Journal of Proteomics 285 (2023) 104955

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

High-throughput proteomics of breast cancer subtypes: Biological characterization and multiple candidate biomarker panels to patients' stratification

Alexandre Luiz Korte Azevedo [a], Talita Helen Bombardelli Gomig [b], Michel Batista [c,d], Fabricio Klerynton Marchini [c,d], Cleverton César Spautz [e], Iris Rabinovich [e], Ana Paula Martins Sebastião [f,g], Jaqueline Carvalho Oliveira [a], Daniela Fiori Gradia [a], Iglenir João Cavalli [a], Enilze Maria de Souza Fonseca Ribeiro [a,*]

[a] Genetics Post-Graduation Program, Genetics Department, Federal University of Parana, Curitiba, Parana, Brazil
[b] Research Institute Pele Pequeno Principe, Curitiba, Parana, Brazil
[c] Laboratory of Applied Sciences and Technologies in Health, Carlos Chagas Institute, Fiocruz, Curitiba, Parana, Brazil
[d] Mass Spectrometry Facility - RPT02H, Carlos Chagas Institute, Fiocruz, Curitiba, Parana, Brazil
[e] Breast Disease Center, Hospital Nossa Senhora das Graças, Curitiba, Parana, Brazil
[f] Medical Pathology Department, Federal University of Parana, Curitiba, Parana, Brazil
[g] Pathology Center, Hospital Nossa Senhora das Graças, Curitiba, Parana, Brazil

ARTICLE INFO

Keywords:
Breast cancer
Machine learning
Mass-spectrometry
Proteome
Support vector machine

ABSTRACT

Background and aims: The actual classification of breast tumors in subtypes represents an attempt to stratify patients into clinically cohesive groups, nevertheless, clinicians still lack reproducible and reliable protein biomarkers for breast cancer subtype discrimination. In this study, we aimed to access the differentially expressed proteins between these tumors and its biological implications, contributing to the subtype's biological and clinical characterization, and with protein panels for subtype discrimination.

Methods: In our study, we applied high-throughput mass spectrometry, bioinformatic, and machine learning approaches to investigate the proteome of different breast cancer subtypes.

Results: We identified that each subtype depends on different protein expression patterns to sustain its malignancy, and also alterations in pathways and processes that can be associated with each subtype and its biological and clinical behaviors. Regarding subtype biomarkers, our panels achieved performances with at least 75% of sensibility and 92% of specificity. In the validation cohort, the panels obtained acceptable to outstanding performances (AUC = 0.740 to 1.00).

Conclusions: In general, our results expand the accuracy of breast cancer subtypes' proteomic landscape and improve the understanding of its biological heterogeneity. In addition, we identified potential protein biomarkers for the stratification of breast cancer patients, improving the repertoire of reliable protein biomarkers.

Significance: Breast cancer is the most diagnosed cancer type worldwide and the most lethal cancer in women. As a heterogeneous disease, breast cancer tumors can be classified into four major subtypes, each presenting particular molecular alterations, clinical behaviors, and treatment responses. Thus, a pivotal step in patient management and clinical decisions is accurately classifying breast tumor subtypes. Currently, this classification is made by the immunohistochemical detection of four classical markers (estrogen receptor, progesterone receptor, HER2 receptor, and the Ki-67 index); however, it is known that these markers alone do not fully discriminate the breast tumor subtypes. Also, the poor understanding of the molecular alterations of each subtype leads to a challenging decision-making process regarding treatment choice and prognostic determination. This study, through high-throughput label-free mass-spectrometry data acquisition and downstream bioinformatic analysis, advances in the proteomic discrimination of breast tumors and achieves an in-depth characterization of the subtype's proteomes. Here, we indicate how the variations in the subtype's proteome can influence the tumor's biological and clinical differences, highlighting the variation in the expression pattern of oncoproteins and tumor

* Corresponding author at: Genetics Department, Federal University of Parana, P.O. Box 19071, CEP: 81531-990, Curitiba, Paraná, Brazil.
E-mail address: eribeiro@ufpr.br (E.M.S.F. Ribeiro).

https://doi.org/10.1016/j.jprot.2023.104955
Received 1 February 2023; Received in revised form 13 June 2023; Accepted 13 June 2023
Available online 28 June 2023

suppressor proteins between subtypes. Also, through our machine-learning approach, we propose multi-protein panels with the potential to discriminate the breast cancer subtypes. Our panels achieved high classification performance in our cohort and in the independent validation cohort, demonstrating their potential to improve the current tumor discrimination system as complements to the classical immunohistochemical classification.

1. Introduction

According to the last cancer statistics of the GLOBOCAN project (https://gco.iarc.fr/), breast cancer overtaken lung cancer and became the most diagnosed cancer worldwide in 2020, with 2.3 million new cases estimated. In women, breast cancer is the leading cause of cancer death besides the most diagnosed cancer type [1].

The breast cancer has a broad phenotypic spectrum, with tumors presenting different clinical behaviors and treatment responses. These differences arise from the wide variability of molecular alterations in the breast tumors, including the proteomic level [2,3]. The molecular classification of Perou and coworkers [4], and the subsequent immunohistochemical (IHC) classification described in the St Gallen International Breast Cancer Conference [5,6], discriminated the mammary tumors according the expression of recognized markers, such as estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and the Marker Of Proliferation Ki-67. These classification includes: Luminal A breast cancer (*LABC*; ER+, PR+, HER2-, KI-67 < 14%), Luminal B HER2+ breast cancer (*LBBC*; ER+, HER2+, any PR or KI-67), Luminal B HER2- breast cancer (*LBBC*; ER+, PR- (or low), KI-67 > 14%), HER2+ enriched breast cancer (*HER2*+; ER-, PR-, HER2+, KI-67 > 14%) and triple-negative/ basal-like breast cancer (*TNBC*; ER-, PR-, HER2- and KI-67 > 14%) subtypes. The luminal tumors (ER+) represents 60-80% of all diagnosed breast tumors. These tumors tend to present low histological grade, good response to endocrine therapy, lower aggressivity, and good prognosis. In contrast, 20-40% of the diagnosed cases are HER2+ and TNBC tumors, which have worse prognosis, poor response to endocrine therapy, high aggressivity, and high metastasis and relapse rates [7].

The IHC classification is the most common approach in clinics: together with clinicopathological features like tumor size, grade, and node commitment, the IHC classification is used to determine the tumor aggressiveness and metastasis probability, and predict patients' survival and relapse time, guiding treatment choice and decision making [8,9]. Nevertheless, the high mortality rates in breast tumor patients suggest that the classification solely based on these features may not be enough to accomplish an adequate stratification of these patients, highlighting the necessity of new biomarkers that could improve the current classification.

In this context, the studies related to breast cancer patients' stratification are focused on the large-scale data generated by the omics approaches, including proteomics [10–12]. In fact, the proteins are the biochemical effectors of most of the biological events involved in cell functions and homeostasis, representing one of the major alterations involved in the malignancy of breast tumors; thus, the identification of expression patterns associated with each subtype could not only contribute to the understanding of the differences in biological and clinical behaviors presented by these subtypes, but also to determine reliable biomarkers for tumor discrimination and patients stratification [13–15].

Mass-spectrometry (MS/MS) is the gold standard method to acquire proteomic expression data, and the machine learning technologies figures among the best suitable approaches to data mining and biomarker selection [10]. MS-based studies have already performed breast cancer proteomic investigations, identifying potential protein biomarkers. However, the low consistency in the repertoire of differentially expressed proteins (DEPs) and biomarkers suggested in these MS-based studies has limited the application of proteins in breast cancer clinics [2,11,12,16,17]. To surpass these limitations, the breast cancer proteomic landscape must be expanded by applying new proteomic approaches in new populations and sample cohorts.

In this study, we applied high-throughput label-free mass-spectrometry, bioinformatic methodologies, and a machine learning approach based on Support Vector Machine (SVM) to investigate the proteome of different tumors representing the breast cancer subtypes. We aimed to identify DEPs among the breast cancer subtypes and its involvement in the biological differences among these subtypes, suggesting multi-protein panels with potential to discriminate breast cancer subtypes.

2. Materials and methods

2.1. Patients and sample

This study was approved by the Comitê de Ética em Pesquisa, Setor de Ciências da Saúde, Universidade Federal do Paraná (CAAE 19870319.3.0000.0102). The tumoral samples were collected under patient's signed informed consent before the surgical procedures at the Nossa Senhora das Graças Hospital in Curitiba, Brazil. The tumor tissues were macrodissected and stored in RNAlater™ (Invitrogen) at -80 °C until protein extraction.

In total, 19 tumor samples were used in this study. Among these, five were classified as LABC, four as LBBC (HER2-), four as HER2+ enriched breast cancer, and six as TNBC (Fig. 1). Immunohistochemical detection of ER, PR, HER2 and the KI-67 marker was obtained from the patients' clinicopathological data and used for subtype classification [5,6]. We selected only tissues with invasive ductal carcinoma (IDC) histopathology (and a unique sample from mixed IDC); tissues with immunohistochemical analysis of the classic markers ER, PR, HER2, and Ki-67; an adequate amount of tissue for chromatographic and mass spectrometry analysis; low variation of patient age and tumor grade, and absence of any cancer treatment. These criteria, along with the selection of matched patients' demography, limited the availability of tissue samples, and resulted in an unequal number of samples from each subtype. Appropriate statistical methods were applied to minimize the impact of these differences. Our study design is summarized in the experimental procedures described in Fig. 1.

2.2. Sample preparation for mass spectrometry analysis

The protein extracts were obtained, quantified, and prepared for mass spectrometry analysis according to a standardized protocol adapted from Gomig and coworkers [18]. Firstly, 25 mg of each sample was lysed in 250 µl lysis buffer (SDS 4%, Tris-HCl 0,1 M; pH 7,5, DTT 0,1 M), homogenized in TissueLyser II sample disruptor (Qiagen Corp. MD, USA) at an oscillation frequency of 30 Hz, in three cycles of 3 min intercalated by heating at 95 °C for 5 min. After, the lysates were sonicated at 25 Hz frequency in three cycles of 30 s, intercalated by intervals of 1 min in ice. Lastly, two centrifugation steps (first 5 mins at 12.880 rcf and 4 °C; second 10 mins at 12.000 rcf and 4 °C) were performed to remove cell debris. Protein concentration was determined by the tryptophan method [19].

The final protein extracts (25 µg) were separated in 1D SDS-PAGE acrylamide gels (12%) and then reduced with DTT 10 mM, alkylated with iodoacetamide 50 mM, and digested overnight with trypsin 12.5 ng/µl solution in ammonium bicarbonate buffer 55 mM at 37 °C. The peptides were extracted through a sequence of twice acetonitrile 30%, trifluoroacetic acid 3%, and twice acetonitrile 30% washes and then

dried in a vacuum centrifuge and desalted with C18 Stage Tips [18]. Protein digestion was performed on the whole gel containing the total tissue protein extract. After, the samples were then quantified on 280 nm nano spectrophotometer, and 0.5 microgram of the sample was injected into the chromatographic column.

2.3. Mass-spectrometry analysis and protein identification

Both liquid chromatography (3000 RSLCnano chromatograph, Thermo Scientific) and mass spectrometry (Orbitrap Fusion Lumos mass spectrometer, Thermo Scientific) were used for peptide analysis. This step was performed in the mass spectrometry facility (RPT02H) of Carlos Chagas Institute - Fiocruz, Paraná. All chromatography and mass spectrometry run parameters are described in the Supplementary Table 1. All runs were performed in duplicate.

Carbamidomethylation of cysteine was set as a fixed modification, and oxidation of methionine and acetylation of protein N-terminal were set as variable modifications. The mass spectra were analyzed in the MaxQuant software (v. 1.6.17), and the protein identification was performed in the human UniProt database (75,777 entries). An FDR of 1% was independently applied for both peptide and protein identification, with at least seven amino acids required for peptide identification. The reverse of peptides was used for FDR estimation. Additional MS/MS data is provided in the Supplementary Table 2.

2.4. Statistical approaches for expression data analysis

The normalized spectral label-free protein intensity (*Label-free quantification*, LFQ intensity) was used to determine protein expression, and all data processing was performed in the Perseus software (v.1.6.15.0) [20]. Initially, the proteins identified as potential contaminants, reverse peptides, and proteins only identified by modified peptides were removed; only proteins identified in at least 70% of the samples were maintained for statistical analysis. The LFQ values were logarithmized (log^2) and normalized by *Z*-score, with missing values being replaced by values from the normal distribution (width = 0.3; down-shift = 1.8), as recommended by the Perseus developers [21].

Two statistical analyses were used to identify differentially expressed proteins. The one *vs.* all *t*-test approach was applied to compare the samples from one subtype against the samples from all other subtypes (*P*-value <0.05; $Log^2FC \pm 0.58$); The ANOVA and Tukey post-hoc test approach was used in the comparison of all subtypes against each other (Permutation-based FDR-value <0.05). The distribution of the DEPs across the subtypes was visualized through hierarchical clustering in Perseus. (The adopted clustering parameters were: Distance: Euclidean; Linkage: Average, and Cluster Preprocessing: K-means).

2.5. Oncoproteins and tumor suppressor proteins identification and functional enrichment analysis in the molecular signatures database (MSigDB)

The Network of Cancer Genes & Healthy Drivers (NCG, v. 7.0) [22] database is a curated collection of cancer genes and healthy drivers. We used the NCG database to verify the presence of putative or canonical oncogenes and tumor suppressor genes among the genes that codify our DEPs. The NCG classification is based on the prevalence of gain-of-function or loss-of-function alterations in The Cancer Genome Atlas (TCGA) data [23]. Some genes were also described by the NCG database as "healthy drivers" or simply as "unclassified mode of action".

The Molecular Signatures Database (MSigDB, v. 7.5.1) [24] was used to perform an *in silico* functional analysis of the DEPs identified by the *t*-tests. The enrichment analyses were performed according to the

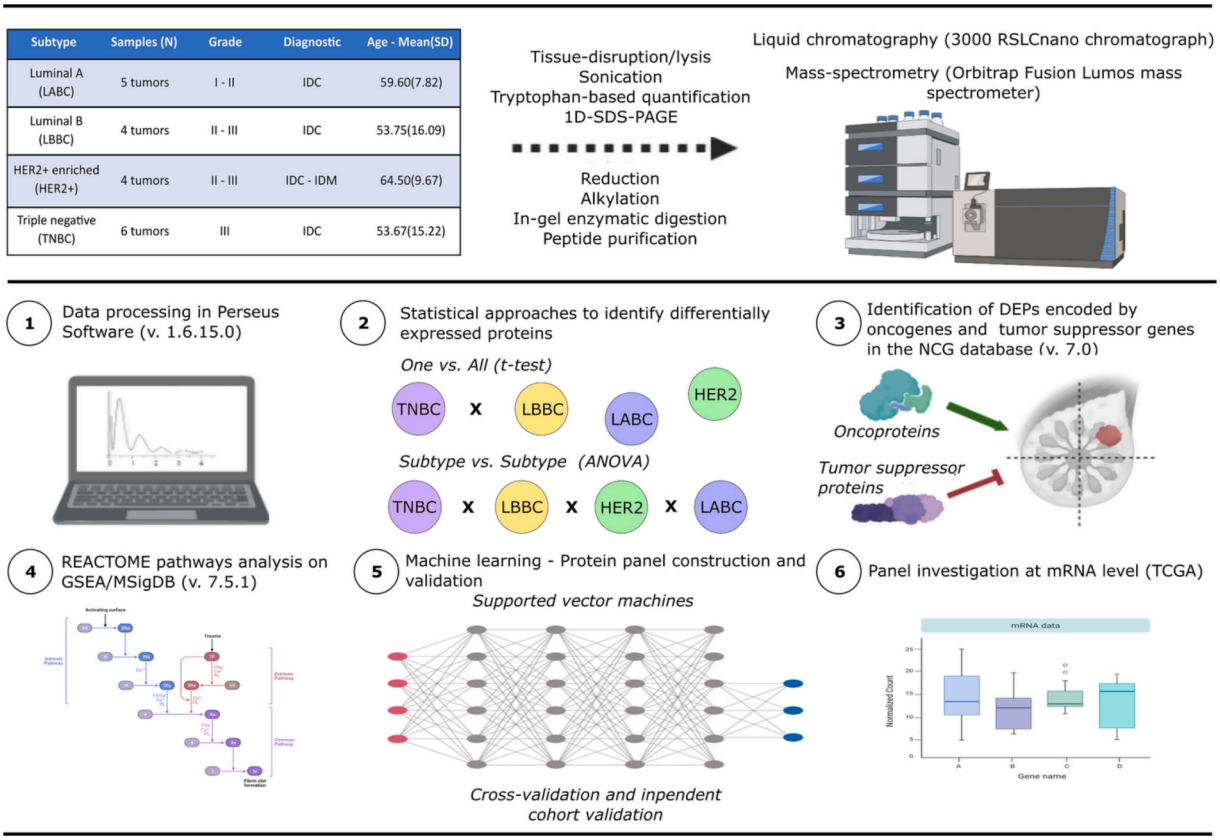

Fig. 1. Workflow of proteomic and bioinformatic analysis of total-proteome data from breast tumor samples. The tumor subtypes were determined immunohistochemically by detection of the estrogen, progesterone, and HER2 receptors, as well the KI-67 index. IDC = Invasive ductal carcinoma. IDM = Mixed invasive ductal lobular carcinoma.

canonical pathway signatures of REACTOME (FDR-value <0.05) to identify the pathways and biological processes enriched in the set of DEPs of each subtype separately. An enrichment *Z*-score was calculated for each significant pathway found [25]. We also used the STRING (v. 11.5) [26] and Cytoscape (v. 3.0.9) [27] tools to visualize the interaction of DEPs enriched in shared pathways/processes. In the STRING analysis, we considered 'co-expression', 'databases', 'co-occurrence', 'experiments', and 'neighborhood' as interaction data sources. We only included interactions with 0.4 or higher interaction scores.

2.6. Construction of the support vector machine models and validation in independent cohorts

A supervised machine learning approach based on Support Vector Machines (SVMs) was applied to construct classification models for each subtype. Each model comprises a protein panel with the potential to discriminate samples of one subtype from others. The SVM was chosen to our analysis due to its already validated effectiveness and reliability in small sample size cohorts [28,29]. Once we needed a unified list of DEPs in this analysis, we used the 122 DEPs originated from the ANOVA test.

We used the SVM machine learning feature implemented in the Perseus (v. 1.6.15.0). Firstly, the "classification parameter optimization" was used to obtain the C and sigma values (<15% classification error) for each model based on the Radial Basis Function (RBF) kernel. A random sampling cross-validation procedure was applied: the samples were separated into test (15%) and train samples (85%), with the random sampling procedure repeated 250 times.

Next, the "classification feature optimization" was applied to rank the DEPs of each subtype according to its discriminative potential. The DEPs were ranked by *P*-value through an ANOVA-based ranking method every cross-validation run. The number of DEPs selected to compose each panel was determined to ensure <15% classification error. In a balance of sensibility and specificity, the 15% error margin is strict enough to guaranty an excellent discrimination performance [30]. Lastly, the "classification" function was used to classify the samples using the previously determined minimum number of the top-ranked DEPs of each subtype. The performance of each model was determined according to the specificity, sensibility, and false-positive rates. Principal Component Analyses (PCA) were performed to visualize the sample segregation pattern.

An independent validation cohort was used to confirm the applicability and reproducibility of our models: Our protein panels were applied to classify the breast cancer subtypes of 35 tumor samples (11 LABC, 7 LBBC, 5 HER2+ enriched, 12 TNBC) available in the LinkedOmics portal [12,31]. We applied the binary regression model implemented in the IBM SPSS Statistics (v.26), using the protein panels as covariates and subtypes (LBBC/non-LBBC, for example) as dependent variables. The performances of the models were obtained by Receiver Operating Characteristic curves (ROC curve; 95% confidence interval; *P*-values <0.05) and quantified by the area under the curve (AUC). We also obtained transcriptomic data from TCGA [23] and applied the same ROC-AUC approach to verify the models' performance at mRNA levels corresponding to the protein panels.

3. Results and discussion

3.1. Proteomic profiling of different breast cancer subtypes

High-throughput label-free mass spectrometry was applied to obtain full-proteomes of breast tumor tissues aiming to clarify the breast cancer proteome landscape and improve the understanding of the breast subtype heterogeneity. We analyzed the proteins expressed by 19 solid breast tumor tissues representing the four IHC subtypes (LABC, LBBC, HER2+, and TNBC), subsequently applying proteomic and bioinformatic approaches to investigate the expression data patterns.

We identified 4324 proteins expressed among all samples at 1% of false discovery rate (FDR) (on average, 1905.263 proteins per sample). Nevertheless, to obtain a more representative proteome, only the 1512 proteins expressed in at least 70% of the samples were maintained for downstream computational analysis and biological interpretation (Supplementary Table 3).

The unsupervised clusterization and principal component analysis (PCA) of the total proteome (1512 proteins), as shown in Fig. 2A-B, do not discriminate the samples into their subtype classification, which corroborates that not all proteome is involved in the variability of breast cancer subtypes [10,32]. In this context, we applied two different statistical approaches to identify DEPs among the subtypes, aiming to identify proteins potentially involved in the biological heterogeneity of the disease.

A one *vs.* all approach (*t*-test: *P*-value <0.05; $Log^2FC \pm 0.58$) was performed to identify proteins consistently differentially expressed in each subtype (Fig. 2C-F). The LABC subtype presented 162 DEPs (43 down- and 119 over-expressed); the LBBC 120 DEPs (106 down- and 14 over-expressed); the HER2+ enriched subtype showed 72 DEPs (53 down- and 19 over-expressed), and the TNBC 464 DEPs (121 down- and 343 over-expressed) (Supplementary Table 3). The HER2+ subtype presented the lowest number of DEPs, which suggest that the discrimination of HER2+ tumors from non-HER2+ tumors by proteome profiling could be challenging.

Next, an ANOVA with Tukey post-hoc test (FDR-value <0.05) was performed to identify DEPs among all subtypes, resulting in 122 proteins with differential expression in at least one of the compared pairs (Supplementary Table 4). As shown in Fig. 2G, most DEPs were found in the comparisons involving TNBC tumors, with an intermediary number of DEPs in the HER2+ comparisons, and few in the LABC *vs.* LBBC comparison, reflecting the degree of biological discrimination that proteome profile can provide for breast cancer subtypes.

3.2. Identification of differentially expressed oncoproteins and tumor suppressor proteins among the subtypes

Oncoproteins and tumor suppressor proteins are known to be involved in the regulation of several biological mechanisms related to cell homeostasis. Alterations in the structure or expression of these proteins can lead to disruption of pathways, processes and, consequently, to carcinogenesis [33–35]. We hypothesized that the different biological behaviors presented by the subtypes could be influenced by differential expression patterns of these proteins.

We used the NCG database (v. 7.0) to identify putative and canonical oncoproteins and tumor suppressor proteins among the DEPs of each subtype. In total, 129 putative and canonical oncoproteins and tumor suppressor proteins were mapped among the DEPs of each subtype. The gene names are used in this and subsequent topics to represent its encoded proteins.

The LABC subtype presented 32 oncoproteins/tumor suppressor proteins differentially expressed, including the over-expressed PPP2R1A (*P*-value = 0,046; $log^2FC = 1.020$) and MAPK1 oncoproteins (P-value = 0.032; $log^2FC = 1.085$), and the down-expressed tumor suppressor protein FEN1 (P-value = 0.031; $log^2FC = -1.241$) (Fig. 3A). Super expression of MAPK1 induces Tamoxifen resistance, a compound typically used in ER-positive tumors treatment [36], highlighting the relevance of the MAPK1 oncoprotein in LABC. The tumor suppressor Flap Structure-Specific Endonuclease 1 (*FEN1*) is related to genome instability and inflammation promotion. Its deficiency is also associated with a higher risk of mutations in tumor suppressor genes and proto-oncogenes [37]. The low expression of FEN1 in LABC tumors regarding other subtypes could indicate a special relevance of this tumor suppressor protein in LABC tumors and the potential impacts of its low/loss of function in LABC tumorigenesis.

LBBC tumors showed 23 oncoproteins/tumor suppressor proteins differentially expressed. Among them, we highlight the over-expressed putative oncoprotein PTPN1 (*P*-value = 0.004; $log^2FC = 1.496$) and

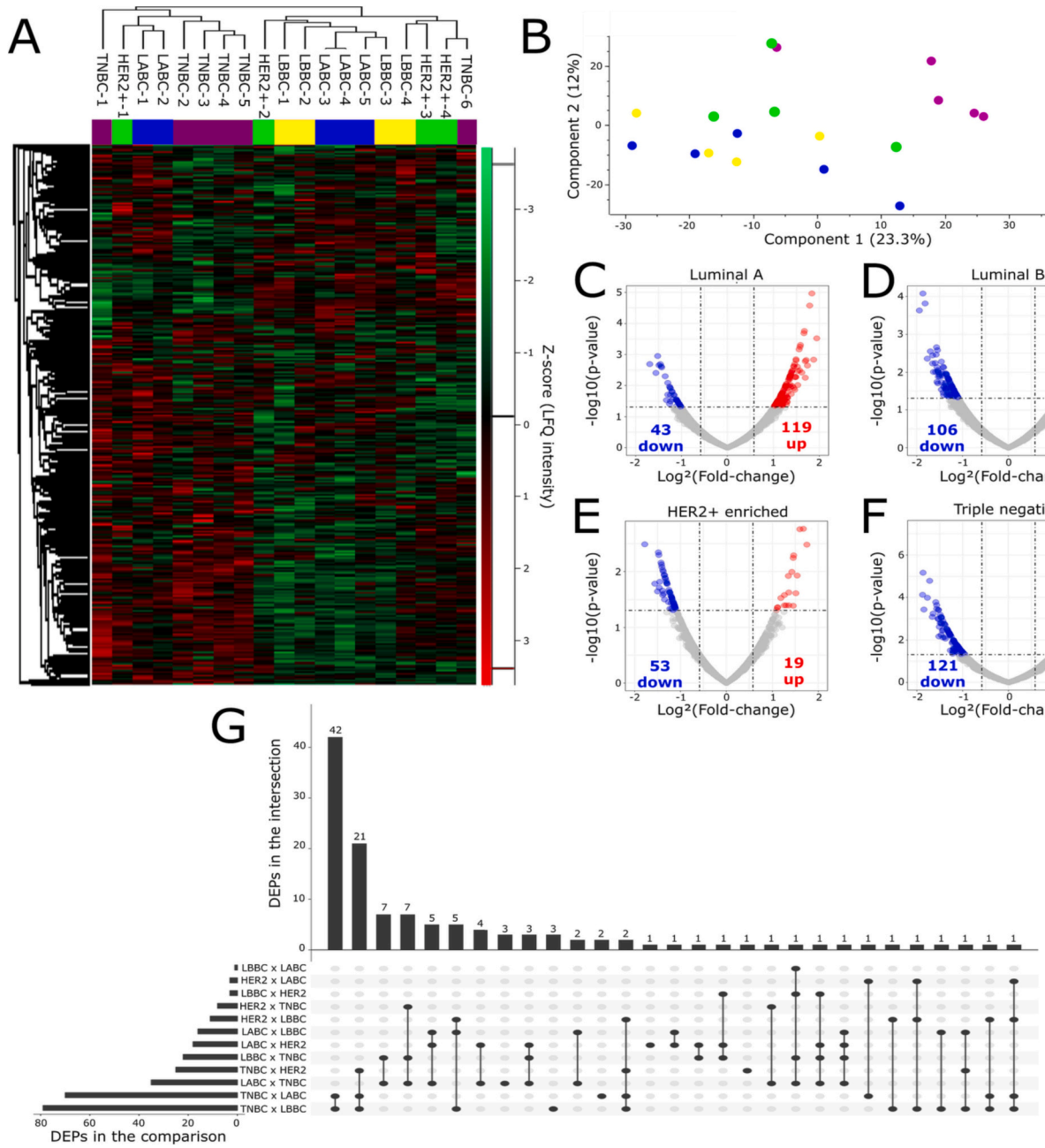

Fig. 2. Proteome profile and differentially expressed proteins in the breast cancer subtypes. The heatmap (A) and PCA (B) indicate the clusterization of breast tumors in mixed clusters when considering the 1512 proteins expressed in at least 70% of the samples. Blue = LABC. Yellow = LBBC. Green = HER2+ enriched. Purple: TNBC. (C—F) Volcano plots representing the DEPs identified per subtype in the one *vs.* all approach (*P*-value <0.05; $Log^2FC \pm 0.58$). (G) Upset plot representing the DEPs identified in the ANOVA comparisons (FDR < 0.05): The number of DEPs per compared pair is shown in the sidebars, and the number of DEPs shared by the pairs is shown in the upper bars and dots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the down-expressed tumor suppressors proteins RPL22 (*P*-value = 0.010; $log^2FC = -1.356$) and FBLN2 (*P*-value = 0.026; $log^2FC = -1.341$) (Fig. 3B). LBBC tumors have an intrinsic relation with *TP53* loss [38]: In several cancer types, down-expression or *in vitro* knockdown of RPL22 is associated with compromised p53 activation and subsequent tumor promotion [39,40]. Likewise, PTPN1 over-expression is related to breast cancer promotion and aggressiveness [41,42]. Our results highlight the relevance of RPL22 and PTPN1 in breast tumorigenesis, with a significant association with LBBC tumors.

The HER2+ subtype presented 20 oncoproteins/tumor suppressor proteins among its DEPs, including the over-expressed oncoprotein IDH1 (*P*-value = 0.005; $log^2FC = 1.452$), NAMPT (P-value = 0.005; $log^2FC = 1.466$) and APMAP (P-value = 0.001; $log^2FC = 1.596$), and the down-expressed putative tumor suppressor protein CCAR1 (*P*-value = 0.016; $log^2FC = -1.556$) (Fig. 4A). The HER2+ subtype is characterized by amplification of ERBB2, yet new protein signatures must be proposed for this subtype. Here, we suggest the over-expression of IDH1, NAMPT, APMAP and CCAR1 as signatures associated with HER2+ tumors. AMPAP and NAMPT expression, for example, were related to HER2 expression [43,44] and inflammation [45] in breast tumors, reinforcing its association with the HER2+ subtype.

We identified 88 oncoproteins/tumor suppressor proteins differentially expressed in the TNBC subtype, with a predominance of high expression of putative and canonical oncoproteins (58/88), including RHOA (*P*-value = 0.021; $log^2FC = 1.098$), CALR (P-value = 0.0006; $log^2FC = 1.494$), PSIP1 (P-value = 0.0009; $log^2FC = 1.623$), IDH2 (P-

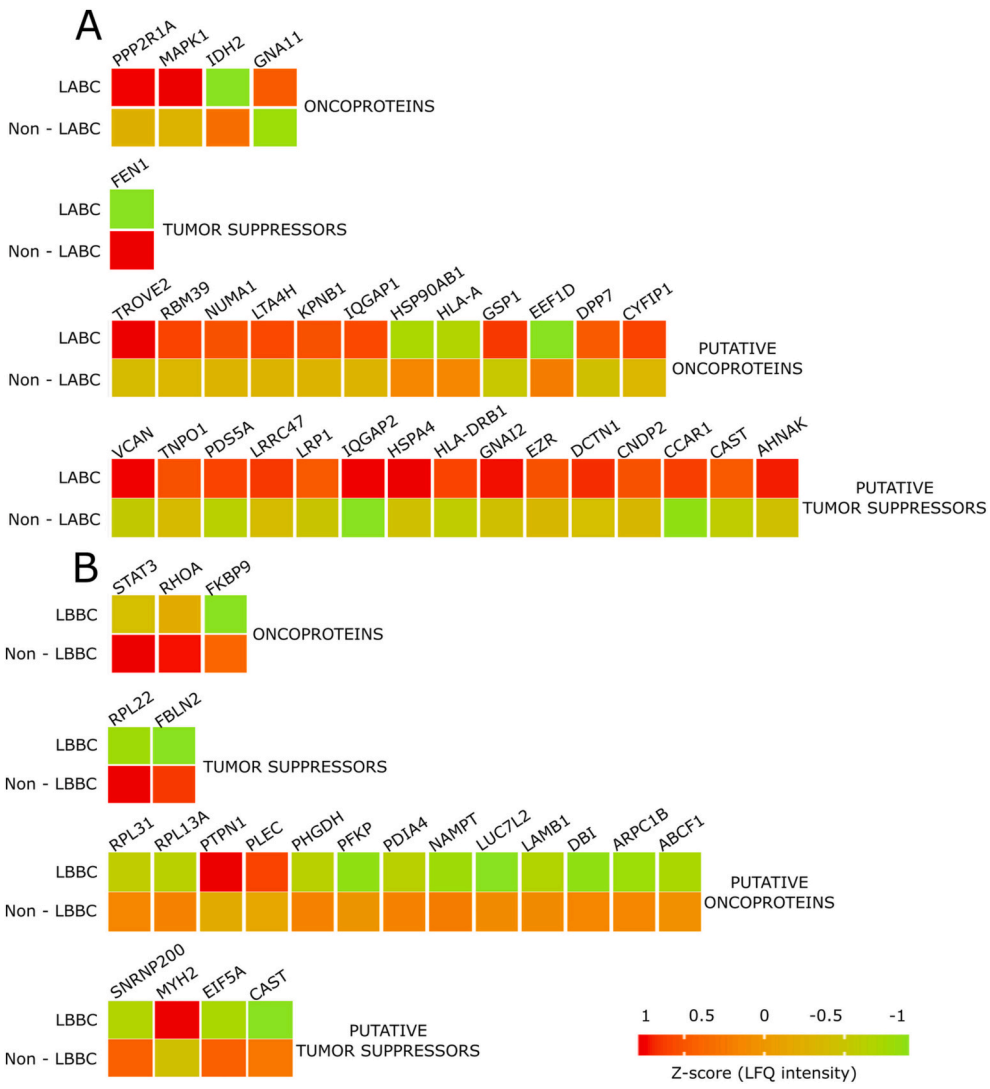

Fig. 3. Differentially expressed oncoproteins and tumor suppressor proteins in the LABC and LBBC subtypes. (A) Putative and canonical oncoproteins and tumor suppressor proteins differentially expressed in LABC tumors *vs.* non-LABC tumors. (B) Putative and canonical oncoproteins and tumor suppressor proteins differentially expressed in LBBC tumors *vs.* non-LBBC tumors. Significance cutoff fixed in P-value <0.05 and $Log^2FC \pm 0.58$. Only the top 15 DEPs of each category are shown.

value = 0.004; log^2FC = 1.305), and EEF1A1 (P-value = 0.00004; log^2FC = 1.66) (Fig. 4B). High mRNA levels of PSIP1 were previously associated with poor prognosis in basal breast tumors, and related to cell cycle dysregulation and proliferation [46]. In hepatocellular carcinoma [47] and prostate cancer [48], EEF1A1 high expression is associated with disease progression, being appointed as a potential therapeutic target. TNBC has a low frequency of activating point mutations in classic proto-oncogenes, which represents a challenge to target therapy implementation [49]. Our results describe TNBC-associated oncoproteins and tumor suppressor proteins, including PSIP1 and EEF1A1, identifying proteomic alterations related to its aggressive malignancy that can also be potential targets for TNBC therapy. A complete view of the putative and canonical differentially expressed oncoproteins and tumor suppressor proteins can be found in Supplementary Table 5.

Several reports indicate that differences among cancer types can arise from variations in the patterns of oncogenes and tumor suppressor alterations [50,51]. In breast cancer, it has been described that each subtype presents a specific landscape of somatic mutations, which defines its different genetic alterations [52]. Consistently, our findings demonstrated that different subtypes present distinct expression patterns of oncoproteins and tumor suppressor proteins, which together with other DEPS, can influence its differential biology and malignant development.

3.3. Determination of biological changes associated with the subtype's proteomes through enrichment analysis

Next, we performed the enrichment analysis on MSigDB (v. 7.5.1) to obtain the pathways and biological processes that are related to our DEPs, and whose functionality could be altered by their expression in the breast cancer subtypes. An enrichment *Z*-score was applied to predict the activation or inhibition of each pathway/process [25]. The identification of these molecular mechanisms can contribute to the understanding of the discrepancies among the subtype behaviors and also may present a relevant clinical application: The knowledge of proteomic alterations and their associated pathways can be used to select the best suitable treatment and therapy in the precision medicine; to understand the molecular complexity associates to resistance mechanisms; and also to predict patients prognostic and disease progression [53–55]. Our enrichment results were divided into processes/pathways enriched in two or more subtypes jointly (but involving different sets of DEPs) and into processes/pathways enriched in specific subtypes and their own set of DEPs (Section 3.4) (Supplementary Table 6).

As shown in Fig. 5, each set of DEPs could be associated with functional alterations related to the biological traits of its corresponding subtype. The luminal subtypes LABC and LBBC, for example, presented typical proteomic alterations regarding membrane trafficking and clathrin-mediated endocytosis: LABC tumors were related to 'activation'

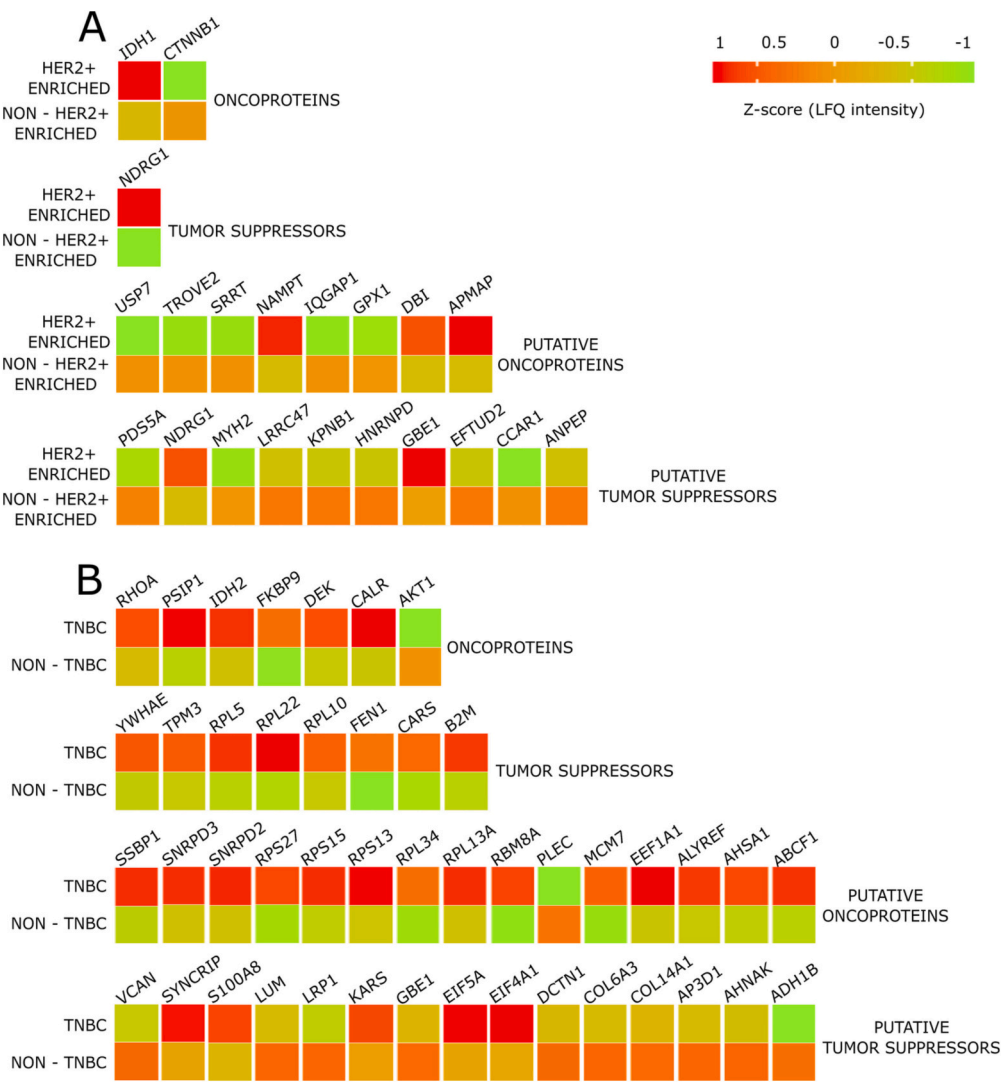

Fig. 4. Top 15 differentially expressed oncoproteins and tumor suppressor proteins in the HER2+ enriched and TNBC subtypes. (A) Putative and canonical oncoproteins and tumor suppressor proteins differentially expressed in the HER2+ enriched *vs.* non-HER2+ tumors. (B) Putative and canonical oncoproteins and tumor suppressor proteins differentially expressed in TNBC *vs.* non-TNBC tumors. Significance cutoff fixed in P-value <0.05 and $Log^2FC \pm 0.58$. Only the top 15 DEPs of each category are shown.

of clathrin-mediated endocytosis (*Z*-score = 2.23) associated with the over-expressed RAB5B, AP2B1, COPS4, and COPS2. In contrast, LBBC tumors were related to 'deactivation' (Z-score = -2.44) of this process by down-regulation of ARPC2, ARPC3, ARPC5, ACTR2, and ACTR3.

The DEPs of LABC and TNBC subtypes allowed the identification of the adaptive immune system (LABC *Z*-score = 1.29; TNBC Z-score = 3.79) and cytokine signaling in the immune system (LABC Z-score = 1.15; TNBC Z-score = 4.64) as pathways potentially involved in their tumorigenesis, among others. These pathways were associated with the over-expression of HLA-DRA, HLA-DRB1, UFL1, PPP2CA, and PPP2R1A in LABC tumors, while these processes were enriched in the TNBC due to the over-expression of SKP1, YWHAZ, and several proteasome subunits.

In LBBC and TNBC subtypes, the DEPs were related to antigen processing and cross-presentation, cellular response to hypoxia and chemical stress, hedgehog ligand biogenesis, RUNX2/3 expression and activity, stabilization of p53, and others. Interestingly, most of these processes were 'deactivated' in LBBC (*Z*-scores <0) and 'activated' in TNBC (Z-scores >0), with relevant involvement of differentially expressed proteasome subunits in both subtypes.

The TNBC and HER2+ enriched subtypes shared protein alteration patterns regarding mRNA splicing and processing. HER2+ tumors were related to 'deactivation' of mRNA splicing (Z-score = -3) and processing (Z-score = -3.16), mainly involving low-expression of HNRNPD, HNRNPM, HNRNPR, and SRRT. In contrast, for the TNBC subtype, we identified the 'activation' of mRNA splicing (Z-score = 4.47) and processing (Z-score = 4.69) through over-expression of MAGOHB, PAPBN1, RBMX, and SRSF1, for example.

Other processes/pathways, such as innate immune system, metabolism of RNA, neutrophil degranulation, apoptosis, and programmed cell death pathways, were enriched in all sets of DEPs (all subtypes), indicating that even biological alterations common to all subtypes can be derived from different alterations in protein expression in a subtype-dependent way.

3.4. Functional characterization of pathways, biological processes, and associated proteins related to specific breast cancer subtypes

The analysis of processes/pathways enriched in only one subtype and its DEPs provides more data to understanding the heterogeneity of the breast tumors and guide the clinical approach. These pathways/processes are not necessarily altered in only a specific subtype, but we propose that these alterations occur in some subtypes through dysregulation of subtype-specific DEPs, representing biological signatures of the subtypes.

The LABC subtype presented a particular expression pattern of proteins related to metabolic changes, such as IQGAP1, SOD2, IDH2, ENO2, HK1, and NDUFA9, which were enriched in processes like integration of energy metabolism (*Z*-score = 2.82), organelle biogenesis and maintenance (*Z*-score = 0.70), citric acid cycle and respiratory electron transport (Z-score 0.37), glucose (Z-score = 2.44) and carbohydrate (Z-score

Fig. 5. Functional enrichment analysis of REACTOME pathways associated with two or more subtypes based on different sets of DEPs. Diamonds' colors are related to the pathways' enrichment *Z*-score in each subtype, while its size varies according to the number of DEPs enriched in each pathway. Red = Pathway 'activation'. Green = Pathway 'deactivation'. R-HAS-174178* = APC/C-Cdh1 mediated degradation of Cdc20 and other APC/C: Cdh1 targeted proteins in late mitosis/early G1. R-HAS-72662** = Activation of the mRNA upon binding of the cap-binding complex and eIFs and subsequent binding to 43 s. Only the most significantly enriched cancer-related pathways were shown in this figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

= 3.16) metabolism, also including DEPs involved in G-alpha I events (Z-score = 2.82) and opioid signaling (Z-score = 2.82) (Fig. 6A). The study of tumor metabolic alterations has emerged as a relevant research area once the metabolic alterations are related to tumor growth and drug sensitivity [56]. Typically, the LABC subtype is associated with a few metabolic alterations [57,58]. However, our result describes a set of proteomic alterations associated with the LABC metabolic landscape, with potential implications in therapy target selection.

The LBBC subtype showed alterations in the expression of TUBA4A, UBA3, and ARPC1B, for example, which were involved in processes such as signaling by hedgehog (*Z*-score = -2.64), ephrin signaling (Z-score = -2.82) and neddylation (Z-score = -2.64) (Fig. 6B). Both hedgehog and ephrin pathways regulate the proliferation and migration of breast tumor cells, representing potential therapeutic targets for patients with breast cancer [59,60]. ARPC1B was identified as a putative oncoprotein in our analysis, while TUBA4A is related to brain metastasis originated from breast primary tumors [61]: We suggest that these pathways and associated proteins can represent signatures of LBBC tumorigenesis.

As shown in Fig. 6C, the HER2+ subtype presented few particular alterations, including changes in tRNA processing (*Z*-score = -2.00) and apoptotic cleavage of cellular proteins (*Z*-score = -1.00), influenced by the down expression of RTRAF, RTCB, DDX1, and FAM98B. In HER2+

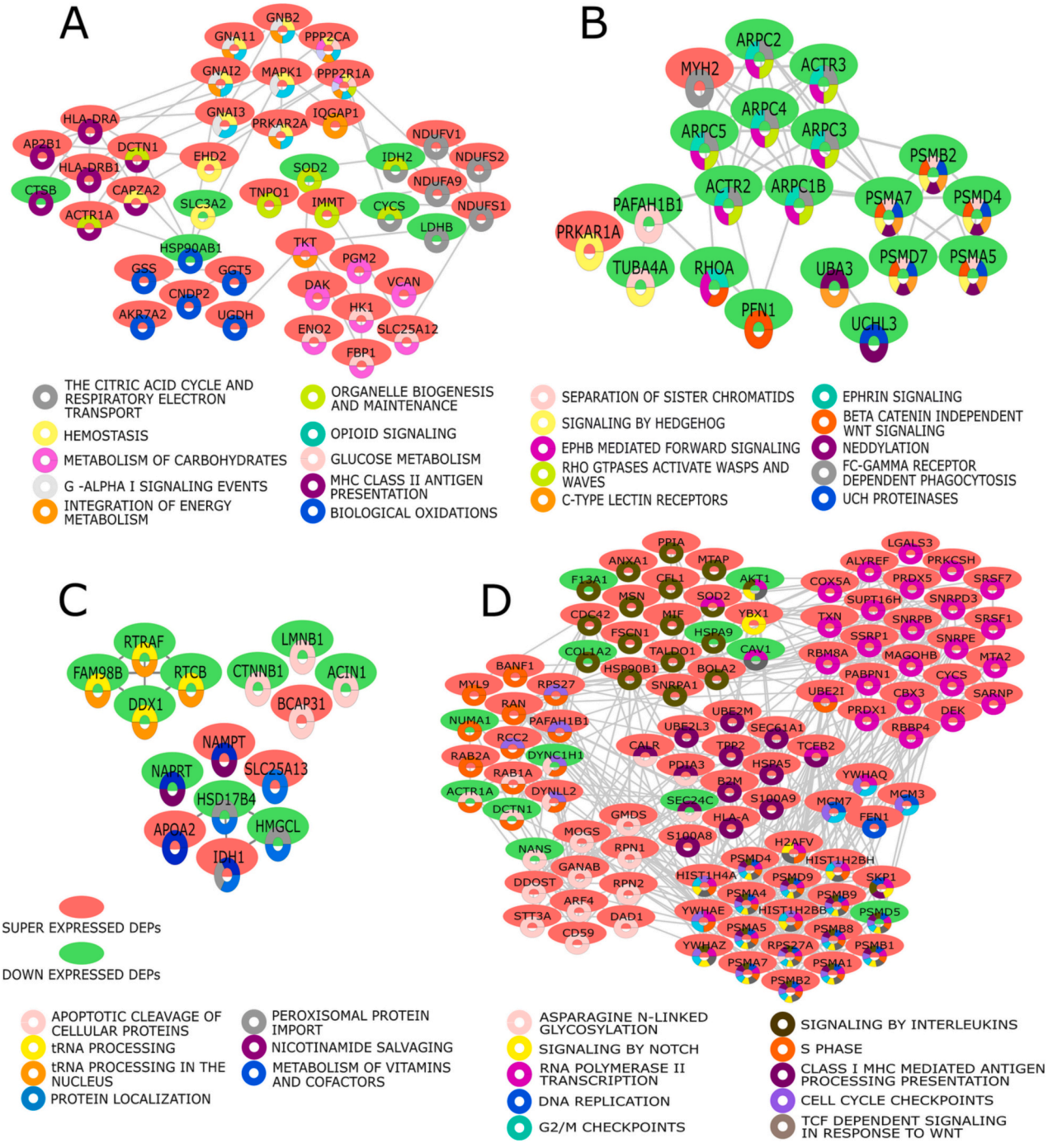

Fig. 6. Functional enrichment analysis of REACTOME pathways associated with specific subtypes. Pathways and processes significantly enriched (FDR <0.05) based on the DEPs of (A) LABC tumors, (B) LBBC tumors, (C) HER2+ tumors, and (D) TNBC tumors. Only the 10 most significant pathways and processes were shown in the figure. The edges represent protein-protein interactions computed by the STRING database (v.11.5).

cell lines, dysregulation in the expression of tRNA fragments is associated with trastuzumab resistance [62]. DDX1 and FAM98B are already associated with tRNA regulation in other cancer types [63], suggesting that further studies on the expression and function of both proteins may be relevant to understanding trastuzumab resistance in the HER2+ enriched breast tumors.

The vast dissimilarities that we found in the TNBC proteome and its related pathways reflect the high diversity of clinical and biological behaviors presented by this subtype compared to the others. There was a predominance of cell cycle-related processes such as G2/M checkpoints

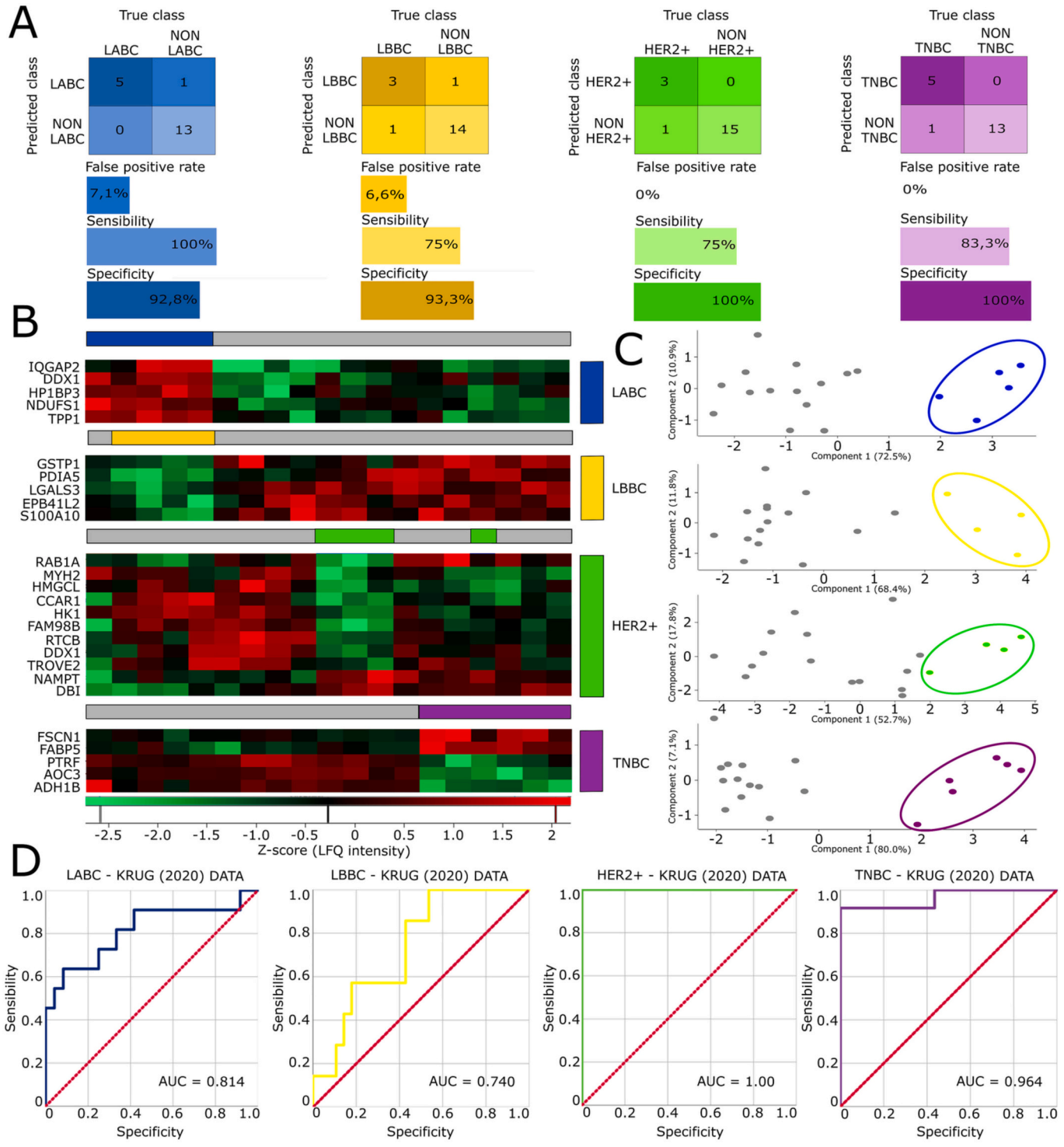

Fig. 7. Proteomic panels and their discriminative performance in the cross-validation and independent cohort validation of breast tumors. (A) Confusion matrices indicate the model's sensitivity, specificity, and false positive rate (B) Heatmaps of proteins expression of each panel. (C) PCA analysis and segregation of samples in subtypes according to each model. (D) ROC and AUC values achieved by each panel in the independent cohort. ROC = Receiver operating characteristic curve. AUC = Area under the curve. Blue = LABC samples. Yellow = LBBC samples. Green = HER2+ enriched samples. Purple: TNBC samples. AUC of 0.5 suggests no discrimination; 0.7 to 0.8 an acceptable performance; 0.8 to 0.9 an excellent performance, and >0.9 an outstanding discriminative performance [30]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Z-score = 4.02), DNA replication (Z-score = 3.50), and S phase (Z-score = 3.29), also including signaling pathways such as signaling by Notch (Z-score = 3.57), signaling by interleukins (Z-score = 3.77) and TCF dependent signaling in response to WNT (Z-score = 2.98), involving the over-expression of several DEPs, including MCM3, MCM7, H2AFY, YBX1 and DYNLL2 (Fig. 6D). These observations can help to clarify the functional proteomic landscape of TNBC, suggesting molecular signatures associated with the subtype.

In general, the enrichment results show that each subtype's proteome is related to its tumor's biological characteristics. Altogether, our observations join previous reports [2,10] to improve the understanding of the particularities of each subtype regarding its functional alterations in biological pathways and processes.

3.5. Construction and validation of proteomic panels to discriminate breast cancer subtypes

We employed a supervised machine learning approach based SVMs to construct predictive models with the potential to discriminate breast tumors of different subtypes. The 122 DEPs obtained from the ANOVA test were ranked for each subtype separately according to their discriminative power, and the top-ranked DEPs were selected according to the minimum number of DEPs necessary to guarantee models with <15% classification error. After, a cross-validation prediction was performed to determine the model's sensibility, specificity, and false-positive rate.

We obtained four models with at least at least 75% of sensibility and 92% of specificity (false positive rate < 7.2%; Fig. 7A). The protein panels of the LABC, LBBC, and TNBC subtypes were composed of five proteins each, while the HER2+ panel was composed of 11 proteins (Fig. 7B). As indicated by our previous results (section 3.1), the HER2+ proteome was the less distinguishable among the subtypes and presented the lower number of DEPs; concordantly, a higher number of DEPs was needed to discriminate the HER2+ samples from non-HER2+ tumors. The PCA analysis indicated that each model leads to a good segregation of tumors according to its subtypes (Fig. 7C).

To confirm the effectiveness of these models, we tested the panels in an independent cohort of proteomic data derived from MS/MS analysis of tumor solid tissues from breast cancer patients [12]. The classification performance of each panel was evaluated by ROC-AUC.

We could conclude that the panels of LABC (AUC = 0.814), HER2+ (AUC = 1.00), and TNBC (AUC = 0.964) subtypes achieved excellent to outstanding performances, and the panel of LBBC (AUC = 0.740) had an acceptable performance [30] when applied to the independent validation cohort (Fig. 7D). Differences in the LBBC samples could explain the inferior performance its panel in the independent cohort: In our cohort, the LBBC tumors were HER2-, while the validation cohort was composed of LBBC HER2+ tumors. Also, the perfect AUC value achieved by the HER2+ panel in the independent cohort (AUC = 1) suggests that further investigations in larger cohorts improve the proteomic panel for this subtype.

We demonstrated in our cohort and the independent cohort the applicability and reproducibility of these panels to discriminate breast cancer tumors and improve patient stratification. The advantages of panels composed of multiple biomarkers over single biomarker panels have already been verified [64,65]. Our panels have consistent or even better performances than previously proposed classification models [10,16,17], reinforcing the application of protein as biomarkers.

3.6. Evaluation of the model's performance at the transcriptomic level

We also explored the transcriptomic data of breast tumor samples from TCGA to evaluate if the expression of the proteins selected by the SVM approach were equivalent at the mRNA level. As shown in the Supplementary Fig. 1A-D, only eight out of 26 mRNAs presented an over/down-expression pattern in the subtype comparisons equivalent to its protein levels. In general, only a partial correspondence between mRNA and protein levels is found in homeostatic cells, with an even smaller correspondence observed in cancer cells [66]. Altogether, this indicates that an entire layer of information regarding cell functioning in cancer can be better accessed through proteomic data mining.

Regarding the performance of the models when using the mRNA data to classify the TCGA breast tumor samples (according to the PAM50 subtype classification), only the TNBC (AUC = 0.925) and HER2+ (AUC = 0.878) panels achieved excellent performances, with LABC (AUC = 0.748) and LBBC (AUC = 0.748) showing only acceptable performances. All models had better protein-level performances, highlighting its relevance and applicability to increase the accuracy of the breast cancer subtype diagnostic. It also shows that the use of proteins as biomarkers to breast cancer subtyping can represent a level of discrimination complementary to the transcriptomic, and that may be used in patient stratification and precision medicine, reinforcing the classical IHC and mRNA markers. We acknowledge that the differences in the IHC and PAM50 classification of the cohorts may influence the protein *vs.* mRNA performance comparisons.

4. Conclusions

Overall, by characterizing the differential proteome of different breast cancer subtypes through mass-spectrometry and downstream bioinformatics, we contributed to the understanding of the subtype's clinical and biological behaviors. Different pathways and biological alterations were associated with each subtype, as well as different expression patterns of oncoproteins and tumor suppressor proteins, providing new data for the characterization of the proteomic alterations involved in the malignancy of each subtype. Through a robust SVM-based machine learning approach, we also propose four new protein panels that achieved satisfactory performances to subtype discrimination. We suggest that these panels may be used to complement the IHC classification of breast subtypes; nevertheless, larger cohorts and validation approaches could be applied to confirm the clinical application of these panels. The relatively small sample set used in this research represents a limitation. However, we believe that our stringent eligibility criteria for sample selection and robust statistical analysis can mitigate this limitation and support the representativeness of our results.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2023.104955.

Ethics approval

This study was approved by the National Commission of Ethics in Research (CAAE 19870319.3.0000.0102).

Consent to participate

Signed informed consent was obtained from all participants included in the study.

Funding

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Araucaria Research Foundation of Paraná State (PRONEX/2012), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Process: 408730/2018-8.

CRediT authorship contribution statement

Alexandre Luiz Korte de Azevedo: Conceptualization, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Talita Helen Bombardelli Gomig:** Conceptualization, Writing – review & editing, Supervision. **Michel Batista:** Formal analysis, Data curation.

Fabricio Klerynton Marchini: Formal analysis, Data curation. **Ana Paula Martins Sebastião:** Data curation. **Jaqueline Carvalho de Oliveira:** Conceptualization, Writing – review & editing, Supervision. **Daniela Fiori Gradia:** Conceptualization, Writing – review & editing, Supervision. **Iglenir João Cavalli:** Conceptualization, Writing – review & editing, Supervision. **Enilze Maria de Souza Fonseca Ribeiro:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

No competing interests were declared.

Data availability

The mass-spectrometry datasets generated in this study are available on request to the corresponding author. Transcriptomic data were obtained online in public databases.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020, GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.

[2] C. Panis, L. Pizzatti, A.C. Herrera, S. Corrêa, R. Binato, E. Abdelhay, Label-free proteomic analysis of breast cancer molecular subtypes, J. Proteome Res. 13 (2014) 4752–4772, https://doi.org/10.1021/pr500676x.

[3] M. Zhang, A.V. Lee, J.M. Rosen, The cellular origin and evolution of breast cancer, Cold Spring Harb. Perspect. Med. 7 (2017), https://doi.org/10.1101/cshperspect.a027128.

[4] Charles M. Perou, Therese Sørlie, Michael B. Eisen, Matt van de Rijn, Stefanie S. Jeffreyk, Christian A. Rees, et al., Molecular portraits of human breast tumours, Lett. Nat. (2000) 747–752. https://doi.org/10.1038/35021093.

[5] A. Goldhirsch, E.P. Winer, A.S. Coates, R.D. Gelber, M. Piccart-Gebhart, B. Thürlimann, et al., Personalizing the treatment of women with early breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013, Ann. Oncol. 24 (2013) 2206–2223, https://doi.org/10.1093/annonc/mdt303.

[6] M. Balic, C. Thomssen, R. Würstlein, M. Gnant, N. Harbeck, St. Gallen/Vienna, A brief summary of the consensus discussion on the optimal primary breast cancer treatment, Breast Care. 14 (2019) (2019) 103–110, https://doi.org/10.1159/000499931.

[7] R.G. do Nascimento, K.M. Otoni, Histological and molecular classification of breast cancer: what do we know? Mastology. 30 (2020). https://doi.org/10.29289/2594539420202020024.

[8] M.C.U. Cheang, S.K. Chia, D. Voduc, D. Gao, S. Leung, J. Snider, et al., Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer, J. Natl. Cancer Inst. 101 (2009) 736–750, https://doi.org/10.1093/jnci/djp082.

[9] X. Dai, T. Li, Z. Bai, Y. Yang, X. Liu, J. Zhan, B. Shi, Breast cancer intrinsic subtype classification, clinical use and future trends, Am. J. Cancer Res. 5 (10) (2015) 2929–2943.

[10] S. Tyanova, R. Albrechtsen, P. Kronqvist, J. Cox, M. Mann, T. Geiger, Proteomic maps of breast cancer subtypes, Nat. Commun. 7 (2016), https://doi.org/10.1038/ncomms10259.

[11] G. Yanovich, H. Agmon, M. Harel, A. Sonnenblick, T. Peretz, T. Geiger, Clinical proteomics of breast cancer reveals a novel layer of breast cancer classification, Cancer Res. 78 (2018) 6001–6010, https://doi.org/10.1158/0008-5472.CAN-18-1079.

[12] K. Krug, E.J. Jaehnig, S. Satpathy, L. Blumenberg, A. Karpova, M. Anurag, G., et al., Proteogenomic landscape of breast Cancer tumorigenesis and targeted therapy, Cell. 183 (2020) 1436–1456.e31, https://doi.org/10.1016/j.cell.2020.10.036.

[13] S.L. Liang, D.W. Chan, Enzymes and related proteins as cancer biomarkers: a proteomic approach, Clin. Chim. Acta 381 (2007) 93–97, https://doi.org/10.1016/j.cca.2007.02.017.

[14] L.-H. Gam, Breast cancer and protein biomarkers, World J. Exp. Med. 2 (2012), https://doi.org/10.5493/wjem.v2.i5.86.

[15] T. Mukama, R.T. Fortner, V. Katzke, L.C. Hynes, A. Petrera, S.M. Hauck, et al., Prospective evaluation of 92 serum protein biomarkers for early detection of ovarian cancer, Br. J. Cancer 126 (2022) 1301–1309, https://doi.org/10.1038/s41416-021-01697-z.

[16] P. Bouchal, O.T. Schubert, J. Faktor, L. Capkova, H. Imrichova, K. Zoufalova, et al., Breast Cancer classification based on Proteotypes obtained by SWATH mass spectrometry, Cell Rep. 28 (2019) 832–843.e7, https://doi.org/10.1016/j.celrep.2019.06.046.

[17] T. Terkelsen, M. Pernemalm, P. Gromov, A.L. Børresen-Dale, A. Krogh, V. D. Haakensen, et al., High-throughput proteomics of breast cancer interstitial fluid: identification of tumor subtype-specific serologically relevant biomarkers, Mol. Oncol. 15 (2021) 429–461, https://doi.org/10.1002/1878-0261.12850.

[18] T.H.B. Gomig, I.J. Cavalli, R.L.R. de Souza, E. Vieira, A.C.R. Lucena, M. Batista, K. C., et al., Quantitative label-free mass spectrometry using contralateral and adjacent breast tissues reveal differentially expressed proteins and their predicted impacts on pathways and cellular functions in breast cancer, J. Proteome 199 (2019) 1–14, https://doi.org/10.1016/j.jprot.2019.02.007.

[19] J.R. Wiśniewski, F.Z. Gaugaz, Fast and sensitive total protein and peptide assays for proteomic analysis, Anal. Chem. 87 (2015) 4110–4116, https://doi.org/10.1021/ac504689z.

[20] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote) omics data, Nat. Methods 13 (2016) 731–740, https://doi.org/10.1038/nmeth.3901.

[21] S. Tyanova, J. Cox, Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research, in: Methods in Molecular Biology, Humana Press Inc., 2018, pp. 133–148, https://doi.org/10.1007/978-1-4939-7493-1_7.

[22] L. Dressler, M. Bortolomeazzi, M.R. Keddar, H. Misetic, G. Sartini, A. Acha-Sagredo, et al., Comparative assessment of genes driving cancer and somatic evolution in non-cancer tissues: an update of the network of Cancer genes (NCG) resource, Genome Biol. 23 (2022), https://doi.org/10.1186/s13059-022-02607-z.

[23] J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, et al., The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (2013) 1113–1120, https://doi.org/10.1038/ng.2764.

[24] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. www.pnas.orgcgidoi10.1073pnas.0506580102, 2005.

[25] W. Walter, F. Sánchez-Cabo, M. Ricote, GOplot: an R package for visually combining expression data with functional analysis, Bioinformatics. 31 (2015) 2912–2914, https://doi.org/10.1093/bioinformatics/btv300.

[26] L.J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, et al., STRING 8 - a global view on proteins and their functional interactions in 630 organisms, Nucleic Acids Res. 37 (2009), https://doi.org/10.1093/nar/gkn760.

[27] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[28] M. Minghui, Z. Chuanfeng, Application of support vector machines to a small-sample prediction, Adv. Petrol. Explorat. Develop. 10 (2015) 72–75, https://doi.org/10.3968/7830.

[29] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine learning algorithm validation with a limited sample size, PLoS One 14 (2019), https://doi.org/10.1371/journal.pone.0224365.

[30] J.N. Mandrekar, Receiver Operating Characteristic Curve in Diagnostic Test Assessment, 2010.

[31] S.V. Vasaikar, P. Straub, J. Wang, B. Zhang, LinkedOmics: analyzing multi-omics data within and across 32 cancer types, Nucleic Acids Res. 46 (2018) D956–D963, https://doi.org/10.1093/nar/gkx1090.

[32] K. Asleh, G.L. Negri, S.E. Spencer Miko, S. Colborne, C.S. Hughes, X.Q. Wang, et al., Proteomic analysis of archival breast cancer clinical specimens identifies biological subtypes with distinct survival outcomes, Nat. Commun. 13 (2022), https://doi.org/10.1038/s41467-022-28524-0.

[33] J.M.A. Moreira, G. Ohlsson, F.E. Rank, J.E. Celis, Down-regulation of the tumor suppressor protein 14-3-3σ is a sporadic event in cancer of the breast, Mol. Cell. Proteomics 4 (2005) 555–569, https://doi.org/10.1074/mcp.M400205-MCP200.

[34] C. Chen, L.G. Zhang, J. Liu, H. Han, N. Chen, A.L. Yao, et al., Bioinformatics analysis of differentially expressed proteins in prostate cancer based on proteomics data, Onco. Targets Ther. 9 (2016) 1545–1557, https://doi.org/10.2147/OTT.S98807.

[35] K. Jacquet, O. Binda, ING proteins: tumour suppressors or oncoproteins, Cancers (Basel). 13 (2021), https://doi.org/10.3390/cancers13092110.

[36] V. Vafeiadou, D. Hany, D. Picard, Hyperactivation of MAPK induces tamoxifen resistance in SPRED2-deficient ERα-positive breast Cancer, Cancers (Basel). 14 (2022), https://doi.org/10.3390/cancers14040954.

[37] L. Zheng, H. Dai, M. Zhou, M. Li, P. Singh, J. Qiu, et al., Fen1 mutations result in autoimmunity, chronic inflammation and cancers, Nat. Med. 13 (2007) 812–819, https://doi.org/10.1038/nm1599.

[38] J.E. Belizario, A.F. Loggulo, Insights into breast cancer phenotying through molecular omics approaches and therapy response, Cancer Drug Resistan. 2 (2019) 527–538, https://doi.org/10.20517/cdr.2018.009.

[39] S. Rao, S.Y. Lee, A. Gutierrez, J. Perrigoue, R.J. Thapa, Z. Tu, et al., Inactivation of ribosomal protein L22 promotes transformation by induction of the stemness factor, Lin28B, Blood. 120 (2012) 3764–3773, https://doi.org/10.1182/blood-2012-03-415349.

[40] B. Cao, Z. Fang, P. Liao, X. Zhou, J. Xiong, S. Zeng, H. Lu, Cancer-Mutated Ribosome Protein L22 (RPL22/eL22) Suppresses cancer Cell Survival by Blocking p53-MDM2 Circuit. www.impactjournals.com/oncotarget, 2017.

[41] X. Liu, Q. Chen, X.G. Hu, X.C. Zhang, T.W. Fu, Q. Liu, et al., PTP1B promotes aggressiveness of breast cancer cells by regulating PTEN but not EMT, Tumor Biol. 37 (2016) 13479–13487, https://doi.org/10.1007/s13277-016-5245-1.

[42] M. Yu, Z. Liu, Y. Liu, X. Zhou, F. Sun, Y. Liu, et al., PTP1B markedly promotes breast cancer progression and is regulated by miR-193a-3p, FEBS J. 286 (2019) 1136–1153, https://doi.org/10.1111/febs.14724.

[43] Y. Zhu, M. Guo, L. Zhang, T. Xu, L. Wang, G. Xu, Biomarker triplet NAMPT/VEGF/HER2 as a de novo detection panel for the diagnosis and prognosis of human breast cancer, Oncol. Rep. 35 (2016) 454–462, https://doi.org/10.3892/or.2015.4391.

[44] M.D.P. Chantada-Vázquez, A.C. López, M. García-Vence, B. Acea-Nebril, S. B. Bravo, C. Núñez, Protein corona gold nanoparticles fingerprinting reveals a profile of blood coagulation proteins in the serum of her2-overexpressing breast

cancer patients, Int. J. Mol. Sci. 21 (2020) 1–18, https://doi.org/10.3390/ijms21228449.
[45] I.J. Suárez-Arroyo, Y.R. Feliz-Mosquea, J. Pérez-Laspiur, R. Arju, S. Giashuddin, G. Maldonado-Martínez, et al., The proteome signature of the inflammatory breast cancer plasma membrane identifies novel molecular markers of disease, Am. J. Cancer Res. 6 (8) (2016) 1720–1740.
[46] D.K. Singh, O. Gholamalamdari, M. Jadaliha, X.L. Li, Y.C. Lin, et al., PSIP1/p75 promotes tumorigenicity in breast cancer cells by promoting the transcription of cell cycle genes, Carcinogenesis. 38 (2017) 966–975, https://doi.org/10.1093/carcin/bgx062.
[47] S.L. Chen, S.X. Lu, L.L. Liu, C.H. Wang, X. Yang, Z. Yi Zhang, et al., eEF1A1 overexpression enhances tumor progression and indicates poor prognosis in hepatocellular carcinoma, Transl. Oncol. 11 (2018) 125–131, https://doi.org/10.1016/j.tranon.2017.11.001.
[48] A. Bosutti, B. Dapas, G. Grassi, R. Bussani, F. Zanconati, F. Giudici, et al., High eEF1A1 protein levels mark aggressive prostate cancers and the in vitro targeting of eEF1A1 reveals the eEF1A1–actin complex as a new potential target for therapy, Int. J. Mol. Sci. 23 (2022), https://doi.org/10.3390/ijms23084143.
[49] H. Liu, C.J. Murphy, F.A. Karreth, K.B. Emdal, K. Yang, F.M. White, et al., Identifying and targeting sporadic oncogenic genetic aberrations in mouse models of triple-negative breast cancer, Cancer Discov. 8 (2018) 354–369, https://doi.org/10.1158/2159-8290.CD-17-0679.
[50] H. Liu, Y. Xing, S. Yang, D. Tian, Remarkable difference of somatic mutation patterns between oncogenes and tumor suppressor genes, Oncol. Rep. 26 (2011) 1539–1546, https://doi.org/10.3892/or.2011.1443.
[51] G. Schneider, M. Schmidt-Supprian, R. Rad, D. Saur, Tissue-specific tumorigenesis: context matters, Nat. Rev. Cancer 17 (2017) 239–253, https://doi.org/10.1038/nrc.2017.5.
[52] Z. Yi, F. Ma, C. Li, R. Chen, L. Yuan, X. Sun, et al., Landscape of somatic mutations in different subtypes of advanced breast cancer with circulating tumor DNA analysis, Sci. Rep. 7 (2017), https://doi.org/10.1038/s41598-017-06327-4.
[53] H. Xie, W. Wang, F. Sun, K. Deng, X. Lu, H. Liu, et al., Proteomics analysis to reveal biological pathways and predictive proteins in the survival of high-grade serous ovarian cancer, Sci. Rep. 7 (2017), https://doi.org/10.1038/s41598-017-10559-9.
[54] S. Romera-Giner, Z. Andreu Martínez, F. García-García, M.R. Hidalgo, Common pathways and functional profiles reveal underlying patterns in breast, kidney and lung cancers, Biol. Direct 16 (2021), https://doi.org/10.1186/s13062-021-00293-8.
[55] H.Y.K. Yip, A. Papa, Signaling pathways in cancer: therapeutic targets, combinatorial treatments, and new developments, Cells. 10 (2021) 1–30, https://doi.org/10.3390/cells10030659.
[56] N. Hammoudi, K.B.R. Ahmed, C. Garcia-Prieto, P. Huang, Metabolic alterations in cancer cells and therapeutic implications, Chin, J. Cancer 30 (2011) 508–525, https://doi.org/10.5732/cjc.011.10267.
[57] V. Cappelletti, E. Iorio, P. Miodini, M. Silvestri, M. Dugo, M.G. Daidone, Metabolic footprints and molecular subtypes in breast Cancer, Dis. Markers 2017 (2017), https://doi.org/10.1155/2017/7687851.
[58] E.A. Serrano-Carbajal, J. Espinal-Enríquez, E. Hernández-Lemus, Targeting metabolic deregulation landscapes in breast Cancer subtypes, Front. Oncol. 10 (2020), https://doi.org/10.3389/fonc.2020.00097.
[59] L. Song, W. Wang, D. Liu, Y. Zhao, J. He, X. Wang, et al., Targeting of sonic hedgehog-Gli signaling: a potential therapeutic target for patients with breast cancer, Oncol. Lett. 12 (2016) 1027–1033, https://doi.org/10.3892/ol.2016.4722.
[60] Z. Magic, J. Sandström, G. Perez-Tenorio, Ephrin-B2 inhibits cell proliferation and motility in vitro and predicts longer metastasis-free survival in breast cancer, Int. J. Oncol. 55 (2019) 1275–1286, https://doi.org/10.3892/ijo.2019.4892.
[61] L. Zhang, M. Fan, F. Napolitano, X. Gao, Y. Xu, L. Li, Transcriptomic analysis identifies organ-specific metastasis genes and pathways across different primary sites, J. Transl. Med. 19 (2021), https://doi.org/10.1186/s12967-020-02696-z.
[62] C. Sun, F. Yang, Y. Zhang, J. Chu, J. Wang, Y. Wang, et al., TRNA-derived fragments as novel predictive biomarkers for Trastuzumab-resistant breast Cancer, Cell. Physiol. Biochem. 49 (2018) 419–431, https://doi.org/10.1159/000492977.
[63] A. Pérez-González, A. Pazo, R. Navajas, S. Ciordia, A. Rodriguez-Frandsen, A. Nieto, HCLE/C14orf166 associates with DDX1-HSPC117-FAM98B in a novel transcription-dependent shuttling RNATransporting complex, PLoS One 9 (2014), https://doi.org/10.1371/journal.pone.0090957.
[64] C. Han, S. Bellone, E.R. Siegel, G. Altwerger, G. Menderes, E. Bonazzoli, et al., A novel multiple biomarker panel for the early detection of high-grade serous ovarian carcinoma, Gynecol. Oncol. 149 (2018) 585–591, https://doi.org/10.1016/j.ygyno.2018.03.050.
[65] T. Muinao, H.P. Deka Boruah, M. Pal, Multi-biomarker panel signature as the key to diagnosis of ovarian cancer, Heliyon. 5 (2019), https://doi.org/10.1016/j.heliyon.2019.e02826.
[66] G. Andrieux, S. Chakraborty, T. Das, M. Boerries, Alteration of Proteotranscriptomic landscape reveals the transcriptional regulatory circuits controlling key-signaling pathways and metabolic reprogramming during tumor evolution, Front. Cell Dev. Biol. 8 (2020), https://doi.org/10.3389/fcell.2020.586479.