

Translational metaproteomics

Analysis of Functions Expressed by Microbiomes
CSIR-IMTech workshop, November 2021

Tim Griffin
University of Minnesota
tgriffin@umn.edu

Learn more at galaxyp.org
z.umn.edu/itcrgalaxyvideo



MICROBIOME

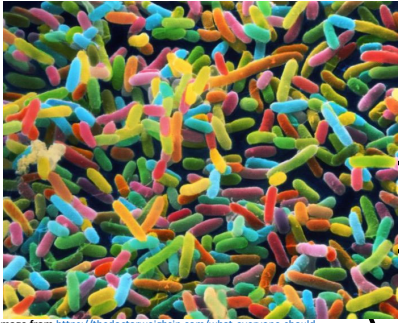
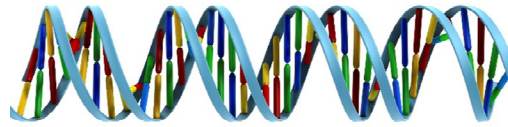


Image from <https://thedoctorweighsin.com/what-everyone-should-know-about-the-infant-microbiome/>

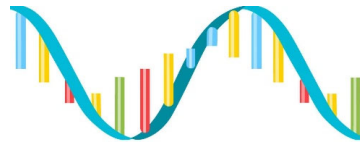
Potential to unravel the mechanistic details of microbial interactions with host / environment by analyzing the functional dynamics of the microbiome.



METAGENOMICS

TAXONOMY

function



METATRANSCRIPTOMICS

TAXONOMY

function



METAPROTEOMICS

TAXONOMY

FUNCTION

Microbial communities and microbiomes contribute to cancer

Oral Oncology 99 (2019) 104453

RESEARCH ARTICLE

The Pa
Oncog
Adapti

Smruti Pushalk
Ankita Mishra^a
Kevin Zhang^{a,1}
Juan Andres Ko
Wei Wang^a, Par
R. Balfour Sart

Department of
Department of
Prostate Cancer
Department of
Integrated Biolo
Department of

ARTICLE

Keywords:
Head and neck
Biomarkers
Microbiome
Microbiota
Oncogenesis
Radiotherapy
Immunotherapy
Toxicity

ABS

are d
prote
of ba
assoc
tion i
TH1

Article

The gut microbio
from tumour-sup

<https://doi.org/10.1038/s41586-020-2541-0> Eilran
Received: 2 April 2019 Ela Ely
Accepted: 1 May 2020 Thorst
Published online: 29 July 2020
Check for updates

Somat
and of
Here v
that et
Csnk1
facilit
effects
oncog
prono
p53 eli
and pr
these
inhibit
mutar
bindin
compl



Article Advance


Cancer Letters



Volume 380, Issue 1, 28 September 2016, Pages 356-358

CANCER LETTERS

Commensal bacteria modulate the tumor microenvironment

Theofilos Poutahidis^{a, b}, Susan E. Erdman^a  

Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.canlet.2015.12.028> Get rights and content

'Secretomics' to identify microbial molecules and host interactions

Proteomics

Proteomics and Systems Biology

Review | [Full Access](#)

Analyzing the Secretome of Gut Microbiota as the Next Strategy For Early Detection of Colorectal Cancer

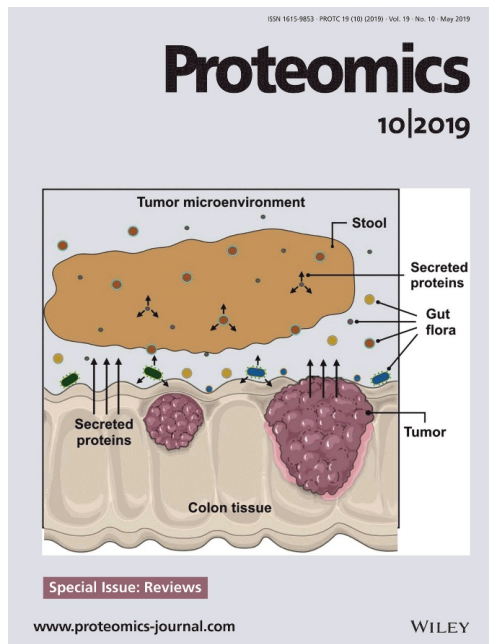
Putri-Intan-Hafizah Megat Mohd Azlan, Siok-Fong Chin , Teck Yew Low, Hui-min Neoh, Rahman Jamal

First published: 17 December 2018 | <https://doi.org/10.1002/pmic.201800176>

- Microbiota-secreted molecules mediate host-bacteria interactions and may regulate host cells

Secreted molecules:

- **Proteins** identified via metaproteomics
- **Peptides** may contribute (metapeptidomics); non-tryptic peptides, LC-MS based quantification
- **Metabolites** are also key: MS-based methods both targeted and untargeted; guided by metaproteomics and/or metatranscriptomics
- **Challenges:** sample prep, abundance of secreted molecules, analytical methods for identification and quantification of proteins/peptides/metabolites



Metaproteomic re-analysis of clinical cancer samples

- Clinical tissue samples collected for quantitative MS-based proteomics of host

STUDY SUMMARY: Proteogenomics of Gastric Cancer - Proteome

80
Cases

160
Aliquots

SUMMARY

PDC Study Identifier	PDC000214	Embargo Release Date	N/A
Study ID	e386be5-693e-11ea-b16f-0aad30af8a83	Analytical Fraction	Proteome
Study Name	Proteogenomics of Gastric Cancer - Proteome	Disease Types	Early Onset Gastric Cancer
Experimental Strategy	ITRAQ4	Project ID	Human Early-Onset Gastric Cancer - Korea University

[Description](#) | [Protocol](#) | [Experimental Design](#) | [Clinical](#) | [Biospecimens](#) | [Workflow](#) | [DUA](#)

Dong-Gi Mun, Jinyuk Bhn, Sangok Kim, Hyunwoo Kim, Jee Hun Jung, et al., *Cancer Cell* (2019) Vol. 35, Issue 1, p111-124 <https://doi.org/10.1016/j.ccr.2018.12.003>

We report proteogenomic analysis of diffuse gastric cancers (GCs) in young population. Phosphoproteome data elucidated signaling pathways associated with somatic mutations based on mutation-phosphorylation correlations. Moreover, correlations between mRNA and protein abundances provided potential oncogenes and tumor suppressors associated with patient survival. Furthermore, integrated clustering of mRNA, protein, phosphorylation, and N-glycosylation data identified four subtypes of diffuse GCs. Distinguishing these subtypes was possible by proteomic data. Four *in vitro* assays associated with proliferation, immune response, metabolism, and invasion, respectively, and associations of the subtypes with immune and immunosuppressive pathways were identified mainly by phosphorylation and N-glycosylation data. Therefore, our proteogenomic analysis provides additional information beyond genomic analyses, which can improve understanding of cancer biology and patient stratification in diffuse GCs.

Mass Spectrometry Data Sets, Provided Below

Peptides from tumor and adjacent normal tissues of early-onset gastric cancer (EOGC) patients were labeled with 4plex ITRAQ reagent (AB Sciex) according to the manufacturer's instructions. For each of the first 50 patients, two independent biological replicates of peptide samples from the pair of tumor and adjacent normal tissues were merged into one peptide pool (i.e., 800 µg/replicate), respectively, and each of two 800 µg samples of normal and tumor peptides was labeled with 114/116 and 115/117 ITRAQ reagents, respectively. For the rest 30 patients, the peptide samples for two pairs of tumor and adjacent normal tissues from two patients were merged into one peptide pool (800 µg each), respectively. The normal and tumor peptides of a patient were labeled with 114 and 115 ITRAQ reagents, respectively, while the normal and tumor peptides of another patient were labeled with 116 and 117 ITRAQ reagents.

There are a total of 65 ITRAQ experiments for global proteome analyses (3.48 TB raw data, 1560 files), 65 ITRAQ experiments for phosphoproteome (1.19 TB raw data, 780 files) and glycoproteome (1.35 TB raw data, 779 files).

Common Data Analysis Pipeline (PDC Harmonization) data		Supplementary data	
Data Category	Files (n=6246)	Data Category	Files (n=4)
Raw Mass Spectra (Proprietary)	1560	Other Metadata (Document)	4
Processed Mass Spectra (Open Standard)	1560		
Peptide Spectral Matches (Open Standard)	1560		
Peptide Spectral Matches (Text)	1560		
Protein Assembly (Text)	4		

Explore protein quantitation from PDC Common Data Analysis pipeline (CDAP) through heatmaps

STUDY SUMMARY: CPTAC HNSCC Discovery Study - Proteome

124
Cases

195
Aliquots

SUMMARY

PDC Study Identifier	PDC000221	Embargo Release Date	N/A
Study ID	95296fd1-694e-11ea-b16f-0aad30af8a83	Analytical Fraction	Proteome
Study Name	CPTAC HNSCC Discovery Study - Proteome	Disease Types	Head and Neck Squamous Cell Carcinoma; Other
Experimental Strategy	TMT11	Project ID	CPTAC3-Discovery

[Description](#) | [Protocol](#) | [Experimental Design](#) | [Clinical](#) | [Biospecimens](#) | [Workflow](#) | [DUA](#)

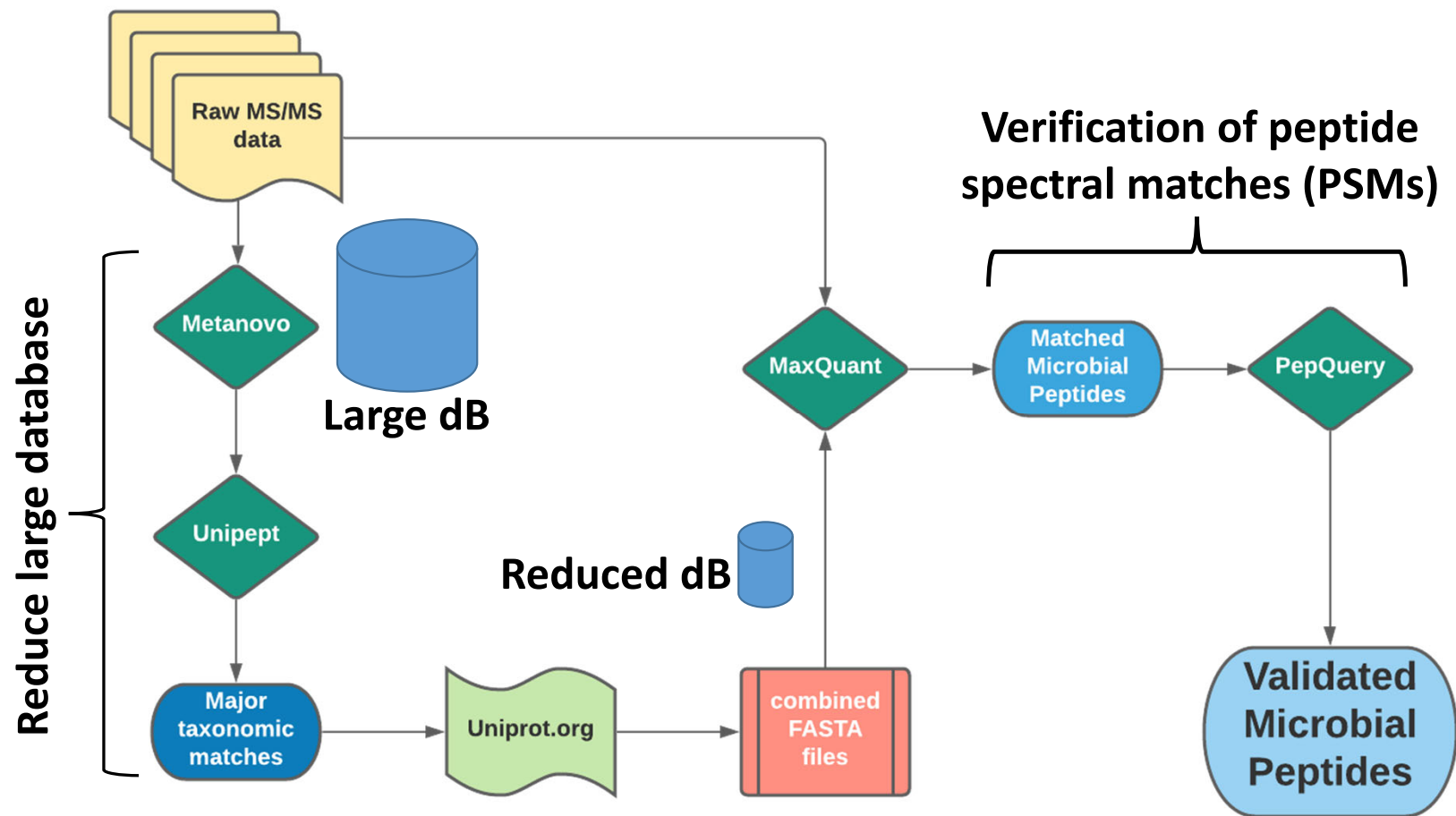
Head and neck squamous cell carcinoma (HNSCC) is the seventh most common cancer worldwide, accounting for approximately 3% of all cancers (Chow, NEJM 2020). In the United States it is estimated that in 2020 there will be 53,000 new cases and over 10,000 deaths (National Cancer Institute, NCI, Surveillance, Epidemiology, and End Results Program; SEERS). These tumors are challenging to treat because of their genetic complexity and heterogeneous nature (Alshafiq et al., Cell Death & Disease 2020). To advance the understanding of HNSCC, the CPTAC program has performed a comprehensive integrated proteogenomic characterization of a 108 human papilloma virus (HPV)-negative HNSCC cohort. Global proteome and phosphoproteome analysis from an optimized workflow mass spectrometry of tissues using isotopic tags (TMT (tandem mass tag)-11) was used (Martins et al., Nature Protocols 2018). Proteome and phosphoproteome data from the HNSCC discovery cohort is available below along with peptide spectrum matches and protein summary reports from the CPTAC common data analysis pipeline (CDAP).

Common Data Analysis Pipeline (PDC Harmonization) data		Supplementary data	
Data Category	Files (n=2008)	Data Category	Files (n=7)
Raw Mass Spectra (Proprietary)	500	Other Metadata (Document)	7
Processed Mass Spectra (Open Standard)	500		
Peptide Spectral Matches (Open Standard)	500		
Peptide Spectral Matches (Text)	500		
Protein Assembly (Text)	6		
Quality Metrics (Text)	1		
Quality Metrics (Web)	1		

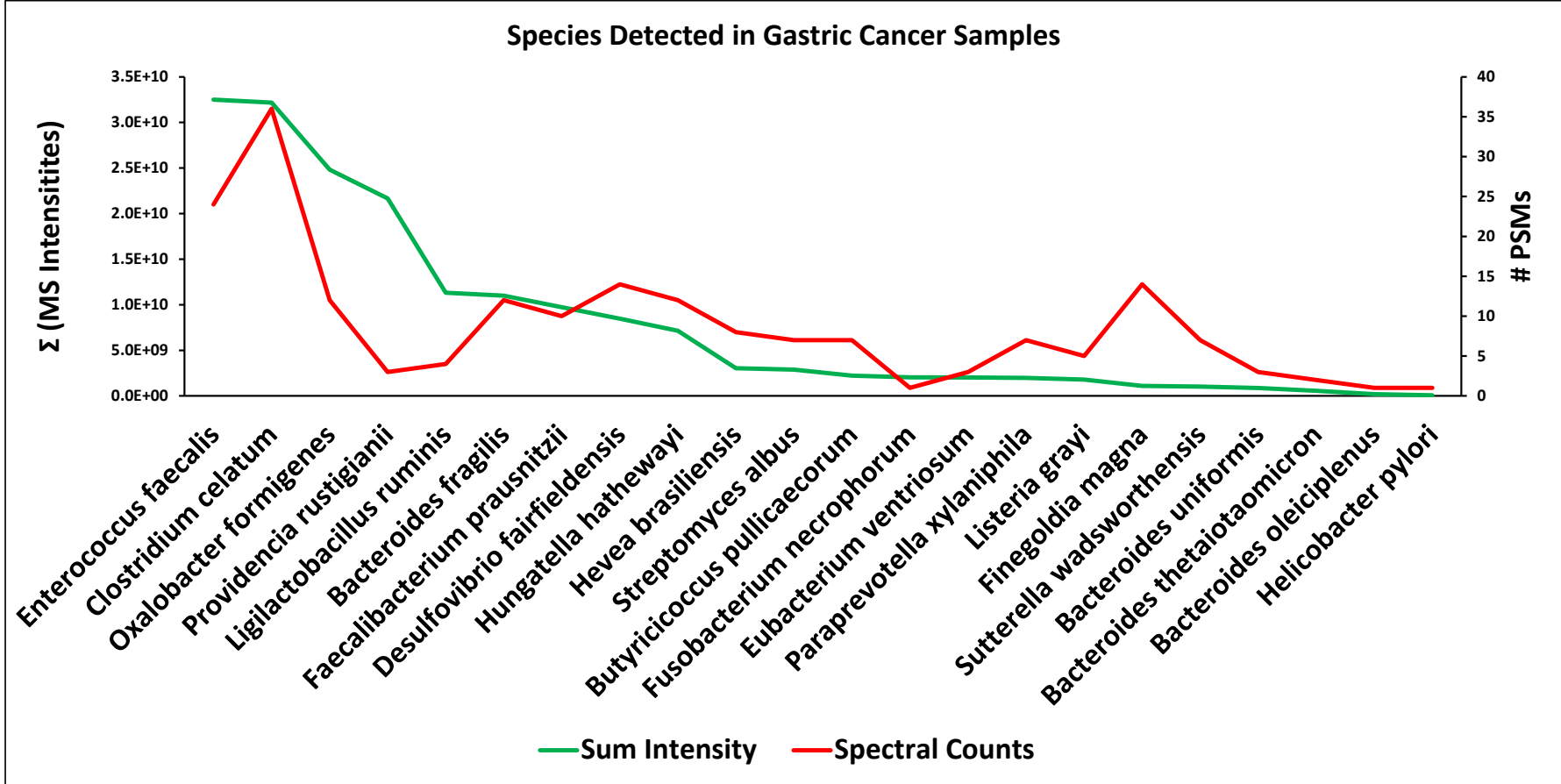
Explore protein quantitation from PDC Common Data Analysis pipeline (CDAP) through heatmaps

(Andrew Rajczewski)

Workflow: reducing protein sequence search space for MS/MS data

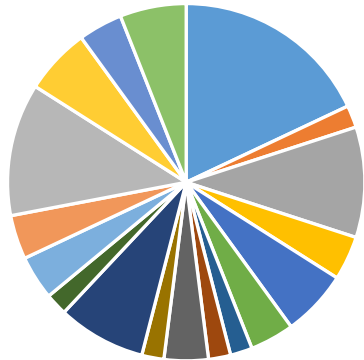


Microbial peptides are present and detectable in gastric tumor samples



Microbial taxa and enriched functions of metaproteins in gastric cancer

Microbial Genera of Validated Peptides

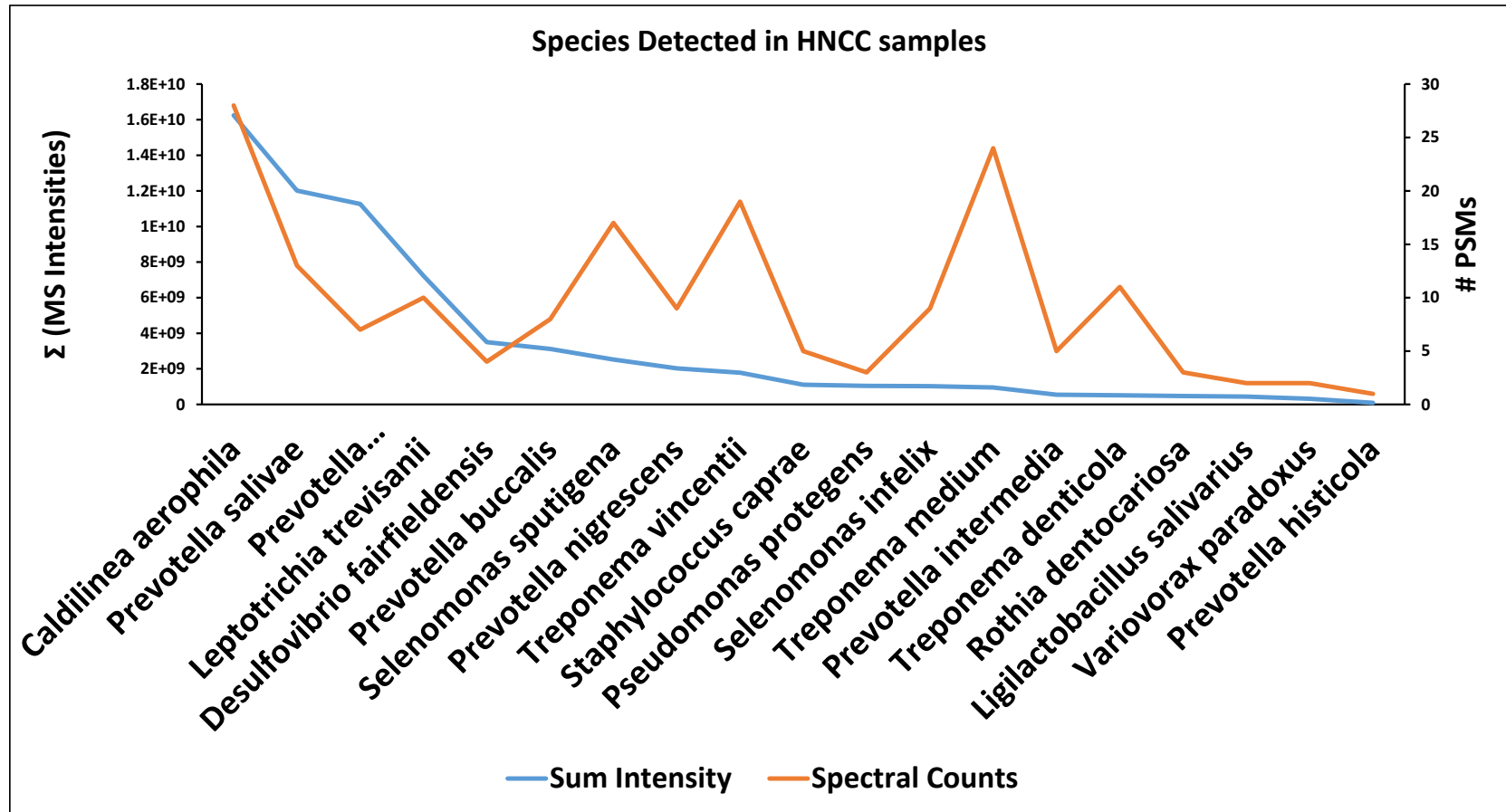


- | | | |
|--------------------|------------------|---------------------|
| ■ Bacteroides | ■ Butyricicoccus | ■ Clostridium |
| ■ Desulfovibrio | ■ Enterococcus | ■ Eubacterium |
| ■ Faecalibacterium | ■ Finegoldia | ■ Fusobacterium |
| ■ Helicobacter | ■ Hungatella | ■ Ligilactobacillus |
| ■ Listeria | ■ Oxalobacter | ■ Paraprevotella |
| ■ Providencia | ■ Streptomyces | ■ Sutterella |

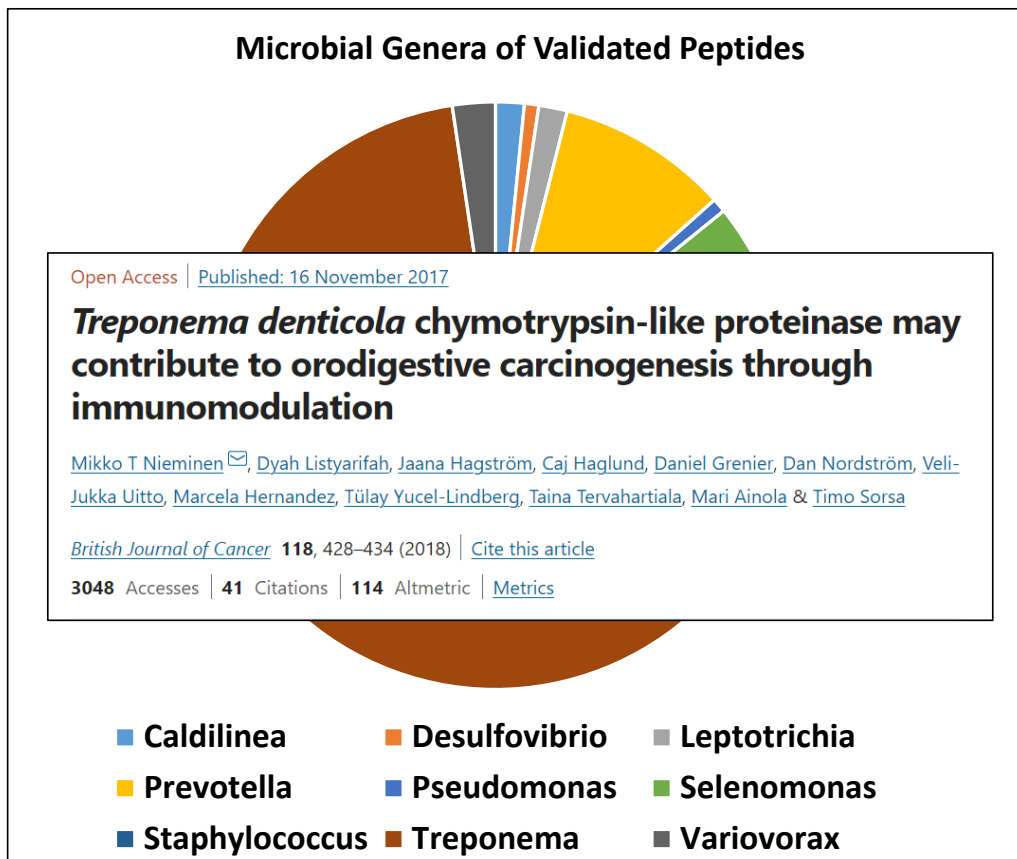
GO Biological Processes in validated microbial peptides of Gastric Cancer data



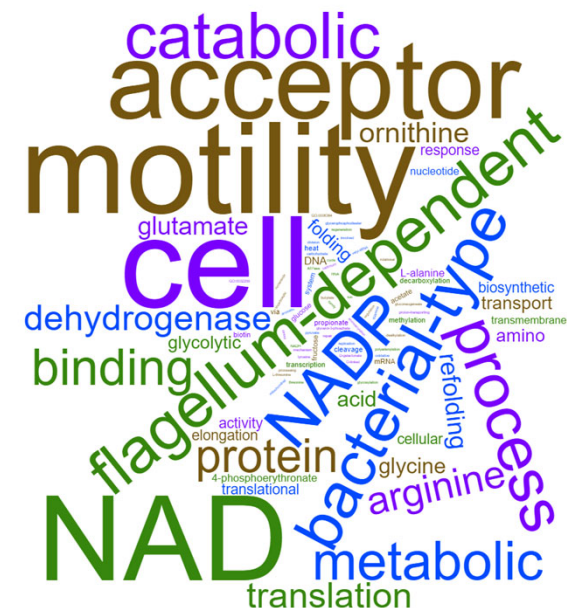
Microbial peptides are present and detectable in head/neck cancer



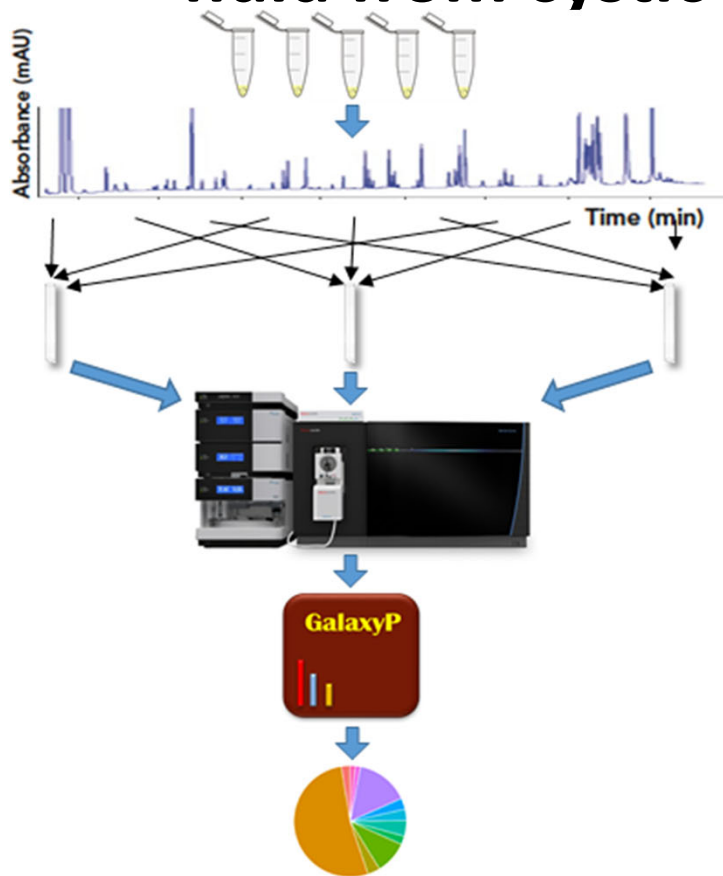
Microbial taxa and enriched functions of metaproteins in HNSCC



GO Biological Processes in validated microbial peptides of HNSCC data



Other translational studies: Microbiota contributions in broncheolavage fluid from cystic fibrosis (CF) patients

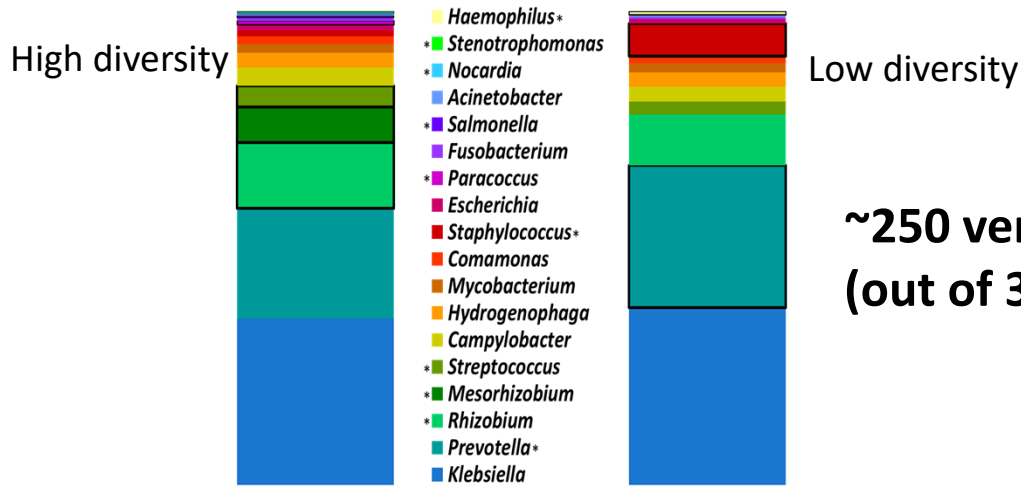


Generating comprehensive databases of CF and DC microbiomes:

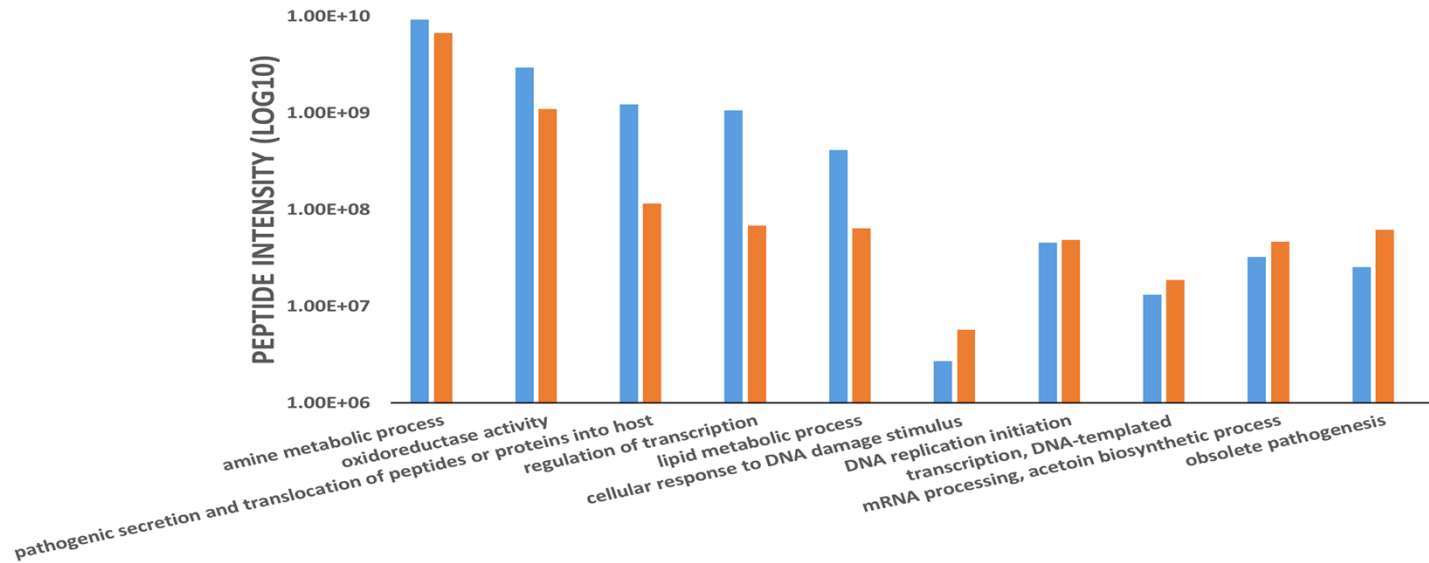
- 1) BALF pelleting, lysis, and digestion;
- 2) High pH RPLC offline fractionation and concatenation (step omitted for individual sample analysis)
- 3) LC-FAIMS-MS/MS on an Orbitrap Eclipse
- 4) Galaxy-P/MetaNovo for matching MS/MS;
- 5) Generation of a compact microbial database with taxonomic and functional information from Galaxy-P/Unipept+metaQuantome.

(Monica Kruk, and collaborators from Lurie's Children's Hospital, Chicago, IL)

Deep profiling of CF BALF cells to identify microbial contributors



~250 verified microbial peptides
(out of 30K+ host peptides)



(Monica Kruk)

Summary: translational metaproteomics offers challenges and opportunities

- Characterizing metaproteins in clinically-relevant samples provides valuable insights into functional microbe-derived molecules that may regulate host phenotypes
- Complementarity to metagenomics/metatranscriptomics; metaproteomics is helped by availability of paired metagenomic or metatranscriptomic sequencing data
- Analytical challenges: in many clinical samples the microbial proteins are rather low in abundance
 - Requires steps to increase depth/sensitivity
 - Enrichment of microbial proteins could be huge benefit

