

RESEARCH ARTICLE

Open Access



CrossMark

# Imprint of parity and age at first pregnancy on the genomic landscape of subsequent breast cancer

Bastien Nguyen<sup>1,2\*</sup> , David Venet<sup>1</sup>, Matteo Lambertini<sup>1,5</sup>, Christine Desmedt<sup>1</sup>, Roberto Salgado<sup>1,3</sup>, Hugo Mark Horlings<sup>4</sup>, Françoise Rothé<sup>1†</sup> and Christos Sotiriou<sup>1†</sup>

## Abstract

**Background:** Although parity and age at first pregnancy are among the most known extrinsic factors that modulate breast cancer risk, their impact on the biology of subsequent breast cancer has never been explored in depth. Recent data suggest that pregnancy-induced tumor protection is different according to breast cancer subtypes, with parity and young age at first pregnancy being associated with a marked reduction in the risk of developing luminal subtype but not triple negative breast cancer. In this study, we investigated the imprint of parity and age at first pregnancy on the pattern of somatic mutations, somatic copy number alterations, transcriptomic profiles, and tumor immune microenvironment by assessing tumor-infiltrating lymphocytes (TILs) levels of subsequent breast cancer.

**Methods:** A total of 313 patients with primary breast cancer with available whole genome, RNA sequencing, and TILs data were included in this study. We used a multivariate analysis adjusted for age at diagnosis, pathological stage, molecular subtypes, and histological subtypes. We compared nulliparous vs. parous, late parous vs. early parous, and nulliparous vs. pregnancy-associated breast cancer (PABC) patients. Late and early parous patients were grouped by using the median age at first pregnancy. PABC was defined as patients diagnosed up to 10 years postpartum.

**Results:** Genomic alterations of breast cancer were associated with age at first pregnancy but not with parity status alone. Independently of clinicopathological features, early parous patients developed tumors characterized by a higher number of Indels ( $P_{adj} = 0.002$ ), a lower frequency of *CDH1* mutations (1.2% vs. 12.7%;  $P_{adj} = 0.013$ ), a higher frequency of *TP53* mutations (50% vs. 22.5%;  $P_{adj} = 0.010$ ), and *MYC* amplification (28% vs. 7%;  $P_{adj} = 0.008$ ). PABC were associated with increased TILs infiltration ( $P_{adj} = 0.0495$ ).

**Conclusions:** These findings highlight an unprecedented link between reproductive history and the genomic landscape of subsequent breast cancer. We further hypothesize that *TP53*-mutant premalignant lesions could be less susceptible to the protective effect of an early parity, which might explain the difference of parity-induced protection according to breast cancer subtypes. This work also advocates that reproductive history should be routinely collected in future large-scale genomic studies addressing the biology of female cancers.

**Keywords:** Breast cancer, Pregnancy, Genomics, SCNAs, Mutational landscape, Whole genome sequencing, PABC

\* Correspondence: [bastien.nguyen@gmail.com](mailto:bastien.nguyen@gmail.com)

†Françoise Rothé and Christos Sotiriou contributed equally to this work.

<sup>1</sup>Breast Cancer Translational Research Laboratory J.-C. Heuson, Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium

<sup>2</sup>Present Address: Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, USA

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

## Background

The effect of parity and age at first pregnancy on the risk of developing breast cancer has been well documented [1–5]. Parity is known to have a dual effect on breast cancer risk with an increased risk during 5 to 10 years after pregnancy, followed by a strong and life-long protective effect [1, 6]. This effect is strongly influenced by age at first pregnancy as pregnancy-induced tumor protection is more pronounced if first pregnancy has occurred early in life. Recent data suggest that pregnancy-induced tumor protection is different according to breast cancer subtypes, with parity and young age at first pregnancy being associated with a marked reduction in the risk of developing luminal subtype tumors [7–10].

Several studies have attempted to investigate the mechanisms underlying this phenomenon [11, 12]. However, although parity and age at first pregnancy are among the most known extrinsic factors that modulate breast cancer risk, their impact on the biology of breast cancer has never been explored in depth. We have previously observed that breast cancer diagnosed during pregnancy has different biology at the genomic levels [13]. In the present study, we used a systematic multivariate analysis to investigate the imprint of parity and age at first pregnancy on the pattern of somatic mutations, somatic copy number alterations (SCNAs), transcriptomic profiles, and tumor-infiltrating lymphocytes (TILs) levels in a series of 313 breast cancer patients with available whole genome, RNA sequencing, and TILs levels data.

## Methods

### Data acquisition

All analyses were performed on a publicly available dataset comprising 560 breast cancer patients referred to as BRCA560 [14]. Clinical data, sequencing coverage, and mutational load were obtained from Additional file 1: Table S1–S3 in that reference. Coding driver mutation events and the contribution of mutational signatures were obtained from Additional file 1: Table S14 and S21 in that reference. Raw count data from RNA sequencing were obtained from the authors. Results from HRDetect classifier were obtained from Additional file 1: Table S4 in reference [15].

### Patients selection

Eligible patients from BRCA560 were those with samples collected from primary tumor only (patients with local recurrence or metastasis samples were excluded,  $n = 8$ ) who had available information on parity. There were only two available HER2+ patients (both parous) in the transcriptomic analysis so we preferred to exclude them from this analysis. For each patient, we determined the

breast cancer intrinsic subtype (PAM50) using geneR/Bioconductor package [16]. Nulliparous patients were defined as women with breast cancer who had no full-term pregnancy at the time of breast cancer diagnosis. Parous patients were defined as women with breast cancer who had at least one full-term pregnancy at the time of breast cancer diagnosis. Early parous patients were defined as  $\leq 25$  years of age at first full-term pregnancy, while late parous patients were defined as  $> 25$  years of age at first full-term pregnancy. PABC patients were defined as women diagnosed not during pregnancy but up to 10 years after the first pregnancy. As referred to the publicly available BRCA560 dataset [14], the Internal Review Boards of each participating institution approved the collection and use of samples of all patients in this study.

### TIL evaluation

The percentage of TILs was independently evaluated by two pathologists (R.S. and H.M.H.) on hematoxylin and eosin slides using the International TILs Working Group 2014 methodology as described before [17]. There were 242 original samples with evaluable TILs from 239 patients. For the three patients with two samples, the arithmetic averages were obtained. We obtain a final set of 231 patients with primary tumor only (patients with local recurrence or metastasis samples only were excluded,  $n = 8$ ). TIL information for patients for which evaluation from only one pathologist was available were discarded ( $n = 3$ ). The TILs were scored in the context of an International Cancer Genome Consortium (ICGC) project on the immune characterization of this series, and the scoring is available for all institutions having the ICGC Data Access Compliance Office (DACO) approval. The data will be made broadly available when this global project will be finished or before upon request to the authors.

### Statistical analysis

Except for age at diagnosis that was considered as a continuous variable and therefore compared using the non-parametric Mann–Whitney  $U$  test, differences in other clinicopathological characteristics of breast cancer between groups were analyzed using the  $\chi^2$  test or the Fisher exact test when appropriate. All statistical tests comparing groups were done using the non-parametric Mann–Whitney  $U$  test and the  $\chi^2$  test or the Fisher exact test when appropriate for continuous and categorical variables, respectively. For the multivariate analysis, we used a linear and logistic regression to assess the independent association of continuous (log transformed) and categorical variables respectively with—parity (nulliparous vs parous) or—age at first pregnancy ( $\leq 25$  years vs.  $> 25$  years) controlling for age at diagnosis,

pathological stage, molecular subtypes by IHC, and histological subtypes. For WGS results, we also corrected for log-transformed sequence coverage of tumor and normal samples (continuous). All interaction and multivariate tests ( $P_{\text{adj}}$ ) were done using the analysis of variance to compare the models with and without the extra term. Because continuous variables contain zeros, the logarithmic transformation was applied as follows:  $\log_{10}(x + 1)$ . The Kruskal–Wallis test was used to test if *MYC* expression originates from the same population according to the status of *MYC/TP53* alterations. All correlations were calculated using the non-parametric Spearman's *rho* coefficient. All reported  $P$  values were two-tailed. Multiple testing correction was performed using the false discovery rate method (FDR) [18], and differences were considered significant when the FDR was  $< 0.05$ . All analyses were done in R software version 3.3.3 (available at [www.r-project.org](http://www.r-project.org)) and Bioconductor version 3.6. Differential expression analysis was performed with DESeq2 v.1.14.1 R/Bioconductor package [19] on raw count data (20,724 genes). Significantly differentially expressed genes were selected with a FDR of  $< 0.1$ , independent filtering was performed using default parameters to select a set of genes for multiple test correction which maximizes the number of adjusted  $P$  values less than a given critical value alpha (by default 0.1) and differentially expressed genes were identified by using the default cutoff of  $P$  value adjusted for multitesting  $< 0.1$ . We used gage v.2.24.0 R/Bioconductor package [20] to identify significantly enriched pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [21] and biological process from Gene Ontology with the log2FoldChange from DEseq2 results as input data.

## Results

### Association between clinicopathological variables, parity, and age at first pregnancy

From a publicly available dataset comprising 560 breast cancer patients [14], a total of 313 with available information on parity were included. We identified 264 (84.3%) parous and 49 (15.7%) nulliparous patients (Additional file 2: Figure S1). In the parous group, 153 patients (57.9%) had available information on age at first pregnancy (median of 25 years, range 16–46 years). Parous patients were divided into two groups: 82 early and 71 late parous patients by using the median age at first pregnancy as a cutoff value. All patients had available somatic mutations and SCNA data, 182 patients (58.1%) had available transcriptomic data, and 170 patients (54.3%) had information on TIL levels (Additional file 2: Figure S2).

Table 1 summarizes the clinicopathological features of patients. Compared to parous patients, nulliparous patients had larger tumors (tumor size  $> 2$  cm, 59.2% vs.

37.1%,  $P = 0.006$ ), higher frequency of lymph node positive disease (40.8% vs. 27.7%;  $P = 0.027$ ), and lower frequency of triple negative breast cancer (TNBC) (4.1% vs. 23.2%;  $P = 0.001$ ). Compared to early parous patients, late parous patients had a younger age at breast cancer diagnosis (median, 49 years; range, 28–81 years vs. median, 59 years; range, 34–81 years;  $P = 2.58 \times 10^{-5}$ ) and were more often pre-menopausal (45.3% vs. 20.9%;  $P = 0.005$ ). Late parous patients also had smaller tumors (tumor pathological size  $> 2$  cm, 40.8% vs. 51.2%,  $P = 0.012$ ), lower frequency of TNBC (19.7% vs. 39%,  $P = 0.026$ ), and higher frequency of lobular histological subtype (14.5% vs. 2.5%,  $P = 0.01$ ).

In the following sections, we investigated the imprint of parity and age at first pregnancy on breast cancer biology by using a systematic multivariate analysis adjusted for potential confounders, namely age at diagnosis, pathological stage, molecular subtypes by IHC, and histological subtypes.

### The influence of parity and age at first pregnancy on the mutational landscape of breast cancer

We first sought to investigate the imprint of parity and age at first pregnancy on somatic mutational load (Additional file 1: Table S1). There was no significant difference in the total number of substitutions (SNVs) according to parity nor age at first pregnancy ( $P_{\text{adj}} = 0.097$ ,  $P_{\text{adj}} = 0.075$ , respectively, Fig. 1a). There was no significant difference in the total number of insertions or deletions (Indels) according to parity ( $P_{\text{adj}} = 0.464$ , Fig. 1a). In contrary, compared to tumors from late parous patients, tumors from early parous patients were significantly associated with a higher Indels load ( $P_{\text{adj}} = 0.002$ , FDR = 0.007, Fig. 1a). There was no significant difference between the total number of rearrangements according to parity nor age at first pregnancy (Additional file 1: Table S1).

We next interrogated the influence of parity and age at first pregnancy on the frequency of mutations in breast cancer driver genes. Among the driver mutated genes, seven had at least one non-silent mutation with a frequency of  $> 5\%$  across the whole cohort (Additional file 1: Table S2). As expected, *PIK3CA* and *TP53* were the most frequently mutated genes (Fig. 1b). None of the driver mutated genes were associated with parity in the multivariate analysis. However, in the parous group, early age at first pregnancy was independently associated with higher frequency of *TP53* mutations (41/82 (50%) vs. 16/71 (22.5%);  $P_{\text{adj}} = 0.010$ ; FDR = 0.046, Fig. 1b) and lower frequency of *CDH1* mutations (1/82 (1.2%) vs. 9/71 (12.7%);  $P_{\text{adj}} = 0.013$ ; FDR = 0.046, Fig. 1b). Considering the distribution of *TP53* mutations type, early parous patients had a significantly higher frequency of truncating mutations as compared to late parous

**Table 1** Clinicopathological features of nulliparous and parous patients

N	Nulliparous	Parous	P	Early parous	Late parous	P
	49	264		82	71	
Age at diagnosis	54 (30–81)	55 (28–81)	0.796 <sup>a</sup>	59 (34–81)	49 (28–81)	$2.6 \times 10^{-5}$ <sup>a</sup>
Menopausal status						
Pre	13 (33.3%)	60 (30.3%)	0.85	14 (20.9%)	29 (45.3%)	0.0053
Post	26 (66.7%)	138 (69.7%)		53 (79.1%)	35 (54.7%)	
Stage						
I	7 (14.9%)	41 (15.9%)	0.019	6 (7.6%)	17 (24.3%)	0.039
II	20 (42.6%)	81 (31.4%)		33 (41.8%)	27 (38.6%)	
III	11 (23.4%)	30 (11.6%)		12 (15.2%)	10 (14.3%)	
IV	0 (0%)	1 (0.4%)		1 (1.3%)	0 (0%)	
Na	9 (19.1%)	105 (40.7%)		27 (34.2%)	16 (22.9%)	
pT						
Tx	9 (18.4%)	105 (39.8%)	0.0062	27 (32.9%)	16 (22.5%)	0.012
≤ 2 cm	11 (22.4%)	61 (23.1%)		13 (15.9%)	26 (36.6%)	
> 2 cm	29 (59.2%)	98 (37.1%)		42 (51.2%)	29 (40.8%)	
pN						
Nx	11 (22.4%)	112 (42.4%)	0.027	30 (36.6%)	17 (23.9%)	0.2
N0	18 (36.7%)	79 (29.9%)		25 (30.5%)	23 (32.4%)	
N1+	20 (40.8%)	73 (27.7%)		27 (32.9%)	31 (43.7%)	
Grade						
I	6 (14%)	29 (12.7%)	0.93	8 (9.8%)	5 (7%)	0.059
II	17 (39.5%)	86 (37.7%)		25 (30.5%)	35 (49.3%)	
III	20 (46.5%)	113 (49.6%)		49 (59.8%)	31 (43.7%)	
Subtype by IHC						
Lum A-like	22 (44.9%)	106 (40.3%)	0.0013	29 (35.4%)	39 (54.9%)	0.026
Lum B-like	19 (38.8%)	54 (20.5%)		20 (24.4%)	17 (23.9%)	
HER2+/HR+	6 (12.2%)	29 (11%)		0 (0%)	0 (0%)	
HER2+/HR-	0 (0%)	13 (4.9%)		1 (1.2%)	1 (1.4%)	
TNBC	2 (4.1%)	61 (23.2%)		32 (39%)	14 (19.7%)	
Histology						
Ductal	36 (76.6%)	203 (81.2%)	0.67	71 (88.8%)	49 (71%)	0.01
Lobular	5 (10.6%)	23 (9.2%)		2 (2.5%)	10 (14.5%)	
Other	6 (12.8%)	24 (9.6%)		7 (8.8%)	10 (14.5%)	

pT pathological tumor size, pN pathological nodal status, HR hormone receptor, P P value derived from  $\chi^2$  test or the Fisher exact test when appropriate

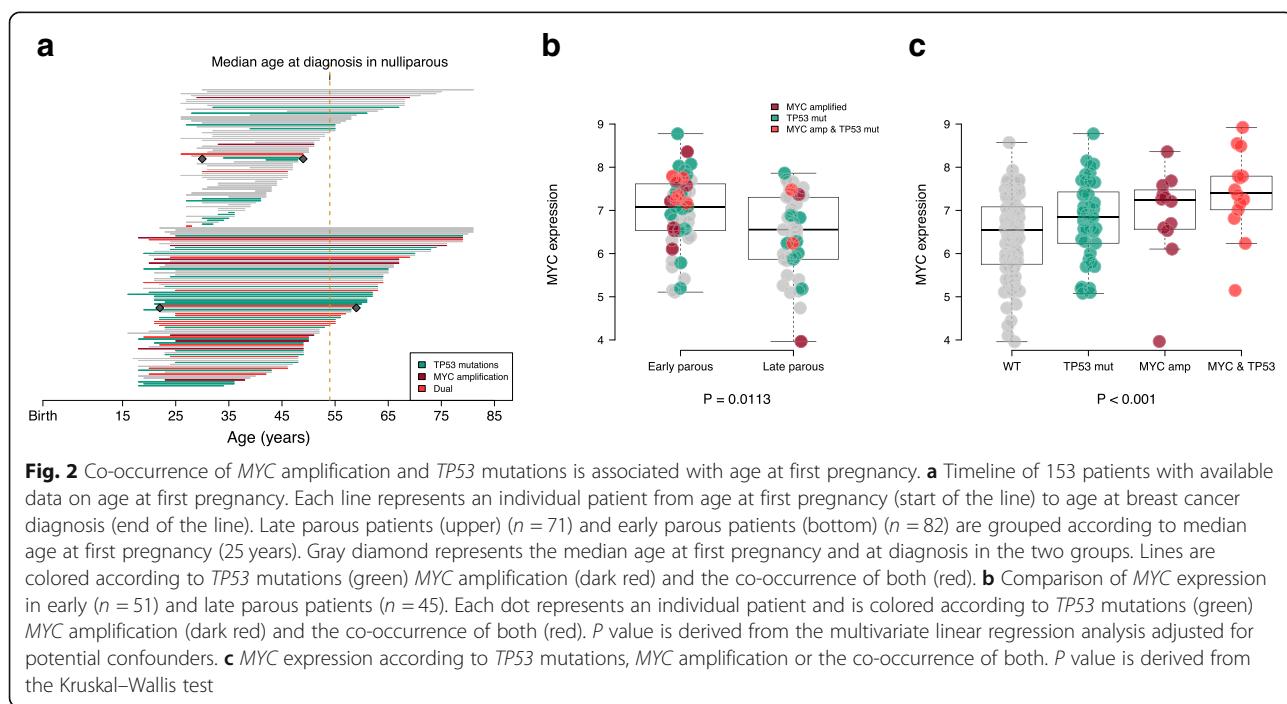
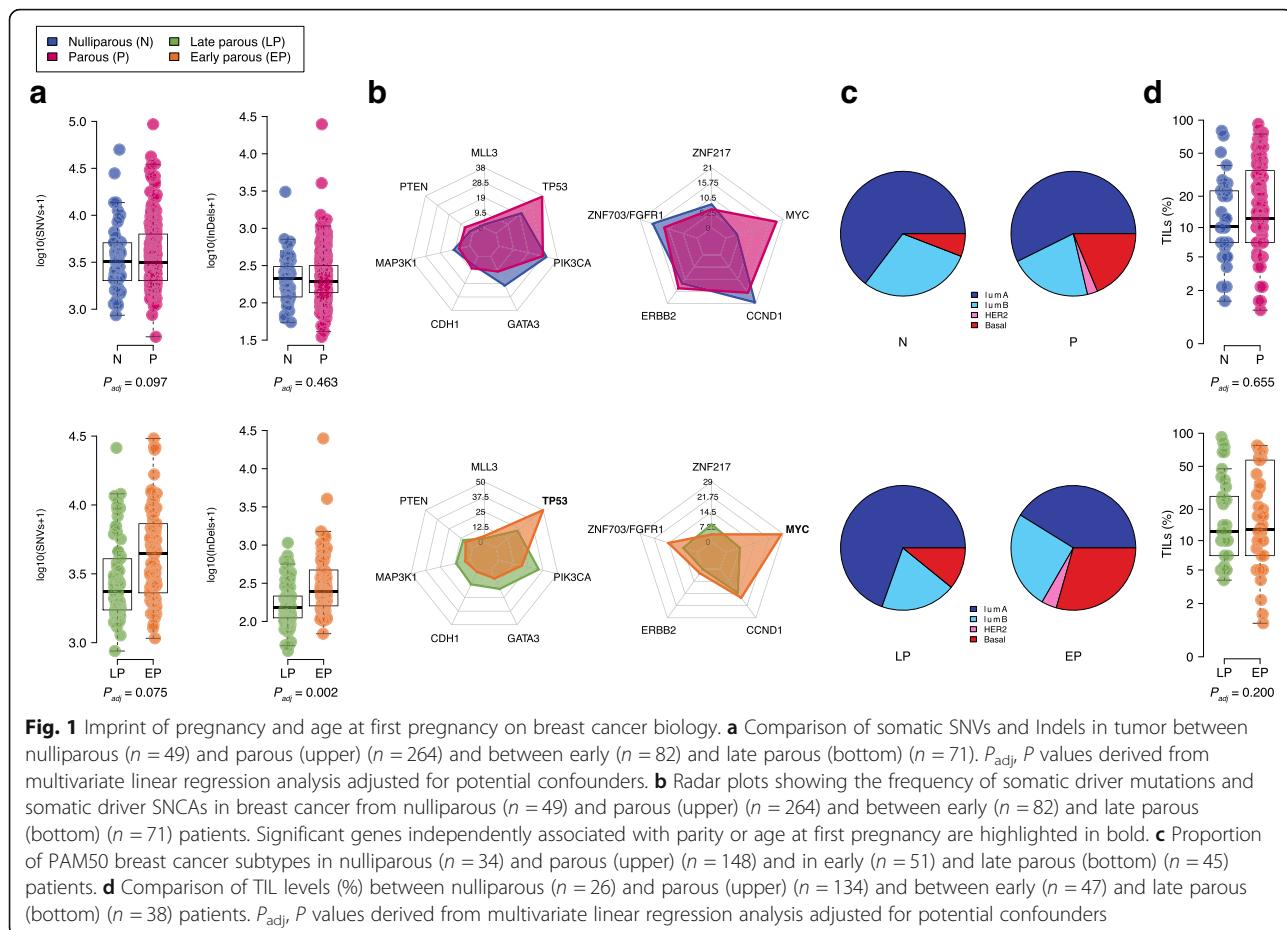
<sup>a</sup>Except continuous variable derived from Mann–Whitney U test

patients (21/82 (25.6%) vs. 5/71 (7%);  $P_{adj} = 0.014$ , Additional file 2 : Figure S3). Altogether, our results show that age at first pregnancy is associated with biological differences in the mutational landscape of subsequent breast tumors with early parity associated with higher Indels burden and higher frequency of deleterious mutations in *TP53* gene.

#### The influence of parity and age at first pregnancy on somatic copy number alterations

Somatic copy number alterations (SCNAs) play a major role in breast cancer biology [22, 23]. We identified five driver genes with a frequency of SCNAs > 5% across all patients (Additional file 1: Table S3). *MYC* tended to be

more frequently amplified in parous than in nulliparous patients (49/264 (18.6%) vs. 2/49 (4.1%);  $P_{adj} = 0.052$ ; FDR = 0.26, Fig. 1b). In the parous group, *MYC* amplification was significantly more frequent in the early parous group than in the late parous group (23/82 (28%) vs. 5/71 (7%);  $P_{adj} = 0.008$ ; FDR = 0.040, Fig. 1b). When evaluating the co-occurrence of SCNAs and somatic mutations, we found that co-occurrence of *MYC* amplification and *TP53* mutations was independently associated with age at first pregnancy, with early parous patients having a higher frequency of simultaneous alterations of *MYC* and *TP53* genes (15/82 (18.3%) vs. 3/71 (4.2%);  $P_{adj} = 0.087$ , Fig. 2a and Additional file 2: Figure S3). Taken together, these results suggest that age at first



pregnancy may also shape the somatic copy number alteration profiles of subsequent breast cancer.

### The influence of parity and age at first pregnancy on BRCA<sup>n</sup>ess

We investigated the BRCA<sup>n</sup>ess status of tumors according to reproductive history. Since germline BRCA1/2 mutation status was not available for all samples, we used HRDetect to identify *BRCA1/BRCA2*-deficient samples [15]. We did not find any significant differences in the proportion of *BRCA1/BRCA2*-deficient patients between nulliparous and parous groups (6/49 (12.2%) vs. 50/264 (18.9%), respectively,  $P_{adj} = 0.386$ ) nor between early and late parous group (25/82 (30.5%) vs. 14/71 (19.7%), respectively,  $P_{adj} = 0.473$ ). Thus, reproductive history and age at first pregnancy do not seem to affect homologous recombination DNA repair capacity in subsequent breast cancer.

### Integrative analysis of the genomic alterations and the transcriptomic profiles associated with parity and age at first pregnancy

RNA sequencing data were available for a subset of 182 patients, of which 34 were nulliparous (Additional file 2: Figure S2 and Additional file 1: Table S4). We first determined the intrinsic molecular subtype distribution of breast cancer using the PAM50 classifier [24]. We did not find a significant difference in the distribution of the PAM50 subtypes between nulliparous and parous (Fig. 1d). In contrast, early parous patients had a higher proportion of basal-like subtype tumors (15/51 (29.4%) vs. 4/45 (8.9%);  $P = 0.009$ , Fig. 1d, Additional file 1: Table S4). In order to identify de novo gene expression profiles that might be associated with parity and age at first pregnancy, we performed a multivariate differential expression analysis using DEseq2 [19] controlling for age at diagnosis, pathological stage, molecular subtypes by IHC, and histological subtypes. A total of 62 genes were differentially expressed between nulliparous and parous (Additional file 1: Table S5). Among these genes, three were associated with mammary development; *OXTR* and *ATP2B2* were downregulated whereas *NRG3* was upregulated in nulliparous. Pathway analysis using the generally applicable gene-set enrichment (GAGE) analysis [20] revealed an enrichment of genes related to extracellular matrix (ECM) receptor interaction in parous (Additional file 1: Table S6). When comparing early and late parous patients, 466 genes were differentially expressed, among which 305 were upregulated in early parous (Additional file 1: Table S7). However, pathway analysis did not reveal any significant enrichment of relevant biological processes (Additional file 1: Table S8).

Due to the higher frequency of *MYC* amplification in early parous patients, we determined if this would also

impact *MYC* at the mRNA expression levels. Early parous patients were independently associated with an upregulation of *MYC* expression ( $P_{adj} = 0.0113$ , Fig. 2b). We also evaluated the expression of *MYC* according to *TP53* mutations and *MYC* amplification and found that *MYC* expression was the highest in tumors harboring concurrent *TP53* mutation and *MYC* amplification (Fig. 2c).

### The influence of parity and age at first pregnancy on tumor immune microenvironment

Previous reports have hypothesized that the pregnancy-induced tumor protection could be attributable to an improved anti-tumor immunity [25–29]. Therefore, we assessed whether reproductive history could be associated with tumor-infiltrating lymphocyte (TIL) levels that is considered as a surrogate of tumor immunogenicity (Additional file 1: Table S9). We did not find any significant difference in the proportion of stromal TILs according to parity or according to the age at first pregnancy ( $P_{adj} = 0.655$ ;  $P_{adj} = 0.200$ , respectively, Fig. 2d). Similarly, no differences were observed when comparing intratumoral TILs according to parity or age at first pregnancy ( $P_{adj} = 0.240$ ;  $P_{adj} = 0.889$ , Additional file 1: Table S9). Thus, reproductive history does not seem to influence breast tumor immune microenvironment.

### Pregnancy-associated breast cancer is associated with increased TIL infiltration

Pregnancy-associated breast cancer (PABC) can be defined as cases diagnosed up to 10 years postpartum [30]. In this cohort, we identified 17 PABC patients and compared them with 49 nulliparous patients. Compared to nulliparous, PABC patients had a younger age at diagnosis (median, 38 years; range, 28–48 years vs. median, 54 years; range, 30–81 years;  $P = 5.79 \times 10^{-6}$ , Additional file 1: Table S10) and higher frequency of TNBC (5/17 (29.4%) vs. 2/49 (4.1%);  $P = 0.021$ , Additional file 1: Table S10). We did not find any significant differences between PABC and nulliparous in the pattern of somatic mutations or somatic copy number alterations (Additional file 1: Table S1-S3). Nine PABC had available gene expression and TIL scoring. At the transcriptomic level, we found that PABC patients were associated with enrichment of biological processes related to immune function (Fig. 3a and Additional file 1: Table S11). Moreover, PABC patients had an increased lymphocytic infiltration both for stromal and intratumoral TIL levels ( $P_{adj} = 0.040$  and  $P_{adj} < 0.0001$ , respectively; Fig. 3b, c). Taken together, these results indicate that cancer occurring in the postpartum breast is associated with increased TIL levels.

## Discussion

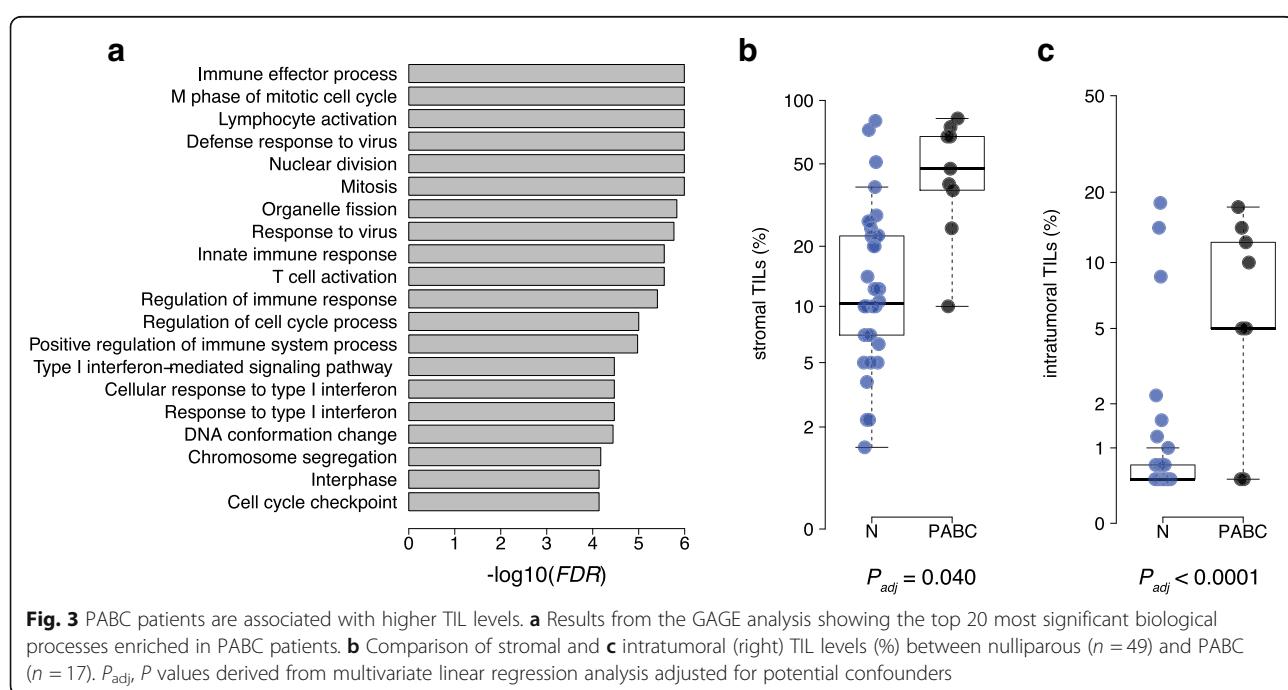
To our knowledge, this is the first study that explores the impact of reproductive history on the genomic landscape and the immune composition of subsequent breast cancer. While previous studies documented the risk of developing breast cancer according to reproductive history [1–5], this analysis provides further insights on the differences at the pathologic, genomic, transcriptomic, and immunologic levels according to prior parity and age at first pregnancy. Independently of clinicopathological features, our findings indicate that age at first pregnancy impacts the genomic makeup of subsequent breast cancer. Early parous patients developed tumors characterized by a higher number of Indels, a lower frequency of *CDH1* mutations, a higher frequency of *TP53* mutations and *MYC* amplification, while PABC patients exhibited higher TIL infiltration.

The higher proportion of TNBC in parous and particularly in early parous patients could be attributed to a differential effect of pregnancy-induced tumor protection according to breast cancer subtypes. We and others have documented that the pregnancy-induced tumor protection is different according to breast cancer subtypes with parity and young age at first pregnancy being associated with a marked reduction in the risk of developing luminal subtype [7–10].

Our study reveals that age at first pregnancy has a bigger imprint on genomic alterations of breast cancer than parity status alone. The apparent lack of impact of parity per se could be due to the relatively low number of nulliparous patients. An alternative explanation could be

that breast cancer from late parous resembles breast cancer from nulliparous women. Therefore, the lack of difference between nulliparous and parous could be related to the fact that the parous group encompasses late parous patients resulting in a possible dilution of the signal.

At the gene level, early parous patients had a higher frequency of *TP53* mutation, *MYC* amplification, and a lower frequency of *CDH1*. Interestingly, the co-occurrence of *TP53* mutations and *MYC* amplification was independently associated with age at first pregnancy, while the proportion of truncating *TP53* mutations was higher in early parous patients. We observed that tumors harboring concurrent *MYC* amplification and *TP53* mutation had the highest *MYC* expression. This observation is in line with a recent investigation of the *MYC* oncogene in pan-cancer data [31]. Previous reports have suggested that *TP53* mutations are a common mechanism that disturbs the apoptotic pathway in *MYC*-driven tumors [32]. It has been hypothesized that overexpression of *MYC* induces *TP53*-dependent apoptosis, and, as a consequence, *MYC*-driven tumors often require dysregulation of the apoptotic pathway to promote proliferation [33]. *TP53* has long been recognized as a potential mediator of pregnancy-induced resistance to mammary carcinogenesis. It has been shown in mice that p53 and its downstream transcriptional target p21 are increased in parous and estrogen/progesterone-treated mammary epithelium in response to a carcinogen [34]. In the absence of p53, the protection given by parity or exogenous hormones is



**Fig. 3** PABC patients are associated with higher TIL levels. **a** Results from the GAGE analysis showing the top 20 most significant biological processes enriched in PABC patients. **b** Comparison of stromal and **c** intratumoral (right) TIL levels (%) between nulliparous ( $n = 49$ ) and PABC ( $n = 17$ ).  $P_{adj}$ ,  $P$  values derived from multivariate linear regression analysis adjusted for potential confounders

lost [35–37]. We hypothesize that the higher frequency of *TP53*-mutant breast cancer observed in early parous women could be explained by the fact that an early pregnancy could protect less effectively against *TP53* mutant pre-malignant lesions (Fig. 4). In breast cancer, *TP53* mutations are highly linked to molecular subtype with a frequency of 80% in basal-like compared to 26% in luminal tumors [23]. The lower susceptibility of *TP53*-mutant premalignant lesions to the protective effect of an early parity might be the underlying cause of the known differential effect of parity-induced protection according to tumor subtypes.

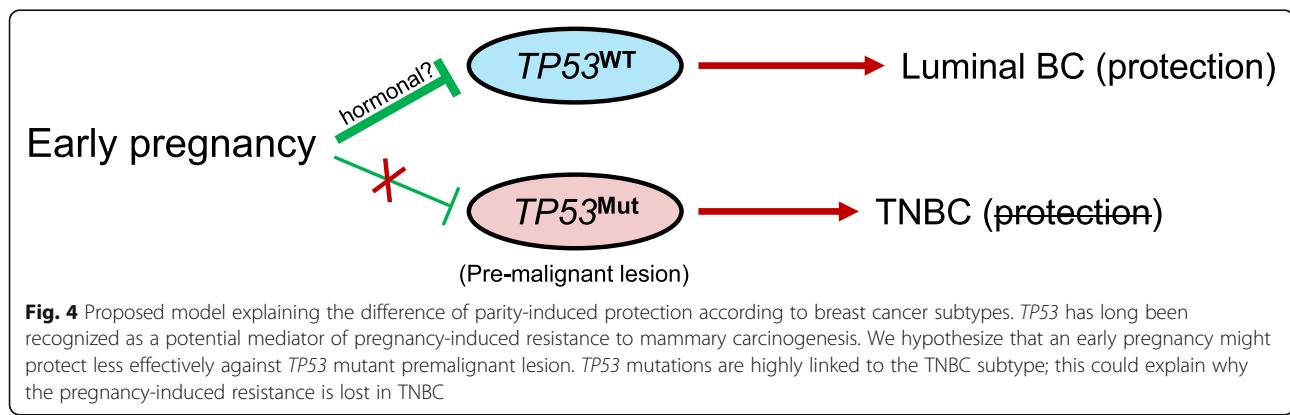
*CDH1* mutations have been associated with invasive lobular breast cancer subtype [38]. As the multivariate analysis was adjusted for histological subtypes, the lower frequency of *CDH1* mutations observed in early parous patients cannot be explained by differences in histological subtypes. Finally, in line with prior studies [39, 40], our data do not show an impact of parity and age at first birth on BRCA-related breast cancer risk.

Another hypothesis underlying the protection associated with an early pregnancy argues that the high level of circulating hormones associated with pregnancy would induce differentiation of the mammary gland while decreasing the tumorigenic potential of breast cells [41]. Gene expression studies comparing nulliparous and parous normal mammary tissue, from both rat and human, have observed upregulation of genes related to cell differentiation in the parous breast [42, 43]. According to this theory, a full-term pregnancy early in reproductive life would induce a molecular switch in mammary stem cells leading to a permanent decrease in their proliferation potential and resistance to oncogenic transformation [11, 44]. The results from our bulk RNA-seq analysis of tumor samples are not in line with this hypothesis, but we believe that future studies using single-cell RNA-seq would improve our understanding of the mechanism of the parity-induced protection against breast cancer. Noteworthy, we found that the expression of *OXTR* was higher in parous compared to

nulliparous patients. In the normal tissue of parous women, the gene encoding oxytocin receptor (*OXTR*) is physiologically upregulated during lactation and has been shown to remain overexpressed later in life [12, 45]. However, due to the lack of functional studies, it is not clear whether this gene is involved in breast cancer tumorigenesis or simply related to physiological changes induced by pregnancy. The enrichment of genes related to ECM receptor interaction in parous patients might be related to involution, a profound physiological change in the mammary gland after pregnancy. Right after breastfeeding the fully differentiated gland regresses to its pre-pregnant state by an innate tissue-remodeling mechanism. Evidence indicates that involution is mediated in part by ECM-degrading proteinases, leading to basement membrane degradation and subsequent apoptosis of the unwanted secretory epithelial cells [46, 47]. The exact role of the enrichment of genes related to ECM in parous patients on human breast cancer biology has still to be determined, but involution, that shares similarities with inflammation and wound healing programs, has been shown to promote breast cancer progression and metastasis in several animal models [46, 48].

Finally, previous reports have hypothesized that the pregnancy-induced tumor protection could be attributable to an improved anti-tumor immunity [25–29]. Our analysis reveals no differences in TILs infiltration levels according to parity or age at first pregnancy. The existence of a more complex immune component related to reproductive history cannot be excluded, but it is not supported by our analysis. Previous reports have documented that the immune milieu of the postpartum mammary gland associated with involution could promote tumorigenesis [49–51]. We observed an increase of TILs levels in PABC patients but the composition of the immune infiltrate has still to be determined to validate this hypothesis.

A potential limitation of our study is the lack of data on other reproductive factors (e.g., breastfeeding, age at menarche, and time since last pregnancy) that could also potentially imprint the biology of breast cancer. Indeed,



breastfeeding and age at menarche have been linked to breast cancer risk but since they are often self-reported, they are more difficult to assess reliably [7]. The lack of validation using an independent cohort is another important limitation. We retrieved the clinical information from the three independent cohorts of breast cancer with publicly available genomic data [23, 52, 53], but unfortunately, reproductive variables were missing, which precluded the validation of our findings. Another limitation is the absence of HER2+ subtype in the transcriptomic analysis.

## Conclusions

In conclusion, our findings highlight an unprecedented link between reproductive factors and the genomic landscape of subsequent breast cancer. We further hypothesize that *TP53*-mutant premalignant lesions could be less susceptible to the protective effect of an early parity, which might explain the difference of parity-induced protection according to breast cancer subtypes. Our results that need to be validated in other studies support that patients' reproductive history should be routinely collected in future large-scale genomic studies addressing the biology of female cancers.

## Additional files

**Additional file 1:** **Table S1.** Systematic multivariate analysis of mutational load comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients. **Table S2.** Systematic multivariate analysis of breast cancer driver SNVs comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients. **Table S3.** Systematic multivariate analysis of breast cancer driver SCNA comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients. **Table S4.** Clinicopathological features of nulliparous and parous patients included in the RNAseq analysis. **Table S5.** Differential expression analysis between nulliparous and parous patients using DEseq2 on raw counts data and controlling for age at diagnosis, pathological stage, molecular subtypes by IHC, histological subtypes. **Table S6.** Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between nulliparous and parous patients. **Table S7.** Differential expression analysis between early and late parous patients using DEseq2 on raw counts data and controlling for age at diagnosis, pathological stage, molecular subtypes by IHC, histological subtypes. **Table S8.** Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between early and late parous patients. **Table S9.** Systematic multivariate analysis of TILs levels comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients. **Table S10.** Clinicopathological features of nulliparous and PABC patients. **Table S11.** Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between nulliparous and PABC patients. (XLSX 366 kb)

**Additional file 2:** **Figure S1.** Flowchart summarizing the number of patients included in the analyses and the reasons for inclusion and exclusion. **Figure S2.** Venn diagram summarizing the number of patients with available data. Figure S3. Genomic landscape of breast cancer according to pregnancy and age at first pregnancy. (PDF 239 kb)

## Abbreviations

TILs: Tumor-infiltrating lymphocytes; PABC: Pregnancy-associated breast cancer; SCNA: Somatic copy number alterations; TNBC: Triple negative breast cancer; IHC: Immunohistochemistry; SNVs: Single-nucleotide variant;

pT: Pathological tumor size; pN: Pathological nodal status; HR: Hormone receptor; P: P value; FDR: False discovery rate

## Acknowledgements

Not applicable.

## Funding

BN and CS are supported by the Télévie and the Fonds National de la Recherche Scientifique (F.R.S.-FNRS). The work was supported by grants from the Breast Cancer Research Foundation (BCRF). ML is supported by an ESMO Translational Research Fellowship.

## Availability of data and materials

Raw data have been available at European-Genome Phenome Archive under the accession number EGAS00001001178. Somatic variants are available at International Cancer Genome Consortium Data Portal (<https://dcc.icgc.org/>).

## Authors' contributions

BN and CS contributed to the conceptualization. BN and DV contributed to the methodology. BN, and DV conducted the formal analysis. BN, RS, and HMH conducted the investigation. RS and HMH contributed to the resources. BN wrote the original draft of the manuscript. BN, DV, FR, CD, ML, and CS contributed in writing, reviewing, and editing of the manuscript. FR and CS did the funding acquisition. CD, FR, and CS did the supervision. All authors reviewed and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable. Since the BRCA560 dataset is publicly available, ethics committee approval was not needed. In addition, neither patient informed consent nor permission to use these data was required to perform this analysis.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Breast Cancer Translational Research Laboratory J.-C. Heuson, Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium. <sup>2</sup>Present Address: Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, USA. <sup>3</sup>Department of Pathology, GZA-ZNA, Antwerp, Belgium. <sup>4</sup>Department of Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. <sup>5</sup>Department of Medical Oncology, Clinica di Oncologia Medica, Ospedale Policlinico San Martino, & Department of Internal Medicine and Medical Specialties (DiMI), School of Medicine, University of Genova, Genova, Italy.

Received: 31 July 2018 Accepted: 31 January 2019

Published online: 15 February 2019

## References

- MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, et al. Age at first birth and breast cancer risk. *Bull World Health Organ.* 1970;43(2):209–21.
- Papatestas AE, Mulvihill M, Josi C, Ioannovich J, Lesnick G, Aufses AH. Parity and prognosis in breast cancer. *Cancer.* 1980;45(1):191–4.
- Trichopoulos D, Hsieh CC, MacMahon B, Lin TM, Lowe CR, Mirra AP, et al. Age at any birth and breast cancer risk. *Int J Cancer.* 1983;31(6):701–4.
- Kroman N, Wohlfahrt J, Andersen KW, Mouridsen HT, Westergaard T, Melbye M. Time since childbirth and prognosis in primary breast cancer: population based study. *BMJ.* 1997;315(7112):851–5.
- Rosenberg L, Thalib L, Adami H-O, Hall P. Childbirth and breast cancer prognosis. *Int J Cancer.* 2004;111(5):772–6.
- Albrektsen G, Heuch I, Hansen S, Kvåle G. Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. *Br J Cancer.* 2005;92(1):167–75.

7. Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: a systematic review and meta-analysis of epidemiological studies. *Cancer Treat Rev.* 2016;49:65–76.
8. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the breast cancer association consortium studies. *J Natl Cancer Inst.* 2011;103(3):250–63.
9. Ellingjord-Dale M, Vos L, Tretli S, Hofvind S, dos-Santos-Silva I, Ursin G. Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. *Breast Cancer Res.* 2017; 19(1):10.
10. Ritte R, Tikk K, Lukanova A, Tjønneland A, Olsen A, Overvad K, et al. Reproductive factors and risk of hormone receptor positive and negative breast cancer: a cohort study. *BMC Cancer.* 2013;13:1–12.
11. Meier-Abt F, Bentires-Alj M. How pregnancy at early age protects against breast cancer. *Trends Mol Med.* 2014;20(3):143–53.
12. Russo J, Santucci-Pereira J, De Cicco RL, Sheriff F, Russo PA, Peri S, et al. Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. *Int J Cancer.* 2012;131(5):1059–70.
13. Nguyen B, Venet D, Azim HA, Brown D, Desmedt C, Lambertini M, et al. Breast cancer diagnosed during pregnancy is associated with enrichment of non-silent mutations, mismatch repair deficiency signature and mucin mutations. *npj Breast Cancer.* 2018;4(1):23.
14. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* 2016;534(7605):1–20.
15. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med (in Press).* 2017.
16. Gendoo DMA, Ratanasirigulchai N, Schröder MS, Paré L, Parker JS, Prat A, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics.* 2016;32(7):1097–9.
17. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol.* 2015;26(2):259–71.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57:289–300.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
20. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics.* 2009;10(1):161.
21. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353–61.
22. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486(7403):346–52.
23. TCGA. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;487(7407):61–70.
24. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27(8):1160–7.
25. Erlebacher A. Mechanisms of T cell tolerance towards the allogeneic fetus. *Nat Rev Immunol.* 2013;13(1):23–33.
26. Agrawal B, Reddish MA, Krantz MJ, Longenecker BM. Does pregnancy immunize against breast cancer? *Cancer Res.* 1995;55(11):2257–61.
27. Finn OJ, Jerome KR, Henderson RA, Pecher G, Domenech N, Magarian-Blander J, et al. MUC-1 epithelial tumor mucin-based immunity and cancer vaccines. *Immunol Rev.* 1995;145:61–89.
28. Jungbluth AA, Silva WA, Iversen K, Frosina D, Zaidi B, Coplan K, et al. Expression of cancer-testis (CT) antigens in placenta. *Cancer Immun.* 2007;7: 15.
29. Arkle J, Taylor-Papadimitriou J, Bodmer W, Egan M, Millis R. Differentiation antigens expressed by epithelial cells in the lactating breast are also detectable in breast cancers. *Int J Cancer.* 1981;28(1):23–9.
30. Lyons TR, Schedin PJ, Borges VF. Pregnancy and breast cancer: when they collide. *J Mammary Gland Biol Neoplasia.* 2009;14(2):87–98.
31. Leiserson MDM, Vandin F, Wu HT, Raphael BJ. Reply: co-occurrence of MYC amplification and TP53 mutations in human cancer. *Nat Genet.* 2016;48(2):106–8.
32. Wolf E, Lin CY, Eilers M, Levens DL. Taming of the beast: shaping Myc-dependent amplification. *Trends Cell Biol.* 2015;25(4):241–8.
33. Hermeking H, Eick D. Mediation of c-Myc-induced apoptosis by p53. *Science.* 1994;265(5181):2091–3.
34. Sivaraman L, Connelly OM, Medina D, O'Malley BW. P53 is a potential mediator of pregnancy and hormone-induced resistance to mammary carcinogenesis. *PNAS.* 2001;98(22):12379–84.
35. Dunphy KA, Blackburn AC, Yan H, O'Connell LR, Jerry DJ. Estrogen and progesterone induce persistent increases in p53-dependent apoptosis and suppress mammary tumors in BALB/c-Trp53<sup>+/−</sup> mice. *Breast Cancer Res.* 2008;10(3):R43.
36. Medina D, Kittrell FS. p53 function is required for hormone-mediated protection of mouse mammary tumorigenesis. *Cancer Res.* 2003;63(19): 6140–3.
37. Jerry DJ, Kittrell FS, Kuperwasser C, Laucirica R, Dickinson ES, Bonilla PJ, et al. A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development. *Oncogene.* 2000;19(8):1052–8.
38. Desmedt C, Zoppoli G, Gundem G, Pruneri G, Larsimont D, Fornili M, et al. Genomic characterization of primary invasive lobular breast cancer. *J Clin Oncol.* 2016;34(16):1872–80.
39. Pan H, He Z, Ling L, Ding Q, Chen L, Zha X, et al. Reproductive factors and breast cancer risk among BRCA1 or BRCA2 mutation carriers: results from ten studies. *Cancer Epidemiol.* 2014;38(1):1–8.
40. Kotsopoulos J, Gronwald J, Lynch HT, Eisen A, Neuhausen SL, Tung N, et al. Age at first full-term birth and breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Res.* 2018;17(2):421–6.
41. Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility to carcinogenesis. *Breast Cancer Res Treat.* 1982 Mar;2(1):5–73.
42. Blakely CM, Stoddard AJ, Belka GK, Dugan KD, Notarfrancesco KL, Moody SE, et al. Hormone-induced protection against mammary tumorigenesis is conserved in multiple rat strains and identifies a core gene expression signature induced by pregnancy. *Cancer Res.* 2006;66(12):6421–31.
43. Russo J, Balogh GA, Russo IH. Full-term pregnancy induces a specific genomic signature in the human breast. *Cancer Epidemiol Biomark Prev.* 2008;17(1):51–66.
44. Medina D. Mammary developmental fate and breast cancer risk. *Endocr Relat Cancer.* 2005;12(3):483–95.
45. Peri S, de Cicco RL, Santucci-Pereira J, Slifker M, Ross EA, Russo IH, et al. Defining the genomic signature of the parous breast. *BMC Med Genet.* 2012;5:46.
46. McDaniel SM, Rumer KK, Biroc SL, Metz RP, Singh M, Porter W, et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. *Am J Pathol.* 2006;168(2):608–20.
47. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer.* 2006;6(4):281–91.
48. Lyons TR, O'Brien J, Borges VF, Conklin MW, Keely PJ, Elceirri KW, et al. Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. *Nat Med.* 2011;17(9):1109–15.
49. Martinson HA, Jindal S, Durand-Rougely C, Borges VF, Schedin P. Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. *Int J Cancer.* 2015;136(8):1803–13.
50. Fornetti J, Martinson HA, Betts CB, Lyons TR, Jindal S, Guo Q, et al. Mammary gland involution as an immunotherapeutic target for postpartum breast cancer. *J Mammary Gland Biol Neoplasia.* 2014;19(2):213–28.
51. Harvell DME, Kim J, O'Brien J, Tan A-C, Borges VF, Schedin P, et al. Genomic signatures of pregnancy-associated breast cancer epithelia and stroma and their regulation by estrogens and progesterone. *Horm Cancer.* 2013;4(3): 140–53.
52. Pereira B, Chin S-F, Rueda OM, Vollan H-KM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.
53. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell.* 2018;34(3):427–438.e6.