSEA

D8 vs S8

- Top 5 enriched pathways for **down-regulation** of genes in Sham 8 group
- This is relative to DOCA 8 group

1.AMIT EGF response - HELA

2. Jechlinger Epithelial mesenchymal transition

3. AMIT serum response

4. AMIT EGF response - MCF10A

5. RASHI response to IONIZING radiation

GSEA

D8 vs I8

- Top 5 enriched pathways for **up-regulation** of genes in DOCA8 group
- Compared to ITF

1. **Altemeier response to lps with mechanical ventilation**
2. **Poola invasive breast cancer**
3. **Mclachlan dental caries up**
4. **Croonquist IL6 deprivation dn**
5. **Croonquist NRAS signaling**

GSEA

D8 vs I8

- Top 5 enriched pathways for **down-regulation** of genes in ITF 8 group
- This is relative to DOCA 8 group

1.REACTOME peptide chain elongation

2. KEGG ribosome

3. AMIT serum response

4. AMIT EGF response - MCF10A

5. RASHI response to IONIZING radiation

GSEA

I8 vs S8

- Top 5 enriched pathways for **up-regulation** of genes in ITF 8 group
- This is relative to Sham 8 group

1.Sotiriou breast cancer grade 1 vs 3 up

2. Whiteford pediatric cancer markers

3. Rosty cervical cancer proliferation cluster

4. Lee early T lymphocyte up

5.Mori Large Pre BII lymphocyte up

GSEA

I8 vs S8

- Top 5 enriched pathways for **down-regulation** of genes in Sham 8 group
- This is relative to ITF 8 group

1.REACTOME peptide chain elongation

2. KEGG ribosome

3. AMIT serum response

4. AMIT EGF response - MCF10A

5. RASHI response to IONIZING radiation

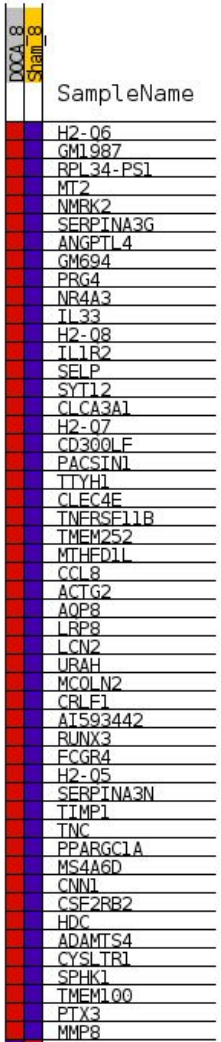
GSEA

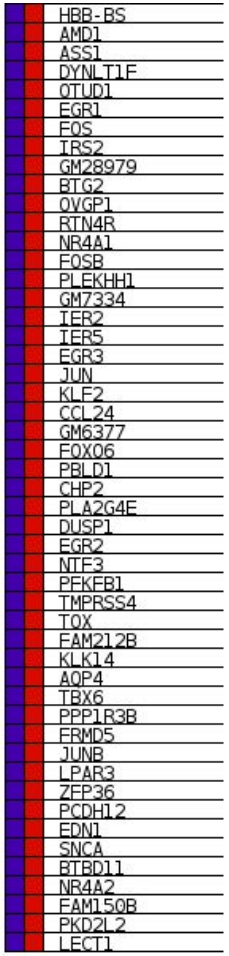
Enrichment in phenotype: DOCA_8 (1 samples)

- 2389 / 3310 gene sets are upregulated in phenotype **DOCA_8**
- 588 gene sets are significant at FDR < 25%
- 271 gene sets are significantly enriched at nominal pvalue < 1%
- 565 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: Sham_8 (1 samples)

- 921 / 3310 gene sets are upregulated in phenotype **Sham_8**
- 50 gene sets are significantly enriched at FDR < 25%
- 44 gene sets are significantly enriched at nominal pvalue < 1%
- 131 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results





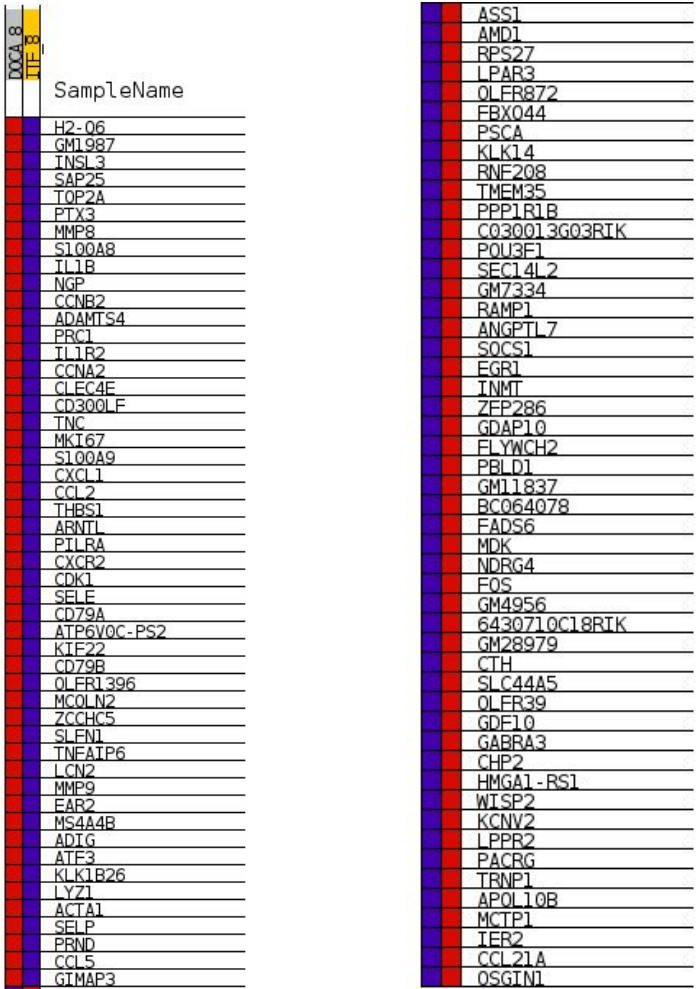
GSEA

Enrichment in phenotype: DOCA_8 (1 samples)

- 2256 / 3310 gene sets are upregulated in phenotype **DOCA_8**
- 1007 gene sets are significant at FDR < 25%
- 439 gene sets are significantly enriched at nominal pvalue < 1%
- 768 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: ITF_8 (1 samples)

- 1054 / 3310 gene sets are upregulated in phenotype **ITF_8**
- 144 gene sets are significantly enriched at FDR < 25%
- 82 gene sets are significantly enriched at nominal pvalue < 1%
- 187 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results



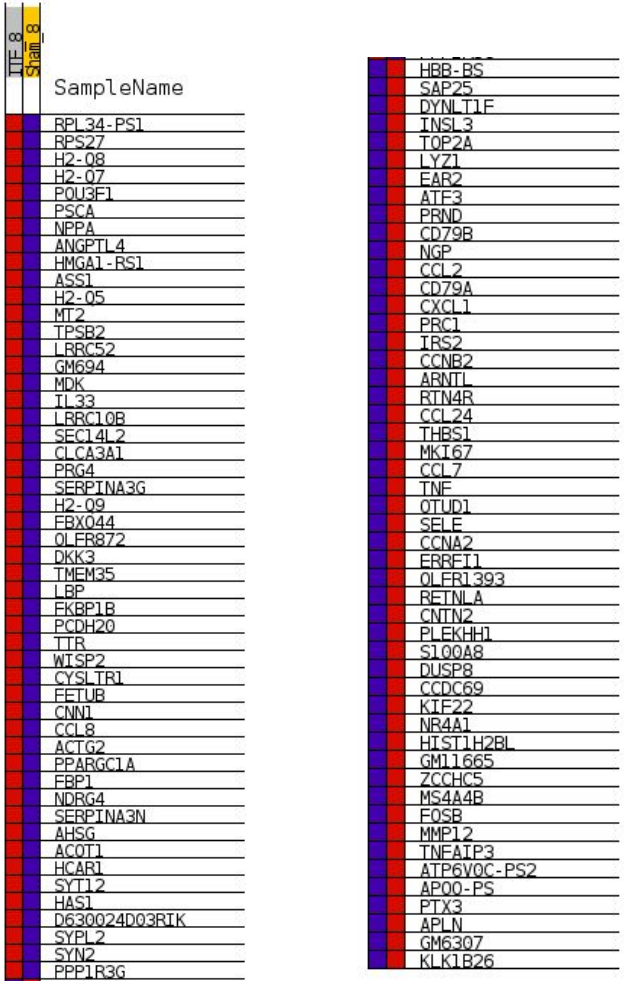
GSEA

Enrichment in phenotype: ITF_8 (1 samples)

- 1658 / 3310 gene sets are upregulated in phenotype **ITF_8**
- 164 gene sets are significant at FDR < 25%
- 125 gene sets are significantly enriched at nominal pvalue < 1%
- 264 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: Sham_8 (1 samples)

- 1652 / 3310 gene sets are upregulated in phenotype **Sham_8**
- 475 gene sets are significantly enriched at FDR < 25%
- 219 gene sets are significantly enriched at nominal pvalue < 1%
- 428 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results



Conclusion

By viewing the index for each group you can determine the quantity of enriched gene sets.

Comparing the two reveals significant up and down regulation of genes for the specified pathway.

This helps narrow down enrichment for pathways that are present in the control groups