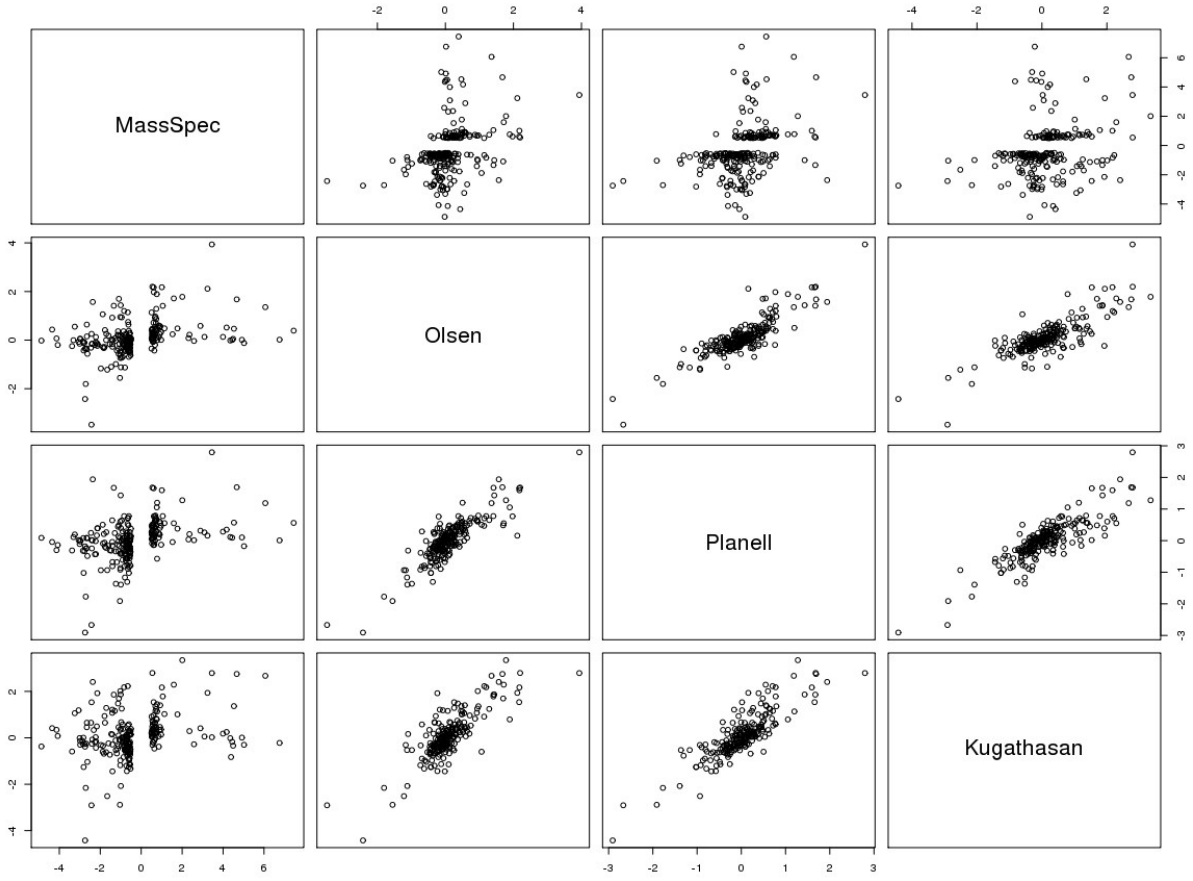
Supplementary Material

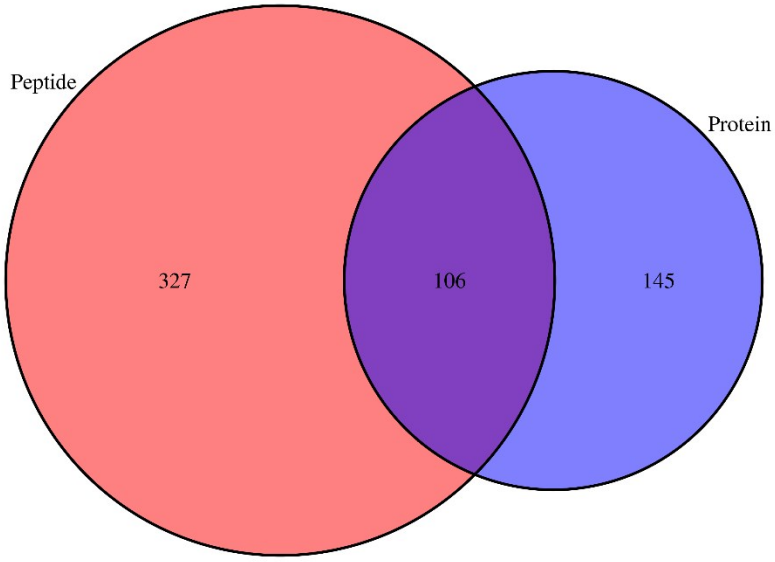
Table of Contents

Supplementary Figures

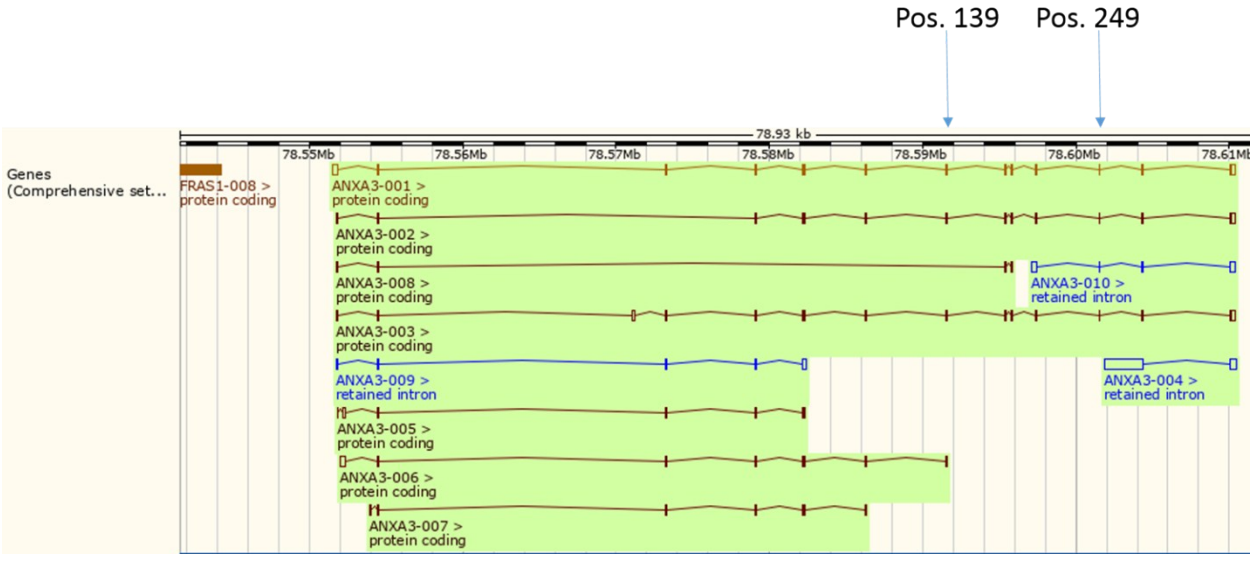
Supplementary figure 1 Scatter plots for proteomics vs transcriptomics datasets. The comparison is based on fold change.



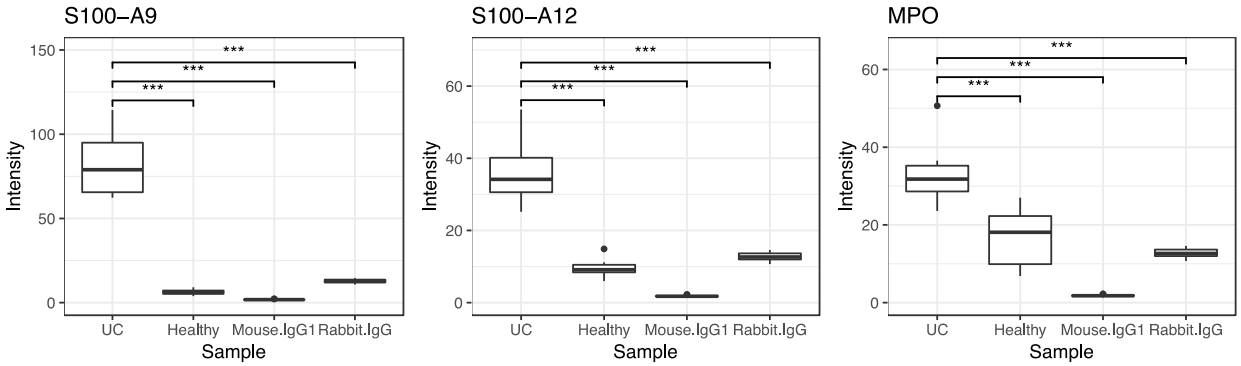
Supplementary figure 2 Overlap between proteins, determined to be differentially regulated when analysis is done at protein and peptide level. Red represents differential expression observed only at peptide level, blue- differential expression observed only at protein level and the overlap are the hits observed both at protein and petide level.



Supplementary figure 3 ANXA3 splicing structure

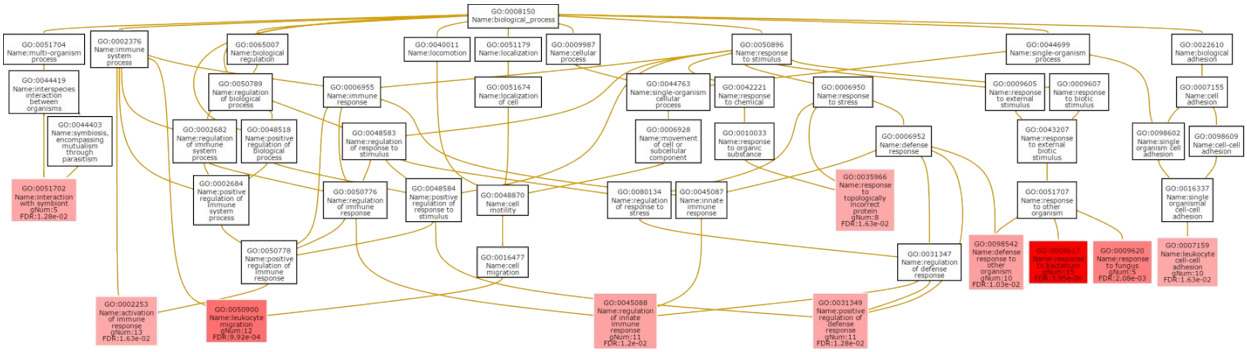


*Supplementary figure 4 Confocal microscopy Immunohistochemistry quantitation. Biopsies from ulcerative colitis patients (UC) and controls were stained with S100-A9, S100-A12, or myeloperoxidase (MPO) respectively. As a negative control, the UC-biopsy was stained with Mouse IgG1 or rabbit Ig. The staining intensity surrounding the 10 cells with the highest intensity was averaged. T-test p-value < *** 0.001, ** 0.01, *0.05.*

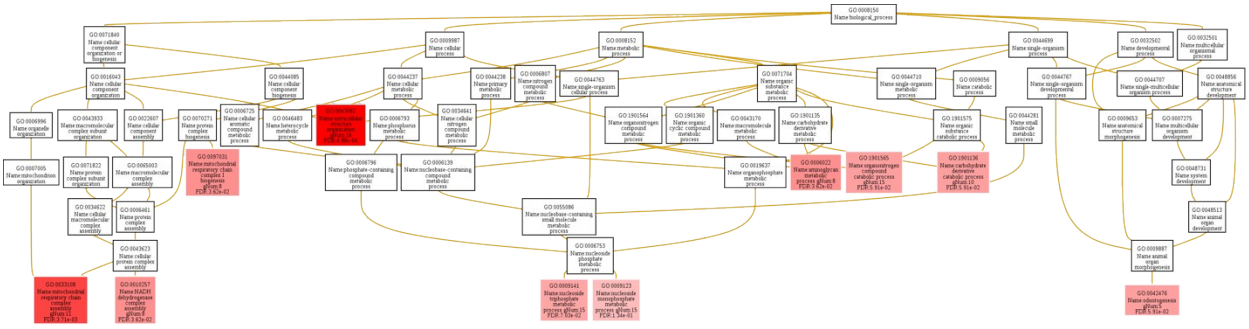


Supplementary figure 5 Webgestalt analysis of the differentially regulated proteins: biological process

A. Upregulated proteins

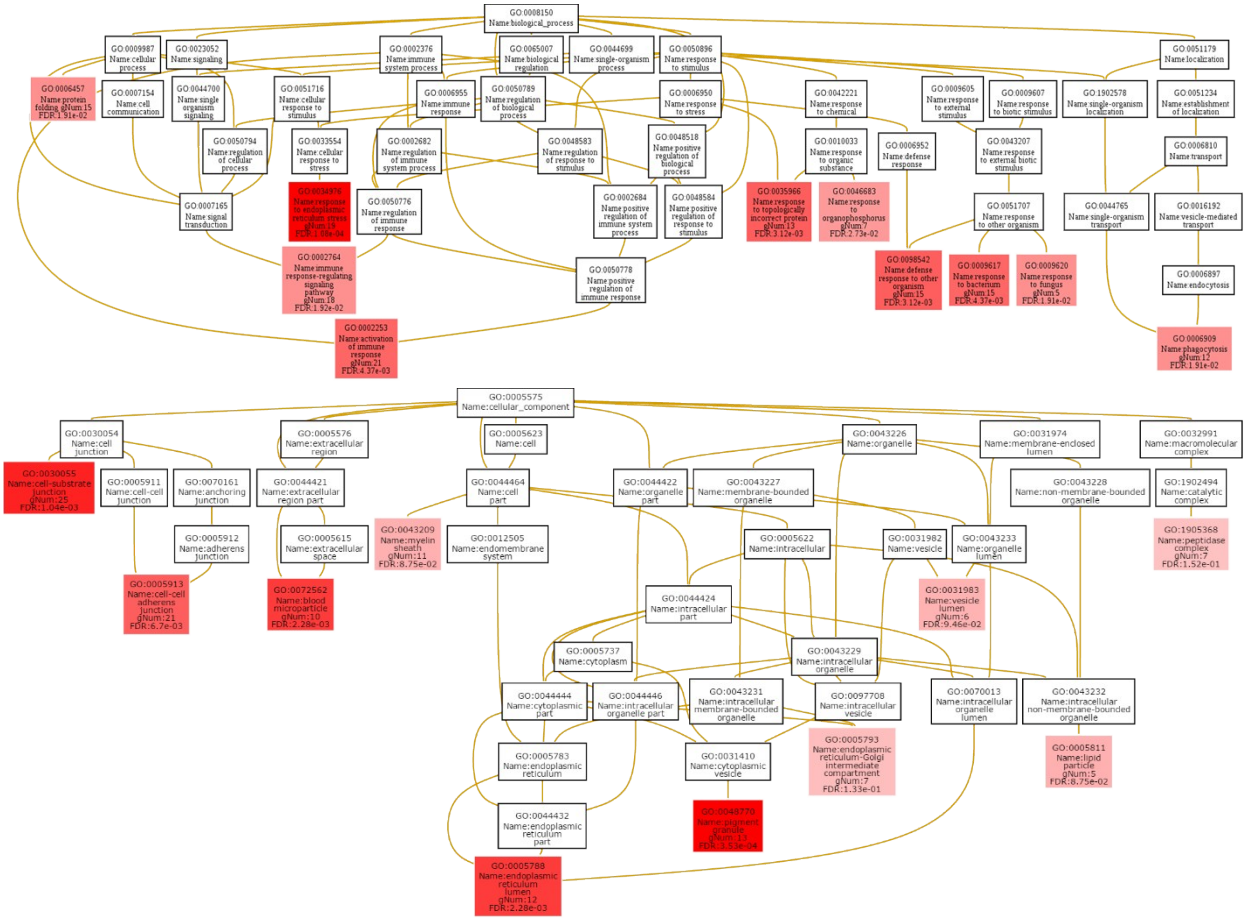


B. Downregulated proteins

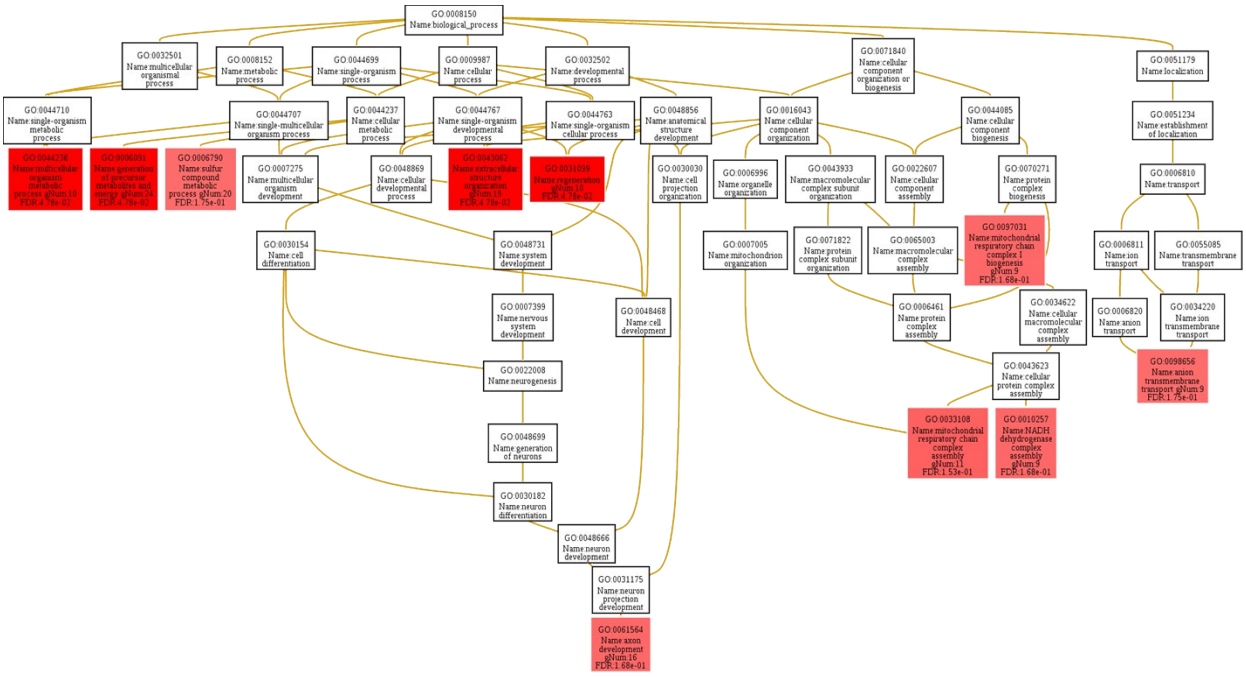


Supplementary figure 6 Over-represented GO process categories, WebGestalt

A. upregulated peptides, WebGestalt



B. Over-represented GO process categories for downregulated peptides, WebGestalt



Supplementary Tables

Supplementary table 1 Chromogranin A peptides in ulcerative colitis (UC) compared to controls. FC: fold change, FDR: false discvery rate.

Peptide	FDR	FC	Start-pos.	End-pos	Region	Overlaps antimicrobial region	Extended region/other name
CIVEVISDTLSK	1.000	2.58	35	47	Vasostatin		Vasostatin1, Vasostatin2
CIVEVISDTLSKPSPMPVSQECFETLR	0.074	3.37	35	62	Vasostatin		Vasostatin1, Vasostatin2
ELQDLALQGAK	**0.019**	**3.89**	**78**	**89**	**Chromofungin**	**Yes**(1)	chromofungin, Vasostatin1, Vasostatin2
KHSGFEDELSEVLENQSSQAELK	1.000	1.66	96	119	Vasostatin2		
HSGFEDELSEVLENQSSQAELK	1.000	3.63	97	119	Vasostatin2		
SGEATDGARPQALPEPMQESK	1.000	1.43	142	163	EA92		
AEGNNQAPGEEEEEEEEATNTHPPASLPSQK	**0.012**	**4.14**	**163**	**194**	**EA-92**		
YPGPQAEGDSEGLSQGLVDR	**<0.001**	**5.67**	**194**	**214**	**EA-92**	**Yes**(2)	
GLSAEPGWQAK	0.850	3.62	216	227	EA-92		
REEEEEEEEEAEAGEEAVPEEEGPTVVLNPHPSLGYK	1.000	2.67	227	264	ES-43		
SEALAVDGAGKPGAEEAQDPEGK	**<0.001**	**4.11**	**272**	**295**	**Pancreastatin**		
GEQEHSQQKEEEEEMAVVPQGLFR	**0.014**	**3.67**	**295**	**319**	**Pancreastatin**		
LEGQEEEEDNRDSSMK	1.000	1.09	358	374	LF-19		
EDSLEAGLPLQVR	1.000	1.89	400	413	GE-25		
RPEDQELESLSAIEAELEK	**0.004**	**4.49**	**428**	**447**	**Serpinin**	**Yes**(2)	**Serpinin, ER-37, CCA**

Supplementary table 2 Fold change and p-values of differentially regulated genes in a RNA-Seq dataset from a similar study, comparing colonic biopsies from ulcerative colitis (UC) patients to controls (3)

Gene Symbol	log2FoldChange	Adjusted p-value
COL6A3	1.10069315	2.57784E-08
COL4A1	1.475241407	3.60649E-14
BGN	2.152792291	2.07091E-20

Supplementary table 3 Over representation analysis of upregulated proteins in UC vs controls- TOP10

Term	logPV	Ontology	Term Description
MARTINELLI IMMATURE NEUTROPHIL UP	6.73	MSigDB - Curated Gene Sets	Interferons (IFNs) are cytokines that possess potent anti-viral and immunoregulatory activities. In contrast, their potential role(s) in anti-bacterial defense and neutrophil activation mechanisms is less well explored. By comparing gene expression patterns between immature and mature human neutrophils, we obtained evidence that intracellular proteases and other anti-bacterial proteins are produced at earlier stages of maturation, whereas the genes for receptors and signaling

			molecules required ...
GAL LEUKEMIC STEM CELL DN	5.12	MSigDB - Curated Gene Sets	Tumors contain a fraction of cancer stem cells that maintain the propagation of the disease. The CD34(+)CD38(-) cells, isolated from acute myeloid leukemia (AML), were shown to be enriched leukemic stem cells (LSC). We isolated the CD34(+)CD38(-) cell fraction from AML and compared their gene expression profiles to the CD34(+)CD38(+) cell fraction, using microarrays. We found 409 genes that were at least twofold over- or underexpressed between the two cell populations. These include underexpress ...
GSE6269 E COLI VS STREP PNEUMO INF PBMC DN	4.9	MSigDB - Immunologic Signatures	Each infectious agent represents a unique combination of pathogen-associated molecular patterns that interact with specific pattern-recognition receptors expressed on immune cells. Therefore, we surmised that the blood immune cells of individuals with different infections might bear discriminative transcriptional signatures. Gene expression profiles were obtained for 131 peripheral blood samples from pediatric patients with acute infections caused by influenza A virus, Gram-negative (Escherichia ...
PMID:2501794	4.55	PubMed	[1989-07] Antibiotic proteins of human polymorphonuclear leukocytes.
GSE4748 CYANOBACTERIUM LPSLIKE VS LPS AND CYANOBACTERIUM LPSLIKE STIM DC 3H DN	4	MSigDB - Immunologic Signatures	A cyanobacterial LPS antagonist prevents endotoxin shock and blocks sustained TLR4 stimulation required for cytokine expression. We report the identification and biologic characterization of an LPS-like molecule extracted from the cyanobacterium Oscillatoria Planktothrix FP1 (CyP)
KAMIKUBO MYELOID CEBPA NETWORK	3.61	MSigDB - Curated Gene Sets	Network of differentially expressed myeloid genes centered around CEBPA [GeneID=1050]
LIAN NEUTROPHIL GRANULE CONSTITUENTS	3.61	MSigDB - Curated Gene Sets	Granule constituents expressed during mouse promyelocytic cell line differentiation to neutrophils
BOYLAN MULTIPLE MYELOMA PCA1 UP	3.29	MSigDB - Curated Gene Sets	Multiple myeloma is an incurable plasma cell malignancy for which existing animal models are limited. We have previously shown that the targeted expression of the transgenes c-Myc and Bcl-X(L) in murine plasma cells produces malignancy that displays features of human myeloma, such as localization of tumor cells to the bone marrow and lytic bone lesions. We have isolated and characterized in vitro cultures and adoptive transfers of tumors from Bcl-xl/Myc transgenic mice. Tumors have a plasmablast ...
BERTUCCI MEDULLARY VS DUCTAL BREAST CANCER UP	3.16	MSigDB - Curated Gene Sets	Medullary breast cancer (MBC) is a rare but enigmatic pathologic type of breast cancer. Despite features of aggressiveness, MBC is associated with a favorable prognosis. Morphologic diagnosis remains difficult in many cases. Very little is known about the molecular alterations involved in MBC. Notably, it is not clear whether MBC and ductal breast cancer (DBC) represent molecularly distinct entities and what genes/proteins might account for their differences. Using whole-genome oligonucleotide m ...

Supplementary table 4 Over representation analysis of downregulated proteins in UC vs controls- TOP10

Term	logPV	Ontology	Term Description
NABA CORE	11.6	MSigDB -	Ensemble of genes encoding core extracellular matrix including

MATRISOME		Curated Gene Sets	ECM glycoproteins, collagens and proteoglycans
PMID:20551380	11.1	PubMed	[2010-06-15] Proteomics characterization of extracellular space components in the human aorta.
NABA PROTEOGLYCANS	6.53	MSigDB - Curated Gene Sets	One hallmark of ECM proteins is their domain-based structure. Exploiting this characteristic, we established a list of diagnostic InterPro domains commonly found in ECM proteins. This domain list was used to screen the UniProt protein database. We know that some of the domains used to select positively for ECM proteins are also found in transmembrane receptors and proteins involved in cell adhesion (growth factor receptors, integrins, etc) that do not belong to the ECM. These families of protein ...
GO:0031012	6.41	GeneOntology	extracellular matrix
NABA MATRISOME	6.31	MSigDB - Curated Gene Sets	One hallmark of ECM proteins is their domain-based structure. Exploiting this characteristic, we established a list of diagnostic InterPro domains commonly found in ECM proteins. This domain list was used to screen the UniProt protein database. We know that some of the domains used to select positively for ECM proteins are also found in transmembrane receptors and proteins involved in cell adhesion (growth factor receptors, integrins, etc) that do not belong to the ECM. These families of protein ...
SABATES COLORECTAL ADENOMA DN	6.28	MSigDB - Curated Gene Sets	Colorectal cancers are believed to arise predominantly from adenomas. Although these precancerous lesions have been subjected to extensive clinical, pathologic, and molecular analyses, little is currently known about the global gene expression changes accompanying their formation. To characterize the molecular processes underlying the transformation of normal colonic epithelium, we compared the transcriptomes of 32 prospectively collected adenomas with those of normal mucosa from the same indivi ...
PMID:22261194	5.86	PubMed	[2012-01-18] Proteomics analysis of cardiac extracellular matrix remodeling in a porcine model of ischemia/reperfusion injury.
BOQUEST STEM CELL UP	5.64	MSigDB - Curated Gene Sets	Genes up-regulated in freshly isolated CD31- [GeneID=5175] (stromal stem cells from adipose tissue) versus the CD31+ (non-stem) counterparts
SCHUETZ BREAST CANCER DUCTAL INVASIVE UP	3.89	MSigDB - Curated Gene Sets	Becoming invasive is a crucial step in breast cancer oncogenesis. At this point, a lesion carries the potential for spreading and metastasis--a process, whose molecular characteristics still remain poorly understood. In this article, we describe a matched-pair analysis of ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC) of nine breast ductal carcinomas to identify novel molecular markers characterizing the transition from DCIS to IDC. The purpose of this study was to better un ...
SMID BREAST CANCER LUMINAL A UP	3.73	MSigDB - Curated Gene Sets	We explored whether the five previously reported molecular subtypes in breast cancer show a preference for organ-specific relapse and searched for molecular pathways involved. The intrinsic gene list describing the subtypes was used to classify 344 primary breast tumors of lymph node-negative patients. Fisher exact tests were used to determine the association between a tumor subtype and a particular site of distant relapse in these patients who only received local treatment. Modulated genes and ...

Supplementary table 5 Top 10 peptides ordered by log ratio

Protein	Symbol	Peptide	pvalue.corrected	Log ratio	Gene
NP_002334	LTF	THYYAVAVVK	4.88E-08	7.55104	4057
NP_000241	MPO	IICDNTGITTVSK	1.05E-08	7.31062	4353
NP_002334	LTF	GGSFQLNELQGLK	1.83E-07	7.27501	4057
NP_002768	PRTN3	LVNVVLGAHNVR	4.02E-05	7.13575	5657
NP_002334	LTF	YLGPQYVAGITNLK	2.22E-07	7.08254	4057
NP_002334	LTF	DVTVLQNTDGNNNEAWAK	5.37E-09	7.00189	4057
NP_002334	LTF	GEADAMSLDGGYVYTAGK	9.83E-08	6.99221	4057
NP_000241	MPO	QALAQISLPR	3.72E-06	6.8142	4353
NP_002334	LTF	SQQSSDPDPNCVDRPVEGYLAVAVVR	1.51E-08	6.79599	4057
NP_002334	LTF	LRPVAAEVYGTER	9.66E-06	6.73663	4057

Peptide level analysis: Results and Discussion

Mass spectometry data inherently measures peptides using a bottom-up strategy. By combining the peptides in proteins, we gain statistical power at the expense of resolution. One clear example where peptide level analysis could be useful is in the presence of unknown proteoforms that significantly differ from the ones currently in the public databases used to search the data(4).

For this reason we decided to analyze the mass spec data that we generated at peptide level.

The overlap between differentially regulated proteins established from the protein level and peptide level analyses is lower than we expected (Supplementary figure 2). We found several explanations for this. First, unlike protein data, we applied a more stringent fold change filter of 2 as we assumed that peptides measurement would exhibit higher variability than protein measurements(5). Next, we were unable to detect ELA2, which is one of the top hits at the protein level. We established that this is due to the relatively spotty detection of each individual peptide. We found that the large number of ELA2 related peptides (11) bring a sufficient number of data points to allow accurate protein quantification. Third, in some cases, peptides from the same protein show high variability and even opposite significant change and as a result, the protein may not necessarily be differentially regulated. An example are two peptides from ANXA3, one mapping to pos. 139 and the other to position 249. According to the Ensembl splicing annotation (Supplementary figure 3), ANXA3-006 can contribute to the peptide at position 139 detection levels, but not to the one at 249. This can potentially explain the discrepancy, but we cannot exclude the contribution of post-translational modifications or protein degradation.

From the top 50 peptide hits based on fold change, 30 are NET associated protein derived peptides. NETs are composed of extracellular DNA coated with a range of granule-derived proteins with microbicidal activity(6,7). These include LTF, MPO, PRTN3, AZU1, ACTG1, S100A8 and S100A9. All top 10 peptides originate from 3 proteins (LTF, MPO and PRTN3), all of which are known parts of the NET granules.

Proteomics data is often explored and reported on the protein level. At the same, time in bottom up proteomics experiments, the readout represents peptides as being much closer(8). Another reason peptide and modification level analysis is important is our lack of complete understanding of individual proteoforms(9). While transcript isoforms can predict proteoforms to some extent, they do not truly represent the complexity of the proteins that can be generated from the same locus.

Protein level analysis has its advantages as well. Combining multiple observed peptides as a single protein eliminates missing datapoints and reduces technical variability. We found that using both peptide and protein level data eliminates some of the false negatives, which is valuable in a hypothesis generation type of experiment.

Overall, we found that there is good overlap between peptide and protein based Gene Set Enrichment Analysis (GSEA). One term that our peptide level analysis adds is blood microparticles. MPs are elevated in patients with active IBD(10,11). Other studies found elevation of MPs in IBD patients vs controls, but no significant correlation with disease activity(12). Microparticles may in fact be causing the increase in neutrophil traps (13). There is evidence that circulating MPs and annexin (+) platelet-derived MPs (PDMPs) are increased in IBD (14).

References

1. Two Chromogranin A-Derived Peptides Induce Calcium Entry in Human Neutrophils by Calmodulin-Regulated Calcium Independent Phospholipase A2 [Internet]. [cited 2017 Apr 17]. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004501

2. Shooshtarizadeh P, Zhang D, Chich J-F, Gasnier C, Schneider F, Haïkel Y, et al. The antimicrobial peptides derived from chromogranin/secretogranin family, new actors of innate immunity. Regul Pept. 2010 Nov 30;165(1):102–10.

3. Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulssen RH. Transcriptomic Landscape of Treatment—Naïve Ulcerative Colitis. J Crohns Colitis. 2018 Feb 28;12(3):327–36.

4. Zhang Q, Fillmore TL, Schepmoes AA, Clauss TRW, Gritsenko MA, Mueller PW, et al. Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. J Exp Med. 2013 Jan 14;210(1):191–203.

5. Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. - PubMed - NCBI [Internet]. [cited 2018 Apr 23]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26423119

6. Neutrophil extracellular traps (NETs) - formation and implications. - PubMed - NCBI [Internet]. [cited 2017 Sep 12]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23819131

7. NETosis: At the crossroads of rheumatoid arthritis, lupus, and vasculitis - ScienceDirect [Internet]. [cited 2017 Sep 12]. Available from: http://www.sciencedirect.com/science/article/pii/S1297319X16301063

8. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. Nat Methods. 2007 Oct;4(10):817–21.

9. Smith LM, Kelleher NL, Proteomics TC for TD. Proteoform: a single term describing protein complexity. Nat Methods. 2013 Mar;10(3):186–7.

10. Elevated Circulating Platelet-Derived Microparticles in Patients with Active Inflammatory Bowel Disease - ProQuest [Internet]. [cited 2017 Aug 4]. Available from: https://search.proquest.com/openview/011b91699c69bbde9f40230988bd3d20/1?pq-origsite=gscholar&cbl=2041977

11. Voudoukis E, Vetsika E-K, Giannakopoulou K, Karmiris K, Theodoropoulou A, Sfiridaki A, et al. Distinct features of circulating microparticles and their relationship with disease activity in inflammatory bowel disease. Ann Gastroenterol Q Publ Hell Soc Gastroenterol. 2016;29(2):180–7.

12. Tziatzios G, Polymeros D, Spathis A, Triantafyllou M, Gkolfakis P, Karakitsos P, et al. Increased levels of circulating platelet derived microparticles in Crohn's disease patients. Scand J Gastroenterol. 2016 Oct 2;51(10):1184–92.

13. He Z, Si Y, Jiang T, Ma R, Zhang Y, Cao M, et al. Phosphotidylserine exposure and neutrophil extracellular traps enhance procoagulant activity in patients with inflammatory bowel disease. Thromb Haemost. 2016;115(4):738–51.

14. Andoh A, Yoshida T, Yagi Y, Bamba S, Hata K, Tsujikawa T, et al. Increased aggregation response of platelets in patients with inflammatory bowel disease. J Gastroenterol. 2006 Jan 1;41(1):47–54.