The expression patterns of different cell types and their interactions in the tumor microenvironment are predictive of breast cancer patient response to neoadjuvant chemotherapy

Saugato Rahman Dhruba[1], Sahil Sahni[1], Binbin Wang[1], Di Wu[2], Padma Sheila Rajagopal[1,3], Yael Schmidt[1], Eldad D. Shulman[1], Sanju Sinha[1,4], Stephen-John Sammut[5,6], Carlos Caldas[7,8], Kun Wang[1,*], Eytan Ruppin[1,*]

[1] Cancer Data Science Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[2] Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[3] Women's Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[4] NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, San Diego, CA, USA.

[5] Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK.

[6] The Royal Marsden Hospital NHS Foundation Trust, London, UK.

[7] Institute of Metabolic Science, School of Clinical Medicine, University of Cambridge, Cambridge, UK.

[8] Department of Clinical Biochemistry, University of Cambridge, Cambridge, UK.

[*] Co-corresponding authors (eytan.ruppin@nih.gov, kun.wang@nih.gov)

ABSTRACT

The tumor microenvironment (TME) is a complex ecosystem of diverse cell types whose interactions govern tumor growth and clinical outcome. While the TME's impact on immunotherapy has been extensively studied, its role in chemotherapy response remains less explored. To address this, we developed DECODEM (DEcoupling Cell-type-specific Outcomes

using DEconvolution and Machine learning), a generic computational framework leveraging cellular deconvolution of *bulk transcriptomics* to associate the gene expression of individual cell types in the TME with clinical response. Employing DECODEM to analyze the gene expression of breast cancer (BC) patients treated with neoadjuvant chemotherapy, we find that the gene expression of specific immune cells (*myeloid*, *plasmablasts*, *B-cells*) and stromal cells (*endothelial*, *normal epithelial*, *CAFs*) are highly predictive of chemotherapy response, going beyond that of the malignant cells. These findings are further tested and validated in a single-cell cohort of triple negative breast cancer. To investigate the possible role of immune cell-cell interactions (CCIs) in mediating chemotherapy response, we extended DECODEM to DECODEMi to identify such CCIs, validated in single-cell data. Our findings highlight the importance of active pre-treatment immune infiltration for chemotherapy success. The tools developed here are made publicly available and are applicable for studying the role of the TME in mediating response from readily available bulk tumor expression in a wide range of cancer treatments and indications.

INTRODUCTION

The tumor microenvironment (TME) is a complex and multifaceted ecosystem that encompasses a plethora of cell types including immune cells (such as T- and B-lymphocytes, macrophages, dendritic cells *etc.*), stromal cells (such as fibroblasts, endothelial cells *etc.*), and the extracellular matrix. It plays crucial roles in regulating tumor progression and response to therapy. Emerging evidence suggest the context-specific behavior of the TME as either tumor-suppressive or supportive, presenting an attractive target for therapeutic intervention [1][2]. Numerous studies have elucidated the critical role of TME in mediating response to immunotherapy, however, insights into how the TME is relevant to chemotherapy response remains limited. Here we focus on studying the influence of the TME on chemotherapy response in breast cancer (BC), where neoadjuvant chemotherapy (NAC) is used prior to primary surgery to enable tumor downstaging and increase the likelihood of breast-conserving surgery [3][4][5].

Recent studies have underscored the relevance of the BC TME in determining patient response to chemotherapy. For instance, chemotherapy can induce dynamic changes in the TME

that lead to an increase in regulatory T-cells and myeloid-derived suppressor-like cells, calling for a balance in immune population for positive treatment outcomes [6]. Lymphocyte density and tumor infiltrating lymphocytes have consistently been associated with BC patients achieving pathological complete response (pCR) following NAC and anti-HER2 therapy [7][8][9][10][11]. Additionally, multiple studies have reported the active involvement of TME in therapy resistance [12][13][14], with macrophages, endothelial, and BC stem cells promoting chemoresistance [15][16][17], and B-cells and plasma cells displaying varying roles in NAC response [18][19][20]. However, to our knowledge, no study has yet attempted to quantify the contribution and predictive power of each individual cell type in the TME to chemotherapy response in a systematic and rigorous manner. We, therefore, aim to investigate three fundamental research questions: (1) Can we apply a rigorous computational approach to delineate the cell-type-specific transcriptome predictive power in the BC TME from bulk transcriptomics data? (2) Which are the cell types in the BC TME and the corresponding pathways whose gene expression are most predictive of chemotherapy response? (3) Does aggregating the inferred expression of multiple predictive cell types in the BC TME lead to a more accurate predictor?

Current advances in precision oncology have highlighted the utility of artificial intelligence [1][2][21][22][23][24] in guiding treatment using diverse data types such as genomics [1][25][26], transcriptomics [29][30][31][32][33], histopathology [34][35][36][37][38][39], or through multimodal approaches [11][27][28][40][41][42]. However, investigations into neoadjuvant therapy response often encounter limitations like small sample size, treatment heterogeneity, or inadequate capturing of the TME's complexity, as highlighted by Sammut et al. [11]. In their pioneering work, Sammut et al. presented a novel multiomic predictor of neoadjuvant therapy response in BC [11]. They have leveraged machine learning (ML) to incorporate 34 features extracted from clinical, DNA, RNA, pathology, and treatment information to develop a single model to predict patient response across all BC subtypes and different treatment regimens with high accuracy. Expanding on their groundwork and addressing the limitations, our study aims to decouple the cell-type-specific effects to NAC response in the BC TME from transcriptomics alone

by focusing on the HER2-negative patients, where NAC is recommended to increase the likelihood of achieving pCR following surgery, especially in triple-negative breast cancer (TNBC) [49].

To systematically explore the association between the TME and chemotherapy response, we developed a computational framework, **DECODEM (DEcoupling Cell-type-specific Outcomes using DEconvolution and Machine learning).** This framework leverages cellular deconvolution with machine learning to elucidate the association of gene expression of diverse cell types in the TME to patient response to a given treatment. Recognizing that these phenotype effects are intricately regulated by the interactions among relevant cell types within the TME, we further extended DECODEM to **DECODEMi**, where 'i' stands for interaction. DECODEMi incorporates the *inferred* cell-to-cell communications to pinpoint the key cell-cell interactions (CCIs) influencing treatment response. As mentioned, we demonstrate the utility of our tools in an example scenario of investigating the role of the BC TME in modulating response to NAC. A recent study by Kester et al. investigated the association of cellular compositions of the TME with patient survival in BC [43], but to our knowledge, the current study is the first to analyze the *cell-type-specific transcriptomic profiles* in the BC TME to objectively quantify their association with treatment response.

Through the application of DECODEM, we illustrate the markedly improved predictive capabilities of cell-type-specific transcriptome over bulk transcriptome, untangling the individual contributions of different cell types in treatment response. Through the application of DECODEMi, we further pinpoint the notable immune interactions among these cell types. Together all these findings point to the necessity of an active pre-treatment immune system to facilitate positive treatment outcome, as supported by Sammut et al. [11]. Therefore, from a methodological perspective, our study provides tools to systematically assess the role of individual cellular components within the TME in anticancer therapy response (DECODEM), and to identify the prominent cell-to-cell communications associated with this response (DECODEMi). These tools will be made publicly available upon publication.

RESULTS

Overview of the analysis

DECODEM and DECODEMi developed and presented here are new computational frameworks designed to explore the predictive powers of individual cell types and relevant cell-cell interactions, respectively, within the TME to response to therapy. They harness the prowess of CODEFACS and LIRICS, two recent tools developed in our laboratory [44]. CODEFACS facilitates the extraction of the cell-type-specific expression profiles from bulk gene expression data for each patient, while LIRICS discerns the activity of cell-type-specific ligand-receptor interactions within the patient's TME. In essence, DECODEM takes the pre-treatment bulk gene expression as input and outputs the quantified predictive power (in terms of AUC and AP *i.e.*, the area under the receiver operating characteristics (ROC) curve and average precision, equivalent to the area under the precision-recall curve, respectively) of each cell type to treatment response. DECODEMi then follows through to extract the most predictive cell-cell interactions among these cell types.

Specifically, DECODEM comprises of two sequential steps: *First*, for each bulk expression cohort, it generates nine cell-type-specific expression profiles (as annotated by Wu et al. [45] which were used for molecular signature generation; see details in **METHODS**: *B-cells*, *Cancer-associated Fibroblasts (CAFs)*, *Cancer Epithelial (malignant)*, *Endothelial*, *Myeloid*, *Normal Epithelial*, *Plasmablasts*, *Perivascular-like Cells (PVL)*, and *T-cells*) in each sample using CODEFACS; *Second*, by leveraging these nine deconvolved expression profiles, it generates nine cell-type-specific predictors of clinical response (responders vs. non-responders *e.g.*, pCR vs. residual disease) using a multi-stage ML pipeline (**Figure 1)**. Within this pipeline, we performed rigorous feature selection to keep the top '*m*' genes per cell type ($2 \leq m \leq 25$; the upper limit is chosen to avoid overfitting and facilitate faster runtime) before feeding their expression into an unweighted ensemble classifier consisting of four ML algorithms: regularized logistic regression, random forest, support vector machine, and XGBoost (see details in **METHODS**). The pipeline was trained with repeated three-fold cross-validation (CV) to optimize model hyperparameters. Additionally, we built a baseline predictor with bulk expression using the same ML pipeline to serve as a control, comparative rod.

To assess the performance of DECODEM and evaluate the contribution of each cell type, we analyzed bulk transcriptomics data from three breast cancer cohorts with different ratios of responders to non-responders (R:NR) to NAC (**Supplementary Figure 1**) in two scenarios: (a) cross-validation with an additional five-fold CV on the TransNEO cohort (*n* = 94, R:NR = 22:72), recently published by Sammut et al. accompanied by a multiomic clinical response predictor [11], and (b) external validation on the ARTemis + PBCP (*n* = 55, R:NR = 8:47) and BrighTNess (*n* = 106, R:NR = 65:41) cohorts [46][47]. TransNEO and ARTemis + PBCP included all major BC subtypes, among which we focused on the HER2-negative patients [11][46] (**Supplementary Figure 1B-C**), whereas the BrighTNess cohort included only TNBC patients from arm B of the trial, treated with a combination of paclitaxel and carboplatin with a placebo for veliparib [47]. Moreover, we established DECODEM's generalizability to single-cell transcriptomics through external validation on an independent cohort from Zhang et al, which consisted of 15 TNBC patients who received either paclitaxel alone (R:NR = 3:3) or in combination with atezolizumab, an anti-PD-L1 agent (R:NR = 4:5) [48]. Notably, we considered the Sammut et al. predictor built with 18 extracted features from clinical and RNA-seq data (e.g. not including the DNA-seq data) as a comparative rod, which achieved an AUC of 0.88 in CV and 0.89 in validation with ARTemis + PBCP [11]. This predictor outperformed a baseline predictor built with seven clinical features including ER status and HER2 status, which are routinely used to guide clinical decisions [49].

To investigate the effect of incorporating the expression profiles of multiple predictive cell types together, as indicated by our third question, we applied two extensions to DECODEM: (i) append the cell-type-specific expression profiles for two and three most predictive cell types together to train a multi-cell-ensemble ML pipeline, and (ii) extend to DECODEMi to incorporate the inferred CCIs to train a CCI-based ML pipeline (**Figure 1**; see details in **METHODS**). The former model maximizes the predictive performance by leveraging the complementary information from different cell types while the latter identifies the most predictive cell-to-cell communications for clinical response, thereby providing deeper insights to the underlying mechanism.

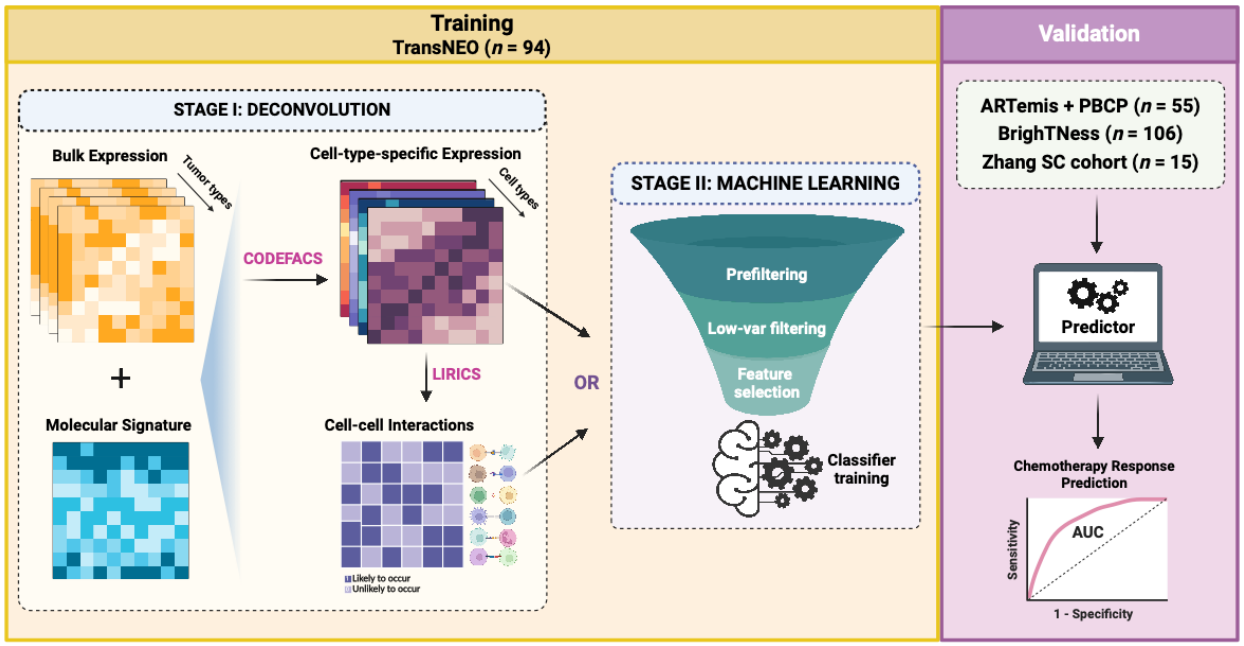


Figure 1: The analysis pipeline, DECODEM and DECODEMi. *First*, we apply CODEFACS on bulk expression to generate nine cell-type-specific expression profiles and further apply LIRICS to infer the cell-cell interactions (CCIs) present in the tumor microenvironment. *Second*, we train a multi-stage machine learning pipeline with either the nine cell-type-specific expression profiles to build nine cell-type-specific clinical response predictors (DECODEM) or the CCI profile to build a CCI-based predictor (DECODEMi). In validation, we deploy the abovementioned first stage to procure the cell-type-specific expression or CCIs and directly feed those as inputs to the trained predictors to evaluate model performance.

Gene expression of immune and stromal cells predicts response to neoadjuvant chemotherapy as accurately as the expression of malignant cells

To assess the impact of individual cell types on chemotherapy response, we applied DECODEM to TransNEO in a CV analysis. Our analysis demonstrated significant differences (one-tailed Wilcoxon rank-sum test p-value ≤ 0.001) in the resulting prediction scores between responders and non-responders across 91 patients (the patients with available predictions from Sammut et al.), with the identified cell types outperforming bulk and Sammut et al. predictors (**Figure 2A**). Seven out of nine cell types exhibited markedly enhanced predictive capabilities in identifying responders.

192 Notably, immune cells (*myeloid*, *plasmablasts*, *B-cells*) contribute as prominently as the malignant
193 (*cancer epithelial*) and stromal cells (*endothelial*, *normal epithelial*, *CAFs*) (**Figure 2B**). Our
194 subsequent analyses, focused on these seven 'prominent' predictive cell types, whose abilities to
195 accurately identify true responders to chemotherapy in terms of the odds ratio and sensitivity
196 values are shown in **Supplementary Figure 2A,C**.
197
198 These cell types have frequently been reported to influence clinical response in BC,
199 particularly in the context of chemotherapy-induced immunogenic cell death as the proposed
200 mechanism [51][11]. Changes in the immune population, such as an increase in regulatory T-cells
201 and myeloid-derived suppressor-like cells, and a decrease in $CD8^+$ T-cells, have been associated
202 with achieving pCR following NAC [6]. Myeloid cells have been implicated in therapy response
203 through interactions with the malignant and endothelial cells, inducing chemoresistance *via* the
204 *CXCL1/2 – S100A8/9* loop, whereas tumor-associated macrophages have been linked to adverse
205 outcomes and chemoresistance through survival factor secretion, anti-apoptotic pathway
206 activation, and modulation of signaling pathways [6][15][16]. *VEGF*-mediated upregulation of
207 survivin in Endothelial cells was further implicated in playing central roles in chemoresistance
208 [15]. Additionally, plasmablasts, as precursors of plasma cells, have been associated with better
209 prognosis [19][20], while B-cells have shown both pro- and anti-tumor roles [18][19][20].
210
211 We further tested and validated the predictive power of these cell types independently in
212 external validation using the ARTemis + PBCP (n = 55) and BrighTNess (n = 106) cohorts, observing
213 consistent results (**Figure 2C-D, Supplementary Figure 2A,C**). In the ARTemis + PBCP cohort
214 (dominated by ER-positive, HER2-negative BC), the prominent contributors to chemotherapy
215 response were CAFs, normal epithelial, and cancer epithelial (**Figure 2C**), whereas in the TNBC-
216 specific BrighTNess cohort, the key players were myeloid, plasmablasts, and cancer epithelial
217 (**Figure 2D**). When stratifying TransNEO and ARTemis + PBCP cohorts by the available BC subtypes,
218 our cell-type-specific models retain good performance across both subtypes (**Supplementary**
219 **Figure 3A-D**). The models for ER-positive, HER2-negative BC were more predictive than TNBC,
220 possibly due to sample size limitations as well as increased heterogeneity in TNBC. These findings

support the prominent involvement of immune and stromal cells alongside the malignant cells in mediating patient response to chemotherapy in the TME, with variations observed likely due to tumor subtype composition and ER/HER2 status differences across cohorts.

Our analysis did not identify T-cells as a prominent predictive cell type, as their expression provide marginal to no enhancement in predictive power over bulk expression (**Figure 2A-B**). This lack of signal may stem from aggregating different T-cell subtypes together (due to limits on the resolution of the deconvolution, which breaks down for low-abundance cell types). To study this further, we performed an enrichment analysis to delineate the contributions of $CD4^+$ and $CD8^+$ T-cells to chemotherapy response. However, the expression of each of these cell types still exhibited limited stratification power in distinguishing responders from non-responders across the three bulk cohorts (**Supplementary Figure 4A-D**). Additionally, we complemented the expression-based analysis with an analysis of the association between cell abundances and clinical response. This revealed that the overall T-cell abundance exhibited predictive power similar to or even higher than $CD4^+$ and $CD8^+$ subtypes (**Supplementary Figure 4E-G**), albeit still significantly lower than that of bulk transcriptomics.

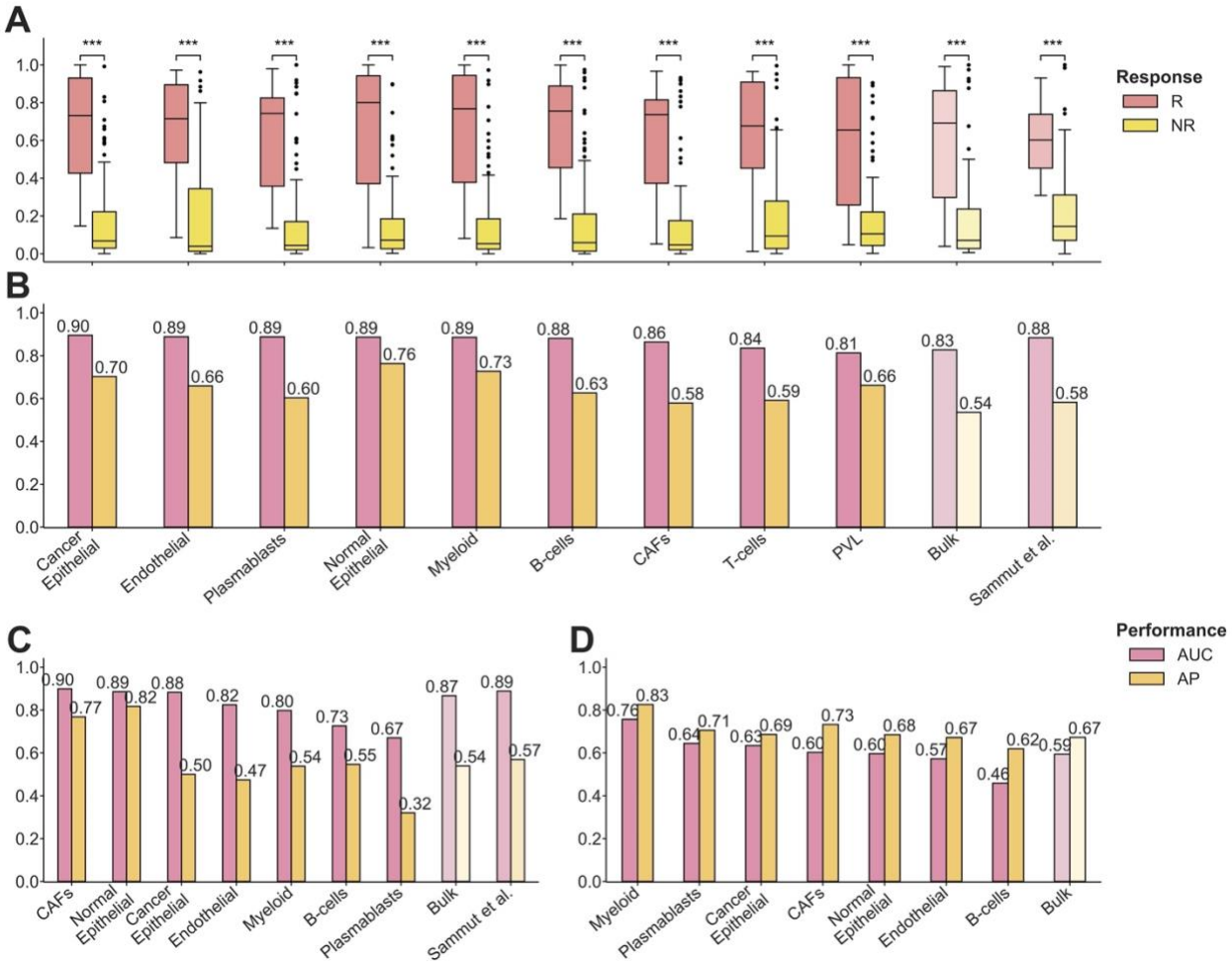


Figure 2: The prominent cell types mediating chemotherapy response in breast cancer tumor microenvironment.

A-B. Comparison of prediction scores (**A**) and model performance (**B**) in cross-validation with TransNEO (n = 91) for nine cell-type-specific, bulk, and Sammut et al. predictors. R and NR stand for responders and non-responders. The differences between the prediction scores were computed by using a one-tailed Wilcoxon rank-sum test. AUC and AP stand for the area under the receiver operating characteristics curve and average precision (equivalent to the area under precision-recall curve), respectively. Cell types are ranked by their AUC values in a descending order.

C. Comparison of model performance in external validation with ARTemis + PBCP (n = 55) for the seven prominent cell types, bulk, and Sammut et al. predictors. Cell types are ranked by their AUC values in a descending order.

251 **D.** Comparison of model performance in external validation with BrighTNess (*n* = 106) for the
252 seven prominent cell types and bulk predictors. Cell types are ranked by their AUC values in a
253 descending order.
254

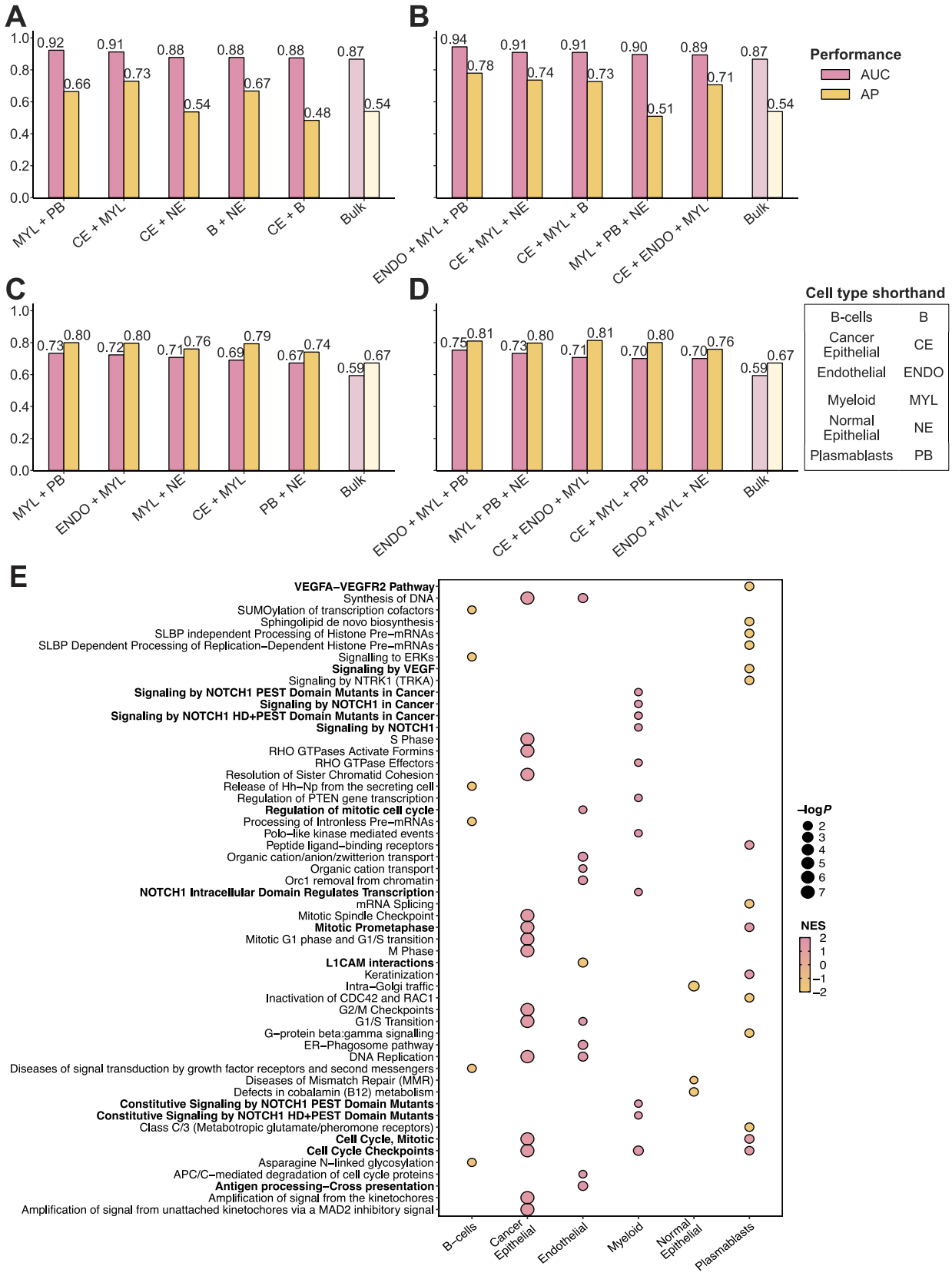
255 An ensemble model incorporating the expression of immune and stromal cells
256 significantly boosts the predictive power of patient response

257 To further enhance DECODEM's performance, we aggregated multiple individual cell types
258 together by appending the corresponding expression profiles as the input features. We explored
259 35 combinations of two to three cell types exhaustively, This comprehensive analysis yielded a
260 boost in predictive performance in both ARTemis + PBCP and BrighTNess cohorts, with significant
261 overlaps observed in the top multi-cell-type combinations across cohorts (**Figure 3A-D,**
262 **Supplementary Figure 2B,D**). Remarkably, the ensemble of immune and stromal cells, specifically
263 the combination of the expressions of endothelial, myeloid, and plasmablast cells exhibited the
264 strongest stratification capability in identifying chemotherapy responders (*ARTemis + PBCP*: AUC
265 = 0.94, AP = 0.78, **Figure 3B**; *BrighTNess*: AUC = 0.75, AP = 0.81, **Figure 3D**). Additionally, the
266 immune-cell-ensemble comprising myeloid and plasmablasts displayed high stratification
267 (*ARTemis + PBCP*: AUC = 0.92, AP = 0.66, **Figure 3A**; *BrighTNess*: AUC = 0.73, AP = 0.80, **Figure 3C**).
268 These findings testify to the complementary contributions of different cell types within the TME
269 to chemotherapy response. However, the top ensemble for the BrighTNess cohort provided a
270 similar performance to the top cell type, myeloid (AUC = 0.76, AP = 0.83, **Figure 2D**), suggesting
271 that the interactions among different cell types are less prominent in this cohort compared to the
272 other cohorts, illustrated further in **Figure 4D**. The subtype-specific stratification of ARTemis +
273 PBCP reinforced the superior predictive capabilities of these ensembles across both available BC
274 subtypes while recapitulating their diminished impact in TNBC (**Supplementary Figure 3E-F**).
275

276 We conducted gene set enrichment analysis (GSEA) with Reactome pathways [52] to
277 explore the functional contributions of these prominent cell types in chemotherapy response
278 prediction. This analysis revealed the enrichment of cell cycle checkpoint and DNA replication
279 pathways across different cell types (false discovery rate (FDR)-adjusted *p*-value ≤ 0.2; **Figure 3E**)

280 as the most common functional contribution. Dysregulation of these pathways are frequent in BC
281 and promotes tumor proliferation [53][54][55], making them valuable prognostic markers and
282 therapeutic targets [56][57][58][59]. Immune cells exhibited enrichment in signal transduction
283 pathways (*e.g.*, *VEGFA* – *VEGFR2* and *VEGF* signaling pathways in plasmablasts and multiple
284 *NOTCH1* signaling pathways in myeloid), implicated in malignant progression and therapy
285 resistance, particularly in TNBC [60][61][62][63][64][65][66]. Stromal endothelial cells showed
286 enrichment in development- and immune-related pathways (*e.g.*, *L1CAM* interactions and
287 antigen processing & presentation), contributing to aggressive BC progression and immune
288 evasion with implications for both chemotherapy and immunotherapy
289 [67][68][69][70][71][72][73].
290
291 To delve deeper into the underlying mechanism, we assessed DECODEM's performance in
292 light of the prevalent chemotherapy regimen within each cohort. TransNEO and ARTemis + PBCP
293 predominantly employed T-FEC (*n* = 61 and 40, respectively; including docetaxel/taxotere, 5-
294 flurouracil, epirubicin, and cyclophosphamide), while BrighTNess used a uniform regimen of
295 paclitaxel and carboplatin (**Supplementary Figure 1C**). All six drugs have known clinical activities
296 stimulating anticancer immunity *via* releasing immunostimulatory molecules from cancer cells
297 (*on-target* effects) or promoting activation of effector cells and diminishing immunosuppressive
298 cells (*off-target* effects) [73]. Remarkably, across all three cohorts spanning two BC subtypes and
299 two regimens, DECODEM identified myeloid and plasmablasts to consistently stratify responders
300 to *immunogenic* drug regimens, alongside malignant cells (**Supplementary Figure 5A,B,D**). For
301 ARTemis + PBCP (dominated by ER-positive, HER2-negative BC), although the bulk mixture
302 displayed comparable or superior predictive power than each of these two individual immune
303 cells, their ensemble markedly boosted the performance, supporting the possibility that there
304 may be prominent CCIs between these cell types that may be associated with the likelihood of
305 response (**Supplementary Figure 5C**). Among the three-cell-ensemble models, the combination
306 of endothelial, myeloid, and plasmablasts retains its highest performance, further underscoring
307 the critical roles of the stromal – immune interactions. Additionally, for BrighTNess, the individual
308 immune cell types showed notably strong predictive abilities, with less pronounced CCI effects

309 (**Supplementary Figure 5E**, **Figure 4D**). Collectively, these findings corroborate the immunogenic
310 effects mediated by chemotherapy, suggestive of *immunogenic cell death* [73] as suggested by
311 Sammut et al. [11].
312


A
B
C
D
E
Performance
AUC
AP
Cell type shorthand
B-cells B
Cancer Epithelial CE
Endothelial ENDO
Myeloid MYL
Normal Epithelial NE
Plasmablasts PB
MYL + PB
CE + MYL
CE + NE
B + NE
CE + B
Bulk
ENDO + MYL + PB
CE + MYL + NE
CE + MYL + B
MYL + PB + NE
CE + ENDO + MYL
ENDO + MYL
PB + NE
VEGFA–VEGFR2 Pathway
Synthesis of DNA
SUMOylation of transcription cofactors
Sphingolipid de novo biosynthesis
SLBP independent Processing of Histone Pre–mRNAs
SLBP Dependent Processing of Replication–Dependent Histone Pre–mRNAs
Signalling to ERKs
Signaling by VEGF
Signaling by NTRK1 (TRKA)
Signaling by NOTCH1 PEST Domain Mutants in Cancer
Signaling by NOTCH1 in Cancer
Signaling by NOTCH1 HD+PEST Domain Mutants in Cancer
Signaling by NOTCH1
S Phase
RHO GTPases Activate Formins
RHO GTPase Effectors
Resolution of Sister Chromatid Cohesion
Release of Hh–Np from the secreting cell
Regulation of PTEN gene transcription
Regulation of mitotic cell cycle
Processing of Intronless Pre–mRNAs
Polo–like kinase mediated events
Peptide ligand–binding receptors
Organic cation/anion/zwitterion transport
Organic cation transport
Orc1 removal from chromatin
NOTCH1 Intracellular Domain Regulates Transcription
mRNA Splicing
Mitotic Spindle Checkpoint
Mitotic Prometaphase
Mitotic G1 phase and G1/S transition
M Phase
L1CAM interactions
Keratinization
Intra–Golgi traffic
Inactivation of CDC42 and RAC1
G2/M Checkpoints
G1/S Transition
G–protein beta:gamma signalling
ER–Phagosome pathway
DNA Replication
Diseases of signal transduction by growth factor receptors and second messengers
Diseases of Mismatch Repair (MMR)
Defects in cobalamin (B12) metabolism
Constitutive Signaling by NOTCH1 PEST Domain Mutants
Constitutive Signaling by NOTCH1 HD+PEST Domain Mutants
Class C/3 (Metabotropic glutamate/pheromone receptors)
Cell Cycle, Mitotic
Cell Cycle Checkpoints
Asparagine N–linked glycosylation
APC/C–mediated degradation of cell cycle proteins
Antigen processing–Cross presentation
Amplification of signal from the kinetochores
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal
B–cells
Cancer Epithelial
Endothelial
Myeloid
Normal Epithelial
Plasmablasts
–logP
NES

313

314 **Figure 3: The prominent multi-cell-type ensembles mediating chemotherapy response in breast**
315 **cancer tumor microenvironment.**
316 **A-B.** Validation performance with ARTemis + PBCP (*n* = 55) for the five most prominent two-cell-
317 ensembles (**A**) and three-cell-ensembles (**B**) along with the bulk. AUC and AP stand for the area
318 under the receiver operating characteristics curve and average precision (equivalent to the area
319 under the precision-recall curve), respectively. Ensembles are ranked by their AUC values in a
320 descending order.
321 **C-D.** Validation performance with BrighTNess (*n* = 106) for the five most prominent two-cell-
322 ensembles (**C**) and three-cell-ensembles (**D**) along with the bulk. Ensembles are ranked by their
323 AUC values in a descending order.
324 **E.** The enriched Reactome pathways (FDR-adjusted *p*-value ≤ 0.2) across the prominent cell types.
325 For each cell type, only the top 10 relevant pathways are displayed. NES and log*P* stand for the
326 normalized enrichment score from GSEA and log-scaled *p*-value, respectively.
327

328 Identifying key cell-cell interactions associated with chemotherapy response

329 To investigate the impact of the presence of interactions among multiple cell types in modulating
330 chemotherapy response, we use DECODEMi to analyze cell-cell interactions (inferred using LIRICS
331 [44], starting with 2,422 ligand – receptor (L-R) pairs from Ramilowski et al. [50]; see details in
332 **METHODS**) that are predictive of patient response. As before, we conducted two analyses: (a)
333 cross-validation on TransNEO (*n* = 94), and (b) external validation on ARTemis + PBCP (*n* = 55) and
334 BrighTNess (*n* = 106). We found that CCIs encompassing only the four most prominent cell types
335 in the TME (*cancer epithelial*, *endothelial*, *myeloid*, and *plasmablasts*) exhibited a comparable
336 classification power to all available CCIs across all three cohorts (one-tailed Wilcoxon rank-sum
337 test *p*-value ≤ 0.05), further testifying to the influential roles of these specific cell types in patient
338 response (**Figure 4A-D**, **Supplementary Figure 6A-D**). The CCIs involving these four cell types
339 markedly improved the predictive power of CCI-based predictors in both TransNEO and ARTemis
340 + PBCP (**Figure 4A-C**). However, the performance of CCIs was generally lower than the top
341 predictive myeloid cells in BrighTNess cohort (**Figure 4D**). This discrepancy can again be attributed
342 to cohort composition and treatment-specific variations. It is worth noting that this cohort

343 included only TNBC patients treated with paclitaxel and carboplatin, for which very few samples
344 were available in training (**Supplementary Figure 1B-C**). These results reinforced that leveraging
345 information from multiple cell types can notably improve clinical response prediction and the
346 inter-cellular crosstalk may involve mediating treatment response in the TME. To further assess
347 the generalizability of DECODEMi beyond bulk cohorts, we assessed its effectiveness on an
348 independent single-cell (SC) TNBC cohort without additional training. To overcome the limitation
349 of small sample size (n = 6), we constructed a SC-TNBC cohort of 200 'pseudopatients' treated
350 with chemotherapy (R:NR = 100:100) by downsampling 28,209 cells across the six TNBC patients
351 who received paclitaxel in Zhang et al.'s study [48]. We focused on the CCIs among the three cell
352 types (encompassing B-cells, myeloid, and T-cells) present in both SC and deconvolved bulk data,
353 and identified the 170 top predictive CCIs (covering 134 L-R pairs) in bulk TransNEO cohort. The
354 prediction scores for the SC-TNBC cohort were computed by counting the number of activated
355 CCIs within each pseudopatient (as inferred by Kumar et al. [103]; see details in **METHODS**). Our
356 analysis demonstrated the significant and robust predictive capabilities of these top CCIs (one-
357 tailed Wilcoxon rank-sum test *p*-value ≤ 0.001, **Figure 4G-H**), affirming DECODEMi's ability to
358 identify the CCIs predictive of patient response.
359
360 By computing feature importance from DECODEMi, we identified the most predictive cell-
361 to-cell communications across the ARTemis + PBCP and BrighTNess cohorts, finding a significant
362 overlap in top CCIs (**Figure 4E-F**). Remarkably, the interaction between '*GDF9*' (growth
363 differentiation factor 9) and '*BMPR1B*' (bone morphogenetic protein receptor 1B), both members
364 of the TGF-β superfamily (the corresponding protein-protein interaction (PPI) network, obtained
365 from STRING [100], is provided in **Supplementary Figure 6E**) with roles in reproductive biology
366 and signal transduction [74][75][76][77], emerged as a key CCI with the same directionality across
367 multiple cell types. These genes have been implicated in BC prognosis and therapeutic outcomes
368 [78][79][80], with *BMPR1B* exhibiting dual roles in chemotherapy response [81][82][83]. Another
369 top CCI involved the renowned *GDNF – RET* signaling pathway interactions between '*NRTN*'
370 (neurturin), or its paralog '*ARTN*' (artemin), and '*GFRA1*' (GDNF family receptor alpha 1) in
371 multiple cell types (**Figure 4E-F**; see PPI network in **Supplementary Figure 6F**), which are all

372 members of the glial cell line-derived neurotrophic factor (GDNF) family within the TGF-β
373 superfamily with roles in neuronal survival, growth, and differentiation [84]. The GDNF family
374 exhibits diverse roles in cancer [85][86], including an association with therapy resistance in ER-
375 positive BC, highlighting *GFRA1* as a potential target [87][88][89][90]. Additionally, the well-
376 studied chemokine interaction between '*CX3CL1*' (C-X3-C chemokine ligand 1) and '*CX3CR1*' (C-
377 X3-C chemokine receptor 1) [91] is prominent in BrighTNess (**Figure 4F**; see PPI network in
378 **Supplementary Figure 6G**), which is known to be involved in progression and metastasis across
379 multiple cancers including BC, presenting a promising pan-cancer target [92][93][94][95][96][97].
380
381 Overall, DECODEMi application results in: (a) a simpler predictor leveraging CCIs that
382 improves patient stratification compared to bulk transcriptomics, and (b) the identification of cell-
383 to-cell communications predictive of treatment response, whose activation is associated with
384 pro-cancerous roles, thus offering potential novel therapeutic targets.
385

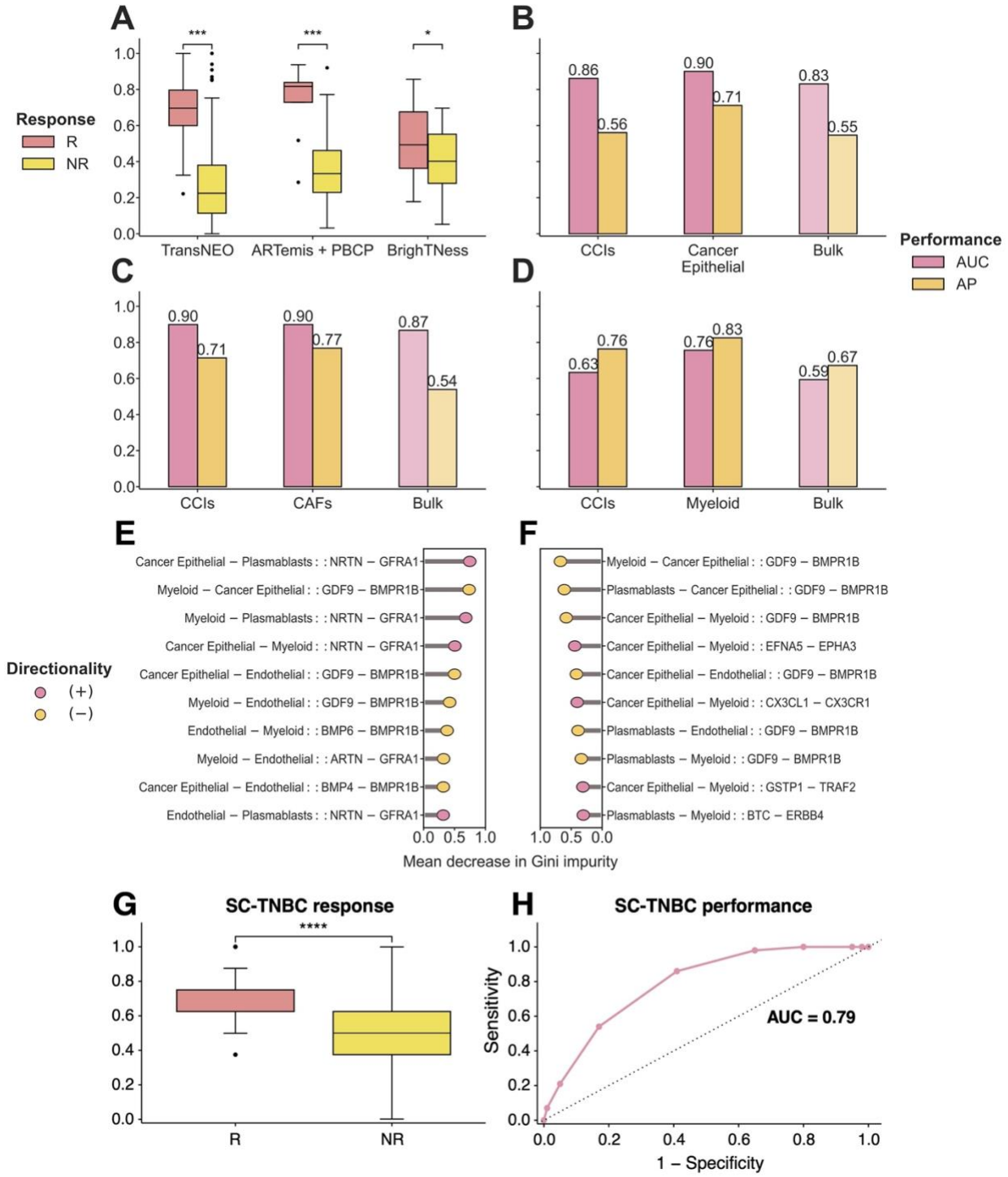


Figure 4: The prominent cell-cell interactions (CCIs) mediating chemotherapy response in breast cancer tumor microenvironment.

A. Prediction scores for DECODEMi in cross-validation with TransNEO (*n* = 94) and in external validation with ARTemis + PBCP (*n* = 55) and BrighTNess (*n* = 106). R and NR stand for responders

391 and non-responders, respectively. The differences between the prediction scores were computed
392 by using a one-tailed Wilcoxon rank-sum test.
393 **B.** Model performance in cross-validation with TransNEO for CCI-based, the top cell-type-specific,
394 and bulk predictors. AUC and AP stand for the area under the receiver operating characteristics
395 curve and average precision (equivalent to the area under precision-recall curve), respectively.
396 **C-D.** Model performance in external validation with ARTemis + PBCP (**C**) and BrighTNess (**D**) for
397 CCI-based, the top cell-type-specific, and bulk predictors.
398 **E-F.** The 10 most predictive CCIs in external validation with ARTemis + PBCP (**E**) and BrighTNess
399 (**F**). Each CCI is displayed as a quadruplet where the first pair containing the ligand and receptor
400 cell types (separated by '–') is separated by '::' from the second pair containing the corresponding
401 ligand and receptor genes (separated by '–'). Mean decrease in Gini impurity is the built-in feature
402 importance measure in random forest where a higher value indicates a higher importance and
403 *vice versa* [98]. Feature directionalities were computed using Fisher's enrichment analysis.
404 **G-H.** Prediction scores (**G**) and AUC value (**H**) for chemotherapy response prediction in external
405 validation with the SC-TNBC cohort (*n* = 200, sourced from Zhang et al. [48]) using the top 170
406 response-relevant CCIs (among B-cells, myeloid, and T-cells) identified in bulk by DECODEMi.
407

408 DECODEM identifies responders to chemotherapy and anti-PD-L1 therapy from 409 single-cell transcriptomics

410 To further assess the generalizability of DECODEM, we evaluated the predictive capabilities of the
411 cell-type-specific models with single-cell transcriptomics. We analyzed a SC cohort containing the
412 expression of B-cells, myeloid, and T-cells for 15 patients with triple negative breast cancer from
413 Zhang et al., who were either treated with chemotherapy (paclitaxel) alone or combined with the
414 anti-PD-L1 agent, atezolizumab [48]. We first stratified the patient cohort into those who received
415 chemotherapy alone (*n* = 6) and those who underwent chemotherapy + immunotherapy (*n* = 9).
416 For each subset, we applied our established DECODEM models for these three cell types which
417 were trained with deconvolved TransNEO data alongside a baseline 'pseudobulk' model
418 (analogous to bulk in previous scenarios; see details in **METHODS**). Our findings revealed that the
419 expression of B-cells exhibited the highest predictive potential within this cohort (**Figure *5*A-B**,

420 **Supplementary Figure 7A-D**). Notably, this B-cells-specific predictor from DECODEM effectively
421 distinguishes the responders from non-responders for both intervention types (one-tailed
422 Wilcoxon rank-sum test *p*-value ≤ 0.05, **Figure 5A-B**). These findings further attest to the
423 generalizability and robustness of DECODEM, highlighting its potential in capturing treatment-
424 invariant properties of the BC TME to predict response to the combination of chemotherapy and
425 immunotherapy through SC transcriptomics.
426

427 ## DECODEM effectively stratifies TCGA patients' survival

428 We finally explored whether DECODEM can stratify the survival of 705 breast cancer patients in
429 The Cancer Genome Atlas (TCGA-BRCA) from their deconvolved expression, obtained from Wang
430 et al. [44]. Focusing on the expression of the prominent cell type of cancer epithelial (and
431 endothelial) cells, DECODEM effectively stratified these patients into low-risk, likely responders
432 (scores > median score, n = 353) and high-risk, unlikely responders (scores ≤ median score, n =
433 352), as displayed by their overall survival and progression-free interval (Log-rank test *p*-value ≤
434 0.05, **Figure 5C-D**, **Supplementary Figure 7E-F**). Remarkably, this stratification was achieved
435 regardless of treatment regimen, attesting that DECODEM captured the treatment-invariant
436 properties of the BC TME that are predictive of patient survival.

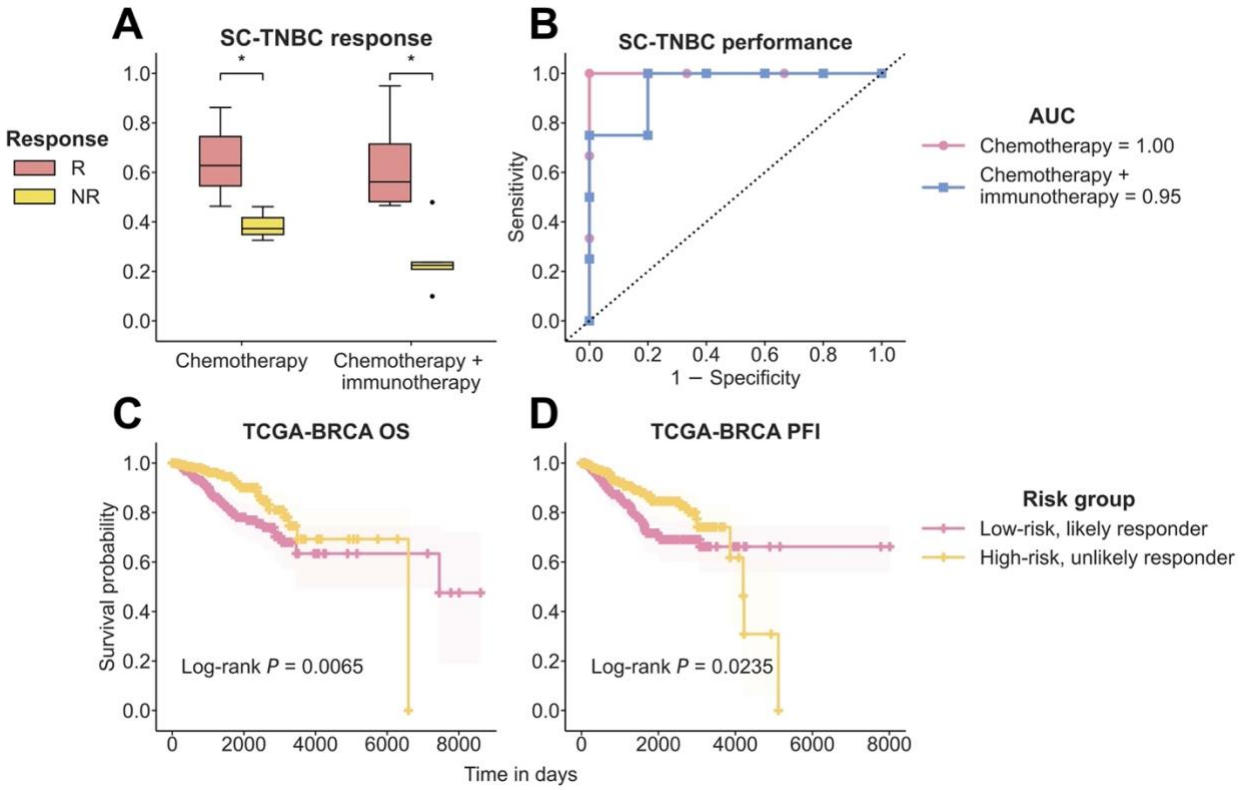


Figure 5: Generalizability of DECODEM to single-cell (SC) transcriptomics and for TCGA patient survival stratification.

A-B. Prediction scores (**A**) and the area under the receiver operating characteristics curve (AUC) values (**B**) in external validation with SC expression of TNBC patients (SC-TNBC) treated with chemotherapy alone (n = 6) or combined with immunotherapy (n = 9) from Zhang et al. [48]. R and NR stand for responders and non-responders, respectively. The differences between the prediction scores were computed by using a one-tailed Wilcoxon rank-sum test.

C-D. Kaplan-Meier plots depicting the overall survival (OS; **C**) and progression-free interval (PFI; **D**) of 705 BC patients from TCGA (TCGA-BRCA). Patients were stratified into low-risk, likely responder (n = 353) and high-risk, unlikely responder (n = 352) groups by their median DECODEM score computed using the expression of cancer epithelial cells, whereby the low-risk and high-risk groups comprised individuals with scores above and below the median, respectively. The differences between the two curves were computed by using a Log-rank test.

453 # DISCUSSION

454 Our study introduces DECODEM, a transcriptomics-based modeling framework employing
455 CODEFACS, a cutting-edge deconvolution tool, with a multi-stage machine learning pipeline to
456 systematically assess the cell-type-specific contributions to predicting clinical response to
457 treatment. As a representative scenario, we apply it to build predictors of patient response to
458 neoadjuvant chemotherapy in breast cancer. We further present DECODEMi, a companion tool
459 that identifies the key cell-cell interactions associated with clinical response. A recent study by
460 Kester et al. investigated the association of cellular compositions of the BC TME with patient
461 survival [43], but to our knowledge, the current study is the first-of-its-kind to analyze the cell-
462 type-specific transcriptomics profiles in the BC TME to quantify their respective contributions to
463 treatment response prediction.
464
465 We applied DECODEM to investigate the predictive power of the expression of various cell
466 types in the BC TME to patient response to NAC. Our findings highlight the prominence of immune
467 cells (*myeloid*, *plasmablasts*, *B-cells*) in chemotherapy response, consistent with Kester et al. [43],
468 elucidating the complementary roles of diverse cell types that allows for the incorporation of top
469 predictive cell types to build accurate clinical response predictors, and showcase DECODEMi's
470 ability to identify key CCIs impacting response across HER2-negative BC subtypes. Furthermore,
471 we illustrated DECODEM's generalizability to single-cell transcriptomics where it captures
472 treatment-invariant properties of the TME, enabling accurate classification of responders for
473 chemotherapy and immunotherapy, and its application for patient survival stratification in TCGA.
474
475 Immune cells have a clear association with prognosis in BC, both without chemotherapy
476 and after chemotherapy in the early-stage setting [104][30][31]. The International TILs Working
477 Group was established for the express purpose of standardizing incorporation of tumor infiltrating
478 lymphocyte pathological review for clinical use, and established an online prognosis calculator for
479 patients with TNBC (https://www.tilsinbreastcancer.org/prognosis-tool/). Plasma cells and naïve
480 B-cells have been associated with better pCR rates in ER-positive tumors, whereas regulatory T-
481 cells have been associated with better pCR rates in TNBC [105]. In our analysis, immune cells such

482 as plasmablasts and myeloid cells exhibit involvement in signal transduction, particularly through
483 *VEGF* signaling and *NOTCH* signaling pathways, which are associated with tumor-associated
484 angiogenesis, enrichment of BC stem cells and epithelial-to-mesenchymal transition, leading to
485 poor prognosis and therapy resistance [60][61][64][66]. Stromal cells such as endothelial cells
486 show enrichments in development- and immune-related pathways including *L1CAM* interactions
487 and antigen processing & presentation, with *L1CAM*'s oncogenic role and defects in antigen
488 processing leading to poor outcomes and therapy resistance [67][68][69][70][71]. Overall, our
489 results paired with the existing clinical evidence reinforce the notion that the presence of an
490 active immune infiltration within the TME prior to treatment, as indicated by the prominent
491 contributions of immune cells, leads to favorable response to chemotherapy, as suggested by
492 Sammut et al. [11].
493
494 It is also important to note that DECODEM and DECODEMi offer to extend upon the
495 existing application of gene expression in the early-stage breast cancer setting. Currently, gene
496 expression-based tests such as Oncotype DX and MammaPrint are used for prediction and
497 prognostication in ER-positive BC patients, reporting if a subset should receive chemotherapy due
498 to worse predicted outcomes [106]. These tests use quantitative RT-PCR or cDNA microarray to
499 assess a known subset of genes, and largely apply only to a subset of early-stage ER-positive BC
500 patients. DECODEM and DECODEMi provide complimentary context to these tools, expanding the
501 applicable subtypes for chemotherapy response prediction as well as greater biological context.
502
503 Although we presented a comprehensive framework to assess the TME in a quantitative
504 manner, certain limitations should be acknowledged and addressed in future studies. *First*, our
505 predictors rely on data-driven algorithms involving deconvolution and ML, necessitating a large
506 sample size to attain statistical robustness and avoid overfitting. Accordingly, the available
507 datasets are limited in size, limiting our ability to perform robust subtype-specific analyses.
508 *Second*, any limitations inherent to CODEFACS, such as discrepancies in RNA-seq processing and
509 reliance on appropriate molecular signatures [44], will affect DECODEM in the downstream
510 analysis. The same goes for DECODEMi and its dependence on LIRICS using an appropriate ligand

511 – receptor database [43]. *Third*, our deconvolution analysis could not be performed separately
512 for the $CD4^+$ and $CD8^+$ T-cell subsets. Alternatively, we provided an enrichment analysis assessing
513 the predictive powers of $CD4^+$ and $CD8^+$ T-cells which corroborated our findings from using the
514 overall T-cells.
515
516 Our work focused on examining the association of the BC TME, specifically in HER2-
517 negative subtypes, to clinical response to NAC. However, both DECODEM and DECODEMi can be
518 applied to investigate diverse treatment regimens in different cancer indications. Although both
519 TransNEO and ARTemis + PBCP cohorts include different chemotherapy combinations, limited
520 sample size prevented us from exploring further. Therefore, an intriguing direction could be to
521 investigate the distinct role of the TME in different chemotherapy families with different
522 mechanisms of actions (MOAs) *e.g.*, taxanes, antimetabolites, platinum agents and so on. With
523 DECODEMi, one could further pinpoint the family-specific significant cell-cell interactions to
524 understand the underlying MOA and uncover novel CCIs with promising therapeutic implications.
525 Additionally, expanding the investigation beyond the scope of cancer and examining the impact
526 of the microenvironment in various indications such as aging-related disorders and autoimmune
527 diseases could provide further valuable insights.
528
529 In summary, our study introduces two data-driven computational frameworks that enable
530 systematic investigation of the roles of the microenvironment in various processes, including
531 treatment response. These frameworks can utilize bulk transcriptomics from healthy and
532 diseased conditions when single-cell information is not easily accessible, providing a valuable
533 approach to study the TME in a comprehensive manner. Their application in breast cancer
534 uncovers the considerable predictive power of diverse cell types within the TME and highlights
535 key stromal – immune cell-cell interactions that are strongly predictive of patient response.
536

537 # METHODS

538 ## Overview of the datasets

539 We analyzed three bulk transcriptomics datasets, encompassing a total of 255 breast cancer
540 patients who received neoadjuvant chemotherapy (**Supplementary Figure 1**). The first dataset,
541 recently published by Sammut et al., was the TransNEO cohort which included 94 HER2-negative
542 breast tumors patients undergoing various NAC regimens [11]. Notably, Sammut et al. presented
543 the complete TransNEO cohort, covering 168 patients with early and locally advanced BC from all
544 major subtypes, and providing DNA-seq, RNA-seq, digital pathology, and treatment information
545 for various subsets of patients [11]. The second dataset was the ARTemis + PBCP cohort, validated
546 externally by Sammut et al., combining the control arm of the ARTemis trial (*n* = 38) received the
547 combination therapy of FEC-T [46] and the HER2-negative cases from the Personalised Breast
548 Cancer Programme (PBCP; *n* = 17) undergoing various NAC regimens [11] (**Supplementary Figure**
549 **1B-C**). The third dataset, the BrighTNess cohort, included 106 patients from arm B of the
550 BrighTNess trial, a randomized, double-blind, placebo-controlled three-arm clinical study for
551 veliparib in stage II/III TNBC [47]. Arm B included only patients who received the NAC combination
552 of paclitaxel, carboplatin, and a placebo for veliparib (**Supplementary Figure 1B-C**). In all three
553 cohorts, patient response was assessed at surgery using the residual cancer burden (RCB) index
554 where patients achieving pathological complete response (pCR; RCB = 0) were considered as
555 responders (R) and patients with residual diseases (RCB-I, II, III) as non-responders (NR) with
556 different R:NR ratios (**Supplementary Figure 1A**).

557

558 Furthermore, we analyzed a recent single-cell cohort presented by Zhang et al. comprising
559 22 patients with advanced triple negative breast cancer. Half of the patients received paclitaxel
560 alone, while the other half received paclitaxel combined with the anti-PD-L1 agent, atezolizumab
561 [48]. This study included 78 tumor biopsies and blood samples collected at three time-points,
562 including baseline. Zhang et al. provided the single-cell expression for four immune cell types (B-
563 cells, T-cells, myeloid, and innate lymphoid cells) and patient response assessed using the
564 Response Evaluation Criteria in Solid Tumors (RECIST) criteria [48]. We categorized the patients
565 achieving complete or partial response as responders (R) and those with stable or progressive

566 diseases as non-responders (NR). Our analysis focused solely on pre-treatment tissue samples,
567 resulting in two treatment groups: (a) chemotherapy (*n* = 6, R:NR = 3:3) and (b) chemotherapy +
568 immunotherapy (*n* = 9, R:NR = 4:5). To align with the output of CODEFACS, which provides the
569 mean cell-type-specific expression profile per patient [44], we harmonized the cell-type-specific
570 SC expression by averaging across all corresponding cells for each patient, while filtering for low-
571 variance genes (threshold = 0.8). Additionally, we calculated the 'pseudobulk' expression as the
572 mean expression across all cells encompassing the three cell types for each patient (post low-
573 variance filtering) to generate a baseline expression profile similar to the bulk mixture.
574

575 ## Deconvolution with CODEFACS

576 The first step of DECODEM involves employing CODEFACS, a cellular deconvolution tool recently
577 developed in our lab [44]. CODEFACS characterizes the TME by reconstructing the cell-type-
578 specific expression in each sample from the input bulk expression, utilizing the corresponding cell
579 type abundance or a cell-type-specific signature profile as another input. The output includes the
580 cell-type-specific expression profiles, the corresponding cell type abundances, and a confidence
581 score matrix that quantifies the confidence level (as a value between 0 and 1) of the estimated
582 expression values for each gene in each cell type across all samples.
583

584 We derived a cell type signature matrix of 1,400 genes representing nine cell types from
585 a single-cell RNA-seq (scRNA-seq) cohort from Wu et al. [45] using an in-house signature
586 derivation tool based on the CIBERSORTx algorithm [101]. We then deconvolved the three bulk
587 expression profiles with this signature into cell-type-specific expression profiles, encompassing
588 the nine cell types: *B-cells*, *Cancer-associated Fibroblasts (CAFs)*, *Cancer Epithelial (malignant)*,
589 *Endothelial*, *Myeloid*, *Normal Epithelial*, *Plasmablasts*, *Perivascular-like Cells (PVL)*, and *T-cells*.
590

591 ## Extraction of cell-cell interactions with LIRICS

592 We extended DECODEM to DECODEMi by incorporating multi-cell-type information through cell-
593 cell interactions. To this end, DECODEMi employs LIRICS [44] on the output of CODEFACS to infer
594 the activity of the corresponding CCIs. LIRICS applies a two-step approach to infer the cell-type-

595 specific ligand – receptor (L-R) pairs that are likely to be 'active' in each sample from deconvolved
596 expression. The first step is to query a database of 49,397 plausible L-R interactions between each
597 pair of cell types in BC TME (see below for details). The second step infers an L-R interaction to
598 be active (indicated by '1') in a pair of cell types in a given sample if the expressions of both ligand
599 and receptor genes exceed their median cell-type-specific expression values or inactive (indicated
600 by '0') otherwise. We used this binary CCI activity matrix output from LIRICS to assess the
601 predictive potential of cell-to-cell communications.
602

603 Curation of cell-type-specific ligand – receptor database for BC TME from single-cell 604 transcriptomics

605 To infer the cell-cell interaction activity using LIRICS, we applied a two-step approach to construct
606 a database of cell-type-specific ligand – receptor interactions (LRIs) to characterize all cell-to-cell
607 communications that are likely to occur in the TME of BC patients. We first obtained a list of 2,422
608 potential L-R pairs, as annotated by Ramilowski et al. [50]. We next evaluated the potential of the
609 L-R pairs occurring between the nine cell-types-of-interest using scRNA-seq data from Wu et al
610 [45]. To discern the activity of the curated L-R list, we pooled together all 130,246 cells for the
611 nine cell types from Wu et al. and adopted an empirical *p*-value based CCI inference approach
612 following Kumar et al. [103], where an LRI with p-value ≤ 0.05 is categorized as 'likely' to occur in
613 the TME of BC patients and 'unlikely' otherwise. We constructed a database of 49,397 cell-type-
614 specific L-R pairs (*i.e.*, CCI quadruplets) that are likely to occur in BC TME. This database composed
615 the initial feature space for LIRICS to yield different numbers of CCIs across three bulk cohorts:
616 TransNEO = 46,755, ARTemis + PBCP = 43,887, and BrighTNess = 38,285.
617

618 Machine learning pipeline with deconvolved expression and CCIs

619 In both DECODEM and DECODEMi, the second step involves building ML predictors of clinical
620 response using cell-type-specific expression profiles and CCIs, respectively. For DECODEM, we
621 built nine cell-type-specific predictors to classify responders and non-responders following
622 treatment using a four-stage ML pipeline (**Figure 1**): (a) prefiltering to remove genes with
623 CODEFACS confidence score < 0.99, (b) low-variance filtering to remove genes with variance < 0.1,

624 (c) feature selection to select the top '*m*' genes maximizing performance, where $2 \leq m \leq 25$ with
625 genes ranked by their ANOVA F-scores for association with response followed by standardization
626 to scale each gene to zero mean and unit variance, and (d) classification with an unweighted
627 ensemble classifier comprised of regularized logistic regression, random forest, support vector
628 machine, and XGBoost (extreme gradient boosting) to predict the probabilities of achieving pCR.
629 Here, the upper limit on the feature size is chosen to avoid model overfitting and longer runtimes
630 (and evaluated to provide similar performance across different feature sizes in cross-validation).
631 Furthermore, each prediction vector was rescaled between 0 and 1 to improve generalization for
632 external validation. We trained this ML pipeline with three-fold cross-validation (CV) for
633 hyperparameter tuning to maximize the AUC for each classifier separately. To avoid overfitting,
634 we repeated the training procedure five times with five different CV splits and took the mean of
635 the five prediction vectors as our final prediction. We analyzed the three bulk expression cohorts
636 with this pipeline in two scenarios: (a) cross-validation with TransNEO (n = 94) where we applied
637 an additional five-fold CV to estimate the performance, and (b) external validation on ARTemis +
638 PBCP (n = 55) and BrighTNess (n = 106) using predictors trained on TransNEO. We further built a
639 baseline predictor using bulk expression in ML pipeline directly and identified the 'prominent' cell
640 types in CV that outperformed the bulk, subsequently validated in external validation.
641
642 To accommodate the sparse nature of SC transcriptomics for external validation with
643 Zhang et al. [48], we slightly modified the pipeline to enable modeling for all available genes post
644 filtering in step (c) *i.e.*, $2 \leq m \leq$ 'all'. Additionally, for CCI-based analysis with DECODEMi, we further
645 updated the low-variance cut-off in step (b) to 0.08 and the classifier in step (d) to a random forest
646 that is better suited for the binary nature of the CCI matrix and includes a built-in measure of
647 feature importance, mean decrease in Gini impurity (similar performance to the permutation-
648 based approaches) [98]. The corresponding CCI directionalities were computed using a two-tailed
649 Fisher's test (followed by false discovery rate (FDR) adjustment for multiple hypothesis correction
650 *i.e.*, FDR-adjusted p-value ≤ 0.05), where an odds ratio (OR) > 1 denotes an enrichment in
651 responders and OR < 1 indicates enrichment in non-responders. Moreover, to filter CCIs for a

652 subset of cell types, we only include the interactions involving the confident L-R genes (CODEFACS
653 confidence level ≥ 0.99) in the corresponding cell type.
654

655 $CD4^+$ / $CD8^+$ T-cells enrichment analysis

656 To analyze the contribution of the individual T-cell subtypes to chemotherapy response, we
657 performed a gene set variation analysis (GSVA) to summarize pathway activity per sample for
658 each T-cell subtype. We employed two single-cell expression signatures derived from Wu et al.
659 [45], collected *via* the Tumor Immune Single Cell Hub 2 (TISCH2) [102]. These signatures consisted
660 of 207 and 424 differentially expressed genes (log-fold-change ≥ 0.5) for $CD4^+$ and $CD8^+$ T-cells,
661 respectively (with an overlap of 151 genes). Subsequently, we evaluated the association of the
662 GSVA enrichment scores (rescaled between 0 and 1) with clinical response across the three bulk
663 cohorts through a one-tailed Wilcoxon rank-sum test (FDR-adjusted *p*-value ≤ 0.05), and also
664 assessed the AUC and AP values independently.
665

666 Cell-type-specific enrichment analysis

667 We conducted gene set enrichment analysis (GSEA) to investigate the association of each cell type
668 with different biological processes. For each cell type, we first ranked the 17,680 protein-coding
669 genes available in TransNEO by their association with positive outcome (*i.e.*, pCR), measured by
670 the *p*-values from a one-tailed Wilcoxon rank-sum test, adjusted by CODEFACS confidence scores.
671 We next performed a separate GSEA analysis using the Reactome pathways [52] for each
672 prominent cell type (except CAFs). This resulted in six different sets of enriched pathways (FDR-
673 adjusted *p*-value ≤ 0.2) across six prominent cell types. From these sets, we reported the top 10
674 relevant pathways for each cell type.
675

676 Single-cell validation of top cell-cell interactions

677 We validated the most predictive cell-cell interactions from DECODEMi using a single-cell cohort
678 from Zhang et al. that includes six TNBC patients treated with chemotherapy [48]. We devised a
679 three-step validation procedure: *First,* for consistency, we retrained DECODEMi using only the
680 interactions among cell types available in both bulk and SC (namely *B-cells*, *myeloid*, and *T-cells*)

681 in TransNEO. This resulted in the extraction of the top predictive CCIs (*m* = 170) involving a total
682 of 134 unique L-R pairs. *Second*, to enhance the statistical power, we pooled the 28,209 cells
683 encompassing these three cell types from the six chemotherapy patients in Zhang et al., and
684 randomly downsampled 30% of the cells 100 times within each treatment group for each cell
685 type. This procedure generated a profile of 200 'pseudopatients' (R:NR = 100:100) with empirical
686 *p*-values encompassing the significance levels of the 804 possible CCIs involving the 134 L-R pairs
687 in three cell types (= 134 × 6) *i.e.*, the 'SC-TNBC cohort'. To identify the activation status of the
688 170 CCIs-of-interest in this cohort, we again adopted the CCI inference approach from Kumar et
689 al. [103], where a CCI is denoted as 'active' if *p*-value ≤ 0.05 and 'inactive' otherwise. *Third*, for
690 each pseudopatient, we predicted the chemotherapy response as the sum of the active CCIs,
691 weighted by their directionalities (*i.e.*, +1 or -1 when OR > 1 or < 1, FDR-adjusted *p*-value ≤ 0.05;
692 and 0 otherwise), followed by a rescaling between 0 and 1 across all pesudopatients. We used
693 these prediction scores to evaluate the predictive performance of these top CCIs learned from
694 bulk on the SC-TNBC cohort through using a one-tailed Wilcoxon rank-sum test (FDR-adjusted *p*-
695 value ≤ 0.05) and receiver operating characteristics (ROC) analysis.
696

697 ## DATA AVAILABILITY

698 For TransNEO cohort, RNA-seq counts and clinical data were acquired from Sammut et al.
699 (https://doi.org/10.1038/s41586-021-04278-5). RNA-seq counts were converted to TPM by using
700 Ensembl gene annotations (GRCh37.87, http://ftp.ensembl.org/pub/grch37/release-
701 87/gtf/homo_sapiens/). For ARTemis + PBCP cohort, clinical data were collected from Sammut et
702 al. (https://doi.org/10.1038/s41586-021-04278-5) and RNA-seq TPM values were obtained
703 directly from the authors. For BrighTNess cohort, RNA-seq FPKM values and clinical data were
704 acquired from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164458). The
705 FPKM values were converted into TPM. For Zhang et al. cohort, scRNA-seq counts and clinical
706 data were obtained from Zhang et al. (https://doi.org/10.1016/j.ccell.2021.09.010). For the
707 single-cell signature for deconvolution, scRNA-seq counts and cell annotations were collected
708 from Wu et al. (https://doi.org/10.1038/s41588-021-00911-1). The ligand-receptor list for LIRICS
709 were obtained from Ramilowski et al. via GitHub (https://github.com/LewisLabUCSD/Ligand-

710 Receptor-Pairs). The deconvolved expression and cell fraction data using CODEFACS on TCGA-
711 BRCA patients were collected from Wang et al. (https://doi.org/10.1158%2F2159-8290.CD-21-
712 0887). The corresponding TCGA-BRCA patients' overall survival and progression-free interval
713 (synonymous to progression-free survival) data were obtained from the UCSC Xena browser
714 (https://xenabrowser.net). All data used in this study are either publicly available or available
715 upon request to the original authors publishing the data. All deconvolved expression data
716 generated in this study will be made publicly available upon publication.
717

718 CODE AVAILABILITY

719 All relevant codes for reproducing the figures in this study will be made publicly available upon
720 publication at https://github.com/ruppinlab/DECODEM/.
721

722 ACKNOWLEDGEMENTS

723 This research is supported in part by the Intramural Research Program of the National Institutes
724 of Health, National Cancer Institute, Center for Cancer Research. This work has utilized the
725 computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). The results
726 presented are in part based on the data generated by The Cancer Genome Atlas Research network
727 (https://www.cancer.gov/tcga/). We would additionally like to acknowledge Dr. E. Michael Gertz
728 and other members of the Cancer Data Science Laboratory for their help with data collection and
729 constructive feedback on this work. Figure 1 has been created with http://biorender.com/.
730

731 AUTHOR CONTRIBUTIONS

732 S.R.D, K.W., and E.R. conceived and designed the study. S.R.D., S.S., and K.W. performed the model
733 development and validation analyses. S.R.D., K.W., S.S., P.S.R., B.W., Y.S., E.S., S-J.S., C.C., and E.R.
734 acquired, analyzed, or interpreted the data. S.R.D., S.S., K.W., D.W., and B.W. worked on data
735 visualization. E.R. and K.W. supervised the study. All authors critically revised the manuscript for
736 important intelligent content.
737

COMPETING INTERESTS

E.R. is a co-founder of Medaware Ltd. (https://www.medaware.com/), Metabomed (https://www.metabomed.com/), and Pangea Biomed (https://pangeamedicine.com/). He has divested and serves as an unpaid scientific consultant to the latter company. The rest of the authors declare no conflicts of interest.

744 # REFERENCES

745 [1] Tsimberidou, A. M., Fountzilas, E., Nikanjam, M., & Kurzrock, R. (2020). Review of precision
746 cancer medicine: Evolution of the treatment paradigm. Cancer treatment reviews, 86,
747 102019.

748 [2] Huang K, Xiao C, Glass LM, Critchlow CM,. Machine learning applications for therapeutic
749 tasks with genomics data. Patterns 2021; 2(10):100328.

750 [3] Fisher, B., Brown, A., Mamounas, E., Wieand, S., Robidoux, A., Margolese, R. G., … &
751 Dimitrov, N. V. (1997). Effect of preoperative chemotherapy on local-regional disease in
752 women with operable breast cancer: findings from National Surgical Adjuvant Breast and
753 Bowel Project B-18. Journal of clinical oncology, 15(7), 2483-2493.

754 [4] Bear, H. D., Anderson, S., Brown, A., Smith, R., Mamounas, E. P., Fisher, B., ... & Wolmark, N.
755 (2003). The effect on tumor response of adding sequential preoperative docetaxel to
756 preoperative doxorubicin and cyclophosphamide: preliminary results from National Surgical
757 Adjuvant Breast and Bowel Project Protocol B-27. Journal of Clinical Oncology, 21(22), 4165-
758 4174.

759 [5] Golshan, M., Loibl, S., Wong, S. M., Huober, J. B., O'Shaughnessy, J., Rugo, H. S., ... & Untch,
760 M. (2020). Breast conservation after neoadjuvant chemotherapy for triple-negative breast
761 cancer: surgical results from the BrighTNess randomized clinical trial. JAMA surgery, 155(3),
762 e195410-e195410.

763 [6] Urueña, C., Lasso, P., Bernal-Estevez, D. et al. The breast cancer immune microenvironment
764 is modified by neoadjuvant chemotherapy. Sci Rep 12, 7981 (2022).
765 https://doi.org/10.1038/s41598-022-12108-5

766 [7] Ali, H. R., Dariush, A., Provenzano, E., Bardwell, H., Abraham, J. E., Iddawela, M., ... & Caldas,
767 C. (2016). Computational pathology of pre-treatment biopsies identifies lymphocyte density
768 as a predictor of response to neoadjuvant chemotherapy in breast cancer. Breast Cancer
769 Research, 18(1), 1-11.

770 [8] Ali, H. R., Dariush, A., Thomas, J., Provenzano, E., Dunn, J., Hiller, L., ... & Caldas, C. (2017).
771 Lymphocyte density determined by computational pathology validated as a predictor of

response to neoadjuvant chemotherapy in breast cancer: secondary analysis of the ARTemis trial. Annals of Oncology, 28(8), 1832-1835.

[9] Dieci, M. V., Prat, A., Tagliafico, E., Paré, L., Ficarra, G., Bisagni, G., ... & Guarneri, V. (2016). Integrated evaluation of PAM50 subtypes and immune modulation of pCR in HER2-positive breast cancer patients treated with chemotherapy and HER2-targeted agents in the CherLOB trial. Annals of Oncology, 27(10), 1867-1873.

[10] Griguolo, G., Serna, G., Pascual, T. et al. Immune microenvironment characterisation and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer. npj Precis. Onc. 5, 23 (2021). HYPERLINK "https://doi.org/10.1038/s41698-021-00163-6. https://doi.org/10.1038/s41698-021-00163-63434

[11] Sammut, SJ., Crispin-Ortuzar, M., Chin, Sf. et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature 601, 623–629 (2022). https://doi.org/10.1038/s41586-021-04278-5

[12] Klemm F, Joyce JA. Microenvironmental regulation of therapeutic response in cancer. Trends Cell Biol. 2015 Apr;25(4):198-213. doi: 10.1016/j.tcb.2014.11.006. Epub 2014 Dec 22. PMID: 25540894; PMCID: PMC5424264.

[13] Son B, Lee S, Youn H, Kim E, Kim W, Youn B. The role of tumor microenvironment in therapeutic resistance. Oncotarget. 2017 Jan 17;8(3):3933-3945. doi: 10.18632/oncotarget.13907. PMID: 27965469; PMCID: PMC5354804.

[14] Wu, P., Gao, W., Su, M., Nice, E. C., Zhang, W., Lin, J., & Xie, N. (2021). Adaptive mechanisms of tumor therapy resistance driven by tumor microenvironment. Frontiers in cell and developmental biology, 9, 641469.

[15] Mehraj, U., Dar, A.H., Wani, N.A. et al. Tumor microenvironment promotes breast cancer chemoresistance. Cancer Chemother Pharmacol 87, 147–158 (2021). https://doi.org/10.1007/s00280-020-04222-w

[16] Ruffell, B., & Coussens, L. M. (2015). Macrophages and therapeutic resistance in cancer. Cancer cell, 27(4), 462-472.

799 [17] Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P. G., ... & Massagué,
800 J. (2012). A CXCL1 paracrine network links cancer chemoresistance and metastasis. Cell,
801 150(1), 165-178.
802 [18] Li, M., Quintana, A., Alberts, E., Hung, M. S., Boulat, V., Ripoll, M. M., & Grigoriadis, A. (2023).
803 B Cells in Breast Cancer Pathology. Cancers, 15(5), 1517.
804 [19] Sakaguchi, A., Horimoto, Y., Onagi, H. et al. Plasma cell infiltration and treatment effect in
805 breast cancer patients treated with neoadjuvant chemotherapy. Breast Cancer Res 23, 99
806 (2021). https://doi.org/10.1186/s13058-021-01477-w
807 [20] Yeong, J., Lim, J. C. T., Lee, B., Li, H., Chia, N., Ong, C. C. H., ... & Iqbal, J. (2018). High densities
808 of tumor-associated plasma cells predict improved prognosis in triple negative breast cancer.
809 Frontiers in immunology, 9, 1209.
810 [21] Bhinder, B., Gilvary, C., Madhukar, N. S., & Elemento, O. (2021). Artificial intelligence in
811 cancer research and precision medicine. Cancer Discovery, 11(4), 900-915.
812 [22] Singla, N., & Singla, S. (2020). Harnessing big data with machine learning in precision
813 oncology. Kidney cancer journal: official journal of the Kidney Cancer Association, 18(3), 83.
814 [23] Senft, D., Leiserson, M. D., Ruppin, E., & Ze'ev, A. R. (2017). Precision oncology: the road
815 ahead. *Trends in molecular medicine*, *23*(10), 874-898.
816 [24] Tsimberidou, A. M., Fountzilas, E., Bleris, L., & Kurzrock, R. (2020, September).
817 Transcriptomics and solid tumors: The next frontier in precision cancer medicine. In Seminars
818 in cancer biology. Academic Press.
819 [25] Siravegna, G., Marsoni, S., Siena, S., & Bardelli, A. (2017). Integrating liquid biopsies into the
820 management of cancer. Nature reviews Clinical oncology, 14(9), 531-548.
821 [26] Heitzer, E., Haque, I. S., Roberts, C. E., & Speicher, M. R. (2019). Current and future
822 perspectives of liquid biopsies in genomics-driven oncology. Nature Reviews Genetics, 20(2),
823 71-88.
824 [27] Sawabata, N. (2020). Circulating tumor cells: From the laboratory to the cancer clinic.
825 Cancers, 12(10), 3065.

826 [28] Beaubier, N., Bontrager, M., Huether, R., Igartua, C., Lau, D., Tell, R., ... & White, K. P. (2019).
827 Integrated genomic profiling expands clinical options for patients with cancer. *Nature*
828 *biotechnology*, *37*(11), 1351-1360.
829 [29] Hayashi, H., Takiguchi, Y., Minami, H., Akiyoshi, K., Segawa, Y., Ueda, H., ... & Nakagawa, K.
830 (2020). Site-specific and targeted therapy based on molecular profiling by next-generation
831 sequencing for cancer of unknown primary site: a nonrandomized phase 2 clinical trial. JAMA
832 oncology, 6(12), 1931-1938.
833 [30] Vaske, O. M., Bjork, I., Salama, S. R., Beale, H., Shah, A. T., Sanders, L., ... & Haussler, D. (2019).
834 Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult
835 patients with cancer. *JAMA network open*, *2*(10), e1913968-e1913968.
836 [31] Lee, J. S., Nair, N. U., Dinstag, G., Chapman, L., Chung, Y., Wang, K., ... & Ruppin, E. (2021).
837 Synthetic lethality-mediated precision oncology via the tumor transcriptome. Cell, 184(9),
838 2487-2502.
839 [32] Sinha, S., Dhruba, S. R., Wu, W., Kerr, D. L., Stroganov, O. V., Grishagin, I., ... & Ruppin, E.
840 (2022). Predicting patient treatment response and resistance via single-cell transcriptomics
841 of their tumors. bioRxiv.
842 [33] Dinstag, G., Shulman, E. D., Elis, E., Ben-Zvi, D. S., Tirosh, O., Maimon, E., ... & Aharonov, R.
843 (2022). Clinically oriented prediction of patient response to targeted and immunotherapies
844 from the tumor transcriptome. bioRxiv.
845 [34] Fu, Y., Jung, A. W., Torne, R. V., Gonzalez, S., Vöhringer, H., Shmatko, A., ... & Gerstung, M.
846 (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and
847 prognosis. Nature Cancer, 1(8), 800-810.
848 [35] Noorbakhsh, J., Farahmand, S., Namburi, S., Caruana, D., Rimm, D., Soltanieh-ha, M., ... &
849 Chuang, J. H. (2020). Deep learning-based cross-classifications reveal conserved spatial
850 behaviors within tumor histological images. Nature communications, 11(1), 1-14.
851 [36] Yu, K. H., Wang, F., Berry, G. J., Re, C., Altman, R. B., Snyder, M., & Kohane, I. S. (2020).
852 Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional
853 neural networks. Journal of the American Medical Informatics Association, 27(5), 757-769.

854 [37] Couture, H. D., Williams, L. A., Geradts, J., Nyante, S. J., Butler, E. N., Marron, J. S., ... &
855 Niethammer, M. (2018). Image analysis with deep learning to predict breast cancer grade,
856 ER status, histologic subtype, and intrinsic subtype. NPJ breast cancer, 4(1), 1-8.
857 [38] Qu, H., Zhou, M., Yan, Z., Wang, H., Rustgi, V. K., Zhang, S., ... & Metaxas, D. N. (2021). Genetic
858 mutation and biological pathway prediction based on whole slide images in breast carcinoma
859 using deep learning. NPJ precision oncology, 5(1), 1-11.
860 [39] Hoang, D. T., Dinstag, G., Hermida, L. C., Ben-Zvi, D. S., Elis, E., Caley, K., ... & Ruppin, E. (2022).
861 Synthetic lethality-based prediction of cancer treatment response from histopathology
862 images. bioRxiv.
863 [40] Tanioka, M., Fan, C., Parker, J. S., Hoadley, K. A., Hu, Z., Li, Y., ... & Perou, C. M. (2018).
864 Integrated Analysis of RNA and DNA from the Phase III Trial CALGB 40601 Identifies
865 Predictors of Response to Trastuzumab-Based Neoadjuvant Chemotherapy in HER2-Positive
866 Breast CancerIntegrated Response Models in HER2-Positive Breast Cancer. Clinical Cancer
867 Research, 24(21), 5292-5304.
868 [41] Wong, M., Mayoh, C., Lau, L., Khuong-Quang, D. A., Pinese, M., Kumar, A., ... & Cowley, M. J.
869 (2020). Whole genome, transcriptome and methylome profiling enhances actionable target
870 discovery in high-risk pediatric cancer. Nature medicine, 26(11), 1742-1753.
871 [42] Rodon, J., Soria, J. C., Berger, R., Miller, W. H., Rubin, E., Kugel, A., ... & Kurzrock, R. (2019).
872 Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial.
873 Nature medicine, 25(5), 751-758.
874 [43] Kester, L., Seinstra, D., van Rossum, A. G., Vennin, C., Hoogstraat, M., van der Velden, D., ...
875 & van Rheenen, J. (2022). Differential Survival and Therapy Benefit of Patients with Breast
876 Cancer Are Characterized by Distinct Epithelial and Immune Cell Microenvironments Tumor
877 Cellular Composition Predicts Benefit to Therapies. *Clinical Cancer Research*, OF1-OF12.
878 [44] Wang, K., Patkar, S., Lee, J.S., Gertz, E.M., Robinson, W., Schischlik, F., Crawford, D.R.,
879 Schäffer, A.A. and Ruppin, E. (2022). Deconvolving Clinically Relevant Cellular Immune Cross-
880 talk from Bulk Gene Expression Using CODEFACS and LIRICS Stratifies Patients with
881 Melanoma to Anti–PD-1 Therapy. Cancer discovery, 12(4), 1088-1105.

882 [45] Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., ... & Swarbrick,
883 A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature*
884 *genetics*, *53*(9), 1334-1347.
885 [46] Earl, H. M., Hiller, L., Dunn, J. A., Blenkinsop, C., Grybowicz, L., Vallier, A. L., ... & Hayward, L.
886 (2015). Efficacy of neoadjuvant bevacizumab added to docetaxel followed by fluorouracil,
887 epirubicin, and cyclophosphamide, for women with HER2-negative early breast cancer
888 (ARTemis): an open-label, randomised, phase 3 trial. The lancet oncology, 16(6), 656-666.
889 [47] Loibl, S., O'Shaughnessy, J., Untch, M., Sikov, W. M., Rugo, H. S., McKee, M. D., ... & Geyer Jr,
890 C. E. (2018). Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to
891 standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a
892 randomised, phase 3 trial. The Lancet Oncology, 19(4), 497-509.
893 [48] Zhang, Y., Chen, H., Mo, H., Hu, X., Gao, R., Zhao, Y., ... & Liu, Z. (2021). Single-cell analyses
894 reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative
895 breast cancer. Cancer Cell, 39(12), 1578-1593.
896 [49] Korde, L. A., Somerfield, M. R., Carey, L. A., Crews, J. R., Denduluri, N., Hwang, E. S., ... &
897 Hershman, D. L. (2021). Neoadjuvant chemotherapy, endocrine therapy, and targeted
898 therapy for breast cancer: ASCO guideline. Journal of Clinical Oncology, 39(13), 1485-1505.
899 [50] Ramilowski, J., Goldberg, T., Harshbarger, J. *et al.* A draft network of ligand–receptor-
900 mediated multicellular signalling in human. *Nat Commun* **6**, 7866 (2015).
901 https://doi.org/10.1038/ncomms8866
902 [51] Green, D. R., Ferguson, T., Zitvogel, L., & Kroemer, G. (2009). Immunogenic and tolerogenic
903 cell death. Nature Reviews Immunology, 9(5), 353-363.
904 [52] Fabregat, A. et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 46, D649–D655
905 (2018).
906 [53] Molinari, M. (2000). Cell cycle checkpoints and their inactivation in human cancer. Cell
907 proliferation, 33(5), 261-274.
908 [54] Bower, J.J., Vance, L.D., Psioda, M. et al. Patterns of cell cycle checkpoint deregulation
909 associated with intrinsic molecular subtypes of human breast cancer cells. npj Breast Cancer
910 3, 9 (2017). https://doi.org/10.1038/s41523-017-0009-7

911 [55] Zhou, K., Sun, Y., Dong, D. et al. EMP3 negatively modulates breast cancer cell DNA
912 replication, DNA damage repair, and stem-like properties. Cell Death Dis 12, 844 (2021).
913 https://doi.org/10.1038/s41419-021-04140-6
914 [56] Repo, H., Löyttyniemi, E., Kurki, S. et al. A prognostic model based on cell-cycle control
915 predicts outcome of breast cancer patients. BMC Cancer 20, 558 (2020).
916 https://doi.org/10.1186/s12885-020-07045-3
917 [57] Vihervuori, H., Korpinen, K., Autere, T. A., Repo, H., Talvinen, K., & Kronqvist, P. (2022).
918 Varying outcomes of triple-negative breast cancer in different age groups–prognostic value
919 of clinical features and proliferation. Breast Cancer Research and Treatment, 1-12.
920 [58] Thu, K. L., Soria-Bretones, I., Mak, T. W., & Cescon, D. W. (2018). Targeting the cell cycle in
921 breast cancer: towards the next phase. Cell Cycle, 17(15), 1871-1885.
922 [59] Zhang, J., Chan, D. W., & Lin, S. Y. (2022). Exploiting DNA Replication Stress as a Therapeutic
923 Strategy for Breast Cancer. Biomedicines, 10(11), 2775.
924 [60] Zhu, X., & Zhou, W. (2015). The emerging regulation of VEGFR-2 in triple-negative breast
925 cancer. Frontiers in Endocrinology, 6, 159.
926 [61] Ceci, C., Atzori, M. G., Lacal, P. M., & Graziani, G. (2020). Role of VEGFs/VEGFR-1 signaling
927 and its inhibition in modulating tumor invasion: Experimental evidence in different
928 metastatic cancer models. International journal of molecular sciences, 21(4), 1388.
929 [62] Zhang, Q., Lu, S., Li, T., Yu, L., Zhang, Y., Zeng, H., ... & Lin, Y. (2019). ACE2 inhibits breast
930 cancer angiogenesis via suppressing the VEGFa/VEGFR2/ERK pathway. Journal of
931 Experimental & Clinical Cancer Research, 38(1), 1-12.
932 [63] Edwards, A., & Brennan, K. (2021). Notch signalling in breast development and cancer.
933 Frontiers in Cell and Developmental Biology, 9, 692173.
934 [64] Zhou, B., Lin, W., Long, Y. *et al.* Notch signaling pathway: architecture, disease, and
935 therapeutics. *Sig Transduct Target Ther* **7**, 95 (2022). https://doi.org/10.1038/s41392-022-
936 00934-y.
937 [65] Miao, K., Lei, J.H., Valecha, M.V. et al. NOTCH1 activation compensates BRCA1 deficiency and
938 promotes triple-negative breast cancer formation. Nat Commun 11, 3256 (2020).
939 https://doi.org/10.1038/s41467-020-16936-9

940 [66] BeLow, M., & Osipo, C. (2020). Notch signaling in breast cancer: a role in drug resistance.
941 Cells, 9(10), 2204.
942 [67] Li, L., Wang, X., Hu, K. et al. ZNF133 is a potent suppressor in breast carcinogenesis through
943 dampening L1CAM, a driver for tumor progression. Oncogene (2023).
944 https://doi.org/10.1038/s41388-023-02731-5
945 [68] van der Maten, M., Reijnen, C., Pijnenborg, J. M., & Zegers, M. M. (2019). L1 cell adhesion
946 molecule in cancer, a systematic review on domain-specific functions. International journal
947 of molecular sciences, 20(17), 4180.
948 [69] Zhang, J., Yang, F., Ding, Y., Zhen, L., Han, X., Jiao, F., & Tang, J. (2015). Overexpression of L1
949 cell adhesion molecule correlates with aggressive tumor progression of patients with breast
950 cancer and promotes motility of breast cancer cells. International journal of clinical and
951 experimental pathology, 8(8), 9240.
952 [70] Bandola-Simon, J., & Roche, P. A. (2019). Dysfunction of antigen processing and presentation
953 by dendritic cells in cancer. Molecular immunology, 113, 31-37.
954 [71] Dhatchinamoorthy, K., Colbert, J. D., & Rock, K. L. (2021). Cancer immune evasion through
955 loss of MHC class I antigen presentation. Frontiers in immunology, 12, 636568.
956 [72] McDonnell, A. M., Lesterhuis, W. J., Khong, A., Nowak, A. K., Lake, R. A., Currie, A. J., &
957 Robinson, B. W. (2015). Tumor-infiltrating dendritic cells exhibit defective cross-presentation
958 of tumor antigens, but is reversed by chemotherapy. European journal of immunology, 45(1),
959 49-59.
960 [73] Galluzzi, L., Humeau, J., Buqué, A., Zitvogel, L., & Kroemer, G. (2020). Immunostimulation
961 with chemotherapy in the era of immune checkpoint inhibitors. Nature reviews Clinical
962 oncology, 17(12), 725-741.
963 [74] Li, J., Li, C., Liu, X. et al. GDF9 concentration in embryo culture medium is linked to human
964 embryo quality and viability. J Assist Reprod Genet 39, 117–125 (2022).
965 https://doi.org/10.1007/s10815-021-02368-x
966 [75] Stocker, W. A., Walton, K. L., Richani, D., Chan, K. L., Beilby, K. H., Finger, B. J., ... & Harrison,
967 C. A. (2020). A variant of human growth differentiation factor-9 that improves oocyte
968 developmental competence. Journal of Biological Chemistry, 295(23), 7981-7991.

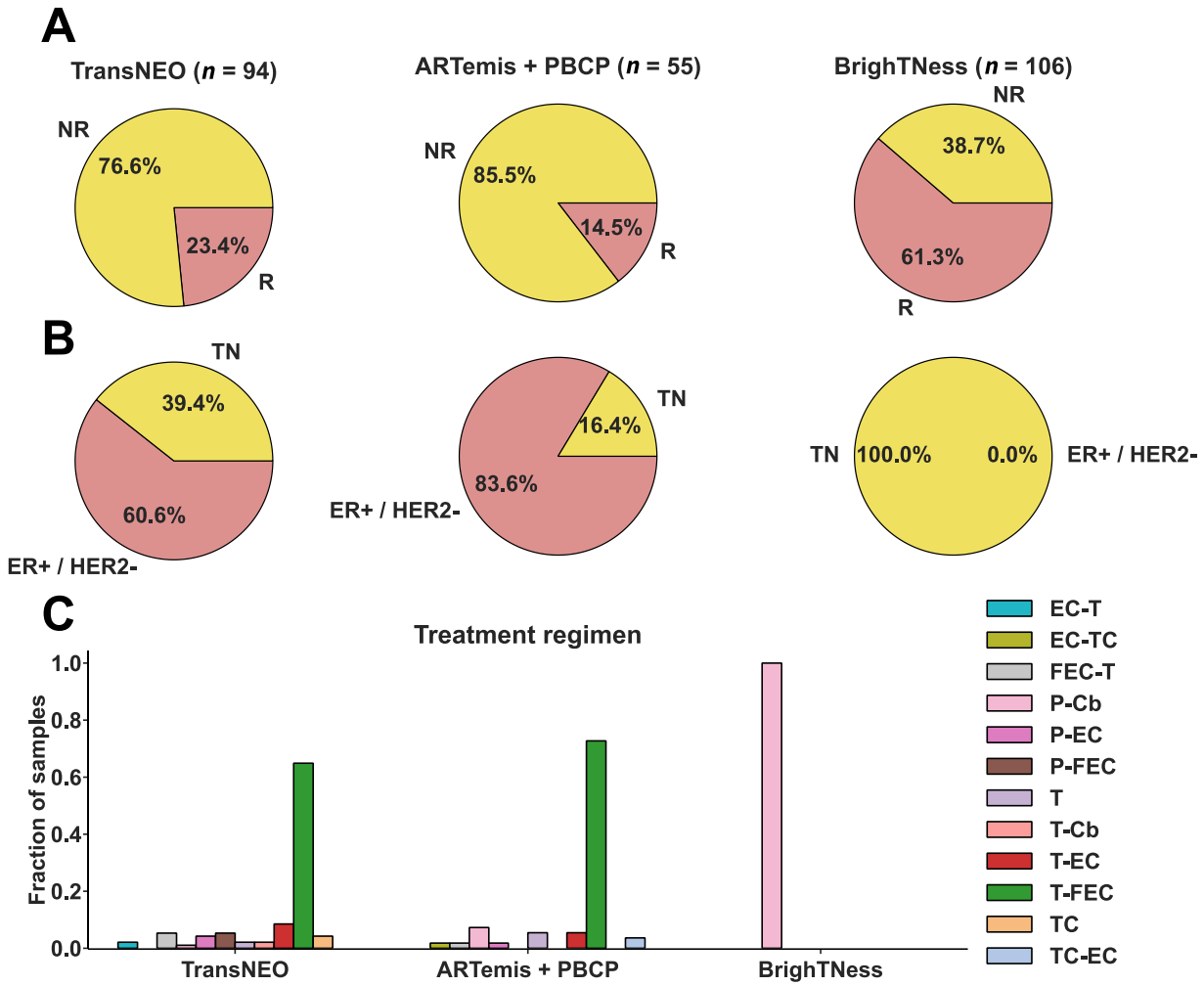
969 [76] Gomez-Puerto, M. C., Iyengar, P. V., García de Vinuesa, A., Ten Dijke, P., & Sanchez-Duffhues,
970 G. (2019). Bone morphogenetic protein receptor signal transduction in human disease. The
971 journal of pathology, 247(1), 9-20.
972 [77] Feary, E. S., Juengel, J. L., Smith, P., French, M. C., O'Connell, A. R., Lawrence, S. B., ... &
973 McNatty, K. P. (2007). Patterns of expression of messenger RNAs encoding GDF9, BMP15,
974 TGFBR1, BMPR1B, and BMPR2 during follicular development and characterization of ovarian
975 follicular populations in ewes carrying the Woodlands FecX2W mutation. Biology of
976 reproduction, 77(6), 990-998.
977 [78] Hanavadi, S., Martin, T. A., Watkins, G., Mansel, R. E., & Jiang, W. G. (2007). The role of
978 growth differentiation factor-9 (GDF-9) and its analog, GDF-9b/BMP-15, in human breast
979 cancer. Annals of surgical oncology, 14, 2159-2166.
980 [79] Harrath, A. H., Jalouli, M., Oueslati, M. H., Farah, M. A., Feriani, A., Aldahmash, W., ... &
981 Alwasel, S. (2021). The flavonoid, kaempferol-3-O-apiofuranosyl-7-O-rhamnopyranosyl, as a
982 potential therapeutic agent for breast cancer with a promoting effect on ovarian function.
983 Phytotherapy Research, 35(11), 6170-6180.
984 [80] Bach, D. H., Park, H. J., & Lee, S. K. (2018). The dual role of bone morphogenetic proteins in
985 cancer. Molecular Therapy-Oncolytics, 8, 1-13.
986 [81] Bokobza, S. M., Ye, L., Kynaston, H. E., Mansel, R. E., & Jiang, W. G. (2009). Reduced
987 expression of BMPR-IB correlates with poor prognosis and increased proliferation of breast
988 cancer cells. Cancer genomics & proteomics, 6(2), 101-108.
989 [82] Zheng, Y., Jiang, X., Wang, M., Yang, S., Deng, Y., Li, Y., ... & Chen, L. (2022). BMPR1B
990 Polymorphisms (rs1434536 and rs1970801) are Associated With Breast Cancer Susceptibility
991 in Northwest Chinese Han Females: A Case-Control Study. Clinical Breast Cancer, 22(5), e641-
992 e646.
993 [83] Dai, K., Qin, F., Zhang, H., Liu, X., Guo, C., Zhang, M., ... & Ma, Y. (2016). Low expression of
994 BMPRIB indicates poor prognosis of breast cancer and is insensitive to taxane-anthracycline
995 chemotherapy. Oncotarget, 7(4), 4770.

996 [84] Jing, S., Wen, D., Yu, Y., Holst, P. L., Luo, Y., Fang, M., ... & Fox, G. M. (1996). GDNF–induced
997 activation of the ret protein tyrosine kinase is mediated by GDNFR-α, a novel receptor for
998 GDNF. Cell, 85(7), 1113-1124.
999 [85] Fielder, G. C., Yang, T. W. S., Razdan, M., Li, Y., Lu, J., Perry, J. K., ... & Liu, D. X. (2018). The
1000 GDNF family: a role in cancer?. Neoplasia, 20(1), 99-117.
1001 [86] Li, Q., Cao, Z., & Zhao, S. (2022). The Emerging Portrait of Glial Cell Line-derived Neurotrophic
1002 Factor Family Receptor Alpha (GFRα) in Cancers. International Journal of Medical Sciences,
1003 19(4), 659.
1004 [87] Pecar, G., Liu, S., Hooda, J. et al. RET signaling in breast cancer therapeutic resistance and
1005 metastasis. Breast Cancer Res 25, 26 (2023). https://doi.org/10.1186/s13058-023-01622-7
1006 [88] Horibata, S., Rice, E. J., Mukai, C., Marks, B. A., Sams, K., Zheng, H., ... & Danko, C. G. (2018).
1007 ER-positive breast cancer cells are poised for RET-mediated endocrine resistance. PLoS One,
1008 13(4), e0194023.
1009 [89] Bhakta, S., Crocker, L. M., Chen, Y., Hazen, M., Schutten, M. M., Li, D., ... & Junutula, J. R.
1010 (2018). An Anti-GDNF Family Receptor Alpha 1 (GFRA1) Antibody–Drug Conjugate for the
1011 Treatment of Hormone Receptor–Positive Breast Cancer Targeting Hormone-Positive Breast
1012 Cancer with a GFRA1 ADC. Molecular cancer therapeutics, 17(3), 638-649.
1013 [90] Kim, M., & Kim, D. J. (2018). GFRA1: a novel molecular target for the prevention of
1014 osteosarcoma chemoresistance. International journal of molecular sciences, 19(4), 1078.
1015 [91] Liu, H., Yang, Z., Lu, W., Chen, Z., Chen, L., Han, S., ... & Cai, Y. (2020). Chemokines and
1016 chemokine receptors: A new strategy for breast cancer therapy. Cancer medicine, 9(11),
1017 3786-3799.
1018 [92] Liang, Y., Yi, L., Liu, P., Jiang, L., Wang, H., Hu, A., … & Dong, J. (2018). CX3CL1 involves in
1019 breast cancer metastasizing to the spine via the Src/FAK signaling pathway. Journal of
1020 Cancer, 9(19), 3603.
1021 [93] Tardáguila, M., Mira, E., García-Cabezas, M. A., Feijoo, A. M., Quintela-Fandino, M., Azcoitia,
1022 I., ... & Manes, S. (2013). CX3CL1 Promotes Breast Cancer via Transactivation of the EGF
1023 PathwayCX3CL1 Promotes Breast Cancer. Cancer research, 73(14), 4461-4473.

1024 [94] Jamieson-Gladney, W.L., Zhang, Y., Fong, A.M. et al. The chemokine receptor CX3CR1 is
1025 directly involved in the arrest of breast cancer cells to the skeleton. Breast Cancer Res 13,
1026 R91 (2011). https://doi.org/10.1186/bcr3016
1027 [95] Rivas-Fuentes, S., Salgado-Aguayo, A., Arratia-Quijada, J., & Gorocica-Rosete, P. (2021).
1028 Regulation and biological functions of the CX3CL1-CX3CR1 axis and its relevance in solid
1029 cancer: A mini-review. Journal of Cancer, 12(2), 571.
1030 [96] Yue, Y., Zhang, Q., & Sun, Z. (2022). CX3CR1 acts as a protective biomarker in the tumor
1031 microenvironment of colorectal Cancer. Frontiers in Immunology, 12, 5972.
1032 [97] Ishida, Y., Kuninaka, Y., Yamamoto, Y., Nosaka, M., Kimura, A., Furukawa, F., ... & Kondo, T.
1033 (2020). Pivotal involvement of the CX3CL1-CX3CR1 axis for the recruitment of M2 tumor-
1034 associated macrophages in skin carcinogenesis. *Journal of Investigative*
1035 *Dermatology*, *140*(10), 1951-1961.
1036 [98] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
1037 [99] Michaud, M., Martins, I., Sukkurwala, A. Q., Adjemian, S., Ma, Y., Pellegatti, P., ... & Kroemer,
1038 G. (2011). Autophagy-dependent anticancer immune responses induced by
1039 chemotherapeutic agents in mice. Science, 334(6062), 1573-1577.
1040 [100] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & Mering, C.
1041 V. (2019). STRING v11: protein–protein association networks with increased coverage,
1042 supporting functional discovery in genome-wide experimental datasets. Nucleic acids
1043 research, 47(D1), D607-D613.
1044 [101] Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., ... &
1045 Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues
1046 with digital cytometry. Nature biotechnology, 37(7), 773-782.
1047 [102] Han, Y., Wang, Y., Dong, X., Sun, D., Liu, Z., Yue, J., ... & Wang, C. (2023). TISCH2: expanded
1048 datasets and new tools for single-cell transcriptome analyses of the tumor
1049 microenvironment. Nucleic acids research, 51(D1), D1425-D1431.
1050 [103] Kumar MP, Du J, Lagoudas G, Jiao Y, Sawyer A, Drummond DC, et al. Analysis of Single-
1051 Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. Cell

1052 Rep [Internet]. 2018;25(6):1458-1468.e4. Available from:
1053 https://www.sciencedirect.com/science/article/pii/S221112471831636X
1054 [104] Leon-Ferre, R. A., Jonas, S. F., Salgado, R., Loi, S., De Jong, V., Carter, J. M., ... &
1055 International Immuno-Oncology Biomarker Working Group. (2024). Tumor-Infiltrating
1056 Lymphocytes in Triple-Negative Breast Cancer. JAMA, 331(13), 1135-1144.
1057 [105] Ali, H. R., Chlon, L., Pharoah, P. D., Markowetz, F., & Caldas, C. (2016). Patterns of immune
1058 infiltration in breast cancer and their clinical implications: a gene-expression-based
1059 retrospective study. PLoS medicine, 13(12), e1002194.
1060 [106] Chowdhury, A., Pharoah, P. D., & Rueda, O. M. (2023). Evaluation and comparison of
1061 different breast cancer prognosis scores based on gene expression data. Breast Cancer
1062 Research, 25(1), 17.
1063

1064 ## SUPPLEMENTARY FIGURES



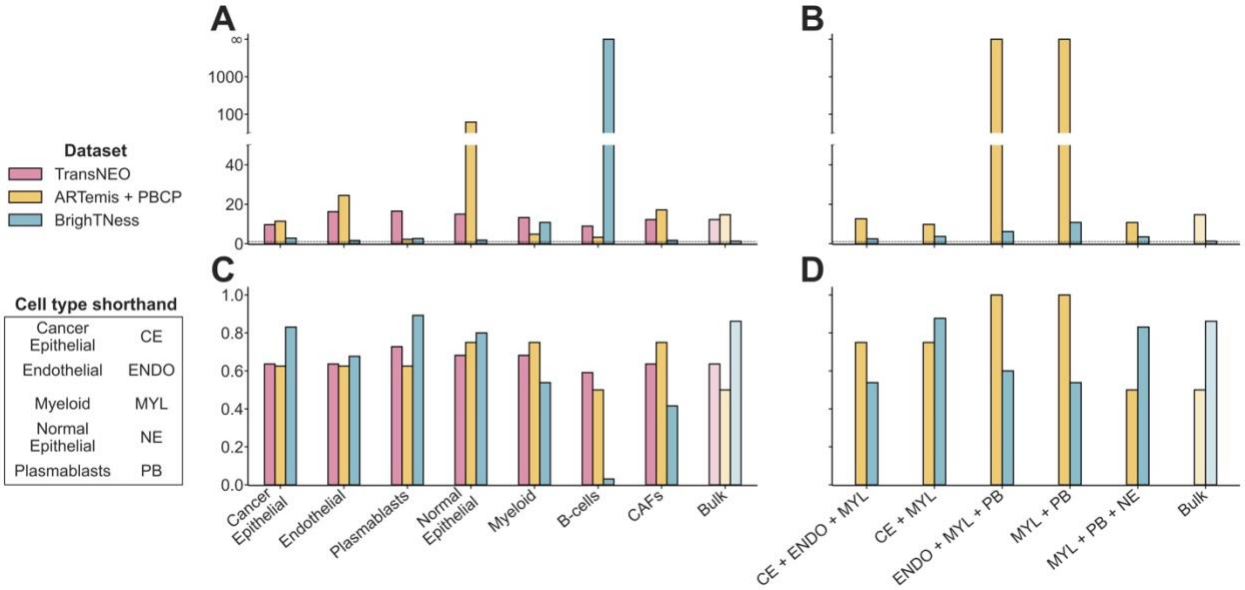
1065

1066 **Supplementary Figure 1: Summary of the three bulk transcriptomics cohorts used.**

1067 **A.** Distribution of responders (R) vs. non-responders (NR) across the three cohorts. R and NR
1068 indicate the patients who achieved pCR following neoadjuvant treatment and those who did not,
1069 respectively.

1070 **B.** Distribution of breast cancer subtypes across the three cohorts. ER+ / HER2- and TN stand for
1071 ER-positive, HER2-negative tumor subtype and triple negative tumor subtype, respectively.

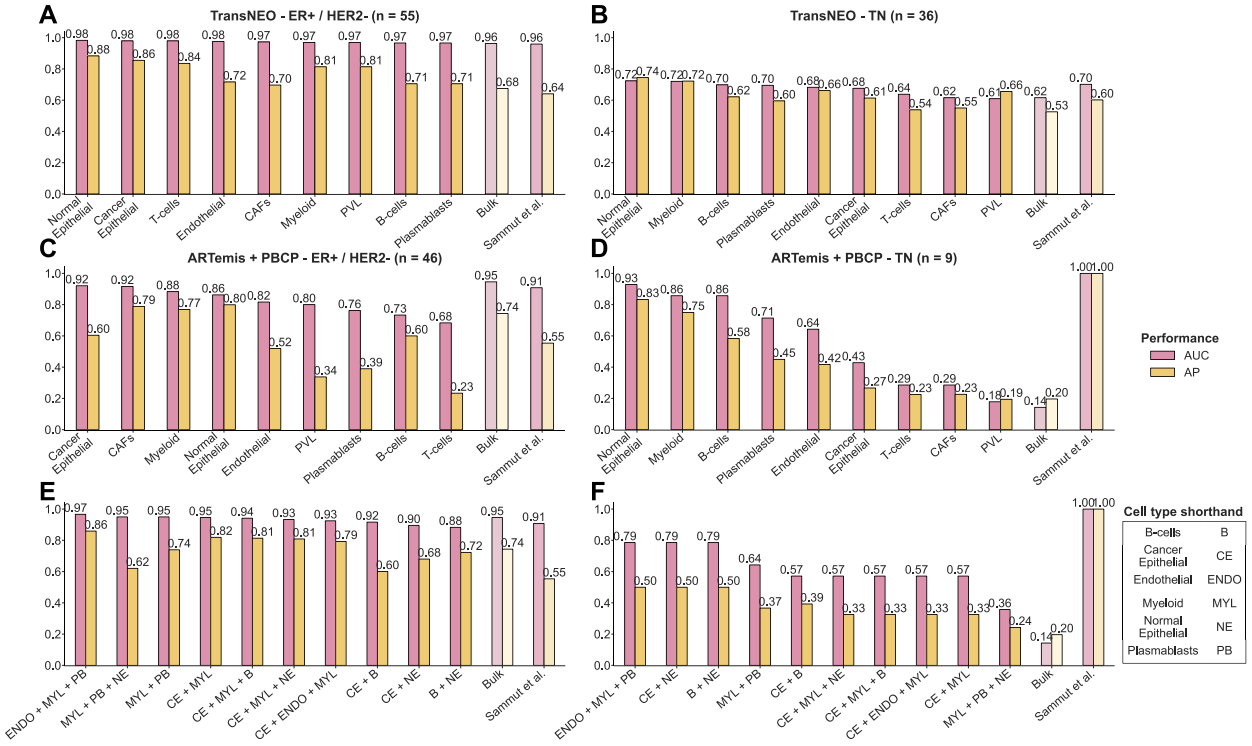
1072 **C.** Fractions of chemotherapy regimens used for neoadjuvant treatment across the three cohorts.
1073 E = Epirubicin, C = Cyclophosphamide, F = 5-Flurouracil, T = Docetaxel (Taxotere), P = Paclitaxel
1074 (Taxol), Cb = Carboplatin.



Supplementary Figure 2: The prominent cell types and ensembles mediating chemotherapy response in breast cancer tumor microenvironment.

A-B. Odds ratios for the prominent cell types (**A**) and multi-cell-type ensembles (**B**) as compared to the bulk across three cohorts. Cell types are ranked by their area under the receiver operating characteristics curve (AUC) values in cross-validation in a descending order (see **Figure 2B**). Ensembles are ranked by their AUC values in external validation in a descending order (see **Figure 3A-D**).

C-D. Sensitivities for the prominent cell types (**C**) and multi-cell-type ensembles (**D**) as compared to the bulk across the three cohorts. Cell types are ranked by their AUC values in cross-validation in a descending order (see **Figure 2B**). Cell ensembles are ranked by their AUC values in external validation in a descending order (see **Figure 3A-D**).

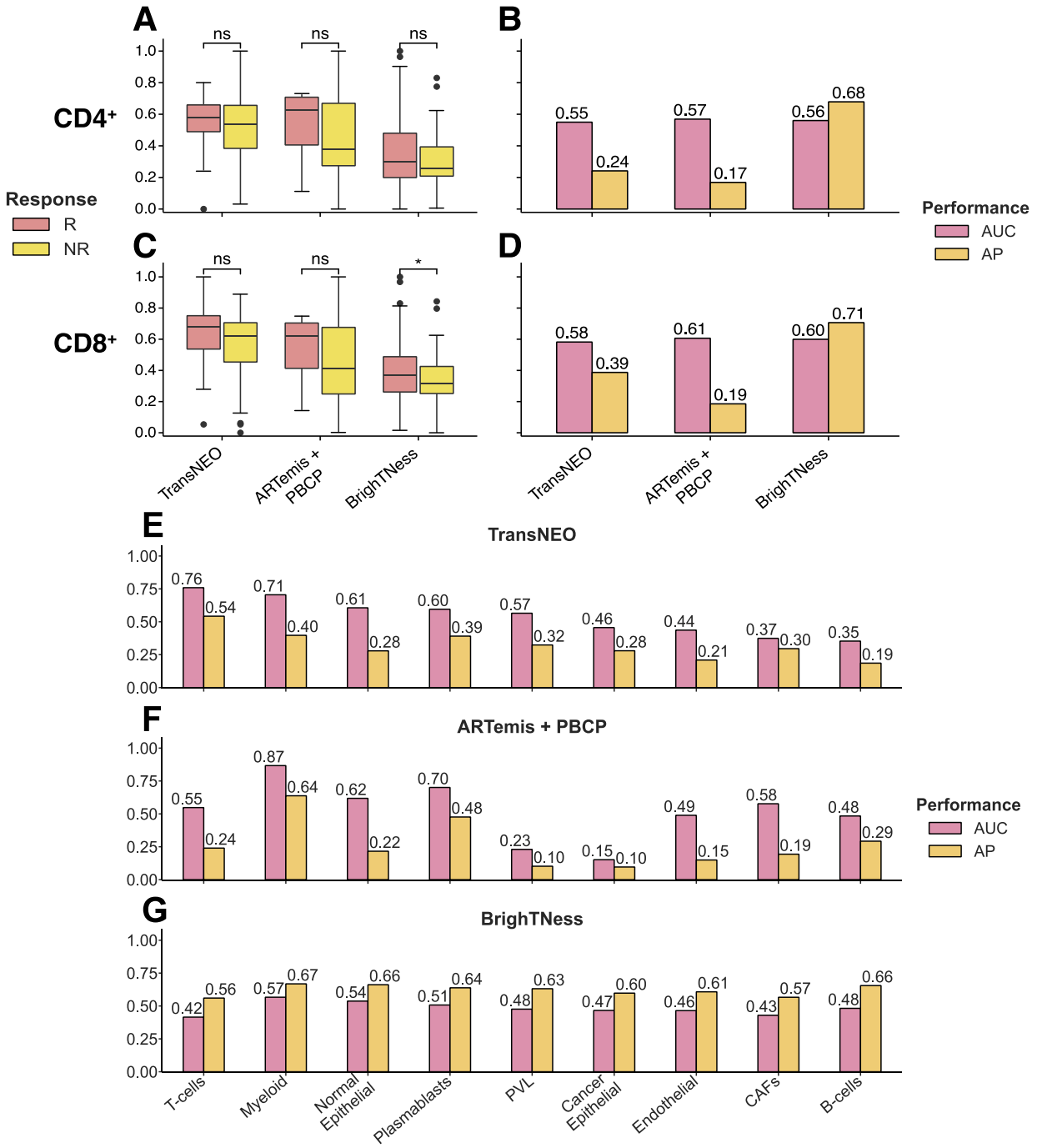


Supplementary Figure 3: The prominent cell types and ensembles mediating chemotherapy response in breast cancer tumor microenvironment in a subtype-specific manner.

A-B. Comparison of model performances in cross-validation with TransNEO for the ER-positive, HER2-negative (ER+ / HER2-; **A**) and triple negative (TN; **B**) subtypes across nine cell-type-specific, bulk, and Sammut et al. predictors. AUC and AP stand for the area under the receiver operating characteristics curve and average precision (equivalent to the area under precision-recall curve), respectively. Cell types are ranked by their AUC values in a descending order.

C-D. Comparison of model performances in external validation with ARTemis + PBCP for the ER+ / HER2- (**C**) and TN (**D**) subtypes across nine cell-type-specific, bulk, and Sammut et al. predictors. Cell types are ranked by their AUC values in a descending order.

E-F. Comparison of model performances in external validation with ARTemis + PBCP for the ER+ / HER2- (**E**) and TN (**F**) subtypes across five most prominent two-cell-ensembles, five most prominent three-cell-ensembles, bulk, and Sammut et al. predictors. Ensembles are ranked by their AUC values in a descending order.



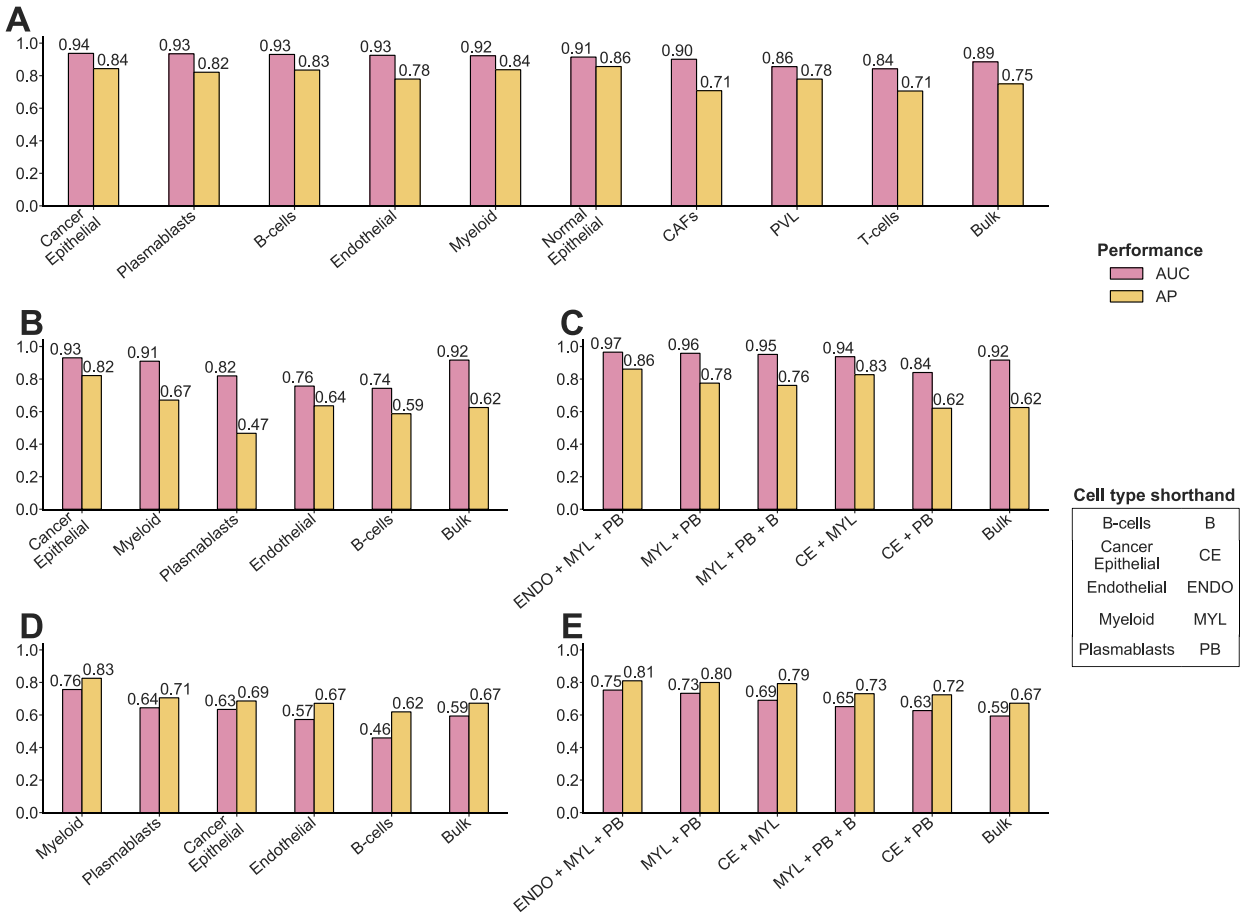
Supplementary Figure 4: The CD4⁺ / CD8⁺ T-cells enrichment and cell abundance – clinical response association analyses reaffirm the poor chemotherapy response stratification by T-cells at the current resolution.

A-B. GSVA enrichment scores (**A**) and predictive performance (**B**) of CD4⁺ T-cells for chemotherapy response stratification across TransNEO (*n* = 94), ARTemis + PBCP (*n* = 55), and BrighTNess (*n* =

106) cohorts. R and NR stand for responders and non-responders, respectively. The differences between the prediction scores were computed by using a one-tailed Wilcoxon rank-sum test. AUC and AP stand for the area under the receiver operating characteristics curve and average precision (equivalent to the area under precision-recall curve), respectively.

C-D. GSVA enrichment scores (**C**) and predictive performance (**D**) of $CD8^+$ T-cells for chemotherapy response stratification across the three cohorts.

E-G. Predictive performance of abundances of the nine cell types for TransNEO (**E**), ARTemis + PBCP (**F**), and BrighTNess (**G**) cohorts. Cell types are ranked by their AUC values for TransNEO in a descending order.

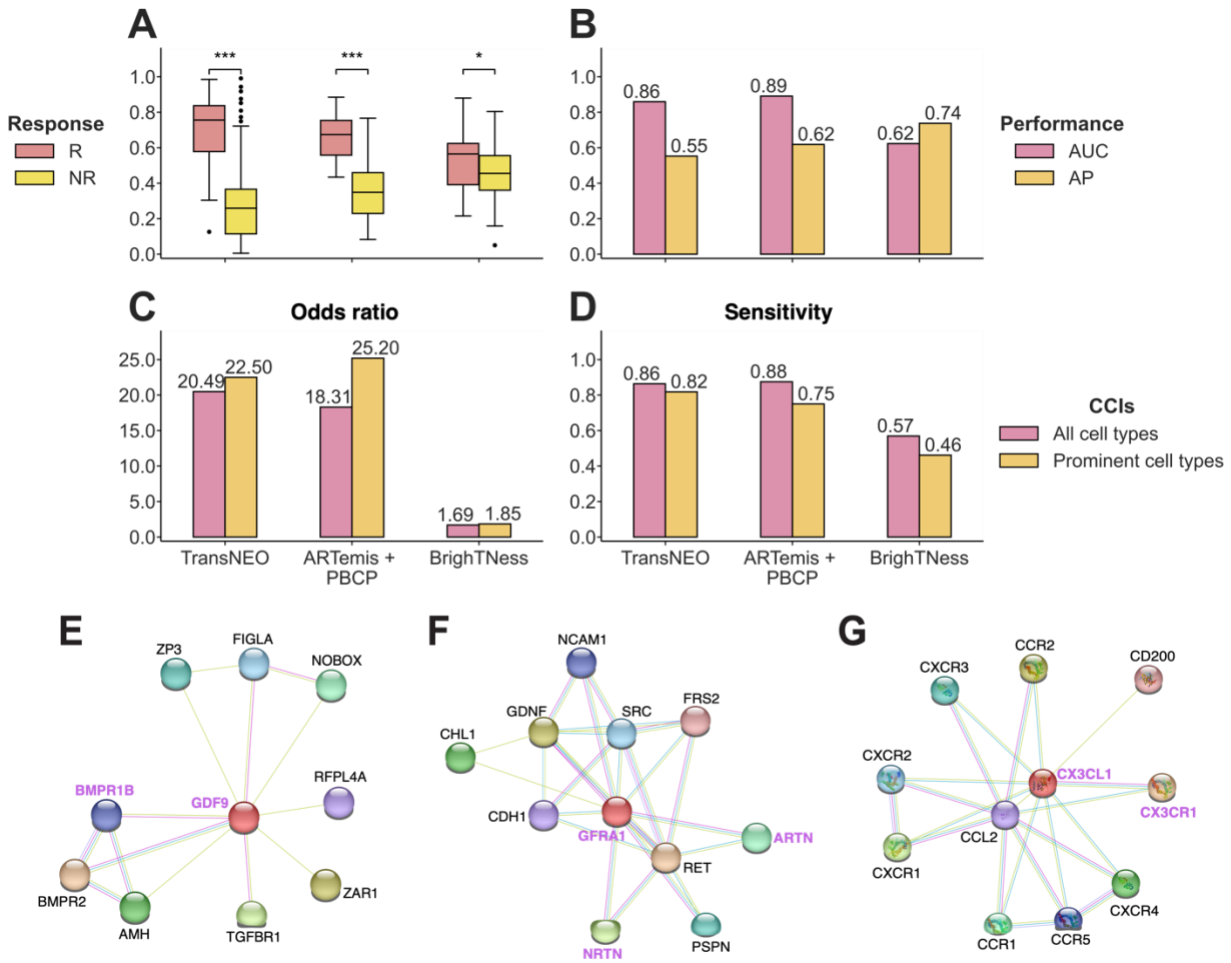


Supplementary Figure 5: The prominent immune cell types and ensembles mediating response to the most prevalent immunogenic chemotherapy regimen in each breast cancer cohort.

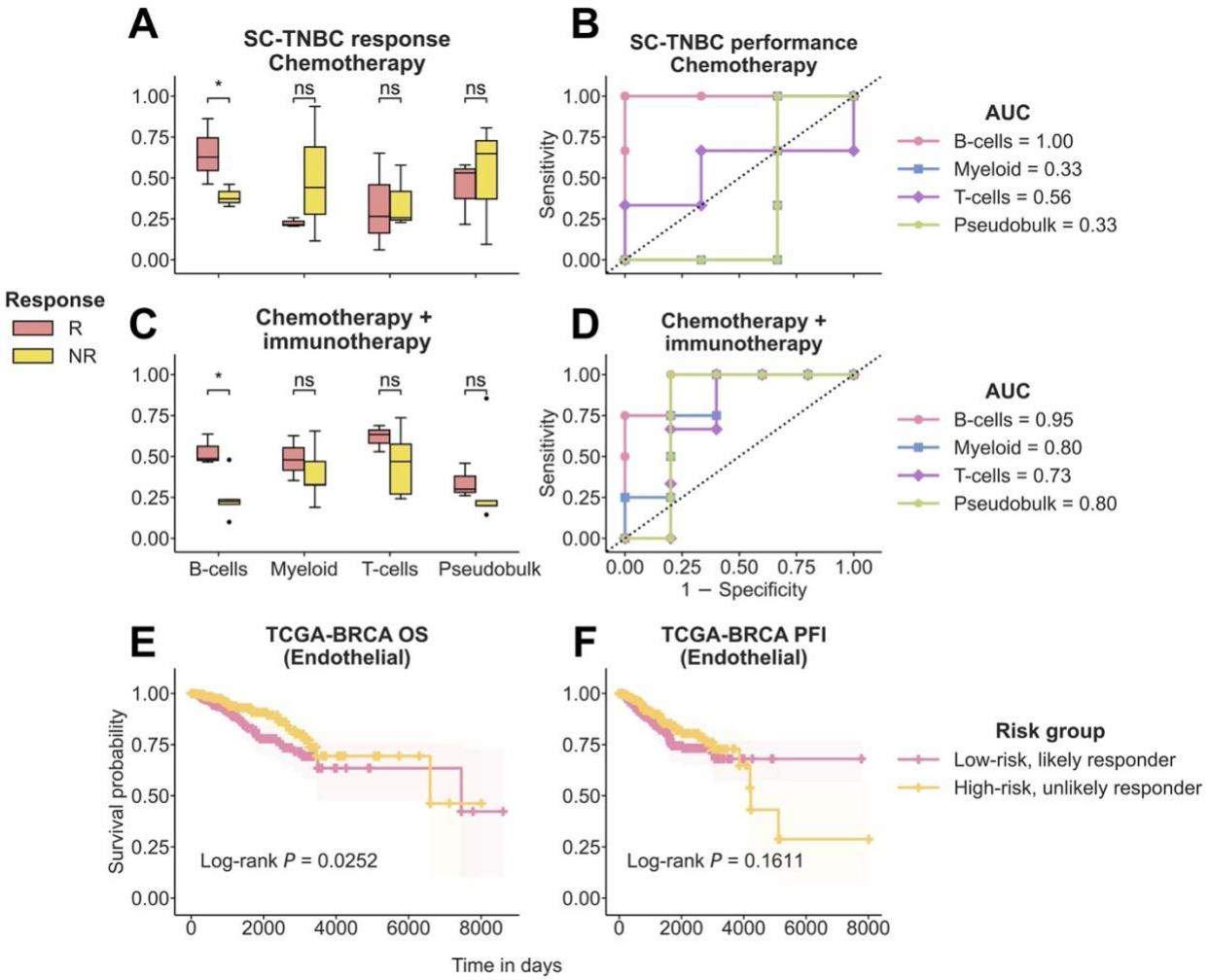
A. Comparison of model performances in cross-validation with TransNEO for the T-FEC regimen (*n* = 61) across nine cell-type-specific and bulk predictors. AUC and AP stand for the area under the receiver operating characteristics curve and average precision (equivalent to the area under precision-recall curve), respectively. Cell types are ranked by their AUC values in a descending order. Here, T = docetaxel/taxotere, F = 5-flurouracil, E = epirubicin, and C = cyclophosphamide.

B-C. Comparison of model performances in external validation with ARTemis + PBCP for the T-FEC regimen (*n* = 40) across the top cell-type-specific models (**B**) and top multi-cell-type ensembles (**C**). Cell types and ensembles are ranked by their AUC values in a descending order.

D-E. Comparison of model performances in external validation with BrighTNess for paclitaxel + carboplatin regimen (*n* = 106) across the top cell-type-specific models (**D**) and top multi-cell-type ensembles (**E**). Cell types and ensembles are ranked by their AUC values in a descending order.



1137 **Supplementary Figure 6: The cell-cell interactions (CCIs) encompassing prominent cell types**
1138 **capture the performance of CCIs using all nine cell types in mediating chemotherapy response**
1139 **and are validated in single-cell (SC) breast cancer tumor microenvironment.**
1140 **A-B.** Prediction scores (**A**) and model performance (**B**) for DECODEMi using all available CCIs in
1141 cross-validation with TransNEO (*n* = 94) and in external validation with ARTemis + PBCP (*n* = 55)
1142 and BrighTNess (*n* = 106). R and NR stand for responders and non-responders, respectively. The
1143 differences between the prediction scores were computed by using a one-tailed Wilcoxon rank-
1144 sum test. AUC and AP stand for the area under the receiver operating characteristics curve and
1145 average precision (equivalent to the area under precision-recall curve), respectively.
1146 **C-D.** Odds ratios (**C**) and sensitivities (**D**) for DECODEMi using different sets of CCIs in cross-
1147 validation with TransNEO (*n* = 94) and in external validation with ARTemis + PBCP (*n* = 55) and
1148 BrighTNess (*n* = 106).
1149 **E-G.** Protein-protein interaction networks obtained from STRING [48] displaying the known
1150 interaction networks for three prominent CCIs *i.e.*, *GDF9 – BMPR1B* (**E**), *NRTN – GFRA1* and *ARTN*
1151 *– GFRA1* (**F**), and *CX3CL1 – CX3CR1* (**G**).
1152



Supplementary Figure 7: The generalizability of DECODEM to single-cell (SC) transcriptomics and TCGA patient survival stratification.

A-B. Prediction scores (**A**) and the area under the receiver operating characteristics curve (AUC) values (**B**) for all available cell types in external validation with SC expression of TNBC patients (SC-TNBC) treated with chemotherapy (n = 6) from Zhang et al. [48]. Pseudobulk is shown as a baseline and computed as the mean expression across all available cells per patient after filtering for low-variance genes (threshold = 0.8). R and NR stand for responders and non-responders, respectively. The differences between the prediction scores were computed by using a one-tailed Wilcoxon rank-sum test.

C-D. Prediction scores (**C**) and the AUC values (**D**) for four cell types in external validation with SC expression of TNBC patients (SC-TNBC) treated with chemotherapy and immunotherapy (n = 9) from Zhang et al. [48].

1166 **E-F.** Kaplan-Meier curves depicting the overall survival (OS; **E**) and progression-free interval (PFI;
1167 **F**) of the 705 BC patients from TCGA (TCGA-BRCA). Patients were stratified into low-risk, likely
1168 responder (*n* = 353) and high-risk, unlikely responder (*n* = 352) groups by their median DECODEM
1169 score computed using the expression of endothelial cells, whereby the low-risk and high-risk
1170 groups comprised individuals with scores above and below the median, respectively. The
1171 differences between the two curves were computed by using a Log-rank test.
1172