Check for updates

Postpartum breast cancer has a distinct molecular profile that predicts poor outcomes

Sonali Jindal [1,2,9], Nathan D. Pennock [1,9], Duanchen Sun[3,9], Wesley Horton[1,3], Michelle K. Ozaki [1], Jayasri Narasimhan[1], Alexandra Q. Bartlett[1], Sheila Weinmann[4], Paul E. Goss[5], Virginia F. Borges[6,7], Zheng Xia [2,3,8,10] & Pepper Schedin [1,2,7,10 ✉]

Young women's breast cancer (YWBC) has poor prognosis and known interactions with parity. Women diagnosed within 5–10 years of childbirth, defined as postpartum breast cancer (PPBC), have poorer prognosis compared to age, stage, and biologic subtype-matched nulliparous patients. Genomic differences that explain this poor prognosis remain unknown. In this study, using RNA expression data from clinically matched estrogen receptor positive (ER+) cases ($n = 16$), we observe that ER+ YWBC can be differentiated based on a postpartum or nulliparous diagnosis. The gene expression signatures of PPBC are consistent with increased cell cycle, T-cell activation and reduced estrogen receptor and TP53 signaling. When applied to a large YWBC cohort, these signatures for ER+ PPBC associate with significantly reduced 15-year survival rates in high compared to low expressing cases. Cumulatively these results provide evidence that PPBC is a unique entity within YWBC with poor prognostic phenotypes.

[1] Department of Cell, Developmental and Cancer Biology, Oregon Health & Science University, Portland, OR, USA. [2] Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. [3] Computational Biology Program, Oregon Health & Science University, Portland, OR, USA. [4] Center for Health Research, Kaiser Permanente Northwest, 3800N. Interstate Ave., Portland, OR, USA. [5] Massachusetts General Hospital Cancer Center, Harvard University, Boston, MA, USA. [6] Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [7] Young Women's Breast Cancer Translational Program, University of Colorado Cancer Center, Aurora, CO, USA. [8] Department of Molecular Microbiology and Immunology, Oregon Health & Science University, Portland, OR, USA. [9] These authors contributed equally: Sonali Jindal, Nathan D. Pennock, Duanchen Sun. [10] These authors jointly supervised this work: Zheng Xia, Pepper Schedin. ✉email: schedin@ohsu.edu

Breast cancer incidence is bimodal, with peaks ~45 and 65 years of age referred to as early and late-onset disease, respectively[1–5]. As breast cancer risk does not increase linearly with age, it is suggested that early and late-onset breast cancer are distinct entities with their own risk factors and molecular signatures[2,4]. Early-onset breast cancer, also known as young women's breast cancer (YWBC), is a global concern. YWBC accounts for ~11% of all new breast cancer diagnoses in the United States[6–8], and the incidence of YWBC in many developing countries is higher[9,10]. Further, the incidence of YWBC is increasing world-wide[11–14]. A recent retrospective SEER registry study representing 25% of the US population reported a 1.62 (1.16–2.09) fold increase in the incidence of YWBC between 2000 and 2015 alone, with increased incidence across all races and ethnicities[13]. In addition, compared with late-onset breast cancers, YWBC is enriched in poor prognostic tumor features[15–18], has high levels of mortality[15,18–21], and has experienced limited gains in treatment efficacy[16,22]. Thus, an improved understanding of the underpinnings of YWBC is needed to effectively combat this poor prognostic disease.

An elevated proportion of poor prognostic hormone receptor (HR)-negative and HER2-positive breast cancers is often cited to account for the adverse outcomes in young patients[15–18]. However, several lines of evidence suggest that differences in intrinsic biologic subtypes—including estrogen receptor (ER) and HER2 status—do not wholly account for the observed increased mortality. For example, in the same US SEER study reporting a 1.62-fold increase in YWBC since 2000, the increase in incidence was attributed exclusively to ER-positive (ER+) disease[13]. Further, contrary to expectations that luminal A and B breast cancers are less deadly in young women, a National Comprehensive Cancer Network study of 17,575 women with stage I–III breast cancer reports higher breast cancer mortality in young women with luminal A (HR 2.1; 95% CI, 1.4–3.2) and B (HR 1.4; 95% CI, 1.1–1.9) cancers compared with young women with triple-negative or HER2+ cancers[23]. Similar trends have also been reported in young Chinese women[24]. These studies provide further rationale to explore early-onset breast cancers as distinct entities whose biology is not fully explained by differing ER or HER2 status.

Since breast cancer incidence is influenced by parity[25–30], one possible explanation for the poor prognosis in young patients is that cancer outcomes are associated with childbirth. A recent meta-analysis of 41 studies addressed whether YWBC outcomes are differentially influenced by a diagnosis during pregnancy or the postpartum period. This analysis revealed a higher risk of death only in women diagnosed postpartum (HR 1.79; 95 % CI 1.39–2.29)[31]. Further, these and other studies found that a diagnosis within 5–10 years of a recent pregnancy, referred to as postpartum breast cancer (PPBC)[32], independently associated with a two- to threefold increased risk of death in both ER+ and ER– disease[33,34]. Conversely, studies find that a diagnosis during pregnancy is not associated with poorer outcomes[35–37]. Combined, these studies implicate the existence of a postpartum event that negatively impacts breast cancer prognosis. In women, the postpartum window coincides with a developmental process known as weaning-induced breast involution, a process demonstrated to promote breast cancer development and metastasis in rodent models[38–41]. Given that ~50% of all YWBC are diagnosed within 10 years of a completed pregnancy[33,34], further investigation into the impact of postpartum breast involution on tumor biology is warranted.

Involution is a physiologically normal process that remodels the epithelial-dense, lactational gland to a pre-pregnant-like, non-secretory state[42–44]. In female rodents, where the involution process has been extensively studied, >80% of the lactational mammary epithelium dies as part of a developmentally regulated tissue remodeling process[42–44]. This process coordinates responses of mucosal immunity, fibroblast activation, lymphangiogenesis, and wound-like extracellular matrix deposition[45–49]. In addition to involution creating a transient stromal microenvironment favorable for the expansion and spread of primary tumor cells, involution also durably alters murine mammary tumors. This is evidenced by features of elevated COX-2 expression, increased lymphangiogenesis-inducing capability, augmentation of a tumor-associated immune milieu, and enhanced tumor growth and dissemination phenotypes, all of which persist beyond the period of weaning-induced gland involution in rodents[38,47,50]. Collectively, preclinical studies of PPBC suggest that YWBC may be durably influenced by the transitory developmental processes of mammary gland involution, which may result in distinct gene expression profiles predictive of poor outcomes.

Here, we address whether YWBC can be delineated into distinct molecular subtypes based on a nulliparous or postpartum diagnosis. We focus on ER+ disease as an under-investigated breast cancer subtype accounting for more deaths overall than ER– disease[34,51,52]. We perform comparative RNA Seq expression analyses on treatment-naive formalin-fixed, paraffin-embedded (FFPE) breast cancer tissues from young patients using tumor stage-matched, ER+ postpartum (PPBC), and nulliparous breast cancers (NPBC). We validate gene expression results using multiplex immunohistochemistry (mIHC). We find that PPBC associates with enhanced signatures of cell cycle control, T-cell activation and exhaustion, decreased ER signaling, and altered P53 signaling compared with matched cases diagnosed in nulliparous women. This study strongly supports the hypothesis that normal postpartum breast involution durably alters breast cancer intrinsic and extrinsic factors predictive of disease progression.

Results

PPBC RNA expression profile is distinct from NPBC. To gain insight into the features that could lead to poorer outcomes in PPBC patients, we focused our analyses on clinically determined ER+ cases, as ≥65% of all young breast cancer patients (≤45 years of age) are diagnosed with ER+ disease[53]. Further, young women's ER+ breast cancers have threefold increased likelihood of progressing to metastatic disease when diagnosed postpartum (PPBC) compared with nulliparous cases (Nulliparous Breast Cancer–NPBC)[34]. To obtain a cohort of age and stage-matched, treatment-naive, ER+ NPBC and PPBC cases, we performed chart review for patient age, pregnancy history, tumor stage, subtype, and treatment history. Of 40 selected cases, 16 ER+ cases (PPBC $n = 9$, NPBC $n = 7$) yielded RNA in sufficient quantity and quality to advance to RNA sequencing and subsequent gene expression analyses. Unsupervised hierarchical clustering of these 16 samples across all 14,830 expressed genes yielded separation of 14 of the 16 samples based on parity status (Fig. 1a, nulliparous (blue) vs postpartum (black)). Of note, these cases did not separate based on clinical stage, suggesting parity history is more predictive of tumor gene expression than tumor clinical stage in this young cohort. We identified the most differentially regulated genes between NPBC and PPBC specimens utilizing DESeq2 bioinformatics program and found 364 genes with a false discovery rate (FDR) of ≤0.1 (adjusted p value). Unsupervised clustering of these 364 genes resulted in only one misalignment between the two parity groups (Fig. 1b). To determine whether these differentially expressed genes represent a coordinated change in tumor biology, we used STRING[54] database analysis, which predicts protein–protein interactions across a
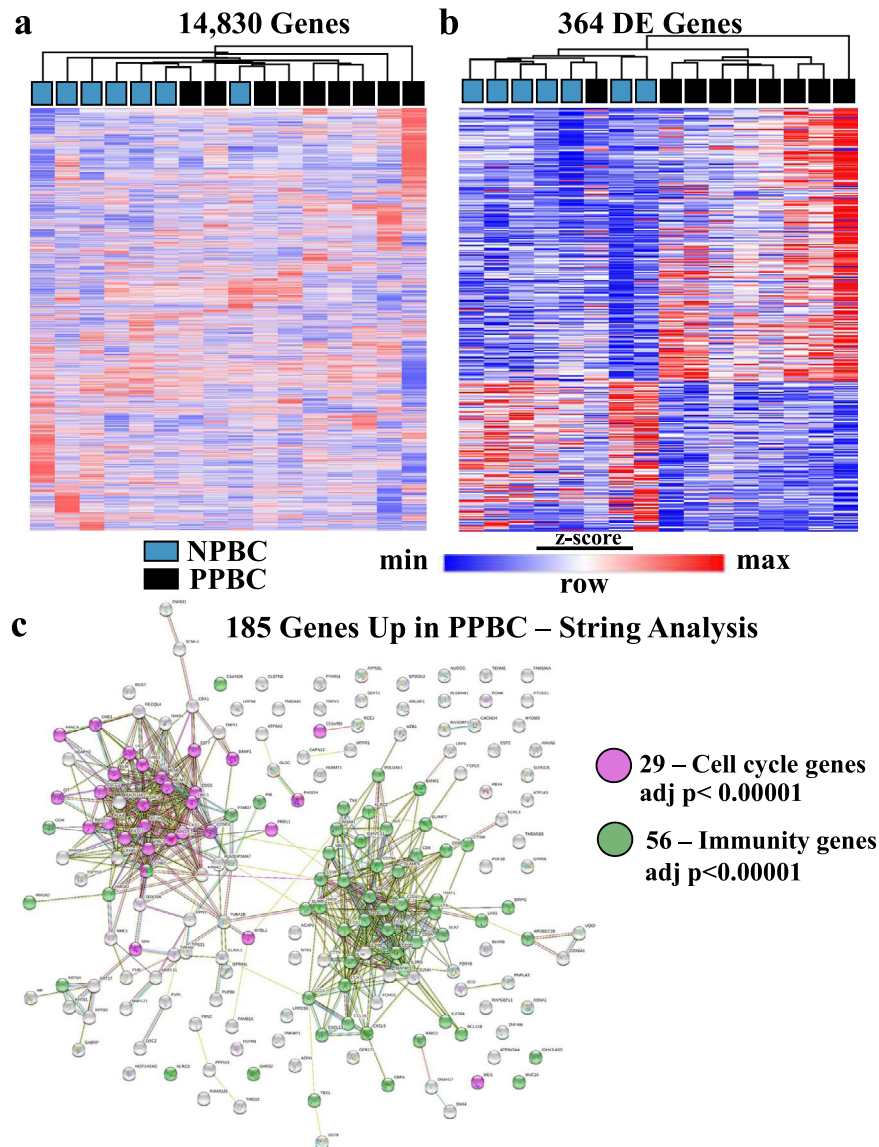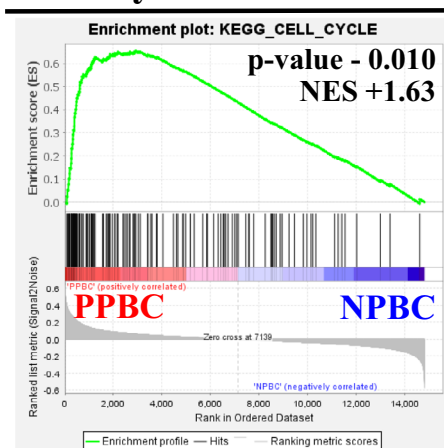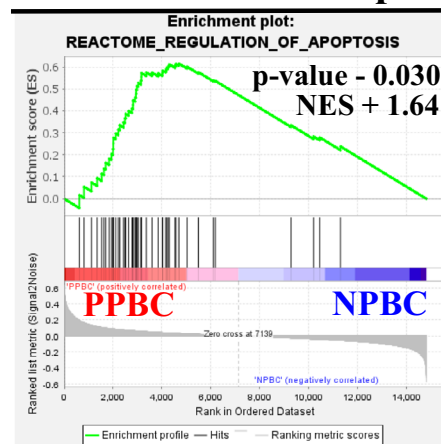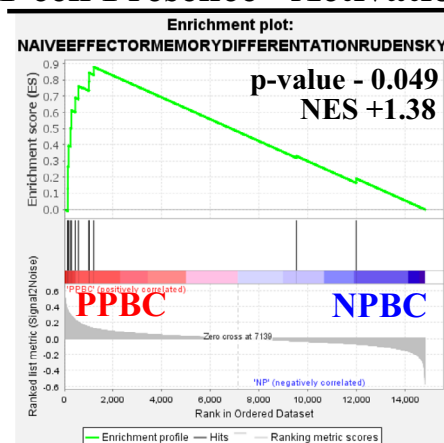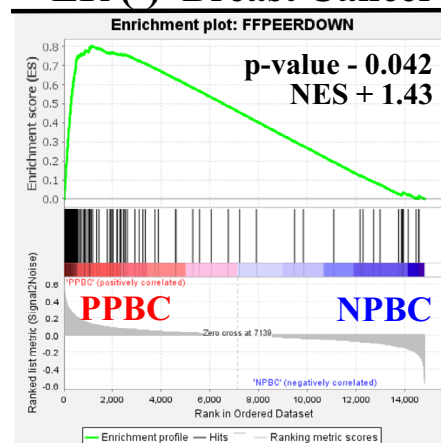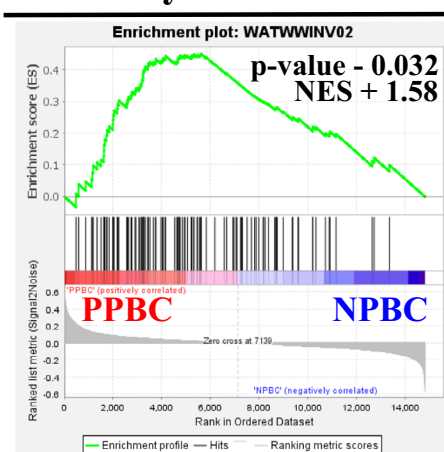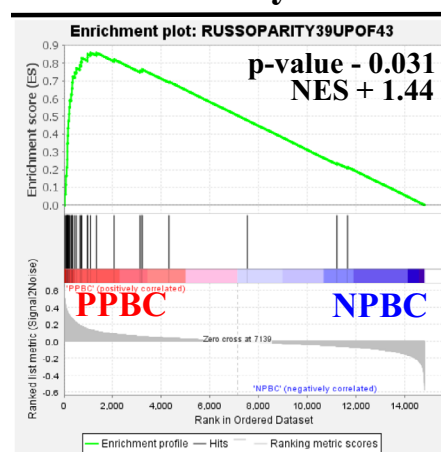
Fig. 1 RNA expression profiling separates postpartum breast cancer (PPBC) from nulliparous breast cancer (NPBC). RNA seq, performed on RNA obtained from FFPE specimens of primary ER+ breast cancer from patients 45 years of age or younger, reveals parity effect. Clustering analysis derived from RNA expression profiles of biologically independent samples of nulliparous breast cancer (NPBC, blue, $n = 7$) and postpartum breast cancer (PPBC, black, $n = 9$). **a** Euclidean hierarchical clustering of the 14,830 genes determined to be expressed above background. **b** Euclidean hierarchical clustering based upon 364 differentially expressed genes between PPBC and NPBC determined by DESeq2 with an FDR < 0.1. **c** STRING database clustering analysis[54] of 185 upregulated PPBC genes generates two distinct biological clusters of statistical significance (adj = adjusted. p values adjusted according to Benjamini–Hochberg for multiple comparisons).

variety of annotated "omics studies". We identified two dominant (p value < 0.00001) clusters of genes that increased in PPBC compared with NPBC. One of these clusters is associated with cell cycle programs (Fig. 1c, purple) and the other with immunity (Fig. 1c, green).

Gene set enrichment characteristics of PPBC. We next performed rank-based gene set enrichment analysis (GSEA) on PPBC compared with NPBC. We observed enrichment in pathways associated with six distinct biological processes in PPBC compared with NPBC (Fig. 2a–f, Supplementary Data 1). Consistent with the STRING analysis (Fig. 1c), we observed enrichment for cell cycle and proliferation signatures (Fig. 2a), as well as signatures associated with cell death and DNA repair (Fig. 2b). We also observed enrichment in the T-cell presence-activation signature (Fig. 2c), an observation that provides cell-specific

insight into the enriched immunity signature detected by STRING analyses. Surprisingly, even though all cases were determined to be definitively ER+ (Supplementary Table 1a), in the PPBC cohort we observed enrichment of gene expression profiles associated with ER-negative breast cancers[55] (Fig. 2d). Further, in PPBC tumors, we observed significant enrichment of gene signatures associated with the normal developmental processes of pregnancy and weaning-induced breast involution, which supports the idea that PPBC tumors are durably influenced by their host environment (Fig. 2e, f). To further investigate the potential role of normal postpartum biology in the imprinting of tumor biology, we next explored the relationship between our PPBC cases and gene expression signatures obtained from whole-transcriptome profiling from breast tissue of healthy patients ($n = 109$)[56]. We analyzed this publicly available data set to focus on gene sets from healthy nulliparous and postpartum subjects

a **Cell Cycle-Proliferation**

p-value - 0.010
NES +1.63

b **Cell Death-DNA Repair**

p-value - 0.030
NES + 1.64

c **T-cell Presence - Activation**

p-value - 0.049
NES +1.38

d **ER (-) Breast Cancer**

p-value - 0.042
NES + 1.43

e **Mammary Gland Involution**

p-value - 0.032
NES + 1.58

f **Parity**

p-value - 0.031
NES + 1.44

within 2 years of their last childbirth. As anticipated from previous reports[56,57], we observed some normal involution signatures in the postpartum normal tissue expression data sets, such as a parity signature (Supplementary Fig. 1a) and the immune infiltrate signature (Supplementary Fig. 1b). However, neither the immune exhaustion signature (Supplementary Fig. 1c), the ER-negative breast cancer signature (Supplementary Fig. 1d) nor the proliferation signatures (Supplementary Fig. 1e, f) were upregulated in normal postpartum tissue, whereas these gene signatures were upregulated in PPBC samples. One interpretation of these data is that PPBC is a convergence between breast cancer and the reproductive milieu.
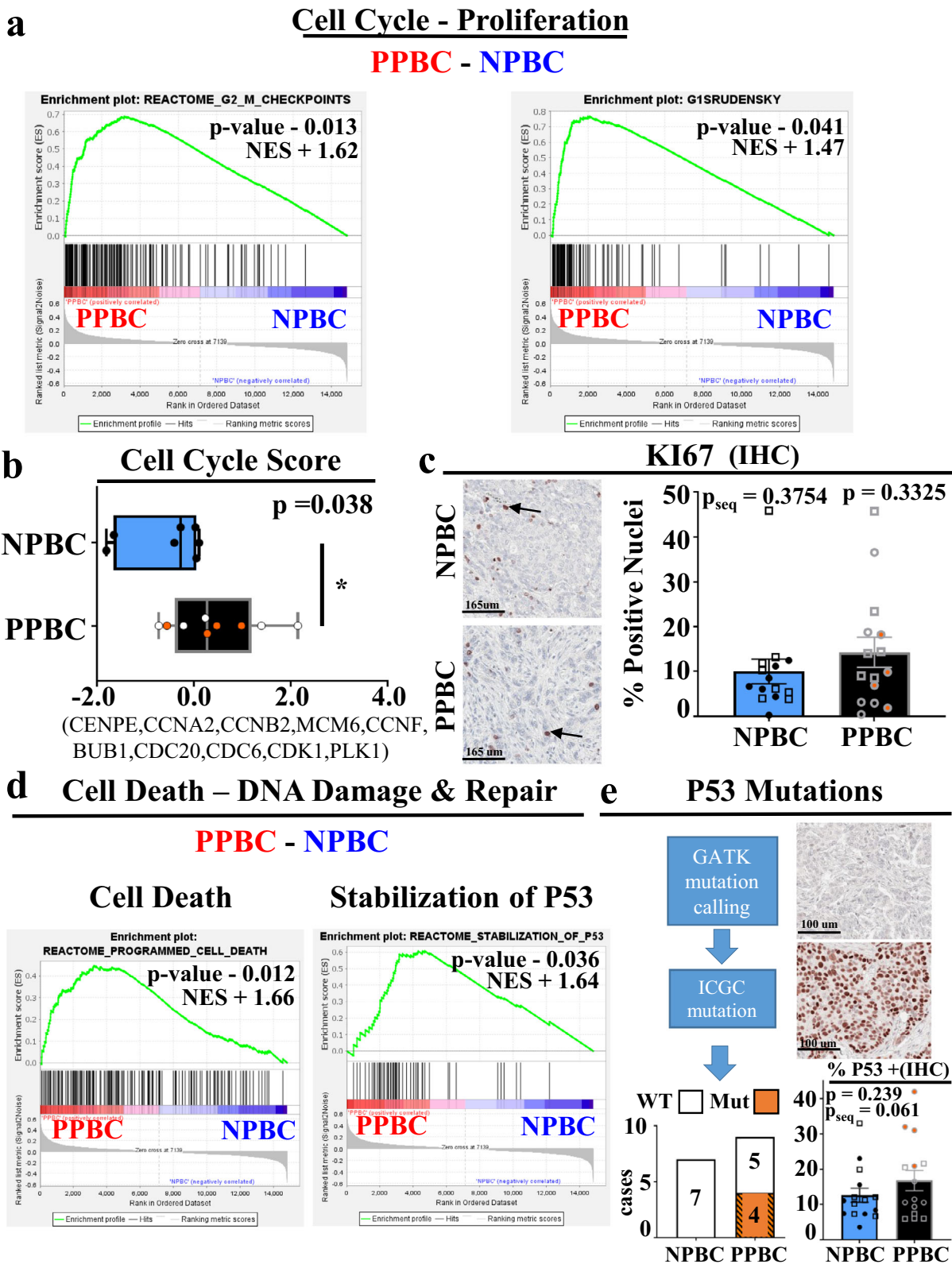
Fig. 2 GSEA identifies cell cycle, cell death, T-cell immunity, estrogen receptor signaling, and mammary gland developmental gene sets as differentially expressed between PPBC and NPBC. Gene Set Enrichment Analysis (GSEA) was performed on normalized RNA Seq expression data from biologically independent samples of postpartum breast cancer (PPBC, red, $n = 9$) and nulliparous breast cancer (NPBC, blue, $n = 7$) patients from Fig. 1, utilizing Molecular Signature Database Collections (V. 7.0) and 100 custom gene lists compiled from the literature review. Gene sets with p values < 0.05 belonging to six biological processes were manually curated: **a** cell cycle and proliferation, **b** cell death and DNA damage repair, **c** T-cell related immunity, **d** estrogen receptor signaling and estrogen receptor-negative breast cancer, **e** post-lactation mammary gland involution in rodents, and **f** parity status in the human breast. Representative enrichment plots from each group are displayed with the determined nominal (non-adjusted) p value and normalized enrichment score (NES)[103].

Proliferation and TP53 characteristics of PPBC. To further explore the relationship between the observed cell cycle gene signature upregulated in PPBC and tumor cell proliferation, we examined additional cell cycle gene sets and performed immunohistochemistry staining for the cell cycle protein KI67. While multiple gene sets (Fig. 3a) and single sample composite gene score analyses (Fig. 3b) confirmed statistically significant enrichment of cell cycle genes in PPBC, IHC staining for KI67 did not differ by parity status in our FFPE RNA Seq samples (Fig. 3c, circles, $pseq = 0.3754$). To more rigorously assess the consistency of our RNA Seq findings, we expanded our IHC cohort to include additional young women's, ER+ PPBC and NPBC, FFPE specimens (Fig. 3c, squares) providing 15 samples in each group. With this expanded cohort we found no statistical significance in KI67 staining between these groups ($p = 0.3325$), depicting a disparity between protein single stain of proliferation (KI67) and composite gene evaluations of proliferative activity. We next explored the signature of increased cell death, DNA damage, and DNA repair gene signatures in PPBC (Fig. 2c), which could suggest increased genetic instability in PPBC tumors. Additional pathway analyses found elevated programmed cell death and TP53 pathways in PPBC (Fig. 3d), data consistent with mutant TP53. To address this possibility, we utilized expression profiling sequences to perform genomic analysis toolkit (GATK) mutational calling, followed by cross-referencing for known TP53 mutations[58] (Fig. 3e, flow chart). These analyses identified four out of nine PPBC samples as containing canonical TP53 mutations (Fig. 3e, bar chart). IHC analyses for P53 on all 30 IHC samples validated these mutation calls. Specifically, the four samples with TP53 mutations displayed enhanced P53 staining consistent with stabilization of P53 protein by mutation (Fig. 3e, inset). Of note, within our entire cohort, we observed significant staining (>10% + nuclei) for P53 in most cases. However, staining was not statistically different between PPBC and NPBC samples. To assess the degree these TP53 mutations were responsible for the increased proliferation signature attributes observed in PPBC, we tracked the position of these bona fide TP53 mutants throughout our analysis (orange-filled circles), and found that TP53 mutational status does not correlate with cell cycle score (Fig. 3b), nor KI67 (Fig. 3c).

PPBC is enriched for T-cell immunity. The most dominant gene signature identified in PPBC is immunity (Fig. 1c), specifically T-cell presence and activation (Fig. 2c). An important direct mechanism of anti-tumor immunity is direct tumor cell lysis by cytotoxic cells. Thus, we evaluated for cytotoxic cells in the individual cases using a validated gene signature[59,60], referred to as an "immune infiltrate" signature, which is reflective of the presence of cytotoxic T-cells or NK cells. We observed PPBC samples were enriched in this immune infiltrate signature (Fig. 4a). We next considered that PPBC tumors might be overall enriched for immune cells; however, examination of CD45 (a pan immune cell gene/protein) by IHC analyses (Supplementary Fig. 2b, $p = 0.108$) or RNA expression (Supplementary Fig. 2c, $p = 0.351$) found that CD45 was not significantly increased in

PPBC tumors. These data are consistent with specific enrichment of cytotoxic immune cells or T-cells within PPBC. To further delineate between these possibilities we performed T-cell receptor (TCR) repertoire analysis to look for T-cell number and evidence of activation. Using RNA Seq expression data, TCR repertoire analysis revealed more unique TCR sequences from PPBC compared with NPBC samples (Fig. 4b). Increased TCR repertoire could be the consequence of increased diversity of tumor resident T-cell clones or the consequence of having increased overall T-cell numbers in PPBC specimens. To address the relative diversity of the repertoire, we performed normalized clonal analyses. The normalized entropy (clonality index) analysis (Fig. 4c) and the Gini index analysis (Supplementary Fig. 2d) are different mathematical models which both assess the diversity of the repertoire relative to overall numbers of unique TCR sequences. Both of these normalized measures of TCR diversity depict a reduced TCR diversity in PPBC specimens, indicative of clonal expansion. Further, we observe the increased clonality to occur in PPBC within the "hyper-expanded" and "small" frequency population of T-cell clones (Supplementary Fig. 2e). Collectively, increased TCR sequences with increased clonality in two different clonal space populations implicate T-cell activation, which could occur through expansion of tissue-resident memory populations as a consequence of inflammation and/or by antigen-specific T-cell responses[61,62]. Overall, these data are consistent with PPBC tumors eliciting a stronger T-cell response (immunologically hotter) when compared with NPBC.

Greater insight as to how a T-cell presence may influence the tumor microenvironment and perhaps contribute to the response to therapy can be gained by a better understanding of attributes of the T-cell pool. Interestingly, in our GSEA analysis, one of the significant signatures to distinguish between PPBC and NPBC samples was derived from molecular distinctions between exhausted and non-exhausted T-cells found to be conserved between chronic viral and tumor murine models (Fig. 4d). Given the potential importance of this exhausted T-cell enrichment profile, we performed additional analyses to further understand the nature of the T-cells in PPBC samples. First, we performed CIBERSORT[63] analyses (Supplementary Data 2), which provides a normalized estimation of specific immune cell populations from mixed population RNA expression data. CIBERSORT analyses reported significantly ($p = 0.009$) increased levels of CD8 T-cells in PPBC cases (Fig. 4e), which we confirmed by IHC analyses (Supplementary Fig. 2f). CIBERSORT also reported a significant increase in T follicular helper cells (Tfh) (Fig. 4f). Interestingly, among the molecules that distinguish Tfh from other T-cell populations is the high expression of PD-1[64]. Although widely utilized, CIBERSORT has demonstrated limitations in accurately predicting differential abundance amongst cell populations with similar features. To more robustly characterize the abundance and identity of T-cells in PPBC compared with NPBC we performed mIHC staining with a specific emphasis on PD-1 (a shared feature of activated, exhausted, and Tfh T-cells) and the exhaustion correlated transcription factor TOX1[65–67]. One distinct advantage to mIHC and image cytometry is the ability

a

Cell Cycle - Proliferation

PPBC - NPBC

b

Cell Cycle Score

p =0.038

(CENPE,CCNA2,CCNB2,MCM6,CCNF,
BUB1,CDC20,CDC6,CDK1,PLK1)

c

KI67 (IHC)

$p_{seq} = 0.3754$ $p = 0.3325$

d

Cell Death – DNA Damage & Repair

PPBC - NPBC

Cell Death   ### Stabilization of P53

e

P53 Mutations

to deepen subset analyses based upon the context of intact tissue. In our samples, we noted a prominent accumulation of T-cells (CD3+) at the tumor border (Fig. 4gi, Source Data) in both PPBC and NPBC. When this tumor border region was interrogated by image cytometry, we identified a statistically significant increase of T-cells—and more specifically of CD4 T-cells—as a fraction of all immune cells (CD45+, Fig. 4gii-v, CD3 $p = 0.046$, CD4 $p = 0.014$). Regarding the relative polarization and activation of T-cells as evaluated by the expression of PD-1 and TOX1, we observed approximately twofold increases in PD-1+ (red or

Fig. 3 Cell Cycle and TP53 gene signatures, TP53 mutational analysis, and immunohistochemical validation. Detailed examination of the proliferation, cell death, and DNA damage pathways, as identified by RNA expression profiling described in Fig. 2, and IHC examination of these pathways. Depiction of **a** two additional GSEA enrichment plots for cell cycle. **b** Single sample cell cycle score determined from RNA expression values from the indicated genes (PPBC $n = 9$, NPBC $n = 7$). Data are presented as a minimum to maximum with median value marked by a line within the depicted interquartile range, and p value determined by Students' unpaired two-tailed t test with Welch correction. **c** Examples of immunohistochemical (IHC) evaluation of KI67-positive (brown color) protein expression (left), with quantification of KI67 signal evaluated as the proportion of nuclei (right, PPBC $n = 15$, NPBC $n = 15$) Data are presented as mean values ±SEM, and p value determined by Students' unpaired two-tailed t test with Welch correction. Samples evaluated by RNA Seq are depicted by circles and pseq refers to p values for these samples only, while expanded cases for IHC are depicted by squares and p values reflect values for the whole cohort. **d** GSEA analysis assessments of cell death (left) or DNA damage and repair associated gene sets (TP53, right) (PPBC $n = 9$, NPBC $n = 7$). **e** Flow diagram outlining computational steps and results for prediction of the presence of wildtype (WT) or mutant (MUT) TP53 genes in PPBC ($n = 9$) and NPBC ($n = 7$) cohorts utilizing RNA Seq expression data (left), and P53 protein expression (brown color) assessed by IHC (PPBC = 15, NPBC = 15), with P53 signal reported as percent positive area (right). Data are presented as mean values ±SEM and p value assessed by students' unpaired two-tailed t test with Welch correction. International Cancer Genome Consortium (ICGC) identified TP53 mutations are noted by orange-filled circles. For GSEA plots, p values and normalized enrichment score (NES) were determined by GSEA software[103] comparing PPBC (red, $n = 9$) and NPBC (blue, $n = 7$) biologically independent samples as described in Fig. 1 and 2.

yellow bars, black star) and PD-1+ TOX1+ T-cells (yellow bar, yellow star) within both the CD4 (Fig. 4giii-iv, white arrows, Fig. 4gvi) and CD8 T-cell (Fig. 4giii-iv, black arrows, Fig. 4gviii) compartments. Intratumoral T-cells were also evaluated; however, in general infiltration beyond the tumor border was sparse. Although these data trended towards the same patterns as observed at the tumor border, the scarcity of populations reduced the numerical power necessary for statistical significance. Combined, these data support the conclusions derived from the RNA expression signatures, chiefly that PPBC has increased levels of activated T-cells that express PD-1 and TOX1, which likely contribute to the enhanced signatures of exhaustion from GSEA analysis and the Tfh profile observed from CIBERSORT analyses.

PPBC regulon activity predicts poor outcomes in YWBC. To further compare differences between PPBC and NPBC, we assessed transcription factor activity networks known as regulons, as prior work relying on FFPE tissues demonstrated enhanced fidelity of RNA pathway analysis through regulon analysis[55]. Consistent with the STRING and GSEA data above, we observed the most upregulated regulons to be transcription factors associated with cell cycle pathways (e.g., E2F1, E2F4) (Fig. 5a). Second, we noted the most downregulated pathways in PPBC to be TP53 and ESR1, data also consistent with our pathway analyses (Fig. 3e, Fig. 2d). Although all tumors in our study are highly ER+ by clinical assessment and do not differ in percent ER positivity between groups (Supplementary Fig. 3a, b), several of the most differentially regulated regulons between PPBC and NPBC are transcription factors that are also differentially regulated between ER− and ER+ breast cancers[55] (Fig. 5a, green boxes). This correlation becomes more evident when we plot regulon activity in PPBC vs NPBC in comparison with regulon activity previously reported between ER− vs ER+ cases (Fig. 5b)[55]. To evaluate ER signaling further, we plotted the single sample regulon activity score for the ER-associated pathway (ESR1) for all samples, which revealed significantly decreased ESR1 signaling in PPBC compared with NPBC (Fig. 5c).
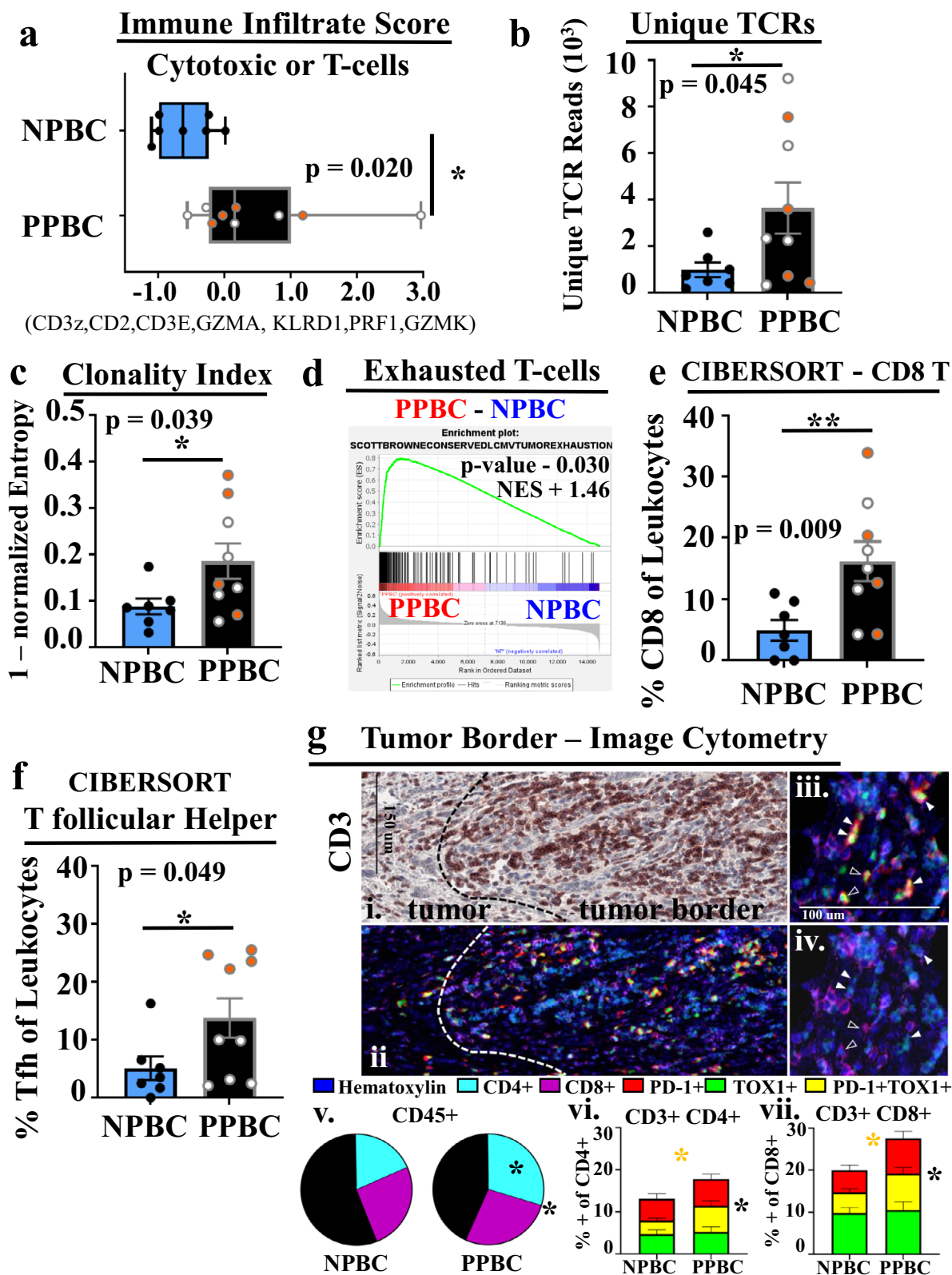
Several gene sets and weighted gene expression algorithms exist for ascribing tumor cell molecular subtype identity and treatment recommendations, which historically have focused on HR activity as a target for therapy and delineator of subtype. We next evaluated whether these validated gene sets could distinguish between PPBC and NPBC cases. First, we performed PAM50® molecular subtype determination on all 16 samples. Unsupervised hierarchical clustering based upon the PAM50® gene expression values did not robustly separate the 16 cases by parity status or molecular subtype (Fig. 5d). However, traditionally good prognostic luminal A cases in the PPBC group clustered with the poorer prognostic luminal B cases in the NPBC group (red boxes), data consistent with the idea that luminal

A PPBC has poorer outcomes than predicted based on their luminal A designation. Next, we used normalized RNA Seq expression values from Oncotype Dx® genes, designed to provide a recurrence score in ER+ tumors[68,69] to compute pseudo-Oncotype Dx® scores[70,71] (Fig. 5e). As predicted, the Oncotype scores were lowest in luminal A (dark purple), increased in luminal B (pink), with further increases in Her2 (green) and finally basal cases (orange, PPBC only). We also observed a statistically significant increase ($p = 0.034$) in overall Oncotype Dx® score in the PPBC cohort compared with the NPBC cohort, data consistent with overall reduced ER signaling in the PPBC tumors. Likewise, we evaluated how genes in the Mammaprint® signature, which is considered to be a tumor cell-intrinsic determination of tumor cell subtype, devoid of stromal-related genes, clustered our PPBC and NPBC cases (Supplementary Fig 2c). We found no association between the expression of these genes and the parity status of samples. Combined, these analyses reveal a need for improved prognostic gene signatures for YWBC. We next utilized our results characterizing PPBC through regulon analysis and immune exhaustion gene sets to establish a PPBC signature for ER+ disease.

To generate a composite PPBC signature, we added together the single sample regulon values for the immune exhaustion and E2F1 regulons (the two most upregulated PPBC regulons), and then subtracted the P53 and ESR1 regulon values (the two most downregulated PPBC regulons). To determine whether this PPBC gene expression signature could predict outcomes in an ER+ YWBC cohort. We assembled a YWBC cohort (≤45 years old) with outcomes data by compiling gene expression data across seven previously published studies[72–78]. Although no parity history was available on these publicly available cases, upon applying our PPBC gene signature to this YWBC cohort ($n = 311$ patients with both ER+ and ER− disease) we observed a highly significant decrease in 15-year overall survival in breast cancer patients with a PPBC Hi signature score (HR 2.134, $p = 0.0011$) compared with those with a low score (PPBC Lo, Fig. 5f). Classically, ER+ breast cancers have a better prognosis than ER− cancers, and this was found to be true in this cohort as well (Supplementary Fig. 3d, HR = 2.455, $p = 0.0001$). To determine whether our PPBC signature was indicative of only ER status, we repeated the analysis on only the ER+ cases ($n = 214$) and again found statistically significant reduced survival in the PPBC signature high group compared with the low group (Fig. 5g, HR = 2.30, $p = 0.0084$).

Discussion

In the present study, we addressed whether PPBC is molecularly distinct from breast cancer diagnosed in nulliparous women. We utilized a small FFPE breast cancer cohort, rigorously controlled

a **Immune Infiltrate Score**

Cytotoxic or T-cells

NPBC

PPBC

p = 0.020

*

-1.0   0.0   1.0   2.0   3.0

(CD3z,CD2,CD3E,GZMA, KLRD1,PRF1,GZMK)

b **Unique TCRs**

*

p = 0.045

NPBC   PPBC

c **Clonality Index**

p = 0.039

*

NPBC   PPBC

d **Exhausted T-cells**

PPBC - NPBC

Enrichment plot:
SCOTT BROWNE CONSERVED LCMV TUMOR EXHAUSTION

p-value - 0.030
NES + 1.46

PPBC   NPBC

— Enrichment profile — Hits — Ranking metric scores

e **CIBERSORT - CD8 T**

**

p = 0.009

NPBC   PPBC

f **CIBERSORT**

T follicular Helper

p = 0.049

*

NPBC   PPBC

g **Tumor Border – Image Cytometry**

i. tumor   tumor border

iii.

iv.

■ Hematoxylin ■ CD4+ ■ CD8+ ■ PD-1+ ■ TOX1+ ■ PD-1+TOX1+

v. CD45+

NPBC   PPBC

vi. CD3+ CD4+

NPBC   PPBC

vii. CD3+ CD8+

NPBC   PPBC

for patient age, BMI, parity history, tumor clinical stage, ER status, and treatment naivety, which permitted us to delineate the role of recent childbirth on tumor gene expression in the absence of potential treatment effects. We observed gene expression signatures of PPBC to include pronounced T-cell presence and T-cell activation/exhaustion signatures, reduced TP53 activity, reduced ER signaling, and increased cell cycle gene signatures. Further, we find PPBC cases in our cohort are characterized by gene expression signatures associated with normal murine mammary gland involution[79–83], as well as recent childbirth in

Fig. 4 PPBC is enriched for activated T-cells compared with NPBC. Characterization of the immune cell presence and T-cell immunity enriched in postpartum breast cancer (PPBC) compared with nulliparous breast cancer (NPBC). **a** RNA Seq expression data was evaluated for genes associated with cytotoxic or T-cell immunity using a gene signature called the single sample immune infiltrate score (PPBC $n = 9$, NPBC $n = 7$). **b** Unique numbers of T-cell receptors (TCR) for each RNA Seq sample, compared between groups (PPBC $n = 9$, NPBC $n = 7$). Relative clonality demonstrated by **c** normalized (norm) entropy (PPBC $n = 9$, NPBC $n = 7$). **d** GSEA profile-derived from exhausted T-cell signature. CIBERSORT deconvolution of RNA Seq data depicts increased **e** CD8 T-cell and **f** T follicular helper (Tfh) presence as a fraction of total leukocytes (PPBC $n = 9$, NPBC $= 7$). **g** multiplex IHC analysis of the tumor border in NPBC ($n = 13$) and PPBC ($n = 14$) cases was subjected to quantification by image cytometry. **gi** Hematoxylin (blue) stain & AMEC (red/brown) for CD3 demonstrating T-cell accumulation in the tumor border region. Dashed lines indicate demarcation of intratumoral and tumor border regions. **gii** Aligned pseudo colored multiplex IHC images depicting staining from hematoxylin (dark blue) and chromagen mediated antibody detection of CD4+ (light blue) CD8+ (purple), PD-1+ (red) or TOX1+ (green) cells. PD-1+/TOX1+ cells **giii** appear yellow due to overlap of red and green coloring. **giv** PD-1+ and TOX1+ cells depicted in giii can be either CD4+ (white arrows) or CD8+ (black arrows) T-cells. **gv** Pie-charts depicting increased CD4+ CD3+ T-cells (light blue, $p = 0.0052$) and total T-cell content (light blue and purple, $p = 0.0225$) as fraction of CD45+ cells in the PPBC cohort. PD-1+ (red), TOX1+ (green) or PD-1+ TOX1+ (yellow) cells as a fraction of the **gvi** CD45+ CD3+ CD4+ (yellow group comparison = yellow star $p = 0.0225$, PD-1/red + yellow comparison = black star $p = 0.05$) or **gvii** CD45+ CD3+ CD8+ (yellow group comparison = yellow star $p = 0.0205$, PD-1/red + yellow comparison = black star $p = 0.028$) T-cell compartments. *P* values determined by GSEA software (nominal, non-adjusted *p* value, **d**) or Students' unpaired two-tailed *t* test with Welch correction (**a–c**, **e**, **f**), or Students' unpaired one-tailed *t* test with Welch correction for confirmatory IHC(*$p \leq 0.05$, **g**). ICGC identified TP53 mutations are noted by orange-filled circles. For box and whisker plot (**a**) data are presented as minimum to maximum with median value marked by a line within the depicted interquartile range, whereas data depicted in bar graphs (**b**, **c**, **e**, **f**, **gvi**, **gvii**) are presented as mean values ±SEM.

healthy women[56,84]. We compiled a signature composed of transcription factor regulons representing the discrete biological pathways differentially expressed in ER+ PPBC and applied this PPBC regulon signature in a large YWBC population. This analysis revealed a significant overall survival disadvantage in young women who had a high PPBC score compared with those with a low score. In sum, these data are consistent with the transient event of normal mammary gland involution durably influencing breast cancer biology, leading to more lethal cancers.

Our data related to a pronounced T-cell presence and activated/exhausted T-cell signatures in PPBC samples is consistent with the idea that normal weaning-induced breast involution impacts the tumor immune milieu. Normal mammary gland involution is characterized by increased T-cell infiltrate[56], which in rodent models includes regulatory (Foxp3, Il-10) and anergized/tolerized T-cell phenotypes[40,45,50]. Physiologically regulated T-cell suppression likely mitigates the potential for self-antigen recognition that could result during the physiologically normal, massive epithelial cell death phase that occurs with cessation of weaning[85,86]. In rodents, PPBC tumors, but not tumors arising in nulliparous hosts, were characterized by an immune milieu consistent with T-cell suppression and tumor cell immune avoidance[50]. This result is consistent with involution durably altering the tumor immune milieu.

Our observation of loss of wildtype TP53—specifically in PPBC tumors—may also reflect normal, weaning-induced involution biology. The P53 tumor suppressor has been studied extensively with respect to its role in maintaining genomic stability[87]. However, P53 is also established as a physiological regulator of involution where its activation initiates apoptosis in the secretory epithelium[88,89]. We speculate that tumor cells present in the involution environment may obtain a survival advantage by suppressing response to this physiologic TP53 dependent cell death pathway. Of potential relevance, studies comparing early and late age at first pregnancy found that early age at first birth associates with long-term protection, whereas late age at first birth is associated with increased risk for breast cancer. In these studies, TP53 mutations were enhanced in late parity cancer cases[90], implicating older maternal age as an additional risk factor for harboring TP53 mutations. Collectively these results and our observations in the present study warrant further investigation into the relationships between parity, maternal age at first childbirth and P53, in conferring poor prognosis.

A dominant molecular distinction in our genomic cohort data was reduced ER signaling in PPBC cases as compared with NPBC. This observation was surprising given that immunohistochemical assessments revealed these tumors to be highly ER-positive. One simple interpretation of these data is that in the postpartum setting, ER-positive breast cancer is more analogous to ER-negative disease with respect to downstream ER signaling pathways. Consistent with our observations of reduced estrogen signaling in PPBC, in a study of postpartum normal and tumor breast tissue[84], the signatures of ER signaling (ESR1) were reduced in postpartum cases compared with their nulliparous counterparts[91]. As with P53, it is possible that the downregulation of ER signaling in tumor cells is a specific adaptation to the involution microenvironment. Signal transducer and activator of transcription (Stat) 5a is a well-established positive regulator of lactation and its suppression is a requisite for the execution of epithelial cell death after weaning[80]. Further, Stat5 expression is under estrogen control in the murine mammary gland[92]. Thus, one untested possibility is that ER+ tumor cells maintain Stat5 survival signaling during involution by downregulating ER signaling. Consistent with this hypothesis, expression of a constitutively activated variant of Stat5 in the murine mammary gland prevented weaning-induced involution and was associated with ER+ adenocarinomas[93]. In sum, our study adds to a growing body of literature reporting poor prognosis in breast cancers expressing classic weaning-induced mammary gland involution gene signatures[56,81,83], and for the first time, extends these studies to demonstrate enrichment of these signatures in breast cancers that have experienced the involution microenvironment.

We also observed a robust increase in cell cycle genes in PPBC. It is noteworthy that these proliferation-associated gene expression signatures did not correlate with increased tumor cell proliferation, as measured by KI67. The lack of increased proliferation in PPBC compared with NPBC is concordant with published data from a large retrospective study showing increased metastasis rates in PPBC compared with nulliparous cases, but similar tumor cell proliferation rates[34], which were also assessed by KI67 protein expression. It is possible that the biology captured in the cell cycle gene sets is, in fact, distinct from cell division biology, and/or that KI67 does not adequately capture cell proliferation[94,95]. Additional research is required to address this apparent conundrum.

Finally, we suggest the gene expression signatures outlined here in human PPBC will provide insight as to why PPBC patients have poorer treatment responses and stimulate interest in alternative treatment approaches. When we considered how our
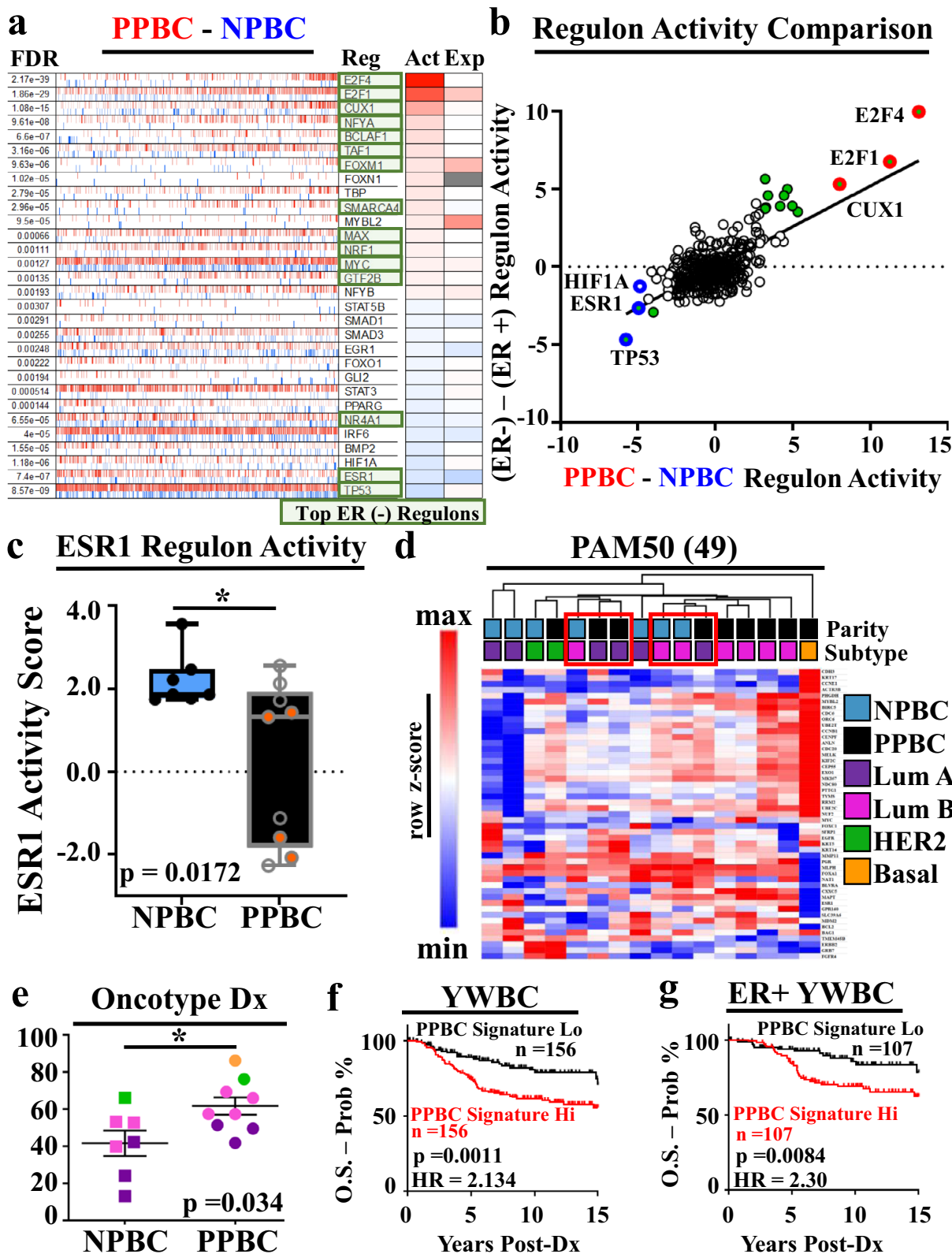
observations fit into existing paradigms of informative gene sets, we found no clear correlation from PAM50® subtype determination nor the Mammaprint® signature. The Oncotype Dx® recurrence score calculation did modestly delineate between nulliparous and PPBC cases, however, the majority of these 16 YWBC cases had high recurrence scores regardless of ER expression or parity status. Thus, further research is needed to determine the best clinical tools capable of delineating low- and high-risk ER+ YWBC and the influence of parity status on those outcomes. By combining parity, treatment, and outcomes data

Fig. 5 Regulon activity signatures identify key biological processes in PPBC that predict poor prognosis in Young Women's Breast Cancer. RNA expression profiles between postpartum breast cancer (PPBC, red, $n = 9$) and nulliparous breast cancer (NPBC, blue, $n = 7$) were evaluated for **a** transcriptional network activity through regulon analysis and **b** compared with regulon results comparing FFPE derived ER-negative to ER-positive breast cancer specimens. Most differentially active ER- regulons are highlighted as green boxes (a) and green circles (b). Most differentially active regulons between PPBC and NPBC are highlighted by bolded circles in red (upregulated) or blue (downregulated). **c** Single sample ESR1 regulon activity scores for PPBC ($n = 9$) and NPBC ($n = 7$) were evaluated. ICGC identified TP53 mutations are noted by orange-filled circles. Data are presented as a minimum to maximum with a median value marked by a line within the depicted interquartile range. **d** Gene expression values for the expressed (49) genes of the PAM50® subtype determination assay were evaluated to determine intrinsic subtype for each sample assessed for determination of sample clustering in PPBC and NPBC samples. **e** Pseudo-Oncotype Dx® recurrence scores were derived from RNA expression values for each sample and compared between cohorts (NPBC $n = 7$, PPBC $n = 9$). Data are presented as mean values ±SEM. **f** A postpartum breast cancer regulon-based gene expression signature was composed incorporating immune exhaustion (Fig. 4d), proliferation (E2F1), P53, and ESR1 regulon activity values and evaluated for prognostic significance from a multi-study accumulated cohort of Young Women's Breast Cancer ($n = 311$) composed of female breast cancer patients whose primary breast cancer diagnosed occurred at the age of 45 or under. **g** Subset analysis in ER-positive cases ($n = 214$) from this YWBC cohort. Cohorts were split into PPBC signature high (hi, red, $n = 107$) or low (lo, black, $n = 107$) cohorts based upon the median value of the group plotted. P values were determined by Students' unpaired two-tailed t test with Welch's correction (**c**, **e**) or by two-tailed log-rank (Mantel–Cox) evaluation for survival plots (**f**, **g**). Log-rank evaluated Hazard Ratios (HR) are depicted.

already available, it may be possible to inform novel treatment strategies for PPBC and determine if any of the existing agents for overcoming ER therapeutic resistance, such as the CDK4/6 inhibitors and their inhibition of the cell cycle, may have added benefit for PPBC, or identify other novel combinations. In addition, given the observation that PPBC, which evolved in the involution environment, has an elevated and activated T-cell compartment with increased expression of PD-1, there may be a select benefit for these patients from checkpoint blockade inhibitors. Already, preclinical data in mouse models depict unique and favorable responses in PPBC tumors to immune modulation via COX-2 suppression[38,96] as well as checkpoint blockade[97].

The chief limitations in this study are the modest size of the NPBC and PPBC cohorts and the reliance on FFPE tissues. As recently highlighted, both of these limitations are predicated on the lack of well-annotated clinical data in YWBC, including time since last pregnancy, as well as the relative rarity of YWBC and PPBC[98]. A further limitation is that the immune milieu profiling by mIHC was focused on a small subset of T-cell activation and exhaustion markers. Future studies are needed to better understand the complexity of the immune milieu in YWBC in general, and in PPBC specifically. Studies of PPBC utilizing fresh, and therefore potentially more informative specimens, necessitate multi-institutional coordination, a worthy objective given the poor prognosis of this disease.

This study utilized an extensive chart review of a single breast cancer repository, spanning 15 years of samples, to build a rigorously controlled FFPE cohort of YWBC with known reproductive histories. This approach demonstrated that ER+ breast cancer in the background of recent childbirth is a molecularly distinct, poor prognostic subtype. This study serves as a molecular anchor point, aligned with extensive epidemiological data, which can support future studies focused on the utilization of fresh samples and larger cohorts. Such studies will undoubtedly provide further insights into the interactions between reproductive history, breast cancer biology, and YWBC patient outcomes, with the potential to improve clinical practice and patient outcomes.

Methods

Ethics approval and consent. The research was conducted on archived FFPE tissues samples collected under IRB-approved protocols at the Kaiser Permanente Northwest Center for Health Research (KPNW IRB) and the Oregon Health & Science University (OHSU IRB). These tissue archives are comprised of clinical samples obtained from women with invasive cancer who were receiving standard of care treatment. The study was retrospective, entailing the use of routinely collected data and archival invasive breast disease tissue and therefore granted a waiver of informed consent by the participating IRBs. All data were fully anonymized before

access by the researchers, labeled only with study-specific identifiers at all points, and the study was approved by the Committee on Clinical Investigations of the OHSU and by the Kaiser Permanente Northwest Biospecimen Review Committee.

Sample description. Archival FFPE breast cancer tissues ($n = 40$) were from primary breast cancers of premenopausal women aged 21–45. Inclusion criteria for the cases section were based on age at cancer diagnosis (≤45), parity status, body mass index (BMI), and availability of necessary clinical data and archived tissue specimens. The study was open to all races and ethnicities, however, based on study site demographics, the majority of the study population was white, nonhispanic, women (73%) (Supplementary Table 1a). Exclusion criteria included unknown time intervals from last childbirth, cases who were pregnant at breast cancer diagnosis, archived tissue specimens unavailable for research use, or from women who did not give consent for use of their tissue or clinical data for future research. As our study specifically used breast tissue that was naive for any treatment including neoadjuvant therapy, if that tissue was unavailable for research we excluded the case from the current study. Further, ER-negative cases ($n = 9$) and cases with DCIS without evidence of invasive cancer on the available tissue section ($n = 1$) were excluded from the current study. Using the above inclusion and exclusion criteria, the selected cases ($n = 30$) included women under the age of 45, ER-positive, who were either diagnosed with invasive breast cancer ≤4 years of last childbirth (PPBC) or were nulliparous (cases with spontaneous and/or elected abortions were excluded) based on reproductive history recorded in clinical charts (NPBC). The clinical characteristics of this cohort are shown in Supplementary Table 1a.

All archived H&E-stained slides from clinically indicated surgery were evaluated by a pathologist for each case. Blocks from slides with >80% tumor content were chosen for RNA extraction (10 μm sections), and sequential sections were used for immunohistochemical analysis (4 μm sections).

RNA isolation. Total RNA was extracted from freshly cut 10 μm FFPE sections using the miRNeasy FFPE kit (Qiagen, Valencia, CA) according to the manufacturer's protocol, using 1–4 sections (10–40 μm) per case[55]. RNA yield was determined by UV absorption on a NanoDrop 1000 spectrophotometer and fragment size was analyzed using the RNA 6000 Nano assay (Agilent Technologies, Santa Clara, CA) run on the 2100 Bioanalyzer. RNA quality was assessed using DV200 values. Of 40 cases meeting our inclusion criteria, 16 ER+ cases (PPBC ($n = 9$), NPBC ($n = 7$)) yielded RNA of quality (DV200 > 27%) needed to advance to RNA sequencing and in-depth RNA expression profiling. The tumor characteristics of these tumors are presented in Table 1b.

Library preparation and sequencing. An input of 75 ng of total FFPE derived RNA was used with the TruSeq RNA Access Library Prep Kit and was prepared according to manufacturer instructions (Illumina, San Diego, CA). Libraries were quantified by real-time PCR using KAPA Library Quantification kits (Kapa Biosystems, Wilmington, MA) on ABI StepOne thermocycler, pooled according to library method (three libraries per lane), and sequenced on a Hi-Seq 2500 (Illumina) using a 100 cycle, single-end protocol providing ~90 million reads per sample. Base call files were converted to fastq format using Bcl2Fastq (Illumina), as described[55].

RNA sequence alignment. All RNA Seq reads were aligned to the human reference genome (GRCh38, release 84) using STAR (version 2.5.2b)[99] with default parameters. The STAR "GeneCounts" module was used to quantify the number of reads mapping to each gene. We also used RSEM (version: v1.2.31) to quantify fragments per kilobase of transcript per million (FPKM) of the gene expressions.

Data processing and significance testing. Gene expressions quantified by read counts from STAR were used as input into DESeq2[100] for differential expression gene (DEG) analysis. Genes with counts per million (cpm) >0.05 in at least three cases for each group were kept (14,830 genes) for subsequent DEG analyses. DEG analysis was performed by comparing the PPBC cases and the nulliparous cases. The differentially expressed genes were called based on the FDR 0.01 and log two-fold change >1. In the DESeq2 package, counts were normalized using the variance stabilizing transformation (VST) module in DESeq2 for downstream analyses.

Breast cancer subtype prediction. All cases included in the study were designated as ER-positive as per clinical immunohistochemical evaluation. Using the PAM50® prediction parameters as described by Parker et al.[101], the tumor biologic subtypes (luminal A, luminal B, Basal, HER2) were predicted for these cases based upon gene expression values derived from whole-exome sequencing.

Mammaprint®, and PAM50® gene set heatmaps. To assess the ability of previously reported cancer gene sets to distinguished cohorts, VST transformed counts by DESeq2 were a subset for all expressed matching genes from the Mammaprint®, and PAM50® gene sets. Dendrograms were produced using hierarchical clustering of the z score transformed Euclidean average linkage distances through the Morpheus software package (https://software.broadinstitute.org/morpheus). For PAM50®, subtypes, as well as proliferation, ER, and HER2 scores, were generated using the original prediction parameters as described by Parker et al.[101].

Pseudo-oncotype Dx® score. Whole-exome sequencing derived rather than clinical diagnostic approved Oncotype Dx® scores (therefore pseudo) were calculated from reported gene expression values utilizing reported normalization equations[70,71,102]. Specifically, all group scores were determined by subtracting the average expression value across control genes (normalized reference) from each target gene value and adding a value of 10 to the difference and scores computed[102].

Gene set enrichment analysis. For GSEA[103] on PPBC compared with NPBC, GSEA version 4.0.3 was used to identify enriched gene sets from the Molecular Signature Database (MSigDB v 7.0, Hallmark, Collection 2, 3, 5–7), as well as 100 customized gene sets prepared from studies relevant to breast cancer, cancer immunity and normal breast biology[56,59,80,81,83,84,90,104–113] (Supplementary Data 01). Gene sets were considered to be enriched if their FDR q value was 0.05. Whole-exome gene expression array data from healthy nulliparous (NP, blue, n = 30) or healthy postpartum breast (PP, red, within 2 years of completed pregnancy, n = 10) tissues was obtained from a previous study (GEO Accession#GSE26457) normalized using Transcriptome Analysis Console software (TAC V.4.02, ThermoFisher Scientific). and used as input values for comparison of GSEA profiles (Supplementary Fig. 1).

Master regulator analysis. In order to infer the activities of transcription factors, we used the master regulator inference algorithm (MARINa)[114] compiled in R 'viper' package[115] to perform the regulon analyses on PPBC and NPBC samples. Two sources of data, gene expression signature, and regulatory network were required as model inputs. In this work, the Student's t test based statistic as suggested in viper manual was used as gene expression signatures. The regulon used for the transcription factor activity inference was curated from four databases[116]. The single sample-based regulon activities were inferred by function "viper", which is an extension of MARINa[114] and transforms a gene expression matrix to a regulatory protein activity matrix. For the model input, we used the FPKM quantification of PPBC and NPBC samples as the expression matrix and the same regulon network described above as the regulatory network.

Clonal entropy and reciprocally related clonal index analyses. We employed MiXCR (Version 3.0.12,MiLaboratory LLC) (Fig. 4d, e) to analyze TCR, which has an option to identify TCRs from standard RNA Seq. Specifically, MiXCR removed out-of-frame TCR sequences and identified unique V-CDR3 (nucleotide sequence)-J seed sequence and clustered identical sequences computing the frequency of each unique TCR clonotype. A number of TCR repertoire metrics were reported by summarizing the results from MiXCR, including the number of unique TCRs, normalized entropy, clonality index, and repertoire occupancy. Diversity was represented by normalized Shannon entropy (H) reflecting a quantitative measure of how many unique TCR clonotypes were present per sample, and simultaneously indicating how evenly they were distributed (p). For diversity measurement, the value of a diversity index increases when the number of unique TCR sequences increases and when evenness increases. For a given number of uniques, the value of a diversity index is maximized when all types of unique TCRs are equally abundant, and calculated using the default entropy function from the entropy R package using the formula: $H = -\sum_{k=1}^{n} f_k \times \ln(f_k)$, where n is the number of unique clonotypes in a sample, k represents a particular clonotype and f is the frequency of the kth clonotype. Clonality or Clonal index (C) reflects the inverse of the normalized Shannon's entropy H, a statistic for how much of the

repertoire is made up of expanded clones calculated by $C = 1 - H/\ln(n)$, where H is the Shannon entropy, and n is the number of unique clonotypes per sample.

CIBERSORT analysis. In order to estimate the abundances of immune cells from the bulk RNA Seq, we utilized CIBERSORT[63] to calculate the proportions of 22 human leukocyte cell subsets defined in the CIBERSORT package for each bulk RNA seq sample. Statistical significance of proportions of each immune cell type between NPBC and PPBC were determined using a two-tailed Student's t test with Welch's correction.

Compilation and analysis of YWBC cohort. In this study, YWBC data sets were collected from 8 studies and downloaded from the Gene Expression Omnibus (GEO) with the following accession number: GSE1992[72], GSE20624[73], GSE21653[74], GSE6532[75], GSE2990[79], GSE4922[76], GSE7390[77], and GSE19615[78]. The GEOquery and biomaRt R packages[117] were used to download the raw expression and meta data. YWBC was defined as a diagnosis at age less than 45, which resulted in 648 YWBC samples in total. The raw data sets with different Affymetrix platforms were merged together and the expressions of all data sets were corrected by ComBat R package to remove the underlying batch effects. The averaged expression profiles of microarray probe IDs that map to the same gene symbols were used to quantify the gene expressions for these 648 samples.

Multiplex IHC, Aperio, and ER quantification. Formalin-fixed, paraffin-embedded (FFPE) tissues were sectioned at 4 µm. Prior to staining, slides were baked for 2 h at 60 °C and then rehydrated through sequential immersion through xylene, graded alcohols, and water. Next, slides were antigen retrieved in a pressure cooker using DAKO Target Retrieval Solution (pH 6) at 125 °C for 5 min and then cyclically probed[50] in the following order with the indicated antibody, dilution and incubation times: Cycle 1 (PD-1, abcam, ab52587, Clone NAT105, 1:100, 1Hr), Cycle 2 (KI67, Thermofisher, RM-9106-S, Clone SP6, 1:300, 1Hr), Cycle 3 (TOX1, abcam, ab237009, Clone NAN448B, 1:800, overnight), Cycle 4 (P53, Thermofisher, MA5-12557, Clone DO-7, 1:100, 2Hr), Cycle 5 (Phospho-Histone H2A.X (Ser139), Cell Signaling, 9718, Clone 20E3, 1:250, 1Hr), Cycle 6 (CD8, BioSB, BSB5174, Clone C8/144B, 1:100, 1Hr), Cycle 7 (CD3, Dako, A0452, 1:400, overnight), Cycle 8 (CD45, Dako, M0701, Clones 2B11+ PD7/26, 1:300, 1Hr). Next, secondary anti-rabbit or anti-mouse Simple Stain MAX PO Histofine Peroxidase Polymer (Nichirei Biochemicals, 414144 or 414134) or anti-rat ImmPRESS Peroxidase Polymer (Vector Laboratories, MP-7444) antibodies were applied, followed by chromogenic detection with peroxidase substrate 3-amino-9-ethylcarbazole (AEC). The stained sections were scanned digitally using Aperio Image Scope AT2 (Leica Biosystems, CA, USA) at 20x magnification. For Aperio analysis, scanned images were visualized on Image scope software (v12.4.3) and the tissue sections were annotated for all tumor areas present per section followed by the semi-quantitative image analysis performed on the entire tumor area using Aperio deconvolution and nuclear algorithms (Leica Biosystems, CA, USA)[57,118]. Further, for the CD45, CD3, CD8, PD-1, and Tox multiplex IHC (mIHC) staining analysis on a per-cell basis, the pixel density of the scanned images necessitates region of interest (ROI) analysis, thus ~3–4 ROIs were selected per case where the immune cell infiltrate was high (based on H&E and CD45 staining review). Regions with high immune cell infiltrate were selected so that sufficient events needed to perform statistically supported single-cell analyses were captured. The selection of ROIs for each case was done by 2 analysts blinded to the reproductive status of the cases, with cases randomly sorted prior to ROI selection. Image processing, alignment of selected regions, and extraction of AEC signals was performed in MATLAB (V9.90.1592791) using the SURF algorithm in the Computer Vision Toolbox (The MathWorks, Inc) and FIJI as reported[55,119]. Pipeline for image processing and cell quantification was performed using FIJI (FIJI v 2.1), CellProfiler Version 4.1.3, and FCS Express Image Cytometry RUO (7.06.0015, De Novo Software, Glendale, CA)[120]. ER staining (ER, Novocastra, NCL-L-ER-6F11, 1:200, 1Hr) and pathological assessment for the intensity and % positive tumor cells for assigning an overall percent positive staining was done by a pathologist. The % positive ER results were independently confirmed by a second observer blinded to the study group. mIHC evaluation was carried out for all cases which passed multiple image alignment and segmentation quality control evaluations (n = 13 NPBC and n = 14 PPBC cases).

Statistics and reproducibility. Statistical significance determined by p values were generated by GraphPad-Prism Software (V9.2.0) (GraphPad Software, San Diego, California USA) unless otherwise stated and was performed as Students' unpaired two-tailed t test with Welch correction (*p ≤ 0.05), or Students' unpaired one-tailed t test with Welch correction for apriori directionality in confirmatory IHC. Survival curves were also plotted with GraphPad-Prism Software and p values reported from log-rank (Mantel–Cox) evaluation and log-rank Hazard Ratios (HR) reported.

To preserve the precious human clinical samples the mIHC staining was conducted once. With each staining run, human breast cancer and tonsil tissue samples were used as negative and positive controls for technical validation of staining and standardization of analysis across cases.

RNA library preparation and sequencing were carried out once per case utilizing previously established methodologies which demonstrated reproducibility

of a single technical replicate through evaluation of isolation and sequencing replicates[55].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The RNA-derived sequencing data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE158854. The publicly available RNA expression data from healthy nulliparous and postpartum breast tissues used in this study are available in the GEO database under accession code GSE26457. The publicly available outcomes data based upon RNA expression profiling used in this study as a YWBC cohort are available from the GEO database under accession codes, GSE1992, GSE20624, GSE21653, GSE6532, GSE2990, GSE4922, GSE7390, and GSE19615. All numerical data used in generating plots of figures are available as Source Data. All remaining data are available within the Article, Supplementary Information, or Source Data files. Source data are provided with this paper.

References

1. Sant, M. et al. Survival and age at diagnosis of breast cancer in a population-based cancer registry. *Eur. J. Cancer* **27**, 981–984 (1991).
2. Anderson, W. F., Pfeiffer, R. M., Dores, G. M. & Sherman, M. E. Comparison of age distribution patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiol. Biomark. Prev.* **15**, 1899–1905 (2006).
3. Matsuno, R. K. et al. Early- and late-onset breast cancer types among women in the United States and Japan. *Cancer Epidemiol. Biomark. Prev.* **16**, 1437–1442 (2007).
4. Allott, E. H. et al. Bimodal age distribution at diagnosis in breast cancer persists across molecular and genomic classifications. *Breast Cancer Res. Treat.* **179**, 185–195 (2020).
5. Dickens, C. et al. Investigation of breast cancer sub-populations in black and white women in South Africa. *Breast Cancer Res. Treat.* **160**, 531–537 (2016).
6. *Cancer Stat Facts: Female Breast Cancer*, https://seer.cancer.gov/statfacts/html/breast.html (2013–2017).
7. Chelmow, D. et al. Executive summary of the early-onset breast cancer evidence review conference. *Obstet. Gynecol.* **135**, 1457–1478 (2020).
8. Borges, V. F., Lyons, T. R., Germain, D. & Schedin, P. Postpartum involution and cancer: an opportunity for targeted breast cancer prevention and treatments? *Cancer Res.* **80**, 1790–1798 (2020).
9. Hadi M. A., Al Madani R., Abu Arida L. & B., A. G. Breast cancer age in developing countries: the narrowing gap. *Clin Surg.* **3**, 2074 (2018).
10. Heer, E. et al. Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *Lancet Glob. Health* **8**, e1027–e1037 (2020).
11. Merlo, D. F. et al. Breast cancer incidence trends in European women aged 20–39 years at diagnosis. *Breast Cancer Res. Treat.* **134**, 363–370 (2012).
12. Keramatinia, A., Mousavi-Jarrahi, S. H., Hiteh, M. & Mosavi-Jarrahi, A. Trends in incidence of breast cancer among women under 40 in Asia. *Asian Pac. J. Cancer Prev.* **15**, 1387–1390 (2014).
13. Thomas, A. et al. Incidence and survival among young women with stage i-iii breast cancer: SEER 2000–2015. *JNCI Cancer Spectr.* **3**, pkz040 (2019).
14. Lima, S. M., Kehm, R. D., Swett, K., Gonsalves, L. & Terry, M. B. Trends in parity and breast cancer incidence in US women younger than 40 years from 1935 to 2015. *JAMA Netw. Open* **3**, e200929 (2020).
15. Anders, C. K. et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J. Clin. Oncol.* **26**, 3324–3330 (2008).
16. Gnerlich, J. L. et al. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. *J. Am. Coll. Surg.* **208**, 341–347 (2009).
17. Anders, C. K., Johnson, R., Litton, J., Phillips, M. & Bleyer, A. Breast cancer before age 40 years. *Semin. Oncol.* **36**, 237–249 (2009).
18. Azim, H. A. Jr. et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clin. Cancer Res.* **18**, 1341–1351 (2012).
19. Fredholm, H. et al. Breast cancer in young women: poor survival despite intensive treatment. *PLoS ONE* **4**, e7695 (2009).
20. Bharat, A., Aft, R. L., Gao, F. & Margenthaler, J. A. Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. *J. Surg. Oncol.* **100**, 248–251 (2009).
21. Copson, E. et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18–40 years at diagnosis: the POSH study. *J. Natl. Cancer Inst.* **105**, 978–988 (2013).
22. Ademuyiwa, F. O. et al. Time-trends in survival in young women with breast cancer in a SEER population-based study. *Breast Cancer Res. Treat.* **138**, 241–248 (2013).
23. Partridge, A. H. et al. Subtype-dependent relationship between young age at diagnosis and breast cancer survival. *J. Clin. Oncol.* **34**, 3308–3314 (2016).
24. Lian, W. et al. The impact of young age for prognosis by subtype in women with early breast cancer. *Sci. Rep.* **7**, 11625 (2017).
25. MacMahon, B., Cole, P. & Brown, J. Etiology of human breast cancer: a review. *J. Natl. Cancer Inst.* **50**, 21–42 (1973).
26. Pike, M. C. et al. The hormonal basis of breast cancer. *Natl. Cancer Inst. Monogr.* 187–193 (1979).
27. Woods, K. L., Smith, S. R. & Morrison, J. M. Parity and breast cancer: evidence of a dual effect. *Br. Med. J.* **281**, 419–421 (1980).
28. Rosner, B., Colditz, G. A. & Willett, W. C. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. *Am. J. Epidemiol.* **139**, 819–835 (1994).
29. Ambrosone, C. B. et al. Parity and breastfeeding among African-American women: differential effects on breast cancer risk by estrogen receptor status in the Women's Circle of Health Study. *Cancer Causes Control* **25**, 259–265 (2014).
30. Nindrea, R. D., Aryandono, T. & Lazuardi, L. Breast cancer risk from modifiable and non-modifiable risk factors among women in southeast asia: a meta-analysis. *Asian Pac. J. Cancer Prev.* **18**, 3201–3206 (2017).
31. Hartman, E. K. & Eslick, G. D. The prognosis of women diagnosed with breast cancer before, during and after pregnancy: a meta-analysis. *Breast Cancer Res. Treat.* **160**, 347–360 (2016).
32. Lyons, T. R., Schedin, P. J. & Borges, V. F. Pregnancy and breast cancer: when they collide. *J. Mammary Gland Biol. Neoplasia* **14**, 87–98 (2009).
33. Callihan, E. B. et al. Postpartum diagnosis demonstrates a high risk for metastasis and merits an expanded definition of pregnancy-associated breast cancer. *Breast Cancer Res. Treat.* **138**, 549–559 (2013).
34. Goddard, E. T. et al. Association between postpartum breast cancer diagnosis and metastasis and the clinical features underlying risk. *JAMA Netw. Open* **2**, e186997 (2019).
35. Amant, F. et al. Prognosis of women with primary breast cancer diagnosed during pregnancy: results from an international collaborative study. *J. Clin. Oncol.* **31**, 2532–2539 (2013).
36. Johansson, A. L., Andersson, T. M., Hsieh, C. C., Cnattingius, S. & Lambe, M. Increased mortality in women with breast cancer detected during pregnancy and different periods postpartum. *Cancer Epidemiol. Biomark. Prev.* **20**, 1865–1872 (2011).
37. Stensheim, H., Moller, B., van Dijk, T. & Fossa, S. D. Cause-specific survival for women diagnosed with cancer during pregnancy or lactation: a registry-based cohort study. *J. Clin. Oncol.* **27**, 45–51 (2009).
38. Lyons, T. R. et al. Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. *Nat. Med.* **17**, 1109–1115 (2011).
39. McDaniel, S. M. et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. *Am. J. Pathol.* **168**, 608–620 (2006).
40. Martinson, H. A., Jindal, S., Durand-Rougely, C., Borges, V. F. & Schedin, P. Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. *Int J. Cancer* **136**, 1803–1813 (2015).
41. Bemis, L. T. & Schedin, P. Reproductive state of rat mammary gland stroma modulates human breast cancer cell migration and invasion. *Cancer Res.* **60**, 3414–3418 (2000).
42. Strange, R., Li, F., Saurer, S., Burkhardt, A. & Friis, R. R. Apoptotic cell death and tissue remodelling during mouse mammary gland involution. *Development* **115**, 49–58 (1992).
43. Werb, Z. et al. Extracellular matrix remodeling and the regulation of epithelial-stromal interactions during differentiation and involution. *Kidney Int. Suppl.* **54**, S68–S74 (1996).
44. Watson, C. J. & Kreuzaler, P. A. Remodeling mechanisms of the mammary gland during involution. *Int. J. Dev. Biol.* **55**, 757–762 (2011).
45. Betts, C. B. et al. Mucosal immunity in the female murine mammary gland. *J. Immunol.* **201**, 734–746 (2018).
46. Guo, Q. et al. Physiologically activated mammary fibroblasts promote postpartum mammary cancer. *JCI Insight* **2**, e89206 (2017).
47. Lyons, T. R. et al. Cyclooxygenase-2-dependent lymphangiogenesis promotes nodal metastasis of postpartum breast cancer. *J. Clin. Investig.* **124**, 3901–3912 (2014).
48. O'Brien, J. et al. Alternatively activated macrophages and collagen remodeling characterize the postpartum involuting mammary gland across species. *Am. J. Pathol.* **176**, 1241–1255, (2010).

49. Goddard, E. T. et al. Quantitative extracellular matrix proteomics to study mammary and liver tissue microenvironments. *Int J. Biochem. Cell Biol.* **81**, 223–232 (2016).

50. Pennock, N. D. et al. Ibuprofen supports macrophage differentiation, T cell recruitment, and tumor suppression in a model of postpartum breast cancer. *J. Immunother. Cancer* **6**, 98 (2018).

51. Yu, K. D., Wu, J., Shen, Z. Z. & Shao, Z. M. Hazard of breast cancer-specific mortality among women with estrogen receptor-positive breast cancer after five years from diagnosis: implication for extended endocrine therapy. *J. Clin. Endocrinol. Metab.* **97**, E2201–E2209 (2012).

52. Narod, S. A., Giannakeas, V. & Sopik, V. Time to death in breast cancer patients as an indicator of treatment response. *Breast Cancer Res. Treat.* **172**, 659–669 (2018).

53. Ademuyiwa, F. O. et al. US breast cancer mortality trends in young women according to race. *Cancer* **121**, 1469–1476 (2015).

54. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

55. Pennock, N. D. et al. RNA-seq from archival FFPE breast cancer samples: molecular pathway fidelity and novel discovery. *BMC Med. Genomics* **12**, 195 (2019).

56. Santucci-Pereira, J. et al. Genomic signature of parity in the breast of premenopausal women. *Breast Cancer Res.* **21**, 46 (2019).

57. Jindal, S., Narasimhan, J., Borges, V. F. & Schedin, P. Characterization of weaning-induced breast involution in women: implications for young women's breast cancer. *NPJ Breast Cancer* **6**, 55 (2020).

58. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

59. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, https://doi.org/10.1126/science.aaf8399 (2017).

60. Foroutan, M. et al. Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018).

61. Oakes, T. et al. Quantitative characterization of the T cell receptor repertoire of naive and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile. *Front Immunol.* **8**, 1267 (2017).

62. Vroman, H. et al. T cell receptor repertoire characteristics both before and following immunotherapy correlate with clinical response in mesothelioma. *J. Immunother. Cancer* **8**, https://doi.org/10.1136/jitc-2019-000251 (2020).

63. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

64. Crotty, S. T follicular helper cell differentiation, function, and roles in disease. *Immunity* **41**, 529–542 (2014).

65. Kim, K. et al. Single-cell transcriptome analysis reveals TOX as a promoting factor for T cell exhaustion and a predictor for anti-PD-1 responses in human cancer. *Genome Med.* **12**, 22 (2020).

66. Khan, O. et al. TOX transcriptionally and epigenetically programs CD8(+) T cell exhaustion. *Nature* **571**, 211–218 (2019).

67. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).

68. Mamounas, E. P. et al. Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. *J. Clin. Oncol.* **28**, 1677–1683 (2010).

69. Tang, G. et al. Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res. Treat.* **127**, 133–142 (2011).

70. Bernhardt, S. M. et al. Hormonal modulation of breast cancer gene expression: implications for intrinsic subtyping in premenopausal women. *Front Oncol.* **6**, 241 (2016).

71. Need, E. F. et al. The unique transcriptional response produced by concurrent estrogen and progesterone treatment in breast cancer cells results in upregulation of growth factor pathways and switching from a Luminal A to a Basal-like subtype. *BMC Cancer* **15**, 791 (2015).

72. Hu, Z. et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).

73. Anders, C. K. et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? *J. Clin. Oncol.* **29**, e18–e20 (2011).

74. Li, Y. et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat. Med.* **16**, 214–218 (2010).

75. Loi, S. et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.* **25**, 1239–1246 (2007).

76. Ivshina, A. V. et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66**, 10292–10301 (2006).

77. Desmedt, C. et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* **13**, 3207–3214 (2007).

78. Zhao, E. et al. Identification of a Six-lncRNA signature with prognostic value for breast cancer patients. *Front. Genet.* **11**, 673 (2020).

79. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272 (2006).

80. Clarkson, R. W. et al. The genes induced by signal transducer and activators of transcription (STAT)3 and STAT5 in mammary epithelial cells define the roles of these STATs in mammary development. *Mol. Endocrinol.* **20**, 675–685 (2006).

81. Stein, T., Salomonis, N., Nuyten, D. S., van de Vijver, M. J. & Gusterson, B. A. A mouse mammary gland involution mRNA signature identifies biological pathways potentially associated with breast cancer metastasis. *J. Mammary Gland Biol. Neoplasia* **14**, 99–116 (2009).

82. Hughes, K. & Watson, C. J. The multifaceted role of STAT3 in mammary gland involution and breast cancer. *Int. J. Mol. Sci.* **19**, https://doi.org/10.3390/ijms19061695 (2018).

83. Bambhroliya, A. et al. Gene set analysis of post-lactational mammary gland involution gene signatures in inflammatory and triple-negative breast cancer. *PLoS ONE* **13**, e0192689 (2018).

84. Asztalos, S. et al. Gene expression patterns in the human breast after pregnancy. *Cancer Prev. Res. (Philos.)* **3**, 301–311 (2010). [pii].

85. Fadok, V. A. et al. Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-beta, PGE2, and PAF. *J. Clin. Investig.* **101**, 890–898 (1998).

86. Stanford, J. C. et al. Efferocytosis produces a prometastatic landscape during postpartum mammary gland involution. *J. Clin. Investig.* **124**, 4737–4752 (2014).

87. Eischen, C. M. Genome stability requires p53. *Cold Spring Harb. Perspect. Med.* **6**, https://doi.org/10.1101/cshperspect.a026096 (2016).

88. Jerry, D. J., Dickinson, E. S., Roberts, A. L. & Said, T. K. Regulation of apoptosis during mammary involution by the p53 tumor suppressor gene. *J. Dairy Sci.* **85**, 1103–1110 (2002).

89. Li, M., Hu, J., Heermeier, K., Hennighausen, L. & Furth, P. A. Apoptosis and remodeling of mammary gland tissue during involution proceeds through p53-independent pathways. *Cell Growth Differ.* **7**, 13–20 (1996).

90. Nguyen, B. et al. Imprint of parity and age at first pregnancy on the genomic landscape of subsequent breast cancer. *Breast Cancer Res.* **21**, 25 (2019).

91. Asztalos, S. et al. High incidence of triple negative breast cancers following pregnancy and an associated gene expression signature. *Springerplus* **4**, 710 (2015).

92. Santos, S. J., Haslam, S. Z. & Conrad, S. E. Estrogen and progesterone are critical regulators of Stat5a expression in the mouse mammary gland. *Endocrinology* **149**, 329–338 (2008).

93. Vafaizadeh, V. et al. Mammary epithelial reconstitution with gene-modified stem cells assigns roles to Stat5 in luminal alveolar cell fate decisions, differentiation, involution, and mammary tumor formation. *Stem Cells* **28**, 928–938 (2010).

94. Miller, I. et al. Ki67 is a graded rather than a binary marker of proliferation versus quiescence. *Cell Rep.* **24**, 1105–1112 e1105 (2018).

95. Gaglia, G. et al. Temporal and spatial topography of cell proliferation in cancer. *bioRxiv*, 2021.2005.2016.443704, https://doi.org/10.1101/2021.05.16.443704 (2021).

96. O'Brien, J. et al. Non-steroidal anti-inflammatory drugs target the pro-tumorigenic extracellular matrix of the postpartum mammary gland. *Int J. Dev. Biol.* **55**, 745–755 (2011).

97. Tamburini, B. A. J. et al. PD-1 blockade during post-partum involution reactivates the anti-tumor response and reduces lymphatic vessel density. *Front Immunol.* **10**, 1313 (2019).

98. Amant, F. et al. The definition of pregnancy-associated breast cancer is outdated and should no longer be used. *Lancet Oncol.* **22**, 753–754 (2021).

99. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

100. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

101. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).

102. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).

103. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

104. Winslow, S., Leandersson, K., Edsjo, A. & Larsson, C. Prognostic stromal gene signatures in breast cancer. *Breast Cancer Res.* **17**, 23 (2015).

105. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 e1236 (2018).

106. Mognol, G. P. et al. Exhaustion-associated regulatory regions in CD8(+) tumor-infiltrating T cells. *Proc. Natl. Acad. Sci. USA* **114**, E2776–E2785 (2017).

107. Seo, H. et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8(+) T cell exhaustion. *Proc. Natl. Acad. Sci. USA* **116**, 12410–12415 (2019).

108. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 e1624 (2017).

109. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

110. Chevrier, S. et al. An immune atlas of clear cell renal cell carcinoma. *Cell* **169**, 736–749 e718 (2017).

111. Philip, M. et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* **545**, 452–456 (2017).

112. Schietinger, A. et al. Tumor-specific T cell dysfunction is a dynamic antigen-driven differentiation program initiated early during tumorigenesis. *Immunity* **45**, 389–401 (2016).

113. Bengsch, B. et al. Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. *Immunity* **48**, 1029–1045 e1025 (2018).

114. Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).

115. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).

116. Robertson, A. G. et al. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer Cell* **33**, 151 (2018).

117. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

118. Jindal, S. et al. Postpartum breast involution reveals regression of secretory lobules mediated by tissue-remodeling. *Breast Cancer Res.* **16**, R31 (2014).

119. Michaelis, K. A. et al. The TLR7/8 agonist R848 remodels tumor and host responses to promote survival in pancreatic cancer. *Nat. Commun.* **10**, 4682 (2019).

120. Link, J. M. et al. Tumor-infiltrating leukocyte phenotypes distinguish outcomes in related patients with pancreatic adenocarcinoma. *JCO Precis. Oncol.* **5**, https://doi.org/10.1200/PO.20.00287 (2021).

Acknowledgements

Author contributions

V.F.B. and P.S. provided the study objectives. N.P., S.J., S.W., P.G., V.F.B., and P.S. designed the study. S.W. and S.J. interrogated the Kaiser Registry for case selection and clinical data set preparation. Z.X. and W.H. performed RNA Seq data alignment, normalization, and quality check. Z.X., N.P., and D.S. performed pathway analyses, and D.S. performed regulon analyses. S.J. and J.N. designed and performed mIHC experiments, J.N and M.O. performed image processing for mIHC. A.B. and M.O. performed image cytometry for single-cell quantification of mIHC data sets and data analysis. N.P., S.J., W.H., A.B., and D.S. generated figures. N.P., S.J., W.H., P.S., and Z.X. interpreted all results, and S. J., N.P., and P.S. composed the manuscript. All authors critically reviewed the manuscript. S.J., N.P., and D.S. contributed equally to the manuscript and without each of their unique contributions, the work could not have been accomplished, and thus share the first author position. P.S. and Z.X. contributed equally to the manuscript and are responsible for data integrity. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information