



A meta-omics workflow to generate a novel host and microbial biomarker panel for early detection of ovarian cancer

Subina Mehta

Galaxy for Proteomics
(Galaxy-P) team

University of Minnesota

Kristin Boylan, Ashley Petersen, Amy Skubitz, Pratik Jagtap, Timothy Griffin

Introduction



OCRA
ovarian cancer
research alliance

woman's lifetime risk of
developing OC is 1 in 78

5-year survival rate

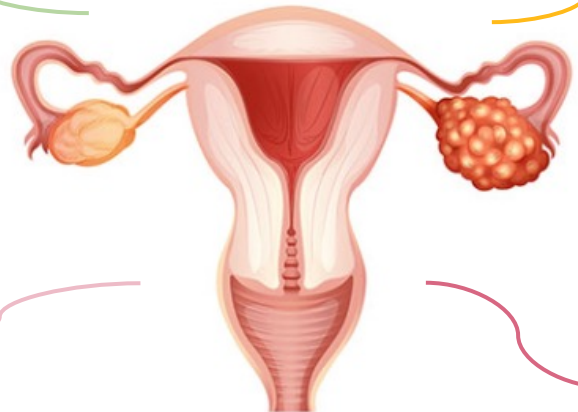
Stage 1- 92.4%
Stage 2 or 3 -71%

Diagnostic Biomarkers



CA-125
HE4

**PAP-TEST FLUID AS
SOURCE?**



Dysbiosis in female
reproductive tract
linked to Ovarian cancer
development

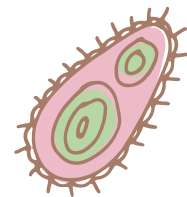
MS-based Bioinformatic
analysis indicates pap test
fluid (PTF) samples are a rich
source of microbial and
human proteins

Microbial attribute?



<https://www.vejthani.com/diseases-conditions/ovarian-cancer/>

OVARIAN CANCER METAPROTEOMICS



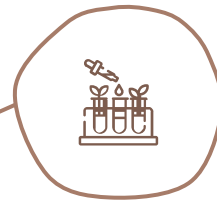
Hypothesis

Microbiome-expressed biomarkers alongside host for early detection?



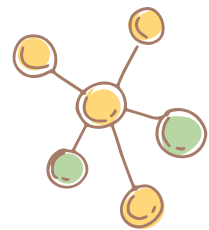
Source

PAP-TEST FLUID AS SOURCE?



Analysis

Aim to use Galaxy platform to investigate and generate host-microbial biomarker panel for early detection of ovarian cancer



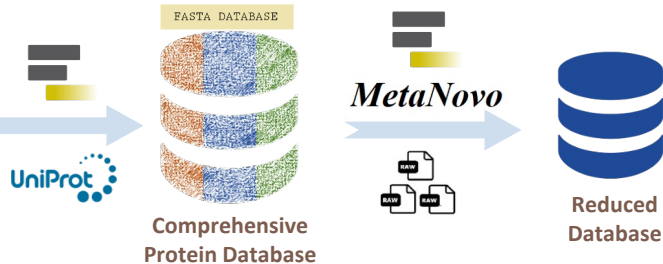
GALAXY BIOINFORMATIC WORKFLOW



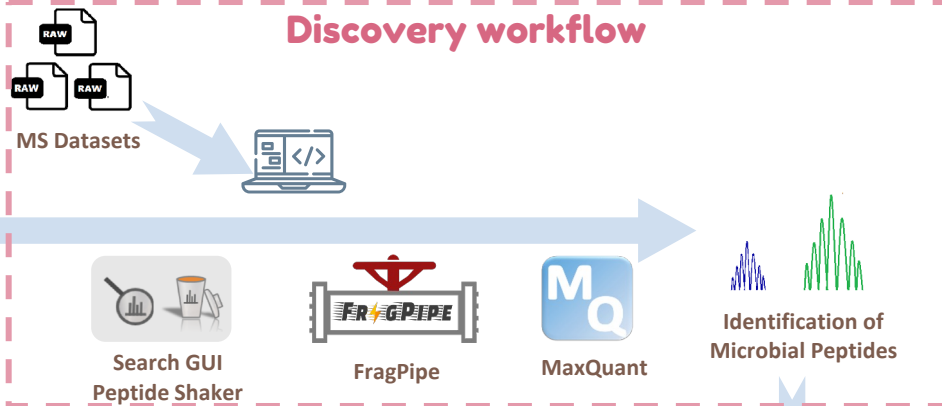
NORMAL
BENIGN
OVARIAN
CANCER

20 HGSOC
and 20 non-
OC women

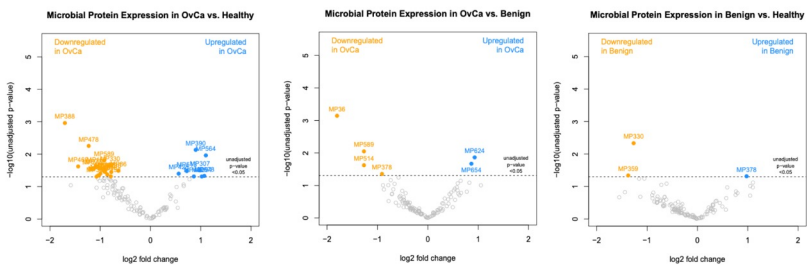
Database generation



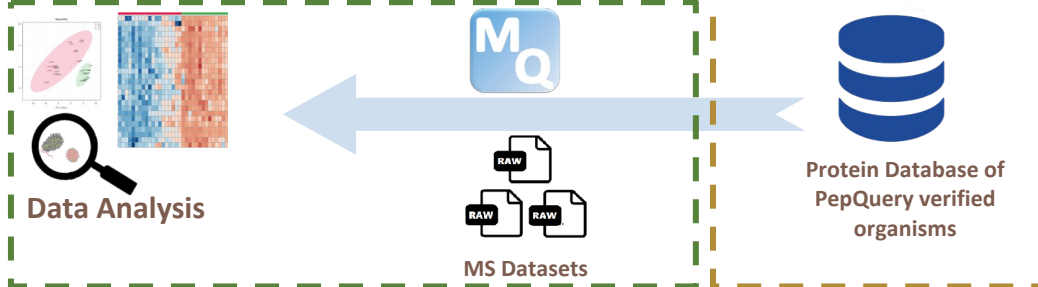
Discovery workflow



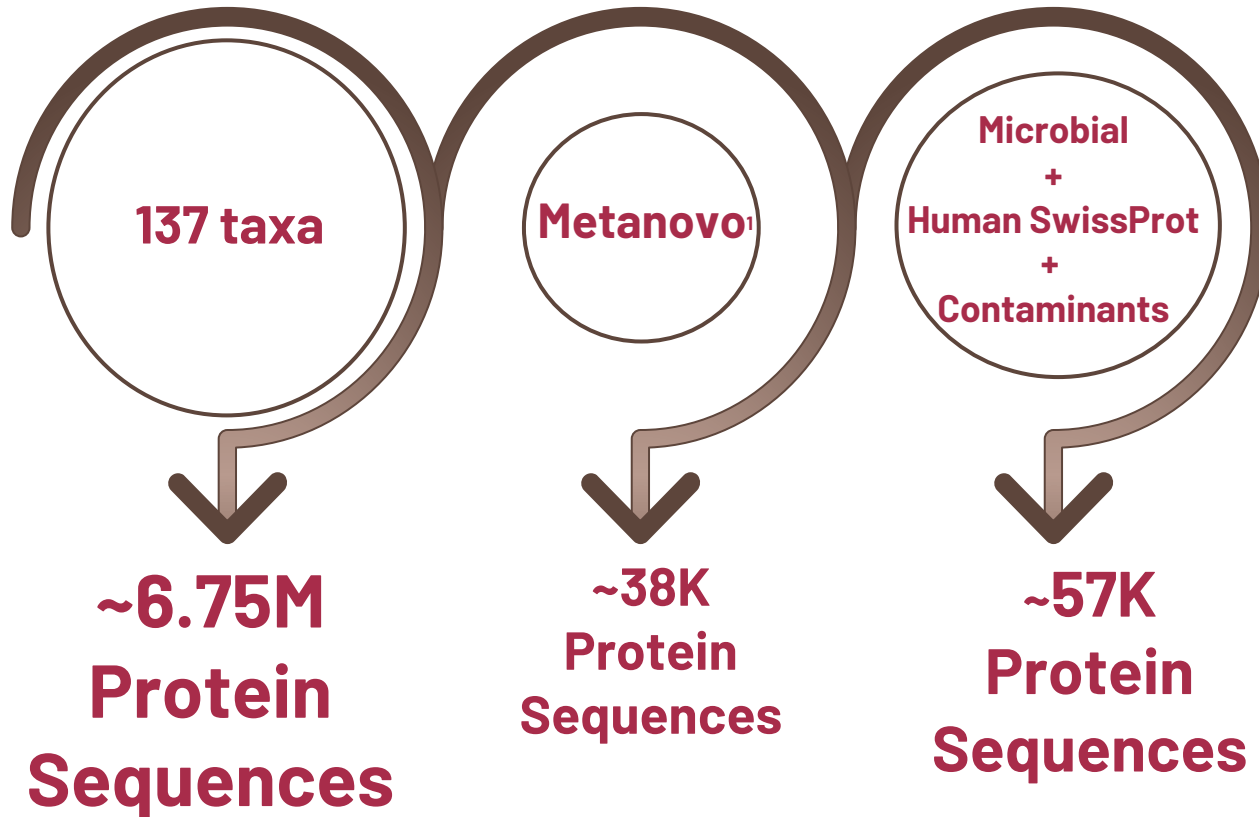
Statistical Analysis

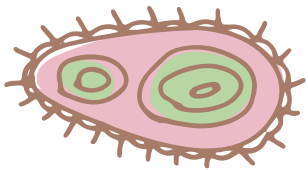


Quantitation

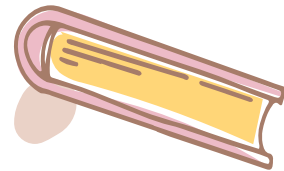
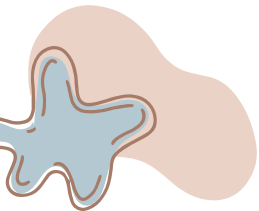
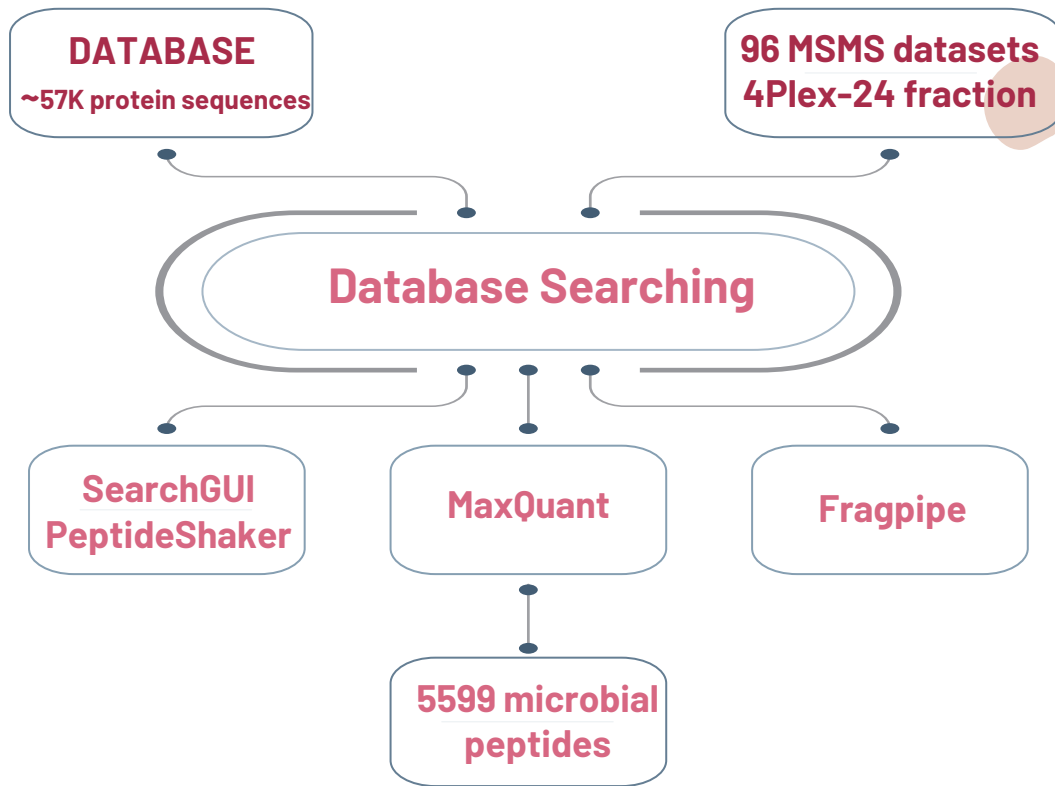


Database Generation





Discovery Workflow

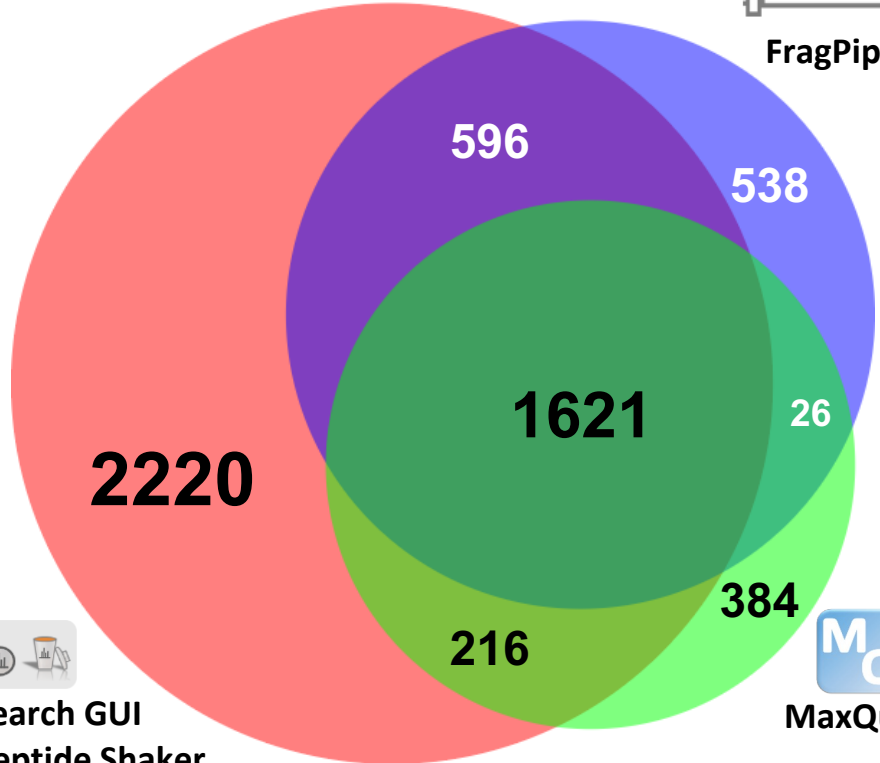


5599 MICROBIAL

DeepVenn
(C) 2020 Tim Hulsen



FragPipe



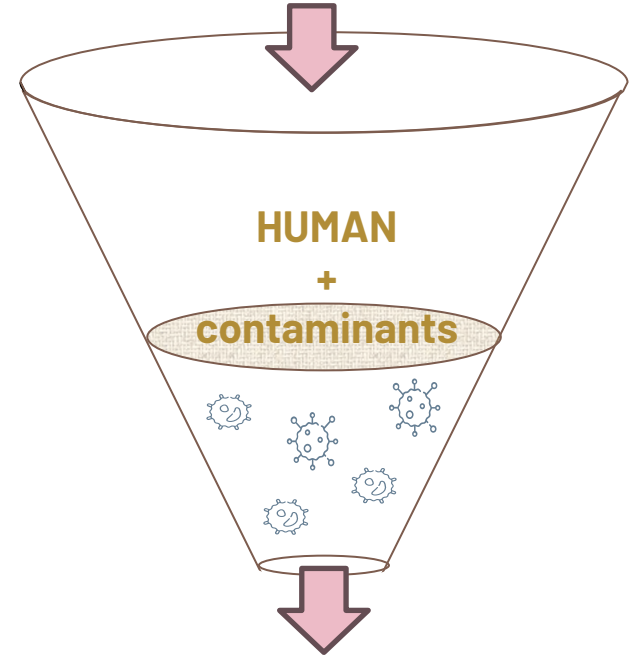
Search GUI
Peptide Shaker



MaxQuant

PepQuery*

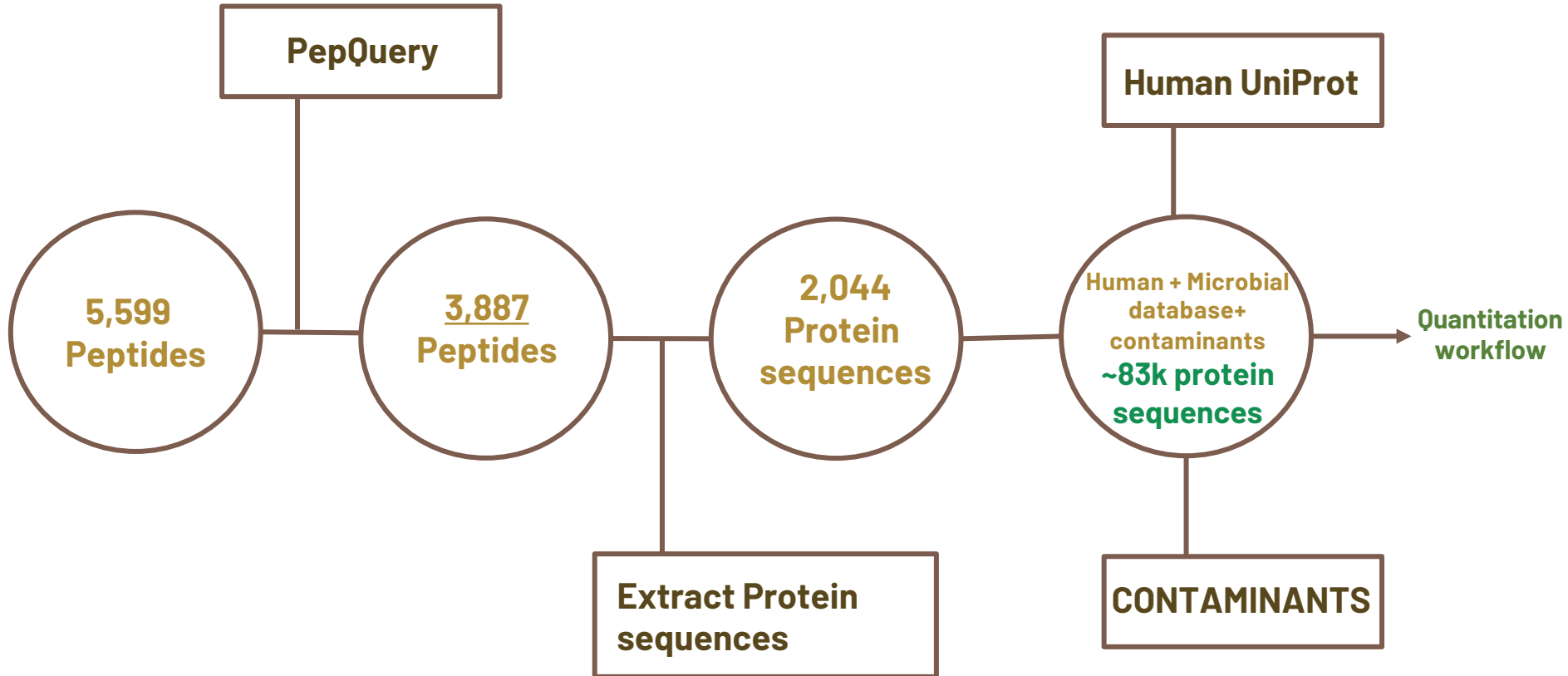
5599
microbial
peptides



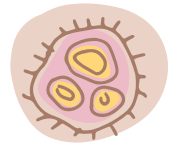
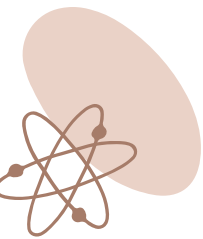
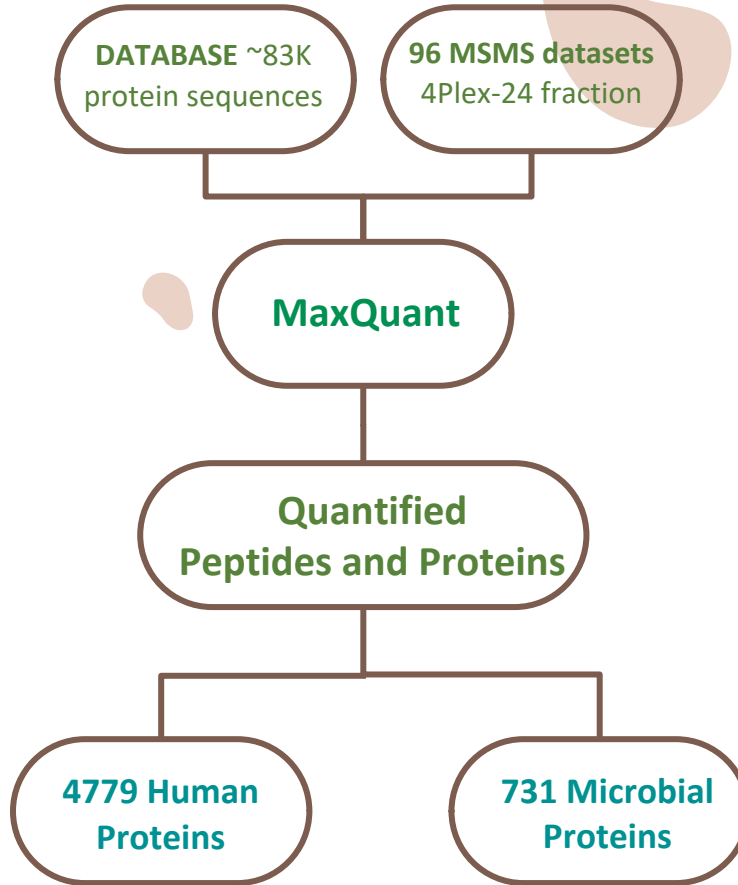
* Wen et.al.(2019): doi:10.1101/gr.235028.118



Verification Workflow



Quantitation workflow



Statistical Analysis



Ashley Petersen, PhD

- Data filtering criteria
- Observed in $\frac{3}{4}$ plex's
- 10% non plex missingness

Missingness observed

ROC curve

Performance Measurement

OVCA-Benign

01

02

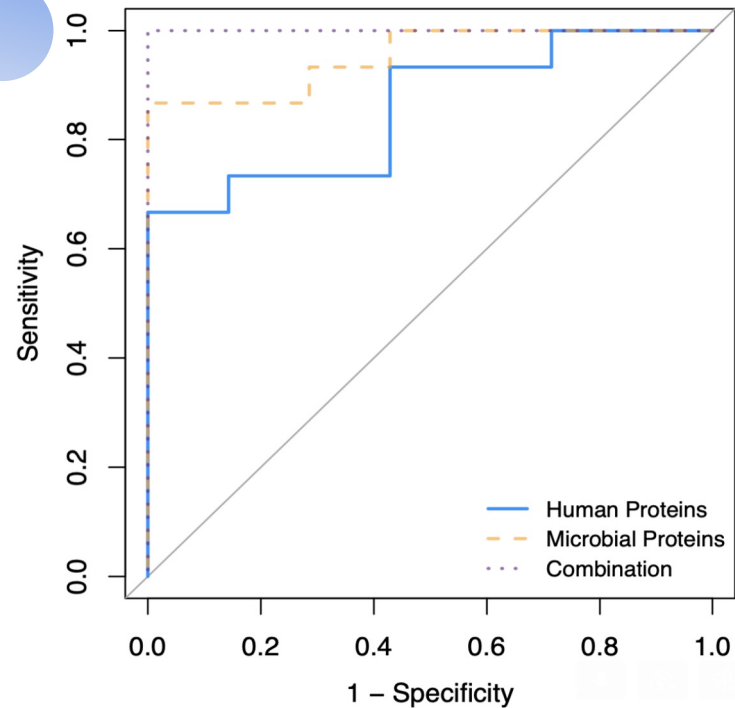
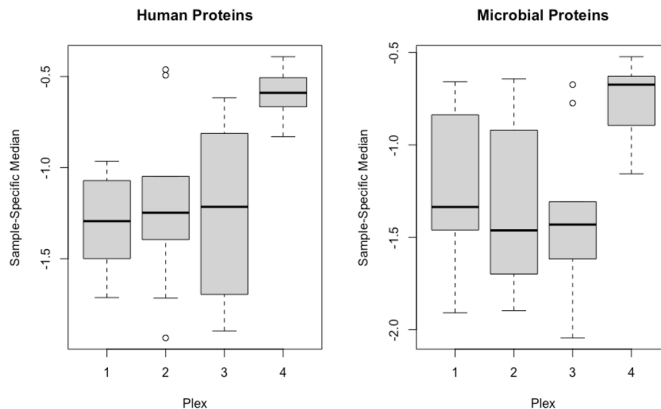
03

04

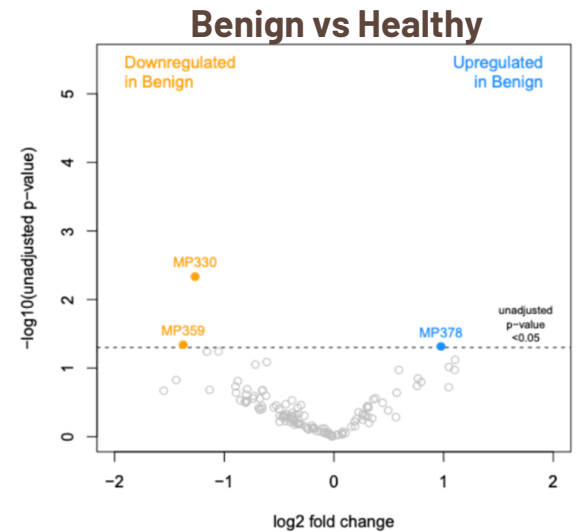
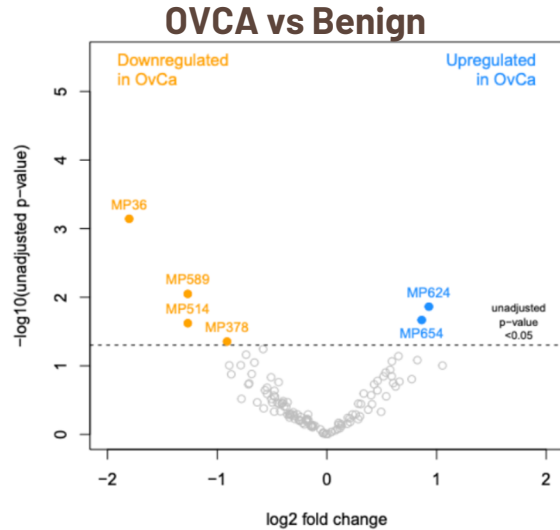
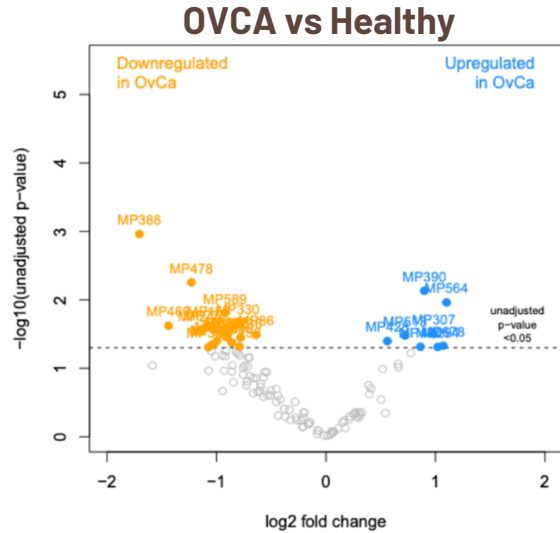
normalization

Data Quality

- Limma based
- Log ratios- median normalized



Microbial Protein Abundance Results

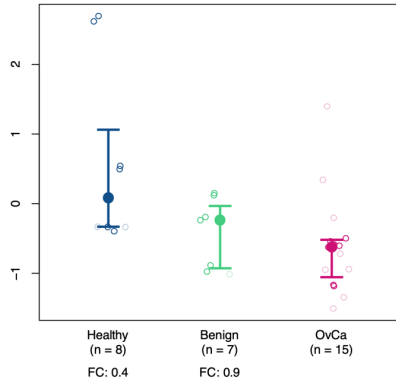


Differentially abundant	OVCA-H	OVCA-B	Benign-H
Microbial Peptides	104	6	8
Microbial Proteins Groups	26 (16↓ and 8↑)	6 (4↓ and 2↑)	3 (2↓ and 1↑)

Microbial Protein Results

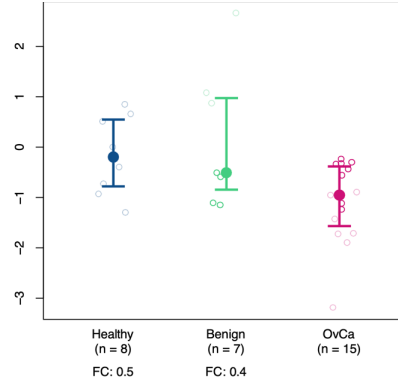
OVCA-Healthy

Surface layer Protein A domain containing protein
Lactobacillus crispatus



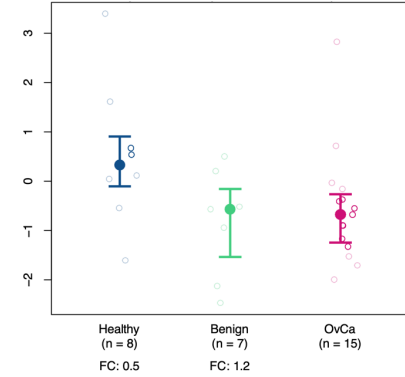
OVCA-Benign

DUF4879 domain containing protein
Bacteroides thetaiotaomicron

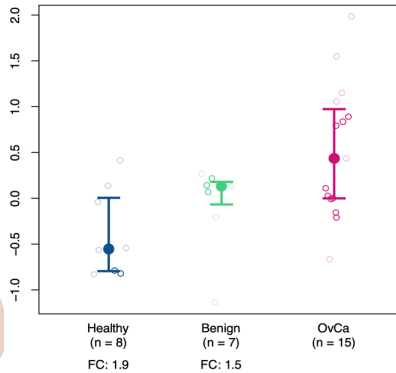


Benign-Healthy

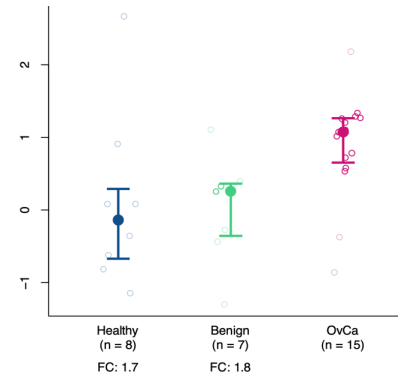
N-acetylneuraminatase
Lactobacillus crispatus



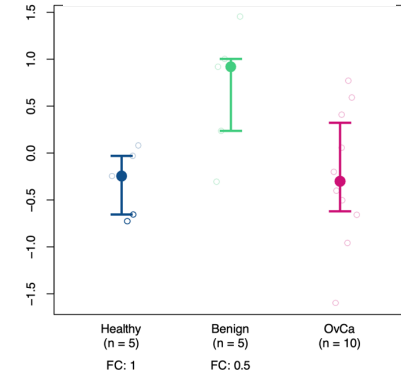
HTH cro/C-1-type domain containing protein
Clostridiales bacterium



Thiol-disulfide interchange protein Dsb-D
Escherichia coli

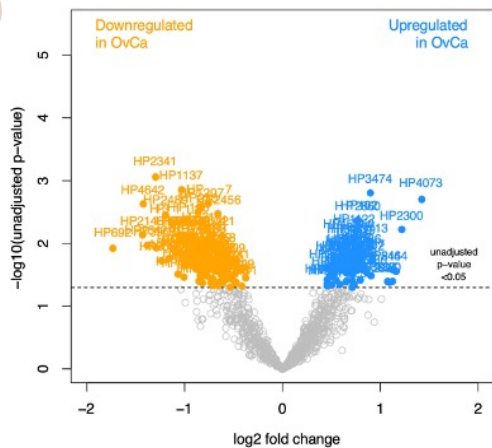


Phage tail tape measure protein
Enterococcus faecalis

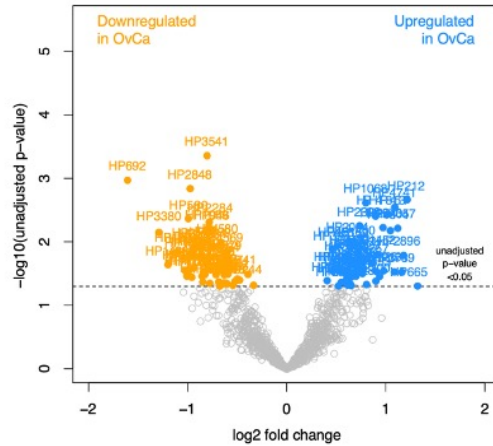


Human proteins

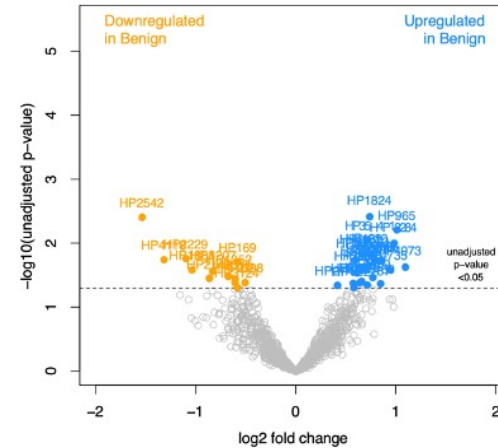
OVCA vs Healthy



OVCA vs Benign



Benign vs Healthy



Differentially abundant	OVCA-H	OVCA-B	Benign-H
Proteins Groups	167	107	39

Conclusion

01

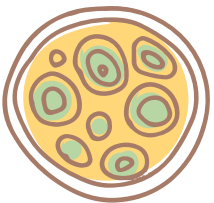
Developed MS-based Galaxy-driven bioinformatics workflow for processing of microbial and host proteins, generating verified microbial peptide candidates suitable for targeted analysis within individual patient samples.

02

Preliminary results show that microbial proteins along with host proteins could improve early detection.

03

Predictive analysis looks promising and plan to follow up with more studies.



Ongoing Plans

	Literature survey database	16S rRNA based database
Taxa identified (genera and species incl)	137	192
Comprehensive protein sequence database	~6.75 million protein sequences	~23.2 million protein sequences
After Metanovo (incl Human SwissProt and contaminants)	~57 thousand protein sequences	~142 thousand protein sequences
Database search microbial Peptides	5599 microbial peptides	



Verification, Quantification and statistical analysis to follow.....



ACKNOWLEDGMENTS



University of Minnesota

Timothy Griffin
Pratik Jagtap
Monica Kruk
Andrew Rajczewski
Danielle Weise
Katherine Do



Minnesota Supercomputing Institute

James Johnson
Reid Wagner

Alexey Nesvizhskii
Fengchao Yu
University of Michigan

Jagroop
Pandhal

Valerie Schiml
Magnus Øverlie Arntzen
NMBU, Oslo, Norway

Anshu Bharadwaj
Chandigarh
India

University of Minnesota

Amy Skubitz
Kristin Boylan
Ashley Petersen

Saskia
Hiltmann

Magnus
Palmlad
Netherlands

Thilo Muth
Germany



Brook Nunn
U of Washington

Jean Armengaud
France

Biologists / collaborators

Chris Wendt
Teresa Laguna
Maneesh Bhargava
Ryan Hunter
David Largaespada

Brian Searle
Ohio State
University

Michael
Shortreed
UW-Madison

Alessandro
Tanca
Italy

Surbhi Bihani
Sanjeeva Srivastava
IIT, Mumbai
India

Bing Zhang
Bo Wen
Baylor College of Medicine

Valdemir Carvalho
Fleury Group, Brazil

Björn Gruening
Melanie Foell
University of Freiburg,
Freiburg, Germany

Matt Chambers
Nashville, TN

Maria Doyle
Melbourne,
Australia

<http://galaxyp.org/contact/>

galaxyp.org

twitter.com/usegalaxyp

Unipept Analysis

