

An Integrated Quantitative Proteogenomics Pipeline Reveals Non-Canonical Variant Peptides in Inflamed Colon Tissue



Masonic Cancer Center

UNIVERSITY OF MINNESOTA

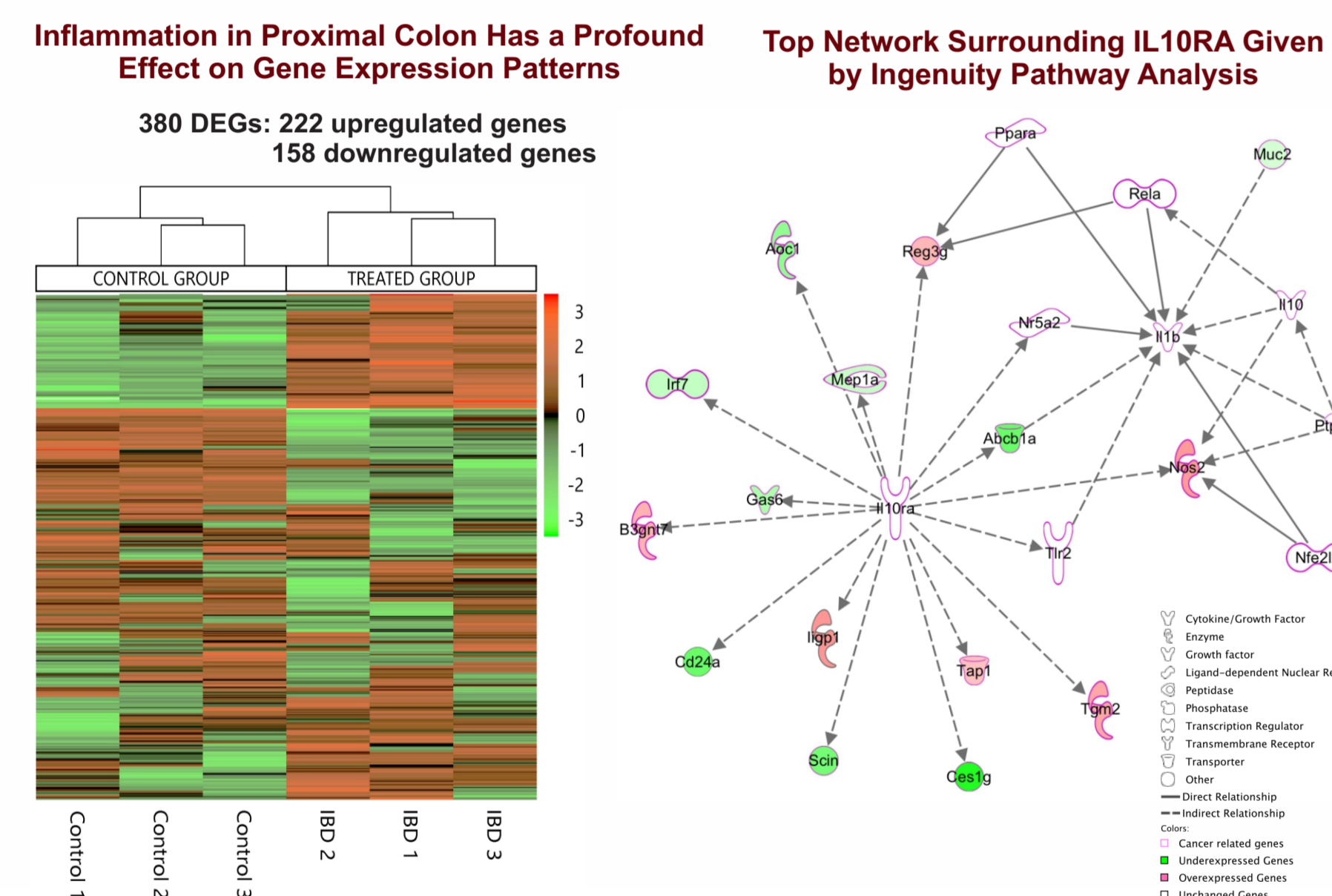
Comprehensive Cancer Center designated by the National Cancer Institute

Andrew Rajczewski,¹ Qiyuan Han,¹ Subina Mehta,¹ Praveen Kumar,¹ Pratik D Jagtap,¹ Timothy J. Griffin,¹ Natalia Y. Tretyakova^{1,2}

Introduction

- When bottom-up proteomics experiments utilize FASTA files derived from genomic data, information on proteoforms unique to the samples under analysis is missed (indels, intron retention, alternative splicing, etc.)
- Proteogenomics workflows utilize custom-built FASTA libraries based on transcriptomic data to inform attempts to detect these protein variants
- We analyzed a mouse inflammatory bowel disease model subjected to infection with *Helicobacter hepaticus* using nano-LC mass spectrometry with a FASTA database generated using RNA-Seq data from these tissues.
- Peptide variants were detected using a bioinformatics workflow in Galaxy-P and subsequently validated computationally using PepQuery and spectrally using targeted MS/MS experiments
- Non-canonical peptide variants were found enriched and depleted in inflamed proximal colon samples.

RNA-Seq Analysis of Inflamed Proximal Colon



- Previously, RNA-Seq analysis was performed on proximal colon samples from IBD model mice after five months of inflammation
- The transcriptomes of mice cluster independently into control and inflamed samples
- Differentially expressed genes in the inflamed samples cluster into 222 upregulated genes and 158 downregulated genes
- Ingenuity pathway analysis of the RNA-Seq data shows many genes showing differential expression are controlled by IL-10RA, which initiates the anti-inflammatory response via STAT3

Hypothesis

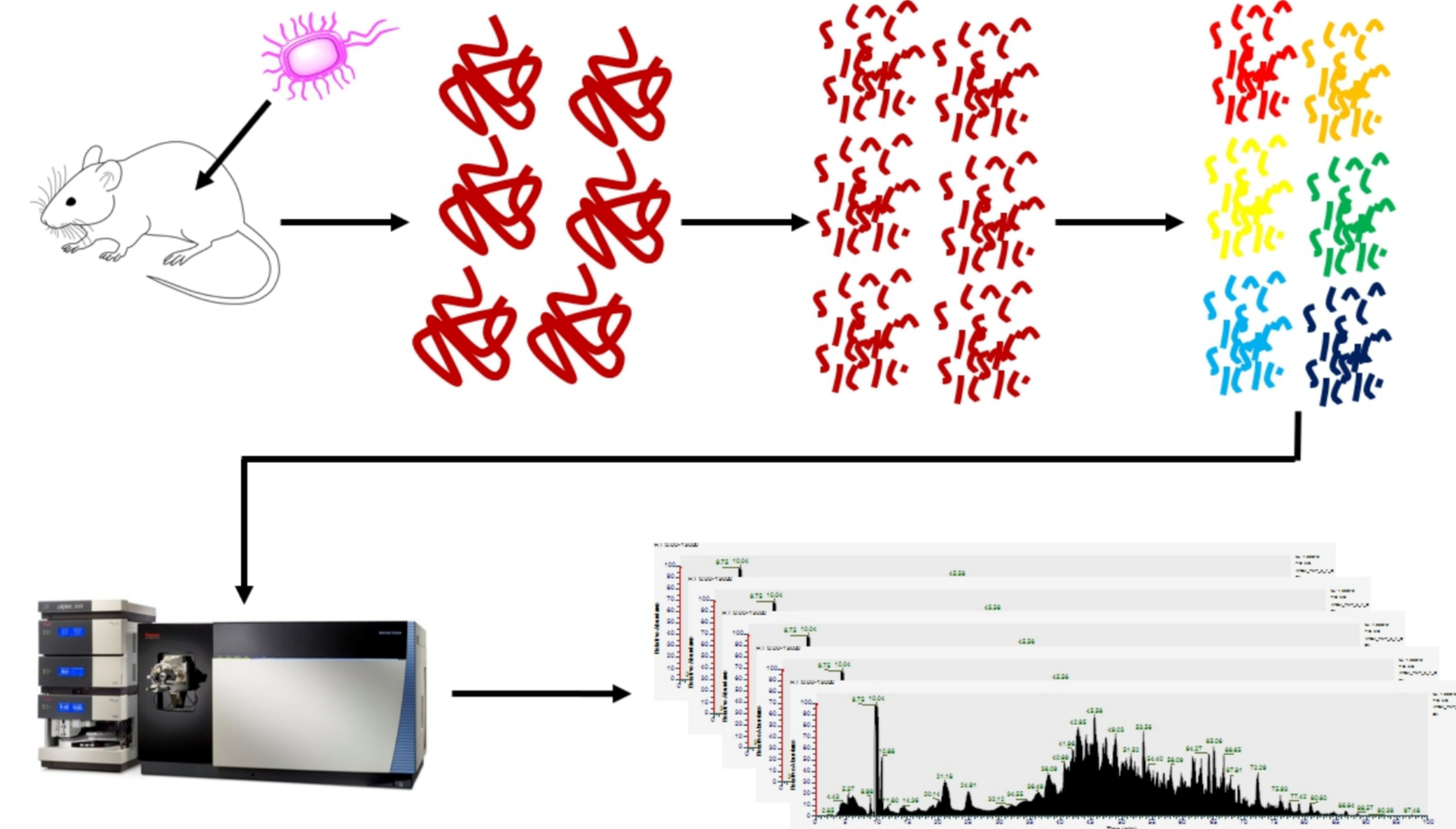
- Proteomic analysis will give a more complete picture of the inflammation phenotype associated with this IBD model
- A proteogenomic analysis of the IBD model during inflammation will reveal non-canonical variant peptides which may serve as biomarkers of IBD

Goals

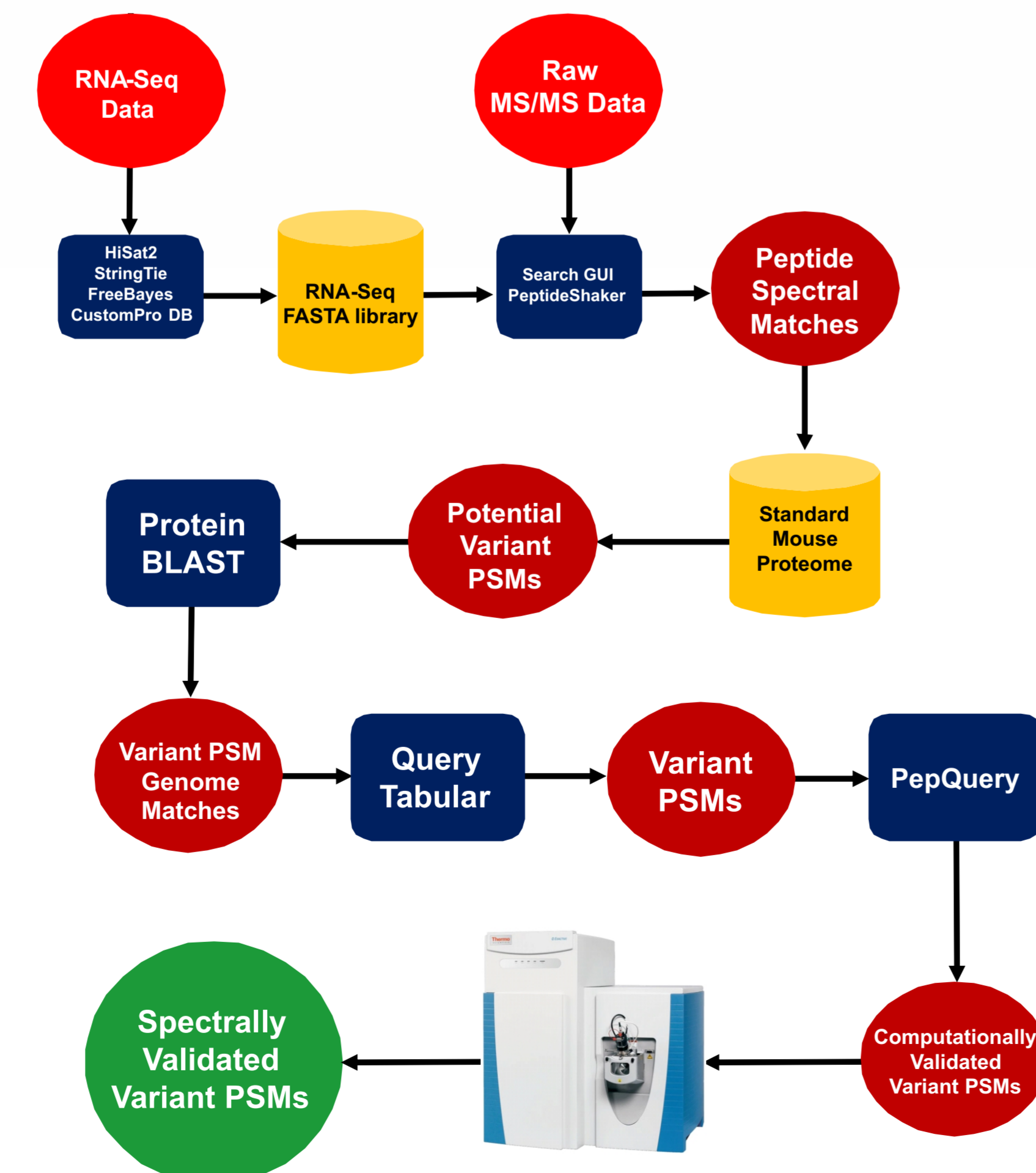
- Assemble the complementary RNA-Seq to create a customized FASTA library
- Isolate proteins from proximal colon samples, process and generate global mass spectrometry data
- Identify non-canonical peptide variants in the treated and control MS/MS data
- Validate these peptides computationally using PepQuery
- Validate and quantitate directly any and all variant peptides using targeted parallel reaction monitoring (PRM) experiments

Generation of MS/MS data

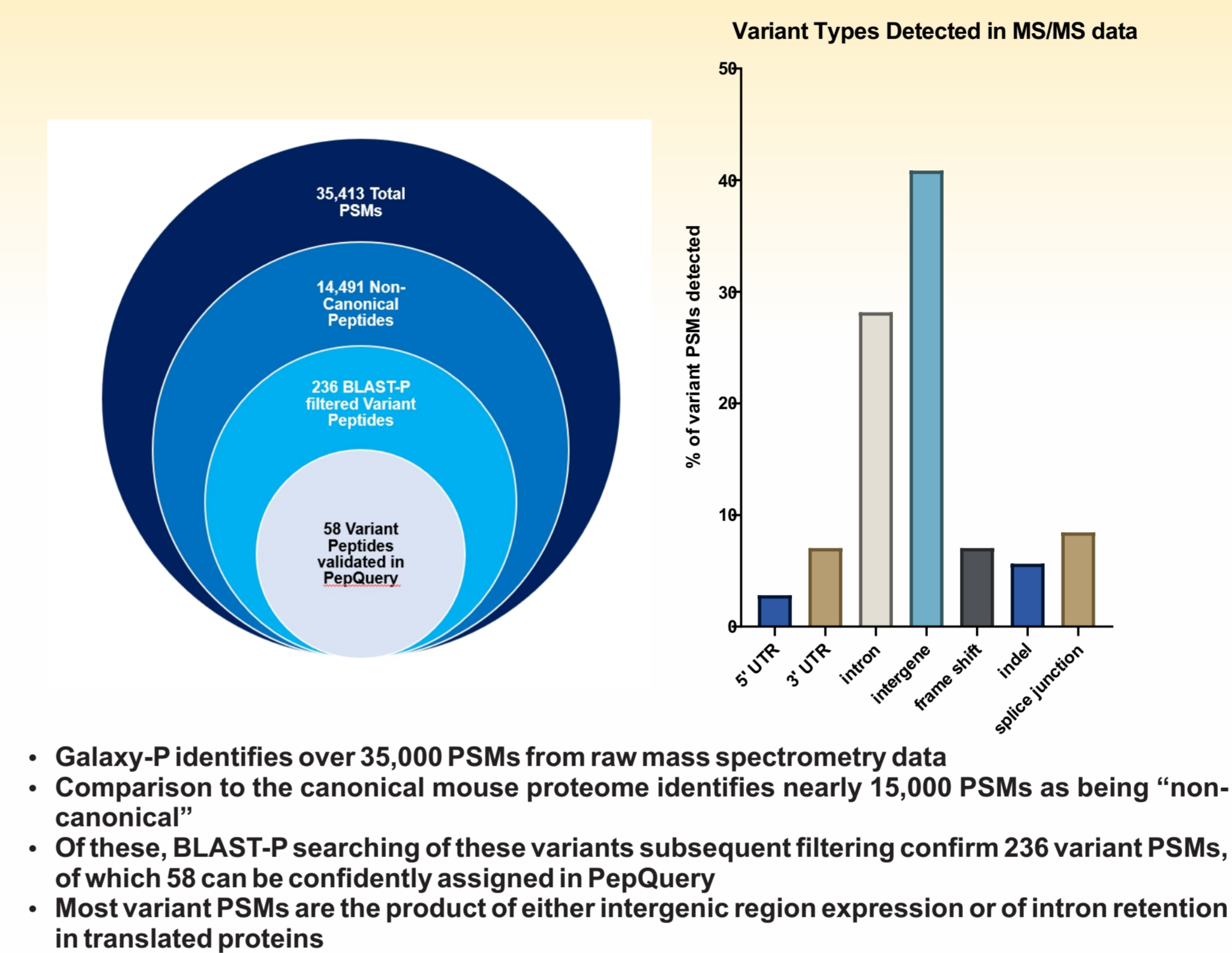
- Samples of control, inflamed (n=3 each) mice obtained from the Tannenbaum lab at MIT
- 10mg of tissue was excised from frozen proximal colon samples
- Proteins were extracted from samples, prepped for MS analysis with reduction, alkylation, and trypsinization
- Samples were labeled with TMT-6 isobaric tags and concatenated in equal amounts
- Concatenated, labeled peptides were fractionated using High pH Reverse Phase fractionation



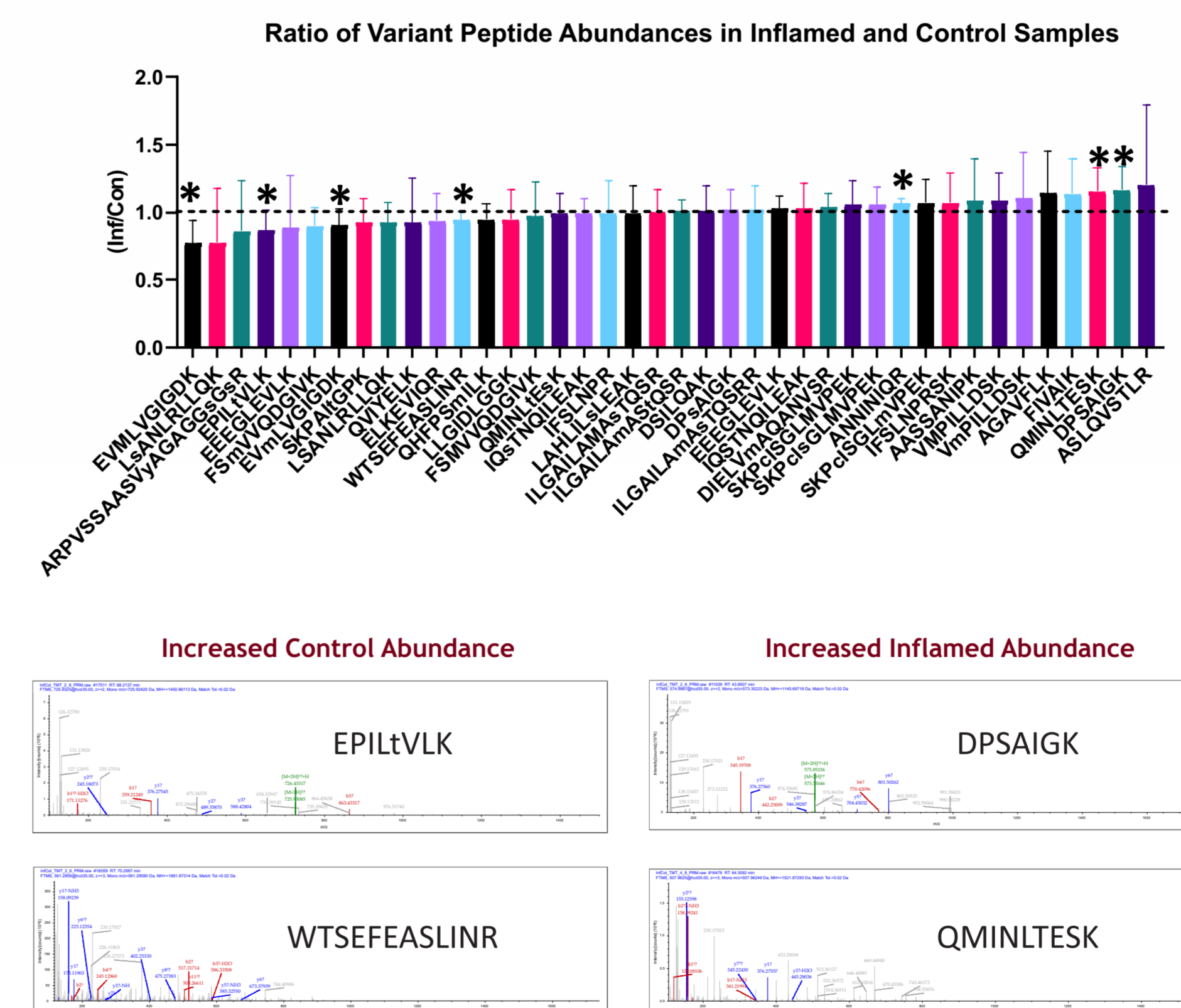
Detection and Validation of Variant PSMs



Computational Validation of Variant PSMs

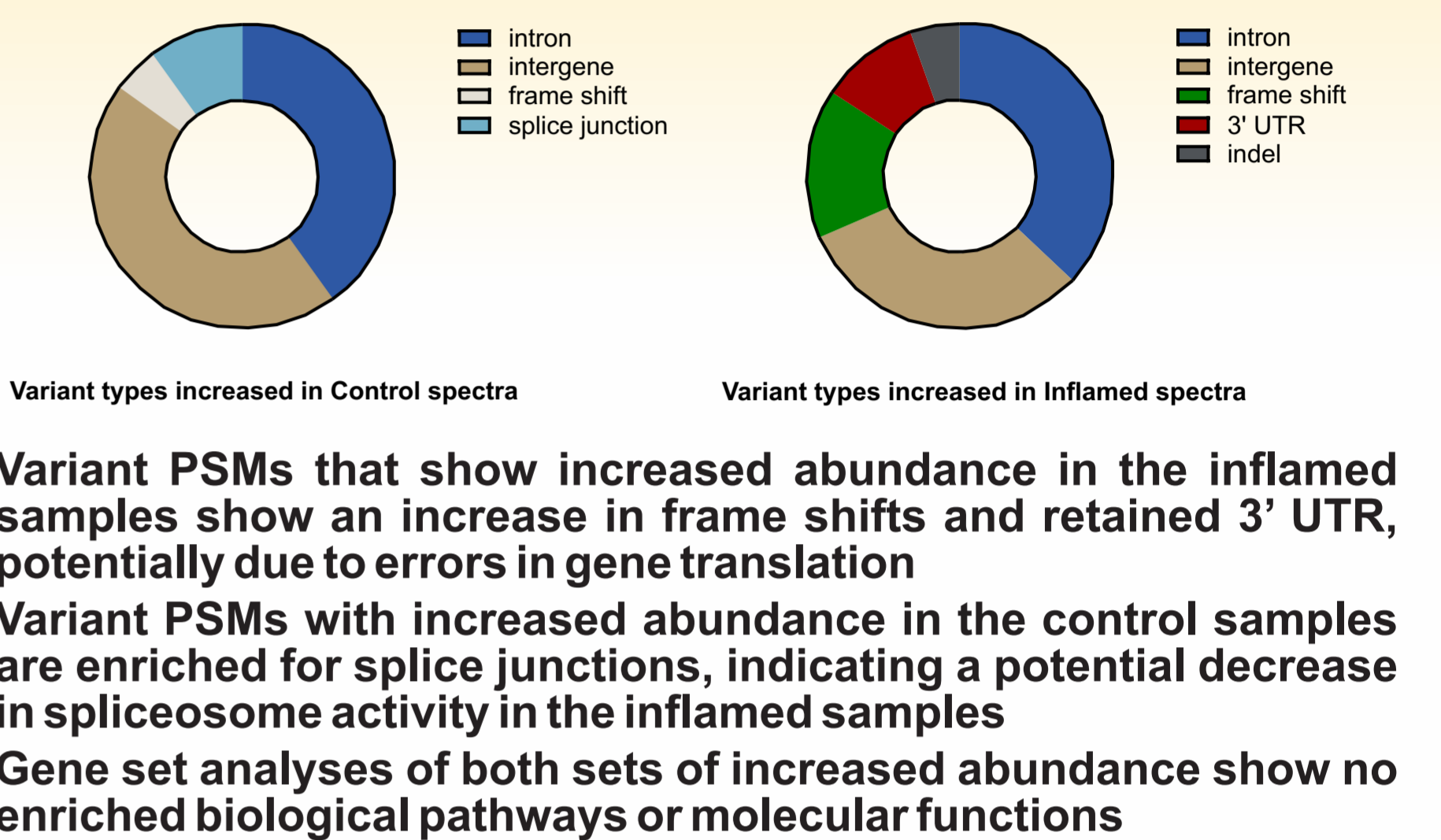


Spectral Validation of Variant PSMs



- Of the 58 variant peptides validated in PepQuery, approximately 40 can be identified directly in targeted experiments
- Most variant peptides were found to have statistically non-significant differences in amounts in both control and inflamed proximal colon samples
- Of those variant peptides identified in targeted mass spectrometry, three peptides were seen to have a significant increase in abundance in the inflamed samples, corresponding to chr4:109689395-109689416, Morc3, and Scaf8
- Four peptides were shown to have an increase in abundance in the control samples, corresponding to chr14:57578727-57578763, Ppil4, and Lamb2

Analysis of Differentially Abundant Variant PSMs



Conclusions

- Our analyses show 29 non-canonical peptide variants in the MS/MS data that can be validated using PepQuery and targeted mass spectrometry.
- Most of these variant peptides stem from transcription and translation of intergenic regions (41%) as well as retention of introns in translated mRNA (29%). Expressed intergenic regions were found in regions throughout the genome.
- While most of variant peptides are found in both control and inflamed spectra in comparable amounts, seven peptides were found to have significantly different abundances in inflamed or control samples

Future Directions

- We will search open-source proteomic datasets of chronically inflamed and cancerous tissues for the presence of the three variant peptides seen to have significantly increased abundance in this study
- We intend to conduct targeted proteomic analyses of epithelial tissues at different durations of chronic inflammation and cancer progression to ascertain the utility of these peptides as biomarkers of oncogenesis
- We are currently integrating all validation tools into a single bioinformatic workflow that automatically detects, validates, and quantifies variant peptides in MS/MS data, as well as guides the development of targeted assays for variant peptides

References

- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chen, A. P. (2012). Proteome. *PLoS one*, 7(10), e46688.
- Din, S., Wong, K., Mueller, M. F., Cristofari, A., Weenon, J., Black, G. J., ... & Aebersold, J. (2018). Clinical cancer research, 24(20), 5133-5142.
- Kisaki, J. B., Klapp, P., Abawi, H. T., Taylor, W. R., Galakosopoulos, M., Sander, T., ... & Anagnostou, A. A. (2019). *Clinical gastroenterology and hepatology*, 17(5), 914-921. E1915.
- Rutledge, S., Claudi, A. M., Cook, W. D., & Schaff, M. D. (1982). *Proceedings of the national academy of sciences*, 79(5), 1979-1983.
- Ullman, T. A., & Itzkowitz, S. H. (2011). Inflammatory inflammation and cancer. *Gastroenterology*, 140(9), 1807-1816. E1801.

Acknowledgements

This research was supported in part by the National Cancer Institute. Andrew Rajczewski was supported by a NIH Biotechnology training grant T32GM008347 from the NIH National Institute of General Medical Sciences. The Galaxy-P platform was supported through NCI grant U24CA199347. RNA-Seq sample preparation was performed by Qiyuan Han. Generation of FASTA database was performed using the Galaxy-P. Samples were obtained from the Tannenbaum lab at the Massachusetts Institute of Technology. We thank Bob Carlson for editorial assistance.



¹DEPARTMENT OF BIOCHEMISTRY, MOLECULAR BIOLOGY, AND BIOPHYSICS, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, MN 55455

²DEPARTMENT OF MEDICAL CHEMISTRY AND MASONIC CANCER CENTER, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, MN 55455