SUPPORTING INFORMATION

Multi-Center Study Using Desorption-Electrospray-Ionization-Mass-Spectrometry Imaging for Breast-Cancer Diagnosis

Andreia M. Porcari[1,2†], Jialing Zhang[3†], Kyana Y. Garza[3ξ], Raquel M. Rodrigues-Peres[4ξ], John Q. Lin[3], Jonathan H. Young[3], Robert Tibshirani[5], Chandandeep Nagi[6], Geisilene R. Paiva[4], Stacey A. Carter[7], Luis Otávio Sarian[4], Marcos N. Eberlin[1,8], and Livia S. Eberlin[3*]

[1] Thomson Mass Spectrometry Laboratory, Department of Chemistry, University of Campinas - UNICAMP, SP, 13083-970, Brazil

[2] Laboratory of Multidisciplinary Research, São Francisco University, Bragança Paulista, São Paulo, 12916-900 Brazil

[3] Department of Chemistry, The University of Texas at Austin, Austin, TX, 78712

[4] Department of Gynecological and Breast Oncology, CAISM Women's Hospital, Faculty of Medical Sciences, University of Campinas, Campinas, São Paulo, 13083-881, Brazil

[5] Departments of Biomedical Data Science and of Statistics, Stanford University, Stanford, CA 94305

[6] Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, 77030

[7] Department of Surgery, Baylor College of Medicine, Houston, TX, 77030

[8] Mackenzie Presbiterian University, School of Engineering, São Paulo, SP 01302-907, Brazil

[†] Authors contributed equally to this work and are co-first authors

[ξ] Authors contributed equally to this work and are co-second authors

* to whom correspondence should be addressed: liviase@utexas.edu

Supporting Materials and Methods

Histopathology and light microscopy

The same tissue sections analyzed by DESI-MS imaging were subjected afterwards to standard H&E staining protocol. Pathologic evaluation was performed by Dr. Geisilene Paiva in Brazil, and Dr. Chandandeep Nagi in the USA using light microscopy. Regions of clear diagnosis were assigned and delineated in the glass slides. Light microscopy images of the H&E stained slides were taken using the EVOS FL Auto Cell Imaging System (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA). ER, PR and HER2 status were established using IHC and/or FISH analyses during clinical workflow. The expression of ER and PR was determined by histopathologic analysis for 77 IDC tissue samples based on the percent of cells stained using IHC. HER2 status was first evaluated by histopathologic analysis using IHC, resulting in negative, indeterminate or positive results. For samples with indeterminate HER2 status by IHC, FISH analysis was performed, yielding a final HER2 status determination for 67 of the IDC tissue samples. For our study, samples with staining higher than 10% were considered positive for ER or PR, whereas samples with no positive stain were considered negative.

Human Breast Cancer Tissues

131 frozen human breast tissue samples were obtained for our study. Demographic and clinicopathologic characteristics of the samples are provided in **Supporting Table 1**. Tissue procurement, handling and shipment were performed under approved IRB protocols at the respective institutions. Samples from Brazil were prospectively collected from a clinic at the CAISM Hospital da Mulher Prof. Dr. José Aristodemo Pinotti, at the University of Campinas (Campinas, Sao Paulo, Brazil). All other samples were obtained in the USA from the Cooperative Human Tissue Network (CHTN) or Asterand Biosciences (Detroit, MI). Tissue samples were sectioned at 16 µm thick sections using a CryoStar^{TM} NX50 cryostat (Thermo Scientific, San Jose, CA) and stored in a -80^{o}C freezer. Prior to MS imaging, all the glass slides

were dried for ~15 min in ambient conditions. After DESI-MSI, the same tissue sections analyzed by DESI-MS imaging were subjected afterwards to standard H&E staining protocol.

For Brazilian samples, both normal and cancer tissue were collected from women undergoing mastectomy or quadrantectomy as part of their cancer treatment in the Division of Gynecological Oncology and Breast Pathology of CAISM. Tissue from the normal glandular area (5 cm away from the tumor edge) of the same breast undergoing a mastectomy or quadrantectomy were resected with a scalpel and used in this study. These samples were further subdivided into two mirrored fragments. Both tumor and normal tissues were later confirmed through histopathology by expert pathologist. Additionally, normal breast samples from American, European, and Asian patients were acquired from contralateral breasts from bilateral mastectomies, breast reduction surgeries of non-neoplastic purposes, and prophylactic mastectomies as well as from non-cancerous regions adjacent to the tumor. Cancer tissue was selected from invasive mammary ductal carcinoma of the breast from patients undergoing mastectomy or lumpectomy.

Tissue samples were composed of areas corresponding to *in situ* ductal carcinoma, invasive ductal carcinoma, fibrosis within neoplasia, pure fibrosis, necrosis, adipose tissue, vessels, normal glands, hyperplasia, cistus, elastosis and inflammatory infiltrates. All these characteristics were annotated per slide by the pathologist and only the regions of interest (well defined as invasive ductal carcinoma or normal glands/fibrosis) were used for statistical evaluation. Although all the samples collected in Brazil and in the USA had already been histologically characterized in the occasion of their collection, all the samples were stained for H&E and reviewed by the pathologists after analysis by DESI-MSI, thus testifying the classification of the samples.

Tissue Collection

Samples from Brazil: Only samples from women who agreed to participate in the study and signed the consent form for this purpose before their surgery were collected. Immediately after the removal of the surgical specimen, the tissue was taken to the Pathology Laboratory which is next to the surgical center. The specimen was then macroscopically assessed, and the tumor area was identified. If the tumor area

had a sufficient size for diagnostic purposes and research (at least 0.8cm), a sample of the tumor as well as one from the normal glandular area from the same breast (5 cm away from tumor edge) were taken with a scalpel. Both tumor and normal tissues were later confirmed through histopathology by expert pathologist. Samples were snap frozen using liquid nitrogen in a maximum of four hours after the surgical removal. Then, the samples were stored at −80°C until they were sectioned in a cryostat for DESI-MSI. The medical records were reviewed to obtain their clinical and epidemiological data. Tissue sections of the Brazilian samples were shipped to the USA in dry ice (2-day international shipment).

Other samples: Samples from European, Asian, and American patients were acquired from the Cooperative Human Tissue Network or Asterand Biosciences. Following surgical resection, samples were immediately transported to Surgical Pathology, where they were microscopically evaluated by pathologist, pathology assistant, or a pathology resident. Sterile instruments and equipment including, towels, gloves, and ink are used by the pathologist during the examination of the specimen. After a quality control diagnosis, the samples were frozen in vapor phase liquid nitrogen, on dry ice, or in liquid nitrogen within 1 to 4 hours of resection. Prior to shipment, all samples are stored in vapor phase liquid nitrogen. Samples were shipped in dry ice and stored at -80°C until sectioned a cryostat and subjected to DESI-MS analysis.

Statistical Analysis

MS data corresponding to the areas of interest were extracted from the ion images using MSiReader software [37]. Data pre-processing and statistical analysis were performed using the same methodology on the DESI mass spectra profiles from Brazil and USA to identify predictive ion markers and evaluate the performance of the classifiers for cancer diagnosis. The *m/z* range was discretized using binning size of 0.01. Peaks appearing in more than 10% of the pixels were kept for analysis. Pixels were then normalized by sum of intensities after any mass range restriction. Logistic regression was performed with Lasso regularization using the “glmnet” package in the R language. Regularization parameters were determined by 5-fold cross-validation analysis. The data acquired in the USA was randomly divided in a training and

validation set of samples, 75%-25% per patient basis. The data acquired in Brazil was independently evaluated as a test set using the same methodology. PCA was performed by centering the pre-processed data to mean zero and computing principal components using the prcomp function in R. The first three principal components were visualized with the rgl and pca3d packages for R. To quantify tissue similarity, the cosine similarity method was used from the lsa package in CRAN. All pixels from each condition (USA normal, USA cancer, Brazil normal, Brazil cancer) were averaged to representative spectra.

The same DESI mass spectra data (*m/z*>700) extracted from cancer regions used to build the classification model for IDC diagnosis were used to build classification models to predict positive or negative status for ER, PR and HER2, using a 5-fold CV approach.

Supporting Results

Predictive features selected by Lasso

A total of 44 predictive *m/z* values were selected for the classification model with assigned mathematical weights related to their importance in distinguishing between IDC and normal tissues. Positive weights were attributed by the Lasso to 17 *m/z* values that were important in characterizing IDC, whereas negative weights were given to 27 values which were important to characterizing normal tissue. Identification of the ions corresponding to the selected *m/z* values were performed using tandem MS experiments and high mass accuracy measurements followed by comparison to lipid databases (**Table 2**). The five largest negative weights were assigned to PE(36:2) of *m/z* 742.539, PG(36:4) of *m/z* 769.502, PC(34:2) of *m/z* 792.530, PE(O/P-38:5) of *m/z* 750.545, and PS(36:2) of *m/z* 786.529, which are the ions indicative of normal breast tissue. For IDC tissue classification, however, the highest five positive weights were assigned by the Lasso to PI(36:2) ions of *m/z* 861.549, PG(34:1) of *m/z* 747.517, PG(40:7) of *m/z* 819.517, PS(38:4) of *m/z* 810.525, and PE(38:4) of *m/z* 766.539. Note also that the weights assigned to these ions corroborate with trends observed in the mass spectra and DESI ion images for IDC and normal tissues, which showcases the power of the Lasso statistical model in selecting predictive features that are potential biomarkers of disease states.

Table S1. Demographic and clinicopathologic characteristics of samples used in our study.

	Normal Samples, n	IDC Samples, n
Country of Collection	45	86
USA	10	7
Brazil	31	50
Eastern Europe	2	16
Vietnam	2	5
Unknown	0	8
Race		
African American	8	8
White	35	64
Asian	2	13
ER Status		
Negative		31
Positive		46
Undetermined		6
PR Status		
Negative		41
Positive		36
Undetermined		6
HER2 Status		
Negative		48
Positive		19
Undetermined		22

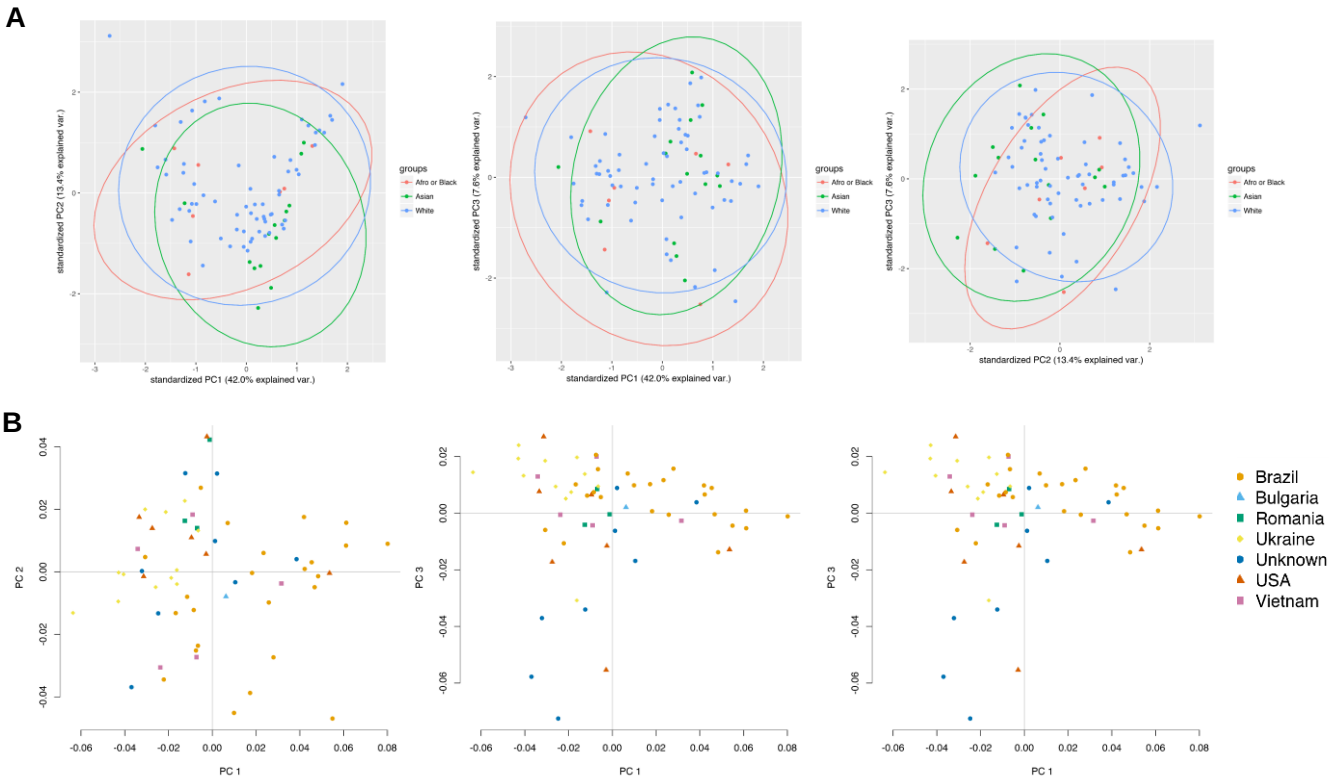

Figure S1. Projections of the mass spectra from normal breast tissue samples onto the first three principal components (PC) do not separate patients by **A)** race, or **B)** country of collection, as observed for the 2D PCA plot of PC1 versus PC2, 2D PCA plot of PC1 versus PC3, and 2D PCA plot of PC2 versus PC3.

Table S2. Cosine similarity values of the data obtained in USA and Brazil for same Brazilian normal and cancer samples.

		USA normal	USA cancer	Brazil normal	Brazil cancer
***m/z* 100-1200**	**USA normal**	1.00	0.73	0.51	0.57
	USA cancer	0.73	1.00	0.69	0.91
	Brazil normal	0.51	0.69	1.00	0.84
	Brazil cancer	0.57	0.91	0.84	1.00
***m/z* 700-1200**	**USA normal**	1.00	0.92	0.85	0.86
	USA cancer	0.92	1.00	0.72	0.92
	Brazil normal	0.85	0.72	1.00	0.82
	Brazil cancer	0.86	0.92	0.82	1.00

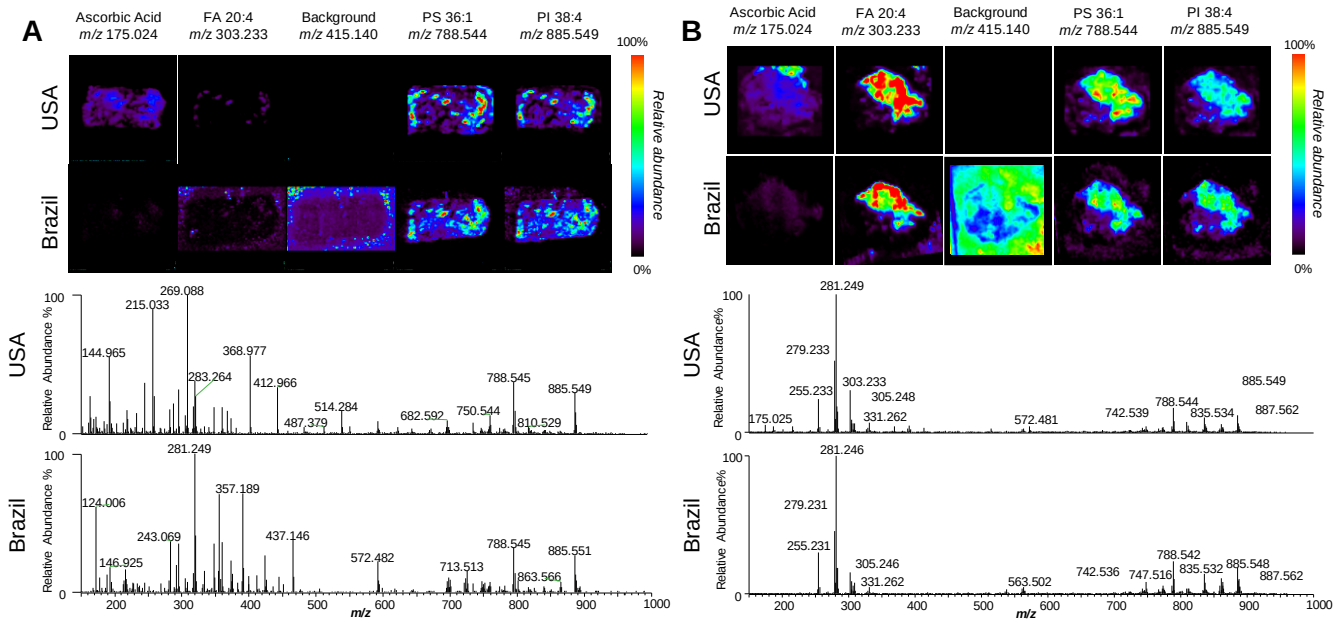

Figure S2. Negative ion mode DESI ion images and mass spectra of adjacent tissue sections of **A)** normal breast and **B)** breast IDC tissue samples analyzed in Brazil and the USA. PI: glycerophosphoinositol, PS: glycerophosphoserine, FA: fatty acid. Lipid species are described by number of fatty acid chain carbons and double bonds.

Table S3. Summary of Lasso prediction results for distinguishing between normal and IDC breast samples for USA training set, USA test set, and Brazil test set.

	Analysis	Pathologic Diagnosis	Lasso Results		Sensitivity (%)	Specificity (%)	Overall Agreement (%)
			Normal	Cancer			
USA Training set	Per Pixel	Normal	1049	36	97.6	96.7	97.6
		Cancer	417	17189			
	Per Patient	Normal	21	0	97.8	100.0	98.5
		Cancer	1	44			
USA Test Set	Per Pixel	Normal	184	1	99.1	99.5	99.2
		Cancer	53	6120			
	Per Patient	Normal	6	0	100.0	100.0	100.0
		Cancer	0	15			
Brazil Test Set	Per Pixel	Normal	2042	45	94.7	97.8	95.3
		Cancer	495	8795			
	Per Patient	Normal	18	0	96.2	100.0	97.7
		Cancer	1	25			