# artMS <br> Analytical R Tools for Mass Spectrometry 

http://artms.org
Bioconductor

David Jimenez-Morales, PhD
敬 Stanford
Cardiovascular Medicine
Ashley Lab
for Mass Spectrometry
http://artms.org Bioconductor

## Goals

Facilitate the analysis of Mass-Spectrometry based Proteomics data using the $\mathbb{R}$ programming language

2
Provide different levels of analysis to ensure data quality, quantification, and frame results in a biological context.

Quantification Methods
Label Free (+/- Fractionation) SILAC

Analysis of

## +Global Proteome <br> +PostTranslational Modifications

Phosphorylation
Ubiquitination
Acetylation $\longleftarrow$ (coming soon)


## Experimental

## Computational



## Peptide Identification

- Tool widely used
- Algorithms yield high mass accuracy and precision
- It covers a wide variety of proteomics experiments
- Simplified user interface
- It can now run on Linux (improved performance)

Experimental

## Computational



## artMS

Analytical R Tools for Mass Spectrometry
http://artms.org

Quality control



## Relative <br> quantification

evidence + keys + contrast

Functional analysis
Data integration Imputation Enrichment analysis
Pathway analysis
Network generation

Miscellaneous

## How does it work?

## $\operatorname{artMS}$

Analytical R Tools
for Mass Spectrometry
http://artms.org Bioconductor
artMS
Analytical R Tools for Mass Spectrometry
http://artms.org

Quality control


Relative quantification

## Miscellaneous

artmsQualityControlEvidenceBasic()
artmsQualityControlEvidenceExtended()
artmsQualityControISummaryExtended()

## artmsQuantification()

artmsAnalysisQuantifications()
artmsEvidenceToSAINTq()
artmsEvidenceToSaintExpress()
artmsPhosfateOutput()
artmsPhotonOutput()
etc

## Basic Functions

artMS
Analytical R Tools for Mass Spectrometry http://artms.org $\sqrt{\text { Bioconductor }}$

Quality control

Functional analysis

Input Files
Configuration

## Experimental design

|  | A | B | C | D | E |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | RawFile | IsotopeLabelType | Condition | BioReplicate | Run |
| 2 | qx006145 | L | Cal33 | Cal33-1 | 1 |
| 3 | qx006146 | L | Cal33 | Cal33-2 | 2 |
| 4 | qx006147 | L | Cal33 | Cal33-3 | 3 |
| 5 | qx006148 | L | Cal33 | Cal33-4 | 4 |
| 6 | qx006154 | L | HSC6 | HSC6-1 | 5 |
| 7 | qx006151 | L | HSC6 | HSC6-2 | 6 |
| 8 | qx006152 | L | HSC6 | HSC6-3 | 7 |
| 9 | qx006153 | L | HSC6 | HSC6-4 | 8 |
| 10 |  |  |  |  |  |

- Two cell cancer cell lines: Cal33, HSC6
- 4 biological replicates each

- Two cell cancer cell lines: Cal33, HSC6
- 4 biological replicates each
artMS
Analytical R Tools
for Mass Spectrometry http://artms.org Bioconductor

Quality control

Functional analysis

Miscellaneous

Input Files
$M_{0}$ evidence.txt + keys.txt + contrast.txt
Configuration

- Two cell cancer cell lines: A, B, C, D, E, F, G
- 4 biological replicates each


## artMS

## Analytical R Tools for Mass Spectrometry

http://artms.org

Quality control

Functional analysis

Miscellaneous
data:
enabled: 1
fractions:
enabled: 0
silac:
enabled: 0
filters:
enabled: 1
contaminants: 1
erotein groups: remove
modifications: PH
sample_plots: 1
msstats:
enabled: 1
msstats inpu :
profilePlots: none
normalization method: equalizeMedians
normalization reference:
summaryMethod: TMP
censoredInt: NA
cutoffcensored: minFeature
MBimpute: 1
feature_subset: all
butput extras:
enabled: 1
annotate:
enabled: 1
specie: HUMAN
plots:
volcano: 1
heatmap: 1
LFC: -1.51 .5
FDR: 0.05
reatman display: log2FC
basic: 1

```
extended: 1
```

```
extended: 1
```

Configuration

## artMS

## Analytical R Tools for Mass Spectrometry

http://artms.org

Quality control

Functional analysis

Miscellaneous

enabled: 1
fractions:
enabled: 0
silac:
enabled: 0
filters:
enabled: 1
orotein groups: remove
modifications: PH
sample_plots: 1
msstats:
enabled: 1
msstats_input:
profilePlots: none
normalization_method: equalizeMedians
normalization_reference:
summaryMethod: TMP
censoredInt: NA
cutoffCensored: minFeature
MBimpute: 1
feature_subset: all
จutput extras:
enabled: 1
annotate:
enabled: 1
specie: HUMAN
plots:
volcano:
heatmap:
LFC: -1.5 1.5
FDR: 0.05
heatinap_ctuster_cols:

Configuration
artMS

## Analytical R Tools for Mass Spectrometry

http://artms.org

Quality control

Functional analysis

Miscellaneous
config.yaml
les :
evidence : '~/ph_ms/evidence.txt'
keys : '~/ph_ms/keys.txt
contrasts : '~/ph_ms/contrast.txt'
output : '~/ph_ms/phglobal/phglobal-results.txt'
basic: 1

```
data:}\begin{array}{l}{\mathrm{ enabled: 1 Proteomics}}\\{\mathrm{ fractions: }}
                                    details
enabled:
silac:
silac:
enabled: 0
filters:
enabled: 1
contaminants: 1
protein_groups: remove
modifications: PH
amole olots: 1
```


## msstats: enabled: 1

msstats_input:
profilePlots: none
normalization method: equalizeMedians
normalization reference:
summaryMethod: TMP
censoredInt: NA
cutoffCensored: minFeature
MBimpute: 1
feature_subset: all
จutput extras:
enabled: 1
annotate:
enabled: 1 specie: HUMAN
plots:
volcano: 1
heatmap: 1
LFC: -1.51 .5
FDR: 0.05
heatmap_cluster_cols: 0

Configuration
artMS

## Analytical R Tools for Mass Spectrometry

http://artms.org

Quality control

Functional analysis

Miscellaneous
config.yaml
les :
evidence : ' $\sim /$ ph_ms/evidence.txt'
keys : '~/ph_ms/keys.txt
contrasts : '~/ph_ms/contrast.txt
output : '~/ph_ms/phglobal/phglobal-results.txt'
qc:
basic: 1
extended: 1
data:
enabled: 1
fractions:
enabled: 0
silac:
enabled: 0
filters:
enabled: 1
contaminants: 1
protein_groups: remove
modifications: PH

## msstats: enabled: 1 msstats_input: $\quad$ Relative Quantification <br> msstats_input:

profilePlots: none
ormalization_method: equalizeMedians
normatization_metnou: equal
sumnaryMethod: TMP
censoredInt: NA
cutofficensored: minFeature
MBimpute: 1
feature_subset: all
output_extras:
enabled: 1
annotate:
enabled: 1
specie: HUMAN
plots:
volcano: 1
heatmap: 1
LFC: -1.51 .5
FDR: 0.05
heatmap_cluster_cols: 0
heatmap_display: log2FC

Configuration

## artMS

## Analytical R Tools for Mass Spectrometry

http://artms.org

Quality control

Functional analysis

Miscellaneous
config yaml
les :
evidence : '~/ph_ms/evidence.txt'
keys : '~/ph_ms/keys.txt
contrasts : '~/ph_ms/contrast.txt
qc :
basic: 1
extended: 1
data:
enabled: 1
fractions:
enabled: 0
sitac:
enabled: 0
filters:
enabled: 1
contaminants: 1
protein_groups: remove
modifications: PH
sample plots: 1
msstats:
enabled: 1
msstats_input:
profilePlots: none
normalization method: equalizeMedians
normalization reference:
summaryMethod: TMP
censoredInt: NA
cutoffCensored: minFeature
MBimpute: 1

```
output_extras:
enabled: 1
annotate:
```

enabled: 1
specie: HUMAN
plots:
volcano: 1
heatmap: 1
LFC: -1.51 .5
FDR: 0.05
heatmap_cluster_cols: 0

Configuration
artMS
Analytical R Tools for Mass Spectrometry
http://artms.org
Bioconductor
$M_{0}$ evidence.txt
keys.txt

Quality control artmsQualityControlEvidenceExtended()<br>artmsQualityControlSummaryExtended()

## Analytical R Tools for Mass Spectrometry

$~ / \sim$
> library(artMS)
> artmsQualityControlEvidenceBasic(evidence_file = "evidence.txt", keys_file = "keys.txt")

QUALITY CONTROL ------------------
>> LOADING FILES
> MERGING FILES
It might take a long time
(depending on the size of the evidence file)
>> GENERATING THE INTENSITY DISTRIBUTION PLOTS
> CONTAMINANTS CON_IREV__ REMOVED
>> GENERATING THE REPRODUCIBILITY PLOTS
(Warning: it might take some time)
|=================================================1100\%
$\gg$ GENERATING CORRELATION MATRICES
--- NO Technical Replicates detected
--- By Biological replicates
--- By Conditions
>> GENERATING INTENSITY STATS PLOTS
--- AB PROCESSED
>> BASIC QUALITY CONTROL ANALYSIS COMPLETED!

Input Files

## keys.txt

artmsQualityControlEvidenceBasic()
Quality control

Analytical R Tools for Mass Spectrometry
http://artms.org
Bioconductor

## $\operatorname{artMS}$

- Reproducibly
- MS Intensity
- Spectral Counts
- Contaminants
- PTMs





## Analytical R Tools for Mass Spectrometry

~/ $\#$
> artmsQualityControlEvidenceExtended(evidence_file = "evidence.txt",
$+$
>>EXTENDED QUALITY CONTROL ANALYSIS (evidence.txt based)-
>> MERGING FILES
It might take a long time
(depending on the size of the evidence file)
>> GENERATING QC PLOTS
--- Plot PSM done
--- Plot IONS done
--- Plot TYPE done
--- Plot PEPTIDES done
--- Plot PROTEINS done
--- Plot Plot Ion Oversampling done
--- Plot Charge State done
--- Plot Mass Error done
--- Plot Mass-over-Charge distribution done
--- Plot Peptide Intensity CV done
--- Plot Peptide Detection (using modified.sequence) done
--- Plot Protein Intensity CV done
--- Plot Protein Detection done
--- Plot ID overlap done
--- Plot Inter-Correlation done
--- Plot Sample Preparation done

## $\operatorname{artMS}$

Analytical R Tools for Mass Spectrometry


Quality control
artmsQualityControlEvidenceExtended()

- Charge State Distribution
- ID overlap
- Peptide Ion Statistics
- Precursor Mass Error
- Precursor m/z error
- Frequency Peptide Detection
- Peptide Intensity CV
- Peptide Statistics
- Peptide Ion oversampling
- Protein detection frequency
- Protein Statistics / CV
- Peptide-spectrum matches
- MaxQuant Type statistics

artMS
Analytical R Tools for Mass Spectrometry
http://artms.org
Bioconductor

Quality control
artmsQualityControlEvidenceBasic()
artmsQualityControlEvidenceExtended()
artmsQualityControISummaryExtended()
artmsQuantification()

Miscellaneous
mo evidence.txt
keys.txt
artmsEvidenceToSAINTq()
artmsEvidenceToSaintExpress()
artmsAnalysisQuantifications()
artmsPhosfateOutput()
artmsPhotonOutput()
etc

artmsQuantification()



Bio
Bioconductor
http://msstats.org/

MSstats is an open-source $R$ package for statistical relative quantification of proteins and peptides in global, targeted, and data-independent proteomics.

It uses a family of linear mixed models that attempts to:

- minimize bias and inefficiencies in spectrometry-based proteomics,
- distinguish the systematic variation from random artifacts,
- maximize the reproducibility of the results

Developed by Meena Choi and Olga Vitek
Used in fllt Skyline

| Analytical R Tools <br> for Mass Spectrometry | keys.txt |
| :---: | :---: |
| http:///artms.org | contrast.txt |

Relative
quantification

Mo. evidence.txt
keys.txt
contrast.txt
config.yaml

- quantification-results.txt
- quantification-results-annotated.txt
- Normalized abundance
* results_ModelQC.txt
* results_RunlevelData.txt
* results-mss-groupQuant.txt
* results-mss-normalized.txt
* results-mss-sampleQuant.txt
- results_sampleSize.txt
- results_experimentPower.txt

PDF FILES

- results-heatmap.pdf
- results-peptidecounts-perBait.pdf
- results-peptidecounts.pdf
- results-sign.pdf
- results-volcano.pdf


Input Files
$\operatorname{artMS}$

## Analytical R Tools for Mass Spectrometry <br> http://artms.org <br> keys.txt <br> contrast.txt <br> config.yaml <br> Mo evidence.txt

Annotations
Summary files in different format (xls, txt) and shapes
Numerous summary plots
Enrichment analysis using Gprofiler
PCA of protein abundance
PCA of quantifications
Clustering analysis
Functional analysis artmsAnalysisQuantifications()

-results-summary.xlsx
-results-abundance-long.txt -results-abundance-wide.txt results-enrich-MAC-allsignificants-corum.txt -results-enrich-MAC-allsignificants.txt -results-enrich-MAC-negatives-corum.txt -results-enrich-MAC-negatives.tx -results-enrich-MAC-positives-corum.txt -results-enrich-MAC-positives.txt -results-log2fc-long.txt -results-log2fc-wide.txt -results.log2fc-clusterheatmap-enriched.txt -results.log2fc-clusterheatmap.txt
-results-enrich-MAC-allsignificants-corum.pdt results-enrich-MAC-negatives-corum.pdf -results-enrich-MAC-positives-corum.pdf -results-pca-correlations.pdf -results-pca-pca01.pdf -results-pca-pca02.pdf -results-pca-pca03.pdf
-results.clustering.abundance.all-overview.pdf -results.clustering.abundance.all-zoom.pd $\ddagger$ -results.clustering.log2fc.all-overview.pdf -results.clustering.log2fc. all-zoom.pdf -results.clustering.log2fcSign.all-overview.pdf -results.clustering.log2fcSign.all-zoom.pdf -results.distributions.pdf
-results.distributionsFil.pdf
-results.imputation.pdf -results.log2fc-clusterheatmap.pdf -results.log2fc-clusters.pdf -results.log2fc-corr.pdf results.log2fc-dendro.pdf -results.log2fc-individuals-pca.pdf -results.TotalQuantifications.pdf -results.correlationConditions.pdf -results.correlationQuantifications.pdf -results.relativeABUNDANCE.pdf -results.reproducibilityAbundance.pdf


Summary files in different format (x|s, txt) and shapes Numerous summary plots
Enrichment analysis using Gprofiler
PCA of protein abundance
PCA of quantifications
Clustering analysis
Functional analysis artmsAnalysisQuantifications()



## artMS

Analytical R Tools for Mass Spectrometry
http://artms.org Bioconductor

## available at:

Bioconductor
https://bioconductor.org/packages/release/bioc/html/artMS.html
( GitHub https://github.com/biodavidjm/artMS

## 気 Documentation: http://artms.org



## UCSF

University California San Francisco


Nevan Krogan Danielle Swaney Jeff Johnson Ruth Huttenhain John Von Dollen


Cyto
GLADSTONE INSTITUTES

Alex Pico


Sara Jimenez-Lopez

