“Mass spectrometry-based proteomics centered multi-omics: applications to clinical lung disease research”

Timothy J. Griffin
University of Minnesota
Professor, Department of Biochemistry, Molecular Biology and Biophysics
Faculty Director, Center for Metabolomics and Proteomics

tgriffin@umn.edu

GalaxyP
galaxyp.org

Acknowledgements

Biochemistry,
Molecular Biology & Biophysics

Pratik Jagtap (Co-leader, Galaxy-P)
Subina Mehta
Andrew Rajczewski
Danielle Weise
Monica Kruk
Kevin Viken

Minnesota Supercomputing Institute

Reid Wagner
James Johnson

Collaborators
Chris Wendt, MD
Maneesh Bhargava, MD,PhD
Terri Laguna, MD
Laurie Parker, PhD
Eric Lock, PhD (Sarah Samorodnitsky, PhD)

COMMUNITY-BASED SOFTWARE DEVELOPMENT

Harald Barsnes and Marc Vaudel
University of Bergen, Bergen, Norway
Bjoern Gruening (Galaxy community…)
University of Freiburg, Freiburg, Germany
Alexey Nesvizhskii
Fengchao Yu
University of Michigan
Bing Zhang
Baylor College of Medicine
Lennart Martens
VIB Department of Medical Protein Research, UGent, Belgium
Lloyd Smith/Michael Shortreed
University of Wisconsin-Madison

GalaxyP
galaxyp.org

NSF

NIH
National Institutes of Health

Masonic Cancer Center
University of Minnesota
Comprehensive Cancer Center designated by the National Cancer Institute

Outline

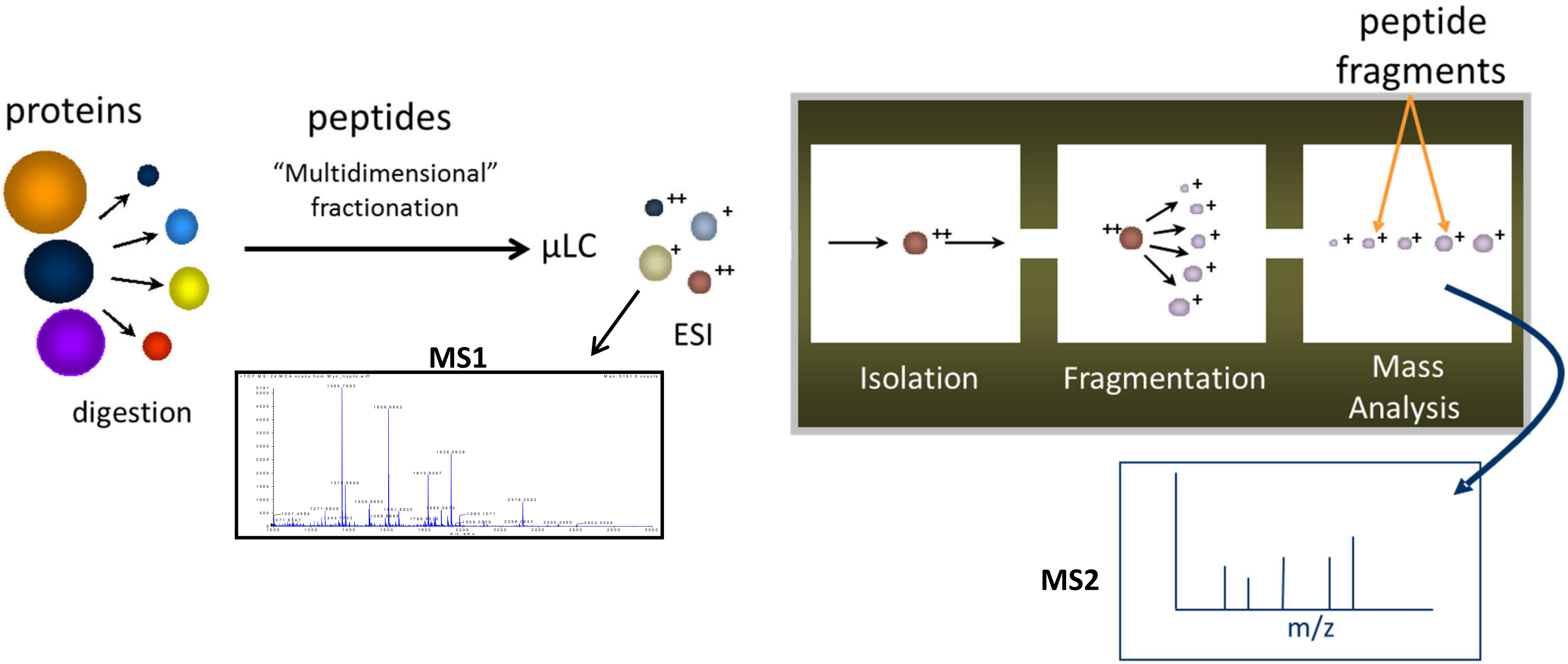
- **Translational mass spectrometry-based proteomics studies of lung disease and health: technology platforms and workflow development**

- **Application to studies of obstructive lung disease at the intersection of HIV:**
    - Quantitative **proteomic** profiling of bronchoalveolar lavage fluid (BALF)
    - **Peptidomics:** characterizing protease activity in lung disease

- **Metaproteomics: microbial contributors to host disease and biology**
  - Value of metaproteomics
  - Challenges and technology platforms
  - Applications to cystic fibrosis

Mass spectrometry-based proteomics: core technology platform

Peptide fractionation coupled to tandem mass spectrometry (MS/MS)



proteins

digestion

peptides

"Multidimensional" fractionation

µLC

ESI

MS1

peptide fragments

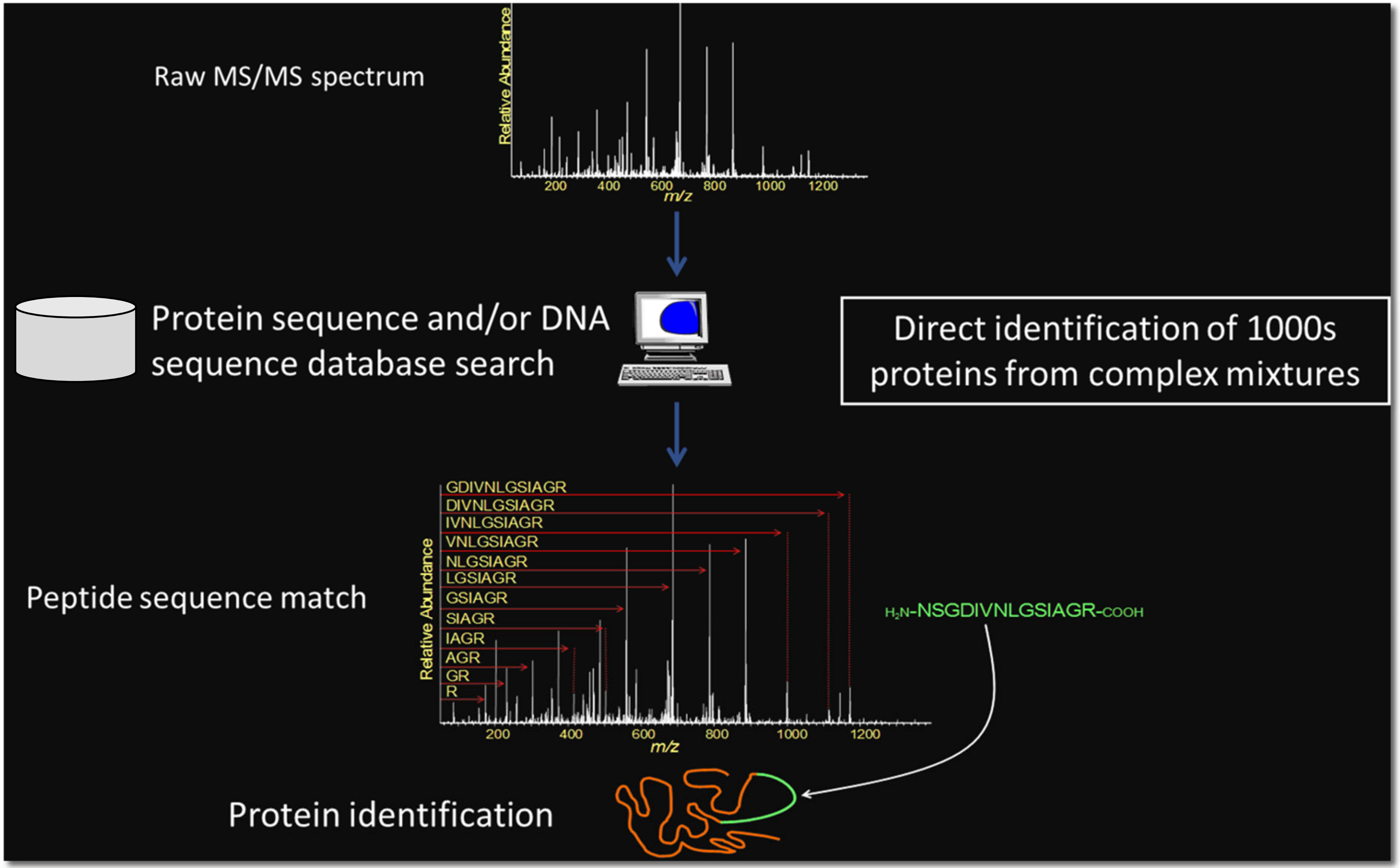
Isolation

Fragmentation

Mass Analysis

MS2

Matching amino acid sequences to MS/MS data



Raw MS/MS spectrum

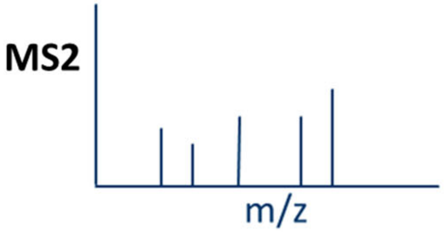
Protein sequence and/or DNA sequence database search

Direct identification of 1000s proteins from complex mixtures

Peptide sequence match

Protein identification

Inferring protein identity for peptide sequences

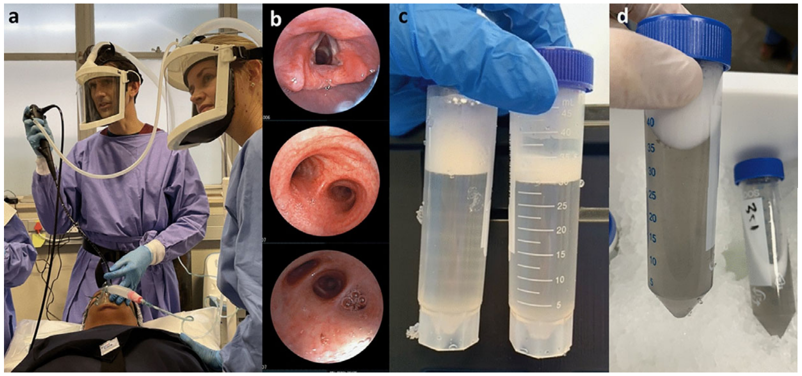


CAQCHTVEK

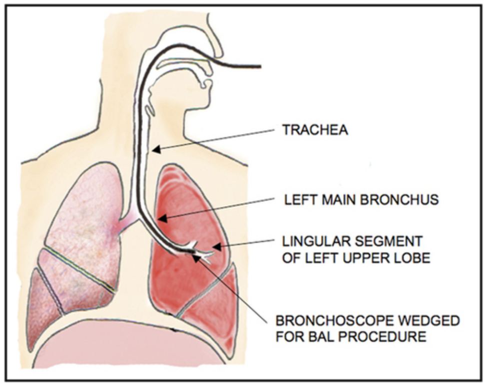
Cytochrome C

NH$_2$GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFT
YTDANKNKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLK
KATNECOOH

Application to clinical lung samples: bronchoalveolar lavage fluid (BALF)



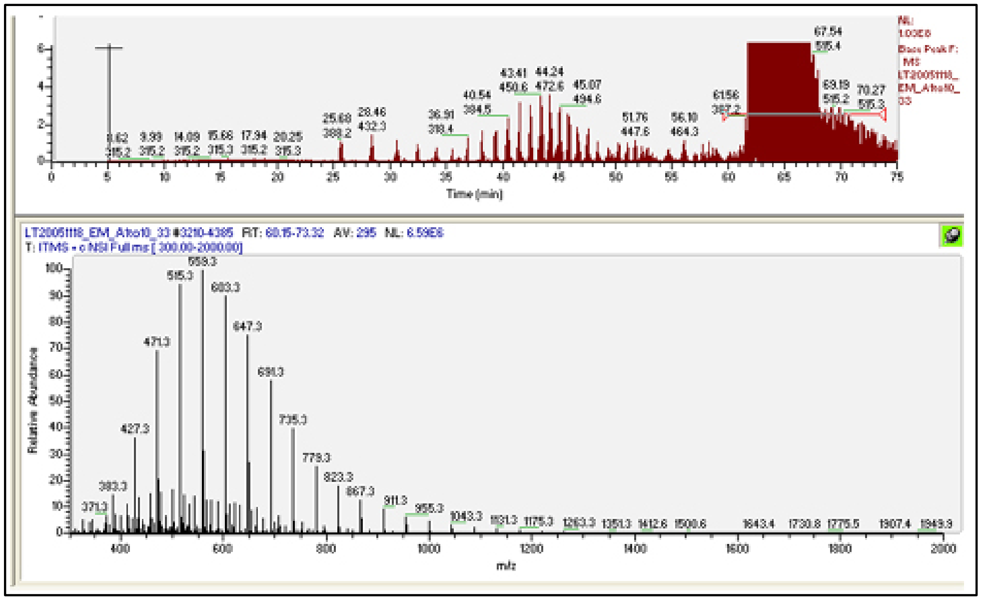
Sci Rep 13, 8859 (2023)



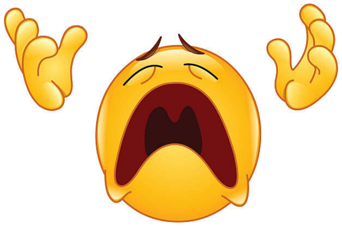
Proteomics Clin Appl. 3, 1044 (2009)

Challenges for MS-based proteomics analysis of BALF

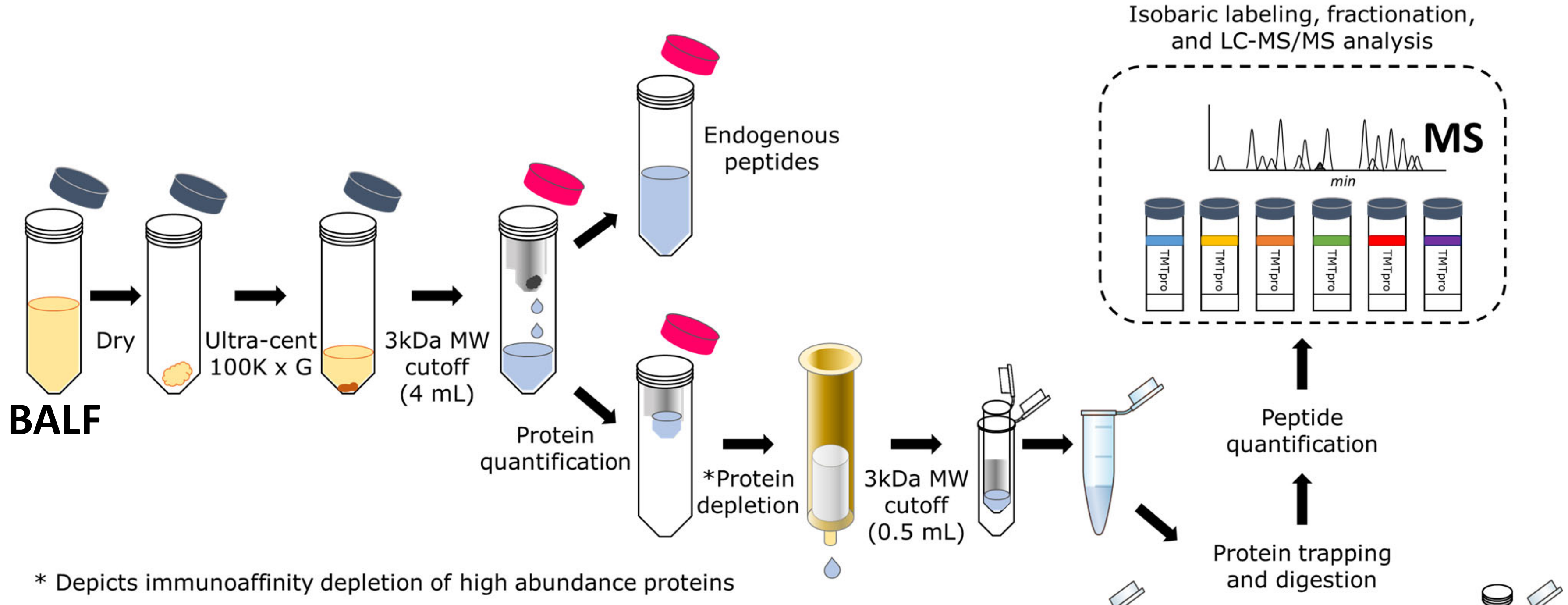
- Dominated by high abundance proteins (primarily from plasma)
- Relatively large volumes of fluid requiring concentration
- Removal of potential contaminants: surfactants, lipids



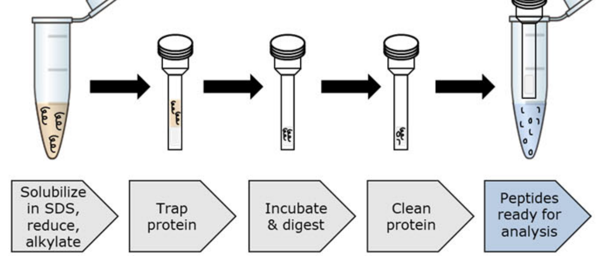
https://www.ohsu.edu/proteomics-shared-resource/sample-preparation

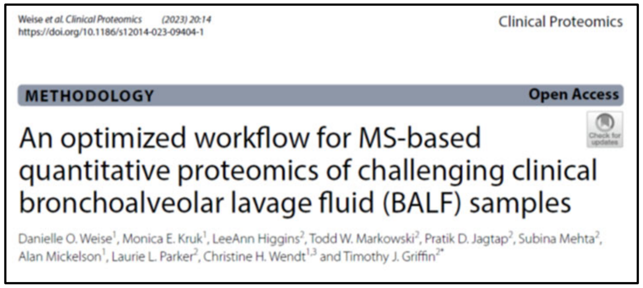


A workflow for MS-based proteomics of BALF: years in the making



* Depicts immunoaffinity depletion of high abundance proteins




Weise *et al. Clinical Proteomics* (2023) 20:14
https://doi.org/10.1186/s12014-023-09404-1

Clinical Proteomics

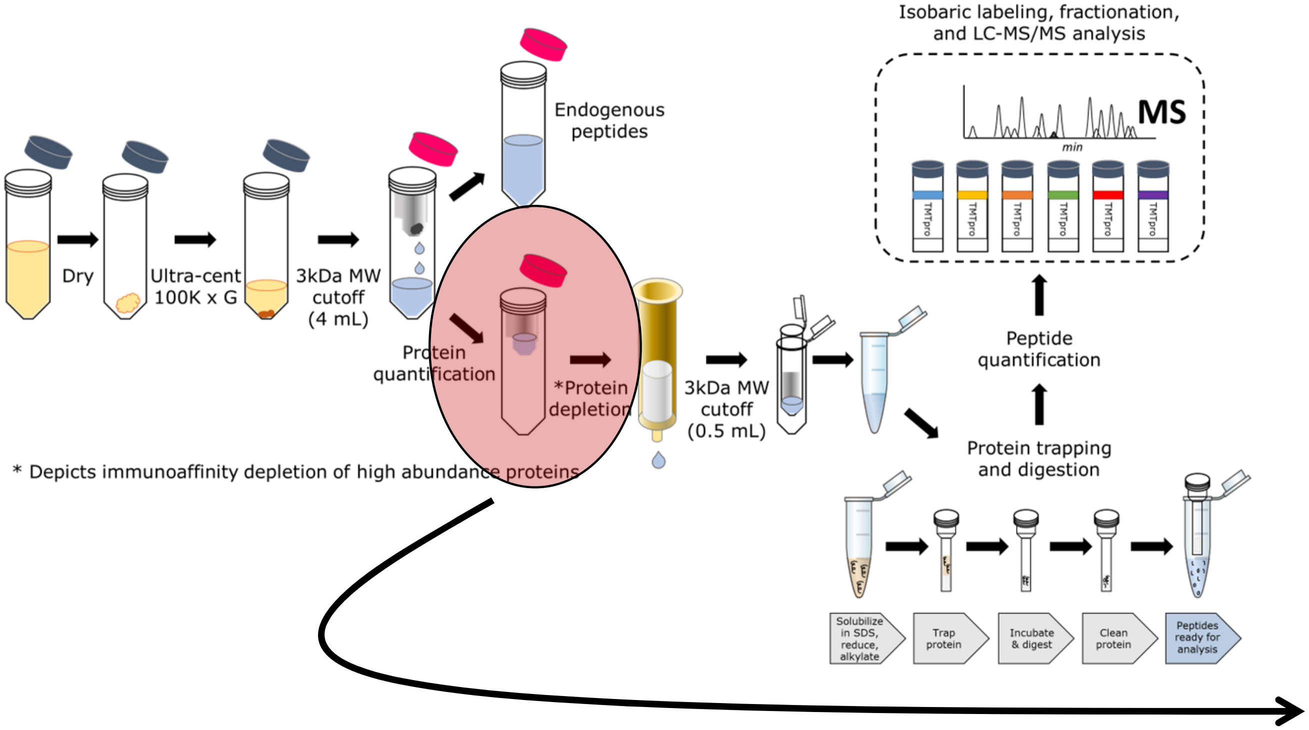
METHODOLOGY **Open Access**

An optimized workflow for MS-based quantitative proteomics of challenging clinical bronchoalveolar lavage fluid (BALF) samples

Danielle O. Weise[1], Monica E. Kruk[1], LeeAnn Higgins[2], Todd W. Markowski[2], Pratik D. Jagtap[2], Subina Mehta[2], Alan Mickelson[1], Laurie L. Parker[2], Christine H. Wendt[1,3] and Timothy J. Griffin[2*]

(Chris Wendt, Maneesh Bhargava)

Challenge 1: High abundant proteins



* Depicts immunoaffinity depletion of high abundance proteins

<u>The “high peaks”– plasma proteins:</u>

- serum albumin (90% of total protein?)
- Transferrin
- alpha-1 antitrypsin
- immunoglobulin A and G

Proteomics Clin Appl. 2014 8:737

Limits protein detection to only a few hundred in BALF

Protein depletion spin columns: works for BALF!

"High Select Top14"

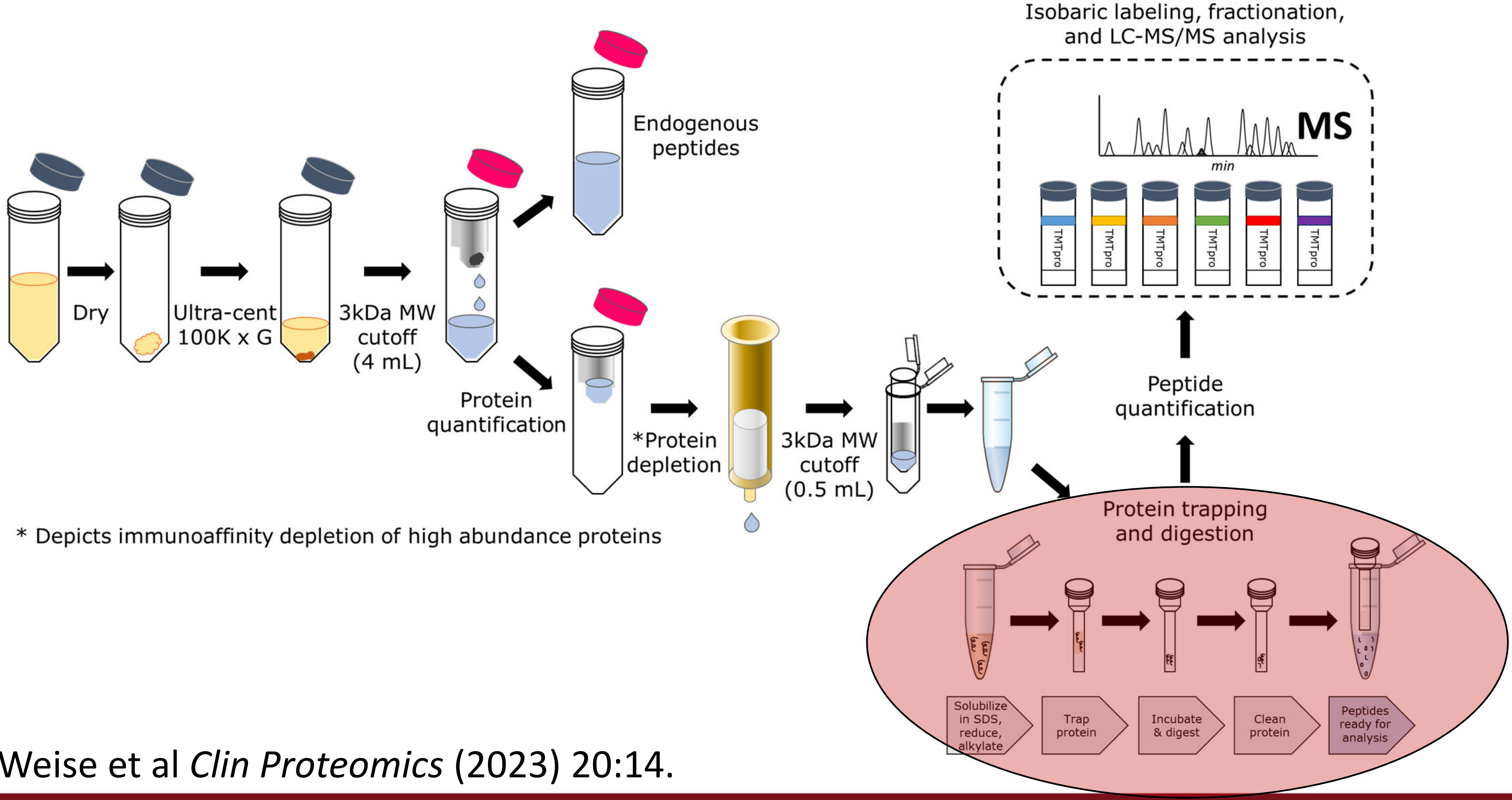
300 proteins

No depletion

2000+ proteins

With depletion

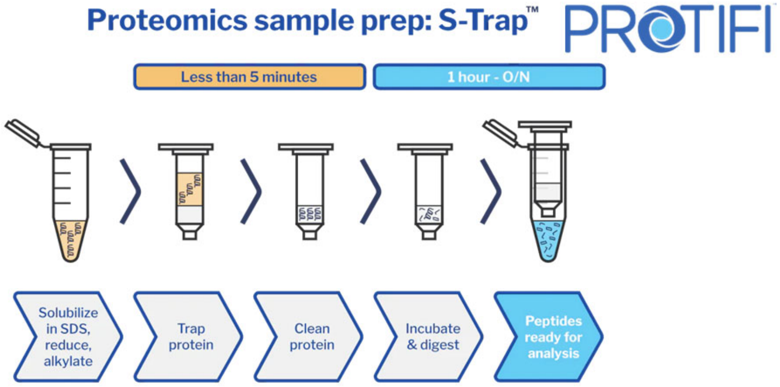
Challenge: MS-incompatible contaminants

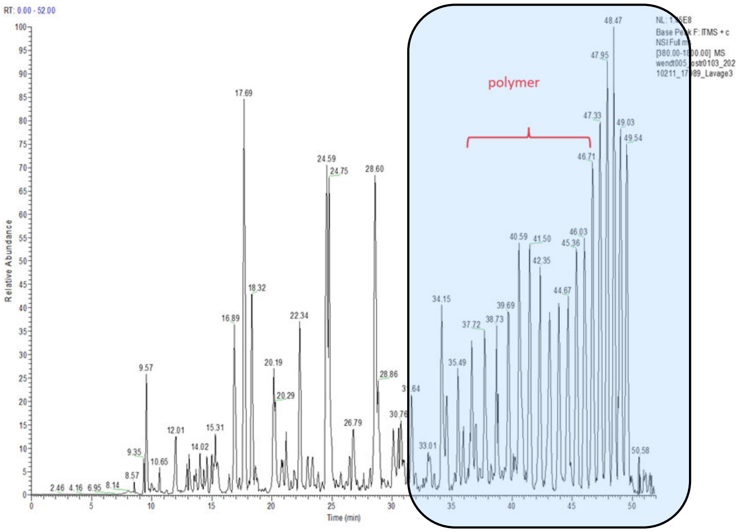


* Depicts immunoaffinity depletion of high abundance proteins

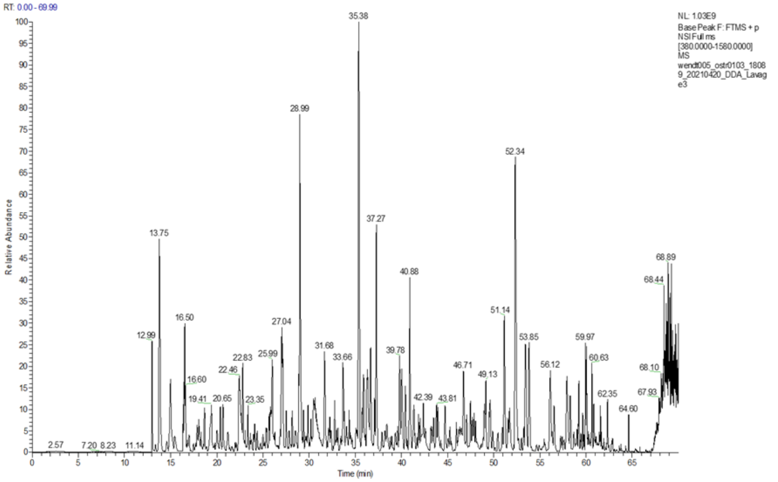
Weise et al *Clin Proteomics* (2023) 20:14.

A workflow to meet the challenges of BALF for MS-based protoemics



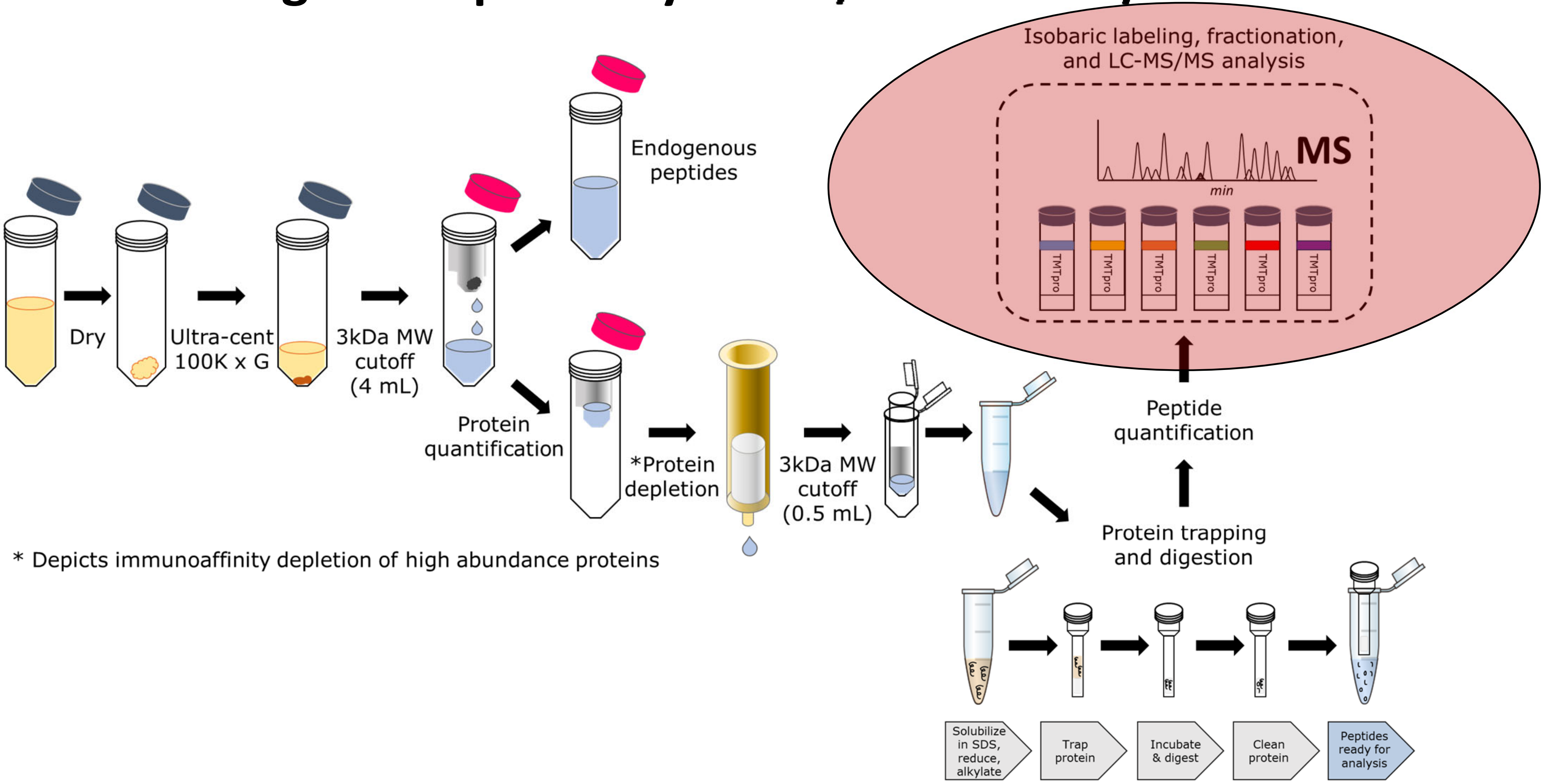


No S-trap



With S-trap

Challenge: Compatibility with quantitative proteomics



Isobaric labeling, fractionation, and LC-MS/MS analysis

Dry

Ultra-cent 100K x G

3kDa MW cutoff (4 mL)

Endogenous peptides

Protein quantification

*Protein depletion

3kDa MW cutoff (0.5 mL)

Protein trapping and digestion

Peptide quantification

Solubilize in SDS, reduce, alkylate

Trap protein

Incubate & digest

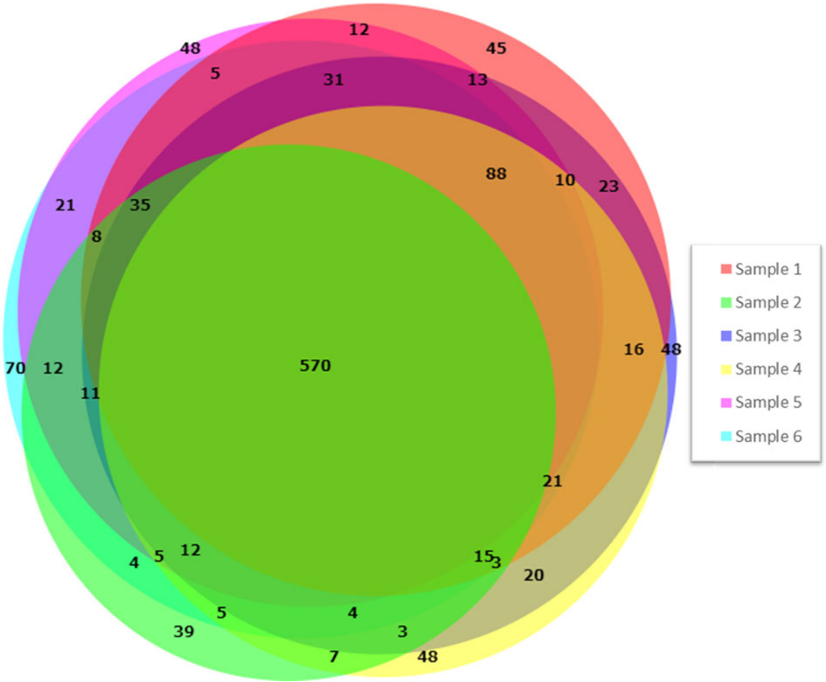
Clean protein

Peptides ready for analysis

* Depicts immunoaffinity depletion of high abundance proteins

Weise et al *Clin Proteomics* (2023) 20:14.

Challenge: Compatibility with quantitative proteomics



Protein identifications across 6 separate BALF samples – pilot study

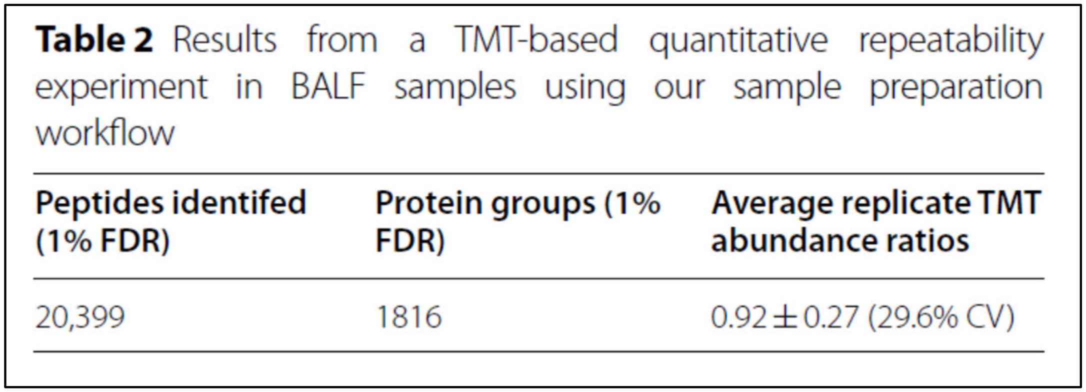


Table 2 Results from a TMT-based quantitative repeatability experiment in BALF samples using our sample preparation workflow

Peptides identifed (1% FDR)	**Protein groups (1% FDR)**	**Average replicate TMT abundance ratios**
20,399	1816	0.92 ± 0.27 (29.6% CV)

GalaxyP
galaxyp.org

Weise et al *Clin Proteomics* (2023) 20:14.

Putting the workflow to use

ERJ open research

THE BEST IN OPEN ACCESS BASIC,
TRANSLATIONAL & CLINICAL
RESPIRATORY RESEARCH

The lung proteome in HIV-associated obstructive lung disease (OLD)

Sarah Samorodnitsky, Danielle Weise, Eric F. Lock, Ken M. Kunisaki, Alison Morris, Janice M. Leung,
Monica Kruk, Laurie Parker, Pratik Jagtap, Timothy J. Griffin, Chris H. Wendt

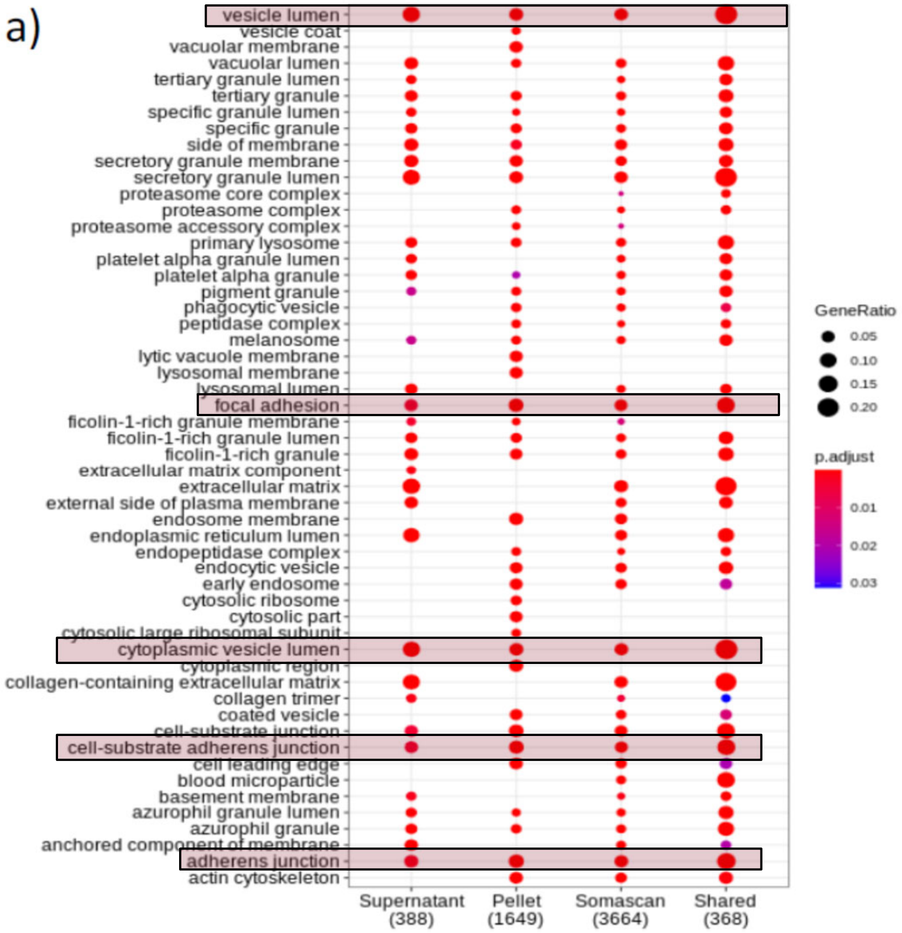
- Quantitative proteomics in BALF samples from persons living with HIV with OLD (n=26) and without OLD (n=26)
- Pearson correlation tests to identify associations between each protein and lung function (FEV1pp)
- **Goal:** Identify pathways and/or protein markers predictive of OLD development

---

Summary of results

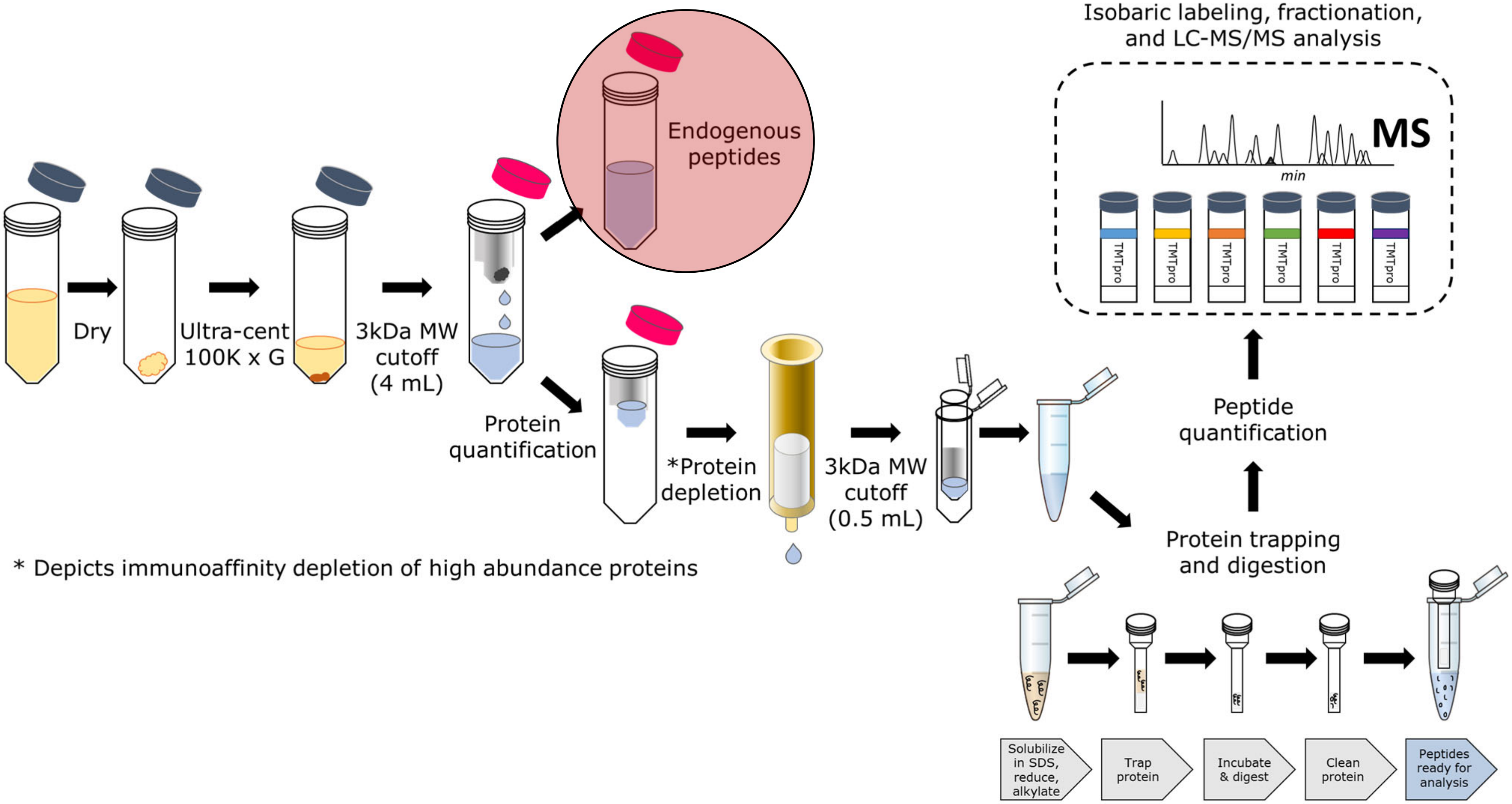
- 3800 proteins identified by MS (cellular and soluble BALF)
- 254 proteins that correlated with FEV1pp controlling for smoking status

Outcome: Cell adhesion proteins as potential predictor of OLD



ERJ Open Research 2024 00204-2024

Application: Peptidomic analysis of BALF



Dry

Ultra-cent 100K x G

3kDa MW cutoff (4 mL)

Endogenous peptides

Protein quantification

*Protein depletion

3kDa MW cutoff (0.5 mL)

Protein trapping and digestion

Peptide quantification

Isobaric labeling, fractionation, and LC-MS/MS analysis

Solubilize in SDS, reduce, alkylate

Trap protein

Incubate & digest

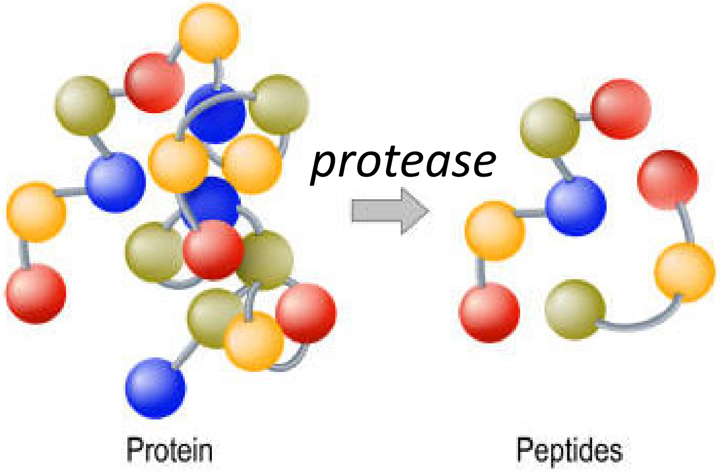
Clean protein

Peptides ready for analysis

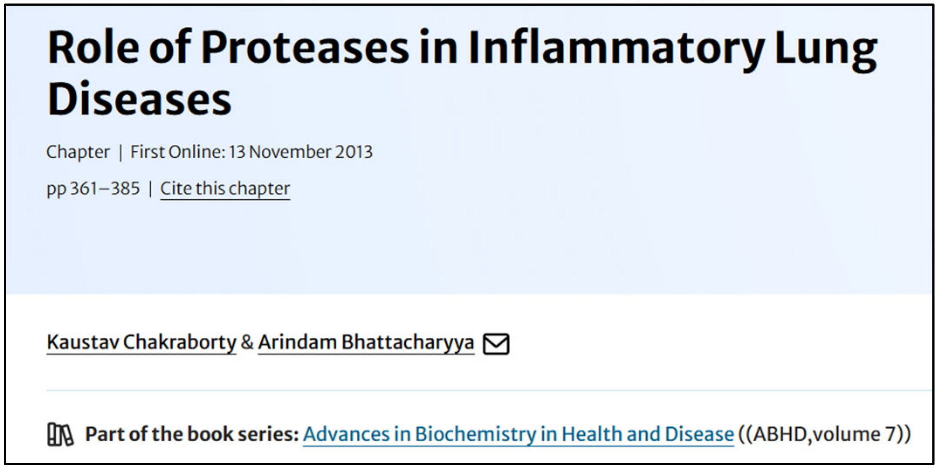
* Depicts immunoaffinity depletion of high abundance proteins

Weise et al *Clin Proteomics* (2023) 20:14.

Value of peptidomics in clinical samples



Identifying endogenous peptides from clinical samples enables understanding of active proteases



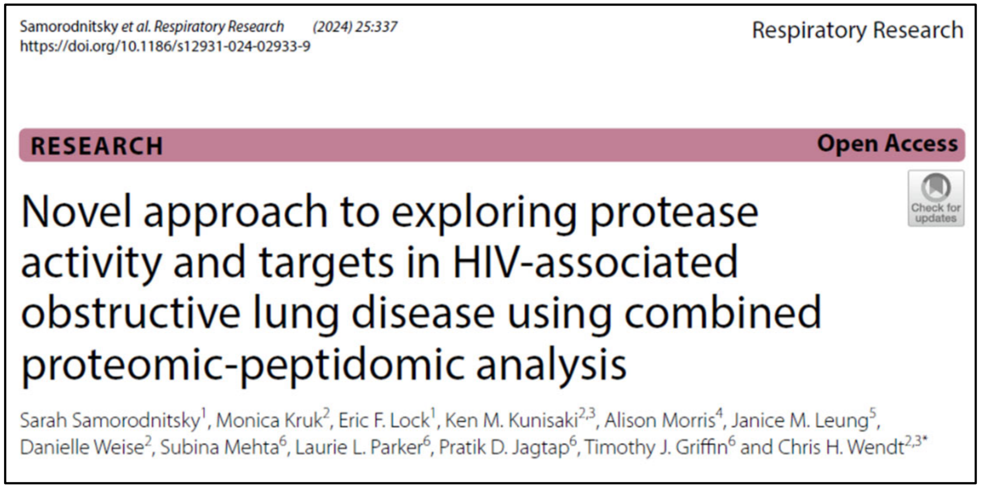
Role of Proteases in Inflammatory Lung Diseases

Chapter | First Online: 13 November 2013

pp 361–385 | Cite this chapter

Kaustav Chakraborty & Arindam Bhattacharyya

Part of the book series: Advances in Biochemistry in Health and Disease ((ABHD,volume 7))



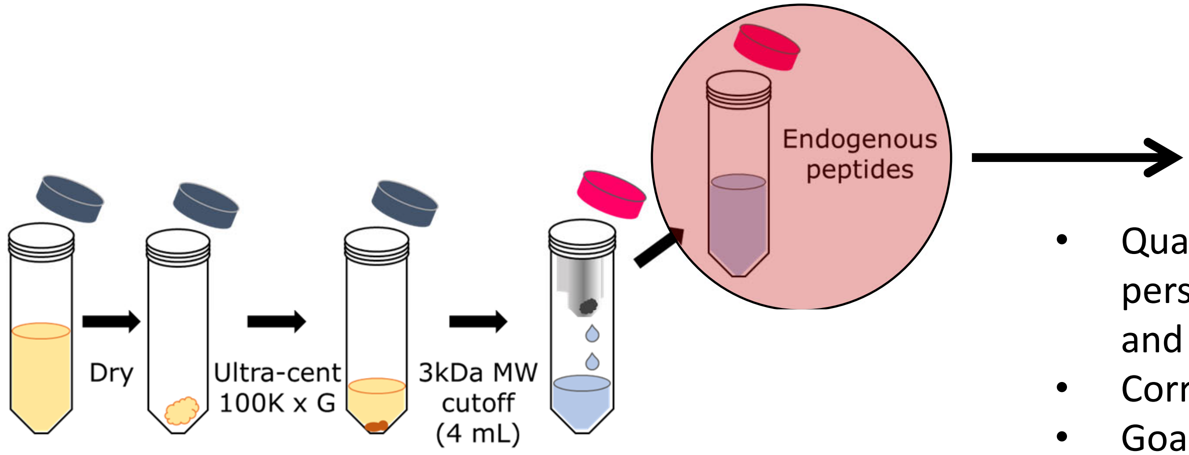
Samorodnitsky *et al. Respiratory Research* (2024) 25:337
https://doi.org/10.1186/s12931-024-02933-9

Respiratory Research

RESEARCH **Open Access**

Novel approach to exploring protease activity and targets in HIV-associated obstructive lung disease using combined proteomic-peptidomic analysis

Sarah Samorodnitsky[1], Monica Kruk[2], Eric F. Lock[1], Ken M. Kunisaki[2,3], Alison Morris[4], Janice M. Leung[5], Danielle Weise[2], Subina Mehta[6], Laurie L. Parker[6], Pratik D. Jagtap[6], Timothy J. Griffin[6] and Chris H. Wendt[2,3*]

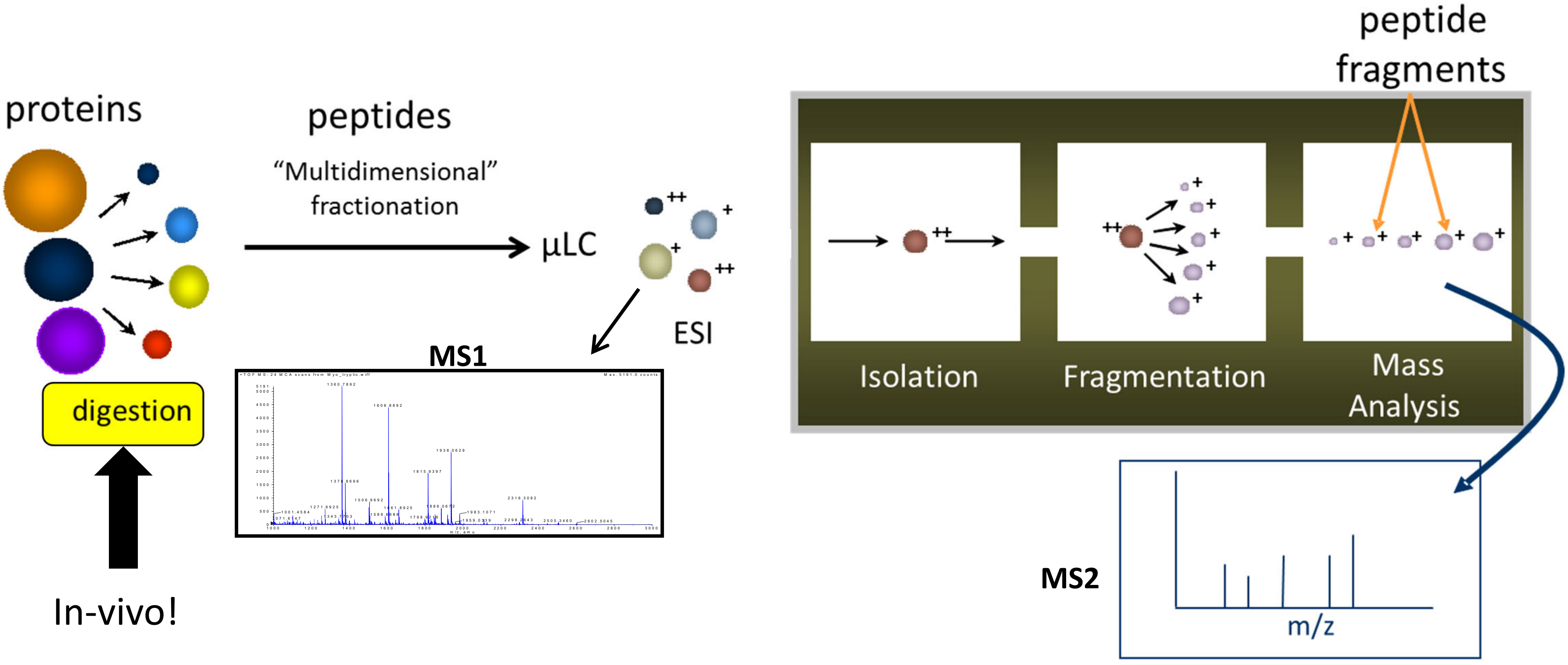


LC-MS/MS

- Quantitative peptidomics in BALF samples from persons living with HIV (PLWH) with OLD (n=25) and without OLD (n=26)
- Correlation to measured lung function (FEV1pp)
- Goal: Identify enriched protease activities associated with OLD in PLWH populations

Mass spectrometry-based proteomics: core technology platform

Peptide fractionation coupled to tandem mass spectrometry (MS/MS)



proteins

digestion

In-vivo!

peptides

“Multidimensional” fractionation

µLC

ESI

MS1

peptide fragments

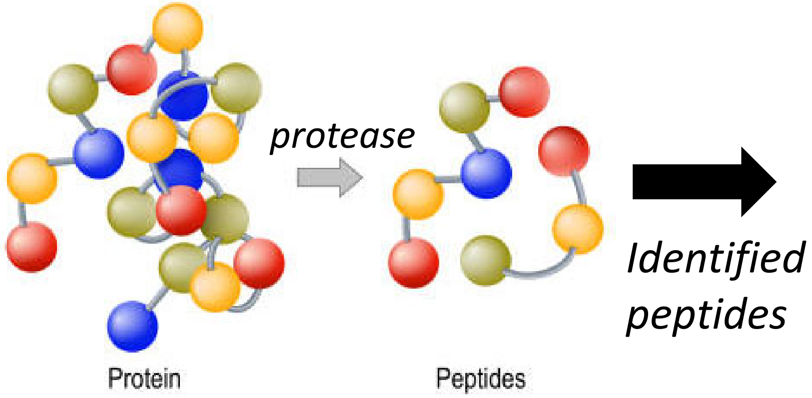
Isolation

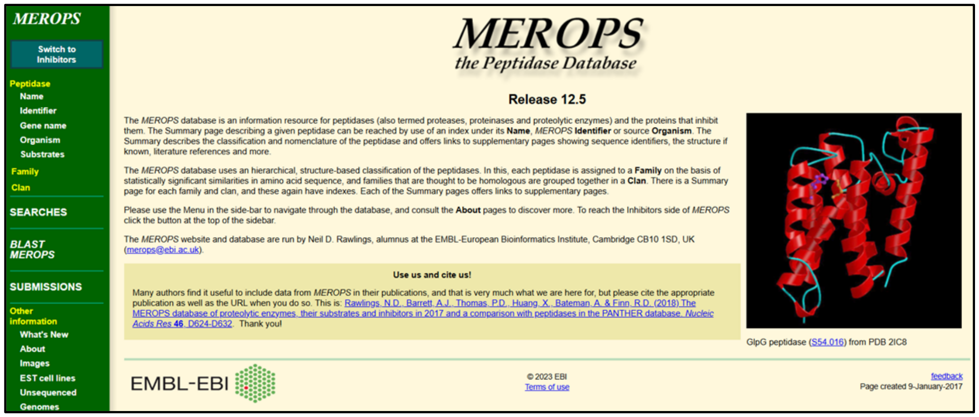
Fragmentation

Mass Analysis

MS2

Bioinformatics tools to identify enriched proteases





MEROPS
the Peptidase Database

Release 12.5

The *MEROPS* database is an information resource for peptidases (also termed proteases, proteinases and proteolytic enzymes) and the proteins that inhibit them. The Summary page describing a given peptidase can be reached by use of an index under its **Name**, *MEROPS* **Identifier** or source **Organism**. The Summary describes the classification and nomenclature of the peptidase and offers links to supplementary pages showing sequence identifiers, the structure if known, literature references and more.

The *MEROPS* database uses an hierarchical, structure-based classification of the peptidases. In this, each peptidase is assigned to a **Family** on the basis of statistically significant similarities in amino acid sequence, and families that are thought to be homologous are grouped together in a **Clan**. There is a Summary page for each family and clan, and these again have indexes. Each of the Summary pages offers links to supplementary pages.

Please use the Menu in the side-bar to navigate through the database, and consult the **About** pages to discover more. To reach the Inhibitors side of *MEROPS* click the button at the top of the sidebar.

The *MEROPS* website and database are run by Neil D. Rawlings, alumnus at the EMBL-European Bioinformatics Institute, Cambridge CB10 1SD, UK (merops@ebi.ac.uk).

Use us and cite us!

Many authors find it useful to include data from *MEROPS* in their publications, and that is very much what we are here for, but please cite the appropriate publication as well as the URL when you do so. This is: Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. & Finn, R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624-D632. Thank you!

GlpG peptidase (S54.016) from PDB 2IC8

EMBL-EBI

© 2023 EBI
Terms of use

feedback
Page created 9-January-2017

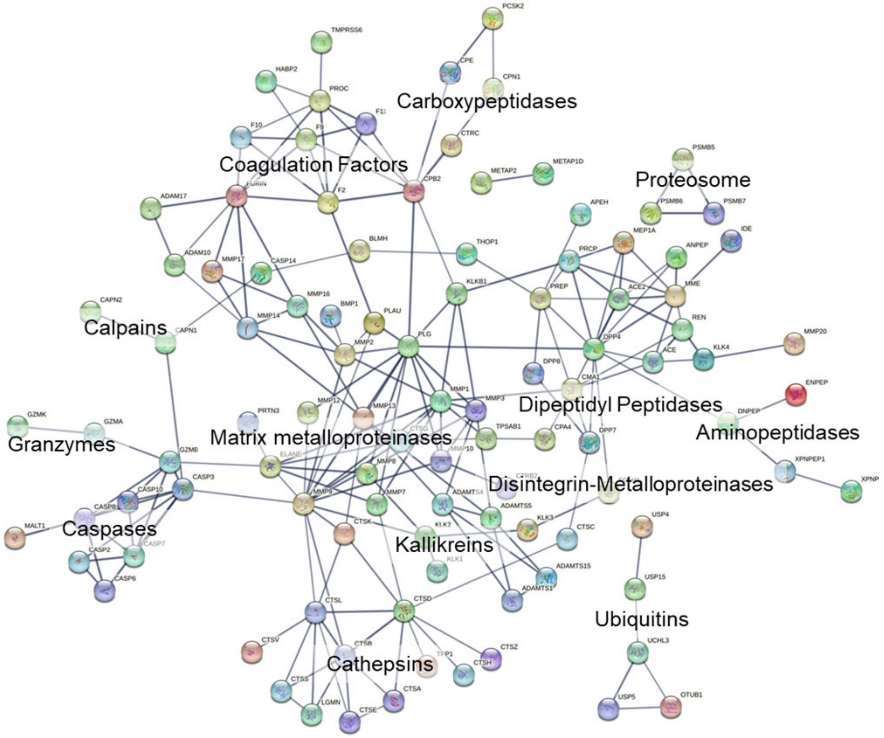
	Protein Description	Peptide Sequence	Cleave	Protease	z-score
Known Associations	Collagen alpha-2(I) chain	QSLQEETVRKGPAG	GPAG|DRGP	MMP13	2.77
	Tubulin beta chain	AILVDLEPGTMDSVR	YVPR|AILV	GZMK	2.67
	Histone H1.4	AAAGYDVEK	KKAL|AAAG	GZMM	2.20
	Histone H2A type 2-C	AGLQFPVGR	RSSR|AGLQ	GZMA	2.06
	Ezrin	LRAKEELERQAV	RMAA|LRAK	MMP3	1.95
	Actin, cytoplasmic 1	APEEHPVLL	PVLL|TEAP	CTSG	1.80
	Kininogen-1	SSRIGEIKEETT	SPFR|SSRI	KLK1	1.81
	Phosphoglycerate kinase 1	SLEPVAVELK	VELK|SLLG	ELANE	1.74
	Vimentin	NLRETNLD	TNLD|SLPL	CASP6	1.61
	Macrophage-capping protein	QVLGPKPALK	PALK|EGNP	CTSS	1.57
Unknown Associations	Ubiquitin-60S ribosomal protein L40	IIEPSLR	LRGG|IIEP	USP4	3.84
	Calreticulin	KSGTIFDNF	FDNF|LITN	CTSA	2.99
	Collagen alpha-2(I) chain	QSLQEETVRKGPAG	GPAG|DRGP	MMP13	2.77
	Prothrombin	ESYIDGR	IDGR|IVEG	F10	2.58
	Bactericidal permeability-increasing protein	NVVLQPHQNF	HQNF|LLFG	CTSA	2.58
	Alpha-1-antitrypsin	LSGVTEEAPLK	APLK|LSKA	CPA4	2.55
	Deleted in malignant brain tumors 1 protein (I2)	DVGSYQEKVDV	RSKR|DVGS	FURIN	2.51
	Apolipoprotein A-IV	EKESQDKTL	STFK|EKES	ATG4B	2.50
	Natural resistance-associated macrophage protein 1	MTGDKGPQR	GPQR|LSGS	MMP8	2.30
	Putative neutrophil cytosol factor 1B	GDTFIRH	- - - M|GDTF	METAP1D	2.25

GalaxyP
galaxyp.org

Monica Kruk

Results: enriched proteases based on analysis of peptidomics data

- Combined analysis of quantified proteases from BALF with analysis of endogenous peptide sequences resulting from protease action



Network view

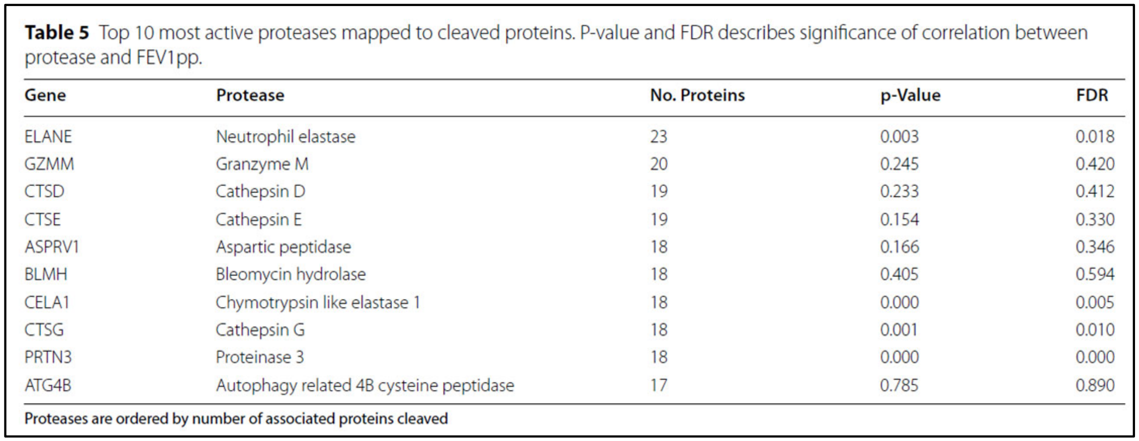


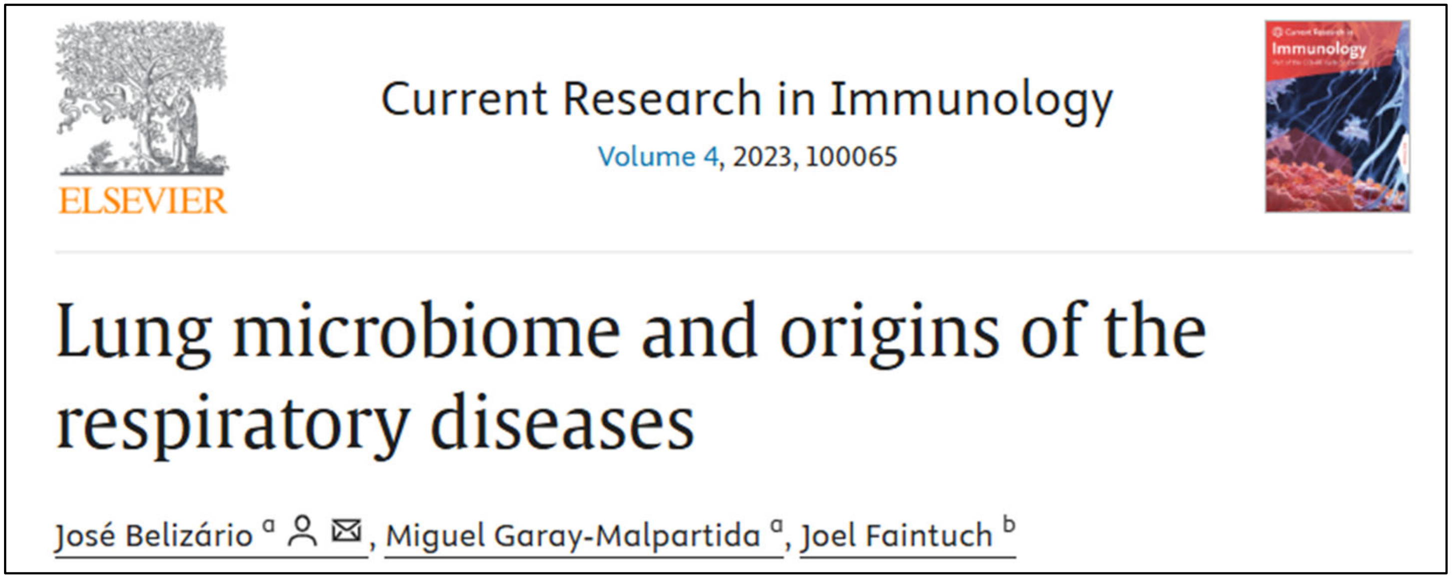
Table 5 Top 10 most active proteases mapped to cleaved proteins. P-value and FDR describes significance of correlation between protease and FEV1pp.

Gene	Protease	No. Proteins	p-Value	FDR
ELANE	Neutrophil elastase	23	0.003	0.018
GZMM	Granzyme M	20	0.245	0.420
CTSD	Cathepsin D	19	0.233	0.412
CTSE	Cathepsin E	19	0.154	0.330
ASPRV1	Aspartic peptidase	18	0.166	0.346
BLMH	Bleomycin hydrolase	18	0.405	0.594
CELA1	Chymotrypsin like elastase 1	18	0.000	0.005
CTSG	Cathepsin G	18	0.001	0.010
PRTN3	Proteinase 3	18	0.000	0.000
ATG4B	Autophagy related 4B cysteine peptidase	17	0.785	0.890

Proteases are ordered by number of associated proteins cleaved

- 22 proteases correlated to FEV1pp function

Using (meta)proteomics to characterize microbial contributors to lung disease and health



ELSEVIER

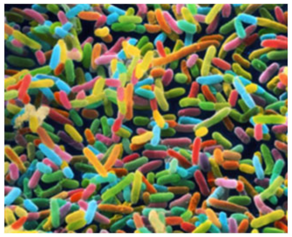
Current Research in Immunology

Volume 4, 2023, 100065

Lung microbiome and origins of the respiratory diseases

José Belizário [a], Miguel Garay-Malpartida [a], Joel Faintuch [b]

- ***Commensal and pathogenic microbes are key contributors to lung health and disease***



MICROBIOME

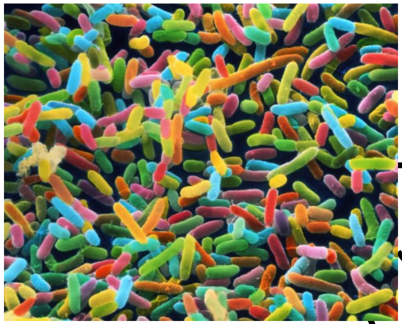
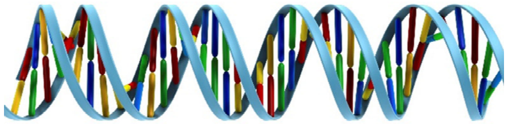


Image from https://thedoctorweighsin.com/what-everyone-should-know-about-the-infant-microbiome/

Potential to unravel the **mechanistic details of microbial interactions with host / environment** by analyzing the **functional dynamics of the microbiome.**



METAGENOMICS

TAXONOMY

function

METATRANSCRIPTOMICS

TAXONOMY

function

METAPROTEOMICS

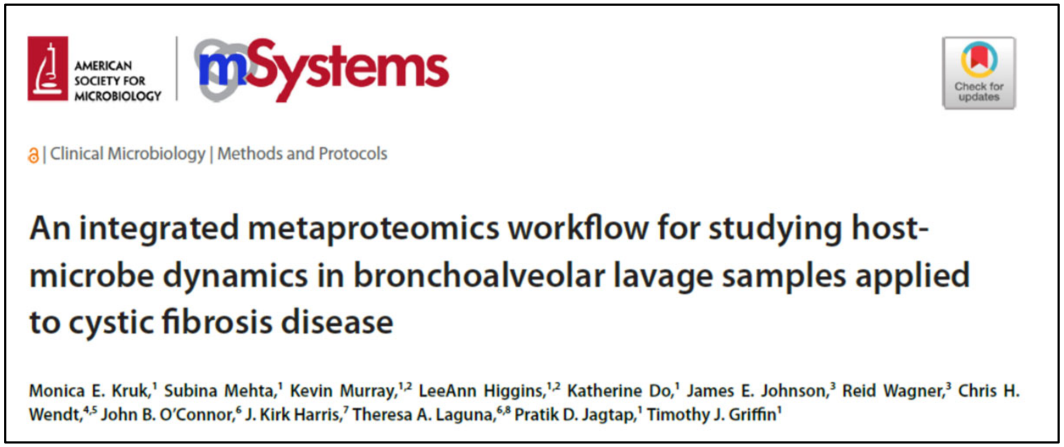
TAXONOMY

FUNCTION

(Slide courtesy of Pratik Jagtap)

Challenges of metaproteomics in clinical human samples (e.g. BALF)

- Detecting relatively low abundance microbial proteins from abundant host (human) proteins
- Knowledge of which microbe proteomes might be present (solved by multiomics – using metagenomic information as a guide)
- Matching MS/MS to very large metaproteomic protein sequence databases (10M+ sequences)
- Mapping identified microbial peptides to specific taxa and annotating protein function

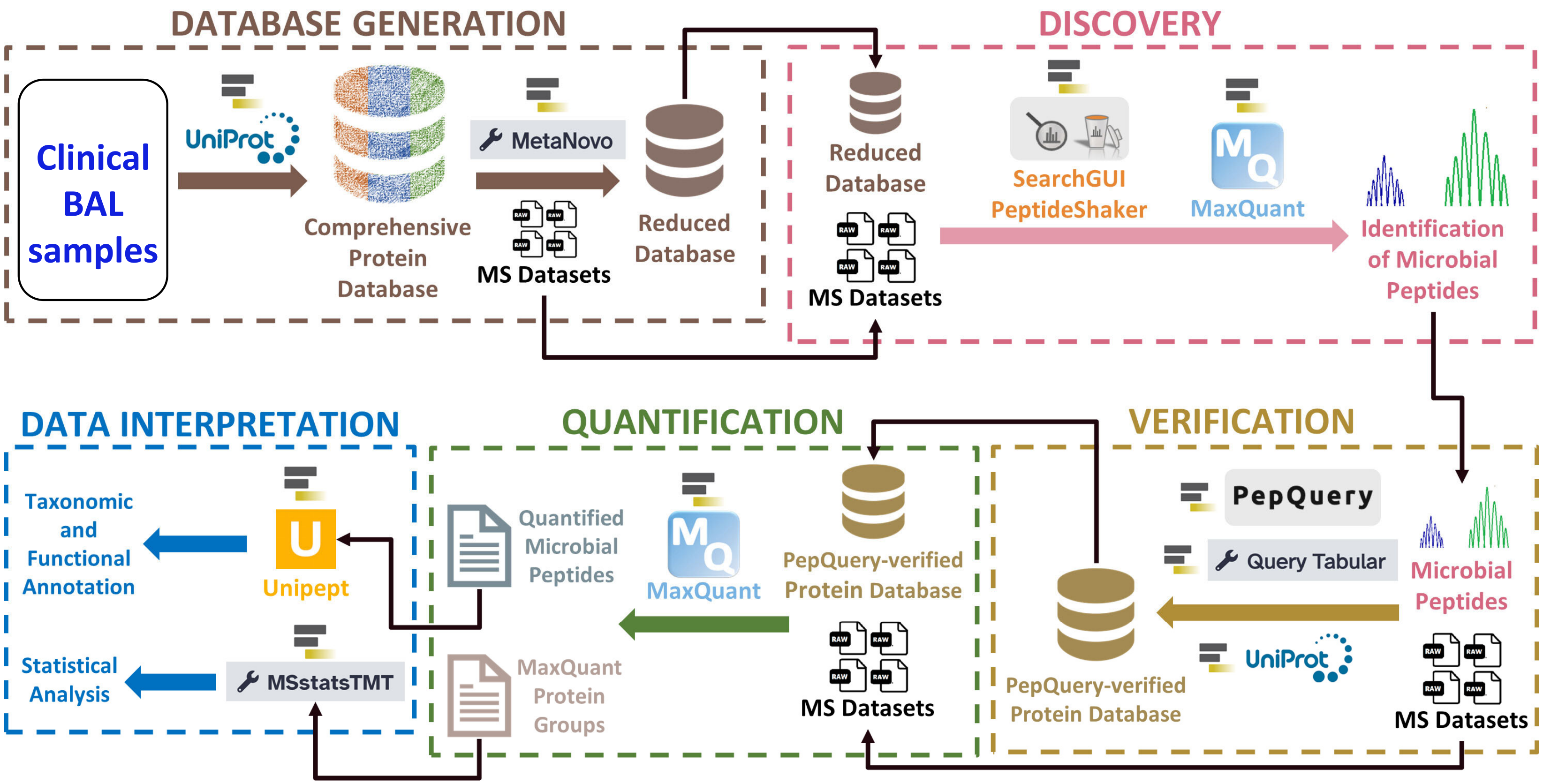


| Clinical Microbiology | Methods and Protocols

An integrated metaproteomics workflow for studying host-microbe dynamics in bronchoalveolar lavage samples applied to cystic fibrosis disease

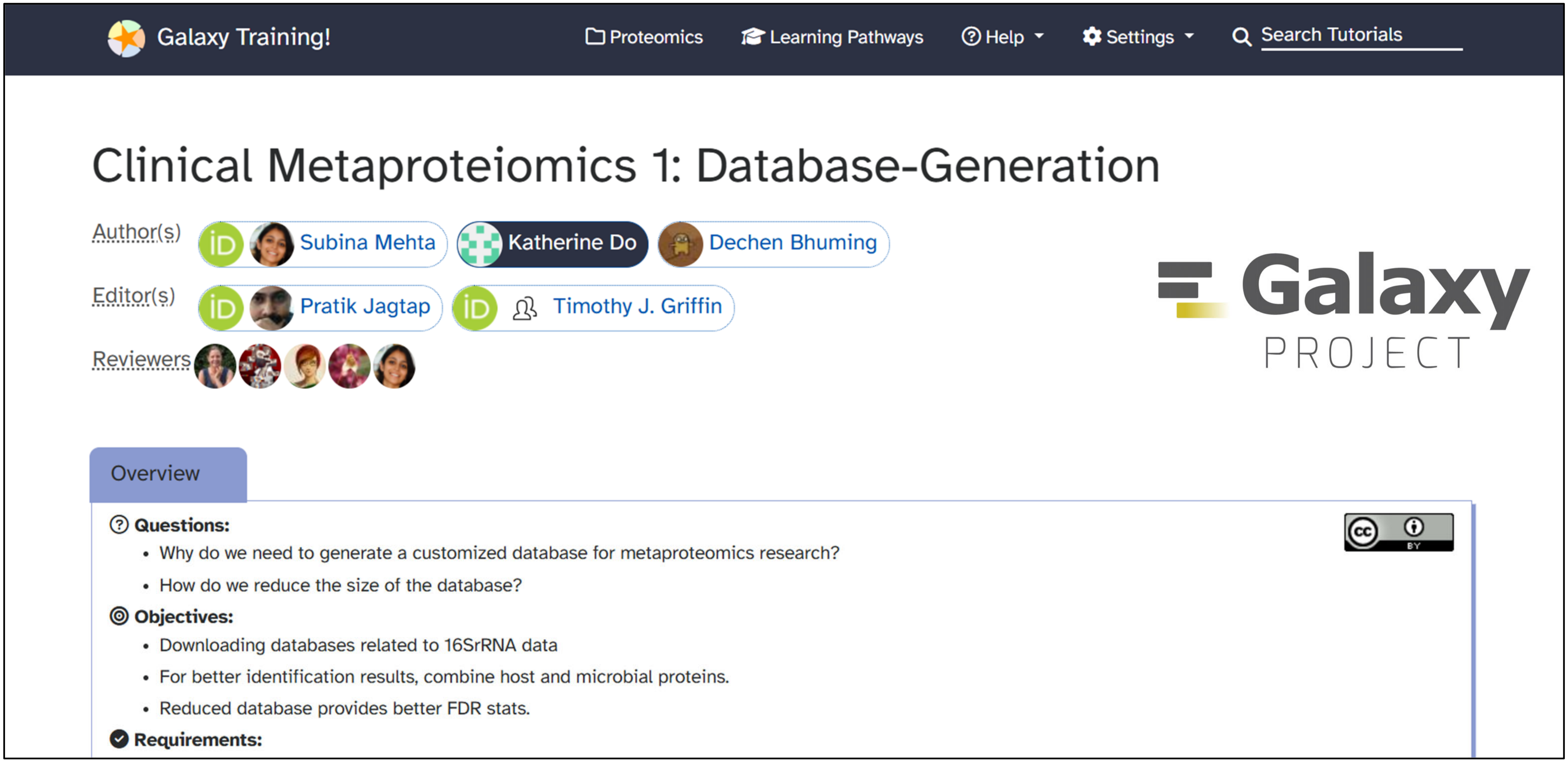
Monica E. Kruk,[1] Subina Mehta,[1] Kevin Murray,[1,2] LeeAnn Higgins,[1,2] Katherine Do,[1] James E. Johnson,[3] Reid Wagner,[3] Chris H. Wendt,[4,5] John B. O'Connor,[6] J. Kirk Harris,[7] Theresa A. Laguna,[6,8] Pratik D. Jagtap,[1] Timothy J. Griffin[1]

A clinical metaproteomics bionformatic workflow



mSphere (2024) 9:e0079323.
(Pratik Jagtap, Galaxy-P team)

Developing accessible software tools for multi-omics analysis



Clinical Metaproteiomics 1: Database-Generation

Author(s) Subina Mehta Katherine Do Dechen Bhuming

Editor(s) Pratik Jagtap Timothy J. Griffin

Reviewers

Questions:

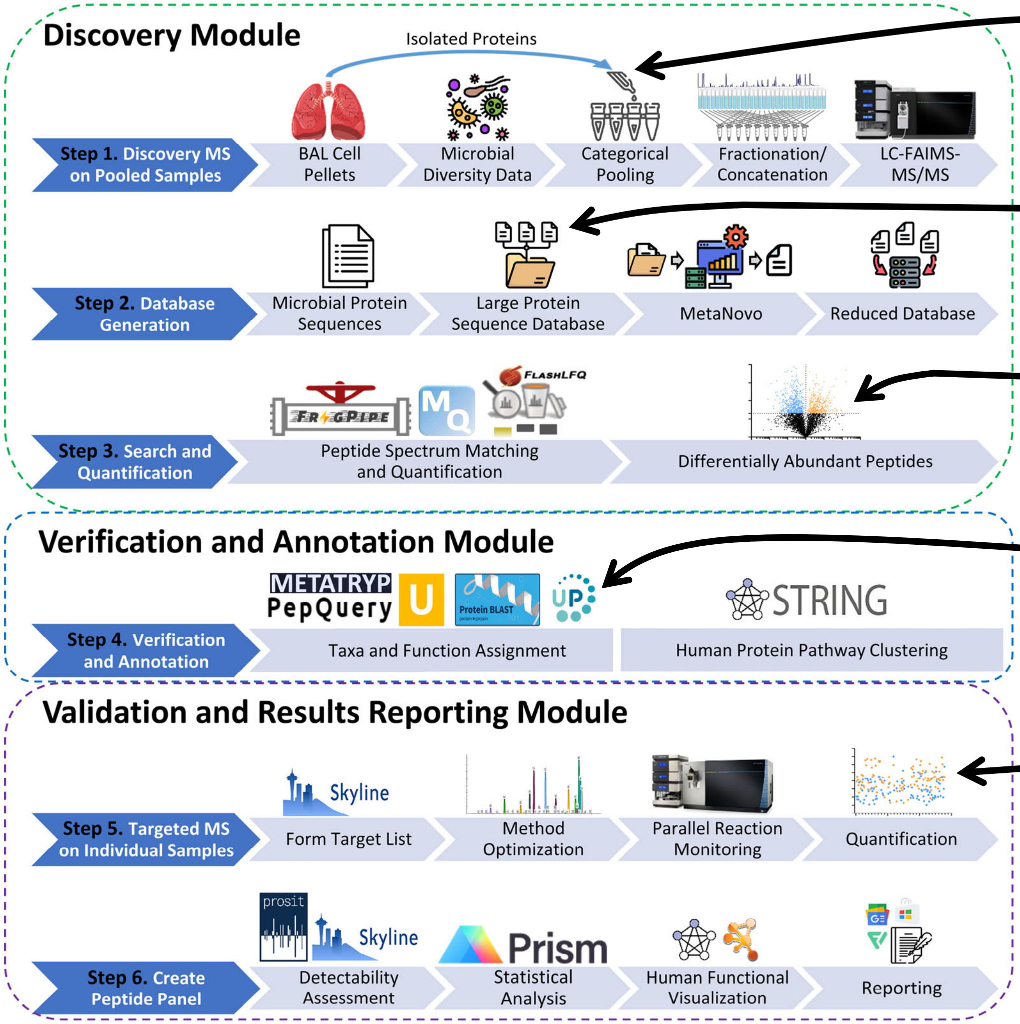
- Why do we need to generate a customized database for metaproteomics research?
- How do we reduce the size of the database?

Objectives:

- Downloading databases related to 16SrRNA data
- For better identification results, combine host and microbial proteins.
- Reduced database provides better FDR stats.

Requirements:

Application to microbial contributors in Cystic Fibrosis



Discovery Module

Isolated Proteins

Step 1. Discovery MS on Pooled Samples → BAL Cell Pellets → Microbial Diversity Data → Categorical Pooling → Fractionation/ Concatenation → LC-FAIMS-MS/MS

Step 2. Database Generation → Microbial Protein Sequences → Large Protein Sequence Database → MetaNovo → Reduced Database

Step 3. Search and Quantification → Peptide Spectrum Matching and Quantification → Differentially Abundant Peptides

Verification and Annotation Module

Step 4. Verification and Annotation → Taxa and Function Assignment → Human Protein Pathway Clustering

Validation and Results Reporting Module

Step 5. Targeted MS on Individual Samples → Form Target List → Method Optimization → Parallel Reaction Monitoring → Quantification

Step 6. Create Peptide Panel → Detectability Assessment → Statistical Analysis → Human Functional Visualization → Reporting

Pooled BAL cells from disease-control and CF patients + 16s rRNA data

Large, 16s rRNA-guided databases (16M+ sequences) reduced by customized software

Rigorous identification and quantification of both human and microbial peptides and proteins

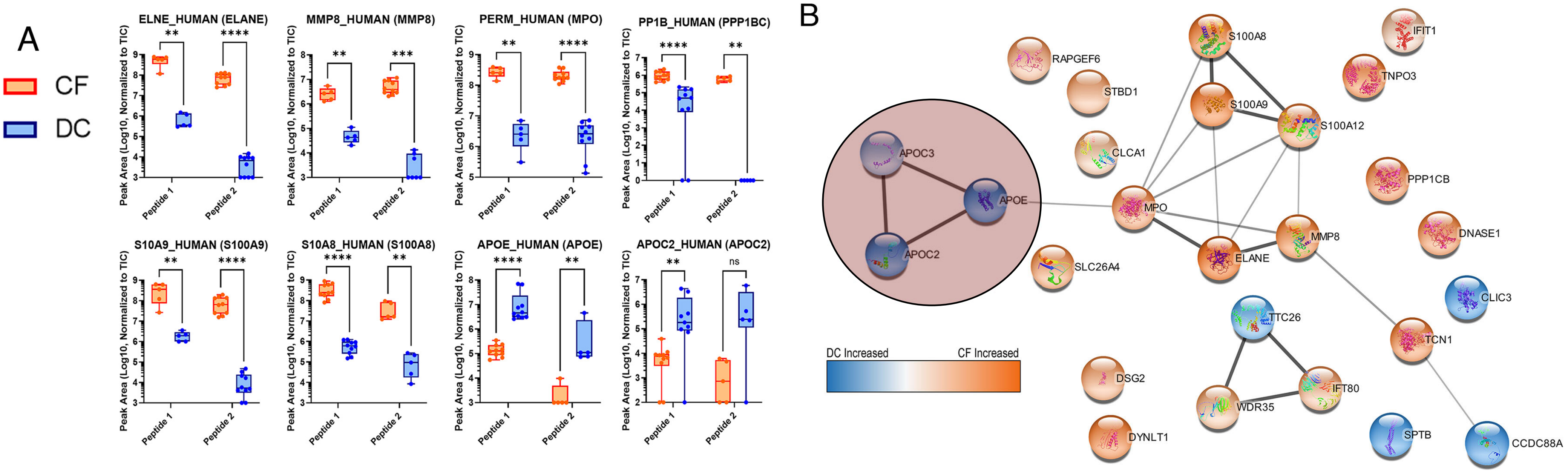
Careful verification and annotation of confident results

Targeted MS-based assays to confirm peptide quantification for human and microbial proteins of highest interest – in individual patient samples

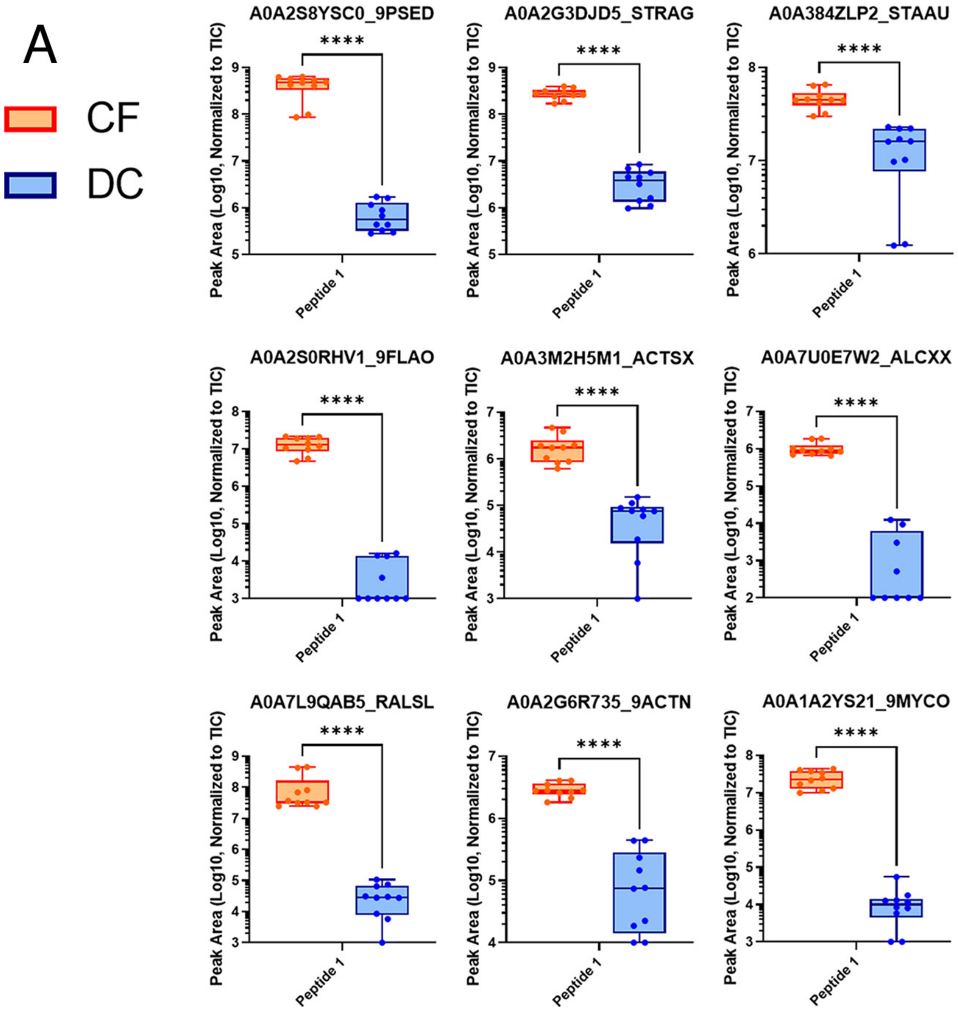
mSphere (2024) 9:e0079323.

Results: Human proteins

- *Discovered from quantitative proteomics data in pooled samples*
- *Validated and quantified in individual patient samples*



Results: Bacterial proteins



Peptides detected from:

Streptococcus

Pseudomonas

Mycobacterium

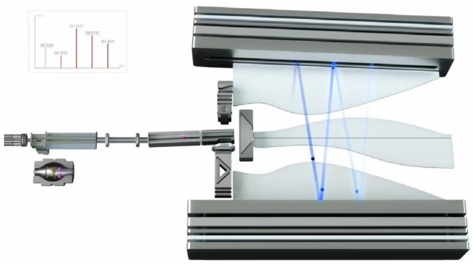
Ralstonia

Outcomes of CF metaproteomics in BAL samples

- Demonstration of combined protein-level analysis of human and microbial contributors to CF
- Demonstration of targeted MS-based assays to characterize individual CF patient samples
- Opens possibility of clinical assays for simultaneous characterization of host-microbial protein markers that may indicate disease state and/or monitor therapy outcomes

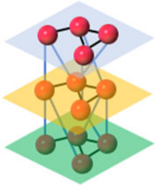
What’s next: improving the state of proteomics and multi-omics

- Improved instrumentation: 10X faster and 10X more sensitive?

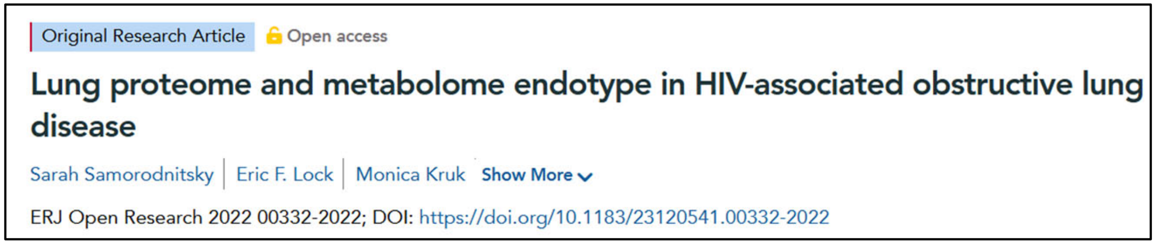


ThermoFisher

- Improved methods to integrated ‘omic levels of information for better molecular understanding of disease in clinical samples (more predictive, more personalized)



- Integration of other MS-based information (metabolomics)



Original Research Article | Open access

Lung proteome and metabolome endotype in HIV-associated obstructive lung disease

Sarah Samorodnitsky | Eric F. Lock | Monica Kruk Show More

ERJ Open Research 2022 00332-2022; DOI: https://doi.org/10.1183/23120541.00332-2022