

# Proteomic Characterization of Serum Small Extracellular Vesicles in Human Breast Cancer

4 Ganfei Xu<sup>1,2,†</sup>, Weiyi Huang<sup>1,†</sup>, Shaoqian Du<sup>1,†</sup>, Minjing Huang<sup>2,†</sup>, Jiacheng  
5 Lyu<sup>2</sup>, Fei Zhou<sup>1</sup>, Rongxuan Zhu<sup>1</sup>, Yuan Cao<sup>1</sup>, Jingxuan Xv<sup>1</sup>, Ning Li<sup>1</sup>, Guoying  
6 Yu<sup>3</sup>, Binghua Jiang<sup>4</sup>, Olivier Gires<sup>5</sup>, Lei Zhou<sup>6</sup>, Hongwei Zhang<sup>7,\*</sup>, Chen  
7 Ding<sup>2,8,\*\*</sup>, Hongxia Wang<sup>1,\*\*\*</sup>

<sup>1</sup> State Key Laboratory of Oncogenes and Related Genes, Department of Oncology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

<sup>2</sup> State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Institute of Biomedical Sciences, Human Phenome Institute, Fudan University, Shanghai 200433, China

<sup>3</sup> State Key Laboratory of Cell Differentiation and Regulation, Henan International Joint Laboratory of Pulmonary Fibrosis, Henan Center for Outstanding Overseas Scientists of Pulmonary Fibrosis, College of Life Science, Institute of Biomedical Science, Henan Normal University, Xinxiang 453007, China

<sup>4</sup> Academy of Medical Science, Zhengzhou University, Zhengzhou 450001, China.

<sup>5</sup> Institute of Molecular Medicine, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

<sup>6</sup> Singapore Eye Research Institute, The Academia, 20 College Road, Discovery Tower  
Level 6, Singapore 169856, Singapore

<sup>7</sup> Department of General Surgery, Zhongshan Hospital, Fudan University. 180 Fenglin Road, Shanghai 200032

<sup>8</sup> Department of Anatomy and Neuroscience Research Institute, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou 450001, China

28 † These authors contributed equally to this work

29 \* Corresponding authors. Tel: 021-63240090; Email: [zhang.hongwei@zs-hospital.sh.cn](mailto:zhang.hongwei@zs-hospital.sh.cn)

30 \*\* Corresponding authors. Tel: 021-31246742; Email: [chend@fudan.edu.cn](mailto:chend@fudan.edu.cn)

31 \*\*\* Corresponding authors. Tel: 021-63240090; Email: [whx365@126.com](mailto:whx365@126.com)

32

33 **Key Words:** Breast cancer; Small extracellular vesicles; Lymph node  
34 metastases; Distant metastasis; Proteomics

35

36 **Abstract**

37 There is a lack of comprehensive understanding of breast cancer (BC) specific  
38 sEVs characteristics and composition on BC unique proteomic information  
39 from human samples. Here, we interrogated the proteomic landscape of sEVs  
40 in 167 serum samples from patients with BC, benign mammary disease (BD)  
41 and from healthy donors (HD). The analysis provides a comprehensive  
42 landscape of serum sEVs with totally 9,589 proteins identified, considerably  
43 expanding the panel of sEVs markers. Of note, serum BC-sEVs protein  
44 signatures were distinct from those of BD and HD, representing stage- and  
45 molecular subtype-specific patterns. We constructed specific sEVs protein  
46 identifiers that could serve as a liquid biopsy tool for diagnosis and  
47 classification of BC from benign mammary disease, molecular subtypes, as  
48 well as assessment of lymph node metastasis. We also identified 11 potential  
49 survival biomarkers for distant metastasis. This work may provide reference  
50 value for the accurate diagnosis and monitoring of BC progression using  
51 serum sEVs.

52

53 **Introduction**

54 Breast cancer (BC) is one of the most common cancers worldwide and  
55 accounts for 30% of female cancers (Kim *et al*, 2012; Liu *et al*, 2021; Siegel *et*  
56 *al*, 2021). A long-term decline in the death rate has been observed since the  
57 mid-1970s due to improvements in treatment protocols, including the  
58 development of chemotherapy, immunotherapy and targeted therapies.  
59 However, improvements in clinical outcomes have slowed over the past  
60 decade, and distant metastasis remains the major cause of mortality (Cassetta  
61 & Pollard, 2017; Liu *et al*, 2019a; Siegel *et al.*, 2021; Yin *et al*, 2014; Zhu *et al*,  
62 2019). The early detection and dynamic assessment of the metastatic status of  
63 BC patients are of great value for the treatment and longitudinal analysis of  
64 cancer evolution in response to therapy. To achieve this, liquid biopsies  
65 utilizing molecular classifiers detected in blood from patients, such as  
66 circulating tumor cells, circulating free DNA, and exosomes, offer minimal  
67 invasiveness, fewer complications, and an increased ability for longitudinal  
68 monitoring compared with traditional tumor tissue biopsies (Wan *et al*, 2017;  
69 Yoneda *et al*, 2019). More importantly, liquid biopsy is more informative than  
70 single locally restricted biopsies, providing unique information about tumor  
71 heterogeneity, clonal evolution, and the potential development of  
72 premetastatic cancer cells (Hoshino *et al*, 2020).

73 Circulating small extracellular vesicles (sEVs), such as exosomes or  
74 exosome-like vesicles (ELVs), are 30-150 nm in size and carry a restricted set  
75 of nucleic acids, lipids, and proteins (Balaj *et al*, 2011; Johnstone *et al*, 1987;  
76 Kim *et al*, 2013; Peinado *et al*, 2011; Raposo & Stoorvogel, 2013; Skog *et al*,  
77 2008; Thakur *et al*, 2014; Thery *et al*, 2009; Valadi *et al*, 2007; Wang & Gires,  
78 2019) that contribute to intercellular communication in normal physiology and  
79 pathology (Johnstone *et al.*, 1987; Maas *et al*, 2017; Skog *et al.*, 2008;  
80 Yanez-Mo *et al*, 2015). The functional importance of sEVs has been

81 intensively studied in multiple human cancers, including BC (Hoshino *et al.*,  
82 2020). Increasing evidence suggests that sEVs are actively released from  
83 cancer cells and markedly affect the tumor microenvironment (TME) as well as  
84 the immune ecosystem (Huber *et al.*, 2005), thereby constructing distant  
85 metastatic niches and facilitating cancer growth (Fang *et al.*, 2018; Kralj-Iglic,  
86 2012; Ozer *et al.*, 2020) and metastasis (Chen *et al.*, 2018; Costa-Silva *et al.*,  
87 2015; Hoshino *et al.*, 2015; Peinado *et al.*, 2012; Zhang & Wang, 2015). Of note,  
88 the membrane encapsulation of sEVs promotes their structural integrity, and  
89 cargos located within sEVs are more stable than other serological proteins  
90 since they have protection against degradation by circulating proteases and  
91 other enzymes (Li *et al.*, 2017a). Considering their facilitated retrieval and their  
92 relatively ubiquitous presence and abundance in serum, sEVs can provide  
93 ample materials for downstream analysis in BC detection, prognosis, and  
94 therapeutic monitoring as a promising, noninvasive liquid biopsy approach  
95 (Choi *et al.*, 2021; Lee *et al.*, 2018; Li *et al.*, 2017b; Wang *et al.*, 2018). For  
96 instance, Peinado *et al.* showed that an “sEV protein signature” could identify  
97 melanoma patients at risk for metastasis to nonspecific distant sites (Peinado  
98 *et al.*, 2012). Hoshino *et al.* identified a specific repertoire of integrins  
99 expressed on cancer-derived sEVs, which were distinct from cancer cells, that  
100 dictated exosome adhesion to specific cell types and ECM molecules in  
101 particular organs (Hoshino *et al.*, 2015).

102 The sEV proteome has been proposed to offer unique advantages as an  
103 informative readout for the detection and stratification of BC (Rontogianni *et al.*,  
104 2019). Nonetheless, the challenge is to optimize a proteomic profiling  
105 approach for sEVs to define and standardize reliable methods. Despite the  
106 availability of several public sEV protein databases (e.g., Vesiclepedia  
107 ([www.microvesicles.org/](http://www.microvesicles.org/)) (Kalra *et al.*, 2012), EVpedia ([www.evpedia.info](http://www.evpedia.info))  
108 (Kim *et al.*, 2012) and ExoCarta ([www.exocarta.org](http://www.exocarta.org)) (Kim *et al.*, 2013)), much

109 remains unknown about the sEV proteomes of BC. This includes the definition  
110 of (1) markers to distinguish BC from benign disease and healthy state, (2)  
111 markers to distinguish diverse molecular subtypes of invasive breast cancer  
112 (IBC), (3) markers to predict lymph node (LN) metastases, and (4) the open  
113 question of whether molecules present on IBC-derived sEVs are “addressing”  
114 them to specific organs. These unresolved problems highlight the need for a  
115 better understanding of the protein composition of BC-derived sEVs that could  
116 qualify them as biomarkers for clinical application, with a specificity and  
117 sensitivity mostly superior to those of traditional serum markers. To address  
118 these aims, mass spectrometry-based proteomic profiling is emerging as a  
119 strategy to gain insight into the biological cargos, functions, and clinical  
120 potential of sEVs ([Wang et al, 2020](#)).

121 Here, we applied a mass spectrometry-based, data-independent acquisition  
122 (DIA) quantitative approach to determine the proteomic features of human  
123 serum sEVs derived from patients with BC, benign mammary disease (BD),  
124 and healthy donors (HDs). In total, we identified 9,589 proteins from 167  
125 analyzed samples with a mean of 1,695 proteins quantified per sEV sample.  
126 Classification of the pathways related to the enriched proteins revealed that  
127 proteins preferentially packaged in BC-sEVs correlated with interferon  
128 γ-mediated signaling as well as pathways associated with immune response  
129 regulation, antigen processing and presentation, glycolysis and angiogenesis.  
130 By examining the sEV proteomes, we constructed specific sEV protein  
131 identifiers that could serve as a liquid biopsy tool for the diagnosis and  
132 classification of BC from BD and its molecular subtypes, as well as the  
133 assessment of LN metastasis. Of note, we found that adipocytes play an  
134 important role in the LN metastasis of BC. We also identified 11 potential  
135 survival markers for distant BC metastasis and 2 potential survival markers for  
136 lung metastasis. This work may provide reference value for the accurate

137 diagnosis and monitoring of BC progression using serum sEVs, and the  
138 identification of novel molecules packaged in sEVs offers an opportunity for  
139 the targeted therapy of BC in the future.

140

## 141 **Results**

### 142 **Proteomic characterization of BC-derived sEVs**

143 To elucidate the proteomic profile of BC-derived sEVs, we purified sEVs from  
144 167 human serum samples derived from BC patients ( $n = 126$ ), BD patients ( $n$   
145 = 17), and HDs ( $n = 24$ ) by differential ultracentrifugation as described in the  
146 Methods and in accordance with previously reported protocols (Colombo *et al*,  
147 2014; Peinado *et al*, 2012; Xu *et al*, 2016) (Fig 1A and C). All samples were  
148 collected prospectively from treatment-naive stage I-IV BC patients (Fig 1B,  
149 Appendix Table 1). Under transmission electron microscopy (TEM) in  
150 combination with nanoparticle tracking analysis (NTA), the isolated sEVs  
151 appeared as morphologically uniform vesicular structures 30-150 nm in size  
152 surrounded by a double-layer membrane (Fig 1D, Appendix Fig S1A). sEV  
153 samples were verified by immunoblotting analyses using the conventional  
154 markers CD9, CD63, TSG101, and ALIX, while we examined 24 sEV markers  
155 in our proteomics data (Hoshino *et al*, 2020) (Fig 1E, Appendix Fig S1B).  
156 Clinical data, including sex, age at diagnosis, tumor staging, BC subtypes, LN  
157 status, distant metastasis, and survival, are summarized in Fig 1B and Table  
158 S1.

159 A proteomic database of serum sEVs was constructed using label-free  
160 LC-MS/MS analysis, identifying 9,589 proteins in total from the 167 analyzed  
161 samples at a protein- and peptide-level FDR of less than 5% (Fig 1F). The  
162 protein abundance was first calculated by iBAQ and then normalized as FOT,  
163 allowing for comparison among different experiments. The mean number of  
164 proteins detected per sEV sample was 1,695 (range 793 to 2,253 proteins)

165 (Fig 1G). In general, 1,924, 187, and 145 unique sEV proteins were identified  
166 in BC, BD, and HD samples, respectively (Fig 1H). Globally, the dynamic  
167 range of proteins detected spanned eight orders of magnitude (Appendix Fig  
168 S1C). Collectively, these data were consistent with previous reports that sEV  
169 protein profiles differ significantly depending on the sample source (Wu *et al*,  
170 2019), and sEVs released by BC cells and from other cancer cells may carry  
171 more encapsulated cargos for signal transfer to induce the malignant  
172 transformation and proliferation of recipient cells (Milane *et al*, 2015).

173

174 **BC-derived sEVs exhibited specific signatures related to immune  
175 response, metabolism, and metastasis**

176 Next, proteomic data were analyzed to determine the characteristics of  
177 BC-derived sEVs. PCA demonstrated a clear distinction among the three  
178 different types of samples, which further highlighted the diverse proteomic  
179 patterns among BC-, BD-, and HD-sEVs that underpinned our stratification  
180 analysis (Appendix Fig S2A).

181 To decipher the protein network associated with BC tumorigenesis, we  
182 identified 287, 602, and 112 proteins that were significantly overrepresented in  
183 the BC ( $BC_{mean}/BD_{mean} > 2$ -fold and  $BC_{mean}/HD_{mean} > 2$ -fold), BD  
184 ( $BD_{mean}/BC_{mean} > 2$ -fold and  $BD_{mean}/HD_{mean} > 2$ -fold) and HD ( $HD_{mean}/BC_{mean} >$   
185  $2$ -fold and  $HD_{mean}/BD_{mean} > 2$ -fold) samples, respectively (see Materials and  
186 Methods). Clustering and cluster-specific enrichment analyses of these  
187 proteins using GOBP and Reactome pathway annotations showed that these  
188 differentially enriched proteins were involved in distinctive biological processes  
189 and pathways (Fig 2A, Appendix Table 2). Specifically, COPI-mediated  
190 anterograde transport (Fisher's exact test,  $p = 3.88e-3$ ), vesicle-mediated  
191 transport (Fisher's exact test,  $p = 1.26e-4$ ), and regulation of actin dynamics  
192 for phagocytic cup formation-related proteins (Fisher's exact test,  $p = 2.48e-5$ )

193 (i.e., ADD2, ARF5, ARPC1A, IGHV3-53, IGHV4-39, SSC5D, and COPE) were  
194 enriched in HD samples (Fig 2A and B, Appendix Table 2). BD-sEVs were  
195 characterized by proteins related to cell-cell adhesion (Fisher's exact test,  $p =$   
196 2.98e-19) (i.e., STAT1, PTPN1, RPL24, and FNBP1L), cholesterol metabolic  
197 process (Fisher's exact test,  $p = 3.34e-7$ ) (i.e., PON1, APOC1, APOA2,  
198 ANGPTL3, and LIPC), and response to estrogen (Fisher's exact test,  $p =$   
199 2.13e-2) (i.e., F7, LDHA, HSP90AA1, IGFBP2, and CTNNA1) (Fig 2A and B,  
200 Appendix Table 2). Of note, BC-sEVs exhibited specific signatures related to  
201 the immune response, metabolism, and metastasis, potentially reflecting the  
202 functional roles and molecular heterogeneity of sEVs during BC tumorigenesis  
203 and progression. Classification of the pathways related to the enriched  
204 proteins from BC-sEVs revealed that these selectively packaged proteins are  
205 involved in the interferon  $\gamma$ -mediated signaling pathway (Fisher's exact test,  $p$   
206 = 4.94e-4) (i.e., HCK, HLA-H, HLA-B, HLA-C, HLA-A, HLA-G, and CD44),  
207 regulation of immune response (Fisher's exact test,  $p = 4.61e-5$ ) (i.e.,  
208 IGLV3-25, COL3A1, CXADR, IGLV3-27, HLA-A, IGLV7-43, and PVR), antigen  
209 processing and presentation (Fisher's exact test,  $p = 1.16e-5$ ) (i.e., ITGB1,  
210 IGLV3-25, CXADR, IGLV3-27, IGLV7-43, PVR, and HLA-G), glycolytic  
211 process (Fisher's exact test,  $p = 1.29e-3$ ) (i.e., GPI, PGK1, PGAM4, PGK2,  
212 and PGM1), and angiogenesis (Fisher's exact test,  $p = 3.92e-2$ ) (i.e., GPI,  
213 RNF213, ANGPTL6, MMP2, PECAM1, CYP1B1, NAA15, and TYMP) (FC > 2,  
214 one-way ANOVA  $p < 0.05$ ) (Fig 2A and B, Appendix Table 2). Notably, in the  
215 Tang et al. BC cohort (Tang et al, 2018), among sEV proteins that were  
216 specifically highly expressed in BC samples, patients with high expression of  
217 MMP2 and TYMP appeared to have poor prognostic outcomes (log rank test,  $p$   
218 < 0.05) (Fig 2C). These findings that BC-, BD-, and HD-sEV cargos are distinct  
219 and related to singular cellular processes suggest that sEV protein packaging  
220 into sEVs is heterogeneous and reflects BC biology.

221

222 **Specific damage-associated molecular pattern (DAMP) molecules are**  
223 **packaged in BC-derived sEVs**

224 Recent advances have indicated that DAMP molecules, such as nucleic acids,  
225 histones, high mobility group box 1, S100, and heat shock proteins, act as  
226 endogenous ligands of innate immune receptors and are linked to the immune  
227 response and cancer progression (Becker *et al*, 2016). In total, we identified  
228 210 different DAMPs in all sEV datasets (Appendix Fig S2B, Appendix Table  
229 2). Specifically, the analysis identified 197, 145, and 157 DAMPs in BC-, BD-,  
230 and HD-sEVs, respectively, suggesting that more DAMPs were enriched in BC  
231 samples than in BD and HD samples (Appendix Fig S2B). Thirty-two of these  
232 DAMPs were identified only in BC-sEVs, 9 DAMPs only in BD-sEVs, and 4  
233 DAMPs only in HD-sEVs (Appendix Fig S2B). Of all DAMPs identified in  
234 BC-sEVs, 27 DAMPs (e.g., ACAN, ANXA11, and CD44) were shared by > 50%  
235 of BC samples and were enriched compared to BD-sEVs and/or HD-sEVs (Fig  
236 2D).

237 Among them, 9 DAMPs, including aggrecan (ACAN), annexin A11  
238 (ANXA11), CD44, fibrinogen gamma chain (FGG), integrin-linked kinase (ILK),  
239 LGALS3, and several ITGs (ITGA6, ITGB1, and ITGB3), were exclusively  
240 present in BC-sEVs versus BD- and HD-sEVs, suggesting that they are  
241 specific sEV markers in BC development and progression (Fig 2D). ITGA6,  
242 ITGB1, and ITGB3 are members of the integrin family of proteins involved in  
243 cell adhesion and recognition in a variety of processes, including tissue repair,  
244 hemostasis, immune response, and metastatic dissemination of cancer cells  
245 (Laudato *et al*, 2017; Wang *et al*, 2019b). ANXA11 and LGALS3 are  
246 associated with the progression of some cancers (Liu *et al*, 2019b; Wang *et al*,  
247 2019a). Another 10 DAMP proteins were highly enriched in both BC- and  
248 BD-sEVs: ANXA1, ANXA2P2, ANXA4, CD14, HABP2, LGALS4, LUM, OMD,

249 S100A9, and TXNDC5, whereas they were rarely detected in HD samples,  
250 suggesting that they represent sEV DAMPs shared across BC and BD (Fig 2E).  
251 Interestingly, our analyses revealed that 8 DAMP molecules (A2M, ANXA5,  
252 CALR, FGB, IL6ST, LGALS3BP, LYVE1, and S100A7) were abundantly  
253 expressed in both BC- and HD-sEVs (Fig 2E). This finding is consistent with  
254 previous studies reporting that the noncancer-derived sEV proteome is as  
255 informative as the cancer-derived sEV proteome in specific cancer types  
256 (Hoshino *et al.*, 2020). It is worth noting that 6 of these molecules (ANXA11,  
257 ILK, ITGA6, ITGB1, LGALS3, and OMD) were highly expressed in BC and  
258 were associated with poor prognosis in the Tang *et al.* BC cohort (Tang *et al.*,  
259 2018) (Fig 2E, Appendix Fig S2C).

260

## 261 **Possible intercellular communication network diagram of BC-driven 262 sEVs in the TME**

263 Previous evidence suggests that sEVs interact with recipient immune cells to  
264 participate in TME remodeling, an effect that is mediated by encapsulated  
265 molecular cargos derived from parent cancer cells (Becker *et al.*, 2016). Thus,  
266 the proteomics profile of BC-sEVs may reflect the status of corresponding  
267 immune cells in the TME. To further map the differentially enriched sEV  
268 proteins to the immune response, we performed cell type deconvolution  
269 analysis using xCell (Aran *et al.*, 2017). A heatmap of overall and type-specific  
270 enrichment scores was constructed to identify the immune landscape of BC  
271 (Fig 2F). Specifically, the enrichment scores of macrophages M2, adipocytes,  
272 epithelial cells, CD4+ T cells,  $\gamma\delta$  T cells (Tgd), Th2 cells and  
273 megakaryocyte-erythroid progenitor cells (MEPs) were significantly elevated in  
274 BC-sEVs compared to HD-sEVs, with FC > 1.3 and Student's *t* test *p* value <  
275 0.05 (Fig 2F, Appendix Table 2). The analysis suggested a possible  
276 intercellular communication network of BC-driven sEVs in the TME when we

277 inferred the relative abundance of various immune cell subtypes in the TME.  
278 MEPs represent a bipotent transitional state that is permissive to the  
279 generation of unipotent progenitors of megakaryocytic or erythroid lineages  
280 (Xavier-Ferrucio *et al*, 2019). Adipocytes in the TME play dynamic and  
281 sophisticated roles in facilitating BC development (Cao, 2019). These  
282 BC-derived sEVs may impact the TME by promoting tumor cell growth and  
283 progression, modulating immune responses, regulating angiogenesis and  
284 inducing metastatic behavior through MEPs, endothelial cells, and mv  
285 endothelial cells (Fig 2G).

286

287 **Eight-protein diagnostic model to distinguish BC from BD and the  
288 healthy population**

289 To further assess whether sEV proteins could be used as a liquid diagnostic  
290 tool to discriminate cancers from noncancers, we next sought to determine  
291 shared and unique sEV proteins by performing pairwise comparisons of  
292 proteomes between BC-, BD-, and HD-sEVs. We applied the XGBoost  
293 classifier, which is robust to noise and overfitting, to verify a distinct sEV  
294 protein subset that can accurately distinguish the BC, BD and HD samples.

295 To train and subsequently test the model, sEV samples were evenly  
296 partitioned based on the sample source, and 70% of samples were used as a  
297 training set, with the remaining 30% used as an independent test set. Applying  
298 5-fold cross-validation to the training set, a combination of 8 sEV proteins  
299 (STAT1, PON1, APOC1, APOC2, MMP2, IGHV4-39, IGHV3-53, and ADD2)  
300 was used to construct a signature that yielded a sensitivity of 100% and  
301 specificity of 100% for discriminating BC from BD and HD (Fig 2H-I, Appendix  
302 Fig S2D). Notably, when applying this eight-protein identifier to sEV samples  
303 of the independent test set, the model achieved 97% sensitivity and 83%  
304 specificity in the diagnosis of BC (Fig 2H).

305

306 **Proteomic characteristics of sEVs derived from four clinical subtypes of**  
307 **BC**

308 IBC is a highly heterogeneous disease that can be categorized into various  
309 intrinsic or molecular subtypes, which are differentially correlated with clinical  
310 presentation, prognosis, distant metastasis, and response to therapy.  
311 Molecular subtypes are defined based on the gene expression signature and  
312 protein expression of estrogen receptor (ER), progesterone receptor (PR),  
313 human epidermal growth factor receptor 2 (Her2), and proliferative cell nuclear  
314 antigen (Ki67) (Li *et al*, 2021; Peng *et al*, 2019; Vallejos *et al*, 2010). We  
315 reasoned that since the biological behavior of IBC cells differs significantly  
316 among IBC subtypes, biological cargos carried by sEVs may vary among  
317 diverse molecular subtypes. To distinguish proteomic landscapes among  
318 diverse molecular subtypes of IBC and identify drivers that boost intertumoral  
319 heterogeneity and cancer evolution, we analyzed sEV samples from luminal A  
320 (ER+/PR+, low-grade and low-Ki67 index, n = 20), luminal B (ER+/PR+ of  
321 higher grade and proliferative index, n = 50), Her2-enriched (Her2+ with or  
322 without ER, n = 21), and triple-negative (ER-PR-Her2-, TNBC, n = 23) IBCs in  
323 our cohort. PCA demonstrated a clear distinction among the different  
324 molecular subtypes, which further highlighted the distinct proteomic patterns  
325 among several clinical subtypes of IBC samples ([Appendix Fig S3A](#)).

326 Next, we applied a *t* test with a nominal *p* value cut-off of < 0.05 and  
327 identified 87, 82, 83, and 104 sEV proteins that were significantly  
328 overrepresented in luminal A (FC (luminal A/any of the other three subtypes) >  
329 2), luminal B (FC (luminal B/ any of the other three subtypes) > 2),  
330 Her2-enriched (FC (Her2-enriched/ any of the other three subtypes) > 2), and  
331 TNBC (FC (TNBC/ any of the other three subtypes) > 2) samples ([see](#)  
332 [Materials and Methods](#)). Clustering and cluster-specific enrichment analyses

333 of the enriched proteins using GOBP and KEGG pathway annotations showed  
334 the distinctive biological processes and pathways represented in luminal A,  
335 luminal B, Her2-enriched, and TNBC samples ([Fig 3A and B, Appendix Table](#)  
336 [3](#)). Specifically, luminal A-derived sEVs were characterized by proteolysis  
337 involved in cellular protein catabolic processes (*i.e.*, PSMB7, PSMB2, FAP,  
338 and CAPN2) (Fisher's exact test,  $p = 1.23e-3$ ) and positive regulation of  
339 protein insertion into mitochondrial membrane involved in apoptotic signaling  
340 pathway (*i.e.*, YWHAB, YWHAG, and YWHAH) (Fisher's exact test,  $p =$   
341  $7.69e-3$ ). Luminal B-derived sEVs were characterized by cellular response to  
342 insulin stimulus (*i.e.*, RAB10, PKLR, GOT1, and STAT1) (Fisher's exact test,  $p$   
343  $= 4.41e-3$ ) and response to hypoxia (*i.e.*, ALAD, VCAM1, PKLR, and HSPD1)  
344 (Fisher's exact test,  $p = 3.77e-2$ ). Her2-enriched sEV-enriched proteins were  
345 related to cellular response to reactive oxygen species (*i.e.*, PRDX1, TXN, and  
346 SOD3) (Fisher's exact test,  $p = 1.34e-2$ ), glucose metabolic process (*i.e.*,  
347 FABP5, GAA, BPGM, and GAPDH) (Fisher's exact test,  $p = 3.47e-3$ ), and  
348 keratinization (*i.e.*, CASP14, KRT17, and TGM3) (Fisher's exact test,  $p =$   
349  $1.99e-2$ ). TNBC samples were characterized by platelet degranulation (*i.e.*,  
350 AHSG, ACTN4, PPBP, TLN1, and PF4) (Fisher's exact test,  $p = 2.67e-3$ ),  
351 blood coagulation (*i.e.*, EHD1, COL1A1, PROC, COL1A2, F11, and PRKACB)  
352 (Fisher's exact test,  $p = 3.74e-3$ ), adaptive immune response (*i.e.*, DBNL,  
353 ANXA1, ERAP2, and ICOSLG) (Fisher's exact test,  $p = 5.03e-2$ ), and platelet  
354 activation (*i.e.*, COL1A1, COL1A2, SAA1, and PF4) (Fisher's exact test,  $p =$   
355  $2.67e-2$ ) ([Fig 3B and C, Appendix Table](#) [3](#)).

356 Collectively, these data suggested that proteomic profiles of serum-derived  
357 sEVs reflect selective packaging, which represents an informative readout and  
358 differs among diverse subtypes of BCs.

359

360 **sEV-based classifier discriminates BC subtypes**

361 To further investigate the clinical significance of the differentially enriched  
362 protein cargos, we addressed whether they could be utilized as a novel liquid  
363 biopsy method to distinguish diverse clinical subtypes in clinical practice.  
364 Employing XGBoost classification, which is robust to noise and overfitting, we  
365 constructed a 61-protein classifier model that can accurately discriminate the  
366 luminal A, luminal B, Her2-enriched, and TNBC subtypes. To train and  
367 subsequently test the model, 70% of samples were used as a training set, with  
368 the remaining 30% used as an independent test set, in the same manner as  
369 previously described. Similar to our analysis of BC versus non-BC-sEVs, we  
370 constructed a 61-protein classifier model using the XGBoost classifier. To test  
371 the 61-sEV protein model, 5-fold cross-validation of the training set was  
372 performed and yielded a sensitivity of 100% and a specificity of 100% for each  
373 molecular subtype ([Fig 3D](#), [Appendix Fig S3B](#)). When applying the 61-protein  
374 classifier to the independent test set, the model achieved 67% sensitivity and  
375 97% specificity in the diagnosis of luminal A, 80% sensitivity and 70%  
376 specificity in diagnosis of luminal B, 57% sensitivity and 89% specificity in  
377 diagnosis of Her2-enriched, and 71% sensitivity and 100% specificity in  
378 diagnosis of TNBC ([Fig 3D](#), [Appendix Fig S3B](#)). The receiver operating  
379 characteristic (ROC) curve derived from the 61-protein signature showed good  
380 sensitivity and specificity, with an area under the curve (AUC) of 1.0 ([Fig 3E](#)).  
381 Then, the 61-protein signature was validated in the test set, resulting in a ROC  
382 curve with an AUC of 0.875 ([Fig 3E](#)).

383 Thus, serum sEV proteomes can be beneficial in determining the BC  
384 subtype for dynamic monitoring in patients during tumor progression, avoiding  
385 repeated tissue biopsies.

386

387 **Adipocytes play an important role in LN metastasis of BC**

388 Furthermore, to elucidate the mechanism of LN metastasis in IBC, we  
389 analyzed sEV proteins of IBC patients with LN metastases (IBC\_LN, n = 51)  
390 and without LN metastases (IBC\_Pure, n = 54). PCA clearly distinguished  
391 between IBC\_LN and IBC\_Pure samples at the protein level, which further  
392 highlighted the diverse proteomic patterns between sEVs from IBC\_LN and  
393 IBC\_Pure samples ([Appendix Fig S4A](#)). We applied Student's *t* test with a  
394 nominal *p* value cut-off of < 0.05 and identified significantly enriched  
395 sEV-derived proteins in IBC\_LN compared with IBC\_Pure (FC > 2). The  
396 results are summarized in the volcano plot shown in Fig S4B, and the most  
397 prominent proteins are indicated ([Appendix Fig S4B](#)).

398 We further performed clustering and cluster-specific enrichment analyses of  
399 the upregulated proteins using gene set enrichment analysis (GSEA). We  
400 found that IBC\_LN samples were characterized by proteins related to  
401 hallmarks of adipogenesis ([Fig 4A](#)). To investigate the immune landscapes of  
402 the IBC\_Pure and IBC\_LN groups, the abundance of 16 different cell types  
403 was computed using xCell based on proteomic data of sEVs retrieved from the  
404 blood of the 105 abovementioned IBC samples ([Fig 4B, Appendix Table 4](#)).  
405 We found that the enrichment scores of B cells, basophils, CD4+ T cells, CD4+  
406 naive T cells, dendritic cells (DCs), mesangial cells, activated dendritic cells  
407 (aDCs), and immature dendritic cells (iDCs) were higher in the IBC\_Pure  
408 group than in the IBC\_LN group. On the other hand, enrichment scores for  
409 adipocytes, CD8+ T cells, CD8+ naive T cells, multipotent progenitors (MPPs),  
410 macrophages, megakaryocytes, platelets, and sebocytes were higher in the  
411 IBC\_LN group than in the IBC\_Pure group (FC > 1.5, Student's *t* test *p* < 0.05)  
412 ([Fig 4B, Appendix Table 4](#)). The enhanced adipocyte enrichment scores in  
413 sEVs from IBC\_LN samples attracted our attention ([Fig 4B, Appendix Fig S4C,](#)  
414 [Table 4](#)). There was a positive correlation between adipogenesis and  
415 adipocytes (Spearman rho = 0.188, *p* = 5.507e-02) ([Appendix Fig S4D](#)).

416 Adipocytes were correlated with the VEGF signaling pathway (Fig 4C), and the  
417 VEGF signaling pathway was upregulated in the IBC\_LN group (Fig 4D). A  
418 previously reported comparative cytokine array analysis of  
419 adipocyte-conditioned medium (ACM) revealed the upregulation of a group of  
420 cytokines belonging to the VEGF signaling pathway in ACM (Sahoo *et al*,  
421 2018).

422 The VEGF signaling pathway was correlated with MPPs (Fig 4E), which  
423 were upregulated in the IBC\_LN group (Appendix Fig S4E) and positively  
424 correlated with the coagulation pathway (Spearman rho = 0.295, p value =  
425 2.216e-03) (Appendix Fig S4F). At the same time, platelets were positively  
426 correlated with the coagulation pathway (Spearman rho = 0.209, p value =  
427 3.225e-02) (Appendix Fig S4G). The enrichment scores of platelets were  
428 upregulated in the IBC\_LN group (Appendix Fig S4H). Experimental evidence  
429 has highlighted platelets as active players in all steps of tumorigenesis,  
430 including cancer growth, cancer cell extravasation and metastasis  
431 (Haemmerle *et al*, 2018). Many of the molecules that are highly associated  
432 with platelets are angiogenesis- and metastasis-related molecules (e.g., KIF5B,  
433 ARHGDIA, ARPC1B, DYNLL2, NUP98, IQGAP2, PTPRJ, PTPRF, MST1L,  
434 and MMP3) (Fig 4F, Appendix Fig S4I). In addition, we found that adipocytes,  
435 MPPs, and MEPs were significantly increased in the tissue samples of 40  
436 additional IBC patients (IBC\_Pure, n = 12; IBC\_LN, n = 28), and the platelet  
437 count in BC patients with LN metastasis (n = 43) was significantly higher than  
438 that in BC patients without LN metastasis (n = 45) in our cohort (Student's *t*  
439 test, *p* < 0.05) (Fig 4H and I, Appendix Fig S4J and K).

440

#### 441 **Twelve-protein diagnostic model for LN metastasis**

442 To generate a protein signature that stratifies patients with or without LN  
443 metastases, we performed random forest classification to identify a subset of

444 sEV proteins that accurately discriminates between IBC\_LN and IBC\_Pure  
445 samples. As before, sEV samples were evenly partitioned based on sample  
446 type (*i.e.*, IBC\_LN samples vs. IBC\_Pure samples), and 70% of samples were  
447 used as a training set, with the remaining 30% used as an independent test set.  
448 By comparing the IBC\_LN- and IBC\_Pure-derived sEV proteomes, we  
449 discovered that the best partition was achieved with 12 sEV proteins (PEPD,  
450 NCL, PARP1, ACTA2, ACTG2, TBCA, TTYH3, MATR3, KPNB1, KRT16,  
451 RANBP2, and CCT6A). Based on this 12-protein signature, applying 5-fold  
452 cross-validation to the training set yielded a sensitivity (true positive rate) of  
453 100% and specificity (true negative rate) of 100% ([Fig 4J and L, Appendix Fig  
454 S4J](#)). When applying the protein signature for discriminating BC patients with  
455 or without LN metastasis to the independent test set samples, it had a  
456 sensitivity of 81% and a specificity of 81% ([Fig 4M and N, Appendix Fig S4E](#)).  
457 In addition, we used the CPTAC breast cancer dataset ( $n = 77$ ) as an external  
458 validation test set and achieved 100% sensitivity and 100% specificity ([Mertins  
459 et al, 2016](#)) ([Fig 4K](#)).

460

#### 461 **Potential sEV survival biomarkers for distant metastases of BC**

462 To identify universal biomarkers associated with distant metastasis, we  
463 performed further analysis based on the proteomic profiles of 7 ductal  
464 carcinoma in situ (DCIS) samples and 21 distant metastasis (D-MET) (*e.g.*,  
465 M-Multiple ( $n = 5$ ), M-Lung ( $n = 3$ ), M-Liver ( $n = 4$ ), M-Bone ( $n = 7$ ), M-Chest  
466 wall ( $n = 1$ ), and M-Soft tissue ( $n = 1$ )) samples in our cohort. Clustering and  
467 cluster-specific enrichment analyses of the upregulated proteins using DAVID  
468 (KEGG gene sets) pathway annotations clearly showed distinctive biological  
469 processes and pathways enriched in D-MET samples compared to DCIS  
470 samples ([Fig 5A](#)). Compared with DCIS samples, D-MET samples showed an  
471 upregulation of focal adhesion (*i.e.*, FLNA and vitronectin (VTN)) (Fisher's

472 exact test,  $p = 5.45\text{e-}03$ ), metabolism-related pathways (e.g., carbon  
473 metabolism (i.e., PKM, G6PD, and TALDO1) (Fisher's exact test,  $p =$   
474  $5.42\text{e-}05$ ), glycolysis/gluconeogenesis (i.e., FBP1, LDHB, and PDHB)  
475 (Fisher's exact test,  $p = 1.33\text{e-}02$ ), fatty acid metabolism (i.e., ACACA,  
476 HSD17B12, and HACD3) (Fisher's exact test,  $p = 1.83\text{e-}02$ )), and complement  
477 and coagulation cascades (i.e., CPB2, alpha-1-antitrypsin (SERPINA1), CFH,  
478 C7, heparin cofactor 2 (SERPIND1), F10, F12, SERPINF2, SERPINE1, F2,  
479 TFPI, and KNG1) (Fisher's exact test,  $p = 3.58\text{e-}02$ ) ([Fig 5A and B](#)). We found  
480 that 24 sEV proteins were significantly overexpressed in distant metastatic  
481 samples ( $\text{D-MET}_{\text{median}}/\text{DCIS}_{\text{median}} > 2\text{-fold}$ , Student's  $t$  test,  $p < 0.05$ ),  
482 suggesting that they may be potential serum sEV protein markers for LN  
483 metastasis of BC ([Fig 5B](#)). Among them, 5 sEV proteins (PDHB, FBP1,  
484 PPP4C, GP1BA, and TFPI) were identified in  $> 75\%$  of D-MET samples ([Fig  
485 5B](#)). Remarkably, 11 sEV proteins (FLNA, VTN, PKM, PDHB, G6PD, TALDO1,  
486 LDHB, ACACA, PPP4C, C7 and F2) were highly expressed in BC and were  
487 associated with poor prognosis in the Tang et al. BC cohort and the Liu et al.  
488 BC cohort ([Liu et al, 2014; Tang et al., 2018](#)) ([Fig 5B and C, Appendix Fig  
489 S5A](#)).

490

## 491 **Potential organ-specific sEV survival biomarkers for distant metastases 492 of BC**

493 Furthermore, we performed pathway enrichment analysis comparing  
494 differentially expressed proteins among three different types of organ  
495 metastasis samples (M-Lung, M-Liver, and M-Bone samples). M-Lung sEVs  
496 showed upregulation of complement and coagulation cascades (i.e., CFD, C6,  
497 and SERPING1) (Fisher's exact test,  $p = 1.36\text{e-}02$ ), focal adhesion (i.e.,  
498 ITGB3, ITGA2B, and VCL) (Fisher's exact test,  $p = 2.63\text{e-}02$ ), and gap  
499 junctions (i.e., TUBB2B, TUBB2A, and TUBB) (Fisher's exact test,  $p =$

500 3.49e-02) ([Fig 5D, Appendix Table 5](#)). This finding is consistent with recent  
501 reports that focal adhesion and regulation of actin cytoskeleton signaling are  
502 involved in lung metastases of BC ([Zeng et al, 2019](#)). Interestingly, we found  
503 that abundant metabolism-related pathways were enriched in M-Liver sEVs,  
504 including fatty acid metabolism (*i.e.*, ACADVL, TECR, and ACSL5) (Fisher's  
505 exact test,  $p = 4.56e-02$ ), galactose metabolism (*i.e.*, GLB1, PGM5, and PGM1)  
506 (Fisher's exact test,  $p = 1.90e-02$ ), and starch and sucrose metabolism (*i.e.*,  
507 AMY2A, AMY1A, and AMY2B) (Fisher's exact test,  $p = 7.99e-05$ ) ([Fig 5D,](#)  
508 [Appendix Table 5](#)). M-Bone sEV samples showed upregulation of protein  
509 processing in the endoplasmic reticulum (*i.e.*, HSPH1, STT3A, RAD23A,  
510 P4HB, and SEC23B) (Fisher's exact test,  $p = 1.90e-02$ ) and nucleotide  
511 excision repair (*i.e.*, RPA1, RAD23A, and CUL4B) (Fisher's exact test,  $p =$   
512  $3.31e-02$ ) ([Fig 5D, Appendix Table 5](#)). These results suggest that although  
513 upregulated expression of adhesion, metabolism, and angiogenesis pathways  
514 are common features of distant metastases, different metastases are biased.  
515 M-Lung was the adhesion type, M-Liver was the metabolism type, and M-Bone  
516 was the repair type.

517 We found that GMDS was specifically highly expressed in M-Liver, P4HB  
518 was specifically highly expressed in M-Bone, and C6, TUBB, SERPING1 and  
519 VCL were specifically highly expressed in M-Lung ([Fig 5E-G](#)). In the Tang et al.  
520 BC cohort, the high expression of C6 and VCL was associated with poor  
521 prognosis, suggesting that they may be survival markers for lung metastasis of  
522 BC ([Tang et al., 2018](#)) ([Fig 5G and H, Appendix Fig S5B](#)).

523

## 524 **Potential BC-derived sEV molecules govern organ-specific metastasis**

525 Metastatic organotropism has remained an enigmatic issue. A recent study  
526 showed that cancer-derived sEV uptake by organ-specific cells may govern  
527 organ-specific metastasis ([Hoshino et al., 2015](#)). To examine whether sEV

528 proteins may guide the colonization of BC cells in specific organs, we  
529 computed the abundance of specific cell types in each of the distant metastatic  
530 samples using xCell ([Fig 6 A-D, Appendix Fig S6A and B, Appendix Table 6](#)).  
531 The analysis showed an enhanced enrichment score of chondrocytes in  
532 M-Bone sEVs, which was 6-fold, 1.75-fold, and 2.75-fold higher than that in  
533 DCIS, M-Lung, and M-Liver sEVs, respectively ([Fig 6B](#)). In contrast, the  
534 enrichment score of myocytes in M-Lung sEVs was upregulated by 2.20-fold,  
535 5.32-fold, and 1.82-fold compared to that in DCIS, M-Liver, and M-Bone sEVs,  
536 respectively ([Fig 6C](#)). Moreover, the enrichment score of fibroblasts in M-Liver  
537 samples was significantly elevated by 4.10-fold, 12.77-fold, and 5.33-fold  
538 compared to DCIS, M-Lung, and M-Bone samples, respectively ([Fig 6D](#))  
539 (Student's *t* test,  $p < 0.05$ ). Therefore, the organ specificity of sEV  
540 biodistribution matched the organotropic distribution of tumor cells.

541 A previous study suggested that specific exosomal integrins were  
542 associated with metastatic organotropism by dictating premetastatic niche  
543 formation ([Hoshino et al., 2015](#)). In our dataset, we identified 25 integrins  
544 enriched in M-Bone, M-Lung and M-Liver sEVs. Further analysis revealed that  
545 ITGA1 was primarily detected in M-Bone sEVs, ITGA7 and ITGA9 were  
546 abundantly enriched in M-Liver sEVs, and ITGB3 and ITGA2B were  
547 abundantly enriched in M-Lung sEVs ([Fig 6E, Appendix Fig S6C and D](#)).

548 In addition to adhesive properties, sEV integrins can upregulate  
549 promigratory and proinflammatory S100 molecules, which influence  
550 premetastatic niche formation ([Hoshino et al., 2015](#)). To determine the pattern  
551 of sEV-S100 molecules in tumor metastasis, we identified 16 S100 molecules  
552 from M-Bone, M-Lung and M-Liver sEVs. The analysis revealed that S100A8  
553 was primarily detected in sEV-derived proteins from M-Bone samples ([Fig 6E](#)).  
554 S100A13 was primarily detected in M-Liver samples ([Appendix Fig S6C](#)).  
555 Interestingly, S100A7A was abundantly present in sEV-derived proteins from

556 M-Lung samples ([Appendix Fig S6D](#)). In addition, we verified that ITGA1 was  
557 significantly increased in M-Bone tissue samples in our additional BC cohort  
558 (DCIS (n = 4), M-Liver (n = 4), M-Lung (n = 4), and M-Bone (n = 8)) (Student's t  
559 test,  $p < 0.05$ ) ([Fig 6F, Appendix Fig S6A](#)). Consistently, S100A8, S100A13,  
560 and S100A7A were significantly increased in M-Bone, M-Liver, and M-Lung  
561 tissue samples in our additional BC cohort (DCIS (n = 3), M-Liver (n = 3),  
562 M-Lung (n = 4), and M-Bone (n = 8)) (Student's  $t$  test,  $p < 0.05$ ) ([Fig 6G,](#)  
563 [Appendix Fig S6B, E and F](#)). Taken together, these results suggested a  
564 correlation between specific sEV integrins and S100 molecules and tissue  
565 organotropism ([Fig 6H](#)).

566

567 **Discussion**

568 Blood tests remain the most readily accessible source for the early detection,  
569 classification, and treatment guidance of BC patients. The billions of sEVs  
570 circulating in blood could represent an essential component of liquid biopsy  
571 ([Miyagi et al, 2021](#)). Despite previous studies on BC-derived sEVs ([Chen et al,](#)  
572 [2017](#)), there is a lack of a comprehensive understanding of BC-specific sEV  
573 characteristics and their composition and consensus on unique BC biomarkers  
574 due to limited sEV proteome data from human samples.

575 Here, we performed a large-scale comprehensive analysis of sEV  
576 proteomes from 167 serum samples obtained from patients with BC, patients  
577 with BD, and healthy individuals. Firstly, we applied this eight-protein (STAT1,  
578 PON1, APOC1, APOC2, MMP2, IGHV4-39, IGHV3-53, and ADD2) identifier to  
579 sEV samples of the independent test set, the model achieved 97% sensitivity  
580 and 83% specificity in the diagnosis of BC. This study may provide reference  
581 value for differentiating benign and malignant breast tumors using serum in the  
582 future.

583 BC is a heterogeneous disease in terms of molecular alterations, cellular  
584 compositions, and clinical outcomes (Wagner *et al*, 2019). Therefore, the  
585 classification of molecular subtype is an important tool for treatment and  
586 prognosis evaluation. Clinically, based on the expression of ER, PR, Her2, and  
587 Ki67 by IHC, BC is categorized into various molecular subtypes (Holm *et al*,  
588 2021). However, the patterns of these biological indicators may change during  
589 the course of BC progression, so they may be used to adjust treatment  
590 strategies accordingly (Ju *et al*, 2018). Thus, we speculated that an sEV-based  
591 in vitro diagnostic strategy is an emerging approach complementary to tissue  
592 pathology. Unfortunately, we failed to confirm the existence of ER, PR, HER2,  
593 and Ki67 in the serum sEV datasets, indicating that they may either have a low  
594 abundance or be lacking in serum sEVs. However, further analyses of  
595 differentially regulated sEV-derived proteins in luminal A, luminal B,  
596 Her2-enriched, and TNBC samples clearly showed significant differences in  
597 the proteins and biological pathways involved. By comparing proteomic  
598 profiles among diverse molecular subtypes of BC, we constructed a 61-protein  
599 classifier. The ROC curve derived from the 61-protein signature showed good  
600 sensitivity and specificity, with an AUC of 1.0. Then, the 61-protein signature  
601 was validated in the test set, resulting in a ROC curve with an AUC of 0.875.  
602 This work may provide reference value for the diagnosis of clinical subtypes of  
603 BC using serum in the future.

604 An accurate preoperative assessment of LN status is one of the most  
605 important prognostic factors determining the long-term outcome (Banerjee *et*  
606 *al*, 2004). Although noninvasive imaging modalities such as ultrasonography,  
607 computed tomography, and magnetic resonance imaging have been widely  
608 adopted for the clinical evaluation of LN status before surgery, the sensitivity of  
609 these modalities is not satisfactory (Song, 2020). In the present study, PCA  
610 demonstrated a clear distinction between IBC\_LN samples and IBC\_Pure

611 samples, which further highlighted the diverse proteomic patterns between  
612 IBC with or without LN metastasis. Hence, we constructed an sEV-based  
613 protein signature that predicted LN metastasis at the serum sEV proteomic  
614 level based on machine learning classification, showing 81% and 81%  
615 specificity and sensitivity, respectively. In addition, we used the CPTAC BC  
616 dataset ( $n = 77$ ) as an external validation test set and achieved 100%  
617 sensitivity and 100% specificity. These data suggest that tumor-associated  
618 sEV proteins can serve as biomarkers for early-stage cancer detection of LN  
619 metastasis.

620 Previous studies showed that adhesion and ECM molecules, such as  
621 integrins, tenascin and periostin, were associated with distant metastasis of  
622 disseminating cancer cells (Fukuda *et al*, 2015; Oskarsson *et al*, 2011;  
623 Radisky *et al*, 2002; Weaver *et al*, 1997). Regarding the research on this  
624 aspect, Hoshino et al. defined a specific repertoire of integrins expressed on  
625 cancer-derived exosomes, distinct from cancer cells, that dictate metastatic  
626 tropism (Hoshino *et al*, 2015). In our study, we identified 25 integrins  
627 abundantly present in human bone-, lung- and liver-tropic metastatic sEVs by  
628 quantitative mass spectrometry. Notably, we found that sEVs expressing  
629 ITGA1 may specifically bind to chondrocytes, which are related to bone  
630 tropism. sEVs expressing ITGB3 and ITGA2B may specifically bind to  
631 lung-resident myocytes, mediating lung tropism. However, sEVs expressing  
632 ITGA7 and ITGA9 may bind liver-resident fibroblasts, governing liver tropism.  
633 Moreover, we revealed that the pattern of sEV-S100 molecules was correlated  
634 with tissue organotropism and could serve as a biomarker for distant  
635 metastasis (Fig 6H).

636 In conclusion, our findings show that proteins carried by BC-derived sEVs  
637 could be used as a novel, minimally invasive liquid biopsy tool for the early  
638 detection of BC, as well as for discriminating molecular subtypes, LN

639 involvement status, and organotropic metastasis. These findings could  
640 advance the implementation of routine serum sEV-based screening in the  
641 clinic.

642

## 643 **Materials and Methods**

### 644 **Sample collection**

645 Serum sample collection was approved by Shanghai General Hospital  
646 Shanghai Jiao Tong University School of Medicine (Shanghai, China, permit  
647 number [2017]KY053), and all patients provided proper consent before  
648 samples were collected. Serum samples were collected between March 2011  
649 and August 2019. Detailed information is shown in Appendix Table 1.

650

### 651 **sEV extraction**

652 Isolation of exosomes was performed by differential ultracentrifugation  
653 following established centrifugation times and parameters (An *et al*, 2018; Gao  
654 *et al*, 2021; Lakhter *et al*, 2018; Takov *et al*, 2019; Thery *et al*, 2006). Firstly, 1  
655 mL serum was thawed on ice and centrifuged at 3,000 g for 10 min at 4°C. The  
656 supernatant was removed, and large vesicles were removed with another  
657 centrifugation step at 10,000 g for 20 min at 4°C and the supernatant was  
658 diluted with 25 mL PBS and filtered through a 0.22 µm centrifugal filter device  
659 to remove any large contaminating vesicles. Secondly, filtered serum was  
660 centrifuge at an overspeed of 150,000 g for 4 h, the milky white floating object  
661 at the top was sucked away. Thirdly, centrifuged material was resuspended  
662 with 25 mL PBS and further centrifuged at 4°C for 150,000 g for 2 h. Fourthly,  
663 supernatant was discarded and 200 µL solution was retained at the bottom to  
664 resuspend the precipitate. Isolation and relative purity of the sEVs were  
665 confirmed by NTA, transmission electron microscopy (TEM) and immunoblot.

666

### 667 **sEVs protein extraction and tryptic digestion**

668 sEV samples (typically 5 µg, adjusted based on BCA measurements) were  
669 dried by vacuum centrifugation and redissolved in 30–50 µL of 8 M urea/50  
670 mM ammonium bicarbonate/10 mM DTT. Following lysis and reduction,  
671 proteins were alkylated using 20 or 30 mM iodoacetamide (Sigma, St. Louis,  
672 MO, USA). Proteins were digested with trypsin (Promega, Madison, WI, USA)  
673 at an enzyme-to-protein mass ratio of 1:50 overnight at 37°C, and peptides  
674 were then extracted and dried (SpeedVac, Eppendorf). Peptides were  
675 desalted and concentrated using Empore C<sub>18</sub>-based solid phase extraction  
676 prior to analysis by high resolution/high mass accuracy reversed-phase (C<sub>18</sub>)  
677 nano-LC-MS/MS.

678

### 679 **Liquid chromatography**

680 We employed an EASY-nLC 1200 ultra-high-pressure system liquid  
681 chromatography system (Thermo Fisher Scientific). Peptides were separated  
682 within 75 min at a flow rate of 600 nL/min on a 150 µm I.D. × 15 cm column  
683 with a laser-pulled electrospray emitter packed with 1.9 µm ReproSil-Pur 120  
684 C<sub>18</sub>-AQ particles (Dr. Maisch). Mobile phases A and B were water and  
685 acetonitrile with 0.1 vol% FA, respectively. The %B was linearly increased  
686 from 15 to 30% within 75 min.

687

### 688 **Mass spectrometry**

689 Samples were analysed on a Q-Exactive-HF mass spectrometer (Thermo  
690 Fisher Scientific) via a nanoelectrospray ion source (Thermo Fisher Scientific).  
691 The mass spectrometer was operated in data-independent mode for ion  
692 mobility-enhanced spectral library generation. Typically, 75% of samples were  
693 injected. The peptides were dissolved in 12 µL of loading buffer (0.1% formic

694 acid), and 9  $\mu$ L was loaded onto a 100  $\mu$ m I.D.  $\times$  2.5 cm C<sub>18</sub> trap column at a  
695 maximum pressure of 280 bar with 14  $\mu$ L of solvent A (0.1% formic acid). The  
696 DIA method consisted of an MS1 scan from 300–1400 m/z at 60 k resolution  
697 (AGC target 4e5 or 50 ms). Then, 30 DIA segments were acquired at 15 k  
698 resolution with an AGC target of 5e4 or 22 ms for maximal injection time. The  
699 setting “injections for all available parallelizable time” was enabled. HCD  
700 fragmentation was set to a normalized collision energy of 30%. The spectra  
701 were recorded in profile mode. The default charge state for the MS2 scan was  
702 set to 3.

703

#### 704 **Peptide identification and protein quantification**

705 All data were processed using Firmiana ([Feng et al, 2017](#)). The DIA data were  
706 searched against the UniProt human protein database using FragPipe (v.12.1)  
707 with MSFragger (2.2) ([Kong et al, 2017](#)). The mass tolerances were 20 ppm  
708 for precursor and 50 mmu for product ions. Up to two missed cleavages were  
709 allowed. The search engine set cysteine carbamidomethylation as a fixed  
710 modification and N-acetylation and oxidation of methionine as variable  
711 modifications. Precursor ion score charges were limited to +2, +3, and +4. The  
712 data were also searched against a decoy database so that protein  
713 identifications were accepted at a false discovery rate (FDR) of 5%.

714 The DIA data was analysed using DIANN (v.1.7.0) ([Demichev et al, 2020](#)).  
715 The quantification of identified peptides was calculated as the average  
716 chromatographic fragment ion peak areas across all reference spectra  
717 libraries. Label-free protein quantifications were calculated using a label-free,  
718 intensity-based absolute quantification (iBAQ) approach ([Zhang et al, 2012](#)).  
719 We calculated the peak area values as parts of the corresponding proteins.  
720 The fraction of total (FOT) was used to represent the normalized abundance of  
721 a particular protein across samples. FOT was defined as a protein’s iBAQ

722 divided by the total iBAQ of all identified proteins within a sample. The FOT  
723 values were multiplied by  $10^5$  for ease of presentation, and missing values  
724 were imputed to  $10^{-5}$ . The raw proteomics data files are hosted by iProX and  
725 can be accessed at <https://www.iprox.cn> (Project ID: IPX0003429000).

726

## 727 **Statistical analysis**

728 To impute the proteomic data, we first screened more than 50% of the  
729 identified proteins in each group and divided the data into two parts. When the  
730 protein detection rate was < 50%, the missing value was replaced with one  
731 tenth of the minimum value. For these proteins, no imputation was applied.  
732 When the protein detection rate was > 0.5, the missing value was probably due  
733 to the detection accuracy limitation of LC/MS. In this case, we first calculated  
734 the missing probability of a protein using the R package “impute”  
735 (<https://git.bioconductor.org/packages/impute>) based on the K-NN algorithm.  
736 Meta-analysis-based discovery and validation of survival biomarkers was  
737 carried out using Kaplan-Meier Plotter (<http://kmplot.com/analysis/>).

738

## 739 **Principal component analysis (PCA)**

740 The imputed data were then normalized using the LogNorm algorithm. The  
741 PCA function of the R package “factoextra”  
742 (<https://cran.r-project.org/web/packages/factoextra/index.html>) was used to  
743 implement unsupervised clustering analysis. The 95% confidence coverage  
744 was represented by a coloured ellipse for each group and was calculated  
745 based on the mean and covariance of points in the different groups. 1,734,  
746 1,038, and 1,116 proteins (features) were used for PCA to illustrate the global  
747 proteomic differences among BC (n = 126), BD (n = 17) and HD (n = 24), the  
748 global proteomic differences among the luminal A (n = 20), luminal B (n = 50),  
749 Her2-enriched (n = 21), and triple-negative breast cancer (TNBC) (n = 23)

750 subtypes, and the global proteomic differences between IBC\_Pure (n = 54)  
751 and IBC\_LN (n = 51) (Appendix Fig S2A, 3A, 4A).

752

### 753 **Global Heatmap**

754 Each gene expression value in the global proteomic expression matrix was  
755 transformed to a z-score across all the samples. The z-score-transformed  
756 matrix was clustered using the R package “pheatmap”  
757 (<https://cran.r-project.org/web/packages/pheatmap/index.html>).

758

### 759 **Pathway enrichment analysis**

760 Pathway enrichment analysis was performed by DAVID (<https://david.ncifcrf.gov>) and ConsensusPathDB (<http://cpdb.molgen.mpg.de>), and  
761 significance in the pathway enrichment analysis was determined by Fisher's  
762 exact test on the basis of Kyoto Encyclopedia of Genes and Genomes (KEGG)  
763 pathways and categorical annotations, including Gene Ontology (GO)  
764 biological process (GOBP) terms and Reactome (<https://reactome.org>).

766

### 767 **Multiplex immunohistochemistry (mIHC) with tyramide signal 768 amplification**

769 Tissues or cells were prepared for detection with kits using standard fixation  
770 and embedding techniques. Each slide was baked in an oven at 65°C for 1 h,  
771 dewaxed with xylene (3 x 10 min) and rehydrated through a graded series of  
772 ethanol solutions (100% ethanol, 95% ethanol, 75% ethanol, 50% ethanol) and  
773 each step took 5 min. After rehydration, immersing the slides in the boiled  
774 appropriate AR buffer, and placed in a microwave for 15 min at 20% power.  
775 After naturally cooling to room temperature, washing the slides with TBST.  
776 Then we used blocking buffer to incubate tissue section for 10 min. The  
777 blocking buffer was drained, and Primary Antibody Working Solution was

778 applied. CD45RA (1:3000; ab755; Abcam), CD34 (1:6000; ab81289; Abcam),  
779 CD38 (1:800; ab108403; Abcam), CD71 (1:800; ab214039; Abcam), and  
780 CDH1 (1:10000; ab181860; Abcam) were used. The slides were incubated at  
781 4°C overnight or at room temperature for 1 h; the time may be adjusted  
782 according to different characteristics of the antibody. After washing the slides  
783 with TBST, incubate them in polymer HRP Ms+Rb for 15 min at room  
784 temperature. Washing the slides twice again. Working Solution (100–300 µL)  
785 was pipetted onto each slide at room temperature for 10 min. And then  
786 immersed in the appropriate AR buffer. This microwave step strips the  
787 primary-secondary-HRP complex, allowing the introduction of the next primary  
788 antibody. For detection of the next target with fluorophores, we restarted the  
789 protocol at blocking. Once all 5 targets were labelled, Opal Polaris 780  
790 labelling was continued.

791 Dropping TSA-DIG Working Solution onto slides and incubating at room  
792 temperature for 10 min. Repeat the previous microwave repair steps after  
793 washing the slides. Polaris 780 Working Solution was pipetted onto each slide  
794 and incubated at room temperature for 1 h. DAPI working solution was applied  
795 for 5 min. The slides were washed twice again. After the slides were slightly  
796 dry, a super quench sealing tablet was added to the slides with a pipette, and  
797 the sample area was immersed.

798

### 799 **Immunohistochemistry (IHC)**

800 Firstly, the sections were baked at 65°C for 1 h and incubated in xylene three  
801 times for 10 min each time. Then, the sections were hydrated by a graded  
802 series of ethanol (100% ethanol, 95% ethanol, 75% ethanol, 50% ethanol and  
803 ddH<sub>2</sub>O), and each step took 5 min. Antigen retrieval was conducted using a  
804 microwave oven: 3 min at 100% power and 15 min at 20% power filled with 10  
805 mM sodium citrate buffer (pH 6.0). After naturally cooling to room temperature

806 and washed in ddH<sub>2</sub>O, we blocked the sections with 5% normal goat serum for  
807 10 min, incubated sections in 3% H<sub>2</sub>O<sub>2</sub> for 10 min at room temperature, and  
808 washed the sections twice in PBS for 5 min. The following antibodies were  
809 diluted in the appropriate concentrations: PPARg (1:10000; ab59256; Abcam),  
810 S100A8 (1:800; 15792-1-AP; Proteintech), S100A13 (1:1200; ab109252;  
811 Abcam), S100A7A (1:400; DF8517; Affinity), and ITGA1 (1:300; 22146-1 AP;  
812 Proteintech). These antibodies were incubated with the sections overnight at  
813 4°C.

814 The next day, after washing the sections twice in PBS, we used an IHC Kit  
815 (ZSGB-BIO, Beijing, China, Cat# SP-9000), incubated the sections with  
816 biotin-labelled secondary antibody for 15 min. After washing sections twice in  
817 PBS, incubating the sections with horseradish enzyme-labelled Streptomyces  
818 ovalbumin working solution for 15 min. Finally, We used DAB solution to stain  
819 the tissues. Then, using haematoxylin to stain nuclears and washing them in  
820 ddH<sub>2</sub>O. Finally, the sections were dehydrated by graded ethanol (50% ethanol,  
821 75% ethanol, 95% ethanol, and 100% ethanol). We dried the slides in a fume  
822 cupboard for at least 20 min and mounted coverslips.

823

## 824 **Acknowledgments**

825 This work was supported by the National Natural Science Funds (grant  
826 numbers 82073269, 81772802 and M-0349), Shanghai Science and  
827 Technology Innovation Action Plan (grant number 20XD1402800), Clinical  
828 Research Plan of SHDC (grant number SHDC2020CR2065B), Clinical  
829 Research Innovation Plan of Shanghai General Hospital (grant number  
830 CTCCR-2016B05), National Key R&D Program of China (grant numbers  
831 2017YFA0505102, 2016YFA0502500, 2018YFA0507501, and  
832 2017YFC0908404), National Natural Science Foundation of China (grant  
833 numbers 31770886, 1972933, and 31700682), Science and Technology

834 Commission of Shanghai Municipality (grant number 2017SHZDZX01), Major  
835 Project of Special Development Funds of Zhangjiang National Independent  
836 Innovation Demonstration Zone (grant number ZJ2019-ZD-004).

837

838 **Author Contributions**

839 H.X.W. contributed to idea, conception, and study design. H.X.W., C.D., and  
840 H.W.Z. wrote the paper and supervised the project. G.F.X., S.Q.D. and M.J.H.  
841 conducted the mass spectrometry analysis. W.Y.H. and the other authors  
842 carried out all the remaining experiments. All authors discussed the results,  
843 commented on the project and approved the manuscript.

844

845 **References**

846 An M, Wu J, Zhu J, Lubman DM (2018) Comparison of an Optimized  
847 Ultracentrifugation Method versus Size-Exclusion Chromatography for  
848 Isolation of Exosomes from Human Serum. *J Proteome Res* 17: 3599-3605  
849 Aran D, Hu Z, Butte AJ (2017) xCell: digitally portraying the tissue cellular  
850 heterogeneity landscape. *Genome Biol* 18: 220  
851 Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, Skog J (2011)  
852 Tumour microvesicles contain retrotransposon elements and amplified  
853 oncogene sequences. *Nat Commun* 2: 180  
854 Banerjee M, George J, Song EY, Roy A, Hryniuk W (2004) Tree-based model  
855 for breast cancer prognostication. *J Clin Oncol* 22: 2567-2575  
856 Becker M, Weinberger T, Chandy A, Schmukler S (2016) Depression During  
857 Pregnancy and Postpartum. *Curr Psychiatry Rep* 18: 32  
858 Cao Y (2019) Adipocyte and lipid metabolism in cancer drug resistance. *J Clin  
859 Invest* 129: 3006-3017  
860 Cassetta L, Pollard JW (2017) Repolarizing macrophages improves breast  
861 cancer therapy. *Cell Res* 27: 963-964  
862 Chen G, Huang AC, Zhang W, Zhang G, Wu M, Xu W, Yu Z, Yang J, Wang B,  
863 Sun H et al (2018) Exosomal PD-L1 contributes to immunosuppression and is  
864 associated with anti-PD-1 response. *Nature* 560: 382-386  
865 Chen IH, Xue L, Hsu CC, Paez JS, Pan L, Andaluz H, Wendt MK, Iliuk AB, Zhu  
866 JK, Tao WA (2017) Phosphoproteins in extracellular vesicles as candidate  
867 markers for breast cancer. *Proc Natl Acad Sci U S A* 114: 3175-3180

868 Choi Y, Park U, Koo HJ, Park JS, Lee DH, Kim K, Choi J (2021)  
869 Exosome-mediated diagnosis of pancreatic cancer using lectin-conjugated  
870 nanoparticles bound to selective glycans. *Biosens Bioelectron* 177: 112980  
871 Colombo M, Raposo G, Thery C (2014) Biogenesis, secretion, and intercellular  
872 interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev  
873 Biol* 30: 255-289

874 Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, Becker A,  
875 Hoshino A, Mark MT, Molina H *et al* (2015) Pancreatic cancer exosomes  
876 initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* 17: 816-826  
877 Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2020) DIA-NN:  
878 neural networks and interference correction enable deep proteome coverage  
879 in high throughput. *Nat Methods* 17: 41-44

880 Fang X, Duan Y, Adkins GB, Pan S, Wang H, Liu Y, Zhong W (2018) Highly  
881 Efficient Exosome Isolation and Protein Analysis by an Integrated  
882 Nanomaterial-Based Platform. *Anal Chem* 90: 2787-2795

883 Feng J, Ding C, Qiu N, Ni X, Zhan D, Liu W, Xia X, Li P, Lu B, Zhao Q *et al*  
884 (2017) Firmiana: towards a one-stop proteomic cloud platform for data  
885 processing and analysis. *Nat Biotechnol* 35: 409-412

886 Fukuda K, Sugihara E, Ohta S, Izuhara K, Funakoshi T, Amagai M, Saya H  
887 (2015) Periostin Is a Key Niche Component for Wound Metastasis of  
888 Melanoma. *PLoS One* 10: e0129704

889 Gao Y, Chen Y, Wang L, Li C, Ge W (2021) Serum-derived extracellular  
890 vesicles inhibit osteoclastogenesis in active-phase patients with SAPHO  
891 syndrome. *Ther Adv Musculoskelet Dis* 13: 1759720X211006966

892 Haemmerle M, Stone RL, Menter DG, Afshar-Kharghan V, Sood AK (2018)  
893 The Platelet Lifeline to Cancer: Challenges and Opportunities. *Cancer Cell* 33:  
894 965-983

895 Holm J, Yu NY, Johansson A, Ploner A, Hall P, Lindstrom LS, Czene K (2021)  
896 Concordance of Immunohistochemistry-Based and Gene Expression-Based  
897 Subtyping in Breast Cancer. *JNCI Cancer Spectr* 5: pkaa087

898 Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M,  
899 Molina H, Kohsaka S, Di Giannatale A, Ceder S *et al* (2015) Tumour exosome  
900 integrins determine organotropic metastasis. *Nature* 527: 329-335

901 Hoshino A, Kim HS, Bojmar L, Gyan KE, Cioffi M, Hernandez J, Zambirinis CP,  
902 Rodrigues G, Molina H, Heissel S *et al* (2020) Extracellular Vesicle and  
903 Particle Biomarkers Define Multiple Human Cancers. *Cell* 182: 1044-1061  
904 e1018

905 Huber V, Fais S, Iero M, Lugini L, Canese P, Squarcina P, Zaccheddu A,  
906 Colone M, Arancia G, Gentile M *et al* (2005) Human colorectal cancer cells  
907 induce T-cell death through release of proapoptotic microvesicles: role in  
908 immune escape. *Gastroenterology* 128: 1796-1804

909 Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C (1987) Vesicle  
910 formation during reticulocyte maturation. Association of plasma membrane  
911 activities with released vesicles (exosomes). *J Biol Chem* 262: 9412-9420  
912 Ju G, Zhu R, Zhao H, Ye F, Zhang L, Lin C, Lu Y, Zhang X, Li N, Xue P *et al*  
913 (2018) The discordance pattern of molecular sub-types between primary and  
914 metastatic sites in Chinese breast cancer patients. *Int J Clin Exp Pathol* 11:  
915 5938-5947  
916 Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, Bond VC,  
917 Borras FE, Breakefield X, Budnik V *et al* (2012) Vesiclepedia: a compendium  
918 for extracellular vesicles with continuous community annotation. *PLoS Biol* 10:  
919 e1001450  
920 Kim DK, Kang B, Kim OY, Choi DS, Lee J, Kim SR, Go G, Yoon YJ, Kim JH,  
921 Jang SC *et al* (2013) EVpedia: an integrated database of high-throughput data  
922 for systemic analyses of extracellular vesicles. *J Extracell Vesicles* 2  
923 Kim JS, Chang JW, Park JK, Hwang SG (2012) Increased aldehyde reductase  
924 expression mediates acquired radioresistance of laryngeal cancer cells via  
925 modulating p53. *Cancer Biol Ther* 13: 638-646  
926 Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI  
927 (2017) MSFagger: ultrafast and comprehensive peptide identification in mass  
928 spectrometry-based proteomics. *Nat Methods* 14: 513-520  
929 Kralj-Iglic V (2012) Stability of membranous nanostructures: a possible key  
930 mechanism in cancer progression. *Int J Nanomedicine* 7: 3579-3596  
931 Lakhter AJ, Pratt RE, Moore RE, Doucette KK, Maier BF, DiMeglio LA, Sims  
932 EK (2018) Beta cell extracellular vesicle miR-21-5p cargo is increased in  
933 response to inflammatory cytokines and serves as a biomarker of type 1  
934 diabetes. *Diabetologia* 61: 1124-1134  
935 Laudato S, Patil N, Abba ML, Leupold JH, Benner A, Gaiser T, Marx A,  
936 Allgayer H (2017) P53-induced miR-30e-5p inhibits colorectal cancer invasion  
937 and metastasis by targeting ITGA6 and ITGB1. *Int J Cancer* 141: 1879-1890  
938 Lee J, Kwon MH, Kim JA, Rhee WJ (2018) Detection of exosome miRNAs  
939 using molecular beacons for diagnosing prostate cancer. *Artif Cells Nanomed  
940 Biotechnol* 46: S52-S63  
941 Li A, Zhang T, Zheng M, Liu Y, Chen Z (2017a) Exosomal proteins as potential  
942 markers of tumor diagnosis. *J Hematol Oncol* 10: 175  
943 Li H, Gao C, Zhuang J, Liu L, Yang J, Liu C, Zhou C, Feng F, Liu R, Sun C  
944 (2021) An mRNA characterization model predicting survival in patients with  
945 invasive breast cancer based on The Cancer Genome Atlas database. *Cancer  
946 Biomark* 30: 417-428  
947 Li W, Li C, Zhou T, Liu X, Liu X, Li X, Chen D (2017b) Role of exosomal  
948 proteins in cancer diagnosis. *Mol Cancer* 16: 145  
949 Liu NQ, Stingl C, Look MP, Smid M, Braakman RB, De Marchi T, Sieuwerts  
950 AM, Span PN, Sweep FC, Linderholm BK *et al* (2014) Comparative proteome

951 analysis revealing an 11-protein signature for aggressive triple-negative breast  
952 cancer. *J Natl Cancer Inst* 106: djt376

953 Liu X, Li J, Cadilha BL, Markota A, Voigt C, Huang Z, Lin PP, Wang DD, Dai J,  
954 Kranz G *et al* (2019a) Epithelial-type systemic breast carcinoma cells with a  
955 restricted mesenchymal transition are a major source of metastasis. *Sci Adv* 5:  
956 eaav4275

957 Liu Y, Wang H, Wang X, Liu J, Li J, Wang X, Zhang Y, Bai Z, Zhou Q, Wu Y *et*  
958 *al* (2021) Prevalence and reclassification of BRCA1 and BRCA2 variants in a  
959 large, unselected Chinese Han breast cancer cohort. *J Hematol Oncol* 14: 18

960 Liu Z, Wang Y, Wang L, Yao B, Sun L, Liu R, Chen T, Niu Y, Tu K, Liu Q  
961 (2019b) Long non-coding RNA AGAP2-AS1, functioning as a competitive  
962 endogenous RNA, upregulates ANXA11 expression by sponging miR-16-5p  
963 and promotes proliferation and metastasis in hepatocellular carcinoma. *J Exp*  
964 *Clin Cancer Res* 38: 194

965 Maas SLN, Breakefield XO, Weaver AM (2017) Extracellular Vesicles: Unique  
966 Intercellular Delivery Vehicles. *Trends Cell Biol* 27: 172-188

967 Mertins P, Mani DR, Ruggles KV, Gillette MA, Clouser KR, Wang P, Wang X,  
968 Qiao JW, Cao S, Petralia F *et al* (2016) Proteogenomics connects somatic  
969 mutations to signalling in breast cancer. *Nature* 534: 55-62

970 Milane L, Singh A, Mattheolabakis G, Suresh M, Amiji MM (2015) Exosome  
971 mediated communication within the tumor microenvironment. *J Control*  
972 *Release* 219: 278-294

973 Miyagi T, Miyata S, Tagami K, Hiratsuka Y, Sato M, Takeda I, Kohata K,  
974 Satake N, Shimokawa H, Inoue A (2021) Prognostic model for patients with  
975 advanced cancer using a combination of routine blood test values. *Support*  
976 *Care Cancer*

977 Oskarsson T, Acharyya S, Zhang XH, Vanharanta S, Tavazoie SF, Morris PG,  
978 Downey RJ, Manova-Todorova K, Brogi E, Massague J (2011) Breast cancer  
979 cells produce tenascin C as a metastatic niche component to colonize the  
980 lungs. *Nat Med* 17: 867-874

981 Ozer T, Geiss BJ, Henry CS (2020) Review-Chemical and Biological Sensors  
982 for Viral Detection. *J Electrochem Soc* 167: 037523

983 Peinado H, Aleckovic M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno  
984 G, Hergueta-Redondo M, Williams C, Garcia-Santos G, Ghajar C *et al* (2012)  
985 Melanoma exosomes educate bone marrow progenitor cells toward a  
986 pro-metastatic phenotype through MET. *Nat Med* 18: 883-891

987 Peinado H, Lavotshkin S, Lyden D (2011) The secreted factors responsible for  
988 pre-metastatic niche formation: old sayings and new thoughts. *Semin Cancer*  
989 *Biol* 21: 139-146

990 Peng J, Chen J, Xie F, Bao W, Xu H, Wang H, Xu Y, Du Z (2019)  
991 Herceptin-conjugated paclitaxel loaded PCL-PEG worm-like nanocrystal

992 micelles for the combinatorial treatment of HER2-positive breast cancer.  
993 *Biomaterials* 222: 119420

994 Radisky D, Muschler J, Bissell MJ (2002) Order and disorder: the role of  
995 extracellular matrix in epithelial cancer. *Cancer Invest* 20: 139-153

996 Raposo G, Stoorvogel W (2013) Extracellular vesicles: exosomes,  
997 microvesicles, and friends. *J Cell Biol* 200: 373-383

998 Rontogianni S, Synadaki E, Li B, Liefaard MC, Lips EH, Wesseling J, Wu W,  
999 Altelaar M (2019) Proteomic profiling of extracellular vesicles allows for human  
1000 breast cancer subtyping. *Commun Biol* 2: 325

1001 Sahoo SS, Lombard JM, Ius Y, O'Sullivan R, Wood LG, Nahar P, Jaaback K,  
1002 Tanwar PS (2018) Adipose-Derived VEGF-mTOR Signaling Promotes  
1003 Endometrial Hyperplasia and Cancer: Implications for Obese Women. *Mol  
1004 Cancer Res* 16: 309-321

1005 Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer Statistics, 2021. *CA: a  
1006 cancer journal for clinicians* 71: 7-33

1007 Skog J, Wurdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M,  
1008 Curry WT, Jr., Carter BS, Krichevsky AM, Breakefield XO (2008) Glioblastoma  
1009 microvesicles transport RNA and proteins that promote tumour growth and  
1010 provide diagnostic biomarkers. *Nat Cell Biol* 10: 1470-1476

1011 Song BI (2020) Nomogram using F-18 fluorodeoxyglucose positron emission  
1012 tomography/computed tomography for preoperative prediction of lymph node  
1013 metastasis in gastric cancer. *World J Gastrointest Oncol* 12: 447-456

1014 Takov K, Yellon DM, Davidson SM (2019) Comparison of small extracellular  
1015 vesicles isolated from plasma by ultracentrifugation or size-exclusion  
1016 chromatography: yield, purity and functional potential. *J Extracell Vesicles* 8:  
1017 1560809

1018 Tang W, Zhou M, Dorsey TH, Prieto DA, Wang XW, Ruppin E, Veenstra TD,  
1019 Ambs S (2018) Integrated proteotranscriptomics of breast cancer reveals  
1020 globally increased protein-mRNA concordance associated with subtypes and  
1021 survival. *Genome Med* 10: 94

1022 Thakur BK, Zhang H, Becker A, Matei I, Huang Y, Costa-Silva B, Zheng Y,  
1023 Hoshino A, Brazier H, Xiang J *et al* (2014) Double-stranded DNA in exosomes:  
1024 a novel biomarker in cancer detection. *Cell Res* 24: 766-769

1025 Thery C, Amigorena S, Raposo G, Clayton A (2006) Isolation and  
1026 characterization of exosomes from cell culture supernatants and biological  
1027 fluids. *Curr Protoc Cell Biol* Chapter 3: Unit 3 22

1028 Thery C, Ostrowski M, Segura E (2009) Membrane vesicles as conveyors of  
1029 immune responses. *Nat Rev Immunol* 9: 581-593

1030 Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO (2007)  
1031 Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism  
1032 of genetic exchange between cells. *Nat Cell Biol* 9: 654-659

1033 Vallejos CS, Gomez HL, Cruz WR, Pinto JA, Dyer RR, Velarde R, Suazo JF,  
1034 Neciosup SP, Leon M, de la Cruz MA *et al* (2010) Breast cancer classification  
1035 according to immunohistochemistry markers: subtypes and association with  
1036 clinicopathologic variables in a peruvian hospital database. *Clin Breast Cancer*  
1037 10: 294-300

1038 Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C,  
1039 Dykgers A, Rees M, Ramaswamy A, Muenst S, Soysal SD *et al* (2019) A  
1040 Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast  
1041 Cancer. *Cell* 177: 1330-1345 e1318

1042 Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C,  
1043 Pacey S, Baird R, Rosenfeld N (2017) Liquid biopsies come of age: towards  
1044 implementation of circulating tumour DNA. *Nat Rev Cancer* 17: 223-238

1045 Wang H, Song X, Huang Q, Xu T, Yun D, Wang Y, Hu L, Yan Y, Chen H, Lu D  
1046 *et al* (2019a) LGALS3 Promotes Treatment Resistance in Glioblastoma and Is  
1047 Associated with Tumor Risk and Prognosis. *Cancer Epidemiol Biomarkers  
Prev* 28: 760-769

1048 Wang HX, Gires O (2019) Tumor-derived extracellular vesicles in breast  
1049 cancer: From bench to bedside. *Cancer Lett* 460: 54-64

1050 Wang J, Liu Z, Zhang S, Wang X, Bai H, Xie M, Dong F, Ema H (2019b)  
1051 Lineage marker expression on mouse hematopoietic stem cells. *Exp Hematol*  
1052 76: 13-23 e12

1053 Wang M, Ji S, Shao G, Zhang J, Zhao K, Wang Z, Wu A (2018) Effect of  
1054 exosome biomarkers for diagnosis and prognosis of breast cancer patients.  
1055 *Clin Transl Oncol* 20: 906-911

1056 Wang YT, Shi T, Srivastava S, Kagan J, Liu T, Rodland KD (2020) Proteomic  
1057 Analysis of Exosomes for Discovery of Protein Biomarkers for Prostate and  
1058 Bladder Cancer. *Cancers (Basel)* 12

1059 Weaver VM, Petersen OW, Wang F, Larabell CA, Briand P, Damsky C, Bissell  
1060 MJ (1997) Reversion of the malignant phenotype of human breast cells in  
1061 three-dimensional culture and in vivo by integrin blocking antibodies. *J Cell  
Biol* 137: 231-245

1062 Wu D, Yan J, Shen X, Sun Y, Thulin M, Cai Y, Wik L, Shen Q, Oelrich J, Qian  
1063 X *et al* (2019) Profiling surface proteins on individual exosomes using a  
1064 proximity barcoding assay. *Nat Commun* 10: 3854

1065 Xavier-Ferrucio J, Scanlon V, Li X, Zhang PX, Lozovatsky L, Ayala-Lopez N,  
1066 Tebaldi T, Halene S, Cao C, Fleming MD *et al* (2019) Low iron promotes  
1067 megakaryocytic commitment of megakaryocytic-erythroid progenitors in  
1068 humans and mice. *Blood* 134: 1547-1557

1069 Xu R, Greening DW, Zhu HJ, Takahashi N, Simpson RJ (2016) Extracellular  
1070 vesicle isolation and characterization: toward clinical application. *J Clin Invest*  
1071 126: 1152-1162

1072

1073

1074 Yanez-Mo M, Siljander PR, Andreu Z, Zavec AB, Borrás FE, Buzas EI, Buzas  
1075 K, Casal E, Cappello F, Carvalho J *et al* (2015) Biological properties of  
1076 extracellular vesicles and their physiological functions. *J Extracell Vesicles* 4:  
1077 27066

1078 Yin ZQ, Liu JJ, Xu YC, Yu J, Ding GH, Yang F, Tang L, Liu BH, Ma Y, Xia YW  
1079 *et al* (2014) A 41-gene signature derived from breast cancer stem cells as a  
1080 predictor of survival. *J Exp Clin Cancer Res* 33: 49

1081 Yoneda K, Imanishi N, Ichiki Y, Tanaka F (2019) A liquid biopsy in primary lung  
1082 cancer. *Surg Today* 49: 1-14

1083 Zeng Y, Cao Y, Liu L, Zhao J, Zhang T, Xiao L, Jia M, Tian Q, Yu H, Chen S *et*  
1084 *al* (2019) SEPT9\_i1 regulates human breast cancer cell motility through  
1085 cytoskeletal and RhoA/FAK signaling pathway regulation. *Cell Death Dis* 10:  
1086 720

1087 Zhang W, Zhang J, Xu C, Li N, Liu H, Ma J, Zhu Y, Xie H (2012) LFQuant: a  
1088 label-free fast quantitative analysis tool for high-resolution LC-MS/MS  
1089 proteomics data. *Proteomics* 12: 3475-3484

1090 Zhang Y, Wang XF (2015) A niche role for cancer exosomes in metastasis.  
1091 *Nat Cell Biol* 17: 709-711

1092 Zhu R, Gires O, Zhu L, Liu J, Li J, Yang H, Ju G, Huang J, Ge W, Chen Y *et al*  
1093 (2019) TSPAN8 promotes cancer cell stemness via activation of sonic  
1094 Hedgehog signaling. *Nat Commun* 10: 2863

1095

1096 **Figure Legends**

1097 **Figure 1. Overview of the proteomic characterization of breast cancer**

1098 **sEVs.**

1099 A Overview of the experimental design and the number of samples for  
1100 proteomics analyses.

1101 B Clinical parameters are indicated in the heatmap.

1102 C Schematic diagram of the extraction process of serum-derived sEVs.

1103 D Representative TEM images of purified EVs. Scale bar–100 nm.

1104 E Immunoblots showing the expression levels of ALIX (PDCD6IP), CD63, CD9,  
1105 and TSG101 in the purified EVs.

1106 F Cumulative number of protein identifications. Red denotes BC samples (n =  
1107 126), yellow denotes BD samples (n = 17), and blue denotes HD samples (n  
1108 = 24).

1109 G The number of proteins identified in 167 samples. Red denotes BC samples  
1110 (n = 126), yellow denotes BD samples (n = 17), and blue denotes HD  
1111 samples (n = 24).

1112 H Venn diagram depicting the numbers of proteins detected in BC-, BD-, and  
1113 HD-sEVs.

1114

1115 **Figure 2. Proteomics features of BC-, BD- and HD-derived sEVs.**

1116 A The bubble plot indicates the distinctive biological pathways of BC, BD, and  
1117 HD. Red box, BC; yellow box, BD; blue box, HD. See Table S2.

1118 B Differentially expressed proteins in the distinctive biological pathways of BC,  
1119 BD, and HD. Fold change > 2 and one-way ANOVA  $p < 0.05$ .

1120 C Two proteins (MMP2 and TYMP) differentially expressed in BC, BD, and HD  
1121 ( $p$  value from Student's  $t$  test) and their association with clinical outcomes in  
1122 BC (2018, Tang et al., BC cohort, n = 118) ( $p$  value from log rank test).

1123 D sEV DAMP molecules were enriched in BC and found in > 50% of BC  
1124 samples, with > 2-fold difference and one-way ANOVA  $p < 0.05$ .

1125 E DAMP molecules enriched in BC-sEVs were significantly associated with  
1126 clinical outcomes in BC (2018, Tang et al., BC cohort, n = 118) ( $p$  value from  
1127 log rank test).

1128 F Distinctive tumour microenvironment in BC. See Table S2.

1129 G Functions of BC-derived sEVs. These sEVs impact the tumour  
1130 microenvironment by promoting tumour cell growth and progression,  
1131 modulating immune responses, regulating angiogenesis and inducing  
1132 metastatic behaviour through MEPs, endothelial cells, and Mv endothelial  
1133 cells.

1134 H Classification error matrix of the training set (70%) and test set (30%) for the  
1135 8 proteins using the random forest classifier. The number of samples  
1136 identified is noted in each box.

1137 I Proteins with the highest predictive values in classifying BC, BD and HD  
1138 samples by XGBoost.

1139

1140 **Figure 3. Proteomic landscapes of four clinical subtypes of BC-derived**  
1141 **sEVs.**

1142 A Differentially expressed proteins in luminal A, luminal B, Her2-enriched, and  
1143 TNBC samples and found in > 50% of the corresponding samples, with >  
1144 2-fold difference from the other three subtypes.

1145 B Gene Ontology biological processes (GOBPs) revealed pathways that were  
1146 significantly enriched in luminal A, luminal B, Her2-enriched and TNBC  
1147 samples (Fisher's exact test,  $p < 0.05$ ). See Table S3.

1148 C Differentially expressed proteins in luminal A, luminal B, Her2-enriched, and  
1149 TNBC samples. See Table S3.

1150 D Classification error matrix of the training set (70%) and test set (30%) for the  
1151 61 proteins using the XGBoost classifier. The number of samples identified  
1152 is noted in each box. The bar chart above represents the predictive  
1153 specificity of each subtype. The bar chart on the right represents the  
1154 predictive sensitivity of each subtype.

1155 **Figure 4. Potential prognostic biomarkers for IBC patients with lymph**  
1156 **node metastases.**

1157 A GSEA of the proteomic data of 105 breast cancer samples revealed that  
1158 adipogenesis was significantly upregulated in IBC\_LN. IBC\_LN: IBC  
1159 patients with lymph node metastases.

1160 B Distinctive tumour microenvironment between IBC\_Pure and IBC\_LN.  
1161 IBC\_Pure: IBC patients without lymph node metastases; IBC\_LN: IBC  
1162 patients with lymph node metastases. See Table S4.

1163 C Correlation between adipocytes and the pathway of positive regulation of  
1164 VEGF production. Spearman rho = 0.412, Wilcoxon rank sum test,  $p$  =  
1165 1.242e-05.

1166 D Comparison of the scores of positive regulation of VEGF production  
1167 between IBC\_LN and IBC\_Pure. The  $p$  value was calculated by the  
1168 Wilcoxon rank sum test. The line and box represent median and upper and  
1169 lower quartiles, respectively.

1170 E Correlation between the pathway of positive regulation of VEGF production  
1171 and MPPs. Correlation coefficients and  $p$  values were calculated by the  
1172 Spearman correlation method.

1173 F Molecules highly associated with platelets were expressed in IBC\_Pure and  
1174 IBC\_LN.

1175 G The pattern diagram shows the process by which adipocytes activate MPPs  
1176 to generate MEPs and MKs through positive regulation of VEGF production

1177 and finally produce platelets. The produced platelets helped breast cancer  
1178 cells migrate to the lymph nodes.

1179 H Representative immunohistochemical images of adipocytes labelled with  
1180 PPRGg. Images revealed that adipocytes proliferically grew in lymph node  
1181 metastases of BC compared to primary breast cancer.

1182 I Representative fluorescence microscopy images of MPPs labelled with  
1183 CD45RA (green), CD34 (red), and CD38 (yellow). Images revealed the  
1184 presence of MPPs in lymph node metastases of BC, which were rare in  
1185 normal lymph nodes and primary breast cancer.

1186 J Classification error matrix of the training set (70%) and test set (30%) for the  
1187 12 proteins using the XGBoost classifier. The number of samples identified  
1188 is noted in each box.

1189 K Classification error matrix of the external validation set (2016, CPTAC, BC  
1190 cohort, n = 77) for the 12 proteins using the XGBoost classifier. The number  
1191 of samples identified is noted in each box.

1192 L Proteins with the highest predictive values in classifying IBC\_Pure and  
1193 IBC\_LN samples by XGBoost.

1194

1195 **Figure 5. Potential sEV survival biomarkers for the distant metastases of**  
1196 **BC.**

1197 A The bubble plot indicates the overrepresented pathways in D-MET  
1198 compared to DCIS. See Table S5.

1199 B Differentially expressed proteins between distant metastases and DCIS  
1200 samples with > 2-fold difference and two-way Student's *t* test *p* < 0.05.

1201 C Potential markers of distant metastasis were significantly associated with  
1202 clinical outcomes in BC (2018, Tang et al., BC cohort, n = 118) (*p* value from  
1203 log rank test).

1204 D DAVID (KEGG gene sets) analyses of the proteomic data of 21 BC patients  
1205 with distant metastases revealed pathways that were significantly altered in  
1206 lung metastases (M-Lung, n = 3), liver metastases (M-Liver, n = 4), and  
1207 bone metastases (M-Bone, n = 7) (Fisher's exact test,  $p < 0.05$ ).  
1208 E GMDS was specifically highly expressed in M-Liver. ns, no significance;  $*p <$   
1209 0.05 by one-way Student's *t* test.  
1210 F P4HB was specifically highly expressed in M-Bone. ns, no significance;  $**p <$   
1211 0.01 by one-way Student's *t* test.  
1212 G C6, TUBB, SERPING1 and VCL were specifically highly expressed in  
1213 M-Lung. ns, no significance;  $*p < 0.05$ ,  $**p < 0.01$  by one-way Student's *t*  
1214 test.  
1215 H High expression of VCL was associated with poor prognosis in BC (2018,  
1216 Tang, et al. BC cohort, n = 126).  
1217

1218 **Figure 6. Potential BC-derived sEV molecules govern organ-specific**  
1219 **metastasis.**

1220 A Distinctive tumour microenvironment of M-Lung, M-Liver, and M-Bone  
1221 samples.  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$  by ANOVA.  
1222 B Boxplot showing the relative abundance of chondrocytes in the distant  
1223 metastases of BC. *P* value from one-way Student's *t* test.  
1224 C Boxplot showing the relative abundance of myocytes in the distant  
1225 metastases of BC. *P* value from one-way Student's *t* test.  
1226 D Boxplot showing the relative abundance of fibroblasts in the distant  
1227 metastases of BC. *P* value from one-way Student's *t* test.  
1228 E sEV ITGA1, S100A8 and S100A11 molecular levels in M-Bone. *P* value from  
1229 one-way Student's *t* test.  
1230 F Protein expression of ITGA1 in DCIS, M-Liver, M-Lung, and M-Bone tissues  
1231 detected by using immunohistochemistry.

1232 G Protein expression of S100A8 in DCIS, M-Liver, M-Lung, and M-Bone  
1233 tissues detected by using immunohistochemistry.

1234 H Model of sEV-mediated organotropic tumour dissemination. BC-derived  
1235 sEVs are taken up by organ-specific resident cells in metastatic organs  
1236 based on integrin expression.

1237

1238 **Appendix Figure 1. Proteomic characterization of BC-derived sEVs,**  
1239 **related to Figure 1**

1240 A NanoSight profiles showing the size distribution of serum-derived sEVs  
1241 isolated from BC, BD, and HD. Red denotes BC-derived sEVs, yellow  
1242 denotes BD-derived sEVs, and blue denotes HD-derived sEVs.

1243 B Identification of 24 sEV protein markers in our proteomic data.

1244 C Distribution of log10-transformed iBAQ abundance of identified proteins in  
1245 167 proteome samples that passed quality control. Red denotes BC  
1246 samples (n = 126), yellow denotes BD samples (n = 17), and blue denotes  
1247 HD samples (n = 24). In the box plots, the middle bar represents the median,  
1248 and the box represents the interquartile range; bars extend to 1.5 × the  
1249 interquartile range.

1250

1251 **Appendix Figure 2. Proteomics features of BC-, BD- and HD-derived**  
1252 **sEVs, related to Figure 2**

1253 A PCA of 1,734 proteins in 167 samples. Red, BC (n = 126); yellow, BD (n =  
1254 17); blue, HD (n = 24).

1255 B Schematic diagram of the structural distribution of damage-associated  
1256 molecular patterns (DAMPs) in sEVs (left). Venn diagram showing the  
1257 number of DAMPs detected in BC, BD, and HD samples (right). See Table  
1258 S2.

1259 C sEV DAMP molecules enriched in BC were significantly associated with  
1260 clinical outcomes in BC (2018, Tang et al., BC cohort, n = 118) (*p* value from  
1261 log rank test).

1262 D The dataset was split randomly into training (70%) and test sets (30%) at the  
1263 patient level. A machine learning algorithm, XGBoost, was used for model  
1264 development, training, and validation. Receiver operating characteristic  
1265 (ROC) analysis was used to evaluate the performance of the classifier on  
1266 the test dataset.

1267

1268 **Appendix Figure 3. Proteomic landscapes of four clinical subtypes of**  
1269 **BC-derived sEVs, related to Figure 3**

1270 A PCA of 1,308 proteins in 114 samples. Orange, luminal A (n = 20); green,  
1271 luminal B (n = 50); purple, Her2-enriched (n = 21); and blue, TNBC (n = 23).

1272 B Proteins with the highest predictive values in classifying luminal A, luminal B,  
1273 Her2-enriched, and TNBC samples by XGBoost.

1274

1275 **Appendix Figure 4. Potential prognostic biomarkers for IBC patients with**  
1276 **lymph node metastases, related to Figure 4**

1277 A PCA of 1,116 proteins in 105 samples. Blue, invasive breast cancer with  
1278 lymph node metastases (IBC\_Pure, n = 54); red, invasive breast cancer  
1279 without lymph node metastases (IBC\_LN, n = 51).

1280 B Differentially expressed proteins between IBC\_Pure and IBC\_LN samples  
1281 that were found in > 50% of the corresponding samples, with > 2-fold  
1282 difference and Student's *t* test *p* < 0.05.

1283 C Comparison of the scores of adipocytes between the IBC\_LN group and the  
1284 IBC\_Pure group. The *p* value was calculated by the Wilcoxon rank sum test.  
1285 The line and box represent median and upper and lower quartiles,  
1286 respectively.

1287 D Correlation between adipogenesis and adipocytes. Spearman rho = 0.188, *p*  
1288 value = 5.507e-02.

1289 E Comparison of the MPP scores between the IBC\_LN group and the  
1290 IBC\_Pure group. The *p* value was calculated by the Wilcoxon rank sum test.  
1291 The line and box represent the median and upper and lower quartiles,  
1292 respectively.

1293 F Correlation between MPPs and the coagulation pathway. Spearman rho =  
1294 0.295, *p* value = 2.216e-03.

1295 G Correlation between platelets and the coagulation pathway. Spearman rho =  
1296 0.209, *p* value = 3.225e-02.

1297 H Comparison of the platelet scores between the IBC\_LN group and the  
1298 IBC\_Pure group. The *p* value was calculated by the Wilcoxon rank sum test.  
1299 The line and box represent the median and upper and lower quartiles,  
1300 respectively.

1301 I Molecules that are highly associated with platelets.

1302 J Representative fluorescence microscopy images of MEPs labelled with  
1303 CD71 (green), CD38 (red), and CD45RA (yellow). Images revealed the  
1304 presence of MEPs in lymph node metastases of BC, which were rare in  
1305 normal lymph nodes and primary breast cancer.

1306 K Platelet counts in the blood of IBC patients with lymph node metastasis (*n* =  
1307 43) and IBC patients without lymph node metastasis (*n* = 45).

1308 L The dataset was randomly split into training (70%) and test sets (30%) at the  
1309 patient level. A machine learning algorithm, XGBoost, was used for model  
1310 development, training, and validation. Receiver operating characteristic  
1311 (ROC) analysis was used to evaluate the performance of the classification  
1312 on the test dataset.

1313

1314 **Appendix Figure 5. Potential sEV survival biomarkers for the distant**  
1315 **metastases of BC, related to Figure 5**

1316 A Potential markers of distant metastasis were significantly associated with  
1317 clinical outcomes in BC (2018, Tang et al., BC cohort, n = 118 and 2014, Liu  
1318 et al., BC cohort, n = 126) (*p* value from log rank test).

1319

1320 **Appendix Figure 6. Potential molecules present on IBC-derived sEVs**  
1321 **target them to specific organs, related to Figure 6**

1322 A IHC score of ITGA1 in DCIS (n = 4), M-Liver (n = 4), M-Lung (n = 4), and  
1323 M-Bone (n = 8). *P* value from two-way Student's *t* test.

1324 B IHC score of S100A8 in DCIS (n = 3), M-Liver (n = 3), M-Lung (n = 4), and  
1325 M-Bone (n = 8). *P* value from two-way Student's *t* test.

1326 C sEV ITGA7, S100A9 and S100A13 molecular levels in M-Liver. *P* value from  
1327 one-way Student's *t* test.

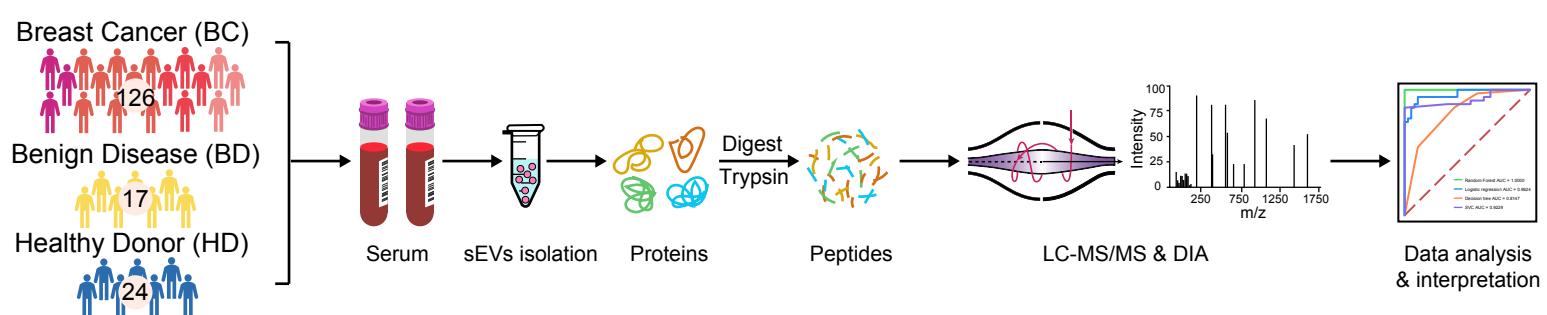
1328 D sEV ITGB3, S100A2B and S100A7A molecular levels in M-Lung. *P* value  
1329 from one-way Student's *t* test.

1330 E The protein expression of S100A13 in DCIS (n = 3), M-Liver (n = 3), M-Lung  
1331 (n = 4), and M-Bone (n = 8) tissues was detected by using  
1332 immunohistochemistry (left); IHC score of S100A13 in M-Liver. *P* value from  
1333 two-way Student's *t* test (right).

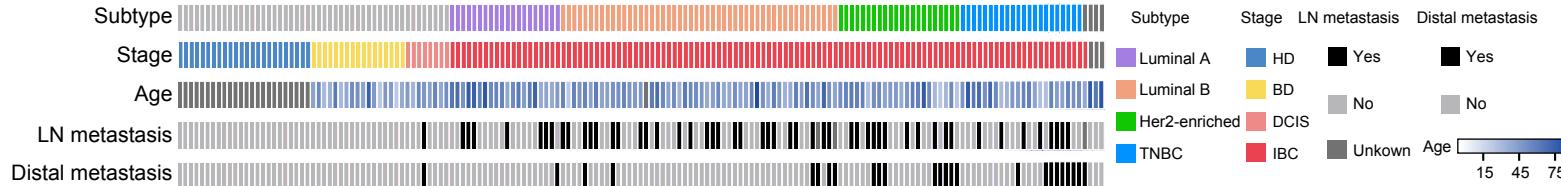
1334 F The protein expression of S100A7A in DCIS (n = 3), M-Liver (n = 3), M-Lung  
1335 (n = 4), and M-Bone (n = 8) tissues was detected by using  
1336 immunohistochemistry (left); IHC score of S100A7A in M-Lung. *P* value from  
1337 two-way Student's *t* test (right).

# Figure 1

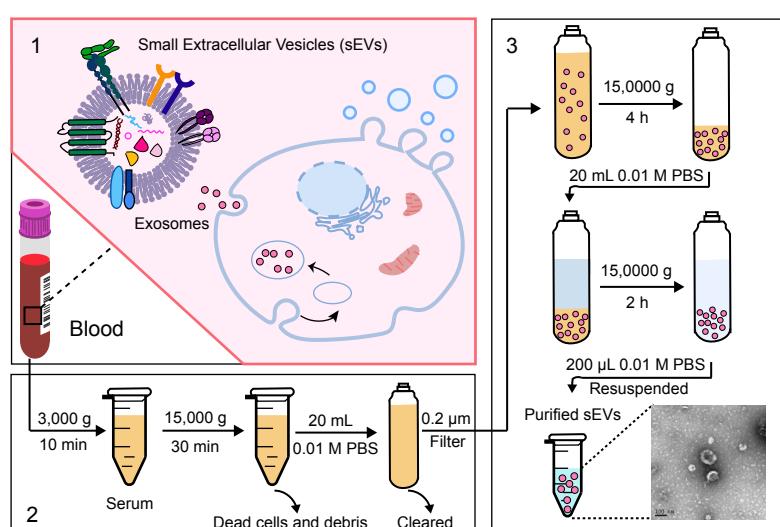
A



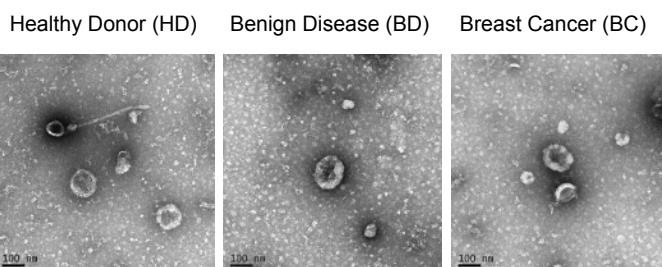
B



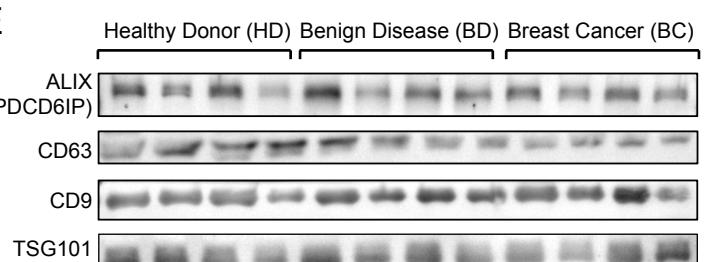
C



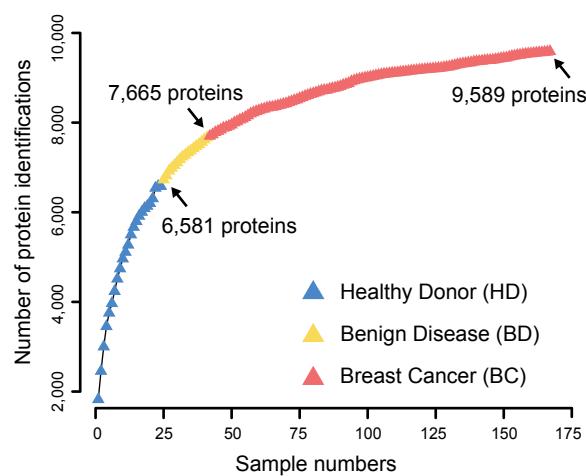
D



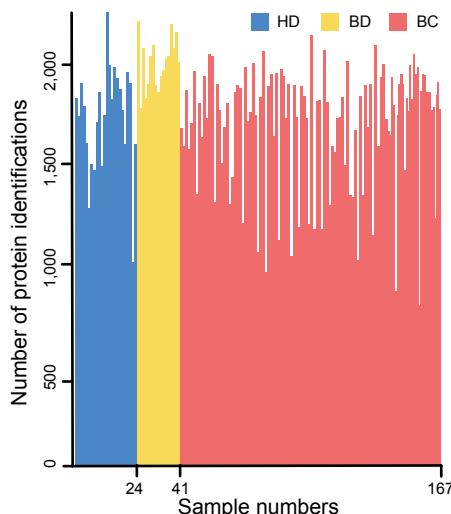
E



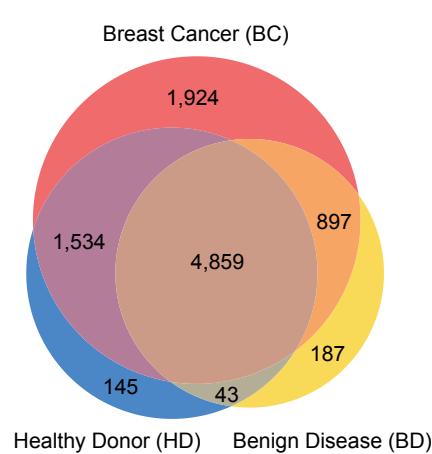
F



G

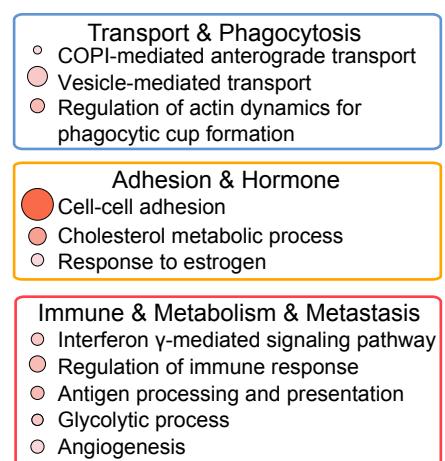


H

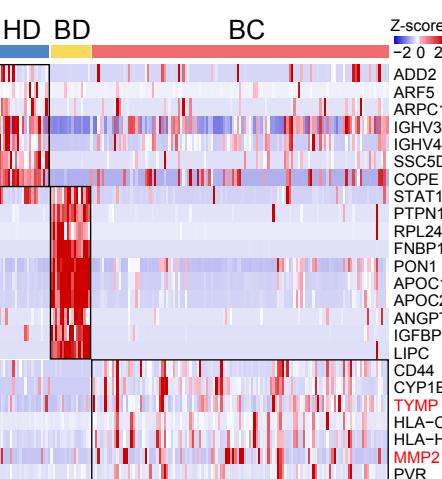


## Figure 2

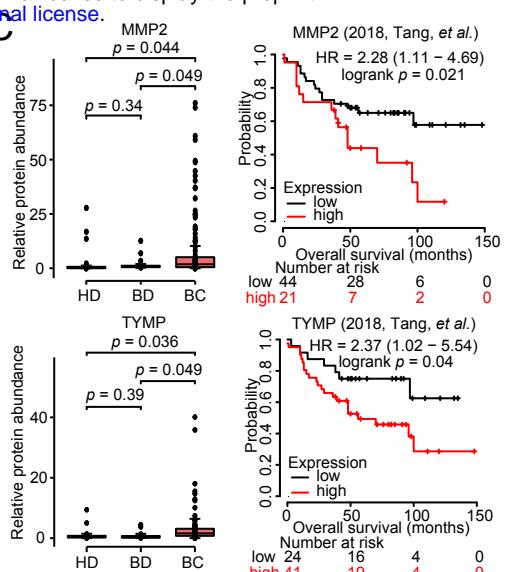
A



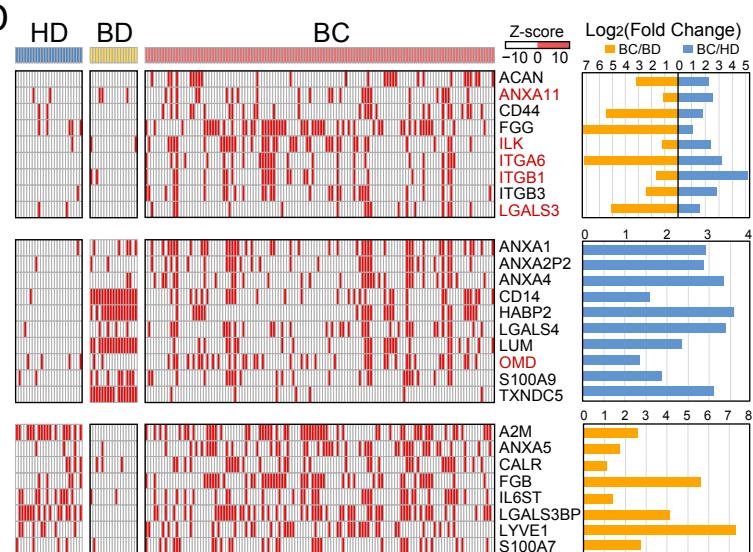
B



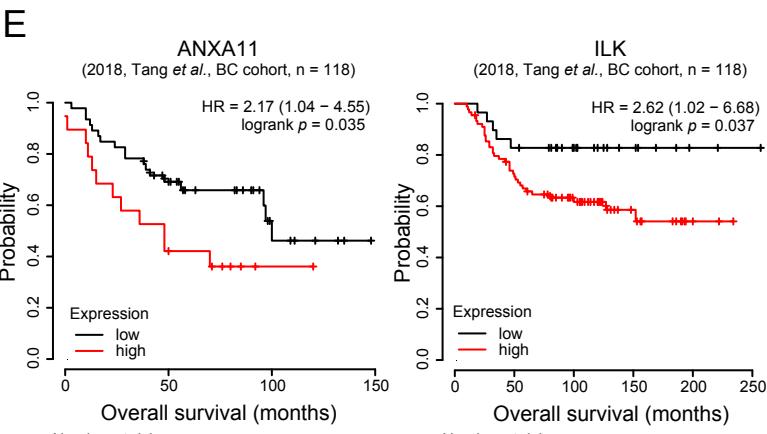
C



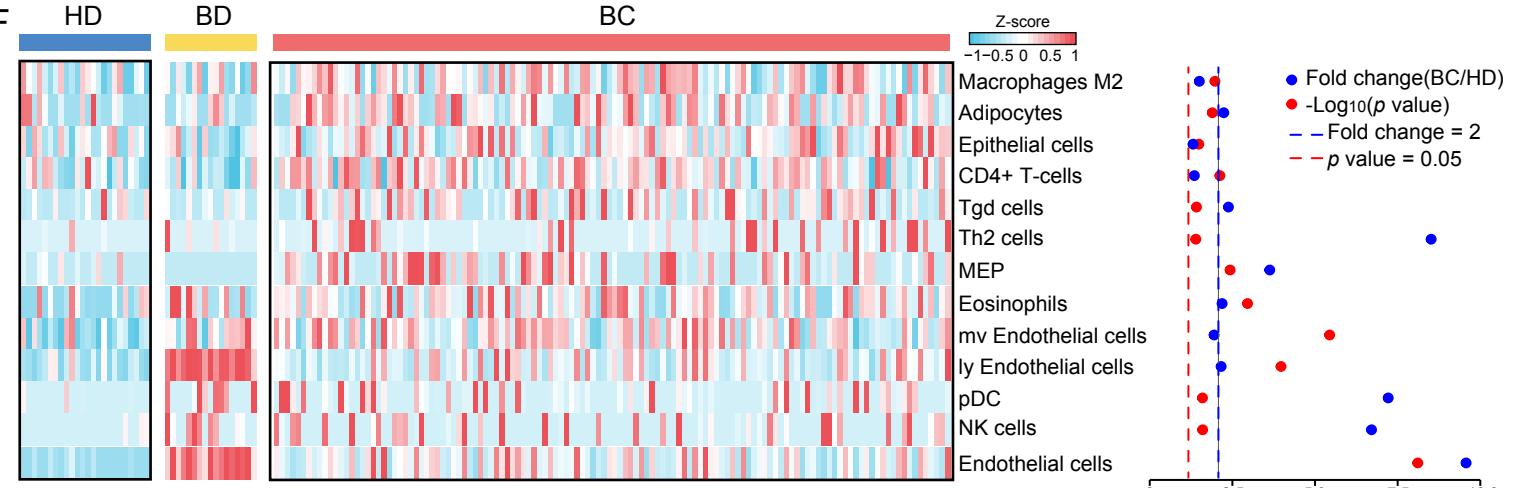
D



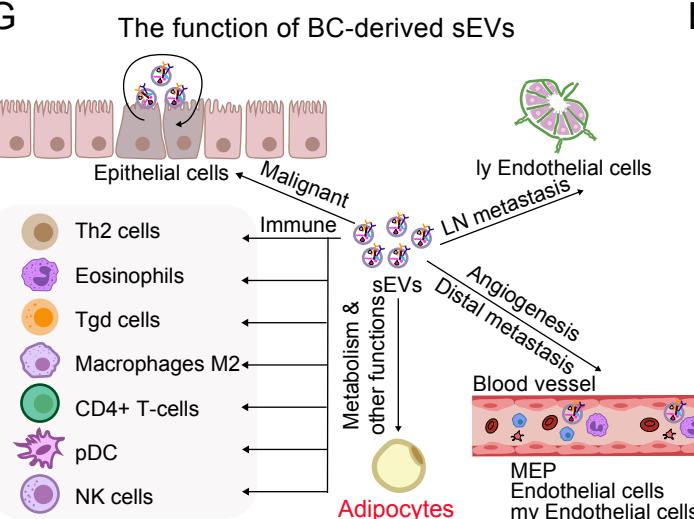
E



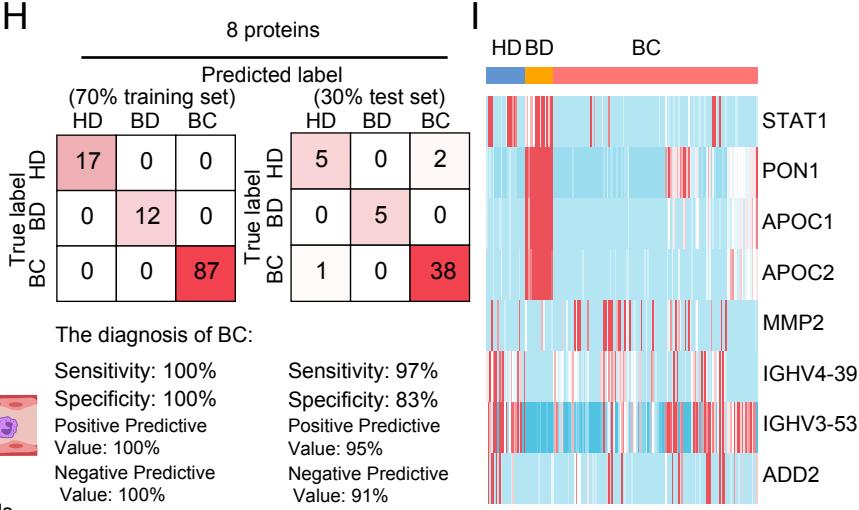
F



G



H



I

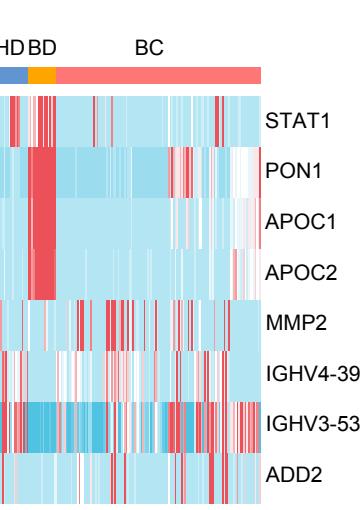
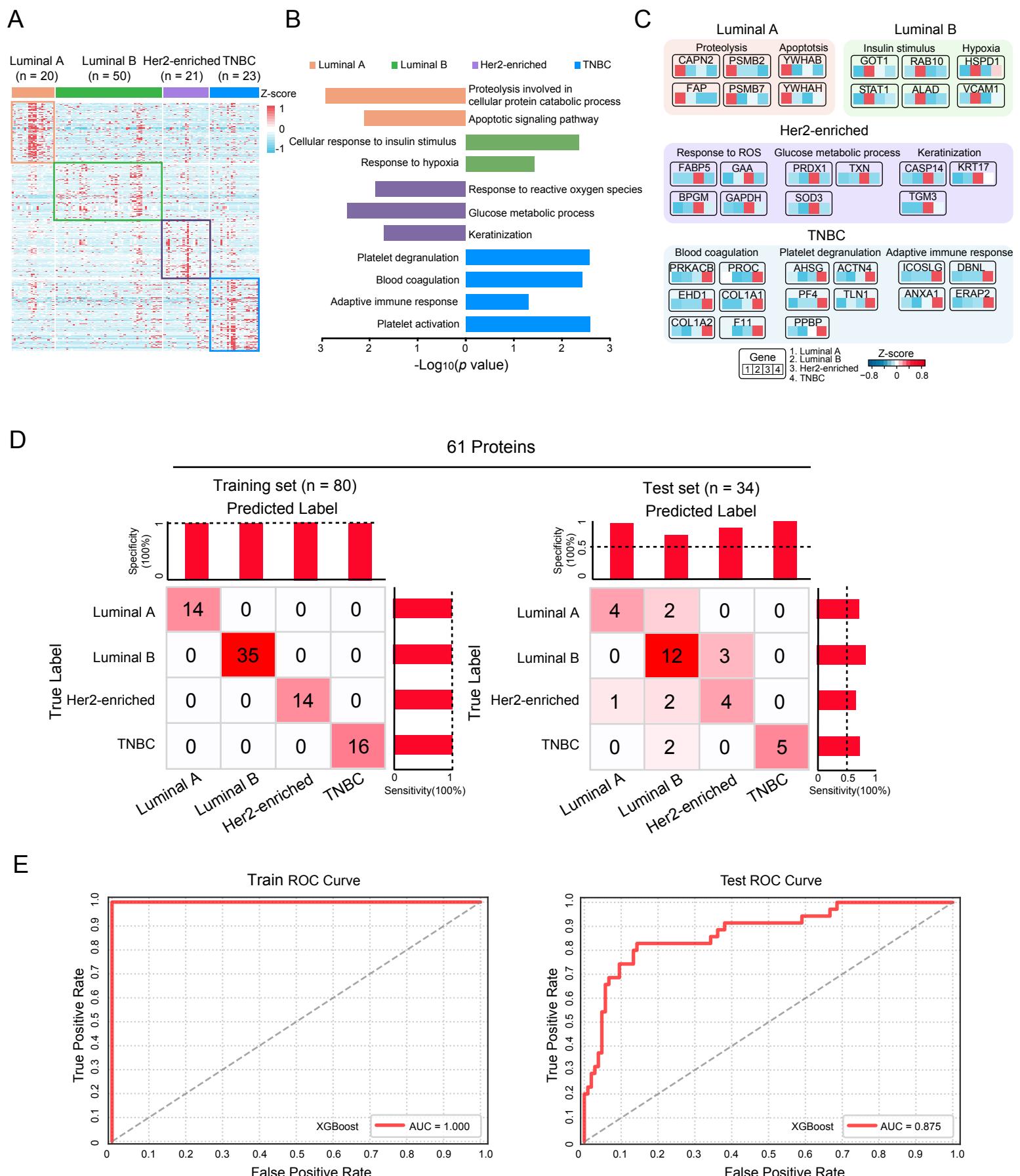
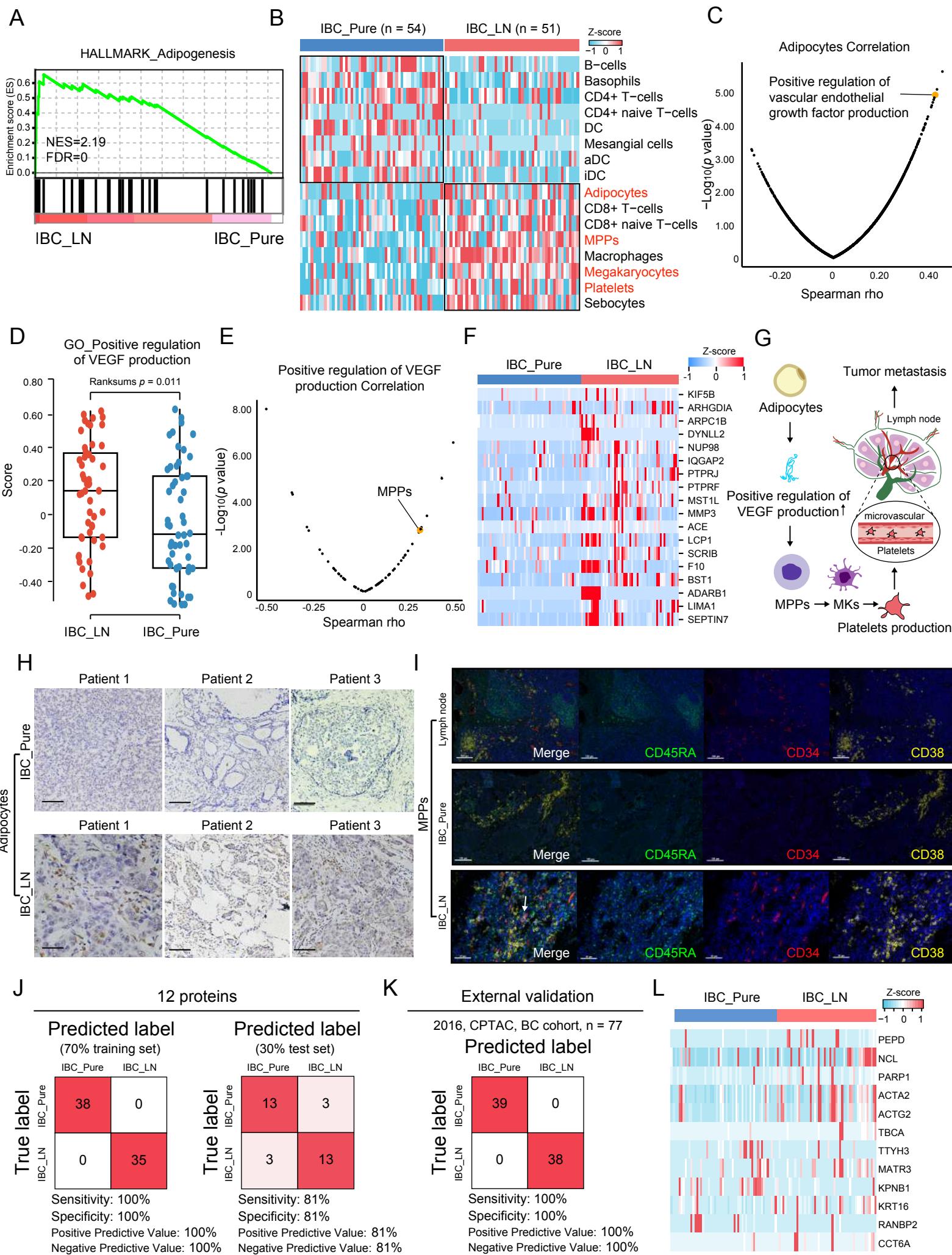


Figure 3

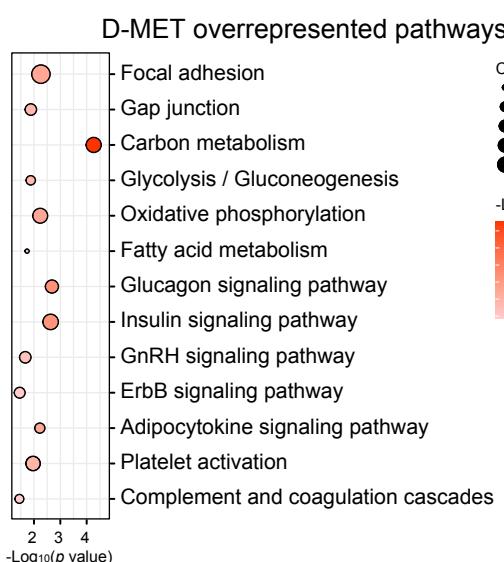


# Figure 4

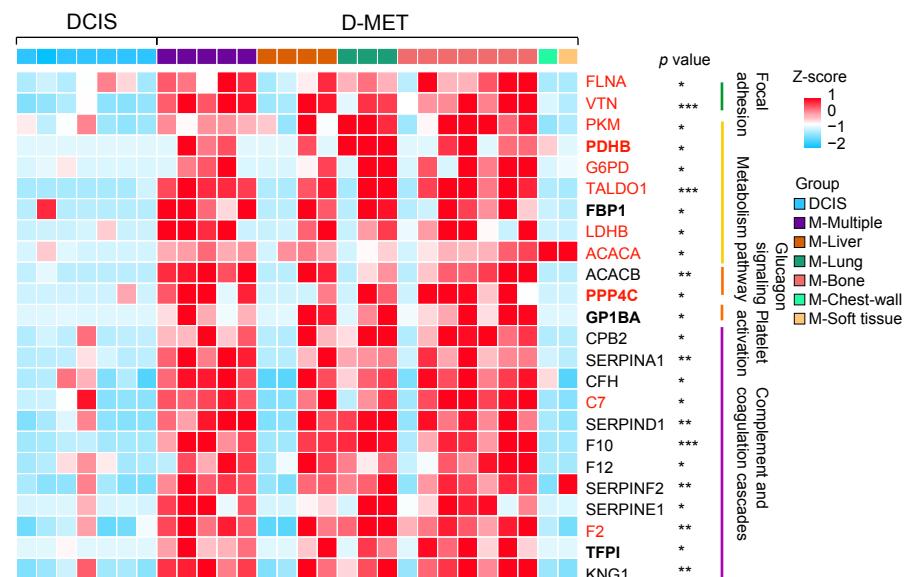


## Figure 3

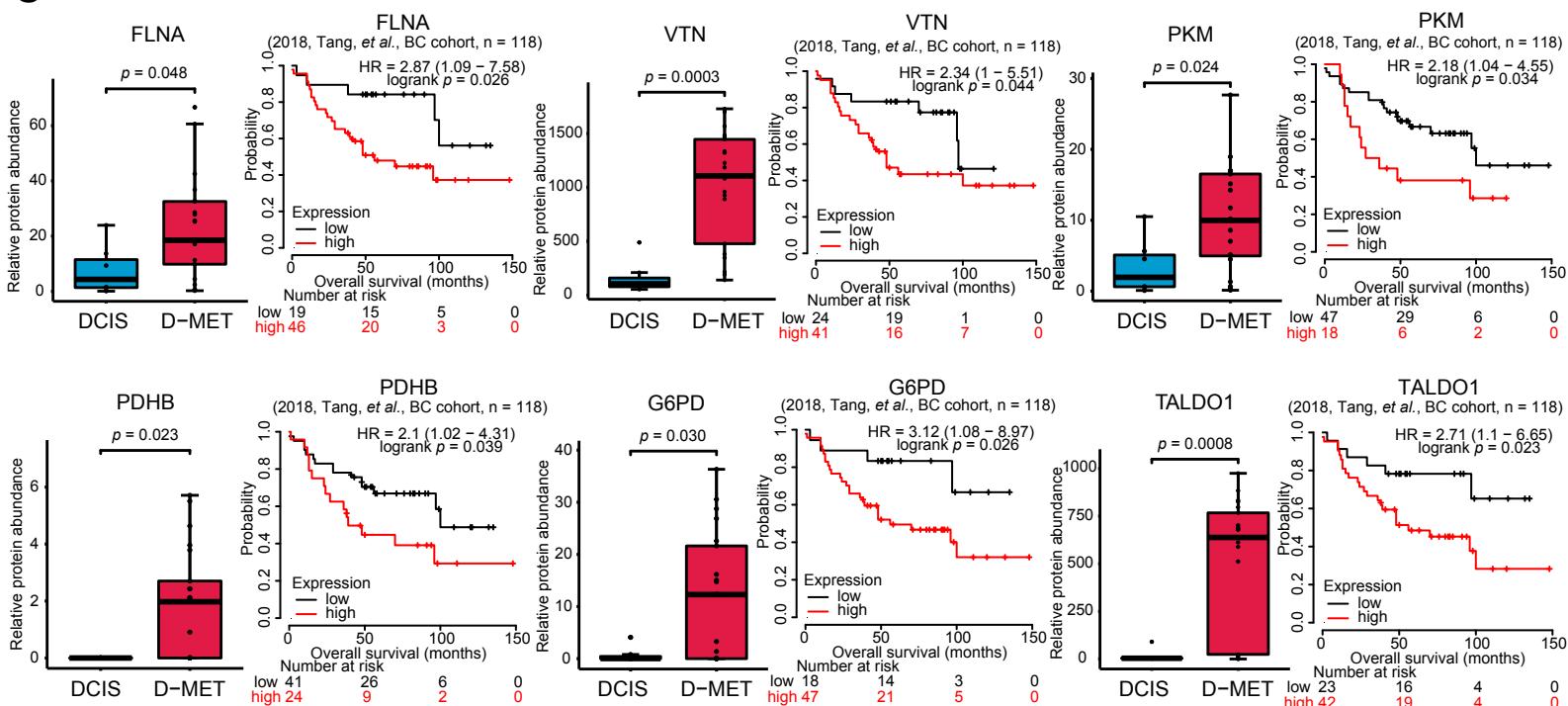
A



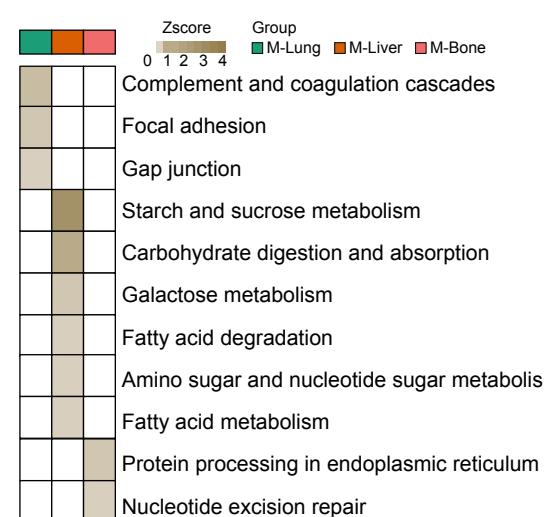
B



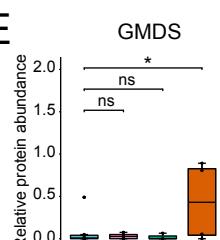
C



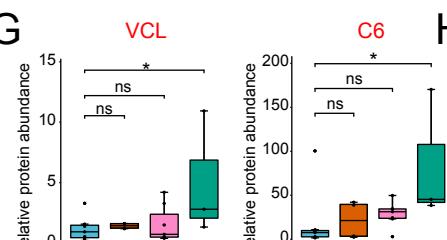
D



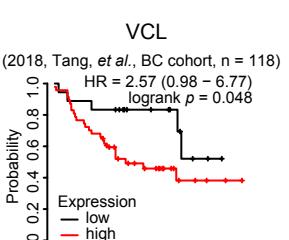
E



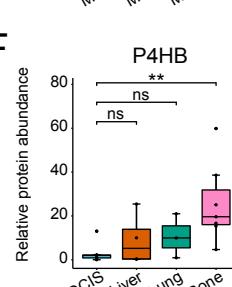
G



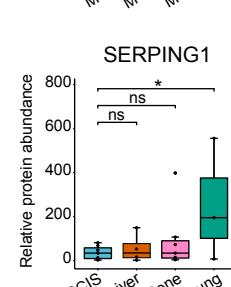
H



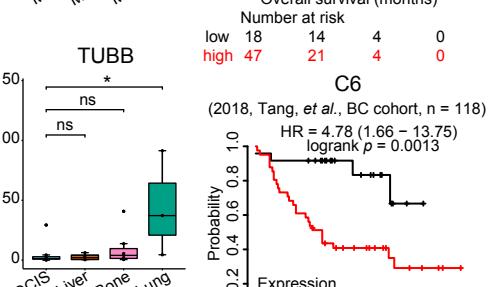
F



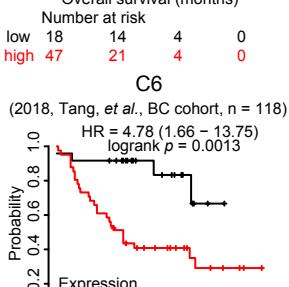
G



H

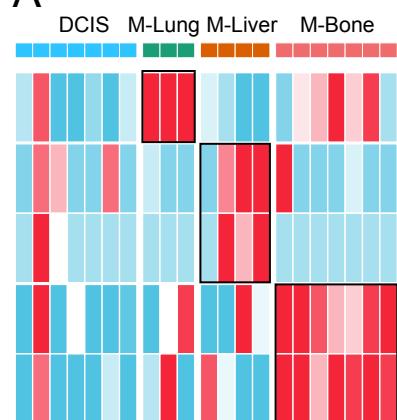


C6

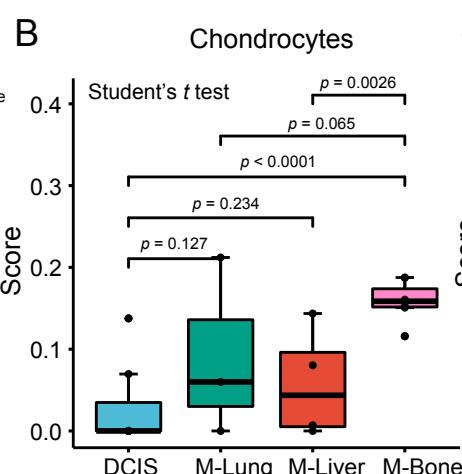


## Figure 6

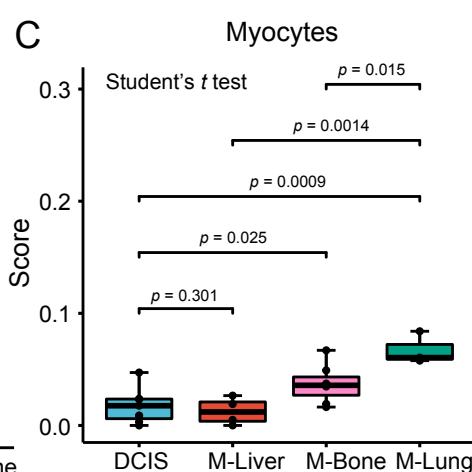
A



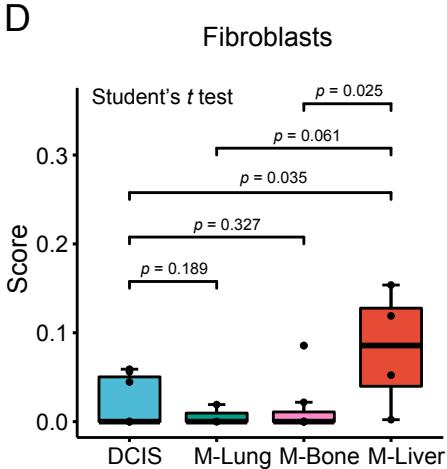
B



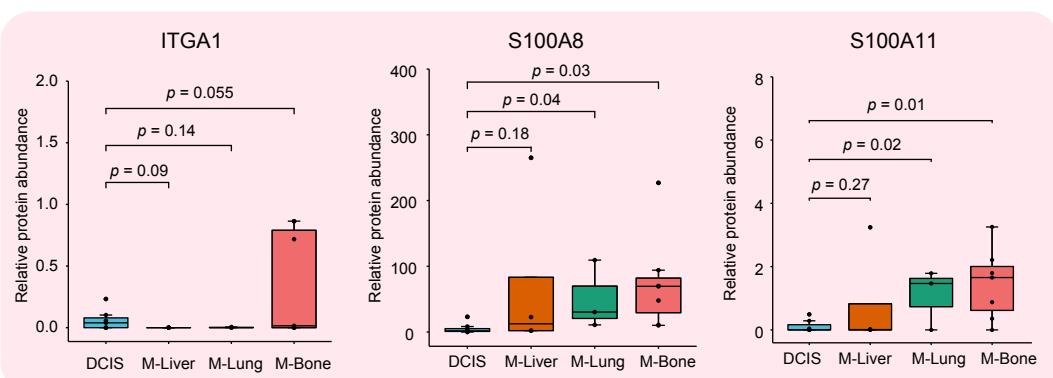
C



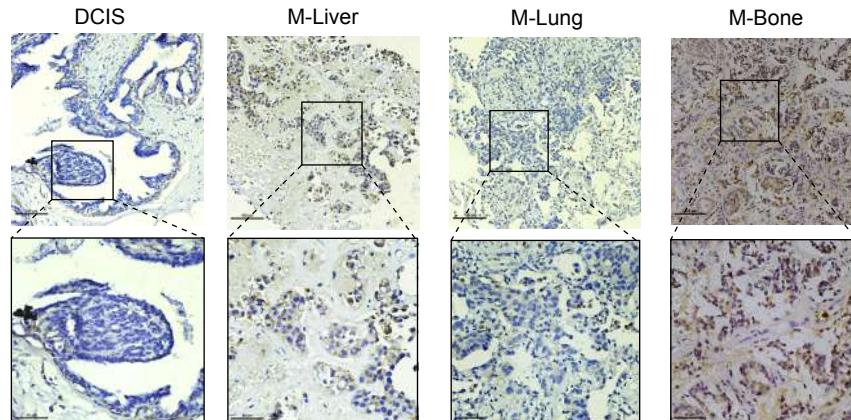
D



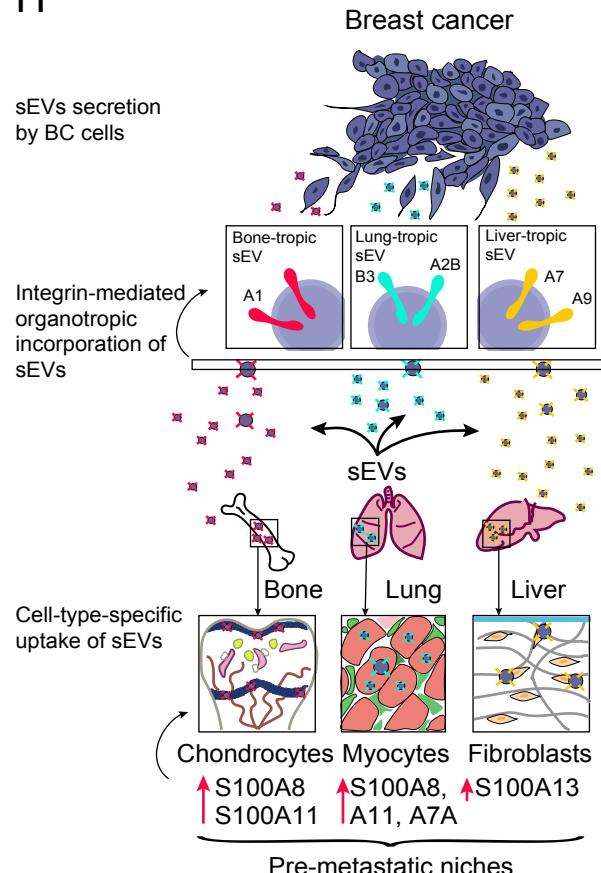
E



F



H



G

