

# Data-driven analysis of immune infiltrate in a large cohort of breast cancer and its association with disease progression, ER activity, and genomic complexity

Ruth Dannenfelser<sup>1,2</sup>, Marianne Nome<sup>3,4</sup>, Andliena Tahiri<sup>3,4</sup>, Josie Ursini-Siegel<sup>5</sup>, Hans Kristian Moen Vollan<sup>4,6,7</sup>, Vilde D. Haakensen<sup>7</sup>, Åslaug Helland<sup>8</sup>, Bjørn Naume<sup>8</sup>, Carlos Caldas<sup>6</sup>, Anne-Lise Børresen-Dale<sup>4,7</sup>, Vessela N. Kristensen<sup>3,4,7</sup> and Olga G. Troyanskaya<sup>1,2,9</sup>

<sup>1</sup>Department of Computer Science, Princeton University, Princeton, New Jersey, United States of America

<sup>2</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, United States of America

<sup>3</sup>Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

<sup>4</sup>Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, Ahus, Norway

<sup>5</sup>Lady Davis Institute for Medical Research, McGill University, Montreal, Quebec, Canada

<sup>6</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

<sup>7</sup>Department of Genetics, Institute for Cancer Research, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway

<sup>8</sup>Department of Oncology, Division for Surgery, Cancer, and Transplantation, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway

<sup>9</sup>Flatiron Institute, Simons Foundation, New York, New York, United States of America

**Correspondence to:** Olga G. Troyanskaya, **email:** ogt@genomics.princeton.edu  
Vessela N. Kristensen, **email:** v.n.kristensen@medisin.uio.no

**Keywords:** *breast cancer, immune profiling, lymphocyte infiltration, normal breast tissue, immune infiltration*

**Received:** March 27, 2017

**Accepted:** June 17, 2017

**Published:** July 07, 2017

Copyright: Dannenfelser et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

The tumor microenvironment is now widely recognized for its role in tumor progression, treatment response, and clinical outcome. The intratumoral immunological landscape, in particular, has been shown to exert both pro-tumorigenic and anti-tumorigenic effects. Identifying immunologically active or silent tumors may be an important indication for administration of therapy, and detecting early infiltration patterns may uncover factors that contribute to early risk. Thus far, direct detailed studies of the cell composition of tumor infiltration have been limited; with some studies giving approximate quantifications using immunohistochemistry and other small studies obtaining detailed measurements by isolating cells from excised tumors and sorting them using flow cytometry. Herein we utilize a machine learning based approach to identify lymphocyte markers with which we can quantify the presence of B cells, cytotoxic T-lymphocytes, T-helper 1, and T-helper 2 cells in any gene expression data set and apply it to studies of breast tissue. By leveraging over 2,100 samples from existing large scale studies, we are able to find an inherent cell heterogeneity in clinically characterized immune infiltrates, a strong link between estrogen receptor activity and infiltration in normal and tumor tissues, changes with genomic complexity, and identify characteristic differences in lymphocyte expression among molecular groupings. With our extendable methodology for capturing cell type specific signal we systematically studied immune infiltration in breast cancer, finding an inverse correlation between beneficial lymphocyte infiltration and estrogen receptor activity in normal breast tissue and reduced infiltration in estrogen receptor negative tumors with high genomic complexity.

## INTRODUCTION

Cancer cells develop in the extracellular matrix (ECM) surrounded by a variety of non-malignant cells, such as fibroblasts, vascular cells, leukocytes, and bioactive substances such as chemokines and cytokines. Together these cells and substances of the host form an environment conducive to carcinogenesis [1–3]. Leukocytes, once thought to be purely beneficial, are now recognized for their functions in tumor promotion as well as inhibition. Many cells of the myeloid lineage contribute to tumor proliferation and cancer development across tumor types. More specifically, tumor associated M2 macrophages, neutrophils, mast cells, immature myeloid cells, and monocytes have been widely shown to support cancer progression through the secretion of growth factors, cytokines, and proteases that promote remodeling of the ECM [4], genomic instability [5–6], angiogenesis [7–11], and suppression of beneficial immune responses [12]. Lymphocytes, while generally having a positive effect, exert pro-tumor or anti-tumor functions in a tissue and cancer specific manner [13]. This effect is largely due to their plasticity [14–15].

There is substantial variability in the number and types of infiltrating lymphocytes in breast tumors across individuals. Recent efforts to profile this landscape have found both of these properties to be indicative of outcome and response to chemotherapy [16–20]. Furthermore, there has been specific interest in finding the component of tumor infiltrating lymphocytes (TILs) that are specifically recruited in the attempt to control tumor growth. These will be of particular importance in the era of immune checkpoint inhibitors such as PD1.

Thus far the established critical factor of the lymphocyte anti-tumor response is the presence of CD8<sup>+</sup> cytotoxic T-cells (CTL), which contribute directly to apoptosis through the secretion of cytotoxins. Numerous breast cancer studies have documented their association with good prognosis and long-term survival [12, 18, 21–23]. T-helper 1 (Th1) cells are also widely thought to contribute to tumor clearance, through the production of interferon gamma (IFNG) which helps curb proliferation, slow angiogenesis, enhance M1 macrophages' tumoricidal ability, and aid in CTL efficacy through the expression of the major histocompatibility complex class I [12, 24]. The exact role of T-helper 2 (Th2) cells in breast cancer is less clear, but generally they are thought to be pro-tumorigenic. In mouse models, Th2 driven expression of IL13 and IL4 contribute to tumor progression and metastasis [12, 25–27]. Additionally, IL4 increases leukocyte recruitment and the promotion of epidermal growth factor by M2 macrophages [28]. However, IL13 has been shown to reduce breast cancer recurrence [25]. The function of B cell infiltration is poorly understood, with some studies highlighting the importance of B cells in good outcome [29–30], some in poor outcome [12], and others deeming B cell infiltration irrelevant [28].

Along with the recognition of the critical role played by the immune system in oncogenesis, tumor progression, and response to therapy, increasing attention has been drawn by the potential prognostic and predictive role of the immune infiltrate in this setting.

Traditional studies profile lymphocyte infiltration utilizing immunohistochemical staining or, less commonly, single cell isolation followed by flow cytometry to assess the composition and abundance of lymphocytes within a tumor sample. Staining based approaches are semi-quantitative at best, while flow based approaches are labor intensive and difficult to perform on a large scale. Additionally, both methods require access to and destruction of a portion of the tumor sample. Herein we propose an alternative *in silico* approach to examine breast tumor infiltration by four major lymphocyte cell types: B cells, CTL, Th1, and Th2 cells. We make use of the fact that gene expression assays of tumor samples capture signal that is representative of tumor cells as well as their microenvironment. Our method is robust, correlates well with experimental measures of immune infiltration, and allows us to retroactively profile lymphocyte abundance in existing breast tumor expression data without a priori knowledge of sample composition. With this approach we show changes in infiltrate levels in estrogen receptor negative normal breast tissue, reproducible patterns of immune activity across molecular groupings, and examine the impact these levels have on patient survival and treatment response. Through this analysis we establish critical links between the molecular features of breast tissue and immune infiltration and demonstrate how our marker set approach can be effectively applied towards this end.

## RESULTS

### *In silico* immune profiling derived from a compendium of 2,171 human samples of cultured and primary human tissues and treatments

We used an *in silico* method to find genes that are preferentially expressed in four lymphocyte cell types: B cells, CD8<sup>+</sup> T cells (CTL), CD4<sup>+</sup> T helper 1 (Th1), and CD4<sup>+</sup> T helper 2 (Th2) cells (Figure 1). For each cell type we expanded a set of literature mined marker genes into a robust set of cell type specific genes using a large compendium of human blood expression data and the nanodissection method (<http://nano.princeton.edu>) we developed previously for computationally predicting cell-type specific genes [31]. Nanodissection uses an iterative machine learning framework with diverse expression data collections, to identify expression patterns specific to marker genes for a given cell type. We validated this approach previously in systematic computational and experimental evaluations [31]. When applied to the four lymphocyte cell types, (see Materials and Methods) the resulting set of markers enable us to robustly identify

immune specific signal (Supplementary Figure 2, Supplementary Figure 3).

The overall performance of nanodissection, for each lymphocyte cell type, was assessed based on classification performance and the relevance of the resulting marker sets. In five-fold cross validation we achieve AUCs above 0.7 for each nanodissection run: 0.8896, 0.822, 0.732, and 0.722 for B cells, CTL, Th1, and Th2 cells respectively (Supplementary Figure 1). We confirmed the biological relevance of our resulting marker sets by analyzing the functional processes they participate in and the correspondence of our predictions with experimental measures of lymphocyte infiltration. To identify overrepresented pathways, we calculated statistical enrichment for each of the resulting marker gene sets. The top significantly enriched terms are predominantly immune related and are representative of the action of each lymphocyte subset (Figure 2A, Supplementary Table 3), with B cell activation enriched for the B cell marker, cell defense terms such as cytosis for CTLs, and many T cell activation terms enriched for the T-helper cells.

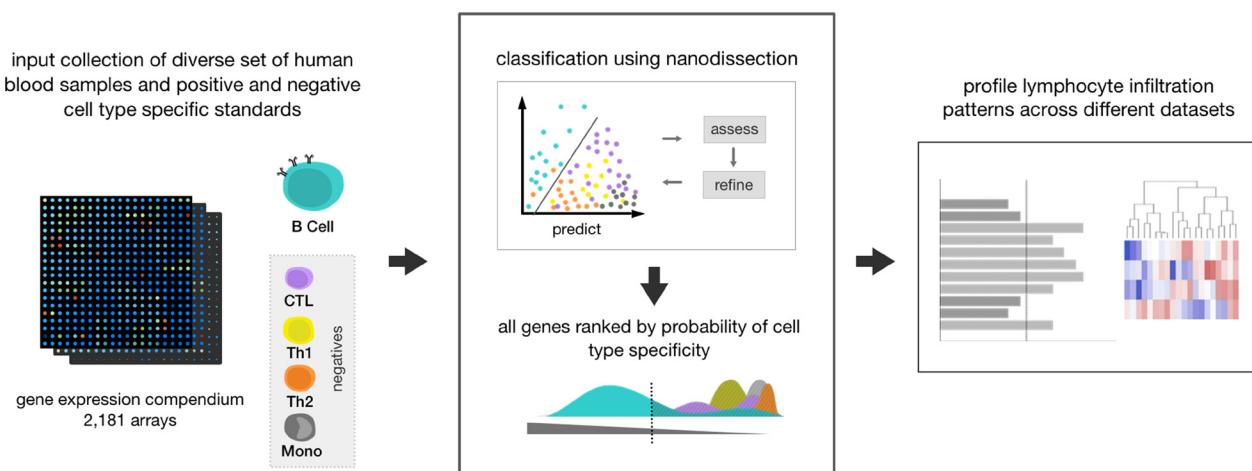
Furthermore, we examined the relationship between the expression level of the immune markers and experimentally derived immune infiltration categories (absent, mild, moderate, severe) in the METABRIC breast cancer cohort [38]. We calculated a B cell, CTL, Th1, and Th2 infiltration score for each sample by averaging the expression of all genes in each marker. With these scores we observe that B cell (*ANOVA*,  $p = 3.7e^{-4}$ ), CTL ( $p < 2.2e^{-16}$ ), Th1 ( $p < 2.2e^{-16}$ ), and Th2 ( $p < 2.2e^{-16}$ ) expression changed significantly across the infiltration categorizations with increases in B cell, CTL, Th1, and Th2

marker expression corresponding with increasing immune infiltration (*Kendall's Tau*, 0.294, 0.290, 0.242, 0.155, Figure 2B). To measure the specificity of our signal we also looked at the enrichment of our marker genes across 185 samples of cell type specific immune expression data, finding that our gene sets are enriched in the corresponding target and related cell types, improving on the specificity of using known standards and T cell surface markers alone (Figure 2C, Supplementary Figure 3).

We also examine the robustness of our signatures to expression data where a subset of our marker genes may not be measured due to platform differences or experimental errors. To test this we randomly removed subsets of genes from each of our markers and recalculated the relationship between the experimentally defined infiltration categories in METABRIC and our predicted B cell, CTL, Th1, and Th2 infiltration scores using Kendall's correlation. With all our signatures we were again significantly able to preserve the infiltration trends up to 50% "lost" genes. Robustness varies only slightly between the cell types, with only the Th2 signature affected with performance degradation around 45% of genes removed (Supplementary Figure 2).

### Lymphocyte infiltration is strongly tied to estrogen receptor status

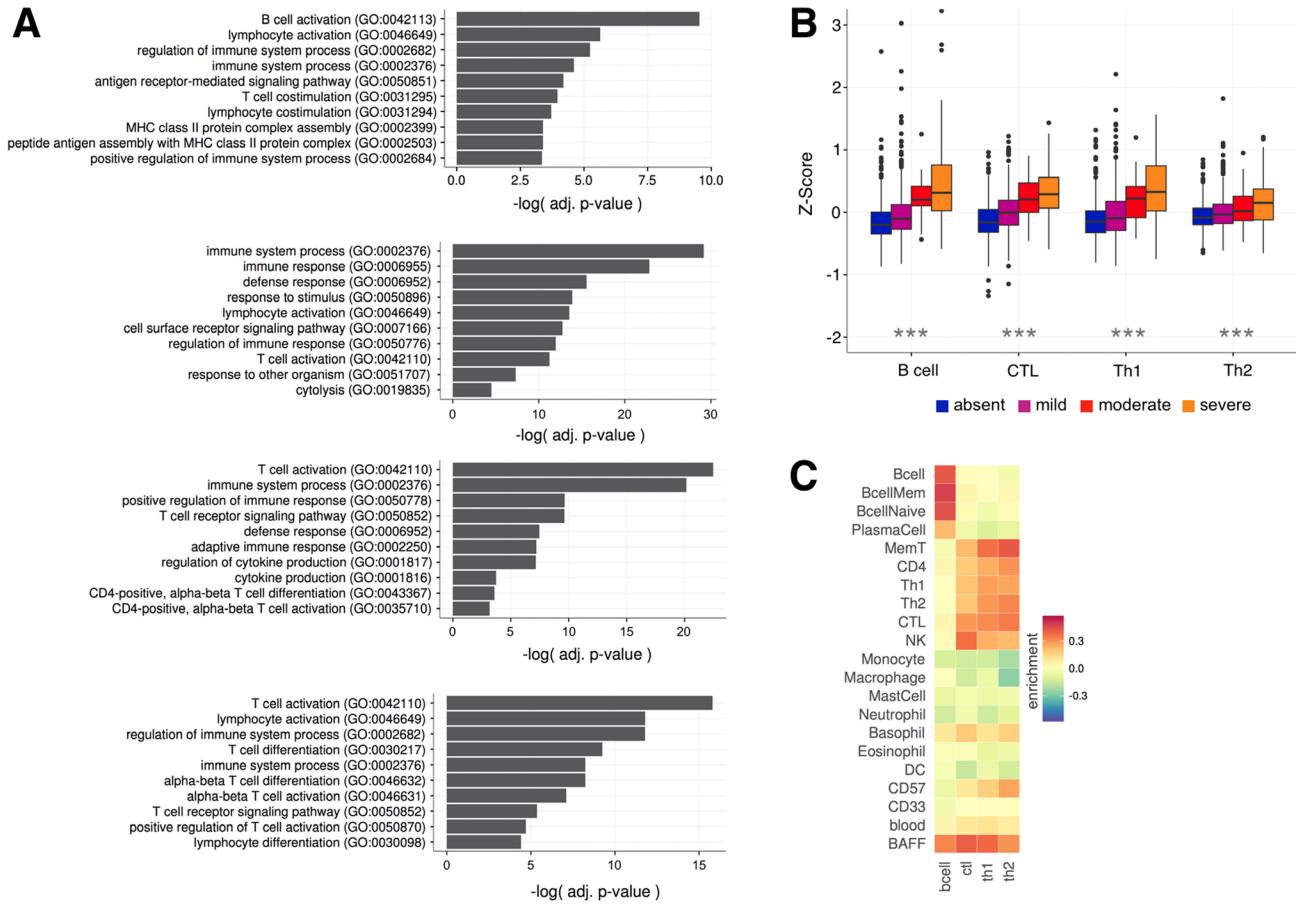
At the most basic level breast tumors are divided based on their ability to respond to estrogen. Thus, we examined whether changes in estrogen receptor (ER) status have any effect on lymphocyte infiltration. In this analysis we used two datasets, MicMa ( $N = 108$ ) [39]



**Figure 1: Method overview.** This diagram shows the process of running our method on one cell type, B cells. First all input data, consisting of over 2,000 human blood microarray samples taken in different conditions and different degrees of resolution, and positive and negative training standards, are assembled. In this example, B cell genes from Gene2MeSH are used as a positive standard and all other lymphocyte cell types (CTL, Th1, Th2, monocytes) in Gene2MeSH are used as negatives. Using the input, nanodissection finds a path that separates the positive training class from the negative ones in the space of the compendium, iteratively improving classification by selecting the best set of standards. Using the classification boundary, all genes assayed are then ranked by their probability to be cell type specific, in this case B cell specific. We then use the top 100 most probable cell type specific genes to estimate lymphocyte infiltration levels in breast cancer datasets.

and the METABRIC ( $N = 1417$ ) cohort [38]. These cohorts together allow us to assess a reasonable number of ER- samples:  $N = 42$  and  $N = 340$  respectively. After partitioning the samples by ER status we observed a significant increase in the expression levels of all lymphocytes in ER- tumors relative to their ER+ counterparts in METABRIC (*Wilcoxon rank sum test*, Bcell, CTL, Th1:  $p < 1.00e^{-7}$  and Th2:  $p = 0.023$ ), and B cell in the much smaller MicMa dataset (B cell:  $p = 0.02$ , CTL:  $p = 0.404$ , Th1:  $p = 0.833$ , Th2: 0.057, Figure 3A, 3B). To further characterize this relationship we devised an ER score using the ER marker genes of van't Veer et al. [37] to capture the activity level of the estrogen receptor. For each cohort we observe a significant negative correlation between ER activity and the level of B cell (*Pearson*,  $R = -0.459$ ,  $p < 2.2e^{-16}$ ;  $R = -0.406$ ,  $p = 1.34e^{-5}$ ), CTL ( $R = -0.292$ ,  $p < 2.2e^{-16}$ ;  $R = -0.343$ ,  $p = 2.83e^{-4}$ ), and Th1 ( $R = -0.401$ ,  $p < 2.2e^{-16}$ ;

$R = -0.247$ ,  $p = 0.023$ ) infiltration as well as the total level of all four lymphocytes ( $R = -0.406$ ,  $p < 2.2e^{-16}$ ;  $R = -0.317$ ,  $p = 5.87e^{-8}$ , Figure 3A, 3B) further supporting our claim. We then extended this analysis to normal tissue, profiling another dataset containing reduction mammoplasties ( $N = 18$ ), adjacent normal tissue from ER+ and ER- patients ( $N = 18$ ), and prophylactic mastectomies ( $N = 6$ ) [40], and found a gradient in the levels of immune infiltration, with a few samples marked by high lymphocyte infiltration. Interestingly, the level of infiltration in these samples can also be explained by estrogen receptor status, as opposed to the sample source. Again using an ER activity score as a proxy for ER status we observed a negative correlation between ER activity and total infiltration (*Pearson*,  $R = -0.541$ ,  $p = 2.29e^{-4}$ , Figure 3C), suggesting an early link between immune infiltration and ER status that is also preserved in benign samples (Supplementary Figure 4).



**Figure 2: Expression of lymphocyte marker genes correspond with experimentally defined measures of immune infiltration.** (A) GO Biological Process enrichment for each of the lymphocyte marker gene sets. Shown here are some of the relevant significant terms (a full list can be found in Supplementary Table 3). Each marker is significantly enriched for immune related terms that are representative of the role of their corresponding cell type. (B) Four categories of immune infiltration (absent, mild, moderate, severe) were defined by H&E staining for the METABRIC cohort [38]. The box plots show that B cell, CTL, Th1, and Th2 marker co-expression correspond with increasing levels of infiltration. All labels were significantly different between the infiltration categories calculated using ANOVA (\*\* $p < 1.0e^{-5}$ ). (C) Average enrichment of marker genes across purified immune cell types from Chantan et al [34], GSE3982, and GSE1133.

## Lymphocyte infiltration is predictive of survival in ER- tumors

With our markers we can examine both a general and detailed picture of the effects of lymphocyte infiltration on survival. We grouped patients into low (bottom three quartiles) and high (upper quartile) infiltration classes for the overall lymphocyte expression score and separately for each of the lymphocyte cell types. We then partitioned samples from the METABRIC cohort based on ER status and examined the effects of expression of our lymphocyte markers on patient survival times using univariate Kaplan Meier analysis [41]. Comparing the low and high classes, we see that the total expression of lymphocytes is an important prognostic factor only in ER- tumors (Figure 4). At higher resolution, we find that ER- samples with an improved outcome tend to have higher levels of lymphocyte infiltration, with the levels of CTL and Th1 infiltration playing a significant role (*log rank*,  $p = 0.067$ ,  $p = 0.05$ ,  $p = 0.032$ ,  $p = 0.132$ ). This result is consistent with several previous studies [18, 21, 23, 42] that have observed an improved treatment response and better prognosis in ER- tumors with increased levels of CTL infiltration.

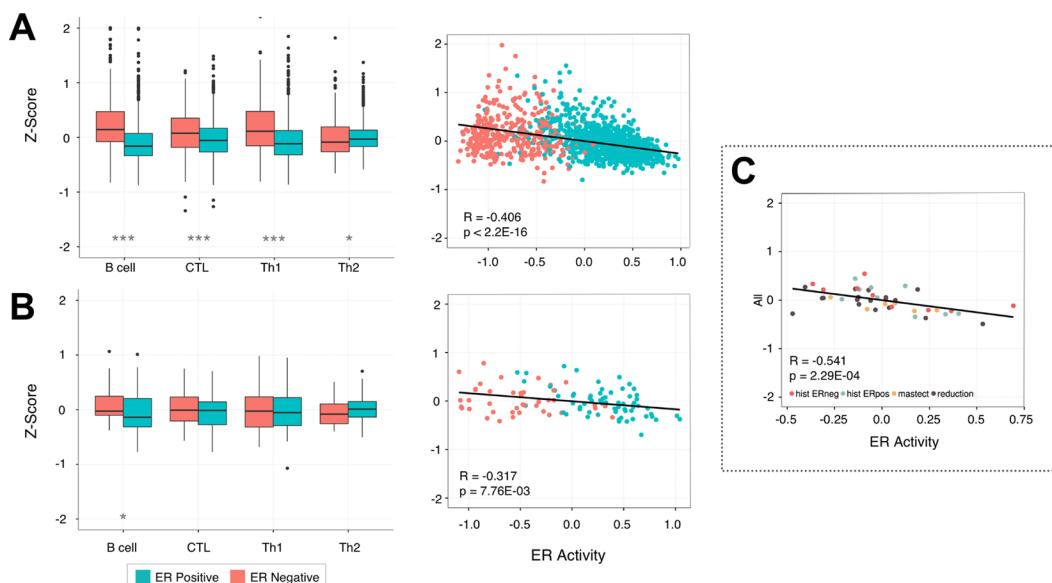
## Unique immune infiltration patterns across breast cancer subtypes

Different breast cancer subtypes are commonly referenced as disparate diseases. As such we expect

them to have very different and reproducible patterns of lymphocyte expression. Examining our B cell, CTL, Th1, and Th2 scores across samples grouped by subtype we saw that very phenomenon in both tumor cohorts: METABRIC and MicMa (Figure 5). Overall lymphocyte infiltration showed the most pronounced difference between the luminal subtypes and the basal, HER2, and normal-like subtypes, recapitulating the patterns observed with ER status. Elevated infiltration in the predominantly ER- subtypes was significantly increased compared to the luminal subtypes for B cell, CTL, and Th1 in METABRIC (*Wilcoxon rank sum test*, B cell, CTL, and Th1:  $p < 2.2e^{-16}$ , Th2:  $p = 0.067$ ) and MicMa (B cell:  $p = 3.56e^{-6}$ , CTL:  $p = 2.02e^{-4}$ , Th1:  $p = 0.013$ , Th2:  $p = 0.635$ ).

## Genomic instability is associated with a reduction in beneficial immune responses

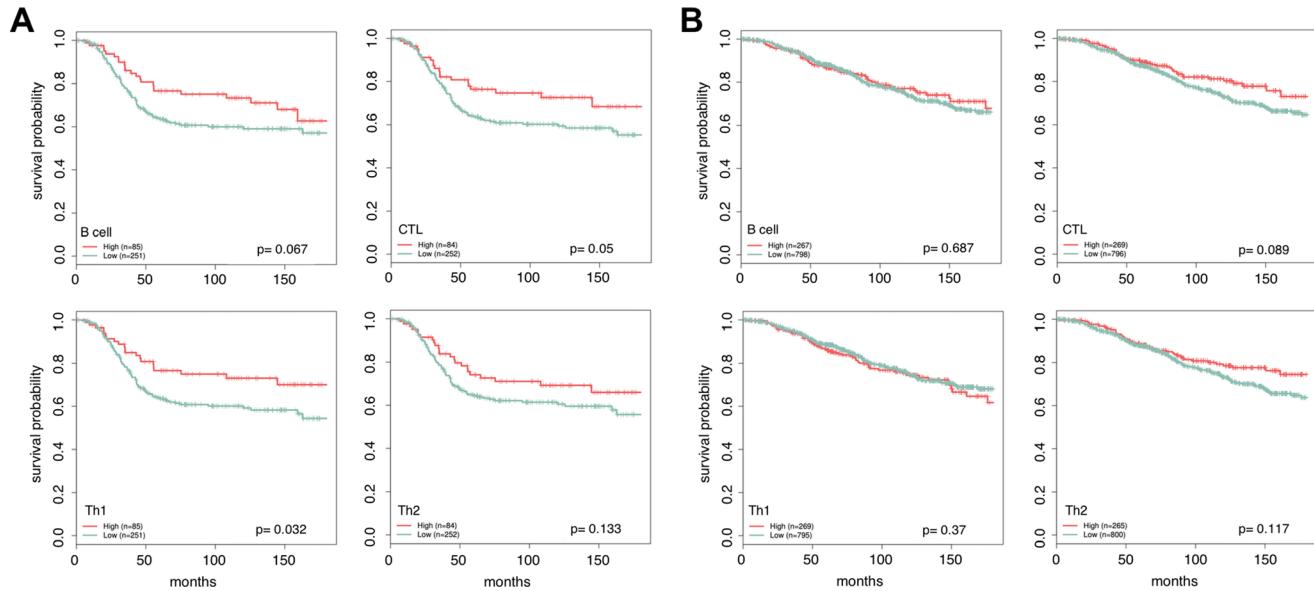
Previously, we have observed that basal-like tumors with wild type copies of TP53 have elevated levels of CTL expression compared to those with one or more mutations of TP53 [43]. Since TP53 is a guardian of genomic stability, we generalized this idea here, examining whether tumors with regions of severe local genomic instability have differing lymphocyte infiltration patterns, using the complex arm-wise aberration index (CAAI) devised by Russnes et al. [44]. CAAI positive samples have many local gains and losses on at least one chromosomal arm and an association with estrogen receptor negativity, poor prognosis, and high tumor grade [44]. We found, when



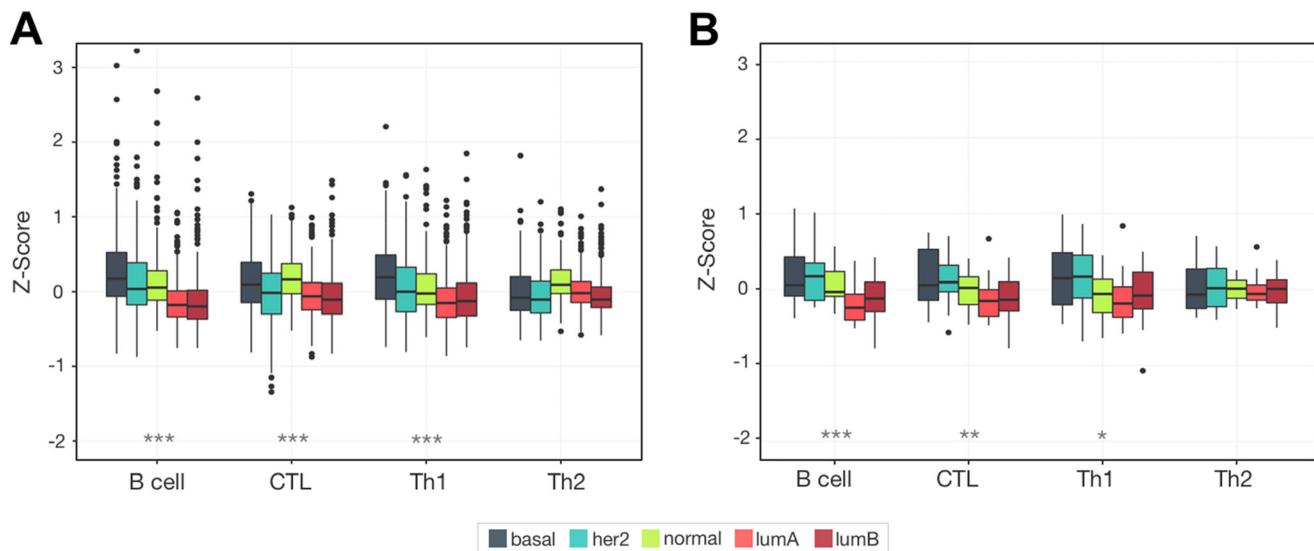
**Figure 3: Lymphocyte infiltration is dependent on the activity of the estrogen receptor.** Box plots show the increased levels of B cell, CTL, Th1, and Th2 infiltration in ER- samples when they are partitioned by estrogen receptor status. The scatterplots show an inverse correlation between the activity of the estrogen receptor and the infiltration of all four lymphocyte subsets in breast tissue. This pattern is reproducible in two separate breast cancer cohorts: METABRIC (A) and MicMa (B), and is also apparent in a set of normal breast tissue samples taken from adjacent tumor tissue, preventative mastectomy patients, and breast reductions (C).  $P$ -values were calculated using the Wilcoxon rank sum test with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.00001$ .

examining the METABRIC cohort, that CAAI positive samples had diminished B cell (*Wilcoxon rank sum test*,  $p = 2.53e^{-5}$ ), CTL ( $p = 3.62e^{-10}$ ), Th1 ( $p = 0.01$ ) and Th2 ( $p = 2.50e^{-11}$ ) infiltration. This effect is markedly more pronounced in ER- tumors but still present in ER+ samples (B cell:  $p = 1.74e^{-6}$ , CTL:  $p = 1.40e^{-5}$ , Th1:  $p = 7.33e^{-5}$ ,

Th2:  $p = 6.58e^{-4}$ ), and the predominantly ER- subtypes (Figure 6, Supplementary Figure 6), basal ( $p = 2.76e^{-5}$ ,  $p = 1.44e^{-4}$ ,  $p = 0.001$ ,  $p = 0.004$ ) and HER2 ( $p = 0.018$ ,  $p = 0.004$ ,  $p = 0.01$ ,  $p = 0.014$ ), and is not immediately evident when the experimentally defined METABRIC immune infiltration classes are used.



**Figure 4: Patients with ER- tumors and high levels of Th1 and CTL infiltration are more likely to have a positive outcome.** For each lymphocyte marker, patients are partitioned into high and low levels using quartiles. Patients with lymphocyte scores in the upper quartile are marked as high (red) and the remaining quartiles are grouped together and labeled low (green). Increased lymphocyte infiltration is associated with better long term survival in ER- tumors (A), and has little effect in ER+ tumors (B). The most extreme difference is seen in patients with high Th1 and CTL infiltration vs. low. Reported  $p$ -values were calculated using the log rank test.



**Figure 5: Unique lymphocyte infiltration patterns are seen for each breast cancer subtype.** Box plots show the distribution of lymphocyte infiltration scores for each subtype. We see a significant reduction in the average level of lymphocyte infiltration in the luminal subtypes (\*\* $p < 1.0e^{-5}$ ), with the highest level of infiltration in basal and normal-like subtypes. This pattern is likely reflective of the strong link between infiltration and ER negativity and is reproducible across two cohorts METABRIC (A) and MicMa (B).

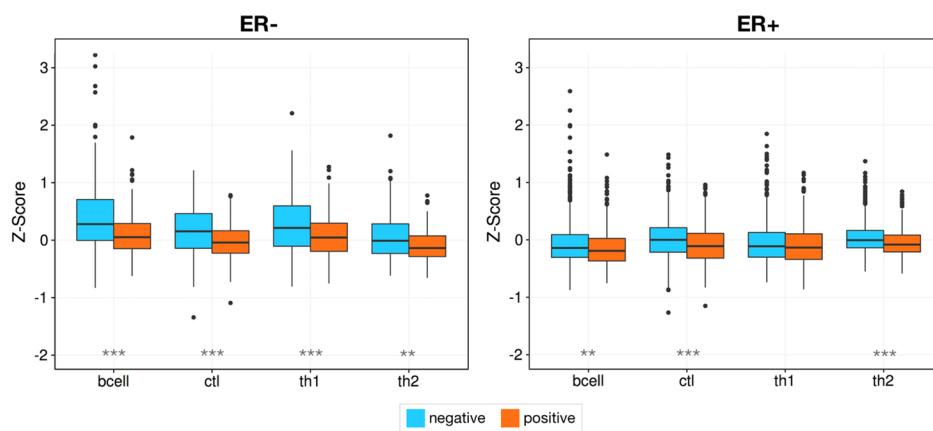
## Aromatase inhibition and increased infiltration

As a preliminary investigation into the relationship between common therapies and lymphocyte infiltration, we examined infiltration levels pre and 90 days into treatment with an aromatase inhibitor. Aromatase inhibitors are a viable treatment option for women with ER+ breast cancer, and we postulated that inducing a more negative ER phenotype would be marked by an increase in lymphocyte infiltration. We tested our hypothesis on a set of previously published data [45] containing gene expression for 58 breast biopsies taken at two time points, pre-treatment and 90 days into treatment with Letrozole. We found that the mean levels of B cell and the T helper subsets increase throughout treatment (Figure 7; *Wilcoxon rank signed test*, B cell:  $p=0.001$ , CTL:  $p = 0.086$ , Th1:  $p = 0.011$ , Th2:  $p = 1.05e^{-4}$ ). This initial insight

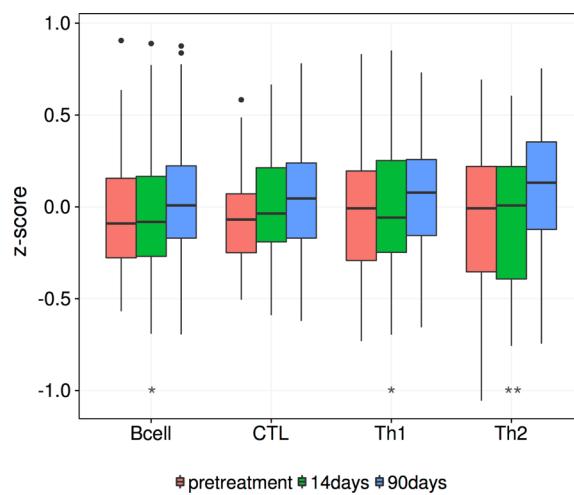
further establishes the link between estrogen as a main determinant of lymphocyte infiltration.

## DISCUSSION

Immune infiltration is now widely recognized as an important component in the development and proliferation of cancer. Experimental methods quantifying the composition and abundance of lymphocyte infiltration within the tumor bed are typically time consuming, costly, and are not easy to perform at a large scale. Additionally, no existing method allows for retroactive absolute profiling while preserving the tumor sample. Given the plethora of existing breast cancer gene expression data, we devised a method that addresses these challenges, without prior knowledge about sample composition, which can be easily used to predict the immune landscape of a



**Figure 6: Genomic instability is associated with a reduction in infiltration.** Higher levels of infiltration are seen in samples with less genomic instability as measured by CAAI status, with a markedly stronger effect in ER- tumors. Asterisks denote Wilcoxon rank sum test  $p$ -values \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.00001$ .



**Figure 7: Lymphocyte infiltration increases over the course of Letrozole treatment.** This box plot shows the increase in lymphocyte infiltration over the course of Letrozole treatment. A Wilcoxon signed rank test was used to assess the significance of these changes between patients at pretreatment and 90 day time points.  $P$ -values are denoted by asterisks, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.00001$ .

sample. We then applied this methodology to improve our understanding of lymphocyte infiltration in breast tissue.

Several other studies [16, 17, 29, 46–48] have also assessed lymphocyte infiltration by profiling gene expression microarrays. The majority of these methods estimate immune infiltration using unsupervised clustering and differential expression analysis of breast cancer cohorts. The remainder perform larger scale meta-analyses expanding immune marker genes using correlation across breast cancer samples. Unlike studies where clustering reveals enriched immune activity, our method does not first require the selection of differential expression patterns in a cohort. Additionally, unlike the correlation based approaches focused on specific studies, we use a very large compendium consisting of multiple human blood samples that are not limited to the breast or cancer, enabling us to robustly identify immune cell markers. This makes our methodology more generalizable to other disease datasets and biological questions, and prioritizes the signal driven by infiltrating lymphocytes, lending more confidence to our results.

Recent deconvolution based approaches [49, 50], such as CIBERSORT [51] have been used with some success profiling the levels of many types of immune cells. However, such approaches require knowledge of all cell types present in a mixture and output relative proportions. Furthermore, CIBERSORT requires purified immune cell expression data to construct their underlying LM22 reference matrix. To assess whether we can achieve similar performance without this potential data limitation we ran CIBERSORT on a small dataset of purified immune cell populations [52] that was not used in the construction of their immune reference matrix. We compared the identification of all overlapping immune cell types profiled in this dataset using our method with CIBERSORT's B cell (memory and naïve B cell) and CD4 (CD4 naïve, CD4 resting memory, and CD4 activated memory T cells) cell types. We found that, both our method and CIBERSORT could correctly label 2/2 of the Bcell samples and 4/4 of the CD4 samples with ours further correctly separating CD4 into the Th1 (2/2) and Th2 (2/2) subsets, making our proposed pipeline the only viable method when purified cell type data are not available.

Any computational method that makes a claim about the mixtures of cells in a sample, whether a differential expression, meta-analysis, or a deconvolution based approach, will suffer decreased performance when the cell types to distinguish are very similar. The lymphocytes we profiled here share many of the same surface markers, which can potentially degrade the quality of computationally derived markers. Additionally, using a large set of the resulting genes for our markers can potentially introduce some weaker markers into the analysis. This, however, is remedied as the large set of genes can robustly capture infiltration signal when one or more genes are missing from the sample profile

(Supplementary Figure 2). Taken together, these *in silico* methods should be seen as exploratory lymphocyte assays that can be readily applied to existing and yet to come gene expression data and not a conclusive replacement for experimental approaches.

One of the most striking findings in our work is how drastically lymphocyte infiltration patterns vary depending on the status of the estrogen receptor. We observed this effect in both tumor and normal tissue, with ER- samples having, on average, higher levels of B cells, CTL, Th1, and Th2 infiltration than ER+ samples. Previous studies have also found strong immune signals in ER- breast tissue [16, 46, 53] that were weak or absent in ER+ tissue, and furthermore observed high CTL infiltration as a positive prognostic factor only in ER- samples [18, 21–23]. Interestingly, no other study has examined this link between ER activity in normal breast tissue. The correlation that we found between ER- and increased infiltration suggests that ER+ breast tissue might be more susceptible to tumor formation, and may be one reason that increased exposure to estrogen increases a woman's overall risk of breast cancer. Several studies [53–55] have linked estrogen to the proliferation of T regulatory (Treg) cells, which suppress B and T lymphocytes. Another recent study by Generali et al. [55] showed a reduction in Treg cells following treatment with the aromatase inhibitor, Letrozole. The inhibition of Treg cells may explain why we observed an increase in lymphocyte expression levels 90 days into treatment with Letrozole, and we hypothesize that Tregs contribute to the decreased infiltration we see in ER+ relative to ER- tumors.

It is still unclear what underlying mechanisms determine the extent of immune infiltration in ER- tumors. Previously, we uncovered one potential component when we observed that mutations in TP53 in ER- tumors are associated with a decrease in the level of CTL infiltration [43]. Here we more generally detect a similar phenomenon with CAAI status in ER- samples, those with low complex genomic instability have higher levels of B cell, CTL, Th1, and Th2 infiltration relative to samples with high instability (Figure 6, Supplementary Figure 6). This echoes the relationship Volland et al. [56] has shown between high CAAI and poorer survival. A rapidly mutating tumor has the potential to adapt and evade the protective mechanisms of the immune system, yet also increase the distress signal and promote immune recruitment to the tumor site. This balance between potentially increasing and actively decreasing immune infiltration might be further influenced by the specific genomic rearrangement pattern the tumor exhibits [57]. CAAI measures a specific type of instability, namely “firestorms” where regions in close proximity are amplified on the same arm. Here we found that a reduction in beneficial immune recruitment occurs particularly with such complex mutations.

Overall, we found that CTL and Th1 infiltration are the main contributors to good outcome, and that the levels

of these lymphocytes tend to increase together along with B cells and to a lesser extent with Th2 cells (Figure 2B, Figure 4). Through univariate survival analysis we observed that lymphocyte infiltration is prognostic only in ER- tumors, with Th1 and CTL infiltration having the most effect. Interestingly, conflicting observations on B cell infiltration and outcome exist, with some suggesting that B cell infiltration is either not functionally significant [14], promotes metastasis [13], and others finding B cells to be an important factor contributing to prolonged survival and long term remission [30, 47–48]. Here we found that elevated levels of B are weakly associated with good outcome in ER- partitions (Figure 4). However, since B cell infiltration correlates strongly with that of the other lymphocytes, including CTL cells, this finding could be confounded with a CTL, or a general infiltration response driven effect. Regardless, increased infiltration of the four lymphocytes we profiled here seem to have positive effects that are not entirely captured when using looking at all lymphocytes in aggregate (Supplementary Figure 5). Reducing immune checkpoints, particularly blocking the mediatory functions of other lymphocytes such as T-regulatory cells, may thus be a fruitful avenue for treatment of breast tumors regardless of estrogen receptor status.

With our methodology for capturing cell type specific signal we were able to leverage existing studies to systematically study immune infiltration without needing physical access to or knowledge of the potential cell composition of the samples. This enabled us to comprehensively examine immune infiltration patterns mentioned in previous studies, reconcile conflicting reports regarding association of B cell infiltration and outcome, and uncover novel biology such as the correlation between CTL and estrogen receptor negativity in normal breast tissue and the decrease in B cell, CTL, Th1, and Th2 infiltration with complex genomic instability. Our method is easily extendable and provides an accessible way to profile the immune landscape in future studies and can be extended to other cancer types.

## MATERIALS AND METHODS

### Generation of cell specific markers

Nanodissection (described formally in [31]) was applied on a large human blood compendium consisting of 2,171 publicly available samples (Supplementary Table 1). Standards of positive and negative example genes (of various “confidence” tiers) were constructed from significant Gene2MeSH (<http://gene2mesh.ncbi.nlm.nih.gov>) annotations for the four lymphocyte cell types and monocyte. For a given cell type, the corresponding Gene2MeSH genes were divided into unique genes and those that overlapped with other cell type gene sets. Unique genes corresponding to the cell type of interest

were used as high-tiered positives, and those that were relevant but overlapping were used as low-tiered positives. All unique genes from the remaining three lymphocyte cell types and monocytes were used as negatives. Samples in the compendium were downloaded from GEO and processed from the raw CEL files, using RMA in “affy” R package [32] with default parameters [33]. The ranked output of each nanodissection run was reduced into a set of high quality marker genes based on the potential markers’ co-expression. To reduce the ranked list into the final signatures we took the top 100 predicted marker genes for each category (Supplementary Table 2). Our findings throughout remain robust to the specific cutoffs used to build these gene sets.

### Clinical data acquisition

Normalized gene expression microarrays for the METABRIC discovery ( $N = 803$ ) and validation ( $N = 614$ ) cohorts were obtained from the European Genome-phenome Archive (accession: EGAS00000000083). Herein we combined both of these partitions and treated them as one cohort. TP53 sequencing and categorization of lymphocyte infiltration by H&E staining is described in Silwal-Pandit et al. [35]. The gene expression data and corresponding annotations for the MicMa ( $N = 108$ ), the set of reduction mammoplasty and prophylactic mastectomies ( $N = 24$ ), and the Letrozole data ( $N = 58$ ), were obtained from Gene Expression Omnibus (GEO) (accessions: GSE3985, GSE20437, GSE20181). The MicMa dataset was used directly and GSE20437 and GSE20181 were processed according to the Affy pipeline described in the above section. After processing all datasets were further z-score transformed across genes using the following formula:

$$z_{gs} = \frac{x_{gs} - \mu_g}{\sigma_g}$$

where  $z_{gs}$  is the z-score for sample  $s$ , gene  $g$ ,  $x_{gs}$  is the normalized intensity for  $g$  in  $s$  and  $\mu_g$  and  $\sigma_g$  are the mean and standard deviation of  $g$  across all samples in the cohort.

### Statistical analysis

A nonparametric Wilcoxon rank sum test, Wilcoxon signed rank test, or two-way ANOVA was used to derive  $p$ -values comparing lymphocyte scores as appropriate. To assess the relationship between the experimentally defined infiltration categories and the lymphocyte infiltration scores, we calculated correlation using Kendall’s rank correlation between the infiltration classes (ranked according to their severity) and the infiltration scores per sample. When assessing robustness, 50 random genes were sampled without replacement 50 times. Gene Ontology enrichment was calculated using PANTHER [58] on a

2017 version of the GO tree. Reported *p*-values have been corrected for multiple hypothesis testing using Bonferroni. Univariate Kaplan-Meier estimates and accompanying log rank tests were calculated using the “survival” R package [36]. Correlations herein were calculated using the Pearson correlation coefficient. All *p*-values less than 0.05 were considered significant.

## Infiltration and ER activity scores

Clinical estimates of ER activity were taken from the publically deposited data. For the MicMa cohort ER activity was determined using monoclonal antibodies against ER, the details of which can be found in [39]. ER classification for METABRIC was calculated computationally using a Gaussian mixture model [38].

Scores measuring both immune infiltration and ER activity were calculated in the same way for a given sample, *s*, by averaging the z-score transformed expression of all *n* genes in the marker.

$$d_s = \frac{1}{n} \sum_{i=1}^n z_{is}$$

Genes in the marker sets not assayed in a given dataset were excluded from averaging. To calculate total lymphocyte infiltration we combined the genes in all four signatures into one before averaging.

Marker genes that capture ER activity were obtained from van’t Veer et al. [37]. Increased activity of the estrogen receptor corresponds with higher ER activity scores. As a rule of thumb, ER activity scores less than zero correspond to ER- samples and scores greater than zero, ER+.

## Comparison with CIBERSORT

GSE3982, obtained and processed as described above without z-score normalization, was uploaded to the CIBERSORT web interface and run with 100 permutations and quantile normalization disabled. Infiltration scores for the nanodissection markers were calculated as described in the previous sections. GSE3982 contained two samples for three cell types, B cell, Th1, Th2. Samples were assigned labels by taking the name of the highest scoring cell type for that sample. CIBERSORT’s multiple B cell types (memory and naïve) were combined into one B cell label, as were the CD4 T cell types (CD4 naïve, CD4 memory resting, CD4 memory activated) for classification of the Th1 and Th2 samples.

## Abbreviations

ECM, extracellular matrix; CTL, CD8+ cytotoxic -T lymphocytes; Th1, CD4+ T helper 1 cells; Th2, CD4+ T helper 2 cells; TIL, tumor infiltrating lymphocytes; ER, estrogen receptor; CAAI, complex arm-wise aberration index.

## Authors’ contributions

RD VNK OGT conceived and designed the research. RD performed the data analysis. AH ALBD AT BN CC HKMV JUS MN VDH contributed materials and analysis tools. RD VNK OGT wrote the manuscript with input and revisions from all coauthors.

## ACKNOWLEDGMENTS

The authors would like to thank the members of the Troyanskaya lab for helpful discussions.

## CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

## FUNDING

This work was supported by the US National Institutes of Health (NIH) grants R01 GM071966 and R01 HG005998 to OGT and the Research Council of Norway grants 193387/V50 and 193387/H10 to ALBD and VNK, the South Eastern Norway Health Authority grant 39346 to ALBD and the annual grants from the Norwegian Cancer Society to ALBD and VNK. The collaboration between the Kristensen and Troyanskaya labs is made possible by the support from the Leiv Eiriksson mobility program from the Research Council of Norway and Radiumhospitalets legater. OGT is a senior fellow of the CIFAR Genetic Networks program

## REFERENCES

1. Sounni NE, Noel A. Targeting the tumor microenvironment for cancer therapy. *Clin Chem.* 2013; 59:85–93.
2. Weber CE, Kuo PC. The tumor microenvironment. *Surg Oncol.* 2012; 21:172–177.
3. Albini A, Sporn MB. The tumour microenvironment as a target for chemoprevention. *Nat Rev Cancer.* 2007; 7:139–147.
4. Vaday GG, Franitzka S, Schor H, Hecht I, Brill A, Cahalon L, Hershkoviz R, Lider O. Combinatorial signals by inflammatory cytokines and chemokines mediate leukocyte interactions with extracellular matrix. *J Leukoc Biol.* 2001; 69:885–892.
5. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis.* 2009; 30:1073–1081.
6. Shimizu T, Marusawa H, Endo Y, Chiba T. Inflammation-mediated genomic instability: roles of activation-induced cytidine deaminase in carcinogenesis. *Cancer Sci.* 2012; 103:1201–1206.

7. Esposito I, Menicagli M, Funel N, Bergmann F, Boggi U, Mosca F, Bevilacqua G, Campani D. Inflammatory cells contribute to the generation of an angiogenic phenotype in pancreatic ductal adenocarcinoma. *J Clin Pathol.* 2004; 57:630–636.
8. Yano H, Kinuta M, Tateishi H, Nakano Y, Matsui S, Monden T, Okamura J, Sakai M, Okamoto S. Mast cell infiltration around gastric cancer cells correlates with tumor angiogenesis and metastasis. *Gastric Cancer.* 1999; 2:26–32.
9. Sparmann A, Bar-Sagi D. Ras-induced interleukin-8 expression plays a critical role in tumor growth and angiogenesis. *Cancer Cell.* 2004; 6:447–458.
10. Coussens LM, Raymond WW, Bergers G, Laig-Webster M, Behrendtsen O, Werb Z, Caughey GH, Hanahan D. Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev.* 1999; 13:1382–1397.
11. Lin EY, Nguyen AV, Russell RG, Pollard JW. Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. *J Exp Med.* 2001; 193:727–740.
12. DeNardo DG, Barreto JB, Andreu P, Vasquez L, Tawfik D, Kolhatkar N, Coussens LM. CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages. *Cancer Cell.* 2009; 16:91–102.
13. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumors: impact on clinical outcome. *Nat Rev Cancer.* 2012; 12:298–306.
14. Coussens LM, Zitvogel L, Palucka AK. Neutralizing tumor-promoting chronic inflammation: a magic bullet? *Science.* 2013; 339:286–291.
15. Protti MP, Monte LD, Lullo GD. Tumor antigen-specific CD4+ T cells in cancer immunity: from antigen identification to tumor prognosis and development of therapeutic strategies. *Tissue Antigens.* 2014; 83:237–246.
16. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol.* 2007; 8:R157.
17. Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E, Solbach C, Hunker L, Ahr A, Metzler D, Engles K, Karn T, Kaufmann M. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res.* 2009; 11:R15.
18. Baker K, Lachapelle J, Zlobec I, Bismar TA, Terracciano L, Foulkes WD. Prognostic significance of CD8+ T lymphocytes in breast cancer depends upon both oestrogen receptor status and histological grade. *Histopathology.* 2011; 58:1107–1116.
19. Denkert C, Loibl S, Noske A, Roller M, Mueller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, von Törne C, Weichert W, Engles K, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol.* 2010; 28:105–113.
20. Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, de Azambuja E, Quinaux E, Di Leo A, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02–98. *J Clin Oncol.* 2013; 31:860–867.
21. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol.* 2011; 29:1949–1955.
22. Clarke B, Tinker AV, Lee CH, Subramanian S, van de Rijn M, Turbin D, Kalloger S, Han G, Ceballos K, Cadungog MG, Huntsman DG, Coukos G, Gilks CB. Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. *Mod Pathol.* 2009; 22:393–402.
23. Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, Earl HM, Poole CJ, Hiller L, Dunn JA, Bowden SJ, Twelves C, Bartlett JM, et al. Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. *Ann Oncol.* 2014; 25:1536–1543.
24. Ruffell B, DeNardo DG, Affara NI, Coussens LM. Lymphocytes in cancer development: polarization towards pro-tumor immunity. *Cytokine Growth Factor Rev.* 2010; 21:3–10.
25. Terabe M, Matsui S, Noben-Trauth N, Chen H, Watson C, Donaldson DD, Carbone DP, Paul WE, Berzofsky JA. NKT cell-mediated repression of tumor immunosurveillance by IL-13 and the IL-4R-STAT6 pathway. *Nat Immunol.* 2000; 1:515–520.
26. Aspord C, Pedroza-Gonzalez A, Gallegos M, Tindle S, Burton EC, Su D, Marches F, Banchereau J, Palucka AK. Breast cancer instructs dendritic cells to prime interleukin 13-secreting CD4+ T cells that facilitate tumor development. *J Exp Med.* 2007; 204:1037–1047.
27. Gocheva V, Wang HW, Gadea BB, Shree T, Hunter KE, Garfall AL, Berman T, Joyce JA. IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. *Genes Dev.* 2010; 24:241–255.
28. DeNardo DG, Coussens LM. Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. *Breast Cancer Res.* 2007; 9:212.
29. Schmidt M, Boehm D, von Toerne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrmann M. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* 2008; 68:5405–5413.
30. Boudreau A, van't Veer LJ, Bissell MJ. An “elite hacker”: breast tumors exploit the normal microenvironment program to instruct their progression and biological diversity. *Cell Adh Migr.* 2012; 6:236–248.

31. Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, Lee YS, Zhu Q, Kehata M, Li M, Jiang S, Rastaldi MP, Cohen CD, et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res.* 2013; 23:1862–1873.
32. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of affymetrix genechip data at the probe level. *Bioinformatics*. 2004; 20:307–315.
33. Zhu Q, Wong AK, Krishnan A, Aure MR, Tadych A, Zhang R, Corney DC, Greene CS, Bongo LA, Kristensen VN, Charikar M, Li K, Troyanskaya OG. Targeted exploration and analysis of large cross-platform human transcriptomic compendia. *Nat Methods*. 2015; 12:211–214.
34. Chtanova T, Tangye SG, Newton R, Frank N, Hodge MR, Rolph MS, Mackay CR. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *J Immunol.* 2004; 173:68–78.
35. Silwal-Pandit L, Volland HK, Chin SF, Rueda OM, McKinney S, Osako T, Quigley DA, Kristensen VN, Aparicio S, Børresen-Dale AL, Caldas C, Langerød A. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clin Cancer Res.* 2014; 20:3569–3580.
36. Therneau TM, Grambsch PM. Modeling survival data: extending the cox model. New York: Springer; 2000.
37. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002; 415:530–536
38. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, et al. The genomic and transcriptomic architecture of 2,000 breast tumors reveals novel subgroups. *Nature*. 2012; 486:346–352.
39. Naume B, Zhao X, Synnestvedt M, Borgen E, Russnes HG, Lingjaerde OC, Strømberg M, Wiedswang G, Kvalheim G, Kåresen R, Nesland JM, Børresen-Dale AL, Sørlie T. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol.* 2007; 1:160–171.
40. Graham K, de las Morenas A, Tripathi A, King C, Kavanah M, Mendez J, Stone M, Slama J, Miller M, Antoine G, Willers H, Sebastiani P, Rosenberg CL. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br J Cancer*. 2010; 102:1284–1293.
41. Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika*. 69:553–566.
42. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JF, Aparicio S, Ellis IO, Brenton JD, Caldas C. A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene*. 2007; 26:1507–1516.
43. Quigley D, Silwal-Pandit L, Dannenfelser R, Langerød A, Volland HK, Vaske C, Siegel JU, Troyanskaya OG, Chin SF, Caldas C, Balmain A, Børresen-Dale AL, Kristensen VN. Lymphocyte invasion in IC10/basal-like breast tumors is associated with wild-type TP53. *Mol Cancer Res.* 2015; 13:493–501.
44. Russnes HG, Volland HK, Lingjaerde OC, Krasnitz A, Lundin P, Naume B, Sørlie T, Borgen E, Rye IH, Langerød A, Chin SF, Teschendorff AE, Stephens PJ, et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med.* 2010; 2:38–47.
45. Miller WR, Larionov A, Anderson TJ, Evans DB, Dixon JM. Sequential changes in gene expression profiles in breast cancers during treatment with the aromatase inhibitor letrozole. *Pharmacogenomics J.* 2012; 12:10–21.
46. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res.* 2008; 14:5158–5165.
47. Ascierto ML, Kmiecik M, Idowu MO, Manjili R, Zhao Y, Grimes M, Dumur C, Wang E, Ramakrishnan V, Wang XY, Bear HD, Marincola FM, Manjili MH. A signature of immune function genes associated with recurrence-free survival in breast cancer patients. *Breast Cancer Res Treat.* 2012; 131:871–880.
48. Bianchini G, Qi Y, Alvarez RH, Iwamoto T, Coutant C, Ibrahim NK, Valero V, Cristofanilli M, Green MC, Radvanyi L, Hatzis C, Hortobagyi GN, Andre F, et al. Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor-positive and –negative cancers. *J Clin Oncol.* 2010; 28:4316–4323.
49. Abbas AR, Wolslegel K, Seshayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One*. 2009; e6098.
50. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*. 2016; 17:218.
51. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–457.
52. Jeffrey KL, Brummer T, Rolph MS, Liu SM, Callejas NA, Grumont RJ, Gillieron C, Mackay F, Grey S, Camps M, Rommel C, Gerondakis SD, Mackay CR. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol.* 2006; 7:274–283.
53. Prieto GA, Rosenstein Y. Oestradiol potentiates the suppressive function of human CD4 CD25 regulatory T cells by promoting their proliferation. *Immunology*. 2006; 118:58–65.

54. Polanczyk MJ, Hopke C, Vandenbark AA, Offner H. Estrogen-mediated immunomodulation involves reduced activation of effector T cells, potentiation of Treg cells, and enhanced expression of the PD-1 costimulatory pathway. *J Neurosci Res.* 2006; 84:370–378.
55. Generali D, Bates G, Berruti A, Brizzi MP, Campo L, Bonardi S, Bersiga A, Allevi G, Milani M, Aguggini S, Dogliotti L, Banham AH, Harris AL, et al. Immunomodulation of FOXP3<sup>+</sup> regulatory T cells by the aromatase inhibitor letrozole in breast cancer patients. *Clin Cancer Res.* 2009. 15:1046–1051.
56. Vollan HK, Rueda OM, Chin SF, Curtis C, Turashvili G, Shah S, Lingjærde OC, Yuan Y, Ng CK, Dunning MJ, Dicks E, Provenzano E, Sammut S, et al. A tumor DNA complex aberration index is an independent predictor of survival in breast and ovarian cancer. *Mol Oncol.* 2015; 9:115–127.
57. Hicks J, Krasnitz A, Lakshmi B, Navin NE, Riggs M, Leibu E, Esposito D, Alexander J, Troge J, Grubor V, Yoon S, Wigler M, Ye K, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res.* 2006; 16:1465–1479.
58. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 2017; 45:183–189.
59. Tahiri A, Leivonen SK, Lüders T, Steinfeld I, Ragel Aure M, Geisler J, Mäkelä R, Nord S, Riis ML, Yakhini Z, Kleivi Sahlberg K, Børresen-Dale AL, Perälä M, et al. Deregulation of cancer-related miRNAs is a common event in both benign and malignant human breast tumors. *Carcinogenesis.* 2014; 35:76–85.