

Integrated proteogenomic analysis of laser capture microdissected breast tumors

Viswanadham Sridhara¹, Tao Liu², Marina A. Gritsenko², Lori A. Sturtz¹, Albert J. Kovatich³, Vladislav A. Petyuk², Brenda Deyarmin¹, Jason E. McDermott², Anil K. Shukla², Ronald J. Moore², Matthew E. Monroe², Bobbie-Jo M. Webb-Robertson², Jeffrey A. Hooke³, Leigh Fantacone-Campbell³, Praveen Kumar Raj Kumar¹, Leonid Kvecher¹, Jianfang Liu¹, Jennifer Kane¹, Jennifer Melley¹, Stella Somiari¹, Joji Iida¹, Stephen C. Benz⁴, Justin Golovato⁴, Shahrooz Rabizadeh⁵, Patrick Soon-Shiong⁵, Richard D. Smith², Richard J. Mural¹, Craig D. Shriver⁶, Hai Hu¹, Karin D. Rodland²

¹Chan Soon-Shiong Institute of Molecular Medicine at Windber, Windber, PA; ²Pacific Northwest National Laboratory, Richland, WA; ³CBCP, Henry Jackson Foundation, Rockville, MD; ⁴NantOmics, LLC, Culver City, CA; ⁵NantWorks, LLC, Culver City, CA; ⁶Murtha Cancer Center, Walter Reed National Military Medical Center/USU, Bethesda, MD

Aim

To investigate the proteogenomic landscape of breast cancer subtypes by integrating comprehensive transcriptomic, proteomic, and phosphoproteomic data, in a cohort with a substantial representation of African Americans, with a secondary objective of identifying druggable proteins that are differentially expressed in specific subtypes.

Findings

Clustering based on protein expression identified a Basal-like cluster and a Luminal-like cluster. The correlation of mRNA and protein expression was higher in the Basal-like cluster than in the Luminal-like cluster.

Computational analysis of proteins over-expressed in basal versus luminal subtypes identified candidate protein markers linked to novel drugs in DrugBank, suggesting a way to prioritize new drugs for clinical trials in basal-like breast cancer.

Phosphopeptide analysis showed an enrichment of phosphorylation on the PXXP motif associated with SH3 binding domains in the basal-like subtype.

Methods

A total of 113 breast tumors were identified from patients enrolled in the Clinical Breast Care Project using a HIPAA compliant, IRB-approved protocol. The cohort included 86 Caucasian Americans (CA) and 16 African Americans (AA), and the age of the patients was 58 ± 13 years. Quantitative global proteomic and phosphoproteomic analyses were performed using isobaric TMT 6-plex labeling with the “universal reference” strategy and IMAC enrichment of phosphopeptides.

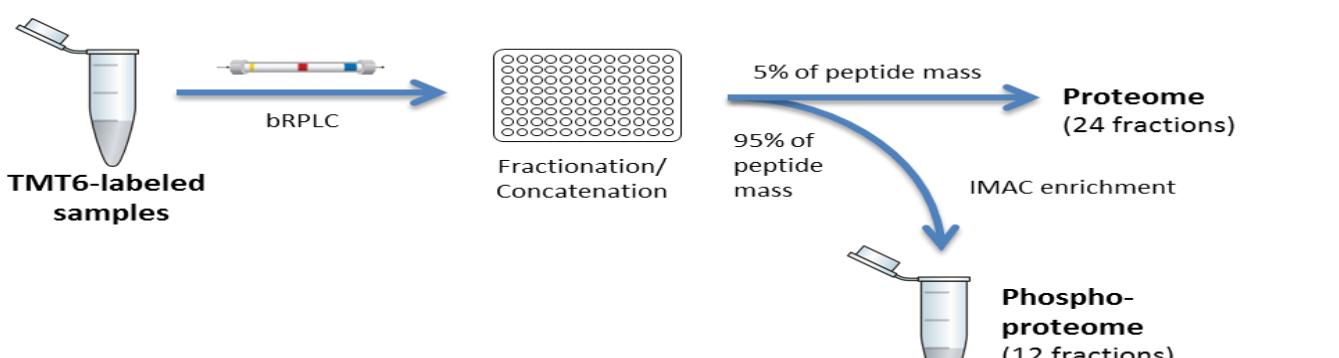


Figure 1. Quantitative TMT 6-plex workflow for global proteomics and phosphoproteomics analysis.

Mass spectrometry data were acquired using a Q-Exactive instrument and analyzed using Proteome Discoverer with Byonic node. In addition to phosphorylation search, phosphoproteomics data was also searched for O-GlcNAc modification. RNA-Seq analyses were done on Illumina HiSeq and the data were analyzed using GSAP. Gene set analysis is performed using GSA R package, and the GO gene sets of the MSigDB database.

Results

Applying the PAM50 algorithm to the RNA-Seq data identified 47 Luminal A, 18 Luminal B, 13 HER2-enriched, and 35 Basal-like subtypes. In the global proteomics data, we were able to quantitate >8600 proteins. Unsupervised clustering using the most highly varying proteins across the samples resulted in two primary clusters; one cluster was enriched for luminal-like, ER+ tumors, while the second main cluster could be subdivided into a basal-like sub-cluster and a mixed sub-cluster, both of which resembled the ER- subtype.

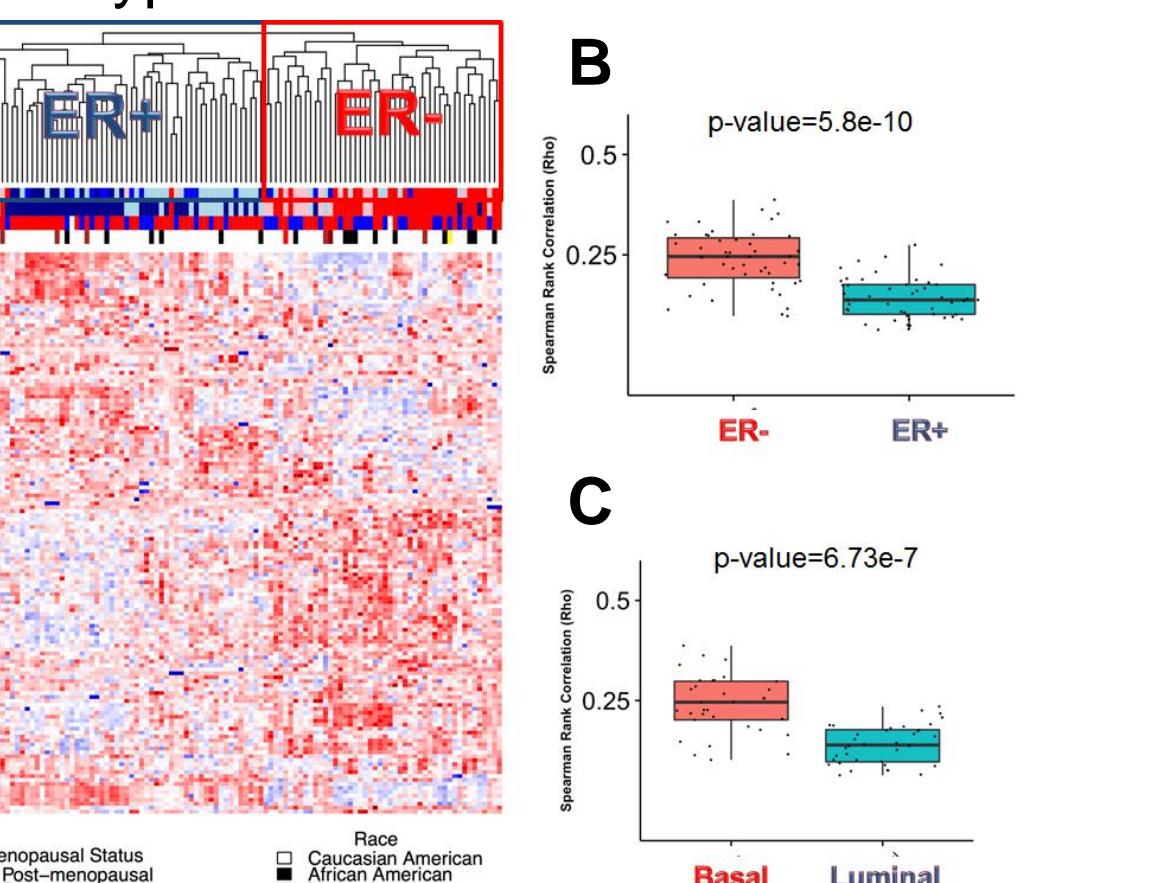


Figure 2. Global proteomics data analyses. [A] Unsupervised clustering of global proteomics data. The column bars indicate the IHC subtyping, PAM50 status derived from RNA-seq data, menopausal status and the race of the breast tumor patients. The rows represent proteins and the columns represent breast tumor samples. [B] mRNA-protein correlations separated by ER status of the tumor subtype, as identified by proteome clustering. ER (-) samples have a higher mRNA-protein correlation compared to ER (+). [C] mRNA-protein correlations of basal- versus luminal-enriched clusters. Basal subtype has a higher correlation compared to luminal.

In general, the ER+ protein cluster showed many known characteristics of the luminal gene expression subtypes, including the increased representation of post-menopausal and CA patients. Most of the pre-menopausal patients and most of the AA patients were in the Basal-like cluster. Differential protein expression analyses between the two primary clusters confirmed several known markers, including overexpression of KRT8/KRT18 in the luminal-enriched cluster. The second major cluster (basal + mixed) was predominantly enriched in ribosomal proteins (i.e., RPS7, RPL12).

Basal and luminal-enriched cluster analyses

In order to further explore the proteomic basis of the IHC subtypes, we focused the rest of our analyses on the samples determined to be “basal-like” and “luminal-like” based on hierarchical clustering of the proteomics data.

The proteomic data confirm the well-known association between the basal subtype and both younger, pre-menopausal patients and African American patients. Since the basal subtype is known to be aggressive and lack a targeted therapy, there is an urgent need to find new drugs for treatment of the basal subtype. We therefore identified the most differentially expressed proteins in the basal-like cluster, and searched those proteins against the list of drug targets in DrugBank.

Differential expression analyses of the basal- and luminal-enriched clusters resulted in 22 markers specific to the basal subtype and 60 markers specific to the luminal-enriched subtype.

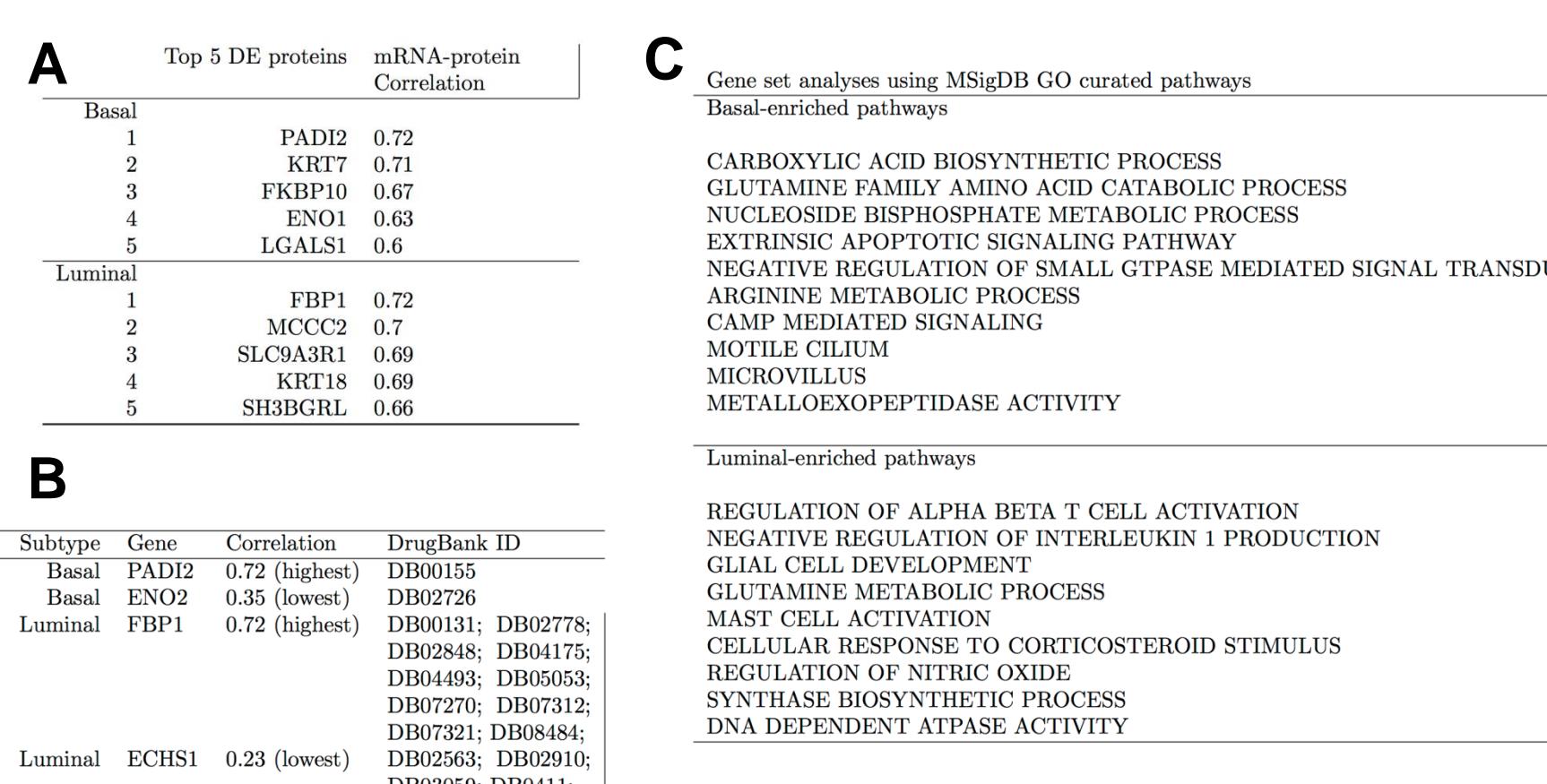


Table 1. [A] List of top 5 differentially expressed proteins between the basal- and luminal- enriched clusters, along with their mRNA-protein expression correlation. [B] Highest and lowest correlated proteins that are potentially druggable targets, along with the DrugBank ID. [C] Subset of basal- and luminal-enriched pathways identified using gene set analyses and GO gene sets of MSigDB curated pathways.

Six out of 22 proteins over-expressed in the basal subtype are targets of drugs in the DrugBank. One of these is HSPA5, a target of the drug DB00945 which is currently under investigation in breast cancer trials. Using gene set analysis to identify pathways enriched in the basal subtype identified 43 significantly enriched pathways. Two proteins, UGP2 and CLIC4 that are categorized in the GO with carboxylic acid biosynthetic process and microvillus cellular structure, overlapped with the 22 basal markers identified earlier.

Phosphopeptide and O-GlcNAc analyses

A similar search of the phosphoproteomic data provided quantitation of >12,750 phosphopeptides. Unsupervised clustering of the phosphoproteins resulted in four primary groups, with one being basal-enriched and another being luminal-enriched [figure not shown]. The other 2 groups were mixed of sub-types. We also observed >50 differentially overexpressed phosphopeptides. While some of these phosphosites have been previously reported (e.g., on RANBP2), other phosphosites appeared to be novel (e.g., on IRF2BP2). Here we focused only on the phosphopeptides that were differentially expressed in basal versus luminal enriched clusters. The list of the identified DE phosphopeptides is shown in Table 2. Interestingly, all of the basal phosphopeptides have a clear PXXP motif, which is known to bind SH3 domains in signal transduction pathways.

	Phosphopeptide	Phosphorylation site	Protein description
Basal	1	SVAEAGLLPQTpPR	RALGPS2 - Ser 329
	2	EMSFpPAQOPGCTAISAR	CEP170B - Ser 1064
	3	GGVGAPGSNpPPTR	PPM1H - Ser 223

Table 2. List of differentially expressed phosphopeptides between the basal- and luminal-enriched clusters. PXXP motif seems to be common among the basal phosphopeptide markers.

There were >700 O-GlcNAc sites for which there is quantitative information. Comparing these with overexpressed phosphopeptides resulted in >10 sequences that seem to be modified with phosphorylation and/or O-linked glycosylation. 19 of the >700 O-GlcNAc sites seem to be present in almost half of the samples, for example, O-GlcNAc sites on RANBP2, CIC, RIPOR1, TECR, PLEKHA6.

Conclusion

Integrating molecular data from different platforms and conducting orthogonal computational methods provides new insights into breast cancer subtypes, and also contribute to identifying candidate drug targets for difficult-to-treat basal-like subtype of breast cancer.

Acknowledgements

Portions of this work were supported by Department of Defense Interagency Agreement MIPR2DO89M2058/183030. The experimental work described herein was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy (DOE) and located at Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL0 1830. The views expressed in this article are those of the authors and do not reflect the official policy of the department of Army/Navy/Air Force, Department of Defense, or U.S. government.