

**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(UNIX) surgery.
-

Negative Control	:	UNIX + veh
Positive Control	:	UNIX/DOCA + veh
Experimental group	:	UNIX/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.

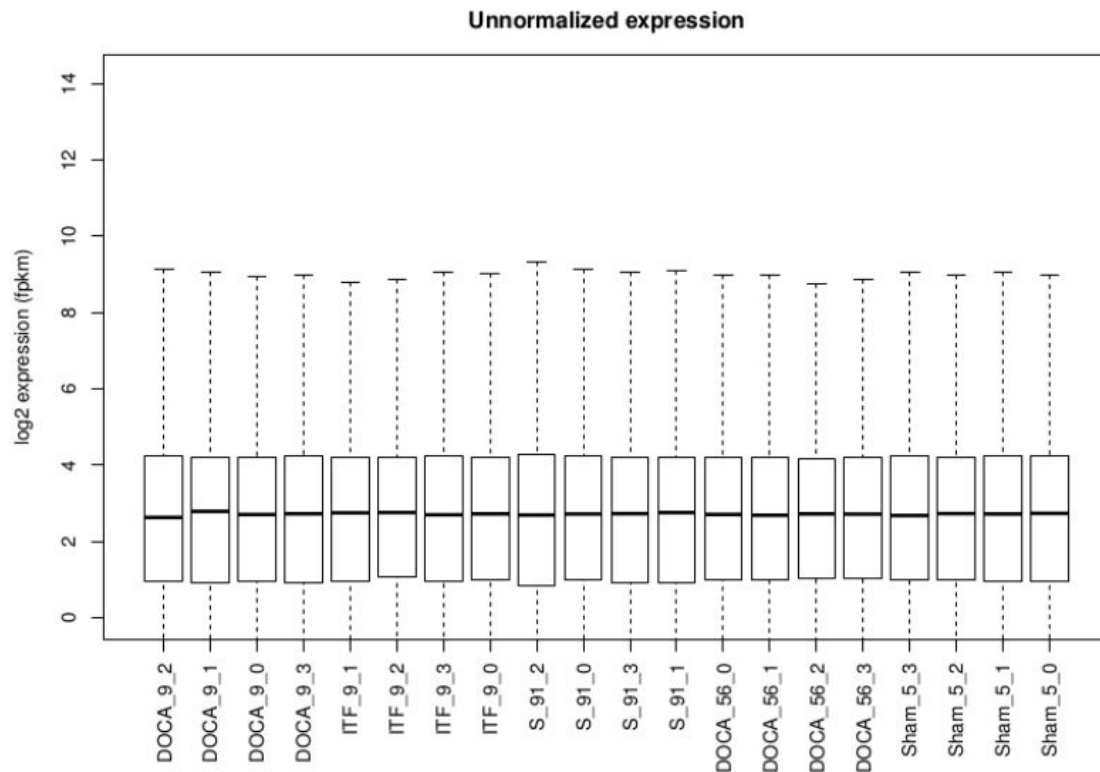
DOCA_5 : +Control mice 4 weeks

DOCA_9 : +Control mice 8 weeks

Sham_5 : -Control mice 4

S_9 : -Control 8 weeks

ITF_9 : exp + inhibitor 8 weeks



Quality Control

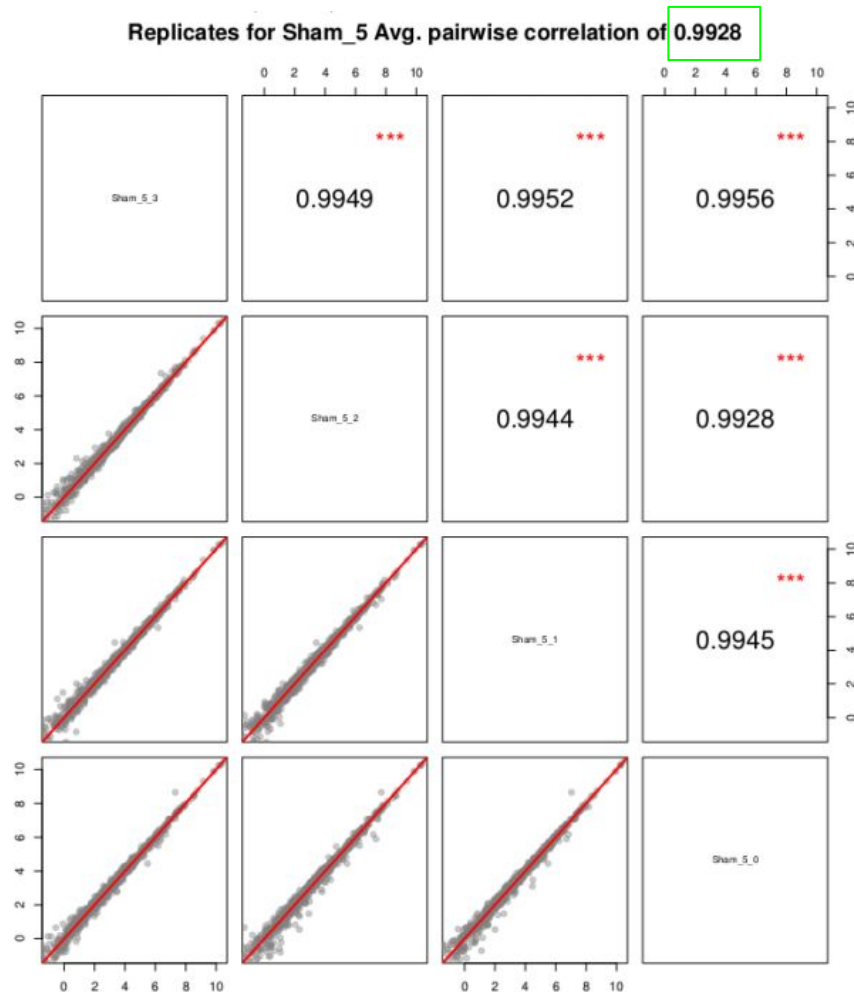
Replicate scatter plot

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.

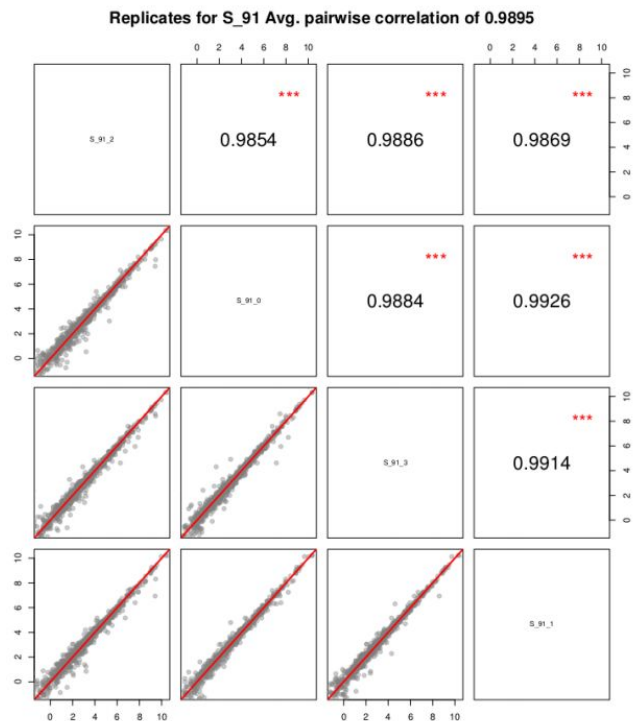
Here we have high similarity among **Sham 4 weeks** replicates at avg 0.9928



Quality Control

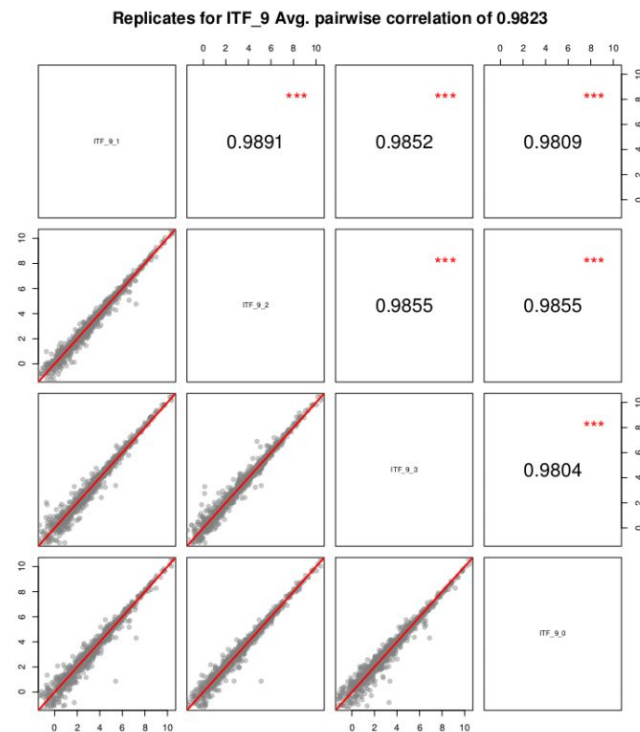
Sham 8

0.9895



ITF

0.9823

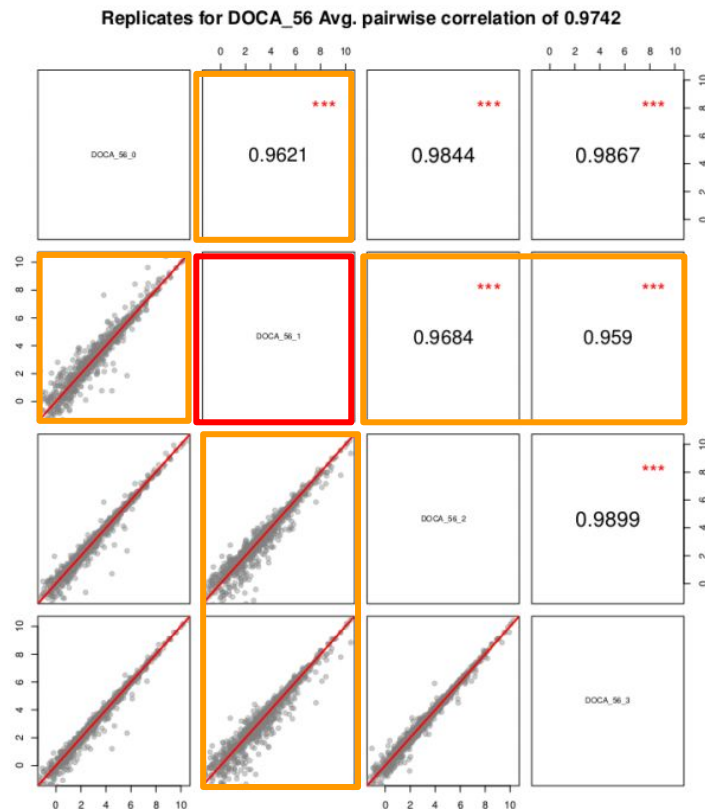


Quality Control

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.

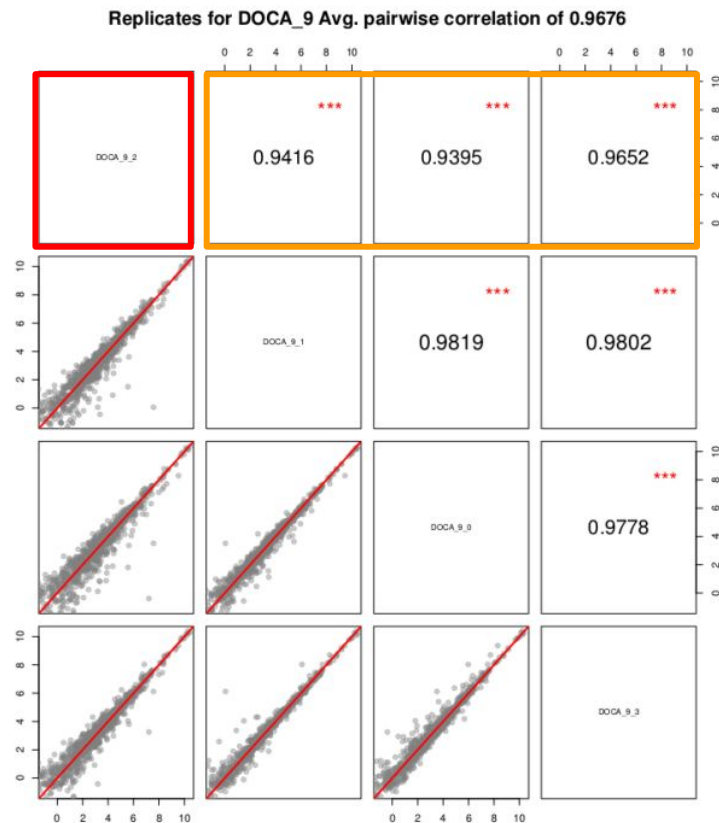


Quality Control

DOCA at 8 weeks

Replicate # 2 shows an even weaker correlation.

Will also be dropped.

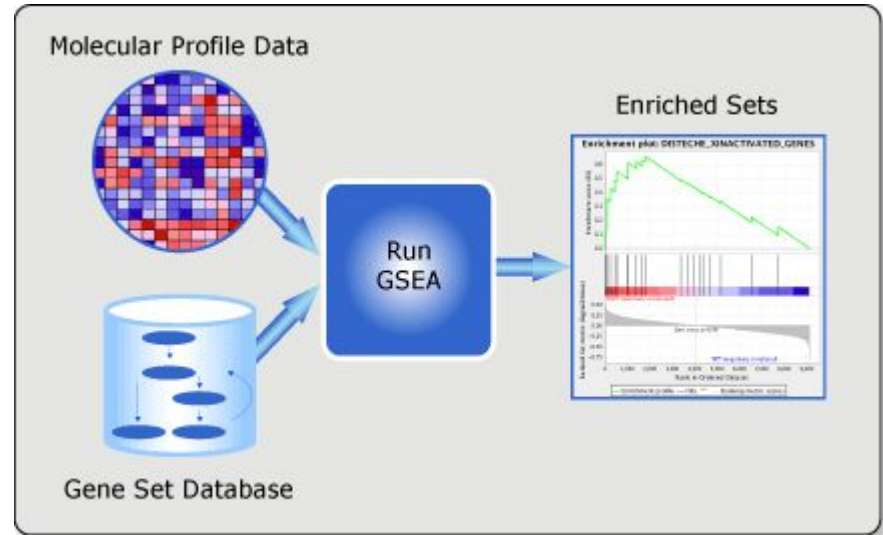


Gene set enrichment

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:

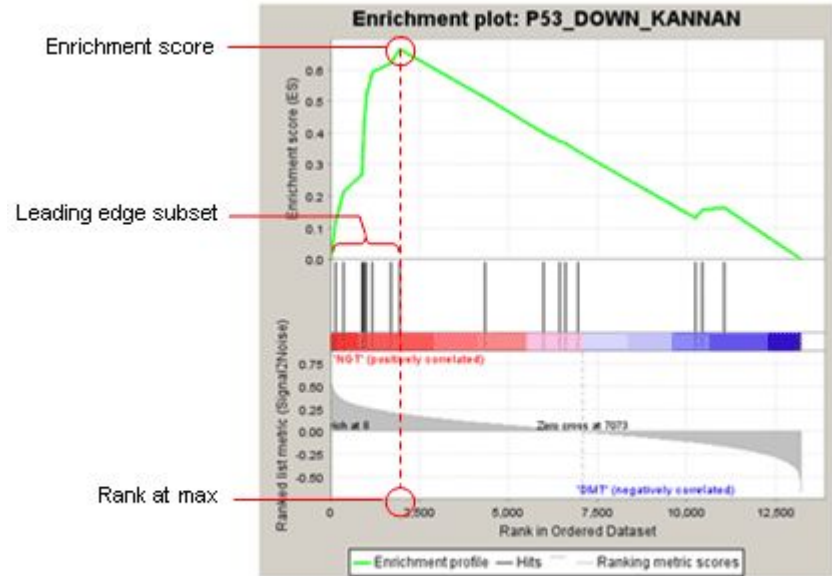


Fig 1: Enrichment plot: P53_DOWN_KANNAN
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

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

This plot shows enrichment for upregulated genes in this pathway

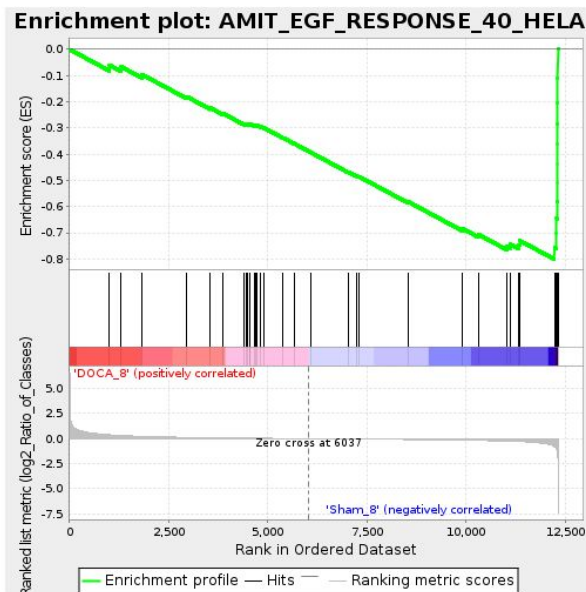


Fig 1: Enrichment plot: AMIT_EGF_RESPONSE_40_HELA

Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

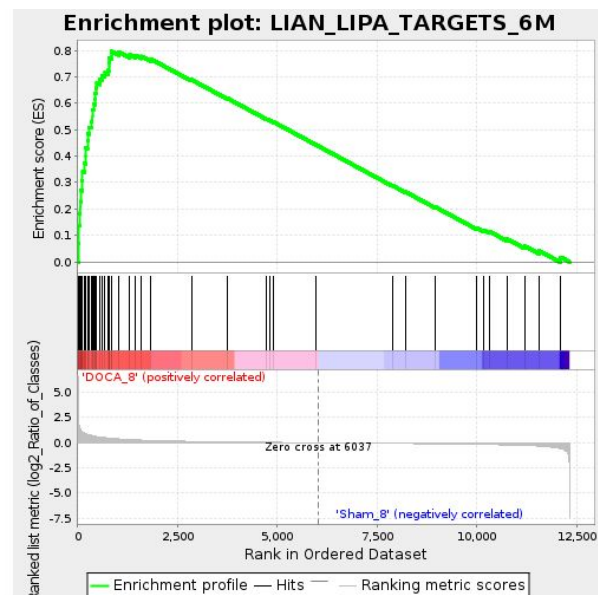


Fig 1: Enrichment plot: LIAN_LIPA_TARGETS_6M

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 Ips 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 significant 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

DOCA_8	Sham_8	SampleName
		H2-06
		GM1987
		RPL34-PS1
		MT2
		NMRK2
		SERPINA3G
		ANGPTL4
		GM694
		PRG4
		NR4A3
		IL33
		H2-08
		IL1R2
		SELP
		SYT12
		CLCA3A1
		H2-07
		CD300LF
		PACSTN1
		TTYH1
		CLEC4E
		TNFRSF11B
		TMEM252
		MTHFD1L
		CCL8
		ACTG2
		AQP8
		LRP8
		LCN2
		URAH
		MCOLN2
		CBLE1
		AT593442
		RUNX3
		FCGR4
		H2-05
		SERPINA3N
		TIMP1
		TNC
		PPARGC1A
		MS4A6D
		CNN1
		CSF2RB2
		HDC
		ADAMTS4
		CYSLTR1
		SPHK1
		TMEM100
		PTX3
		MMP8
		HBB-BS
		AMD1
		ASS1
		DYNLT1F
		OTUD1
		EGR1
		FOS
		IRS2
		GM28979
		BTG2
		OVGP1
		RTN4R
		NR4A1
		FOSB
		PLEKHH1
		GM7334
		IER2
		IER5
		EGR3
		JUN
		KLF2
		CCL24
		GM6377
		FOXO6
		PBLD1
		CHP2
		PLA2G4E
		DIUSP1
		EGR2
		NTE3
		PEKFB1
		TMPRSS4
		TOX
		FAM212B
		KLK14
		AQP4
		TBX6
		PPP1R3B
		FRMD5
		JUNB
		LPAR3
		ZFP36
		PCDH12
		EDN1
		SNCA
		RTBD11
		NR4A2
		FAM150B
		PKD2L2
		LECT1

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 significant 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

DOCA_8	ITF_8	SampleName
		H2-06
		GM1987
		INSL3
		SAP25
		TOP2A
		PTX3
		MMP8
		S100A8
		IL1B
		NGP
		CCNB2
		ADAMTS4
		PRC1
		IL1R2
		CCNA2
		CLEC4E
		CD300LF
		TNC
		MKI67
		S100A9
		CXCL1
		CCL2
		THBS1
		ARNTL
		PILRA
		CXCR2
		CDK1
		SELE
		CD79A
		ATP6V0C-PS2
		KIF22
		CD79B
		OLFR1396
		MCOLN2
		ZCCHC5
		SLFN1
		TNFAIP6
		L CN2
		MMP9
		EAR2
		MS4A4B
		ADIG
		ATF3
		KLK1B26
		LYZ1
		ACTA1
		SELP
		PRND
		CCL5
		GIMAP3
		ASS1
		AMD1
		RPS27
		LPAR3
		OLFR872
		FBX044
		PSCA
		KLK14
		RNF208
		TMEM35
		PPP1R1B
		C030013G03RTK
		POU3F1
		SEC14L2
		GM7334
		RAMP1
		ANGPTL7
		SOCS1
		EGR1
		INMT
		ZFP286
		GDAP10
		FLYWCH2
		PBLD1
		GM11837
		BC064078
		FADS6
		MDK
		NDRG4
		FOS
		GM4956
		6430710C18RTK
		GM28979
		CTH
		SLC44A5
		OLFR39
		GDF10
		GABRA3
		CHP2
		HMGAI-RS1
		WISP2
		KCNV2
		LPPR2
		PACRG
		TRNP1
		APOL10B
		MCTP1
		IER2
		CCL21A
		OSGIN1

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

ITF_8	Sham_8	SampleName
		HBB-BS
		SAP25
		DYNLT1F
		INSL3
		TOP2A
		LYZ1
		EAR2
		ATF3
		PRND
		CD79B
		NGP
		CCL2
		CD79A
		CXCL1
		PRC1
		IRS2
		CCNB2
		ARNTL
		RTN4R
		CCL24
		THBS1
		MKI67
		CCL7
		TNF
		OTUD1
		SELE
		CCNA2
		EBRF11
		OLFR1393
		RETNLA
		CNTN2
		PLEKH1
		S100A8
		DUSP8
		CCDC69
		KIF22
		NR4A1
		HIST1H2BL
		GM11665
		ZCCHC5
		MS4A4B
		F0SB
		MMP12
		TNFAIP3
		ATP6VOC-PS2
		AP00-PS
		PTX3
		APLN
		GM6307
		KLK1B26
		RPL34-PS1
		RPS27
		H2-08
		H2-07
		PQU3F1
		PSCA
		NPPA
		ANGPTL4
		HMGAI-RS1
		ASS1
		H2-05
		MT2
		TPSB2
		LRRC52
		GM694
		MDK
		IL33
		LRRC10B
		SEC14L2
		CLCA3A1
		PRG4
		SERPINA3G
		H2-09
		FBX044
		OLFR872
		DKK3
		TMEM35
		LBP
		EKBP1B
		PCDH20
		TTR
		WISP2
		CYSLTR1
		FETUB
		CNN1
		CCL8
		ACTG2
		PPARGC1A
		FBP1
		NDRG4
		SERPINA3N
		AHSG
		ACOT1
		HCAR1
		SYT12
		HAS1
		D630024D03RTK
		SYPL2
		SYN2
		PPP1R3G

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