

# Peripheral Blood Monocyte Abundance Predicts Outcomes in Breast Cancer Patients

**Justin Balko** (✉ [justin.balko@vumc.org](mailto:justin.balko@vumc.org))

Vanderbilt University Medical Center <https://orcid.org/0000-0002-4263-5974>

**Margaret Axelrod**

Vanderbilt University Medical Center <https://orcid.org/0000-0003-4713-8798>

**Yu Wang**

Vanderbilt University Medical Center <https://orcid.org/0000-0002-2819-4551>

**Yaomin Xu**

Vanderbilt University Medical Center <https://orcid.org/0000-0002-3752-4006>

**Xiaopeng Sun**

Vanderbilt University Medical Center

**Cosmin Bejan**

Vanderbilt University Medical Center

**Paula González-Ericsson**

Vanderbilt University Medical Center <https://orcid.org/0000-0002-6292-6963>

**Sara Nunnery**

Vanderbilt University Medical Center

**Riley Berhman**

Vanderbilt University Medical Center

**Joshua Donaldson**

Vanderbilt University Medical Center

**Angel Guerrero-Zotano**

Instituto Valenciano

**Chiara Massa**

Institute of Medical Immunology, Martin Luther University Halle-Wittenberg

**Barbara Seliger**

Institute of Medical Immunology, Martin Luther University Halle-Wittenberg

**Melinda Sanders**

Vanderbilt University Medical Center

**Ingrid Mayer**

Vanderbilt University Medical Center

**Keywords:** Biomarkers, breast cancer patients, Peripheral blood, monocyte abundance, outcomes, toxicity

**Posted Date:** November 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1017242/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Biomarkers of response are needed in breast cancer to stratify patients to appropriate therapies and avoid unnecessary toxicity. Peripheral blood gene expression and cell type abundance were used to identify biomarkers of response and recurrence in neoadjuvant chemotherapy treated breast cancer patients. Higher peripheral blood monocyte abundance after neoadjuvant chemotherapy was associated with improved prognosis in multiple independent cohorts of breast cancer patients.

## Main

Neoadjuvant chemotherapy (NAC), the standard of care for many breast cancer patients, is known to have systemic immunologic effects and is increasingly being used in clinical trials in combination with immunotherapeutics. Currently, there are few biomarkers to predict NAC or immunotherapy response, although response to NAC is known to be associated with long term outcome in breast cancer<sup>1</sup>. Thus, biomarkers are needed to identify patients who will benefit from combination therapy compared to those who are likely to respond to NAC alone, and thereby avoid the added risk of toxicity and financial burden. Peripheral blood is an attractive site of biomarker development due to the relative ease of longitudinal sampling. We have previously shown that high expression of a cytotoxicity gene signature in the blood following NAC is associated with the presence of residual disease (RD) and future breast cancer recurrence, demonstrating the feasibility of using blood-based transcriptional biomarkers<sup>2</sup>.

RNA sequencing was performed on whole blood of 53 breast cancer patients after completion of NAC (if received) and prior to definitive surgery (Fig. 1a; n=23 RD, 9 pathologic complete response (pCR), 21 no NAC; Table 1). We stratified patients with RD by whether they experienced a breast cancer recurrence within three years of surgery (RD-R) or remained free of recurrence for three years (RD-nR). Using DeSeq, we identified 1,238 (FDR corrected q-value <0.1) differentially expressed genes between pCR and RD samples (Supplemental Table 1)<sup>3</sup>. Using gene set enrichment analysis (GSEA), we collapsed differentially expressed genes into pathways using the Molecular Signatures Database hallmark gene sets<sup>4,5</sup>. Hallmark Interferon (IFN) Gamma Response (q-value <0.0001; normalized enrichment score (NES)=3.32), Hallmark Interferon Alpha Response (q-value <0.0001; NES=3.14), and Hallmark Complement (q-value=0.000111; NES=2.29) pathways were significantly enriched in the blood of patients experiencing pCR compared to those with RD (Fig. 1b). No pathways were statistically significantly upregulated in RD samples relative to pCR samples. To evaluate the genes involved in these pathways, we identified the leading-edge genes from each pathway (IFN gamma = 49 genes; IFN alpha = 26 genes; complement = 15 genes) and selected only the unique genes (n=60 genes). There is strong, uniform upregulation of many of these genes in many of the pCR samples, regardless of TNBC status (Supplementary Fig. 1). We combined expression of these genes into an IFN/complement score, calculated as sum of z-scores divided by number of genes in the signature (n=60 genes). We compared expression of the IFN/complement score to a previously published 8 gene cytotoxic score (*FGFBP2* + *GNLY* + *GZMB* + *GZMH* + *NKG7* + *LAG3* + *PDCD1* - *HLA-G*)<sup>2</sup>. No genes overlapped between the two signatures. Samples with the highest expression of the

IFN/complement had low expression of the cytotoxic score and tended to be pCR samples. Conversely, those with highest expression of the cytotoxic score tended to have low expression of the IFN/complement score and be RD samples (Fig. 1c). A combination peripheral immunologic response score (PIRS) of IFN/complement score minus cytotoxic score had improved predictive power compared to either signature alone (Fig. 1d). To examine which cell types predominately express each signature, we used single cell RNA sequencing data from whole peripheral blood mononuclear cells (PBMCs) from two breast cancer patients post-NAC, prior to surgery<sup>2</sup>. Expression of the cytotoxic score was the highest in CD8+ T cells and natural killer cells, while the IFN/complement score was the highest in a subset of monocytes. There was very little co-expression of the signatures across cells (Fig. 1e).

Given the partitioning of the gene expression scores into cell types, we next aimed to identify whether there were differences in cell type abundances between the outcome groups. CIBERSORTx was used to deconvolute relative cell type abundance from the RNA sequencing data<sup>6</sup> (Fig. 2a). Relative monocyte abundance was highest in samples with pCR, intermediate in those with RD, and lowest in samples not receiving NAC (Fig. 2b). Naïve B cells were also statistically significantly different across groups, being highest in no NAC samples and lowest in pCR samples (Supplementary Fig. 2a). However, only monocytes followed the trend of increases from no NAC to RD-R to RD-nR to pCR (Supplementary Fig. 2b). Additionally, the routine nature of clinically measuring monocytes made monocyte values an intriguing metric for further study. We reviewed electronic medical records and extracted monocyte values from complete blood counts for patients receiving NAC in this cohort. 23 out of 32 (72%) of NAC-treated patients (n=7 pCR, 12 RD-nR, 4 RD-R) had a complete blood count with differential (which includes monocyte values) in the 30-day interval prior to surgery (following completion of NAC), indicating the commonality of collecting this information clinically. Clinically measured monocyte values were significantly positively correlated with monocyte values inferred by CIBERSORTx ( $R=0.51$ ,  $p=0.012$ ; Supplementary Fig. 2c), even though monocyte values were not always collected on the same day as the blood for RNA sequencing (though in the same 30-day window). Post-NAC, but not pre-NAC, monocytes were significantly higher in patients with pCR compared to those with RD (Fig. 2c; Supplementary Fig. 2d).

Next, we sought to assess whether monocyte abundance was associated with outcome in independent cohorts. Higher monocytes were also seen with pCR in an additional cohort of 41 TNBC patients (n=18 RD, 23 pCR; placebo arm of the GeparNuevo study; Table1), though this association was not statistically significant ( $p=0.0638$  for absolute monocyte counts,  $p=0.186$  for relative monocyte frequencies, one-tailed Wilcoxon; Supplementary Fig. 2e)<sup>7</sup>. In this TNBC only dataset PIRS, measured by NanoString, was not associated with outcome, indicating independence of monocytes and PIRS measurements (Supplementary Fig. 2f). In an additional independent cohort of 14 hormone receptor positive (HR+) HER2- breast cancer patients from the Instituto Valenciano de Oncología, monocytes tended to be higher in patients without metastatic recurrence, with at least four years of follow-up time for each patient ( $p=0.0949$ , one-tailed Wilcoxon; n= 5 with metastasis, 9 without metastasis; Supplementary Fig. 2g; Table 1). Using a de-identified medical record database called the synthetic derivative (SD), we identified 110

breast cancer patients (VICC-SD; n=35 pCR, 75 RD; Table 1) who had been treated with NAC, had a breast surgery, and had a monocyte value within 30 days prior to surgery. In the VICC-SD cohort, relative frequencies of monocytes were statistically significantly higher in patients with pCR compared to those with RD (Fig. 2d). This effect was more pronounced when considering only the TNBC patients (n=50), which may be reflective of underlying TNBC-specific biology, or the more uniform treatment options for TNBC (chemotherapy rather than targeted therapy agents) (Supplementary Fig. 2h). In all cohorts, patients who had received GM-CSF products in the 30-day window prior to surgery were excluded from analysis as this may affect monocyte counts. Taken together, these data suggest that higher blood monocyte levels post-NAC may be indicative of superior outcomes in breast cancer patients.

Peripheral blood gene expression scores and cell type abundance may be useful biomarkers of NAC response and outcomes in breast cancer. We identified an immunologic gene signature (PIRS) that was highest in patients with the best outcomes (pCR) and lowest in those with the worst outcome (RD with recurrence). However, PIRS was not associated with outcome in a separate cohort of TNBC only patients. Additional studies are needed to test whether PIRS or other gene expression scores may be useful biomarkers in breast cancer patients. Higher peripheral monocytes, a standard clinical assay performed on most breast cancer patients, was associated with improved patient outcomes (pCR or lack of recurrence) in four independent breast cancer patient cohorts. Additional efforts are needed to explore whether there might be a causal link between chemotherapy induced monocyte mobilization and improved response. Taken together, these results suggest that peripheral blood biomarkers following NAC may be useful in predicting long-term outcome. Future work will explore the utility of peripheral blood biomarkers in predicting immunotherapy response.

## Materials & Methods

*Patients.* For all cohorts, data use was approved by the relevant ethics committee, institutional review board, and national competent authority and adheres to the ethical principles of the Declaration of Helsinki. For the VICC-1 cohort, clinical and pathologic data were retrieved from medical records under an institutionally approved protocol (VICC IRB 030747). For the VICC-SD cohort, clinical and pathologic data were retrieved from the deidentified synthetic derivative medical record under and institutionally approved protocol (VICC IRB 202207). First, all female patients with International Classification of Diseases, 9th/10th Revision, Clinical Modification (ICD-9/10-CM) billing codes for malignant neoplasm of breast were identified. Next, from this cohort, patients with at least one breast cancer surgical procedure using Current Procedural Terminology (CPT) codes and exposure to cyclophosphamide and doxorubicin within 12 months preceding the surgery date were automatically selected. The ICD and CPT codes used for this purpose are listed in the Supplementary Tables 2 and 3, respectively. Finally, the electronic health records of the selected patients were manually reviewed to extract the VICC-SD cohort used in this study. To guide the manual chart review process, additional clinical information was automatically extracted from the SD. This includes breast cancer surgery dates, dates when patients received cyclophosphamide and doxorubicin, laboratory measurements, other chemotherapy or cytokine-related medications and date of administration, pathology reports and operative notes. The GeparNuevo cohort consisted of patients from

the placebo arm of the GeparNuevo trial (registration number NCT02685059)<sup>7</sup>. The Instituto Valenciano de Oncología cohort consisted of HR+HER2- patients treated with neoadjuvant chemotherapy. All data use was approved by the relevant ethics committees.

*RNA Sequencing.* RNA was extracted from 0.5-2mL of whole blood or processed PBMCs using the Promega Maxwell RSC simplyRNA Blood kit (AS1380, Promega). Total RNA quality was assessed using the 2200 TapeStation (Agilent). Library preparation was done with a ribo-depletion total RNA library preparation kit and 150 bp paired-end sequencing was performed on the Illumina NovaSeq 6000 targeting an average of 10M reads per sample. Quality control was evaluated at different levels, including RNA quality, raw read data, alignment, and gene expression. Raw RNA-seq paired-ends were mapped to the human reference genome hg19 using STAR 2.7.3. Raw reads count matrix was calculated by featureCounts and used for downstream analysis. 25 genes were removed from the RNA sequencing analysis and transcripts-per-million (TPM) calculation due to over representation. These genes represent red blood cell contamination of the PBMC. The removed genes are: "RN7SL1", "RN7SL2", "HBA1", "HBA2", "HBB", "HBQ1", "HBZ", "HBD", "HBG2", "HBE1", "HBG1", "HBM", "MIR3648-1", "MIR3648-2", "AC104389.6", "AC010507.1", "SLC25A37", "SLC4A1", "NRGN", "SNCA", "BNIP3L", "EPB42", "ALAS2", "BPGM", "OSBP2". DESeq2 was used to identify differentially expressed genes and apeglm was used for log fold change reduction<sup>3,8</sup>. GSEA was used to identify pathways using the Molecular Signatures Database hallmark gene sets<sup>4,5</sup>. CIBERSORTx was used in relative mode with 500 permutations and the LM22 reference matrix<sup>6</sup>. Simplified cell type categories were collapsed as follows: CD4 T cells = Memory activated CD4 T cells + memory resting CD4 T cells + naïve CD4 T cells + T regulatory cells. B cells = Naïve B cells + Memory B cells + plasma cells. NK cells = activated NK cells + resting NK cells. Other Myeloid = M0 macrophages + M1 macrophages + M2 macrophages + activated dendritic cells + resting dendritic cells. Notably, these populations were very low as these are all cell types not commonly seen in the peripheral blood. Granulocytes = activated mast cells + resting mast cells + eosinophils+ neutrophils.

*NanoString nCounter Analysis.* Gene expression was assessed on the GeparNuevo cohort using a custom NanoString Elements panel to measure PIRS genes according to the manufacturer's standard protocol. Briefly, RNA was extracted from processed PBMC pellets using the Promega Maxwell RSC simplyRNA Blood kit and 50 ng of total RNA was used for input into nCounter hybridizations. Data were normalized according to positive and negative spike-in controls, then endogenous housekeeper controls, and transcript counts were log transformed for downstream analyses.

*Statistical analysis.* All statistical analyses were performed in R. Single-cell statistical analyses were calculated in R using the Seurat package<sup>9,10</sup>. Shared Nearest Neighbors were calculated using the Harmony reduction, and clusters were identified at a resolution of 0.3. UMAP was performed for visualization, and missing values were imputed using ALRA<sup>11</sup>. Cell types were assigned to individual cells using SingleR<sup>12</sup>. BlueprintEncodeData was used as a reference<sup>13,14</sup>. Some heatmaps were generated using the R package Complex Heatmap<sup>15</sup>. P-value cut-offs displayed on plots correspond to "ns" equals

p>0.05, \* equals 0.01<p< 0.05, \*\* equals 0.001<p<0.01, \*\*\* equals 0.0001<p<0.001, \*\*\*\* equals p<0.0001. Code used to generate figures is available on request.

## Declarations

### Data Availability:

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Acknowledgements:

The authors would like to thank Andrew Consroe for assistance with data extraction. Funding for this work was provided by Susan G. Komen Career Catalyst Grant CCR14299052 (to J.M. Balko), NIH/NCI R00CA181491 (to J.M. Balko), NIH/NCI SPORE 2P50CA098131–17 (to J.M. Balko), Department of Defense Era of Hope Award BC170037 (to J.M. Balko), and the Vanderbilt-Ingram Cancer Center Support Grant P30 CA68485. Additional funding was provided by NIH T32GM007347 (to M.L. Axelrod), F30CA236157 (to M.L. Axelrod), T32CA217834-04 (to S. Nunnery), and by Celgene (to B. Seliger). Fig. 1a was created with Biorender.com.

### Author Contributions:

M.L. Axelrod: Conceptualization, data curation, formal analysis, investigation, visualization, writing-original draft. Y. Wang: data curation, formal analysis, investigation, writing-review and editing. Y. Xu: data curation, formal analysis, investigation, writing-review and editing. X. Sun: data curation, formal analysis, investigation, writing-review and editing. C. A. Bejan: data curation, formal analysis, investigation, writing-review and editing. P.I. Gonzalez-Ericsson: Formal analysis, writing-review and editing. S. Nunnery: Data curation. R.E. Bergman: Data curation. J. Donaldson: Data curation, writing-review and editing. A. L. Guerrero-Zotano: Data curation, writing-review and editing. C. Massa: data curation, formal analysis, investigation, writing-review and editing. B. Seliger: Resources, data curation, formal analysis, investigation, writing-review and editing. M. Sanders: Formal analysis, writing-review and editing. I.A. Mayer: Resources, data curation, writing-review and editing. J.M. Balko: Conceptualization, formal analysis, supervision, funding acquisition, visualization, writing-original draft.

### Competing Interests:

M.L. Axelrod is listed as a coinventor on a provisional patent application on methods to predict therapeutic outcome using blood-based gene expression patterns, that is owned by Vanderbilt University Medical Center, and is currently unlicensed. B. Seliger receives research funding from Celgene. I.A. Mayer receives research support from Pfizer, Genentech, Puma, and Polyphor, and advisory board consultation for Pfizer, Genentech, Puma, Polyphor, Novartis, Lilly, Abbvie, Immunomedics, MacroGenics, SeaGen, AstraZeneca, Puma, GSK, Cyclacell, and Blueprint. J.M. Balko receives research support from Genentech/Roche, Bristol Myers Squibb, and Incyte Corporation, has received consulting/expert witness

fees from Novartis, and is listed as a coinventor on a provisional patent application on methods to predict therapeutic outcome using blood-based gene expression patterns, that is owned by Vanderbilt University Medical Center and is currently unlicensed. No potential conflicts of interest were disclosed by the other authors.

## References

1. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol*. 2008;26(8):1275-1281. doi:10.1200/JCO.2007.14.4147
2. Axelrod ML, Nixon MJ, Gonzalez-Ericsson PI, et al. Changes in Peripheral and Local Tumor Immunity after Neoadjuvant Chemotherapy Reshape Clinical Outcomes in Patients with Breast Cancer. *Clin Cancer Res*. 2020;26(21):5668-5681. doi:10.1158/1078-0432.ccr-19-3685
3. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106. doi:10.1186/gb-2010-11-10-r106
4. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov M, Sergushichev A. Fast gene set enrichment analysis. *bioRxiv*. February 2016:060012. doi:10.1101/060012
5. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
6. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782. doi:10.1038/s41587-019-0114-2
7. Massa C, Karn T, Denkert C, et al. Differential effect on different immune subsets of neoadjuvant chemotherapy in patients with TNBC. *J Immunother Cancer*. 2020;8(2):1261. doi:10.1136/jitc-2020-001261
8. Zhu A, Ibrahim JG, Love MI. Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*. 2019;35(12):2084-2092. doi:10.1093/bioinformatics/bty895
9. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420. doi:10.1038/nbt.4096
10. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*. 2015;33(5):495-502. doi:10.1038/nbt.3192
11. Linderman GC, Zhao J, Kluger Y. Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv*. 2018;(C):397588. doi:10.1101/397588
12. D A, AP L, L L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019;20(2):163-172. doi:10.1038/S41590-018-0276-Y



13. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nat* 2012 4897414. 2012;489(7414):57-74. doi:10.1038/nature11247
14. Martens JHA, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica*. 2013;98(10):1487. doi:10.3324/HAEMATOL.2013.094243
15. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849. doi:10.1093/bioinformatics/btw313

## Table

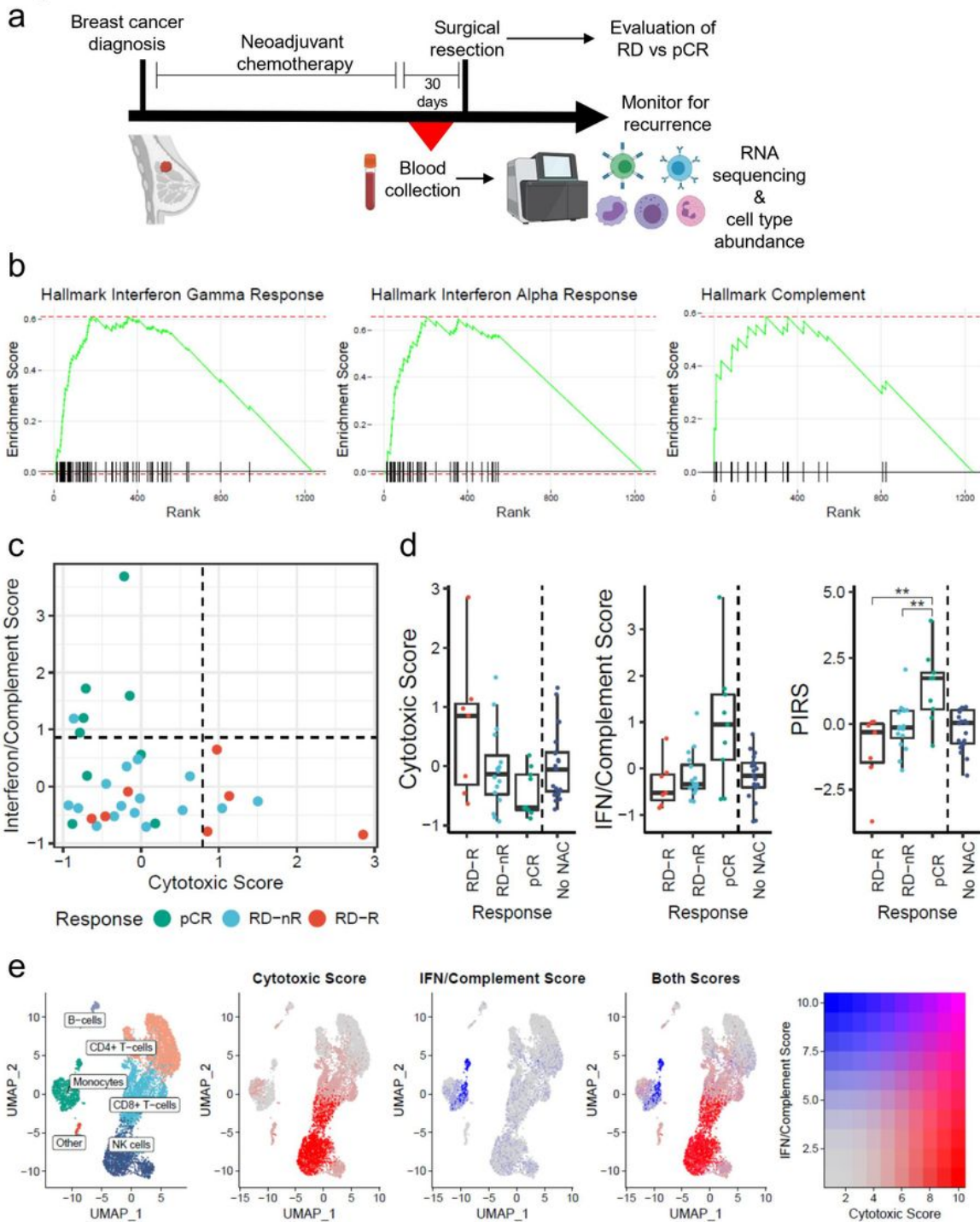
Table 1  
Cohort Descriptions

<b>VICC-1</b>	
<i>Response</i>	
No NAC	21
pCR	9
RD-nR	16
RD-R	7
<i>TNBC</i>	
Yes	17
No	36
<b>GeparNuevo</b>	
<i>Response</i>	
RD	18
pCR	23
<i>TNBC</i>	
Yes	41
No	0
<b>Instituto Valenciano</b>	
<i>Metastasis</i>	
Yes	5
No	9
<i>Hormone Receptor +</i>	
Yes	14
No	0
<b>VICC-SD</b>	
<i>Response</i>	
RD	75
pCR	35
<i>TNBC</i>	

VICC-1	
Yes	50
No	60

# Figures

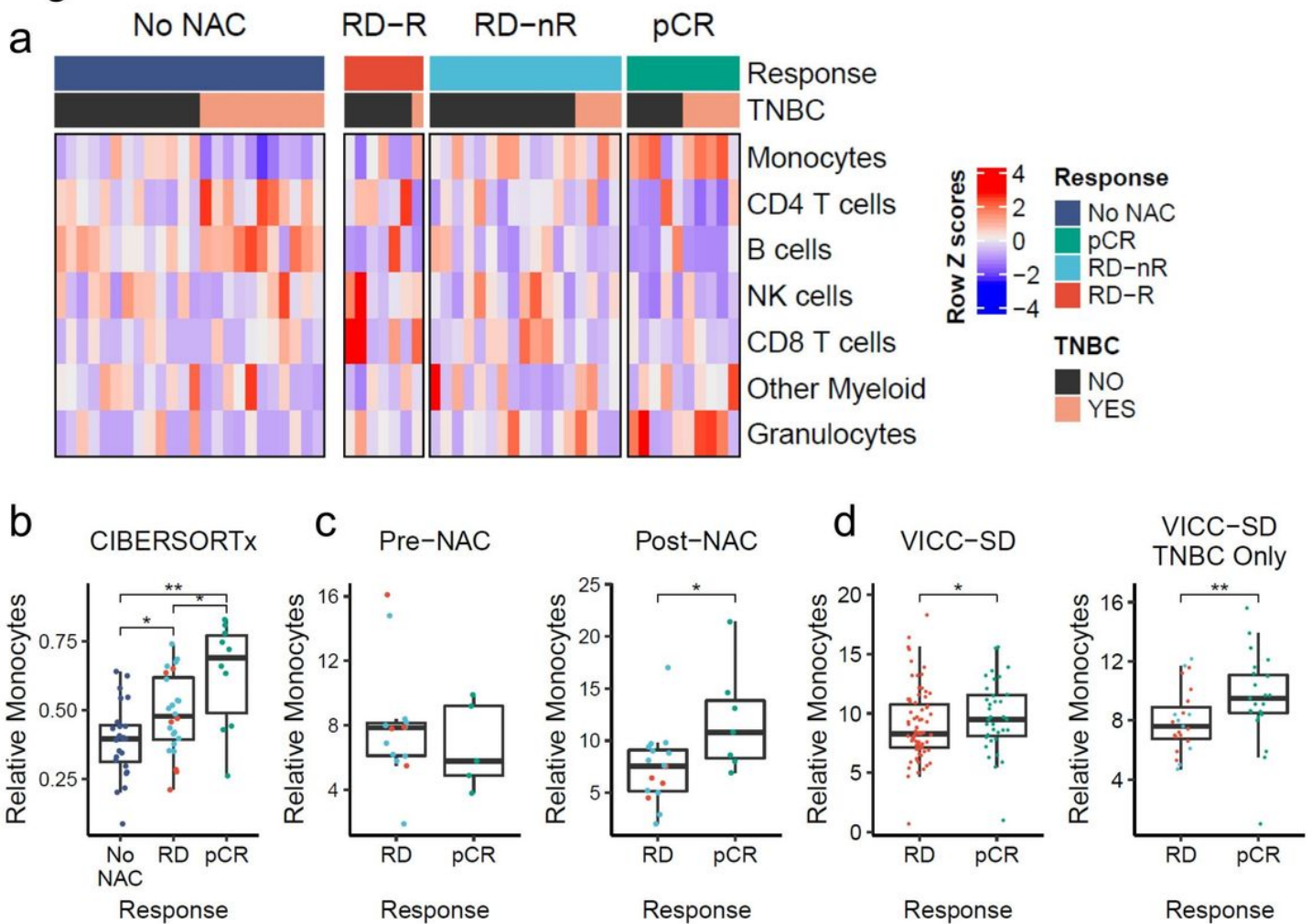
Fig. 1



**Figure 1**

Expression of immune related genes in the peripheral blood is associated with good outcome following NAC. A) Schematic describing blood collection timing and downstream analyses. B) GSEA plots for gene sets enriched in blood of patients with pCR relative to RD. C) Expression of IFN/Complement and cytotoxicity scores is shown for each sample. D) Cytotoxic, IFN/complement, and combined peripheral immunologic response scores (PIRS; IFN/Complement score minus cytotoxic score) are shown for each sample, stratified by outcome. Box plots show the median, first and third quartiles. P-values represent FDR corrected Wilcox tests. E) Single cell sequencing data shows cell type specific expression of each score. Cytotoxic score is shown in red. IFN/Complement score is shown in blue. Co-expression would be shown in pink.

**Fig. 2**



**Figure 2**

Monocytes are most abundant in blood of patients with good outcomes following NAC. A) Heatmap showing row standardized (z-score) of CIBERSORTx inferred cell type abundance. B) CIBERSORTx

monocyte values, stratified by outcome. C) Clinical lab values for relative monocytes pre- and post- NAC. D) Relative monocytes in full and TNBC only synthetic derivative cohort. Box plots show the median, first and third quartiles. P-values represent one-tailed Wilcoxon tests (FDR corrected where applicable).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.csv](#)
- [SupplementaryTables23.docx](#)
- [FigS1.jpg](#)
- [FigS2.jpg](#)