Electronic Supplementary Material (ESI) for ChemComm.

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL DATA

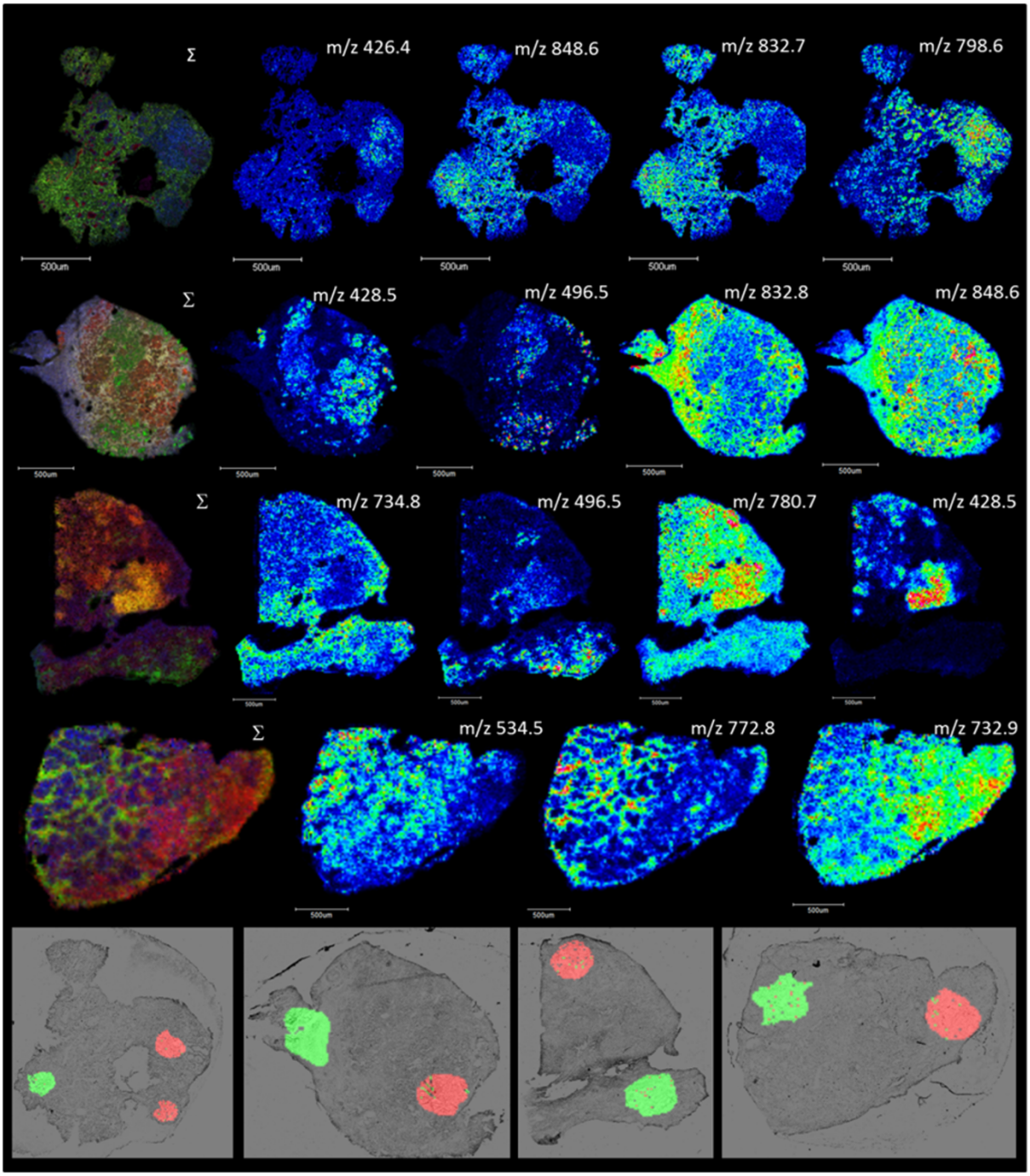

Figure S1. Sample MS images of lipids and resultant clustering images of prostate tissue sections. The image on the left of each panel is the composite image of the selected peaks. The bottom panel shows the resultant clustering images, with the green and red clusters corresponding to benign and tumor areas, respectively.

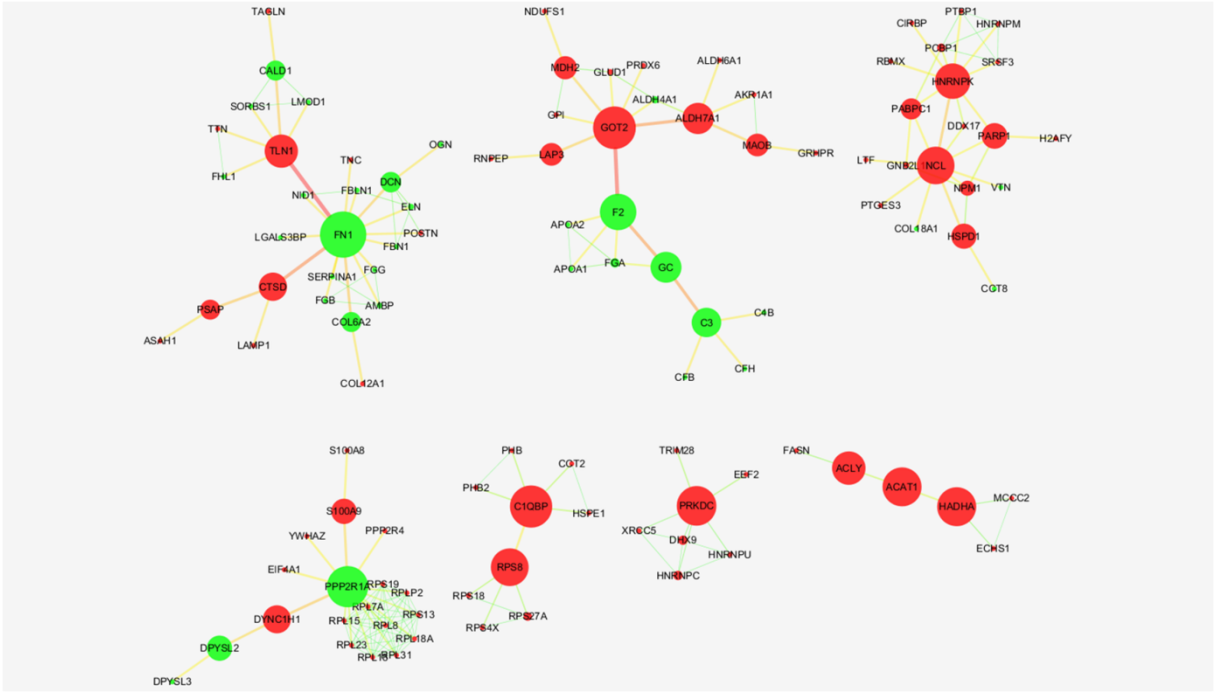

Figure S2. Top MCL clusters obtained from the PPI network, showing the hubs where most nodes intersect. Betweenness centrality and normalized log_2Fold change were used to plot the node size and color, respectively. Green and red nodes indicate under- and overregulated proteins, respectively.

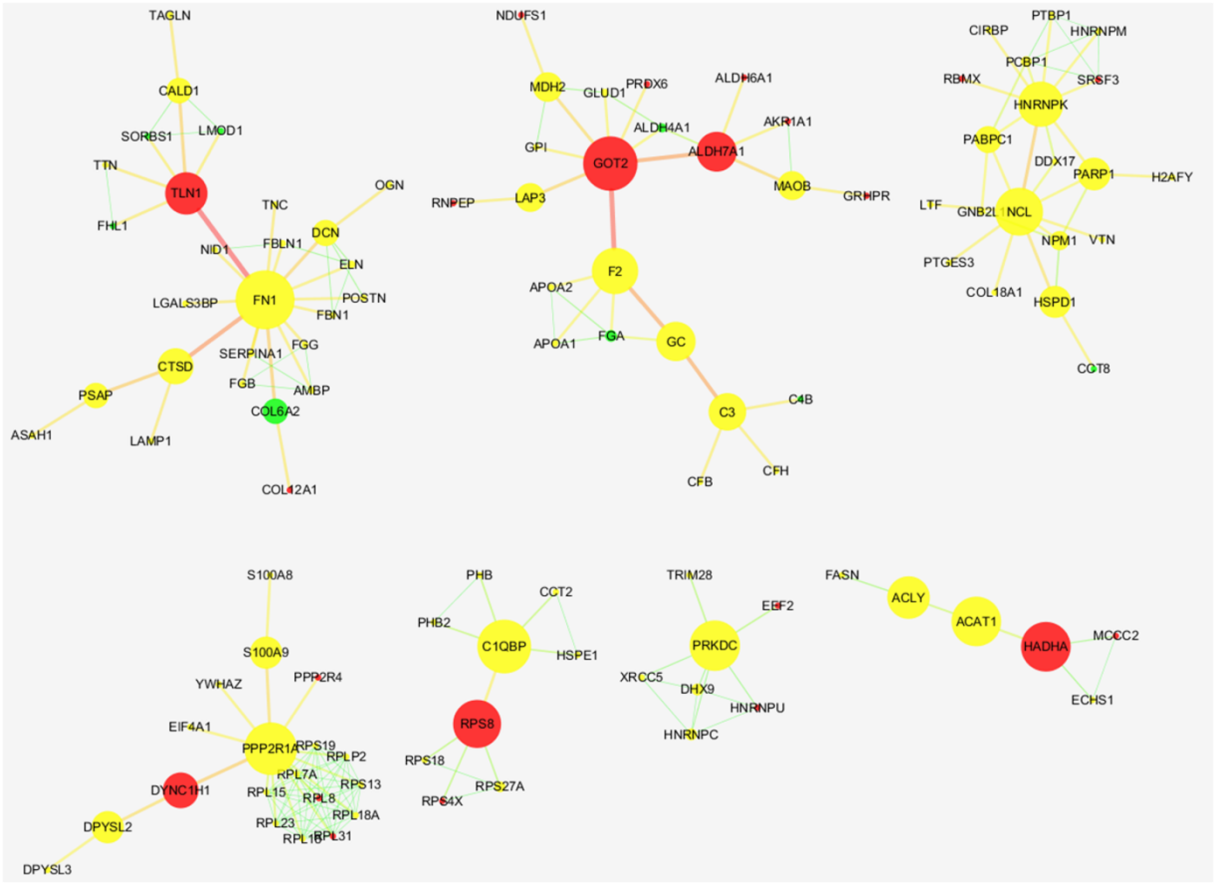

Figure S3. Mapping of Cancer Gene Index-annotated proteins using the term "neoplasm" (nodes in yellow) in the modules. GOT2 is the only hub that does not have CGI annotations.

Table S1. Over- and under-expressed proteins and their $\log_2$ fold change values (Log2 Fold Change), Related to Figure 3. Negative values denote under-expression.

Protein Name	Log2 Fold	Protein Name	Log2 Fold	Protein Name	Log2 Fold	Protein Name	Log2 Fold
GOLM1	4	RPSA	1	ECH1	1	COL6A2	-1
PRKDC	4	FASN	1	PTBP1	1	C3	-1
POSTN	4	PGLS	1	ANXA3	1	COL18A1	-1
RPL23	4	LAMP1	1	RPS19	1	GNAI2	-1
GDF15	4	FSTL1	1	RPS27A	1	CP	-1
HYOU1	4	SUB1	1	MDH2	1	ACPP	-1
CTNNA1	4	SRSF3	1	RPS4X	1	OGN	-1
PSME2	3	HSPD1	1	CKAP4	1	CRYZ	-1
PARP1	3	DBNL	1	RPLP2	1	VTN	-1
HADHA	3	TTN	1	RPL15	1	FGG	-1
DAK	3	NPM1	1	PSAP	1	LMOD1	-1
TRIM28	3	COPG	1	MT1G	1	PALLD	-1
ST13	2	UBA1	1	RPL18	1	HMGB1	-1
RPL31	2	CCT2	1	P4HB	1	DPYSL2	-1
RPN1	2	S100A9	1	LTF	1	FGA	-1
LAMP2	2	PPP2CA	1	IQGAP2	1	APOH	-1
DECR1	2	RPL18A	1	DHRS7	1	CORO1A	-1
LAP3	2	SET	1	PRDX3	1	LAMB2	-1
DYNC1H1	2	VCP	1	PHB	1	EMILIN1	-1
ANXA6	2	COPA	1	DDAH2	1	SERPINA1	-1
GAA	2	HSPE1	1	MAOB	1	PTRF	-1
CIRBP	2	FAM129A	1	LMNA	1	GDI1	-1
ATIC	2	TXN	1	H3F3B	1	LAMA4	-1
GRHPR	2	AHNAK	1	LCP1	1	HSPB1	-1
C1QBP	2	TPI1	1	CDC42	1	CFH	-1
FBP1	2	FUCA1	1	RPS27A	1	NID1	-1
ETFB	2	HSD17B4	1	ALDH6A1	1	ELN	-1
puf	2	CANX	1	HNRNPM	1	HSPB6	-1
PPP2R4	2	GPI	1	COPB2	1	EHD2	-1
RPL13	2	RPL7A	1	RPS13	1	LASP1	-1
UQCRC1	2	NCL	1	ECHS1	1	SERPINC1	-1
MCCC2	2	RBMX	1	CALR	1	C4B	-1
ASAH1	2	RPL14	1	TLN1	1	PTGR1	-1
ANP32A	2	RNPEP	1	RAN	1	CLU	-1
PRDX5	2	PRKCSH	1	PDIA6	1	AMBP	-1
RPL8	2	XRCC5	1	ATP1A1	1	HRG	-1
C4A	2	ACAT1	1	YWHAZ	1	CAV1	-1
NUMA1	2	PARK7	1	PCBP1	1	A1BG	-1
TKT	2	LPP	1	EIF4A1	1	CCT8	-1
APEX1	2	CPE	1	HSPA8	1	EPHX2	-1
NACA	2	COL12A1	1	RDH11	1	FHL1	-1
SEPT9	2	TAGLN	1	GLUD1	1	FBLN1	-1
DHX9	2	GOT2	1	ANXA1	-1	CFB	-1
TGFBI	2	RPS12	1	SYNPO2	-1	ARMC5	-1
MARCKS	2	DDX39B	1	SORBS1	-1	COL14A1	-1
NANS	1	RPS8	1	CSRP1	-1	SYNE1	-1
TPP1	1	HMGN2	1	MYO1C	-1	VWA1	-1
PRDX6	1	HNRNPC	1	APOA1	-1	SERPINA3	-1
IQGAP1	1	VCL	1	DPYSL3	-1	LAMA5	-1
EEF2	1	ERP29	1	CALD1	-1	DSP	-1
HEXB	1	HNRNPU	1	TNS1	-1	PLG	-2
CTSD	1	LTA4H	1	IGL@	-1	KNG1	-2
PDIA4	1	H2AFY	1	PHGDH	-1	FN1	-2
PTMS	1	PABPC1	1	GC	-1	DCN	-2
SCARB2	1	TNC	1	LGALS3BP	-1	EFEMP1	-2
S100A8	1	RPL6	1	APOA2	-1	CAPG	-2
NDUFS1	1	EEF1G	1	SYNM	-1	FBLN5	-2
ACLY	1	UGDH	1	RSU1	-1	KANK2	-2
PDLIM5	1	CMPK1	1	TPSB2	-1	COL3A1	-2
COPB1	1	CORO1B	1	FERMT2	-1	TINAGL1	-2
RPL23A	1	TXNDC17	1	HSPG2	-1	F2	-2
XRCC6	1	GNB2L1	1	PDLIM7	-1	ADIRF	-2
MSN	1	HSP90AA1	1	PPP2R1A	-1	GSTM3	-3
RPS18	1	APRT	1	FBN1	-1	OLFML1	-3
DDX17	1	PHB2	1	GSTP1	-1	A2M	-3
HNRNPK	1	ALDH7A1	1	ALDH4A1	-1	NID2	-3
UQCRC2	1	RPL12	1	FGB	-1	MAP1B	-5
PTGES3	1	AKR1A1	1	LAMC1	-1	MSMB	-5

Table S2. Top 8 groupings observed after GO pathway analysis using ClueGO, Related to Figure 3.

Function	Groups	Group Genes
Formation of Fibrin Clot (Clotting Cascade)	Group0	A2M, APOH, C1QBP, CAV1, F2, FGA, FGB, FGG, HRG, KNG1, PLG, SERPINC1,VTN
Cytoplasmic Ribosomal Proteins	Group1	APEX1, CALR, CANX, CAPG, CCT2, CCT8, EEF1G, EEF2, EIF4A1, GDI1, GNB2L1, HSP90AA1, LAMA5, MYO1C, PABPC1, PPP2CA, PPP2R1A, PRDX3, RAN, RPL12, RPL13, RPL14, RPL15, RPL18, RPL18A, RPL23, RPL23A, RPL31, RPL6, RPL7A, RPL8, RPLP2, RPN1, RPS12, RPS13, RPS18, RPS19, RPS27A, RPS4X, RPS8, RPSA, SEPT9, SORBS1, YWHAZ
Complement and coagulation cascades	Group2	A2M, ANXA1, APOA1, APOA2, APOH, C1QBP, C3, C4A, C4B, CALR, CAV1, CFB, CFH, CLU, COL18A1, F2, FBN1, FGA, FGB, FGG, FN1, GNAI2, GSTP1, HNRNPC, HRG, HSP90AA1, HSPG2, HYOU1, KNG1, LAMP2, LTF, PLG, PSAP, SERPINA1, SERPINC1, TLN1, TNC, VCL, VTN
Laminin interactions	Group3	COL18A1, COL3A1, COL6A2, DCN, FN1, HSPG2, LAMA4, LAMA5, LAMB2, LAMC1, NID1, NID2, TNC, VTN
regulation of nuclease activity	Group4	CALR, HYOU1, LMNA, NPM1, PDIA6, TLN1, TPP1, VCP
cytoplasmic membrane-bounded vesicle lumen	Group5	A2M, ANXA1, APOA1, APOA2, APOH, C1QBP, C3, C4A, C4B, CALR, CAV1, CFB, CFH, CLU, COL18A1, F2, FBN1, FGA, FGB, FGG, FN1, GNAI2, GSTP1, HNRNPC, HRG, HSP90AA1, HSPG2, HYOU1, KNG1, LAMP2, LTF, PLG, PSAP, SERPINA1, SERPINC1, TLN1, TNC, VCL, VTN
Valine, leucine and isoleucine degradation	Group6	ACAT1, ALDH6A1, ALDH7A1, DECR1, ECHS1, HADHA, MAOB, MCCC2, TPI1
Glucose metabolism	Group7	AKR1A1, ALDH7A1, FBP1, GOT2, GPI, MDH2, PPP2CA, PPP2R1A, RPS27A, TPI1

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Reagents

HPLC grade methanol (MeOH), acetonitrile (ACN), water and AR grade trifluoroacetic acid (TFA) were obtained from Biosolve B. V. (Valkensvaard, Netherlands). Ammonium bicarbonate (NH_4HCO_3), ethanol (EtOH), α-cyano-4-hydroxycinnamic acid (HCCA), 2,5-dihydrobenzoic acid (2,5-DHB), DL-dithiothreitol (DTT), iodoacetamide (IAA), cetrimonium bromide (CTAB), benzalkonium chloride (BAC), octyl β-D-glucopyranoside (OGP) and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were obtained from Sigma (Saint-Quentin Fallavier, France). Tetramethylethylenediamine (TEMED), ammonium persulfate (APS), sodium dodecyl sulfate (SDS) and Triton-X 100 were purchased from Bio-Rad (Marnes La Cocquette, France). Sequencing grade, modified porcine trypsin was obtained from Promega (Charbonnieres, France).

Tissue Specimens

Fresh frozen tumors embedded on optimal cutting temperature (OCT) polymer were retrieved from archived specimens of a previous study [1]. These specimens have been obtained with informed consent from patients in the Centre Hospitalier of University of Sherbooke and the tumor and benign regions have been identified by an experienced pathologist. These were confirmed, upon receipt of the samples, by dissecting 10-µm sections using a cryomicrotome (Leica Microsystems, Nantere, France), subjecting them to HES staining, and verifying the actual positions of the regions.

MALDI MS Imaging

The consecutive 10-µm section adjacent to the one subjected to hematoxylin erythrosine saffron (HES) staining was used for MALDI MSI analysis of lipids using methods previously described [2]. The section was mounted on an ITO-coated conductive slide and dried under vacuum for 5 min. 2,5-Dihydroxybenzoic acid (DHB) was used as matrix and prepared at a concentration of 20 µg/mL in 70 :30 ethanol/0.1 % TFA. The matrix solution was sprayed uniformly throughout the entire tissue section using a discarded electrospray nebulizer. The nebulizer is attached to a 500-µL syringe that is driven by a pump at a flow rate of 2.5 µL/min. The fine mist generated by the spray head ensures the deposition of micrometer-sized matrix droplets onto the tissue surface. Spraying was performed for 10 min and the formation of matrix crystals was confirmed by examination under a light microscope.

The prepared sections were examined using an UltraFlex II instrument equipped with a Smart beam (Nd:YAG, 355nm) laser having a repetition rate up to 200 Hz (BrukerDaltonics, Bremen, Germany). The images were acquired in positive reflector mode at 100-µm resolution at a mass range of m/z 200-1500, and the obtained spectra were averaged from 500 laser shots per pixel. Images obtained were analyzed using Fleximaging 2.1 (BrukerDaltonics, Bremen, Germany) to map the entre section. Regions of interest (ROIs) were defined on the benign and tumor regions basing from the HES optical images, and the spectra from these regions were exported to ClinPro Tools version 3.0 (Bruker Daltonics, Bremen, Germany). A limit of less than 1000 spectra was exported per region. These representative spectra were used for the hierarchical clustering analysis using the following parameters. The spectra were initially prepared by setting the resolution to 800 and the m/z range from 200 to 900. Savitsky-Golay smoothing was employed using 2.0 m/z spectral width and 5 cycles, and the signal-to-noise (S/N) ratio of the total average spectrum was set 3.0. Unsupervised clustering

was selected with Correlation and Average as distance and linkage methods, respectively. The results were then exported back to FlexImaging and the spectra of the branches of the dendrogram overlapped with the optical image of the tissue section. This method verifies the visually-defined benign and tumor regions as well as made them more precise.

Protein Extraction and LC-MS

The consecutive section adjacent to the one subjected to MALDI MS imaging was used for the proteomics experiments. These sections were mounted on parafilm-covered glass slides, dried under vacuum, and subjected to the PAM procedure, trypsin digestion and desalting previously described [2]. The extracts were dried and resuspended in 5% ACN/0.1% FA prior to injecting to the LC-MS instrument.

Separation of the sample components was done using an online reversed-phase chromatographic system (Thermo Scientific Proxeon Easy-nLC II) equipped with a Proxeon trap column (100 μm ID x 2 cm, Thermo Scientific) and C18 packed tip column (100 μm ID x 15 cm, NikkyoTechnos Co. Ltd). Elution was carried out using an increasing gradient of AcN (5% to 40% over 110 min) and a flow rate of 300 nL/min. A voltage of 1.7 kV was applied via the liquid junction of the nanospray source. This was interfaced to a Thermo Scientific Orbitrap Elite mass spectrometer programmed to acquire in data-dependent mode. The survey scans were acquired in the Orbitrap mass analyzer operated at 120,000 (FWHM) resolving power. A mass range of 400 to 2000 m/z and a target of 1E6 ions were used for the survey scans. Precursor ions observed with an intensity over 500 counts were selected “on the fly” for ion trap collision-induced dissociation (CID) fragmentation with an isolation window of 2 u and a normalized collision energy of 35%. A target of 5000 ions and a maximum injection

time of 200 ms were used for MS2 spectra. The method was set to analyze the top 20 most intense ions from the survey scan and a dynamic exclusion was enabled for 20 s.

Samples analyzed using the UPLC Q-Exactive system, on the other hand, were separated using a 10-cm EASY-column (75 μm ID x 10 cm, Thermo Scientific) and the peptides were eluted following a 2-h gradient of of AcN starting from 5% to 30% over 120 min at a flow rate of 250 nL/min. The Q-Exactive instrument was set to acquire top 10 MS^2 in data-dependent mode. The survey scans were taken at 70,000 FWHM (at m/z 400) resolving power in positive mode and using a target of 3E6 and default charge state of 2. Unassigned and +1 charge states were rejected, and dynamic exclusion was enabled for 20s. The scan range was set to 300-1600 m/z. For the MS^2, 1 microscan was obtained at 17,500 FWHM and isolation window of 4.0 m/z, using a scan range between 200-2000 m/z.

Protein Identification and Bioinformatic Analysis

The MS/MS spectra were analyzed with Sequest using Proteome Discoverer Software (version 1.4.1.14, Thermo Fisher Scientific, San Jose, CA, USA). Spectra were searched against the Uniprot database (version November 2013) filtered with *Homo sapiens* (122786 sequences) taxonomy using trypsin as digestion enzyme (one missed cleavage). Sequest was searched with a fragment ion tolerance of 0.100 Da and a parent ion tolerance of 10.0 ppm. Carbamidomethylation of cysteine and oxidation of methionine were set as fixed and variable modifications, respectively.

Scaffold (version 4.0.5, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at a given probability to achieve an FDR less than 5.0 % by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be

established at a given probability to achieve an FDR less than 2.0 % and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [3]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

The protein IDs for each sample were loaded into Perspectives version 1.0.3 (Proteome Software Inc., Portland, Oregon) using the same FDR thresholds used in Scaffold and the decoys were removed from the final list. Clustering was performed to group together proteins with any shared evidence and for which the peptide IDs cannot be discerned. Doing so produced 1221 and 1227 protein groups for data generated by the Orbitrap Elite and Q-Exactive instruments, respectively. The total weighted spectral counts of the protein groups were then obtained and subjected to Fisher's Exact Test. Only proteins with $p<0.05$ were considered to have significant differential expression between the benign and tumor group datasets. The protein list was then exported in Excel and the $\log_2$ fold change was computed from the weighted spectral counts (tumor/benign). The protein list was further filtered by removing IDs with $\log_2$ fold change = 0 and selecting IDs which were detected in at least 60% of the benign or tumor samples (for samples with replicates, the spectral count was averaged). For the dataset obtained using the Orbitrap Elite instrument, this corresponds to detection in at least 3 of 4 benign or tumor samples, while for the dataset obtained using the Q-Exactive, this corresponds to at least 6 out of 9 benign or tumor samples.

Protein-Protein Interaction (PPI) Network

The differentially expressed protein list identified from the two datasets were combined and used as seed terms for the construction of the PPI network using STRING (Search Tool for the Retrieval of Interacting Genes, version 9.1). STRING is a database that generates and

scores both physical and functional PPIs from various sources based on their neighborhood, gene fusions, co-occurrence, co-expression, experiments and literature mining. The extended network was constructed from the list using a confidence score = 0.7, corresponding to high level of confidence.

Network Construction and Analysis

The PPI network was visualized using Cytoscape network visualization software version 3.1.1 [4] and analyzed using the Network Analyzer plug-in [5]. The topology of the network was visualized by mapping the Degree and Betweenness centrality parameters, corresponding to node size and node color, respectively. Degree pertains to the number of edges that links a given node, so that nodes possessing a high Degree value may represent hub genes. Betweenness centrality, on the other hand, reflects the importance of a node based on the number of shortest path lengths that passes this node. Meanwhile, the thickness of the edges was depicted using the edge betweenness values.

The constructed network was then subjected to Markov Clustering (MCL) algorithm using the ClusterMaker plug-in to determine the major clusters. The granularity parameter was set to default value (2.5) and the EdgeBetweenness parameter was used as the source for array data. With this parameter, the edge cut-off was set to 73.042. Major clusters identified in this manner were visualized using Betweenness centrality for node size and expression value for node color. For the expression values, a binary scale was used to represent the $\log_2$ fold change values: 0 for $\log_2$ fold change less than 0 (underexpressed), and 1 for those greater than 0 (overexpressed).

Gene Ontology Analysis

The differentially expressed proteins identified from the two datasets were uploaded to the ClueGO application (version 2.1.2, [6]) to check for enriched pathways in the benign and tumor regions. Functionally-grouped annotated networks were generated using the following settings. The organism was set to *Homo sapiens*. The gene ontology (GO, [7]) terms were accessed from the following ontologies/pathways : GO_Biological Process, GO_Molecular Function, GO_Cellular Component and GO_Immune System Process (ontology updated 5/23/2014), Kyoto Encyclopedia of Genes and Genomes (KEGG, updated 5/24/2014), Reactome Pathway database (updated 5/24/2014), and WikiPathways (updated 5/24/2014). All types of evidence were used (see http://www.geneontology.org/GO/evidence.shtml). The Use GO Term Fusion option was selected, and only pathways with $pV \leq 0.05$ were accepted. The significance of each term or group was calculated using the right-sided hypergeometric test corrected using Bonferroni step-down correction. The kappa score was set to 0.5 and for network specificity, the GO tree levels were restricted at 6-13 (medium-detailed specificity) and for each cluster a minimum of 5 genes and 7% of the gene population was set. GO terms were grouped with an initial group size of 3 and 50% for group merge. The remaining parameters were set to default.

The gene identifiers of the major clusters identified by MCL clustering during network analysis were also examined using the Cancer Gene Index feature available in the Reactome FI plug-in [8]. The tree of National Cancer Institute (NCI) disease terms was loaded and the term “neoplasm” was selected to highlight gene identifiers that have been annotated for this term and its sub-terms. To map selected over-represented pathways to Reactome events, The Pathway Analysis tool was used.

SUPPLEMENTAL REFERENCES

1. F. D'Anjou, S. Routhier, J. P. Perreault, A. Latil, D. Bonnel, I. Fournier, M. Salzet and R. Day, *Transl Oncol*, 2011, **4**, 157-172.
2. J. Franck, J. Quanico, M. Wisztorski, R. Day, M. Salzet and I. Fournier, *Anal Chem*, 2013, **85**, 8127-8134.
3. A. I. Nesvizhskii, A. Keller, E. Kolker and R. Aebersold, *Anal Chem*, 2003, **75**, 4646-4658.
4. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res*, 2003, **13**, 2498-2504.
5. Y. Assenov, F. Ramirez, S. E. Schelhorn, T. Lengauer and M. Albrecht, *Bioinformatics*, 2008, **24**, 282-284.
6. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W. H. Fridman, F. Pages, Z. Trajanoski and J. Galon, *Bioinformatics*, 2009, **25**, 1091-1093.
7. Consortium, *Nucleic Acids Res*, 2010, **38**, D331-335.
8. M. Milacic, R. Haw, K. Rothfels, G. Wu, D. Croft, H. Hermjakob, P. D'Eustachio and L. Stein, *Cancers (Basel)*, 2012, **4**, 1180-1211.