HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

STEPUPIORS

Amyloid typing by mass spectrometry

Dr Julie Courraud

2nd of June 2023

J. Courraud, P.E. Nikolaou, D. Canetti, G. Médard, J. Gilbertson, A. Wechalekar,
I. Andreadou, N. Thomaidis, E. Kastritis

1

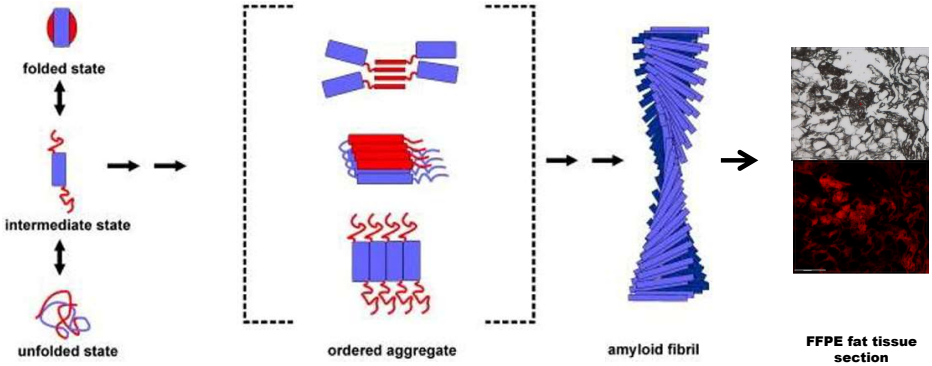
I will talk about…

1. Intro on amyloidosis and why we need this assay
2. Sample workflow and modifications
3. Results: DDA vs DIA – data complexity
4. Why you might want to divorce Excel
5. Take home messages

2

Amyloidosis is the disease of misfolded proteins

In amyloidosis, misfolded proteins of **different origin** lead to the production of **amyloid fibrils** which are **deposited in the extracellular matrix of various tissues**.



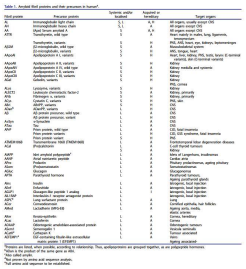
Modified from Iannuzzi C, et al., Front Mol Biosci. 2014;1:9

Typing of amyloid is critical to define treatment

- Symptoms may be subtle and non-specific.
- The diagnostic can take several years.

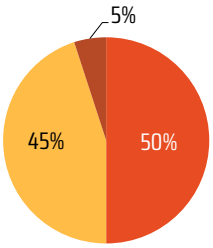
- Biopsies from **subcutaneous fat aspirates** or target organ
  ➔ Amyloid typing
  = recognition of the **precursor protein** causing the formation of amyloid fibrils

42 proteins



Wrong typing = wrong treatment

Amyloid types at the Referral center in Athens



AL (immunoglobulin light chain)
ATTR (transthyretin)
other types

Department of Clinical Therapeutics, Alexandra Hospital, Athens, 2023

100 new patients every year

- A patient with ATTR should receive a **liver transplant**.
- A patient with AL should receive **chemotherapy.**

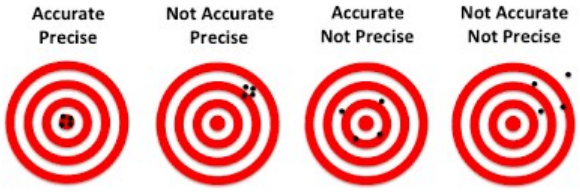
5

CONTEXT

Diagnostics ≠ Research

- **The main research behind this assay has been done already.**
- **We are not comparing individuals (yet).**
- We want the assay to be
  - **Robust** = will not fail easily
  - **Sensitive** = will not miss a positive patient
  - **Specific** = will not mistake a negative patient for a positive patient
  - **Repeatable / reproducible** = will give consistent results for the same samples (on the same day / different days)
  - **Precise & Accurate**

Quantitation is difficult



6

2. Sample workflow & modifications

- Proteomic typing of amyloid protein pioneered at the Mayo Clinic, Rochester, USA (Vrana, Julie A., et al. *Blood*, 2009).
- Few centers in the world
- Collaboration with the National Amyloidosis Center in London, UK (Canetti, Diana, et al. *Molecules*, 2021)

1,200-1,500 USD

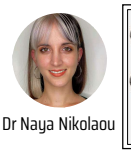
Bruker timsTOF flex coupled with a nanoElute

Implementation + modification of method

- ✓ Gain of ~3 hours in sample prep+analysis time
- ✓ Double acquisition for now: DDA PASEF + DIA PASEF

7

Congo red staining



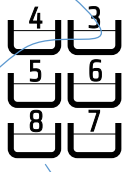
Dr Naya Nikolaou

Poly-L-lysine slides

6-µm sections of FFPE tissue

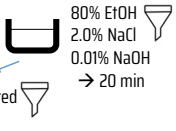
Drying 1 h at 37°C

Dewax in xylene
→ 2 x 5 min

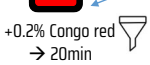


Hydrate in a gradient 100% to 50% EtOH in tap water
→ 2 min each

100% EtOH



80% EtOH
2.0% NaCl
0.01% NaOH
→ 20 min



+0.2% Congo red
→ 20min

Laser microdissection

Zeiss PALM MicroBeam

Fridge/Freezer

0.5-mL

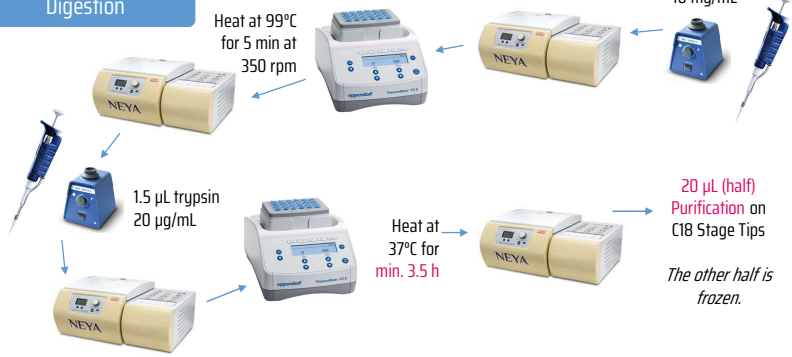
35 µL

Tris 10 mM
EDTA 1 mM
Zwittergent 3-16 0.002% w/v

8

Dissolution

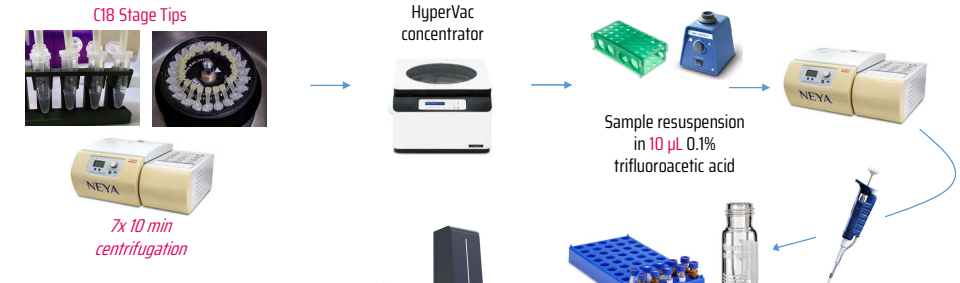
Digestion



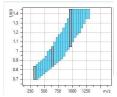
STEPUPIORS J. Courraud 02.06.2023

9

Purification & LC-MS/MS

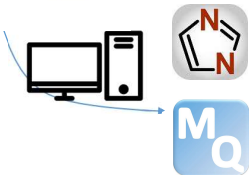


LC-MS/MS and TIMS analysis using a Bruker timsTOF flex coupled with a nanoElute:
DDA vs. DIA



12min run
A = 0.5% **AcOH*** in water
B = 0.5% **AcOH*** in acetonitrile

Data processing



Data preprocessing using DIA-NN & MaxQuant (identification with SwissProt Human database)

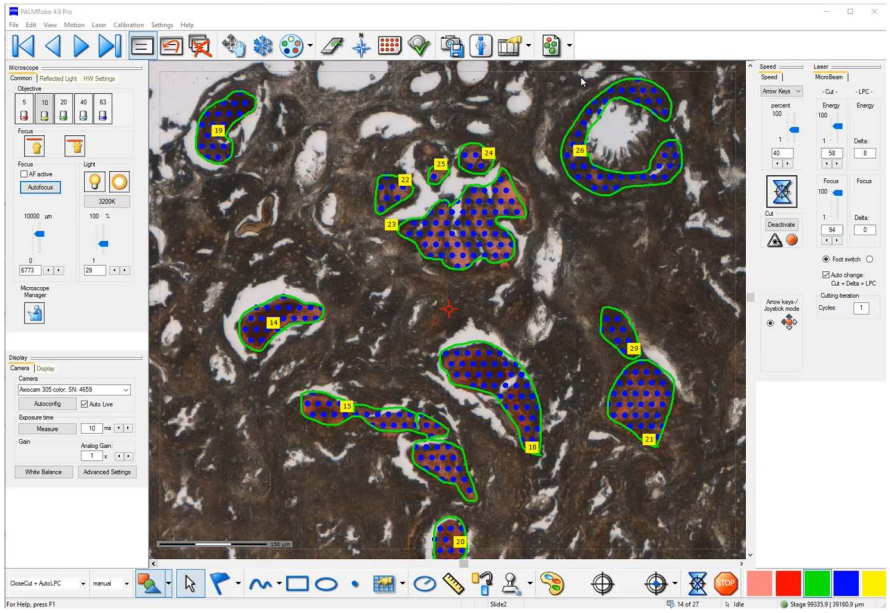
Data processing & reporting

Battellino et al., J. of Proteome Research, 2023

STEPUPIORS J. Courraud 02.06.2023

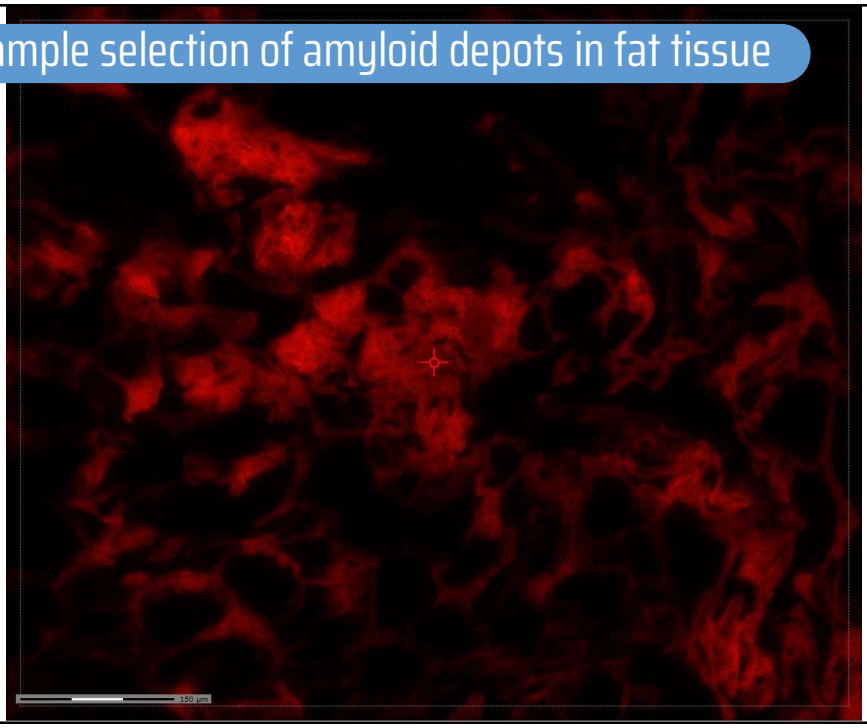
10

Video showing the dissection of amyloid depots in a renal biopsy



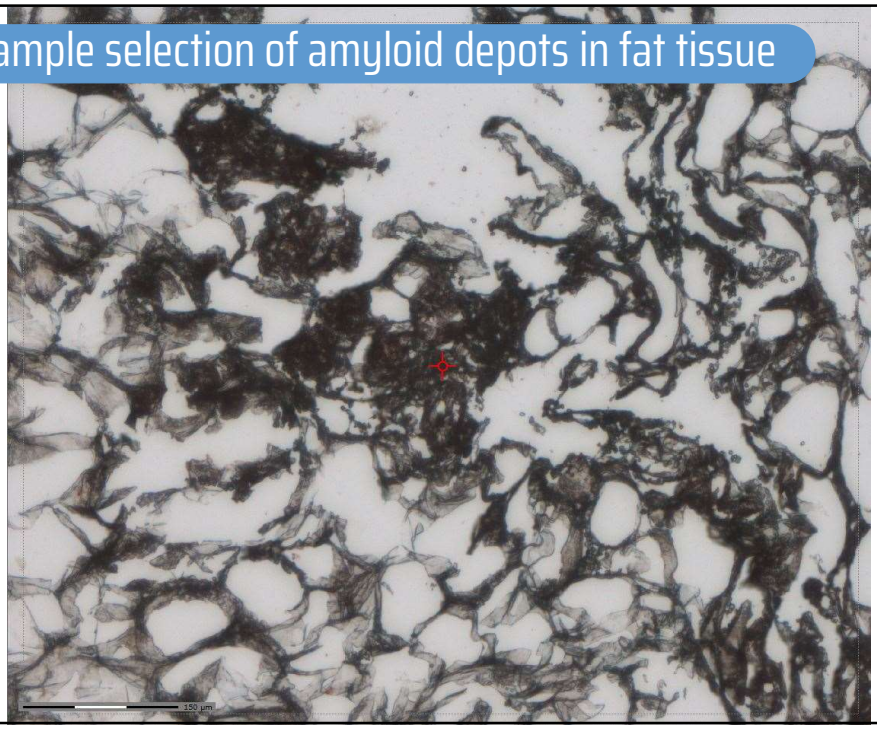
11

Example selection of amyloid depots in fat tissue



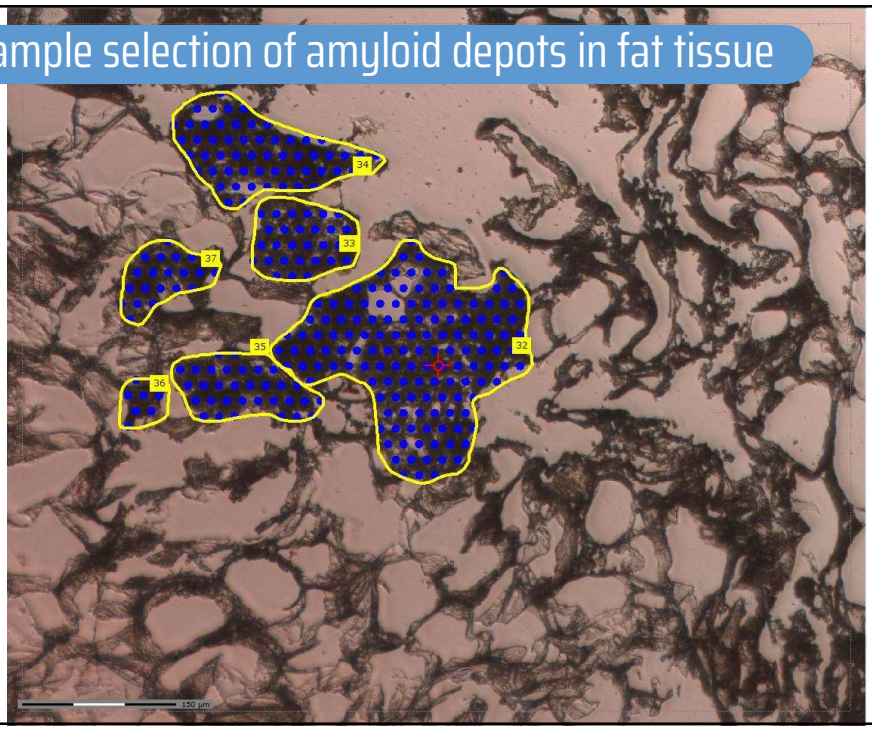
12

Example selection of amyloid depots in fat tissue



13

Example selection of amyloid depots in fat tissue



14

ShareX

To record videos of your screen

Shotcut

To cut videos

15

3. Analysis of 16 fat aspirate samples

- Cases of AL (Ig light chain kappa or lambda) amyloidosis randomly picked in the database
- 14 were analyzed in London in 2019.

Average values (n=14)

	# of proteins identified	# of 'relevant' proteins...	...of which of high confidence
London2019 DDA reprocessed	65	11	7*
Athens2023 DDA pasef	84+	11	8*
Athens2023 DIA pasef	485	28	21**

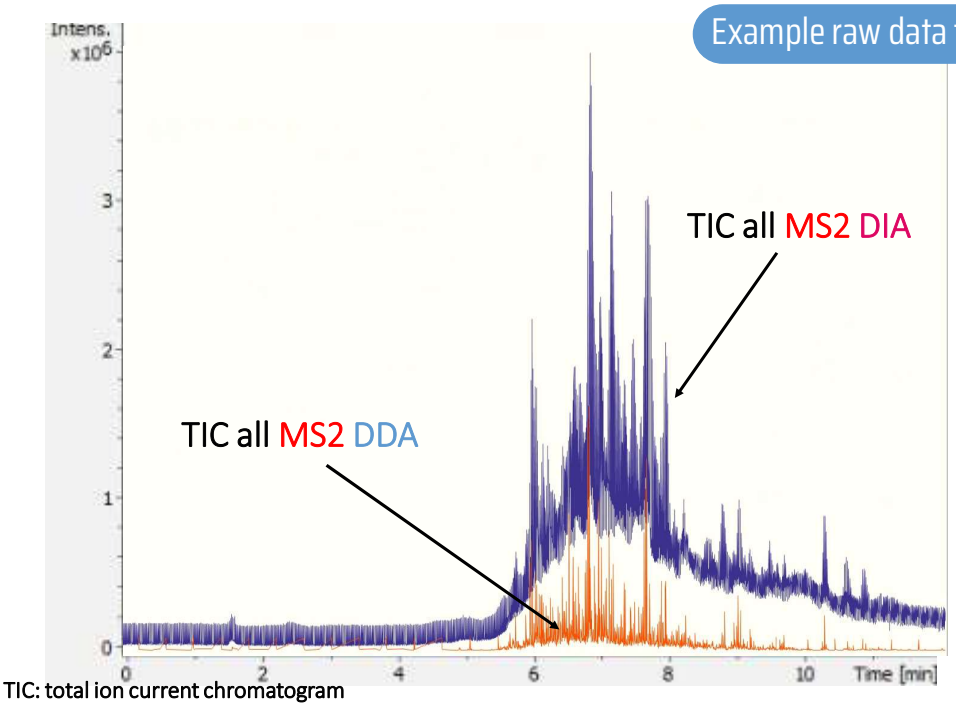
*for DDA: score >10 in MaxQuant
**for DIA: filtered at 1% FDR, using global q-values for protein groups and both global and run-specific q-values for precursors in DIA-NN

In fat / kidney biopsies:
- ❑ Signature SAP
- ❑ Signature APOE
- ❑ Signature APOA4
- ❑ Ig light chains λ / κ
- ❑ Fibrinogen A, B, G
- ❑ Serum amyloid A
- ❑ Transthyretin
- ❑ Other amyloidogenic proteins...
- ❑ Few proteins under investigation

216 genes for now

16

Example raw data file



TIC: total ion current chromatogram

17

London2019 DDA reprocessed	Athens2023 DDA pasef	Athens2023 DIA pasef
3 sig + lambda	same	same + kappa?
2 sig + lambda + kappa	kappa missing?	same but low lambda
3 sig + lambda	missing 1 sig	same + kappa?
3 sig + lambda	same	same
3 sig + lambda	same	same
Only ApoE	same	same
3 sig + lambda	same	same + kappa?
3 sig + lambda	same	same
No sig	same	same
3 sig + kappa	missing 1 sig	same
3 sig + lambda	same	same + kappa?
2 sig + kappa	same	same
1 sig + low lambda	same	same
2 sig + lambda	same	same + low kappa?
NA	kappa	kappa
NA	kappa	kappa

- Only one biological replicate
- Cutting pattern is important
- in DIA, I actually see all 3 signatures for all samples ➔ New diagnosis flowchart to develop for DIA with **adapted confidence and quantity thresholds**

18

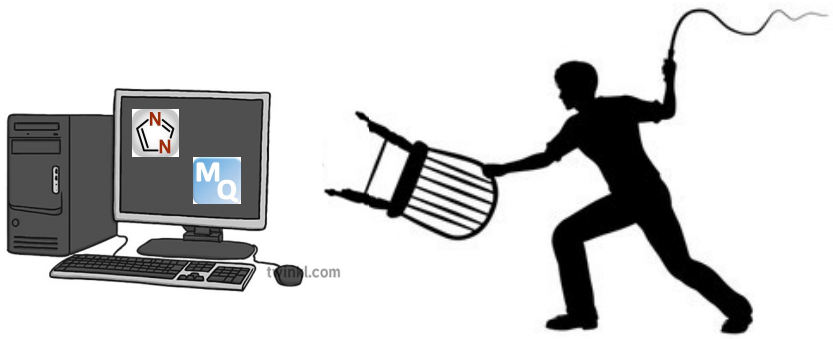
Some lessons…

- DIA acquisition is promising and allows to see peptides of lower abundance
  ➔ we need to get experience to define thresholds and appropriate confidence scores
- Cutting at several places + several biological replicates is important to get confidence in the diagnosis.
- Light chain sequences are easily missed / low confidence ➔ database expansion?

…and next steps

- **Optimizing data processing pipelines to allow better comparison between DDA and DIA outputs**
- Digging into the possibility to use an internal standard (non-human purified protein) and other proteases such as Glu-C
- Comparing results obtained from renal + fat biopsies for known patients (n=11+) and negative controls.

Taming software output



shutterstock.com

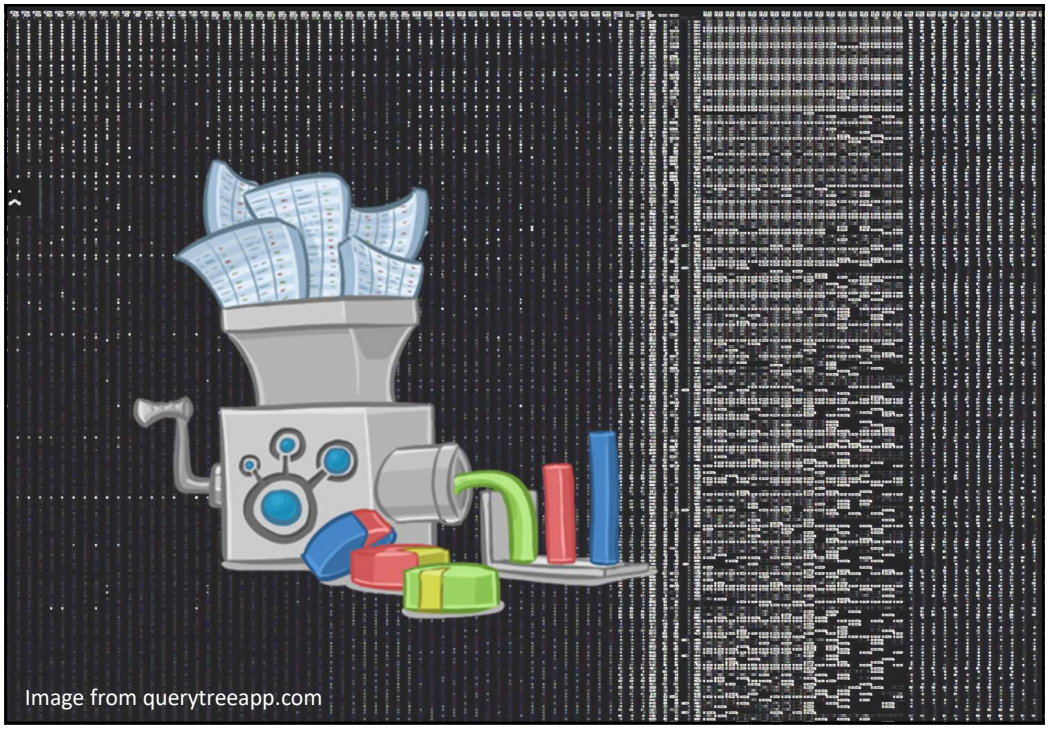
Software and data output for DDA (MaxQuant) and DIA (DIA-NN)

	MaxQuant	DIA-NN
• Summary	summary.txt **18** col. x **1**/s line	report.stats.tsv **51** col. x **1**/s line
• 'Protein' list	proteinGroups.txt **37+6**/s col. x **500-3,000** lines	report.pg_matrix.tsv **5+1**/s col. x **4,000** lines
• Peptide list	evidence.txt **73** col. x **5-10,000**/s lines	report.tsv **64** col. x **10-20,000**/s lines

Not completely equivalent!
Read the documentation, *again and again*,
and watch Youtube videos!

STEPUPIORS J. Courraud 02.06.2023

21



22

4. Excel and my data, an unstable relationship

- Quick look at data, filtering of columns super useful
- Searching specific entries
- Quick plots
- For repetitive actions: macros possible but difficult to edit

- Automatically changes some values/strings thinking it knows best – into dates especially
- Crops some values with 'scientific format' when exporting to .csv
- Bugs in some formulas
- Slow…. and cannot handle more than 1,048,576 rows by 16,384 columns
- Does not offer certain types of charts (boxplots, etc.)
- Copy-pasting errors

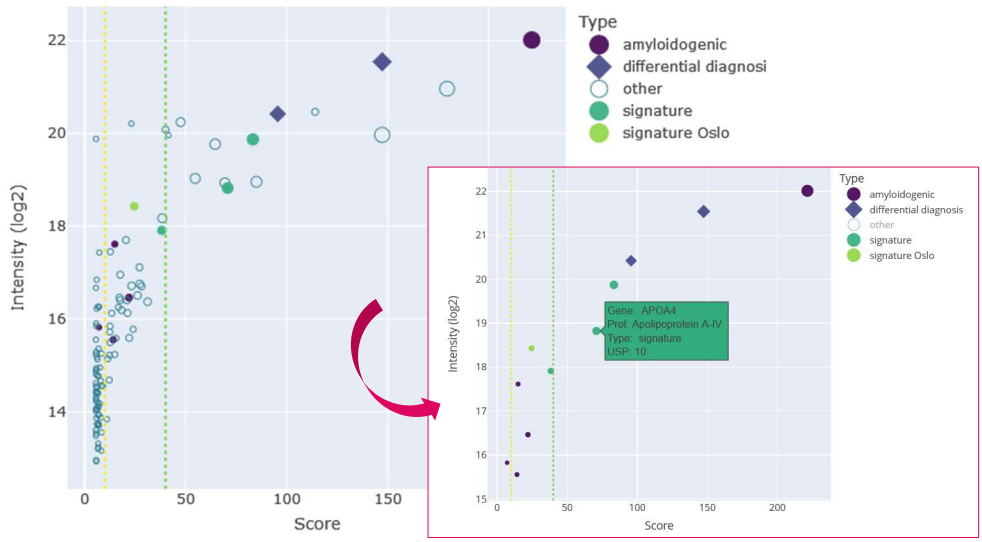
"What would you do differently?"

"I would learn to use R much earlier!"

Bjarke Feenstra
Senior Scientist in genomics and metabolomics
at SSI, Copenhagen

Processing in R markdown: example for DDA (MaxQuant) output

Html reports: Interactive plot using plotly R package



25

Processing in R markdown: example for DDA (MaxQuant) output

Html reports: Interactive table

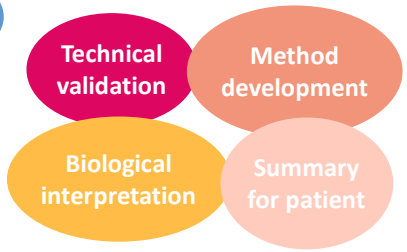
Search:

Type	Gene	Protein	Am. fibrils	Intensity	UP	Score	Seq. cov
signature	APOE	Apolipoprotein E [ApoE]		717 850	9	67.1	25.2
signature	APOA4	Apolipoprotein A IV, wild type	AApoAIV	600 110	7	49.2	15.7
signature	APCS	serum amyloid P component [SAP]		144 070	2	13.2	9
differential diagnosis	FGB	Fibrinogen beta chain		430 070	2	13.6	3.5
differential diagnosis	FGG	Fibrinogen gamma chain		177 930	6	37.1	12.8
amyloidogenic	FGA	Fibrinogen alpha, variants	AFib	462 220	12	82.7	15.6
amyloidogenic	APOA1	Apolipoprotein A I, variants	AApoAI	266 060	5	32.5	22.1
amyloidogenic	IGLC1	Immunoglobulin light chain lambda constant	AL (lambda)	139 960	2	16.6	23.6
amyloidogenic	LYZ	Lysozyme, variants	ALys	12 272	1	5.8	6.1

26

Creating interactive reports in 30 seconds

Define the objective & audience of the report

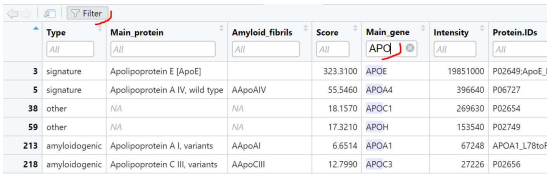


Example steps to visualize data

1. Reading data output files .txt or .tsv or else
2. Optional: create new 'tables' with the rows/columns you want in the order you want
3. Filter contaminants and/or extract peptides/proteins of interest
4. Compute some descriptive statistics: # of 'relevant' proteins, median scoring, modifications detected…
5. Generate plots of interest
6. Save your new data automatically

R studio user interface: in 1 min you can

- ✓ Visualize your table right away
- ✓ Filter the columns
- ✓ Check the dimensions of your data



	Type	Main_protein	Amyloid_fibrils	Score	Main_gene	Intensity	Protein.IDs
	All	All	All	All	APO	All	All
3	signature	Apolipoprotein E [ApoE]		323.3100	APOE	19851000	P02649;ApoE_I
5	signature	Apolipoprotein A IV, wild type	AApoAIV	55.5460	APOA4	396640	P06727
38	other	NA	NA	18.1570	APOC1	269630	P02654
59	other	NA	NA	17.3210	APOH	153540	P02749
213	amyloidogenic	Apolipoprotein A I, variants	AApoAI	6.6514	APOA1	67248	APOA1_L78toF
218	amyloidogenic	Apolipoprotein C III, variants	AApoCIII	12.7990	APOC3	27226	P02656

R is worth the time you spend learning it!

"I don't have the time, I'll go with what I know."

My advice: spend the time now. You will save a lot!

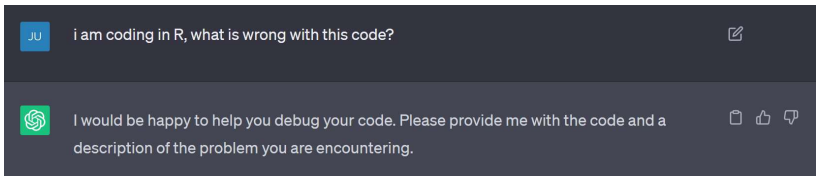
- ❑ Find an 'easy' task to do, but that is actually **useful/fun**.
- ❑ Great tutorial on internet or even mobile app. Workshops are the best to get launched.
- ❑ **Reproducibility**: makes great and reliable material to share when you publish your research, and to remember what you have done in the past.
- ❑ It might make the difference when you apply for a job…

image from @_NTTL

R is accessible to non-bioinformaticians

"I cannot code, it's too complicated."

- ❑ Many scripts are available online.
- ❑ The community is super friendly and reactive.
- ❑ You'll google everything in the beginning, **that's ok!**
- ❑ It doesn't matter if you need 10 steps to do something simple.
- ❑ chatGPT can help you find what's wrong in your code but be careful.



29

A few tips from experience

- ❑ Avoid '-' and spaces in your raw file names during acquisition
- ❑ Keep track of which versions of R/MaxQuant/DIA-NN you have used in a specific project
- ❑ Document your code with #, ideally in English directly!!!

```
#Searching Relevant Proteins in proteingroups.txt
protG <-  left_join(protG, relevant_prot, by = "Main_gene")
```

- ❑ If your code doesn't work, break it down to smaller pieces.
- ❑ Keep it readable.

```
protG <- protG_all[, c("Score","Intensity","Protein.IDs",
"Protein.names","Gene.names","Number.of.proteins","Peptides",
"Unique.peptides","Sequence.coverage....","Only.identified.by.site",
"Reverse","Potential.contaminant","id")]
```

```
protG <- protG_all[, c(
  "Score",
  "Intensity",
  "Protein.IDs",
  "Protein.names",
  "Gene.names",
  "Number.of.proteins",
  "Peptides",
  "Unique.peptides",
  "Sequence.coverage....",
  "Only.identified.by.site",
  "Reverse",
  "Potential.contaminant",
  "id"
  )]
```

- ❑ Use column **names**, not column numbers

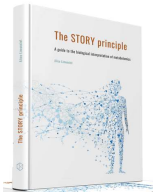
➔ More flexible with different pre-processing settings generating extra columns

30

5. Take home messages

- Agree *with yourself* & others to define the objectives and research questions clearly. *Don't try to answer 5 questions at the same time!*
- Remember that diagnostics ≠ research
  ➔ Think of it ahead when doing clinical research
- Take the time to design your experiments, you'll save yourself time and pain

THE METABOLOMIST
by Alice Limonciel



- Write things down, it will help you sort your ideas.
- Ask for feedback quickly. *Don't try to make it perfect!*

Better is the enemy of good.

- Take/ask for the time to learn a programming language for (big or small) data processing, as you are the bridge between the objectives, the sample story, and the interpretation of results.

31

Acknowledgments

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Dep. of Clinical Therapeutics, School of Medicine
E. Kastritis

Laboratory of Pharmacology, Faculty of Pharmacy
P.E. Nikolaou
I. Andreadou

Proteomics core facility & Dep. of Chemistry
G. Médard
N. Thomaidis

UCL

Centre for Amyloidosis, University College London, and Royal Free London NHS Foundation Trust

D. Canetti
J. Gilbertson
A. Wechalekar

STEPUPIORS

32