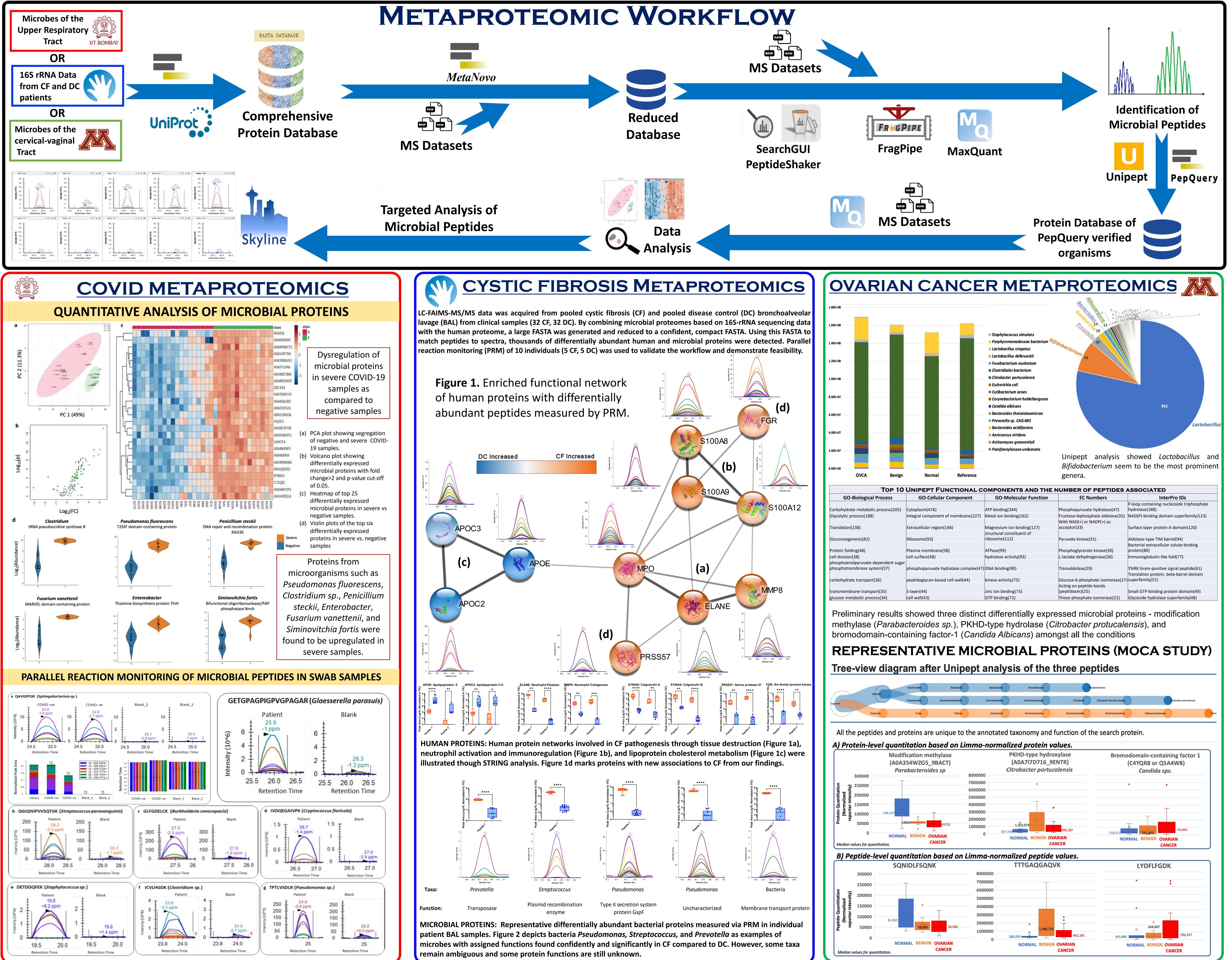# BIOINFORMATIC WORKFLOW FOR METAPROTEOMIC ANALYSIS OF HOST-MICROBE DYNAMICS IN CLINICAL SAMPLES

Pratik Jagtap[1], Surbhi Bihani[2], Monica Kruk[1], Subina Mehta[1], Aryan Gupta[2], Kevin Murray[1], Andrew Rajczewski[1], James Johnson[3], Reid Wagner[3], Kristin Boylan[1], Amy Skubitz[1], Theresa Laguna[4], Sanjeeva Srivastava[2], Timothy Griffin[1]

[1]Biochemistry, Mol. Biology and Biophysics, University of Minnesota, Minneapolis, USA; [2]Indian Institute of Technology Bombay, Mumbai, India ; [3]Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, USA
[4]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA ; [5]Pulmonary and Sleep Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, USA

Clinical metaproteomics has the potential to offer insights into host-microbiome interactions. However, the field faces challenges in the detection of low-abundance microbial proteins. As a solution, we have developed an integrated workflow coupling mass spectrometry-based analysis with customized bioinformatic processing of microbial proteins. We have utilized this workflow in ongoing projects to identify microbial peptide panels for: A) co-infection status during COVID-19 pandemic waves, B) cystic fibrosis (CF) disease progression studies, and C) ovarian cancer biomarker discovery.

METAPROTEOMIC WORKFLOW

Microbes of the Upper Respiratory Tract (IIT BOMBAY) OR 16S rRNA Data from CF and DC patients OR Microbes of the cervical-vaginal Tract → UniProt → Comprehensive Protein Database → FASTA DATABASE → MetaNovo → MS Datasets → Reduced Database → SearchGUI PeptideShaker → FragPipe → MaxQuant → MS Datasets → Identification of Microbial Peptides → Unipept / PepQuery → Protein Database of PepQuery verified organisms → MS Datasets (MaxQuant) → Data Analysis → Targeted Analysis of Microbial Peptides (Skyline)

COVID METAPROTEOMICS

QUANTITATIVE ANALYSIS OF MICROBIAL PROTEINS

Dysregulation of microbial proteins in severe COVID-19 samples as compared to negative samples

(a) PCA plot showing segregation of negative and severe COVID-19 samples.
(b) Volcano plot showing differentially expressed microbial proteins with fold change>2 and p-value cut-off of 0.05.
(c) Heatmap of top 25 differentially expressed microbial proteins in severe vs negative samples.
(d) Violin plots of the top six differentially expressed proteins in severe vs. negative samples

Clostridium - tRNA pseudouridine synthase B
Pseudomonas fluorescens - T2SSF domain-containing protein
Penicillium steckii - DNA repair and recombination protein RAD26
Fusarium vanettenii - MARVEL domain-containing protein
Enterobacter - Thiamine biosynthesis protein ThiH
Siminovitchia fortis - Bifunctional oligoribonuclease/PAP phosphatase NrnA

Proteins from microorganisms such as Pseudomonas fluorescens, Clostridium sp., Penicillium steckii, Enterobacter, Fusarium vanettenii, and Siminovitchia fortis were found to be upregulated in severe samples.

PARALLEL REACTION MONITORING OF MICROBIAL PEPTIDES IN SWAB SAMPLES

(a) QAVGEVGPR (Sphingobacterium sp.)
GETGPAGPIGPVGPAGAR (Glaesserella parasuis)
(b) QGIQNIPVVSQTGK (Streptococcus parasanguinis)
(c) GLFGDELCK (Burkholderia cenocepacia)
(d) IVDQEGAIVPK (Cryptococcus floricola)
(e) DETDDQFEK (Staphylococcus sp.)
(f) ICVLHGDK (Clostridium sp.)
(g) TPTLVSDLR (Pseudomonas sp.)

CYSTIC FIBROSIS METAPROTEOMICS

LC-FAIMS-MS/MS data was acquired from pooled cystic fibrosis (CF) and pooled disease control (DC) bronchoalveolar lavage (BAL) from clinical samples (32 CF, 32 DC). By combining microbial proteomes based on 16S-rRNA sequencing data with the human proteome, a large FASTA was generated and reduced to a confident, compact FASTA. Using this FASTA to match peptides to spectra, thousands of differentially abundant human and microbial proteins were detected. Parallel reaction monitoring (PRM) of 10 individuals (5 CF, 5 DC) was used to validate the workflow and demonstrate feasibility.

Figure 1. Enriched functional network of human proteins with differentially abundant peptides measured by PRM.

DC Increased — CF Increased

APOC3, APOE, APOC2, MPO, MMP8, ELANE, PRSS57, FGR, S100A8, S100A9, S100A12

(a), (b), (c), (d)

HUMAN PROTEINS: Human protein networks involved in CF pathogenesis through tissue destruction (Figure 1a), neutrophil activation and immunoregulation (Figure 1b), and lipoprotein cholesterol metabolism (Figure 1c) were illustrated though STRING analysis. Figure 1d marks proteins with new associations to CF from our findings.

APOE: Apolipoprotein E ; APOC3: Apolipoprotein C-II ; ELANE: Neutrophil Elastase ; MMP8: Neutrophil Collagenase ; S100A8: Calgranulin A ; S100A9: Calgranulin B ; PRSS57: Serine protease 57 ; FGR: Src-family tyrosine kinase

Taxa: Prevotella ; Streptococcus ; Pseudomonas ; Pseudomonas ; Bacteria
Function: Transposase ; Plasmid recombination enzyme ; Type II secretion system protein GspF ; Uncharacterized ; Membrane transport protein

MICROBIAL PROTEINS: Representative differentially abundant bacterial proteins measured via PRM in individual patient BAL samples. Figure 2 depicts bacteria Pseudomonas, Streptococcus, and Prevotella as examples of microbes with assigned functions found confidently and significantly in CF compared to DC. However, some taxa remain ambiguous and some protein functions are still unknown.

OVARIAN CANCER METAPROTEOMICS

Staphylococcus simulans, Porphyromonadaceae bacterium, Lactobacillus crispatus, Lactobacillus delbrueckii, Fusobacterium nucleatum, Clostridioides bacterium, Citrobacter portucalensis, Escherichia coli, Cutibacterium acnes, Corynebacterium heidelbergense, Candida albicans, Bacteroides thetaiotaomicron, Prevotella sp. CAG:485, Bacteroides acidifaciens, Aerococcus viridans, Actinomyces graevenitzii, Parafannyhessea umbonata

Unipept analysis showed Lactobacillus and Bifidobacterium seem to be the most prominent genera.

Top 10 Unipept Functional Components and the Number of Peptides Associated

GO-Biological Process	GO-Cellular Component	GO-Molecular Function	EC Numbers	InterPro IDs
Carbohydrate metabolic process(205)	Cytoplasm(476)	ATP binding(344)	Phosphopyruvate hydratase(47)	P-loop containing nucleoside triphosphate hydrolase(188)
Glycolytic process(188)	Integral component of membrane(227)	Metal ion binding(162)	Fructose-bisphosphate aldolase(35)	NAD(P)-binding domain superfamily(123)
Translation(138)	Extracellular region(146)	Magnesium ion binding(127)	With NAD(+) or NADP(+) as acceptor(33)	Aldolase-type TIM barrel(94)
Gluconeogenesis(82)	Ribosome(93)	structural constituent of ribosome(112)	Pyruvate kinase(31)	Bacterial extracellular solute-binding protein(89)
Protein folding(48)	Plasma membrane(58)	ATPase(99)	Phosphoglycerate kinase(30)	Immunoglobulin-like fold(77)
cell division(38)	cell surface(48)	hydrolase activity(93)	L-lactate dehydrogenase(30)	
phosphoenolpyruvate-dependent sugar phosphotransferase system(37)	phosphopyruvate hydratase complex(47)	DNA binding(90)	Transaldolase(29)	YSIRK Gram-positive signal peptide(61) Translation, beta-barrel domain superfamily(51)
carbohydrate transport(37)	peptidoglycan-based cell wall(44)	kinase activity(75)	Glucose-6-phosphate isomerase(27)	
transmembrane transport(35)	S-layer(44)	zinc ion binding(73)	Acting on peptide bonds (peptidases)(25)	Small GTP-binding protein domain(49)
glucose metabolic process(34)	cell wall(43)	GTP binding(72)	Triose-phosphate isomerase(21)	Glycoside hydrolase superfamily(48)

Preliminary results showed three distinct differentially expressed microbial proteins - modification methylase (Parabacteroides sp.), PKHD-type hydrolase (Citrobacter protucalensis), and bromodomain-containing factor-1 (Candida Albicans) amongst all the conditions.

REPRESENTATIVE MICROBIAL PROTEINS (MOCA STUDY)

Tree-view diagram after Unipept analysis of the three peptides

All the peptides and proteins are unique to the annotated taxonomy and function of the search protein.

A) Protein-level quantitation based on Limma-normalized protein values.

Modification methylase (A0A354WZG5_9BACT) Parabacteroides sp ; PKHD-type hydroxylase (A0A717D716_9ENTR) Citrobacter portucalensis ; Bromodomain-containing factor 1 (C4YQR8 or Q5A4W8) Candida sps.

NORMAL, BENIGN, OVARIAN CANCER

B) Peptide-level quantitation based on Limma-normalized peptide values.

SQNIDLFSQNK ; TTTGAQGAQVK ; LYDFLFGDK