AN INTEGRATED WORKFLOW FOR HOST-MICROBIOTA PROTEOMICS AND TARGETED ASSAY DEVELOPMENT TO STUDY HOST-MICROBE DYNAMICS IN CLINICAL CYSTIC FIBROSIS SAMPLES

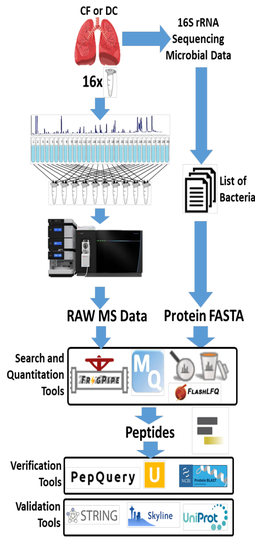
Monica E. Kruk[1], Subina Mehta[1], Katherine Do[1], James E. Johnson[2], Reid Wagner[2], Chris H. Wendt[4,5], John B. O'Connor[6], Theresa Laguna[6,7], Pratik D. Jagtap[1], and Timothy J. Griffin[1]

[1]Biochemistry, Mol. Biology and Biophysics, University of Minnesota; [2]Minnesota Supercomputing Institute, University of Minnesota; [3]Department of Infectious Disease, Imperial College London, London, UK; [4]Pulmonary, Allergy, Critical Care and Sleep Medicine Section, Minneapolis Veterans Administration Health Care System; [5]Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Minnesota; [6]Pulmonary and Sleep Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago; [7]Pediatrics, Northwestern University Feinberg School of Medicine

Introduction

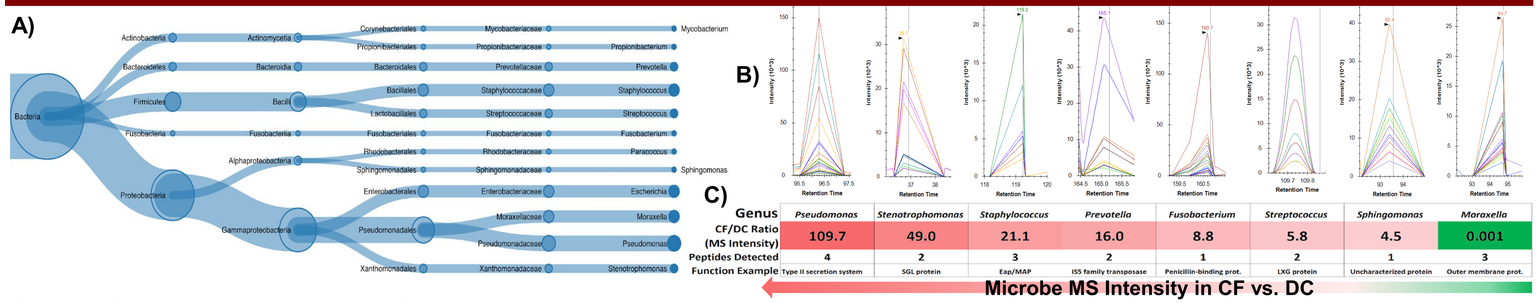
- Airway microbiota composition correlates with cystic fibrosis (CF) progression, but microbial drivers of disease remain unclear.
- Mass spectrometry (MS)-based metaproteomics of bronchoalveolar lavage fluid (BALF) offers insights into host-microbe dynamics & potential interactions.
- We have developed a MS-based BALF analysis/bioinformatics processing of both host and microbial proteins, generating verified host and microbe peptide candidates suitable for targeted analysis within individual patient samples.
- We have utilized this workflow in our ongoing work to identify a promising host and microbe peptide panel for application to CF disease progression studies by comparison to disease control (DC).

Methods



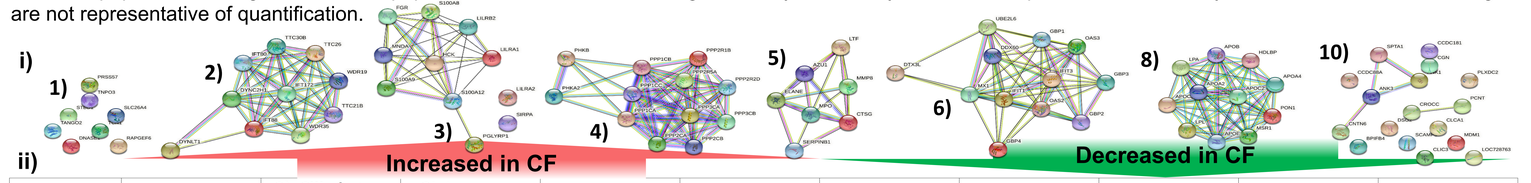
1. Pool and trypsinize pediatric BALF pellets based on disease status and 16S rRNA seq. data; 16 samples each category of CF or DC and high or low microbial diversities.
2. High pH RPLC fractionation and concatenation.
3. LC-FAIMS-MS/MS data acquisition on an Orbitrap Eclipse.
4. Created and search MS data with a customized ~17M microbial sequence database created with Metanovo.
5. Analyze and quantify mass spec data with Fragpipe, MaxQuant, and SearchGui/ PeptideShaker/FlashLFQ.
6. Generate human and microbial peptide targets for future PRM experiments and verify with PepQuery, UniPept, and NCBI Blast.
7. Validate and characterize with STRING-db.org, Skyline, and UniProt.

Results



Genus	*Pseudomonas*	*Stenotrophomonas*	*Staphylococcus*	*Prevotella*	*Fusobacterium*	*Streptococcus*	*Sphingomonas*	*Moraxella*
CF/DC Ratio (MS Intensity)	109.7	49.0	21.1	16.0	8.8	5.8	4.5	0.001
Peptides Detected	4	2	3	2	1	2	1	3
Function Example	Type II secretion system	SGL protein	Eap/MAP	IS5 family transposase	Penicillin-binding prot.	LXG protein	Uncharacterized protein	Outer membrane prot.

Figure 1. Examples of the microbial PRM peptide targets. A total of **87** non-human peptides with apparent differences in prevalence depending on lung disease status were detected within pooled clinical human BALF samples and will be used as microbial targets in PRM analysis of individual samples. **A)** Phylogenetic tree of confident microbial peptides. **B)** Representative chromatograms to show detectability of microbial peptides and **C)** CF/DC ratios and functional data of highly confident microbial peptide PRM targets. These 87 peptides will be used in PRM targeted analysis to study the microbial presence and host dynamics within CF. Chromatograms are not representative of quantification.



STRING-Db Pathway	1) Top CF Enriched	2) Cilia/Intra-Flagellar Regulation	3) Immuno-Regulatory Interactions	4) Phosphatase Complex	5) Neutrophil Cytotoxicity/Neg. Regulation of Cilium	6) Interferon Alpha/Beta	7) Chloride Channel	8) Lipoprotein	9) Ciliogenesis	10) Top DC Enriched
CF/DC Ratio (MS Intensity)	196.54	13.99	6.62	5.69	4.56	0.25	0.16	0.12	0.07	0.00
Protein Examples	Deoxyribonuclease-1	Dynein light chain Tctex-type 1	Protein S100-A8	Isoform 3 of Phosphorylase b kinase regulatory subunit beta	Neutrophil elastase	Guanylate-binding protein 2	Calcium-activated chloride channel regulator 1	Apolipoprotein E	Rootletin	Ankyrin-3
	Pendrin	Intraflagellar transport protein 56	Neutrophil collagenase	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	Azurocidin	Ubiquitin/ISG15-conjugating enzyme E2 L6	Chloride intracellular channel protein 3	Serum paraoxonase/arylesterase 1	Ciliary rootlet coiled-coil protein 2	Cingulin

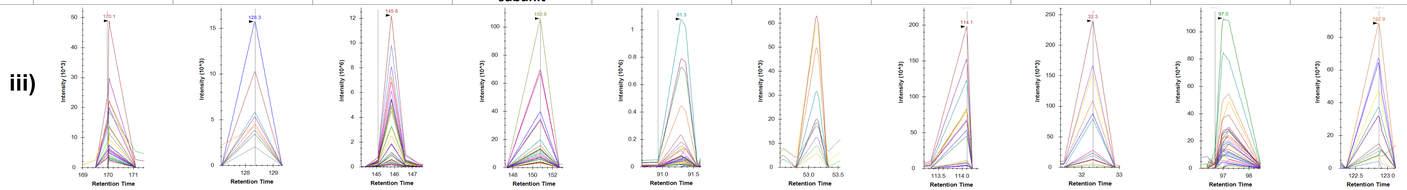


Figure 2. Examples of the human PRM peptide targets. Thousands of human proteins detected within pooled clinical human BALF samples with altered expression in the CF. Up to 106 of those will be used to study the human impact and microbe dynamics in CF as the host targets in PRM experiments of individual samples. **i)** Protein pathways identified with STRING-Db, **ii)** CF/DC ratios and functional data of the top 5 CF and DC pathways. **iii)** Representative chromatograms of peptides from the top protein pathways. Chromatograms are not representative of quantification.

Discussion & Conclusion

- PepQuery validated 680 microbial peptides out of 2292 total non-human peptides identified with Fragpipe, MaxQuant, and SearchGui.
- Bioinformatics analysis of these microbial peptides including detailed taxonomic, functional and quantitative analysis generated a peptide panel of 87 microbial peptides.
- The microbial peptides includes 24 taxonomy-specific peptides, 20 peptides from four protein clusters groups, and the rest of the peptides with ambiguous taxonomy but known functions.
- 106 human peptides spanning 8 pathways
- Proteins enriched in CF are involved in immune response and inflammation, as expected, as well as transport and cell motility proteins which have intriguing, but not fully understood potential roles in CF.
- Proteins enriched in DC include membrane proteins involved in cytoskeletal and lipoprotein regulation that may contribute to CF progression.

Future Directions

This peptide panel will be used for targeted quantification within the individual samples to further characterize the host-microbe dynamics in human pediatric CF.

References

Yang, F., et al. (2014) Exp. Review of Proteomics 9: 129
Potgieter, M. G., et al. (2019). bioRxiv (doi.org/10.1101/605550v6).
Kumar, P., et al. (2020). J Proteome Res. 19:2772
Mehta, S., et al. (2021). J Proteome Res. 20:2130.

Acknowledgements

This work was funded by NIH grant 1R01HL136499. We thank the Center for Mass Spectrometry and Proteomics (CMSP) at the University of Minnesota for assistance generating mass spectrometry data. We also thank the Minnesota Supercomputing Institute (MSI) for resources supporting the bioinformatic tools used in this study.