# Transcriptome Data Analysis

Federico Marini (marinif@uni-mainz.de)


IMBEI@


2022/07/20-21

@FedeBioinfo

# Questions

What are the steps to process RNA-Seq data?

- How to convert RNA-seq reads into counts?
- How to perform quality control (QC) of RNA-seq reads?

How to identify differentially expressed genes across multiple experimental conditions?

- How to properly analyze RNA count data using DESeq2?
- How to perform quality control (QC) and exploratory data analysis (EDA) of RNA-seq count data?

What are the biological functions impacted by the differential expression of genes?

- How can I perform a gene ontology enrichment analysis?

How can I create neat visualizations of the data?

- How can I visualize the results for my enrichment analysis?

How can I generate interactive reports to summarise my analyses?

# Questions

**What are the steps to process RNA-Seq data?**

- How to convert RNA-seq reads into counts?
- How to perform quality control (QC) of RNA-seq reads?

**How to identify differentially expressed genes across multiple experimental conditions?**

- How to properly analyze RNA count data using DESeq2?
- How to perform quality control (QC) and exploratory data analysis (EDA) of RNA-seq count data?

What are the biological functions impacted by the differential expression of genes?

- How can I perform a gene ontology enrichment analysis?

How can I create neat visualizations of the data?

- How can I visualize the results for my enrichment analysis?

**How can I generate interactive reports to summarise my analyses?**

# What you will learn

- the basics of RNA-seq data
- the basics of RNA-seq data analysis
- to get familiar with the concepts of gene expression, high-dimensional data, expression quantification, differential expression analysis
- the importance to pose the right question, in order to get the right answer :)

This material has been developed together with Charlotte Soneson in the scope of the GTIPI Summer School (https://imbeimainz.github.io/GTIPI2022)

# Setup for practical sessions

Got R/RStudio?

Latest versions highly recommended!

See https://imbeimainz.github.io/GTIPI2022/material.html for details!

# Decomposing the title

RNA

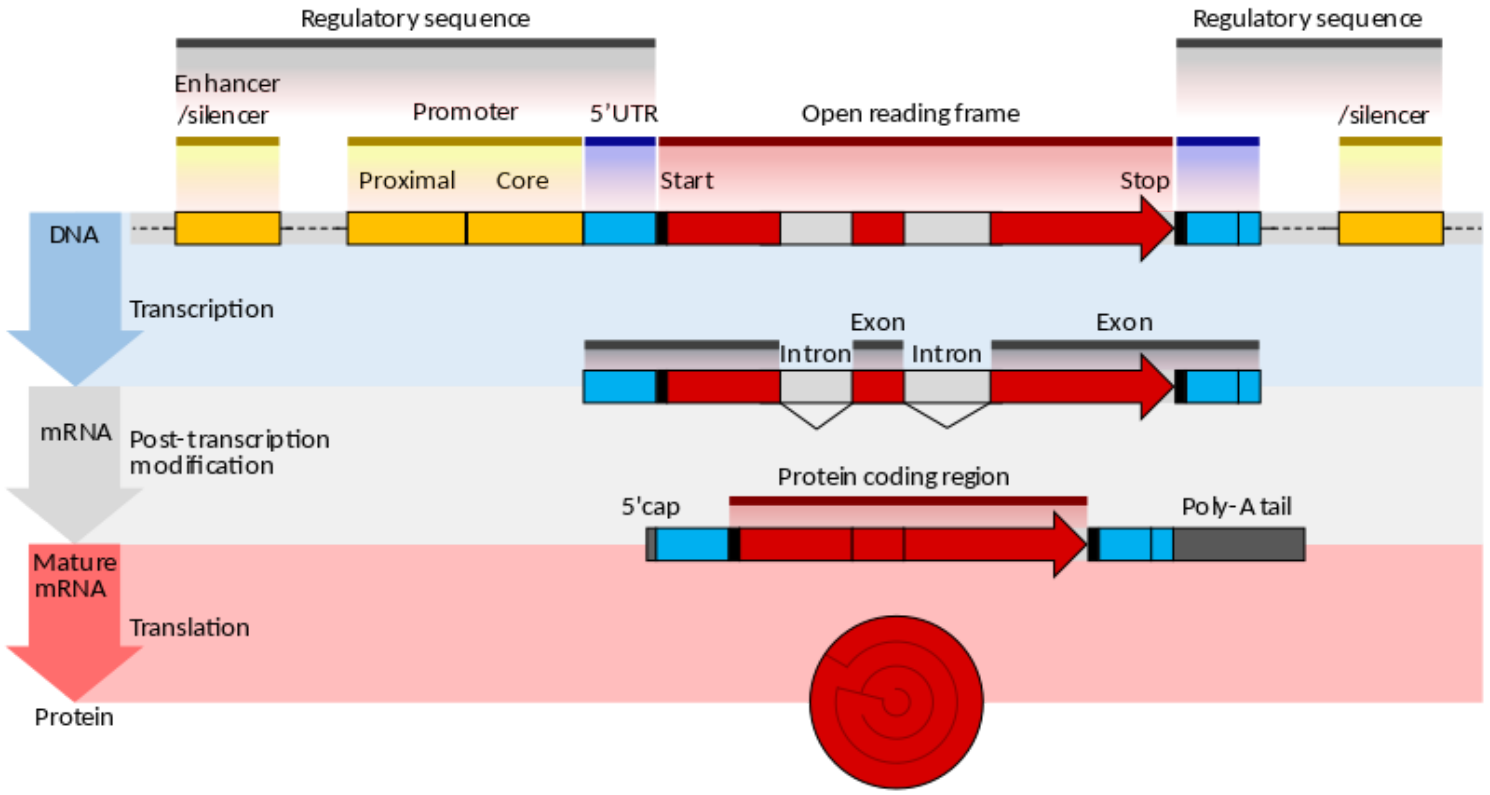
Sequencing

Bioinformatics

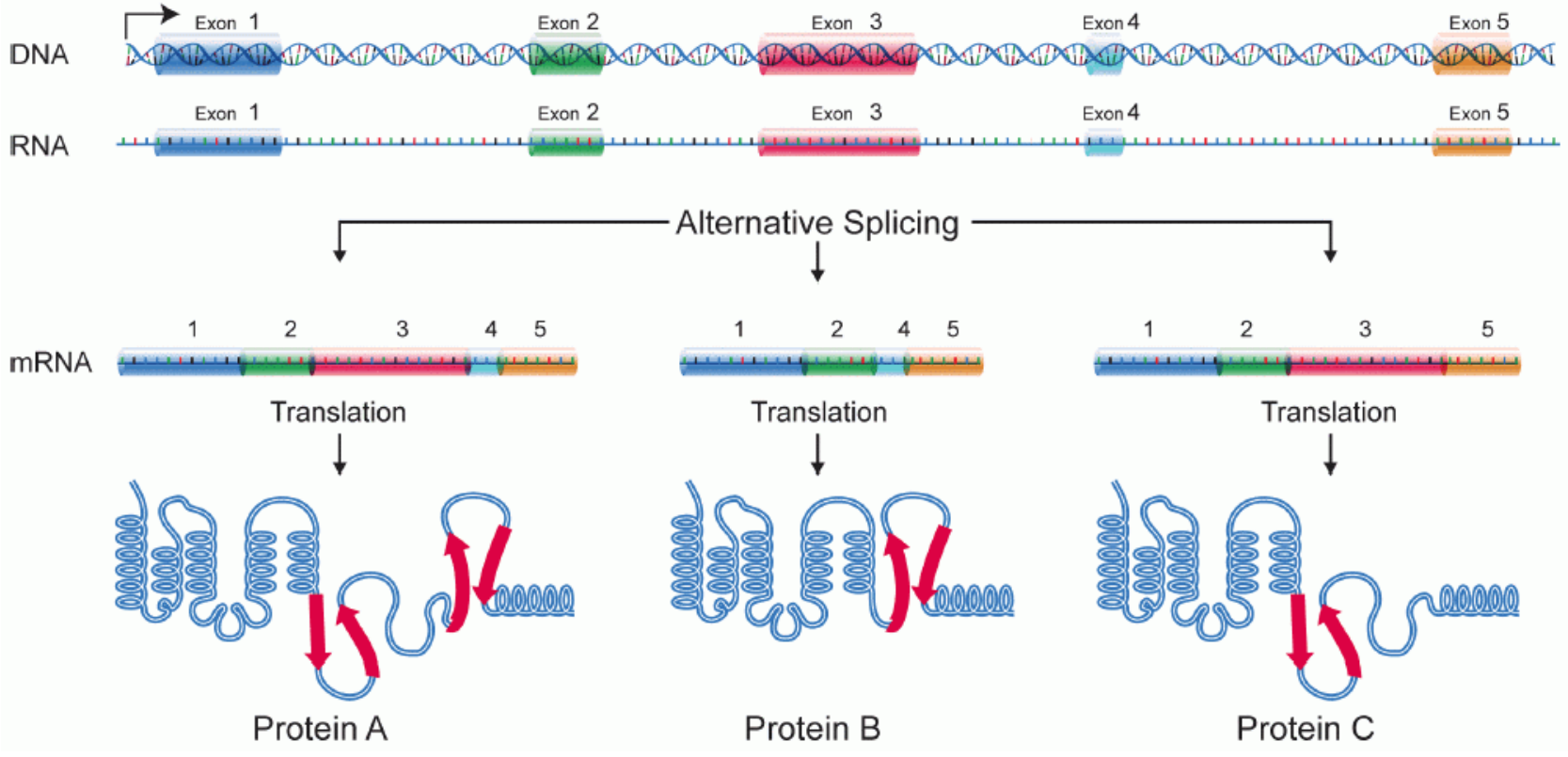
Transcriptome

Analysis

# (messenger) RNA

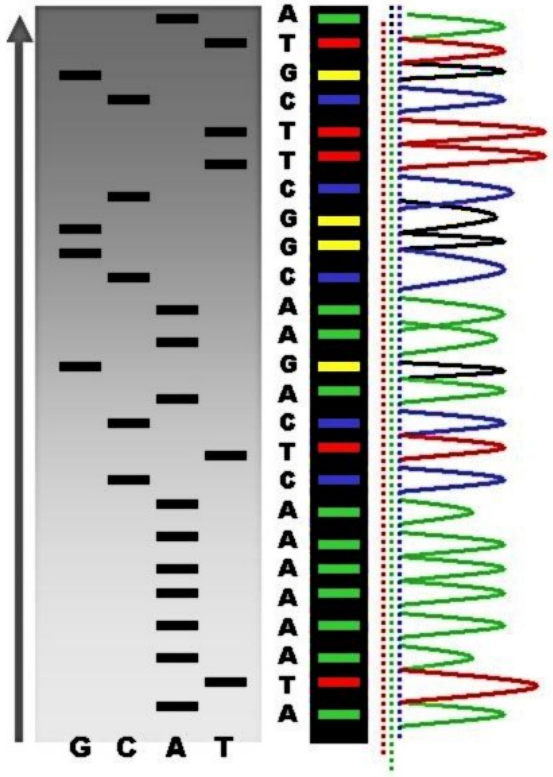


# (messenger) RNA




Exons, introns, transcripts, isoforms

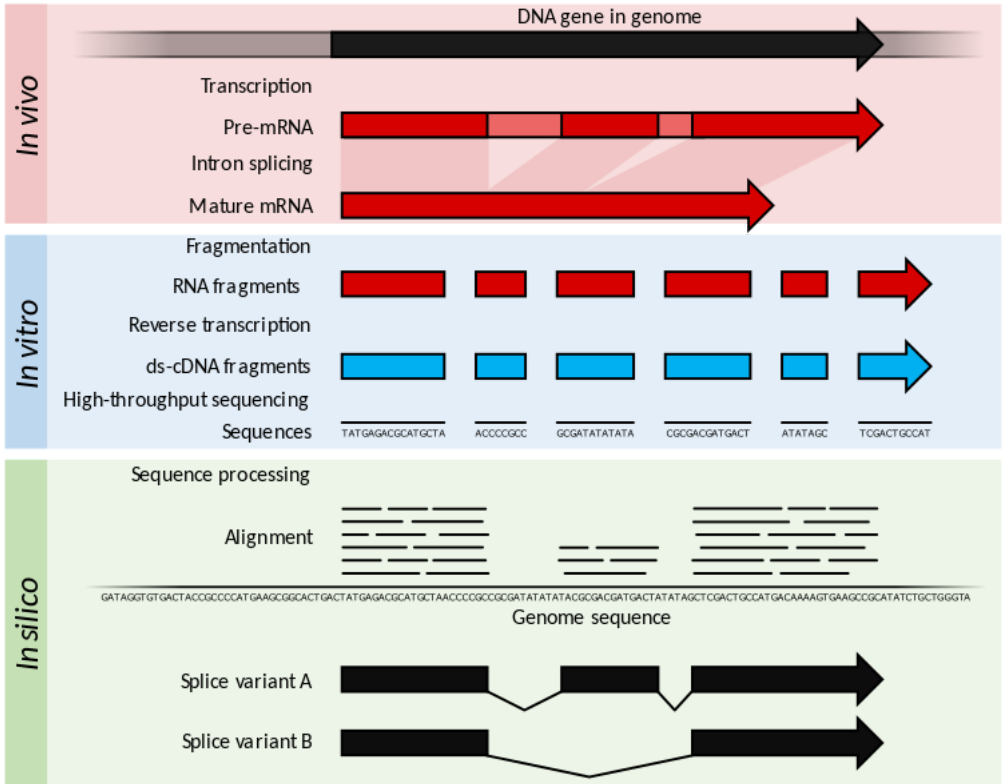
# Sequencing




4 base pairs, 21 aminoacids

Excellent review on next-generation sequencing: https://www.nature.com/articles/nrg.2016.49

# RNA-sequencing




- RNA quantification at single base resolution
- Cost efficient analysis of the whole transcriptome in a high-throughput manner

# Challenges in RNA-seq

- Different origin for the sample RNA and the reference genome
- Presence of incompletely processed RNAs or transcriptional background noise
- Sequencing biases (e.g. PCR library preparation)

Benefits

- sensitive
- specific
- high-throughput
- cost-efficient
- basepair resolution

You can see: transcripts, splicing, lncRNA, circRNA, gene fusions

# Bioinformatics

> "an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex."

A combination of

- biology
- computer science
- information engineering
- mathematics
- statistics

... to analyze and interpret the biological data

# Transcriptome

Gene expression is a fundamental level at which the results of various genetic and regulatory programs are observable.

RNA sequencing (RNA-seq) provides a quantitative and open system for profiling transcriptional outcomes on a large scale

Much has been learned about the characteristics of the RNA-seq data sets, as well as the performance of the myriad of methods developed

"RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis" ->
https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-072018-021255

# Analysis

There's data involved!

and these datasets have particular properties (how they are generated, ...)

How to make sense out of it?

There is a large diversity of applications to deal with

Among the most widely adopted workflows:

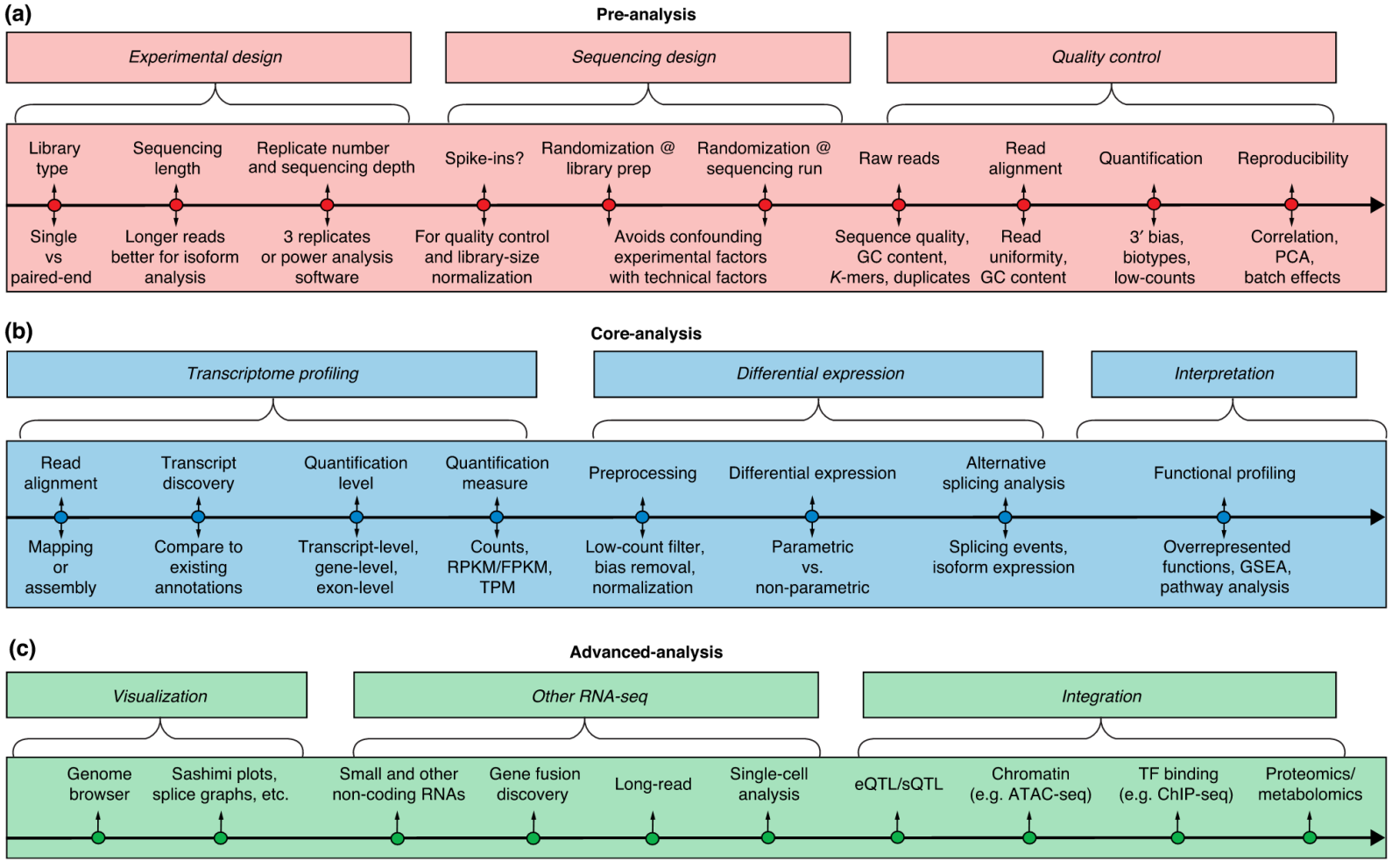
- Transcript discovery

  > *Which RNA molecules are in my sample?*
  >
  > Novel isoforms and alternative splicing, Non-coding RNAs, Single nucleotide variations, Fusion genes

- RNA quantification

  > *What is the concentration of RNAs?*
  >
  > Absolute gene expression (within sample), Differential expression (between biological samples)

# Analysis - a bird's eye view



# Differential analysis types for RNA-seq

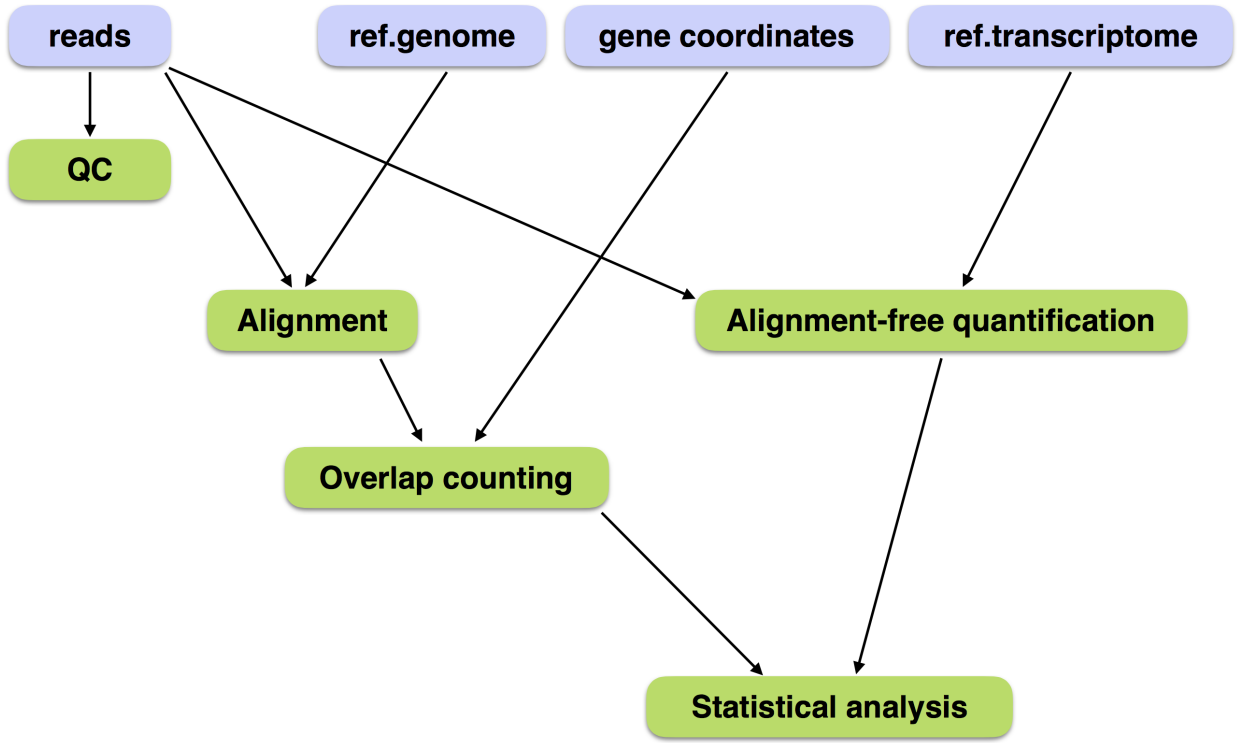
No single available standardized workflow
Multiple possible best practices for every dataset

To get the right answer, you have to pose the right question.

- Does the total output of a gene change between conditions? Differential Gene Expression (DGE)
- Does the expression of individual transcripts change? Differential Transcript Expression (DTE)
- Does any isoform of a given gene change? DTE+G
- Does the isoform composition for a given gene change? Differential Transcript Usage/Differential Exon Usage (DTU/DEU)

Each needs different computational approaches (quantifications + tests)

# Overview of the processing workflow




Ingredients + operations

# The raw data: sequencing reads

FASTQ files: sequence + base quality (phred score)

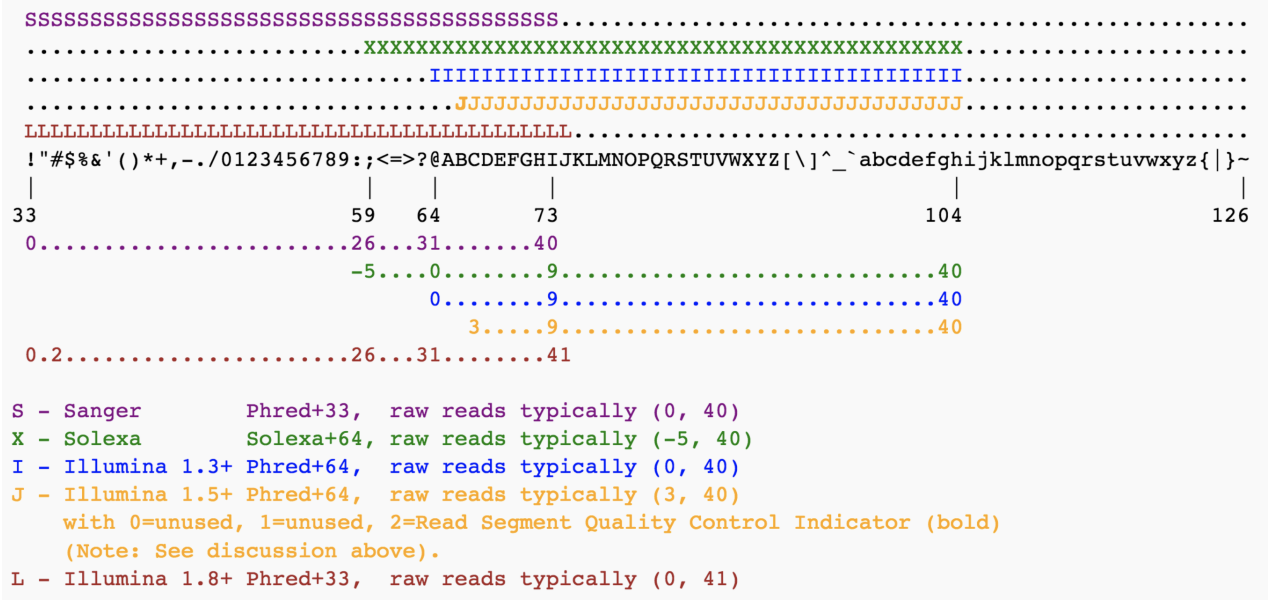
First lines of a FASTQ file

```
@SRR1055095.1 HWI-ST156:397:D09NJACXX:5:1101:1222:1915/1
NGCTGCTGGACTCCGAAGATGGGCGGTATATCATCCCACTGCTGACTCTN
+
#1=DDFFFHFHHHJJHIIGHIJJGIJAFDDHGGGIJJGJIJJJIJJIIH#
@SRR1055095.2 HWI-ST156:397:D09NJACXX:5:1101:1245:1920/1
NCTTTTCTTTGTTCTCATCATCTTCAGGAGGAGGAGGGTCATCCTTGTGN
+
#1=BDB:?FFDFFDF?EF<FFF>B>?C@CF<1??CFB:09;09BFE9DB#
```

Repeat this tens of millions of times, and you'll have *one* sample

# The raw data: sequencing reads

Different quality encodings exist



# The raw data

Demo: quality control report, from FastQC

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html

Your next best friend: MultiQC

Common sequence artefacts in NGS data:

- read errors
- base calling errors
- small insertions and deletions
- poor quality reads
- primer/adapter contamination

Solutions: Quality trimming & filtering (wide range of QC tools available)

# Reference files

Reference genome sequences (in `fasta` format), required for genome alignment.

Think of the alignment as the address of each read in a 3-billion houses street, where the elements along the street can also end up repeating themselves.

**What's out there?**

Ensembl: http://www.ensembl.org/info/data/ftp/index.html

Gencode (human & mouse): https://www.gencodegenes.org/

UCSC: http://hgdownload.cse.ucsc.edu/downloads.html

iGenome: http://support.illumina.com/sequencing/sequencing_software/igenome.html

# Reference files

Some critical points:

Be consistent!

Different chromosome identifiers!

Reference genomes and annotations are continuously refined, extended and improved

Keep track of version and be consistent!

Naming of genes can vary across versions, in some databases!

# Reference files

An example:

http://ftp.ensembl.org/pub/release-106/fasta/homo_sapiens/dna/

... and one level up...

http://ftp.ensembl.org/pub/release-106/fasta/homo_sapiens/

```
>1 dna:chromosome chromosome:GRCh38:1:1:248956422:1 REF
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
...
TTGGTGCCAGTTCCTCCAAGTCGATGGCACCTCCCTCCCTCTCAACCACTTGAGCAAACT
CCAAGACATCTTCTACCCCAACACCAGCAATTGTGCCAAGGGCCATTAGGCTCTCAGCAT
GACTATTTTTAGAGACCCCGTGTCTGTCACTGAAACCTTTTTTGTGGGAGACTATTCCTC
CCATCTGCAACAGCTGCCCCTGCTGACTGCCCTTCTCTCCTCCCTCTCATCCCAGAGAAA
CAGGTCAGCTGGGAGCTTCTGCCCCCACTGCCTAGGGACCAACAGGGGCAGGAGGCAGTC
```

# Reference files

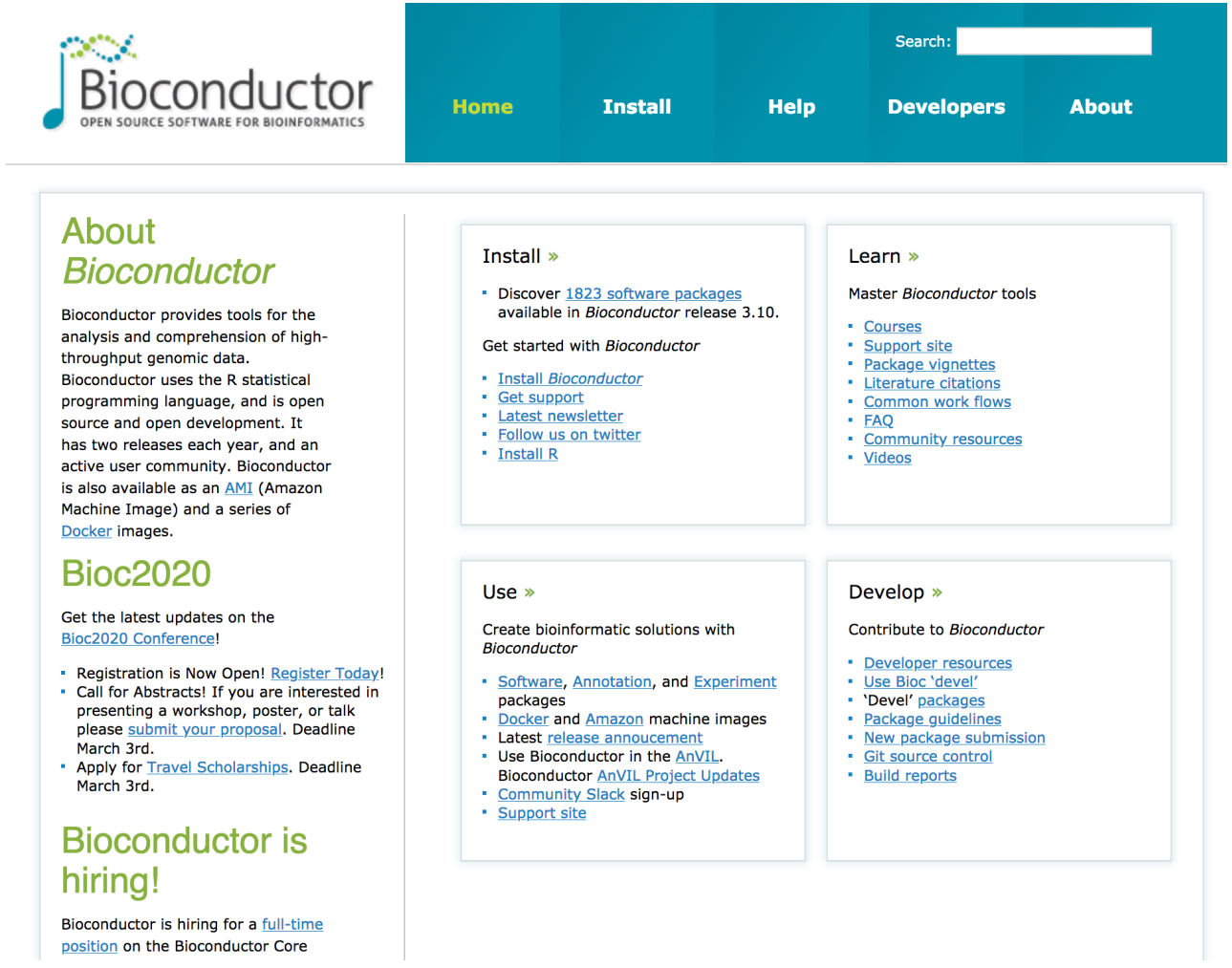
The GTF format

```
chr1    unknown exon    11874   12227   .       +       .       gene_id "DDX11L1"; gene_name "DDX11L1";
transcript_id "NR_046018"; tss_id "TSS16107";
chr1    unknown CDS     3427347 3427466 .       -       2       gene_id "MEGF6"; gene_name "MEGF6";
p_id "P34437"; transcript_id "NM_001409"; tss_id "TSS31177";
```

- One line per "feature" (exon, transcript, gene, CDS, 3'UTR, 5'UTR, ...)
- One feature = 9 columns of data, plus optional track definition lines
- Essential for releasing annotation information

```
seqname - name of chromosome/scaffold
source - data/program source
feature - feature type name
start - positions of the feature
end
score - floating point value
strand - forward or reverse
frame -  0|1|2
attribute - semicolon-separated list of tag-value pairs, providing
additional information
```

# Bioconductor



# Bioconductor - soon your best friend

- an open source project
- a repository of packages, focused on bioinformatics/computational biology
- a open development platform and community

Currently (May 2022)

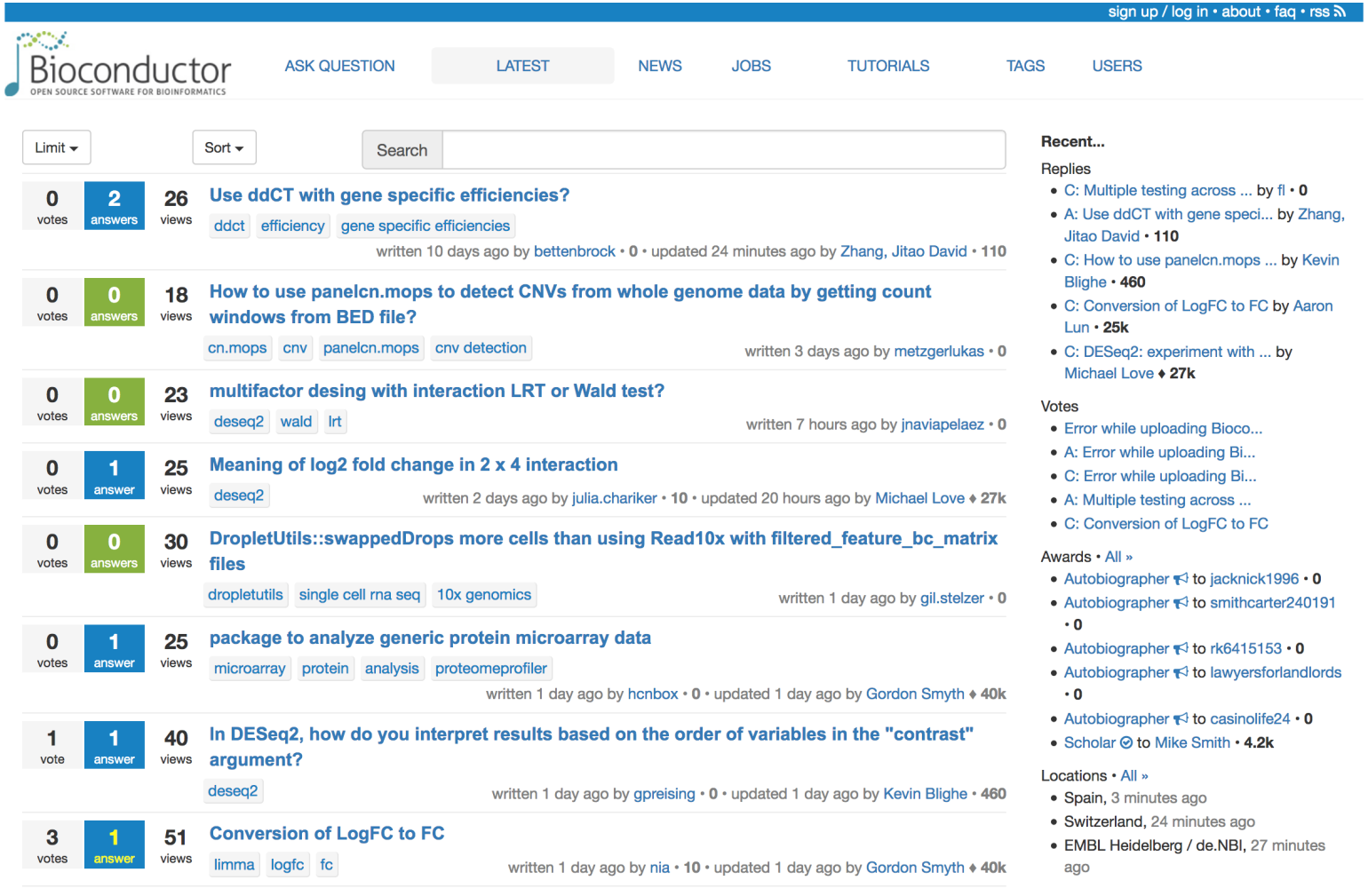
- 2140 software packages
- 909 annotation packages
- 410 experiment packages
- 29 workflows
- 8 books

**Aim**: interdisciplinary research, collaboration and rapid development of scientific software

## Documentation

- function manual pages, most of them with runnable examples
- package vignettes - mandatory here!
- workflows, documenting full analyses spanning multiple tools
- a very active support site

# The real deal: Bioconductor's community



# Data processing

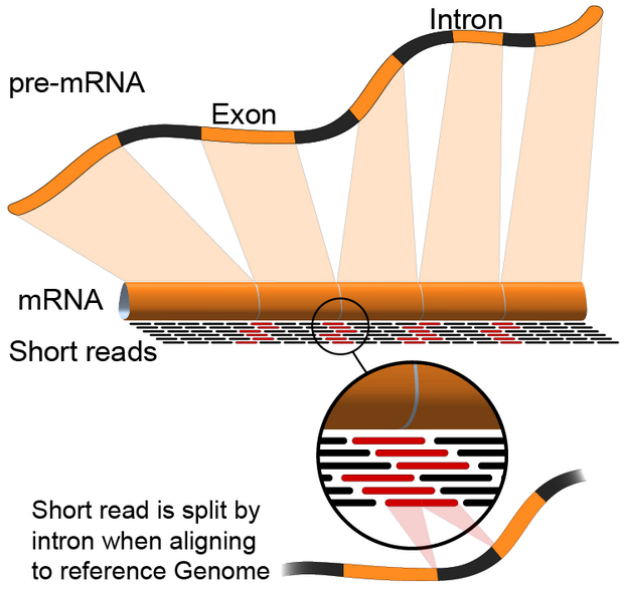
# Data processing

Turning millions of text lines into properly structured abundance tables

Our aim: we often want to compare abundance (expression) of genes or other features between conditions

Splice-aware genome alignment vs "direct" transcript mapping and quantification

# Alignment: not just simple mapping




For RNA-seq data, we need a splice-aware aligner

Common choices:

- STAR
- HISAT2

# Alignment

File-format: sam (compressed into bam)

```
SRR1055095.6079377  353  chr1    11167    0    50M    =    11751    634    CGCCCCTTGCTTGCAGCCGGGCAC
TACAGGACCCGCTTGCTCACGGTGAA    CCCFFFFFHHHHHJJJJJJJJJJJJIGJJJIJJJJJJJJJJJDHJJCEHH
AS:i:-10    XN:i:0    XM:i:2    XO:i:0    XG:i:0    NM:i:2    MD:Z:48C0T0    YT:Z:UU    NH:i:20
CC:Z:chrY    CP:i:59361513    HI:i:0
```

Again: repeat this, one read at a time!

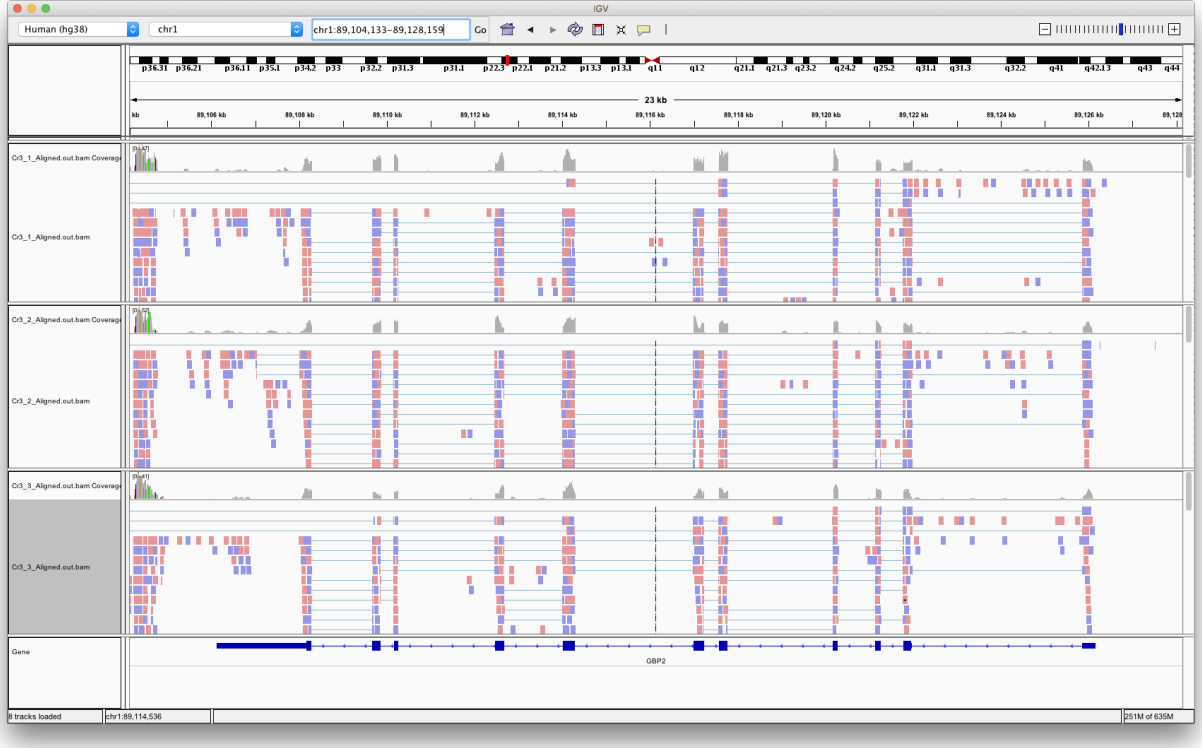
Entries:

```
QNAME - Query NAME of the read or the read pair
FLAG - Bitwise FLAG (pairing, strand, mate strand, etc.)
RNAME - Reference sequence NAME
POS - 1-Based leftmost POSition of clipped alignment
MAPQ - MAPping Quality (Phred-scaled)
CIGAR - Extended CIGAR string (operations: MIDNSHP)
MRNM - Mate Reference NaMe ('=' if same as RNAME)
MPOS - 1-Based leftmost Mate POSition
ISIZE - Inferred Insert SIZE
SEQ - Query SEQuence on the same strand as the reference
QUAL - Query QUALity (ASCII-33=Phred base quality)
Tags: used to store info about alignment
```

# Before quantification

… and actually, always: Do visualize your data!

Options: UCSC Genome Browser, IGV, IGB - http://software.broadinstitute.org/software/igv/



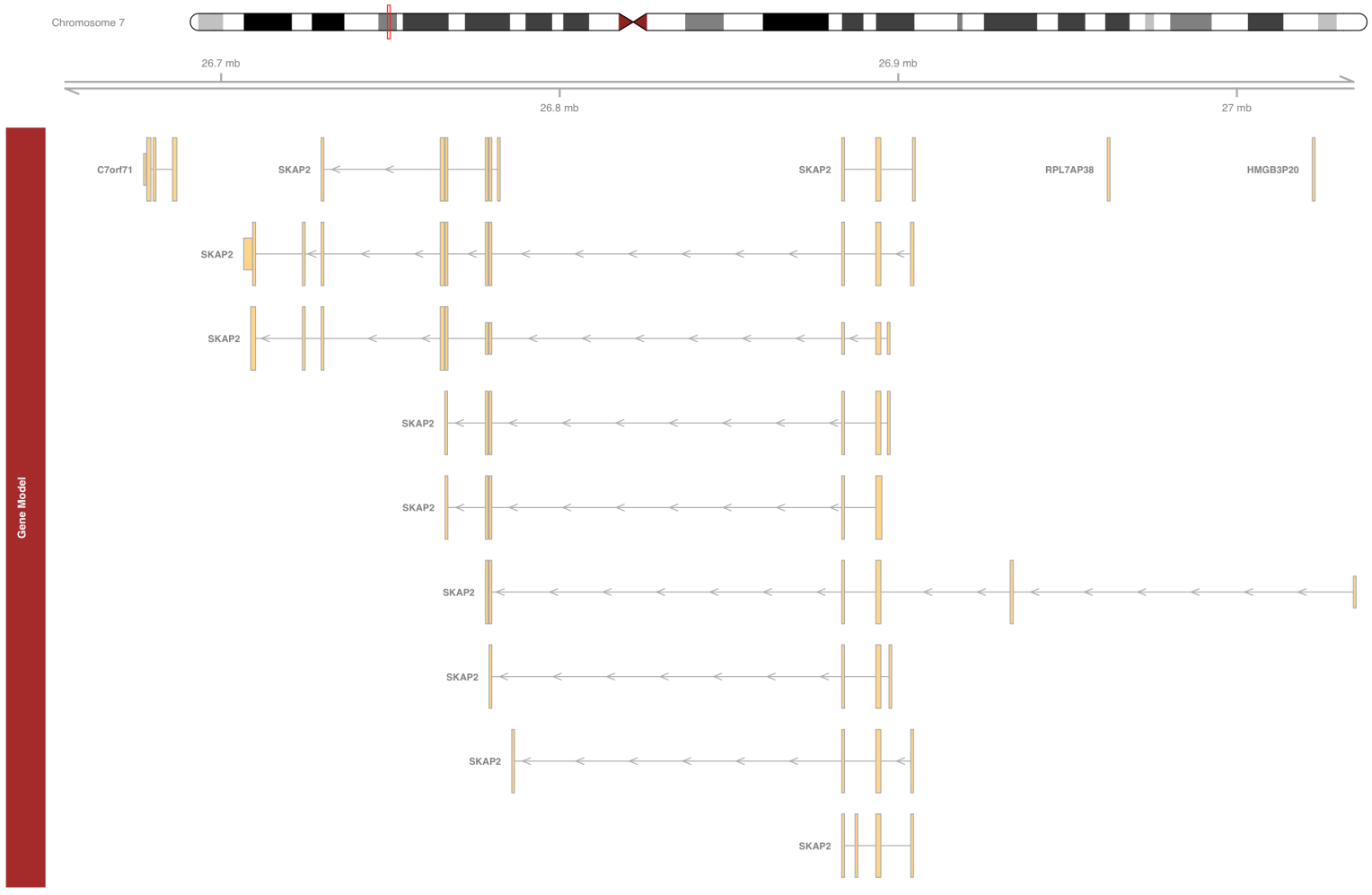
# Quantification

Ingredients: BAM data + GTF annotation file

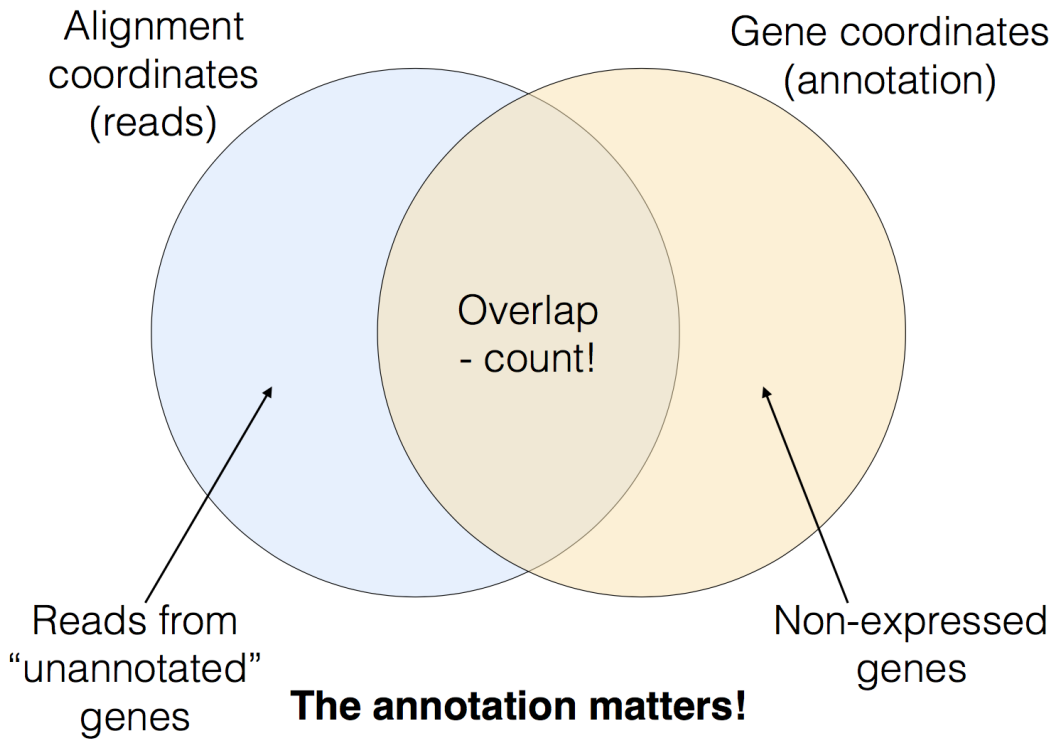
Output: number of reads overlapping known features (discrete, positive, skewed)

Gene-level counts, often obtained by genome alignment + overlap counting

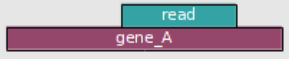
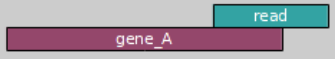
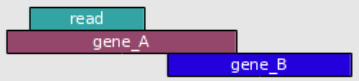
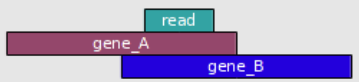
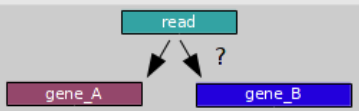
# The role of annotation



# The annotation matters!



# The annotation matters!

| | union | intersection_strict | intersection_nonempty |
|---|---|---|---|
|  | gene_A | gene_A | gene_A |
|  | gene_A | no_feature | gene_A |
|  | gene_A | no_feature | gene_A |
|  | gene_A | gene_A | gene_A |
|  | gene_A | gene_A | gene_A |
|  | ambiguous (both genes with --nonunique all) | gene_A | gene_A |
|  | ambiguous (both genes with --nonunique all) | | |
|  | alignment_not_unique (both genes with --nonunique all) | | |

# Alignment-free quantifications

Some recently developed methods:

- salmon (Patro et al, Nat Methods 2017)
- kallisto (Bray et al, Nat Biotechnol 2016)

return…

- Transcript-level counts and TPM (transcripts-per-million) estimates, which can be summed up to get
- Gene-level counts and TPM estimates

Pros & cons

- considerably faster than traditional alignment+counting -> allow bootstrapping
- more highly resolved estimates (transcripts rather than gene) + can be aggregated
- can use a slightly larger fraction of the reads since multi-mapping reads are not excluded
- don't return precise alignments (bam files, for e.g. visualization in genome browser)

# Which way to go?

Based on genome alignment - mainly gene-level quantification: combine exons, "ignoring" splice variants

- Simple, powerful, yet in some cases inaccurate
- Tools:
  - `htseq-count`, `featureCounts` for estimating expression levels (counts)
  - `edgeR`, `DESeq2`, `voom+limma` for statistical modeling

Based on transcriptome mapping - transcript- and gene-level quantification: 'assign' reads (or rather, estimate most likely expression level) based on probabilistic modeling

- Potentially cleaner, but high degree of uncertainty on the transcript level!
- Tools:
  - `bitSeq`, `RSEM`, `salmon`, `kallisto` for (pseudo)alignment/quantification
  - `DESeq2`, `edgeR`, `voom+limma`, `swish`, `DRIMseq`, `DEXSeq`, `sleuth` for modeling (depending on the question of interest)

# What it would look like - STAR + featureCounts:

... due to time constraints

Index the genome

```
$ STAR --runThreadN 24 \
       --runMode genomeGenerate \
       --genomeDir my_genome \
       --genomeFastaFiles my_genome.fa \
       --sjdbGTFfile my_genes.gtf \
       --sjdbOverhang 99
```

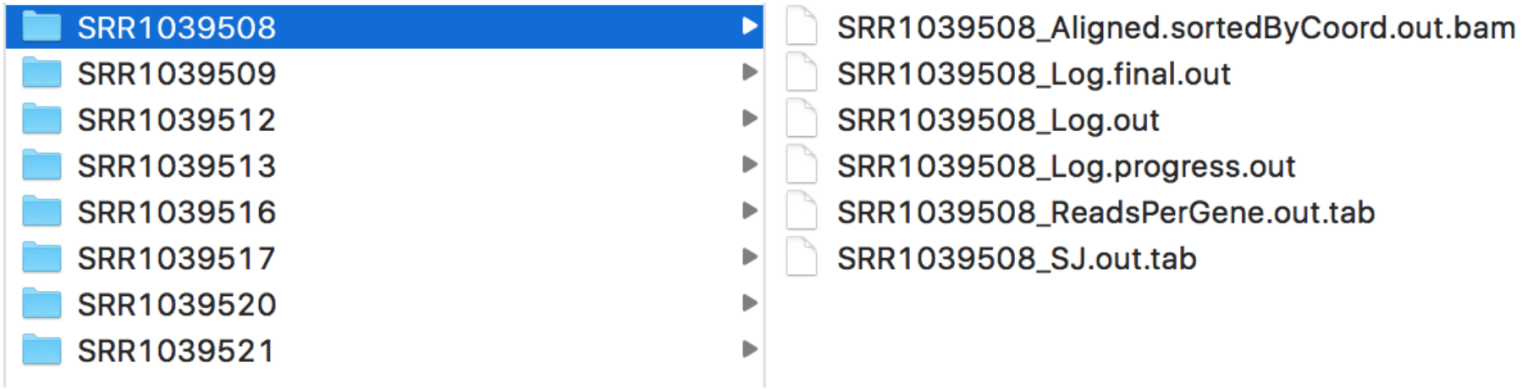
# What it would look like - STAR + featureCounts:

... due to time constraints

Map each file

```
$ STAR --runThreadN 24 \
       --runMode alignReads \
       --genomeDir my_genome \
       --readFilesIn my_sample_read1.fastq.gz \
                     my_sample_read2.fastq.gz \
       --readFilesCommand zcat \
       --outFileNamePrefix output/S1/ \
       --outSAMtype BAM SortedByCoordinate \
       --quantMode GeneCounts
```

# What it would look like - STAR + featureCounts:



# What it would look like - STAR + featureCounts:

... due to time constraints

Quantify

```
featureCounts(files = bamfiles,
              annot.ext = "my_genes.gtf",
              isGTFAnnotationFile = TRUE,
              GTF.featureType = "exon",
              GTF.attrType = "gene_id",
              useMetaFeatures = TRUE,
              isPairedEnd = TRUE,
              strandSpecific = 0)
```

Directly generates a count matrix in your R session.

# What it would look like - salmon

... due to time constraints

Create an index of the transcriptome

```
$ salmon index -i my_transcripts.idx \
               -t <(cat my_transcripts.fasta my_genome.fasta) \
               -d chromosome_names.txt
```

The genome acts as a 'decoy' sequence, to collect reads truly arising from intronic or intergenic locations.

# What it would look like - salmon

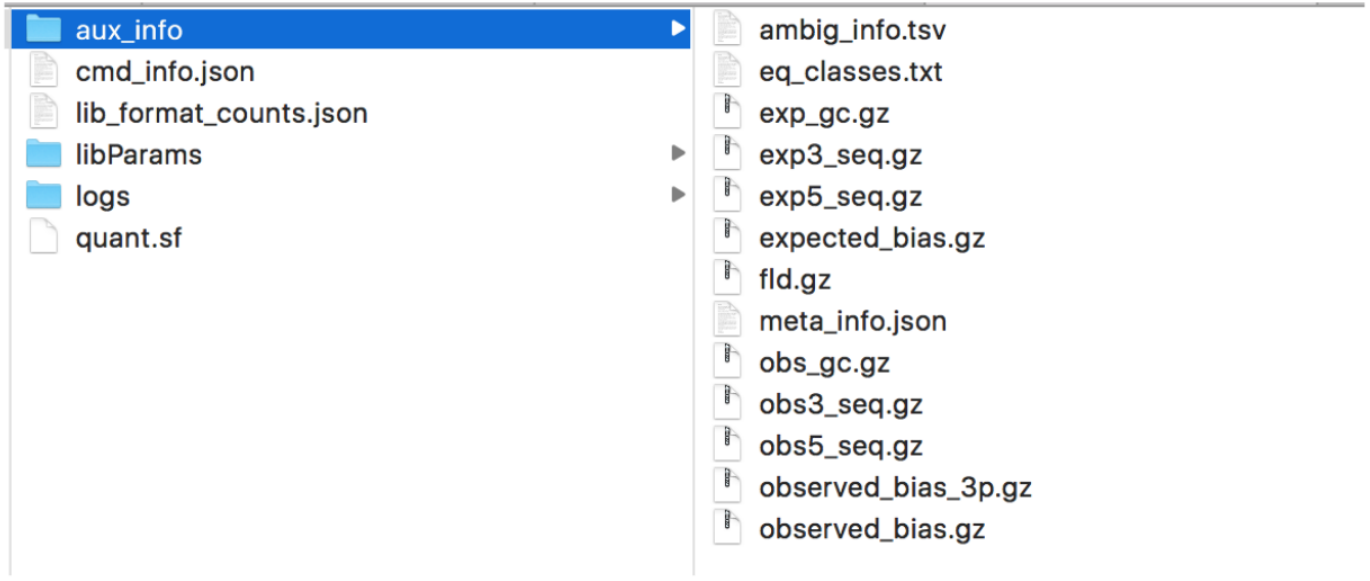
... due to time constraints

Quantify a sample at the transcript level

```
$ salmon quant -i my_transcripts.idx -l A \
               -1 my_sample_read1.fastq.gz -2 my_sample_read2.fastq.gz \
               -p 10 -o results/sample1 --validateMappings \
               --numBootstraps 30 --seqBias --gcBias
```

# What it would look like - salmon

… due to time constraints



# What it would look like - salmon

… due to time constraints

Salmon

[quant.sf]

| Name | Length | EffectiveLength | TPM | NumReads |
|---|---|---|---|---|
| ENST00000406070 | 2025 | 1869.81 | 0 | 0 |
| ENST00000446844 | 2227 | 2071.81 | 0.137334 | 3.71695 |
| ENST00000599620 | 686 | 530.936 | 0 | 0 |
| ENST00000471557 | 505 | 350.256 | 0.731211 | 3.3457 |
| ENST00000338761 | 1456 | 1300.81 | 0 | 0 |
| ENST00000417509 | 1444 | 1288.81 | 7.58582e-08 | 1.27717e-06 |
| ENST00000484946 | 610 | 455.039 | 2.87905 | 17.1142 |
| ENST00000490656 | 660 | 504.969 | 1.46703 | 9.67744 |
| ENST00000439537 | 1161 | 1005.81 | 1.47611 | 19.3952 |
| ENST00000493251 | 641 | 485.994 | 0.597774 | 3.79512 |
| ENST00000460127 | 408 | 253.708 | 0 | 0 |

# Importing salmon quantifications into R

You can follow (offline) the instructions of the `tximport` package - https://bioconductor.org/packages/tximport/

`tximeta`: another precious assistant on the way to be consistent and to keep track of provenance identification (we'll see it in action during the exercises)

# What does our data look like now?

| | SRR1039508 | SRR1039509 | SRR1039512 | SRR1039513 | SRR1039516 | SRR1039517 | SRR1039520 | SRR1039521 |
|---|---|---|---|---|---|---|---|---|
| ENSG00000000003 | 679 | 448 | 873 | 408 | 1138 | 1047 | 770 | 572 |
| ENSG00000000005 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ENSG00000000419 | 467 | 515 | 621 | 365 | 587 | 799 | 417 | 508 |
| ENSG00000000457 | 260 | 211 | 263 | 164 | 245 | 331 | 233 | 229 |
| ENSG00000000460 | 60 | 55 | 40 | 35 | 78 | 63 | 76 | 60 |
| ENSG00000000938 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 |
| ENSG00000000971 | 3251 | 3679 | 6177 | 4252 | 6721 | 11027 | 5176 | 7995 |
| ENSG00000001036 | 1433 | 1062 | 1733 | 881 | 1424 | 1439 | 1359 | 1109 |
| ENSG00000001084 | 519 | 380 | 595 | 493 | 820 | 714 | 696 | 704 |
| ENSG00000001167 | 394 | 236 | 464 | 175 | 658 | 584 | 360 | 269 |
| ENSG00000001460 | 172 | 168 | 264 | 118 | 241 | 210 | 155 | 177 |
| ENSG00000001461 | 2112 | 1867 | 5137 | 2657 | 2735 | 2751 | 2467 | 2905 |
| ENSG00000001497 | 524 | 488 | 638 | 357 | 676 | 806 | 493 | 475 |
| ENSG00000001561 | 71 | 51 | 211 | 156 | 23 | 38 | 134 | 172 |
| ENSG00000001617 | 555 | 394 | 905 | 415 | 727 | 697 | 618 | 599 |
| ENSG00000001626 | 10 | 2 | 9 | 2 | 10 | 6 | 5 | 5 |
| ENSG00000001629 | 1660 | 1251 | 2259 | 1079 | 2462 | 2514 | 1888 | 1660 |
| ENSG00000001630 | 59 | 54 | 66 | 23 | 84 | 87 | 31 | 59 |
| ENSG00000001631 | 729 | 692 | 943 | 475 | 1034 | 1163 | 731 | 744 |
| ENSG00000002016 | 201 | 161 | 256 | 99 | 268 | 257 | 160 | 137 |
| ENSG00000002079 | 3 | 0 | 3 | 1 | 4 | 0 | 0 | 1 |
| ENSG00000002330 | 206 | 174 | 184 | 111 | 194 | 260 | 156 | 177 |

# Some challenges in RNA-seq data analysis

1 - Choosing an appropriate statistical distribution

2 - Normalization between samples

3 - Few samples available make it difficult to estimate parameters (e.g., variance)

4 - Many genes, many tests - high dimensionality

# Some challenges in RNA-seq data analysis - 1

**Choosing an appropriate statistical distribution**

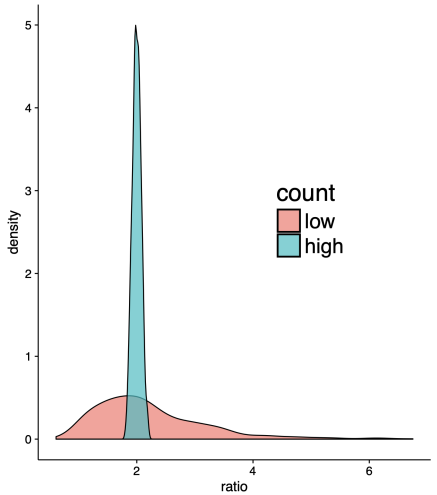
Variance depends on the mean count

Counts are non-negative and often highly skewed
This means you can't just use t-tests, ANOVA - no prob, `glm`s to the rescue!

Poisson -> negative binomial, better captures variability across biological replicates

# "Why do we not just take the ratios?"

Fold changes, relative abundances




Ex: ratio between two Poisson distributed variables

Low: mean = 20 vs mean = 10
High: mean = 2000 vs mean = 1000

Which one would you trust more? Why?

This goes back to having appropriate statistical frameworks that nicely model your datasets (and how these get generated)

# Some challenges in RNA-seq data analysis - 2

**Normalization between samples**

Observed counts depend on:

- abundance level
- gene/transcript length
- sequencing depth
- sequencing biases

"As-is" estimates not directly comparable across samples

Normalization aims to ensure our expression estimates are

- comparable across features (genes, isoforms, etc)
- comparable across libraries (different samples)
- on a human-friendly scale (interpretable magnitude)

Most RNA-seq methods (e.g., edgeR, DESeq2, voom) need raw counts (or equivalent) as input
Don’t provide these methods with (e.g.) RPKMs, FPKMs, TPMs, CPMs, log-transformed counts, normalized counts, ...
Read documentation carefully!

# Digression: Normalization expression units

- RPKM/FPKM (Reads/Fragments per kilobase of transcript per million reads of library)
    - Corrects for total library coverage
    - Corrects for gene length
    - Comparable between different genes within the same dataset
- TPM (transcripts per million)
    - normalizes to transcript copies instead of reads - gives an idea of the proportion of transcripts
    - Corrects for cases where the total RNA output differs between samples
    - More appropriate for between sample comparisons (the sum of all TPMs in each sample are the same)

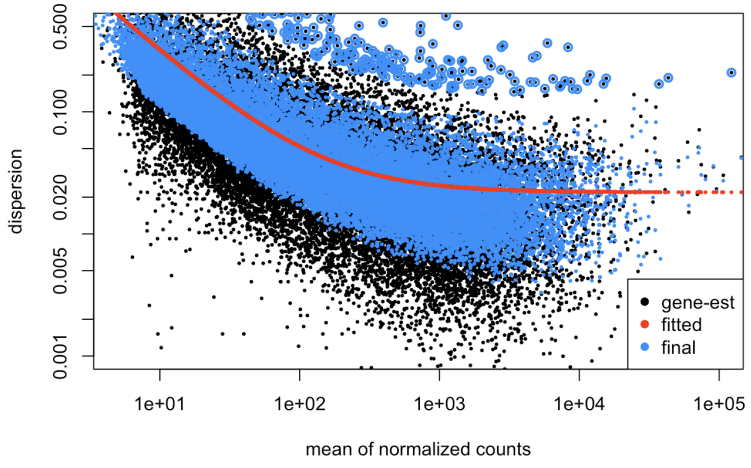
For DE analysis you have to work with discrete counts…

… and for comparisons you can use normalized counts (median ratio/TMM methods are robust across all genes!)

# Some challenges in RNA-seq data analysis - 3

**Few samples available make it difficult to estimate parameters (e.g., variance)**

You can take advantage of the large number of genes




-> Shrink the gene-wise estimates towards a center value defined by the observed distribution of dispersions across...

- all genes (common dispersion estimate)
- genes with similar expression (trended dispersion estimate)

# Some challenges in RNA-seq data analysis - 4

**Many genes, many tests - high dimensionality**

FDR for multiple test correction

Some more ideas:

- filter out genes that have little chance of showing significance (without looking at the test results)
- independent hypothesis weighting

... all nicely implemented for the DESeq framework

Do not test AND filter on logFC post-hoc!

**Think** of the null you're testing against - $\beta = 0$ by default, useful to adapt if you want to focus on larger effect sizes

# Exploratory analysis and visualization

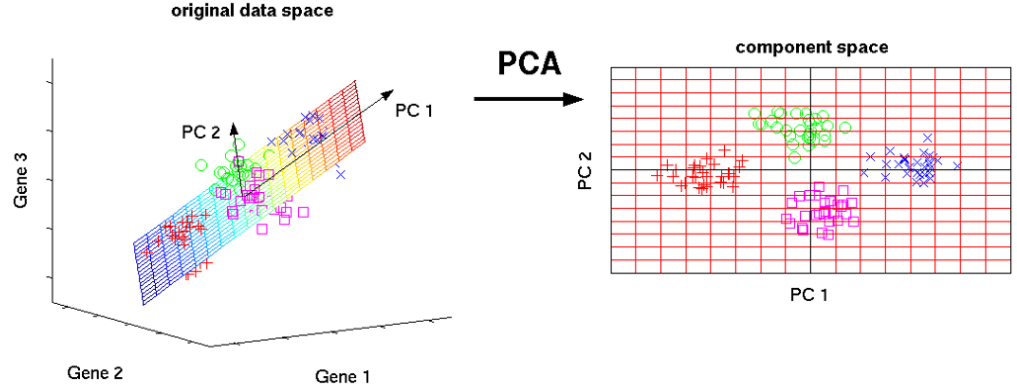
Is the data of good quality?

Quality as "fitness for purpose": DE

Very important: Transforming the data is often required for better further explorations (here: variance stabilization, regularized logarithm…)

One notable example, commonly used: Principal Components Analysis.

```
The data points (here, the samples) are projected onto the 2D plane
such that they spread out in the two directions that explain most
of the differences (the variability)
```



# Exploratory analysis and visualization

"Perspective matters"

# Gene identifiers

Take for example **GBP2**, guanylate binding protein 2 (4 transcripts available)

Ensembl ID: ENSG00000162645 (human), ENSMUSG00000028270 (mouse)

http://www.ensembl.org/id/ENSG00000162645

Entrez ID: 2634

HGNC ID: 4183

RefSeq ID: NM_004120

UCSC ID: uc001dmz.3

Official symbol: GBP2

Synonyms: - (sometimes makes the ambiguity even bigger!!!)

Gene symbols can change over time!

Typically, no 1:1 mapping between different ID types

# Working with Excel

Don't do it - and not just because I do like R

Correspondence | Open Access | Published: 23 June 2004

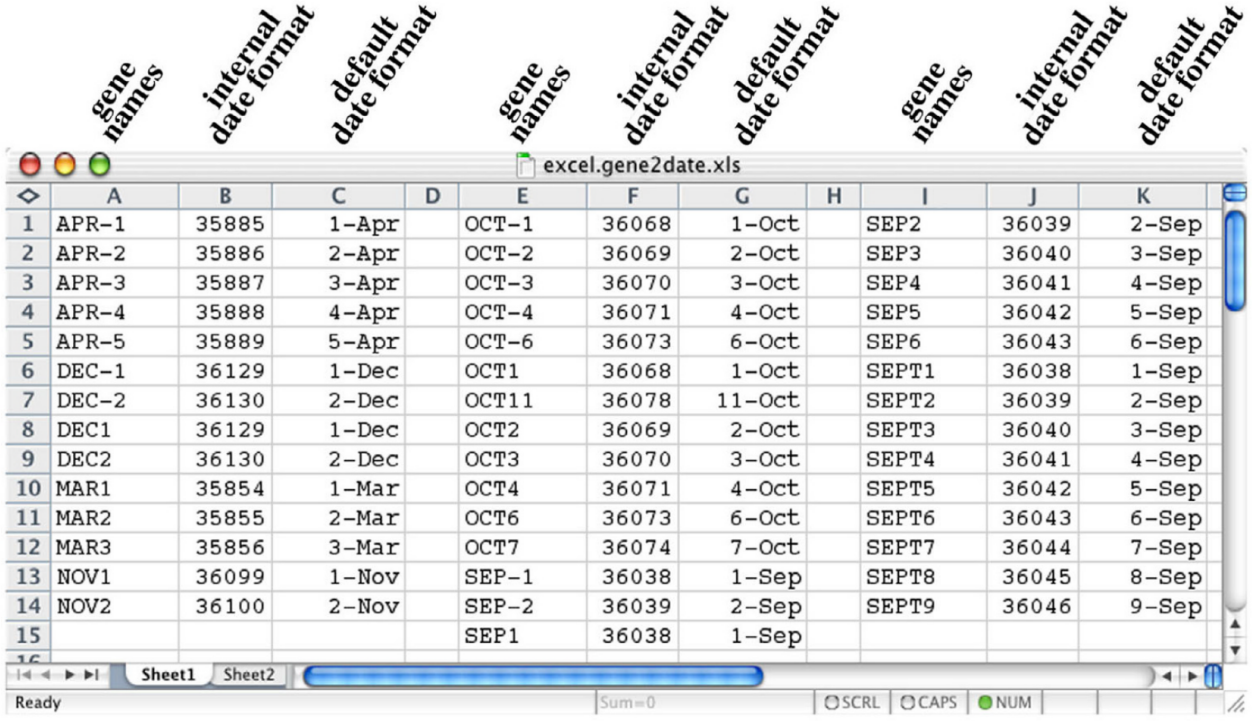
## Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

Barry R Zeeberg, Joseph Riss, David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett & John N Weinstein

*BMC Bioinformatics* **5**, Article number: 80 (2004) | Cite this article

# Working with Excel

There are excellent reasons *not* to do it

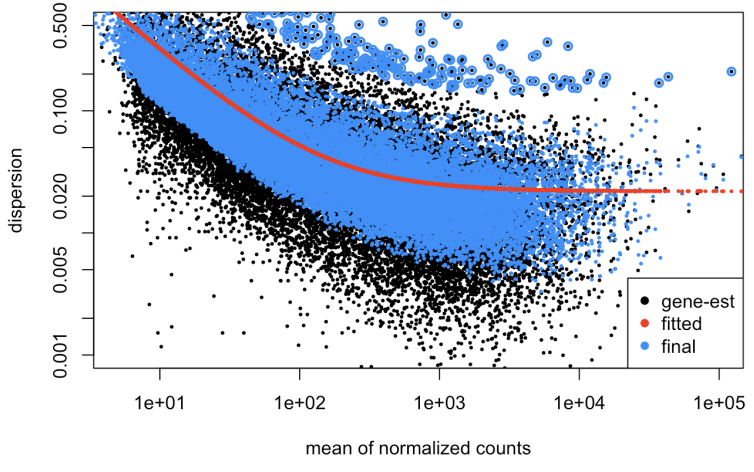



| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | APR-1 | 35885 | 1-Apr | | OCT-1 | 36068 | 1-Oct | | SEP2 | 36039 | 2-Sep |
| 2 | APR-2 | 35886 | 2-Apr | | OCT-2 | 36069 | 2-Oct | | SEP3 | 36040 | 3-Sep |
| 3 | APR-3 | 35887 | 3-Apr | | OCT-3 | 36070 | 3-Oct | | SEP4 | 36041 | 4-Sep |
| 4 | APR-4 | 35888 | 4-Apr | | OCT-4 | 36071 | 4-Oct | | SEP5 | 36042 | 5-Sep |
| 5 | APR-5 | 35889 | 5-Apr | | OCT-6 | 36073 | 6-Oct | | SEP6 | 36043 | 6-Sep |
| 6 | DEC-1 | 36129 | 1-Dec | | OCT1 | 36068 | 1-Oct | | SEPT1 | 36038 | 1-Sep |
| 7 | DEC-2 | 36130 | 2-Dec | | OCT11 | 36078 | 11-Oct | | SEPT2 | 36039 | 2-Sep |
| 8 | DEC1 | 36129 | 1-Dec | | OCT2 | 36069 | 2-Oct | | SEPT3 | 36040 | 3-Sep |
| 9 | DEC2 | 36130 | 2-Dec | | OCT3 | 36070 | 3-Oct | | SEPT4 | 36041 | 4-Sep |
| 10 | MAR1 | 35854 | 1-Mar | | OCT4 | 36071 | 4-Oct | | SEPT5 | 36042 | 5-Sep |
| 11 | MAR2 | 35855 | 2-Mar | | OCT6 | 36073 | 6-Oct | | SEPT6 | 36043 | 6-Sep |
| 12 | MAR3 | 35856 | 3-Mar | | OCT7 | 36074 | 7-Oct | | SEPT7 | 36044 | 7-Sep |
| 13 | NOV1 | 36099 | 1-Nov | | SEP-1 | 36038 | 1-Sep | | SEPT8 | 36045 | 8-Sep |
| 14 | NOV2 | 36100 | 2-Nov | | SEP-2 | 36039 | 2-Sep | | SEPT9 | 36046 | 9-Sep |
| 15 | | | | | SEP1 | 36038 | 1-Sep | | | | |

# Differential expression analysis, with DESeq2

There is one main function, `DESeq`, that does the following

- estimation of size factors
- estimation of the dispersion values for each gene
- fitting of the generalized linear model and performing statistical testing



# What do our results look like?

| | id | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|---|
| 1 | ENSG00000152583 | 997.43977 | -4.5749187 | 0.18405609 | -24.856111 | 2.220933e-136 | 4.003898e-132 |
| 2 | ENSG00000165995 | 495.09291 | -3.2910618 | 0.13317370 | -24.712551 | 7.839410e-135 | 7.066444e-131 |
| 3 | ENSG00000120129 | 3409.02938 | -2.9478099 | 0.12143769 | -24.274258 | 3.666925e-130 | 2.203577e-126 |
| 4 | ENSG00000101347 | 12703.38706 | -3.7669954 | 0.15543799 | -24.234715 | 9.583815e-130 | 4.319425e-126 |
| 5 | ENSG00000189221 | 2341.76725 | -3.3535799 | 0.14178235 | -23.653014 | 1.098955e-123 | 3.962392e-120 |
| 6 | ENSG00000211445 | 12285.61515 | -3.7304027 | 0.16583058 | -22.495263 | 4.618318e-112 | 1.387651e-108 |
| 7 | ENSG00000157214 | 3009.26322 | -1.9767725 | 0.08998232 | -21.968454 | 5.769975e-107 | 1.486016e-103 |
| 8 | ENSG00000162614 | 5393.10168 | -2.0356649 | 0.09418231 | -21.614091 | 1.323849e-103 | 2.983294e-100 |
| 9 | ENSG00000125148 | 3656.25278 | -2.2109786 | 0.10564078 | -20.929214 | 2.902397e-97 | 5.813824e-94 |
| 10 | ENSG00000154734 | 30315.13547 | -2.3456037 | 0.11580806 | -20.254235 | 3.259713e-91 | 5.876611e-88 |
| 11 | ENSG00000139132 | 1223.46614 | -2.2289030 | 0.11283944 | -19.752872 | 7.578284e-87 | 1.242012e-83 |
| 12 | ENSG00000162493 | 1099.62587 | -1.8912170 | 0.09577959 | -19.745511 | 8.767336e-87 | 1.317146e-83 |
| 13 | ENSG00000134243 | 5510.95826 | -2.1957116 | 0.11200768 | -19.603224 | 1.451321e-85 | 2.012647e-82 |
| 14 | ENSG00000179094 | 776.59667 | -3.1917499 | 0.16396292 | -19.466291 | 2.120834e-84 | 2.731028e-81 |
| 15 | ENSG00000162692 | 508.17023 | 3.6926606 | 0.19018590 | 19.416059 | 5.646022e-84 | 6.785765e-81 |
| 16 | ENSG00000163884 | 561.10717 | -4.4591282 | 0.23486369 | -18.986027 | 2.225442e-80 | 2.507517e-77 |
| 17 | ENSG00000178695 | 2649.85015 | 2.5281746 | 0.13443012 | 18.806607 | 6.667027e-79 | 7.070186e-76 |
| 18 | ENSG00000198624 | 2057.19173 | -2.9184357 | 0.16129634 | -18.093626 | 3.577533e-73 | 3.583098e-70 |
| 19 | ENSG00000107562 | 25136.30757 | 1.9116698 | 0.10574810 | 18.077582 | 4.786167e-73 | 4.541317e-70 |
| 20 | ENSG00000148848 | 1365.17399 | 1.8145431 | 0.10048365 | 18.058094 | 6.813522e-73 | 6.141709e-70 |

I would like to understand more what is going on here

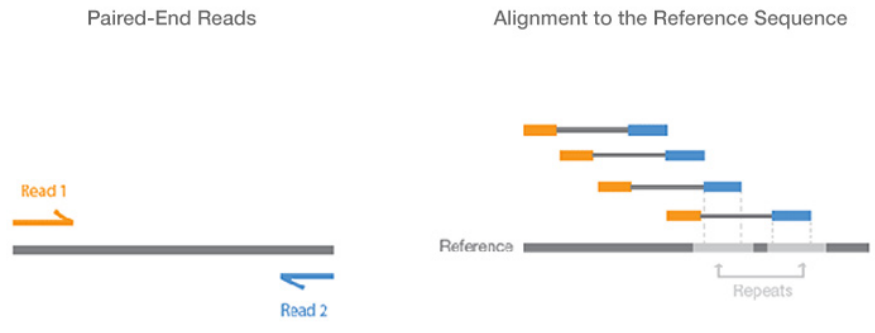
# Practical tips for your (next) RNA-seq dataset

## Single- vs paired-end sequencing

Each fragment can be sequenced from one end only, or from both ends

Single-end cheaper and faster

Paired-end provide improved ability to localize the fragment in the genome and resolve mapping close to repeat regions - less multimapping reads




Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

# Practical tips for your (next) RNA-seq dataset

## Strand-specificity

In “standard” protocols, we don’t know from which strand a read stems

Various “strand-specific” protocols allow us to keep this information

Strand-specificity leads to lower number of ambiguous reads (overlapping multiple genes)

# Practical tips for your (next) RNA-seq dataset

## Sequencing depth and number of replicates

Recommendation: At least 3 biological replicates - Conesa et al, Genome Biol, 2016

Recommendation: At least 6 biological replicates - Schurch

Diminishing returns in increasing the read depth - go for replicates given a fixed budget!

# I am not kidding you

# Practical tips for your (next) RNA-seq dataset

## Bulk or single cell?

That depends on your question!

Did someone say heterogeneity?

# Questions

What are the steps to process RNA-Seq data?

- How to convert RNA-seq reads into counts?
- How to perform quality control (QC) of RNA-seq reads?

How to identify differentially expressed genes across multiple experimental conditions?

- How to properly analyze RNA count data using DESeq2?
- How to perform quality control (QC) and exploratory data analysis (EDA) of RNA-seq count data?

What are the biological functions impacted by the differential expression of genes?

- How can I perform a gene ontology enrichment analysis?

How can I create neat visualizations of the data?

- How can I visualize the results for my enrichment analysis?

How can I generate interactive reports to summarise my analyses?

# Questions

What are the steps to process RNA-Seq data?

- How to convert RNA-seq reads into counts?
- How to perform quality control (QC) of RNA-seq reads?

How to identify differentially expressed genes across multiple experimental conditions?

- How to properly analyze RNA count data using DESeq2?
- How to perform quality control (QC) and exploratory data analysis (EDA) of RNA-seq count data?

**What are the biological functions impacted by the differential expression of genes?**

- How can I perform a gene ontology enrichment analysis?

**How can I create neat visualizations of the data?**

- How can I visualize the results for my enrichment analysis?

**How can I generate interactive reports to summarise my analyses?**

# Goals

after constructing and running a differential gene expression analysis...

- Perform a functional enrichment analysis
- Perform and visualize the enrichment analysis for Gene Ontology terms
- Get familiar with representations of the relationship between genes and gene sets

## Prerequisites

- a basic understanding of R
- familiarity with gene expression data (RNA-seq)
- familiarity with the concept of differential expression analysis

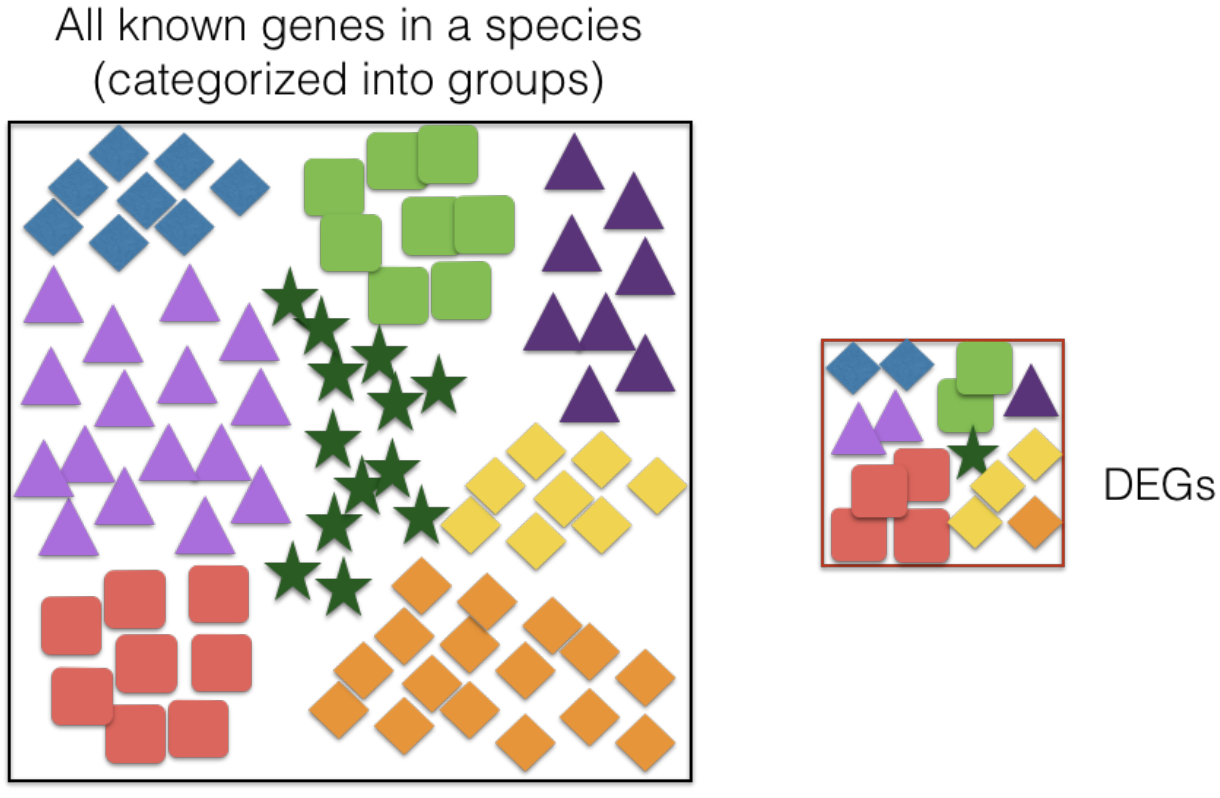
# What do our results look like?

| | id | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|---|
| 1 | ENSG00000152583 | 997.43977 | -4.5749187 | 0.18405609 | -24.856111 | 2.220933e-136 | 4.003898e-132 |
| 2 | ENSG00000165995 | 495.09291 | -3.2910618 | 0.13317370 | -24.712551 | 7.839410e-135 | 7.066444e-131 |
| 3 | ENSG00000120129 | 3409.02938 | -2.9478099 | 0.12143769 | -24.274258 | 3.666925e-130 | 2.203577e-126 |
| 4 | ENSG00000101347 | 12703.38706 | -3.7669954 | 0.15543799 | -24.234715 | 9.583815e-130 | 4.319425e-126 |
| 5 | ENSG00000189221 | 2341.76725 | -3.3535799 | 0.14178235 | -23.653014 | 1.098955e-123 | 3.962392e-120 |
| 6 | ENSG00000211445 | 12285.61515 | -3.7304027 | 0.16583058 | -22.495263 | 4.618318e-112 | 1.387651e-108 |
| 7 | ENSG00000157214 | 3009.26322 | -1.9767725 | 0.08998232 | -21.968454 | 5.769975e-107 | 1.486016e-103 |
| 8 | ENSG00000162614 | 5393.10168 | -2.0356649 | 0.09418231 | -21.614091 | 1.323849e-103 | 2.983294e-100 |
| 9 | ENSG00000125148 | 3656.25278 | -2.2109786 | 0.10564078 | -20.929214 | 2.902397e-97 | 5.813824e-94 |
| 10 | ENSG00000154734 | 30315.13547 | -2.3456037 | 0.11580806 | -20.254235 | 3.259713e-91 | 5.876611e-88 |
| 11 | ENSG00000139132 | 1223.46614 | -2.2289030 | 0.11283944 | -19.752872 | 7.578284e-87 | 1.242012e-83 |
| 12 | ENSG00000162493 | 1099.62587 | -1.8912170 | 0.09577959 | -19.745511 | 8.767336e-87 | 1.317146e-83 |
| 13 | ENSG00000134243 | 5510.95826 | -2.1957116 | 0.11200768 | -19.603224 | 1.451321e-85 | 2.012647e-82 |
| 14 | ENSG00000179094 | 776.59667 | -3.1917499 | 0.16396292 | -19.466291 | 2.120834e-84 | 2.731028e-81 |
| 15 | ENSG00000162692 | 508.17023 | 3.6926606 | 0.19018590 | 19.416059 | 5.646022e-84 | 6.785765e-81 |
| 16 | ENSG00000163884 | 561.10717 | -4.4591282 | 0.23486369 | -18.986027 | 2.225442e-80 | 2.507517e-77 |
| 17 | ENSG00000178695 | 2649.85015 | 2.5281746 | 0.13443012 | 18.806607 | 6.667027e-79 | 7.070186e-76 |
| 18 | ENSG00000198624 | 2057.19173 | -2.9184357 | 0.16129634 | -18.093626 | 3.577533e-73 | 3.583098e-70 |
| 19 | ENSG00000107562 | 25136.30757 | 1.9116698 | 0.10574810 | 18.077582 | 4.786167e-73 | 4.541317e-70 |
| 20 | ENSG00000148848 | 1365.17399 | 1.8145431 | 0.10048365 | 18.058094 | 6.813522e-73 | 6.141709e-70 |

I need to understand more what is going on here

# Functional enrichment analysis

Our problem, at a glance



# Functional enrichment analysis

Test whether known biological functions or processes are over-represented (= enriched) in an experimentally-derived gene list, e.g. a list of differentially expressed (DE) genes.

**Example:** Transcriptomic study, in which 12671 genes have been tested for differential expression between two sample conditions and 529 genes were found DE

Among the DE genes, 28 are annotated to a specific functional gene set, which contains in total 170 genes. This setup corresponds to a 2x2 contingency table:

```
deTable <- matrix(
  c(28, 142, 501, 12000),
  nrow = 2,
  dimnames = list(c("DE", "Not_DE"),
                  c("In_gene_set", "Not_in_gene_set")))
deTable
#        In_gene_set Not_in_gene_set
# DE              28             501
# Not_DE         142           12000
```

```
fisher.test(deTable, alternative = "greater")
#
#     Fisher's Exact Test for Count Data
```

# Some terminology: Gene sets, pathways, networks

**Gene sets** are simple lists of usually functionally related genes without further specification of relationships between genes. **Any** *a priori* classification of genes into biologically relevant groups. Sets do not need to be exhaustive or disjoint

**Pathways** can be interpreted as specific gene sets, typically representing a group of genes that work together in a biological process.

**Gene regulatory networks** describe the interplay and effects of regulatory factors (such as transcription factors and microRNAs) on the expression of their target genes.

Is the expression of genes in a gene set associated with the experimental conditions?

| Genes categories | Organism-specific Background | DE results | Over-represented? |
|---|---|---|---|
| Functional category 1 | 35/13000 | 25/1000 | Likely |
| Functional category 2 | 56/13000 | 4/1000 | Unlikely |
| Functional category 3 | 90/13000 | 8/1000 | Unlikely |
| Functional category 4 | 15/13000 | 10/1000 | Likely |
| ... | | | |
| ... | | | |

# Functional enrichment analysis

Primarily developed and applied on transcriptomic data, now extended and applied also in other fields of genomic and biomedical research (proteomic and metabolomic data, genomic regions, ...)

Many methods available!

- Over-representation analysis (ORA) - are differentially expressed (DE) genes in the set more common than expected? Based on (variations) of the 2x2 contingency table method
- Functional class scoring (FCS) - summarize gene set (= functional class) scores by summarizing statistic of DE of genes in a set, and compare to null
- Pathway topology (PT) - explicitly taking into account interactions between genes as defined in signaling pathways and gene regulatory networks

Topology-based methods appear to be most realistic, but: features that are not-detectable on the transcriptional level + insufficient network knowledge

**Cautious interpretation of results is required to derive valid conclusions!**

# Collections of gene sets

Gene Ontology (GO) Annotation (GOA)

- CC Cellular Components
- BP Biological Processes
- MF Molecular Function

Pathways

- MSigDb
- KEGG
- reactome
- PantherDB
- ...

GO and KEGG annotations are most frequently used for the enrichment analysis of functional gene sets - likely due to their long-standing curation and availability for a wide range of species

# Software for functional enrichment

- DAVID
- GSEA and its variations
- inside Bioconductor: `reactome`, `topGO`, `pathview`, `GOSeq`, ...
- Ingenuity Pathway Analysis

Not straightforward, but amazing way of helping generating hypothesis for the bench scientist

**Friendly tip:** wrappers for topGO and goseq are in `pcaExplorer` and in `ideal`!

## General recommendations

Never forget to visualize your data!

Pick candidates, sets of candidates, plot instead of guessing!

The choice of a background matters!

# Background matters

... with a slight modification of the example above

```
deTable_detected <- matrix(
  c(8, 142, 501, 12000),
  nrow = 2,
  dimnames = list(c("DE", "Not_DE"),
                  c("In_gene_set", "Not_in_gene_set")))
fisher.test(deTable_detected, alternative = "greater")
```

versus...

```
deTable_allgenes <- matrix(
  c(8, 142, 501, 32000),
  nrow = 2,
  dimnames = list(c("DE", "Not_DE"),
                  c("In_gene_set", "Not_in_gene_set")))
fisher.test(deTable_allgenes, alternative = "greater")
```

# Data analysis: where's the real bottleneck?

✅ Efficient methods to process your data (entire workflows available, see e.g. the Bioconductor workflow packages)

✅ Compelling ways of visualizing your data (better QC, better hypotheses, better answers → see `http://bioconductor.org/packages/iSEE/` & `http://bioconductor.org/packages/iSEEu/` )

✅ Powerful framework to communicate, interactively and reproducibly

🔥 **Data interpretation**

... just not in plain sight

Show me your data...

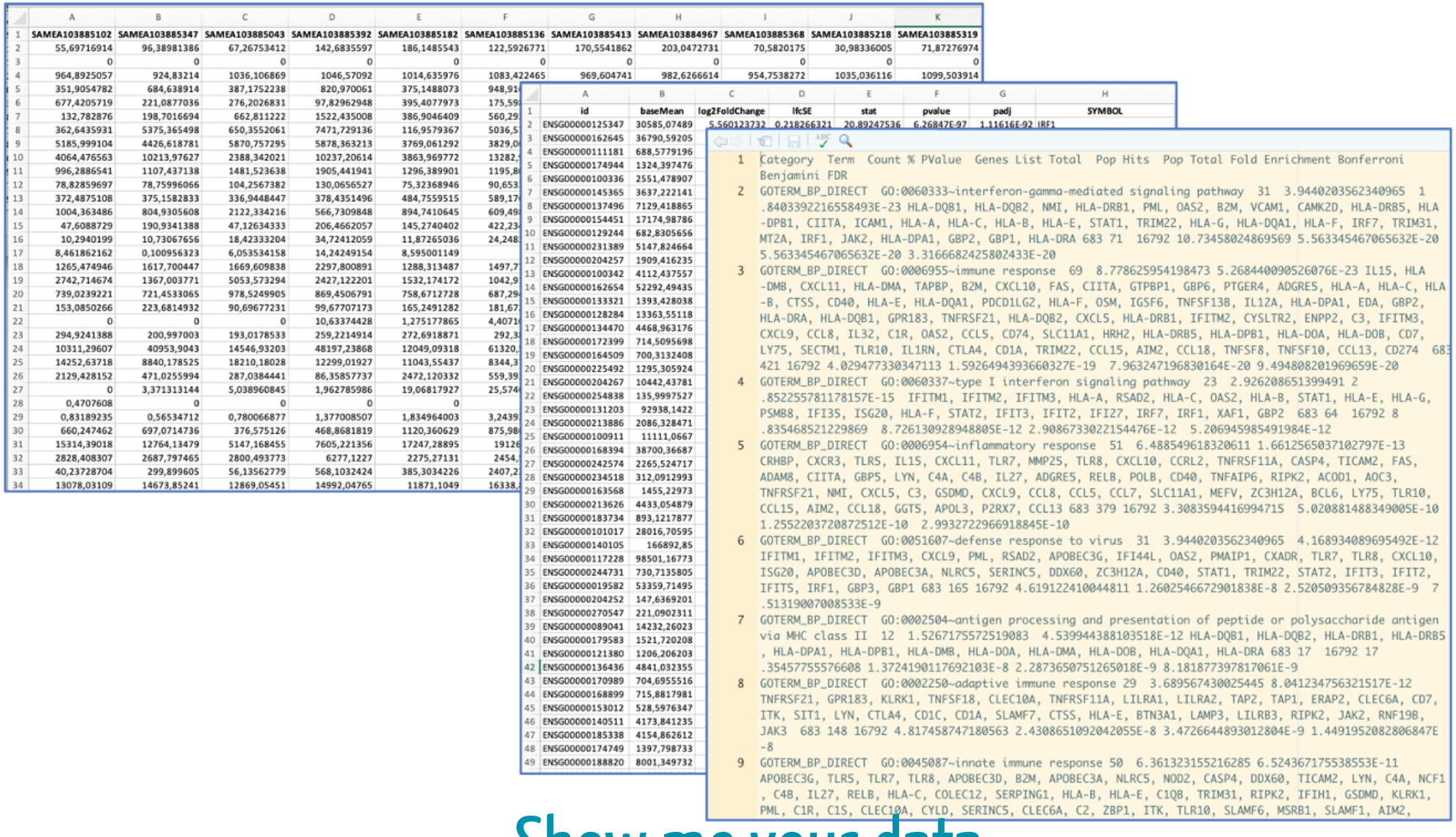
| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAMEA103885102 | SAMEA103885347 | SAMEA103885043 | SAMEA103885392 | SAMEA103885182 | SAMEA103885136 | SAMEA103885413 | SAMEA103884967 | SAMEA103885368 | SAMEA103885218 | SAMEA103885319 |
| 2 | 55,69716914 | 96,38981386 | 67,26753412 | 142,6835597 | 186,1485543 | 122,5926771 | 170,5541862 | 203,0472731 | 70,5820175 | 30,98336005 | 71,87276974 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 964,8925057 | 924,83214 | 1036,106869 | 1046,57092 | 1014,635976 | 1083,422465 | 969,604741 | 982,6266614 | 954,7538272 | 1035,036116 | 1099,503914 |
| 5 | 351,9054782 | 684,638914 | 387,1752238 | 820,970061 | 375,1488073 | 948,9162924 | 503,55028 | 1231,24557 | 300,7527914 | 911,6795373 | 889,3737712 |
| 6 | 677,4205719 | 221,0877036 | 276,2026831 | 97,82962948 | 395,4077973 | 175,5987827 | 179,7898584 | 133,993082 | 675,5371526 | 153,7212284 | 319,7823597 |
| 7 | 132,782876 | 198,7016694 | 662,811222 | 1522,435008 | 386,9046409 | 560,2933294 | 1257,995559 | 5812,826356 | 909,7667115 | 2395,917116 | 2048,353449 |
| 8 | 362,6435931 | 5375,365498 | 650,3552061 | 7471,729136 | 116,9579367 | 5036,571099 | 223,3816218 | 5507,177186 | 66,70469095 | 5451,589642 | 480,5118025 |
| 9 | 5185,999104 | 4426,618781 | 5870,757295 | 5878,363213 | 3769,061292 | 3829,066212 | 2882,333961 | 3365,101977 | 3953,382129 | 3625,84689 | 4932,212822 |
| 10 | 4064,476563 | 10213,97627 | 2388,342021 | 10237,20614 | 3863,969772 | 13282,99842 | 2921,831249 | 11699,4626 | 2120,792061 | 6985,444675 | 2949,988682 |
| 11 | 996,2886541 | 1107,437138 | 1481,523638 | 1905,441941 | 1296,389901 | 1195,809922 | 1705,234892 | 1943,638311 | 981,0211341 | 1280,849815 | 1550,697703 |
| 12 | 78,82859697 | 78,75996066 | 104,2567382 | 130,0656527 | 75,32368946 | 90,65332179 | 69,80311374 | 172,8622743 | 73,59486838 | 47,6298952 | 86,1652492 |
| 13 | 372,4875108 | 375,1582833 | 336,9448447 | 378,4351496 | 484,7559515 | 589,1766985 | 548,4876464 | 415,3602864 | 281,9806048 | 533,7955594 | 265,9931667 |
| 14 | 1004,363486 | 804,9305608 | 2122,334216 | 566,7309848 | 894,7410645 | 609,4980038 | 1607,317126 | 355,0574378 | 1234,02879 | 589,4720116 | 1880,361894 |
| 15 | 47,6088729 | 190,9341388 | 47,12634333 | 206,4662057 | 145,2740402 | 422,2343534 | 150,387129 | 287,7915443 | 53,8085684 | 313,9134375 | 92,61472368 |
| 16 | 10,2940199 | 10,73067656 | 18,42333204 | 34,72412059 | 11,87265036 | 24,24830727 | 21,76451809 | 77,31764486 | 6,892555164 | 4,824180733 | 11,23090998 |
| 17 | 8,461862162 | 0,100956323 | 6,053534158 | 14,24249154 | 8,595001149 | 0 | 0,22640779 | 17,29978818 | 4,094109524 | 5,538925825 | 21,58950506 |
| 18 | 1265,474946 | 1617,700447 | 1669,609838 | 2297,800891 | 1288,313487 | 1497,771426 | 1576,57678 | 2747,84197 | 890,4337923 | 1037,283091 | 1936,845356 |
| 19 | 2742,714674 | 1367,003771 | 5053,573294 | 2427,122201 | 1532,174172 | 1042,918499 | 2445,006435 | 1099,595415 | 1781,289062 | 913,0693244 | 2293,72266 |
| 20 | 739,0239221 | 721,4533065 | 978,5249905 | 869,4506791 | 758,6712728 | 687,2941888 | 958,8405469 | 1131,158723 | 723,1258083 | 891,5201189 | 1172,627644 |
| 21 | 153,0850266 | 223,6814932 | 90,69677231 | 99,67707173 | 165,2491282 | 181,6727726 | 165,8323252 | 124,8435309 | 135,0806311 | 172,4092075 | 154,5518966 |
| 22 | 0 | 0 | 0 | 10,63374428 | 1,275177865 | 4,407102394 | 0,415747599 | 0,615201563 | 4,740536892 | 0 | 2,063830113 |
| 23 | 294,9241388 | 200,997003 | 193,0178533 | 259,2214914 | 272,6918871 | 292,386033 | 180,4896341 | 179,036124 | 382,9993779 | 271,7909482 | 229,4676362 |
| 24 | 10311,29607 | 40953,9043 | 14546,93203 | 48197,23868 | 12049,09318 | 61320,54376 | 15556,91082 | 48401,85979 | 6945,254193 | 62689,66844 | 24693,50787 |
| 25 | 14252,63718 | 8840,178525 | 18210,18028 | 12299,01927 | 11043,55437 | 8344,372672 | 11131,06854 | 8353,885536 | 9092,021277 | 6232,986277 | 11739,0848 |
| 26 | 2129,428152 | 471,0255994 | 287,0384441 | 86,35857737 | 2472,120332 | 559,3936962 | 486,4665709 | 66,4070027 | 5038,903217 | 1219,154589 | 445,1902436 |
| 27 | 0 | 3,371313144 | 5,038960845 | 1,962785986 | 19,06817927 | 25,57466411 | 19,61201231 | 32,07074799 | 2,764897276 | 7,007072669 | 3,240100076 |
| 28 | 0,4707608 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 0,83189235 | 0,56534712 | 0,780066877 | 1,377008507 | 1,834964003 | 3,243938183 | 1,613445267 | 35,93939273 | 21,80390352 | 0,697877051 | 0,497286622 |
| 30 | 660,247462 | 697,0714736 | 376,575126 | 468,8681819 | 1120,360629 | 875,9863134 | 416,0590986 | 737,4089693 | 1126,114691 | 873,0030589 | 707,054169 |
| 31 | 15314,39018 | 12764,13479 | 5147,168455 | 7605,221356 | 17247,28895 | 19126,9456 | 5236,812531 | 11044,68781 | 16448,27488 | 19508,61894 | 5852,803153 |
| 32 | 2828,408307 | 2687,797465 | 2800,493773 | 6277,1227 | 2275,27131 | 2454,51924 | 1339,12723 | 2779,114942 | 2188,442779 | 2400,119315 | 1476,802287 |
| 33 | 40,23728704 | 299,899605 | 56,13562779 | 568,1032424 | 385,3034226 | 2407,236342 | 279,7861222 | 2043,400943 | 69,42871741 | 1174,143223 | 83,05725953 |
| 34 | 13078,03109 | 14673,85241 | 12869,05451 | 14992,04765 | 11871,1049 | 16338,89938 | 10146,36181 | 10380,14905 | 10500,81202 | 14966,33552 | 10032,07621 |

# Show me your data...

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAMEA103885102 | SAMEA103885347 | SAMEA103885043 | SAMEA103885392 | SAMEA103885182 | SAMEA103885136 | SAMEA103885413 | SAMEA103884967 | SAMEA103885368 | SAMEA103885218 | SAMEA103885319 |
| 2 | 55,69716914 | 96,38981386 | 67,26753412 | 142,6835597 | 186,1485543 | 122,5926771 | 170,5541862 | 203,0472731 | 70,5820175 | 30,98336005 | 71,87276974 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 964,8925057 | 924,83214 | 1036,106869 | 1046,57092 | 1014,635976 | 1083,422465 | 969,604741 | 982,6266614 | 954,7538272 | 1035,036116 | 1099,503914 |
| 5 | 351,9054782 | 684,638914 | 387,1752238 | 820,970061 | 375,1488073 | 948,91 | | | | | |
| 6 | 677,4205719 | 221,0877036 | 276,2026831 | 97,82962948 | 395,4077973 | 175,59 | | | | | |
| 7 | 132,782876 | 198,7016694 | 662,811222 | 1522,435008 | 386,9046409 | 560,29 | | | | | |
| 8 | 362,6435931 | 5375,365498 | 650,3552061 | 7471,729136 | 116,9579367 | 5036,5 | | | | | |
| 9 | 5185,999104 | 4426,618781 | 5870,757295 | 5878,363213 | 3769,061292 | 3829,0 | | | | | |
| 10 | 4064,476563 | 10213,97627 | 2388,342021 | 10237,20614 | 3863,969772 | 13282, | | | | | |
| 11 | 996,2886541 | 1107,437138 | 1481,523638 | 1905,441941 | 1296,389901 | 1195,8 | | | | | |
| 12 | 78,82859697 | 78,75996066 | 104,2567382 | 130,0656527 | 75,32368946 | 90,653 | | | | | |
| 13 | 372,4875108 | 375,1582833 | 336,9448447 | 378,4351496 | 484,7559515 | 589,17 | | | | | |
| 14 | 1004,363486 | 804,9305608 | 2122,334216 | 566,7309848 | 894,7410645 | 609,49 | | | | | |
| 15 | 47,6088729 | 190,9341388 | 47,12634333 | 206,4662057 | 145,2740402 | 422,23 | | | | | |
| 16 | 10,2940199 | 10,73067656 | 18,42333204 | 34,72412059 | 11,87265036 | 24,248 | | | | | |
| 17 | 8,461862162 | 0,100956323 | 6,053534158 | 14,24249154 | 8,595001149 | | | | | | |
| 18 | 1265,474946 | 1617,700447 | 1669,609838 | 2297,800891 | 1288,313487 | 1497,7 | | | | | |
| 19 | 2742,714674 | 1367,003771 | 5053,573294 | 2427,122201 | 1532,174172 | 1042,9 | | | | | |
| 20 | 739,0239221 | 721,4533065 | 978,5249905 | 869,4506791 | 758,6712728 | 687,29 | | | | | |
| 21 | 153,0850266 | 223,6814932 | 90,69677231 | 99,67707173 | 165,2491282 | 181,67 | | | | | |
| 22 | 0 | 0 | 0 | 10,63374428 | 1,275177865 | 4,4071 | | | | | |
| 23 | 294,9241388 | 200,997003 | 193,0178533 | 259,2214914 | 272,6918871 | 292,3 | | | | | |
| 24 | 10311,29607 | 40953,9043 | 14546,93203 | 48197,23868 | 12049,09318 | 61320, | | | | | |
| 25 | 14252,63718 | 8840,178525 | 18210,18028 | 12299,01927 | 11043,55437 | 8344,3 | | | | | |
| 26 | 2129,428152 | 471,0255994 | 287,0384441 | 86,35857737 | 2472,120332 | 559,39 | | | | | |
| 27 | 0 | 3,371313144 | 5,038960845 | 1,962785986 | 19,06817927 | 25,574 | | | | | |
| 28 | 0,4707608 | 0 | 0 | 0 | 0 | | | | | | |
| 29 | 0,83189235 | 0,56534712 | 0,780066877 | 1,377008507 | 1,834964003 | 3,2439 | | | | | |
| 30 | 660,247462 | 697,0714736 | 376,575126 | 468,8681819 | 1120,360629 | 875,98 | | | | | |
| 31 | 15314,39018 | 12764,13479 | 5147,168455 | 7605,221356 | 17247,28895 | 19126 | | | | | |
| 32 | 2828,408307 | 2687,797465 | 2800,493773 | 6277,1227 | 2275,27131 | 2454, | | | | | |
| 33 | 40,23728704 | 299,899605 | 56,13562779 | 568,1032424 | 385,3034226 | 2407,2 | | | | | |
| 34 | 13078,03109 | 14673,85241 | 12869,05451 | 14992,04765 | 11871,1049 | 16338, | | | | | |

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | id | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj | SYMBOL |
| 2 | ENSG00000125347 | 30585,07489 | 5,560123732 | 0,218266321 | 20,89247536 | 6,26847E-97 | 1,11616E-92 | IRF1 |
| 3 | ENSG00000162645 | 36790,59205 | 6,665803772 | 0,286573165 | 19,77088043 | 5,30422E-87 | 4,72235E-83 | GBP2 |
| 4 | ENSG00000111181 | 688,5779196 | 4,710591861 | 0,196102275 | 18,92171759 | 7,55518E-80 | 4,48425E-76 | SLC6A12 |
| 5 | ENSG00000174944 | 1324,397476 | 9,863903776 | 0,471512737 | 18,7988639 | 7,71512E-79 | 3,43439E-75 | P2RY14 |
| 6 | ENSG00000100336 | 2551,478907 | 8,708550505 | 0,421409592 | 18,29229959 | 9,53083E-75 | 3,39412E-71 | APOL4 |
| 7 | ENSG00000145365 | 3637,222141 | 5,193447761 | 0,236854233 | 17,70476176 | 3,85308E-70 | 1,14346E-66 | TIFA |
| 8 | ENSG00000137496 | 7129,418865 | 4,05866209 | 0,177364585 | 17,24505541 | 1,21896E-66 | 3,10068E-63 | IL18BP |
| 9 | ENSG00000154451 | 17174,98786 | 10,37647789 | 0,551586848 | 16,9990962 | 8,33958E-65 | 1,85618E-61 | GBP5 |
| 10 | ENSG00000129244 | 682,8305656 | 7,191566714 | 0,381506097 | 16,2292733 | 3,13145E-59 | 6,19539E-56 | ATP1B2 |
| 11 | ENSG00000231389 | 5147,824664 | 5,593758623 | 0,28477475 | 16,13120061 | 1,53997E-58 | 2,74208E-55 | HLA-DPA1 |
| 12 | ENSG00000204257 | 1909,416235 | 4,071574682 | 0,190921436 | 16,08816042 | 3,08879E-58 | 4,99991E-55 | HLA-DMA |
| 13 | ENSG00000100342 | 4112,437557 | 7,13529667 | 0,39590635 | 15,49683826 | 3,64388E-54 | 5,40691E-51 | APOL1 |
| 14 | ENSG00000162654 | 52292,49435 | 9,752769157 | 0,569217591 | 15,37684234 | 2,34089E-53 | 3,2063E-50 | GBP4 |
| 15 | ENSG00000133321 | 1393,428038 | 7,448819999 | 0,419830443 | 15,36053451 | 3,01082E-53 | 3,82933E-50 | PLAAT4 |
| 16 | ENSG00000128284 | 13363,55118 | 7,977215573 | 0,454494945 | 15,35158013 | 3,45663E-53 | 4,10325E-50 | APOL3 |
| 17 | ENSG00000134470 | 4468,963176 | 4,300572585 | 0,220226433 | 14,98717724 | 8,90586E-51 | 9,91111E-48 | IL15RA |
| 18 | ENSG00000172399 | 714,5095698 | -7,829885987 | 0,463112753 | -14,74778213 | 3,17943E-49 | 3,33017E-46 | MYOZ2 |
| 19 | ENSG00000164509 | 700,3132408 | 8,395184784 | 0,536171751 | 13,79256698 | 2,82525E-43 | 2,7948E-40 | IL31RA |
| 20 | ENSG00000225492 | 1295,305924 | 6,141301184 | 0,374065413 | 13,74439071 | 5,50344E-43 | 5,15759E-40 | GBP1P1 |
| 21 | ENSG00000204267 | 10442,43781 | 3,458265659 | 0,179235202 | 13,71530609 | 8,22187E-43 | 7,31993E-40 | TAP2 |
| 22 | ENSG00000254838 | 135,9997527 | 5,591075237 | 0,335403469 | 13,68821633 | 1,19404E-42 | 1,01244E-39 | GVINP1 |
| 23 | ENSG00000131203 | 92938,1422 | 9,545152592 | 0,644384499 | 13,26095306 | 3,89909E-40 | 3,15578E-37 | IDO1 |
| 24 | ENSG00000213886 | 2086,328471 | 10,26938285 | 0,703626324 | 13,17372947 | 1,24302E-39 | 9,62317E-37 | UBD |
| 25 | ENSG00000100911 | 11111,0667 | 3,358971492 | 0,179410258 | 13,14847609 | 1,73639E-39 | 1,28825E-36 | PSME2 |
| 26 | ENSG00000168394 | 38700,36687 | 5,356767076 | 0,337264864 | 12,91793941 | 3,56563E-38 | 2,53958E-35 | TAP1 |
| 27 | ENSG00000242574 | 2265,524717 | 3,819583423 | 0,219870422 | 12,82384142 | 1,20575E-37 | 8,25756E-35 | HLA-DMB |
| 28 | ENSG00000234518 | 312,0912993 | 8,276859324 | 0,576720452 | 12,61765436 | 1,68774E-36 | 1,11303E-33 | |
| 29 | ENSG00000163568 | 1455,22973 | 7,804253083 | 0,543387364 | 12,52191997 | 5,6648E-36 | 3,60241E-33 | AIM2 |
| 30 | ENSG00000213626 | 4433,054879 | -4,42792907 | 0,275004543 | -12,46499068 | 1,15885E-35 | 7,11536E-33 | LBH |
| 31 | ENSG00000183734 | 893,1217877 | 4,237240545 | 0,260041233 | 12,4489509 | 1,41695E-35 | 8,4101E-33 | ASCL2 |
| 32 | ENSG00000101017 | 28016,70595 | 3,478496822 | 0,200633386 | 12,3533619 | 4,67175E-35 | 2,6834E-32 | CD40 |
| 33 | ENSG00000140105 | 166892,85 | 5,82727503 | 0,391766026 | 12,32183169 | 6,91066E-35 | 3,84535E-32 | WARS1 |
| 34 | ENSG00000117228 | 98501,16773 | 7,686166048 | 0,552174564 | 12,10879039 | 9,48831E-34 | 5,11966E-31 | GBP1 |
| 35 | ENSG00000244731 | 730,7135805 | 5,027915974 | 0,334446879 | 12,04351489 | 2,09821E-33 | 1,09885E-30 | C4A |
| 36 | ENSG00000019582 | 53359,71495 | 4,15277183 | 0,264525014 | 11,91861512 | 9,46691E-33 | 4,81622E-30 | CD74 |
| 37 | ENSG00000204252 | 147,6369201 | 5,782657331 | 0,402965448 | 11,86865362 | 1,72217E-32 | 8,51804E-30 | HLA-DOA |
| 38 | ENSG00000270547 | 221,0902311 | -5,370582108 | 0,370389876 | -11,79995024 | 3,90542E-32 | 1,87946E-29 | LINC01235 |
| 39 | ENSG00000089041 | 14232,26023 | 4,656176735 | 0,316525516 | 11,55096999 | 7,29919E-31 | 3,42025E-28 | P2RX7 |
| 40 | ENSG00000179583 | 1521,720208 | 6,112248838 | 0,449889179 | 11,36335141 | 6,36549E-30 | 2,90625E-27 | CIITA |
| 41 | ENSG00000121380 | 1206,206203 | 5,996911594 | 0,439943587 | 11,35807349 | 6,76193E-30 | 3,01008E-27 | BCL2L14 |
| 42 | ENSG00000136436 | 4841,032355 | 2,828125249 | 0,161138537 | 11,34505304 | 7,84786E-30 | 3,40827E-27 | CALCOCO2 |
| 43 | ENSG00000170989 | 704,6955516 | -5,365857478 | 0,386740233 | -11,28886293 | 1,48951E-29 | 6,31482E-27 | S1PR1 |
| 44 | ENSG00000168899 | 715,8817981 | 4,743931938 | 0,333167369 | 11,23739082 | 2,67162E-29 | 1,1063E-26 | VAMP5 |
| 45 | ENSG00000153012 | 528,5976347 | -6,370842542 | 0,483211518 | -11,11488931 | 1,06182E-28 | 4,297E-26 | LGI2 |
| 46 | ENSG00000140511 | 4173,841235 | 7,091736603 | 0,55079617 | 11,05987466 | 1,96372E-28 | 7,60129E-26 | HAPLN3 |
| 47 | ENSG00000185338 | 4154,862612 | 7,689753422 | 0,604771596 | 11,06161974 | 1,92588E-28 | 7,60129E-26 | SOCS1 |
| 48 | ENSG00000174749 | 1397,798733 | 2,925585998 | 0,174477721 | 11,03628579 | 2,55372E-28 | 9,67481E-26 | FAM241A |
| 49 | ENSG00000188820 | 8001,349732 | 5,681784849 | 0,426951103 | 10,96562302 | 5,59141E-28 | 2,07418E-25 | CALHM6 |

# Show me your data...



| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAMEA103885102 | SAMEA103885347 | SAMEA103885043 | SAMEA103885392 | SAMEA103885182 | SAMEA103885136 | SAMEA103885413 | SAMEA103884967 | SAMEA103885368 | SAMEA103885218 | SAMEA103885319 |
| 2 | 55,69716914 | 96,38981386 | 67,26753412 | 142,6835597 | 186,1485543 | 122,5926771 | 170,5541862 | 203,0472731 | 70,5820175 | 30,98336005 | 71,87276974 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 964,8925057 | 924,83214 | 1036,106869 | 1046,57092 | 1014,635976 | 1083,422465 | 969,604741 | 982,6266614 | 954,7538272 | 1035,036116 | 1099,503914 |
| 5 | 351,9054782 | 684,638914 | 387,1752238 | 820,970061 | 375,1488073 | 948,91 | | | | | |
| 6 | 677,4205719 | 221,0877036 | 276,2026831 | 97,82962948 | 395,4077973 | 175,59 | | | | | |
| 7 | 132,782876 | 198,7016694 | 662,811222 | 1522,435008 | 386,9046409 | 560,29 | | | | | |
| 8 | 362,6435931 | 5375,365498 | 650,3552061 | 7471,729136 | 116,9579367 | 5036,5 | | | | | |
| 9 | 5185,999104 | 4426,618781 | 5870,757295 | 5878,363213 | 3769,061292 | 3829,0 | | | | | |
| 10 | 4064,476563 | 10213,97627 | 2388,342021 | 10237,20614 | 3863,969772 | 13282, | | | | | |
| 11 | 996,2886541 | 1107,437138 | 1481,523638 | 1905,441941 | 1296,389901 | 1195,8 | | | | | |
| 12 | 78,82859697 | 78,75996066 | 104,2567382 | 130,0656527 | 75,32368946 | 90,653 | | | | | |
| 13 | 372,4875108 | 375,1582833 | 336,9448447 | 378,4351496 | 484,7559515 | 589,17 | | | | | |
| 14 | 1004,363486 | 804,9305608 | 2122,334216 | 566,7309848 | 894,7410645 | 609,49 | | | | | |
| 15 | 47,6088729 | 190,9341388 | 47,12634333 | 206,4662057 | 145,2740402 | 422,23 | | | | | |
| 16 | 10,2940199 | 10,73067656 | 18,42333204 | 34,72412059 | 11,87265036 | 24,248 | | | | | |
| 17 | 8,461862162 | 0,100956323 | 6,053534158 | 14,24249154 | 8,595001149 | | | | | | |
| 18 | 1265,474946 | 1617,700447 | 1669,609838 | 2297,800891 | 1288,313487 | 1497,7 | | | | | |
| 19 | 2742,714674 | 1367,003771 | 5053,573294 | 2427,122201 | 1532,174172 | 1042,9 | | | | | |
| 20 | 739,0239221 | 721,4533065 | 978,5249905 | 869,4506791 | 758,6712728 | 687,29 | | | | | |
| 21 | 153,0850266 | 223,6814932 | 90,69677231 | 99,67707173 | 165,2491282 | 181,67 | | | | | |
| 22 | 0 | 0 | 0 | 10,63374428 | 1,275177865 | 4,4071 | | | | | |
| 23 | 294,9241388 | 200,997003 | 193,0178533 | 259,2214914 | 272,6918871 | 292,3 | | | | | |
| 24 | 10311,29607 | 40953,9043 | 14546,93203 | 48197,23868 | 12049,09318 | 61320, | | | | | |
| 25 | 14252,63718 | 8840,178525 | 18210,18028 | 12299,01927 | 11043,55437 | 8344,3 | | | | | |
| 26 | 2129,428152 | 471,0255994 | 287,0384441 | 86,35857737 | 2472,120332 | 559,39 | | | | | |
| 27 | 0 | 3,371313144 | 5,038960845 | 1,962785986 | 19,06817927 | 25,574 | | | | | |
| 28 | 0,4707608 | 0 | 0 | 0 | 0 | | | | | | |
| 29 | 0,83189235 | 0,56534712 | 0,780066877 | 1,377008507 | 1,834964003 | 3,2439 | | | | | |
| 30 | 660,247462 | 697,0714736 | 376,575126 | 468,8681819 | 1120,360629 | 875,98 | | | | | |
| 31 | 15314,39018 | 12764,13479 | 5147,168455 | 7605,221356 | 17247,28895 | 19126 | | | | | |
| 32 | 2828,408307 | 2687,797465 | 2800,493773 | 6277,1227 | 2275,27131 | 2454, | | | | | |
| 33 | 40,23728704 | 299,899605 | 56,13562779 | 568,1032424 | 385,3034226 | 2407,2 | | | | | |
| 34 | 13078,03109 | 14673,85241 | 12869,05451 | 14992,04765 | 11871,1049 | 16338, | | | | | |

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | id | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj | SYMBOL |
| 2 | ENSG00000125347 | 30585,07489 | 5,560123732 | 0,218266321 | 20,89247536 | 6,26847E-97 | 1,11616E-92 | IRF1 |
| 3 | ENSG00000162645 | 36790,59205 | | | | | | |
| 4 | ENSG00000111181 | 688,5779196 | | | | | | |
| 5 | ENSG00000174944 | 1324,397476 | | | | | | |
| 6 | ENSG00000100336 | 2551,478907 | | | | | | |
| 7 | ENSG00000145365 | 3637,222141 | | | | | | |
| 8 | ENSG00000137496 | 7129,418865 | | | | | | |
| 9 | ENSG00000154451 | 17174,98786 | | | | | | |
| 10 | ENSG00000129244 | 682,8305656 | | | | | | |
| 11 | ENSG00000231389 | 5147,824664 | | | | | | |
| 12 | ENSG00000204257 | 1909,416235 | | | | | | |
| 13 | ENSG00000100342 | 4112,437557 | | | | | | |
| 14 | ENSG00000162654 | 52292,49435 | | | | | | |
| 15 | ENSG00000133321 | 1393,428038 | | | | | | |
| 16 | ENSG00000128284 | 13363,55118 | | | | | | |
| 17 | ENSG00000134470 | 4468,963176 | | | | | | |
| 18 | ENSG00000172399 | 714,5095698 | | | | | | |
| 19 | ENSG00000164509 | 700,3132408 | | | | | | |
| 20 | ENSG00000225492 | 1295,305924 | | | | | | |
| 21 | ENSG00000204267 | 10442,43781 | | | | | | |
| 22 | ENSG00000254838 | 135,9997527 | | | | | | |
| 23 | ENSG00000131203 | 92938,1422 | | | | | | |
| 24 | ENSG00000213886 | 2086,328471 | | | | | | |
| 25 | ENSG00000100911 | 11111,0667 | | | | | | |
| 26 | ENSG00000168394 | 38700,36687 | | | | | | |
| 27 | ENSG00000242574 | 2265,524717 | | | | | | |
| 28 | ENSG00000234518 | 312,0912993 | | | | | | |
| 29 | ENSG00000163568 | 1455,22973 | | | | | | |
| 30 | ENSG00000213626 | 4433,054879 | | | | | | |
| 31 | ENSG00000183734 | 893,1217877 | | | | | | |
| 32 | ENSG00000101017 | 28016,70595 | | | | | | |
| 33 | ENSG00000140105 | 166892,85 | | | | | | |
| 34 | ENSG00000117228 | 98501,16773 | | | | | | |
| 35 | ENSG00000244731 | 730,7135805 | | | | | | |
| 36 | ENSG00000019582 | 53359,71495 | | | | | | |
| 37 | ENSG00000204252 | 147,6369201 | | | | | | |
| 38 | ENSG00000270547 | 221,0902311 | | | | | | |
| 39 | ENSG00000089041 | 14232,26023 | | | | | | |
| 40 | ENSG00000179583 | 1521,720208 | | | | | | |
| 41 | ENSG00000121380 | 1206,206203 | | | | | | |
| 42 | ENSG00000136436 | 4841,032355 | | | | | | |
| 43 | ENSG00000170989 | 704,6955516 | | | | | | |
| 44 | ENSG00000168899 | 715,8817981 | | | | | | |
| 45 | ENSG00000153012 | 528,5976347 | | | | | | |
| 46 | ENSG00000140511 | 4173,841235 | | | | | | |
| 47 | ENSG00000185338 | 4154,862612 | | | | | | |
| 48 | ENSG00000174749 | 1397,798733 | | | | | | |
| 49 | ENSG00000188820 | 8001,349732 | | | | | | |

1 Category Term Count % PValue Genes List Total Pop Hits Pop Total Fold Enrichment Bonferroni Benjamini FDR

2 GOTERM_BP_DIRECT GO:0060333~interferon-gamma-mediated signaling pathway 31 3.9440203562340965 1 .84033922165584 93E-23 HLA-DQB1, HLA-DQB2, NMI, HLA-DRB1, PML, OAS2, B2M, VCAM1, CAMK2D, HLA-DRB5, HLA -DPB1, CIITA, ICAM1, HLA-A, HLA-C, HLA-B, HLA-E, STAT1, TRIM22, HLA-G, HLA-DQA1, HLA-F, IRF7, TRIM31, MT2A, IRF1, JAK2, HLA-DPA1, GBP2, GBP1, HLA-DRA 683 71 16792 10.73458024869569 5.563345467065632E-20 5.563345467065632E-20 3.3166682425802433E-20

3 GOTERM_BP_DIRECT GO:0006955~immune response 69 8.778625954198473 5.268440090526076E-23 IL15, HLA -DMB, CXCL11, HLA-DMA, TAPBP, B2M, CXCL10, FAS, CIITA, GTPBP1, GBP6, PTGER4, ADGRE5, HLA-A, HLA-C, HLA -B, CTSS, CD40, HLA-E, HLA-DQA1, PDCD1LG2, HLA-F, OSM, IGSF6, TNFSF13B, IL12A, HLA-DPA1, EDA, GBP2, HLA-DRA, HLA-DQB1, GPR183, TNFRSF21, HLA-DQB2, CXCL5, HLA-DRB1, IFITM2, CYSLTR2, ENPP2, C3, IFITM3, CXCL9, CCL8, IL32, C1R, OAS2, CCL5, CD74, SLC11A1, HRH2, HLA-DRB5, HLA-DPB1, HLA-DOA, HLA-DOB, CD7, LY75, SECTM1, TLR10, IL1RN, CTLA4, CD1A, TRIM22, CCL15, AIM2, CCL18, TNFSF8, TNFSF10, CCL13, CD274 683 421 16792 4.029477330347113 1.5926494393660327E-19 7.963247196830164E-20 9.494808201969659E-20

4 GOTERM_BP_DIRECT GO:0060337~type I interferon signaling pathway 23 2.926208651399491 2 .852255781178157E-15 IFITM1, IFITM2, IFITM3, HLA-A, RSAD2, HLA-C, OAS2, HLA-B, STAT1, HLA-E, HLA-G, PSMB8, IFI35, ISG20, HLA-F, STAT2, IFIT3, IFIT2, IFI27, IRF7, IRF1, XAF1, GBP2 683 64 16792 8 .835468521229869 8.726130928948805E-12 2.9086733022154476E-12 5.206945985491984E-12

5 GOTERM_BP_DIRECT GO:0006954~inflammatory response 51 6.488549618320611 1.6612565037102797E-13 CRHBP, CXCR3, TLR5, IL15, CXCL11, TLR7, MMP25, TLR8, CXCL10, CCRL2, TNFRSF11A, CASP4, TICAM2, FAS, ADAM8, CIITA, GBP5, LYN, C4A, C4B, IL27, ADGRE5, RELB, POLB, CD40, TNFAIP6, RIPK2, ACOD1, AOC3, TNFRSF21, NMI, CXCL5, C3, GSDMD, CXCL9, CCL8, CCL5, CCL7, SLC11A1, MEFV, ZC3H12A, BCL6, LY75, TLR10, CCL15, AIM2, CCL18, GGT5, APOL3, P2RX7, CCL13 683 379 16792 3.3083594416994715 5.020881488349005E-10 1.2552203720872512E-10 2.9932722966918845E-10

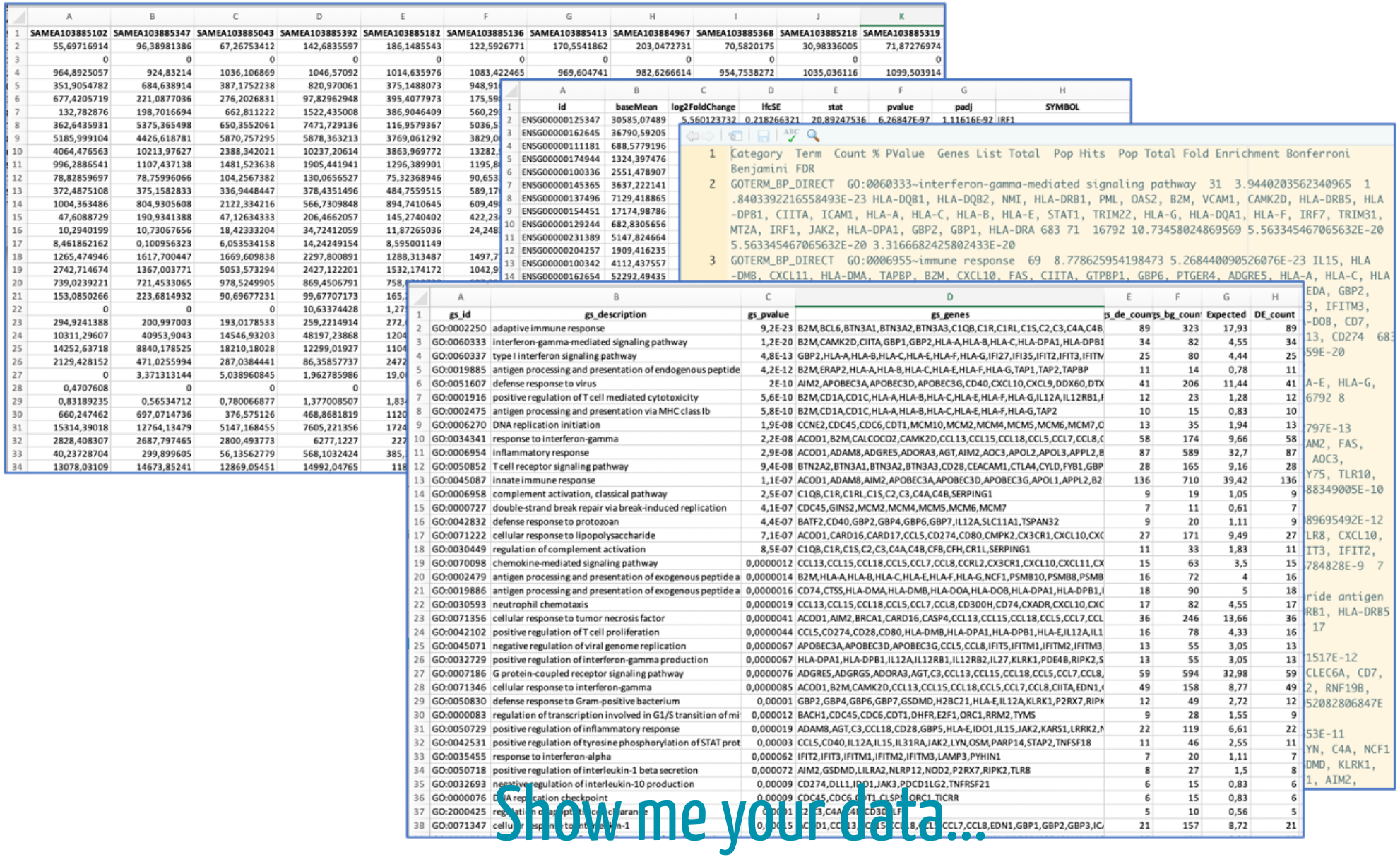
6 GOTERM_BP_DIRECT GO:0051607~defense response to virus 31 3.9440203562340965 4.168934089695492E-12 IFITM1, IFITM2, IFITM3, CXCL9, PML, RSAD2, APOBEC3G, IFI44L, OAS2, PMAIP1, CXADR, TLR7, TLR8, CXCL10, ISG20, APOBEC3D, APOBEC3A, NLRC5, SERINC5, DDX60, ZC3H12A, CD40, STAT1, TRIM22, STAT2, IFIT3, IFIT2, IFIT5, IRF1, GBP3, GBP1 683 165 16792 4.619122410044811 1.260254667290183 8E-8 2.520509356784828E-9 7 .51319007008533E-9

7 GOTERM_BP_DIRECT GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II 12 1.5267175572519083 4.539944388103518E-12 HLA-DQB1, HLA-DQB2, HLA-DRB1, HLA-DRB5 , HLA-DPA1, HLA-DPB1, HLA-DMB, HLA-DOA, HLA-DMA, HLA-DOB, HLA-DQA1, HLA-DRA 683 17 16792 17 .35457755576608 1.3724190117692103E-8 2.2873650751265018E-9 8.181877397817061E-9

8 GOTERM_BP_DIRECT GO:0002250~adaptive immune response 29 3.689567430025445 8.041234756321517E-12 TNFRSF21, GPR183, KLRK1, TNFSF18, CLEC10A, TNFRSF11A, LILRA1, LILRA2, TAP2, TAP1, ERAP2, CLEC6A, CD7, ITK, SIT1, LYN, CTLA4, CD1C, CD1A, SLAMF7, CTSS, HLA-E, BTN3A1, LAMP3, LILRB3, RIPK2, JAK2, RNF19B, JAK3 683 148 16792 4.817458747180563 2.4308651092042055E-8 3.4726644893012804E-9 1.4491952082806847E -8

9 GOTERM_BP_DIRECT GO:0045087~innate immune response 50 6.361323155216285 6.524367175538553E-11 APOBEC3G, TLR5, TLR7, TLR8, APOBEC3D, B2M, APOBEC3A, NLRC5, NOD2, CASP4, DDX60, TICAM2, LYN, C4A, NCF1 , C4B, IL27, RELB, HLA-C, COLEC12, SERPING1, HLA-B, HLA-E, C1QB, TRIM31, RIPK2, IFIH1, GSDMD, KLRK1, PML, C1R, C1S, CLEC10A, CYLD, SERINC5, CLEC6A, C2, ZBP1, ITK, TLR10, SLAMF6, MSRB1, SLAMF1, AIM2,

# Show me your data...



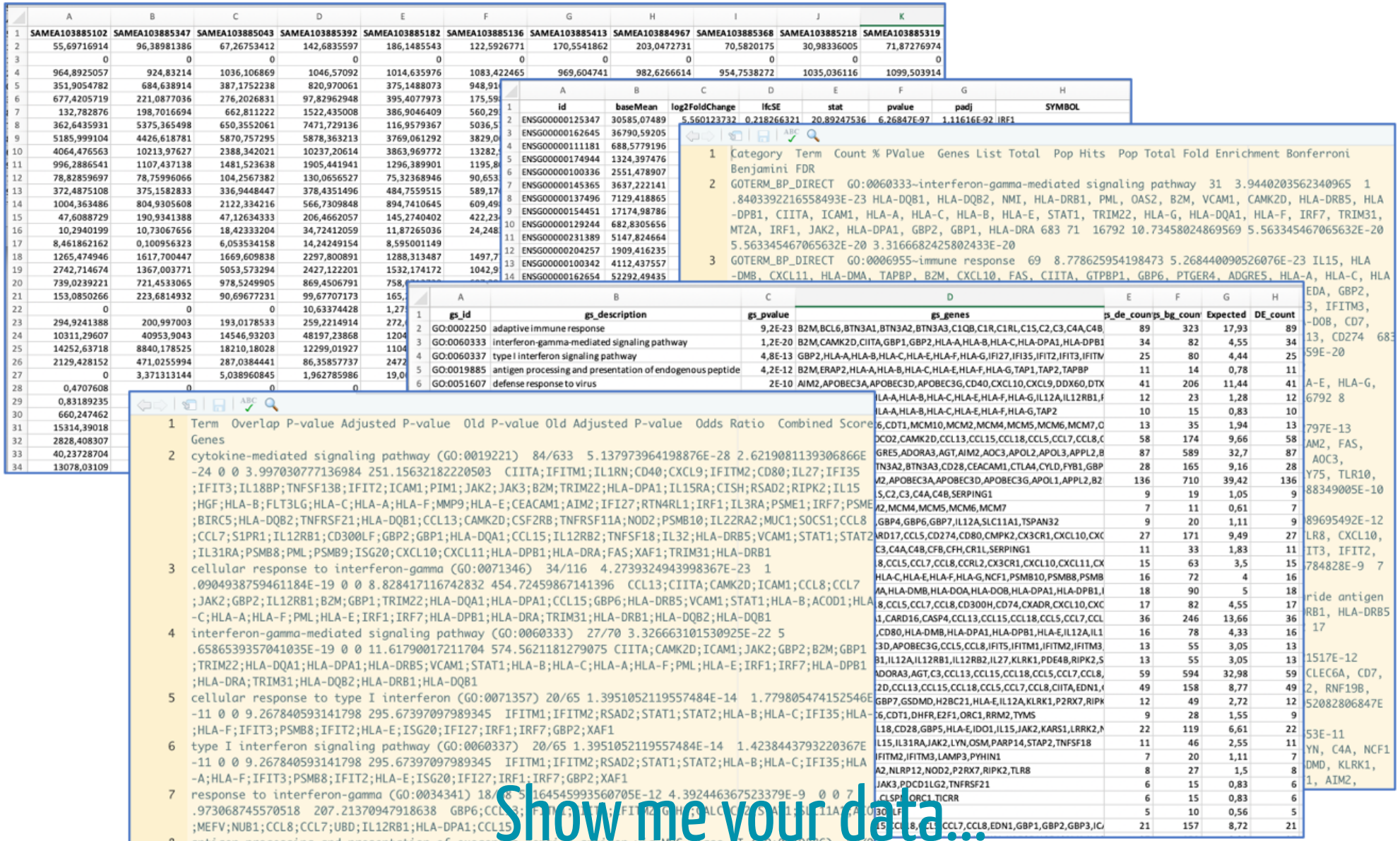
| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAMEA103885102 | SAMEA103885347 | SAMEA103885043 | SAMEA103885392 | SAMEA103885182 | SAMEA103885136 | SAMEA103885413 | SAMEA103884967 | SAMEA103885368 | SAMEA103885218 | SAMEA103885319 |
| 2 | 55,69716914 | 96,38981386 | 67,26753412 | 142,6835597 | 186,1485543 | 122,5926771 | 170,5541862 | 203,0472731 | 70,5820175 | 30,98336005 | 71,87276974 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 964,8925057 | 924,83214 | 1036,106869 | 1046,57092 | 1014,635976 | 1083,422465 | 969,604741 | 982,6266614 | 954,7538272 | 1035,036116 | 1099,503914 |
| 5 | 351,9054782 | 684,638914 | 387,1752238 | 820,970061 | 375,1488073 | 948,91 | | | | | |
| 6 | 677,4205719 | 221,0877036 | 276,2026831 | 97,82962948 | 395,4077973 | 175,59 | | | | | |
| 7 | 132,782876 | 198,7016694 | 662,811222 | 1522,435008 | 386,9046409 | 560,29 | | | | | |
| 8 | 362,6435931 | 5375,365498 | 650,3552061 | 7471,729136 | 116,9579367 | 5036,5 | | | | | |
| 9 | 5185,999104 | 4426,618781 | 5870,757295 | 5878,363213 | 3769,061292 | 3829,0 | | | | | |
| 10 | 4064,476563 | 10213,97627 | 2388,342021 | 10237,20614 | 3863,969772 | 13282, | | | | | |
| 11 | 996,2886541 | 1107,437138 | 1481,523638 | 1905,441941 | 1296,389901 | 1195,8 | | | | | |
| 12 | 78,82859697 | 78,75996066 | 104,2567382 | 130,0656527 | 75,32368946 | 90,653 | | | | | |
| 13 | 372,4875108 | 375,1582833 | 336,9448447 | 378,4351496 | 484,7559515 | 589,17 | | | | | |
| 14 | 1004,363486 | 804,9305608 | 2122,334216 | 566,7309848 | 894,7410645 | 609,49 | | | | | |
| 15 | 47,6088729 | 190,9341388 | 47,12634333 | 206,4662057 | 145,2740402 | 422,23 | | | | | |
| 16 | 10,2940199 | 10,73067656 | 18,42333204 | 34,72412059 | 11,87265036 | 24,248 | | | | | |
| 17 | 8,461862162 | 0,100956323 | 6,053534158 | 14,24249154 | 8,595001149 | | | | | | |
| 18 | 1265,474946 | 1617,700447 | 1669,609838 | 2297,800891 | 1288,313487 | 1497,7 | | | | | |
| 19 | 2742,714674 | 1367,003771 | 5053,573294 | 2427,122201 | 1532,174172 | 1042,9 | | | | | |
| 20 | 739,0239221 | 721,4533065 | 978,5249905 | 869,4506791 | 758, | | | | | | |
| 21 | 153,0850266 | 223,6814932 | 90,69677231 | 99,67707173 | 165, | | | | | | |
| 22 | 0 | 0 | 0 | 10,63374428 | 1,27 | | | | | | |
| 23 | 294,9241388 | 200,997003 | 193,0178533 | 259,2214914 | 272, | | | | | | |
| 24 | 10311,29607 | 40953,9043 | 14546,93203 | 48197,23868 | 1204 | | | | | | |
| 25 | 14252,63718 | 8840,178525 | 18210,18028 | 12299,01927 | 1104 | | | | | | |
| 26 | 2129,428152 | 471,0255994 | 287,0384441 | 86,35857737 | 2472 | | | | | | |
| 27 | 0 | 3,371313144 | 5,038960845 | 1,962785986 | 19,0 | | | | | | |
| 28 | 0,4707608 | 0 | 0 | 0 | | | | | | | |
| 29 | 0,83189235 | 0,56534712 | 0,780066877 | 1,377008507 | 1,83 | | | | | | |
| 30 | 660,247462 | 697,0714736 | 376,575126 | 468,8681819 | 1120 | | | | | | |
| 31 | 15314,39018 | 12764,13479 | 5147,168455 | 7605,221356 | 1724 | | | | | | |
| 32 | 2828,408307 | 2687,797465 | 2800,493773 | 6277,1227 | 227 | | | | | | |
| 33 | 40,23728704 | 299,899605 | 56,13562779 | 568,1032424 | 385, | | | | | | |
| 34 | 13078,03109 | 14673,85241 | 12869,05451 | 14992,04765 | 118 | | | | | | |

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | id | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj | SYMBOL |
| 2 | ENSG00000125347 | 30585,07489 | 5,560123732 | 0,218266321 | 20,89247536 | 6,26847E-97 | 1,11616E-92 | IRF1 |
| 3 | ENSG00000162645 | 36790,59205 | | | | | | |
| 4 | ENSG00000111181 | 688,5779196 | | | | | | |
| 5 | ENSG00000174944 | 1324,397476 | | | | | | |
| 6 | ENSG00000100336 | 2551,478907 | | | | | | |
| 7 | ENSG00000145365 | 3637,222141 | | | | | | |
| 8 | ENSG00000137496 | 7129,418865 | | | | | | |
| 9 | ENSG00000154451 | 17174,98786 | | | | | | |
| 10 | ENSG00000129244 | 682,8305656 | | | | | | |
| 11 | ENSG00000231389 | 5147,824664 | | | | | | |
| 12 | ENSG00000204257 | 1909,416235 | | | | | | |
| 13 | ENSG00000100342 | 4112,437557 | | | | | | |
| 14 | ENSG00000162654 | 52292,49435 | | | | | | |

```
Category  Term  Count % PValue  Genes List Total  Pop Hits  Pop Total Fold Enrichment Bonferroni
Benjamini FDR
GOTERM_BP_DIRECT  GO:0060333~interferon-gamma-mediated signaling pathway  31  3.9440203562340965  1
.84033922165584 93E-23 HLA-DQB1, HLA-DQB2, NMI, HLA-DRB1, PML, OAS2, B2M, VCAM1, CAMK2D, HLA-DRB5, HLA
-DPB1, CIITA, ICAM1, HLA-A, HLA-C, HLA-B, HLA-E, STAT1, TRIM22, HLA-G, HLA-DQA1, HLA-F, IRF7, TRIM31,
MT2A, IRF1, JAK2, HLA-DPA1, GBP2, GBP1, HLA-DRA 683 71  16792 10.73458024869569 5.563345467065632E-20
5.563345467065632E-20 3.3166682425802433E-20
GOTERM_BP_DIRECT  GO:0006955~immune response  69  8.778625954198473 5.268440090526076E-23 IL15, HLA
-DMB, CXCL11, HLA-DMA, TAPBP, B2M, CXCL10, FAS, CIITA, GTPBP1, GBP6, PTGER4, ADGRE5, HLA-A, HLA-C, HLA
                                                                                                EDA, GBP2,
                                                                                                3, IFITM3,
                                                                                                -DOB, CD7,
                                                                                                13, CD274  683
                                                                                                59E-20

                                                                                                A-E, HLA-G,
                                                                                                6792 8

                                                                                                797E-13
                                                                                                AM2, FAS,
                                                                                                 AOC3,
                                                                                                Y75, TLR10,
                                                                                                8834900 5E-10

                                                                                                89695492E-12
                                                                                                LR8, CXCL10,
                                                                                                IT3, IFIT2,
                                                                                                784828E-9  7

                                                                                                ride antigen
                                                                                                RB1, HLA-DRB5
                                                                                                 17

                                                                                                1517E-12
                                                                                                CLEC6A, CD7,
                                                                                                2, RNF19B,
                                                                                                52082806847E

                                                                                                53E-11
                                                                                                YN, C4A, NCF1
                                                                                                DMD, KLRK1,
                                                                                                1, AIM2,
```

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | gs_id | gs_description | gs_pvalue | gs_genes | s_de_coun | s_bg_coun | Expected | DE_count |
| 2 | GO:0002250 | adaptive immune response | 9,2E-23 | B2M,BCL6,BTN3A1,BTN3A2,BTN3A3,C1QB,C1R,C1RL,C1S,C2,C3,C4A,C4B | 89 | 323 | 17,93 | 89 |
| 3 | GO:0060333 | interferon-gamma-mediated signaling pathway | 1,2E-20 | B2M,CAMK2D,CIITA,GBP1,GBP2,HLA-A,HLA-B,HLA-C,HLA-DPA1,HLA-DPB1 | 34 | 82 | 4,55 | 34 |
| 4 | GO:0060337 | type I interferon signaling pathway | 4,8E-13 | GBP2,HLA-A,HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,IFI27,IFI35,IFIT2,IFIT3,IFITM | 25 | 80 | 4,44 | 25 |
| 5 | GO:0019885 | antigen processing and presentation of endogenous peptide | 4,2E-12 | B2M,ERAP2,HLA-A,HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,TAP1,TAP2,TAPBP | 11 | 14 | 0,78 | 11 |
| 6 | GO:0051607 | defense response to virus | 2E-10 | AIM2,APOBEC3A,APOBEC3D,APOBEC3G,CD40,CXCL10,CXCL9,DDX60,DTX | 41 | 206 | 11,44 | 41 |
| 7 | GO:0001916 | positive regulation of T cell mediated cytotoxicity | 5,6E-10 | B2M,CD1A,CD1C,HLA-A,HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,IL12A,IL12RB1,I | 12 | 23 | 1,28 | 12 |
| 8 | GO:0002475 | antigen processing and presentation via MHC class Ib | 5,8E-10 | B2M,CD1A,CD1C,HLA-A,HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,TAP2 | 10 | 15 | 0,83 | 10 |
| 9 | GO:0006270 | DNA replication initiation | 1,9E-08 | CCNE2,CDC45,CDC6,CDT1,MCM10,MCM2,MCM4,MCM5,MCM6,MCM7,C | 13 | 35 | 1,94 | 13 |
| 10 | GO:0034341 | response to interferon-gamma | 2,2E-08 | ACOD1,B2M,CALCOCO2,CAMK2D,CCL13,CCL15,CCL18,CCL5,CCL7,CCL8,C | 58 | 174 | 9,66 | 58 |
| 11 | GO:0006954 | inflammatory response | 2,9E-08 | ACOD1,ADAM8,ADGRE5,ADORA3,AGT,AIM2,AOC3,APOL2,APOL3,APPL2,B | 87 | 589 | 32,7 | 87 |
| 12 | GO:0050852 | T cell receptor signaling pathway | 9,4E-08 | BTN2A2,BTN3A1,BTN3A2,BTN3A3,CD28,CEACAM1,CTLA4,CYLD,FYB1,GBP | 28 | 165 | 9,16 | 28 |
| 13 | GO:0045087 | innate immune response | 1,1E-07 | ACOD1,ADAM8,AIM2,APOBEC3A,APOBEC3D,APOBEC3G,APOL1,APPL2,B2 | 136 | 710 | 39,42 | 136 |
| 14 | GO:0006958 | complement activation, classical pathway | 2,5E-07 | C1QB,C1R,C1RL,C1S,C2,C3,C4A,C4B,SERPING1 | 9 | 19 | 1,05 | 9 |
| 15 | GO:0000727 | double-strand break repair via break-induced replication | 4,1E-07 | CDC45,GINS2,MCM2,MCM4,MCM5,MCM6,MCM7 | 7 | 11 | 0,61 | 7 |
| 16 | GO:0042832 | defense response to protozoan | 4,4E-07 | BATF2,CD40,GBP2,GBP4,GBP6,GBP7,IL12A,SLC11A1,TSPAN32 | 9 | 20 | 1,11 | 9 |
| 17 | GO:0071222 | cellular response to lipopolysaccharide | 7,1E-07 | ACOD1,CARD16,CARD17,CCL5,CD274,CD80,CMPK2,CX3CR1,CXCL10,CXC | 27 | 171 | 9,49 | 27 |
| 18 | GO:0030449 | regulation of complement activation | 8,5E-07 | C1QB,C1R,C1S,C2,C3,C4A,C4B,CFB,CFH,CR1L,SERPING1 | 11 | 33 | 1,83 | 11 |
| 19 | GO:0070098 | chemokine-mediated signaling pathway | 0,0000012 | CCL13,CCL15,CCL18,CCL5,CCL7,CCL8,CCRL2,CX3CR1,CXCL10,CXCL11,CX | 15 | 63 | 3,5 | 15 |
| 20 | GO:0002479 | antigen processing and presentation of exogenous peptide a | 0,0000014 | B2M,HLA-A,HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,NCF1,PSMB10,PSMB8,PSMB | 16 | 72 | 4 | 16 |
| 21 | GO:0019886 | antigen processing and presentation of exogenous peptide a | 0,0000016 | CD74,CTSS,HLA-DMA,HLA-DMB,HLA-DOA,HLA-DOB,HLA-DPA1,HLA-DPB1,I | 18 | 90 | 5 | 18 |
| 22 | GO:0030593 | neutrophil chemotaxis | 0,0000019 | CCL13,CCL15,CCL18,CCL5,CCL7,CCL8,CD300H,CD74,CXADR,CXCL10,CXC | 17 | 82 | 4,55 | 17 |
| 23 | GO:0071356 | cellular response to tumor necrosis factor | 0,0000041 | ACOD1,AIM2,BRCA1,CARD16,CASP4,CCL13,CCL15,CCL18,CCL5,CCL7,CCL | 36 | 246 | 13,66 | 36 |
| 24 | GO:0042102 | positive regulation of T cell proliferation | 0,0000044 | CCL5,CD274,CD28,CD80,HLA-DMB,HLA-DPA1,HLA-DPB1,HLA-E,IL12A,IL1 | 16 | 78 | 4,33 | 16 |
| 25 | GO:0045071 | negative regulation of viral genome replication | 0,0000067 | APOBEC3A,APOBEC3D,APOBEC3G,CCL5,CCL8,IFIT5,IFITM1,IFITM2,IFITM3 | 13 | 55 | 3,05 | 13 |
| 26 | GO:0032729 | positive regulation of interferon-gamma production | 0,0000067 | HLA-DPA1,HLA-DPB1,IL12A,IL12RB1,IL12RB2,IL27,KLRK1,PDE4B,RIPK2,S | 13 | 55 | 3,05 | 13 |
| 27 | GO:0007186 | G protein-coupled receptor signaling pathway | 0,0000076 | ADGRE5,ADGRG5,ADORA3,AGT,C3,CCL13,CCL15,CCL18,CCL5,CCL7,CCL8, | 59 | 594 | 32,98 | 59 |
| 28 | GO:0071346 | cellular response to interferon-gamma | 0,0000085 | ACOD1,B2M,CAMK2D,CCL13,CCL15,CCL18,CCL5,CCL7,CCL8,CIITA,EDN1,( | 49 | 158 | 8,77 | 49 |
| 29 | GO:0050830 | defense response to Gram-positive bacterium | 0,00001 | GBP2,GBP4,GBP6,GBP7,GSDMD,H2BC21,HLA-E,IL12A,KLRK1,P2RX7,RIPK | 12 | 49 | 2,72 | 12 |
| 30 | GO:0000083 | regulation of transcription involved in G1/S transition of mi | 0,000012 | BACH1,CDC45,CDC6,CDT1,DHFR,E2F1,ORC1,RRM2,TYMS | 9 | 28 | 1,55 | 9 |
| 31 | GO:0050729 | positive regulation of inflammatory response | 0,000019 | ADAM8,AGT,C3,CCL18,CD28,GBP5,HLA-E,IDO1,IL15,JAK2,KARS1,LRRK2,N | 22 | 119 | 6,61 | 22 |
| 32 | GO:0042531 | positive regulation of tyrosine phosphorylation of STAT prot | 0,00003 | CCL5,CD40,IL12A,IL15,IL31RA,JAK2,LYN,OSM,PARP14,STAP2,TNFSF18 | 11 | 46 | 2,55 | 11 |
| 33 | GO:0035455 | response to interferon-alpha | 0,000062 | IFIT2,IFIT3,IFITM1,IFITM2,IFITM3,LAMP3,PYHIN1 | 7 | 20 | 1,11 | 7 |
| 34 | GO:0050718 | positive regulation of interleukin-1 beta secretion | 0,000072 | AIM2,GSDMD,LILRA2,NLRP12,NOD2,P2RX7,RIPK2,TLR8 | 8 | 27 | 1,5 | 8 |
| 35 | GO:0032693 | negative regulation of interleukin-10 production | 0,00009 | CD274,DLL1,IDO1,JAK3,PDCD1LG2,TNFRSF21 | 6 | 15 | 0,83 | 6 |
| 36 | GO:0000076 | DNA replication checkpoint | 0,00009 | CDC45,CDC6,CDT1,CLSPN,ORC1,TICRR | 6 | 15 | 0,83 | 6 |
| 37 | GO:2000425 | regulation of apoptotic cell clearance | 0,0001 | C3,C4A,C4B,CD300LF | 5 | 10 | 0,56 | 5 |
| 38 | GO:0071347 | cellular response to interleukin-1 | 0,00015 | CCL1,CCL13,CCL15,CCL18,CCL5,CCL7,CCL8,EDN1,GBP1,GBP2,GBP3,ICA | 21 | 157 | 8,72 | 21 |

# Show me your data...


Show me your data...

# A cocktail recipe

*You might (and should) have a set of standardized objects for your datasets and results (Excel does **not** count)*

# A cocktail recipe

*A container for the expression matrix,* `dds`

# A cocktail recipe

*A container for the results of differential expression,* `res_de`

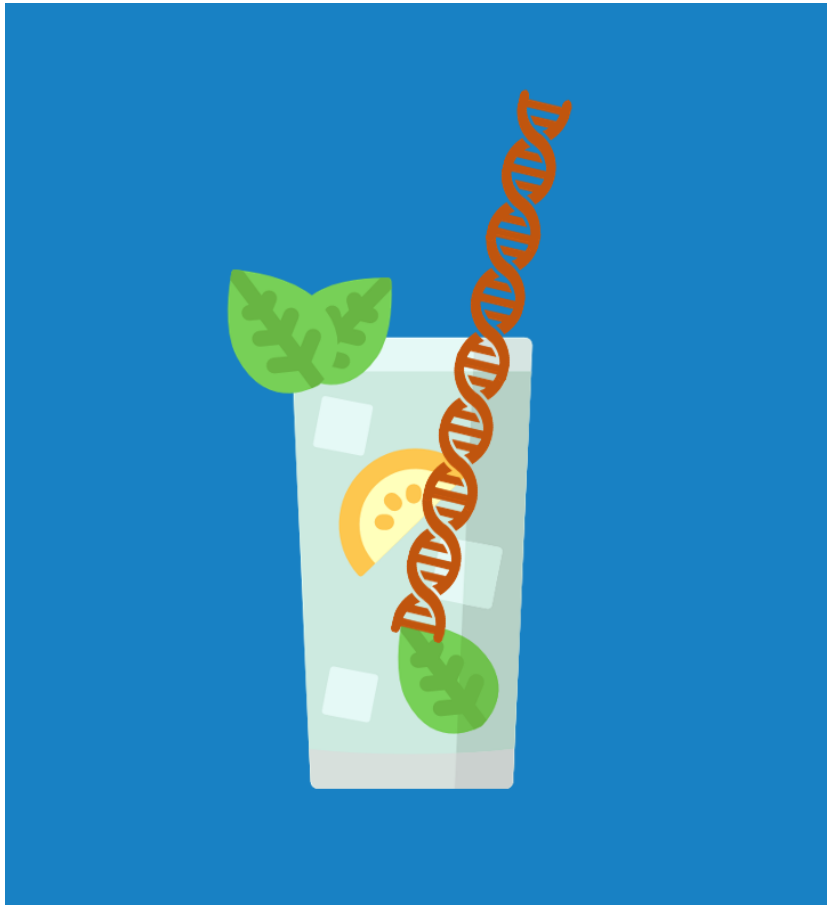
# A cocktail recipe

*A table for the result of functional enrichment,* `res_enrich`

# A cocktail recipe

*Decorate with an annotation table,* `anno_df`

# A cocktail recipe

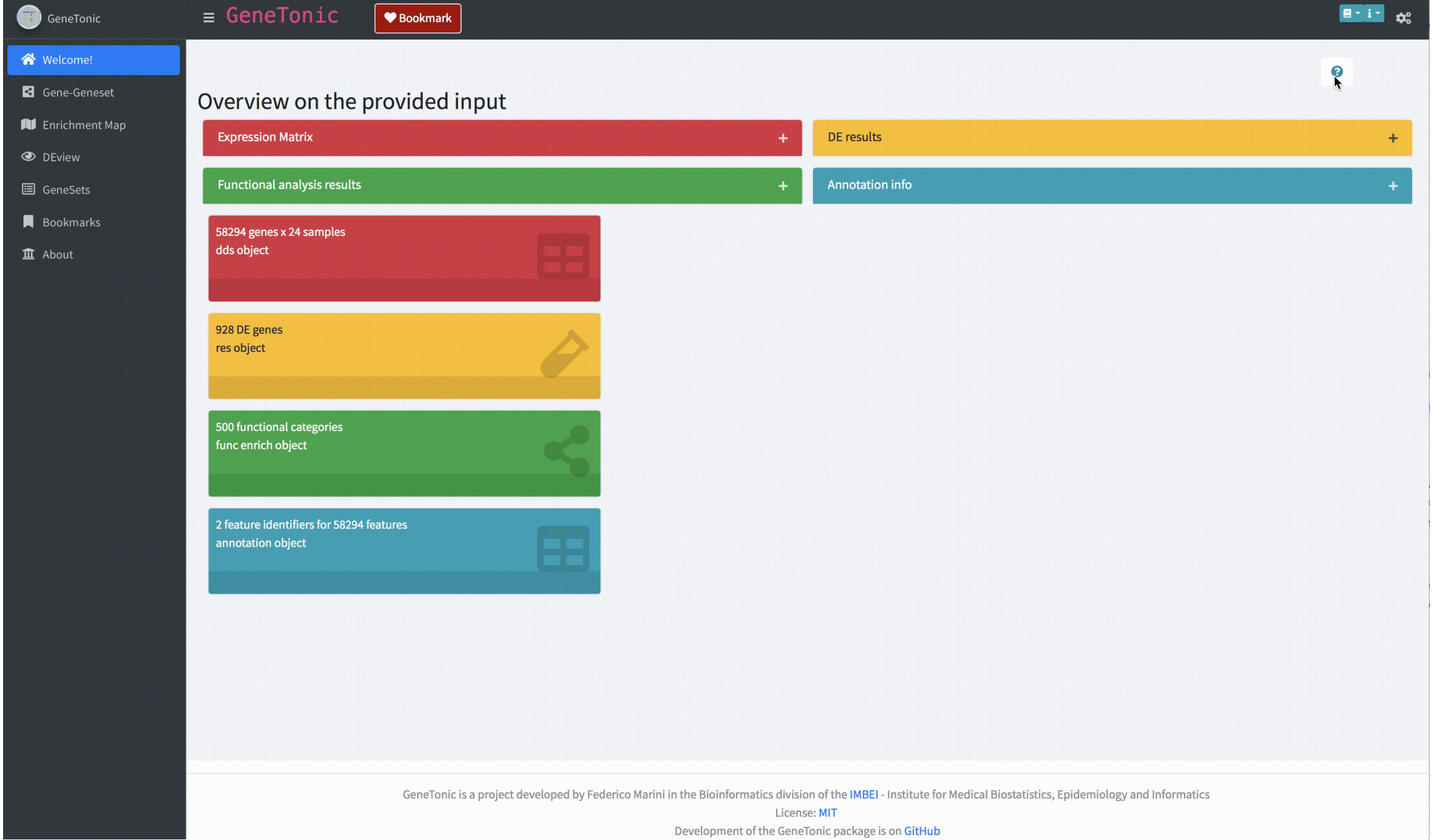


*Shaken, not stirred...*

# A cocktail recipe


`GeneTonic` is now on Bioconductor!


~/Development/GeneTonic - Shiny
http://127.0.0.1:5184
Open in Browser
Publish
GeneTonic
GeneTonic
Bookmark
First steps with a full glass
Welcome!
Gene-Geneset
Enrichment Map
Overview
GSViz
Bookmarks
Overview on the provided input
Expression Matrix
DE results
Functional analysis results
Annotation info
58294 genes x 24 samples
dds object
928 DE genes
res object
500 functional categories
func enrich object
2 feature identifiers for 58294 features
annotation object
GeneTonic is a project developed by Federico Marini in the Bioinformatics division of the IMBEI - Institute for Medical Biostatistics, Epidemiology and Informatics
License: MIT - The GeneTonic package is developed and available on GitHub

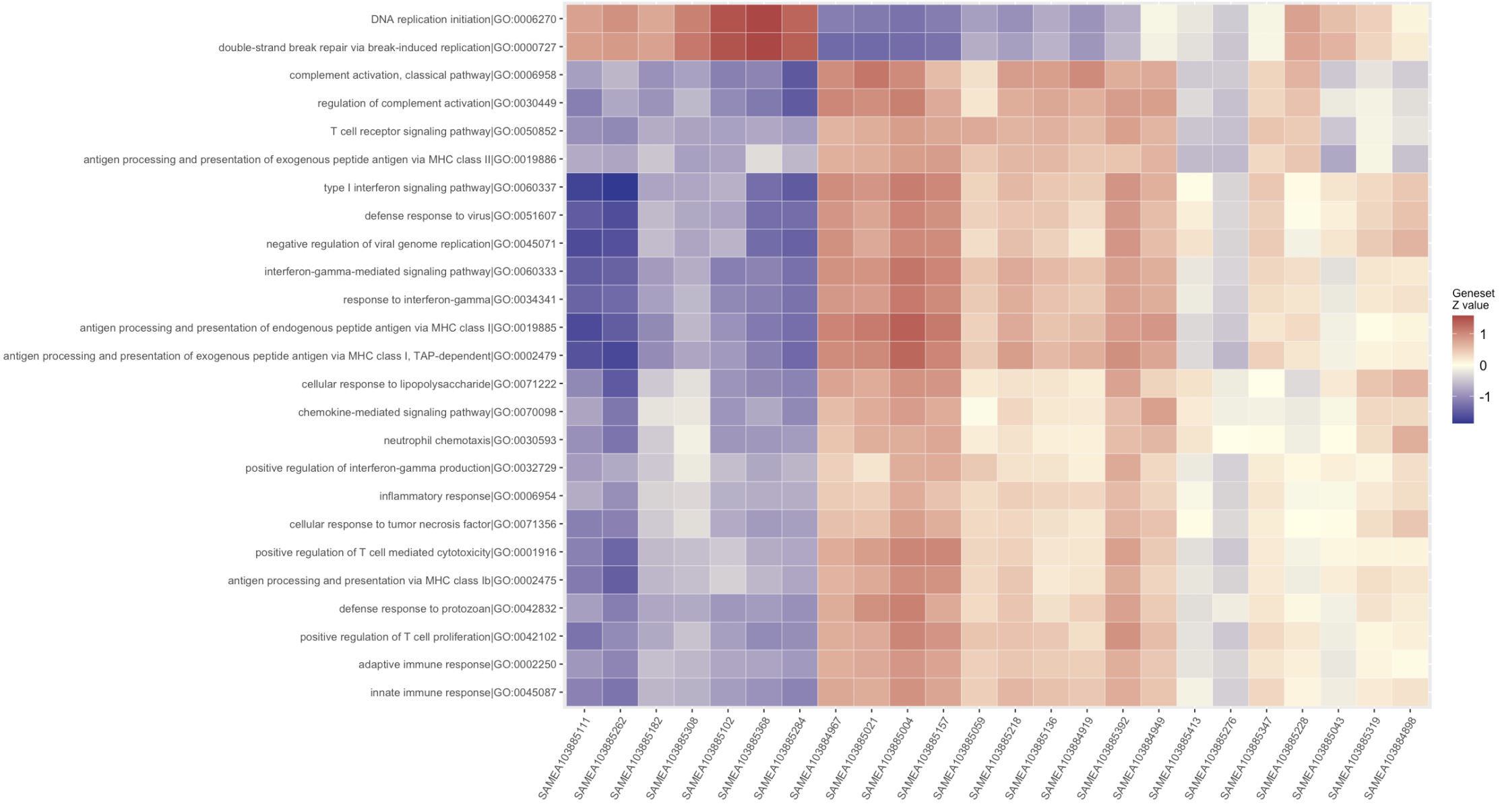

GeneTonic
GeneTonic
Bookmark
Welcome!
Gene-Geneset
Enrichment Map
DEview
GeneSets
Bookmarks
About
Overview on the provided input
Expression Matrix
DE results
Functional analysis results
Annotation info
58294 genes x 24 samples
dds object
928 DE genes
res object
500 functional categories
func enrich object
2 feature identifiers for 58294 features
annotation object
GeneTonic is a project developed by Federico Marini in the Bioinformatics division of the IMBEI - Institute for Medical Biostatistics, Epidemiology and Informatics
License: MIT
Development of the GeneTonic package is on GitHub

GSViz

Scores Heatmap | Alluvial Plot | Summary Heatmap | Geneset MDS | **Summary Overview** | Geneset Radar | Geneset Dendrogram

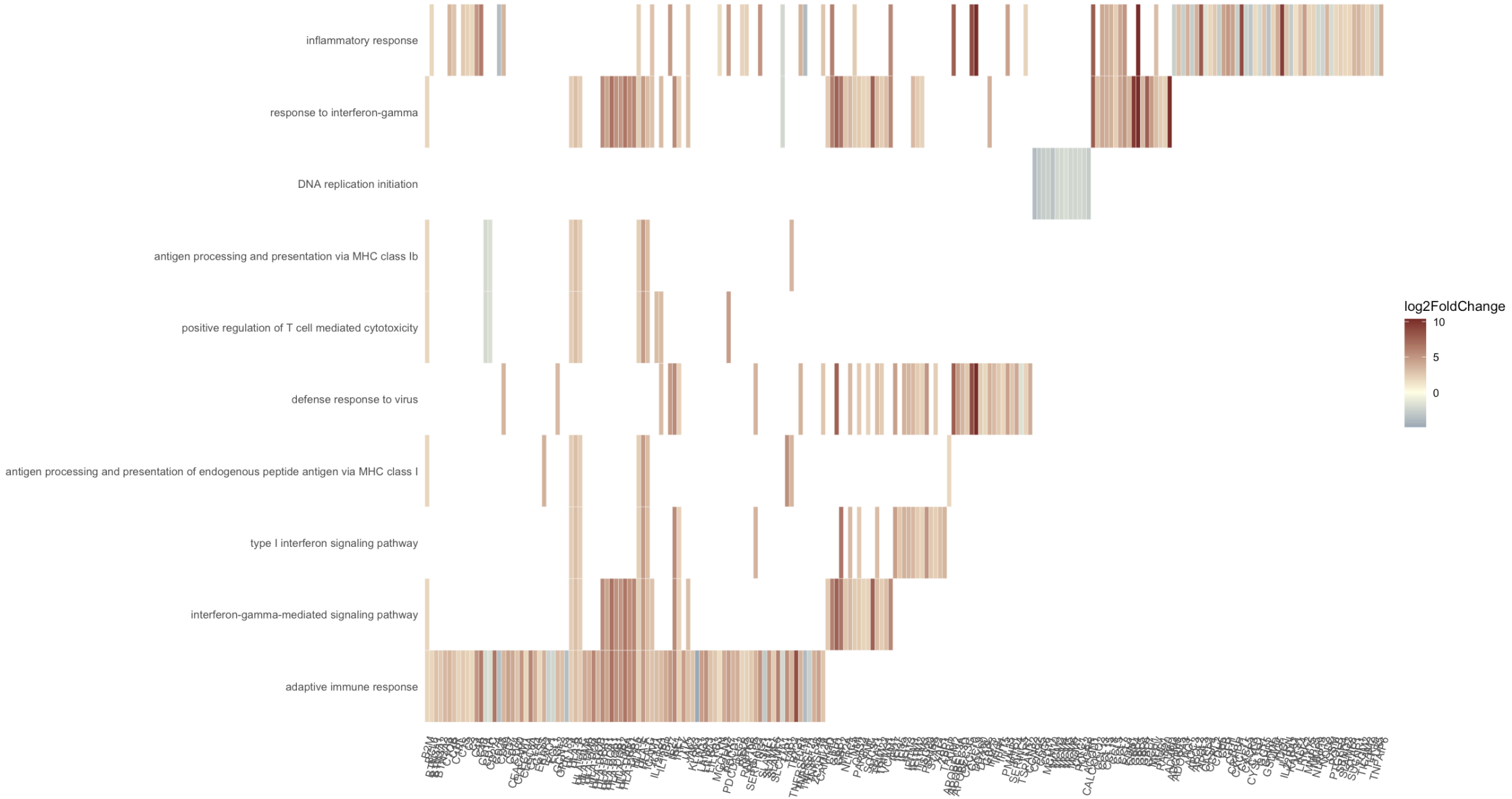
GSViz

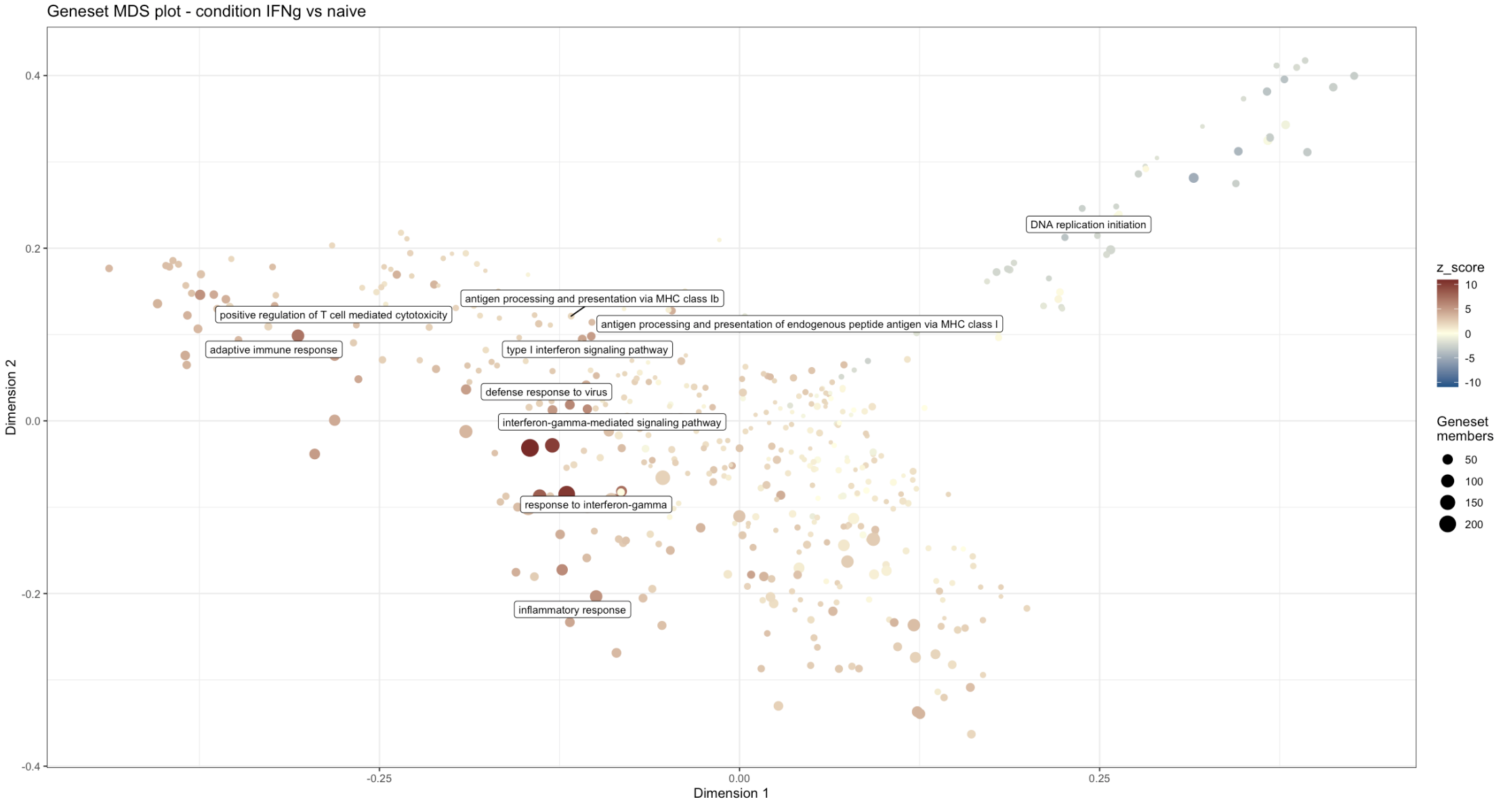
Scores Heatmap | Alluvial Plot | Summary Heatmap | Geneset MDS | Summary Overview | Geneset Radar | Geneset Dendrogram

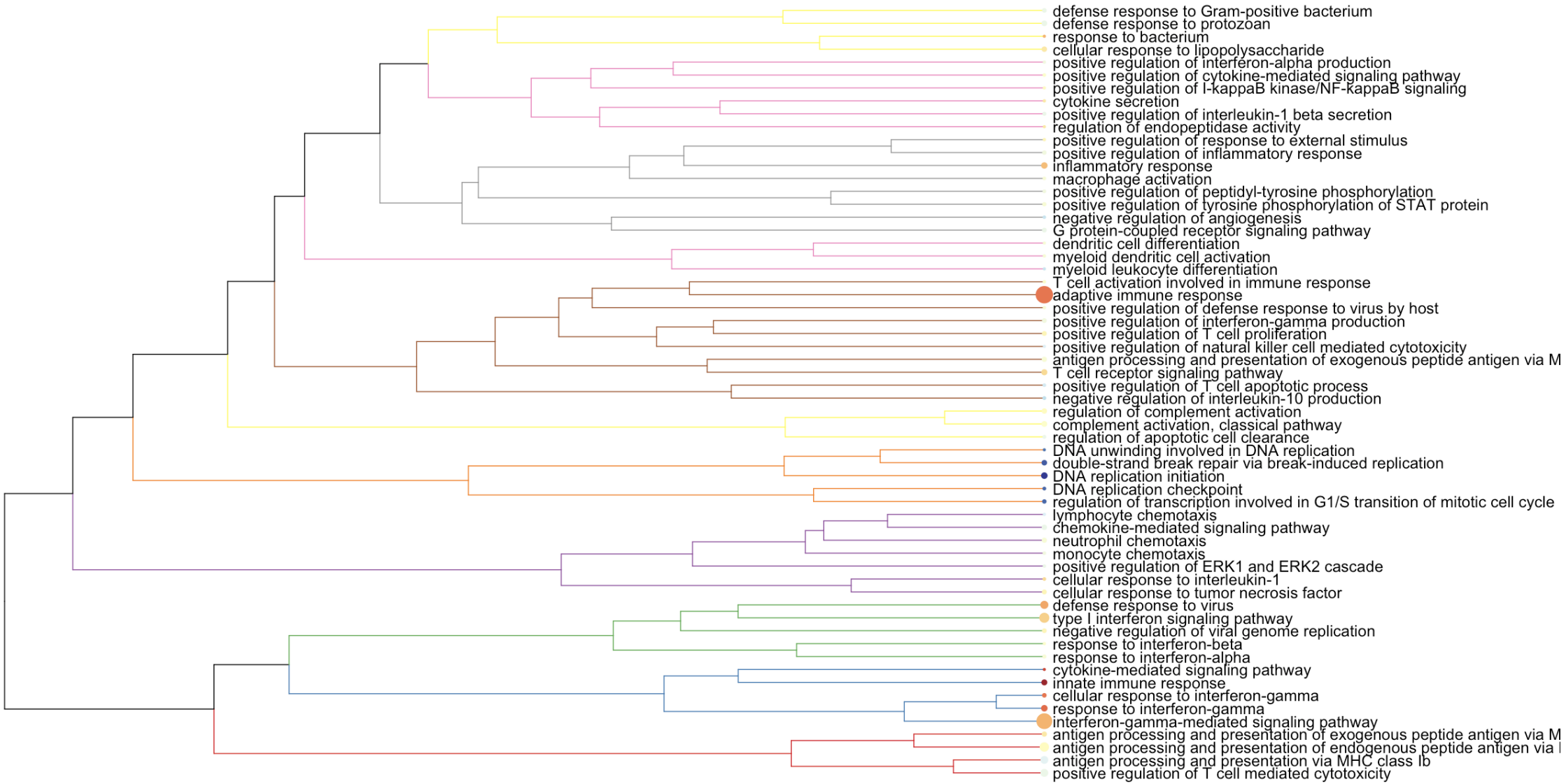


GSViz

Scores Heatmap | Alluvial Plot | **Summary Heatmap** | Geneset MDS | Summary Overview | Geneset Radar | Geneset Dendrogram

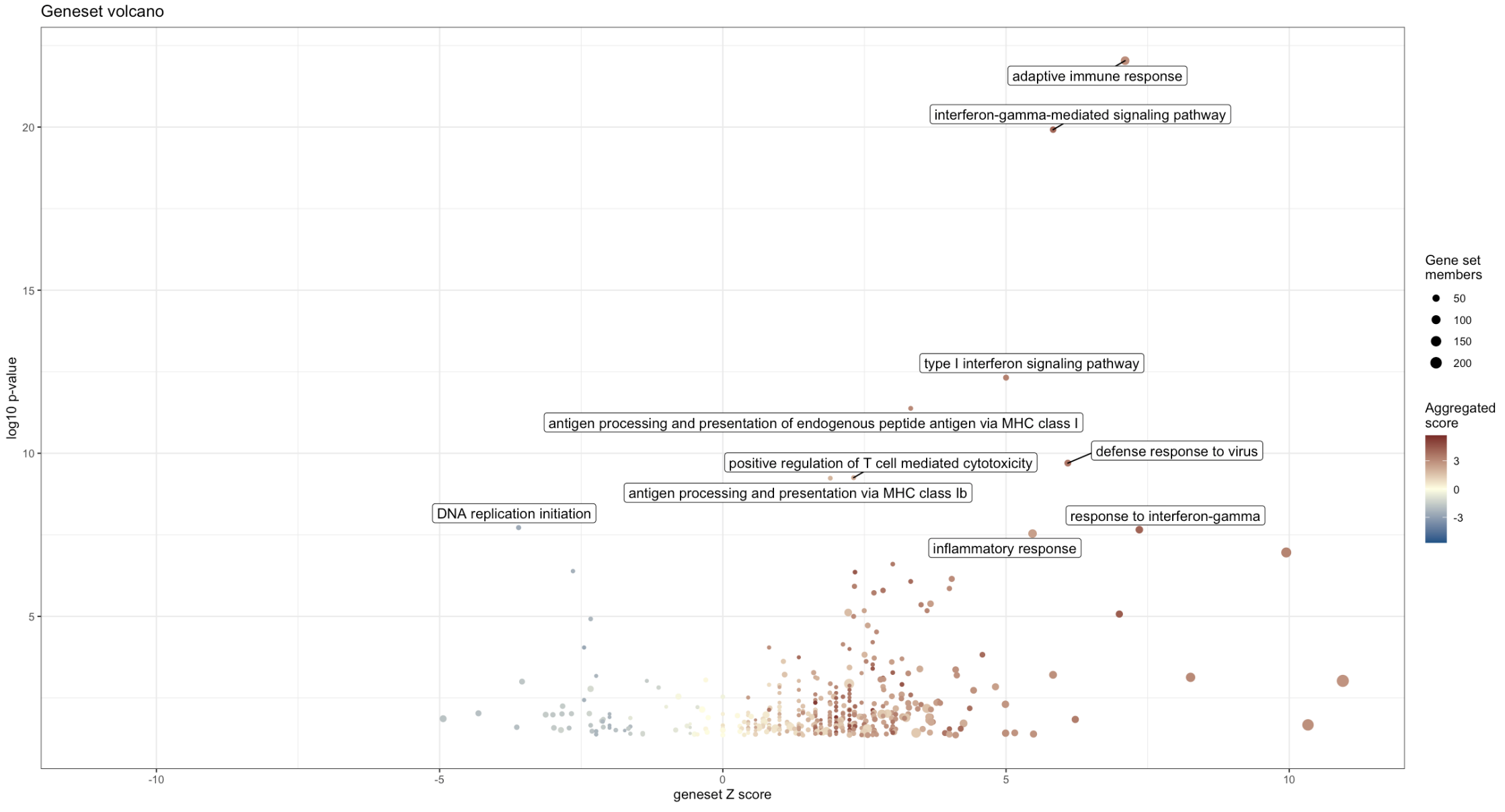



Geneset MDS plot - condition IFNg vs naive
DNA replication initiation
antigen processing and presentation via MHC class Ib
positive regulation of T cell mediated cytotoxicity
antigen processing and presentation of endogenous peptide antigen via MHC class I
adaptive immune response
type I interferon signaling pathway
defense response to virus
interferon-gamma-mediated signaling pathway
response to interferon-gamma
inflammatory response
Dimension 2
Dimension 1
0.4
0.2
0.0
-0.2
-0.4
-0.25
0.00
0.25
z_score
10
5
0
-5
-10
Geneset members
50
100
150
200


defense response to Gram-positive bacterium
defense response to protozoan
response to bacterium
cellular response to lipopolysaccharide
positive regulation of interferon-alpha production
positive regulation of cytokine-mediated signaling pathway
positive regulation of I-kappaB kinase/NF-kappaB signaling
cytokine secretion
positive regulation of interleukin-1 beta secretion
regulation of endopeptidase activity
positive regulation of response to external stimulus
positive regulation of inflammatory response
inflammatory response
macrophage activation
positive regulation of peptidyl-tyrosine phosphorylation
positive regulation of tyrosine phosphorylation of STAT protein
negative regulation of angiogenesis
G protein-coupled receptor signaling pathway
dendritic cell differentiation
myeloid dendritic cell activation
myeloid leukocyte differentiation
T cell activation involved in immune response
adaptive immune response
positive regulation of defense response to virus by host
positive regulation of interferon-gamma production
positive regulation of T cell proliferation
positive regulation of natural killer cell mediated cytotoxicity
antigen processing and presentation of exogenous peptide antigen via M
T cell receptor signaling pathway
positive regulation of T cell apoptotic process
negative regulation of interleukin-10 production
regulation of complement activation
complement activation, classical pathway
regulation of apoptotic cell clearance
DNA unwinding involved in DNA replication
double-strand break repair via break-induced replication
DNA replication initiation
DNA replication checkpoint
regulation of transcription involved in G1/S transition of mitotic cell cycle
lymphocyte chemotaxis
chemokine-mediated signaling pathway
neutrophil chemotaxis
monocyte chemotaxis
positive regulation of ERK1 and ERK2 cascade
cellular response to interleukin-1
cellular response to tumor necrosis factor
defense response to virus
type I interferon signaling pathway
negative regulation of viral genome replication
response to interferon-beta
response to interferon-alpha
cytokine-mediated signaling pathway
innate immune response
cellular response to interferon-gamma
response to interferon-gamma
interferon-gamma-mediated signaling pathway
antigen processing and presentation of exogenous peptide antigen via M
antigen processing and presentation of endogenous peptide antigen via
antigen processing and presentation via MHC class Ib
positive regulation of T cell mediated cytotoxicity

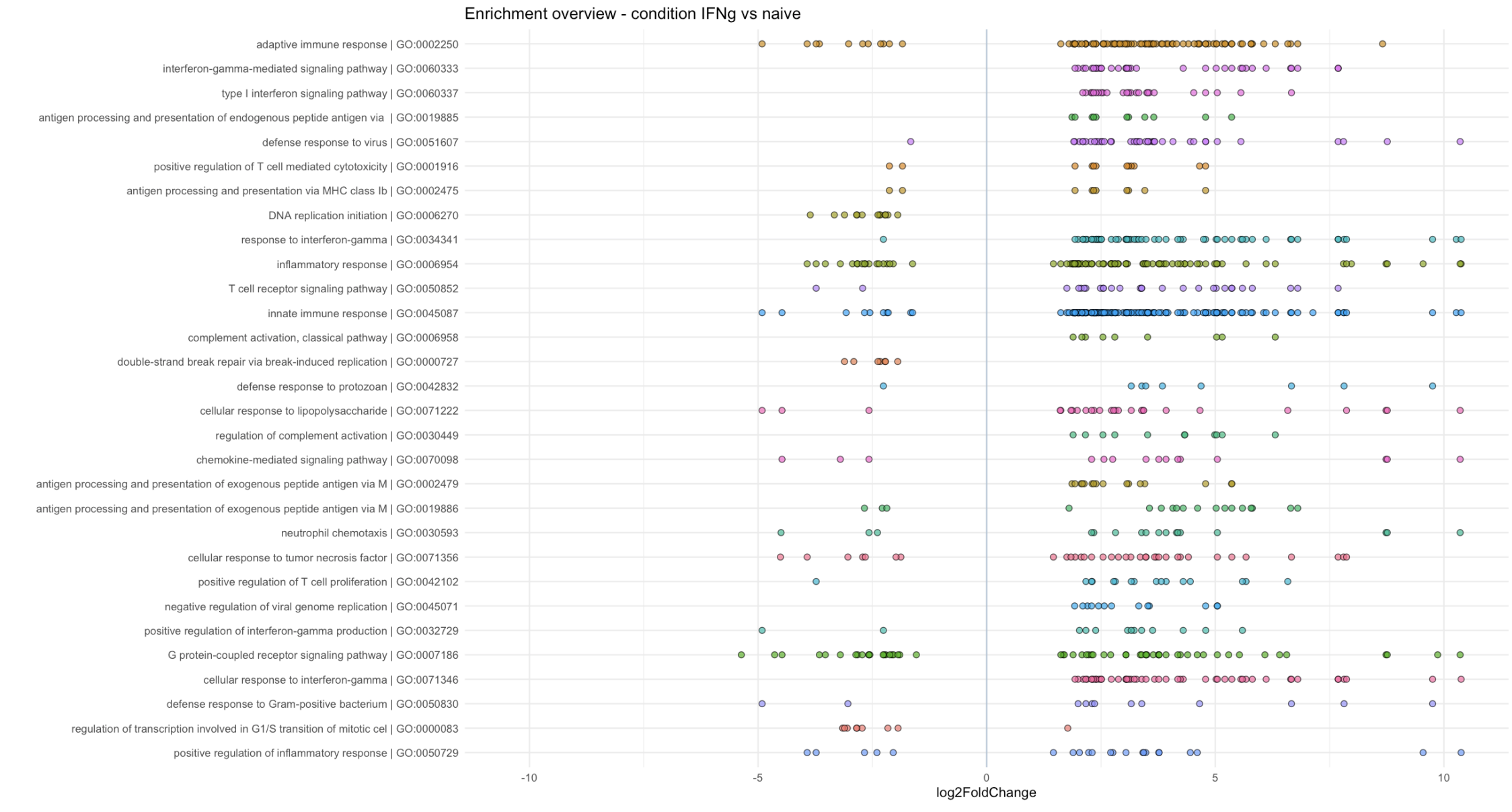
Overview

Geneset Volcano | Geneset Volcano - simplified | Enhanced Table | Enhanced Table - interactive



Overview

Geneset Volcano | Geneset Volcano - simplified | **Enhanced Table** | Enhanced Table - interactive



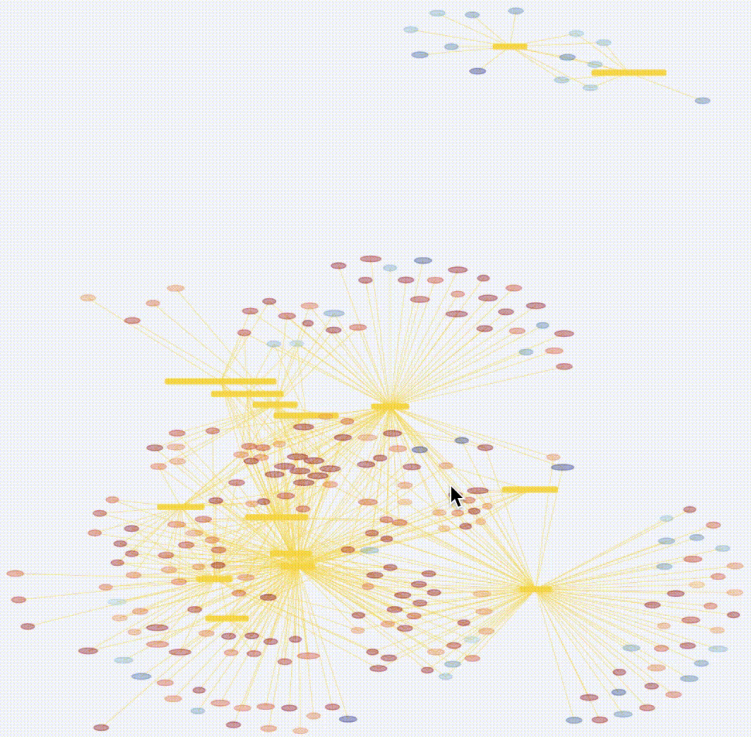
GeneTonic

GeneTonic

Bookmark

- Welcome!
- Gene-Geneset
- Enrichment Map
- DEview
- GeneSets
- Bookmarks
- About

Select by id



Genesetbox

Please select a gene set.

Genebox

Please select a gene/feature.

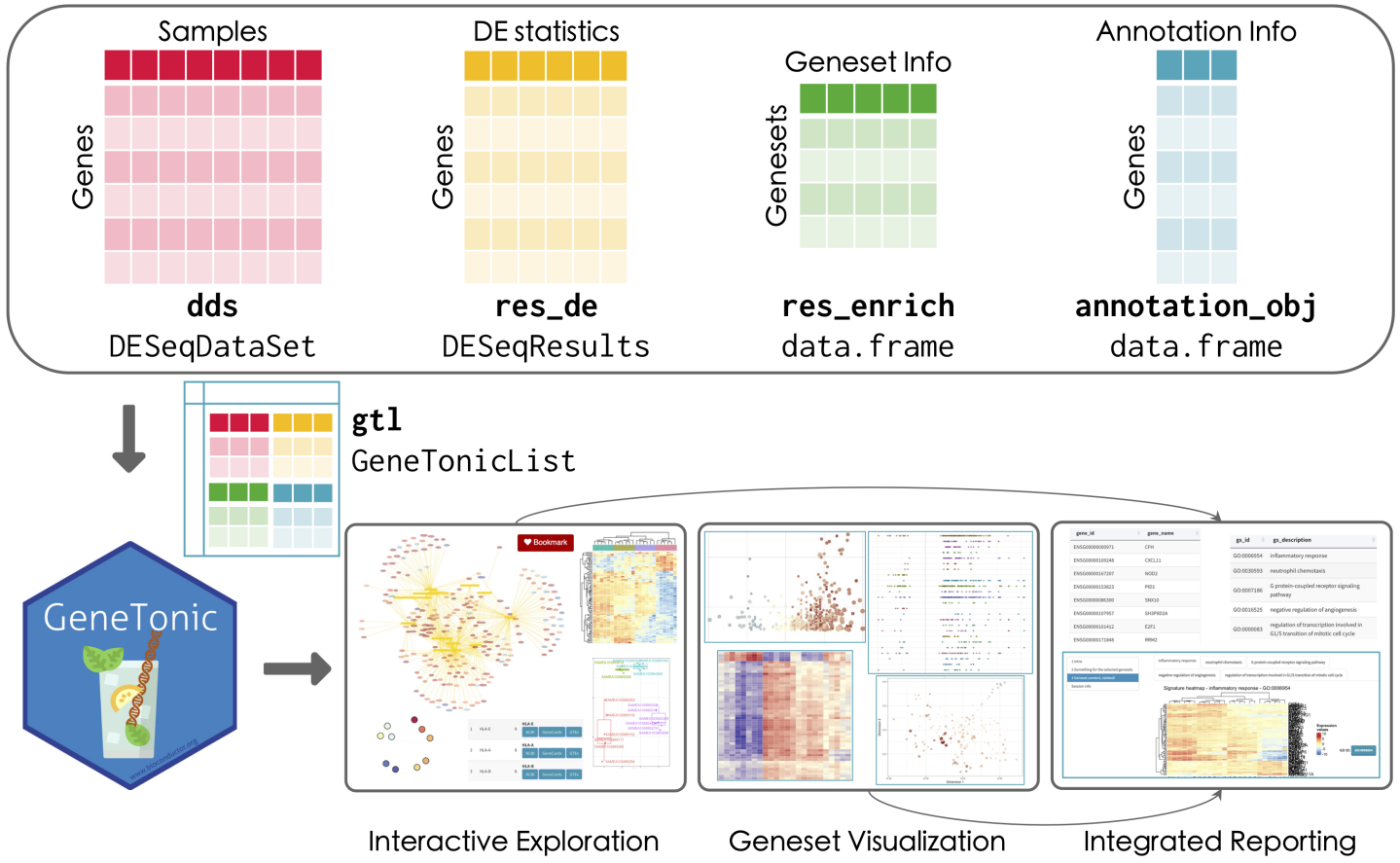
# The interplay of genes and genesets

# A quick demo

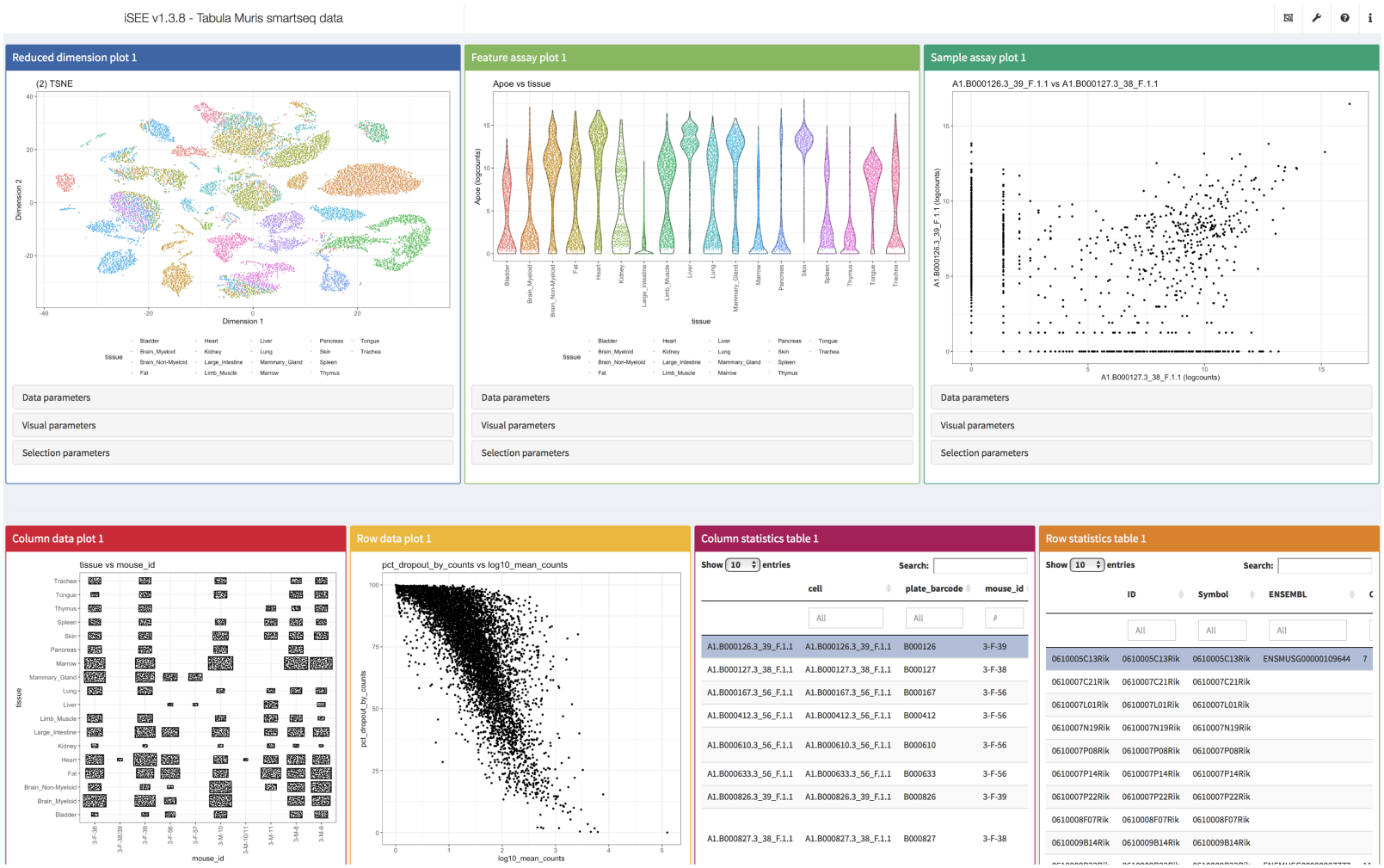
```
BiocManager::install("GeneTonic")
library("GeneTonic")
example(GeneTonic, ask = FALSE)
```

# GeneTonic

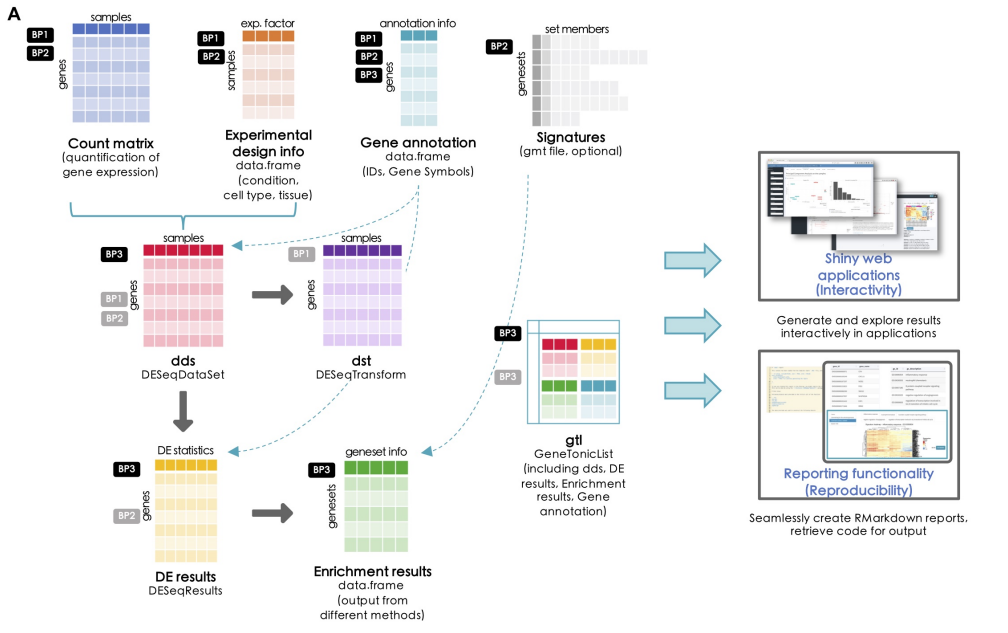

A way to enjoy the data interpretation...



# Suddenly iSEE



# Putting it all together



# Interfaces for RNA-seq

**Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective** FREE

Alicia Poplawski, Federico Marini, Moritz Hess, Tanja Zeller, Johanna Mazur, Harald Binder

*Briefings in Bioinformatics*, Volume 17, Issue 2, March 2016, Pages 213–223,
https://doi.org/10.1093/bib/bbv036
**Published:** 23 June 2015 **Article history** ▾

Interfaces have been developed for making such analysis steps accessible to life scientists without extensive knowledge of command line tools.

Systematic search and evaluation of such interfaces

Definition of criteria for evaluation (ease of configuration, documentation, usability, computational demand, reporting)

There's no `one tool fits all` winner!


Does the analysis/exploration/interpretation of your data spark joy?

# Lessons learnt for me (so far)

**It takes two to tango**

Results & Interpretation

Bioinformatician & Experimental scientist

Reproducibility & Interactivity

*Stop using Excel for performing Bioinformatics data analysis :)*

**Our contributions to the Bioconductor project:**

`pcaExplorer` - Exploratory Data Analysis

`ideal` - Differential Expression analysis

`GeneTonic` - Interpretation of DE results

`iSEE` - Exploration of single cell datasets

...

`countsimQC` - comparing count data sets

`iCOBRA` - calculation and visualization of performance metrics

`ExploreModelMatrix`

`tximeta` - tx importing with metadata, *for free*

`alevinQC`

# Practical session

# References:


- Marini and Binder (2016) - Development of Applications for Interactive and Reproducible Research: a Case Study (Genomics Computational Biology) 10.18547/gcb.2017.vol3.iss1.e39
- Marini and Binder (2019) - pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components (BMC Bioinformatics) 10.1186/s12859-019-2879-1
- Marini, Linke, Binder (2020) - ideal: an R/Bioconductor package for Interactive Differential Expression Analysis (BMC Bioinformatics) 10.1186/s12859-020-03819-5
- Rue-Albrecht, Marini, Soneson, Lun (2018) - iSEE: Interactive SummarizedExperiment Explorer (F1000 Research) https://doi.org/10.12688/f1000research.14966.1
- Marini, Ludt, Linke, Strauch (2021) - GeneTonic: an R/Bioconductor package for streamlining the interpretation of RNA-seq data (BMC Bioinformatics) https://doi.org/10.1186/s12859-021-04461-5
- Ludt, Ustjanzew, Binder, Strauch, Marini (2022) - Interactive and Reproducible Workflows for Exploring and Modeling RNA-seq Data with pcaExplorer, Ideal, and GeneTonic (Current Protocols) https://doi.org/10.1002/cpz1.411

## References: (cont'd)


- Srivastava, Malik, Sarkar, Zakeri, Almodaresi, Soneson, Love, Kingsford, Patro (2020) - Alignment and mapping methodology influence transcript abundance estimation (Genome Biology) https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02151-8
- Conesa, Madrigal, Tarazona, Gomez-Cabrero, Cervera, McPherson, Szcześniak, Gaffney, Elo, Zhang, Mortazavi (2016) - A survey of best practices for RNA-seq data analysis (Genome Biology) `https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8`


## Resources

The whole `rnaseqGene` workflow is detailed here: https://www.bioconductor.org/packages/rnaseqGene/

# ... thank you for your attention!

marinif@uni-mainz.de - @FedeBioinfo charlotte.soneson@fmi.ch - @CSoneson