

PRIDE Peptidome: A hub for peptide evidences from public proteomics datasets for ENSEMBL and Uniprot



Yasset Perez-Riverol ¹, Chakradhar Reddy Bandla ¹, Selvakumar Kamatchinathan ¹, Juan Antonio Vizcaíno ¹

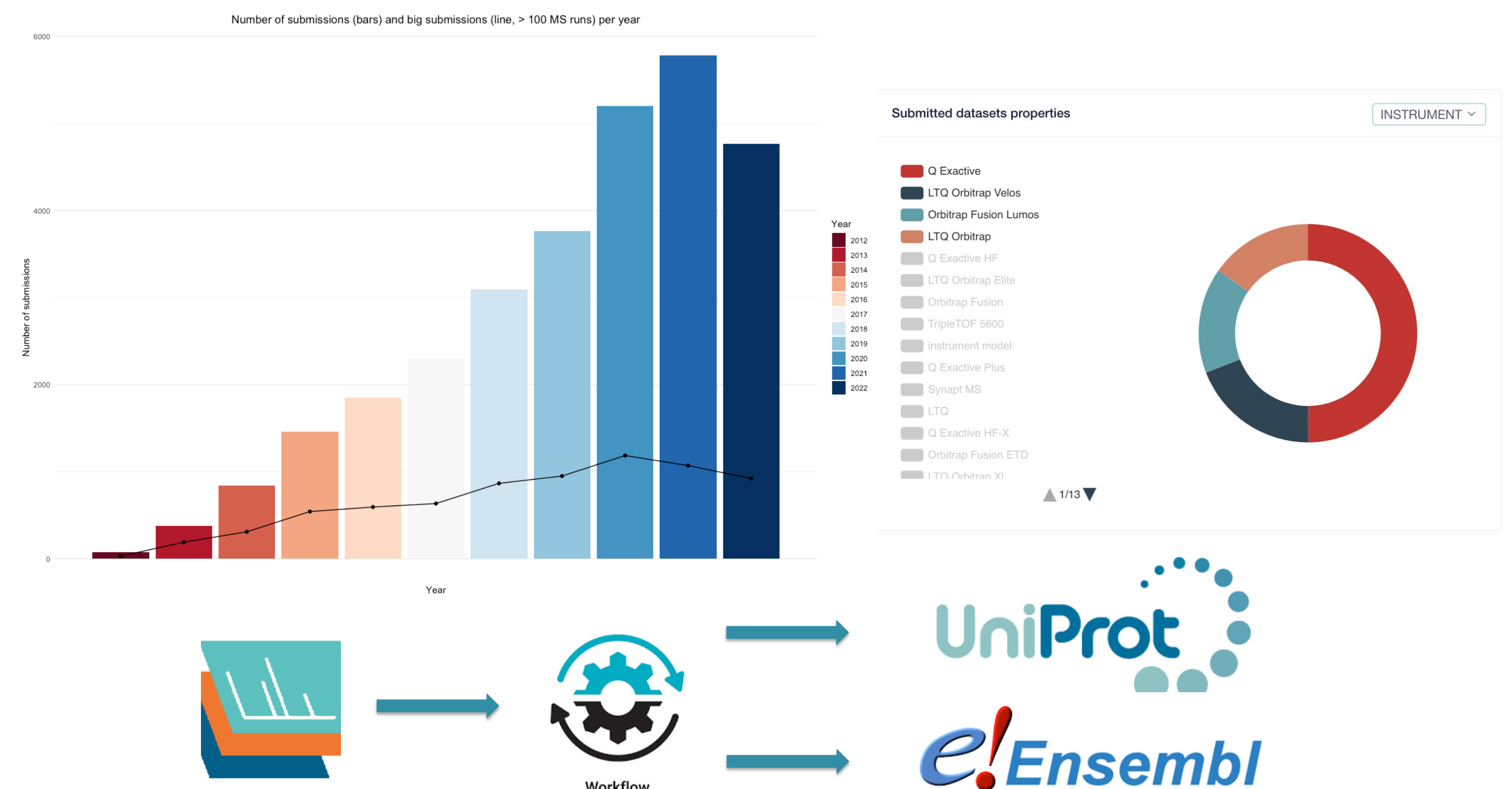
(1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome

Campus, Hinxton, Cambridge, CB10 1SD, UK.

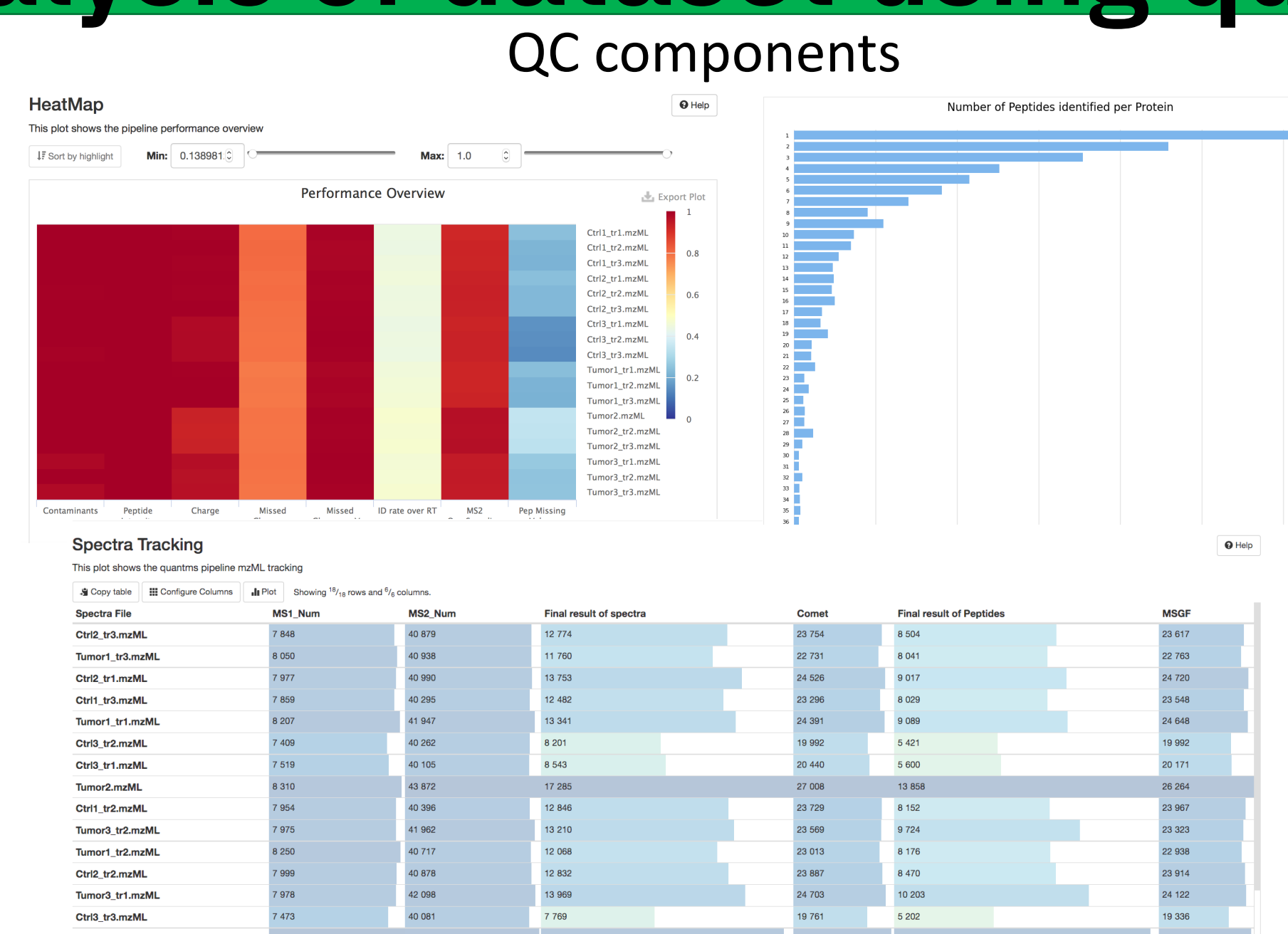
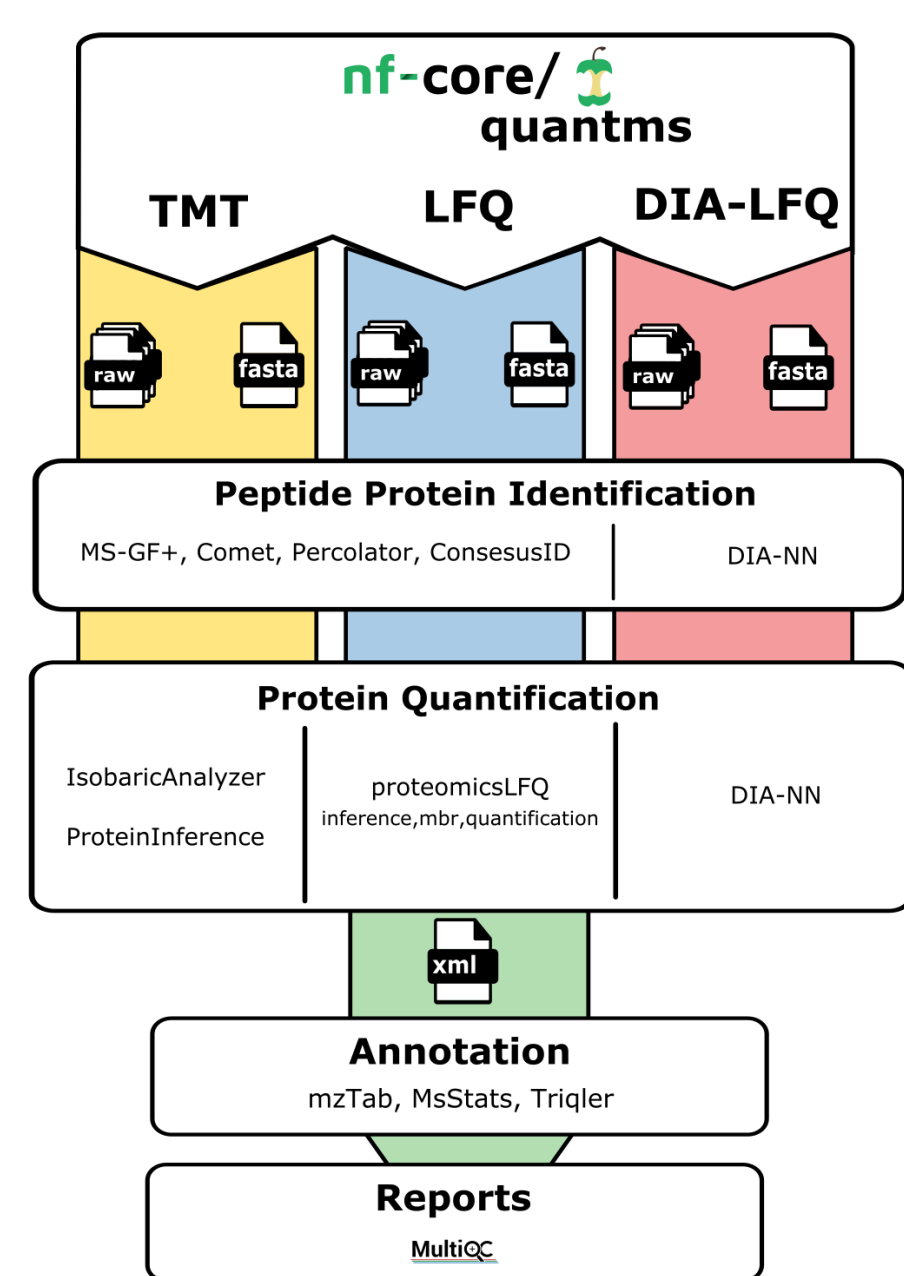
<https://www.ebi.ac.uk/pride/>

The PRIDE database – Public data reuse: Bringing proteomics data to biologists

- EMBL-EBI's PRIDE database is the largest data resource of MS proteomics data worldwide.
- During 2021 alone, 490 datasets were submitted per month on average to PRIDE.
- Reuse of public proteomics datasets is increasing dramatically, with multiple applications.
- Rather than in the creation of new bioinformatic resources, for sustainability reasons, our focus is to disseminate and integrate PRIDE proteomics data into added-value EMBL-EBI resources, namely **UniProt** (PTM data, protein sequences), **Expression Atlas** (protein expression information, the focus of this poster), **Ensembl** (proteogenomics data) and **MGnify** (metaproteomics and metagenomics data).
- The goal is to make proteomics data more accessible to life science scientists, especially those non-expert in proteomics.



Internal reanalysis of dataset using quantms (DDA-LFQ, DDA-TMT, DIA)



Major Data-related features:

- Use multiple search engines in combination (msgf+, comet), boost the number of peptides by using percolator.
- Perform multiple protein quantification methods: LFQ-intensity based, spectral counting, TMT intensity, DIA-LFQ intensity based.
- False localization rate analysis for PTMs including Phosphorylation and acetylation.
- Based on HUPO-PSI standard file formats including mzTab, mzML, SDRF.

Major cloud features:

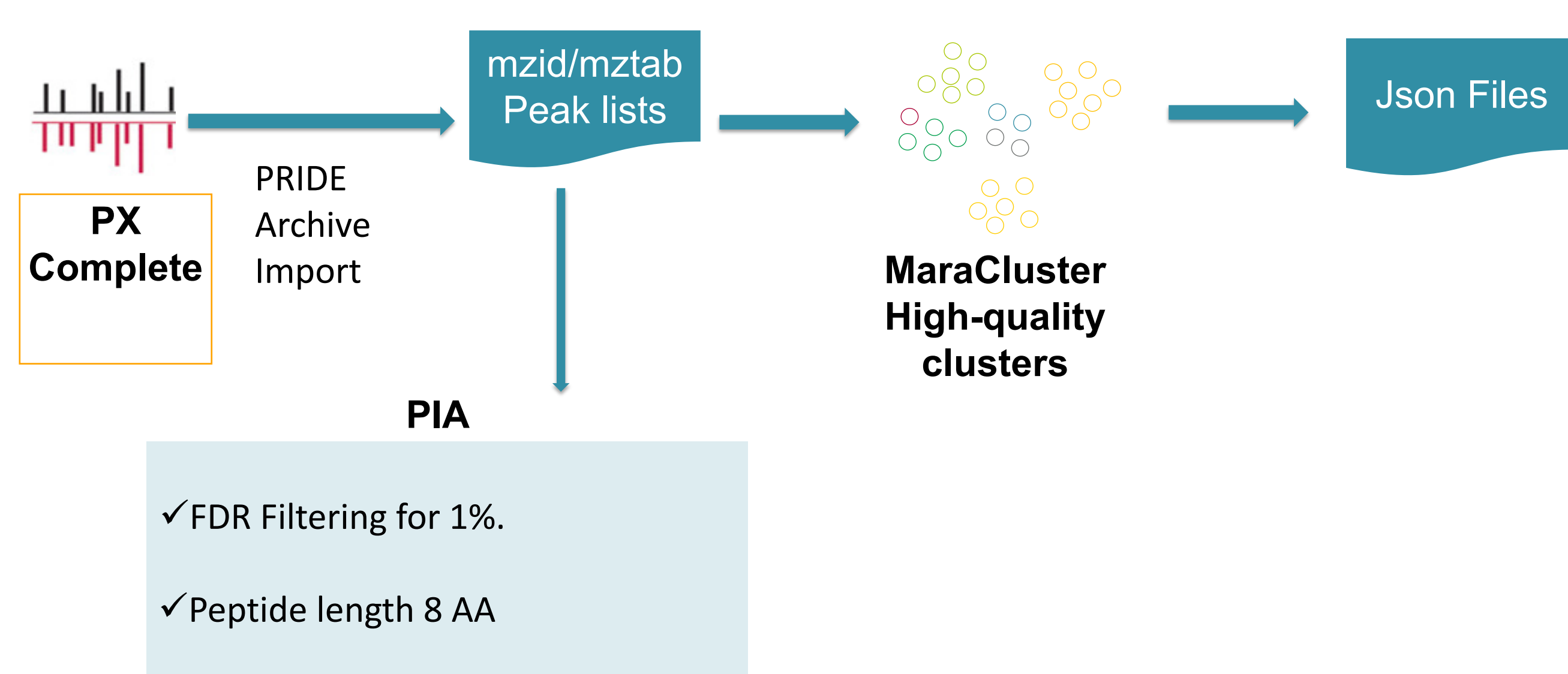
- Can be executed in LSF, AWS, Google Cloud, Kubernetes, local PC, HPC. Based on BioContainers.
- Open-source, open access.
- Include features for compute resource allocation, automatic fail-tolerance and resume workflows.

References:

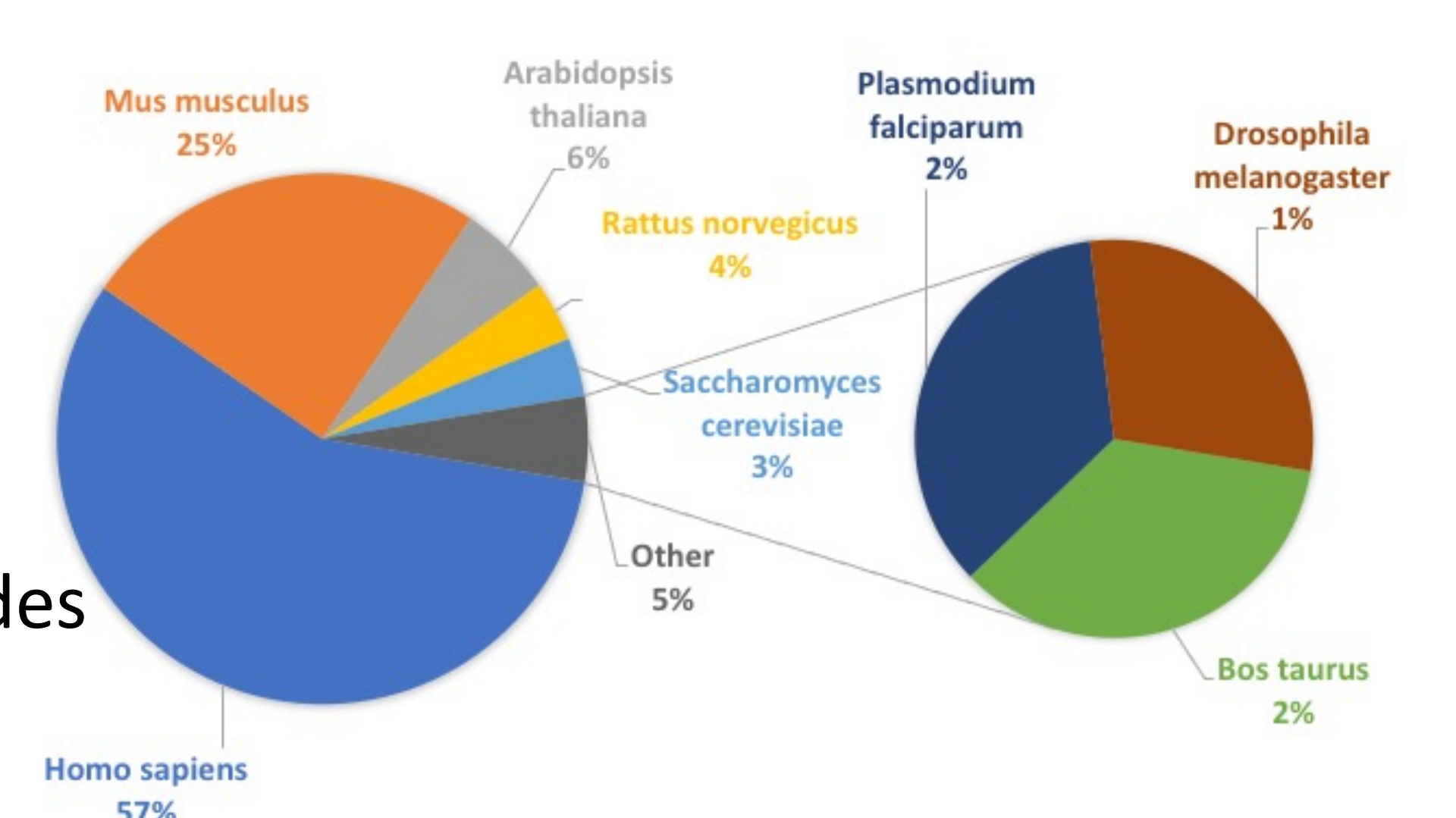
- Umer, Husen M., et al. Bioinformatics 38.5 (2022): 1470-1472.
- Bai, Mingze, et al. chemRxiv (2022). 10.26434/chemrxiv-2022-x18d5

<https://quantms.readthedocs.io/>

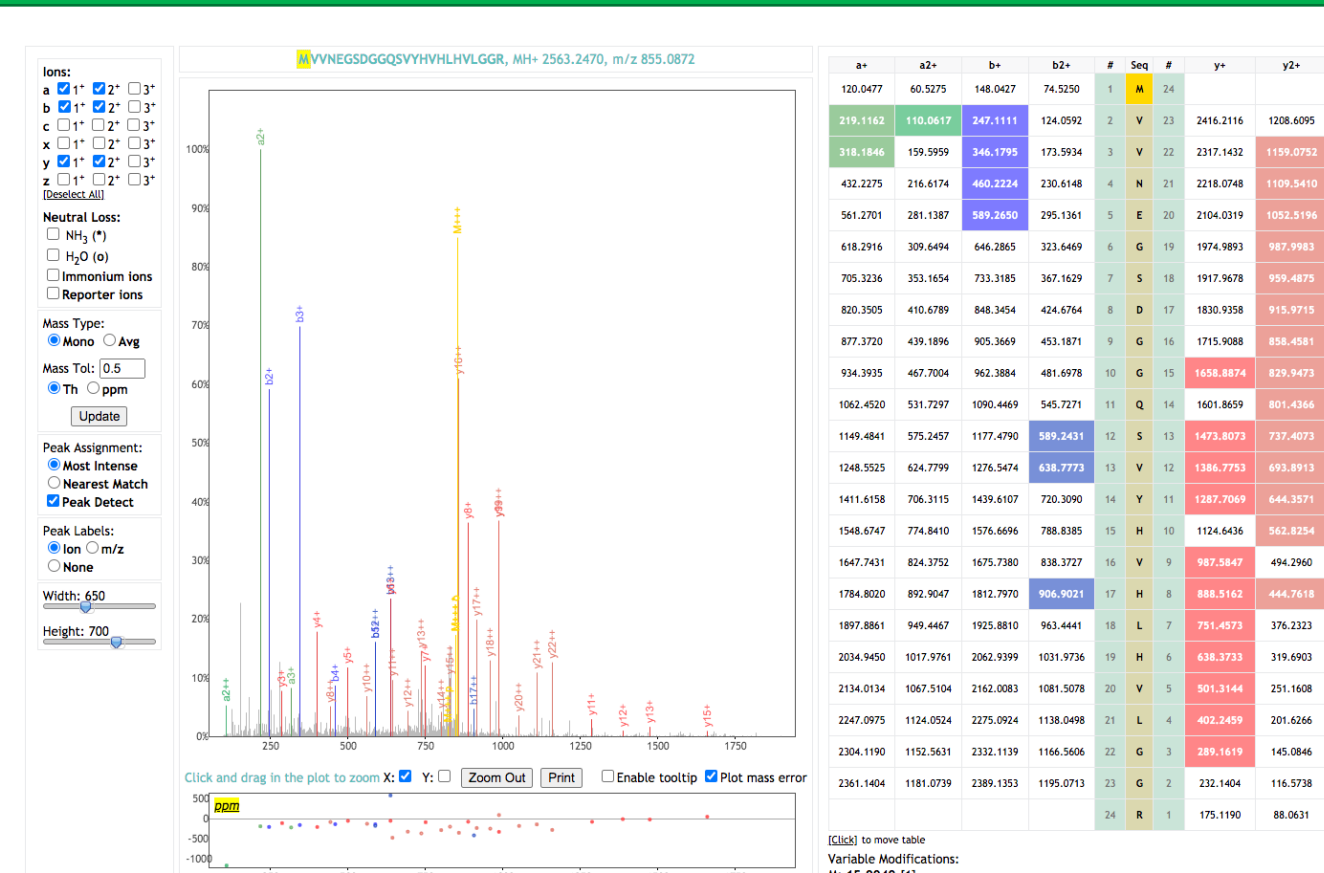
From complete submissions to peptide evidences



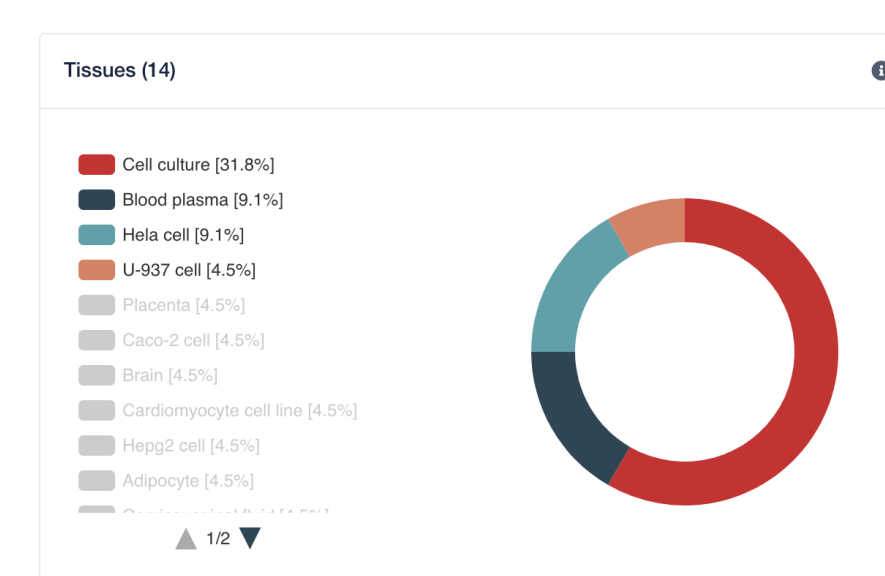
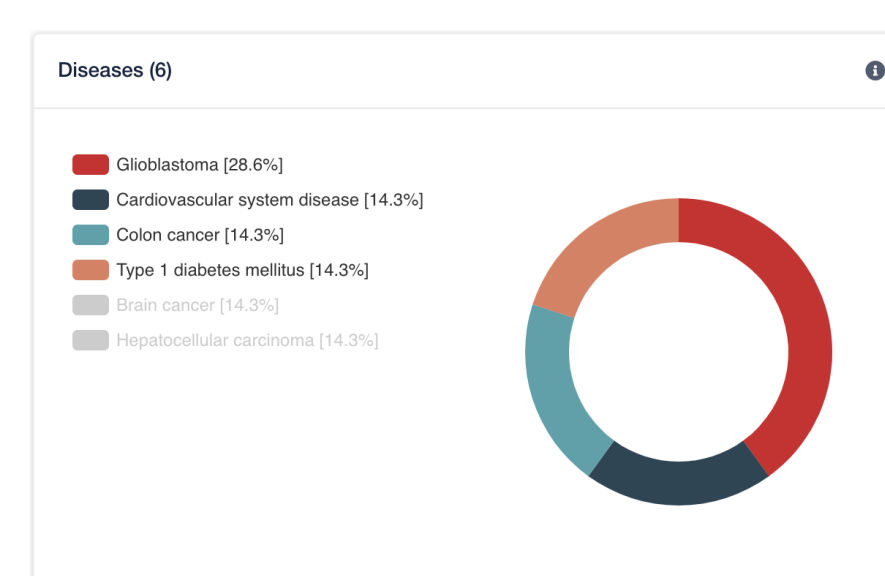
- ✓1294 Projects
- ✓217'261 Proteins
- ✓2'300'389 Peptides
- ✓1'922'855 Unique Peptides
- ✓150'658'201 PSMs



Peptidome and spectra archive



Peptide: TNEEAGDGTITATVLAR
Uniprot Accession: P10809
Gene: HSPD1
Protein Name: 60 kDa heat shock protein, mitochondrial
Organism: Homo sapiens
Evidences: 135 PSM, 12 Projects
Status: 0.00001



- ✓Search by peptides, protein accessions, names, gene names.
- ✓Peptide + protein visualization including names, number of psm and projects.
- ✓ENSEMBL and Uniprot links to the corresponding spectra that validate a specific sequence feature (e.g., PTM or variant)

References:

- Yasset Perez-Riverol, et. al. Nucleic Acids Research (2022), Pages D543–D552,
- Deutsch, E. W., Perez-Riverol, et. Al. (2021). Nature methods, 18(7), 768-770.

<https://www.ebi.ac.uk/pride/peptidome>