A METAPROTEOMICS BIOINFORMATICS WORKFLOW TO STUDY HOST-MICROBE DYNAMICS IN CLINICAL SAMPLES

Pratik Jagtap

Galaxy for Proteomics (Galaxy-P) team

University of Minnesota

Co-authors:
Monica E. Kruk, Subina Mehta, Katherine Do, James E. Johnson, Reid Wagner, Chris H. Wendt, John B. O’Connor, Theresa Laguna and Timothy J. Griffin

IMSC 2022
MAASTRICHT, THE NETHERLANDS

27th August - 2nd September
Cystic Fibrosis and the Microbiome

- Life-shortening Mendelian disease.
- Mutations in the CF transmembrane conductance regulator (CFTR) gene, which encodes an epithelial anion channel.
- Abnormal anion transport across epithelia of secretory glands including lung.
- Thick mucus which predisposes patients to chronic bacterial infections and airway inflammation.
- Progressive and irreversible airway damage.
- Reduced quality and length of life in CF.

- Understanding CF respiratory microbiome (‘song along with the singers’) may help in improved therapeutics.
- Understanding of host–microbiome interactions may help in defining relationships between microbiome, disease status, and treatment response.

CLINICAL METAPROTEOMIC STUDIES DETECT LIMITED NUMBER OF MICROBIAL PEPTIDES

- **Cystic Fibrosis dataset**
  - Microbial: 2292
  - Human: 85713

- **Salivary dataset**
  - Microbial
  - Human

- **Cervical vaginal dataset**
  - Microbial
  - Human


BALF from Cystic Fibrosis or Disease Control

Sample Optimization, LC-FAIMS-MS/MS

RAW MS Data

16S rRNA Sequencing Microbial Data

List of Bacteria

Large Protein FASTA (17 M)

MetaNovo

Compact Protein FASTA (232 K)

Search and Quantitation Tools

PEPTIDES
Overview of Microbial Peptide Verification

Peptides and Proteins Detection & Quantitation

Microbial Peptides
2292 microbial peptides

Microbial Peptide Verification

PepQuery

PepQuery Verified Peptides

VERIFIED: 680
NON-VERIFIED: 1612
Ensuring confidence in MS/MS matches to novel peptide sequences: important!

Verified Microbial Peptides

Peptides and Proteins Detection & Quantitation

Microbial Peptide Verification

Taxonomy & Functional Analysis

PepQuery
Quantitation of Microbial Peptides

**Pseudomonas**
GO:0016021 integral component of membrane
IPR:IPR042094 Type II secretion system GspF domain superfamily

**Moraxella catarrhalis**
GO:009279 cell outer membrane
GO:0005509 calcium ion binding
GO:0006811 ion transport
IPR006664 Outer membrane protein

**Moraxella**
GO:009279 cell outer membrane
GO:0005509 calcium ion binding
GO:0006811 ion transport
IPR011250 Outer membrane protein / outer membrane enzyme PagP beta-barrel
**Verified Microbial Peptides**

![Diagram showing 87 verified microbial peptides, with 24 taxonomy-specific candidate peptides (known taxonomy), 20 peptides of ambiguous taxonomy (3 proteins), and additional peptides for targeted analysis.]

- **Pseudomonas**
  - DWLDSSLQR
  - NALQQALAAVPHP
  - MSVVIAGER
  - WMYSADDTOTPLAGER
- **Moraxella**
  - FKPVVLVQAGQSK
  - QSQDQVQILTQK
  - IKVNAIDYTAEVANGQIAEY
- **Escherichia**
  - TAATLTDKGEIR
  - NEKMQRSVER
  - IMVEALKSGDAI
- **Streptococcus**
  - MKIGNLGAVYR
  - QDPPPSSSYISSLR
- **Prevotella**
  - VDEGDFCASWGGGIGR
  - KTVYDINGM
- **Stenotrophomonas**
  - TARFQHYSYLR
  - AAMGSGDFSIAK
- **Mycobacterium**
  - ERWADTPPDINNSPLAWLR
- **Fusobacterium**
  - TSLNRRXVMTINDQLQFIL
- **Sphingomonas**
  - NANIIVEGR
- **Propionibacterium**
  - TGKDFIDIVADR
- **20 Peptides Of Ambiguous Taxonomy (3 Proteins)**
  - GTLTVSSASTK
  - SSAKTPSVPFLAPSKS
  - TVSSASTKPSVPFLAPSKS
  - LESHGGLVQPGSGSLR
  - LLESHGGLVQPGSGSLR
  - EVQLLESHGGLVQPGSGSLR
  - FGGMGTKLTVLQGPQK
  - LTVLGPGQK
  - VGFGGKTVLQGPQK
  - STFQMQWISKEEYDESGPSVHR
  - ASLTFQMQWISKEEYDESGPSVHR
  - SLSTFQMQWISKEEYDESGPSVHR
  - STFQMQWISKEEYDESGPSVHR
  - STFQMQWISKEEYDESGPSVHR
  - TFQMQWISKEEYDESGPSVHR
  - SLSTFQMQWISKEEYDESGPSVHR
  - QMWMQWISKEEYDESGPSVHR
  - FQMQWISKEEYDESGPSVHR
- **Additional Peptides for Targeted Analysis**
  - AKADSYPYTI
  - ILDPGYWNSTLR
  - TGGDFSINR
  - VMDSDGVHTVPIYEQYALPHAIR
  - WGSFDDYKGDLSR
  - ATVEELHLEG
  - TTGVMDSDGVHTVPIYEQYALPHAIR
  - LGQDVFPGVTR
  - LVADSITSQLER
  - GITINTSHYEDTPTR
  - QAQAVGWWGADK
  - LADEEIEIR
  - AYLTIAK
  - LIPNNOQK
  - IPAPSGHEEG
  - RAETELOFCQR
  - ALGMQSGAEIH
  - YLVSMVNGOIQK
  - FIVPTDPK
  - MAGDGQFALFQPSTQQQGQK
  - VIPELDIKL
  - AALGAYDLR

Katherine Do
Microbial candidates (known and novel) were detected in this analysis.
Peptides with ambiguous taxonomic assignment

**Functional Domains (BLAST-P)**

  - LTVLDTKLAATMGMWSCIIFLVATATGHSVLEQLESGGGLVQPQGGLSLRLSCEASGLF
    - FTFSYTMWMVVQAPGKGLWLSLYGSSGSYFAYSDVGRFTISREDNSANAYLY
    - QMNSLRAEDTAVYVACILRGRGSLMSGDLWQGGLTIVVSASTKGPSSPVFPLAPSSK
    - STSGTAAALCGCLVDKYPEPFTFVSTWNSGALTSGVHPFVAPVSLGGVLSLSYTVVP
    - SSSLGTQTVICNHYKPSNVTKVKEPKCSDKTHTPCPAAPPELLGGPVPFLLPPFK
    - PKPDMLISRTPEVTCVVVSHEDPEVKVFVNYGDVEVHINAKTPREEQYN
    - TYRVSGLTLHQLWGNYKCLKVSNKLAPPIEKTIKACGQPREPCVFYTLPPS
    - RDELTKNQVSLTCVKGFPYPSIAWESNGQPENNYYKTPPVLDOSGSSFYKSL
    - TVDKSR

- **Sequences 20 – 111, 137 – 231: Immunoglobulin-like domains**
  - MGWSCIIFLVATATGHSVQSVLTQPSASGTPQQRVTISCTGSSSNIAGNYVNY
    - WYQQGPLGTPAKLLYGINRNPSSGVDRSFSGKSQTSALSAGLRSEDEADYYCAG
    - WDKSISGIKAAAPSVTLFPPSESSELQANKTIVLCLISDFYPGLA
    - VTVAWKADSSSPKAVGIVETTPKQSNKYYYYSLSLTEQWKSHRSYCQVTH
    - EGSTVEKTVAPTECS

- **Actin-like activity**
  - MEGEDVQALVIDNDSGMCKAGFAFDADAPRAVPSISVGRPHTGVMVGVGMGQK
    - DSYVDVGAQQSKRGITLKYEHGIVNTWVDMEKIIWHFTYFENLWAEFEEHPVLL
    - TEAPLNPANKRENKMTIQMVETFTNPMAYIAQVLSYASRTRTTGVMDSGDGV
    - SHTVPYEYCALHARLDLADRLGARLDTLICFTGYSSTFTAEIREVIREDIEKLAY
    - VALDFEQEMQTAASSSALEKSYLPDPDIQTVIGNERFRCPEALSFQPSFLGQEMSGAGI
    - HETTNYSIMKCDVIRDKLIVNVLSGTTMTMFPIADMROMQKELTALAPSTMKII
    - APPERKYSVWIGGSIL

**Protein Sequence Database (FASTA db)**

- LTVLDTKLAATMGMWSCIIFLVATATGHSVLEQLESGGGLVQPQGGLSLRLSCEASGLF
- FTFSYTMWMVVQAPGKGLWLSLYGSSGSYFAYSDVGRFTISREDNSANAYLY
- QMNSLRAEDTAVYVACILRGRGSLMSGDLWQGGLTIVVSASTKGPSSPVFPLAPSSK
- STSGTAAALCGCLVDKYPEPFTFVSTWNSGALTSGVHPFVAPVSLGGVLSLSYTVVP
- SSSLGTQTVICNHYKPSNVTKVKEPKCSDKTHTPCPAAPPELLGGPVPFLLPPFK
- PKPDMLISRTPEVTCVVVSHEDPEVKVFVNYGDVEVHINAKTPREEQYN
- TYRVSGLTLHQLWGNYKCLKVSNKLAPPIEKTIKACGQPREPCVFYTLPPS
- RDELTKNQVSLTCVKGFPYPSIAWESNGQPENNYYKTPPVLDOSGSSFYKSL
- TVDKSR

**20 Peptides of Ambiguous Taxonomy**

- LESGGGLVQPQGGLSLRL
- EVCLEQLESGGGLVQPQGGLSLRL

**3 Protein Matches (FASTA db)**

- **A0A246E3YS_9MICO**
  - Uncharacterized protein (Fragment)
  - **Organism:** Microbacterium sp. AIS03

- **A0A2D0ABQ9_9MICO**
  - Uncharacterized protein (Fragment)
  - **Organism:** Microbacterium sp. AIS03

- **A0A228ZS16_ECOLX**
  - Actin, cytoplasmic 2
  - **Organism:** Escherichia coli
Validation of microbial peptide targets detected using PRM

- Pseudomonas
- Stenotrophomonas
- Staphylococcus
- Prevotella
- Fusobacterium
- Streptococcus
- Sphingomonas

<table>
<thead>
<tr>
<th>Genus</th>
<th>Pseudomonas</th>
<th>Stenotrophomonas</th>
<th>Staphylococcus</th>
<th>Prevotella</th>
<th>Fusobacterium</th>
<th>Streptococcus</th>
<th>Sphingomonas</th>
<th>Maraxella</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF/DC Ratio (MS Intensity)</td>
<td>109.7</td>
<td>49.0</td>
<td>21.1</td>
<td>16.0</td>
<td>8.8</td>
<td>5.8</td>
<td>4.5</td>
<td>0.001</td>
</tr>
<tr>
<td>Peptides Detected</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Function Example</td>
<td>Type II secretion system</td>
<td>SGL protein</td>
<td>Esp/MAP</td>
<td>ISS family transposase</td>
<td>Penicillin-binding prot.</td>
<td>LKG protein</td>
<td>Uncharacterized protein</td>
<td>Outer membrane prot.</td>
</tr>
</tbody>
</table>

Microbe MS Intensity in CF vs. DC

Monica Kruk
What about the human proteins?

### STRING-Db Pathway

<table>
<thead>
<tr>
<th>STRING-Db Pathway</th>
<th>CF/DC Ratio (MS intensity)</th>
<th>Increased in CF</th>
<th>Decreased in CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pendrin</td>
<td>196.54</td>
<td>13.99</td>
<td>6.62</td>
</tr>
<tr>
<td>Intraflagellar transport protein 56</td>
<td>5.69</td>
<td>4.56</td>
<td>0.25</td>
</tr>
<tr>
<td>Neutrophil Collagenase</td>
<td>0.16</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td>Serine/threonine-protein phosphatase</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Neutrophil Elastase</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Protein Examples

<table>
<thead>
<tr>
<th>Deoxyribonuclease-1</th>
<th>Dynein light chain Tctex-type 1</th>
<th>Protein S100-A8</th>
<th>Isoform 3 of Phosphorylase b kinase regulatory subunit beta</th>
<th>Neutrophil elastase</th>
<th>Guanylate-binding protein 2</th>
<th>Calcium-activated chloride channel regulator 1</th>
<th>Apolipoprotein E</th>
<th>Rootletin</th>
<th>Ankyrin-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pendrin</td>
<td>Intraflagellar transport protein 56</td>
<td>Neutrophil collagenase</td>
<td>Serine/threonine-protein phosphatase PPI-b alpha catalytic subunit</td>
<td>Azurocidin</td>
<td>Ubiquitin/ISG15-conjugating enzyme E2 L6</td>
<td>Chloride intracellular channel protein 3</td>
<td>Serum paraoxonase/arylesterase 1</td>
<td>Ciliary rootlet coiled-coil protein 2</td>
<td>Cingulin</td>
</tr>
</tbody>
</table>

Monica Kruk
Developing A Quantitative assay for host-microbial protein dynamics in clinical CF samples

Targeted mass spectrometry-based assays for quantifying levels of host-microbe proteins of interest

MICROBIAL PEPTIDES
HUMAN PEPTIDES
Conclusions & Next Steps

• We have developed a MS-based BALF analysis and Galaxy-driven bioinformatics workflow for processing of microbial and host proteins, generating verified microbial peptide candidates suitable for targeted analysis within individual patient samples.

• We have utilized this workflow to detect a promising microbial and host peptide panel for application to CF disease progression studies by comparing it with disease control (DC).

• Our goal is to quantify microbial and host peptides in individual samples, as an assay for characterizing bacterial and human dynamics in CF.

• We have been using this workflow in other clinical metaproteomic studies such as ovarian cancer microbiome study and COVID-19 co-infection analysis.