A Galaxy of informatics resources for MS-based proteomics

Subina Mehta[1], Matthias Bernt[2], Matthew Chambers[3], Matthias Fahrner[4,5], Melanie Christine Föll[4,5,6], Bjoern Gruening[7], Carlos Horro[8,9], James E. Johnson[10], Valentin Loux[11,12], Andrew T. Rajczewski[1], Oliver Schilling[4,5], Yves Vandenbrouck[13], Ove Johan Ragnar Gustafsson[14], W. C. Mike Thang[15,17], Cameron Hyde[15,16], Gareth Price[15,17], Pratik D. Jagtap[1]*, Timothy J. Griffin[1]*

[1]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA
[2]Helmholtz Centre for Environmental Research – UFZ, Department Computational Biology, Permoserstraße 15, D-04318 Leipzig, Germany
[3]Bioinformatics Consultant, Stamford, CT, USA
[4]Institute for Surgical Pathology, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany
[5]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Heidelberg, Germany
[6]Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA
[7]Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Freiburg, Germany
[8]Proteomics Unit, Department of Biomedicine, University of Bergen, Bergen, Norway
[9]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway
[10]Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN, USA
[11]Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
[12]Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France
[13]Proteomics French Infrastructure, CEA, 38000, Grenoble, France
[14]Australian BioCommons, University of Melbourne, Melbourne, Australia
[15] Queensland Cyber Infrastructure Foundation (QCIF), St Lucia, Australia
[16]University of the Sunshine Coast, Sippy Downs, Australia
[17] University of Queensland, St Lucia, Australia
*corresponding authors

Abstract:

Introduction:
Continuous advances in mass spectrometry (MS) technologies have enabled deeper and more reproducible proteome characterization, and a better understanding of biological systems when integrated with other 'omics data. Bioinformatic resources meeting analysis requirements of increasingly complex MS-based proteomic data, and associated multi-omic data, are critically needed. These requirements included availability of software spanning diverse types of analyses, along with scalability for large-scale, compute-intensive applications and mechanisms to ease adoption of the software.

Areas covered:
The Galaxy ecosystem meets these requirements by offering a multitude of open-source tools for MS-based proteomics analyses and applications, all in an adaptable, scalable, and accessible computing environment. A thriving global community maintains these software and associated training resources to empower researcher-driven analyses.

Expert opinion:
The community-supported Galaxy ecosystem remains a crucial contributor to basic biological and clinical studies using MS-based proteomics. In addition to the current status of Galaxy-based resources, we describe ongoing developments for meeting emerging challenges in MS-based proteomic informatics. We hope this review will catalyze increased use of Galaxy by researchers employing MS-based proteomics and inspire software developers to join the community and implement new tools, workflows, and associated training content that will add further value to this already rich ecosystem.

Keywords
Bioinformatics, Computational workflows, Galaxy platform, Mass-spectrometry, Multi-omics, Proteomics, Reproducibility

Article highlights:

- The Galaxy bioinformatics ecosystem provides a flexible and scalable informatics resource for scientists to access software analysis tools, process data, and visualize results customized to their needs.
- MS-based methods in proteomic research generate complex data types, that are computationally challenging for their processing, analysis, and interpretation.
- The provenance tracking in Galaxy allows users to save and share complete analysis histories and workflows in shotgun proteomics, data independent acquisition (DIA) MS-based quantitation, multi-omics, MS imaging, and results visualization and interpretation.
- Current workflows within Galaxy have enabled research in various fields such as COVID-19 pandemic research, proteogenomics, metaproteomics, MS-imaging, and MS-based proteomic clinical and translational studies in patient-derived samples.
- Galaxy ecosystem also offers access to training resources which promotes awareness and empowers adoption of these tools by the research community.

1. Introduction: The Galaxy Ecosystem

The constantly evolving mass spectrometry (MS)-based methods and technologies that drive proteomic research generate intricate and diverse data types which in turn create numerous computational and informatics obstacles [1]. To overcome these challenges, researchers need access to adaptable, scalable, and reproducible informatics resources that cater to the specific requirements of end-users and their research projects. Researchers worldwide have come together to utilize the Galaxy bioinformatics ecosystem as a powerful solution [2,3]. Galaxy is a user-friendly workbench for scientific computing, deployable on scalable computing infrastructure, and accessible through a web graphical user interface. With minimal technical barriers, scientists are able to access software analysis tools and workflows, process their data, and explore, visualize, and portray their results in a manner customized to their needs [4]. Provenance tracking is a key feature, allowing users to save complete "Histories", which record and archive all steps of an analysis, as well as intermediate and final result files. "Workflows" (**Figure 1**) can be extracted from a history and shared. These workflows record the validated settings for each software tool utilized and can be modified and applied to any input data set with compatible formats. These Histories and Workflows can be shared with other users, assigned DOIs (across versions), and adopt best practice metadata on platforms like WorkflowHub [5] and Dockstore [6], thereby promoting transparent and reproducible analyses. These workflows also encapsulate a series of interconnected software tools packaged within Galaxy, streamlining the entire analysis process, and promoting reproducibility by ensuring consistent methodology across experiments. This modular approach enhances flexibility, as researchers can easily modify, add, or remove specific components to tailor workflows to their unique research questions [7]. Moreover, the accessibility to extendable dynamic memory, made possible through Galaxy's integration with High-Performance Computing (HPC) platforms[8], has been demonstrated as a key feature towards enabling large-scale, compute-intensive analyses [9–11].

Since its inception in 2005 [3,12], the community of Galaxy users and developers has continuously grown and facilitated the development of a comprehensive and open bioinformatics ecosystem with broad application across an expanding range of scientific domains [3].The framework has been so impactful that more than 130 public instances of Galaxy are deployed worldwide on numerous scalable computing platforms [13]. Galaxy has contributed to over 10,000 published studies by researchers across the globe, a number that continues to grow [14]. Access to Galaxy resources is further enhanced by a global network of large, freely available servers that offer both new and experienced users access to validated tools that span the 'omic domains: these are part of the usegalaxy.* network of publicly accessible instances (https://galaxyproject.org/use/). The Galaxy Tool Shed [15] underpins the ecosystem by making thousands of wrapped and validated software tools for Galaxy available [16].

Notably, an active global community of researchers has created numerous publicly accessible, online, and on-demand training resources, including for MS-based proteomics, as part of the innovative Galaxy Training Network (GTN) [17]. These resources - which include guided walkthroughs of diverse analyses based on small exemplary datasets, recorded videos, and pre-prepared workflows - offer a unique means

for non-expert users to learn and access sophisticated software tools and optimized workflows and subsequently adopt them for their research questions.

2. MS-based proteomic informatics in Galaxy

Galaxy offers a consolidated platform where MS-proteomics data and related data from other 'omic' domains can be analyzed with a flexible combination of tools and workflows that cover many of the mainstream contemporary applications of proteomics (**Table 1**). Importantly, Galaxy can be deployed across multiple different types of backend infrastructure, including locally maintained and accessed servers with advanced high-performance computing systems, as well as scalable cloud environments for web-based access. A single Galaxy instance can distribute its jobs across compute clusters using the Pulsar system [18], such that computationally intensive tools can be directed to Pulsar nodes with many cores. The Total Perspective Vortex (TPV) [19] has empowered Galaxy to install, deploy and configure specific tools with the unique requirement of different infrastructures.

Over the last decade, tools and resources have been established in Galaxy that encompass the main analysis approaches and requirements of MS-based proteomic applications: 1) shotgun, data-dependent acquisition (DDA), 2) data-independent acquisition (DIA), 3) multi-omics, centered around MS-based proteomics, 4) MS imaging (MSI), and 5) visualization and biological interpretation of processed results. **Table 1** shows a selection of the more well-known tools [20] available in Galaxy for various analysis applications, which are described in more detail below. In total, there are approximately 400 tools for MS-based proteomics within the Galaxy Tool Shed [15] as well as dedicated GitHub repositories (e.g. github.com/galaxyproteomics/tools-galaxyp), including both standalone software designed by leading labs that specialize in MS-based proteomic software development, and also specialized tools designed for specific applications and creation of integrated customized workflows. Galaxy's extensive tool repository available through either the Tool Shed [https://toolshed.g2.bx.psu.edu/] or GitHub [https://github.com/galaxyproject] enables researchers to access and incorporate a diverse array of tools that can be tailored to their specific needs. Comprehensive tool suites (e.g. OpenMS [21], TransProteome Pipeline [22], CompOmics [23], and ongoing implementation of FragPipe [24]) are also accommodated by wrapping individual functions to create a modular set of independent Galaxy tools that can be executed one-by-one, or chained into complex workflows to suit bespoke analysis tasks [25]. Access to many of these tools and workflows is further facilitated by resources in the GTN and deployment on the usegalaxy.* network of publicly available instances.

This review offers an overview of the Galaxy ecosystem and its value as a comprehensive solution for MS-based proteomic informatics. We focus on highlighting the established, major usage applications in Galaxy (see **Figure 2**), as well as complementary, value-added characteristics of the Galaxy ecosystem that distinguish it from other MS-based proteomic informatics platforms. Finally, we provide some thoughts on Galaxy's role in meeting informatics challenges related to emerging technological advances in MS-based proteomics.

2.1. Shotgun Proteomics

From the outset, the implementation of MS-based proteomics tools in Galaxy focused on datatype specifications and software tools for liquid chromatography-tandem mass spectrometry (LC-MS/MS) shotgun proteomics [26]. LC-MS/MS of complex peptide mixtures derived from tryptic digestion of proteins is a mainstay for proteomic analysis, using so-called "data dependent", or DDA methods, for detecting and selecting individual peptides for fragmentation and generation of MS/MS spectra. Sequence database search programs utilize FASTA-formatted protein sequence databases to match peptide sequences to these spectra, generating peptide spectrum matches (PSMs), which are used to infer protein identities in the starting sample.

Over the last decade, several core sequence database search tools have been implemented and used extensively in Galaxy. These have followed a foundational philosophy of the Galaxy community, to deploy well-validated, Galaxy-compatible command-line software tools within the platform that are of the highest value to the user community, while also offering a choice of tools where possible. For shotgun DDA MS/MS data, available programs include SearchGUI/PeptideShaker [27,28], MaxQuant, and appropriate tools from the OpenMS [21], Trans Proteomic Pipeline (TPP) [22] and CompOmics [23] suites have been implemented. More recently, the powerful and increasingly popular FragPipe tool suite [24] is being deployed and validated, which should offer new possibilities for efficient PSM and protein identification across many studies. For all these tools, the recorded histories can be easily re-run on the same input data, modifying settings as needed to explore new questions (e.g., searching for novel proteoforms, microbial peptides or new post-translational modifications etc.). Many of these software takes advantage of the scalability of Galaxy on HPCs, making it possible to distribute jobs and analyze hundreds of samples in feasible timeframes, as has been shown for other compute-intensive analyses. Galaxy offers a distinctive advantage over stand-alone software tools as it removes the need for researchers to install each piece of software and manage the compute system that the software runs on, it allows researchers to construct tailored workflows for sequence database searching that employ several of these software, leveraging the individual strengths of each tool and the complementary results they produce [29,30].

Beyond qualitative peptide and protein identification of LC-MS/MS data, tools for quantitative analysis of this data have been emphasized. These include software for popular methods in quantitative proteomics that utilize both stable-isotope labeling and label-free methods [31]. The software MaxQuant and Fragpipe tool suite[32] provides rich functionality for the analysis of stable-isotope labeled data such as stable isotope labeling in cell culture (SILAC)[33], and isobaric peptide tagging (e.g. iTRAQ [34], and tandem mass tag, TMT[33]. For label-free quantification (LFQ), the tools FlashLFQ and moFF have been implemented and rigorously tested in Galaxy[35], complementing tools available in MaxQuant [36] and Fragpipe [37] for similar LFQ analyses. Importantly, the community-standard MSStats platform [38,39] for statistical analysis of quantitative MS-based proteomics data is available within Galaxy[40].

2.2. *Data independent acquisition (DIA) MS-based proteomics*

Although shotgun DDA-based LC-MS/MS remains a mainstay for many proteomics researchers, DIA methods are quickly gaining popularity. Rather than rely on stochastic detection of individual peptides in DDA acquisition, DIA collects MS/MS data from all detectable peptides by rapidly scanning discrete mass-to-charge (m/z) windows across the entire usable range of peptide masses as they elute from the LC [41–43]. Using customized software tools, such as EncyclopeDIA [44], DIA-NN [45], or OpenSWATH [46], co-eluting peptide fragments from a set mass window are extracted and used to verify the presence of the sequence and quantify its abundance by area-under-the-curve measurements. DIA offers many potential advantages compared to DDA, including more reproducible detection and quantification of peptides and their inferred proteins, associated post-translational modification (PTM) events such as phosphorylation [47–49], and amenability to deep and accurate quantification of complex samples in a high throughput manner[50–52].

Given its potential and increasing popularity, Galaxy community members have focused attention on deploying tools to enable DIA-based analyses. These have included the OpenSwath tool suite [46,53,54] and associated tools such as diapysef, PyProphet [55], TRIC [56] and SWATH2stats [57] and DIA-focused analysis functionalities now available in FragPipe [58]. The EncyclopeDIA software, which enables efficient and comprehensive DIA analysis using chromatogram library information, has recently been installed[59]. Outputs from these DIA analysis tools are also amenable to statistical analysis using the MSStats tools in Galaxy [60]. Although the Galaxy platform currently lacks dedicated tools for Selected Reaction Monitoring(SRM)/Parallel Reaction Monitoring(PRM), the DIA and DDA workflows create results for identified peptides and proteins that lend themselves development of such targeted methods using the popular Skyline platform [61]. The intensive computing and memory requirements typical of complex DIA analyses can be met by Galaxy's amenability to deployment on scalable high-performance computing infrastructure.

2.3. *Multi-omics*

Given its initial development as a genomics-centric bioinformatics platform, Galaxy houses a large selection of contemporary tools for the analysis of next-generation sequencing (NGS) data. With the addition of MS-based proteomics tools, it quickly became apparent that Galaxy offered a unique solution for integrative, multi-omic informatics combining NGS and proteomic data and software [62–66]. One such application, proteogenomics [67], combines DNA and/or RNA NGS information with MS-based proteomics data. This approach is well-represented in Galaxy. A central aspect of proteogenomics is the ability to confirm the translation of novel protein products that are predicted by the assembly and annotation of expressed transcripts. This is accomplished by generating custom protein sequence databases that incorporate predicted and canonical reference sequences, as well as sequences translated from non-normal gene or mRNA sequences indicated by NGS analysis. For proteogenomics, customized software wrapped for Galaxy is used to identify non-normal sequences from assembled NGS data and generate corresponding novel protein sequences which can be included in FASTA-formatted sequence databases. These tools include CustomProDB [68] and the community-standard tools HISAT2 [69] and

StringTie [70,71] for assembling and annotating NGS data. The identification of non-normal sequences is supported by efforts in the Galaxy community to enable efficient, programmatic access to reference sequence repositories [72]. The PepQuery software [73,74] has been implemented as well, serving as a means to verify the confidence of putative novel peptides, by rigorously evaluating PSMs to novel sequences against other possible reference sequence matches (including those carrying PTMs). Galaxy also houses the QuanTP tool [75] to compare the expression response of RNA transcripts and their corresponding, encoded proteins, offering a means to ascertain potential post-transcriptional regulation events.

Another growing, MS-based proteomics-centered multi-omics approach is metaproteomics, which seeks to characterize the functional proteins expressed by the microbiome of microorganism communities [76,77]. Metaproteomics combines metagenomic and proteomic information to understand the biochemical response and functional properties of complex microbiomes, complementing information offered by metagenomic information alone. Despite its power, metaproteomics offers a number of bioinformatic challenges that make it unique compared to single-organism proteomics such as spectrum-to-peptide-to-protein-to-species inference, large databases with peptide overlap due to homologous proteins in closely related species, horizontal gene transfer (HGT) of entire functional modules, and sample complexity leading to lower identification rates per protein and per species.

The flexibility of Galaxy allows researchers to meet many of these challenges. Recently, complete applications for integrative “meta-omic” analysis in Galaxy have been described [78]. A number of tools also are available to address the challenges of MS-based metaproteomics. For example, approaches have been developed [79–81] to handle PSM generation using the very large protein sequence databases comprising all proteomes within a community (composed of millions of protein sequences). For taxonomic and functional analysis of peptide-level metaproteomics data, tools such as Unipept[82] have been implemented within Galaxy and rigorously evaluated [30]. Quantitative statistics of metaproteomics data are also enabled by the metaQuantome software suite which can analyze metaproteomics MS data to determine those taxa and functions that are differentially abundant. The tool uses an expand-filter function that leverages Enzyme Commission (EC) number, Gene Ontology (GO), and NCBI RefSeq databases to assign taxonomic and functional annotation to the statistically analyzed proteins. These serve as input to generate publication-ready Principle Component Analysis (PCA) plots, clustered heatmaps, and tabular outputs at the level of proteins, taxa, protein function, and taxon-function relationships [83,84]. The PepQuery tool has proven highly valuable for verifying the accuracy of PSMs matching microbial sequences, especially when analyzing samples dominated by non-microbial host sequences (e.g. human) [85]. The verified peptides can be quantified using Label-free Quantitation (LFQ) tools available in Galaxy and then passed on for further functional and taxonomic annotation [83,84], statistical analysis [40], and visualization [86] via modular workflows. This modularity not only expedites workflow development but also empowers researchers to customize analyses according to their research questions. However, this modularity can occasionally lead to workflow complexity, making it essential for users to possess a solid understanding of the tools and their interconnections. Additionally, as workflows grow in sophistication, they might become challenging to manage, requiring clear organizational practices, thorough documentation, version control, and registration (e.g. WorkflowHub [5], Dockstore

[6]). Fortunately, the Galaxy ecosystem offers resources to accommodate these needs [10], as described below.

2.4. MS imaging

Mass spectrometry imaging (MSI) is an MS technique that specializes in measuring molecular spatial distributions from complex samples such as thin tissue sections. MSI-based peptide and protein imaging differ from immunohistochemistry in that it is specific, untargeted, multiplexed, and label-free analyses. MSI enables numerous applications across diverse research fields: bacterial biofilm characterization [87], understanding spatial plant and animal biology [88,89], identification of disease-related biomarkers [90], food quality control [91], and forensic applications such as detection of blood in fingerprints. MSI acquires tens of thousands of mass spectra in a grid pattern across the sample with step sizes between 5 and 200 µm. This results in complex and often large raw data requiring specialized MSI software for quality control, pre-processing, co-registration, statistical analysis and visualization[92].

All these typical MSI data analysis steps can be performed within Galaxy. Analysis methods from the Cardinal [93] and MALDIquant [94] R packages are implemented into Galaxy as modular tools [95]. At the same time, a unique "MSI Quality control" tool that generates a comprehensive quality report with information on all important data properties has been developed as part of the MSI tool suite in Galaxy [95]. MSI data are often accompanied by optical images e.g., of stained tissue sections, and additional shotgun proteomics data to identify and validate the observed m/z features. Galaxy is uniquely suited for such multimodal imaging and multi-omics experiments because it provides nearly 100 general image analysis tools as well as a variety of shotgun proteomics software mentioned above, enabling such complex analysis within a single platform.

2.5. Results visualization and interpretation

Galaxy offers rich functionalities that allow a user to view, interactively explore, and export processed results, streamlining this essential analysis step and supporting interpretation and hypothesis generation from 'omic studies. An underrated functionality is the ability to transform outputted results, either in a final or intermediate form, to generate results data in formats compatible with downstream visualization and interpretation tools. This includes core Galaxy tools for manipulating generic text and tabular formatted outputs, as well as a more sophisticated tool [96] which uses an SQLite database to transform complex outputs into customized formats for downstream processing.

Galaxy also offers a sophisticated visualization registry (galaxyproject.org/visualizations-registry/) allowing for the development of customized plugin tools. A prominent set of these tools, for manipulating, visualizing, annotating, and running pathway enrichment analysis has been developed by the ProteoRE group [97–99] for MS-based proteomics results. Sets of protein IDs of interest serve as inputs to ProteoRE, which then offers numerous options for assessing enriched functional classes, pathways, and interaction networks, along with visualizations of these results (see **Figure 3** for an example). ProteoRE

also provides a means to access the SRM Atlas [100] repository to aid in developing targeted assays for proteins deemed of highest priority and interest based on interpretation of results.

Multi-omics applications in Galaxy also leverage its visualization capabilities. For proteogenomics, the Multiomics Visualization Platform (MVP)[86] integrates genome- and proteome-level knowledge to evaluate the quality of PSMs to novel peptide sequences, understand the nature of the sequence variation, and map these to the coding regions of the genome and/or transcriptome. The QuanTP tools also offer visualization of RNA-protein quantitative response that helps determine influential data points[75]. A Galaxy implementation of the Cancer Related Analysis of Variants [101] tool (called CRAVAT-P [102]), retrieves and visualizes information on the cancer-associated impact of non-normal peptide sequences identified by proteogenomics. For metaproteomics, tools such as Unipept[82] provide a means to annotate microbial peptides with taxonomy and function and visualize the phylogenetic properties of the sample indicated by these annotations. metaQuantome [83,84] is a customized Galaxy tool that further analyzes quantitative metaproteomics data, offering statistical analysis of differentially abundant peptides and their represented functions and taxonomies. It also visualizes these results, offering unique looks at taxonomy-function relationships indicated by the quantitative metaproteomic results. More recently, the interactive tool LFQ-Analyst was released for Galaxy. This interactive tool simplifies and standardizes complex downstream analysis and visualization of LFQ datasets, and is aimed at making these complex datasets more approachable and easier to interpret for less experienced users [103].

3. Access and Training

From its inception, the Galaxy community has focused on the democratization of advanced bioinformatic tools and emphasized the need to empower end-user biologists to utilize these tools in their research, without worrying about the technical details of how these tools are executed on the underlying computational infrastructure. To this end, both easy access by the research community and straightforward onboarding for software use are key. The globally distributed and interoperable usegalaxy.* network (galaxyproject.org/usegalaxy/) offers free and open access to Galaxy resources, deployed on a scalable infrastructure, and tailored to all experience levels. This network even includes a gateway specifically for MS-based proteomic tools (proteomics.usegalaxy.eu), with a similar gateway under development on the Australian-maintained public instance (proteomics.usegalaxy.org.au). These gateways are powerful because a Galaxy instance can host many thousands of tools spanning many different domains. Domain specificity allows for a focused view of the Galaxy ecosystem for users, and the opportunity to support the proteomics community more directly.

Training of end-users is also key to promoting the adoption of bioinformatics tools. The Galaxy community has intensely focused on developing high-value, easily accessed training materials [104]. As a result, the GTN was developed to support online, on-demand training material that guides trainees using tools and workflows while pointing them to the accessible Galaxy gateways where these can be utilized on their own datasets. Currently, 27 different tutorials exist related to MS-based proteomic tools and workflows. The GTN entry page for proteomics tutorials presents a listing of resources grouped by application area [17,104]. The GTN materials are also used for workshops, either held in-person or via online formats, both

live and recorded for on-demand access. **Figure 4** shows examples of training activities relevant to metaproteomics offered over recent years that leveraged GTN resources. Training materials for MS-based proteomics are also showcased during the annual Galaxy Community Conference [105] or during worldwide "Smorgasbord" events, a free series of online, self-paced workshops held annually that reaches thousands of end-users [106,107]. Collectively, the community-driven training activities offer a multitude of powerful onboarding mechanisms which lower the entry barrier for new users and empower adoption of Galaxy for MS-based proteomics.

4. Example applications of Galaxy for MS-based proteomics-focused research

As an increasing number of tools and training materials have been implemented within the ecosystem, numerous groups have made use of Galaxy to drive their research projects and make new discoveries across several fields. A selection of these studies is summarized below.

4.1. COVID Pandemic Work

Amidst the worldwide outbreak of coronavirus disease 2019 (COVID-19), multiple workflows for data analysis spanning 'omic domains were published by the global community on the Galaxy Europe instance [108]. This included MS-based proteomics methods and workflows (covid19.galaxyproject.org/proteomics/) to detect and characterize proteins expressed by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of the COVID-19 disease, and thereby facilitate the development of better therapeutic measures and diagnostic tools. One of the first studies applied Galaxy workflows to published clinical and cell culture datasets to detect peptides that are specific to the SARS-CoV-2 virus. These peptides were predicted to have great value for clinical proteomics applications seeking to detect COVID-19 from patient samples [109]. Galaxy workflows were also used to reanalyze published MS datasets generated from clinical samples to determine the coinfection status of individuals infected with SARS-CoV-2 coronavirus [85]. This led to the detection of opportunistic pathogens which may aid in better diagnosis and treatment of COVID-19 patients. These workflows were further extended to study the co-infection status of COVID-19 patients during two pandemic waves from India [29]. In this study, opportunistic pathogens such as *Streptococcus pneumoniae*, *Rhizopus microsporus*, *Enterobacter*, and *Clostridium* were detected in COVID-19 patients and validated by targeted proteomics analysis. Galaxy's provenance-tracking architecture enabled COVID-19 data analysis workflows and histories to be accessed through public gateways. This facilitated rapid collaborative research that analyzed mutating strains during pandemic waves all over the world [110]. Workflows were even developed in this study to detect variant-specific viral peptide sequences from MS data derived from published clinical data during different pandemic waves. The analysis identified six SARS-CoV-2 variant-specific peptides suitable for confident detection by MS in commonly collected clinical samples. This study highlights the strengths of Galaxy's modular workflows and underscores its notable reanalysis capabilities which have also been highlighted in a number of other published studies [29,83,85,109–

116]. These capabilities empower researchers to effectively reexamine datasets using the latest tools and techniques, facilitating new discoveries leveraging a dynamic bioinformatics landscape.

4.2. Proteogenomics

Galaxy's appeal as a multi-omics platform prompted the integration of MS-based proteomics tools which synergize with the array of genomic and transcriptomic software already hosted by the platform[62]. As such, Galaxy has been applied to a number of proteogenomic studies. One study utilized a "Proteomics informed by transcriptomics" approach in Galaxy to identify active transposable elements and further annotate the genome of the pathogen *Aedes aegypti* [117]. Another utilized the Peptimapper proteogenomic tools to annotate the genome of the marine brown algae *Ectocarpus* [118]. The tools comprising the PROTEOFORMER Galaxy workflow [66], were used to discover and create a database of small open reading frame (sORF) sequences via the integration of ribosome-protected transcripts (Ribo-Seq) and MS-based proteomics data [119]. Co-authors on this review leading the Galaxy for proteomics (Galaxy-P) project have applied Galaxy-based proteogenomics tools to study hibernation in the non-model, 13-lined ground squirrel *Ictidomys tridecemlineatus* [120,121]; detecting peptides corresponding to potential novel proteoforms from human saliva [116] and more recently these tools were applied to the proteogenomic analysis of inflamed colon-tissue as a means to discover mechanisms underlying cancer progression[122].

4.3. Metaproteomics

Galaxy has been a leading platform for metaproteomic studies. Both workflows [111,113] and specific, Galaxy-implemented software tools [30,83] have provided an in-depth analysis of the taxonomic composition of the cervical-vaginal microbiome [115], the Broncho-alveolar lavage fluid metaproteome in Acute Respiratory Failure [114] and metaproteomics analysis of SARS-CoV-2-infected patient samples for secondary infections [85]. Galaxy-based metaproteomics workflows were also used to analyze protein relative abundance patterns associated with sucrose-induced dysbiosis within oral microcosm biofilm models of dental caries [123]. The Galaxy-based metaQuantome suite has been used for the analysis of MS data acquired from irritable bowel syndrome mice revealing the regulation of host luminal proteases as a disease-relevant mechanism of host-microbial interaction that maintains protease homeostasis in the gut [124]; it also was a key tool for in-depth metaproteomics analysis of the oral microbiome in lung cancer to reveal taxonomy-function relationships [125]. Galaxy-based software and workflows also have played a key role in the Metaproteomics Initiative's [126] Critical Assessment of MetaProteome Investigation (CAMPI) study which compared metaproteomic workflows and platforms used across multiple laboratories [127]. The Galaxy workflows have also been used in an inter-laboratory comparison study as part of the ocean metaproteomics community initiative [128].

In a powerful example of its value to meta-omic analyses, the Arntzen lab has developed three Galaxy-based workflows to integrate metagenomics, metatranscriptomics, and metaproteomics. The workflow for metagenomics applies trimming and quality control of metagenomic reads followed by read assembly. Contigs can be phylogenetically binned into metagenome-assembled genomes

(MAGs), de-replicated if needed, and all genes are further annotated with functional data from InterProScan [129], KEGG [130], and CAZy [131]. Nucleotide and protein sequences serve as inputs for metatranscriptomics and metaproteomics, respectively. The workflow for metatranscriptomics performs trimming and quality control of the reads, removal of rRNAs, and finally quantification of mRNA by mapping to the metagenomics data using the pseudoaligner Kallisto [132]. The workflow for metaproteomics works similarly by data processing and filtering of mass spectra before matching to the proteins predicted by the metagenomics workflow using MaxQuant. The outputs from all three Galaxy- based workflows are integrated and can be visualized through their in-house R-shiny-based web application, ViMO [133], to study complex microbial communities' metabolic processes.

4.4. MS imaging (MSI)

MSI is a powerful technique for cancer research because it allows the spatial analysis of cancer tissues, which consists of a complex tumor microenvironment and often display molecular heterogeneity within macroscopically homogeneous cancer areas [134–136]. Co-authors of this review have applied tryptic peptide imaging to the bladder and colorectal cancer tissues and analyzed the MSI data in Galaxy [137,138]. In the first study, the tryptic peptide profiles of 39 tissues clearly distinguished the tumor from the surrounding stroma. In addition, the tryptic peptide profiles were used to build a classifier to distinguish tumor areas of muscle-invasive bladder cancer from non-muscle-invasive bladder cancer [137]. Apart from the biological findings, this study stands out in being the first fully transparent and reproducible tryptic peptide MSI study of a patient tissue cohort. This was made possible by performing all analysis steps on a single platform, the European Galaxy Server, and by sharing all Galaxy histories containing raw, meta-, and intermediate data according to the FAIR (findability, accessibility, interoperability, and reusability) principles. This study highlights the benefits of Galaxy for the MSI field, where data analysis is predominantly carried out using proprietary software or in-house scripts with insufficient reporting to facilitate reproducibility [139]. In a second study, conducted and shared via the European Galaxy Server, tryptic peptide MSI revealed similar intratumor heterogeneity within six patient-matched primary colorectal cancers and liver metastases [138]. In addition, the study found peptide features specific to tumor areas of both entities, but the metastatic tumors showed greater variability between the patients.

4.5. MS-based proteomic clinical and translational studies in patient-derived samples

Quantitative proteomics in patient-derived samples contribute complementary biological information in various diseases that have predominantly been studied on a genomic and transcriptomic level. In a multi-omic clinical investigation, co-authors of this review have performed quantitative proteomics in patient-derived skin samples to investigate underlying molecular pathomechanisms in a rare genetic skin disease called Netherton syndrome [140]. For robust and reproducible proteome quantification, the data-independent acquisition (DIA) strategy was applied, yielding comprehensive and complex quantitative proteomic data. The multi-omic approach enabled the detection of a shared immune signature in Netherton syndrome and distinct allergic responses between two clinical subtypes. Ultimately, the study proposes a pathophysiologic model that paves

the way toward novel therapeutic targets and improved medical treatment. The complete proteomic data analysis was performed using DIA analysis tools in Galaxy (including diapysef, OpenSwath tools, and PyProphet). Furthermore, the complete analysis and datasets have been published as Galaxy histories, promoting transparent and reproducible data analysis. The DIA strategy combined with the published, complete analysis history also allows for future reanalysis e.g., when additional samples are included, which is particularly interesting in rare diseases such as Netherton syndrome. This is one example of Galaxy playing an important role in clinical proteomics by empowering researchers to perform and share proteomic analysis of patient-derived samples.

5. Conclusions

MS-based proteomics will undoubtedly continue to advance technologically, and end-user biologists will need flexible, user-friendly informatics platforms that can easily adapt to these changes and facilitate onboarding and straightforward application to research questions that may require tailored informatics solutions. Researchers employing MS-based proteomics can use the Galaxy bioinformatics system as a comprehensive solution for processing protein-level data across many applications. The multifaceted Galaxy ecosystem, supported by a thriving community, should continue to play a significant role in addressing future challenges in MS-based proteomics and, more importantly, in accelerating new findings in biological and clinical research.

6. Expert Opinion

The continued emergence of new MS-based proteomics technologies will only increase the value of bioinformatics platforms such as Galaxy, which can democratize access to best-practice bioinformatics and meet evolving data analysis and informatics requirements. Here, we provide thoughts on areas focused on emerging approaches in MS-based proteomic over the next five years where Galaxy platform could be particularly valuable.

6.1. Scalability and adaptability. Galaxy democratizes access to computational proteomics so that a researcher does not need to be concerned with the technical execution of software but can instead focus on obtaining results to advance their research. With new MS instruments capable of rapid data generation from thousands of samples (e.g. ion mobility coupled with DIA data generation [141]), the volume of data that needs processing and management will grow exponentially - a trend observed in genomics research [142]. Furthermore, multi-step workflows need to keep pace with new software requirements for this data [143]. Galaxy will continue to offer adaptable workflows deployed on scalable high-performance computing resources to meet these requirements. Data analysis accomplished within Galaxy can be readily shared and downloaded by users. Although the current capability to directly upload data to public repositories (e.g., ProteomeXchange repositories such as PRIDE, MassIVE) is absent, there exists an alternative approach. The publicly accessible URLs (Uniform Resource Locator) linked with Galaxy histories and workflows can be seamlessly incorporated into a repository submission, along with raw data and processed results. These histories contain all raw input data, intermediate and results data, as well as records of all analyses and software parameters used,

promoting transparency and reproducibility. The workflows contain all software and optimized settings, facilitating their use for analysis of raw data generated by others.

6.2. Deep-learning-based spectral predictions and spectral library searching. In recent years, approaches based on spectral library searching of MS/MS data have gained momentum. Due to a significantly smaller search space and simpler matching offered by annotated MS/MS spectra, spectral library searching can be orders of magnitude faster and more accurate than conventional sequence database searching [144]. To avoid the laborious creation of spectral libraries from empirical MS/MS datasets, deep learning tools, notably PROSIT [145], have been developed to predict the fragmentation spectra of peptide sequences. These predicted spectral libraries can be matched to empirically generated MS/MS DDA data, using tools such as SCRIBE [146], or to DIA data using tools such as EncyclopeDIA [44] and DIA-NN [75]. Given the reliance of these methods on multiple, interoperable tools, Galaxy should provide a platform well-suited for spectral library analysis coupled with deep learning methods.

6.3. Microbe-host analysis in clinical samples. Over the past decade, various contributions of the microbiome in human disease have been explored and accepted in clinical research [147]. With tools accommodating end-to-end analysis of clinical datasets - from raw data processing to identification and quantification of host and microbial-expressed proteins, to functional characterization and visualization of results [148], we see Galaxy as an ideal platform for facilitating “clinical metaproteomics” studies (e.g. similar to COVID-19-focused studies [29]. Leveraging Galaxy's extensive metaproteomics workflows will help researchers better understand the role of microbiome with respect to disease progression.

6.4. Environmental metaproteomics. With increasing application of metaproteomics to characterize microbiome contributions to ocean samples [128] and soil samples [149–151], there is a need for robust bioinformatics workflows that can address questions in microbial ecology in these complex systems. The Metaproteomics Initiative [126] recently announced the CAMPI2 study [152] to establish a collaborative research focus on methods for best practice sample handling to preserve proteins contained in different environmental sample types. Galaxy’s established capabilities for metaproteomics should prove useful for these emerging applications studying microbial ecology in environmentally relevant systems.

6.5. Deeper and more accurate proteome quantitation. Recent advances in MS instrumentation, in particular, the inclusion of ion mobility for mixture fractionation, signal extraction and scoring of identified peptides, has enabled researchers to achieve deep proteome coverage and improved quantitative accuracy and reproducibility [153]. For example, combining Zeno trap technology with orthogonal quadrupole time-of-flight, called the ZenoTOF system™, enables high acquisition rates in MS1 and MS/MS mode [154], improving sensitivity without loss in acquisition speed or spectral resolution [155]. Other new MS platforms (Thermo Scientific™ Orbitrap™ Astral™ and Bruker timsTOF™ Ultra mass spectrometer) have significantly increased the depth of quantifiable

peptides generated with short analysis times [156,157]. Galaxy should offer a platform to implement new software for these emerging data types [75].

6.6. **MS Imaging (MSI).** Tremendous advancement in MSI has opened new avenues leading to the integration of single-cell omics and imaging approaches into large multi-omics/multimodal experiments. Galaxy offers a single platform to analyze this diverse data, already having tools for common 'omics and imaging techniques focused on proteomic and metabolomics [158], single-cell transcriptomics [159], traditional imaging, and multiplexed tissue imaging [160,161]. Accessible, cloud-based Galaxy public instances provide ample computing power, required not only for the ever-increasing file sizes due to higher mass and spatial resolution, but also for utilizing emerging machine and deep learning methods for MSI analysis [162].

6.7. **Single-cell proteomics.** Analysis of the molecular phenotypes of specific cell types (or 'single cell analysis') within complex, heterogeneous tissue samples holds great power in understanding disease pathogenesis and treatment [163,164]. Galaxy already offers a suite of tools for single-cell RNA-Seq data analysis [165]. Although still difficult, single-cell proteomics using MS and specialized sample preparation [166,167] and analysis approaches have begun to emerge [168]. We foresee a need for access to specialized single-cell proteomics analysis tools, integrated with other single-cell genomic tools, which Galaxy can facilitate.

6.8. **Strengthening community participation and value for MS-based proteomics informatics in Galaxy.** The GTN resources will expand as new MS-based proteomics applications and technologies emerge. Domain-specific gateways will guide users to available training material and software. Underlying infrastructure is being developed to ensure software tools and reference databases can be shared seamlessly across public servers. The Intergalactic Utilities Commission (IUC) and the Intergalactic Workflows Commission (IWC) seek to aid developers by vetting tools and workflows prior to public release. Guided by online documentation, developers can contribute their own tools and become part of the thriving global community which sustains these resources. We expect steady growth in users and contributors to Galaxy for MS-based proteomic informatics, collectively enabling more discoveries advancing studies in biology and medicine.

Declaration of Interest

T.J.G was supported by NIH 1U24CA199347; **T.J.G.** and **P.D.J.** were also supported by the Masonic Cancer Center at the University of Minnesota, an NCI-designated Comprehensive Cancer Center. **M.C.F.** was supported by the Hans A. Krebs Medical Scientist Programme, Faculty of Medicine, University of Freiburg. **Y.V:** Investissement d'Avenir Infrastructures Nationales en Biologie et Santé ANR-11-INBS-0013 (French Institute of Bioinformatics) and ANR-10-INBS-08 (Proteomics French Infrastructure - ProFI). **O.S** acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, projects 446058856, 466359513, 444936968 ,405351425, 431 336276, 431984000 (SFB 1453 "NephGen"), 441891347 (SFB 1479 "OncoEscape"), 423813989 (GRK 2606 "ProtPath"), 322977937 (GRK 2344 "MeInBio")), the ERA PerMed program (BMBF, 01KU1916, 01KU1915A), the German Consortium for Translational Cancer Research

(project Impro-Rec), the MatrixCode research group, FRIAS, Freiburg, the investBW program BW1_1198/03, the ERA TransCan program (project 01KT2201, “PREDICO”), the BMBF KMUi program (project 13GW0603E, project ESTHER), and the BMBF Cluster4Future program (nanodiag). **Carlos.H** is supported by the Bergen Research Foundation. **O.J.R.G** is supported by the Australian BioCommons which is enabled by NCRIS via Bioplatforms Australia funding. **C.H** is supported by the Australian BioCommons and the Queensland Cyber Infrastructure Foundation.

Funding

This paper was not funded.

Tables and Figures:

Table 1: Examples of MS-based Proteomics tools in Galaxy classified into various analysis categories.

Analysis Category	*Software Tools available in Galaxy*
DDA proteomics	SearchGUI/Peptide Shaker[27,28,169], MaxQuant[36], OpenMS[21], TransProteomic Pipeline[22], MSFragger/Fragipe (in development)[170]
DIA proteomics	OpenSWATH[46], diapysef[171], pyprophet[55], EncyclopeDIA[44], DIA Umpire[172]
Peptide/Protein Quantitation	FlashLFQ[173], moFF[174,175], MaxQuant[36], MSstats[176]
Peptide Verification	PepQuery[73]
MS Imaging	Cardinal[93], MALDIquant[94]
Multi-omics	metaQuantome[83,84], QuanTP[75], CustomProDB[68], MetaNovo[81], MT2MQ[177]
Taxonomic/Functional Annotation	Unipept[82], Blast-P[178], eggNOGmapper[179,180], MetaProteomeAnalyzer[181]
Statistical & Functional Analysis/Visualization	MSstats[176], TRIC[56], Multiomics Visualization platform[86], ProteoRE Enrichment & Pathways analysis[97–99]

Figure 1: Workflow extraction process from a Galaxy history. A: History with input data files and analysis steps. B: Select the "Extract workflow option" from the dropdown menu. C. List of tools that will be extracted to create the workflow. D. Editable workflow naming field. E. Preview of the extracted workflow which can be archived, shared, and further customized as desired.

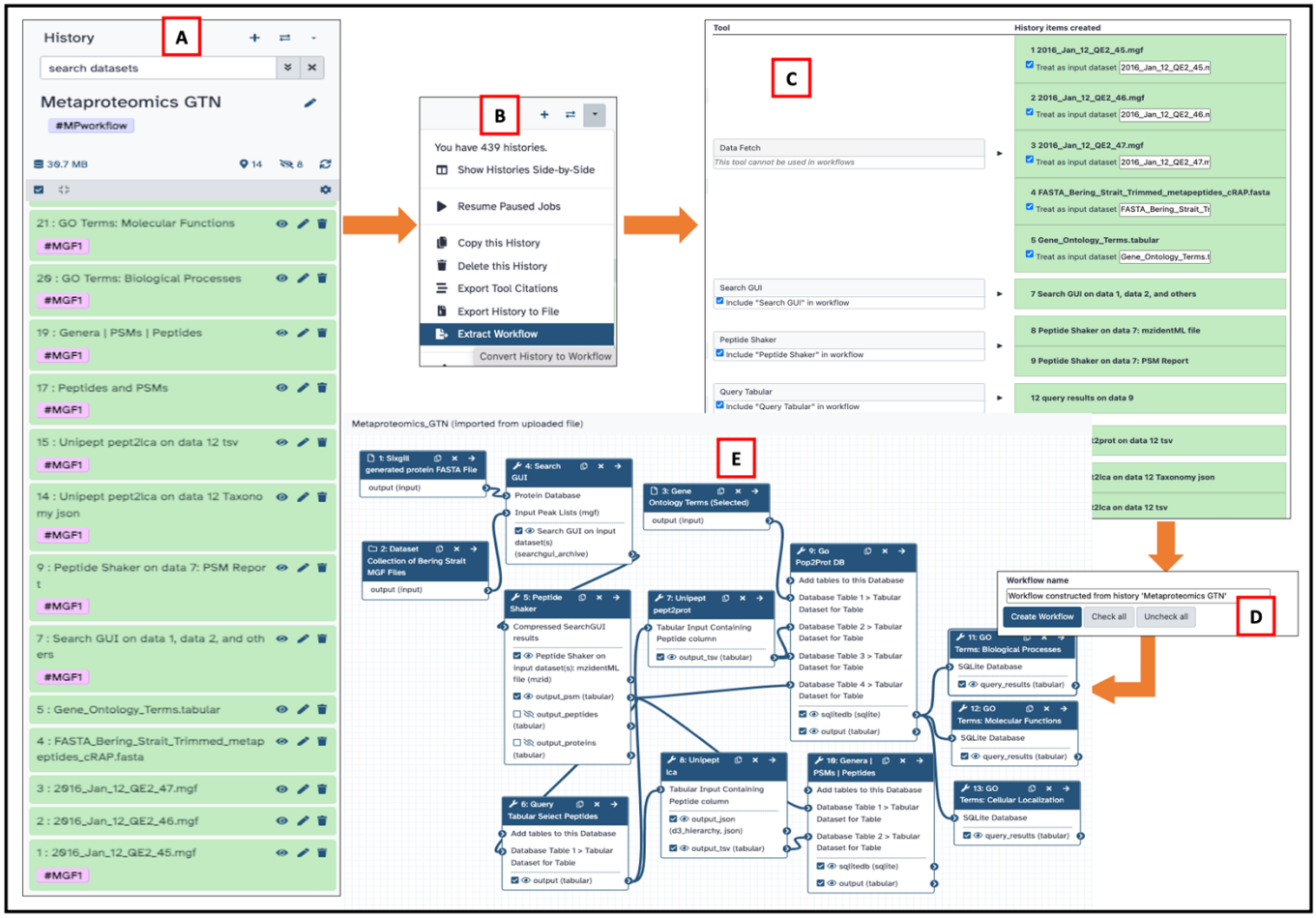


Figure 2: Overview of the Galaxy Ecosystem for MS-based proteomics studies

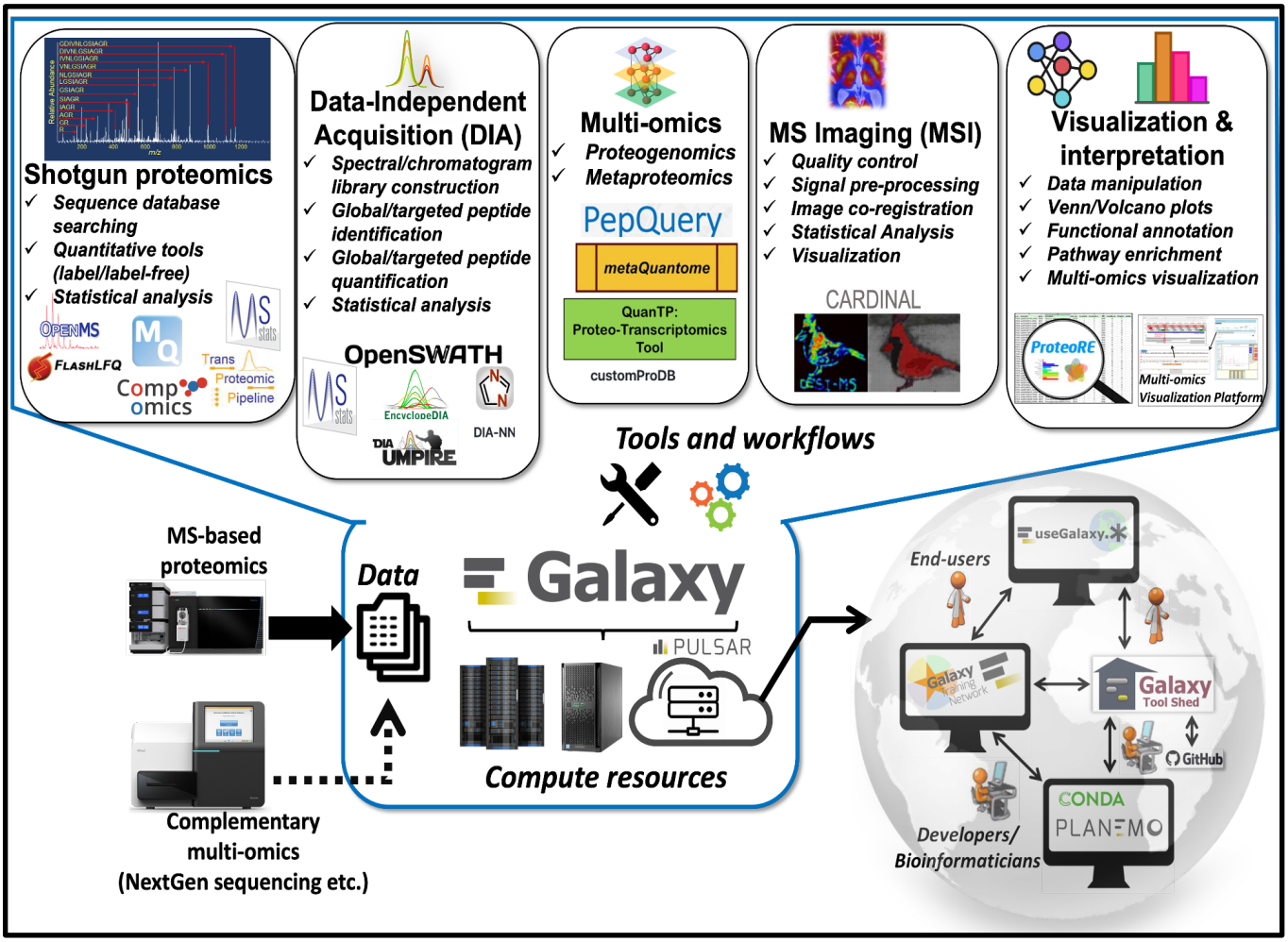


Figure 3: ProteoRE's Galaxy workspace with visualization components. (A.) **Navigation Bar:** provides tabs to switch between current workspace, workflows, visualization options, shared data libraries, and user repositories. (B.) **Tools Panel:** consists of ProteoRE and Galaxy proteomics tools along with text and data manipulation tools. (C.) **Central Panel:** Displays information from the user-selected tools. for e.g: (i) GO term enrichment analysis between up and down-regulated proteins. (ii) Reactome pathways analysis displaying the pathways topology of the proteins.(iii) Venn Diagram to show the protein overlap between two conditions. (D.) **History Panel:** shows results from the data analysis performed by the user along with metadata.

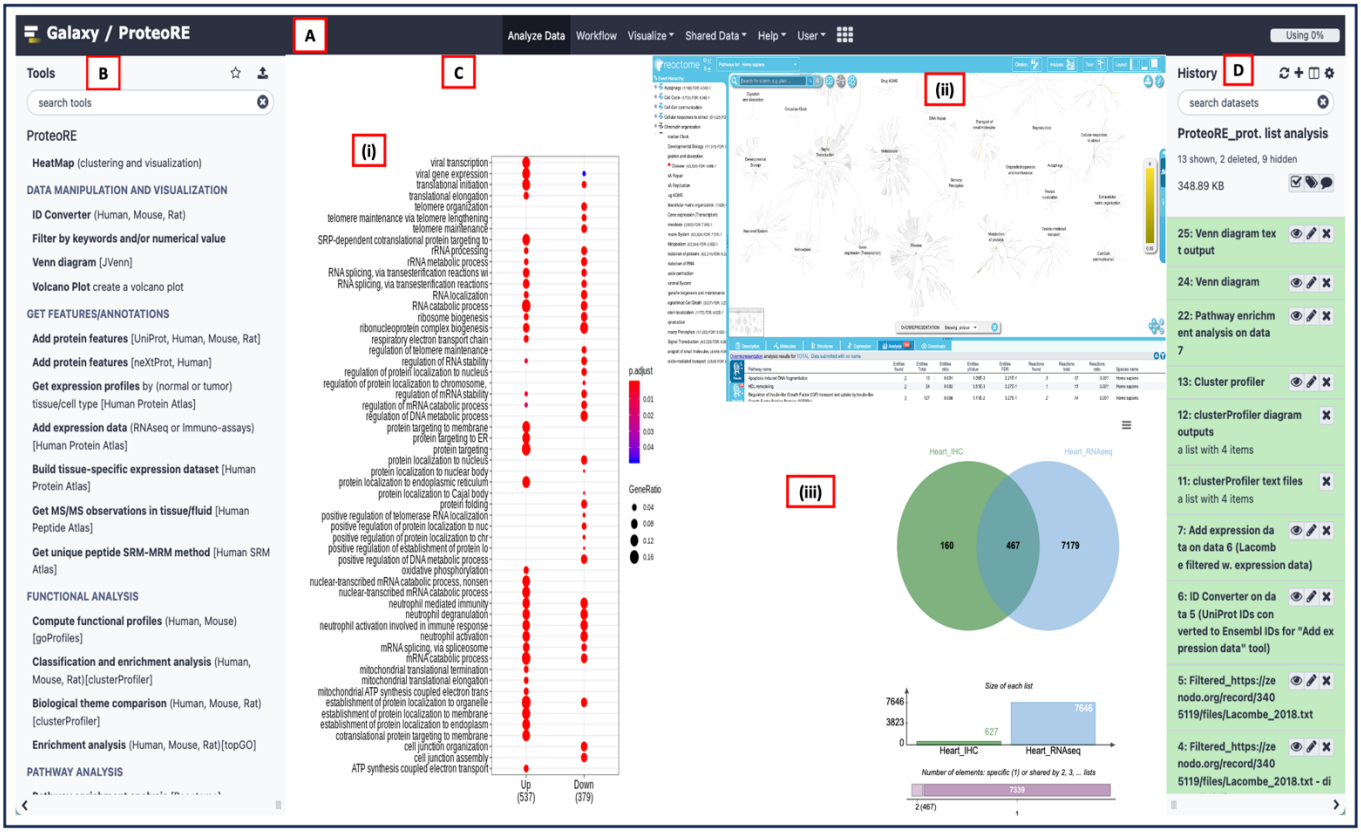
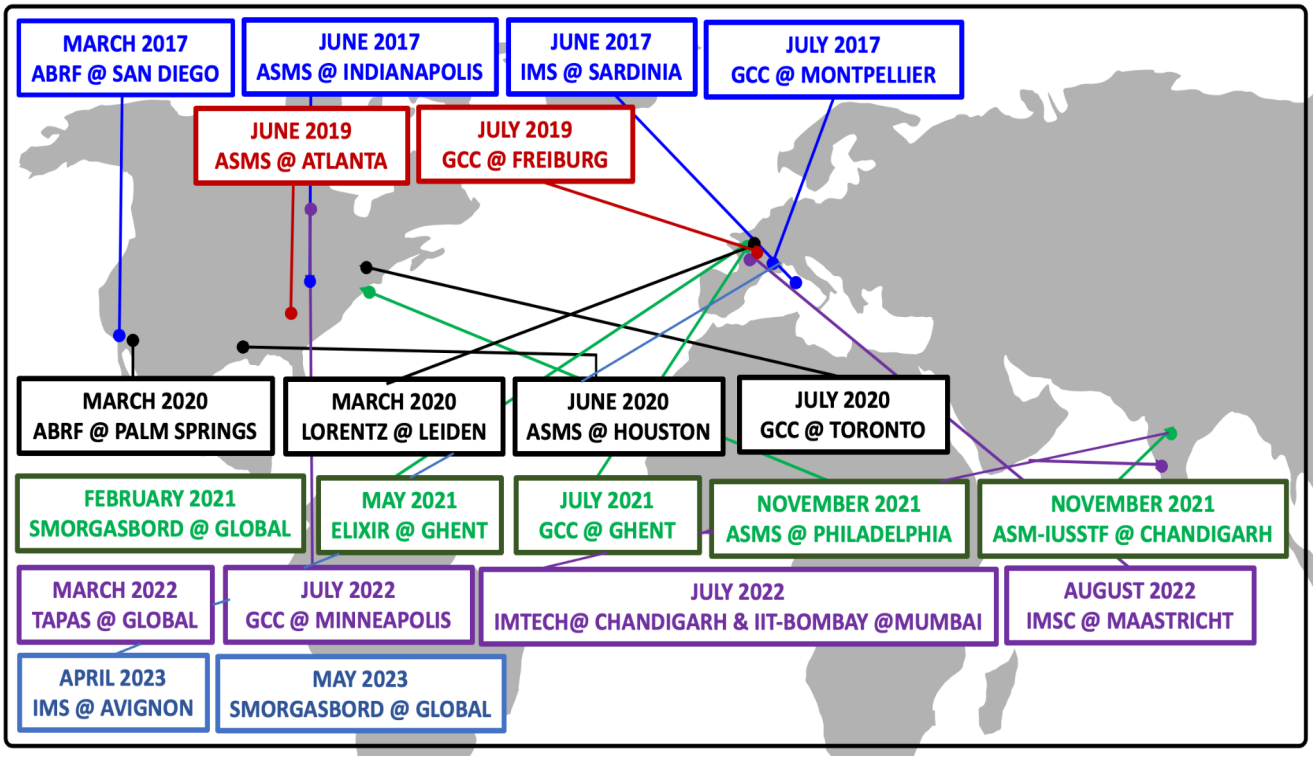


Figure 4: Galaxy MS-based metaproteomics workshops conducted over the years globally, exemplifying the use of GTN resources to support training activities aimed at new users.



References

[1] Sinitcyn P, Rudolph JD, Cox J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. https://doi.org/101146/annurev-biodatasci-080917-013516 [Internet]. 2018 [cited 2023 Jun 27];1:207–234. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-080917-013516.

[2] Jalili V, Afgan E, Gu Q, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 2021;48:W395–W402.

[3] Afgan E, Nekrutenko A, Grüning BABA, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res [Internet]. 2022 [cited 2022 Jun 24];50:W345–W351. Available from: https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkac247/6572001.

[4] Erratum: Correction to “The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update” (Nucleic acids research (2022)). Nucleic Acids Res. 2022;50:8999.

[5] Goble C, Soiland-Reyes S, Bacall F, et al. Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. 2021 [cited 2023 Aug 17]; Available from: https://zenodo.org/record/4605654.

[6] O’Connor BD, Yuen D, Chung V, et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. F1000Research 2017 652 [Internet]. 2017 [cited 2023 Aug 17];6:52. Available from: https://f1000research.com/articles/6-52.

[7] Lamprecht AL, Palmblad M, Ison J, et al. Perspectives on automated composition of workflows in the life sciences. F1000Research [Internet]. 2021 [cited 2023 Aug 15];10. Available from: https://pubmed.ncbi.nlm.nih.gov/34804501/.

[8] Connecting Galaxy to a compute cluster [Internet]. [cited 2023 Aug 15]. Available from: https://training.galaxyproject.org/training-material/topics/admin/tutorials/connect-to-compute-cluster/tutorial.html.

[9] Chappell K, Francou B, Habib C, et al. Galaxy Is a Suitable Bioinformatics Platform for the Molecular Diagnosis of Human Genetic Disorders Using High-Throughput Sequencing Data Analysis: Five Years of Experience in a Clinical Laboratory. Clin Chem. 2022;68:313–321.

[10] Bray SA, Senapathi T, Barnett CB, et al. Intuitive, reproducible high-throughput molecular dynamics in Galaxy: a tutorial. J Cheminform [Internet]. 2020 [cited 2023 Aug 15];12:54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33431030.

[11] Gangiredla J, Rand H, Benisatto D, et al. GalaxyTrakr: a distributed analysis tool for public health whole genome sequence data accessible to non-bioinformaticians. BMC Genomics. 2021;22.

[12] Giardine B, Riemer C, Hardison RC, et al. Galaxy: A platform for interactive large-scale genome analysis. Genome Res [Internet]. 2005 [cited 2020 Dec 7];15:1451–1455. Available from: https://pubmed.ncbi.nlm.nih.gov/16169926/.

[13] Galaxy Platform Directory: Servers, Clouds, and Deployable Resources - Galaxy Community Hub [Internet]. [cited 2023 Jun 27]. Available from: https://galaxyproject.org/use/.

[14] Galaxy: the first 10,000 pubs - Galaxy Community Hub [Internet]. [cited 2023 Jun 27]. Available from: https://galaxyproject.org/blog/2020-08-10k-pubs/.

[15] Galaxy | Tool Shed [Internet]. [cited 2022 Jul 3]. Available from: https://toolshed.g2.bx.psu.edu/.

[16] Blankenberg D, Von Kuster G, Bouvier E, et al. Dissemination of scientific software with Galaxy ToolShed. Genome Biol. 2014;15.

[17] Hiltemann S, Rasche H, Gladman S, et al. Galaxy Training: A powerful framework for teaching! PLoS Comput Biol. 2023;19:e1010752.

[18] Welcome to Pulsar’s documentation! — Pulsar 0.15.0.dev0 documentation [Internet]. [cited 2023 Jun 27]. Available from: https://pulsar.readthedocs.io/en/latest/.

[19] How it works — Total Perspective Vortex 2.2.4 documentation [Internet]. [cited 2023 Jun 27]. Available from: https://total-perspective-vortex.readthedocs.io/en/latest/.
[20] Schwämmle V, Harrow J, Ienasescu H. Proteomics Software in bio.tools: Coverage and Annotations. J Proteome Res [Internet]. 2021 [cited 2023 Jun 29];20:1821–1825. Available from: https://pubs.acs.org/doi/abs/10.1021/acs.jproteome.0c00978.
[21] Pfeuffer J, Sachsenberg T, Alka O, et al. OpenMS – A platform for reproducible analysis of mass spectrometry data. J Biotechnol. 2017;261:142–148.
[22] Deutsch EW, Mendoza L, Shteynberg D, et al. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics - Clin Appl. 2015;9:745–754.
[23] Barsnes H, Vaudel M, Colaert N, et al. Compomics-utilities: An open-source Java library for computational proteomics. BMC Bioinformatics. 2011;12.
[24] FragPipe | A complete proteomics pipeline with the MSFragger search engine at heart [Internet]. [cited 2023 Feb 23]. Available from: https://fragpipe.nesvilab.org/.
[25] Peptide and Protein ID using OpenMS tools [Internet]. [cited 2023 Jun 27]. Available from: https://training.galaxyproject.org/training-material/topics/proteomics/tutorials/protein-id-oms/tutorial.html.
[26] Zhang Y, Fonslow BR, Shan B, et al. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013;113:2343–2394.
[27] Barsnes H, Vaudel M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. J Proteome Res. 2018;17:2552–2555.
[28] Farag YM, Horro C, Vaudel M, et al. PeptideShaker Online: A User-Friendly Web-Based Framework for the Identification of Mass Spectrometry-Based Proteomics Data. J Proteome Res. 2021;20:5419–5423.
[29] Bihani S, Gupta A, Mehta S, et al. Metaproteomic Analysis of Nasopharyngeal Swab Samples to Identify Microbial Peptides in COVID-19 Patients. J Proteome Res [Internet]. 2023 [cited 2023 Aug 15];22:2608–2619. Available from: http://www.ncbi.nlm.nih.gov/pubmed/37450889.
[30] Sajulga R, Easterly C, Riffle M, et al. Survey of metaproteomics software tools for functional microbiome analysis. PLoS One. 2020;15:e0241503.
[31] Van Riper SK, de Jong EP, Carlis J V., et al. Mass spectrometry-based proteomics: basic principles and emerging technologies and directions. Adv Exp Med Biol. 2013;990:1–35.
[32] He T, Liu Y, Zhou Y, et al. Comparative Evaluation of Proteome Discoverer and FragPipe for the TMT-Based Proteome Quantification. J Proteome Res. 2022;21:3007–3015.
[33] Thompson A, Schäfer J, Kuhn K, et al. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003;75:1895–1904.
[34] Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3:1154–1169.
[35] Mehta S, Easterly CWCW, Sajulga R, et al. Precursor Intensity-Based Label-Free Quantification Software Tools for Proteomic and Multi-Omic Analysis within the Galaxy Platform. Proteomes [Internet]. 2020 [cited 2023 Mar 19];8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32650610.
[36] Cox J, Hein MY, Luber CA, et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics. 2014;13:2513–2526.
[37] Yu F, Haynes SE, Nesvizhskii AI. IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. Mol Cell Proteomics. 2021;20.
[38] Choi M, Chang CY, Clough T, et al. MSstats: an R package for statistical analysis of quantitative

mass spectrometry-based proteomic experiments. Bioinformatics [Internet]. 2014 [cited 2023 Feb 23];30:2524–2526. Available from: https://dx.doi.org/10.1093/bioinformatics/btu305.
[39] Huang T, Choi M, Tzouros M, et al. MSstatsTMT: Statistical Detection of Differentially Abundant Proteins in Experiments with Isobaric Labeling and Multiple Mixtures. Mol Cell Proteomics [Internet]. 2020 [cited 2023 Jun 27];19:1706–1723. Available from: http://www.mcponline.org/article/S1535947620351148/fulltext.
[40] Pinter N, Glätzer D, Fahrner M, et al. MaxQuant and MSstats in Galaxy Enable Reproducible Cloud-Based Analysis of Quantitative Proteomics Experiments for Everyone. J Proteome Res. 2022;21:1558–1565.
[41] Pino LK, Just SC, MacCoss MJ, et al. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. Mol Cell Proteomics. 2020;19:1088–1103.
[42] Ludwig C, Gillet L, Rosenberger G, et al. Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial . Mol Syst Biol. 2018;14.
[43] Doerr A. DIA mass spectrometry. Nat Methods 2015 121 [Internet]. 2014 [cited 2023 Jun 27];12:35–35. Available from: https://www.nature.com/articles/nmeth.3234.
[44] Searle BC, Pino LK, Egertson JD, et al. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nat Commun. 2018;9.
[45] Demichev V, Messner CB, Vernardis SI, et al. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods [Internet]. 2020 [cited 2023 Jun 27];17:41. Available from: /pmc/articles/PMC6949130/.
[46] Röst HL, Rosenberger G, Navarro P, et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol. 2014;32:219–223.
[47] Bruderer R, Bernhardt OM, Gandhi T, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics. 2015;14:1400–1410.
[48] Kelstrup CD, Bekker-Jensen DB, Arrey TN, et al. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. J Proteome Res. 2018;17:727–738.
[49] Bekker-Jensen DB, Bernhardt OM, Hogrebe A, et al. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. Nat Commun. 2020;11.
[50] Muntel J, Gandhi T, Verbeke L, et al. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. Mol Omi. 2019;15:348–360.
[51] Lin L, Zheng J, Yu Q, et al. High throughput and accurate serum proteome profiling by integrated sample preparation technology and single-run data independent mass spectrometry analysis. J Proteomics. 2018;174:9–16.
[52] Skowronek P, Thielert M, Voytik E, et al. Rapid and In-Depth Coverage of the (Phospho-) Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. Mol Cell Proteomics. 2022;21.
[53] Röst HL, Aebersold R, Schubert OT. Automated swath data analysis using targeted extraction of ion chromatograms. Methods Mol Biol. 2017;1550:289–307.
[54] Schubert OT, Gillet LC, Collins BC, et al. Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc. 2015;10:426–441.
[55] Teleman J, Röst HL, Rosenberger G, et al. DIANA-algorithmic improvements for analysis of data-independent acquisition MS data. Bioinformatics. 2015;31:555–562.
[56] Röst HL, Liu Y, D'Agostino G, et al. TRIC: An automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat Methods. 2016;13:777–783.
[57] Blattmann P, Heusel M, Aebersold R. SWATH2stats: An R/bioconductor package to process and

convert quantitative SWATH-MS proteomics data for downstream analysis tools. PLoS One. 2016;11.
[58] Demichev V, Szyrwiel L, Yu F, et al. dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. Nat Commun. 2022;13.
[59] Jagtap P, Johnson J, Yang T-Y, et al. Exploring chromatogram library-based data-independent acquisition analysis using EncyclopeDIA within Galaxy framework. . F1000Research. 2020;9.
[60] Fahrner M, Föll MC, Grüning BA, et al. Democratizing data-independent acquisition proteomics analysis on public cloud infrastructures via the Galaxy framework. Gigascience. 2022;11.
[61] MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics [Internet]. 2010 [cited 2023 Aug 15];26:966–968. Available from: https://pubmed.ncbi.nlm.nih.gov/20147306/.
[62] Boekel J, Chilton JM, Cooke IR, et al. Multi-omic data analysis using Galaxy. Nat Biotechnol. 2015;33:137–139.
[63] Fan J, Saha S, Barker G, et al. Galaxy integrated omics: Web-based standards-compliant workflows for proteomics informed by transcriptomics. Mol Cell Proteomics. 2015;14:3087–3093.
[64] Chambers MC, Jagtap PD, Johnson JE, et al. An Accessible Proteogenomics Informatics Resource for Cancer Researchers. 2017 [cited 2017 Dec 18];77:e43–e46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29092937.
[65] Sheynkman GM, Johnson JE, Jagtap PD, et al. Using Galaxy-P to leverage RNA-Seq for the discovery of novel protein variations. BMC Genomics. 2014;15.
[66] Crappé J, Ndah E, Koch A, et al. PROTEOFORMER: Deep proteome coverage through ribosome profiling and MS integration. Nucleic Acids Res. 2015;43.
[67] Nesvizhskii AI. Proteogenomics: Concepts, applications and computational strategies. Nat Methods. 2014;11:1114–1125.
[68] Wang X, Zhang B, Wren J. CustomProDB: An R package to generate customized protein databases from RNA-Seq data for proteomics search. Bioinformatics. 2013;29:3235–3237.
[69] Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–915.
[70] Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–295.
[71] Shumate A, Wong B, Pertea G, et al. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLoS Comput Biol. 2022;18.
[72] Reference Data with CVMFS [Internet]. [cited 2023 Jun 27]. Available from: https://training.galaxyproject.org/training-material/topics/admin/tutorials/cvmfs/tutorial.html.
[73] Wen B, Wang X, Zhang B. PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations. Genome Res [Internet]. 2019 [cited 2022 Jun 24];29:485–493. Available from: https://genome.cshlp.org/content/29/3/485.full.
[74] Hari PS, Balakrishnan L, Kotyada C, et al. Proteogenomic Analysis of Breast Cancer Transcriptomic and Proteomic Data, Using De Novo Transcript Assembly: Genome-Wide Identification of Novel Peptides and Clinical Implications. Mol Cell Proteomics. 2022;21.
[75] Kumar P, Panigrahi P, Johnson J, et al. QuanTP: A Software Resource for Quantitative Proteo-Transcriptomic Comparative Data Analysis and Informatics. J Proteome Res [Internet]. 2019 [cited 2020 Nov 30];18:782–790. Available from: https://pubmed.ncbi.nlm.nih.gov/30582332/.
[76] Wang Z, Maschera B, Lea S, et al. Airway host-microbiome interactions in chronic obstructive pulmonary disease. Respir Res. 2019;20.
[77] Van T, Bossche D, Kunath BJ, et al. Critical Assessment of Metaproteome Investigation (CAMPI): A Multi-Lab Comparison of Established Workflows. bioRxiv [Internet]. 2021 [cited 2023 Jun

27];2021.03.05.433915. Available from: https://www.biorxiv.org/content/10.1101/2021.03.05.433915v1.
[78] Galaxy Europe | Integrative meta-omics analysis - Metagenomics, Metatranscriptomics, Metaproteomics [Internet]. [cited 2023 Jun 29]. Available from: https://usegalaxy-eu.github.io/posts/2020/04/14/integrative-meta-omics/.
[79] Jagtap P, Goslinga J, Kooren JA, et al. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. Proteomics. 2013;13:1352–1357.
[80] Kumar P, Johnson JEJE, Easterly C, et al. A Sectioning and Database Enrichment Approach for Improved Peptide Spectrum Matching in Large, Genome-Guided Protein Sequence Databases. J Proteome Res. 2020;19:2772–2785.
[81] Potgieter MG, Nel AJM, Fortuin S, et al. MetaNovo: An open-source pipeline for probabilistic peptide discovery in complex metaproteomic datasets. Ioshikhes I, editor. PLoS Comput Biol [Internet]. 2023 [cited 2023 Jun 20];19:e1011163. Available from: https://pubmed.ncbi.nlm.nih.gov/37327214/.
[82] Gurdeep Singh R, Tanca A, Palomba A, et al. Unipept 4.0: Functional Analysis of Metaproteome Data. J Proteome Res. 2019;18:606–615.
[83] Easterly CW, Sajulga R, Mehta S, et al. MetaQuantome: An integrated, quantitative metaproteomics approach reveals connections between taxonomy and protein function in complex microbiomes. Mol Cell Proteomics. 2019;18:S82–S91.
[84] Mehta S, Kumar P, Crane M, et al. Updates on metaQuantome Software for Quantitative Metaproteomics. J Proteome Res [Internet]. 2021 [cited 2021 Mar 24];20. Available from: https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00960.
[85] Thuy-Boun PS, Mehta S, Gruening B, et al. Metaproteomics Analysis of SARS-CoV-2-Infected Patient Samples Reveals Presence of Potential Coinfecting Microorganisms. J Proteome Res [Internet]. 2021 [cited 2022 Jun 24];20:1451–1454. Available from: /pmc/articles/PMC7805602/.
[86] Mcgowan T, Johnson JEJE, Kumar P, et al. Multi-omics Visualization Platform: An extensible Galaxy plug-in for multi-omics data visualization and exploration. Gigascience. 2020;9.
[87] Si T, Li B, Zhang K, et al. Characterization of Bacillus subtilis Colony Biofilms via Mass Spectrometry and Fluorescence Imaging. J Proteome Res [Internet]. 2016 [cited 2023 Jun 29];15:1955–1962. Available from: https://pubmed.ncbi.nlm.nih.gov/27136705/.
[88] Poth AG, Mylne JS, Grassl J, et al. Cyclotides associate with leaf vasculature and are the products of a novel precursor in Petunia (Solanaceae). J Biol Chem [Internet]. 2012 [cited 2023 Jun 27];287:27033–27046. Available from: http://www.jbc.org/article/S0021925820478974/fulltext.
[89] Ong TH, Romanova E V., Roberts-Galbraith RH, et al. Mass spectrometry imaging and identification of peptides associated with cephalic ganglia regeneration in Schmidtea mediterranea. J Biol Chem [Internet]. 2016 [cited 2023 Jun 27];291:8109–8120. Available from: http://www.jbc.org/article/S0021925820407756/fulltext.
[90] Vaysse PM, Heeren RMA, Porta T, et al. Mass spectrometry imaging for clinical research – latest developments, applications, and current limitations. Analyst [Internet]. 2017 [cited 2023 Jun 27];142:2690–2712. Available from: https://pubs.rsc.org/en/content/articlehtml/2017/an/c7an00565b.
[91] Rešetar Maslov D, Svirkova A, Allmaier G, et al. Optimization of MALDI-TOF mass spectrometry imaging for the visualization and comparison of peptide distributions in dry-cured ham muscle fibers. Food Chem. 2019;283:275–286.
[92] Buchberger AR, DeLaney K, Johnson J, et al. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. Anal Chem. 2018;90:240–265.
[93] Bemis KD, Harry A, Eberlin LS, et al. Cardinal: An R package for statistical analysis of mass

spectrometry-based imaging experiments. Bioinformatics. 2015;31:2418–2420.
[94] Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. Bioinformatics. 2012;28:2270–2271.
[95] Föll MC, Moritz L, Wollmann T, et al. Accessible and reproducible mass spectrometry imaging data analysis in Galaxy. Gigascience [Internet]. 2019 [cited 2023 Jun 27];8:1–12. Available from: https://dx.doi.org/10.1093/gigascience/giz143.
[96] Johnson JEJE, Kumar P, Easterly C, et al. Improve your Galaxy text life: The Query Tabular Tool [version 1; referees: 1 approved, 2 approved with reservations]. F1000Research. 2018;7.
[97] Combes F, Loux V, Vandenbrouck Y. GO Enrichment Analysis for Differential Proteomics Using ProteoRE. Methods Mol Biol. 2021;2361:179–196.
[98] Nguyen L, Brun V, Combes F, et al. Designing an in silico strategy to select tissue-leakage biomarkers using the galaxy framework. Methods Mol Biol. 2019;1959:275–289.
[99] Galaxy | ProteoRE [Internet]. [cited 2023 Jun 27]. Available from: https://proteore.org/.
[100] Kusebauch U, Campbell DS, Deutsch EW, et al. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. Cell. 2016;166:766–778.
[101] Masica DL, Douville C, Tokheim C, et al. CRAVAT 4: Cancer-related analysis of variants toolkit. Cancer Res. 2017;77:e35–e38.
[102] Sajulga R, Mehta S, Kumar P, et al. Bridging the Chromosome-centric and Biology/Disease-driven Human Proteome Projects: Accessible and Automated Tools for Interpreting the Biological and Pathological Impact of Protein Sequence Variants Detected via Proteogenomics. J Proteome Res. 2018;17:4329–4336.
[103] Shah AD, Goode RJA, Huang C, et al. Lfq-Analyst: An easy-To-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with maxquant. J Proteome Res [Internet]. 2019 [cited 2023 Jun 29];204–211. Available from: https://pubs.acs.org/doi/abs/10.1021/acs.jproteome.9b00496.
[104] Galaxy Training! [Internet]. [cited 2023 Jun 27]. Available from: https://training.galaxyproject.org/training-material/topics/proteomics/.
[105] Galaxy Community Conferences (GCCs) - Galaxy Community Hub [Internet]. [cited 2023 Jun 27]. Available from: https://galaxyproject.org/.
[106] GTN Smörgåsbord: A Global Galaxy Course - Galaxy Community Hub [Internet]. [cited 2023 Mar 20]. Available from: https://galaxyproject.org/events/2021-02-smorgasbord/.
[107] GTN Smörgåsbord 2: 14-18 March | Gallantries [Internet]. [cited 2023 Mar 20]. Available from: https://gallantries.github.io/posts/2021/12/14/smorgasbord2-tapas/.
[108] COVID-19 analysis on usegalaxy.★ [Internet]. [cited 2022 Jul 3]. Available from: https://covid19.galaxyproject.org/.
[109] Rajczewski AT, Mehta S, Nguyen DDA, et al. A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19). Clin Proteomics [Internet]. 2021 [cited 2023 Aug 15];18:15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33971807.
[110] Mehta S, Carvalho VM, Rajczewski AT, et al. Catching the Wave: Detecting Strain-Specific SARS-CoV-2 Peptides in Clinical Samples Collected during Infection Waves from Diverse Geographical Locations. Viruses. 2022;14.
[111] Jagtap PD, Blakely A, Murray K, et al. Metaproteomic analysis using the Galaxy framework. Proteomics. 2015;15:3553–3565.
[112] Chambers MC, Jagtap PD, Johnson JE, et al. An accessible proteogenomics informatics resource for cancer researchers. Cancer Res. 2017;77:e43–e46.
[113] Blank C, Easterly C, Gruening B, et al. Disseminating metaproteomic informatics capabilities and knowledge using the galaxy-P framework. Proteomes [Internet]. 2018 [cited 2019 Oct 8];6.

Available from: http://www.ncbi.nlm.nih.gov/pubmed/29385081.
[114] Jagtap PD, Viken KJ, Johnson J, et al. BAL fluid metaproteome in acute respiratory failure. Am J Respir Cell Mol Biol. 2018;59:648–652.
[115] Afiuni-Zadeh S, Boylan KLM, Jagtap PD, et al. Evaluating the potential of residual Pap test fluid as a resource for the metaproteomic analysis of the cervical-vaginal microbiome. Sci Rep. 2018;8.
[116] Jagtap PD, Johnson JE, Onsongo G, et al. Flexible and accessible workflows for improved proteogenomic analysis using the galaxy framework. J Proteome Res. 2014;13:5898–5908.
[117] Maringer K, Yousuf A, Heesom KJ, et al. Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in Aedes aegypti. BMC Genomics. 2017;18.
[118] Guillot L, Delage L, Viari A, et al. Peptimapper: Proteogenomics workflow for the expert annotation of eukaryotic genomes. BMC Genomics. 2019;20.
[119] Olexiouk V, Crappé J, Verbruggen S, et al. SORFs.org: A repository of small ORFs identified by ribosome profiling. Nucleic Acids Res. 2016;44:D324–D329.
[120] Vermillion KL, Jagtap P, Johnson JE, et al. Characterizing cardiac molecular mechanisms of mammalian hibernation via quantitative proteogenomics. J Proteome Res. 2015;14:4792–4804.
[121] Anderson KJ, Vermillion KL, Jagtap P, et al. Proteogenomic Analysis of a Hibernating Mammal Indicates Contribution of Skeletal Muscle Physiology to the Hibernation Phenotype. J Proteome Res. 2016;15:1253–1261.
[122] Rajczewski AT, Han Q, Mehta S, et al. Quantitative Proteogenomic Characterization of Inflamed Murine Colon Tissue Using an Integrated Discovery, Verification, and Validation Proteogenomic Workflow. Proteomes [Internet]. 2022 [cited 2023 Feb 23];10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35466239.
[123] Rudney JD, Jagtap PD, Reilly CS, et al. Protein relative abundance patterns associated with sucrose-induced dysbiosis are conserved across taxonomically diverse oral microcosm biofilm models of dental caries. Microbiome. 2015;3:69.
[124] Edwinson AL, Yang L, Peters S, et al. Gut microbial β-glucuronidases regulate host luminal proteases and are depleted in irritable bowel syndrome. Nat Microbiol. 2022;7:680–694.
[125] Jiang X, Zhang Y, Wang H, et al. In-Depth Metaproteomics Analysis of Oral Microbiome for Lung Cancer. 2022 [cited 2023 Feb 23];2022:9781578. Available from: /pmc/articles/PMC9590273/.
[126] Van Den Bossche T, Arntzen M, Becher D, et al. The Metaproteomics Initiative: a coordinated approach for propelling the functional characterization of microbiomes. Microbiome. 2021;9.
[127] Van Den Bossche T, Kunath BJ, Schallert K, et al. Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. Nat Commun. 2021;12.
[128] Saito MA, Bertrand EM, Duffy ME, et al. Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. J Proteome Res. 2019;18:1461–1476.
[129] Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: Protein domains identifier. Nucleic Acids Res. 2005;33.
[130] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353–D361.
[131] Cantarel BI, Coutinho PM, Rancurel C, et al. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 2009;37.
[132] Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol [Internet]. 2016 [cited 2023 Jun 29];34:525–527. Available from: https://pubmed.ncbi.nlm.nih.gov/27043002/.
[133] magnusarntzen/ViMO: Visualizer for Meta-omics [Internet]. [cited 2023 Jun 29]. Available from: https://github.com/magnusarntzen/ViMO.

[134] Stillger M, Li M, Hönscheid P, et al. Advancing rare cancer research by MALDI mass spectrometry imaging: Applications, challenges and future perspectives in sarcoma. Authorea Prepr [Internet]. 2023 [cited 2023 Jun 27]; Available from: https://www.authorea.com/users/628120/articles/648844-advancing-rare-cancer-research-by-maldi-mass-spectrometry-imaging-applications-challenges-and-future-perspectives-in-sarcoma?commit=d585085c8db0fcd6c820cd9bb8c9b2b6ed45e34e.
[135] Berghmans E, Boonen K, Maes E, et al. Implementation of MALDI Mass Spectrometry Imaging in Cancer Proteomics Research: Applications and Challenges. J Pers Med 2020, Vol 10, Page 54 [Internet]. 2020 [cited 2023 Jun 27];10:54. Available from: https://www.mdpi.com/2075-4426/10/2/54/htm.
[136] Arentz G, Mittal P, Zhang C, et al. Applications of Mass Spectrometry Imaging to Cancer. Adv Cancer Res. 2017;134:27–66.
[137] Föll MC, Volkmann V, Enderle-Ammour K, et al. Moving translational mass spectrometry imaging towards transparent and reproducible data analyses: a case study of an urothelial cancer cohort analyzed in the Galaxy framework. Clin Proteomics [Internet]. 2022 [cited 2023 Jun 27];19:1–14. Available from: https://clinicalproteomicsjournal.biomedcentral.com/articles/10.1186/s12014-022-09347-z.
[138] Moritz L, Stillger MN, Takács F, et al. Characterization of Spatial Heterogeneity in Metastasized Colorectal Cancer by MALDI Imaging. 2023 [cited 2023 Jun 27]; Available from: https://www.preprints.org/manuscript/202302.0363/v1.
[139] Gustafsson OJR, Winderbaum LJ, Condina MR, et al. Balancing sufficiency and impact in reporting standards for mass spectrometry imaging experiments. Gigascience [Internet]. 2018 [cited 2023 Jun 27];7:1–13. Available from: https://dx.doi.org/10.1093/gigascience/giy102.
[140] Barbieux C, Bonnet des Claustres M, Fahrner M, et al. Netherton syndrome subtypes share IL-17/IL-36 signature with distinct IFN-α and allergic responses. J Allergy Clin Immunol. 2022;149:1358–1372.
[141] Messner CB, Demichev V, Wang Z, et al. Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. Proteomics. 2022;
[142] Stephens ZD, Lee SY, Faghri F, et al. Big Data: Astronomical or Genomical? PLoS Biol [Internet]. 2015 [cited 2023 Jun 27];13. Available from: /pmc/articles/PMC4494865/.
[143] Perez-Riverol Y, Moreno P. Scalable data analysis in proteomics and metabolomics using BioContainers and workflows engines. bioRxiv [Internet]. 2019 [cited 2023 Jun 27];604413. Available from: https://www.biorxiv.org/content/10.1101/604413v1.
[144] Frewen BE, Merrihew GE, Wu CC, et al. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal Chem. 2006;78:5678–5684.
[145] Gessulat S, Schmidt T, Zolg DP, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat Methods. 2019;16:509–518.
[146] Searle BC, Shannon AE, Wilburn DB. Scribe: Next Generation Library Searching for DDA Experiments. J Proteome Res. 2023;22:482–490.
[147] Mohajeri MH, Brummer RJM, Rastall RA, et al. The role of the microbiome for human health: from basic science to clinical applications. Eur J Nutr [Internet]. 2018 [cited 2020 Nov 5];57:1–14. Available from: https://pubmed.ncbi.nlm.nih.gov/29748817/.
[148] Jagtap P. A METAPROTEOMICS BIOINFORMATICS WORKFLOW TO STUDY HOST- MICROBE DYNAMICS IN CLINICAL SAMPLES. 2023 [cited 2023 Feb 23]; Available from: https://zenodo.org/record/7671203.
[149] Jouffret V, Miotello G, Culotta K, et al. Increasing the power of interpretation for soil metaproteomics data. Microbiome. 2021;9.
[150] Fernandes MLP, Bastida F, Jehmlich N, et al. Functional soil mycobiome across ecosystems. J

Proteomics. 2022;252.
[151] Armengaud J. Unique Insights into How Plants and Soil Microbiomes Interact Are at Our Fingertips. mSystems [Internet]. 2022 [cited 2023 Jun 27];7:e0058922. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35975919.
[152] CAMPI 2 - Metaproteomics Initiative [Internet]. [cited 2023 Jun 27]. Available from: https://metaproteomics.org/campi/campi2/.
[153] Meier F, Brunner AD, Frank M, et al. diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. Nat Methods. 2020;17:1229–1236.
[154] Patel A, McGrosso D, Hefner Y, et al. Proteome allocation is linked to transcriptional regulation through a modularized transcriptome. bioRxiv Prepr Serv Biol [Internet]. 2023 [cited 2023 Mar 21]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/36865326.
[155] Wang Z, Mülleder M, Batruch I, et al. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. Elife. 2022;11.
[156] Meier F, Brunner AD, Koch S, et al. Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Mol Cell Proteomics. 2018;17:2534–2545.
[157] Heil LR, Damoc E, Arrey TN, et al. Evaluating the performance of the Astral mass analyzer for quantitative proteomics using data independent acquisition. bioRxiv [Internet]. 2023 [cited 2023 Jun 27];2023.06.03.543570. Available from: https://www.biorxiv.org/content/10.1101/2023.06.03.543570v1.
[158] Peters K, Bradbury J, Bergmann S, et al. PhenoMeNal: processing and analysis of metabolomics data in the cloud. Gigascience [Internet]. 2019 [cited 2023 Jun 27];8:1–12. Available from: https://dx.doi.org/10.1093/gigascience/giy149.
[159] Tekman M, Batut B, Ostrovsky A, et al. A single-cell RNA-sequencing training and analysis suite using the Galaxy framework. Gigascience [Internet]. 2020 [cited 2023 Jun 27];9:1–9. Available from: https://dx.doi.org/10.1093/gigascience/giaa102.
[160] Creason AL, Watson C, Gu Q, et al. A Web-based Software Resource for Interactive Analysis of Multiplex Tissue Imaging Datasets. bioRxiv [Internet]. 2022 [cited 2023 Jun 27];2022.08.18.504436. Available from: https://www.biorxiv.org/content/10.1101/2022.08.18.504436v2.
[161] Schapiro D, Sokolov A, Yapp C, et al. MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. Nat Methods [Internet]. 2022 [cited 2023 Jun 27];19:311–315. Available from: https://pubmed.ncbi.nlm.nih.gov/34824477/.
[162] Alexandrov T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. https://doi.org/101146/annurev-biodatasci-011420-031537 [Internet]. 2020 [cited 2023 Jun 27];3:61–87. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-biodatasci-011420-031537.
[163] Fernandez DM, Rahman AH, Fernandez NF, et al. Single-cell immune landscape of human atherosclerotic plaques. Nat Med. 2019;25:1576–1588.
[164] Li Y, Ma C, Liao S, et al. Combined proteomics and single cell RNA-sequencing analysis to identify biomarkers of disease diagnosis and disease exacerbation for systemic lupus erythematosus. Front Immunol. 2022;13.
[165] Tekman M, Batut B, Ostrovsky A, et al. A single-cell RNA-seq Training and Analysis Suite using the Galaxy Framework. bioRxiv [Internet]. 2020 [cited 2023 Jun 27];2020:2020.06.06.137570. Available from: https://www.biorxiv.org/content/10.1101/2020.06.06.137570v4.
[166] Eshghi A, Xie X, Hardie D, et al. Sample Preparation Methods for Targeted Single-Cell Proteomics. J Proteome Res. 2022;
[167] Swennenhuis JF, Terstappen L. Sample preparation methods following cellsearch approach compatible of single-cell whole-genome amplification: An overview. Whole Genome Amplif

Methods Protoc. 2015;57–67.
[168] Schoof EM, Furtwängler B, Üresin N, et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat Commun 2021 121 [Internet]. 2021 [cited 2023 Jun 27];12:1–15. Available from: https://www.nature.com/articles/s41467-021-23667-y.
[169] Vaudel M, Burkhart JM, Zahedi RP, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets: To the editor [Internet]. Nat. Biotechnol. Nature Publishing Group; Jan 1, 2015 p. 22–24. Available from: https://pubmed.ncbi.nlm.nih.gov/25574629/.
[170] Kong AT, Leprevost F V., Avtonomov DM, et al. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods. 2017;14:513–520.
[171] Diapysef :: Anaconda.org [Internet]. [cited 2023 Jun 29]. Available from: https://anaconda.org/bioconda/diapysef.
[172] Zhang F, Ge W, Ruan G, et al. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. Proteomics. 2020;20.
[173] Millikin RJ, Solntsev SK, Shortreed MR, et al. Ultrafast Peptide Label-Free Quantification with FlashLFQ. J Proteome Res. 2018;17:386–391.
[174] Argentini A, Goeminne LJE, Verheggen K, et al. MoFF: A robust and automated approach to extract peptide ion intensities. Nat. Methods. Nature Publishing Group; 2016. p. 964–966.
[175] Argentini A, Staes AA, Grüning B, et al. Update on the moFF Algorithm for Label-Free Quantitative Proteomics. J Proteome Res. 2019;18:728–731.
[176] Kohler D, Staniak M, Tsai TH, et al. MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. J Proteome Res [Internet]. 2022 [cited 2023 Jun 27]; Available from: https://pubs.acs.org/doi/full/10.1021/acs.jproteome.2c00834.
[177] Mehta S, Crane M, Leith E, et al. ASaiM-MT: a validated and optimized ASaiM workflow for metatranscriptomics analysis within Galaxy framework. F1000Research [Internet]. 2021 [cited 2021 Feb 17];10:103. Available from: https://f1000research.com/articles/10-103/v1.
[178] Altschul SF. A protein alignment scoring system sensitive at all evolutionary distances. J Mol Evol [Internet]. 1993 [cited 2023 Jun 29];36:290–300. Available from: https://pubmed.ncbi.nlm.nih.gov/8483166/.
[179] Huerta-Cepas J, Szklarczyk D, Heller D, et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res [Internet]. 2019 [cited 2020 Dec 17];47:D309–D314. Available from: https://pubmed.ncbi.nlm.nih.gov/30418610/.
[180] Cantalapiedra CP, Hern̨andez-Plaza A, Letunic I, et al. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38:5825–5829.
[181] Muth T, Behne A, Heyer R, et al. The MetaProteomeAnalyzer: A powerful open-source software suite for metaproteomics data analysis and interpretation. J Proteome Res. 2015;14:1557–1565.