



# *GLM for testing differential expression and GSEA in Galaxy*

**GCC2013 Lightning talk**

***Ross Lazarus, Mark Ziemann and Antony Kaspi.***

# Motivation

- Context: RNA-seq → GLM → DEG
- eg: Paired (L/R mouse legs) Rx/Ctrl
- Not available with current (RPKM) tools
- Negative binomial GLM (eg edgeR)
- Cuff\* → absolute isoform abundance
- Bias using RPKM for DEG models?
- Dillies et al., <http://bib.oxfordjournals.org/content/early/2012/09/15/bib.bbs046.long>

# Models for raw counts

- Replicates essential (even DESeq!)
- edgeR – NB; TMM; GLM
- DESeq – NB; local dispersion
- VOOM transformation → limma
- (others not yet implemented..)

# HTSeq to prepare count matrix

- Supply a GTF for the gene model
- User choices for counting
- eg: exonic reads → gene totals
- Any number of samples – bam/sam
- Phenodata in fastq/bam/sam names

Galaxy / GCC2012Training      Using 66.4 MB

| Contigname | FC044_8_K562_15LSF_sequence.txt_bwa.sam.bam | FC076_2_Index3_HUVEC_48hrLSF_rep2_trimmed.txt_bwa.sam.bam | FC076_1_Index5_HUVEC_48hrLSF_rep1_trimmed.txt_bwa.sam.bam | FC068_7_Index5_K562_L... |
|------------|---|---|---|--------------------------|
| A1BG       | 94  | 15  | 47  |                          |
| A1BG-AS1   | 14  | 3   | 2   |                          |
| A1CF       | 0   | 0   | 0   |                          |
| A2LD1      | 3   | 1   | 0   |                          |
| A2M        | 3   | 11  | 5   |                          |
| A2ML1      | 2   | 0   | 0   |                          |
| A4GALT     | 0   | 161   | 219   |                          |
| A4GNT      | 0   | 0   | 0   |                          |
| AA06       | 0   | 0   | 0   |                          |
| AAA1       | 0   | 0   | 0   |                          |
| AAAS       | 1061  | 66  | 88  |                          |
| AACS       | 222   | 85  | 124   |                          |
| AACSP1     | 47  | 0   | 0   |                          |
| AADAC      | 1   | 1   | 0   |                          |
| AADACL2    | 1   | 0   | 0   |                          |
| AADACL3    | 0   | 0   | 0   |                          |
| AADACL4    | 0   | 0   | 0   |                          |
| AADAT      | 0   | 3   | 2   |                          |
| AAGAB      | 2478  | 119   | 71  |                          |
| AAK1       | 869   | 238   | 154   |                          |
| AAMP       | 3171  | 303   | 346   |                          |
| AANAT      | 3   | 0   | 0   |                          |
| AARS       | 6323  | 255   | 402   |                          |
| AARS2      | 1318  | 57  | 79  |                          |
| AARSD1     | 850   | 35  | 35  |                          |
| AASDH      | 361   | 18  | 3   |                          |
| AASDHPTT   | 3606  | 121   | 54  |                          |
| AASS       | 1   | 8   | 7   |                          |
| AATF       | 4669  | 118   | 109   |                          |
| AATK       | 199   | 18  | 38  |                          |
| AATK-AS1   | 3   | 0   | 0   |                          |
| ABAT       | 39  | 3   | 0   |                          |
| ABC41      | 4   | 95  | 50  |                          |
| ABC410     | 2   | 0   | 0   |                          |
| ABC411P    | 123   | 5   | 6   |                          |
| ABC412     | 1   | 0   | 0   |                          |
| ABC413     | 26  | 0   | 1   |                          |
| ABC417P    | 14  | 0   | 0   |                          |
| ABC42      | 194   | 426   | 594   |                          |
| ABC43      | 420   | 254   | 450   |                          |
| ABC44      | 0   | 11  | 9   |                          |
| ABC45      | 27  | 12  | 11  |                          |

History

- lightning (6.5 MB)
- 3: edgeR.html
- 2: edgeR.xls
- 1: TomK Andersson count matrix\_bams 2mx.xls

# EdgeR - a Tool Factory Tool

- Per sample, per contig matrix input
- Compare edgeR/DESeq/VOOM !
- Html outputs to present images
- Venn, MDS, heatmap, QQ,...
- Tabular “topTable” stats outputs
- Ready for GSEA

Galaxy / @BakerIDI

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 50%

**edgeR paired (version 0.18)**

Select an input matrix - rows are contigs, columns are counts for each sample:

1: TomK\_Andersson\_co.bams2mx.xls

Use the HTSeq based count matrix preparation tool to create these matrices from BAM/SAM files and a GTF file of genomic features

Title for job outputs:

edgeR

Supply a meaningful name here to remind you what the outputs contain

Treatment Name:

Treatment

Select columns containing treatment:

Select All  Unselect All

c2: FC044\_8\_K562\_15LSF\_sequence.txt\_bwa.sam.bam  
 c3: FC076\_2\_Index3\_HUVEC\_48hr\_SF\_rep2\_trimmed.txt\_bwa.sam.bam  
 c4: FC076\_1\_Index5\_HUVEC\_48hr\_SF\_rep1\_trimmed.txt\_bwa.sam.bam  
 c5: FC068\_7\_Index5\_K562\_LSF48hr\_rep1.txt\_bwa.sam.bam  
 c6: FC076\_2\_Index4\_HUVEC\_untr\_rep3\_trimmed.txt\_bwa.sam.bam  
 c7: FC076\_1\_Index6\_HUVEC\_untr\_rep2\_trimmed.txt\_bwa.sam.bam  
 c8: FC068\_8\_Index3\_K562\_LSF48hr\_rep2.txt\_bwa.sam.bam  
 c9: FC044\_5\_HUVEC\_UNTR\_sequence.txt\_bwa.sam.bam  
 c10: FC068\_7\_Index5\_K562\_untr\_rep2.txt\_bwa.sam.bam  
 c11: FC068\_7\_Index5\_K562\_untr\_rep1.txt\_bwa.sam.bam  
 c12: FC042\_6\_K562\_UNTR\_sequence.txt\_bwa.sam.bam  
 c13: FC076\_2\_Index7\_HUVEC\_48hr\_SF\_rep3\_trimmed.txt\_bwa.sam.bam  
 c14: FC076\_1\_Index2\_HUVEC\_untr\_rep1\_trimmed.txt\_bwa.sam.bam  
 c15: FC044\_6\_HUVEC\_15LSF\_sequence.txt\_bwa.sam.bam

Control Name:

Control

Select columns containing control:

Select All  Unselect All

c2: FC044\_8\_K562\_15LSF\_sequence.txt\_bwa.sam.bam  
 c3: FC076\_2\_Index3\_HUVEC\_48hr\_SF\_rep2\_trimmed.txt\_bwa.sam.bam  
 c4: FC076\_1\_Index5\_HUVEC\_48hr\_SF\_rep1\_trimmed.txt\_bwa.sam.bam  
 c5: FC068\_7\_Index5\_K562\_LSF48hr\_rep1.txt\_bwa.sam.bam  
 c6: FC076\_2\_Index4\_HUVEC\_untr\_rep3\_trimmed.txt\_bwa.sam.bam  
 c7: FC076\_1\_Index6\_HUVEC\_untr\_rep2\_trimmed.txt\_bwa.sam.bam  
 c8: FC068\_8\_Index3\_K562\_LSF48hr\_rep2.txt\_bwa.sam.bam  
 c9: FC044\_5\_HUVEC\_UNTR\_sequence.txt\_bwa.sam.bam  
 c10: FC068\_7\_Index5\_K562\_untr\_rep2.txt\_bwa.sam.bam  
 c11: FC068\_7\_Index2\_K562\_untr\_rep1.txt\_bwa.sam.bam  
 c12: FC042\_6\_K562\_UNTR\_sequence.txt\_bwa.sam.bam  
 c13: FC076\_2\_Index7\_HUVEC\_48hr\_SF\_rep3\_trimmed.txt\_bwa.sam.bam  
 c14: FC076\_1\_Index2\_HUVEC\_untr\_rep1\_trimmed.txt\_bwa.sam.bam  
 c15: FC044\_6\_HUVEC\_15LSF\_sequence.txt\_bwa.sam.bam

IF SUBJECTS NOT ALL INDEPENDENT! Enter integers to indicate sample pairing for every column in input:

History

549.9 MB

17: edgeR.html

54.0 KB format: html, database: hg19

16: edgeR.xls

15: K562\_DESeq\_msigdb\_gsea.html

14: TomK\_Andersson\_count\_matrix\_bams2mx.xls K562\_ALL\_DSeq.html

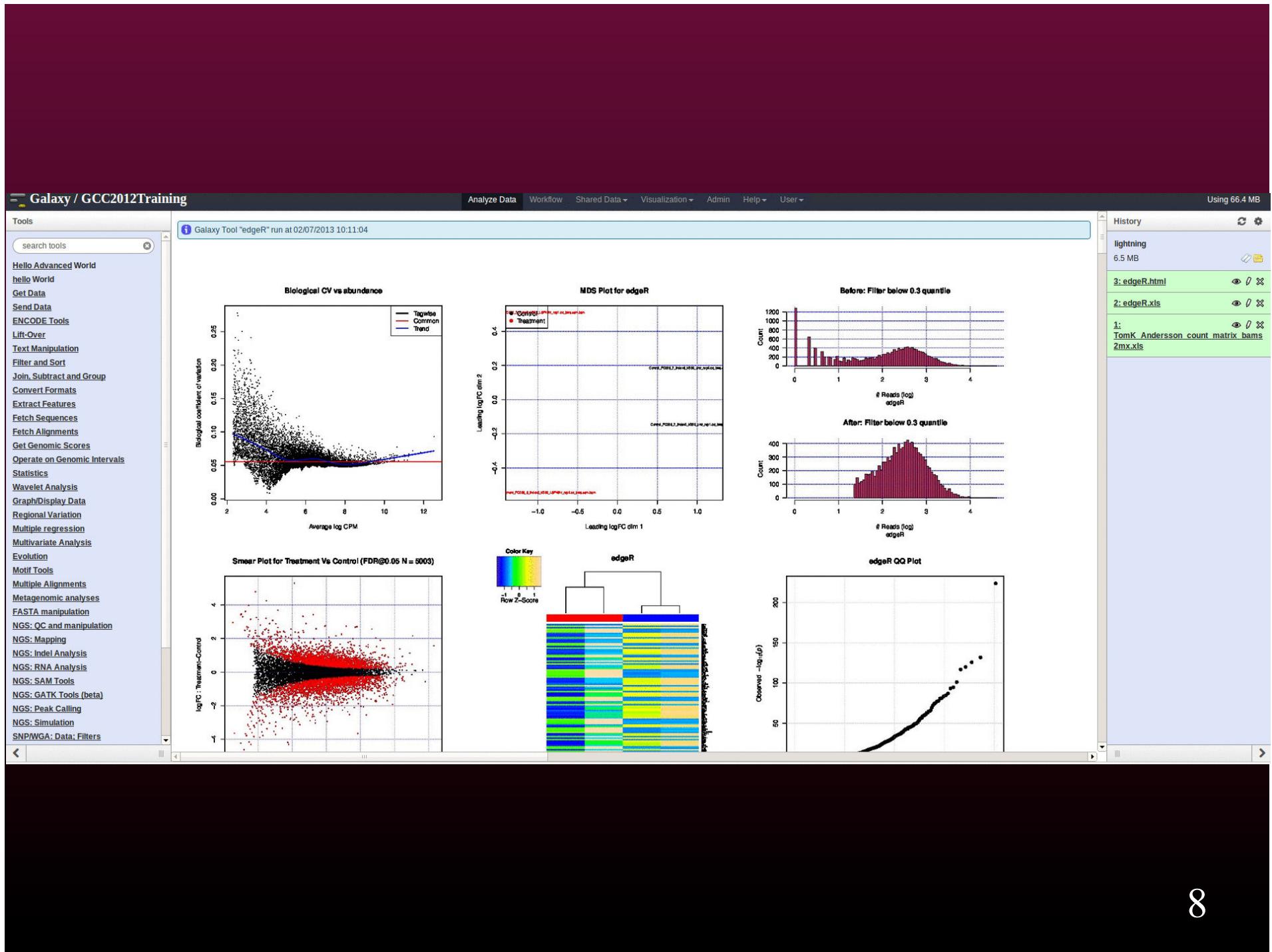
13: TomK\_Andersson\_count\_matrix\_bams2mx.xls K562\_ALL\_DSeq.xls

12: HUVEC\_DESeq\_msigdb\_gsea.html

11: TomK\_Andersson\_count\_matrix\_bams2mx.xls HUVEC\_A\_LL\_DSeq.html

10: TomK\_Andersson\_count\_matrix\_bams2mx.xls HUVEC\_A\_LL\_DSeq.xls

9: K562\_ALL\_priorit1\_0\_encodeTF\_GSEA.html



| Galaxy / GCC2012Training     |         |                    |                  |                      |                       |                       |                      |         |  |           |             | Using 66.4 MB |
|------------------------------|---------|--------------------|------------------|----------------------|-----------------------|-----------------------|----------------------|---------|--|-----------|-------------|---------------|
| Tools                        | Name    | logFC              | logCPM           | LR                   | PValue                | adj.p.value           | Dispersion           | tobread | Treatment_FC068_7_Index5_K562_LSF48hr_rep1.txt_bwa.sam | History   |             |               |
| Hello Advanced World         | DDIT4   | -3.63953568196737  | 8.77891934518699 | 1023.43492319872     | 1.44682751988544e-224 | 1.53768828813424e-220 | 0.00251051413122621  | 2640    | 44.30563361  | lightning | 6.5 MB      |               |
| hello World                  | NQO1    | 2.5886952440081    | 9.18551416435936 | 600.182495193243     | 1.52793502371099e-132 | 8.11944671600018e-129 | 0.00315986457704603  | 3797    | 499.3154830  |           | 64.45628693 |               |
| Get Data                     | BCAT1   | -2.53055956691928  | 8.8483215238664  | 571.788603448303     | 2.2921784055525e-126  | 8.12042415540706e-123 | 0.00290252295625431  | 2812    |  |           |             |               |
| Send Data                    | SQSTM1  | 2.69558961541079   | 8.25641063657078 | 545.592959315402     | 1.14474167853167e-120 | 3.04157863985864e-117 | 0.00261312827990848  | 1989    |  |           | 266.6547761 |               |
| ENCODE Tools                 | C5orf45 | 2.71974801927341   | 8.36627519576198 | 531.486827966047     | 1.34116493702163e-117 | 2.85078019013318e-114 | 0.00307925348593036  | 2213    |  |           | 261.7984805 |               |
| Lift-Over                    | AKR1C1  | 2.7306405200461    | 7.7482309850483  | 459.432185544493     | 6.38936446411588e-102 | 1.13176942541039e-98  | 0.00263276815655338  | 1420    |  |           | 179.2414554 |               |
| Text Manipulation            | MAP1B   | -2.5014777737748   | 7.85388272127184 | 430.03777425605      | 1.5946499129336e-95   | 2.42113419174998e-92  | 0.00263525014891064  | 1409    |  |           | 66.67172032 |               |
| Filter and Sort              | PRKDC   | -1.90069965486032  | 10.1717488535625 | 422.747536730584     | 6.15767821397982e-94  | 8.18047550727219e-91  | 0.00314186435362538  | 7089    |  |           | 319.2433326 |               |
| Join, Subtract and Group     | MTHFD2  | -2.049834473005223 | 8.51837918451033 | 394.512872394413     | 8.618541258080098e-88 | 1.01775396100983e-84  | 0.00257783474973037  | 2242    |  |           | 144.3031754 |               |
| Convert Formats              | FTL     | 2.1846200002484    | 11.4415246187052 | 382.73620708645      | 3.15683232060268e-82  | 3.35508139034035e-82  | 0.00517735762411276  | 18746   |  |           | 3408.058107 |               |
| Fetch Sequences              | PSAT1   | -2.3527944029264   | 7.6961929192371  | 377.242455903649     | 4.95824071657864e-84  | 4.7905620352077e-81   | 0.00245369417461906  | 1258    |  |           | 70.32496527 |               |
| Fetch Alignments             | MDN1    | -1.9968233387689   | 8.36856907457732 | 367.648594567323     | 6.08376327978809e-82  | 5.38818634479898e-79  | 0.00248557310563671  | 2017    |  |           | 84.36278181 |               |
| Get Genomic Scores           | CHML    | -2.10012459616817  | 8.00405266756395 | 365.336917147423     | 1.93869272959798e-81  | 1.58495587127485e-78  | 0.00229353048956088  | 1563    |  |           | 47.67999307 |               |
| Operate on Genomic Intervals | OSGIN1  | 5.32278452977903   | 5.84970479277874 | 350.605028317132     | 3.12886221865603e-78  | 3.27525340427688e-75  | 0.00220882613013133  | 379     |  |           | 87.97874791 |               |
| Statistics                   | FTH1    | 2.06581774972737   | 9.25197448574919 | 348.8708692975645    | 7.4655761204148e-78   | 5.289605953385123e-75 | 0.00406759880906692  | 4098    |  |           | 431.7688262 |               |
| Wavelet Analysis             | FADS1   | -1.68108205516268  | 9.80636322503291 | 339.987621299716e-76 | 6.42046472377916e-76  | 4.26479369277031e-73  | 0.0029224008178788   | 5486    |  |           | 355.6429047 |               |
| Graph/Display Data           | G6PD    | 1.91777604448109   | 9.4667067629884  | 335.357167938631     | 6.54661811382036e-72  | 4.09279160668722e-72  | 0.0037799926004857   | 4728    |  |           | 691.896196  |               |
| Regional Variation           | SLC38A1 | -1.70379393677178  | 9.09874174946178 | 327.166095513655     | 3.9812870625842e-73   | 2.35072882784151e-70  | 0.0026412974674811   | 3355    |  |           | 178.4363874 |               |
| Multiple regression          | CCND2   | -2.86874357412734  | 6.7087886326904  | 314.7744129895306    | 1.991736365697117e-70 | 1.11412969190998e-67  | 0.00261674872575284  | 630     |  |           | 17.8939826  |               |
| Multivariate Analysis        | SLC38A2 | -1.713392781495    | 9.16933819765736 | 298.22974902504      | 8.00669302956963e-74  | 4.2547566759133e-64   | 0.00326082230369119  | 3521    |  |           | 138.6251650 |               |
| Evolution                    | WARS    | -1.65939231866156  | 9.12481867288411 | 293.275419540371     | 9.6135903282074e-66   | 4.86539228610184e-63  | 0.00292679531556379  | 3439    |  |           | 130.6784995 |               |
| Motif Tools                  | CTSL1   | 2.02092359702583   | 7.90158593301493 | 292.005160631735     | 1.81825614090872e-65  | 8.78383012071722e-63  | 0.00291171617806341  | 1594    |  |           | 170.8533085 |               |
| Multiple Alignments          | SRGN    | 2.2817791360807    | 7.31620286250315 | 285.083829802284     | 5.85836503602385e-64  | 2.7070460968963e-61   | 0.00277237851512884  | 1036    |  |           | 130.6784995 |               |
| Metagenomic analyses         | XIST    | -1.58394867331339  | 8.82795735940876 | 283.025876585504     | 1.645195926193156e-63 | 7.28547595982532e-61  | 0.00243690828631519  | 2798    |  |           | 107.7214658 |               |
| FASTA manipulation           | MACF1   | -1.5305633875621   | 9.54442443667167 | 274.58417886778      | 1.13720290381916e-61  | 4.83447698471602e-59  | 0.0029101568337277   | 4576    |  |           | 326.6551829 |               |
| NGS: QC and manipulation     | GSR     | 1.53880307728024   | 9.14329544305892 | 271.677843087334     | 4.8891659965261e-61   | 1.99854062350305e-58  | 0.00266695902276778  | 3693    |  |           | 828.373292  |               |
| NGS: Mapping                 | HOXB9   | -2.05818471108305  | 7.4377748534257  | 269.60353892116      | 1.38457837178696e-61  | 5.45011071679985e-58  | 0.00256734894635065  | 1060    |  |           | 29.13777354 |               |
| NGS: Indel Analysis          | AKR1C3  | 2.776922773792708  | 6.5563738920406  | 266.335664235522     | 7.137648298317e-60    | 2.70924736123265e-57  | 0.00262639826092613  | 609     |  |           | 111.0675472 |               |
| NGS: RNA Analysis            | PDE4DIP | -1.8279799548123   | 8.1647647375783  | 265.065988241511     | 1.34986947330644e-59  | 4.94703888355202e-57  | 0.00307353722584722  | 1775    |  |           | 112.3372821 |               |
| NGS: SAM Tools               | LEPR    | -2.24451797391058  | 7.36005262647778 | 254.956698052        | 2.15712984697824e-57  | 7.64199200456158e-55  | 0.00379279682192281  | 1005    |  |           | 54.7986742  |               |
| NGS: GATK Tools (beta)       | SLC7A5  | -1.54578421566973  | 8.58525262250276 | 252.174703922101     | 8.7165105258566e-57   | 2.98835722157432e-54  | 0.00245426276367508  | 2371    |  |           | 152.8444349 |               |
| NGS: Peak Calling            | AKR1C2  | 2.72603495724937   | 6.50576162979877 | 251.684498764384     | 1.1482986003099e-56   | 3.7026867262791e-54   | 0.00271453133724073  | 590     |  |           | 130.4769566 |               |
| NGS: Simulation              | SLC02B1 | 3.50403058847961   | 6.1173125596877  | 248.275561031172     | 6.17146160972676e-56  | 1.98758466630836e-53  | 0.004618775282239283 | 464     |  |           | 91.70666096 |               |
| SNP/WGA: Data; Filters       | PKM2    | 1.3953851491037    | 10.4836747189962 | 240.351710897101     | 3.29622675826709e-54  | 1.03036170549596e-51  | 0.00322402934302576  | 9266    |  |           | 1026.444295 |               |
| SNP/WGA: QC; LD; Plots       | MYB     | -1.57190941417847  | 8.10553904453639 | 221.631464525543     | 3.98589097992694e-50  | 1.05905123336659e-47  | 0.00246262570397775  | 1691    |  |           | 68.87110111 |               |
| SNP/WGA: Statistical Models  | PLCD1   | -2.57224177301186  | 6.43972406248483 | 220.450568683698     | 7.21279384322187e-50  | 1.869069010395e-47    | 0.0031045018278032   | 524     |  |           | 19.38514784 |               |

# Gene Set Enrichment Analysis

- Uses all DEG evidence and direction
- Tab input: p value/direction columns
- Unexpected up/down enrichment
- A priori gene sets – MSigDB/user
- Permutation for FWER control
- Biologist friendly outputs

Galaxy / @BakerIDI

Analyze Data Workflow Shared Data Visualization Admin Help User Using 50%

**GSEA Report for Dataset K562\_ALL\_priordf10\_edgeR.xls**

**Enrichment in phenotype: na**

- 2521 / 6007 gene sets are upregulated in phenotype na\_pos
- 427 gene sets are significant at FDR < 25%
- 244 gene sets are significantly enriched at nominal pvalue < 1%
- 522 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: na**

- 3486 / 6007 gene sets are upregulated in phenotype na\_neg
- 480 gene sets are significantly enriched at FDR < 25%
- 311 gene sets are significantly enriched at nominal pvalue < 1%
- 699 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

**Dataset details**

- The dataset has 11871 features (genes)
- No probe set => gene symbol collapsing was requested, so all 11871 features were used

**Gene set details**

- Gene set size filters (min=15, max=1000) resulted in filtering out 2506 / 8513 gene sets
- The remaining 6007 gene sets were used in the analysis
- List of gene sets used and their sizes (restricted to features in the specified dataset)

**Gene markers for the na\_pos versus na\_neg comparison**

- The dataset has 11871 features (genes)
- Detailed rank ordered gene

History

- 17: edgeR.html
- 16: edgeR.xls
- 15: K562 DESeq msigdb gsea.html
- 14: TomK Andersson count matrix ba ms2mx.xls K562 ALL DESeq.html
- 13: TomK Andersson count matrix ba ms2mx.xls K562 ALL DESeq.xls
- 12: HUVEC DESeq msigdb gsea.html
- 11: TomK Andersson count matrix ba ms2mx.xls HUVEC ALL DESeq.htm !
- 10: TomK Andersson count matrix ba ms2mx.xls HUVEC ALL DESeq.xls
- 9: K562 ALL priordf10 encodeTF GSE A.html
- 8: HUVEC ALL priordf10 encodeTF G SEA.html
- 7: HUVEC ALL priordf10 msigdb GSE A.html
- 6: K562 ALL priordf10 msigdb GSEA.html
- 5: K562 ALL priordf10 edgeR.html
- 4: K562 ALL priordf10 edgeR.xls
- 3:

Galaxy / @BakerIDI

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 50%

Tools

search tools

**Gene Expression**

BakerIDI

SR Test/Repair/BWA Tools

Local unreliable SR Quality Tools

Get Data

Send Data

Repeats and Complexity

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Wavelet Analysis

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NGS: QC and manipulation

Phenotype Association

NGS: Assembly

NGS: Mapping

NGS: Indel Analysis

NGS: RNA Analysis

NGS: SAM Tools

NGS: Picard (beta)

Table: Gene sets enriched in phenotype na [plain text format]

GS follow link to MSigDB

|    |   | GS DETAILS                  | SIZE | ES   | NES  | NOM p-val | FDR q-val | FWER p-val | R |
|----|---|-----------------------------|------|------|------|-----------|-----------|------------|---|
| 1  | <a href="#">NFE2L2_V2</a>   | <a href="#">Details ...</a> | 224  | 0.63 | 2.33 | 0.000     | 0.000     | 0.000      | 4 |
| 2  | <a href="#">CONCANNON_APOPTOSIS_BY_EPOXOMICIN_UP</a>                              | <a href="#">Details ...</a> | 185  | 0.61 | 2.22 | 0.000     | 0.001     | 0.002      | 7 |
| 3  | <a href="#">PODAR_RESPONSE_TO_ADAPHOSTIN_UP</a>                                   | <a href="#">Details ...</a> | 127  | 0.64 | 2.21 | 0.000     | 0.001     | 0.002      | 7 |
| 4  | <a href="#">ENGELMANN_CANCER_PROGENITORS_DN</a>                                   | <a href="#">Details ...</a> | 42   | 0.74 | 2.17 | 0.000     | 0.002     | 0.006      | 7 |
| 5  | <a href="#">KEGG_METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450</a>                 | <a href="#">Details ...</a> | 21   | 0.85 | 2.16 | 0.000     | 0.002     | 0.010      | 4 |
| 6  | <a href="#">CHIBA_RESPONSE_TO_TSA</a>   | <a href="#">Details ...</a> | 41   | 0.73 | 2.12 | 0.000     | 0.008     | 0.045      | 1 |
| 7  | <a href="#">REGULATION_OF_GROWTH</a>  | <a href="#">Details ...</a> | 42   | 0.73 | 2.08 | 0.000     | 0.013     | 0.083      | 1 |
| 8  | <a href="#">COLIN_PILOCYTIC_ASTROCYTOMA_VS_GLIOMA_DN</a>                          | <a href="#">Details ...</a> | 18   | 0.84 | 2.07 | 0.000     | 0.013     | 0.100      | 3 |
| 9  | <a href="#">OXIDOREDUCTASE_ACTIVITY_GO_0016616</a>                                | <a href="#">Details ...</a> | 38   | 0.74 | 2.07 | 0.000     | 0.014     | 0.121      | 4 |
| 10 | <a href="#">GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_BLUE_UP</a>             | <a href="#">Details ...</a> | 114  | 0.60 | 2.05 | 0.000     | 0.017     | 0.159      | 1 |
| 11 | <a href="#">OXIDOREDUCTASE_ACTIVITY_ACTING_ON_CH_OH_GROUP_OF_DONORS</a>           | <a href="#">Details ...</a> | 42   | 0.72 | 2.04 | 0.000     | 0.019     | 0.193      | 4 |
| 12 | <a href="#">REGULATION_OF_CELL_GROWTH</a>   | <a href="#">Details ...</a> | 35   | 0.75 | 2.04 | 0.000     | 0.020     | 0.213      | 1 |
| 13 | <a href="#">TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_SUSTAINED_IN_ERYTHROCYTE_UP</a> | <a href="#">Details ...</a> | 27   | 0.77 | 2.03 | 0.002     | 0.020     | 0.233      | 7 |
| 14 | <a href="#">HOUSTIS_ROS</a>   | <a href="#">Details ...</a> | 25   | 0.79 | 2.03 | 0.000     | 0.020     | 0.247      | 4 |

History 549.9 MB

17: edgeR.html

16: edgeR.xls

15: K562 DESeq\_msig db gsea.html

14: TomK Andersson count matrix bams 2mx.xls K562 ALL DESeq.html

13: TomK Andersson count matrix bams 2mx.xls K562 ALL DESeq.xls

12: HUVEC DESeq\_ms igdb gsea.html

11: TomK Andersson count matrix bams 2mx.xls HUVEC A LL DESeq.html

10: TomK Andersson count matrix bams 2mx.xls HUVEC A LL DESeq.xls

9: K562 ALL priordf1 0 encodeTF GSEA.html

8: HUVEC ALL priordf10 encodeTF GSE A.html

7:

Galaxy / @BakerIDI

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 50%

History 549.9 MB

17: edgeR.html  
16: edgeR.xls  
15: K562 DESeq\_msig db gsea.html  
14: TomK Andersson count matrix bams 2mx.xls K562 ALL DESeq.html  
13: TomK Andersson count matrix bams 2mx.xls K562 ALL DESeq.xls  
12: HUVEC DESeq\_ms igdb\_gsea.html  
11: TomK Andersson count matrix bams 2mx.xls HUVEC A LL DESeq.html  
10: TomK Andersson count matrix bams 2mx.xls HUVEC A LL DESeq.xls  
9: K562 ALL priordf1\_0 encodeTF GSEA.html  
8: HUVEC ALL priorf10 encodeTF GSE A.html  
7:

**Table: GSEA Results Summary**

|                                   |                              |
|-----------------------------------|------------------------------|
| Dataset                           | K562_ALL_priordf10_edgeR.xls |
| Phenotype                         | NoPhenotypeAvailable         |
| Upregulated in class              | na_pos                       |
| GeneSet                           | NFE2L2.V2                    |
| Enrichment Score (ES)             | 0.631594                     |
| Normalized Enrichment Score (NES) | 2.3260202                    |
| Nominal p-value                   | 0.0                          |
| FDR q-value                       | 0.0                          |
| FWER p-Value                      | 0.0                          |

**Enrichment plot: NFE2L2.V2**

Enrichment score (ES)

Ranked list metric (preRanked)

'na\_pos' (positively correlated)

Zero cross at 5643

'na\_neg' (negatively correlated)

Rank in Ordered Dataset

— Enrichment profile — Hits — Ranking metric scores

galaxy.iaas1.bhri.internal/galaxy/

News shop linux bioinf Other python Teach Google Transmission W...

Analyze Data Workflow Shared Data Visualization Admin Help User

login register BROAD INSTITUTE

Other Bookmarks Using 50%

Galaxy / @BakerIDI

Tools search tools Gene Expression BakerIDI SR Test/Repair/BWA Tools Local unreliable SR Quality Tools Get Data Send Data Repeats and Complexity ENCODE Tools Lift-Over Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Fetch Alignments Get Genomic Scores Operate on Genomic Intervals Statistics Wavelet Analysis Graph/Display Data Regional Variation Multiple regression Multivariate Analysis Evolution Motif Tools Multiple Alignments Metagenomic analyses FASTA manipulation NGS: QC and manipulation Phenotype Association NGS: Assembly NGS: Mapping NGS: Indel Analysis NGS: RNA Analysis NGS: SAM Tools NGS: Picard (beta)

GSEA Gene Set Enrichment Analysis

Molecular Signatures Database Documentation Contact

MSigDB Home Downloads

Gene Set: NFE2L2.V2

**Standard name** NFE2L2.V2  
**Systematic name** M2870  
**Brief description** Genes up-regulated in MEF cells (embryonic fibroblasts) with knockout of NFE2L2 [GeneID=4780] gene.

**Full description or abstract** The Nr2 (nuclear factor E2 p45-related factor 2) transcription factor responds to diverse oxidative and electrophilic environmental stresses by circumventing repression by Keap1, translocating to the nucleus, and activating cytoprotective genes. Nr2 responses provide protection against chemical carcinogenesis, chronic inflammation, neurodegeneration, emphysema, asthma and sepsis in murine models. Nr2 regulates the expression of a plethora of genes that detoxify oxidants and electrophiles and repair or modulate damage induced by such stressors through multiple mechanisms. However, many direct targets of Nr2 remain undefined. Here we use embryonic fibroblasts (MEF) with either constitutive nuclear accumulation (Keap1<sup>-/-</sup>) or depletion (Nr2<sup>-/-</sup>) of Nr2 were utilized to perform chromatin-immunoprecipitation with parallel sequencing (ChIP-Seq) and global transcription profiling. This unique Nr2 ChIP-Seq dataset is highly enriched for Nr2-binding motifs. Integrating ChIP-Seq and microarray analyses, we identified 645 basal and 654 inducible direct targets of Nr2, with 244 genes at the intersection. Modulated pathways in stress response and cell proliferation distinguish the inducible and basal programs. Results were confirmed in an in vivo stress model of cigarette smoke-exposed mice. This study reveals global circuitry of the Nr2 stress response emphasizing Nr2 as a central node in cell survival response.

**Collection** C6: oncogenic signatures

**Source publication** Pubmed 20460467 Authors: Malhotra D,Portales-Casamar E,Singh A,Srivastava S,Arenillas D,Happel C,Shyr C,Wakabayashi N,Kensler TW,Wasserman WW,Biswal S

**Exact source** NERF2 vs control; up-regulated genes

**Related gene sets** (show gene sets from the same authors)

**External links**

**Organism** Mus musculus

**Contributed by** Pablo Tamayo (Broad Institute)

**Source platform** HUMAN\_GENE\_SYMBOL

**Dataset references**

**Download gene set** format: grp | text | gmt | gmx | xml

**Compute overlaps** ?

C1: positional gene sets  
C2: curated gene sets  
C3: gene sets based on genetic perturbations  
CP: Canonical pathways  
CR: Biocarta-BioCarta gene sets  
CP:KEGG: KEGG gene sets  
CP:REACTOME: Reactome gene sets  
C3: motif gene sets  
MIR: microRNA targets  
TET: transcription factor targets  
C4: computational gene sets  
CGN: cancer gene neighborhoods  
CM: cancer modules  
C5: GO gene sets  
BP: GO biological process  
CC: GO cellular component  
MF: GO molecular function  
C6: oncogenic signatures  
C7: immunologic signatures

History 549.9 MB 17: edgeR.html 16: edgeR.xls 15: K562 DESeq msig db gsea.html 14: TomK Andersson count\_matrix\_bams 2mx.xls K562 ALL DESeq.html 13: TomK Andersson count\_matrix\_bams 2mx.xls K562 ALL DESeq.xls 12: HUVEC DESeq ms igdb gsea.html 11: TomK Andersson count\_matrix\_bams 2mx.xls HUVEC A LL DESeq.html 10: TomK Andersson count\_matrix\_bams 2mx.xls HUVEC A LL DESeq.xls 9: K562 ALL prior1 0 encodeTF GSEA.html 8: HUVEC ALL prior1 f10 encodeTF GSE A.html 7:

# In test toolshed

- GSEA is in good shape
- EdgeR/htseq still in beta
- Feedback welcomed...



**VIVA LA EVOLUCIÓN**

**GALAXY**

**<http://usegalaxy.org>**