MaxQuant data analysis in Galaxy

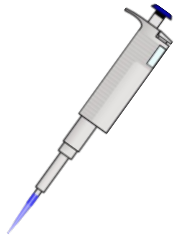
Schedule

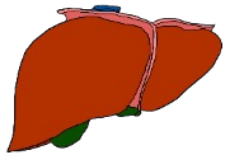
- Demo: Using MaxQuant in Galaxy
- **Hands-on**: Label-free proteomics analysis from skin cancer samples

Hands-on analysis / homework

- Tutorial: MaxQuant and Msstats for the analysis of label-free data https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/maxquant-label-free/tutorial.html
- Share the Galaxy History:
  - Gear wheel on the upper right corner of the history
  - share or publish
  - Enable: make history accessible
  - Copy the link that is shown below

Sample preparation | Mass spectrometry | Data analysis





Protein extraction

Protein(s)

Reduction& Alkylation

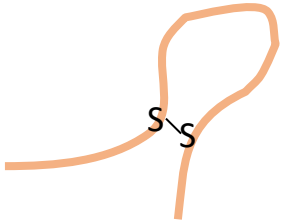
Linear Protein(s)

Tryptic digestion

Peptides

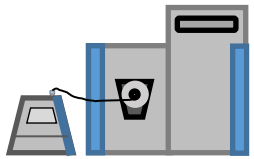
Desalting Drying

Peptide Pellet



DAHSFGBKDFBVKNXSRMGLX

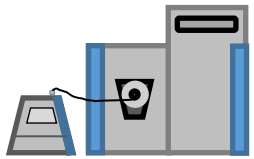
DAHSFGBK
DFBVKNXSR
MGLX



Peptide Pellet

Solving in acidic buffer

Liquid Chromato-graphy (LC) → Ion source → Mass Analyzer → Detector

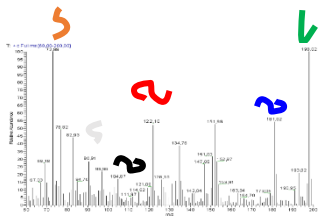


Tandem mass spectrometry (LC-MS/MS) using data dependent acquisition (Top3)

MS1: Mass spectra

MS2: Mass spectra

LC + Ion source

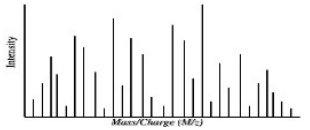


Filter for most abundant m/z

Fragmentation of filtered m/z

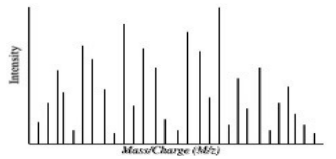
Filter for 2nd most abundant m/z

Fragmentation of filtered m/z



Filter for 3rd most abundant m/z

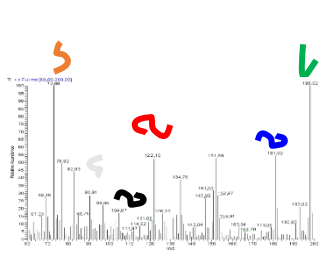
Fragmentation of filtered m/z



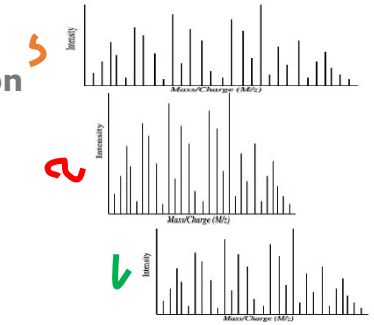
Next MS1 spectrum etc…

MS1: Mass spectra

MS2: Mass spectra



Fragmentation (top 3)

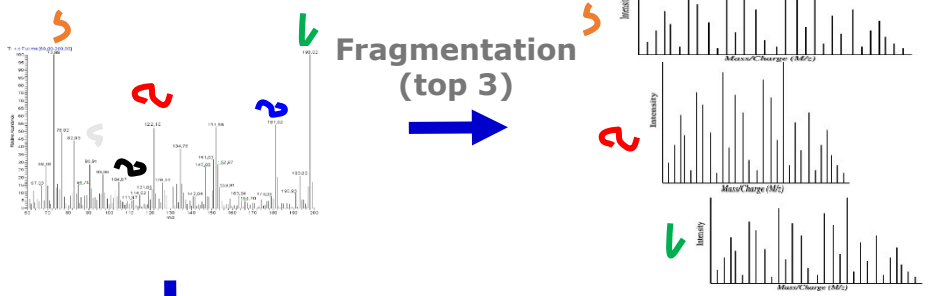


Peptide quantification

Peptide identification

MS1: Mass spectra

MS2: Mass spectra



Peptide quantification

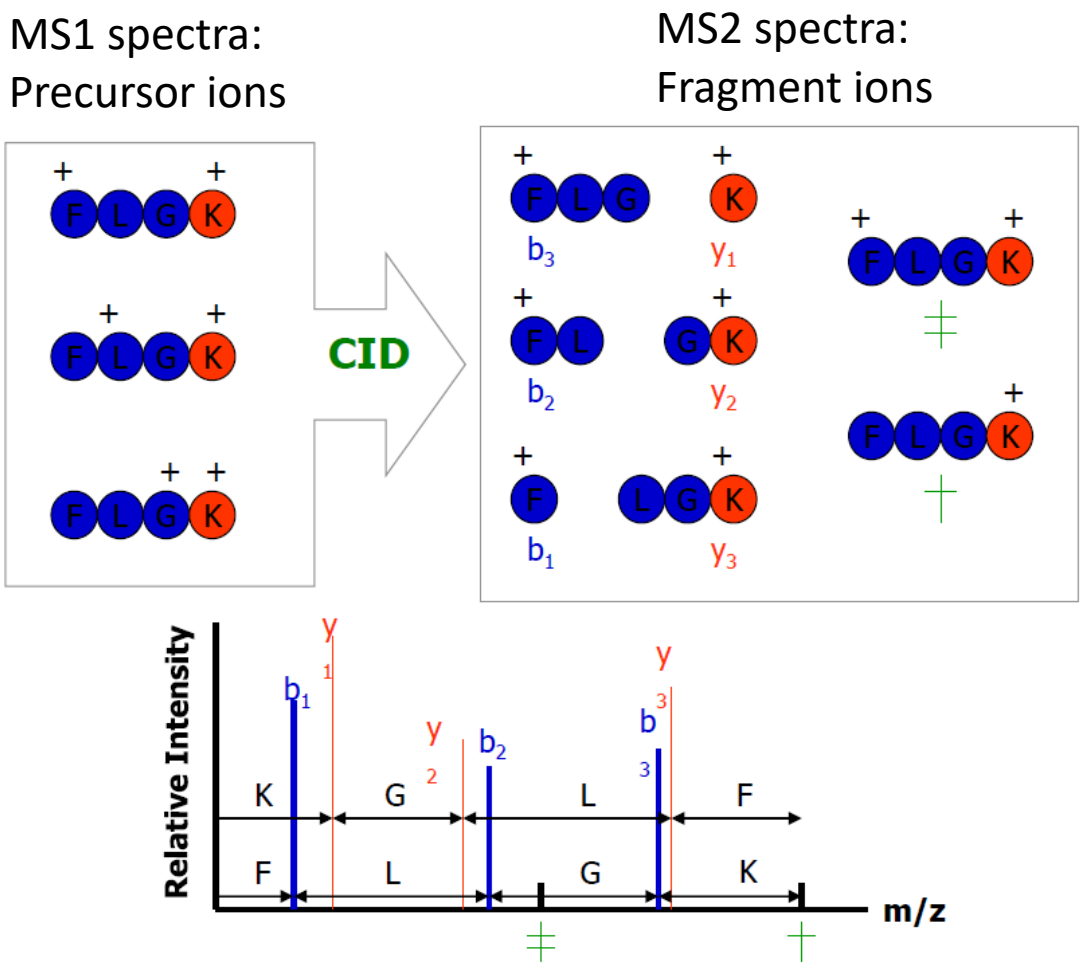
Peptide identification

Protein identification

Protein quantification

Statistical Analysis

Peptide identification with MS2 fragment spectra



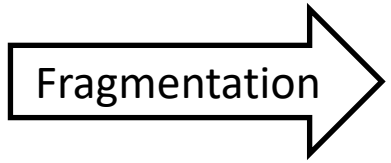
Identification options:

1) Manual interpretation
2) De novo sequencing
3) Matching to in silico spectra (generated from fasta database)

Peptide identification with MS2 fragment spectra

MS1 spectra: Precursor ions

V-H-L-T-P-E-E-K



MS2 spectra: Precursor ions

b-ions	y-ions	
b7V-H-L-T-P-E-E	K	y1
b6V-H-L-T-P-E	E-K	y2
b5V-H-L-T-P	E-E-K	y3
b4V-H-L-T	P-E-E-K	y4
b3V-H-L	T-P-E-E-K	y5
b2V-H	L-T-P-E-E-K	y6
b1V	H-L-T-P-E-E-K	y7

Peptide identification with MS2 fragment spectra

MS1 spectra: Precursor ions

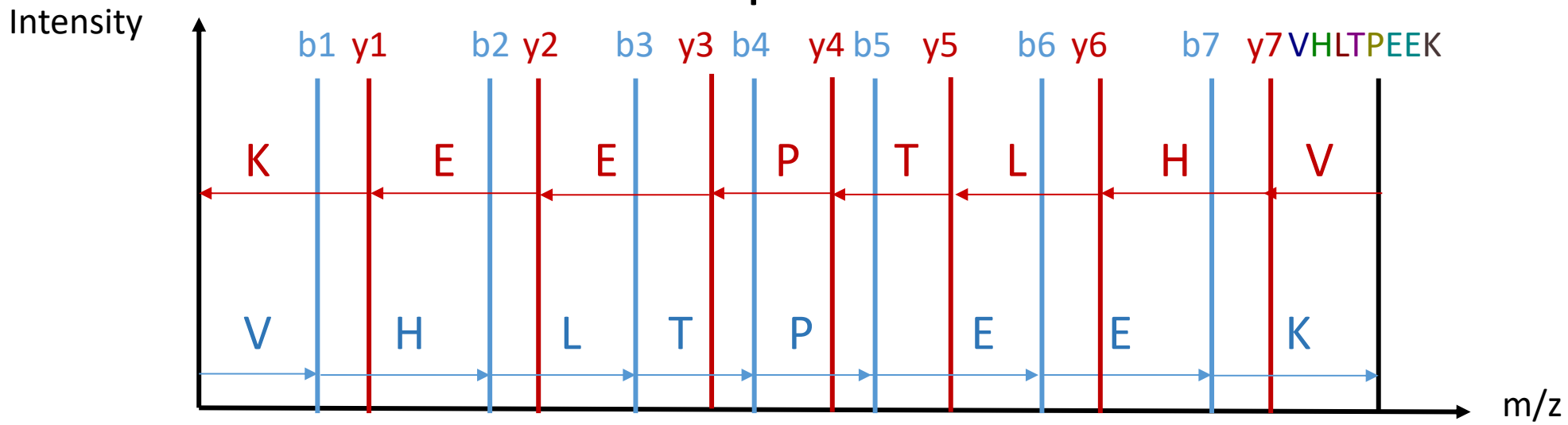
V-H-L-T-P-E-E-K

Fragmentation

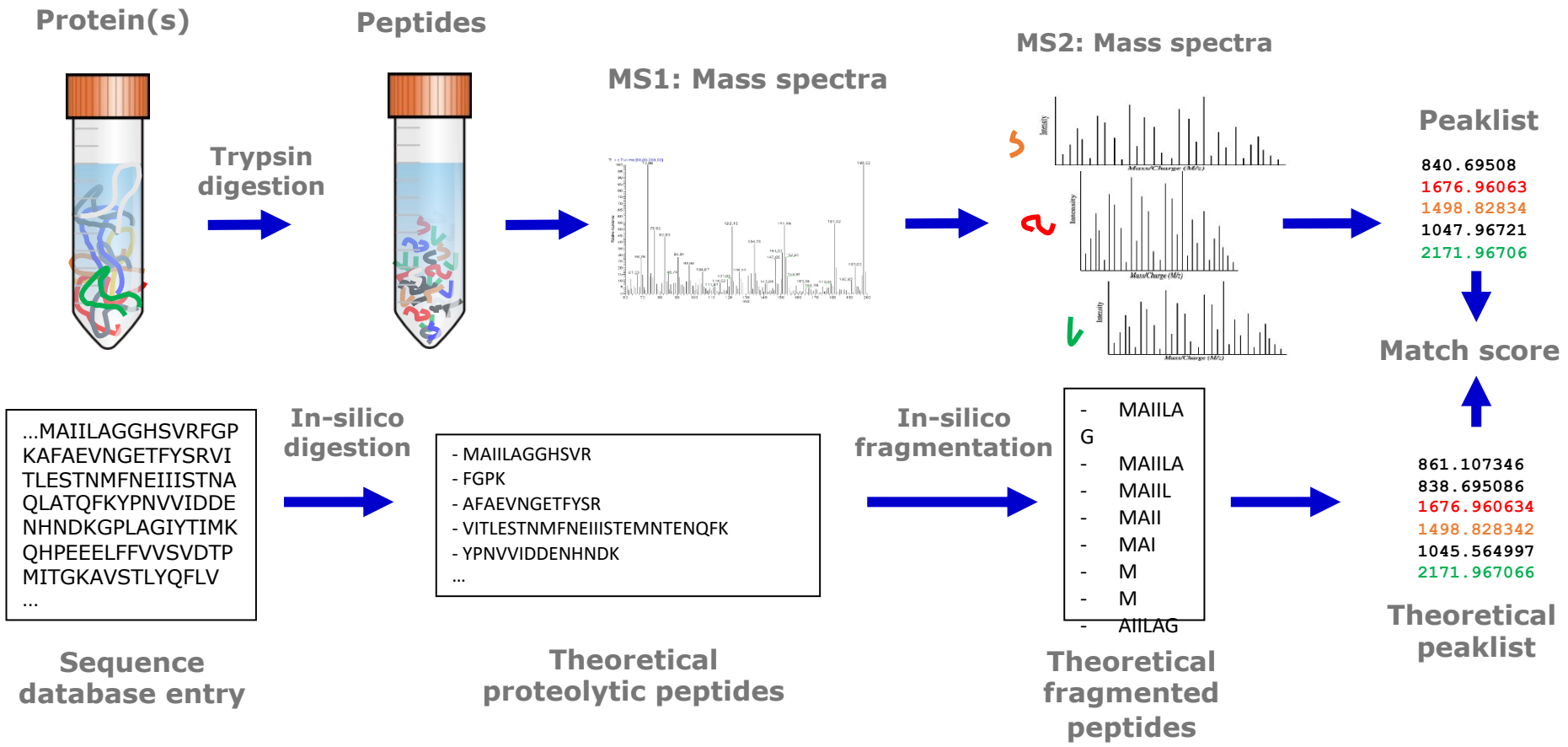
MS2 spectra: Precursor ions

b7	V-H-L-T-P-E-E	K	y1
b6	V-H-L-T-P-E	E-K	y2
b5	V-H-L-T-P	E-E-K	y3
b4	V-H-L-T	P-E-E-K	y4
b3	V-H-L	T-P-E-E-K	y5
b2	V-H	L-T-P-E-E-K	y6
b1	V	H-L-T-P-E-E-K	y7

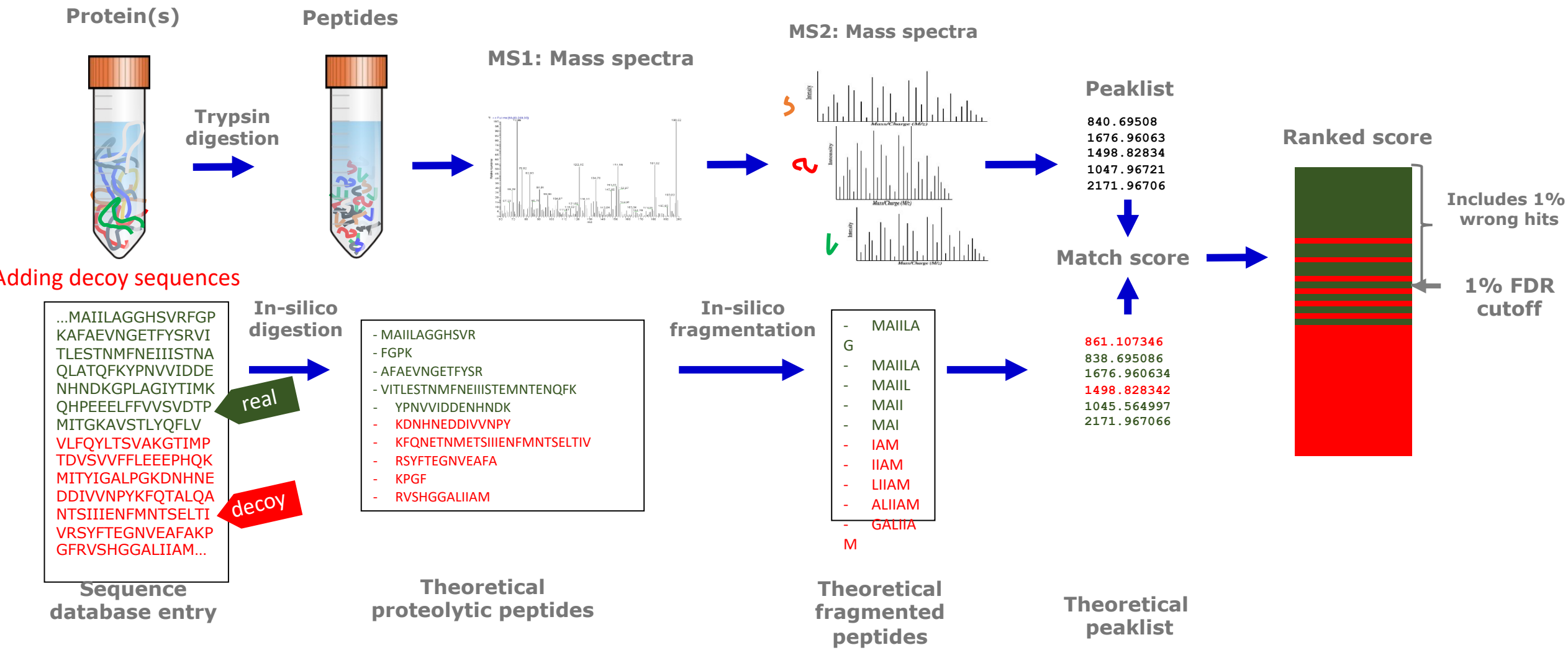
MS2 spectrum



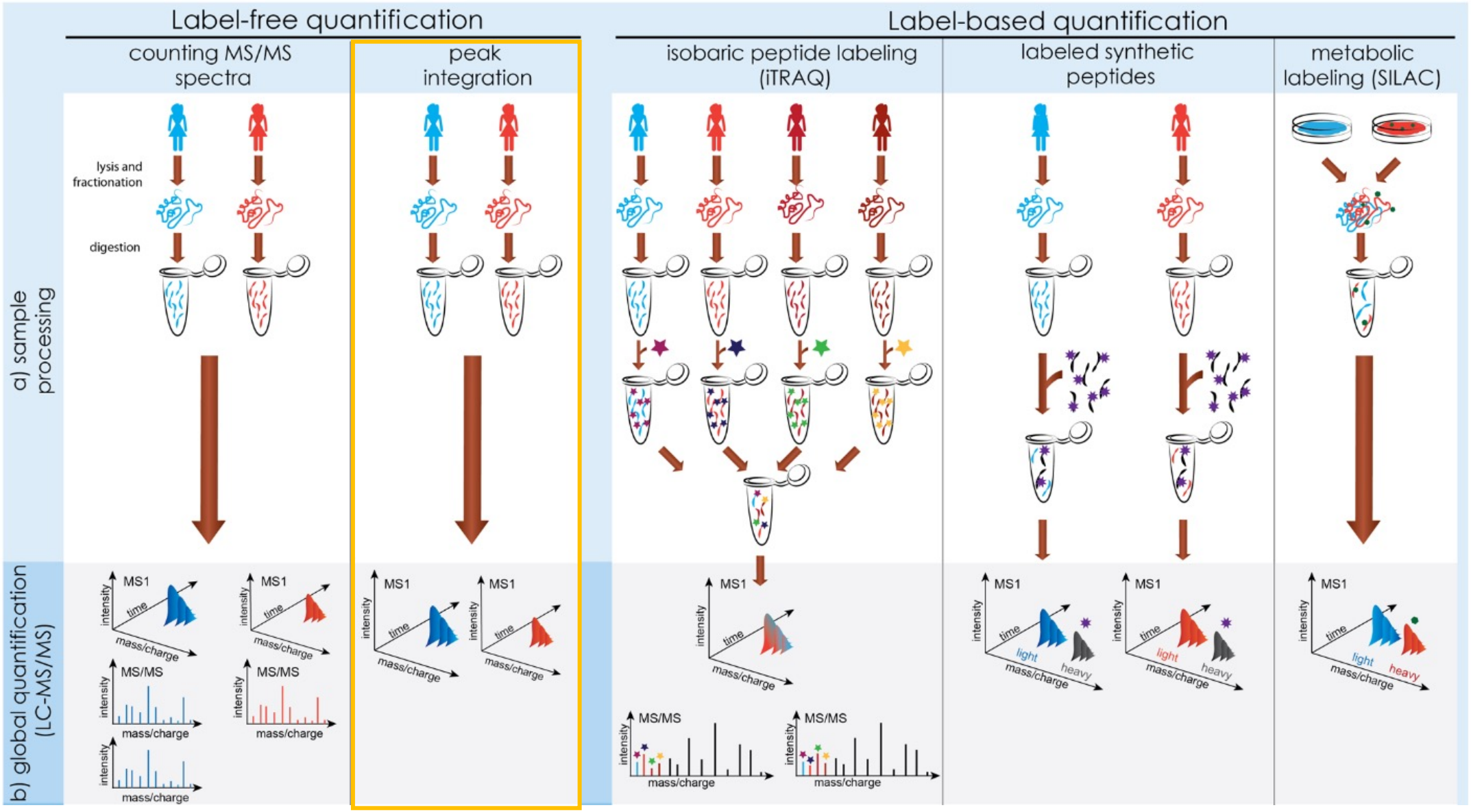
Peptide identification via an in-silico database



Peptide identification via an in-silico database

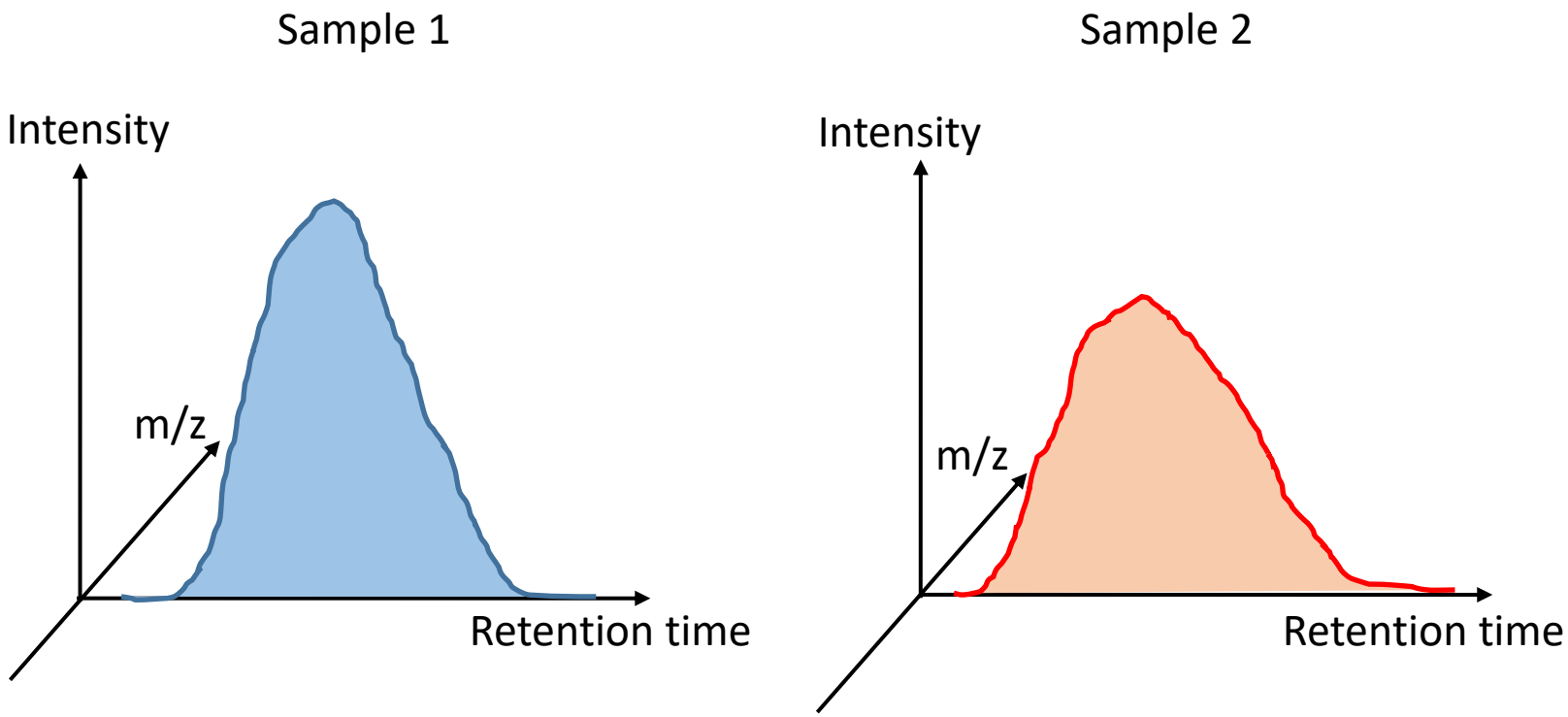


Quantification methods in proteomics



Label-free peptide quantification

MS1 spectra: for each peptide feature peak quantify area under curve to obtain peptide abundance



MaxQuant software

Published: 30 November 2008

MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification

Jürgen Cox & Matthias Mann

Nature Biotechnology **26**, 1367–1372(2008) | Cite this article

Freeware, “ Black box“
Popular non-commercial proteomics software
MaxQuant videos on youtube:
https://www.youtube.com/channel/UCKYzYTm1cnmc0CFAMhxDO8w

- **Raw data import**
- **Protein Identification:**
  - Andromeda Search Engine
  - Speciality: Match-between-runs
- **Protein Quantification:**
  - Label-free
  - Label-based (SILAC, Dimethyl, …)
  - Reporter ion MS2 (TMT, iTRAQ)

MSstats software

MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments FREE

Meena Choi, Ching-Yun Chang, Timothy Clough, Daniel Broudy, Trevor Killeen, Brendan MacLean, Olga Vitek ✉ Author Notes

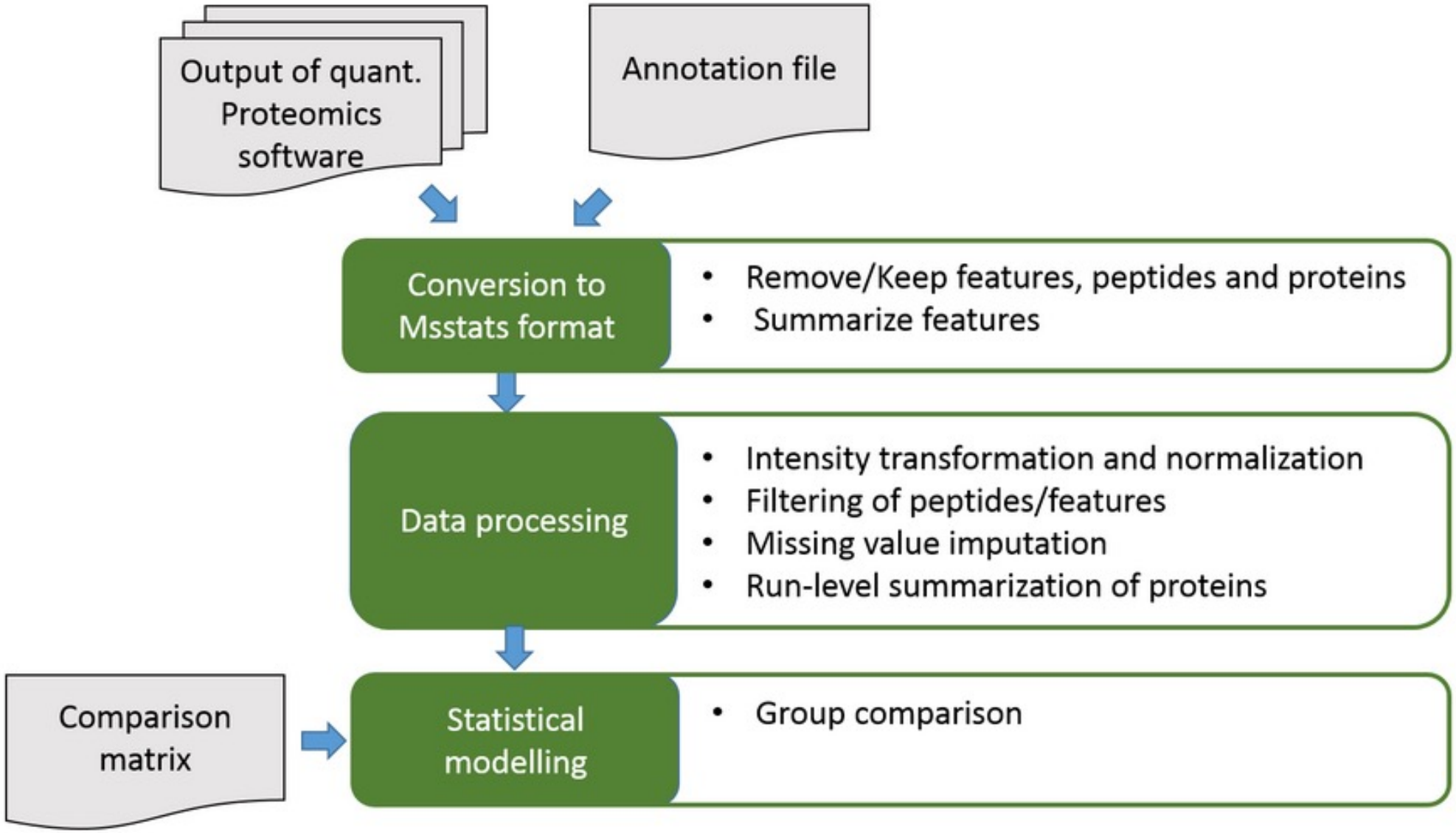
Bioinformatics, Volume 30, Issue 17, 1 September 2014, Pages 2524–2526, https://doi.org/10.1093/bioinformatics/btu305

Published: 02 May 2014 **Article history** ▾

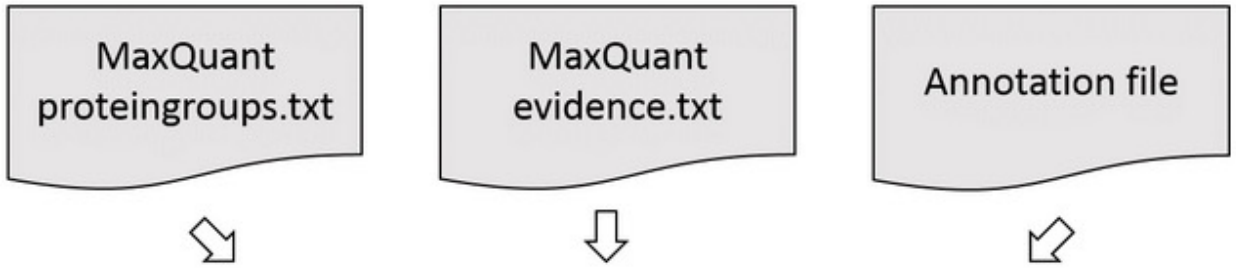
- **Import results from common proteomics software:**
  - MaxQuant, Skyline, OpenMS
- **Transformation and normalization**
- **Missing value imputation**
- **Protein summarization and quantification**
- **Statistical modelling**
  - linear models to detect differentially abundant proteins in label-free and isobaric labeled experiments

Two Bioconductor R packages: Msstats & MSstatsTMT
Popular open-source statistical proteomics software
MSstats videos on youtube:
https://www.youtube.com/c/MayInstituteNEU/search?query=msstats

Statistical Analysis with MSstats



Conversion from MaxQuant to Msstats format



Protein Name	Peptide Sequence	Precursor Charge	Fragm entIon	Product Charge	Isotope LabelType	Condition	Bio Replicate	Run	Intensity
ProteinA	TPAVLK	3	NA	NA	L	Cond1	1	1	2636791
ProteinA	TPAVLK	3	NA	NA	L	Cond1	2	2	5019594
ProteinA	TPAVLK	3	NA	NA	L	Cond1	3	3	4560462
ProteinA	TPAVLK	3	NA	NA	L	Cond2	4	4	2918293
ProteinA	TPAVLK	3	NA	NA	L	Cond2	5	5	4534487

Peptide filtering and missing value imputation

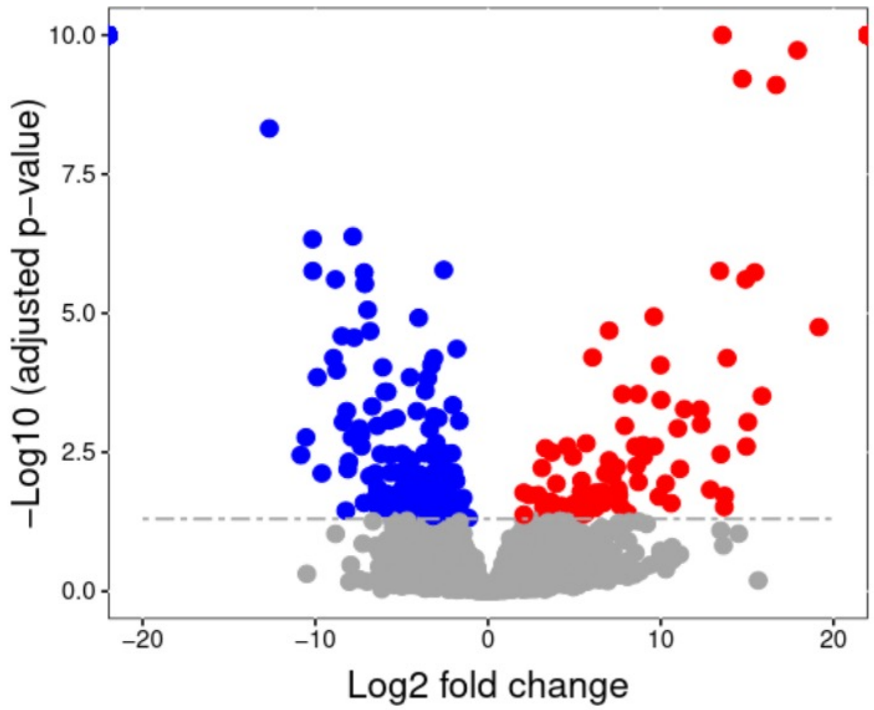
- **Feature selection**: which peptides should be kept for protein quantification
- **Missing value imputation**: Imputation of NA and very low abundant intensities
- **Protein summarization**: calculates new protein abundances after data processing via TMP model

	Condition1			**Condition2**		
	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6
Protein1	Run1	Run2	Run3	Run4	Run5	Run6
Peptide1						
Peptide2						
Peptide3						
Peptide4						
Peptide5						

Statistical modelling

- Uses run-level summarized data for hypothesis testing
- Needs comparison matrix to specify comparisons
- Adjusts the linear model according to information from annotation file

name	Cond1	Cond2	Cond3	Cond4
cond1-cond3	1	0	-1	0
cond1-cond4	1	0	0	-1
cond2-cond3	0	1	-1	0
cond2-cond4	0	1	0	-1

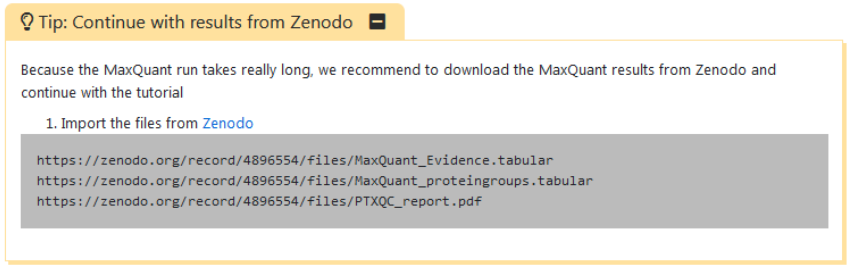


Hands-on

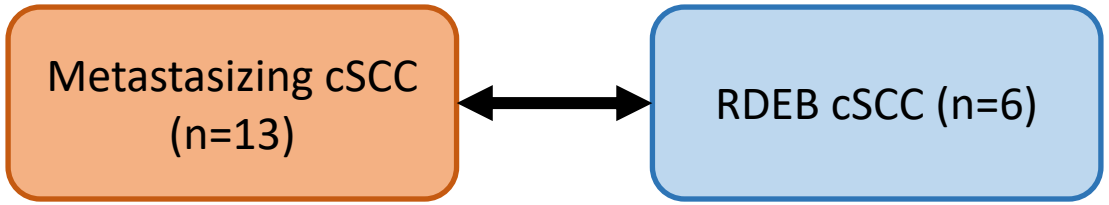
Tutorial: MaxQuant and Msstats for the analysis of label-free data
https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/maxquant-label-free/tutorial.html

Video with demonstration of tutorial in youtube: https://www.youtube.com/watch?v=IXdLAt2PAT4
A few things might be outdated however, please stick to the recommendations of the training material which is up to date

In the first hands-on skip step 4 to 7 (don‘t load the raw data from PRIDE)
Instead of running MaxQuant, load the MaxQuant result files from Zenodo.
The links can be obtained by opening the box:
Continue with the hands-on: MSstats Analysis



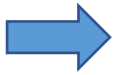
Aim: Find proteins that are differentially abundant between two different types of skin cancer: metastasizing cutaneeous squamous cell carcinoma (cSCC) and recessive dystrophic epidermolysis bullosa cSCC



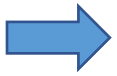
Per patient: 1-2 10µm slices of FFPE tissues

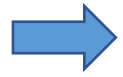
HANDS-ON SESSION

Instructions








Please **Register** for creating an account with a valid email ID and Password at usegalaxy.eu.

Once Registered, click on TIAAS to join the GCC 2022 Galaxy session. https://usegalaxy.eu/join-training/quant-meta

Go to Shared Data

Go to Shared Data

Run the workflow on active history

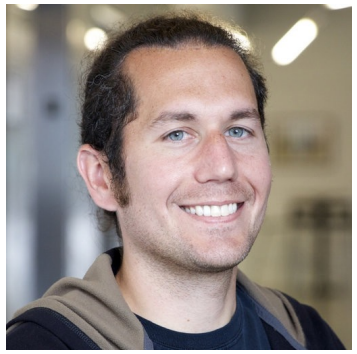
Further resources

- Summary of proteomics data analysis: https://www.youtube.com/watch?v=2C96AvrFT38
- More proteomics tutorials in Galaxy: https://training.galaxyproject.org/training-material/topics/proteomics
- UC Davis Proteomics: https://video.ucdavis.edu/playlist/details/0_4jkc4swu
- Global online Galaxy course in March (much more than proteomics) https://gallantries.github.io/posts/2021/12/14/smorgasbord2-tapas/

Cloud Computing Workshop (2022)

The iPRG will conduct a series of online video tutorials about the use of cloud computing resources for MS-based proteomics, focusing on Nextflow, the Trans-Proteomic Pipeline (TPP) and Galaxy Platform.

September 2022



Trans Proteomic Pipeline

Michael Hoopmann - *ISB, Seattle, WA*

- Instructions on how to use TPP to analyze MS data.
- Answer questions from the participants

October 2022

Galaxy

Melanie Foell - *Freiburg University, Germany*

- Instructions on how to use Galaxy to analyze MS data.
- Answer questions from the participants

November 2022

nextflow

Yasset Perez-Riverol - *EBI, Hinxton, UK*

- Instructions on how to use Nextflow to analyze MS data.
- Answer questions from the participants