UNIVERSITÉ LIBRE DE BRUXELLES

Faculté de Médecine

Impact of reproductive history and pregnancy on breast cancer biology

Bastien NGUYEN

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences biomédicales et pharmaceutiques

Promoteur de thèse : Professeur Christos Sotiriou

Août 2018

REMERCIEMENTS

Je remercie toutes les personnes qui m'ont soutenu durant ces quatre années. En particulier :

Mon promoteur, le professeur Christos Sotiriou, qui m'a permis de mûrir en tant que scientifique et qui a su canaliser mon optimisme.

Hatem, qui n'a cessé de m'accompagner. J'espère lui avoir fait honneur en continuant ses projets.

Tous mes collègues du BCTL, Delphine, Ghizlane, Samira, Jeanne, Julie, Vinu, Floriane Yacine, François, Marion, Françoise, David, Laurence, Christine, Giulia et Matteo, qui m'ont appris énormément et qui ont participé à faire du labo un endroit chaleureux, joyeux, et amical.

Ma famille, pour son soutien et ses encouragements.

Claire, pour son amour et sa confiance en moi.

Les patients anonymes qui ont participé à cette thèse et sans qui rien n'aurait été possible.

ABSTRACT

It is estimated that four in five women will give birth while one in eight women will be diagnosed with breast cancer at some point in her lifetime. It is also known that pregnancy at a young age is associated with a marked decrease in the risk of breast cancer and that this protection is different according to breast cancer subtypes. This thesis explores the impact of reproductive history on breast cancer biology and provides the molecular characterization of breast cancer diagnosed during pregnancy. The last part investigates the effect of RANKL inhibition on the biology of breast cancer in young women.

In the first study, we investigated the impact of parity and age at first pregnancy on the clinicopathological features, the genomic and transcriptomic landscape, and the immune microenvironment of 313 breast cancers. For the first time, we highlighted a link between reproductive history and the genomic landscape of subsequent breast cancer. We demonstrated that, independently of clinicopathological features, age at first birth is associated with specific genomic alterations that could explain the differences in risk reduction associated with pregnancy according to breast cancer subtypes. This study represents a first step toward the recognition that reproductive factors matter in order to fully understand breast cancer biology and advocates that reproductive history should be routinely collected in future studies addressing the biology of breast cancer but also of other female cancers.

The second study is focused on the molecular characterization of breast cancer diagnosed during pregnancy (BCP). We conducted a comparative analysis of a unique cohort of BCP patients and non-pregnant control patients by integrating gene expression, copy number alterations, and whole-genome sequencing data. We showed that BCP has unique molecular characteristics including an enrichment of non-silent mutations, a higher frequency of mutations in mucin gene family and an enrichment of mismatch repair deficiency mutational signature. This provides important insights into the biology of BCP and suggests that these features may be implicated in promoting tumor progression during pregnancy. In addition, it provides an unprecedented resource for further understanding of the biology of breast cancer in young women and how pregnancy could modulate tumor biology.

In a previous study, the laboratory had reported up-regulation of RANKL in young and pregnant breast cancer patients. Therefore, in the last chapter, we investigated the biological effect of denosumab, a RANKL inhibitor, in a preoperative study including 27 young primary breast cancer patients. We demonstrated evidence that denosumab induces modulation of the tumor immune microenvironment with an increased level of tumor-infiltrating lymphocytes. This effect was likely due to upregulation of inflammatory cytokines and depletion of immunosuppressive regulatory T cells within the tumor microenvironment. These findings suggest a role for denosumab in reshaping the tumor immune microenvironment of breast cancer and that its use in combination could improve immunotherapy efficacy.

RÉSUMÉ

On estime que quatre femmes sur cinq donneront un jour naissance tandis qu'une femme sur huit sera diagnostiquée d'un cancer du sein au cours de sa vie. Il est également établi qu'une grossesse à un jeune âge est le facteur le plus fortement associé à une réduction du risque de cancer du sein et que cette protection est différente selon les sous-types. Cette thèse explore l'influence de l'histoire génésique sur la biologie du cancer du sein ainsi que la caractérisation moléculaire du cancer du sein durant la grossesse. Enfin, le troisième volet de cette thèse porte sur l'effet de l'inhibition de RANKL sur la biologie du cancer du sein chez les femmes jeunes.

Dans la première étude, nous avons étudié l'effet de la parité et de l'âge de la première grossesse sur les caractéristiques cliniques, le profil génomique et transcriptionnel ainsi que le microenvironnement immunitaire de 313 cancers du sein. Pour la première fois, nous avons démontré un lien entre l'histoire génésique et le profil génomique du cancer du sein. Nous avons ainsi observé qu'indépendamment des caractéristiques cliniques, l'âge de la première grossesse est associé à des altérations génomiques spécifiques pouvant expliquer les différences de réduction du risque de cancer du sein attribué à la parité en fonction du sous-type moléculaire. Cette étude est une première étape dans la reconnaissance du fait que les variables gynéco-obstétriques sont essentielles pour mieux comprendre la biologie du cancer du sein. Elle recommande également que les antécédents gynécologiques et obstétricaux soient collectés dans les futures études visant à caractériser la biologie des cancers féminins.

La seconde étude de cette thèse se concentre sur la caractérisation moléculaire des cancers du sein durant la grossesse (CSG). Nous avons mené une analyse comparative de patients atteints de CSG et de patients contrôle (atteints de cancers du sein non diagnostiqués durant la grossesse) en utilisant les dernières technologies de séquençage d'ADN à haut débit et en intégrant leur profil transcriptomique et génomique. Nous avons montré que les CSG ont des caractéristiques moléculaires uniques, associées à un enrichissement de mutations non-silencieuses, une plus grande fréquence de mutations dans les gènes codants pour les mucines et une signature mutationnelle propre à un défaut dans la machinerie de réparation de l'ADN. Ces résultats ont mis en évidence de nouveaux mécanismes spécifiquement impliqués dans la tumorigenèse des cancers du sein durant

la grossesse. De plus, cette étude constitue une nouvelle base de données unique permettant une meilleure compréhension de la biologie des cancers du sein chez la femme jeune.

Dans une étude précédente, le laboratoire a mis en évidence que RANKL était surexprimé dans les cancers du sein des femmes jeunes et enceintes. C'est pourquoi, nous avons étudié l'effet de l'inhibition de RANKL par le denosumab dans une étude préopératoire incluant 27 jeunes patientes atteintes d'un cancer du sein primaire. Nous avons mis en évidence qu'une exposition au denosumab induit une modulation du microenvironnement immunitaire accompagnée d'une augmentation de l'infiltration lymphocytaire tumorale. L'étude du microenvironnement de ces tumeurs a montré que cet effet est probablement dû à une régulation positive de certaines cytokines inflammatoires ainsi qu'une déplétion de lymphocytes T régulateurs immunosuppressifs. Ces résultats suggèrent que le denosumab joue un rôle dans le remodelage du microenvironnement immunitaire du cancer du sein et que son utilisation combinée à l'immunothérapie pourrait améliorer l'efficacité de ces traitements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ADDITIONAL DATA

Supporting materials that cannot be included in the printed version of this thesis for reasons of space are available online.

1 Supplementary Data File 1

Available online at: `https://doi.org/10.6084/m9.figshare.6934907`

- *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 SNVs comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients*
- *Table S3. Systematic multivariate analysis of breast cancer SCNAs comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients*
- *Table S4. Systematic multivariate analysis of breast cancer mutational signatures comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients*
- *Table S5. Clinicopathological features of nulliparous and parous patients included in the RNAseq analysis*
- *Table S6. 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 S7. Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between nulliparous and parous patients*
- *Table S8. 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 S9. Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between early and late parous patients*
- *Table S10. Systematic multivariate analysis of TILs levels comparing nulliparous vs. parous, early parous vs. late parous and PABC vs. nulliparous patients*
- *Table S11. Clinicopathological features of nulliparous and PABC patients*

- *Table S12. Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between nulliparous and PABC patients*

2 Supplementary Data File 2

Available online at: `https://doi.org/10.6084/m9.figshare.6934913`

- *Table S1. Individual patients data with complete clinicopathological features*
- *Table S2. Results from the CNAs analysis by segments comparing BCP and control and by molecular subgroups*
- *Table S3. Catalogues of every non-silent mutations found in the whole cohort annotated by SnpEff and VEP*

3 Supplementary Data File 3

Available online at: `https://figshare.com/s/53e12f942093378ca90d`

- *Table S1. All non-serious adverse events sorted by MedDRA SOC and MedDRA PT*
- *Table S2. Differentially expressed genes between pre- and post-treatment*
- *Table S3. Gene ontology enrichment analysis of differentially expressed genes between pre- and post-treatment*
- *Table S4. Pathway analysis using the generally applicable gene-set enrichment (GAGE) method to identify significantly enriched pathways between pre- and post-treatment samples*
- *Table S5. Potential predictive biomarkers at baseline associated with a TIME response induced by denosumab*
- *Table S6. Differentially expressed genes between non-responders and responders*
- *Table S7. Gene ontology enrichment analysis of differentially expressed genes between non-responders and responders*
- *Table S8. List of antibodies used for dual IHC*

LIST OF ABBREVIATIONS

AEs	Adverse events
BC	Breast cancer
BCP	Breast cancer diagnosed during pregnancy
CCF	Cancer cell fraction
CNAs	Copy number alterations
COSMIC	Catalogue of Somatic Mutations in Cancer
CSG	Cancers du sein durant la grossesse
DFS	Disease-free survival
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GAGE	Generally applicable gene-set enrichment
GalNAc	N-acetylgalactosamine
HE	Hematoxylin and eosin
HER2	Human epidermal growth factor receptor 2
HR	Hormone receptors
IDC	Invasive ductal carcinoma
IHC	Immunohistochemistry
IJB	Institut Jules Bordet
ILC	Invasive Lobular Carcinoma
Indels	Insertions and deletions
KEGG	Kyoto Encyclopaedia of Genes and Genomes
METABRIC	Molecular taxonomy of breast cancer international consortium
mIHC	Multiplex IHC
MMR	DNA mismatch repair
NMF	Non-negative matrix factorization
OS	Overall survival
PABC	Pregnancy-associated breast cancer
pN	Pathological nodal status
PR	Progesterone receptor
pT	Pathological tumor size
RT	Radiation therapy
Sig20	Signature 20
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SREs	Skeletal-related events
TCGA	The Cancer Genome Atlas
TCR	T cell receptor
TILs	Tumor-infiltrating lymphocytes
TNBC	Triple-negative breast cancer
TPM	Transcripts per million
Treg	Regulatory T cells
VAFs	Variant allele fractions
WEG	Whole exome sequencing
WGS	Whole genome sequencing

INTRODUCTION

1 Introduction and rationale

Breast cancer (BC) is the most frequently diagnosed cancer and the second leading cause of cancer death among women worldwide. In 2018, it is estimated that 266,120 women will be diagnosed with BC and 40,920 women will die from this disease in the US alone (Siegel et al., 2018). However, from 1989 to 2015, overall BC death rates have dropped by 39 percent (Siegel et al., 2018). The decline in BC mortality has been attributed to improvements in treatment (e.g., adjuvant chemotherapy and hormonal therapy in the 1980s, targeted therapy in the 1990s) and early detection by mammography.

Invasive breast cancer is a heterogeneous disease with distinct subtypes, defined mainly by the expression of hormone receptors (HR) and human epidermal growth factor receptor 2 (HER2) (i.e., HR-positive, HER2-positive and triple-negative subtypes) assessed by immunohistochemistry (IHC) and fluorescence or dual in situ hybridization in routine clinical practice. HR-positive/HER2-negative subtype can be further sub-classified into luminal A and luminal B based on PAM50 classifier, histological grade and proliferation markers such as Ki-67. These subtypes differ in their biology, prognosis, treatment strategies and pattern of metastasis.

Several international consensus guidelines have been published on the clinical management of early primary breast cancer (Coates et al., 2015; Senkus et al., 2015), advanced breast cancer (Cardoso et al., 2017), breast cancer in young women (Paluch-Shimon et al., 2017) and breast cancer during pregnancy (Loibl et al., 2015; Peccatori et al., 2013). Of note, the clinical management of breast cancer diagnosed during pregnancy is further detailed in section 5.3. Therapeutic options can be broadly distinguished into local therapies, such as surgery and radiation therapy, and systemic therapies, such as chemotherapy, hormone therapy, and targeted therapy. Systemic therapies can be used before surgery (neoadjuvant setting) to reduce tumor size and allow breast-conserving surgery and to assess *in vivo* sensitivity. Systemic therapies can be used after surgery (adjuvant setting), to eliminate micrometastases and reduce risks of recurrence. Prognosis and treatment decisions are individualized based on tumor stage, histological grade, HR status, and HER2 status. HR-positive BC patients benefit from endocrine therapy, and

HER2-positive BC patients benefit from targeted therapies such as trastuzumab and more recently pertuzumab and neratinib. Currently, the only systemic option for triple-negative subtypes remains chemotherapy (Figure 0.1). We refer the reader to the above-mentioned clinical guidelines for a detailed review of the current management of breast cancer.

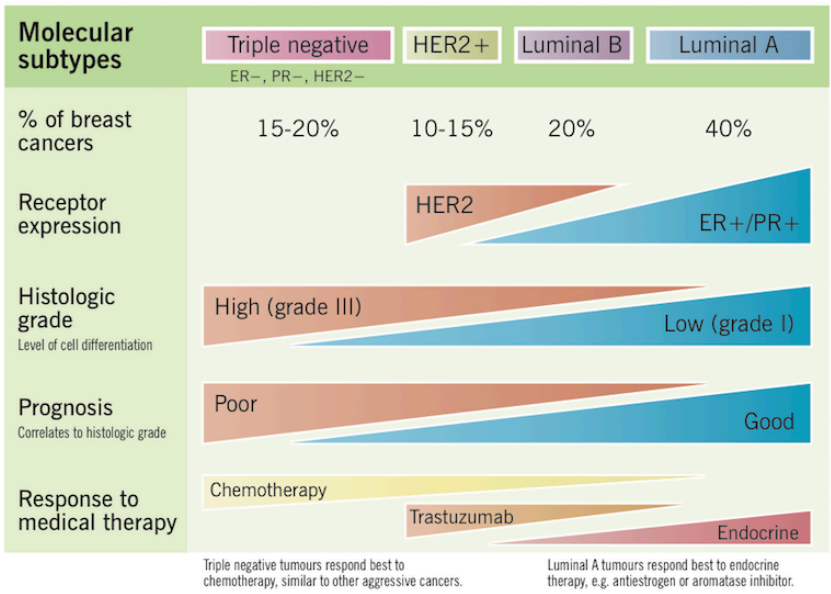


Figure 0.1 – The four molecular subtypes of breast cancer

Breast tumors are classified into four molecular subtypes according to the expression of several receptors. This molecular classification is critical in clinical decision-making and therapeutic approaches. Abbreviations: ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2. Figure adapted from http://www.pathophys.org/breast-cancer/.

A woman's risk of breast cancer is due to a combination of factors. Women's age is the main non-modifiable factor influencing breast cancer risk and explains why most breast cancers are found in women age 55 and older. In contrary, pregnancy is the most important modifiable factor influencing breast cancer risk, with late or no pregnancy associated with increased risk of breast cancer.

In this thesis, we will present three studies. First, we will explore the complex relationship between reproductive history and breast cancer biology in a reanalysis of a publicly available dataset of 560 breast cancer. Next, we will present the first and unique study aiming to uncover the genomic alterations associated with breast cancer diagnosed during pregnancy. Finally, we will investigate the effect of RANKL inhibition on the biology of breast cancer in young women.

2 The role of omics technologies in reshaping our understanding of breast cancer biology

Scientific research and technological advances have always been tightly linked. In this section, we will cover the tools that have facilitated our understanding of breast cancer biology. Since the apparition of microarrays that allowed the evaluation of the expression of thousands of genes simultaneously to the high-throughput whole genome sequencing, the technological advances that have been made in the "omics" field have greatly improved our understanding of breast cancer biology.

2.1 Transcriptomic profiling of breast cancer and its clinical utility

Gene expression profiling of breast cancer with DNA microarray, measuring thousands of transcripts simultaneously, have confirmed the existence of interpatient heterogeneity. Breast cancer is now seen as a group of molecularly distinct subtypes. The "intrinsic" molecular classification based on unsupervised hierarchical clustering of gene expression profiles proposes four distinct subtypes (Perou et al., 2000; Sorlie et al., 2001; Sotiriou et al., 2003):

(1) Luminal-A, which are mostly HR-positive and histologically low-grade/low proliferative

(2) Luminal-B, which are mostly HR-positive but may express low levels of hormone receptors and are often high-grade/high proliferative

(3) HER2-enriched, which show amplification and high expression of HER2

(4) Basal-like, which mostly correspond to HR-negative and HER2-negative tumors ("triple-negative breast cancer", TNBC)

These subgroups correspond well to pathological classification by HR and HER2 status, as well as on proliferation markers and histologic grade. Besides, this classification is associated with clinical outcome. The Basal-like and HER2-enriched subtypes are associated with the worse survival whereas the Luminal-A is associated with the best survival. Finally, the Luminal-B is associated with worse survival than Luminal-A but a better survival than the Basal-like and HER2-enriched (Figure 0.2).

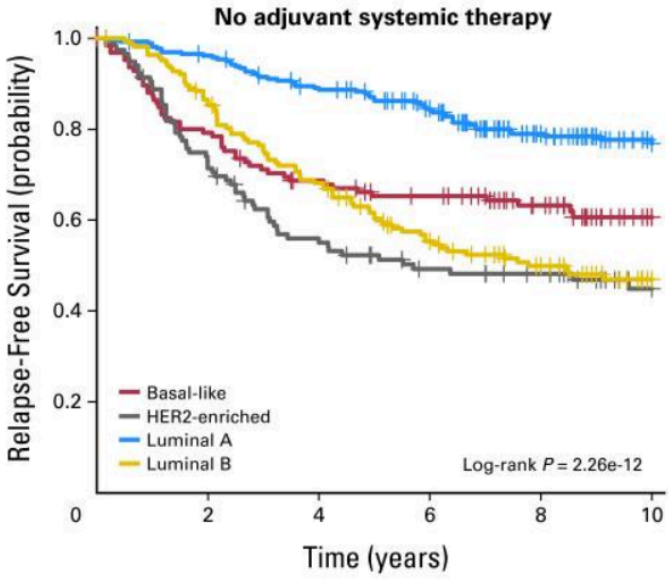


Figure 0.2 – Relapse-free survival of breast cancer patients according to intrinsic molecular classification

Outcome predictions according to intrinsic BC subtypes (PAM50) in a set of 710 node-negative, non-adjuvant treated patients. Reprinted from (Parker et al., 2009).

Gene expression profiling has led to the development of several diagnostic tests useful for clinical decision making. Supervised analyses have led to the development of gene expression signatures designed to predict survival and/or treatment response (Prat et al., 2012; Sotiriou and Pusztai, 2009). Several methods such as PAM50 (Parker et al., 2009), MammaPrint (van 't Veer et al., 2002), OncoType DX (Paik et al., 2004), Genomic Grade Index (Sotiriou et al., 2006) and others (Ma et al., 2008), have been shown to add prognostic and predictive information to standard clinicopathological features. The clinical utility of these assays was recently confirmed in large prospective trials (Cardoso et al., 2016; Sparano et al., 2015, 2018).

2.2 The genomic revolution in breast cancer

The technological advances, coupled with the shrinking cost of high-throughput sequencing, characterize the genomic revolution. In breast cancer research, this adventure started in 2009 when Shah et al. used whole-genome sequencing (WGS) to reveal the presence of heterogeneity between the mutational landscape of primary and metastasis pairs from a single patient (Shah et al., 2009). Ding et al. confirmed a year later the substantial heterogeneity in the distribution of single nucleotide variants (SNVs) and structural variants between metastasis and primary (Ding et al., 2010). In 2012 alone, seven back-to-back studies reported the genomic landscape of breast cancer and the

catalogue of driver mutations (Banerji et al., 2012; Curtis et al., 2012; Ellis et al., 2012; Nik-Zainal et al., 2012a; Shah et al., 2012; Stephens et al., 2012; TCGA, 2012). Among these, a landmark study from The Cancer Genome Atlas (TCGA) including 825 patients, provided the foundation of breast cancer genomics by integrating multiple data at the DNA (genome and epigenome), RNA and protein levels (TCGA, 2012). This study revealed that the four intrinsic breast cancer subtypes are associated with distinct genomic features. For example, mutations in *PIK3CA, GATA3*, and *MAP3K1* are associated with the luminal-A subtype whereas *TP53* mutation is present in the vast majority of the basal-like subtype. This publicly available dataset is an important resource for our understanding of the disease and is extremely useful in providing a comprehensive overview of breast cancer biology. The same year, Curtis et al. investigated a set of more than 2,000 breast tumors with long-term survival data (from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)) for somatic copy number alterations (SCNAs), single nucleotide variants and gene expression. The joint analysis of gene expression and SCNAs revealed that SCNAs are the dominant feature affecting the breast cancer transcriptome and further refined the molecular stratification of breast tumors (Curtis et al., 2012). This genome-driven classification identified ten subtypes, termed integrative clusters (IntClust), characterized by distinct genomic drivers, pattern of survival, and response to neoadjuvant chemotherapy demonstrating potential clinical utility (Ali et al., 2014a; Dawson et al., 2013).

These studies were not only done to provide a comprehensive catalogue of driver mutations. For example, sequencing of samples accrued from neoadjuvant aromatase inhibitor clinical trials led to the identification of mechanisms underlying aromatase inhibitor resistance (Ellis et al., 2012). Additionally, these studies brought to our attention the thousands of passenger mutations buried in the cancer genome and the recognition that these passenger mutations were not completely random. In 2012, Nik-Zainal et al. presented the concept of mutational signature in a study of 21 whole breast cancer genomes (Nik-Zainal et al., 2012b). Passenger mutations are the result of both DNA damage and DNA repair mechanisms that have occurred during the course of cancer and leave a characteristic imprint, or mutational signature, on the cancer genome. Each substitution can be classified by taking into account the sequence context immediately 5' and 3' to each mutated base. As there are six classes of base substitution and 16 possible sequence contexts for each mutated base (A, C, G, or T at the 5' base and A, C, G, or T

at the 3' base), there are 96 possible mutated trinucleotides for each tumor. A mathematical approach, called non-negative matrix factorization (NMF), was used to extract the first five substitution signatures identified in breast cancer (Nik-Zainal et al., 2012b). Following this discovery, a pan-cancer analysis was conducted in which this approach was applied across 30 cancer types and revealed the existence of 21 distinct substitution signatures (http://cancer.sanger.ac.uk/cosmic/signatures) (Alexandrov et al., 2013). More recently, the same group identified twelve base substitution signatures in the largest breast cancer genomic dataset published so far, comprising 560 whole breast cancer genomes (Nik-Zainal et al., 2016). Of these twelve substitution signatures documented in breast cancer, signatures 1 and 5 are associated with age of diagnosis; signatures 2 and 13 are associated with the activity of the APOBEC cytidine deaminases; signatures 3 and 8 are associated with BRCA1/BRCA2 deficiency; signatures 6, 20, and 26 are associated with DNA mismatch repair deficiency; and signatures 17, 18, and 30 are still of unknown etiology (Nik-Zainal et al., 2016). In the first chapter, we will mine this publicly available dataset to explore the relationship between reproductive history and breast cancer biology.

We have summarized important studies on breast cancer genomics and their main findings in Table 0.1. Most of these studies were performed using common breast cancers. However, the genome of less common cancers, such as special histological subtypes, metastatic, male or diagnosed during pregnancy, might be distinct from the general breast cancer population and further effort in these groups are warranted.

Table 0.1 – Studies associated with the genomic revolution in breast cancer

Study	BC subtypes and number of patients	Technology and number of patients	Main findings
(Shah et al., 2009)	ILC, ER+	WGS (1)	- Substantial SNVs heterogeneity between metastasis and primary
(Ding et al., 2010)	TNBC only	WGS (1)	- Substantial heterogeneity in SNVs and structural variation between metastasis and primary
(Nik-Zainal et al., 2012b, 2012a)	ER+ (9) TNBC (8) HER2 (4)	WGS (21)	- First catalogue of somatic mutations from breast cancer genomes - First identification of substitution signatures using NMF
(Ellis et al., 2012)	ER+ only (77)	WGS (46) WES (31)	- Mutational landscape of luminal BC - Identification of mechanisms underlying aromatase inhibitor resistance
(Shah et al., 2012)	TNBC only (104)	Affymetrix SNP6.0 (106) RNA-seq (80) WES (54) WGS (15)	- Interpatient heterogeneity within TNBC and difference in clonally
(Banerji et al., 2012)	All major subtypes (108)	WEG (86) WGS (22)	- Identification of new recurrent mutations and fusion genes
(Stephens et al., 2012)	ER + (79) ER – (21)	WEG (100)	- Identification of new mutations in breast cancer driver genes
(TCGA, 2012)	All major subtypes (825)	WEG (507) DNA methylation (802) SNP arrays (773) mRNA microarrays (547) miRNA sequencing (697) RPPA (403)	- Seminal study integrating multiple data at the DNA (genome and epigenome), RNA and protein levels - The four main breast cancer subtypes are associated with distinct genomic features
(Curtis et al., 2012) (METABRIC)	All major subtypes (2000+)	Affymetrix SNP 6.0 (2000+) Microarray Illumina HT-12 v3 (2000+)	- Largest population-based study on breast cancer genomic architecture - Breast cancer is mainly driven by SCNAs - Identification of 10 new integrative subgroups based on SCNAs and its relation to gene expression
(Ciriello et al., 2015)	IDC (490) ILC (127) Mixed (88)	WEG (507)	- First molecular portraits of invasive lobular breast cancer - Invasive lobular carcinoma (ILC) is a clinically and molecularly distinct disease - ILCs show CDH1 and PTEN loss, AKT activation, and mutations in TBX3 and FOXA1
(Desmedt et al., 2016)	ILC of all major subtypes (630)	Targeted sequencing of 360 genes (630)	- Confirmation of the high mutation frequency of CDH1 - Identification of high-prevalence therapeutic targets HER2, HER3, AKT1 - Identification of distinct genomic alterations according to ILC histologic subtypes
(Pereira et al., 2016) (METABRIC)	All major subtypes (2000+)	Targeted sequencing of 170 genes (2433)	- Extension of the first METABRIC study - Associations between PIK3CA mutations and reduced survival are identified in three subgroups of ER-positive cancer - Intra-tumour heterogeneity is generally associated with worse outcome
(Davies et al., 2017a; Morganella et al., 2016; Nik-Zainal et al., 2016)	All major subtypes (560)	WGS (560) RNA-seq (260) DNA methylation miRNA	- Identification of 12 base substitution signatures in breast cancer, 2 insertion/deletion (indel) and 6 rearrangement signatures - Catalogue of 93 breast cancer driver genes - Development of HRDetect, a method for identifying BRCAness - Systematic characterization of mutational signatures revealed an association with APOBEC mutagenesis, DNA repair deficiency

Abbreviations: ILC, invasive lobular carcinoma; ER, estrogen receptor; WGS, whole genome sequencing; NMF, non-negative matrix factorization; SNVs, single nucleotide variants; SCNAs, somatic copy number alterations; WEG, whole exome sequencing; IDC, invasive ductal carcinoma; TNBC, triple-negative breast cancer.

3 Beyond the cancer genome and the importance of the immune microenvironment of breast cancer

In 2011, Hanahan and Weinberg updated their six well-established hallmarks of cancer by adding two emerging hallmarks which are "deregulating cellular energetics" and "avoiding immune destruction" (Hanahan and Weinberg, 2011).

It has been observed that some tumors rewire their metabolism to promote growth, survival, proliferation, and long-term maintenance. The common feature of this metabolic switch is increased glucose uptake and fermentation of glucose to lactate. However, the function of this phenomenon, known as the Warburg Effect, remains unclear (Liberti and Locasale, 2016).

The concept of immune surveillance of tumors proposes that cells and tissues are constantly monitored by the immune system. This surveillance is responsible for the primary defense against cancer to identify and eliminate nascent tumor cells. According to this logic, solid tumors that do appear have somehow managed to escape immune surveillance. It is hypothesized that this occurs after an intermediate phase, known as immunoediting, in which there is a selection of non-immunogenic tumor cell (Zitvogel et al., 2006) (Figure 0.3).

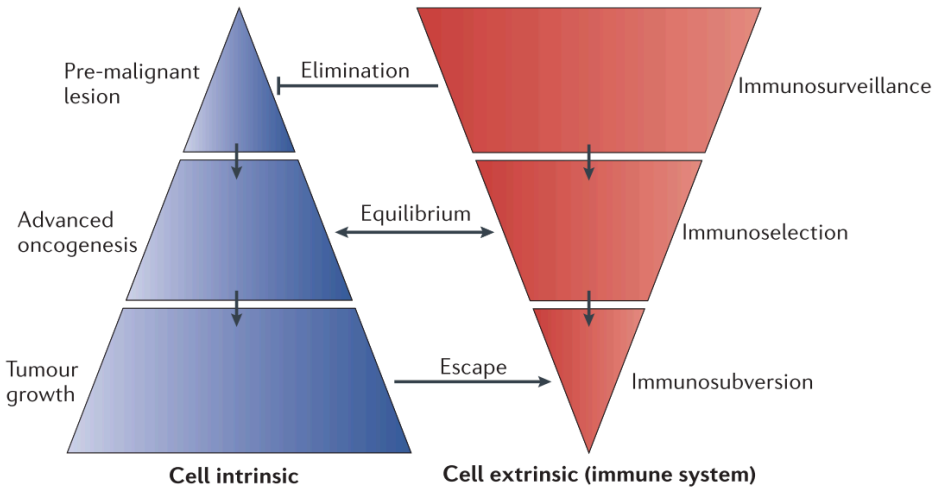


Figure 0.3 – The concept of immune surveillance

This figure illustrates the concept of immune surveillance in solid tumors. Tumorigenesis results from crosstalk between cancer cell-intrinsic effects and host immune system (cell-extrinsic) effects. Reprinted from (Zitvogel et al., 2006).

Compared to other cancers, breast cancer has not been traditionally considered as a highly immunogenic tumor. However, since 2010, a series of studies have evaluated the presence of tumor-infiltrating lymphocytes (TILs) in numerous breast cancer clinical trials, revealing the prognostic and predictive value of TILs as a biomarker (Denkert et al., 2010; Loi et al., 2013, 2014; Mahmoud et al., 2011). TILs are mononuclear immune cells that infiltrate tumor tissue and are commonly evaluated in hematoxylin and eosin (HE) stained histological slides through light microscopy. As with any biomarker, the utility of TILs assessment relies on the development of a standardized and reproducible scoring methodology. That is why, Salgado, R. et al. have published the international guidelines for the evaluation of TILs in breast cancer (Salgado et al., 2015). Savas et al., have reviewed the clinical relevance of host antitumor immunity and the role of TILs in breast cancer (Savas et al., 2016). Numerous studies have found that a greater number of TILs in the tumor stroma is associated with a better prognosis in patients with TNBC and HER2-positive BC (Adams et al., 2014; Ali et al., 2014b; Loi et al., 2013, 2014; Luen et al., 2017). A higher level of TILs is associated with response to therapy and predicts survival in patients treated with anthracycline-based chemotherapy and trastuzumab (Loi et al., 2014) and a potential benefit to immune checkpoint inhibitor (Loi et al., 2018).

Not all breast cancers have a high presence of TILs, and this depends greatly on BC subtypes. The current key question is to determine how to induce TILs infiltration, for example in luminal breast cancer, in which TILs levels were reported to be low. Some evidence regarding this question will be presented in **chapter 3** of this work.

4 The link between reproductive history and breast cancer

As previously mentioned, a woman's reproductive history is the most important modifiable factor influencing her risk of breast cancer. In the following section, we will review the current evidence on the impact of parity and age at first birth on breast cancer risk and prognosis. Next, we will discuss current data underlying the biological mechanisms putatively associated with pregnancy-induced breast cancer protection. This relates particularly to the first chapter of this thesis.

4.1 The impact of parity and age at first birth on breast cancer risk

The effect of parity and age at first birth on the risk of developing breast cancer has been well documented in several epidemiological studies (Kroman et al., 1997a; MacMahon et al., 1970; Papatestas et al., 1980; Rosenberg et al., 2004; Trichopoulos et al., 1983). Parity is known to have a dual effect on breast cancer risk with an increased risk from 5 to 10 years after pregnancy followed by a strong and life-long protective effect (Albrektsen et al., 2005; MacMahon et al., 1970). This protective effect is strongly influenced by age at first birth as pregnancy-induced tumor protection is more pronounced if the first birth has occurred early in life. For example, women having a first birth before 20 years are associated with a 50 percent decreased risk of breast cancer compared to nulliparous. On the contrary, women having a first birth after 35 years do not benefit from the pregnancy-induced protection and are even associated with an increased risk of BC compared to nulliparous (Figure 0.4). We and others have documented that pregnancy-induced tumor protection is different according to breast cancer subtypes with parity and young age at first birth been associated with a marked reduction in the risk of developing HR-positive BC subtype (Ellingjord-Dale et al., 2017; Lambertini et al., 2016; Ritte et al., 2013; Yang et al., 2011). Of note, breastfeeding and multiple pregnancies also decrease breast cancer risk. Longer time of breastfeeding (at least one year) is associated with a lower breast cancer incidence, and each additional birth increases protection (Collaborative Group on Hormonal Factors in Breast Cancer, 2002; Faupel-Badger et al., 2013; Islami et al., 2015).

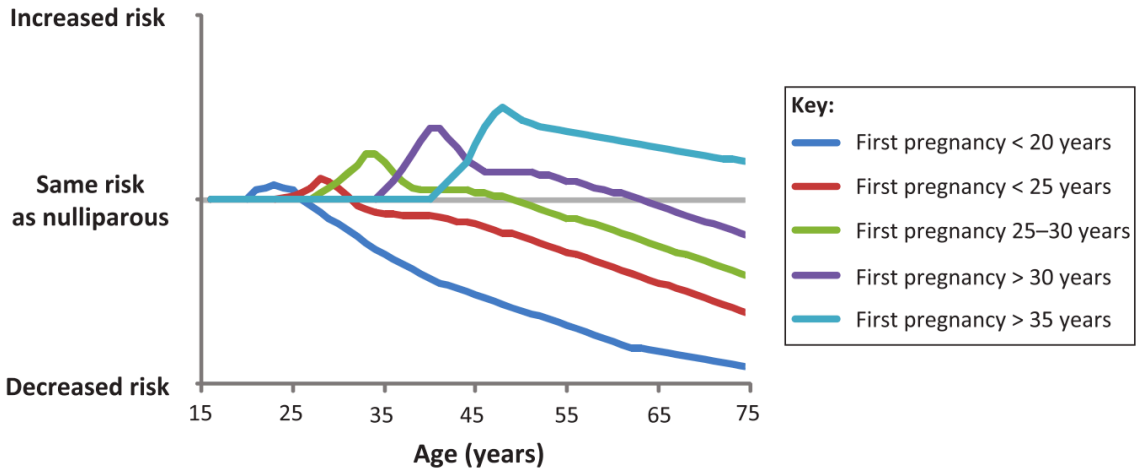


Figure 0.4 – Effect of pregnancy and age at first birth on the risk of breast cancer

This figure highlights lifetime risk of breast cancer according to age at first birth (colored lines) compared to nulliparous (straight grey line). This demonstrates that: (1) early pregnancy decreases breast cancer risk in the long term; (2) if the first pregnancy has occurred earlier, the breast cancer protective effect of pregnancy is greater; (3) pregnancy leads to a transient increase in breast cancer risk following parturition; and (4) pregnancy-associated increase in breast cancer risk becomes more pronounced with increasing age at first pregnancy. Reprinted from (Meier-Abt and Bentires-Alj, 2014).

4.2 The impact of parity and age at first birth on breast cancer prognosis

The impact of parity and age at first birth on the prognosis of BC remains unclear. Some reports claim no association between parity and prognosis (Bladström et al., 2003; Kroman et al., 1997b; Rosenberg et al., 2004), others claim that parity is associated with a poorer prognosis (Alsaker et al., 2013; Halmin et al., 2008; Phillips et al., 2004). These discrepancies may be attributed to different distributions of potential confounders such as race, menopausal status and intrinsic subtypes across study populations.

In contrast, breast cancer diagnosed shortly after pregnancy has been associated with poorer prognosis across multiple studies (Barnett et al., 2008; Nagatsuma et al., 2013; Rosenberg et al., 2004; Sun et al., 2016). Schedin et al. have published several experiments investigating the impact of postpartum mammary changes on breast cancer initiation and progression (Schedin, 2006). Right after breastfeeding the fully differentiated gland regresses to its pre-pregnant state by an innate tissue-remodeling mechanism called involution (Strange et al., 1992). The scale and rapidity of this phenomenon is unique to the mammary gland, and shares similarities with inflammation and wound healing programs that are known to be pro-tumorigenic. Evidence indicates that involution promotes breast cancer progression and metastasis in several animal

models (Martinson et al., 2015). Lyons et al. found that the inhibition of the proinflammatory Cox-2 by ibuprofen resulted in a reduction in tumor size and metastasis in rodent models of postpartum breast cancer (Lyons et al., 2011, 2014).

4.3 Biological mechanisms underlying pregnancy-induced breast cancer protection

Several hypotheses, relying on both cell autonomous and non-autonomous mechanisms, have been proposed to explain the protection against breast cancer conferred by pregnancy. Similar to humans, rodents exhibit parity-induced protection against mammary tumorigenesis. Thus, most of the studies have been conducted on animal models.

4.3.1 Cell-autonomous mechanisms

One of the first hypothesis underlying the breast cancer protection associated with an early pregnancy argued 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 (Russo et al., 1982). Gene expression studies comparing nulliparous and parous mammary tissue, from both rat and human, have observed upregulation of genes related to cell differentiation in the parous breast (Blakely et al., 2006; Russo et al., 2008). 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 (Medina, 2005; Meier-Abt and Bentires-Alj, 2014). Downregulation of the Wnt/Notch signaling, the major pathway that maintains stemness, is thought to be one of the pathways responsible for this change in cell fate determination (Meier-Abt et al., 2013, 2014, 2015). Several animal studies have also demonstrated a role for p53 in linking gestational hormones and the protection against mammary tumor. It has been shown that expression of p53 and its downstream transcriptional target p21, were increased in parous and estrogen/progesterone-treated mammary epithelium in response to carcinogen (Sivaraman et al., 2001). In the absence of p53, the protection given by parity or exogenous hormones was lost (Dunphy et al., 2008; Jerry et al., 2000; Medina and Kittrell, 2003).

A second hypothesis relies on a mechanism in which pregnancy induces a decrease of hormone responsiveness within the mammary gland. A study, comparing ER and PR staining of normal breast tissue from 26 premenopausal women and 30 pregnant women, showed that PR expression was significantly decreased during pregnancy. They also compared the level of PR between 16 nulliparous and 10 parous from the premenopausal group and found that PR was lower in parous women (Taylor et al., 2009). Consistent results were observed in another study comparing normal tissue from 20 nulliparous and 32 parous women (Asztalos et al., 2010).

4.3.2 Cell non-autonomous mechanisms

Persistent changes in circulating hormone levels induced by pregnancy have been associated with breast cancer risk reduction. Prolactin, a peptide hormone involved in mammary development and lactation, has been associated with an increased risk of breast cancer through an increase in cell proliferation and inhibition of apoptosis (Clevenger et al., 2003). In a cross-sectional study including 2400 women, it has been proposed that the link between parity and breast cancer risk was mediated, at least in part, by a long-lasting reduction of serum prolactin levels induced by pregnancy (Eliassen et al., 2007).

It has also been hypothesized that the pregnancy-induced protection against breast cancer could be related to the modification of the extracellular matrix (ECM) composition after pregnancy (Polyak, 2006). One study has shown that parity induces change of fibrillar collagen organization, in favor of a tumor-suppressive extracellular matrix (Maller et al., 2013).

To our knowledge, no studies have explored the link between reproductive history and the genomic landscape of breast cancer, a topic that will be covered in the first chapter of this thesis.

5 Breast cancer diagnosed during pregnancy

Breast cancer diagnosed during pregnancy (BCP) is a tragic situation that taints one the happiest moment of a woman's life. In the following section, we will summarize several aspects of this rare disease such as its epidemiology, prognosis and clinical management. We will review the current data on how pregnancy affects breast cancer biology. This relates particularly to the second chapter of this work.

5.1 Epidemiology

Breast cancer diagnosed during pregnancy is a rare disease with an incidence ranging from 2.4 to 7.3 per 100,000 pregnancies in population-based investigations (Andersson et al., 2009; Eibye et al., 2013; Lee et al., 2012; Stensheim et al., 2009). During the last few decades, the incidence of BCP has increased (Andersson et al., 2009; Eibye et al., 2013) (Figure 0.5). As breast cancer occurrence increases with age, it is hypothesized that this increase is due to the general trend of delaying childbearing since the 1970s in developed countries.

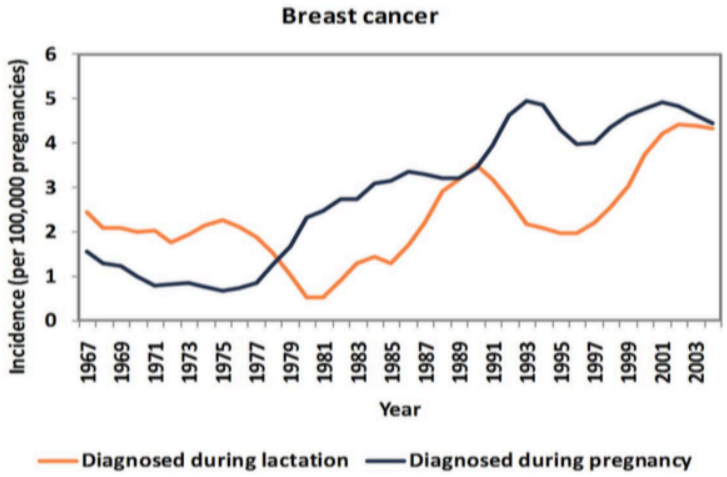


Figure 0.5 – Incidence of breast cancer during pregnancy and lactation in the last decades

Change in the incidence of breast cancer diagnosed during pregnancy and lactation based on data extracted from the cancer registry and the medical birth registry of Norway. The figure shows the annual incidence of BCP (blue line) for the period 1967-2004, using 5 years moving averages, showed as proportions per year per 100,000 pregnancies. For the period 1990-2004, the incidence of BC during pregnancy were 4.6/100,000, which is about 1/20,000. Reprinted from (Loibl et al., 2015a).

5.2 Prognosis

Given the rarity of BCP, most of the studies have been focused on pregnancy-associated breast cancer (PABC) which can be defined as breast cancer diagnosed within one year

following pregnancy. So far, two meta-analyses have addressed the prognosis of PABC. In 2012, Azim et al. have conducted a meta-analysis on 30 studies comparing the overall survival (OS) and the disease-free survival (DFS) of 3,628 patients diagnosed with PABC and 37,100 control breast cancer patients (Azim et al., 2012a). Overall, this study reported an inferior OS and DFS for women diagnosed with PABC compared with controls (pooled HR 1.44, CI 1.27–1.63 and pooled HR 1.60, CI 1.19–2.16, for OS and DFS, respectively). More recently, Harman et al. have conducted a meta-analysis including 60 case-controlled/cohort studies that confirmed the poorer prognosis associated with PABC (HR, 1.54; 95% CI, 1.16–2.04 and HR, 1.47; 95% CI, 1.04–2.08 respectively for OS and DFS) (Hartman and Eslick, 2016). However, they used a much larger definition of PABC that included patients diagnosed up to 5 years postpartum. Of the 60 studies included, 13 were focused on patients diagnosed during pregnancy and included 906 BCP cases. There was an increased risk of death for patients diagnosed during pregnancy compared to non-pregnant control (pooled HR 1.47, 95% CI 1.04–2.08). The unique biological characteristics of breast cancer during pregnancy, as well as the delay in diagnosis with a more advanced stage at presentation, may represent possible explanations for these findings. These conclusions should be nuanced as the largest international study, in terms of the number of BCP (311 eligible cases), showed similar prognosis between BCP compared with non-pregnant patients (HR, 1.19; 95% CI, 0.73–1.93 and HR, 1.34; 95% CI, 0.93–1.91, respectively for OS and DFS) (Amant et al., 2013). This highlights the importance of careful management of this challenging situation in accordance with the available international guidelines.

5.3 Clinical management

International guidelines on the clinical management of BCP have been extensively reviewed by Loibl et al. and Peccatori et al. (Loibl et al., 2015b; Peccatori et al., 2013). General recommendations indicate that BCP should be treated as similar to young non-pregnant breast cancer patients with adaptations according to gestational age and the use of some treatments (Figure 0.6). These patients require adequate expertise and should be preferably managed in a comprehensive cancer center involving a specialized multidisciplinary team.

Surgery can be safely performed during all trimesters of pregnancy with similar indications for radical or conservative surgery (Toesca et al., 2014). If breast-conserving

surgery is not a reasonable option, breast reconstruction is an essential component of the clinical management of breast cancer patients. Of note, one study showed that immediate breast reconstruction after mastectomy does not seem to be associated with considerable morbidity to the patient or the fetus (Lohsiriwat et al., 2013).

Sentinel lymph node biopsy involving the injection of Technetium, a relatively low radioactive tracer with fast clearance, appears to be safe for the fetus and should not be discouraged (Gentilini et al., 2004, 2010).

Radiation therapy (RT) has teratogenic and carcinogenic effects and, if possible, should be postponed until after the delivery. However, if RT is done during the first half of gestation, the uterus is still contained in the true pelvis and the fetus would be exposed to a relatively low radiation dose (Van der Giessen, 1997). Hence, if the risk of delaying RT is too high for the patient, RT combined with appropriate abdominal shielding might be considered in the first trimester (Kal and Struikmans, 2005).

Chemotherapy is contraindicated during the first trimester due to the high risk of congenital malformations (National Toxicology Program, 2013). However, the prevalence of malformations in women treated with chemotherapy after the first trimester is comparable to the general population. The standard combination of anthracyclines, cyclophosphamide, and taxanes recommended in non-pregnant breast cancer patients is safe after the first trimester in BCP patients either in the adjuvant or neoadjuvant setting (Cardonick et al., 2012; Zagouri et al., 2013).

Endocrine therapy and anti-HER2 treatments should be avoided during pregnancy and postponed after delivery (Loibl et al., 2015b; Peccatori et al., 2013). Tamoxifen use in pregnant patients has been associated with fetal malformation such as ambiguous genitalia and craniofacial malformations. Trastuzumab use during pregnancy has been associated with increased risk of oligohydramnios with a subsequent predisposition to preterm labor, fetal morbidity, and mortality (Lambertini et al., 2015).

Finally, iatrogenic preterm delivery should be avoided at all times. Amant and al. have shown that compared to the use of chemotherapy during pregnancy, prematurity alone is associated with worse cognitive outcome for the newborn (Amant et al., 2015).

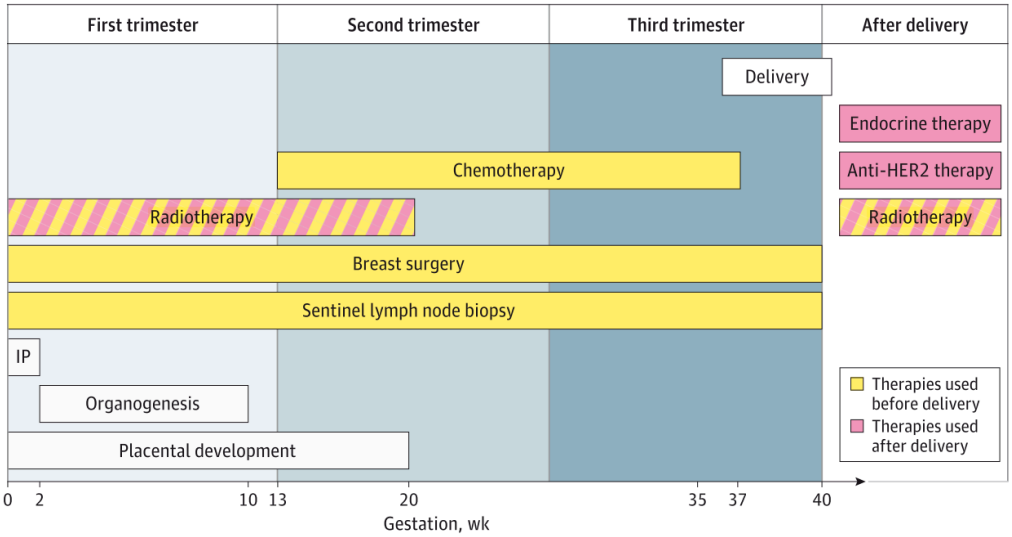


Figure 0.6 – Therapeutic options for breast cancer according to the gestational period

The timeline summarizes the therapeutic options for breast cancer during pregnancy. Surgery and sentinel lymph node biopsy are considered safe throughout all trimester of pregnancy. Chemotherapy can be used after the first trimester. If indicated, radiotherapy can be applied from the first until early second trimester. Reprinted from (Loibl et al., 2015a).

5.4 Biology of BCP

To understand the biology of BCP one has to understand the physiology of pregnancy. Indeed, the considerable biological changes that occur during pregnancy both at the systemic level and within the mammary gland might influence the biology of breast cancer. For example, the increased levels of gestational hormones, the morphological changes, and the immunosuppressive state associated with pregnancy could influence and even promote BCP. That is why one of the main research themes of the *J.-C. Heuson Breast Cancer Translational Laboratory* (BCTL) is to understand the effect of pregnancy on breast cancer biology.

BCTL research project lead by Azim et al. reported the first study on the distribution of breast cancer subtypes in a case-control study. There was no difference in the distribution of the four main subtypes as assessed by IHC between BCP and control (luminal-A, luminal-B, HER2, TNBC) (Azim et al., 2012b). In a subsequent study, the same investigators performed gene expression profiling to assess the intrinsic molecular subtypes with the PAM50 classifier and did not find any difference as well (Azim et al., 2014). However, they found that pregnancy impacted the breast cancer transcriptome with enrichment of G protein-coupled receptor pathway and the serotonin receptor pathway. In a first attempt of reporting the pattern of selected mutations they performed

genotyping of 84 hotspot SNVs, but this analysis revealed no difference between pregnant and non-pregnant patients.

RANKL, a protein that appears to play a key role in breast carcinogenesis in young women (see section 6.2), has been shown to be more expressed in BCP than control non-pregnant young BC patients (Azim et al., 2015a). Finally, focusing on the tumor immune microenvironment, TILs levels were found to be lower in BCP than in BC from non-pregnant controls (Azim et al., 2015b). Taken together, pregnancy does appear to alter the biology of breast cancer and its microenvironment.

Complementing the work of Azim et al., we will uncover the mutational landscape of breast cancer diagnosed during pregnancy using genome-wide copy number alterations profiling and whole genome sequencing in the second chapter of this thesis.

6 RANKL as a potential target in breast cancer patients

As previously stated, RANKL up-regulation has been associated with breast cancer in young and during pregnancy. In the following section, we will review some recent advances on the role of RANK/RANKL signaling in homeostasis and its emerging importance in cancer biology. This relates particularly to the last chapter of this work.

6.1 The role of RANK/RANKL in homeostasis

6.1.1 Bone physiology

The receptor activator of nuclear factor-kB ligand RANKL, its receptor RANK and its natural decoy receptor osteoprotegerin (OPG) were first identified as key regulators of bone metabolism (Wada et al., 2006). Bone tissue is not in a fixed state but is constantly remodeled during the life. Two major cell types are involved in bone remodeling: osteoblasts, derived from mesenchymal stem cells, are responsible for bone formation, on the contrary, osteoclasts, derived from hematopoietic stem cells, are responsible for bone resorption (Charles and Aliprantis, 2014; Long, 2012). The balance between osteoblasts and osteoclasts is tightly regulated by the RANKL/RANK/OPG system to maintain proper bone homeostasis (Leibbrandt and Penninger, 2008). RANKL is expressed by osteoblasts whereas its receptor RANK is expressed at the external surface membrane of osteoclast precursors. Stimulation by RANKL induces osteoclast differentiation and subsequent bone resorption (Wada et al., 2006). As its name suggests, osteoprotegerin protects from bone breakdown by serving as a decoy receptor for RANKL, thus inhibiting osteoclast stimulation. OPG expression is regulated by estrogen. The natural decline of estrogen levels associated with menopause explains why osteoporosis is a common condition in postmenopausal women (Leibbrandt and Penninger, 2008). In elderly populations, osteoporosis represents a major health problem, and several drugs have been developed and approved for its treatment.

Denosumab, a human monoclonal antibody against RANKL, has shown to be highly effective for the management of osteoporosis, but also for the prevention of skeletal-related events (SREs) due to bone metastasis in breast and prostate cancers (Lacey et al., 2012). Bone is the most common metastatic site in breast cancer patients and SREs induced by metastases cause severe morbidity and pain (Costa et al., 2008; Kennecke et

al., 2010). Denosumab is also approved in hormone receptor positive postmenopausal BC patients receiving aromatase inhibitor therapy for reducing bone loss and the risk of fracture. Aromatase inhibitors suppress the conversion of androgens to estrogens, resulting in estrogen depletion, which in turn leads to lower bone mineral density and increase the risk of fracture (Forbes et al., 2008). The efficacy of denosumab in reducing therapy-induced osteoporosis associated with aromatase inhibitors was reported in the ABCSG-18 trial, a large prospective, double-blind, placebo-controlled, multicenter, phase III study (patients in the denosumab group had a significantly delayed time to first clinical fracture; HR = 0.50, 95% CI 0.39–0.65, $P < 0.0001$) (Gnant et al., 2015).

6.1.2 Immune system

As reviewed by Cheng et al. and Ferrari-Lacraz et al., the role of RANK/RANKL signaling in the immune system is indisputable. Pre-clinically, RANK/RANKL signaling has been implicated in various physiological immune processes such as lymph node organogenesis and immune-tolerance control (Cheng and Fong, 2014; Ferrari-Lacraz and Ferrari, 2011).

It has been shown that RANK and RANKL knockout mice do not develop lymph node and have an impaired lymphocyte development, revealing a developmental role of RANK/RANKL system in immune tissue (Dougall et al., 1999; Kong et al., 1999). While RANKL is expressed by activated T cells, it has been shown that RANK is expressed by myeloid cells including monocytes, macrophages and dendritic cells (DC). Activated T cell expresses RANKL to promote DC survival and proliferation (Anderson et al., 1997). Thereby, RANK/RANKL crosstalk appears to be involved in enhancing global memory T cell responses.

As opposed to its positive effect on immune response, RANK/RANKL signaling is crucial to maintaining central and peripheral immune tolerance (Cheng and Fong, 2014; Ferrari-Lacraz and Ferrari, 2011). In the thymus, this pathway regulates AIRE medullary thymic epithelial cell maturation, a cell type involved in central immunotolerance (Akiyama et al., 2008). In the gastrointestinal tract, RANKL induces microfold cells (M cells) that contribute to the establishment of peripheral T cell tolerance to commensal bacteria (Knoop et al., 2009). In the pancreas, RANK/RANKL interaction is also involved in the regulation of anti-islet autoreactive T cells. As shown in a mouse model

of type 1 diabetes, RANK/RANKL blockade resulted in a reduction of regulatory T cells and diabetes progression (Green et al., 2002).

6.1.3 Mammary gland development during pregnancy

As previously stated, the mammary gland undergoes dramatic changes during pregnancy to prepare for lactation. RANK/RANKL signaling is crucial for the formation of a lactating mammary gland. It has been shown that RANK or RANKL knockout mice failed to form lobuloalveolar milk-secreting structures during pregnancy (Fata et al., 2000). Progesterone, the most important pregnancy hormone, induces proliferation and differentiation of mammary epithelial cells into milk-secreting acini (Hennighausen and Robinson, 2001). Mechanistically, progesterone induces RANKL expression in luminal epithelial cells, which in turn induces proliferation of basal RANK expressing progenitor cells in a paracrine fashion (Rao et al., 2017).

6.2 The role of RANK/RANKL in cancer

The link between RANK/RANKL signaling and cancer has been demonstrated in three independent mechanisms, each relying on different cell types (cancer cell itself, immune cells, and bone cells), and influencing all stages of tumor progression, from initiation to metastasis (Figure 0.7). Given the critical importance of RANK/RANKL signaling in mammary gland, immunity, and bone metastasis, it has been proposed that RANKL inhibition could represent a therapeutic opportunity in breast cancer. The excellent safety profile of denosumab established in thousands of patients makes it a good candidate to study RANKL inhibition in breast cancer patients.

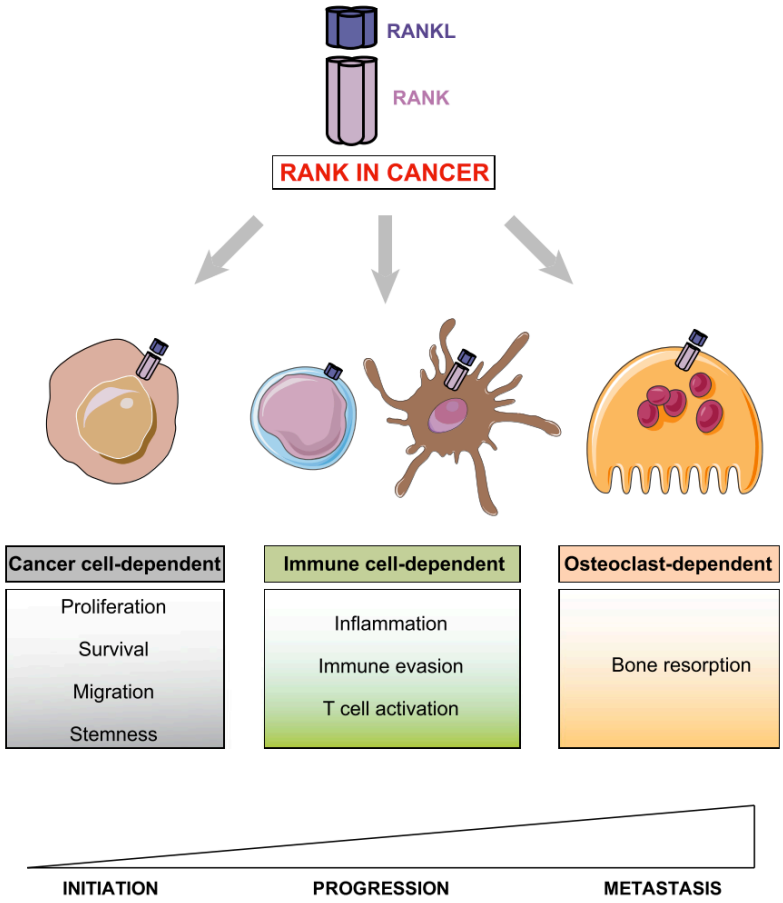


Figure 0.7 – The role of RANK/RANKL signaling in cancer

The drawing illustrates the multiple effects of the RANK/RANKL signaling according to the targeted cell population. Activation of the RANK/RANKL pathway has been reported in three different cell populations (cancer cells, immune cells, and osteoclasts) influencing all stages of tumor progression from initiation to metastasis. Reprinted from (Gonzalez-Suarez and Sanz-Moreno, 2016).

6.2.1 Cancer cell dependent

As previously stated, RANK/RANKL signaling is crucial to transduce progesterone's proliferative signal for mammary gland expansion during pregnancy. In 2010, two back-to-back studies showed that RANKL inhibition slowed down tumor progression in two distinct progestin-driven mammary cancer models (Gonzalez-Suarez et al., 2010a; Schramek et al., 2010a). Schramek et al. used a progestin- and carcinogen-induced mouse cancer models and showed that inhibition of RANKL using RANK-Fc attenuated breast tumor progression. Gonzalez-Suarez et al. confirmed these results and further showed that RANK-Fc decreases spontaneous mammary tumorigenesis and lung metastasis in a transgenic tumor model (MMTV-*neu* mice) (Figure 0.8). Of note, administration of medroxyprogesterone acetate, a synthetic progestin, resulted in a 3000-fold upregulation of RANKL in mammary epithelial cells. Overexpression of RANK has been involved in

the induction of epithelial-mesenchymal transition (EMT) and stemness in normal and tumor cells (Palafox et al., 2012; Tsubaki et al., 2013). RANK deletion in tumor cells increases the tumor latency and decreases tumor initiating ability, by inducing differentiation and reducing survival of tumor cells (Yoldi et al., 2016). More recently it has been shown that RANK and RANKL are also critical regulators of BRCA1-mutation-driven breast cancer and that anti-RANKL therapy could be an attractive preventive strategy for women carrying BRCA1 mutations (Nolan et al., 2016; Sigl et al., 2016b, 2016a).

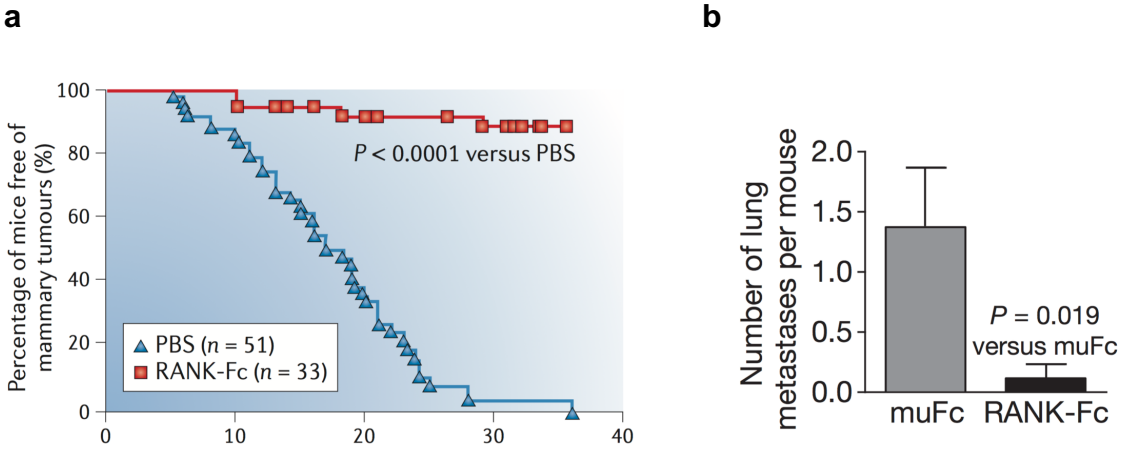


Figure 0.8 – The anti-tumor effect of RANKL blockade in preclinical models

***a**, RANKL blockade inhibits tumor progression in a mouse model of progestin-driven mammary cancer. **b**, RANKL blockade inhibits spontaneous lung metastasis in a transgenic tumor model (MMTV-neu mice). Reprinted from (Gonzalez-Suarez et al., 2010a; Schramek et al., 2010a).*

Although not primarily designed to study an effect on disease-free survival (DFS), two recent large phase III study of adjuvant denosumab in postmenopausal early breast cancer population revealed inconsistent DFS benefit (Coleman et al., 2018; Gnant et al., 2018). While analysis of the D-CARE study showed no benefit (HR = 1.04, 95%CI 0.91-1.19, P = 0.57), analysis of ABCSG-18 trial showed an improved DFS (HR = 0.823, 95% CI 0.69-0.98, P = 0.026) in breast cancer patients treated with denosumab. The possible explanations of these discrepancies could be differences in primary endpoints, a higher-risk population and higher dose regimen in the D-CARE study.

6.2.2 Immune cell dependent

Immunosuppressive regulatory T cells (Treg) play an essential role in the immunological tolerance and prevention of autoimmunity (Josefowicz et al., 2012). In breast cancer, it

has been shown that increased presence of Treg in tumor is associated with worse DFS and OS (Bates et al., 2006; Ohara et al., 2009). In a mice model, Tan W et al. have shown that pulmonary metastases were promoted by RANKL-expressing Treg. The exogenous administration of RANKL further enhanced metastasis, whereas inhibition of RANKL using RANK-Fc inhibits pulmonary metastasis (Tan et al., 2011). Clinical studies on the immunomodulatory potential of anti-RANKL are limited, but it is worth to mention a case report describing an exceptional response upon treatment to combination anti-CTLA4 ipilimumab and denosumab in a metastatic melanoma patient (Smyth et al., 2013) (Figure 0.9). Following this observation, the same group reported that RANKL blockade improved the efficacy of both anti-CTLA4 and PD1/PDL1 blockade in several tumor mouse models (Ahern et al., 2017, 2018a). Because of its emerging role in antitumour immunity, the efficacy of denosumab in combination with immune checkpoint inhibitors is currently being assessed in several clinical trials including melanoma, non-small cell lung cancer, and breast cancer (Ahern et al., 2018b).

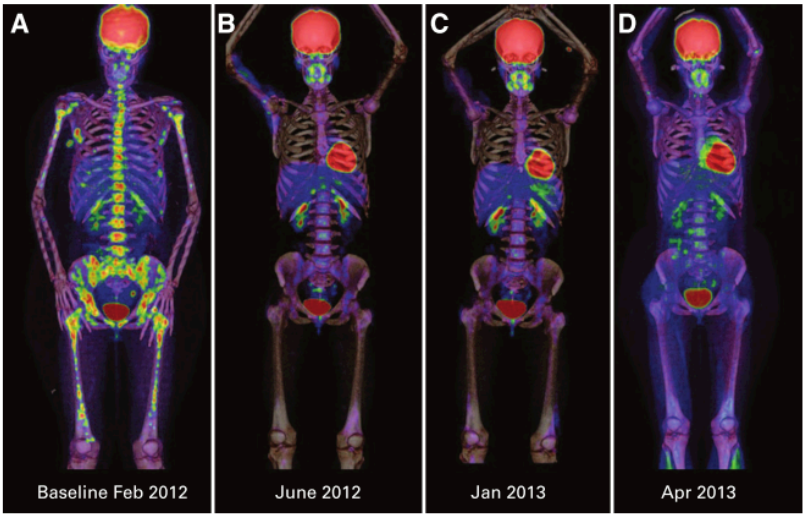


Figure 0.9 – Pathological response to combination anti-CTLA-4 and anti-RANKL in a case report of metastatic melanoma

Fluorodeoxyglucose positron emission tomography (FDG-PET) scan performed 16 weeks after the first dose of ipilimumab and 18 weeks after the first dose of denosumab showed a dramatic partial response. The patient remained free of disease progression until 48 weeks after starting ipilimumab when the patient developed a single cerebral metastasis that was resected. There was no evidence of any residual melanoma on the FDG-PET scan at the time of publication of the case report. Reprinted from (Smyth et al., 2013).

6.2.3 Osteoclast dependent

As previously stated, bone is the most common metastatic site in breast cancer patients, and bone metastases can cause severe morbidity and pain by inducing SREs (Costa et al.,

2008; Kennecke et al., 2010). Bone metastases induce bone destruction leading to the release of growth factors, such as TGF-beta and insulin-like growth factors, which in turn have a pro-tumorigenic effect. This phenomenon, usually referred to as the 'vicious cycle' of bone metastasis, rely at least in part on RANK/RANKL signaling (Figure 0.10). Importantly, co-culture experiments revealed that breast cancer cell lines could stimulate RANKL expression by osteoblasts, leading to increased osteolysis (Thomas et al., 1999). Moreover, it has been shown that overexpression of RANK in breast cancer cells was sufficient to increase their metastasis potential to the bone (Blake et al., 2014). This evidence led to the development of the D-CARE study aiming to assess the efficacy of denosumab in decreasing the risk of bone metastasis in early-stage breast cancer patients. Disappointingly, this trial, recently presented at the 2018 ASCO annual meeting, did not meet its primary endpoint of bone metastasis-free survival (HR = 0.97, 95%CI 0.82-1.14, P = 0.70).

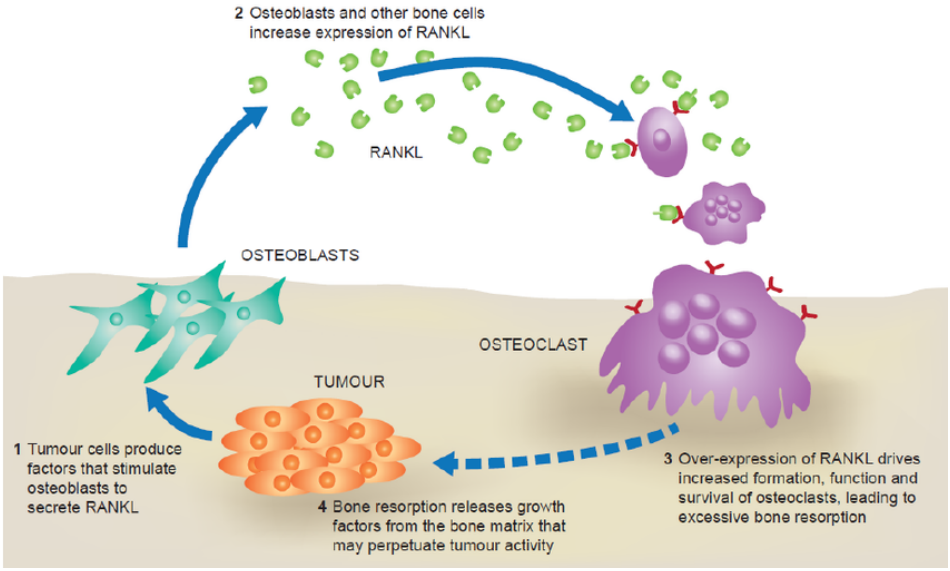


Figure 0.10 – The 'vicious cycle' of bone metastasis

Schematic illustration of the role of RANK/RANKL signaling in the 'vicious cycle' of bone metastasis. Tumor cells induce up-regulation of RANKL by osteoblasts. Then, RANKL stimulates osteoclasts which promotes bone resorption. Finally, bone resorption is accompanied by the release of growth factors that promote tumor cell growth. Reprinted from (von Moos and Haynes, 2013).

JUSTIFICATION OF THE THESIS

Early pregnancy is one of the most effective ways of decreasing breast cancer risk, and many studies have attempted to investigate the underlying mechanisms behind this phenomenon. However, no studies have explored the link between reproductive history and the genomic landscape of subsequent breast cancer. This will be the topic of the first chapter of this thesis.

The diagnosis of breast cancer during pregnancy (BCP) taints one the happiest moment of a woman's life. Due to the general trend of delaying childbirth, the incidence of BCP has increased during the last decades. In the second chapter, we will present the molecular alterations that characterize breast cancer diagnosed during pregnancy.

Finally, in the last chapter, we will explore the biological effect of RANKL inhibition on breast cancer in young women through a window-of-opportunity trial.

As previously stated, when starting this work;

- The link between parity and age at first birth on breast cancer risk was well documented, and new studies shed light on the differential effect according to breast cancer subtypes. However, no studies have explored the link between reproductive history and the genomic landscape of subsequent breast cancer.
- International guidelines were published on the clinical management of BCP but the genomic landscape of BCP was still unknown.
- Only pre-clinical studies indicated a strong link between RANK/RANKL signaling and breast cancer. In addition, the immunomodulatory effect of denosumab on human breast cancers was unknown.

In **chapter 1** of this thesis, we sought to address the following research questions to examine the possible associations between reproductive history and the biology of subsequent breast cancer;

<u>Main research question:</u>

- What is the imprint of pregnancy and age at first birth on the biology of subsequent breast cancer?

Specific research questions:

- What is the influence of parity and age at first birth on somatic mutations, somatic copy number alterations, and mutational signatures?
- What is the influence of parity and age at first birth on the breast cancer transcriptome including the distribution of the intrinsic molecular subtypes?
- What is the influence of parity and age at first birth on the tumor immune microenvironment?

In **chapter 2**, we sought to address the following research questions to better understand the biology of breast cancer diagnosed during pregnancy;

Main research question:

- What is the effect of pregnancy on the biology of breast cancer?

Specific research questions:

- Are there differences in the copy number alterations profiles between BCP and age/stage-matched breast cancer patients?
- Are there differences in the mutational landscape between BCP and age/stage-matched breast cancer patients?
- Are there differences in the mutational signatures distribution between BCP and age/stage-matched breast cancer patients?

In **chapter 3**, we sought to address the following research questions to investigate the effects of RANKL inhibition in primary breast cancer in young women;

Main research question:

- What is the effect of RANKL inhibition on the biology of breast tumors?

Specific research questions:

- Is a short of RANKL inhibition can induce a decrease in tumor proliferation rates as determined by Ki-67 IHC?
- What is the effect of RANKL inhibition on the tumor immune microenvironment?
- What is the effect of RANKL inhibition on breast cancer transcriptome?

CHAPTER 1: THE IMPRINT OF REPRODUCTIVE HISTORY ON BREAST CANCER BIOLOGY

This research work is related to the following publication:

Nguyen, B., Venet, D., Lambertini, M., Desmedt, C., Salgado, R., Horlings, H., Rothe, F., and Sotiriou, C. (2018). Imprint of parity and age at first pregnancy on the genomic landscape of subsequent breast cancer.

Preprint available online at: https://doi.org/10.1101/351205
Manuscript submitted in July 2018.

My contribution to this study involves:

- Conceptualization and Design
- Methodology
- Formal Analysis
- Literature search and Interpretation
- Statistical analyses
- Visualization
- Writing of the manuscript
- Presentation of the results at the following meeting:
    - San Antonio Breast Cancer Symposium, December 2018, San Antonio, TX, US

1 Introduction

The effect of parity and age at first pregnancy on the risk of developing breast cancer has been well documented (Kroman et al., 1997; MacMahon et al., 1970; Papatestas et al., 1980; Rosenberg et al., 2004; Trichopoulos et al., 1983). 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 (Albrektsen et al., 2005; MacMahon et al., 1970). 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 (Ellingjord-Dale et al., 2017; Lambertini et al., 2016; Ritte et al., 2013; Yang et al., 2011).

Several studies have attempted to investigate the mechanisms underlying this phenomenon (Meier-Abt and Bentires-Alj, 2014; Russo et al., 2012). 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. 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.

2 Methods

2.1 Data acquisition

All analyses were performed on a publicly available dataset comprising 560 breast cancer patients referred to as BRCA560 (Nik-Zainal et al., 2016). Clinical data, sequencing coverage, and mutational load were obtained from Supplementary Tables 1-3 in that reference. Coding driver mutation events and the contribution of mutational signatures were obtained from Supplementary Tables 14 and 21 in that reference. Raw count data from RNA sequencing were obtained from the authors. Results from HRDetect classifier were obtained from Supplementary Table 4 in reference (Davies et al., 2017a).

2.2 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 genefu R/Bioconductor package (Gendoo et al., 2016). Nulliparous patients were defined as women with breast cancer who have never given birth. Parous patients were defined as women with breast cancer who had at least one full term pregnancy. Early parous patients were defined as ≤ 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. 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 were required to perform this analysis.

2.3 TILs 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 (Salgado et al., 2015). 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). TILs information for patients for which evaluation from only one pathologist was available was discarded (N=3).

2.4 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, 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_{adj}) were done using 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 genomic status of *MYC*/*TP53* alterations. All correlations were measured using the non-parametric Spearman's *rho* coefficient. All reported P-values were two-tailed. Multiple testing correction was done using the false discovery rate method (FDR) (Benjamini and Hochberg, 1995) 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) and Bioconductor version 3.6. Differential expression analysis was performed with DESeq2 v.1.14.1 R/Bioconductor package (Love et al., 2014) on raw count data. Significantly differentially expressed genes were selected with an FDR of < 0.1. We used gage v.2.24.0 R/Bioconductor package (Luo et al., 2009) to identify significantly enriched pathways form the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017) and biological process from Gene Ontology with the log2FoldChange from DEseq2 results as input data.

3 Results

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

From a publicly available dataset comprising 560 breast cancer patients (Nik-Zainal et al., 2016), a total of 313 with available information on parity were included. We identified 264 (84.3%) parous and 49 (15.7%) nulliparous patients (Supplementary Figure 1.1). 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 cut-off value. All patients had available somatic mutations and SCNAs data, 182 patients (58.1%) had available transcriptomic data and 170 patients (54.3%) had information on TILs levels (Supplementary Figure 1.2).

Table 1.1 summarizes the clinicopathological features of patients. Compared to parous patients, nulliparous patients had significantly larger tumors (tumor size > 2cm, 59.2% vs. 37.1%, $P = 0.006$), higher frequency of node involvement (40.8% vs. 27.7%; $P = 0.027$) and lower frequency of triple negative disease (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 premenopausal (45.3% vs. 20.9%; $P = 0.005$). Late parous patients also had a lower frequency of large tumors (tumor pathological size > 2cm, 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 parous patients, we found a negative correlation between age at first pregnancy and age at breast cancer diagnosis (*rho* = -0.25, $P = 0.001$, Supplementary Figure 1.3), this association was even stronger when restricting to ER positive patients (*rho* = -0.37, $P < 0.001$, Supplementary Figure 1.3b). In a linear regression analysis adjusted for potential cofounders including pathological stage, molecular subtypes by IHC and histological subtypes we found that age at first pregnancy was independently and negatively associated with age at diagnosis ($P = 0.0004$).

Table 1.1 – Clinicopathological features of nulliparous and parous patients

		Nulliparous	Parous	P		Early parous	Late parous	P
N		49	264			82	71	
Age at diagnosis	**Median years (range)**	54 (30-81)	55 (28-81)	0.796[a]		59 (34-81)	49 (28-81)	$2.6x10^{-5}$ [a]
Menopausal status	**Pre**	13 (33.3%)	60 (30.3%)			14 (20.9%)	29 (45.3%)	
	Post	26 (66.7%)	138 (69.7%)	0.85		53 (79.1%)	35 (54.7%)	0.0053
Stage	**I**	7 (14.9%)	41 (15.9%)			6 (7.6%)	17 (24.3%)	
	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%)	0.019		27 (34.2%)	16 (22.9%)	0.039
pT	**Tx**	9 (18.4%)	105 (39.8%)			27 (32.9%)	16 (22.5%)	
	≤ 2cm	11 (22.4%)	61 (23.1%)			13 (15.9%)	26 (36.6%)	
	> 2cm	29 (59.2%)	98 (37.1%)	0.0062		42 (51.2%)	29 (40.8%)	0.012
pN	**Nx**	11 (22.4%)	112 (42.4%)			30 (36.6%)	17 (23.9%)	
	N0	18 (36.7%)	79 (29.9%)			25 (30.5%)	23 (32.4%)	
	N1+	20 (40.8%)	73 (27.7%)	0.027		27 (32.9%)	31 (43.7%)	0.2
Grade	**1**	6 (14%)	29 (12.7%)			8 (9.8%)	5 (7%)	
	2	17 (39.5%)	86 (37.7%)			25 (30.5%)	35 (49.3%)	
	3	20 (46.5%)	113 (49.6%)	0.93		49 (59.8%)	31 (43.7%)	0.059
Subtype by IHC	**Lum A-like**	22 (44.9%)	106 (40.3%)			29 (35.4%)	39 (54.9%)	
	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%)	0.0013		32 (39%)	14 (19.7%)	0.026
Histology	**Ductal**	36 (76.6%)	203 (81.2%)			71 (88.8%)	49 (71%)	
	Lobular	5 (10.6%)	23 (9.2%)			2 (2.5%)	10 (14.5%)	
	Other	6 (12.8%)	24 (9.6%)	0.67		7 (8.8%)	10 (14.5%)	0.01

Abbreviations: pT, pathological tumor size; pN, pathological nodal status; HR, hormone receptor; P, p-value derived from the $\chi 2$ test or the Fisher exact test when appropriate ([a]except continuous variable derived from Mann–Whitney U test).

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.

3.2 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 (Supplementary Data 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_{adj} = 0.097, P_{adj} = 0.075, respectively, Figure 1.1a). There was no significant difference in the total number of insertions or deletions (Indels) according to parity (P_{adj} = 0.464, Figure 1.1a). In contrary, compared to tumors from late parous patients, tumors from early parous patients were significantly associated with a higher Indels load (P_{adj} = 0.002, *FDR* = 0.007, Figure 1.1a). There was no significant difference between the total number of rearrangements according to parity nor age at first pregnancy.

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 (Supplementary Data File 1: Table S2). As expected, *PIK3CA* and *TP53* were the most frequently mutated genes (Figure 1.1b). 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 (50% vs. 22.5%; P_{adj} = 0.010; *FDR* = 0.046, Figure 1.1b) and lower frequency of *CDH1* mutations (1.2% vs. 12.7%; P_{adj} = 0.013; *FDR* = 0.046, Figure 1.1b). Considering the distribution of *TP53* mutations type, early parous patients had a significantly higher frequency of truncating mutations as compared to late parous patients (25.6% vs. 7%; P_{adj} = 0.014, Supplementary Figure 1.4). 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.

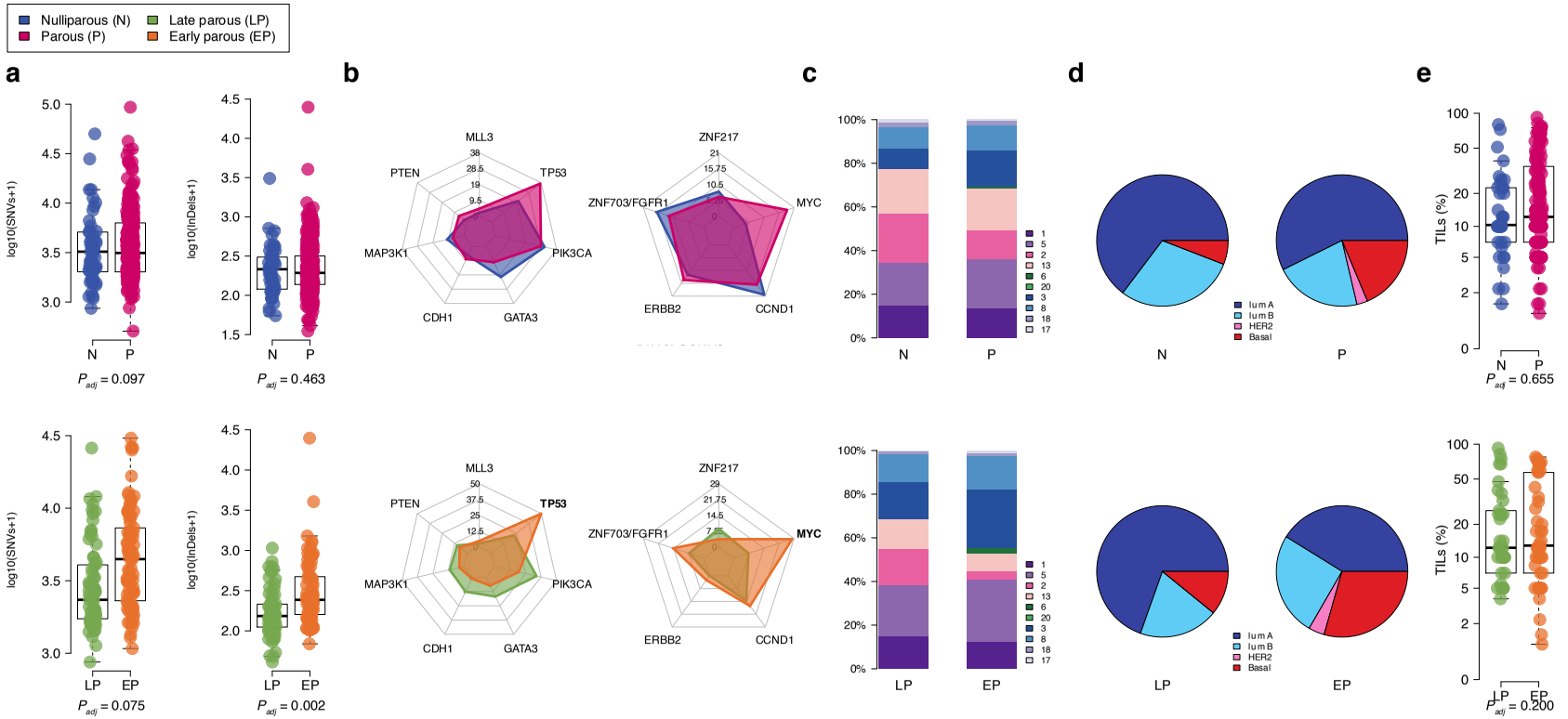


Figure 1.1 – Imprint of pregnancy and age at first pregnancy on breast cancer biology.

a, *Comparison of SNVs and Indels between nulliparous and parous (upper) and between early and late parous (bottom).* ***b,*** *Radar plots showing the frequency of driver mutations and driver SNCAs between nulliparous and parous (upper) and between early and late parous (bottom) patients. Significant genes independently associated with parity or age at first pregnancy are highlighted in bold.* ***c,*** *Proportion of breast cancer substitution signatures in nulliparous and parous (upper) and in early and late parous (bottom) patients.* ***d,*** *Proportion of PAM50 breast cancer subtypes in nulliparous and parous (upper) and in early and late parous (bottom) patients.* ***e,*** *Comparison of TILs levels (%) between nulliparous and parous (upper) and between early and late parous (bottom) patients. Padj, P-values derived from multivariate linear regression analysis adjusted for potential confounders.*

3.3 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 (Curtis et al., 2012; TCGA, 2012). We identified five driver genes with a frequency of SCNAs > 5% across all patients (Supplementary Data File 1: Table S3). *MYC* tended to be more frequently amplified in parous than in nulliparous patients (18.6% vs. 4.1%; P_{adj} = 0.052; *FDR* = 0.26, Figure 1.1b). In the parous group, *MYC* amplification was significantly more frequent in the early parous group than in the late parous group (28% vs. 7%; P_{adj} = 0.008; FDR = 0.040, Figure 1.1b). 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 (18.3% vs. 4.2%; P_{adj} = 0.087, Figure 1.2a and Supplementary Figure 1.4). Taken together, these results suggest that age at first pregnancy may also shape the somatic copy number alterations profiles of subsequent breast cancer.

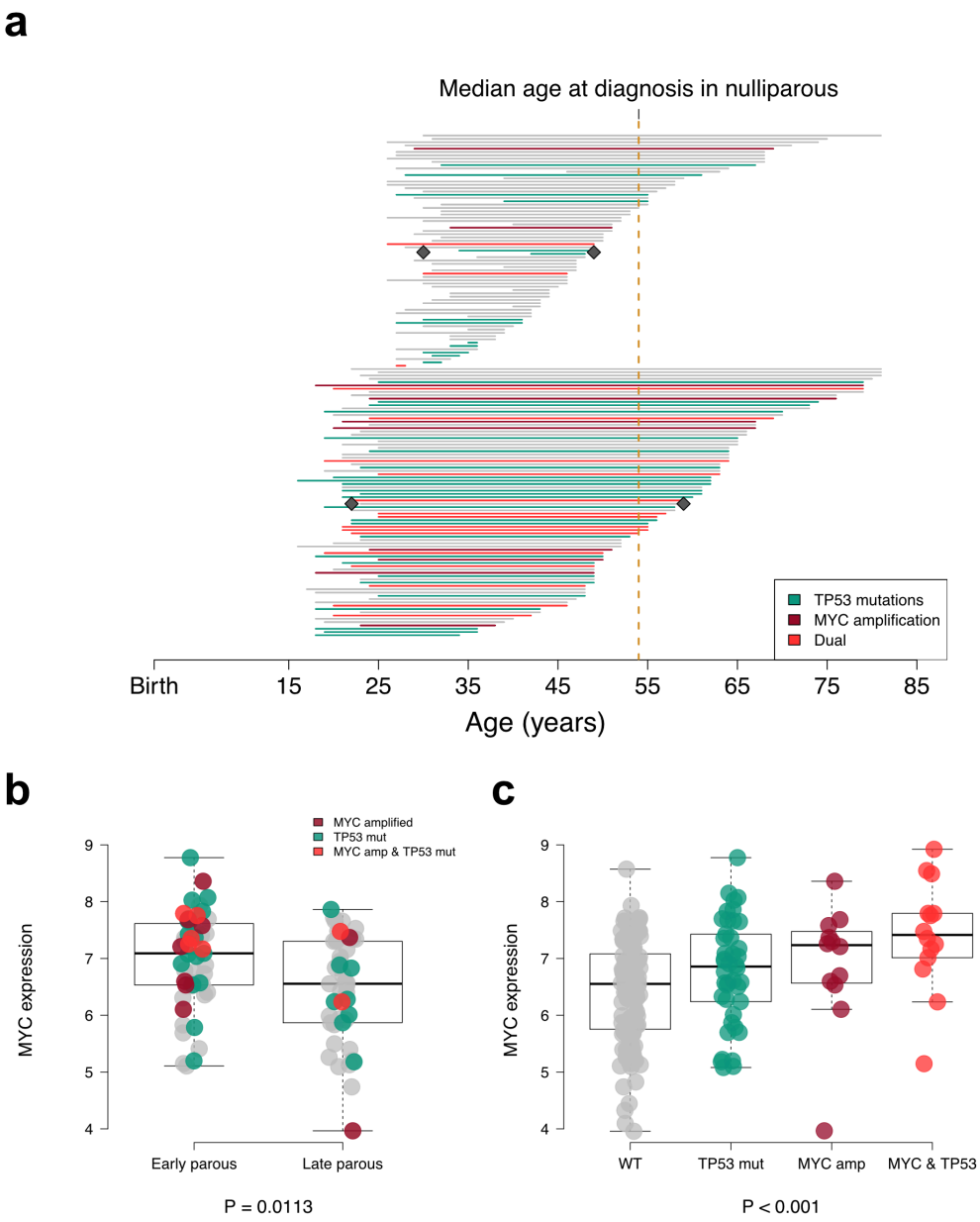


Figure 1.2 – Co-occurrence of MYC amplification and TP53 mutations is associated with age at first pregnancy

a, *Timeline of 153 patients with available data on age at first pregnancy. Each line represents an individual patient from age at first pregnancy (start of the line) to age at breast cancer diagnosis (end of the line). Late parous patients (upper) and early parous patients (bottom) are grouped according to median age at first pregnancy. Grey diamond represents the median age at first pregnancy and at diagnosis in the two groups. Lines are colored according to TP53 mutations (green) MYC amplification (dark red) and the co-occurrence of both (red).* ***b,*** *Comparison of MYC expression in early and late parous patients. Each dot represents an individual patient and is colored according to TP53 mutations (green) MYC amplification (dark red) and the co-occurrence of both (red). P-value is derived from multivariate linear regression analysis adjusted for potential confounders.* ***c,*** *MYC expression according to TP53 mutations, MYC amplification or the co-occurrence of both. P-value is derived from the Kruskal–Wallis test.*

3.4 The influence of parity and age at first pregnancy on mutational signatures

To have a better understanding of the mutational processes that have occurred during the course of cancer according to reproductive history, we examined the contribution of mutational substitution signatures known to occur in breast cancer (Nik-Zainal et al., 2016) (Supplementary Data File 1: Table S4). The distribution of mutational signatures was similar between parous and nulliparous patients. However, in the parous group, signatures 2 was more prevalent in late parous patients (P_{adj} = 0.001; *FDR* = 0.011, Figure 1.1c). Of interest, two early parous patients with an exceptionally high number of Indels were associated with mutational processes 6 and 26, attributable to mismatch repair deficiency (Alexandrov et al., 2013) (Supplementary Figure 1.4). Similarly, to our previous findings, these results highlight that age at first pregnancy may be associated with specific mutational processes in subsequent breast cancer.

3.5 The influence of parity and age at first pregnancy on BRCAness

Here, we investigated the BRCAness status of tumors according to reproductive history. Since germline BRCA1/2 mutation status was not available for all samples, we used HRDetect score to identify *BRCA1*/*BRCA2*-deficient samples (Davies et al., 2017a). We did not find any significant differences in the proportion of *BRCA1*/*BRCA2*-deficient patients between nulliparous and parous groups (12.24% vs. 18.94%, respectively, P_{adj} = 0.473), nor between early and late parous group (30.49% vs. 19.71%, respectively, P_{adj} = 0.386). Thus, reproductive history and age at first pregnancy do not seem to affect homologous recombination DNA repair capacity in subsequent breast cancer.

3.6 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 (Supplementary Figure 1.2 and Supplementary Data File 1: Table S5). We first determined the intrinsic molecular subtypes distribution of breast cancer using the PAM50 classifier (Parker et al., 2009). We did not find a significant difference in the distribution of the PAM50 subtypes between nulliparous and parous (Figure 1.1d). In contrast, early parous patients had a higher proportion of basal-like subtype tumors (29.4% vs. 8.9%; *P* = 0.009, Figure 1.1d, Supplementary Data File 1: Table S5). 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 (Love et al., 2014) 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 (Supplementary Data File 1: Table S6). Among these genes, three were associated with mammary development; *OXTR* and *ATP2B2* were down-regulated whereas *NRG3* was up-regulated in nulliparous. Pathway analysis using the generally applicable gene-set enrichment (GAGE) analysis (Luo et al., 2009) revealed an enrichment of genes related to extracellular matrix (ECM) receptor interaction (Supplementary Data File 1: Table S7). When comparing early and late parous patients, 466 genes were differentially expressed, among which 305 were up-regulated in early parous (Supplementary Data File 1: Table S8). However, pathway analysis did not reveal any significant enrichment of relevant biological processes (Supplementary Data File 1: Table S9).

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 up-regulation of *MYC* expression (P_{adj} = 0.0113, Figure 1.2b). 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 (Figure 1.2c).

Signature 2 and 13 have been attributed to the activity of the AID/APOBEC family of cytidine deaminases converting cytosine to uracil (Alexandrov et al., 2013). Thus, we inspected if the higher prevalence of signature 2 observed in late parous patients was associated with higher expression of AID/APOBEC expression. As opposed to signature 13, which was positively correlated with all APOBEC3s family members and AID expression, signature 2 was not significantly associated with the expression of any of these deaminases (Supplementary Figure 1.5).

3.7 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 (Agrawal et al., 1995; Arklie et al., 1981; Erlebacher, 2013; Finn et al., 1995; Jungbluth et al., 2007). Therefore, we assessed

whether reproductive history could be associated with tumor infiltrating lymphocyte (TILs) level that is considered as a surrogate of tumor immunogenicity (Supplementary Data File 1: Table S10). 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} = 0.325, respectively, Figure 1.1e). 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). Thus, reproductive history does not seem to influence breast tumor immunogenicity.

3.8 Pregnancy-associated breast cancers are associated with increased TILs infiltration

Pregnancy-associated breast cancer (PABC) can be defined as cases diagnosed up to 10 years postpartum (Lyons et al., 2009). In this cohort, we identified 17 PABC patients and compared them with 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; $P = 5.79 \times 10^{-6}$, Supplementary Data File 1: Table S11) and higher frequency of TNBC (29.4% vs. 4.1%; P = 0.021, Supplementary Data File 1: Table S11). We did not find any significant differences in the pattern of somatic mutations, somatic copy number alterations (SCNAs) nor in the distribution of mutational signatures (Supplementary Data File 1: Table S1-S4). Nine PABC had available gene expression and TILs scoring. At the transcriptomic level, we found that PABC patients were associated with enrichment of biological processes related to immune function (Figure 1.3a and Supplementary Data File 1: Table S12). Moreover, PABC patients had an increased lymphocytic infiltration both for stromal and intratumoral TILs levels (P_{adj} = 0.040 and $P_{adj} < 0.0001$, respectively; Figure 1.3b,c). Taken together these results indicate that cancer occurring in the postpartum mammary gland is associated with increased inflammation.

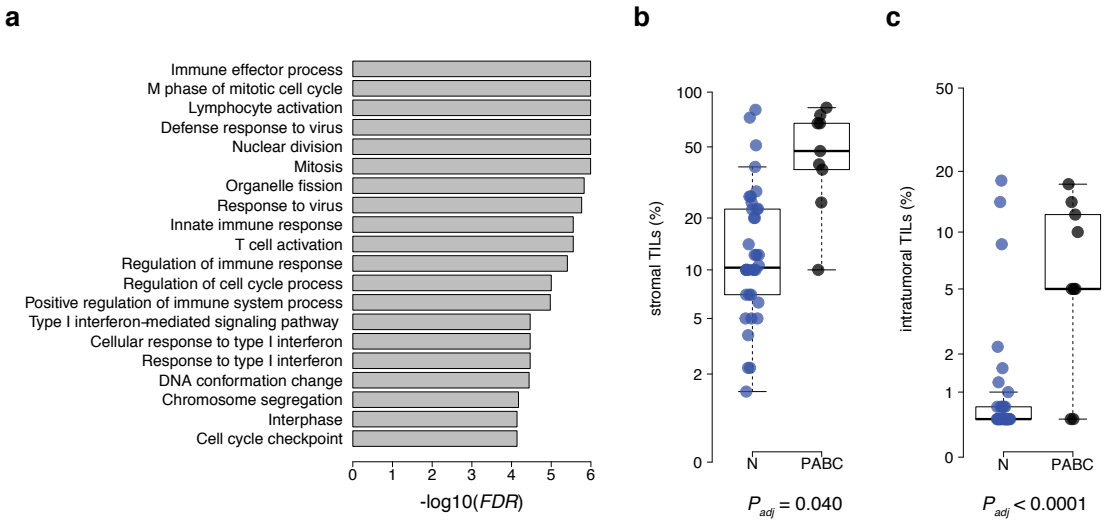


Figure 1.3 – PABC patients are associated with higher TILs levels

a, *Results form the GAGE analysis showing the top 20 most significant biological processes enriched in PABC patients.* ***b,*** *Comparison of stromal and* ***(c)*** *intratumoral (right) TILs levels (%) between nulliparous (N) and PABC.* P_{adj}*, P-values derived from multivariate linear regression analysis adjusted for potential confounders.*

4 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 (Kroman et al., 1997; MacMahon et al., 1970; Papatestas et al., 1980; Rosenberg et al., 2004; Trichopoulos et al., 1983), this analysis provides further insights on the differences at the pathologic, genomic, transcriptomic and immunogenic 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 and a lower prevalence of mutational signature 2, while PABC patients exhibited higher TILs 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 shown 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 (Ellingjord-Dale et al., 2017; Lambertini et al., 2016; Ritte et al., 2013; Yang et al., 2011). For the first time to the best of our knowledge, we have documented that age at first pregnancy is negatively associated with age at diagnosis irrespective of classical clinicopathological features. This observation, which needs to be validated in larger cohorts, is in line with the reported protective effect of early pregnancy on breast cancer risk.

Our study reveals that age at first pregnancy has a bigger imprint on genomic alterations of breast cancer than parity status alone. However, the apparent lack of impact of parity could also be related to the relatively low number of nulliparous patients.

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 (Leiserson et al., 2016). Previous reports have suggested that *TP53* mutations are a common mechanism that disturbs the apoptotic pathway in *MYC*-driven tumors (Wolf et al., 2015). 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 (Hermeking and Eick, 1994). *TP53* has long been recognized as a potential mediator of pregnancy-induced resistance to mammary carcinogenesis. It has been shown that p53 and its downstream transcriptional target p21 are increased in parous and estrogen/progesterone-treated mammary epithelium in response to carcinogen (Sivaraman et al., 2001). In the absence of p53, the protection given by parity or exogenous hormones is lost (Dunphy et al., 2008; Jerry et al., 2000; Medina and Kittrell, 2003). We hypothesized that the higher frequency of *TP53* mutation observed in breast cancer from early parous woman could be explained by the fact that an early pregnancy might protect less effectively against *TP53* mutated breast cancer. 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 (TCGA, 2012). The differential effect of parity-induced protection according to *TP53* mutational status might also explain the differential effect of parity-induced protection according to tumor subtypes.

CDH1 mutations have been associated with invasive lobular breast cancer subtype (Desmedt et al., 2016). 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. The lower level of mutational signature 2 seen in early parous is not explained by a difference in the expression of AID/APOBEC family but could be related to other factors that remained to be determined.

At the mRNA level, we found that age at first pregnancy had a stronger impact on the transcriptome than parity status alone, indicating again that age at first pregnancy might be the most critical factor. In the normal tissue of parous women, the gene encoding oxytocin receptor (*OXTR*) is physiologically up-regulated during lactation and has been shown to remain overexpressed later in life (Peri et al., 2012; Russo et al., 2012).

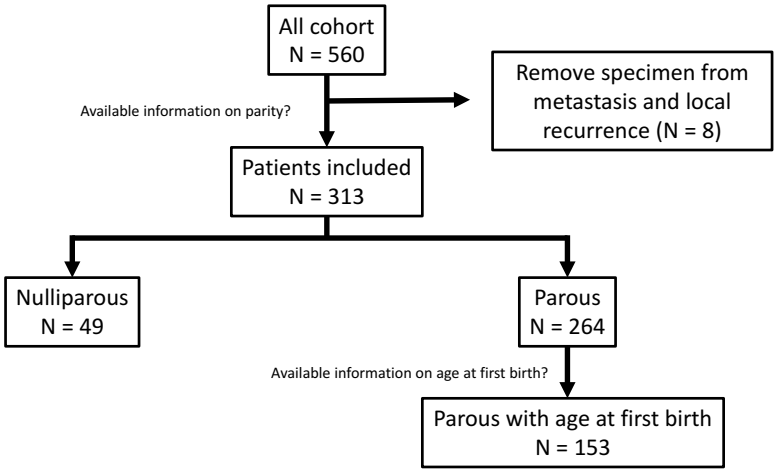
Noteworthy, the expression of *OXTR* was higher in parous compared to nulliparous patients. However, due to the lack of functional studies, it is not clear whether this gene is involved in the tumorigenesis of breast cancer or simply related to physiological changes induced by pregnancy. The enrichment 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 (McDaniel et al., 2006; Schedin, 2006). 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 (Lyons et al., 2011; McDaniel et al., 2006).

Finally, previous reports have hypothesized that the pregnancy-induced tumor protection could be attributable to an improved anti-tumor immunity (Agrawal et al., 1995; Arklie et al., 1981; Erlebacher, 2013; Finn et al., 1995; Jungbluth et al., 2007). 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 caused by involution could contribute to tumor promotion (Fornetti et al., 2014; Harvell et al., 2013; Martinson et al., 2015). We observed an increase of TILs levels in PABC patients but the composition of the immune infiltrate has still be determined to validate this hypothesis.

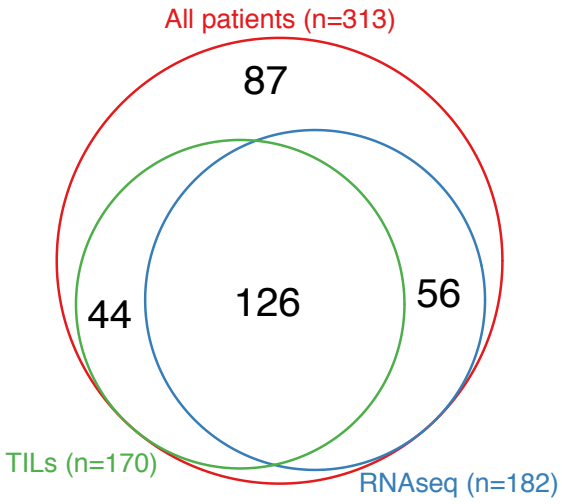
A potential limitation of our study is the lack of data on other reproductive factors (e.g., breastfeeding, age at menarche and time since last pregnancy) that could also potentially imprint the genomic alterations of breast cancer. Indeed, breastfeeding and age at menarche have also been linked to breast cancer risk, but since they are often self-reported, they are more difficult to assess reliably (Lambertini et al., 2016). Another limitation is the absence of HER2+ subtype in the transcriptomic analysis.

In conclusion, our findings highlight an unprecedented link between reproductive factors and the genomic landscape of subsequent breast cancer. Specifically, our analysis suggests that age at first pregnancy, a known breast cancer risk factor, adds a layer of biological complexity to subsequent breast tumors. 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.

5 Supplementary Materials

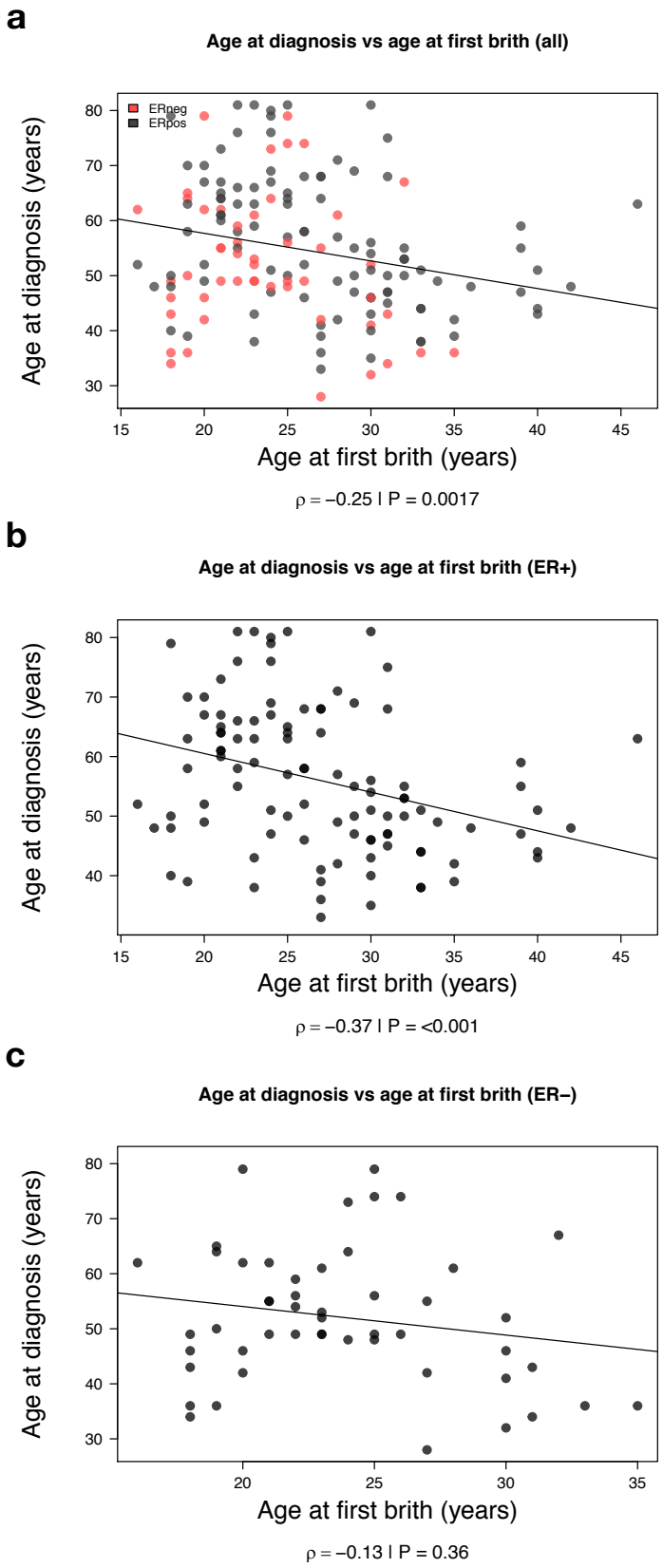


Supplementary Figure 1.1 – Flowchart summarizing the number of patients included in the analyses and the reasons for inclusion and exclusion



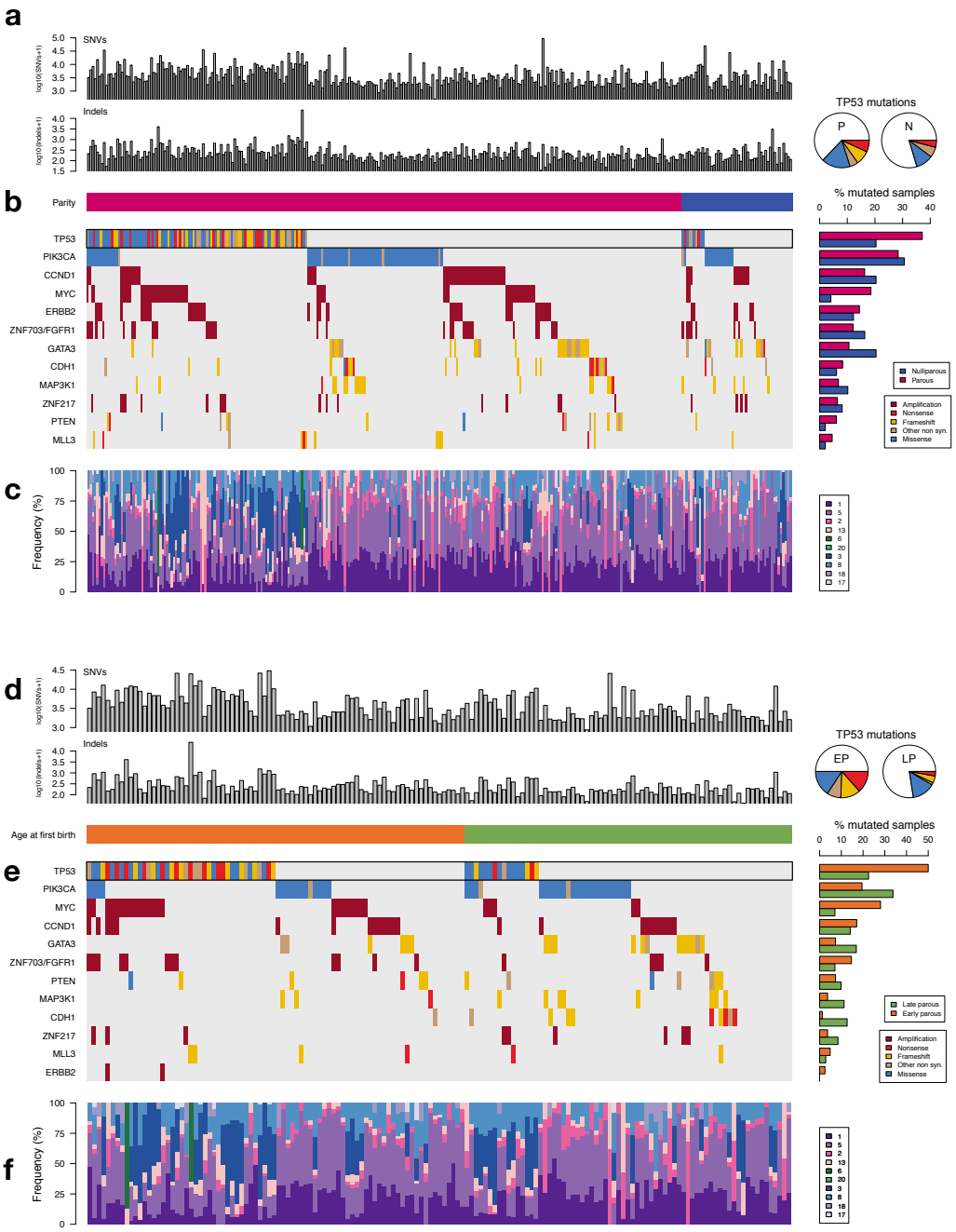
Supplementary Figure 1.2 – Venn diagram summarizing the number of patients with available data

A total of 313 patients (100%) had available somatic mutations and somatic copy number alterations (SCNAs) data (red circle) while 182 patients (58.1%) had available transcriptomic data (blue circle) and 170 patients (54.3%) had available information on TILs (green circle).



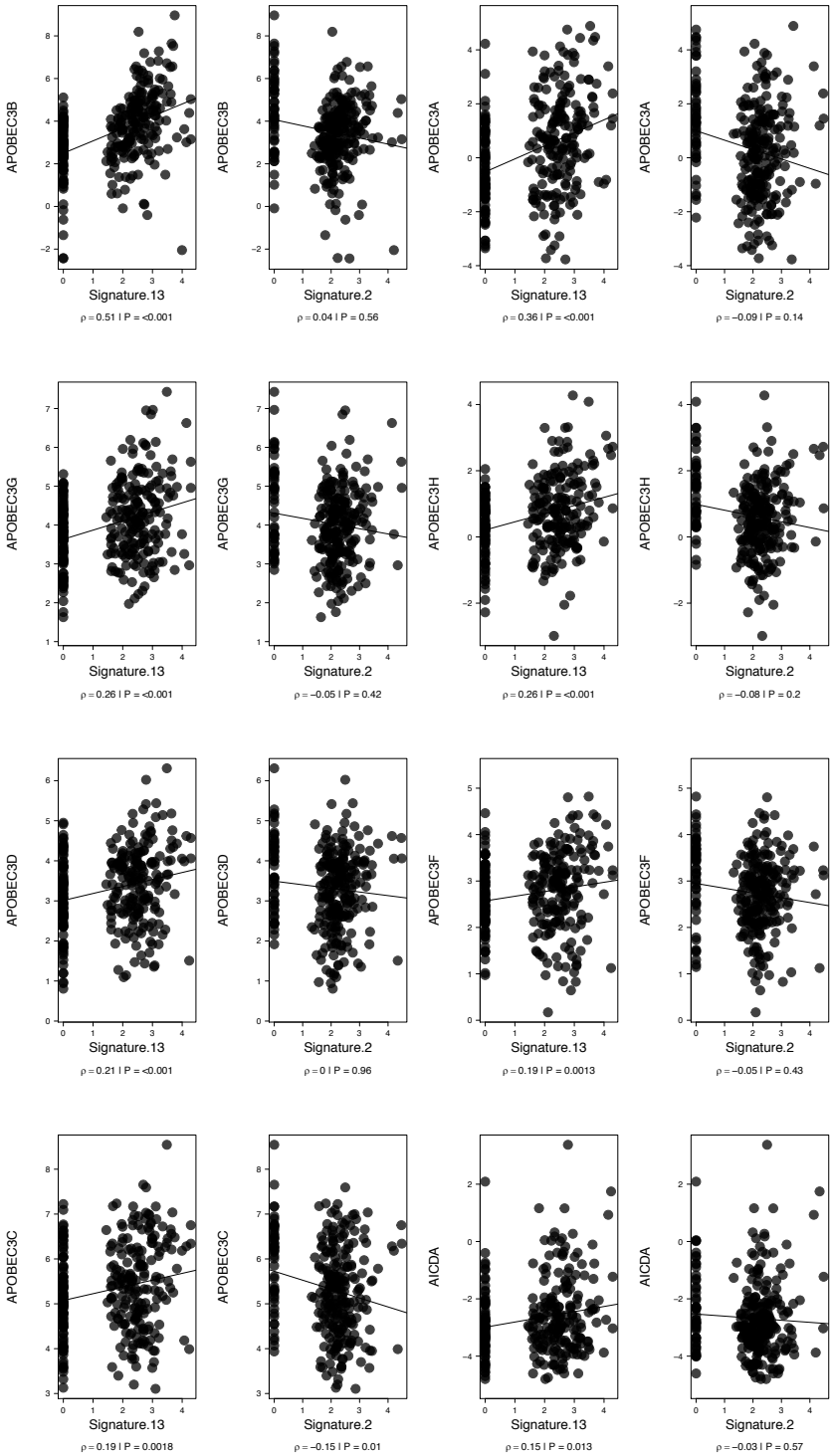
Supplementary Figure 1.3 – Scatterplot showing the correlation between age at diagnosis and age at first pregnancy

a, *All patients.* ***b,*** *ER positive.* ***c,*** *ER negative. P; P-values are derived from the Spearman's rank correlation.*



Supplementary Figure 1.4 – Genomic landscape of breast cancer according to pregnancy and age at first pregnancy

a, *Bar chart representing the absolute number of substitutions and indels in parous and nulliparous patients.* ***b,*** *Co-mutation plot showing genes harboring at least one non-silent mutation with a frequency of at least 5% across the whole cohort, and their corresponding frequencies in parous and nulliparous patients.* ***c,*** *Proportion of breast cancer substitution signatures in each sample. Signatures are colored according to broad biological groups: 1 and 5 are associated with clock-like processes, 2 and 13 are APOBEC-related, 20 and 26 are associated with mismatch- repair deficiency, 3 and 8 are associated with homologous-recombination deficiency.* ***d,*** *Bar chart representing the absolute number of substitutions and indels in early and late parous patients.* ***e,*** *Co-mutation plot showing genes harboring at least one non-silent mutation with a frequency of at least 5% across the whole cohort, and their corresponding frequencies in early and late parous patients.* ***f,*** *Proportion of breast cancer substitution signatures in each sample. Signatures are colored according to broad biological groups: 1 and 5 are associated with clock-like processes, 2 and 13 are APOBEC-related, 20 and 26 are associated with mismatch- repair deficiency, 3 and 8 are associated with homologous-recombination deficiency.*



Supplementary Figure 1.5 – Correlation between AID/APOBEC3s family of cytidine deaminases expression and presence of mutational signature 2 and 13

P; P-values are derived from the Spearman's rank correlation.

CHAPTER 2: THE MOLECULAR CHARACTERIZATION OF BREAST CANCER DIAGNOSED DURING PREGNANCY

This research work is related to the following publication:

Nguyen, B., Venet, D., Azim, H.A., Brown, D., Desmedt, C., Lambertini, M., Majjaj, S., Pruneri, G., Peccatori, F., Piccart, M., Rothé, F. and Sotiriou, C. (2018). Breast cancer diagnosed during pregnancy is associated with enrichment of non-silent mutations, mismatch repair deficiency signature, and mucin mutations. Npj Breast Cancer *4*, 23.

Available online at: https://doi.org/10.1038/s41523-018-0077-3

My contribution to this study involves:

- Conceptualization and Design
- Methodology
- Formal Analysis
- Literature search and Interpretation
- Statistical analyses
- Visualization
- Writing of the manuscript
- Presentation of the results at the following meetings:
    - Annual BSMO Meeting, January 2016, Brussels, BELGIUM (selected for an oral abstract session)
    - San Antonio Breast Cancer Symposium, December 2016, San Antonio, TX, US
    - AACR Annual Meeting, April 2017, Washington, DC, US
    - San Antonio Breast Cancer Symposium, December 2017, San Antonio, TX, US (awarded with a Basic Science Scholar Award)
    - Televie's Researchers Seminar, February 2018, Brussels, BELGIUM (awarded with the Best Oral Presentation Award)

1 Introduction

Breast cancer is the most frequently diagnosed malignancy during pregnancy (Anderson, 1979). Its incidence is increasing given the rising trend of delayed childbearing (Loibl et al., 2015a). Given its rarity, few dedicated studies were performed so far; hence, our understanding of these tumors remains poor. The clinical management of these patients follows standard guidelines with only minor adaptations according to gestational age, maternal wishes and fetal considerations (Loibl et al., 2015a). Therefore, the molecular characterization of BCP goes beyond academic curiosity as it is of utmost clinical interest to determine if these patients should be treated similarly to non-pregnant breast cancer patients. In this report, we aimed to identify specific molecular alterations characterizing BCP by combining whole genome sequencing, copy number alteration, and gene expression data.

2 Methods

2.1 Patients and samples

A total of 167 patients with primary breast cancer were retrospectively included in this study, 54 of whom were diagnosed during pregnancy. All patients were diagnosed and followed up at the European Institute of Oncology (IEO, Milan, Italy) from 1996 to 2010. As previously described (Azim et al., 2012b), this is a case-control study, in which pregnant breast cancer patients and controls were matched according to age, tumor size, nodal status, and date of diagnosis. For the current genomic analysis, we opted to exclude patients who received neoadjuvant therapy to avoid potential impact of treatment on the obtained results. The majority were treated with anthracycline-based regimen (individual patients data are presented in Supplementary Data File 2: Table S1). All patients had available formalin-fixed paraffin-embedded (FFPE) tissue from the primary tumor resection, and there was only one tumor sample per patient. All control patients were premenopausal at the time of diagnosis. ER/PR-status were defined by ASCO-CAP. For the classification of Luminal A and B we used a cut-off of Ki-67 > 20% according to St Gallen 2015 Consensus Meetings (Coates et al., 2015). Matched normal tissues were collected from histologically confirmed tumor-free axillary lymph nodes or tumor-adjacent normal tissue and there was only one normal sample per patient. FFPE tissue sections were deparaffinized by xylene followed by a 100% ethanol wash. DNA extraction was performed using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations. The quantity of double-stranded DNA was evaluated using the Qubit dsDNA BR Assay Kit. For the WGS, we selected 18 control patients based on major clinicopathological features of the 35 BCP patients, namely age at diagnosis, ER status, and grade. All patients provided written informed consent for the use of tissue samples for research purposes as per the IEO institutional policies. This study was approved by the Ethics Committee of Institut Jules Bordet (Number 1782). The validation of the enrichment of mutations in the mucin family genes in BCP were done by comparing the frequency of these mutations in BCP patients with putatively non-pregnant patients retrieved from the TCGA dataset (TCGA, 2012) and selected to have similar age, estrogen receptor (ER) and progesterone receptor (PR) distribution ($N = 56$) (Supplementary Table 2.1). The validation of the enrichment of signature 20 in BCP were done by comparing the frequency of this signature in BCP patients with putatively non-pregnant patients retrieved from the 560 breast cancer

dataset (Nik-Zainal et al., 2016) (referred to as BRCA560) and selected to have similar age, ER and PR distribution ($N = 64$).

2.2 Transcriptomic profiling

All samples were hybridized on Affymetrix Human Genome U219 array plates following the manufacturer's protocol, as described before (Azim et al., 2014). The metagene signature MUCsig was calculated by taking the mean expression level of all genes present in the mucin family, scaled to a standard deviation of one and centered around zero. The publicly available murine dataset derived from the normal breast of pregnant mice (GEO ID: GSE8191 (Anderson et al., 2007)) was used to evaluate mucin expression in the normal breast during pregnancy. Ensembl database was used to convert mouse gene names to the human equivalent.

2.3 Genome-wide copy number analysis

Hematoxylin and eosin slides from the archived FFPE blocks were reviewed by a pathologist (G.P.) to confirm diagnosis and evaluate tumor content. Samples with tumor purity below 60% were macrodissected ($N = 56$). DNA was extracted as described above. A total of 80 ng of DNA was used for copy number profiling using the Affymetrix OncoScan® FFPE Assay Kit according to the manufacturer's instructions. The raw intensity values from the scanned chips were normalized to obtain Log2 ratios, B Allele Frequencies (BAF) and genotyping calls (AA/AB/BB) using Affymetrix Power Tools. We used release NA.33 of the NetAffx library for the reference model and annotation. We computed the Median Absolute Pairwise Deviation (MAPD) and the Median Auto-Correlation (MAC) from the normalized Log2 ratios as quality control metrics and used a threshold of 0.30 and 0.5, respectively, to flag failed arrays.

We used two parallel approaches involving (a) allele-specific copy number analysis using heterozygous SNP probes and (b) total copy number analysis using the full set of 200 K markers and parameters from (a) to control for the cancer cell fraction (CCF) and genomic mass. From the BAF and genotyping calls, only informative SNP probes displaying heterozygous genotype (AB) and $0.1 < \text{BAF} < 0.9$ were kept for analysis at (a). The Log2 ratios and BAF were smoothed using the median absolute deviation and segmented jointly using a multitrack segmentation algorithm from the library copynumber (Nilsen et al., 2012) to determine common breakpoints. Estimates of CCF

and genomic mass were obtained using GAP (Popova et al., 2009). Samples with a CCF lower than 30% were further excluded. For analysis at (b), the Log2 ratios for the same samples analyzed at (a) were segmented by penalized least square regression as above and non-rounded estimates of copy numbers *y* were obtained as

$$y = \frac{1}{a}\left(2^{\frac{x}{c}}\left(\psi a + 2(1-a)\right) - 2(1-a)\right)$$

where a is the CCF and ψ is the genomic mass, both estimated at (a). c = 0.8 is a constant representing the compression ratio of the array and finally x is the observed Log2 ratio of a given segment. The copy numbers were categorized as deletions (−1) if $y < \psi - 0.5$, gains (+1) if $y > \psi + 0.5$, amplifications (+2) if $y > \psi + 2.5$, and copy neutral (0) otherwise. Unless otherwise stated, all parameter settings were kept at default values and all computations were done using R/Bioconductor. Segmented data were used as input for Genomic Identification of Significant Targets in Cancer, version 2.0 (GISTIC 2.0) (Mermel et al., 2011) and version 6.2 on the Broad Institute GenePattern cloud server to obtain somatic focal and broad CNA events. These were then parsed in R. For focal events, only "high-level" focal amplification events, defined as log2 ratio >0.9 were retained, whereas focal losses were retained with a log2 ratio >0.3 and with a Q value <0.25. Broad events, defined as arm-level events encompassing 98% or more of a chromosome arm, were computed using GISTIC as well. For gene-levels analysis, we also used the gene level output given by GISTIC analysis. Unless otherwise stated, all parameter settings were kept at default values and all computations were done using R/Bioconductor.

2.4 Library preparation and whole genome sequencing

For each of 53 patients, two samples of 1 μg genomic DNA from tumor and histologically normal axillary lymph nodes were whole genome sequenced at The McDonnell Genome Institute at Washington University (St Louis MO, USA) on an Illumina HiSeqX platform. Briefly, manual dual-indexed libraries were constructed with 1 ug of FFPE genomic DNA for the 53 tumor/normal pairs using the Accel-NGS 2S Plus Library Kit (Swift, MI, USA). Samples were fragmented on the Covaris LE220 instrument with 350bp target insert size. PCR cycle optimization was performed to prevent over-amplification of the libraries. The concentration of each library was determined through qPCR (Kapa

Biosystems, MA, USA). For the normal samples, each library was loaded on one lane of a HiSeqX flow cell, whereas for tumor samples, each library was loaded across two lanes of a HiSeqX flow cell. 2x150 paired-end sequence data were generated at a target depth of 30x (normal) and 60x (tumor) haploid genome coverage. All sequencing data are available in EGA under accession "EGAS00001002685".

Adapters were trimmed using Trimmomatic (Bolger et al., 2014). Paired sequence reads were aligned to reference human genome build hg19/GRCh37 using bwa mem (Li and Durbin, 2009). Marking and removal of duplicates were done using biobambam (Tischler and Leonard, 2014) while bases in overlapping reads from the same read pair were removed with BamUtil clipOverlap. Somatic mutations were called using Strelka (Saunders et al., 2012). Except for mutations identified in the COSMIC database (Forbes et al., 2017) related to breast cancer, mutations were filtered using default quality thresholds with a QSS_NT > 15 for SNVs and a QSI_NT> 30 for Indels. We also filtered out SNVs when <5 sequence reads reported a variant allele in the tumor. Variant annotation and effect prediction were carried out using SnpEff (v. 4.3p) (Zhang et al., 2012) and Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Mutations with putative impact were defined as those assigned a high or moderate impact from SnpEff. To predict the pathogenicity of non-synonymous SNVs we also used a battery of in silico algorithms (SIFT, PolyPhen and ConDel) (Adzhubei et al., 2010; González-Pérez and López-Bigas, 2011; Sim et al., 2012).

2.5 Tumor heterogeneity

To quantify the level of intra-tumor heterogeneity present in a sample, we used the MATH score as previously described (Mroz and Rocco, 2013);

$$MATH = \frac{\text{MAD}(VAFs)}{\text{median}(VAFs)}$$

where MAD(*VAFs*) is the median absolute deviation of the variant allele fractions (*VAFs*) of all the mutations (coding and noncoding) in a tumor.

2.6 Mutational signature

All samples were analyzed using deconstructSigs (Rosenthal et al., 2016) to extract signatures based on the Wellcome Trust Sanger Institute Mutational Signature Framework.

2.7 Significance of the missense mutation in mucins producing a serine

Because (i) the distribution of bases is not uniform, (ii) the distribution of the different missense mutations is not uniform, and (iii) the distribution of bases within each mucin varies, we used an empirical approach to test the significance of the missense mutation in mucins producing a serine. First, we defined the probability of mutation at a base *b* (*b* in A, C, G, T) P_b*:*

$$P_b = \frac{P_b^n}{P_b^g}$$

where P_b^n is the proportion of the bases *b* that are affected by a missense mutation in the whole cohort and P_b^g is the proportion of the base *b* in the coding bases of the whole genome. For each mucin, we determine P_{muc} as the probability of having a mutation in base *b* in a mucin *muc*:

$$P_b^{muc} = \frac{\left(P_b \cdot \acute{P}_b^{\mathrm{muc}}\right)}{\sum_{b\prime}\left(P_{b\prime} \cdot \acute{P}_{b\prime}^{\mathrm{muc}}\right)}$$

where P_b is the baseline probability above-mentioned and $\acute{P}_b^{\mathrm{muc}}$ is the proportion of the coding bases present in mucin *muc*. Using P_b^{muc} we artificially generated 1000 random missense mutations in each mucin and calculated the proportion of altered codon producing a serine. For each random mutation, we first drew randomly the type of base mutated base on P_b^{muc}. The resulting nucleotide was randomly chosen based on the observed mutations. The precise nucleotide mutated among those corresponding to the base mutated in the mucin was chosen randomly. If the resulting mutation did not lead to a missense mutation it was discarded. The number of serine obtained after 1000 random mutations was used to derive a probability to obtain a serine by the play of chance in each mucin: P_{muc}^{ser}.

To estimate the probability of observing *N* serines in total in the mucin mutated, we again used a Monte-Carlo method. We simulated 10^5 scenarios, each with the same mucin mutated as in the real dataset and drawn randomly whether each mutation led to a serine, using P_{muc}^{ser}.

Finally, the empirical p-value was calculated by using the Monte-Carlo procedure (North et al., 2002)

$$p = \frac{(r+1)}{(n+1)}$$

where *r* is the number of simulations that produced at least *N* serines and *n* is the number simulations (source code available upon request).

2.8 Statistical analysis and survival analysis

Except for age and date of diagnosis that were considered as continuous variables and therefore compared using the non-parametric Mann–Whitney *U* test, differences in other clinicopathological characteristics between BCP and controls were analyzed using the χ^2 test or the Fisher exact test when appropriate. All statistical tests comparing BCP and controls were done using the non-parametric Mann–Whitney *U* test and the Fisher exact test for continuous and categorical variables, respectively. Independent association between continuous and binary variables with BCP vs. controls was investigated using linear and logistic regressions, respectively. All multivariate tests were adjusted for age at diagnosis, date of diagnosis, pathological stage, and molecular subtypes by IHC. All interaction and multivariate tests were done using analysis of variance to compare the models with and without the extra term.

All correlations were measured using the non-parametric Spearman's *rho* coefficient. Reported *P*-values were two-tailed, and differences were considered significant when the *P*-value was less than 0.05. When applicable, multiple testing correction was done using the false discovery rate method (FDR) (Benjamini and Hochberg, 1995), FDR below 0.05 being considered significant. All analyses were done in R software version 3.3.2 (available at www.r-project.org) and Bioconductor version 3.4.

Survival endpoint was disease-free survival (DFS) and calculated from the date of surgery to any loco-regional or distant recurrence, contralateral BC, other primary tumor

or death from any cause, whichever occurred first. In the absence of any of the above-mentioned events, survival was censored at the last follow-up visit or phone call with the patient. Survival curves were estimated using the Kaplan-Meier method and compared by the log-rank test. The prognostic impact of pregnancy on survival was evaluated using univariate and multivariate Cox proportional hazards regression models and expressed as hazard ratio (HR) with 95% confidence intervals (CI). Multivariate analysis was adjusted for standard clinical prognostic factors (age at diagnosis, date of diagnosis, pathological stage, and molecular subtypes by IHC).

3 Results

A total of 167 patients with primary breast cancer were retrospectively included in this study, 54 of whom were diagnosed during pregnancy. Detailed patient characteristics were previously published (Azim et al., 2012b). At a median follow-up of 9 years, median disease-free survival (DFS) time of BCP was 9.8 years vs. 12.5 years in controls (P = 0.041, log rank test, Supplementary Figure 2.1a). Observed 5-year overall survival (OS) rate was 95.5% vs. 85.1% in BCP and control, respectively; median OS time was not reached within the time frame of the study (Supplementary Figure 2.1b). In a multivariable Cox proportional hazards regression of DFS and OS, adjusted for age at diagnosis, date of diagnosis, pathological stage and molecular subtypes by IHC, we found that BCP was associated with worse DFS (multivariable hazard ratio [mHR] 1.81; 95% CI 1.09-3.01, P = 0.024) and OS (mHR 2.53; 95% CI 1.20-5.36, P = 0.017) (detailed survival data is provided in Supplementary Data File 2: Table S1).

3.1 BCP and controls have similar somatic copy number alteration profiles

We first sought to investigate whether tumors from BCP patients show distinct copy number alterations (CNAs) compared to tumors from matched non-pregnant breast cancer patients (controls). Hence, we performed genome-wide copy number alterations profiling on 160 formalin-fixed paraffin-embedded (FFPE) primary tumor samples from 52 BCP patients and 108 controls. Of note, gene expression data were available for all patients as previously described (Azim et al., 2014). After quality control, CNA profiles were obtained for 125 tumor samples (78%) from 38 BCP and 87 controls. The main reason for exclusion was low cancer cell fraction (CCF < 30%) as estimated with the Genome Alteration Print algorithm (Popova et al., 2009) (Supplementary Figure 2.2). No differences in clinicopathological features were observed between BCP and controls (Supplementary Table 2.2). We found no significant differences between BCP and controls in terms of cancer cell fraction, ploidy, and fraction of genome altered (Figure 2.1a-c). Moreover, no significant differences were observed between the CNA profiles of the two groups neither at the segment nor at the chromosome arm levels, including the gains of 1q and 8q and loss of 8p, reported to frequently occur in breast cancer (Mika et al., 1998) (Figure 2.1d). We also compared CNAs profiles by intrinsic subtypes as

defined by PAM50 and found no significant differences (Supplementary Figure 2.3 and Supplementary Data File 2: Table S2).

We next focused our analysis on the 35 genes that were previously identified as CNA drivers in breast cancer (Nik-Zainal et al., 2016). As expected, *MYC* oncogene was the most frequently gained/amplified whereas *TP53* tumor suppressor gene was the most frequently lost/deleted across the whole cohort (Figure 2.1e). Using GISTIC2.0 (Mermel et al., 2011), we identified 22 focal amplifications and 23 focal deletions and found no differences between their prevalence in the two groups (Supplementary Figure 2.4). Taken together, these results suggest that the CNA profiles of BCP and controls are similar.

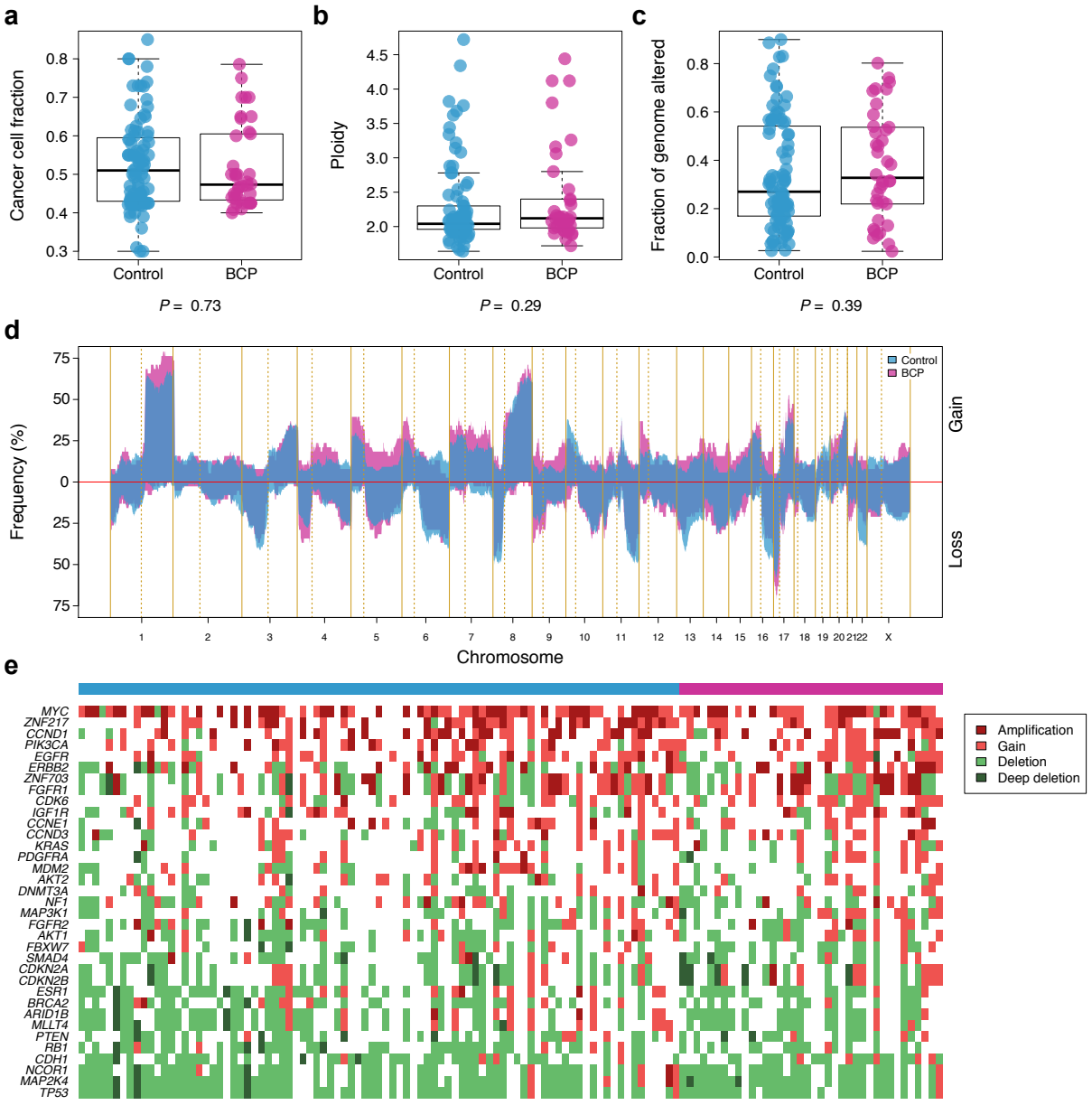


Figure 2.1 – Summary of the genome-wide copy number analysis of 87 controls and 38 BCP tumor samples

a-c, *Comparison of cancer cell fraction, ploidy and fraction of genome altered between controls and BCP.* ***d,*** *Comparison of the CNA frequencies of controls (blue) and BCP (pink).* ***e,*** *Heatmap of 35 CNA breast cancer driver genes according to their alterations; controls (blue) and BCP (pink). P, p-value derived from the Mann-Whitney U test.*

3.2 BCP shows a higher number of non-silent mutations

To identify potential genomic differences between BCP patients and controls, we performed whole-genome sequencing (WGS) on paired DNA samples extracted from FFPE blocks (i.e. primary tumors and histologically normal axillary lymph nodes) in a subset of 53 breast cancer patients from our initial series, 35 of whom were BCP (Supplementary Figure 2.2 and Supplementary Table 2.2). We achieved 32X and 19X median haploid genome coverage for tumor and normal samples respectively, which is similar in range to previous studies (Nik-Zainal et al., 2016) (Supplementary Figure 2.5). We detected a median of 13,829 and 10,084 single nucleotide variants (SNVs) and a median of 21 and 26 small insertions and deletions (Indels) in BCP and controls, respectively, and found no difference between the two groups (Figure 2.2a and Supplementary Figure 2.6a-c). Moreover, there was no difference in structural variations (insertions, deletions, duplications) nor tumor heterogeneity as assessed by the MATH score (Mroz and Rocco, 2013) (Supplementary Figure 2.6d-f).

We identified a median of 14 non-silent mutations per tumor which is comparable to another large-scale breast cancer cohort study (Nik-Zainal et al., 2016) (Supplementary Data File 2: Table S3). Interestingly, BCP had a significantly higher number of non-silent mutations than controls (median: 20 vs. 12, $P = 0.027$, Figure 2.2b and Supplementary Figure 2.6g-h). This observation remained consistent after correcting for potential confounding factors including age at diagnosis, date of diagnosis, pathological stage and molecular subtypes by IHC ($P = 0.019$, Supplementary Figure 2.6g). Compared to controls, BCP had also a significantly higher number of mutations previously reported in breast cancer in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Forbes et al., 2017) ($P = 0.018$, Supplementary Figure 2.6i). At the gene level, we identified 17 genes harboring at least one non-silent mutation with a frequency of at least 5% across all patients. Of those, *TP53* and *PIK3CA* were the most frequently mutated genes without any significant difference between the two groups (Figure 2.2c).

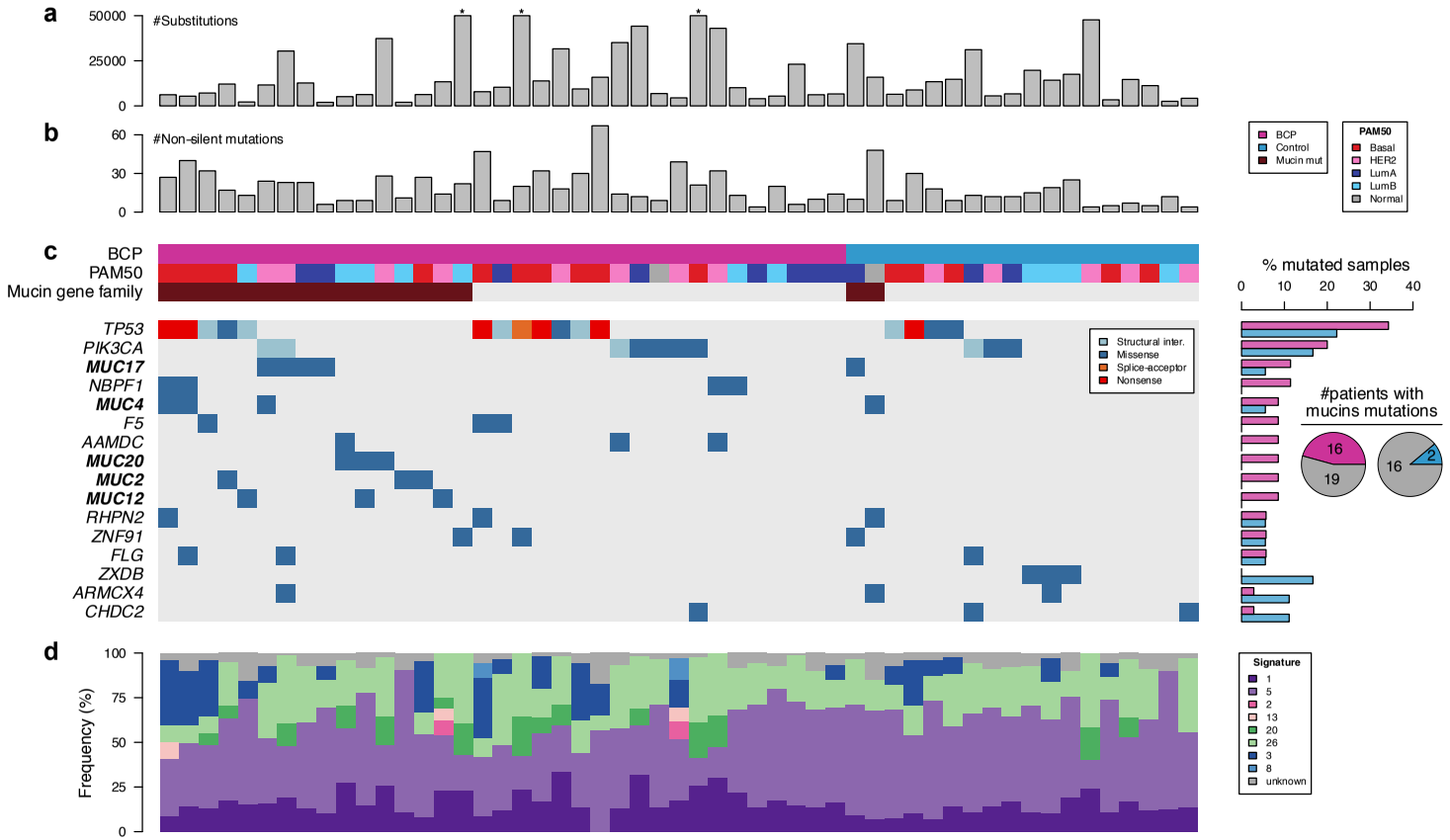


Figure 2.2 – Mutational landscape of individual BCP and controls

a, *Bar chart representing the absolute number of substitutions in BCP and controls, y-axis limited to 50,000 indicated by (*).* ***b,*** *Bar chart representing the absolute number of non-silent mutations in BCP and controls (median: 20 vs. 12, P = 0.027, respectively).* ***c,*** *Co-mutation plot showing genes harboring at least one non-silent mutation with a frequency of at least 5% across the whole cohort, and their corresponding frequencies in BCP and controls (right).* ***d,*** *Proportion of breast cancer substitution signatures in each sample. Signatures are colored according to broad biological groups: 1 and 5 are associated with clock-like processes, 2 and 13 are APOBEC-related, 20 and 26 are associated with mismatch-repair deficiency, 3 and 8 are associated with homologous-recombination deficiency.*

3.3 BCP is associated with a higher frequency of mutations in mucin gene family

MUC17 was the third most mutated gene and four other mucin gene family members namely *MUC2*, *MUC4*, *MUC12,* and *MUC20*, were among the most frequently mutated genes in BCP (Figure 2.2c). Within the mucin gene family, we identified 20 missense mutations and one nonsense mutation in BCP compared to only two missense mutations in controls. Among these 20 mucin variants, 10 were present in the COSMIC database (Forbes et al., 2017), which was higher than expected by chance ($P = 0.006$, Monte-Carlo test, Supplementary Data File 2: Table S3). Altogether, we found a significantly higher number of BCP with non-silent mutations in the mucin gene family compared to controls (45.7% vs. 11.1% respectively, $P = 0.015$, Figure 2.2c). This observation remained consistent after correcting for classical clinicopathological features ($P = 0.008$). Similar findings were observed by comparing BCP with 56 matched controls taken from the TCGA dataset (45.7% vs. 23.1% respectively, $P = 0.034$). Acknowledging that some mucins (*MUC4, MUC16*) are known to give rise to false positive calls due to technical artifacts (Lawrence et al., 2013), we removed these two genes and confirmed the above-mentioned results (37.1% vs. 5.5%, $P = 0.020$ and 37.1% vs. 14.3%, $P = 0.020$, using controls and TCGA controls, respectively).

We did not find any differences in clinicopathological features or survival according to mucin mutational status (Supplementary Table 2.3 and Supplementary Figure 2.7). There were three hotspots mutations (i.e. present in two distinct patients) two in *MUC17* and one in *MUC20,* and five missense mutations were clustered within 260 base pairs of *MUC2* (Figure 2.3a). None of these mutations were in annotated protein domains. Since the glycosylation of mucins is known to play a major role in producing a chemical barrier at the epithelium of tubular organs for protection and lubrication, we interrogated whether these mutations could affect glycosylation acceptor sites. The mucin O-glycosylation is characterized by the addition of N-acetylgalactosamine (GalNAc) to the hydroxyl group of serine or threonine residues (Hanisch, 2001). Remarkably, 40.9% of missense mutations affecting mucins resulted in an amino acid change to a serine residue, which was significantly higher than expected by chance ($P = 0.0002$, Monte-Carlo test), suggesting mucin hyperglycosylation in BCP. We also found that the frequency of missense mutations resulting in a gain of serine site in mucins in the TCGA dataset was significantly lower compared with BCP (6.3% in TCGA vs. 40.9% in BCP, $P < 0.001$).

Since mucins expression is known to increase throughout gestation in mice (Anderson et al., 2007), we expected that mucins were also upregulated in BCP. We therefore derived a metagene signature comprising all members of the mucin gene family (called "MUCsig") from the corresponding gene expression data and found higher expression of MUCsig in BCP than in controls (P = 0.017, Figure 2.3b-c). Altogether, these results show that BCP is associated with an increased expression of mucins as well as a higher frequency of mutations in mucin gene family that may potentially lead to mucin hyperglycosylation.

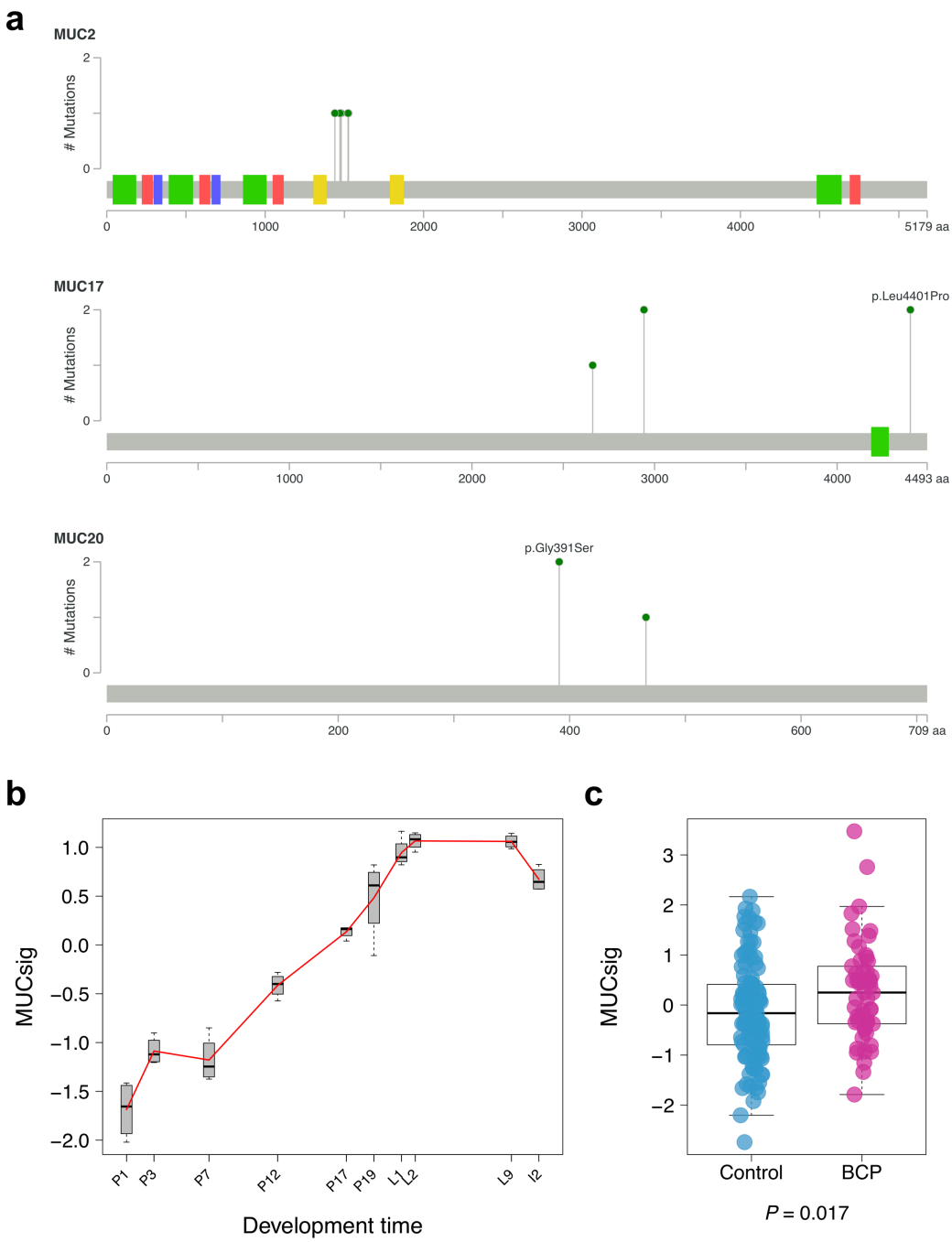


Figure 2.3 – Enrichment of mucin mutations and upregulation in BCP

a, *Lollipop plots were generated using cBioPortal Mutation Mapper. Each lollipop denotes a unique missense mutation for MUC2, MUC17, and MUC20 in BCP.* ***b,*** *MUCsig according to normal adult mouse mammary development (from pregnancy day 1 to involution day 2).* ***c,*** *Comparison of MUCsig between controls and BCP. P, p-value derived from the Mann-Whitney U test.*

3.4 BCP is enriched in mutational signature related to mismatch repair deficiency

To have a better understanding of the etiology of BCP, we interrogated the contribution of base-substitution signatures known to occur in breast cancer (Nik-Zainal et al., 2016). When evaluating the proportion of each signature present in each sample, we found that signature 1 was more prevalent in BCP compared to controls whereas signature 5 was more prevalent in controls (P = 0.013, FDR = 0.053 and P = 0.01, FDR = 0.053 respectively, Figure 2.2d). These results remained consistent after controlling for clinicopathological features (P = 0.002, FDR = 0.014 and P = 0.004, FDR = 0.016, respectively). When evaluating the presence or absence of mutational signatures we found that signature 20 (Sig20) was found in 13 out of 35 BCP (37.1%), as compared to only 2 out of 18 controls (11.1%) (P = 0.059, FDR = 0.410, Figure 2.2d). When controlling for clinicopathological features, this observation was significant (P = 0.004, FDR = 0.029). Signature 1 is known to be associated with age at diagnosis while the etiology of signature 5 is still unclear. Sig20, previously found in stomach and breast cancers, is related to DNA mismatch repair (MMR) deficiency (Alexandrov et al., 2013). Of interest, this signature remained significantly enriched in BCP when increasing the number of controls with 64 matched cases derived from the BRCA560 dataset (37.1% vs. 3.1%, $P < 0.001$). No classical clinicopathological features were associated with BCP Sig20-positive tumors except progesterone receptor negative status (Supplementary Table 2.4). We found that Sig20 frequency was strongly correlated with SNV mutational load (ρ = 0.56, $P < 0.001$, Supplementary Figure 2.8a) with Sig20-positive tumors harboring a median of 31,632 SNVs, as compared to 7,352 SNVs in Sig20-negative tumors ($P < 0.001$, Figure 2.4a). Next, we interrogated if Sig20 could be caused by alteration of genes involved in the MMR machinery either at the expression or copy number levels. The first step of MMR is the recognition of replication errors mediated by MutS homologue complexes; *MSH2* and *MSH6* (Jiricny, 2006). We found a significantly lower expression of *MSH2* in patients harboring Sig20 (P = 0.047, Figure 2.4b) corroborated by a negative correlation between *MSH2* expression and Sig20 frequency (ρ = -0.27, P = 0.024, Supplementary Figure 2.8b). This could be partially caused by CNA in *MSH2* since 5 out of 15 Sig20-positive versus 1 out of 38 Sig20-negative tumors harbored *MSH2* deletions (33.3% vs. 2.6%, P = 0.01). Finally, we interrogated the impact of Sig20 on survival and found that BCP Sig20-positive patients had a shorter DFS than

BCP Sig20-negative patients (median DFS time of 2.9 years vs. 10.2 years respectively $P = 0.091$, log-rank test, Figure 2.4c). In Sig20-positive patients the median OS was 6.72 years while the median OS was not reached in BCP Sig20-negative patients ($P = 0.009$, log-rank test, Supplementary Figure 2.9). This was not significant in a multivariate model (DFS mHR 1.06; 95% CI 0.21-4.27, $P = 0.31$; OS mHR 0.8; 95% CI 0.12-5.07, $P = 0.81$, respectively). Overall, these results suggest that some BCP patients show a defective MMR due to copy number loss of *MSH2*.

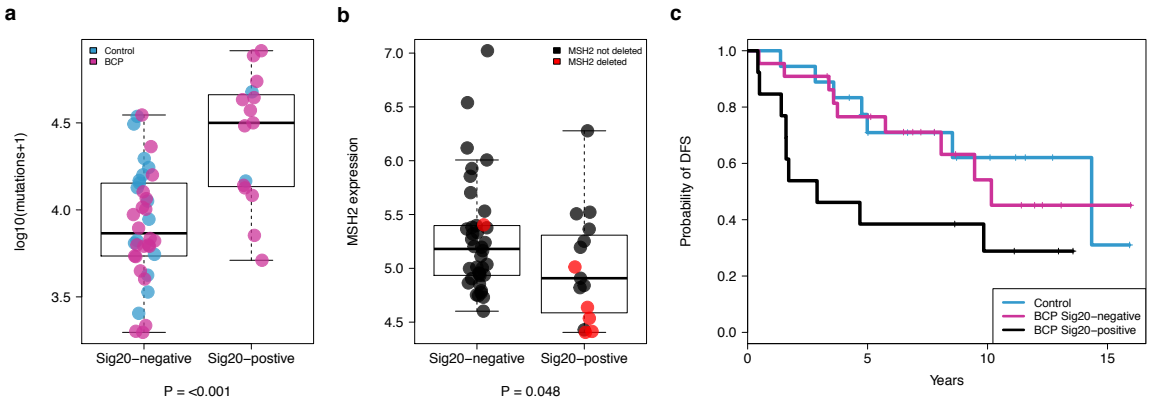


Figure 2.4 – Association of signature 20 with mutational load and clinical outcome

***a** Comparison of SNV mutational load between Sig20 negative and Sig20 positive tumors.* ***b,** Comparison of MSH2 expression between Sig20 negative and Sig20 positive tumors.* ***c,** Kaplan-Meier plot showing the difference in DFS between control patients (N = 18), BCP patients with Sig20 negative tumors (N = 22) and BCP patients with Sig20 positive tumors (N = 13). P, p-value derived from the Mann-Whitney U test.*

4 Discussion

This study reveals important molecular differences characterizing BCP that may potentially represent a biologic explanation for their rather aggressive clinical behavior. First, BCP was enriched in non-silent mutations that could have potential oncogenic functions. Second, 45% of BCP harbored a mutation in mucin gene family in addition to upregulation of mucins at the mRNA level. Like in mice (Anderson et al., 2007), this could be due to physiological change induced by pregnancy to prepare the breast for lactation. Our hypothesis is that some preexisting subclones carrying mucin mutations could have a growth advantage under pregnancy state. Another argument in favor of this hypothesis is the fact that most mucin mutations resulted in an amino acid change to a serine residue and that some of them are in hotspot regions. It has been previously found that in breast cancer, alterations in mucin expression or glycosylation influence tumor growth, adhesion, invasion, and immune surveillance (Hollingsworth and Swanson, 2004; Mukhopadhyay et al., 2011). The impact of missense mutations resulting in an amino acid change to a serine residue on the glycosylation status of mucins is unknown, but it is tempting to speculate that these alterations could influence their function, stability and secretion. More investigations are required to determine the exact effect of mucin mutations in BCP and in breast cancer in general, but these alterations could play a role in BCP biology.

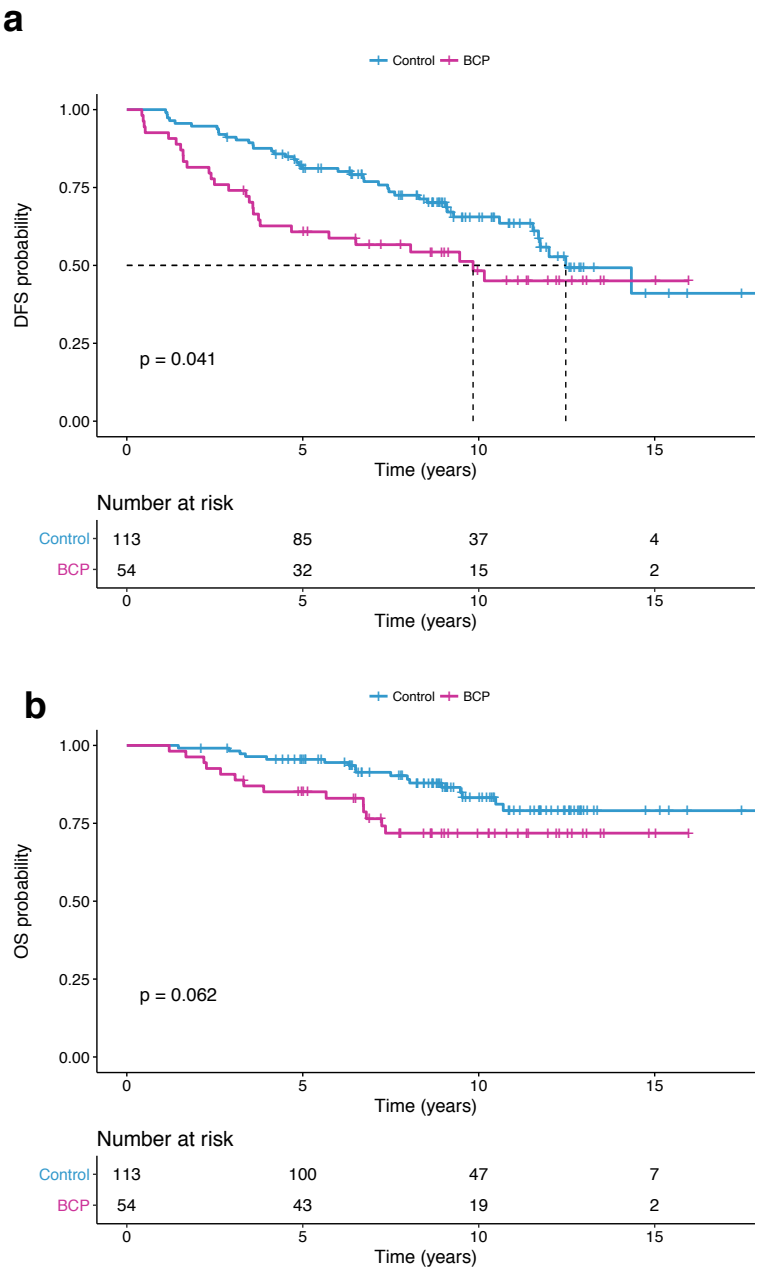
Moreover, BCP showed a higher prevalence of signatures 1 and 20 and a lower prevalence of signature 5. The etiology of signature 5 is not well understood (Alexandrov et al., 2015). The high prevalence of signature 1 cannot be explained by a difference in age at diagnosis or age of the blocks since similar results were found in a multivariate analysis after adjusting for both variables. 37.1% of BCP were associated with signature 20 (Sig20), attributable to DNA mismatch repair deficiency. This is surprising given the low frequency (1-2%) of MMR deficiency recently reported in breast cancer (Davies et al., 2017b). Mechanistically, this could be explained in part by the deletion of *MSH2,* a key gene involved in MMR. Survival analysis showed that BCP Sig20-positive patients had the worst prognosis whereas BCP Sig20-negative patients had DFS comparable to controls. MMR deficiency and high mutational burden have been shown to predict clinical benefit to immune checkpoint blockade in colorectal and other types of highly immunogenic cancers (Le et al., 2015; Rizvi et al., 2015). To date, the role of checkpoint

inhibitors in the treatment of breast cancer is under intensive investigation and the results are still awaited (Emens, 2017). While the feasibility of investigating new agents in such peculiar disease is rather complex, these results could potentially open the door to identify high-risk BCP patients who could benefit from immunotherapy.

A potential limitation of our study is that we used archived FFPE samples that are known to be challenging for WGS due to DNA degradation and induction of artefacts. Indeed, the higher proportion of signature 1 and 5 observed in our study could be due to C > T artefacts induced by formalin fixation. Nonetheless, BCP and controls were processed in the same way with no difference in the age of the blocks and the sequencing coverage reached in normal and in tumor tissues was comparable to other studies (Nik-Zainal et al., 2016). Another limitation of our study is the lack of epigenetic profiling analysis. As it is known that pregnancy induces epigenetic changes in epithelial cells to support mammary development (Huh et al., 2015), we can hypothesize that these modifications could impact breast cancer biology. Therefore, the study of such modifications in BCP is worth further investigation. In conclusion, we believe that our work provides important insights into the biology of BCP and a unique resource to study the biology of breast cancer in young women and how pregnancy could modulate tumor biology.

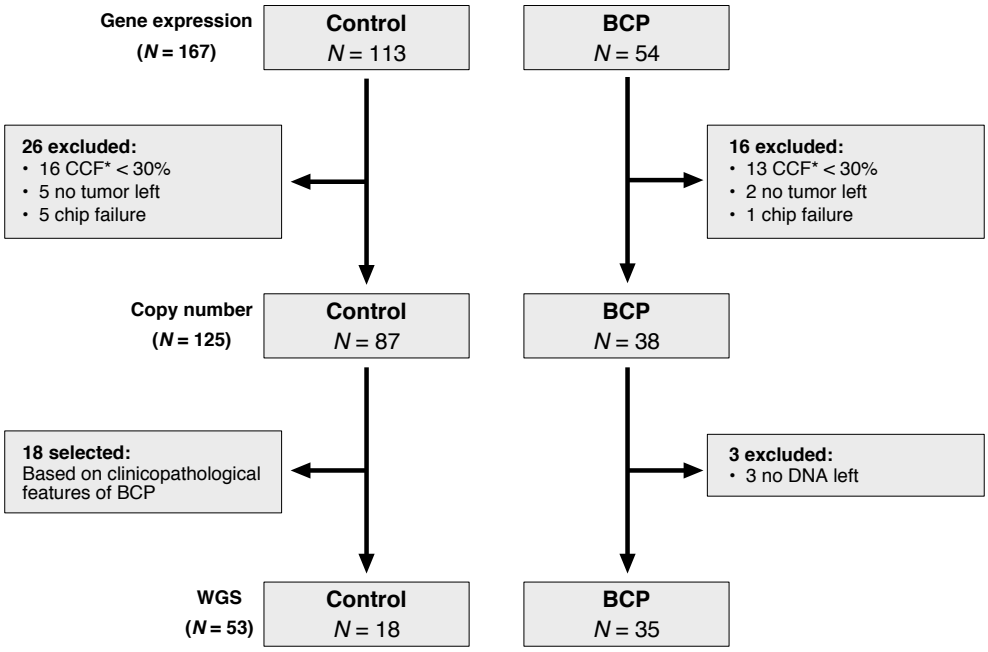
5 Supplementary Materials

5.1 Supplementary Figures



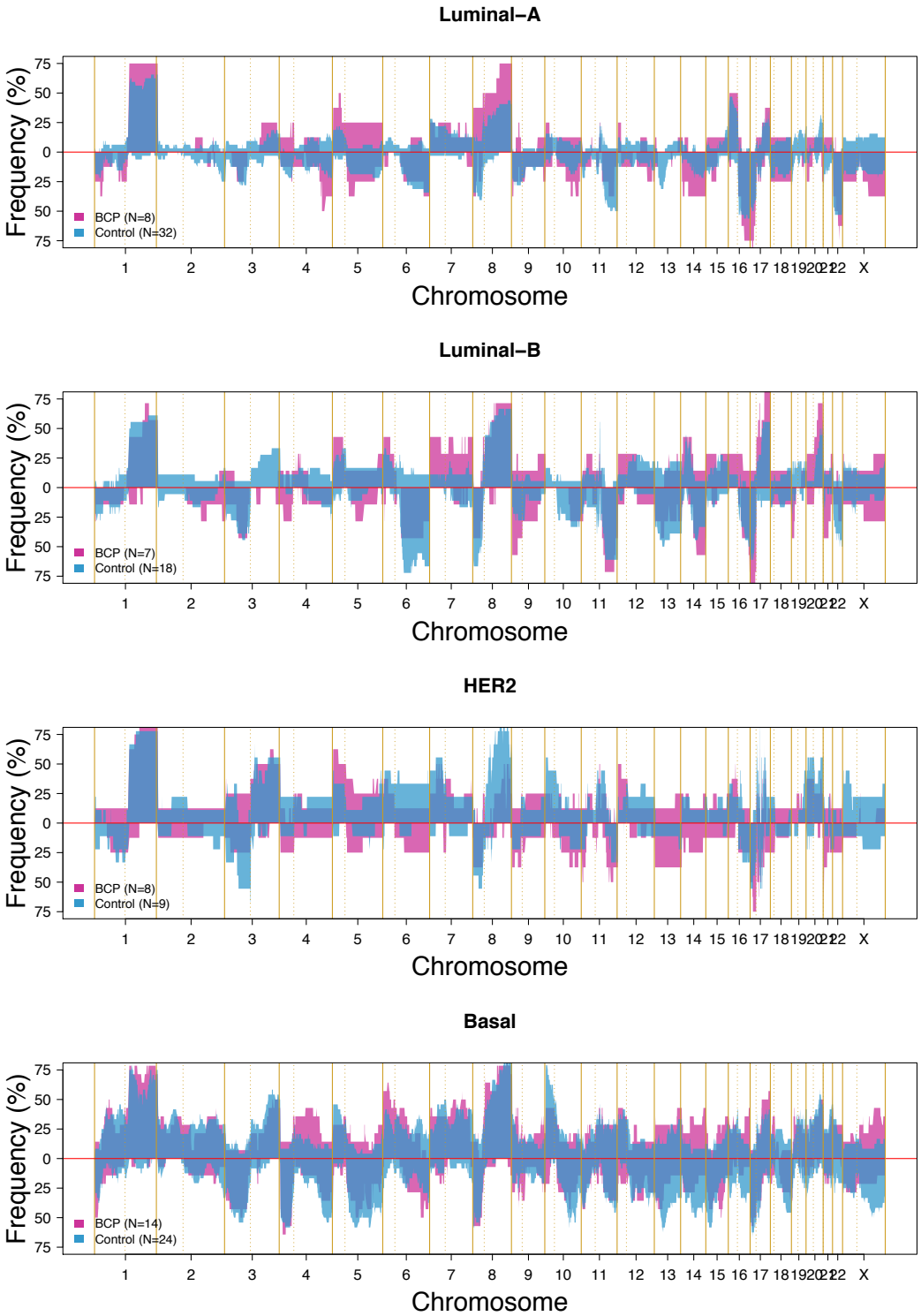
Supplementary Figure 2.1 – Updated survival analysis of BCP and controls

a, *Kaplan-Meier plot showing the DFS probability between control and BCP.* ***b,*** *Kaplan-Meier plot showing the OS probability between control and BCP.*

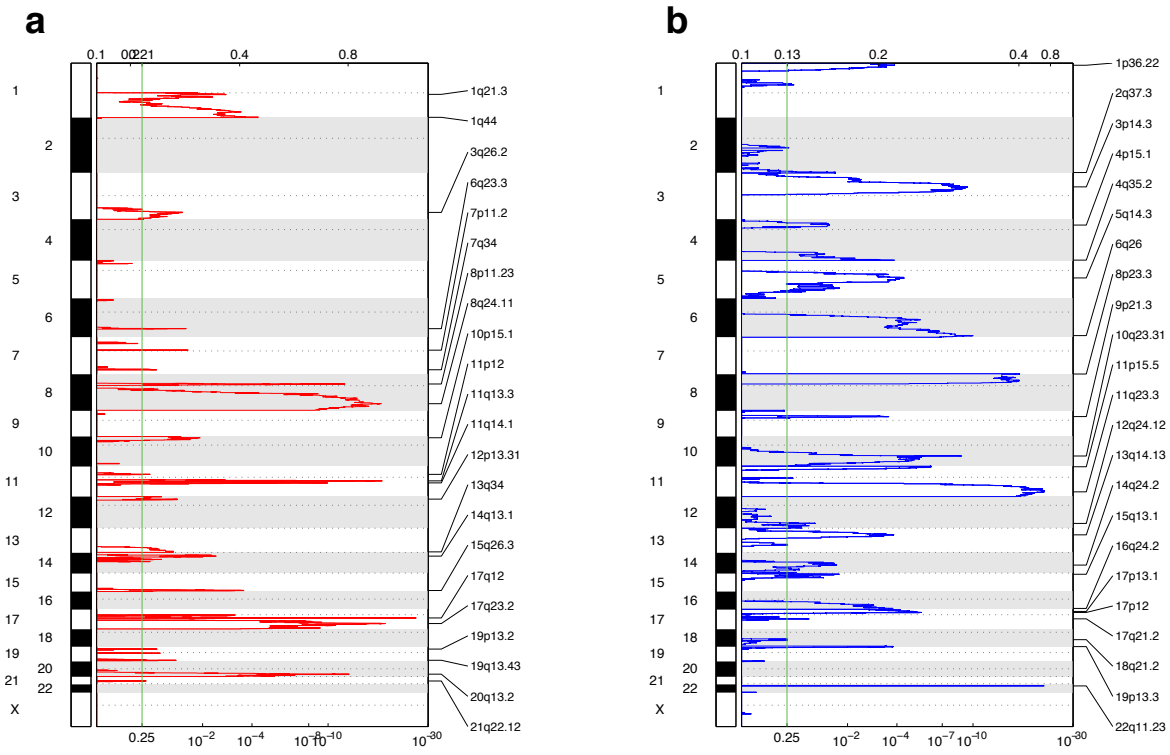


Supplementary Figure 2.2 – Flowchart summarizing the number of patients included in the analyses and the reasons for inclusion and exclusion

*There were 167 patients with available gene expression profiles in the original cohort. Seven patients were excluded because the tumor FFPE blocks were exhausted. We performed genome-wide copy number profiling on a total of 160 patients comprising 108 breast cancer controls and 52 breast cancers diagnosed during pregnancy (BCP). After quality control, a total of 125 patients were analyzed. Whole genome sequencing (WGS) was performed on 53 matched normal and tumor samples, 35 of which were BCP. *CCF, Cancer cell fraction estimated with the Genome Alteration Print algorithm.*

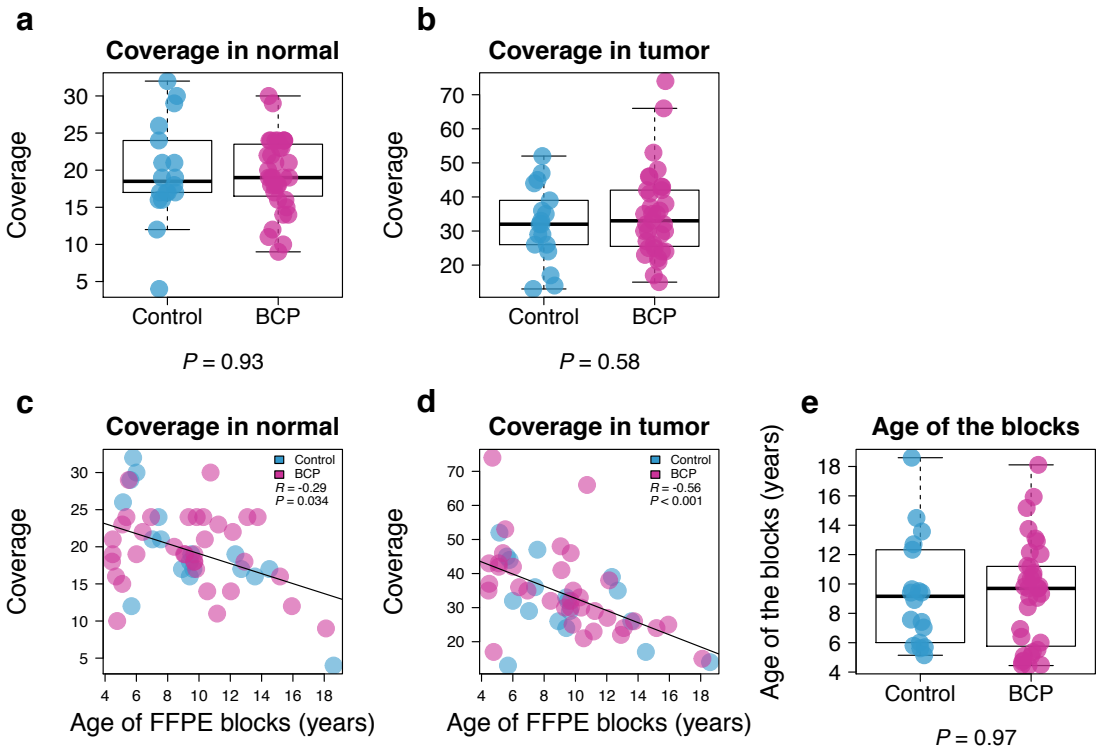


Supplementary Figure 2.3 – Comparison of the CNA frequencies of controls (blue) and BCP (pink) by intrinsic subtypes as defined by PAM50



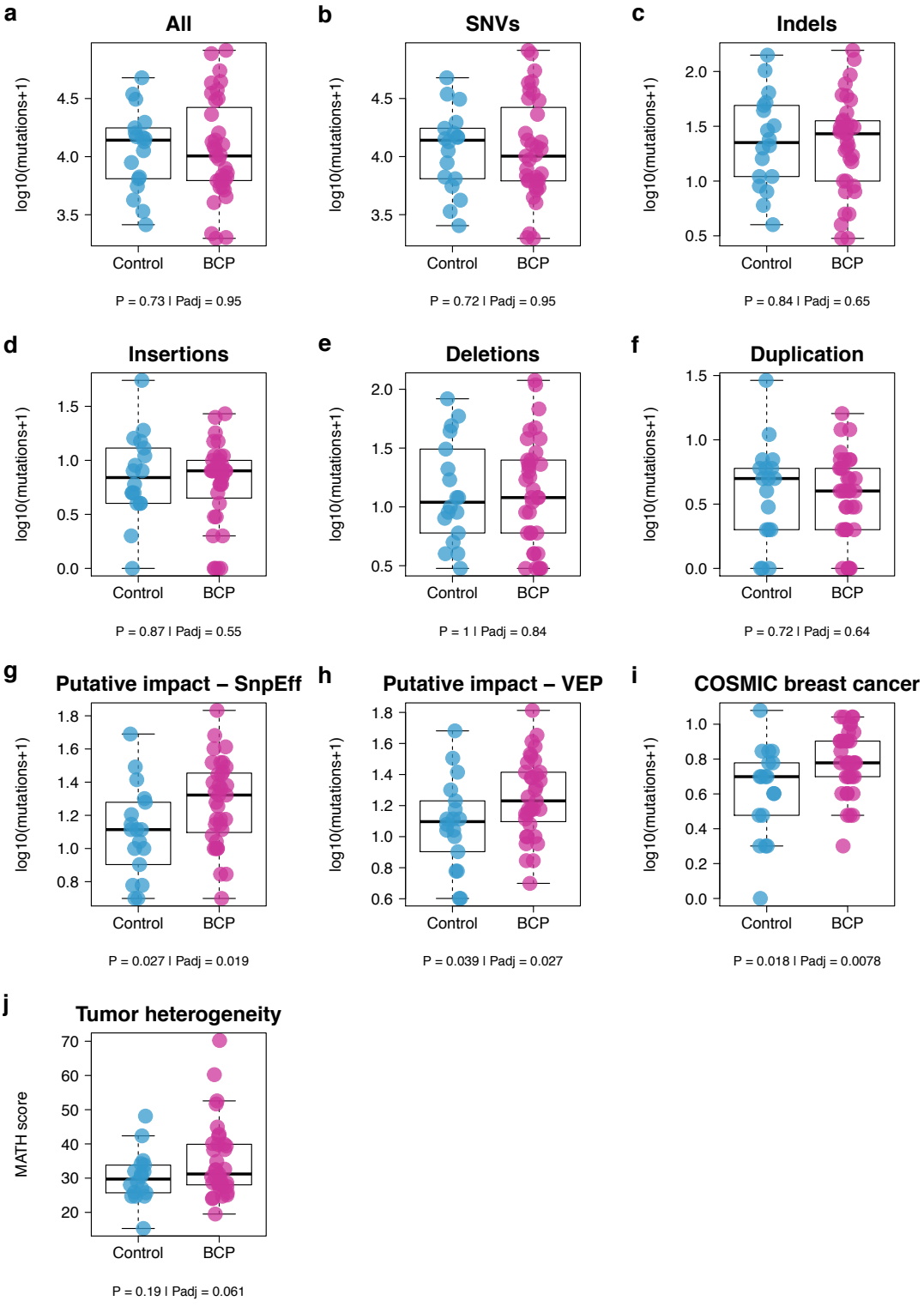
Supplementary Figure 2.4 – Genomic identification of significant targets in cancer (GISTIC) analysis

a, *GISTIC plot showing 22 amplification peaks.* ***b,*** *GISTIC plot showing 23 deletion peaks.*



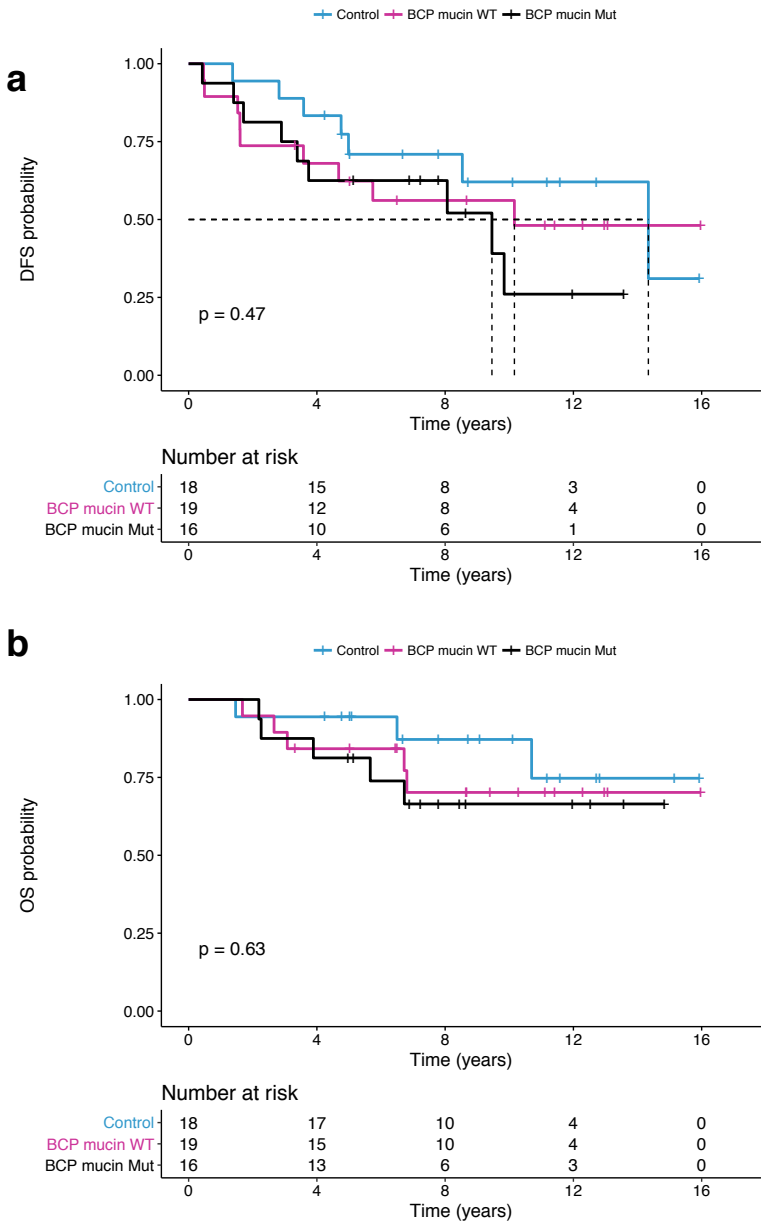
Supplementary Figure 2.5 – Coverage and age of the FFPE blocks

a-b, *Comparison of average coverage between BCP and controls in normal* ***(a)*** *and in tumor* ***(b)*** *P, p-value derived from the Mann-Whitney U test.* ***c-d,*** *Correlation between average coverage and the age of the blocks (years) in normal* ***(c)*** *and tumor. P, p-value derived from the non-parametric Spearman's rho (R) coefficient.* ***(d)***. ***e,*** *Comparison of the age of the blocks between controls and BCP.). P, p-value derived from the Mann-Whitney U test.*



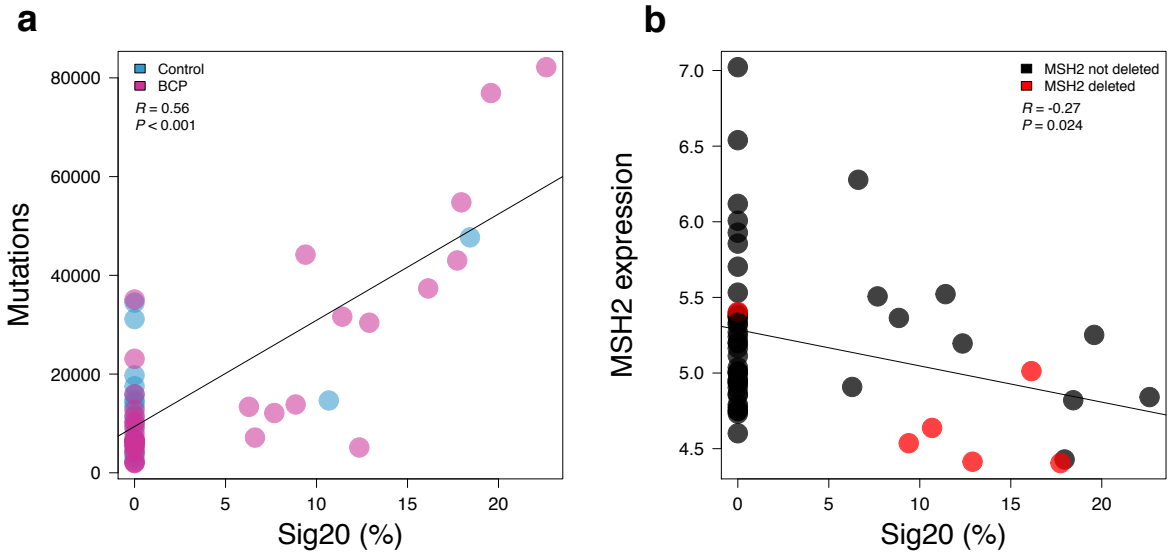
Supplementary Figure 2.6 – Mutational load and tumor heterogeneity

a-i, *Comparison of mutational load according to mutation types between controls and BCP.* ***j,*** *Comparison of tumor heterogeneity as assessed by the MATH score between controls and BCP. P, P-value comparing BCP and controls using the non-parametric Mann–Whitney U test;* P_{adj}*, P-value adjusted for age at diagnosis, date of diagnosis, pathological stage and molecular subtypes by IHC using a linear regression.*



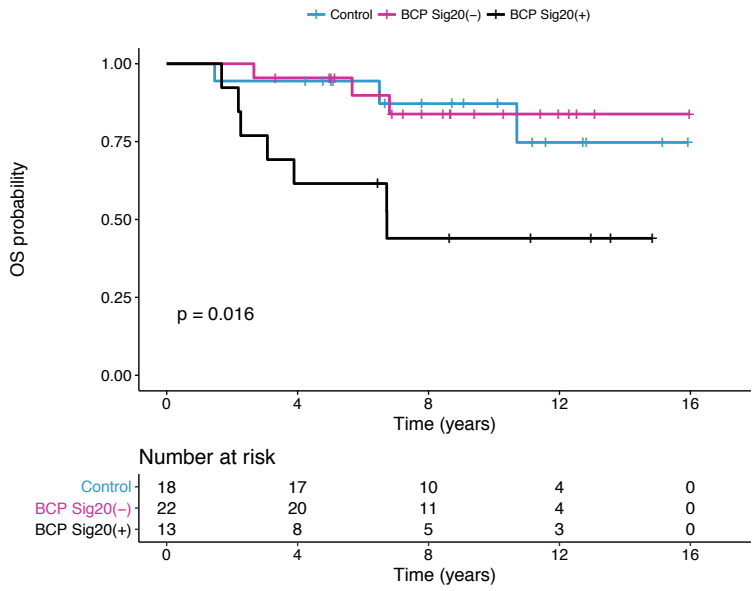
Supplementary Figure 2.7 – Survival analysis of controls and BCP according to mutational status of mucins

a, *Kaplan-Meier plot showing the DFS probability between control and BCP WT (pink) or mutated (black) for mucin.* ***b,*** *Kaplan-Meier plot showing the OS probability between control and BCP WT (pink) or mutated (black) for mucin. p; log-rank p-value comparing the three groups.*



Supplementary Figure 2.8 – Relationship between SNV mutational load and MSH2 expression with signature 20 (Sig20) frequency

a, *Correlation between SNV mutational load and signature 20 frequency.* ***b,*** *Correlation between MSH2 expression and signature 20 frequency.*



Supplementary Figure 2.9 – Survival analysis of BCP and controls

Kaplan-Meier plot showing the OS probability between control and BCP with (black) or without (pink) signature 20. p; log-rank p-value comparing the three groups.

5.2 Supplementary Tables

Supplementary Table 2.1 – Clinicopathological features of BCP and TCGA controls

		BCP	TCGA controls	P
N		35	56	
Age at diagnosis	**Median years (range)**	36.5 (30-39)	36 (29-44)	0.957[a]
Date of diagnosis	**Median years (range)**	2007 (1995-2011)	2005 (1996-2010)	0.132[a]
Tumor size	**≤ 2cm**	21 (60%)	45 (80.4%)	
	> 2cm	14 (40%)	11 (19.6%)	0.061
Nodal status	**Negative**	17 (48.6%)	21 (37.5%)	
	Positive	18 (51.4%)	35 (62.5%)	0.41
ER status	**Negative**	13 (37.1%)	16 (28.6%)	
	Positive	22 (62.9%)	40 (71.4%)	0.53
PR status	**Negative**	15 (42.9%)	16 (28.6%)	
	Positive	20 (57.1%)	40 (71.4%)	0.24
HER2 status	**Negative**	29 (82.9%)	48 (85.7%)	
	Positive	6 (17.1%)	8 (14.3%)	0.95

Abbreviations: BCP, breast cancer diagnosed during pregnancy; ER, Estrogen receptor; PR, Progesterone receptor; P, p-value derived from the $\chi 2$ test or the Fisher exact test when appropriate ([a]except continuous variable derived from Mann–Whitney U test).

Supplementary Table 2.2 – Clinicopathological features of control and BCP patients included in the different analyses

		Gene expression (N=167)			CNAs (N=125)			WGS (N=53)		
		Control	BCP	P	Control	BCP	P	Control	BCP	P
N		113	54		87	38		18	35	
Median DFS time (years)		12.47	9.84	0.041[a]	12	9.46	0.112[a]	14.34	9.46	0.289[a]
5-year OS rate (95% CI)		95.5% (91.7% - 99.4%)	85.1% (76.1% - 95.2%)		94.2% (89.4% - 99.3%)	81.4% (69.9% - 94.8%)		94.4% (84.4% - 100%)	82.8% (71.1% - 96.3%)	
Age at diagnosis	Median years (range)	36 (28-47)	36 (28-47)	0.677b	36 (28-47)	36 (29-44)	0.796b	35 (30-43)	36 (29-44)	0.202b
Date of diagnosis	Median years (range)	2005 (1996-2009)	2005 (1996-2010)	0.743b	2005 (1996-2009)	2005 (1996-2010)	0.776b	2005 (1996-2009)	2005 (1996-2010)	0.97b
Gestational age at diagnosis (trimester)	First		26 (48.1%)			18 (47.4%)			17 (48.6%)	
	Second		15 (27.8%)			11 (28.9%)			11 (31.4%)	
	Third		13 (24.1%)			9 (23.7%)			7 (20%)	
Stage	I	27 (24.3%)	14 (26.4%)		19 (22.1%)	7 (18.4%)		2 (11.1%)	7 (20%)	
	II	55 (49.5%)	27 (50.9%)		47 (54.7%)	21 (55.3%)		12 (66.7%)	20 (57.1%)	
	III	29 (26.1%)	12 (22.6%)	0.88	20 (23.3%)	10 (26.3%)	0.87	4 (22.2%)	8 (22.9%)	0.79
Tumor size	≤ 2cm	62 (56.9%)	28 (53.8%)		52 (61.2%)	24 (63.2%)		12 (66.7%)	21 (60%)	
	> 2cm	47 (43.1%)	24 (46.2%)	0.85	33 (38.8%)	14 (36.8%)	0.99	6 (33.3%)	14 (40%)	0.86
Nodal status	Negative	52 (46%)	26 (48.1%)		40 (46%)	18 (47.4%)		8 (44.4%)	17 (48.6%)	
	Positive	61 (54%)	28 (51.9%)	0.93	47 (54%)	20 (52.6%)	1	10 (55.6%)	18 (51.4%)	1
Grade	1	4 (3.7%)	3 (5.8%)		3 (3.5%)	3 (7.9%)		2 (11.1%)	3 (8.6%)	
	2	44 (40.4%)	17 (32.7%)		31 (36.5%)	9 (23.7%)		4 (22.2%)	9 (25.7%)	
	3	61 (56%)	32 (61.5%)	0.56	51 (60%)	26 (68.4%)	0.24	12 (66.7%)	23 (65.7%)	1
ER	Positive	88 (77.9%)	36 (66.7%)		67 (77%)	23 (60.5%)		13 (72.2%)	22 (62.9%)	
	Negative	25 (22.1%)	18 (33.3%)	0.17	20 (23%)	15 (39.5%)	0.095	5 (27.8%)	13 (37.1%)	0.71
PR	Positive	79 (69.9%)	35 (64.8%)		62 (71.3%)	22 (57.9%)		12 (66.7%)	20 (57.1%)	
	Negative	34 (30.1%)	19 (35.2%)	0.63	25 (28.7%)	16 (42.1%)	0.21	6 (33.3%)	15 (42.9%)	0.71
HER2	Negative	91 (80.5%)	45 (83.3%)		71 (81.6%)	31 (81.6%)		11 (61.1%)	29 (82.9%)	
	Positive	22 (19.5%)	9 (16.7%)	0.82	16 (18.4%)	7 (18.4%)	1	7 (38.9%)	6 (17.1%)	0.1
Molecular subtypes by IHC	Lum A-like	28 (24.8%)	15 (27.8%)		21 (24.1%)	9 (23.7%)		3 (16.7%)	9 (25.7%)	
	Lum B-like	42 (37.2%)	18 (33.3%)		33 (37.9%)	12 (31.6%)		5 (27.8%)	11 (31.4%)	
	Lum HER2	18 (15.9%)	5 (9.3%)		13 (14.9%)	4 (10.5%)		5 (27.8%)	3 (8.6%)	
	ERneg HER2-like	4 (3.5%)	4 (7.4%)		3 (3.4%)	3 (7.9%)		2 (11.1%)	3 (8.6%)	
	TNBC	21 (18.6%)	12 (22.2%)	0.58	17 (19.5%)	10 (26.3%)	0.66	3 (16.7%)	9 (25.7%)	0.45
PAM50	Luminal-A	40 (35.4%)	13 (24.1%)		32 (36.8%)	8 (21.1%)		3 (16.7%)	8 (22.9%)	
	Luminal-B	23 (20.4%)	10 (18.5%)		18 (20.7%)	7 (18.4%)		4 (22.2%)	7 (20%)	
	HER2	12 (10.6%)	10 (18.5%)		9 (10.3%)	8 (21.1%)		5 (27.8%)	8 (22.9%)	
	Basal	32 (28.3%)	18 (33.3%)		24 (27.6%)	14 (36.8%)		5 (27.8%)	11 (31.4%)	
	Normal-like	6 (5.3%)	3 (5.6%)	0.45	4 (4.6%)	1 (2.6%)	0.25	1 (5.6%)	1 (2.9%)	0.96

Abbreviations: CNAs, copy number alterations; WGS, whole genome sequencing; BCP, breast cancer diagnosed during pregnancy; ER, Estrogen receptor; PR, Progesterone receptor; IHC, Immunohistochemistry; P, p-value derived from χ2 test or the Fisher exact test when appropriate ([a]except DFS derived from the log-rank test and [b]continuous variable derived from Mann–Whitney U test).

Supplementary Table 2.3 – Clinicopathological features between control, BCP without mucins mutations, and BCP with mucins mutations

		Control Mucin WT	Control Mucin mut	P		BCP Mucin WT	BCP Mucin mut	P
N		16	2			19	16	
Age at diagnosis	Median years (range)	35 (30-40)	42.5 (42-43)	0.0275[a]		37 (30-44)	36 (29-41)	0.894[a]
Stage	I	1 (6.2%)	1 (50%)			4 (21.1%)	3 (18.8%)	
	II	11 (68.8%)	1 (50%)			11 (57.9%)	9 (56.2%)	
	III	4 (25%)	0 (0%)	0.25		4 (21.1%)	4 (25%)	1
Tumor size	≤ 2cm	11 (68.8%)	1 (50%)			11 (57.9%)	10 (62.5%)	
	> 2cm	5 (31.2%)	1 (50%)	1		8 (42.1%)	6 (37.5%)	1
Nodal status	Negative	6 (37.5%)	2 (100%)			10 (52.6%)	7 (43.8%)	
	Positive	10 (62.5%)	0 (0%)	0.18		9 (47.4%)	9 (56.2%)	0.85
Grade	1	1 (6.2%)	1 (50%)			2 (10.5%)	1 (6.2%)	
	2	3 (18.8%)	1 (50%)			5 (26.3%)	4 (25%)	
	3	12 (75%)	0 (0%)	0.098		12 (63.2%)	11 (68.8%)	1
ER	Positive	11 (68.8%)	2 (100%)			12 (63.2%)	10 (62.5%)	
	Negative	5 (31.2%)	0 (0%)	1		7 (36.8%)	6 (37.5%)	1
PR	Positive	10 (62.5%)	2 (100%)			12 (63.2%)	8 (50%)	
	Negative	6 (37.5%)	0 (0%)	0.53		7 (36.8%)	8 (50%)	0.66
HER2	Negative	9 (56.2%)	2 (100%)			17 (89.5%)	12 (75%)	
	Positive	7 (43.8%)	0 (0%)	0.5		2 (10.5%)	4 (25%)	0.38
Molecular subtypes	Lum A-like	2 (12.5%)	1 (50%)			5 (26.3%)	4 (25%)	
	Lum B-like	4 (25%)	1 (50%)			7 (36.8%)	4 (25%)	
	Lum HER2	5 (31.2%)	0 (0%)			1 (5.3%)	2 (12.5%)	
	ERneg HER2-like	2 (12.5%)	0 (0%)			1 (5.3%)	2 (12.5%)	
	TNBC	3 (18.8%)	0 (0%)	0.84		5 (26.3%)	4 (25%)	0.92
PAM50	Luminal-A	2 (12.5%)	1 (50%)			6 (31.6%)	2 (12.5%)	
	Luminal-B	4 (25%)	0 (0%)			2 (10.5%)	5 (31.2%)	
	HER2	5 (31.2%)	0 (0%)			4 (21.1%)	4 (25%)	
	Basal	5 (31.2%)	0 (0%)			6 (31.6%)	5 (31.2%)	
	Normal-like	0 (0%)	1 (50%)	0.039		1 (5.3%)	0 (0%)	0.42

Abbreviations: BCP, breast cancer diagnosed during pregnancy; ER, Estrogen receptor; PR, Progesterone receptor; IHC, Immunohistochemistry; P, p-value derived from the $\chi 2$ test or the Fisher exact test when appropriate ([a]except continuous variable derived from Mann–Whitney U test).

Supplementary Table 2.4 – Clinicopathological features between control, BCP Sig20-negative patients, and BCP Sig20-positive patients

		Control Sig20-	Control Sig20+	P		BCP Sig20-	BCP Sig20+	P
N		16	2			22	13	
Age at diagnosis	Median years (range)	35 (30-43)	34.5 (33-36)	0.887[a]		36 (30-44)	37 (29-41)	0.38[a]
Stage	I	1 (6.2%)	1 (50%)			5 (22.7%)	2 (15.4%)	
	II	11 (68.8%)	1 (50%)			13 (59.1%)	7 (53.8%)	
	III	4 (25%)	0 (0%)	0.25		4 (18.2%)	4 (30.8%)	0.71
Tumor size	≤ 2cm	11 (68.8%)	1 (50%)			13 (59.1%)	8 (61.5%)	
	> 2cm	5 (31.2%)	1 (50%)	1		9 (40.9%)	5 (38.5%)	1
Nodal status	Negative	6 (37.5%)	2 (100%)			11 (50%)	6 (46.2%)	
	Positive	10 (62.5%)	0 (0%)	0.18		11 (50%)	7 (53.8%)	1
Grade	1	2 (12.5%)	0 (0%)			2 (9.1%)	1 (7.7%)	
	2	4 (25%)	0 (0%)			7 (31.8%)	2 (15.4%)	
	3	10 (62.5%)	2 (100%)	1		13 (59.1%)	10 (76.9%)	0.73
ER	Positive	11 (68.8%)	2 (100%)			16 (72.7%)	6 (46.2%)	
	Negative	5 (31.2%)	0 (0%)	1		6 (27.3%)	7 (53.8%)	0.16
PR	Positive	10 (62.5%)	2 (100%)			16 (72.7%)	4 (30.8%)	
	Negative	6 (37.5%)	0 (0%)	0.53		6 (27.3%)	9 (69.2%)	0.038
HER2	Negative	11 (68.8%)	0 (0%)			20 (90.9%)	9 (69.2%)	
	Positive	5 (31.2%)	2 (100%)	0.14		2 (9.1%)	4 (30.8%)	0.17
Molecular subtypes	Lum A-like	3 (18.8%)	0 (0%)			8 (36.4%)	1 (7.7%)	
	Lum B-like	5 (31.2%)	0 (0%)			7 (31.8%)	4 (30.8%)	
	Lum HER2	3 (18.8%)	2 (100%)			2 (9.1%)	1 (7.7%)	
	ERneg HER2-like	2 (12.5%)	0 (0%)			0 (0%)	3 (23.1%)	
	TNBC	3 (18.8%)	0 (0%)	0.44		5 (22.7%)	4 (30.8%)	0.091
PAM50	Luminal-A	3 (18.8%)	0 (0%)			7 (31.8%)	1 (7.7%)	
	Luminal-B	4 (25%)	0 (0%)			5 (22.7%)	2 (15.4%)	
	HER2	3 (18.8%)	2 (100%)			3 (13.6%)	5 (38.5%)	
	Basal	5 (31.2%)	0 (0%)			6 (27.3%)	5 (38.5%)	
	Normal-like	1 (6.2%)	0 (0%)	0.3		1 (4.5%)	0 (0%)	0.24

Abbreviations: BCP, breast cancer diagnosed during pregnancy; ER, Estrogen receptor; PR, Progesterone receptor; IHC, Immunohistochemistry; P, p-value derived from the $\chi 2$ test or the Fisher exact test when appropriate ([a]except continuous variable derived from Mann–Whitney U test).

CHAPTER 3: THE IMMUNOMODULATORY POTENTIAL OF DENOSUMAB IN BREAST CANCER

This research work is related to the following publication:

Nguyen B., Maetens M., Salgado R., Venet D., Garaud S., Vuylsteke P., Polastro L., Wieldiers H., Simon P., Lindeman G., Larsimont D., Van den Eynden G., Velghe C., Rothé F., Michiels S., Willard-Gallo K., Hatem A. Azim Jr, Loi S., Piccart M. and Sotiriou C. (2018). The immunomodulatory potential of denosumab in breast cancer: results from a window-of-opportunity trial evaluating RANKL inhibition and its biological effects in young premenopausal women with early breast cancer.

Manuscript in preparation

My contribution to this study involves:

- Conceptualization and Design
- Methodology
- Formal Analysis
- Literature search and Interpretation
- Statistical analyses
- Visualization
- Writing of the manuscript
- Presentation of the results at the following meetings:
    - AACR Annual Meeting, April 2018, Chicago, IL, US
    - San Antonio Breast Cancer Symposium, December 2018, TX, US (selected for a poster-discussion spotlight presentation)

1 Introduction

Breast cancer in young women has unique biology and poor prognosis. Previous reports suggest that they often express RANKL (Azim et al., 2012c, 2015a), which has shown to play an important role in mammary tumorigenesis and antitumor immunity control. Denosumab is a human monoclonal antibody against RANKL, approved for the management of treatment-induced bone loss in early postmenopausal breast cancer and prevention of skeletal morbidity associated with metastatic bone disease. Preclinical data have reinforced the potential value of RANKL inhibition in breast cancer prevention (Gonzalez-Suarez et al., 2010b; Nolan et al., 2016; Schramek et al., 2010b; Sigl et al., 2016a) and its therapeutic potential on established tumors for its ability to reduce recurrence and metastasis (Yoldi et al., 2016).

RANK/RANKL axis is involved in various immune processes, including lymph node development (Dougall et al., 1999) and the establishment of central and peripheral tolerance (Akiyama et al., 2008; Barbaroux et al., 2008; Knoop et al., 2009; Rossi et al., 2007). RANK and RANKL are expressed in a wide variety of immune cells such as monocytes, macrophages, dendritic cells and activated T lymphocytes (de Groot et al., 2018) and are involved in T regulatory-induced self-tolerance (Anderson et al., 1997; Biswas and Lewis, 2010; Farrell et al., 2003; Seshasayee et al., 2004). Thus, the RANK/RANKL pathway regulates innate and adaptive immune responses and may promote or suppress immune responses depending on the context.

Tumor cells develop multiple strategies to evade immune surveillance; they create a microenvironment enriched in anti-inflammatory cytokines that will decrease the infiltration of cytotoxic T lymphocytes or natural killer cells, increase the recruitment of immunosuppressive cells such as T regulatory cells (Treg) or myeloid derived suppressor cells, and impose an anti-inflammatory profile in $CD4^+$ cells (Th2) and macrophages (M2) (DeNardo and Coussens, 2007; Mantovani et al., 2007). Immune checkpoint inhibitors (mainly antibodies against CTLA4 and PD-1/PD-L1) have emerged as a potent therapy against some solid tumors such as melanoma and advanced non-small cell lung cancer (Hodi et al., 2010; Sgambato et al., 2016). Nevertheless, in breast cancer and other poorly immunogenic tumors, the efficacy of immunotherapy remains limited (Solinas et al., 2017).

Given the critical importance of RANK/RANKL signaling in the mammary gland and antitumor immunity, we hypothesized that denosumab would have anti-tumor efficacy in breast cancer. To test this, we conducted a prospective, phase IIa, single-arm, multicenter study to evaluate its safety and biological effects in premenopausal patients with early breast cancer.

2 Methods

2.1 Patients

Eligible patients were premenopausal women with histologically confirmed, newly diagnosed, operable primary invasive carcinoma of the breast who had not undergone previous treatment for invasive breast cancer or being considered for neoadjuvant therapy. Other key eligibility criteria included tumor size > 1.5 cm, any nodal status, and known estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) status. Key exclusion criteria included bilateral invasive tumors, current or previous osteonecrosis or osteomyelitis of the jaw, and known hypersensitivity to denosumab. Evaluation of conventional breast cancer markers including ER, PR, HER2 and Ki-67 was performed centrally at Institut Jules Bordet (IJB). ER/PR-status were defined by ASCO-CAP guidelines. Breast cancer subtypes were defined according to St Gallen 2015 Consensus Meetings (Coates et al., 2015) using immunohistochemical surrogates as follows: Luminal A, ER and/or PR(+), HER2(–), Ki-67 < 20%; Luminal B, ER and/or PR(+), HER2(–), Ki-67 ≥ 20; Basal, ER, PR and HER2(–), irrespective of Ki-67 score; and HER2, HER2(+), irrespective of ER, PR or Ki-67.

2.2 Study design

This was a prospective, single arm, multi-center, open-label, phase IIa trial (D-BEYOND, NCT01864798). This study was approved by the Ethics Committee of Institut Jules Bordet (N°: 2064) and all patients provided written informed consent form. D-BEYOND was a preoperative "window-of-opportunity" study in which all patients received two injections of denosumab 120mg subcutaneously administered one week apart (minimum of 7 days to a maximum of 12 days) prior to surgical intervention. Surgery had to be performed within 10-21 days after the first dose of denosumab. Post-study treatment was at the discretion of the investigator. Snap-frozen tumor tissues embedded in optimal cutting temperature and formalin-fixed paraffin embedded (FFPE) tumor tissues were collected at baseline (pre-treatment) and at surgery (post-treatment).

Serious adverse events (AEs) and non-serious AEs were collected from the day of signed informed consent until one month after the last administration of study drug except for the project-specific AEs for which the reporting was extended to 3 months after the last dose of denosumab. Safety data were evaluated using the National Cancer Institute

Common Terminology Criteria for Adverse Events (NCI-CTCAE v 4.0). Adverse effects were coded with the use of the Medical Dictionary for Regulatory Activities (version 20.1).

The primary study endpoint was geometric mean decrease in the percentage of Ki-67-positive cells assessed by IHC. Keys secondary endpoints included absolute Ki-67 responders (defined as <2.7% Ki-67 IHC staining in the post-treatment tumor tissue), decrease in serum C-terminal telopeptide (CTX) levels measured by ELISA, change in IHC expression of RANK/RANKL and change in tumor infiltrating lymphocytes (TIL) percentage assessed on hematoxylin and eosin (HE) slides. Because the primary endpoint was a change in the percentage of Ki-67 positive cell after treatment, paired samples of breast tumor tissue at baseline and at surgery were required.

2.3 ELISA

Serum concentrations of sRANKL (the soluble homotrimeric form of RANKL) was centrally assessed at IJB in triplicate, using an enzyme-linked immunosorbent assay (ELISA) according to the manufacturer's instructions (Biomedica, Austria). Serum CTX levels were routinely evaluated locally in each center by ELISA.

2.4 Immunohistochemistry staining

Tissue sections (4 μm) from formalin-fixed paraffin embedded (FFPE) tissues of primary breast tissue were used for RANK and RANKL evaluation. For each patient, an hematoxylin and eosin-stained slide along with representative slides of the primary tumor were shipped to NeoGenomics Laboratories (California, USA) for immunohistochemical staining of RANK (N-1H8) and RANKL (M366) as described previously (Pfitzner et al., 2014; Wood et al., 2013), blinded to clinical information. For each epitope, the staining score for tumor cells and adjacent normal epithelial cells was recorded separately. The percentage of immunostaining and the staining intensity (0, negative; 1+, weak; 2+, moderate; and 3+, strong) were recorded. An H-score was calculated using the following formula:

$$\begin{aligned} H-score \; = & (\%\, of\; cells\; of\; weak\; intensity \times 1) \\ & + (\%\, of\; cells\; of\; moderate\; intensity \times 2) \\ & + (\%\, of\; cells\; of\; strong\; intensity \times 3) \end{aligned}$$

The maximum H-score would be 300, corresponding to 100% of cells with strong intensity.

Serial FFPE tissue sections (4 μm) were immunohistochemically stained for CD3/CD20, CD4/CD8 and FOXP3/CD4 double staining on a Ventana Benchmark XT automated staining instrument (Ventana Medical Systems) as described previously (Buisseret et al., 2017). The antibodies used for dual IHC are detailed in Supplementary Data File 3: Table S8.

For multiplex IHC (mIHC), FFPE tissue sections (4 μm) were processed manually. Briefly, slides were heated at 37°C overnight, then deparaffinized and fixed in neutral-buffered 10% formalin. Slides were labeled for CD4 (helper T cells), CD8 (cytotoxic T cells), CD20 (B cells), FOXP3 (regulatory T cells), CD68 (macrophages), pan-cytokeratin (cancer cells), and DAPI (all nuclei) using a serial same-species fluorescence-labeling approach that employs tyramide signal amplification and microwave-based antigen-retrieval and antibody stripping according to the manufacturer's instructions (Opal Multiplex IHC, Perkin Elmer). Samples were visualized on a Zeiss LSM 710 confocal microscope equipped with PMT spectral 34-Channel QUASAR (Carl Zeiss). A single pathologist (R.S.) reviewed all IHC slides.

2.5 Pathological assessment

For both FFPE and frozen tumor tissue, evaluation of tumor cellularity was assessed by a single pathologist (R.S.) on hematoxylin and eosin stained tissue sections. For patients with multiple samples, the sample with the highest tumor content was chosen. The percentage of TILs was independently evaluated by two trained pathologists (R.S. and G.V.D.E.) who were blinded to the clinical and experimental data on HE slides using the International TILs Working Group 2014 methodology, as described before (Salgado et al., 2015).

2.6 RNA Extraction and sequencing

RNA was extracted from frozen tumor tissue using the AllPrep DNA/RNA Mini kit (Qiagen, Germany) according to the manufacturer's instructions. RNA quality was assessed using a Bioanalyzer 2100 (Agilent technologies, US). Indexed cDNA libraries were obtained using the TruSeq Stranded Total RNA Kit (Illumina, US) following

manufacturer recommendation. The multiplexed libraries were loaded on a NovaSeq 6000 (Illumina, US) using an S2 flow cell and sequences were produced using a 200 Cycle Kit (Illumina, US). Read pairs were trimmed using Trimmomatic (Bolger et al., 2014). Alignment was performed using STAR (Dobin et al., 2013). The number of reads mapping to each gene was then assessed with the R statistical software with the Rsamtools package. Fragments per kilobase of transcript per million mapped reads (FPKM) was defined as the number of fragments (1 or both members of a read pair) mapping a gene per kilobase of transcript per million mapped reads, using the most common gene isoform as the transcript.

2.7 Bioinformatics analysis

Because it has been shown that the gene expression profiles of tissues taken at biopsy and surgery are impacted by difference of tissue handling procedures (López-Knowles et al., 2016), we used a publically available dataset from the no-treatment arm of The PeriOperative Endocrine Therapy - Individualizing Care (POETIC) study to filter-out differentially expressed genes. In that study, there were a total of 57 pairs of samples from untreated patients taken at diagnosis and surgery (GEO ID: GSE73235 (López-Knowles et al., 2016)). We filter out 3270/21931 (14.9%) genes that were differentially expressed between diagnosis and surgery by using a strict cut off of raw p-value of less than 0.05 using the non-parametric Mann–Whitney U test. Differential expression analysis was performed with DESeq2 v.1.14.1 R/Bioconductor package (Love et al., 2014) on raw count data. Significantly differentially expressed genes were selected with a qval of < 0.05 and an absolute log2 fold change of > 0.5. We used GAGE v.2.24.0 R/Bioconductor package (Luo et al., 2009) to identify significantly enriched biological process from the Gene Ontology (GO) database. Cytolytic activity was determined as the geometric mean expression of *GZMA* and *PRF1* as expressed in transcripts per million (TPM), as previously described (Rooney et al., 2015). To further refine subsets of immune cells that were present in each sample, CIBERSORT software was used (Newman et al., 2015). RPKM expression data were uploaded on https://cibersort.stanford.edu and CIBERSORT was run using LM22 as a reference matrix and, as recommended for RNA-seq data, quantile normalization was disabled. All other parameters were set to default. Output files were downloaded as tab-delimited text files, and immune cell subsets that were not present in at least 10 samples were discarded.

2.8 Statistical analysis

For the sample size estimation, we used data from a previous preoperative study (Bedard et al., 2011; Miller et al., 2009) to estimate the expected variability in the change in Ki-67 value. In that study, the geometric mean decreases in Ki-67 values after 14 days of letrozole alone was given by 29% (95% CI 22%-38%) among 51 patients diagnosed with estrogen receptor-positive breast cancer. The observed standard deviation was 0.98. In order to estimate a mean decrease in Ki-67 assessed by IHC with a 95% confidence interval with a width of 0.66, a total of 34 evaluable patients were needed. All statistical tests comparing pre- and post-treatment pairs value were done using the sign test or the paired Student's t-test. All IHC value were logged transformed, and because variables contain zeros, the logarithmic transformation was applied as follows: $\log10(x + 1)$. For the comparison between non-responders and responders, non-parametric Mann–Whitney U test and the Fisher exact test was used for continuous and categorical variables, respectively. All correlations were measured using the non-parametric Spearman's *rho* coefficient. All reported P-values were two-tailed. All analyses were done in R software version 3.3.3 (available at www.r-project.org) and Bioconductor version 3.6. There was no correction for multiple testing for exploratory analyses, except for the gene expression analysis for which the false discovery rate (FDR) was used.

3 Results

3.1 Patient characteristics and safety

A total of 27 out of 34 patients initially planned were enrolled in the study between October 2013 and July 2016 (the study was discontinued due to poor recruitment). All patients received two doses of denosumab 120 mg subcutaneously one week apart followed by surgery, and all were included in the safety analysis. The median time interval between the first administration of denosumab and surgery was 13 days (range: 9-21 days). No serious adverse events were reported. All non-serious adverse events are summarized in Table S1 (Supplementary Data File 3), the most frequent was arthralgia (4/27, 14.8%). One patient was excluded because it has a ductal carcinoma *in situ* and two patients were excluded because of lack of available tumor tissue. Table 3.1 summarizes the clinicopathological features of the remaining 24 patients. In brief, the median age at diagnosis was 45 years (range 35-51 years), tumors of 19 patients were hormone receptor positive (79.2%), 4 were HER2 positive (16.7%), and one was triple negative (4.2%).

Table 3.1 – Clinicopathological features of the 24 evaluable patients

N		24
Interval surgery-Dmab	**Median days (range)**	13 (9-21)
Age	**Median years (range)**	44 (35-51)
Size	**≤ 2cm**	13 (54.2%)
	> 2cm	11 (45.8%)
Nodal status	**Negative**	20 (83.3%)
	Positive	4 (16.7%)
Histologic grade	**Low/Intermediate**	16 (66.7%)
	High	8 (33.3%)
Molecular subtypes	**LumA**	14 (58.3%)
	LumB	5 (20.8%)
	HER2	4 (16.7%)
	TNBC	1 (4.2%)

3.2 Denosumab activity and its effects on breast cancer

To confirm adequate target inhibition by denosumab, we measured the serum concentrations of sRANKL, the direct target of denosumab and CTX, a bone resorption marker considered as a surrogate marker for denosumab activity. After treatment, serum levels of sRANKL and CTX decreased in all patients evaluated ($P < 0.001$, Figure 3.1a,b).

The primary endpoint of this study was a decrease in tumor proliferation, as assessed by the percentage of Ki-67 positive cells, between baseline and two weeks of treatment. There was no significant reduction in the percentage of Ki-67 positive cells after a short course of denosumab (geometric mean [GM] change from baseline; 1.07, 95% CI 0.87–1.33; $P = 0.485$, Figure 3.1c) and no absolute Ki-67 responders were identified. We also interrogated the effect of denosumab on the expression level of RANKL and RANK by immunohistochemistry (IHC) and found no significant change ($P = 0.891$ and $P = 0.061$, respectively, Figure 3.1d,e).

Collectively, these data confirm that short course of denosumab was associated with an effective systemic RANKL inhibition. However, this intervention was not associated with a reduction of tumor proliferation as assessed by Ki-67 staining or change in RANK/RANKL protein expression.

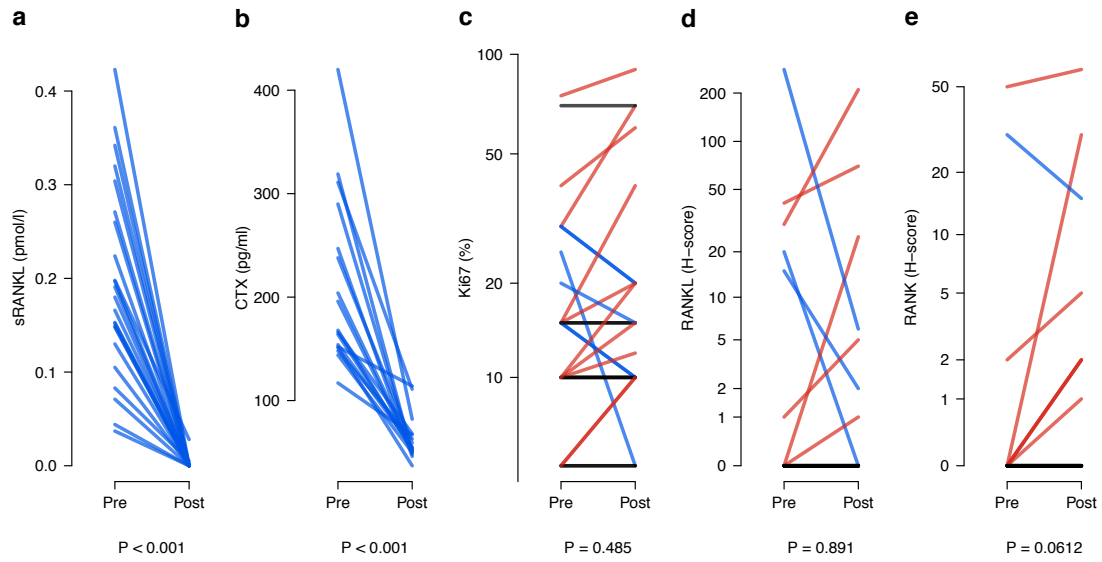


Figure 3.1 – Biological effect of denosumab in breast cancer

a-b, *Change from baseline in serum levels of sRANKL (N=23) and CTX (N = 17). P, p-value derived from the sign test.* ***c,*** *Change from baseline in the percentage of Ki-67 positive cell (N = 24). P, p-value derived from the paired Student's t-test.* ***d-e,*** *Change from baseline of RANKL (N = 23) and RANK H-score (N = 24). P, p-value derived from the paired Student's t-test. Lines are color-coded according to increase (red), decrease (blue) and no change (black) from baseline.*

3.3 Exploration of the immunomodulatory role of denosumab

Next, we interrogated the effect of denosumab on the tumor immune infiltrate and found a significant increased level of both stromal and intratumoral TILs (GM change from baseline; 1.75, 95% CI 1.28–2.39; $P = 0.001$ and 1.52, 95% CI 1.09–2.13, P = 0.016,

respectively, Figure 3.2a and Supplementary Figure 3.1). We identified 11/24 patients (45.8%) associated with an immunomodulatory response, defined as patients showing a ≥ 10 percentage points increase in stromal TILs. We next interrogated the expression of multiple immune markers by IHC from 23 available pairs of pre- and post-treatment tumor FFPE tissue (Figure 3.2b-e and Supplementary Figure 3.1). The remaining patient was excluded due to tissue exhaustion. Denosumab was associated with a significant increased percentage of CD3- and CD20-positive cells, which are T and B cell markers, respectively (GM change from baseline; 1.68, 95% CI 1.18–2.40; P = 0.006 and 1.62, 95% CI 1.09–2.40; P = 0.019, respectively, Figure 3.2b,c). Next, we evaluated the proportion of CD8 positive cytotoxic T cells and found a significant increase after treatment (GM change from baseline; 1.59, 95% CI 1.14–2.21; P = 0.008, Figure 3.2d). On the contrary, we found a significant decreased level of FOXP3/CD4 double positive regulatory T cells (GM change from baseline; 0.63, 95% CI 0.49–0.83; P = 0.002, Figure 3.2e). We also interrogated the percentage of CD68-positive cells, a macrophage marker, but did not find any significant difference from baseline (GM change from baseline; 1.50, 95% CI 0.91–2.46; P = 0.11, Supplementary Figure 3.1). A potential mechanism of action of denosumab responsible for the observed increased level of TILs could be the enhancement of TILs proliferation. However, we did not find any significant change in the percentage of Ki-67 positive TILs (Figure 3.2f). The immunomodulatory response associated with denosumab of a representative patient is shown in Supplementary Figure 3.2. To illustrate these findings, multiplex IHC (mIHC) was also performed on the top four tumors associated with the highest immunomodulatory response (Supplementary Figure 3.3).

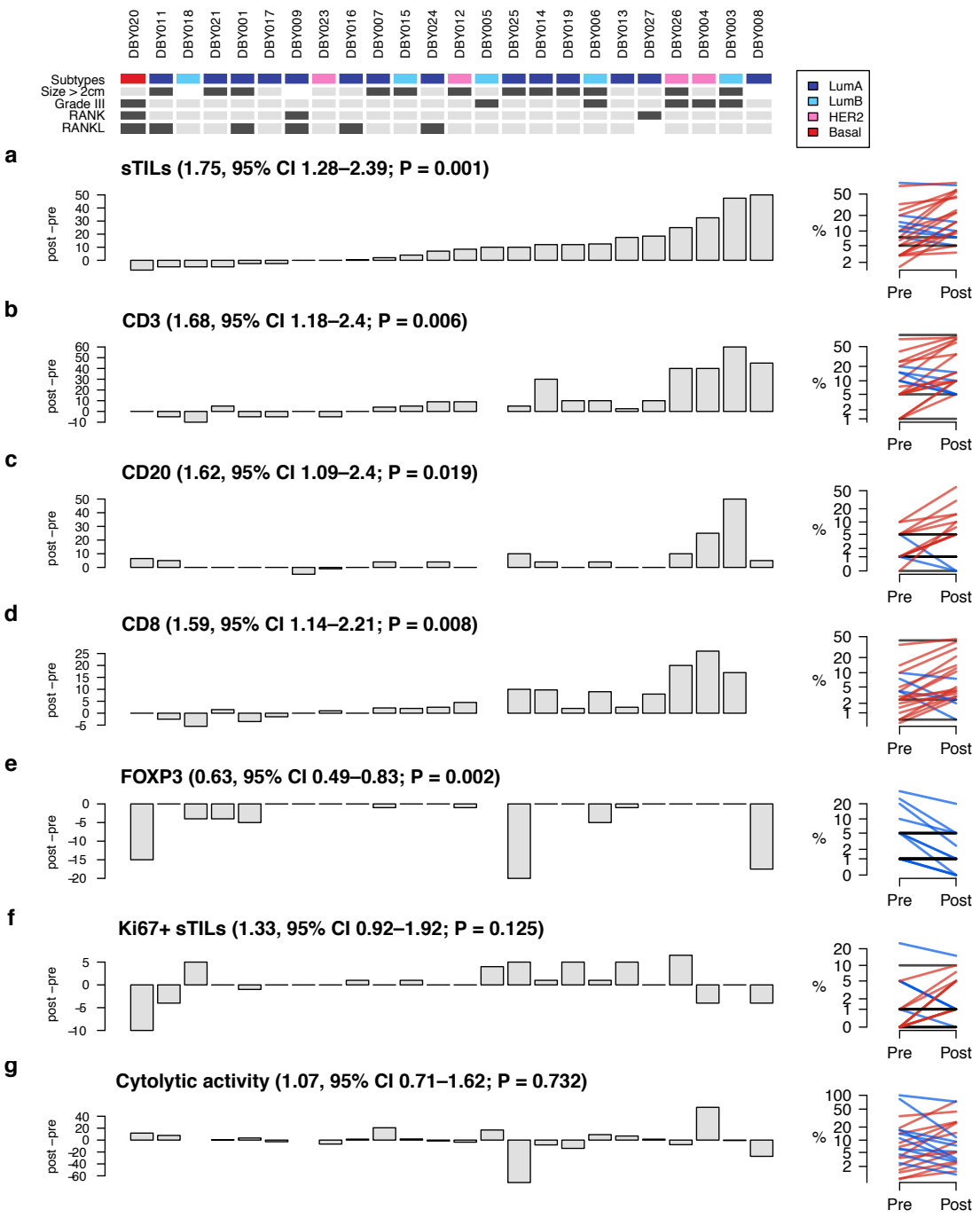


Figure 3.2 – The immunomodulatory role of denosumab in breast cancer

a-g, *Barplot showing the change (post- minus pre-treatment values) of the measured immune parameters for each patient. For each measured parameter, the corresponding ladder plot is displayed on the right-hand side. Lines are color-coded according to increase (red), decrease (blue) and no change (black) from baseline. Patients are ranked from low to high degree of increase in stromal TILs level and corresponding tumor characteristics at baseline are shown above. Patients positive for RANK and RANKL were defined as having a H-score > 0. Geometric mean changes values, 95% CI are shown above each barplot. P, p-value derived from the paired Student's t-test.*

To further investigate the biological effect of denosumab in breast cancer, we performed RNA sequencing on RNA extracted from matched fresh frozen tissues from 22 available pairs of pre- and post-treatment samples. The two remaining patients were excluded due to lack of tumor samples at baseline. We identified 379 differentially expressed genes between pre- and post-treatment (Supplementary Data File 3: Table S2). Gene ontology enrichment analysis revealed enrichment of genes related to immune cell migration and cytokine-mediated signaling pathways, including the well-known CD8 T cell attractant *CXCL10* (Dufour et al., 2002) (Figure 3.3a and Supplementary Data File 3: Table S3). Pathway analysis using the generally applicable gene-set enrichment (GAGE) analysis revealed an enrichment of genes related to numerous immune processes (Figure 3.3b and Supplementary Data File 3: Table S4).

To further explore the impact of denosumab on the immune cell landscape of breast cancer, we used CIBERSORT (Newman et al., 2015), a deconvolution method to infer immune cell content from gene expression data. Consistent with the IHC results, CIBERSORT analysis confirmed a significant increase in $CD8^+$ cytotoxic T cells, naïve B cells and memory CD4 T cells together with a decrease in regulatory T cells. A decrease in the proportion of Macrophages "M0" and neutrophils was also observed (Supplementary Figure 3.4).

In order to assess T cell activity, we measured the cytolytic activity score, which measures the mRNA levels of granzyme A (*GZMA*) and perforin (*PRF1*), two genes upregulated upon $CD8^+$ T cell activation (Rooney et al., 2015), but we did not find a significant change between pre- and post-treatment tumors (GM change from baseline; 1.07, 95% CI 0.71–1.62; $P = 0.732$, Figure 3.2g).

Taken together, these results indicate that short course of denosumab induces immunomodulation of the tumor microenvironment with an increased level of TILs, B, T, and cytotoxic T cells. This effect does not appear to be due to an enhancement of TILs proliferation but rather due to upregulation of inflammatory cytokines, regulatory T depletion and a subsequent influx of immune cells.

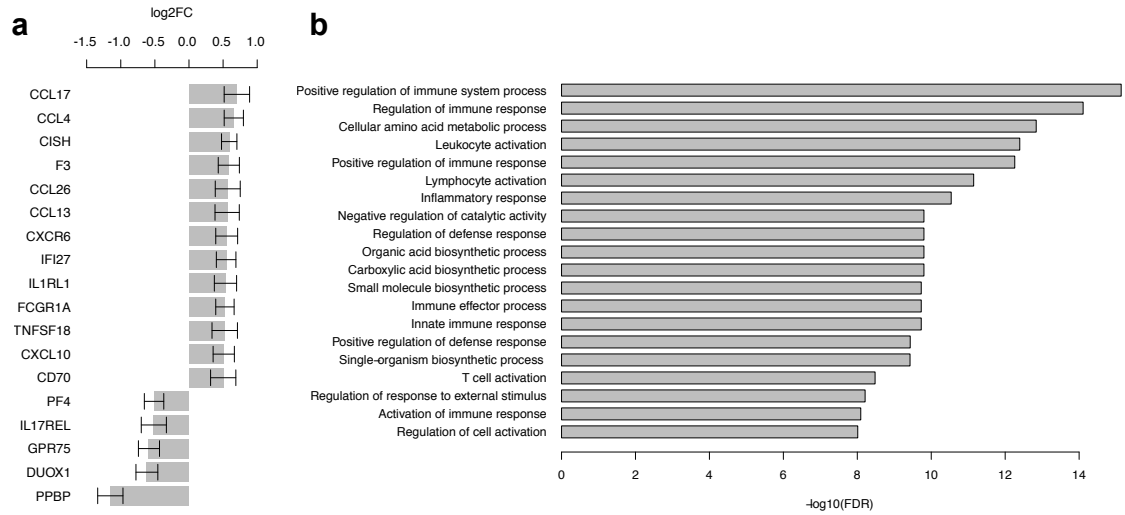


Figure 3.3 – RNA-seq analysis of pre- and post-treatment samples

a, *Log2FC of the differentially expressed genes annotated to "GO:0019221 cytokine-mediated signaling pathway".* ***b,*** *Top 20 significantly enriched pathways between pre- and post-treatment samples.*

3.4 Predictive value of baseline biomarkers associated with tumor immune microenvironment response

We sought to identify potential baseline predictive biomarkers associated with the immunomodulatory response induced by denosumab. No associations were found with any clinicopathological features, tumoral RANK/RANKL expression, nor with the percentage of TILs at baseline (Supplementary Data File 3: Table S5). Among the surrogate markers of denosumab activity at baseline, higher serum levels of sRANKL were significantly associated with an immunomodulatory response (P = 0.037, Figure 3.4a,b). Among the immune cell markers at baseline, a higher percentage of CD20-positive B cells (> 10%) and the presence of intratumoral FOXP3-positive regulatory T cells were associated with higher frequency of immunomodulatory-responsive patients (75% vs. 26.7%, P = 0.039 and 83.3% vs. 29.4%, P = 0.052, respectively, Supplementary Data File 3: Table S5). We used CIBERSORT data to confirm the above results. While the presence of B-naïve cells was not associated with higher frequency of immunomodulatory-responsive patients, the presence of regulatory T cells was significantly associated with higher frequency of immunomodulatory-responsive patients (75% vs. 20%, P = 0.032, Figure 3.4c and Supplementary Data File 3: Table S5). RNA sequencing analysis of samples at baseline revealed 42 genes differentially expressed genes between non-responsive and responsive patients (Supplementary Data File 3: Table S6). Gene ontology enrichment analysis revealed enrichment of genes related to

lymphocyte activation, including *FOXP3*, *CD28*, *IL7R*, *BANK1*, *RAC2* and *IFNG*, all up-regulated in pre-treated tumors from responsive patients (Supplementary Data File 3: Table S7 and, Figure 3.4d). Altogether, these data indicate that higher serum sRANKL concentration, the presence of regulatory T cells and higher *FOXP3* mRNA expression at baseline could predict the immunomodulatory response induced by denosumab in breast cancer.

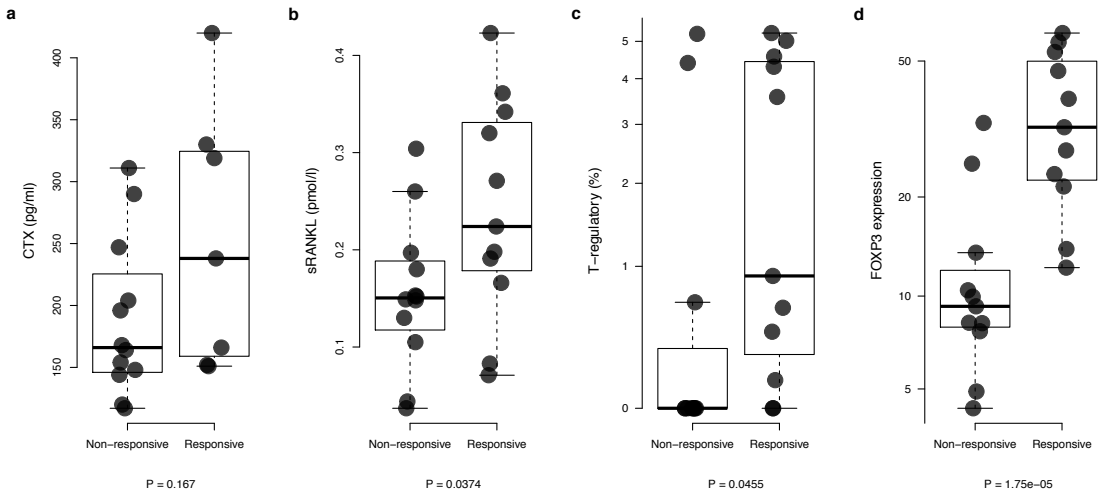


Figure 3.4 – Potential predictive biomarkers at baseline associated with an immunomodulatory response

a, *Comparison of baseline serum levels of CTX between non-responsive (N=12) vs. responsive (N=7). P, p-value derived from the Mann–Whitney U test.* ***b,*** *Comparison of baseline serum levels of sRANKL between non-responsive (N=13) vs. responsive (N=11) P, p-value derived from the Mann–Whitney U test.* ***c,*** *Comparison of baseline percentage of regulatory T cell as inferred by CIBERSORT between non-responsive (N=11) vs. responsive (N=11). P, p-value derived from the Mann–Whitney U test.* ***d,*** *Comparison of baseline expression level of FOXP3 (normalized counts) between non-responsive (N=11) vs. responsive (N=11). P, p-value from DEseq2.*

4 Discussion

Based on preclinical evidence supporting the importance of RANK/RANKL signaling in mammary tumors, we have assessed the biological effects of short-term single agent denosumab treatment in premenopausal patients with early breast cancer. Denosumab was well-tolerated overall but did not meet the primary efficacy endpoint with no decrease in the percentage of Ki-67 positive cells in this premenopausal early breast cancer population.

As expected, denosumab treatment was associated with a marked decrease in serum levels of sRANKL and CTX, demonstrating adequate target inhibition. However, a short course of denosumab had no apparent effect on the protein level of RANK/RANKL in the tumor tissue.

RANK/RANKL have been implicated in various physiological immune processes such as lymph node organogenesis and immune-tolerance control (Cheng and Fong, 2014; Ferrari-Lacraz and Ferrari, 2011). To investigate the role of RANKL inhibition on the tumor immune microenvironment, we explored the effect of denosumab on the immune infiltrate. The prognostic and predictive value of TILs has been demonstrated in numerous studies, especially in HER2 and triple-negative breast cancer (Salgado and Loi, 2018; Savas et al., 2016). However, not all breast cancer has a high presence of TILs. The identification of a therapy that can convert immune cold tumors into hot ones without adding toxicity is considered as a “holy grail” in the field. This question is particularly relevant for luminal breast cancer in which TILs levels were reported to be low. Strikingly, we found that short course of denosumab was associated with immunomodulation of the breast tumor microenvironment with an increased level of TILs, B, T and cytotoxic T cells and depletion of regulatory T cells. We also measured Ki-67-positive TILs and found no change after treatment. RNA-seq data obtained from matched pairs of frozen tumor tissue confirmed these results and differential gene expression analysis showed that denosumab induced upregulation of inflammatory cytokines. Thus, we hypothesized that the observed increased levels of TILs could be due to an influx of immune cells from the circulation. However, as there was no evidence of increased cytolytic activity, an important question that remains is whether these immune cells have the ability to recognize tumor cells and initiate an appropriate antitumor response.

Exploratory analysis of potential baseline biomarkers associated with the tumor immunomodulatory response revealed that the presence of regulatory T cells and higher FOXP3 expression were associated with a response, which further supports that the influx of TILs might also be mediated in part by depletion of regulatory T cells. High serum sRANKL concentration was also associated with response and could represent a practical approach to identify patients likely to benefit from the immunomodulation associated with denosumab in future clinical studies. More functional studies are still required to conclude that the immunomodulatory response associated with RANKL-inhibition is mediated through a tumor cell-dependent, immune cell-dependent or through a more complex mechanism involving both tumor and stromal cells interplay.

In light of our results, it would be worth to reassess the clinical outcome of the two recent large phase III trials of adjuvant denosumab in postmenopausal early breast cancer, namely the D-CARE study which reported negative results and the ABCSG-18 trial which showed DFS improvement as secondary endpoint (Coleman et al., 2018; Gnant et al., 2018).

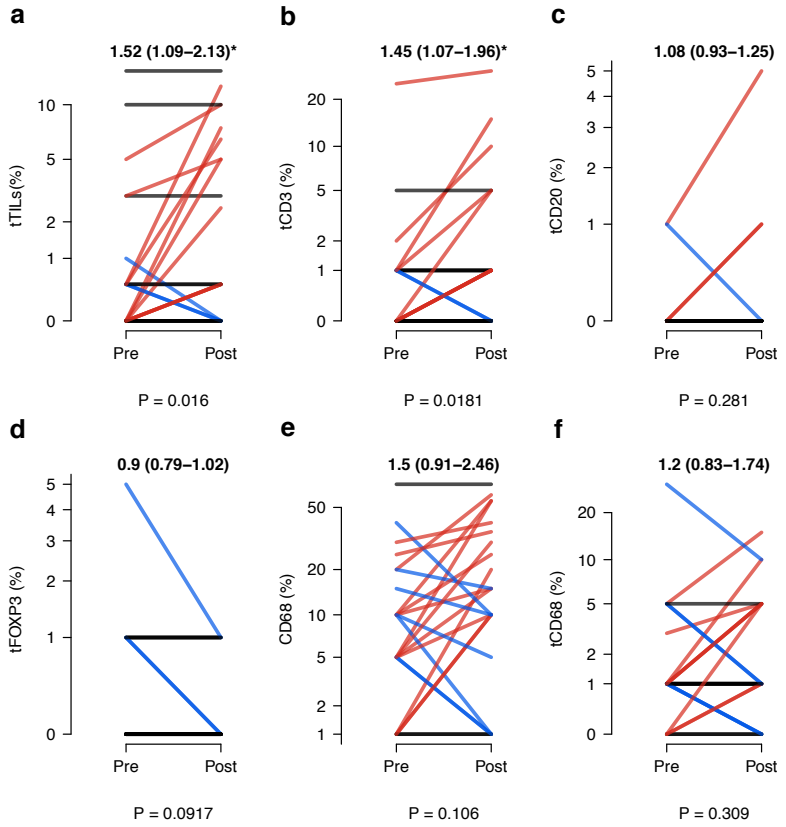
Clinical studies of the immunomodulatory potential of anti-RANKL are limited, but it is worth to mention a case report describing an exceptional response upon treatment to combination anti-CTLA4 ipilimumab and denosumab in a metastatic melanoma patient (Smyth et al., 2013). Following this observation, the same group reported that RANKL blockade improved the efficacy of both anti-CTLA4 and PD1/PDL1 blockade in several tumor mouse models (Ahern et al., 2017, 2018a). The combinatorial effect of anti-RANKL and immune checkpoint inhibitors will be investigated in the CHARLI trial (NCT03161756), a phase I/II study examining the effect of denosumab in combination with nivolumab (an anti-PD1) with or without ipilimumab in metastatic melanoma patients.

Our study has some limitations, such as the small sample size and the lack of a control group. The number of TNBC and HER2-positive cases in our study was too small to perform subgroup analysis. Moreover, our study included only premenopausal patients, a question that remains is whether the immunomodulatory response associated with RANKL inhibition could also be effective in postmenopausal patients. This will be tested in the randomized PERIDENO trial (NCT03532087) in which the effect of denosumab

on the systemic and local immune environment will be determined in postmenopausal patients with early breast cancer.

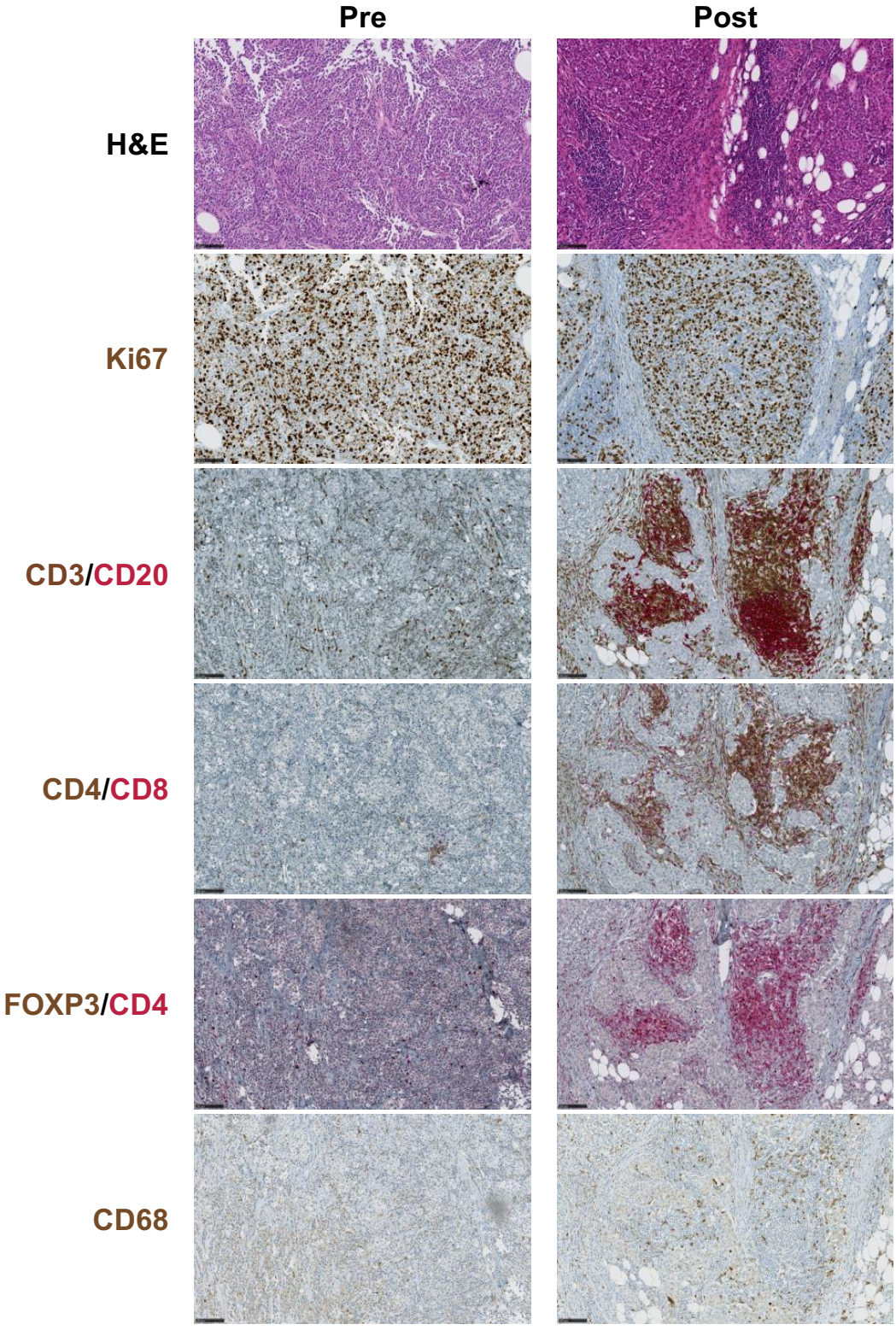
In conclusion, this is the first study on the effect of RANKL-inhibition on breast cancer biology in the preoperative setting. We have been able to demonstrate that a median of 13 days of denosumab treatment is associated with an immunomodulatory effect on breast tumor microenvironment. This work encourages further research on the immunomodulatory potential of RANKL inhibition. To have an effective antitumor immunity response, tumor cells and tumor antigens have to be recognized by immune cells. Despite the fact that our study did not demonstrate evidence of cytolytic activity, we believe that this could be due to the very short timing, inherent of preoperative window-of-opportunity trials or that anti-RANKL might simply induce an influx of non-effective immune cells. In the latter case, an interesting starting point could be to explore the synergistic effect of denosumab with drug inducing neoantigens or immune checkpoints inhibitors.

5 Supplementary Materials



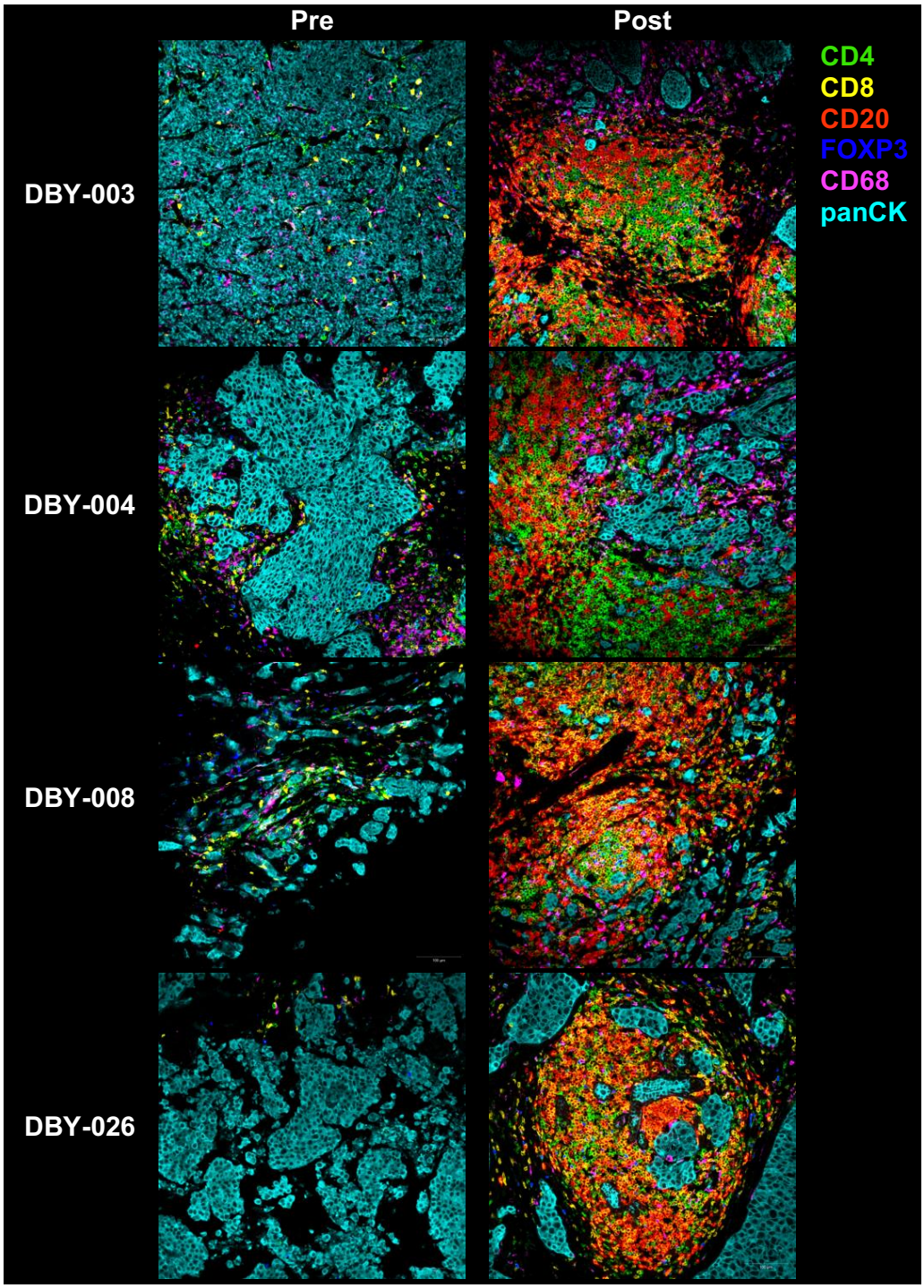
Supplementary Figure 3.1 – The immunomodulatory role of denosumab in breast cancer at the intratumoral level

a, *Change from baseline in the percentage of intratumoral TILs (N = 24).* ***b-f,*** *Change from baseline in the percentage of intratumoral CD3-positive, intratumoral CD20-positive, intratumoral FOXP3/CD4 double positive, stromal CD68-positive and intratumoral CD68-positive cells (N = 23). Geometric mean changes and 95% CI values are shown above each ladder plot and significant changes are indicated with "*". P, p-value derived from the paired Student's t-test. Lines are color-coded according to increase (red), decrease (blue) and no change (black) from baseline.*



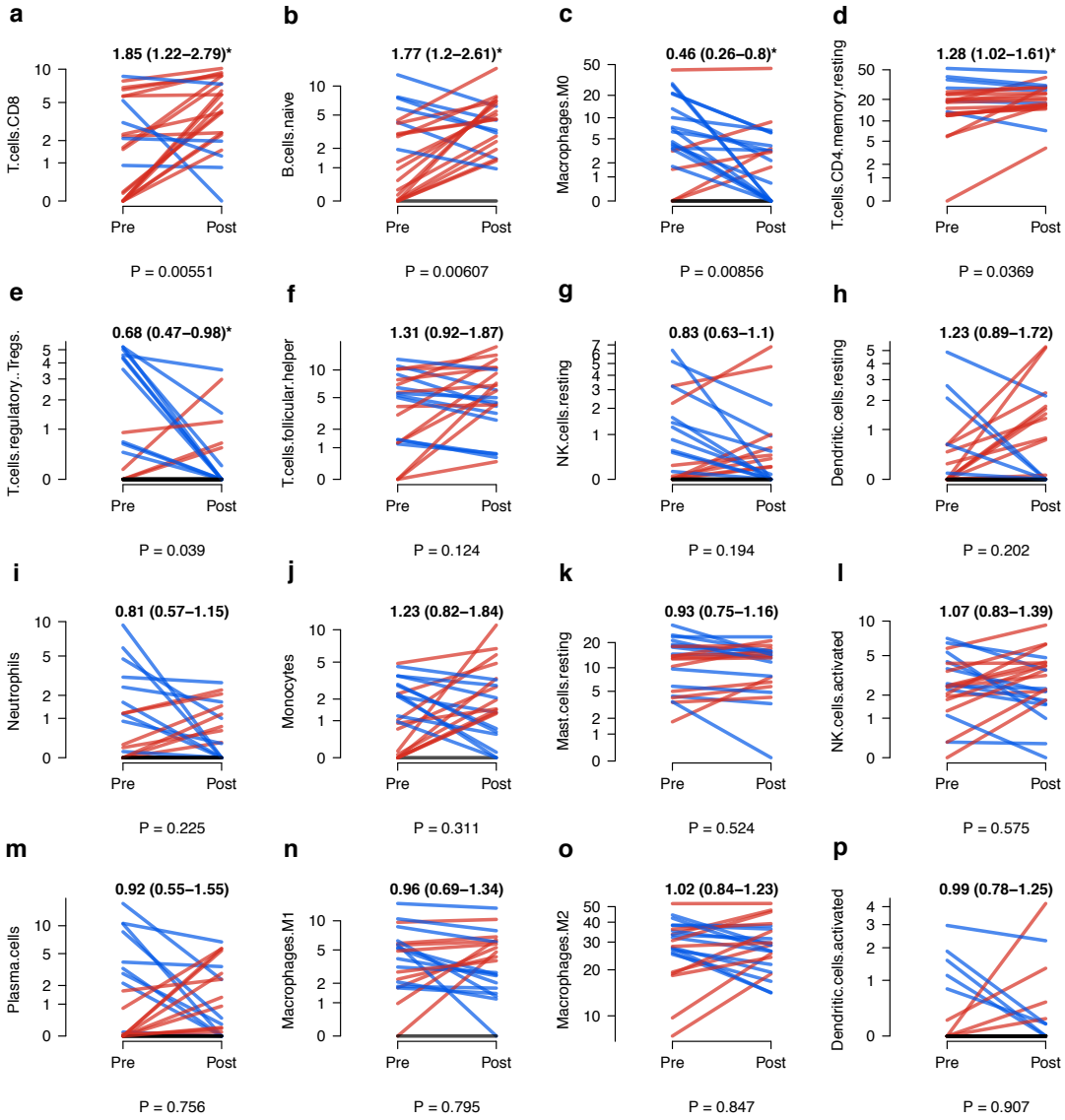
Supplementary Figure 3.2 – Representative micrographs of pre- and post-treatment tumor sections from one patient (DBY003)

Corresponding stainings are shown on the left-hand side. Scale bar indicates 100 µm.



Supplementary Figure 3.3 – Representative micrographs of multiplex IHC of pre- and post-treatment tumor sections from the top four patients associated with the highest immunomodulatory response

Corresponding stainings are shown on the upper right corner. Scale bar indicates 100 μm.



Supplementary Figure 3.4 – Comparison of 16 immune cell fractions, as inferred by CIBERSORT, between pre- and post-treatment tumor samples

Geometric mean changes and 95% CI values are shown above each ladder plot and significant changes are indicated with "" (N = 22). P, p-value derived from the paired Student's t-test. Lines are color-coded according to increase (red), decrease (blue) and no change (black) from baseline.*

CONCLUDING REMARKS AND PERSPECTIVES

Since long, epidemiological studies have established a link between women's reproductive history and the risk of developing breast cancer. However, to date, there were no studies addressing the association between reproductive history and the genomic alterations of breast tumors. The advent of next-generation sequencing and other "omics" tools, coupled with the public availability of the data generated, have allowed us to carry out the first study of this thesis. The main aim of this research was to explore the association of parity status and age at first pregnancy with clinical, biological and molecular characteristics of subsequent breast cancer. In addition to the publicly available data already present in the BRCA560 dataset (Nik-Zainal et al., 2016) (e.i., clinicopathological features, WGS, RNAseq data), we also evaluated the level of tumor-infiltrating lymphocytes (TILs) using international guidelines.

At the clinical level, we found a higher proportion of triple-negative breast cancer (TNBC) in parous and particularly in early parous patients. This could be attributed to a differential effect of pregnancy-induced tumor protection according to breast cancer subtypes that we and others have shown (Lambertini et al., 2016; Ellingjord-Dale et al., 2017; Ritte et al., 2013; Yang et al., 2011). We also documented that, irrespective of clinicopathological features, age at first pregnancy is negatively associated with age at breast cancer diagnosis. This observation, which needs to be validated in larger cohorts, is in line with the reported reduced risk of breast cancer associated with early pregnancy. Previous reports have hypothesized that tumor protection associated with early parity could be attributable to an improved antitumor immunity (Agrawal et al., 1995; Arklie et al., 1981; Erlebacher, 2013; Finn et al., 1995; Jungbluth et al., 2007). Neither parity status nor age at first birth was associated with tumor-infiltrating lymphocytes. However, an increased level of tumor-infiltrating lymphocytes was observed in tumors diagnosed shortly after pregnancy. At the genomic level, we demonstrated that age at first birth, but not parity status was associated with the genomic landscape of subsequent breast cancer. However, because there was a relatively small number of nulliparous patients, we cannot definitively conclude that parous and nulliparous patients have no difference in their genomic makeup. 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, and a lower prevalence of mutational signature 2. The

higher frequency of *TP53* mutations observed in early parous patients is of particular relevance given the known role of *TP53* in mediating the pregnancy-induced resistance to breast cancer. This finding may explain the differential effect of parity-induced protection according to tumor subtypes. Indeed, the beneficial effect of early pregnancy on breast cancer risk might be less effective against *TP53* mutated progenitor cells that are more prone to give rise to TNBC.

The lack of data on other reproductive factors such as age at menarche, time since last pregnancy and breastfeeding, precluded the possibility of conducting an extensive analysis of the complete reproductive history on the biology of breast cancer. Nonetheless, we hope that this study represents a first step toward the recognition that reproductive variables matter in order to understand the biology of female cancers fully. Of course, this study needs to be validated in others large-scale genomic studies of breast cancer. However, publicly available genomic studies, such as TCGA, usually lack reproductive history data. Therefore, this study advocates that reproductive history should be routinely collected in future studies addressing the biology of breast cancer but also of other female cancers. Indeed, the association between reproductive history and the risk of gynecological cancers such as ovarian (Hankinson et al., 1995; La Vecchia et al., 1993) and endometrial cancers (Titus-Ernstoff et al., 2001) have been established since long. Therefore, this work may open the door to further understand the link between reproductive history and the genomic landscape of other female cancers.

Complementing works of Azim et al. (Azim et al., 2012b, 2014, 2015a, 2015b), the second study of this thesis concludes one central research theme of the *J.-C. Heuson Breast Cancer Translational Laboratory*. Because we hypothesized that pregnancy could impact breast cancer biology, we characterized a unique cohort of patients with breast cancer diagnosed during pregnancy (BCP) and age/stage-matched non-pregnant control by performing whole-genome sequencing and copy number alteration profiling.

In this study, we identified important molecular differences characterizing BCP that may potentially represent a biologic explanation for their rather aggressive clinical behavior. Genome-wide copy number alteration profiling shown no striking difference between BCP and controls. However, multiple themes emerged from the whole-genome sequencing analysis. First, BCP was enriched in non-silent mutations that could have important oncogenic functions. Second, BCP was associated with a higher frequency of

mutation in mucin gene family. Particularly, the gain of serine residues, in addition to upregulation of mucins at the mRNA level, tempted us to speculate that BCP could be associated with mucins hyperglycosylation. Of course, the exact biological relevance of mucin mutations in breast cancer is still unknown and should be explored in future functional studies. Finally, analysis of breast cancer mutational signatures revealed that BCP were enriched in a process related to DNA mismatch repair deficiency. This is of particular clinical relevance as this feature has been shown to predict clinical benefit from immune checkpoint blockade in other types of cancers (Le et al., 2015; Rizvi et al., 2015).

This study had some technical considerations to take into account. Due to the rarity of this disease, we used archived FFPE samples that are known to be challenging for WGS due to DNA degradation and induction of artifacts. Nonetheless, BCP and controls were processed together without difference in the age of the samples or the sequencing coverage. Thus, an ideal validation study would be to carry out the same analysis on samples collected from fresh frozen tissues. As it is known that pregnancy induces epigenetic changes in epithelial cells to support mammary development (Huh et al., 2015), DNA methylation profiling could represent an attractive approach for future studies on the impact of pregnancy on breast cancer biology. We believe that this study represents a unique opportunity and provides new data to study the impact of pregnancy on breast cancer biology. The massive amount of sequencing data generated by this project is publicly available. It is our hope and expectation that this study might be considered as a useful resource aiming to shed new light on the biology of BCP and on the biology of breast cancer in young women in general.

In the last chapter of this thesis, we attempted to translate some biological insights gained by previous studies into improving the clinical management of breast cancer diagnosed in young women. Independently of classical clinicopathological features, it has been shown that young age is associated with worse prognosis and that breast tumor of young women might be considered as a unique biological entity. Previous work of Azim at al. and others have found that RANKL expression was associated with young age at breast cancer diagnosis. Given the preclinical evidence on the role of RANKL in mammary tumorigenesis and antitumor immunity, we initiated a preoperative window-of-opportunity trial evaluating the effect of RANKL inhibition by denosumab in young premenopausal breast cancer patients.

We found that denosumab was well-tolerated overall but did not meet its primary efficacy endpoint, with no decrease in the percentage of Ki-67 positive tumor cells. However, this intervention was associated with a tumor microenvironment response with an increased level of TILs, B, T and cytotoxic T cells and depletion of regulatory T cells. RNAseq analysis revealed that denosumab induces upregulation of inflammatory cytokines within the tumor. Thus, we hypothesized that the increased levels of TILs could be due to an influx of immune cells from the circulation. The study of potential biomarkers at baseline revealed that the presence of regulatory T cells and higher *FOXP3* expression was associated with an immunomodulatory response which supports that the influx of TILs might also be mediated by depletion of regulatory T cells.

The relatively small sample size precluded the possibility to perform subgroup analysis according to breast cancer subtypes. Moreover, this study included only premenopausal patients, a question that remains is whether the immunomodulatory response associated with denosumab could also be effective in postmenopausal patients. The lack of a control group is a second limitation that should be taken into consideration. To date, this is the first study on the effect of RANKL inhibition on breast cancer biology in the preoperative setting. Even if this study did not meet its primary endpoint, many new biological insights on the effect of a short course of denosumab on breast tumors have been gathered. The tumor immune microenvironment response is encouraging and is in line with previous preclinical observations.

In conclusion, it is our belief and hope that the different research projects conducted throughout this thesis represent a step forward toward a better understanding of the complex interplay between women's reproductive life and breast cancer biology. Breast cancer mortality has been decreased in the last decades, but this disease still represents a major health concern. Much remains to be learned to understand the disease fully but understanding its molecular pathology is a key element toward ultimately improving patients' lives. Finally, as the famous quote by Benjamin Franklin goes – "An ounce of prevention is worth a pound of cure", the in-depth study of natural mechanisms associated with breast cancer prevention, such as early pregnancy, could also open new avenues for contributing to the fight against cancer.

REFERENCES

van ’t Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Adams, S., Gray, R.J., Demaria, S., Goldstein, L., Perez, E.A., Shulman, L.N., Martino, S., Wang, M., Jones, V.E., Saphner, T.J., et al. (2014). Prognostic Value of Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancers From Two Phase III Randomized Adjuvant Breast Cancer Trials: ECOG 2197 and ECOG 1199. J. Clin. Oncol. *32*, 2959–2966.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

Agrawal, B., Reddish, M.A., Krantz, M.J., and Longenecker, B.M. (1995). Does pregnancy immunize against breast cancer? Cancer Res. *55*, 2257–2261.

Ahern, E., Harjunpää, H., Barkauskas, D., Allen, S., Takeda, K., Yagita, H., Wyld, D., Dougall, W.C., Teng, M.W.L., and Smyth, M.J. (2017). Co-administration of RANKL and CTLA4 antibodies enhances lymphocyte-mediated antitumor immunity in mice. Clin. Cancer Res. *23*, 5789–5801.

Ahern, E., Harjunpää, H., O’Donnell, J.S., Allen, S., Dougall, W.C., Teng, M.W.L., and Smyth, M.J. (2018a). RANKL blockade improves efficacy of PD1-PD-L1 blockade or dual PD1-PD-L1 and CTLA4 blockade in mouse models of cancer. Oncoimmunology.

Ahern, E., Smyth, M.J., Dougall, W.C., and Teng, M.W.L. (2018b). Roles of the RANKL–RANK axis in antitumour immunity — implications for therapy. Nat. Rev. Clin. Oncol.

Akiyama, T., Shimo, Y., Yanai, H., Qin, J., Ohshima, D., Maruyama, Y., Asaumi, Y., Kitazawa, J., Takayanagi, H., Penninger, J.M., et al. (2008). The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance. Immunity *29*, 423–437.

Albrektsen, G., Heuch, I., Hansen, S., and Kvåle, G. (2005). Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. Br. J. Cancer *92*, 167–175.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S. a J.R., Behjati, S., Biankin, A. V, Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat. Genet. *47*, 1402–1407.

Ali, H.R., Rueda, O.M., Chin, S.-F., Curtis, C., Dunning, M.J., Aparicio, S.A., and Caldas, C. (2014a). Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biol. *15*, 431.

Ali, H.R., Provenzano, E., Dawson, S.-J., Blows, F.M., Liu, B., Shah, M., Earl, H.M., Poole, C.J., Hiller, L., Dunn, J.A., et al. (2014b). Association between CD8+ T-cell infiltration and breast cancer survival in 12 439 patients. Ann. Oncol. *25*, 1536–1543.

Amant, F., von Minckwitz, G., Han, S.N., Bontenbal, M., Ring, A.E., Giermek, J., Wildiers, H., Fehm, T., Linn, S.C., Schlehe, B., et al. (2013). Prognosis of Women With Primary Breast Cancer Diagnosed During Pregnancy: Results From an International Collaborative Study. J. Clin. Oncol. *31*, 2532–2539.

Amant, F., Vandenbroucke, T., Verheecke, M., Fumagalli, M., Halaska, M.J., Boere, I., Han, S., Gziri, M.M., Peccatori, F., Rob, L., et al. (2015). Pediatric Outcome after Maternal Cancer Diagnosed during Pregnancy. N. Engl. J. Med. *373*, 1824–1834.

Anderson, J.M. (1979). Mammary cancers and pregnancy. Br. Med. J. *1*, 1124–1127.

Anderson, D.M., Maraskovsky, E., Billingsley, W.L., Dougall, W.C., Tometsko, M.E., Roux, E.R., Teepe, M.C., DuBose, R.F., Cosman, D., and Galibert, L. (1997). A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. Nature *390*, 175–179.

Anderson, S.M., Rudolph, M.C., McManaman, J.L., and Neville, M.C. (2007). Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Res. *9*, 204.

Andersson, T.M.-L., Johansson, A.L. V., Hsieh, C.-C., Cnattingius, S., and Lambe, M. (2009). Increasing Incidence of Pregnancy-Associated Breast Cancer in Sweden. Obstet. Gynecol. *114*, 568–572.

Arklie, J., Taylor-Papadimitrious, J., Bodmer, W., Egan, M., and Millis, R. (1981). Differentiation antigens expressed by epithelial cells in the lactating breast are also detectable in breast cancers. Int. J. Cancer *28*, 23–29.

Asztalos, S., Gann, P.H., Hayes, M.K., Nonn, L., Beam, C.A., Dai, Y., Wiley, E.L., and Tonetti, D.A. (2010). Gene expression patterns in the human breast after pregnancy. Cancer Prev. Res. (Phila). *3*, 301–311.

Azim, H.A., Santoro, L., Russell-Edu, W., Pentheroudakis, G., Pavlidis, N., and Peccatori, F.A. (2012a). Prognosis of pregnancy-associated breast cancer: a meta-analysis of 30 studies. Cancer Treat. Rev. *38*, 834–842.

Azim, H.A., Botteri, E., Renne, G., Dell'orto, P., Rotmensz, N., Gentilini, O., Sangalli, C., Pruneri, G., Di Nubila, B., Locatelli, M., et al. (2012b). The biological features and prognosis of breast cancer diagnosed during pregnancy: a case-control study. Acta Oncol. *51*, 653–661.

Azim, H.A., Michiels, S., Bedard, P.L., Singhal, S.K., Criscitiello, C., Ignatiadis, M., Haibe-Kains, B., Piccart, M.J., Sotiriou, C., and Loi, S. (2012c). Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. Clin. Cancer Res. *18*, 1341–1351.

Azim, H.A., Brohee, S., Peccatori, F.A., Desmedt, C., Loi, S., Lambrechts, D., Dell'Orto, P., Majjaj, S., Jose, V., Rotmensz, N., et al. (2014). Biology of breast cancer during pregnancy using genomic profiling. Endocr. Relat. Cancer *21*, 545–554.

Azim, H.A., Peccatori, F.A., Brohée, S., Branstetter, D., Loi, S., Viale, G., Piccart, M., Dougall, W.C., Pruneri, G., and Sotiriou, C. (2015a). RANK-ligand (RANKL) expression in young breast cancer patients and during pregnancy. Breast Cancer Res. *17*, 24.

Azim, H.A., Vingiani, A., Peccatori, F., Viale, G., Loi, S., and Pruneri, G. (2015b). Tumour infiltrating lymphocytes (TILs) in breast cancer during pregnancy. Breast *24*, 290–293.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. Nature *486*, 405–409.

Barbaroux, J.-B.O., Beleut, M., Brisken, C., Mueller, C.G., and Groves, R.W. (2008). Ligand Controls Langerhans Cells Numbers B κ Epidermal Receptor Activator of NF- Epidermal Receptor Activator of NF-□B Ligand Controls Langerhans Cells Numbers and Proliferation. J Immunol J. Immunol. by Guest January J. Immunol. *181*, 1103–1108.

Barnett, G.C., Shah, M., Redman, K., Easton, D.F., Ponder, B.A.J., and Pharoah, P.D.P. (2008). Risk factors for the incidence of breast cancer: do they affect survival from the disease? J. Clin. Oncol. *26*, 3310–3316.

Bates, G.J., Fox, S.B., Han, C., Leek, R.D., Garcia, J.F., Harris, A.L., and Banham, A.H. (2006). Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. J. Clin. Oncol. *24*, 5373–5380.

Bedard, P.L., Singhal, S.K., Ignatiadis, M., Bradbury, I., Haibe-Kains, B., Desmedt, C., Loi, S., Evans, D.B., Michiels, S., Dixon, J.M., et al. (2011). Low residual proliferation after short-term letrozole therapy is an early predictive marker of response in high proliferative ER-positive breast cancer. Endocr. Relat. Cancer *18*, 721–730.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

Biswas, S.K., and Lewis, C.E. (2010). NF-κB as a central regulator of macrophage function in tumors. J. Leukoc. Biol. *88*, 877–884.

Blake, M.L., Tometsko, M., Miller, R., Jones, J.C., and Dougall, W.C. (2014). RANK expression on breast cancer cells promotes skeletal metastasis. Clin. Exp. Metastasis *31*, 233–245.

Blakely, C.M., Stoddard, A.J., Belka, G.K., Dugan, K.D., Notarfrancesco, K.L., Moody, S.E., D'Cruz, C.M., and Chodosh, L.A. (2006). Hormone-induced protection against mammary tumorigenesis is conserved in multiple rat strains and identifies a core gene expression signature induced by pregnancy. Cancer Res. *66*, 6421–6431.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Buisseret, L., Garaud, S., de Wind, A., Van den Eynden, G., Boisson, A., Solinas, C., Gu-Trantien, C., Naveaux, C., Lodewyckx, J.-N., Duvillier, H., et al. (2017). Tumor-infiltrating lymphocyte composition, organization and PD-1/ PD-L1 expression are linked in breast cancer. Oncoimmunology *6*, e1257452.

Cardonick, E., Bhat, A., Gilmandyar, D., and Somer, R. (2012). Maternal and fetal outcomes of taxane chemotherapy in breast and ovarian cancer during pregnancy: case series and review of the literature. Ann. Oncol. *23*, 3016–3023.

Cardoso, F., van't Veer, L.J., Bogaerts, J., Slaets, L., Viale, G., Delaloge, S., Pierga, J.-Y., Brain, E., Causeret, S., DeLorenzi, M., et al. (2016). 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N. Engl. J. Med. *375*, 717–729.

Charles, J.F., and Aliprantis, A.O. (2014). Osteoclasts: more than 'bone eaters.' Trends Mol. Med. *20*, 449–459.

Cheng, M.L., and Fong, L. (2014). Effects of RANKL-Targeted Therapy in Immunity and Cancer. Front. Oncol. *3*, 1–8.

Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell *163*, 506–519.

Clevenger, C. V., Furth, P.A., Hankinson, S.E., and Schuler, L.A. (2003). The Role of Prolactin in Mammary Carcinoma. Endocr. Rev. *24*, 1–27.

Coates, A.S., Winer, E.P., Goldhirsch, A., Gelber, R.D., Gnant, M., Piccart-Gebhart, M., Thürlimann, B., Senn, H.-J., André, F., Baselga, J., et al. (2015). Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann. Oncol. *26*, 1533–1546.

Coleman, R.E., Finkelstein, D., Barrios, C.H., Martin, M., Iwata, H., Glaspy, J.A., Zhou, Y., Jandial, D., and Chan, A. (2018). Adjuvant denosumab in early breast cancer: First results from the international multicenter randomized phase III placebo controlled D-CARE study. J. Clin. Oncol. *36*, 501.

Collaborative Group on Hormonal Factors in Breast Cancer (2002). Breast cancer and

breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet (London, England) *360*, 187–195.

Costa, L., Badia, X., Chow, E., Lipton, A., and Wardley, A. (2008). Impact of skeletal complications on patients' quality of life, mobility, and functional independence. Support. Care Cancer *16*, 879–889.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Davies, H., Glodzik, D., Morganella, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M., et al. (2017a). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat. Med. (in Press.

Davies, H., Morganella, S., Purdie, C.A., Jang, S.J., Borgen, E., Russnes, H., Glodzik, D., Zou, X., Viari, A., Richardson, A.L., et al. (2017b). Whole-Genome Sequencing Reveals Breast Cancers with Mismatch Repair Deficiency. Cancer Res. 4755–4763.

Dawson, S.-J., Rueda, O.M., Aparicio, S., and Caldas, C. (2013). A new genome-driven integrated classification of breast cancer and its implications. EMBO J. *32*, 617–628.

DeNardo, D.G., and Coussens, L.M. (2007). Inflammation and breast cancer. Balancing immune response: crosstalk between adaptive and innate immune cells during breast cancer progression. Breast Cancer Res. *9*, 212.

Denkert, C., Loibl, S., Noske, A., Roller, M., Müller, B.M., Komor, M., Budczies, J., Darb-Esfahani, S., Kronenwett, R., Hanusch, C., et al. (2010). Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. *28*, 105–113.

Desmedt, C., Zoppoli, G., Gundem, G., Pruneri, G., Larsimont, D., Fornili, M., Fumagalli, D., Brown, D., Rothé, F., Vincent, D., et al. (2016). Genomic Characterization of Primary Invasive Lobular Breast Cancer. J. Clin. Oncol. *34*, 1872–1880.

Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., et al. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature *464*, 999–1005.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dougall, W.C., Glaccum, M., Charrier, K., Rohrbach, K., Brasel, K., De Smedt, T., Daro, E., Smith, J., Tometsko, M.E., Maliszewski, C.R., et al. (1999). RANK is essential for osteoclast and lymph node development. Genes Dev. *13*, 2412–2424.

Dufour, J.H., Dziejman, M., Liu, M.T., Leung, J.H., Lane, T.E., and Luster, A.D. (2002). IFN- -Inducible Protein 10 (IP-10; CXCL10)-Deficient Mice Reveal a Role for IP-10 in Effector T Cell Generation and Trafficking. J. Immunol. *168*, 3195–3204.

Dunphy, K.A., Blackburn, A.C., Yan, H., O'Connell, L.R., and Jerry, D.J. (2008). Estrogen and progesterone induce persistent increases in p53-dependent apoptosis and suppress mammary tumors in BALB/c-Trp53+/-mice. Breast Cancer Res. *10*, R43.

Eibye, S., Kjær, S.K., and Mellemkjær, L. (2013). Incidence of Pregnancy-Associated Cancer in Denmark, 1977–2006. Obstet. Gynecol. *122*, 608–617.

Eliassen, A.H., Tworoger, S.S., and Hankinson, S.E. (2007). Reproductive factors and family history of breast cancer in relation to plasma prolactin levels in premenopausal and postmenopausal women. Int. J. Cancer *120*, 1536–1541.

Ellingjord-Dale, M., Vos, L., Tretli, S., Hofvind, S., dos-Santos-Silva, I., and Ursin, G. (2017). Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. Breast Cancer Res. *19*, 10.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C., et al. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature *486*, 353–360.

Emens, L.A. (2017). Breast Cancer Immunotherapy: Facts and Hopes. Clin. Cancer Res. clincanres.3001.2017.

Erlebacher, A. (2013). Mechanisms of T cell tolerance towards the allogeneic fetus. Nat. Rev. Immunol. *13*, 23–33.

Farrell, J.P., Padigel, U.M., Kim, N., and Choi, Y. (2003). Infection in CD40L-Deficient Mice Leishmania major Th1-Type Response to for IL-12 Production and the Initiation of a TRANCE-RANK Costimulation is Required TRANCE-RANK Costimulation is Required for IL-12 Production and the Initiation of a Th1-Type Response. J Immunol Ref. *171*, 5437–5441.

Fata, J.E., Kong, Y.-Y.Y., Li, J., Sasaki, T., Irie-Sasaki, J., Moorehead, R.A., Elliott, R., Scully, S., Voura, E.B., Lacey, D.L., et al. (2000). The Osteoclast Differentiation Factor Osteoprotegerin-Ligand Is Essential for Mammary Gland Development. Cell *103*, 41–50.

Faupel-Badger, J.M., Arcaro, K.F., Balkam, J.J., Eliassen, A.H., Hassiotou, F., Lebrilla, C.B., Michels, K.B., Palmer, J.R., Schedin, P., Stuebe, A.M., et al. (2013). Postpartum Remodeling, Lactation, and Breast Cancer Risk: Summary of a National Cancer Institute–Sponsored Workshop. JNCI J. Natl. Cancer Inst. *105*, 166–174.

Ferrari-Lacraz, S., and Ferrari, S. (2011). Do RANKL inhibitors (denosumab) affect inflammation and immunity? Osteoporos. Int. *22*, 435–446.

Finn, O.J., Jerome, K.R., Henderson, R.A., Pecher, G., Domenech, N., Magarian-Blander, J., and Barratt-Boyes, S.M. (1995). MUC-1 epithelial tumor mucin-based immunity and cancer vaccines. Immunol. Rev. *145*, 61–89.

Forbes, J.F., Cuzick, J., Buzdar, A., Howell, A., Tobias, J.S., and Baum, M. (2008). Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 100-month analysis of the ATAC trial. Lancet Oncol. *9*, 45–53.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: Somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

Fornetti, J., Martinson, H.A., Betts, C.B., Lyons, T.R., Jindal, S., Guo, Q., Coussens, L.M., Borges, V.F., and Schedin, P. (2014). Mammary Gland Involution as an Immunotherapeutic Target for Postpartum Breast Cancer. J. Mammary Gland Biol. Neoplasia *19*, 213–228.

Gendoo, D.M.A., Ratanasirigulchai, N., Schröder, M.S., Paré, L., Parker, J.S., Prat, A., and Haibe-Kains, B. (2016). Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics *32*, 1097–1099.

Gentilini, O., Cremonesi, M., Trifirò, G., Ferrari, M., Baio, S.M., Caracciolo, M., Rossi, A., Smeets, A., Galimberti, V., Luini, A., et al. (2004). Safety of sentinel node biopsy in pregnant patients with breast cancer. Ann. Oncol. *15*, 1348–1351.

Gentilini, O., Cremonesi, M., Toesca, A., Colombo, N., Peccatori, F., Sironi, R., Sangalli, C., Rotmensz, N., Pedroli, G., Viale, G., et al. (2010). Sentinel lymph node biopsy in pregnant patients with breast cancer. Eur. J. Nucl. Med. Mol. Imaging *37*, 78–83.

Van der Giessen, P.H. (1997). Measurement of the peripheral dose for the tangential breast treatment technique with Co-60 gamma radiation and high energy X-rays. Radiother. Oncol. *42*, 257–264.

Gnant, M., Pfeiler, G., Dubsky, P.C., Hubalek, M., Greil, R., Jakesz, R., Wette, V., Balic, M., Haslbauer, F., Melbinger, E., et al. (2015). Adjuvant denosumab in breast cancer (ABCSG-18): A multicentre, randomised, double-blind, placebo-controlled trial. Lancet *386*, 433–443.

Gnant, M., Pfeiler, G., Steger, G.G., Egle, D., Greil, R., Fitzal, F., Wette, V., Balic, M., Haslbauer, F., Melbinger-Zeinitzer, E., et al. (2018). Adjuvant denosumab in early breast cancer: Disease-free survival analysis of 3,425 postmenopausal patients in the ABCSG-18 trial. J. Clin. Oncol. *36*, 500.

González-Pérez, A., and López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am. J. Hum. Genet. *88*, 440–449.

Gonzalez-Suarez, E., and Sanz-Moreno, A. (2016). RANK as a therapeutic target in cancer. Febs J *283*, 2018–2033.

Gonzalez-Suarez, E., Jacob, A.P., Jones, J., Miller, R., Roudier-Meyer, M.P., Erwert, R., Pinkas, J., Branstetter, D., and Dougall, W.C. (2010a). RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature *468*, 103–107.

Gonzalez-Suarez, E., Jacob, A.P., Jones, J., Miller, R., Roudier-Meyer, M.P., Erwert, R., Pinkas, J., Branstetter, D., and Dougall, W.C. (2010b). RANK ligand mediates progestin-induced mammary epithelial proliferation and carcinogenesis. Nature *468*, 103–107.

Green, E.A., Choi, Y., and Flavell, R.A. (2002). Pancreatic lymph node-derived CD4+CD25+ treg cells: Highly potent regulators of diabetes that require TRANCE-RANK signals. Immunity *16*, 183–191.

de Groot, A.F., Appelman-Dijkstra, N.M., van der Burg, S.H., and Kroep, J.R. (2018). The anti-tumor effect of RANKL inhibition in malignant solid tumors – A systematic review. Cancer Treat. Rev. *62*, 18–28.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Hanisch, F.-G. (2001). O-Glycosylation of the Mucin Type. Biol. Chem. *382*, 143–149.

Hankinson, S.E., Colditz, G.A., Hunter, D.J., Willett, W.C., Stampfer, M.J., Rosner, B., Hennekens, C.H., and Speizer, F.E. (1995). A prospective study of reproductive factors and risk of epithelial ovarian cancer. Cancer *76*, 284–290.

Hartman, E.K., and Eslick, G.D. (2016). The prognosis of women diagnosed with breast cancer before, during and after pregnancy: a meta-analysis. Breast Cancer Res. Treat. *160*, 347–360.

Harvell, D.M.E., Kim, J., O'Brien, J., Tan, A.-C., Borges, V.F., Schedin, P., Jacobsen, B.M., and Horwitz, K.B. (2013). Genomic signatures of pregnancy-associated breast cancer epithelia and stroma and their regulation by estrogens and progesterone. Horm. Cancer *4*, 140–153.

Hennighausen, L., and Robinson, G.W. (2001). Signaling Pathways in Mammary Gland Development. Dev. Cell *1*, 467–475.

Hermeking, H., and Eick, D. (1994). Mediation of c-Myc-induced apoptosis by p53. Science (80-.). *265*, 2091–2093.

Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N. Engl. J. Med. *363*, 711–723.

Hollingsworth, M.A., and Swanson, B.J. (2004). Mucins in cancer: protection and control of the cell surface. Nat. Rev. Cancer *4*, 45–60.

Huh, S.J., Clement, K., Jee, D., Merlini, A., Choudhury, S., Maruyama, R., Yoo, R., Chytil, A., Boyle, P., Ran, F.A., et al. (2015). Age- and pregnancy-associated dna methylation changes in mammary epithelial cells. Stem Cell Reports *4*, 297–311.

Islami, F., Liu, Y., Jemal, A., Zhou, J., Weiderpass, E., Colditz, G., Boffetta, P., and Weiss, M. (2015). Breastfeeding and breast cancer risk by receptor status—a systematic review and meta-analysis. Ann. Oncol. *26*, mdv379.

Jerry, D.J., Kittrell, F.S., Kuperwasser, C., Laucirica, R., Dickinson, E.S., Bonilla, P.J., Butel, J.S., and Medina, D. (2000). A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development. Oncogene *19*, 1052–1058.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat. Rev. Mol. Cell Biol. *7*, 335–346.

Josefowicz, S.Z., Lu, L.-F., and Rudensky, A.Y. (2012). Regulatory T Cells: Mechanisms of Differentiation and Function. Annu. Rev. Immunol. *30*, 531–564.

Jungbluth, A.A., Silva, W.A., Iversen, K., Frosina, D., Zaidi, B., Coplan, K., Eastlake-Wade, S.K., Castelli, S.B., Spagnoli, G.C., Old, L.J., et al. (2007). Expression of cancer-testis (CT) antigens in placenta. Cancer Immun. *7*, 15.

Kal, H.B., and Struikmans, H. (2005). Radiotherapy during pregnancy: fact and fiction. Lancet Oncol. *6*, 328–333.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.

Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M.C.U., Voduc, D., Speers, C.H., Nielsen, T.O., and Gelmon, K. (2010). Metastatic behavior of breast cancer subtypes. J. Clin. Oncol. *28*, 3271–3277.

Knoop, K.A., Kumar, N., Butler, B.R., Sakthivel, S.K., Taylor, R.T., Nochi, T., Akiba, H., Yagita, H., Kiyono, H., and Williams, I.R. (2009). RANKL Is Necessary and Sufficient to Initiate Development of Antigen-Sampling M Cells in the Intestinal Epithelium. J. Immunol. *183*, 5738–5747.

Kong, Y.-Y., Yoshida, H., Sarosi, I., Tan, H.-L., Timms, E., Capparelli, C., Morony, S., Oliveira-dos-Santos, A.J., Van, G., Itie, A., et al. (1999). OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. Nature *397*, 315–323.

Kroman, N., Wohlfahrt, J., Andersen, K.W., Mouridsen, H.T., Westergaard, T., and Melbye, M. (1997). Time since childbirth and prognosis in primary breast cancer: population based study. BMJ *315*, 851–855.

Lacey, D.L., Boyle, W.J., Simonet, W.S., Kostenuik, P.J., Dougall, W.C., Sullivan, J.K., Martin, J.S., and Dansey, R. (2012). Bench to bedside: elucidation of the OPG–RANK–RANKL pathway and the development of denosumab. Nat. Rev. Drug Discov. *11*, 401–419.

Lambertini, M., Peccatori, F.A., and Azim, H.A. (2015). Targeted agents for cancer treatment during pregnancy. Cancer Treat. Rev. *41*, 301–309.

Lambertini, M., Santoro, L., Del Mastro, L., Nguyen, B., Livraghi, L., Ugolini, D., Peccatori, F.A., and Azim, H.A. (2016). Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat. Rev. *49*, 65–76.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. *372*, 2509–2520.

Lee, Y., Roberts, C., Dobbins, T., Stavrou, E., Black, K., Morris, J., and Young, J. (2012). Incidence and outcomes of pregnancy-associated cancer in Australia, 1994-2008: a population-

based linkage study. BJOG An Int. J. Obstet. Gynaecol. *119*, 1572–1582.

Leibbrandt, A., and Penninger, J.M. (2008). RANK/RANKL: Regulators of Immune Responses and Bone Physiology. Ann. N. Y. Acad. Sci. *1143*, 123–150.

Leiserson, M.D.M., Vandin, F., Wu, H.T., and Raphael, B.J. (2016). Reply: Co-occurrence of MYC amplification and TP53 mutations in human cancer. Nat. Genet. *48*, 106–108.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liberti, M. V, and Locasale, J.W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem. Sci. *41*, 211–218.

Lohsiriwat, V., Peccatori, F.A., Martella, S., Azim, H.A., Sarno, M.A., Galimberti, V., De Lorenzi, F., Intra, M., Sangalli, C., Rotmensz, N., et al. (2013). Immediate breast reconstruction with expander in pregnant breast cancer patients. The Breast *22*, 657–660.

Loi, S., Sirtaine, N., Piette, F., Salgado, R., Viale, G., Van Eenoo, F., Rouas, G., Francis, P., Crown, J.P.A., Hitre, E., et al. (2013). Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. J. Clin. Oncol. *31*, 860–867.

Loi, S., Michiels, S., Salgado, R., Sirtaine, N., Jose, V., Fumagalli, D., Kellokumpu-Lehtinen, P.-L., Bono, P., Kataja, V., Desmedt, C., et al. (2014). Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. Ann. Oncol. *25*, 1544–1550.

Loi, S., Giobbe-Hurder, A., Gombos, A., Bachelot, T., Hui, R., Curigliano, G., Campone, M., Biganzoli, L., Bonnefoi, H., Jerusalem, G., et al. (2018). Abstract GS2-06: Phase Ib/II study evaluating safety and efficacy of pembrolizumab and trastuzumab in patients with trastuzumab-resistant HER2-positive metastatic breast cancer: Results from the PANACEA (IBCSG 45-13/BIG 4-13/KEYNOTE-014) study. Cancer Res. *78*, GS2-06-GS2-06.

Loibl, S., Schmidt, A., Gentilini, O., Kaufman, B., Kuhl, C., Denkert, C., von Minckwitz, G., Parokonnaya, A., Stensheim, H., Thomssen, C., et al. (2015a). Breast Cancer Diagnosed During Pregnancy. JAMA Oncol. *1*, 1145.

Loibl, S., Schmidt, A., Gentilini, O., Kaufman, B., Kuhl, C., Denkert, C., von Minckwitz, G., Parokonnaya, A., Stensheim, H., Thomssen, C., et al. (2015b). Breast Cancer Diagnosed During Pregnancy: Adapting Recent Advances in Breast Cancer Care for Pregnant Patients. JAMA Oncol. *1*, 1145–1153.

Long, F. (2012). Building strong bones: molecular regulation of the osteoblast lineage. Nat. Rev. Mol. Cell Biol. *13*, 27–38.

López-Knowles, E., Gao, Q., Cheang, M.C.U., Morden, J., Parker, J., Martin, L.A., Pinhel, I., McNeill, F., Hills, M., Detre, S., et al. (2016). Heterogeneity in global gene expression profiles between biopsy specimens taken peri-surgically from primary ER-positive breast carcinomas. Breast Cancer Res. *18*.

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

Luen, S.J., Salgado, R., Fox, S., Savas, P., Eng-Wong, J., Clark, E., Kiermaier, A., Swain, S.M., Baselga, J., Michiels, S., et al. (2017). Tumour-infiltrating lymphocytes in advanced HER2-positive breast cancer treated with pertuzumab or placebo in addition to trastuzumab and docetaxel: a retrospective analysis of the CLEOPATRA study. Lancet Oncol. *18*, 52–62.

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., and Woolf, P.J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics *10*, 161.

Lyons, T.R., Schedin, P.J., and Borges, V.F. (2009). Pregnancy and breast cancer: When they collide. J. Mammary Gland Biol. Neoplasia *14*, 87–98.

Lyons, T.R., O'Brien, J., Borges, V.F., Conklin, M.W., Keely, P.J., Eliceiri, K.W., Marusyk, A., Tan, A.-C., and Schedin, P. (2011). Postpartum mammary gland involution drives progression of ductal carcinoma in situ through collagen and COX-2. Nat. Med. *17*, 1109–1115.

Lyons, T.R., Borges, V.F., Betts, C.B., Guo, Q., Kapoor, P., Martinson, H.A., Jindal, S., and Schedin, P. (2014). Cyclooxygenase-2–dependent lymphangiogenesis promotes nodal metastasis of postpartum breast cancer. J. Clin. Invest. *124*, 3901–3912.

Ma, X.-J., Salunga, R., Dahiya, S., Wang, W., Carney, E., Durbecq, V., Harris, A., Goss, P., Sotiriou, C., Erlander, M., et al. (2008). A Five-Gene Molecular Grade Index and HOXB13:IL17BR Are Complementary Prognostic Factors in Early Stage Breast Cancer. Clin. Cancer Res. *14*, 2601–2608.

MacMahon, B., Cole, P., Lin, T.M., Lowe, C.R., Mirra, A.P., Ravnihar, B., Salber, E.J., Valaoras, V.G., and Yuasa, S. (1970). Age at first birth and breast cancer risk. Bull. World Health Organ. *43*, 209–221.

Mahmoud, S.M.A., Paish, E.C., Powe, D.G., Macmillan, R.D., Grainge, M.J., Lee, A.H.S., Ellis, I.O., and Green, A.R. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. J. Clin. Oncol. *29*, 1949–1955.

Maller, O., Hansen, K.C., Lyons, T.R., Acerbi, I., Weaver, V.M., Prekeris, R., Tan, A.-C., and Schedin, P. (2013). Collagen architecture in pregnancy-induced protection from breast cancer. J. Cell Sci. *126*, 4108–4110.

Mantovani, A., Sica, A., and Locati, M. (2007). New vistas on macrophage differentiation and activation. Eur. J. Immunol. *37*, 14–16.

Martinson, H.A., Jindal, S., Durand-Rougely, C., Borges, V.F., and Schedin, P. (2015). Wound healing-like immune program facilitates postpartum mammary gland involution and tumor progression. Int. J. Cancer *136*, 1803–1813.

McDaniel, S.M., Rumer, K.K., Biroc, S.L., Metz, R.P., Singh, M., Porter, W., and Schedin, P. (2006). Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. Am. J. Pathol. *168*, 608–620.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. 042374.

Medina, D. (2005). Mammary developmental fate and breast cancer risk. Endocr. Relat. Cancer *12*, 483–495.

Medina, D., and Kittrell, F.S. (2003). p53 function is required for hormone-mediated protection of mouse mammary tumorigenesis. Cancer Res. *63*, 6140–6143.

Meier-Abt, F., and Bentires-Alj, M. (2014). How pregnancy at early age protects against breast cancer. Trends Mol. Med. *20*, 143–153.

Meier-Abt, F., Milani, E., Roloff, T., Brinkhaus, H., Duss, S., Meyer, D.S., Klebba, I., Balwierz, P.J., van Nimwegen, E., and Bentires-Alj, M. (2013). Parity induces differentiation and reduces Wnt/Notch signaling ratio and proliferation potential of basal stem/progenitor cells isolated from mouse mammary epithelium. Breast Cancer Res. *15*, R36.

Meier-Abt, F., Brinkhaus, H., and Bentires-Alj, M. (2014). Early but not late pregnancy induces lifelong reductions in the proportion of mammary progesterone sensing cells and epithelial Wnt signaling. Breast Cancer Res. *16*, 402.

Meier-Abt, F., Bentires-Alj, M., and Rochlitz, C. (2015). Breast cancer prevention: Lessons to be learned from mechanisms of early pregnancy-mediated breast cancer protection. Cancer Res. *75*, 803–807.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41.

Mika, T., Minna, T., R, K., A, K., J, I., and OP., K. (1998). Molecular cytogenetics of primary breast cancer by CGH. Genes Chromosom. Cancer. *21*, 177–184.

Miller, W.R., Larionov, A., Renshaw, L., Anderson, T.J., Walker, J.R., Krause, A., Sing, T., Evans, D.B., and Dixon, J.M. (2009). Gene expression profiles differentiating between breast cancers clinically responsive or resistant to letrozole. J. Clin. Oncol. *27*, 1382–1387.

von Moos, R., and Haynes, I. (2013). Where Do Bone-Targeted Agents RANK in Breast Cancer Treatment? J. Clin. Med. *2*, 89–102.

Morganella, S., Alexandrov, L.B., Glodzik, D., Zou, X., Davies, H., Staaf, J., Sieuwerts, A.M., Brinkman, A.B., Martin, S., Ramakrishna, M., et al. (2016). The topography of mutational processes in breast cancer genomes. Nat. Commun. *7*, 11383.

Mroz, E.A., and Rocco, J.W. (2013). MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol. *49*, 211–215.

Mukhopadhyay, P., Chakraborty, S., Ponnusamy, M.P., Lakshmanan, I., Jain, M., and Batra, S.K. (2011). Mucins in the pathogenesis of breast cancer: implications in diagnosis, prognosis and therapy. Biochim. Biophys. Acta *1815*, 224–240.

Nagatsuma, A.K., Shimizu, C., Takahashi, F., Tsuda, H., Saji, S., Hojo, T., Sugano, K., Takeuchi, M., Fujii, H., and Fujiwara, Y. (2013). Impact of recent parity on histopathological tumor features and breast cancer outcome in premenopausal Japanese women. Breast Cancer Res. Treat. *138*, 941–950.

National Toxicology Program (2013). NTP Monograph: Developmental Effects and Pregnancy Outcomes Associated With Cancer Chemotherapy Use During Pregnancy. NTP Monogr. i-214.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012a). The life history of 21 breast cancers. Cell *149*, 994–1007.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012b). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 1–20.

Nilsen, G., Liestøl, K., Van Loo, P., Moen Vollan, H.K., Eide, M.B., Rueda, O.M., Chin, S.-F., Russell, R., Baumbusch, L.O., Caldas, C., et al. (2012). Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics *13*, 591.

Nolan, E., Vaillant, F., Branstetter, D., Pal, B., Giner, G., Whitehead, L., Lok, S.W., Mann, G.B., Rohrbach, K., Huang, L.-Y., et al. (2016). RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers. Nat. Med. *22*, 933–939.

North, B. V, Curtis, D., and Sham, P.C. (2002). A note on the calculation of empirical P values from Monte Carlo procedures. Am. J. Hum. Genet. *71*, 439–441.

Ohara, M., Yamaguchi, Y., Matsuura, K., Murakami, S., Arihiro, K., and Okada, M. (2009). Possible involvement of regulatory T cells in tumor onset and progression in primary breast

cancer. Cancer Immunol. Immunother. *58*, 441–447.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. N. Engl. J. Med. *351*, 2817–2826.

Palafox, M., Ferrer, I., Pellegrini, P., Vila, S., Hernandez-Ortega, S., Urruticoechea, A., Climent, F., Soler, M.T., Muñoz, P., Viñals, F., et al. (2012). RANK induces epithelial-mesenchymal transition and stemness in human mammary epithelial cells and promotes tumorigenesis and metastasis. Cancer Res. *72*, 2879–2888.

Papatestas, A.E., Mulvihill, M., Josi, C., Ioannovich, J., Lesnick, G., and Aufses, A.H. (1980). Parity and prognosis in breast cancer. Cancer *45*, 191–194.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J. Clin. Oncol. *27*, 1160–1167.

Peccatori, F.A., Azim, H.A., Orecchia, R., Hoekstra, H.J., Pavlidis, N., Kesic, V., and Pentheroudakis, G. (2013). Cancer, pregnancy and fertility: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. / ESMO *24 Suppl 6*, vi160-170.

Pereira, B., Chin, S.-F., Rueda, O.M., Vollan, H.-K.M., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat. Commun. *7*, 11479.

Peri, S., de Cicco, R.L., Santucci-Pereira, J., Slifker, M., Ross, E.A., Russo, I.H., Russo, P.A., Arslan, A.A., Belitskaya-Lévy, I., Zeleniuch-Jacquotte, A., et al. (2012). Defining the genomic signature of the parous breast. BMC Med. Genomics *5*, 46.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Pfitzner, B.M., Branstetter, D., Loibl, S., Denkert, C., Lederer, B., Schmitt, W.D., Dombrowski, F., Werner, M., Rüdiger, T., Dougall, W.C., et al. (2014). RANK expression as a prognostic and predictive marker in breast cancer. Breast Cancer Res. Treat. *145*, 307–315.

Polyak, K. (2006). Pregnancy and breast cancer: The other side of the coin. Cancer Cell *9*, 151–153.

Popova, T., Manié, E., Stoppa-Lyonnet, D., Rigaill, G., Barillot, E., and Stern, M.H. (2009). Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome Biol. *10*, R128.

Prat, A., Ellis, M.J., and Perou, C.M. (2012). Practical implications of gene-expression-based assays for breast oncologists. Nat. Rev. Clin. Oncol. *9*, 48–57.

Rao, S., Cronin, S.J.F., Sigl, V., and Penninger, J.M. (2017). RANKL and RANK: From Mammalian Physiology to Cancer Treatment. Trends Cell Biol. *0*, 1–11.

Ritte, R., Tikk, K., Lukanova, A., Tjønneland, A., Olsen, A., Overvad, K., Dossus, L., Fournier, A., Clavel-Chapelon, F., Grote, V., et al. (2013). Reproductive factors and risk of hormone receptor positive and negative breast cancer: A cohort study. BMC Cancer *13*, 1–12.

Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (80-.). *348*, 124–128.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48–61.

Rosenberg, L., Thalib, L., Adami, H.-O., and Hall, P. (2004). Childbirth and breast cancer

prognosis. Int. J. Cancer *111*, 772–776.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31.

Rossi, S.W., Kim, M.-Y., Leibbrandt, A., Parnell, S.M., Jenkinson, W.E., Glanville, S.H., McConnell, F.M., Scott, H.S., Penninger, J.M., Jenkinson, E.J., et al. (2007). RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. J. Exp. Med. *204*, 1267–1272.

Russo, J., Tay, L.K., and Russo, I.H. (1982). Differentiation of the mammary gland and susceptibility to carcinogenesis. Breast Cancer Res. Treat. *2*, 5–73.

Russo, J., Balogh, G.A., and Russo, I.H. (2008). Full-term Pregnancy Induces a Specific Genomic Signature in the Human Breast. Cancer Epidemiol. Biomarkers Prev. *17*, 51–66.

Russo, J., Santucci-Pereira, J., De Cicco, R.L., Sheriff, F., Russo, P.A., Peri, S., Slifker, M., Ross, E., Mello, M.L.S., Vidal, B.C., et al. (2012). Pregnancy-induced chromatin remodeling in the breast of postmenopausal women. Int. J. Cancer *131*, 1059–1070.

Salgado, R., and Loi, S. (2018). Tumour infiltrating lymphocytes in breast cancer: increasing clinical relevance. Lancet Oncol. *19*, 3–5.

Salgado, R., Denkert, C., Demaria, S., Sirtaine, N., Klauschen, F., Pruneri, G., Wienert, S., Van den Eynden, G., Baehner, F.L., Penault-Llorca, F., et al. (2015). The evaluation of tumor-infiltrating lymphocytes (TILS) in breast cancer: Recommendations by an International TILS Working Group 2014. Ann. Oncol. *26*, 259–271.

Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics *28*, 1811–1817.

Savas, P., Salgado, R., Denkert, C., Sotiriou, C., Darcy, P.K., Smyth, M.J., and Loi, S. (2016). Clinical relevance of host immunity in breast cancer: From TILs to the clinic. Nat. Rev. Clin. Oncol. *13*, 228–241.

Schedin, P. (2006). Pregnancy-associated breast cancer and metastasis. Nat. Rev. Cancer *6*, 281–291.

Schramek, D., Leibbrandt, A., Sigl, V., Kenner, L., Pospisilik, J.A., Lee, H.J., Hanada, R., Joshi, P.A., Aliprantis, A., Glimcher, L., et al. (2010a). Osteoclast differentiation factor RANKL controls development of progestin-driven mammary cancer. Nature *468*, 98–102.

Schramek, D., Leibbrandt, A., Sigl, V., Kenner, L., Pospisilik, J.A., Lee, H.J., Hanada, R., Joshi, P.A., Aliprantis, A., Glimcher, L., et al. (2010b). Osteoclast differentiation factor RANKL controls development of progestin-driven mammary cancer. Nature *468*, 98–102.

Seshasayee, D., Wang, H., Lee, W.P., Gribling, P., Ross, J., Van Bruggen, N., Carano, R., and Grewal, I.S. (2004). A novel in vivo role for osteoprotegerin ligand in activation of monocyte effector function and inflammatory response. J. Biol. Chem. *279*, 30202–30209.

Sgambato, A., Casaluce, F., Sacco, P.C., Palazzolo, G., Maione, P., Rossi, A., Ciardiello, F., and Gridelli, C. (2016). Anti PD-1 and PDL-1 Immunotherapy in the Treatment of Advanced Non-Small Cell Lung Cancer (NSCLC): A Review on Toxicity Profile and its Management. Curr. Drug Saf. *11*, 62–68.

Shah, S.P., Morin, R.D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., et al. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature *461*, 809–813.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative

breast cancers. Nature *486*, 395–399.

Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. CA. Cancer J. Clin. *68*, 7–30.

Sigl, V., Owusu-Boaitey, K., Joshi, P. a, Kavirayani, A., Wirnsberger, G., Novatchkova, M., Kozieradzki, I., Schramek, D., Edokobi, N., Hersl, J., et al. (2016a). RANKL/RANK control Brca1 mutation-driven mammary tumors. Cell Res. *26*, 761–774.

Sigl, V., Jones, L.P., and Penninger, J.M. (2016b). RANKL/RANK: from bone loss to the prevention of breast cancer. Open Biol. *6*, 160230.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452-7.

Sivaraman, L., Conneely, O.M., Medina, D., and O'Malley, B.W. (2001). P53 Is a Potential Mediator of Pregnancy and Hormone-Induced Resistance To Mammary Carcinogenesis. Pnas *98*, 12379–12384.

Smyth, M.J., Yagita, H., and McArthur, G.A. (2013). Combination Anti-CTLA-4 and Anti-RANKL in Metastatic Melanoma. J. Clin. Oncol. *31*, 2013–2015.

Solinas, C., Gombos, A., Latifyan, S., Piccart-Gebhart, M., Kok, M., and Buisseret, L. (2017). Targeting immune checkpoints in breast cancer: an update of early results. ESMO Open *2*, e000255.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. *98*, 10869–10874.

Sotiriou, C., and Pusztai, L. (2009). Gene-Expression Signatures in Breast Cancer. N. Engl. J. Med. *360*, 790–800.

Sotiriou, C., Neo, S.-Y.-Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L., and Liu, E.T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc. Natl. Acad. Sci. U. S. A. *100*, 10393–10398.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., et al. (2006). Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. JNCI J. Natl. Cancer Inst. *98*, 262–272.

Sparano, J.A., Gray, R.J., Makower, D.F., Pritchard, K.I., Albain, K.S., Hayes, D.F., Geyer, C.E., Dees, E.C., Perez, E.A., Olson, J.A., et al. (2015). Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. N. Engl. J. Med. *373*, 2005–2014.

Sparano, J.A., Gray, R.J., Makower, D.F., Pritchard, K.I., Albain, K.S., Hayes, D.F., Geyer, C.E., Dees, E.C., Goetz, M.P., Olson, J.A., et al. (2018). Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N. Engl. J. Med. *379*, 111–121.

Stensheim, H., Møller, B., van Dijk, T., and Fosså, S.D. (2009). Cause-Specific Survival for Women Diagnosed With Cancer During Pregnancy or Lactation: A Registry-Based Cohort Study. J. Clin. Oncol. *27*, 45–51.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400–404.

Strange, R., Li, F., Saurer, S., Burkhardt, A., and Friis, R.R. (1992). Apoptotic cell death and tissue remodelling during mouse mammary gland involution. Development *115*, 49–58.

Sun, X., Nichols, H.B., Tse, C.-K., Bell, M.B., Robinson, W.R., Sherman, M.E., Olshan, A.F.,

and Troester, M.A. (2016). Association of Parity and Time since Last Birth with Breast Cancer Prognosis by Intrinsic Subtype. Cancer Epidemiol. Biomarkers Prev. *25*, 60–67.

Tan, W., Zhang, W., Strasner, A., Grivennikov, S., Cheng, J.Q., Hoffman, R.M., and Karin, M. (2011). Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. Nature *470*, 548–553.

Taylor, D., Pearce, C.L., Hovanessian-Larsen, L., Downey, S., Spicer, D. V., Bartow, S., Pike, M.C., Wu, A.H., and Hawes, D. (2009). Progesterone and estrogen receptors in pregnant and premenopausal non-pregnant normal human breast. Breast Cancer Res. Treat. *118*, 161–168.

TCGA (2012). Comprehensive molecular portraits of human breast tumours. Nature *487*, 61–70.

Thomas, R.J., Guise, T.A., Yin, J.J., Elliott, J., Horwood, N.J., Martin, T.J., and Gillespie, M.T. (1999). Breast Cancer Cells Interact with Osteoblasts to Support Osteoclast Formation. Endocrinology *140*, 4451–4458.

Tischler, G., and Leonard, S. (2014). biobambam: tools for read pair collation based algorithms on BAM files. Source Code Biol. Med. *9*, 13.

Titus-Ernstoff, L., Perez, K., Cramer, D.W., Harlow, B.L., Baron, J.A., and Greenberg, E.R. (2001). Menstrual and reproductive factors in relation to ovarian cancer risk. Br. J. Cancer *84*, 714–721.

Toesca, A., Gentilini, O., Peccatori, F., Azim, H.A., Amant, F., and Amant, F. (2014). Locoregional treatment of breast cancer during pregnancy. Gynecol. Surg. *11*, 279–284.

Trichopoulos, D., Hsieh, C.C., MacMahon, B., Lin, T.M., Lowe, C.R., Mirra, a P., Ravnihar, B., Salber, E.J., Valaoras, V.G., and Yuasa, S. (1983). Age at any birth and breast cancer risk. Int J Cancer *31*, 701–704.

Tsubaki, M., Komai, M., Fujimoto, S., Itoh, T., Imano, M., Sakamoto, K., Shimaoka, H., Takeda, T., Ogawa, N., Mashimo, K., et al. (2013). Activation of NF-κB by the RANKL/RANK system up-regulates snail and twist expressions and induces epithelial-to-mesenchymal transition in mammary tumor cell lines. J. Exp. Clin. Cancer Res. *32*, 62.

La Vecchia, C., Negri, E., Franceschi, S., and Parazzini, F. (1993). Long-term impact of reproductive factors on cancer risk. Int. J. Cancer *53*, 215–219.

Wada, T., Nakashima, T., Hiroshi, N., and Penninger, J.M. (2006). RANKL–RANK signaling in osteoclastogenesis and bone disease. Trends Mol. Med. *12*, 17–25.

Wolf, E., Lin, C.Y., Eilers, M., and Levens, D.L. (2015). Taming of the beast: Shaping Myc-dependent amplification. Trends Cell Biol. *25*, 241–248.

Wood, C.E., Branstetter, D., Jacob, A.P., Cline, J.M., Register, T.C., Rohrbach, K., Huang, L.-Y., Borgerink, H., and Dougall, W.C. (2013). Progestin effects on cell proliferation pathways in the postmenopausal mammary gland. Breast Cancer Res. *15*, R62.

Yang, X.R., Chang-Claude, J., Goode, E.L., Couch, F.J., Nevanlinna, H., Milne, R.L., Gaudet, M., Schmidt, M.K., Broeks, A., Cox, A., et al. (2011). Associations of breast cancer risk factors with tumor subtypes: A pooled analysis from the breast cancer association consortium studies. J. Natl. Cancer Inst. *103*, 250–263.

Yoldi, G., Pellegrini, P., Trinidad, E.M., Cordero, A., Gomez-Miragaya, J., Serra-Musach, J., Dougall, W.C., Muñoz, P., Pujana, M.A., Planelles, L., et al. (2016). RANK signaling blockade reduces breast cancer recurrence by inducing tumor cell differentiation. Cancer Res. *76*, 5857–5869.

Zagouri, F., Sergentanis, T.N., Chrysikos, D., Dimitrakakis, C., Tsigginou, A., Zografos, C.G., Dimopoulos, M.-A., and Papadimitriou, C.A. (2013). Taxanes for breast cancer during pregnancy: a systematic review. Clin. Breast Cancer *13*, 16–23.

Zhang, D., Zhao, X., Cui, Y., Dong, X., Zhao, J., Meng, J., Jia, X., and Chi, J. (2012). A program

for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118 ; iso-2; iso-3. Fly (Austin). 321–329.

Zitvogel, L., Tesniere, A., and Kroemer, G. (2006). Cancer despite immunosurveillance: Immunoselection and immunosubversion. Nat. Rev. Immunol. *6*, 715–727.

ANNEX

Annex 1: Published version of research articles derived from this thesis as a lead author

Nguyen, B., Venet, D., Azim, H.A., Brown, D., Desmedt, C., Lambertini, M., Majjaj, S., Pruneri, G., Peccatori, F., Piccart, M., Rothé, F. and Sotiriou, C. (2018). Breast cancer diagnosed during pregnancy is associated with enrichment of non-silent mutations, mismatch repair deficiency signature and mucin mutations. Npj Breast Cancer *4*, 23.

Annex 2: Published version of research articles derived from this thesis as a co-author

Azim, H.A., **Nguyen, B.**, Brohée, S., Zoppoli, G., and Sotiriou, C. (2015). Genomic aberrations in young and elderly breast cancer patients. BMC Med. 13, 266.

Lambertini, M., Santoro, L., Del Mastro, L., **Nguyen, B.**, Livraghi, L., Ugolini, D., Peccatori, F.A., and Azim, H.A. (2016). Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat. Rev. 49, 65–76.

npj | Breast Cancer

www.nature.com/npjbcancer

ARTICLE OPEN

Breast cancer diagnosed during pregnancy is associated with enrichment of non-silent mutations, mismatch repair deficiency signature and mucin mutations

Bastien Nguyen[1], David Venet[1], Hatem A. Azim Jr [1,2], David Brown[1,3], Christine Desmedt[1], Matteo Lambertini[1], Samira Majjaj[1], Giancarlo Pruneri[4], Fedro Peccatori[5], Martine Piccart[1], Françoise Rothé[1] and Christos Sotiriou[1]

Breast cancer diagnosed during pregnancy (BCP) is a rare and highly challenging disease. To investigate the impact of pregnancy on the biology of breast cancer, we conducted a comparative analysis of a cohort of BCP patients and non-pregnant control patients by integrating gene expression, copy number alterations and whole genome sequencing data. We showed that BCP exhibit unique molecular characteristics including an enrichment of non-silent mutations, a higher frequency of mutations in mucin gene family and an enrichment of mismatch repair deficiency mutational signature. This provides important insights into the biology of BCP and suggests that these features may be implicated in promoting tumor progression during pregnancy. In addition, it provides an unprecedented resource for further understanding the biology of breast cancer in young women and how pregnancy could modulate tumor biology.

npj Breast Cancer (2018)4:23 ; doi:10.1038/s41523-018-0077-3

INTRODUCTION

Breast cancer is the most frequently diagnosed malignancy during pregnancy.[1] Its incidence is increasing given the rising trend of delayed childbearing.[2] Given its rarity, few dedicated studies were performed so far; hence, our understanding of these tumors remains poor. The clinical management of these patients follows standard guidelines with only minor adaptations according to gestational age, maternal wishes and fetal considerations.[2] Therefore, the molecular characterization of BCP goes beyond academic curiosity as it is of utmost clinical interest to determine if these patients should be treated similarly to non-pregnant breast cancer patients. In this report, we aimed to identify specific molecular alterations characterizing BCP by combining whole genome sequencing, copy number alteration and gene expression data.

RESULTS

A total of 167 patients with primary breast cancer were retrospectively included in this study, 54 of whom were diagnosed during pregnancy. Detailed patient characteristics were previously published.[3] At a median follow-up of 9 years, median disease-free survival (DFS) time of BCP was 9.8 years vs. 12.5 years in controls ($P = 0.041$, log rank test, Supplementary Fig. S1a). Observed 5-year overall survival (OS) rate was 95.5% vs. 85.1% in BCP and control, respectively; median OS time was not reached within the time frame of the study (Supplementary Fig. S1b). In a multivariable Cox proportional hazards regression of DFS and OS, adjusted for age at diagnosis, date of diagnosis, pathological stage and molecular subtypes by IHC, we found that BCP was associated with worse DFS (multivariable hazard ratio [mHR] 1.81; 95% CI 1.09–3.01, $P = 0.024$) and OS (mHR 2.53; 95% CI 1.20–5.36, $P = 0.017$) (detailed survival data is provided in Supplementary Table S1).

BCP and controls have similar somatic copy number alteration profiles

We first sought to investigate whether tumors from BCP patients show distinct copy number alterations (CNAs) compared to tumors from matched non-pregnant breast cancer patients (controls). Hence, we performed genome-wide copy number alterations profiling on 160 formalin-fixed paraffin-embedded (FFPE) primary tumor samples from 52 BCP patients and 108 controls. Of note, gene expression data were available for all patients as previously described.[4] After quality control, CNA profiles were obtained for 125 tumor samples (78%) from 38 BCP and 87 controls. The main reason for exclusion was low cancer cell fraction (CCF < 30%) as estimated with the Genome Alteration Print algorithm[5] (Supplementary Fig. S2). No differences in clinicopathological features were observed between BCP and controls (Supplementary Table S2). We found no significant differences between BCP and controls in terms of cancer cell fraction, ploidy, and fraction of genome altered (Fig. 1a-c). Moreover, no significant differences were observed between the CNA profiles of the two groups neither at the segment nor at the chromosome arm levels, including the gains of 1q and 8q and loss

[1]Breast Cancer Translational Research Laboratory J.-C. Heuson, Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium; [2]Department of Internal Medicine, American University of Beirut (AUB), Beirut, Lebanon; [3]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, USA; [4]Department of Pathology, IRCCS Istituto Nazionale Tumori, Milan and University of Milan, School of Medicine, Milan, Italy and [5]European Institute of Oncology, Milan, Italy
Correspondence: Christos Sotiriou (christos.sotiriou@bordet.be)
These authors contributed equally: Bastien Nguyen, David Venet, Hatem A. Azim.
These authors jointly supervised this work: Françoise Rothé, Christos Sotiriou.

Received: 23 March 2018 Revised: 4 July 2018 Accepted: 9 July 2018
Published online: 06 August 2018

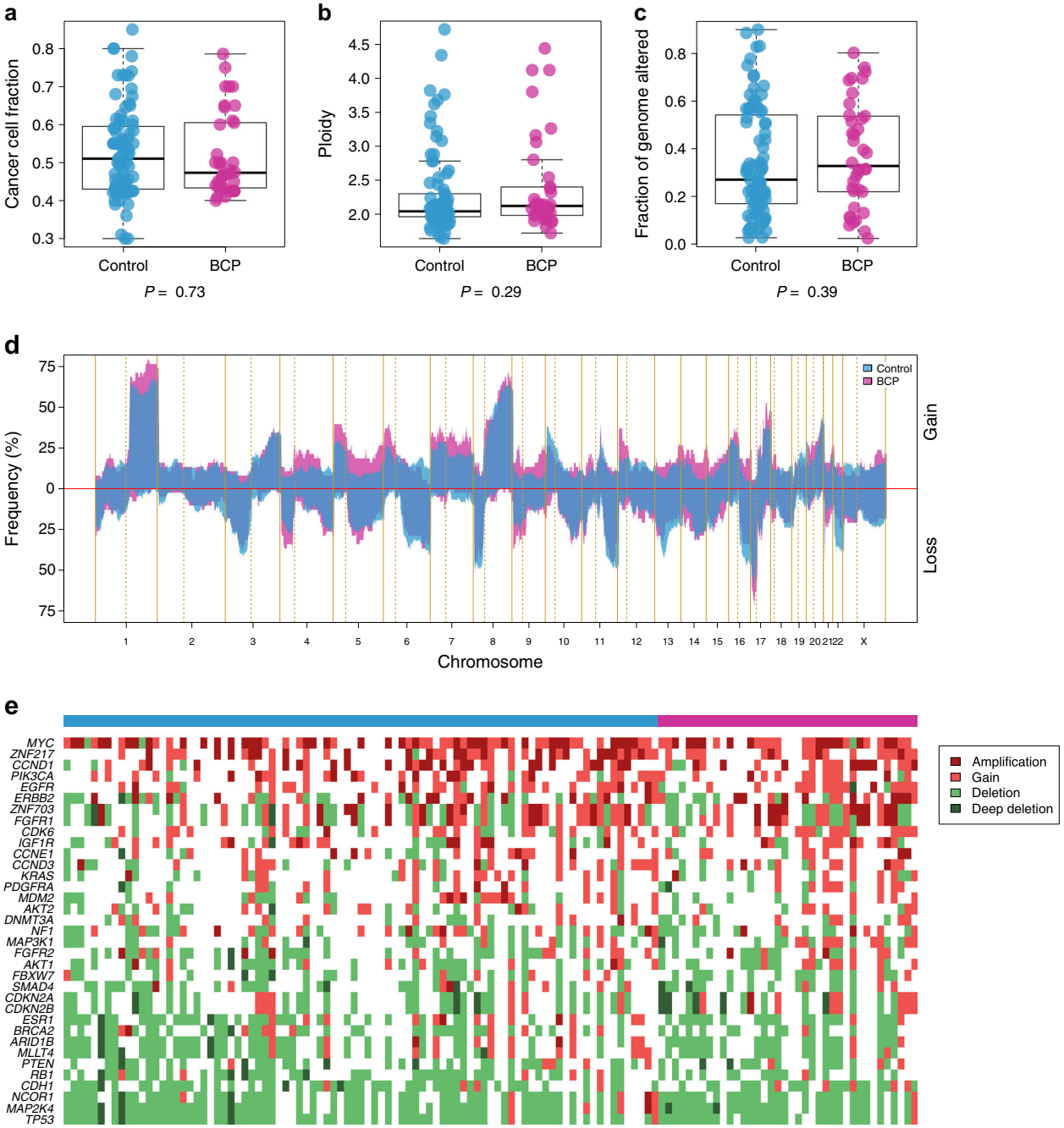


Fig. 1 Summary of the genome-wide copy number analysis of 87 controls and 38 BCP tumor samples. **a**–**c** Comparison of cancer cell fraction, ploidy and fraction of genome altered between controls and BCP. **d** Comparison of the CNA frequencies of controls (blue) and BCP (pink). **e** Heatmap of 35 CNA breast cancer driver genes according to their alterations; controls (blue) and BCP (pink). *P*, *p*-value derived for the Mann–Whitney *U* test

of 8p, reported to frequently occur in breast cancer[6] (Fig. 1d). We also compared CNAs profiles by intrinsic subtypes as defined by PAM50 and found no significant differences (Supplementary Fig. S3 and Supplementary Table S3).

We next focused our analysis on the 35 genes that were previously identified as CNA drivers in breast cancer.[7] As expected, *MYC* oncogene was the most frequently gained/amplified whereas *TP53* tumor suppressor gene was the most frequently lost/deleted across the whole cohort (Fig. 1e). Using GISTIC2.0,[8] we identified 22 focal amplifications and 23 focal deletions and found no differences between their prevalence in the two groups (Supplementary Fig. S4). Taken together, these results suggest that the CNA profiles of BCP and controls are similar.

BCP shows a higher number of non-silent mutations

To identify potential genomic differences between BCP patients and controls, we performed whole genome sequencing (WGS) on paired DNA samples extracted from FFPE blocks (i.e., primary tumors and histologically normal axillary lymph nodes) in a subset of 53 breast cancer patients from our initial series, 35 of whom were BCP (Supplementary Fig. S2 and Supplementary Table S2). We achieved 32X and 19X median haploid genome coverage for

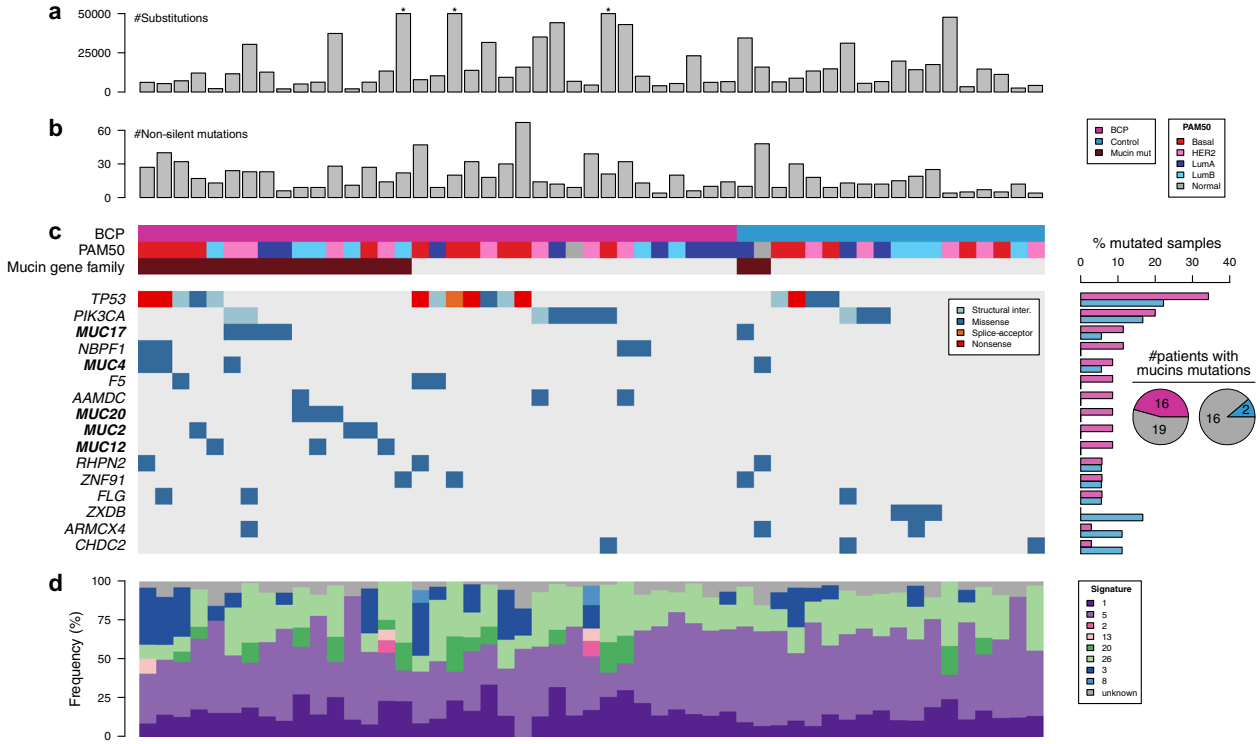


Fig. 2 Mutational landscape of individual BCP and controls. **a** Bar chart representing the absolute number of substitutions in BCP and controls, y-axis limited to 50,000 indicated by (*). **b** Bar chart representing the absolute number of non-silent mutations in BCP and controls (median: 20 vs. 12, $P = 0.027$, respectively). **c** Co-mutation plot showing genes harboring at least one non-silent mutation with a frequency of at least 5% across the whole cohort, and their corresponding frequencies in BCP and controls (right). **d** Proportion of breast cancer substitution signatures in each sample. Signatures are colored according to broad biological groups: 1 and 5 are associated with clock-like processes, 2 and 13 are APOBEC-related, 20 and 26 are associated with mismatch-repair deficiency, 3 and 8 are associated with homologous-recombination deficiency

tumor and normal samples respectively, which is similar in range to previous studies[7] (Supplementary Fig. S5). We detected a median of 13,829 and 10,084 single nucleotide variants (SNVs) and a median of 21 and 26 small insertions and deletions (Indels) in BCP and controls, respectively, and found no difference between the two groups (Fig. 2a and Supplementary Fig. S6a-c). Moreover, there was no difference in structural variations (insertions, deletions, duplications) nor tumor heterogeneity as assessed by the MATH score[9] (Supplementary Fig. S6d-f).

We identified a median of 14 non-silent mutations per tumor which is comparable to another large-scale breast cancer cohort study[7] (Supplementary Table S4). Interestingly, BCP had a significantly higher number of non-silent mutations than controls (median: 20 vs. 12, $P = 0.027$, Fig. 2b and Supplementary Fig. S6g-h). This observation remained consistent after correcting for potential confounding factors including age at diagnosis, date of diagnosis, pathological stage and molecular subtypes by IHC ($P = 0.019$, Fig. S6g). Compared to controls, BCP had also a significantly higher number of mutations previously reported in breast cancer in the Catalog of Somatic Mutations in Cancer (COSMIC) database[10] ($P = 0.018$, Supplementary Fig. S6i). At the gene level, we identified 17 genes harboring at least one non-silent mutation with a frequency of at least 5% across all patients. Of those, *TP53* and *PIK3CA* were the most frequently mutated genes without any significant difference between the two groups (Fig. 2c).

BCP is associated with a higher frequency of mutations in mucin gene family

MUC17 was the third most mutated gene and four other mucin gene family members namely *MUC2*, *MUC4*, *MUC12*, and *MUC20*, were among the most frequently mutated genes in BCP (Fig. 2c). Within the mucin gene family, we identified 20 missense mutations and one nonsense mutation in BCP compared to only two missense mutations in controls. Among these 20 mucin variants, 10 were present in the COSMIC database,[10] which was higher than expected by chance ($P = 0.006$, Monte-Carlo test, Supplementary Table S4). Altogether, we found a significantly higher number of BCP with non-silent mutations in the mucin gene family compared to controls (45.7 vs. 11.1% respectively, $P = 0.015$, Fig. 2c). This observation remained consistent after correcting for classical clinicopathological features ($P = 0.008$). Similar findings were observed by comparing BCP with 56 matched controls taken from the TCGA dataset (45.7 vs. 23.1% respectively, $P = 0.034$). Acknowledging that some mucins (*MUC4, MUC16*) are known to give rise to false positive calls due to technical artifacts,[11] we removed these two genes and confirmed the above-mentioned results (37.1 vs. 5.5%, $P = 0.020$ and 37.1 vs. 14.3%, $P = 0.020$, using controls and TCGA controls, respectively).

We did not find any differences in clinicopathological features or survival according to mucin mutational status (Supplementary Table S5 and Supplementary Fig. S7). There were three hotspots mutations (i.e., present in two distinct patients) two in *MUC17* and one in *MUC20*, and five missense mutations were clustered within 260 base pairs of *MUC2* (Fig. 3a). None of these mutations were in annotated protein domains. Since the glycosylation of mucins is known to play a major role in producing a chemical barrier at the epithelium of tubular organs for protection and lubrication, we interrogated whether these mutations could affect glycosylation acceptor sites. The mucin O-glycosylation is characterized by the addition of N-acetylgalactosamine (GalNAc) to the hydroxyl group of serine or threonine residues.[12] Remarkably, 40.9% of missense mutations affecting mucins resulted in an amino acid change to a

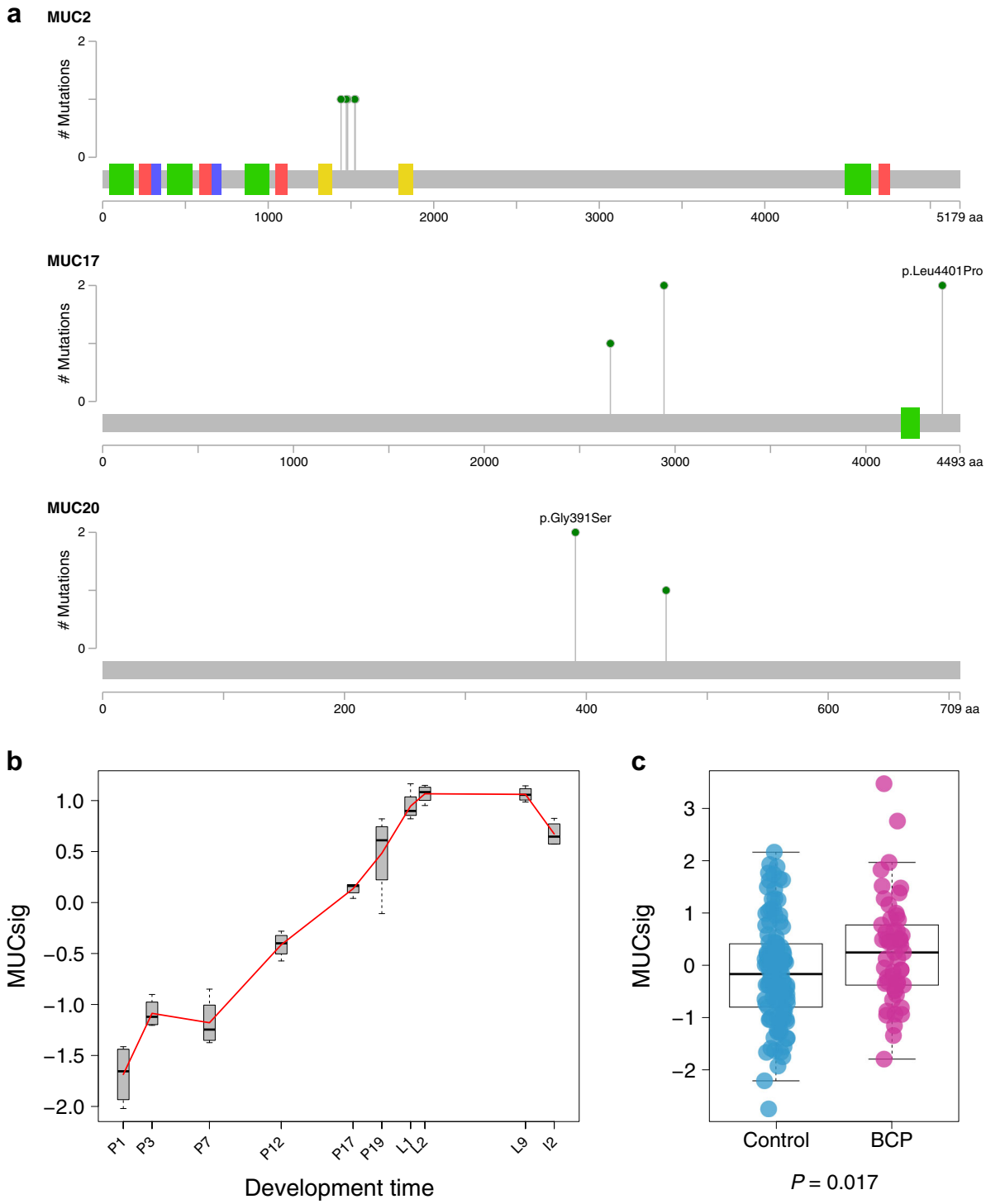


Fig. 3 Enrichment of mucin mutations and upregulation in BCP. **a** Lollipop plots were generated using cBioPortal Mutation Mapper. Each lollipop denotes a unique missense mutation for *MUC2*, *MUC17*, and *MUC20* in BCP. **b** MUCsig according to normal adult mouse mammary development (from pregnancy day 1 to involution day 2). **c** Comparison of MUCsig between controls and BCP. *P*, *p*-value derived for the Mann–Whitney *U* test

serine residue, which was significantly higher than expected by chance ($P = 0.0002$, Monte-Carlo test), suggesting mucin hyperglycosylation in BCP. We also found that the frequency of missense mutations resulting in a gain of serine site in mucins in the TCGA dataset was significantly lower compared with BCP (6.3% in TCGA vs. 40.9% in BCP, $P < 0.001$). Since mucins expression is known to increase throughout gestation in mice,[13] we expected that mucins were also upregulated in BCP. We therefore derived a metagene signature comprising all members of the mucin gene family (called "MUCsig") from the corresponding gene expression data and found higher expression of MUCsig in BCP than in controls ($P = 0.017$, Fig. 3b-c). Altogether, these results show that BCP is associated with an increased expression of mucins as well as a higher frequency of mutations in mucin

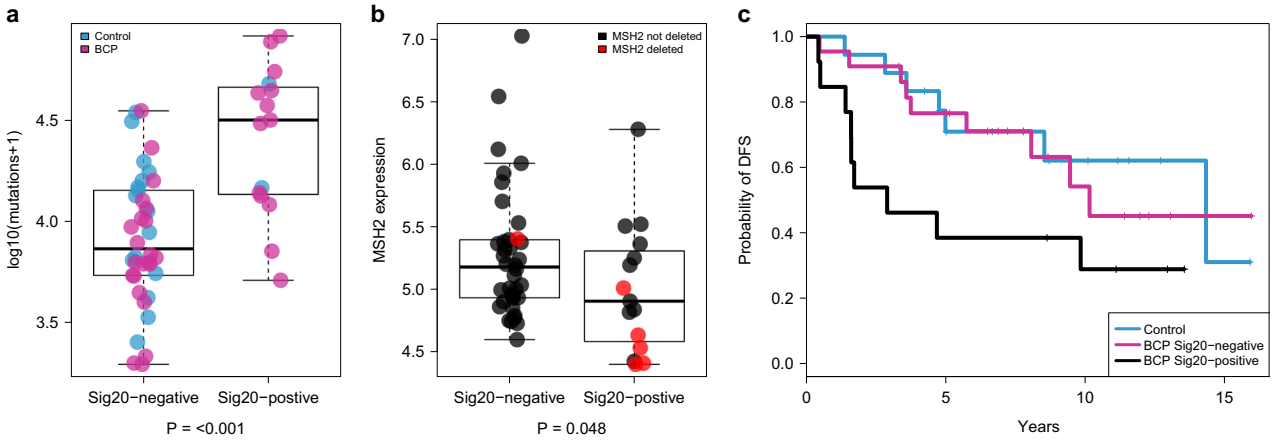


Fig. 4 Association of signature 20 with mutational load and clinical outcome. **a** Comparison of SNV mutational load between Sig20 negative and Sig20 positive tumors. **b** Comparison of MSH2 expression between Sig20 negative and Sig20 positive tumors. **c** Kaplan–Meier plot showing the difference in DFS between control patients ($N = 18$), BCP patients with Sig20 negative tumors ($N = 22$) and BCP patients with Sig20 positive tumors ($N = 13$). *P*, *p*-value derived for the Mann–Whitney *U* test

gene family that may potentially lead to mucin hyperglycosylation.

BCP is enriched in mutational signature related to mismatch repair deficiency

To have a better understanding of the etiology of BCP, we interrogated the contribution of base-substitution signatures known to occur in breast cancer.[7] When evaluating the proportion of each signature present in each sample, we found that signature 1 was more prevalent in BCP compared to controls whereas signature 5 was more prevalent in controls ($P = 0.013$, FDR = 0.053 and $P = 0.01$, FDR = 0.053, respectively, Fig. 2d). These results remained consistent after controlling for clinicopathological features ($P = 0.002$, FDR = 0.014 and $P = 0.004$, FDR = 0.016, respectively). When evaluating the presence or absence of mutational signatures we found that signature 20 (Sig20) was found in 13 out of 35 BCP (37.1%), as compared to only 2 out of 18 controls (11.1%) ($P = 0.059$, FDR = 0.410, Fig. 2d). When controlling for clinicopathological features, this observation was significant ($P = 0.004$, FDR = 0.029). Signature 1 is known to be associated with age at diagnosis while the etiology of signature 5 is still unclear. Sig20, previously found in stomach and breast cancers, is related to DNA mismatch repair (MMR) deficiency.[14] Of interest, this signature remained significantly enriched in BCP when increasing the number of controls with 64 matched cases derived from the BRCA560 dataset (37.1 vs. 3.1%, $P < 0.001$). No classical clinicopathological features were associated with BCP Sig20-positive tumors except progesterone receptor negative status (Supplementary Table S6). We found that Sig20 frequency was strongly correlated with SNV mutational load ($\rho = 0.56$, $P < 0.001$, Supplementary Fig. S8a) with Sig20-positive tumors harboring a median of 31,632 SNVs, as compared to 7,352 SNVs in Sig20-negative tumors ($P < 0.001$, Fig. 4a). Next, we interrogated if Sig20 could be caused by alteration of genes involved in the MMR machinery either at the expression or copy number levels. The first step of MMR is the recognition of replication errors mediated by MutS homolog complexes; *MSH2* and *MSH6*.[15] We found a significantly lower expression of *MSH2* in patients harboring Sig20 ($P = 0.047$, Fig. 4b) corroborated by a negative correlation between *MSH2* expression and Sig20 frequency ($\rho = -0.27$, $P = 0.024$, Supplementary Fig. S8b). This could be partially caused by CNA in *MSH2* since 5 out of 15 Sig20-positive versus 1 out of 38 Sig20-negative tumors harbored *MSH2* deletions (33.3 vs. 2.6%, $P = 0.01$). Finally, we interrogated the impact of Sig20 on survival and found that BCP Sig20-positive patients had a shorter DFS than BCP Sig20-negative patients (median DFS time of 2.9 years vs. 10.2 years respectively $P = 0.091$, log rank test, Fig. 4c). In Sig20-positive patients the median OS was 6.72 years while the median OS was not reached in BCP Sig20-negative patients (P = 0.009, log rank test, Supplementary Fig. S9). This was not significant in a multivariate model (DFS mHR 1.06; 95% CI 0.21–4.27, $P = 0.31$; OS mHR 0.8; 95% CI 0.12–5.07, $P = 0.81$, respectively). Overall, these results suggest that some BCP patients show a defective MMR due to copy number loss of *MSH2*.

DISCUSSION

This study reveals important molecular differences characterizing BCP that may potentially represent a biologic explanation for their rather aggressive clinical behavior. First, BCP was enriched in non-silent mutations that could have potential oncogenic functiond. Second, 45% of BCP harbored a mutation in mucin gene family in addition to an upregulation of mucins at the mRNA level. Like in mice,[13] this could be due to physiological change induced by pregnancy to prepare the breast for lactation. Our hypothesis is that some preexisting subclones carrying mucin mutations could have a growth advantage under pregnancy state. Another argument in favor of this hypothesis is the fact that most mucin mutations resulted in an amino acid change to a serine residue and that some of them are in hotspot regions. It has been previously found that in breast cancer, alterations in mucin expression or glycosylation influence tumor growth, adhesion, invasion, and immune surveillance.[16,17] The impact of missense mutations resulting in an amino acid change to a serine residue on the glycosylation status of mucins is unknown, but it is tempting to speculate that these alterations could influence their function, stability and secretion. More investigations are required to determine the exact effect of mucin mutations in BCP and in breast cancer in general, but these alterations could play a role in BCP biology.

Moreover, BCP showed a higher prevalence of signatures 1 and 20 and a lower prevalence of signature 5. The etiology of signature 5 is not well understood.[18] The high prevalence of signature 1 cannot be explained by a difference in age at diagnosis or age of the blocks since similar results were found in a multivariate analysis after adjusting for both variables. 37.1% of BCP were associated with signature 20 (Sig20), attributable to DNA mismatch repair deficiency. This is surprising given the low frequency (1–2%) of MMR deficiency recently reported in breast cancer.[19] Mechanistically, this could be explained in part by the deletion of *MSH2*, a key gene involved in MMR. Survival analysis

showed that BCP Sig20-positive patients had the worst prognosis whereas BCP Sig20-negative patients had DFS comparable to controls. MMR deficiency and high mutational burden have been shown to predict clinical benefit to immune checkpoint blockade in colorectal and other types of highly immunogenic cancers.[20,21] To date, the role of checkpoint inhibitors in the treatment of breast cancer is under intensive investigation and the results are still awaited.[22] While the feasibility of investigating new agents in such peculiar disease is rather complex, these results could potentially open the door to identify high-risk BCP patients who could benefit from immunotherapy.

A potential limitation of our study is that we used archived FFPE samples that are known to be challenging for WGS due to DNA degradation and induction of artefacts. Indeed, the higher proportion of signature 1 and 5 observed in our study could be due to C > T artefacts induced by formalin fixation. Nonetheless, BCP and controls were processed in the same way with no difference in the age of the blocks and the sequencing coverage reached in normal and in tumor tissues was comparable to other studies.[7] Another limitation of our study is the lack of epigenetic profiling analysis. As it is known that pregnancy induce epigenetic changes in epithelial cells to support mammary development,[23] we can hypothesize that these modifications could impact breast cancer biology. Therefore, the study of such modifications in BCP is worthy further investigation. In conclusion, we believe that our work provides important insights into the biology of BCP and a unique resource to study the biology of breast cancer in young women and how pregnancy could modulate tumor biology.

METHODS

Patients and samples

A total of 167 patients with primary breast cancer were retrospectively included in this study, 54 of whom were diagnosed during pregnancy. All patients were diagnosed and followed up at the European Institute of Oncology (IEO, Milan, Italy) from 1996 to 2010. As previously described,[3] this is a case-control study, in which pregnant breast cancer patients and controls were matched according to age, tumor size, nodal status, and date of diagnosis. For the current genomic analysis, we opted to exclude patients who received neoadjuvant therapy to avoid potential impact of treatment on the obtained results. The majority were treated with anthracycline-based regimen (individual patients data are presented in Supplementary Table S1). All patients had available FFPE tissue from the primary tumor resection and there was only one tumor sample per patient. All control patients were pre-menopausal at time of diagnosis. ER/PR-status were defined by ASCO-CAP. For the classification of Luminal A and B we used a cut-off of Ki67 > 20% according to the St Gallen 2015 Consensus Meetings.[24] Matched normal tissues were collected from histologically confirmed tumor-free axillary lymph nodes or tumor-adjacent normal tissue and there was only one normal sample per patient. FFPE tissue sections were deparaffinized by xylene followed by a 100% ethanol wash. DNA extraction was performed using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations. The quantity of double-stranded DNA was evaluated using the Qubit dsDNA BR Assay Kit. For the WGS, we selected 18 control patients based on major clinicopathological features of the 35 BCP patients, namely age at diagnosis, ER status, and grade. All patients provided written informed consent for the use of tissue samples for research purposes as per the IEO institutional policies. This study was approved by the Ethics Committee of Institut Jules Bordet (Number 1782). The validation of the enrichment of mutations in the mucin family genes in BCP were done by comparing the frequency of these mutations in BCP patients with putatively non-pregnant patients retrieved from the TCGA dataset[25] and selected to have similar age, estrogen receptor (ER) and progesterone receptor (PR) distribution ($N = 56$) (Supplementary Table S7). The validation of the enrichment of signature 20 in BCP were done by comparing the frequency of this signature in BCP patients with putatively non-pregnant patients retrieved from the 560 breast cancer dataset[7] (referred to as BRCA560) and selected to have similar age, ER and PR distribution ($N = 64$).

Transcriptomic profiling

All samples were hybridized on Affymetrix Human Genome U219 array plates following the manufacturer's protocol, as described before.[4] The metagene signature MUCsig was calculated by taking the mean expression level of all genes present in the mucin family, scaled to a standard deviation of one and centered around zero. The publicly available murine data set derived from normal breast of pregnant mice (GEO ID: GSE8191[13]) was used to evaluate mucin expression in the normal breast during pregnancy. Ensembl database was used to convert mouse gene names to the human equivalent.

Genome-wide copy number analysis

Hematoxylin and eosin slides from the archived FFPE blocks were reviewed by a pathologist (G.P.) to confirm diagnosis and evaluate tumor content. Samples with tumor purity below 60% were macrodissected ($N = 56$). DNA was extracted as described above. A total of 80 ng of DNA was used for copy number profiling using the Affymetrix OncoScan® FFPE Assay Kit according to the manufacturer's instructions. The raw intensity values from the scanned chips were normalized to obtain Log2 ratios, B allele frequencies and genotyping calls (AA/AB/BB) using Affymetrix Power Tools. We used release NA.33 of the NetAffx library for the reference model and annotation. We computed the median absolute pairwise deviation and the median auto-correlation from the normalized log2 ratios as quality control metrics and used a threshold of 0.30 and 0.5, respectively, to flag failed arrays. Further details are provided in the Supplementary Methods.

Library preparation and whole genome sequencing

For each of 53 patients, two samples of 1µg genomic DNA from tumor and histologically normal axillary lymph nodes were whole genome sequenced at The McDonnell Genome Institute at Washington University (St Louis MO, USA) on an Illumina HiSeqX platform. Briefly, manual dual indexed libraries were constructed with 1µg of FFPE genomic DNA for the 53 tumor/normal pairs using the Accel-NGS 2S Plus Library Kit (Swift, MI, USA). Samples were fragmented on the Covaris LE220 instrument with 350bp target insert size. PCR cycle optimization was performed to prevent over-amplification of the libraries. The concentration of each library was determined through qPCR (Kapa Biosystems, MA, USA). For the normal samples, each library was loaded on one lane of a HiSeqX flow cell, whereas for tumor samples, each library was loaded across two lanes of a HiSeqX flow cell. 2×150 paired-end sequence data were generated at a target depth of 30×(normal) and 60 (tumor) haploid genome coverage. All sequencing data are available in EGA under accession "EGAS00001002685". Further details are provided in the Supplementary Methods.

Tumor heterogeneity

To quantify the level of intra-tumor heterogeneity present in a sample, we used the MATH score as previously described[9];

$$MATH = \frac{\mathrm{MAD}(VAFs)}{\mathrm{median}(VAFs)}$$

where MAD(VAFs) is the median absolute deviation of the variant allele fractions (VAFs) of all the mutations (coding and noncoding) in a tumor.

Mutational signature

All samples were analyzed using deconstructSigs[26] to extract signatures based on the Wellcome Trust Sanger Institute Mutational Signature Framework.

Statistical analysis and survival analysis

Except for age and date of diagnosis that were considered as continuous variables and therefore compared using the non-parametric Mann–Whitney U test, differences in other clinicopathological characteristics between BCP and controls were analyzed using the χ^2 test or the Fisher exact test when appropriate. All statistical tests comparing BCP and controls were done using the non-parametric Mann–Whitney U test and the Fisher exact test for continuous and categorical variables, respectively. Independent association between continuous and binary variables with BCP vs. controls was investigated using linear and logistic regressions, respectively. All multivariate tests were adjusted for age at diagnosis, date of diagnosis, pathological stage, and molecular subtypes by IHC. All

interaction and multivariate tests were done using analysis of variance to compare the models with and without the extra term.

All correlations were measured using the non-parametric Spearman's *rho* coefficient. Reported *P*-values were two-tailed, and differences were considered significant when the *P*-value was less than 0.05. When applicable, multiple testing correction was done using the false discovery rate method (FDR),[27] FDR below 0.05 being considered significant. All analyses were done in R software version 3.3.2 (available at www.r-project.org) and Bioconductor version 3.4.

Survival endpoint was DFS and calculated from the date of surgery to any loco-regional or distant recurrence, contralateral BC, other primary tumor or death from any cause, whichever occurred first. In the absence of any of the above-mentioned events, survival was censored at the last follow-up visit or phone call with the patient. Survival curves were estimated using the Kaplan–Meier method and compared by the log-rank test. The prognostic impact of pregnancy on survival was evaluated using univariate and multivariate Cox proportional hazards regression models and expressed as hazard ratio (HR) with 95% CI. Multivariate analysis was adjusted for standard clinical prognostic factors (age at diagnosis, date of diagnosis, pathological stage, and molecular subtypes by IHC). Further details are provided in the Supplementary Methods.

Data availability

Raw gene expression data, together with patients' characteristics, are publicly available on GEO http://www.ncbi.nlm.nih.gov/geo/, under accession number GSE53031. Sequencing data have been deposited at the European Genome-Phenome Archive (http://www.ebi.ac.uk/ega/), under accession number EGAS00001002685.

ACKNOWLEDGEMENTS

We thank the patients who contributed to this study by providing tissue samples and acknowledge the clinical staff at European Institute of Oncology for their dedication. We acknowledge whole genome sequencing data generation by the production sequencing group at the McDonnell Genome Institute at Washington University in St. Louis, and the oversight of this work by Elaine Mardis.

AUTHOR CONTRIBUTIONS

Conceptualization, B.N., H.A., and C.S.; Methodology, B.N., D.V., and D.B.; Formal Analysis, B.N., D.V., and D.B.; Investigation, B.N. and S.M.; Resources, G.P. and F.P.; Writing—Original Draft, B.N.; Writing—Review & Editing, B.N., D.V., F.R., C.D., M.L, D.B., H.A., M.P., and C.S., Funding Acquisition, F.R., H.A., M.P and C.S., Supervision, F.R., H.A. and C.S. All authors reviewed and approved the final manuscript. The work was supported by grants from the Breast Cancer Research Foundation (BCRF). C. Sotiriou is a ULB Professor funded by the Fonds National de la Recherche Scientifique (FNRS). M Lambertini is supported by an ESMO Translational Research Fellowship. B. Nguyen is supported by a Télévie fellowship.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Breast Cancer* website (https://doi.org/10.1038/s41523-018-0077-3).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Anderson, J. M. Mammary cancers and pregnancy. *Br. Med. J.* **1**, 1124–1127 (1979).
2. Loibl, S. et al. Breast cancer diagnosed during pregnancy. *JAMA Oncol.* **1**, 1145 (2015).
3. Azim, H. A. et al. The biological features and prognosis of breast cancer diagnosed during pregnancy: a case-control study. *Acta Oncol.* **51**, 653–661 (2012).
4. Azim, H. A. et al. Biology of breast cancer during pregnancy using genomic profiling. *Endocr. Relat. Cancer* **21**, 545–554 (2014).
5. Popova, T. et al. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol.* **10**, R128 (2009).
6. Mika, T. et al. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosom. Cancer* **21**, 177–184 (1998).
7. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 1–20 (2016).
8. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
9. Mroz, E. A. & Rocco, J. W. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral. Oncol.* **49**, 211–215 (2013).
10. Forbes, S. A. et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
11. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
12. Hanisch, F.-G. O-Glycosylation of the mucin type. *Biol. Chem.* **382**, 143–149 (2001).
13. Anderson, S. M., Rudolph, M. C., McManaman, J. L. & Neville, M. C. Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! *Breast Cancer Res.* **9**, 204 (2007).
14. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
15. Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7**, 335–346 (2006).
16. Mukhopadhyay, P. et al. Mucins in the pathogenesis of breast cancer: implications in diagnosis, prognosis and therapy. *Biochim. Biophys. Acta* **1815**, 224–240 (2011).
17. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60 (2004).
18. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
19. Davies, H. *et al.* Whole-genome sequencing reveals breast cancers with mismatch repair deficiency. *Cancer Res.* https://doi.org/10.1158/0008-5472.CAN-17-1083 (2017).
20. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
21. Rizvi, N. A. et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
22. Emens, L. A. Breast cancer immunotherapy: facts and hopes. *Clin. Cancer Res.* **3001**, 2017, https://doi.org/10.1158/1078-0432.CCR-16-3001 (2017).
23. Huh, S. J. et al. Age and pregnancy-associated dna methylation changes in mammary epithelial cells. *Stem Cell Rep.* **4**, 297–311 (2015).
24. Coates, A. S. et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann. Oncol.* **26**, 1533–1546 (2015).
25. TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* **487**, 61–70 (2012).
26. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
27. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

Azim et al. BMC Medicine (2015) 13:266
DOI 10.1186/s12916-015-0504-3

Spotlight on breast cancer

BMC Medicine

RESEARCH ARTICLE

Open Access

Genomic aberrations in young and elderly breast cancer patients

Hatem A. Azim Jr[1*], Bastien Nguyen[1], Sylvain Brohée[1], Gabriele Zoppoli[2] and Christos Sotiriou[1]

Abstract

Background: Age at breast cancer diagnosis is a known prognostic factor. Previously, several groups including ours have shown that young age at diagnosis is associated with higher prevalence of basal-like tumors and aggressive tumor phenotypes. Yet the impact of age at diagnosis on the genomic landscape of breast cancer remains unclear. In this study, we examined the pattern of somatic mutations, chromosomal copy number variations (CNVs) and transcriptomic profiles in young and elderly breast cancer patients.

Methods: Analyses were performed on The Cancer Genome Atlas (TCGA) dataset. Patients with metastatic disease at diagnosis, classified as normal-like by PAM50 or had missing clinical information were excluded. Young patients were defined as ≤45 years of age, while elderly patients were those ≥70 years of age at breast cancer diagnosis. The remaining patients were classified as "intermediate". We evaluated the association between age at diagnosis and somatic mutations, CNV and gene expression in a logistic regression model adjusting for tumor size, nodal status, histology and breast cancer subtype. All analyses were corrected for multiple testing using the Benjamini–Hochberg approach.

Results: In this study, 125, 486 and 169 patients were ≤45, 46–69 and ≥70 years of age, respectively. Older patients had more somatic mutations (n = 44 versus 35 versus 31; $P = 0.0009$) and more CNVs, especially in ductal tumors ($P = 0.02$). Eleven mutations were independently associated with age at diagnosis, of which only GATA3 was associated with young age (15.2 % versus 8.2 % versus 9 %; $P = 0.003$). Only two CNV events were independently associated with age, with more chr18p losses in older patients and more chr6q27 deletions in younger ones. Younger age at diagnosis was associated with higher expression of gene signatures related to proliferation, stem cell features and endocrine resistance.

Conclusions: Age adds a layer of biological complexity beyond breast cancer molecular subtypes, classic pathological and clinical variables, worthy of further consideration in future drug development as we seek to refine therapeutic strategies in the era of personalized medicine.

Keywords: Age, Breast cancer biology, Breast cancer in the elderly, Breast cancer in young patients, GATA3, Gene expression, Mutations

Background

Young age at breast cancer diagnosis is a known poor prognostic factor [1, 2]. Previous studies have indicated higher prevalence of poorly differentiated, estrogen receptor (ER)-negative and human epidermal growth factor receptor 2 (HER2)-positive tumors in women diagnosed at a young age [3, 4]. Further genomic characterization has revealed enrichment with basal-like tumors [5, 6]. While these observations could well explain the poorer outcome of young breast cancer patients compared to their older counterparts, younger age remains an independent poor determinant of long-term outcome [5]. This underscores the need to further refine our understanding of the impact of age on cancer biology, which could have relevant implications on patient management.

On the other hand, few data are available with respect to the biological features of tumors arising in the elderly. Currently, around 30–35 % of breast cancer patients are over 70 years of age at the time of diagnosis and this is expected to increase in the coming years [7]. While these patients appear to develop relatively more

* Correspondence: hatem.azim@bordet.be
[1]Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Université Libre de Bruxelles, Boulevard de Waterloo, 121, Brussels, Belgium
Full list of author information is available at the end of the article

"indolent" tumors characterized by high endocrine receptor expression [8], the late onset of these tumors may also suggest accumulation of several genomic aberrations over time, due to the stochastic nature of DNA damage in eukaryotic cells during the replication process. Acknowledging that morbidities other than cancer itself often contribute to mortality of older patients [9], it is very important to refine our understanding of the biology of these tumors in an attempt to optimize their management.

Previously, our group and others have published on the differences at the transcriptomic level according to age at diagnosis, investigating selected genes or pathways [5, 6, 10]. However, we lack studies that evaluate the differences at the DNA level. In the current study,we investigated for the first time the differences in somatic mutations and copy number variations (CNVs) between young and older breast cancer patients. In addition, we evaluated the expression of thousands of relevant genomic signatures at the RNA level.

Methods

Eligible patients

All analyses were performed on The Cancer Genome Atlas (TCGA) publicly available dataset. Eligible patients were those with non-metastatic disease who had complete information on age at breast cancer diagnosis, tumor histology, tumor size and lymph node status. For each patient, we determined the breast cancer molecular subtype using PAM50 [11]. PAM50 classes were determined from the TCGA RNA-Seq gene expression data using the genefu package of the R/Bioconductor statistical package. Samples of patients classified as normal-like were excluded, as they often represent an artifact due to limited tumor cellularity and a large background of normal breast cells in the sample [12].

Young patients were defined as ≤45 years of age, while elderly patients were defined as those ≥70 years of age at breast cancer diagnosis. The remaining patients were classified as "intermediate". Since the TCGA dataset is publicly available, ethics committee approval was not needed. In addition, neither patient informed consent nor permission to use this data was required to perform this analysis.

Genomic analysis

We evaluated three parameters: 1) somatic mutations using exome sequencing; 2) somatic CNV; and 3) transcriptomic profiles. We downloaded the data from the TCGA online repository in February 2015.

In the current analysis, all somatic mutations were considered apart from those referred to as "silent" mutations. Somatic CNV was evaluated using array comparative genomic hybridization (CGH) data, available as pre-processed, publicly available information and not validated by any other methodology. Segmented data were used as input for Genomic Identification of Significant Targets in Cancer, version 2.0 (GISTIC 2.0) and version 6.2 on the Broad Institute GenePattern cloud server to obtain somatic focal and broad CNV events [13]. These were then parsed in R. For focal events, only "high-level" focal amplification events, defined as log2 ratio >0.9 were retained, whereas focal losses were retained with log2 ratio >0.3 and with a Q value <0.25. Broad events, defined as arm-level events encompassing 98 % or more of a chromosome arm, were computed using GISTIC as well.

For transcriptomic profiling, we used the RNA sequencing data to evaluate differences in transcriptomic profiles according to age. Data were downloaded from the TCGA online repository and RNA-Seq absolute expression values were log2 transformed before performing the analyses.

Statistical analyses

The association between age groups, that is, young (≤45 years), intermediate (46–49 years) and elderly patients (≥70 years), with clinicopathological characteristics was evaluated using Pearson's chi-squared test. The Kruskal–Wallis test was used to compare the number of mutations and CNVs according to age group. For mutations that were represented in at least 5 % in any age group, we evaluated their independent association with age at diagnosis (as a continuous variable) in a logistic regression model adjusting for tumor size (≤2 cm versus >2 cm), nodal status (negative versus positive), tumor histology (ductal versus lobular) and breast cancer subtype (luminal-A versus luminal-B versus HER2 versus basal). A similar model was used to evaluate the independent association between age, CNV and gene expression using the Molecular Signatures Database (MSigDB; PMID: 16199517). All analyses were corrected for multiple testing using the Benjamini–Hochberg approach [14].

Results

A total of 780 patients from the TCGA dataset where included, of whom 125, 486 and 169 were ≤45, 46–69 and ≥70 years of age, respectively. Transcriptomic data was available for all patients, while 722 (92.5 %) and 713 (91.4 %) had available somatic mutation and CNV data, respectively.

Table 1 summarizes the main characteristics of patients. As expected, young patients had less lobular cancer (7 % versus 24 % versus 29 %; P <0.001), fewer node-negative tumors (38 % versus 49 % versus 49 %; P = 0.05) and a trend of more basal-like tumors (20 % versus 18 % versus 14 %; P = 0.16).

Table 1 Main characteristics of patients

Characteristic	The Cancer Genome Atlas (N = 780)			
	≤45 years of age	46–69 years of age	≥70 years of age	*P* value
Number	125	486	169	
Tumor size				
≤2 cm	30 (24 %)	135 (28 %)	43 (26 %)	0.64
>2 cm	95 (76 %)	351 (72 %)	126 (74 %)	
Nodal status				
Negative	47 (38 %)	241 (49 %)	83 (49 %)	0.05
Positive	78 (62 %)	245 (51 %)	86 (51 %)	
Histology				
Ductal	116 (93 %)	371 (76 %)	121 (71 %)	<0.001
Lobular	9 (7 %)	95 (24 %)	48 (29 %)	
PAM50 subtype				
Luminal-A	44 (35 %)	200 (41 %)	70 (41 %)	0.16
Luminal-B	41 (33 %)	140 (29 %)	64 (38 %)	
HER2	15 (12 %)	57 (12 %)	12 (7 %)	
Basal	25 (20 %)	89 (18 %)	23 (14 %)	

Somatic mutations according to age

We found a significant association between age at diagnosis and the prevalence of somatic mutations. Median number of somatic mutations in the young group was 31, compared to 35 and 44 in the intermediate and older patient groups, respectively (*P* value = 0.0009). Figure 1 shows the four most prevalent somatic mutations in the different age groups. PIK3CA and TP53 were the most common somatic mutations, constituting around 50–60 % of all mutations across the different age groups. The striking difference between the three age groups was for GATA3, which was the third most common somatic mutation in young patients, constituting 15.2 %, while TTN mutation was the third most frequent mutation in the intermediate (15.1 %) and older patient groups (29 %).

To evaluate the independent effect of age on the prevalence of somatic mutations, we performed a logistic regression analysis adjusted for tumor size, nodal status, histology and breast cancer molecular subtype. We found 11 mutations to be independently associated with age at diagnosis (Table 2). All were associated with older age at diagnosis, except GATA3, which was independently associated with breast cancer arising in young women (15.2 % versus 8.2 % versus 9 %; *P* = 0.003, false discovery rate (FDR) = 0.033).

Somatic CNV events according to age

We evaluated the prevalence of CNV events according to age. We found a tendency of higher focal and broad CNV in older patients (mean = 15), compared to 13.9 and 13.5 in the intermediate and younger age groups, respectively (*P* = 0.2). The differences were more apparent when restricting the analysis to patients with ductal carcinoma (mean CNV in older patients = 16.4 versus 14.9 in intermediate versus 13.8 in young patients; *P* = 0.05). In a logistic regression model, we found 13 CNV events to be independently associated with age (Fig. 2, Additional file 1). However, upon adjusting for multiple testing, only two CNV events maintained a *P* value <0.05: chr18p loss and chr6q27 deletion; the former was associated with tumors diagnosed in older patients, while the latter was more common in younger patients.

Gene expression differences according to age

We evaluated the association between age at diagnosis and the expression of 10,296 gene expression signatures. In a logistic regression model adjusted for tumor size, nodal status, histology and breast cancer molecular subtype, we found around 1,200 gene signatures to be independently associated with age at diagnosis (FDR <0.05), mainly in younger patients (Additional file 2). The main themes that emerged from this analysis are summarized in Table 3 and indicated higher expression of signatures related to proliferation, stem cell and endocrine resistance in tumors arising at young age.

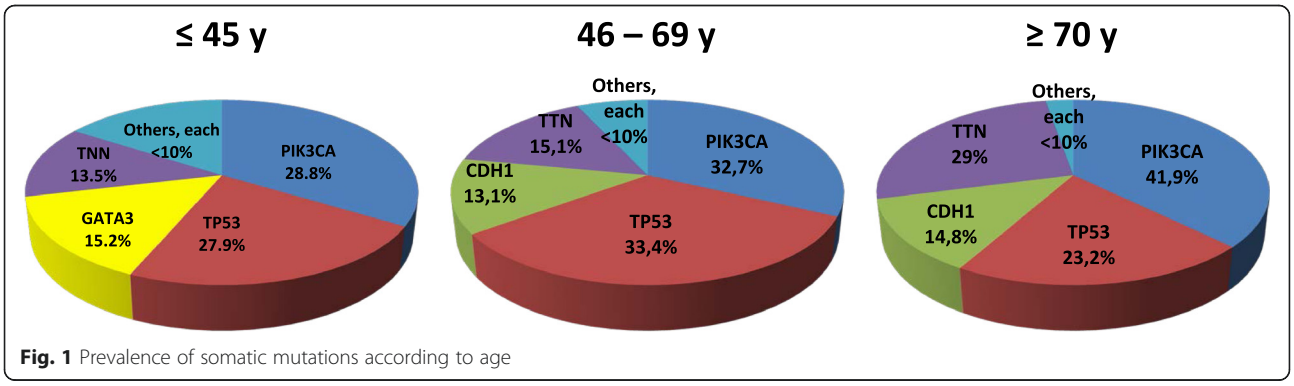


Fig. 1 Prevalence of somatic mutations according to age

Table 2 The independent association between age at diagnosis and somatic mutations

	Young age (≤45 years, n = 118)	Intermediate age (46–69 years, n = 449)	Older age (≥70 years, n = 155)	Logistic model (*P* value)[a]	FDR
Mutations independently associated with young age at diagnosis					
GATA3	18 (15.2 %)	37 (8.2 %)	14 (9 %)	0.003	0.033
Mutations independently associated with older age at diagnosis					
TTN	16 (13.5 %)	68 (15.1 %)	45 (29 %)	0.0003	0.01
KMT2D	1 (0.8 %)	9 (2 %)	9 (5.8 %)	0.0003	0.01
CSPP1	0	3 (0.6 %)	8 (5.1 %)	0.0002	0.01
FOXA1	1 (0.8 %)	6 (1.3 %)	9 (5.8 %)	0.0009	0.013
XIST	0	6 (1.3 %)	9 (5.8 %)	0.0008	0.013
KMT2C	4 (3.3 %)	26 (5.7 %)	18 (11.6 %)	0.002	0.027
SYNE2	3 (2.5 %)	16 (3.5 %)	13 (8.3 %)	0.005	0.033
SPEN	2 (1.6 %)	13 (2.8 %)	12 (7.7 %)	0.005	0.033
USP34	1 (0.8 %)	12 (2.6 %)	9 (5.8 %)	0.004	0.033
ANK2	0	11 (2.4 %)	9 (5.8 %)	0.007	0.043

[a]Analysis adjusted for age, tumor size, nodal status, histology and breast cancer subtype. Only mutations with a minimum prevalence of 5 % in at least one age group is included. FDR, false discovery rate

Discussion

This is the first analysis to explore the prevalence of somatic mutations and CNV according to age. Our findings indicate that age is associated with unique biological features at the DNA level, independent of tumor stage, histology and breast cancer molecular subtype. In addition, age at diagnosis appears to impact the tumor transcriptome confirming previous observations, but also highlighting novel findings. While previous studies provide ample information on the differences at the pathological level according to age [2, 15], this study provides further insights on differences at the genomic

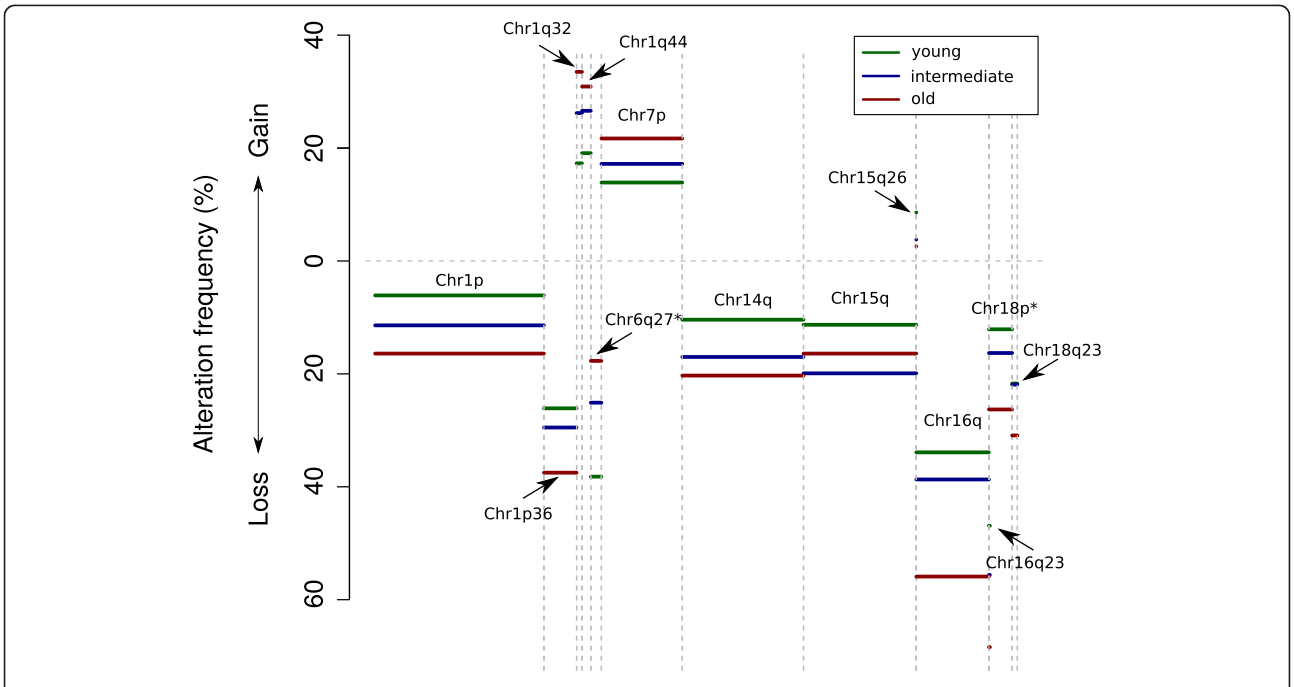


Fig. 2 Copy number variation (CNV) events that are significantly different according to age (*P* <0.05 in the adjusted logistic regression model). Green represents younger patients (≤45 years), blue represents intermediate (46–69 years) and red represents elderly patients (≥70 years). The Y access shows the percentage and indicates the direction of CNV gain (above 0) or loss (below 0). *Aberrations that show a false discovery rate (FDR) <0.05

Table 3 Selected gene expression signatures that are highly expressed in young breast cancer patients

Signature	Logistic regression[b]	FDR
Endocrine resistance		
Creighton_Endocrine_Therapy_Resistance[a]	1.03E-18	1.07E-14
Massarweh_Tamaxifen_Resistance	1.79E-14	1.84-10
Masri_resistance_to Tamoxifen_And_ Aromatase_Inhibitor	7.58E-09	7.66E-05
Proliferation		
Regulation of cell cycle	1.43E-11	1.47E-07
Regulation of Mitotic_Cell_Cycle	8.43E-10	8.59E-06
Stem cell		
Lim_Mammary_Stem_Cell	3.62E-08	0.0003
PID_Notch_Pathway	1.52E-11	1.56E-07
Benporath_SOX2_Targets	1.58E-10	1.62E-06
Galie_Tumor_Stemness_Genes	1.62E-07	0.001
Notch_Signaling_Pathway	1.00E_06	0.009
Nguyen_Notch_Targets_Up	1.12E-06	0.01

[a]More than one signature; [b]model adjusted for tumor size, nodal status, tumor histology and breast cancer subtype. FDR, false discovery rate

level as well. This is also in line with previous studies that showed changes in the normal breast at both the genomic and epigenetic level between young and older women, including changes in genes that are known to be relevant in breast carcinogenesis [16, 17]. Such evidence may suggest the need to explore treatment strategies in patients diagnosed at extremes of age based on their unique molecular makeup.

Different themes emerged from our analysis. First, older patients have more mutations and CNV events. This is likely a reflection of more genomic errors accumulated in the DNA as women age. We found that several somatic mutations were independently associated with older age at diagnosis. Of particular relevance, the high prevalence of KMT2D mutations. Since this gene was recently shown to be involved in tumor proliferation and cell migration [18], we speculate that KMT2D mutations may alter breast cancer behavior. Another finding is the high prevalence of FOXA1 mutations. The latter is required for ER-alpha as a cofactor for chromatin binding and constitutes a major proliferative and survival pathway for luminal-A tumors [19], which are common in older patients [20]. Nevertheless, it is yet to be determined whether these mutations and/or others represent key driver mutations of tumors arising in older patients and the optimal way of targeting them.

On the other hand, GATA3 mutation was the main somatic event that characterized tumors arising at a younger age, which could have relevant clinical implications. GATA3 is an essential component of the ER complex and its mutations are likely to affect ER-regulated transcriptional activity [21, 22]. GATA3 directly upregulates ER-alpha and other proto-oncogenes suggesting that it may promote tumorigenesis in luminal cancer [23]. Preclinical data indicate that mutations in GATA3 also affect ER binding to DNA [22, 24], modulate response of breast cancer cells to estrogen signaling [25], could promote tumor growth [21, 26] and could be associated with endocrine resistance [25]. This is of extreme relevance, since the poor prognosis associated with younger age at diagnosis has been mainly observed in patients with ER-positive breast cancers [3, 5]. We could speculate that the higher prevalence of GATA3 mutations in these patients may render these patients more resistant to endocrine therapy. Our transcriptomic analyses also highlights the high expression of endocrine resistance signatures in younger patients, thus suggesting that endocrine resistance is an important hallmark of tumors arising in young women, worthy of further exploration. Of note, previous preclinical studies have shown that GATA expression (not mutation) results in reversal of the epithelial-mesenchymal transition (EMT) and induction of differentiation in basal-like tumors [27, 28]. Therefore, it is the loss of GATA3 expression that was suggested to contribute to the aggressiveness of basal-like tumors. Using our dataset, we found that GATA3 expression is higher in patients with GATA3 mutation (data not shown). These mutations were mostly exclusive in patients with ER-positive breast cancer. Thus, based on our findings, we cannot assume that the higher rate of GATA3 mutations observed in younger patients is linked to the known increased incidence of basal-like tumors in these patients.

CNVs are genomic events that are regarded as highly biologically relevant in breast cancer [29] and we found two events, more chr18p losses and chr6q27 deletions, to be independently associated with age at diagnosis. chr18p loss was more common in older patients and previous data indicated that it is associated with higher risk of recurrence [30]. Of note, chr18 also harbors SMAD4, which is a known tumor suppressor gene and has been shown to be associated with poor prognosis in several tumor types when lost [31–33]. On the other hand, very little is known on its significance in breast cancer. A previous study showed that chromosome 6 is frequently rearranged in breast cancer, particularly at three regions, including 6q27 [34]. In addition, chr6q27 deletion appears to be more prevalent in tumors with aggressive features [34]. This may suggest that this region could harbor relevant tumor suppressor genes that may contribute to the aggressive nature of tumors arising in younger patients.

Another key point emerging from our study is the existence of relevant gene expression differences according

to age. Previously, we showed that tumors arising in young women are enriched with stem cell-related genes [5]. In addition, Pirone et al. have shown that pathways implicated in maintaining stem cell dynamics, Wnt/β-catenin and ephrin receptor signaling [35, 36] were differentially expressed in the normal breast between young and older women [16]. The current analysis corroborates this association and suggests that targeting the stem cell component is a strategy that deserves exploration in young breast cancer patients. Currently, there are several drugs in development, such as Notch inhibitors that are known to target the stem cell compartment [37]. Of note, in the current analysis, we found high expression of signatures related to Notch signaling pathways (Table 3) in young breast cancer patients, which may suggest the potential relevance of exploring such strategies in younger patients.

We recently initiated a preoperative window trial evaluating the role of targeting RANKL, a known stem cell regulator [38] and in which we have previously shown to be highly expressed in tumors arising at a young age [5, 39]. In this trial (D-BEYOND; NCT01864798), all patients are premenopausal and receive the anti-RANKL monoclonal antibody denosumab before surgery. The aim is to evaluate the impact of RANKL inhibition on several biological processes, including proliferation, stem cell markers, immune-related markers, and many others. The trial has recruited >50 % of its target accrual and represents a proof of concept that could open the door for designing future trials in women diagnosed at extremes of age, based on a better understanding of the biology of their tumors.

Conclusion

In conclusion, the present work shows that tumors arising at different ages are biologically distinct, not only at the protein level, as previously shown, but also at the RNA and DNA levels. This includes aberrations in relevant cancer-related genes. While current treatment decision-making is mainly based on tumor stage and breast cancer subtype, our analysis suggests that age adds a layer of biological complexity, worthy of investigating tailored therapeutic strategies in specific age groups. This could further result in refining therapeutic strategies as we embark on an era of personalized medicine.

Additional files

Additional file 1: The independent association between age at diagnosis and chromosomal copy number variation (CNV) events. (DOCX 15 kb)

Additional file 2: The independent association between age at diagnosis and gene expression signatures (>10,000) in a logistic regression model adjusted for tumor size, nodal status, tumor histology and breast cancer subtype. (PDF 1780 kb)

Abbreviations

CGH: Comparative genomic hybridization; CNV: Copy number variation; EMT: Epithelial-mesenchymal transition; ER: Estrogen receptor; FDR: False discovery rate; GISTIC 2.0: Genomic Identification of Significant Targets in Cancer, version 2.0; HER2: Human epidermal growth factor receptor 2; MSigDB: Molecular Signatures Database; TCGA: The Cancer Genome Atlas.

Competing interests

None of the authors have any competing interests.

Authors' contributions

HAA Jr, BN and SB produced the study concept and design. SB and GZ collected and assembled data. All authors performed data analysis and interpretation. All authors contributed to manuscript writing. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank all patients who donated samples for research purposes. This work was partly supported by research grants from *Le Fonds de la Recherche Scientifique and the Breast Cancer Research Foundation (BCRF)*.

Author details

[1]Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Université Libre de Bruxelles, Boulevard de Waterloo, 121, Brussels, Belgium. [2]Department of Internal Medicine, University of Genova and IRCCS AOU San Martino - IST, Genoa, Italy.

Received: 3 August 2015 Accepted: 1 October 2015
Published online: 15 October 2015

References

1. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. PLoS One. 2009;4, e7695.
2. Gnerlich JL, Deshpande AD, Jeffe DB, Sweet A, White N, Margenthaler JA. 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. 2009;208:341–7.
3. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Ann Oncol. 2010;21:1974–81.
4. Azim HA Jr, Partridge AH. Biology of breast cancer in young women. Breast Cancer Res. 2014;16:427.
5. Azim HA Jr, Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. Clin Cancer Res. 2012;18:1341–51.
6. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-Demore N, et al. Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? J Clin Oncol. 2011;29:e18–20.
7. Wildiers H. Issues in the adjuvant treatment of common tumors (with a focus on breast cancer) in older adults (age >70). Ann Oncol. 2012;23:x339–341.
8. Morrison DH, Rahardja D, King E, Peng Y, Sarode VR. Tumour biomarker expression relative to age and molecular subtypes of invasive breast cancer. Br J Cancer. 2012;107:382–7.
9. van de Water W, Markopoulos C, van de Velde CJ, Seynaeve C, Hasenburg A, Rea D, et al. Association between age at diagnosis and disease-specific mortality among postmenopausal women with hormone receptor-positive breast cancer. JAMA. 2012;307:590–7.
10. Johnson RH, Hu P, Fan C, Anders CK. Gene expression in "young adult type" breast cancer: a retrospective analysis. Oncotarget. 2015;6:13688–702.
11. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160–7.
12. Bastien RR, Rodriguez-Lescure A, Ebbert MT, Prat A, Munarriz B, Rowe L, et al. PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med Genet. 2012;5:44.
13. GenePattern. http://genepattern.broadinstitute.org/gp/pages/index.jsf.

14. 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.
15. El Saghir NS, Seoud M, Khalil MK, Charafeddine M, Salem ZK, Geara FB, et al. Effects of young age at presentation on survival in breast cancer. BMC Cancer. 2006;6:194.
16. Pirone JR, D'Arcy M, Stewart DA, Hines WC, Johnson M, Gould MN, et al. Age-associated gene expression in normal breast tissue mirrors qualitative age-at-incidence patterns for breast cancer. Cancer Epidemiol Biomarkers Prev. 2012;21:1735–44.
17. Johnson KC, Koestler DC, Cheng C, Christensen BC. Age-related DNA methylation in normal breast tissue and its relationship with invasive breast tumor methylation. Epigenetics. 2014;9:268–75.
18. Guo C, Chen LH, Huang Y, Chang CC, Wang P, Pirozzi CJ, et al. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. Oncotarget. 2013;4:2144–53.
19. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005;122:33–43.
20. Azim HA Jr, Azim H. Breast cancer arising at a young age: do we need to define a cut-off? Breast. 2013;22:1007–8.
21. Chou J, Provot S, Werb Z. GATA3 in development and cancer differentiation: cells GATA have it! J Cell Physiol. 2010;222:42–9.
22. Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman MJ, et al. Enhancer activation requires trans-recruitment of a mega transcription factor complex. Cell. 2014;159:358–73.
23. Cohen H, Ben-Hamo R, Gidoni M, Yitzhaki I, Kozol R, Zilberberg A, et al. Shift in GATA3 functions, and GATA3 mutations, control progression and clinical presentation in breast cancer. Breast Cancer Res. 2014;16:464.
24. Gaynor KU, Grigorieva IV, Allen MD, Esapa CT, Head RA, Gopinath P, et al. GATA3 mutations found in breast cancers may be associated with aberrant nuclear localization, reduced transactivation and cell invasiveness. Horm Cancer. 2013;4:123–39.
25. Adomas AB, Grimm SA, Malone C, Takaku M, Sims JK, Wade PA. Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. BMC Cancer. 2014;14:278.
26. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, et al. Mutation of GATA3 in human breast tumors. Oncogene. 2004;23:7669–78.
27. Kouros-Mehr H, Bechis SK, Slorach EM, Littlepage LE, Egeblad M, Ewald AJ, et al. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. Cancer Cell. 2008;13:141–52.
28. Yan W, Cao QJ, Arenas RB, Bentley B, Shao R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. J Biol Chem. 2010;285:14042–51.
29. Ueno T, Emi M, Sato H, Ito N, Muta M, Kuroi K, et al. Genome-wide copy number analysis in primary breast cancer. Expert Opin Ther Targets. 2012;16:S31–35.
30. Climent J, Martinez-Climent JA, Blesa D, Garcia-Barchino MJ, Saez R, Sanchez-Izquierdo D, et al. Genomic loss of 18p predicts an adverse clinical outcome in patients with high-risk breast cancer. Clin Cancer Res. 2002;8:3863–9.
31. Singhi AD, Foxwell TJ, Nason K, Cressman KL, McGrath KM, Sun W, et al. Smad4 loss in esophageal adenocarcinoma is associated with an increased propensity for disease recurrence and poor survival. Am J Surg Pathol. 2015;39:487–95.
32. Kozak MM, von Eyben R, Pai J, Vossler SR, Limaye M, Jayachandran P, et al. Smad4 inactivation predicts for worse prognosis and response to fluorouracil-based treatment in colorectal cancer. J Clin Pathol. 2015;68:341–5.
33. Liu NN, Xi Y, Callaghan MU, Fribley A, Moore-Smith L, Zimmerman JW, et al. SMAD4 is a potential prognostic marker in human breast carcinomas. Tumour Biol. 2014;35:641–50.
34. Noviello C, Courjal F, Theillet C. Loss of heterozygosity on the long arm of chromosome 6 in breast cancer: possibly four regions of deletion. Clin Cancer Res. 1996;2:1601–6.
35. Genander M, Frisen J. Ephrins and Eph receptors in stem cells and cancer. Curr Opin Cell Biol. 2010;22:611–6.
36. Yang L, Tang H, Kong Y, Xie X, Chen J, Song C, et al. LGR5 promotes breast cancer progression and maintains stem-like cells through activation of Wnt/beta-catenin signaling. Stem Cells. 2015;33:2913–24.
37. Han J, Hendzel MJ, Allalunis-Turner J. Notch signaling as a therapeutic target for breast cancer treatment? Breast Cancer Res. 2011;13:210.
38. Asselin-Labat ML, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, et al. Control of mammary stem cell function by steroid hormone signalling. Nature. 2010;465:798–802.
39. Azim HA Jr, Peccatori FA, Brohee S, Branstetter D, Loi S, Viale G, et al. RANK-ligand (RANKL) expression in young breast cancer patients and during pregnancy. Breast Cancer Res. 2015;17:24.

Cancer Treatment Reviews 49 (2016) 65–76

Contents lists available at ScienceDirect

Cancer Treatment Reviews

journal homepage: www.elsevierhealth.com/journals/ctrv

Systematic or Meta-analysis Studies

Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies

Matteo Lambertini [a,b], Luigi Santoro [c], Lucia Del Mastro [d], Bastien Nguyen [e], Luca Livraghi [f], Donatella Ugolini [g], Fedro A. Peccatori [h], Hatem A. Azim Jr. [a,e,*]

[a] BrEAST Data Centre, Department of Medicine, Institut Jules Bordet, and l'Université Libre de Bruxelles (U.L.B.), Brussels, Belgium
[b] Department of Medical Oncology, U.O. Oncologia Medica 2, IRCCS AOU San Martino-IST, Genoa, Italy
[c] Division of Epidemiology and Biostatistics, European Institute of Oncology, Milan, Italy
[d] Department of Medical Oncology, U.O. Sviluppo Terapie Innovative, IRCCS AOU San Martino-IST, Genoa, Italy
[e] Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Brussels, Belgium
[f] USC Oncologia Medica, Azienda Ospedaliera "Papa Giovanni XXIII", Bergamo, Italy
[g] Department of Internal Medicine, University of Genoa, Unit of Clinical Epidemiology, IRCCS AOU San Martino-IST, Genoa, Italy
[h] Fertility and Procreation Unit, Division of Gynecologic Oncology, European Institute of Oncology, Milan, Italy

ARTICLE INFO

Article history:
Received 23 May 2016
Received in revised form 19 July 2016
Accepted 21 July 2016

Keywords:
Breast cancer risk
Breast cancer subtypes
Parity
Age at first birth
Breastfeeding

ABSTRACT

Background: Breast cancer is composed of distinct subtypes defined mainly based on the expression of hormone receptors (HR) and HER2. For years, reproductive factors were shown to impact breast cancer risk but it is unclear whether this differs according to tumor subtype. In this meta-analysis we evaluated the association between parity, age at first birth, breastfeeding and the risk of developing breast cancer according to tumor subtype.

Methods: PubMed and Embase were searched to identify epidemiological studies that evaluated the impact of parity and/or age at first birth and/or breastfeeding on breast cancer risk with available information on HR and HER2. Tumor subtypes were defined as: luminal (HR-positive, HER2-negative or HER2-positive), HER2 (HR-negative, HER2-positive) and triple-negative (HR-negative, HER2-negative). Summary risk estimates (pooled OR [pOR]) and 95% confidence intervals (CI) were calculated using random effects models. The MOOSE guidelines were applied.

Results: This meta-analysis evaluated 15 studies, including 21,941 breast cancer patients and 864,177 controls. Parity was associated with a 25% reduced risk of developing luminal subtype (pOR 0.75; 95% CI, 0.70–0.81; $p < 0.0001$). Advanced age at first birth was associated with an increased risk of developing luminal subtype (pOR 1.15; 95% CI, 1.00–1.32; $p = 0.05$). Ever breastfeeding was associated with a reduced risk of developing both luminal (pOR 0.77; 95% CI, 0.66–0.88; $p = 0.003$) and triple-negative (pOR 0.79, 95% CI, 0.66–0.94; $p = 0.01$) subtypes.

Conclusions: The reproductive behaviors impact the risk of developing breast cancer but this varies according to subtype.

Introduction

Breast cancer is the most frequent tumor in women with an estimated 1.67 million new cases diagnosed worldwide in 2012 [1]. For a decade or more, it has been recognized that breast cancer is a heterogeneous disease with distinct tumor subtypes defined mainly based on the expression of hormone receptors (HR) and HER2 (i.e. luminal, HER2 and triple-negative subtypes) [2]. These subtypes differ in their biology, prognosis, treatment strategies and pattern of metastasis.

Several studies have indicated a clear relationship between the risk of developing breast cancer and the pattern of reproductive behaviors like pregnancy and breastfeeding [3–5]. Yet to date, it is unclear whether the different reproductive factors predispose or protect against certain subtype of breast cancer over the other. A prior meta-analysis showed that parity and young age at first

* Corresponding author at: BrEAST Data Centre, Department of Medicine, Institut Jules Bordet, Boulevard de Waterloo, 121, 1000 Brussels, Belgium. Fax: +32 2 541 3477.
E-mail address: hatemazim@icloud.com (H.A. Azim Jr.).

http://dx.doi.org/10.1016/j.ctrv.2016.07.006
0305-7372/

birth protects against the risk of developing HR-positive breast cancer, while breastfeeding decreases the risk of both HR-positive and HR-negative tumors [6]. However, the studies included in the meta-analysis lacked information on HER2 expression, hence it was not possible to assess the association between these reproductive factors and the risk of developing the different breast cancer subtypes.

Recently, several epidemiological studies tried to address this question evaluating the impact of reproductive behaviors on tumor subtypes defined by HR and HER2 expression, yet with conflicting results [7]. In this meta-analysis we evaluated the impact of parity, age at first birth, breastfeeding on the risk of developing breast cancer according to tumor subtype. Reaching solid conclusions in this regard is rather important as it could have important implications on understanding disease etiology and suggest potential preventive strategies.

Methods

A systematic review and meta-analysis was performed and conducted according to the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines [8].

Study objectives

The primary objective was to evaluate the association between reproductive factors (parity, age at first birth and breastfeeding) and the risk of developing breast cancer according to tumor subtype. The different subtypes of breast cancer were defined based on the expression of HR and HER2 as follows: luminal (HR-positive, HER2-negative or HER2-positive), HER2 (HR-negative, HER2-positive) and triple-negative (HR-negative, HER2 negative).

Subgroup analyses were performed to evaluate the impact of year of study publication, type of estimate (from publication or from crude data), location (United States of America [USA], Europe or Asia), type of study (case–control study, population-based case–control study, prospective cohort study), and cut-off for age at first birth (20 years, 25 years or 30 years).

Data sources and search strategy

A literature search using PubMed and Embase was performed with no date restriction up to October 31st, 2014 using the following keywords: "breast cancer", "risk factors", "parity", "age at first birth", "breastfeeding", "estrogen receptor", "progesterone receptor", "HER2" and "tumor subtype". Boolean operators were used to combine specific keywords for each database and free text terms. According to study protocol, one reviewer (D.U.) designed and set up the effective combination of search terms and discussed with other two reviewers (M.L. and L.L.). The titles and abstracts of the identified studies were independently evaluated by two reviewers (M.L. and L.L.) and consensus was reached by discussion with L.S. and H.A.A. Jr. Cross-referencing from relevant studies and review articles on the topic was conducted to confirm retrieval of all possible pertinent studies.

Selection of the articles

Eligible studies had to fulfill the following criteria: (a) epidemiological studies (cohort studies or case–control studies) that evaluated the associations of parity and/or age at first birth and/or breastfeeding with the risk of developing breast cancer; (b) available information on HR status (estrogen receptor and progesterone receptor) and HER2 status; (c) English-language published studies; (d) the odds ratio (OR) for risk of developing breast cancer had to be reported or could be computed from the data reported in the manuscript.

Studies excluded from the analysis were those with the following characteristics: (a) epidemiological studies other than cohort studies or case–control studies (e.g. case–case studies, case series, etc) that evaluated the associations of parity and/or age at first birth and/or breastfeeding with the risk of developing breast cancer; (b) epidemiological studies (cohort studies and case–control studies) that evaluated the associations of parity and/or age at first birth and/or breastfeeding with the risk of developing breast cancer without information on tumor subtypes; (c) epidemiological studies (cohort studies and case–control studies) that evaluated the associations of reproductive factors other than parity, age at first birth, or breastfeeding (e.g. menarche and menopause, treatment of infertility, use of oral contraceptives) with the risk of developing breast cancer; (d) ongoing studies and studies which were only presented at conferences and not fully published nor available online at the time of the literature search.

For each eligible study, the following data were collected independently by two authors (M.L. and L.L.): first author, year of publication, study design, sample size and source of study subjects, mean age of the subjects, source of information on HR and HER2, type of reproductive factors evaluated (parity and/or age at first birth and/or breastfeeding), adjustments for potential confounders in data analysis, number of study subject and number of breast cancer cases for each tumor subtype.

Statistical analysis

Study-specific estimates of OR of breast cancer subtype (luminal, HER2 and triple-negative) and the 95% confidence intervals (CI) for parity (parous versus nulliparous), age at first birth (advanced age group versus youngest age group) and breastfeeding (ever versus never) were extracted or calculated from each eligible study.

The OR for parity was calculated as the odds of parous women who developed breast cancer divided by the odds of nulliparous women who developed breast cancer. An OR < 1 indicates parous yielded lower probability of developing breast cancer according to tumor subtype.

The OR for age at first birth was calculated as the odds of women at advanced age who developed breast cancer divided by the odds of women at youngest age who developed breast cancer. An OR < 1 indicates advanced age yielded lower probability of developing breast cancer according to tumor subtype.

The OR for breastfeeding was calculated as the odds of ever breastfeeding women who developed breast cancer divided by the odds of never breastfeeding women who developed breast cancer. An OR < 1 indicates ever breastfeeding yielded lower probability of developing breast cancer according to tumor subtype.

For each point estimate, 95% CI were computed.

The association between reproductive factors and risk of breast cancer according to tumor subtype was computed as pooled OR (pOR) with 95% CI. The pOR was considered statistically significant if the 95% CIs did not include 1.0. The pOR was estimated by pooling the study-specific estimates by the Der Simonian & Laird's random effect models [9] fitted using SAS (proc Mixed) with maximum likelihood estimate. These models provided estimates adjusted for the potential correlation within studies as well as the heterogeneity between studies.

The homogeneity of the effect across studies was assessed by using the large sample test based on the Cochrane's Q statistics, which is approximately distributed as a I^2 statistics. A p value <0.10 was used to indicate lack of homogeneity among effects. I^2 statistics was also provided to quantify the percentage of total variation across studies that were attributable to heterogeneity rather

than to chance [10]. The method of Macaskill et al. was used for assessing publication bias [11]. It consists of a funnel-plot regression of the log OR on the sample size, weighted by the inverse of the pooled variance.

To search for potential source of heterogeneity, subgroup analyses have been performed. We considered only study characteristics (country, type of study, year of publication, type of estimate) because the papers included in the present meta-analysis did not provide sufficient subgroup analysis and/or interaction tests based on patient's characteristics to be pooled together.

To assess whether the pooled OR estimates were stable or strongly dependent on one or few of the studies included in the meta-analysis, sensitivity analyses were conducted by iteratively recalculating the pOR estimates after exclusion of the largest studies [12,13].

All reported p values were two-sided. All statistical analyses were performed by using Microsoft Excel and SAS software Version 9.2 for Windows (SAS Institute, Cary, NC). The generation of forest plot was performed using the R software (http://cran.r-project.org/).

Results

A total of 4615 entries were returned with the search strategy and 2 potentially eligible studies were retrieved from the screened papers. After applying eligibility criteria, a total of 15 studies were deemed eligible and were included in the meta-analysis (Fig. 1) [12–26]. Three were prospective cohort studies [12,13,26], 10 were case–control studies [14,16–24], and 2 were pooled analyses (one from 2 cohort studies and 2 population-based case–control studies [25] the other were from 2 population-based case–control studies [15]). A total of 21,941 breast cancer patients (17,307 luminal, 1073 HER2 and 3561 triple-negative) and 864,177 controls (i.e. women who did not develop breast cancer) were included (Table 1).

Ten studies evaluated all the reproductive factors (parity, age at first birth and breastfeeding) [13,15–23], while 3 evaluated parity and age at first birth only [12,14,26], 1 study evaluated parity and breastfeeding [25], and 1 study evaluated age at first birth only [24] (Table 1).

Parity

A total of 14 studies evaluated the association between parity and the risk of developing breast cancer according to tumor subtype [12–23,25,26]. Parous women were those with one or more pregnancies, nulliparous those without pregnancies.

Twelve studies including 863,656 women assessed the risk of luminal breast cancer subtype (Supplementary Table A.1) [12–17,19–23,26]. A highly significant reduction in the risk of developing luminal breast cancer subtype (pOR 0.75; 95% CI, 0.70–0.81; $p < 0.0001$) was observed in parous women, although with significant heterogeneity ($I^2 = 46.2\%$, $P_{heterogeneity} = 0.04$) (Fig. 2A).

Eleven studies including 697,884 women assessed the risk of HER2 breast cancer subtype (Supplementary Table A.2) [12,14–17,19–23,26]. No difference in the risk of developing HER2 breast cancer subtype (pOR 0.90; 95% CI, 0.69–1.16; $p = 0.36$; $I^2 = 33.2\%$, $P_{heterogeneity} = 0.13$) was observed between parous and nulliparous women (Fig. 2B).

Fourteen studies including 865,725 women assessed the risk of triple-negative breast cancer subtype (Supplementary Table A.3) [12–23,25,26]. No difference in the risk of developing triple-negative breast cancer subtype (pOR 1.01; 95% CI, 0.87–1.17; $p = 0.89$; $I^2 = 30.3\%$, $P_{heterogeneity} = 0.13$) was observed between parous and nulliparous women (Fig. 2C).

The sensitivity analysis performed by excluding the largest studies [12,13] provided similar results (Supplementary Table A.4).

The subgroup analyses did not show any significance difference in the risk of developing the different breast cancer subtypes according to type of estimate, location or type of study (Supplementary Table A.4).

Age at first birth

A total of 14 studies evaluated the association between age at first birth and the risk of developing breast cancer according to tumor subtype [12–24,26]. Most studies defined young age at first birth as ⩽24 years and advanced age at first birth as >24 years.

Twelve studies including 676,386 women assessed the risk of luminal breast cancer subtype (Supplementary Table A.5) [12–17,19–23,26]. A significant increase in the risk of developing luminal breast cancer subtype (pOR 1.15; 95% CI, 1.00–1.32; $p = 0.05$) was observed in the advanced age group, although with significant heterogeneity ($I^2 = 86.9\%$, $P_{heterogeneity} < 0.001$) (Fig. 3A).

Eleven studies including 545,216 women assessed the risk of HER2 breast cancer subtype (Supplementary Table A.6) [12,14–17,19–23,26]. No difference in the risk of developing HER2 breast cancer subtype (pOR 0.91; 95% CI, 0.72–1.16; $p = 0.41$; $I^2 = 64.3\%$, $P_{heterogeneity} = 0.002$) was observed between women in the advanced and youngest groups (Fig. 3B).

Fourteen studies including 667,156 women assessed the risk of triple-negative breast cancer subtype (Supplementary Table A.7) [12–24,26]. No difference in the risk of developing triple-negative breast cancer subtype (pOR 0.94; 95% CI, 0.80–1.11; $p = 0.45$; $I^2 = 64.5\%$, $P_{heterogeneity} < 0.001$) was observed between women in the advanced and youngest groups (Fig. 3C).

The sensitivity analysis performed by excluding the largest studies [12,13] provided similar results (Supplementary Table A.8).

Subgroup analysis indicated that the increased risk of developing HER2 breast cancer subtype seemed to be positively correlated to the increased of age at first birth. In the study in which "advanced age" was defined as subjects ">30 years" [12], their risk of developing breast cancer was significantly higher than women who were <30 at first birth (OR 1.83; 95% CI, 1.31–2.56). This increasing trend of the risk was statistically significant ($p = 0.02$) (Supplementary Table A.8).

Breastfeeding

A total of 11 studies evaluated the association between breastfeeding and the risk of developing breast cancer according to tumor subtype [13,15–23,25]. Women were divided between those who ever breastfed regardless of its duration, and those who never breastfed.

Nine studies including 169,870 women assessed the risk of luminal breast cancer subtype (Supplementary Table A.9) [13,15–17,19–23]. A highly significant reduction in the risk of developing luminal breast cancer subtype (pOR 0.77; 95% CI, 0.66–0.88; $p = 0.003$) was observed in women who ever breastfed, although with significant heterogeneity ($I^2 = 79.1$, $P_{heterogeneity} < 0.001$) (Fig. 4A).

Eight studies including 14,266 women assessed the risk of HER2 breast cancer subtype (Supplementary Table A.10) [15–17,19–23]. No significant difference in the risk of developing HER2 breast cancer subtype (pOR 0.78; 95% CI, 0.59–1.03; $p = 0.07$; $I^2 = 45.6\%$, $P_{heterogeneity} = 0.07$) was observed between women who ever or never breastfed (Fig. 4B).

Eleven studies including 176,340 women assessed the risk of triple-negative breast cancer subtype (Supplementary Table A.11) [13,15–23,25]. A significant reduction in the risk of developing triple-negative breast cancer subtype (pOR 0.79; 95% CI,

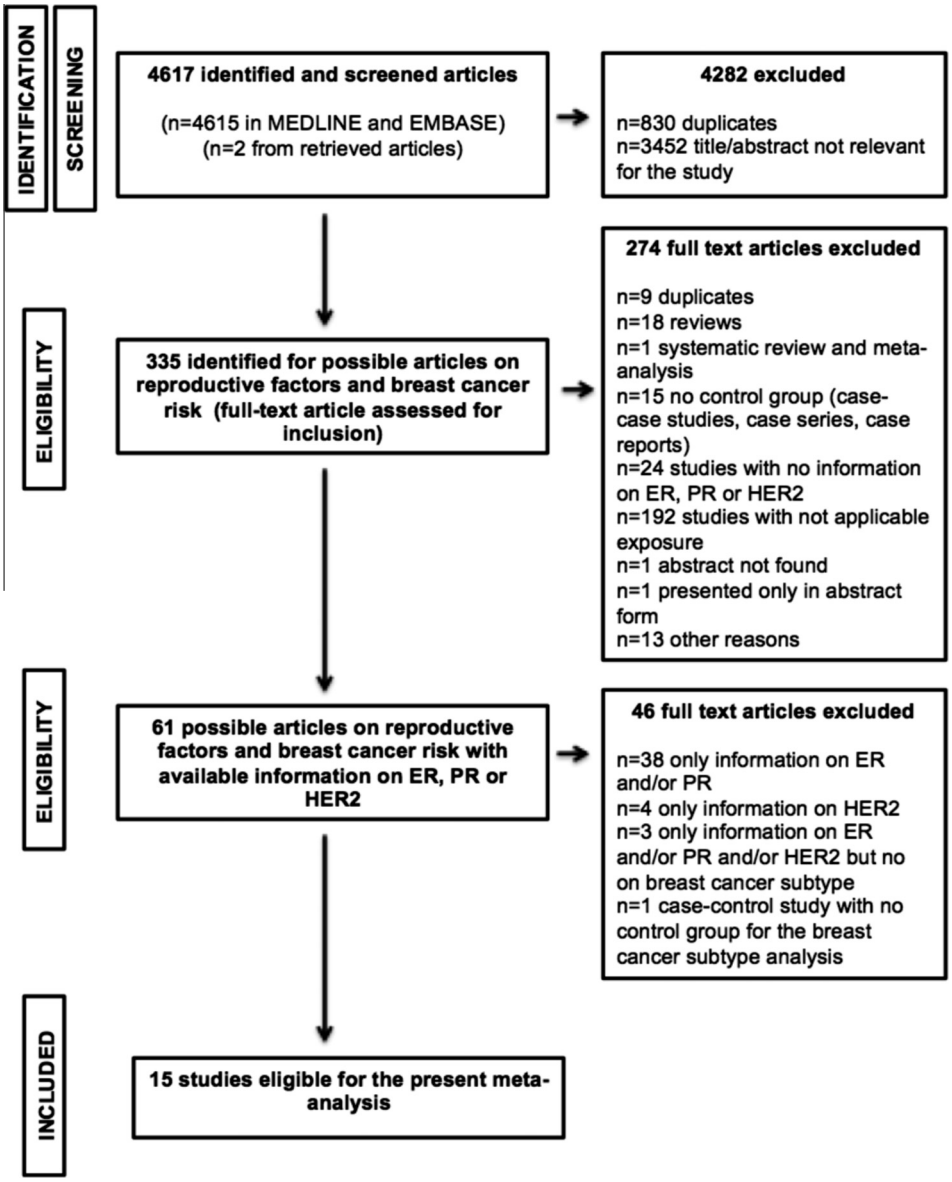


Fig. 1. The flow chart summarizing the process for the identification of the eligible studies. Abbreviations, ER, estrogen receptor; PR, progesterone receptor.

0.66–0.94; $p = 0.01$) was observed in women who ever breastfed, although with significant heterogeneity ($I^2 = 65.1\%$, $P_{\text{heterogeneity}} = 0.001$) (Fig. 4C).

The sensitivity analysis performed by excluding the largest studies [12,13] provided similar results (Supplementary Table A.12).

The subgroup analysis showed that risk of breastfeeding seemed to be estimated differently depending on the type of the study. The protective effect shown for women who breastfed, appeared to be higher within the case–control studies than prospective cohort studies. The difference shown between subgroups was not statistically significant for HER2 breast cancer subtype ($p = 0.06$), while it was statistically significant for luminal breast cancer ($p = 0.03$) and triple-negative breast cancer ($p = 0.015$) subtypes (Supplementary Table A.12). Of note, only two case–control studies were performed and both were conducted in Asia while all the other studies were conducted in the US. Hence, it is possible that the results obtained are partly confounded by the location of the trials (Supplementary Table A.12).

Discussion

Data from recent epidemiological studies suggested possible differential effects of reproductive risk factors on the risk of developing breast cancer according to HR status [6,27,28]. However, no solid conclusions were made partly because HER2 status was not considered in a large fraction of these studies. To our knowledge, this is the first meta-analysis aiming to evaluate the association between parity, age at first birth, breastfeeding, and the risk of developing breast cancer according to tumor subtype. Differently from prior studies, our meta-analysis defined breast cancer subtypes based on the expression of both HR and HER2 status, which remains highly relevant for nowadays practice.

In the present study, a 25% reduction in the risk of developing luminal breast cancer subtype was observed in parous women. Advanced age at first birth showed to be associated with a 15% increase in the risk of developing luminal breast cancer subtype. Importantly, ever breastfeeding was associated with a reduced risk of developing both luminal and triple-negative breast cancer subtypes. However, our results were associated with significant

Table 1
Main characteristics of the study included in the meta-analysis.

Author	Year	Type of Study	Location	Age	Source of information on HR and HER2	Adjustment for potential confounders in data analysis	Categories for age at first birth (years)
Ambrosone et al. [24]	2014	Case–control	USA	20–75	Abstracted from pathology reports	Age, site, education, age at menarche, age at menopause, history of BBD, family history of breast cancer, HRT use, and country of origin	<20, 20–24, 25–29, ⩾30
Palmer et al. [25]	2014	Pooled data from2 prospective cohort study and 2 population-based case–control studies	USA	21–69	Abstracted from pathology reports and cancer registry records	Age in five-year categories, study, geographic region, period, family history of breast cancer, age at menarche, oral contraceptive use, BMI, years of education, alcohol consumption, and cigarette smoking	NA
Horn et al. [26]	2014	Prospective cohort	Norway	40–58	IHC on TMAs constructed form original specimens	Age, birth cohort, age at first birth, number of births, age at menarche, parity and duration of breastfeeding, and socioeconomical status	<25, ⩾25
Li et al. [23]	2013	Population based case–control	USA	20–44	Abstracted from pathology reports and cancer registry records	Age at first live birth, number of live births, age, race/ethnicity, education, annual household income, first-degree family history of breast cancer, BMI, and duration of contraceptive use	<20, 20–24, 25–29, 30–34, ⩾35
Islam et al. [22]	2012	Case–control	Japan	20–79	Original pathology reports	Age, smoking habit, BMI, drinking habit, daily physical activity of any intensity, and family history of breast cancer	<25, 25–29, ⩾30
Phipps et al. [12]	2011	Prospective cohort	USA	40–84	Abstracted from pathology reports	Race, family history of breast cancer in first-degree female relative, and personal history of BBD	<30, ⩾30
Phipps et al. [13]	2011	Prospective cohort	USA	50–79	Abstracted from pathology reports	Age, study arm, race, educational level, family history of breast cancer in first-degree relatives, BMI, HT use, smoking history, and history of mammography use during study	<20, 20–29, ⩾30
Gaudet et al. [21]	2011	Population based case–control	USA	20–56	Central review of original tumor blocks	Age at reference, study site, menopausal status, age at menarche, parity, age at first birth, duration of breastfeeding, BMI, use of oral contraceptives, history of BBD, and family history of first-degree relatives with breast cancer	<20, 20–22, 23–24, >25
Xing et al. [20]	2010	Case–control	China	30–72	Original pathology reports	None	⩽24, 25–29, ⩾30
Ma et al. [19]	2010	Population based case–control	USA	<80	Central review of original tumor blocks	Race, family history of breast cancer in a first degree relative, age at menarche, HRT use, BMI, number of full-term pregnancies, age at first full-term pregnancy, and duration of breastfeeding	⩽19, 20–24, 25–29, ⩾30
Trivers et al. [17]	2009	Population based case–control	USA	20–54	Central review of original tumor blocks	Race, age, education, socio-economic status, insurance, smoking status, alcohol consumption, age at menarche, parity, age at first birth, recency of birth, breastfeeding, physical activity, BMI, and waist-to–hips ratio	<18, ⩾18
Millikan et al. [16]	2008	Population based case–control	USA	20–74	Central review of original tumor blocks	Family history, reproductive history, measures of body size, weight gain, physical activity, environmental exposures, HT use, and socioeconomic status (education and family income)	<26, ⩾26
Phipps et al. [15]	2008	Pooled data from 2 population-based case–control studies	USA	55–79	Abstracted from cancer registry, central review for a restricted sample	Education level, smoking status, alcohol consumption, family history of breast cancer in first-degree relatives, age at first live birth, breastfeeding history, type of menopause, HT use, and age at menopause	<20, 20–24, 25–29, ⩾30
Dolle et al. [18]	2009	Population based case–control	USA	21–45	Central review of original tumor blocks	Age, race, education, annual income, family history of breast cancer, BMI, smoking history, alcohol consumption, age at menarche, number of live births, age at first birth, lactation history, and oral contraceptive use	<20, 20–29, ⩾30
Yang et al. [14]	2007	Population based case–control	Poland	20–74	Abstracted from clinical reports and independent evaluation by the study pathologist	Age, study site, education level, age at menarche, age at menopause, menopausal status, number of full-term births, age at first full-term birth, current/recent oral HRT use among postmenopausal women, previous breast disease, mammographic screening, family history of breast cancer among first-degree relatives, and BMI	Reported as per 5-year increase

Abbreviations: USA, United States of America; HR, hormone receptors; BBD, benign breast disease; HRT, hormone replacement therapy; BMI, body mass index; IHC, immunohistochemistry; TMA, tissue microarray; HT, hormone therapy; NA, not applicable.

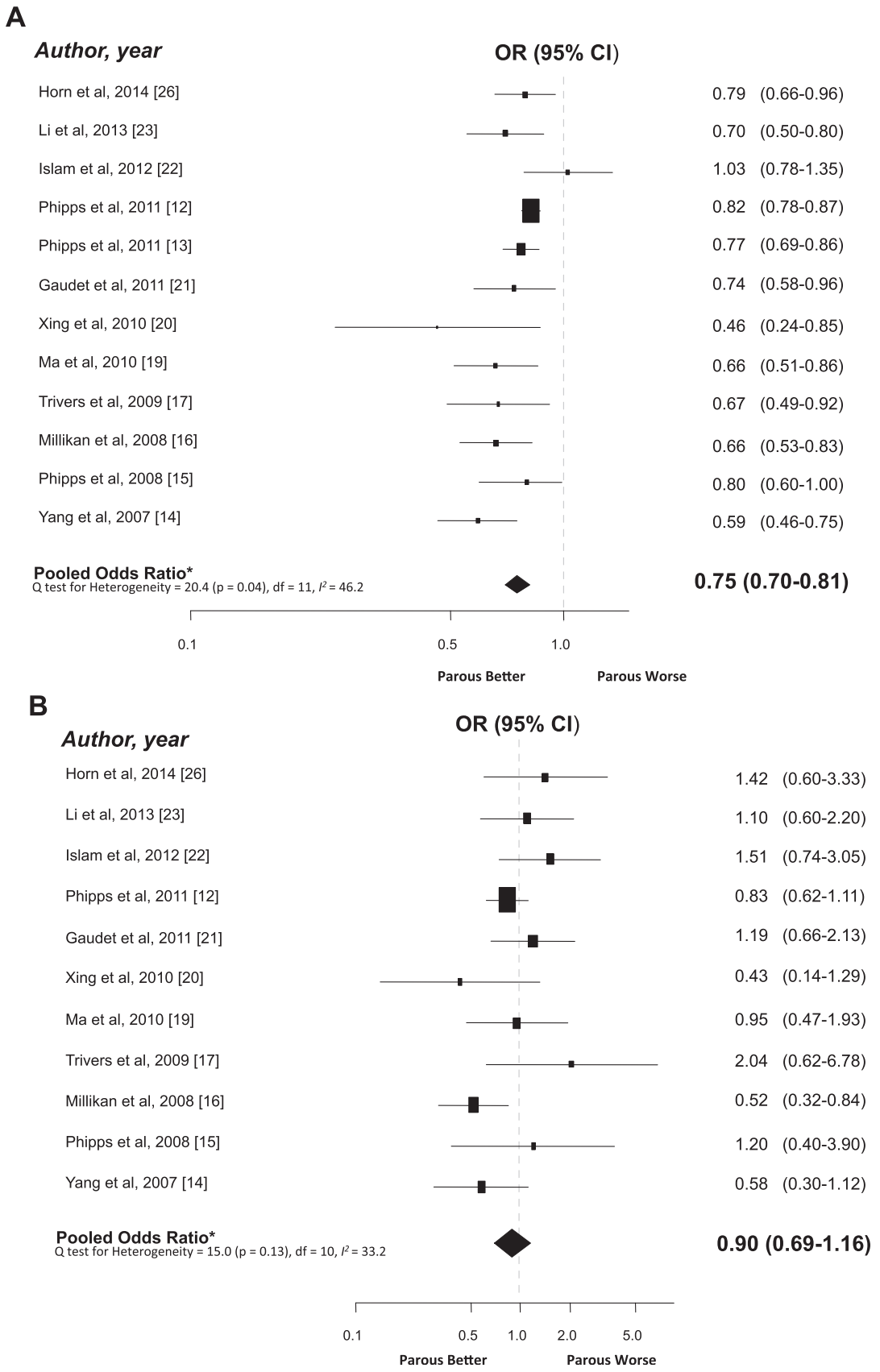


Fig. 2. Parity and risk of breast cancer subtypes. (A) Risk of luminal breast cancer; no publication bias, Macaskill test p value = 0.06. (B) Risk of HER2 positive breast cancer subtype; no publication bias, Macaskill test p value = 0.77. (C) Risk of triple-negative breast cancer subtype; no publication bias, Macaskill test p value = 0.99. *Mixed effect model: estimates adjusted for the correlation within studies and heterogeneity between studies. Abbreviations: OR, odds ratio; CI, confidence intervals.

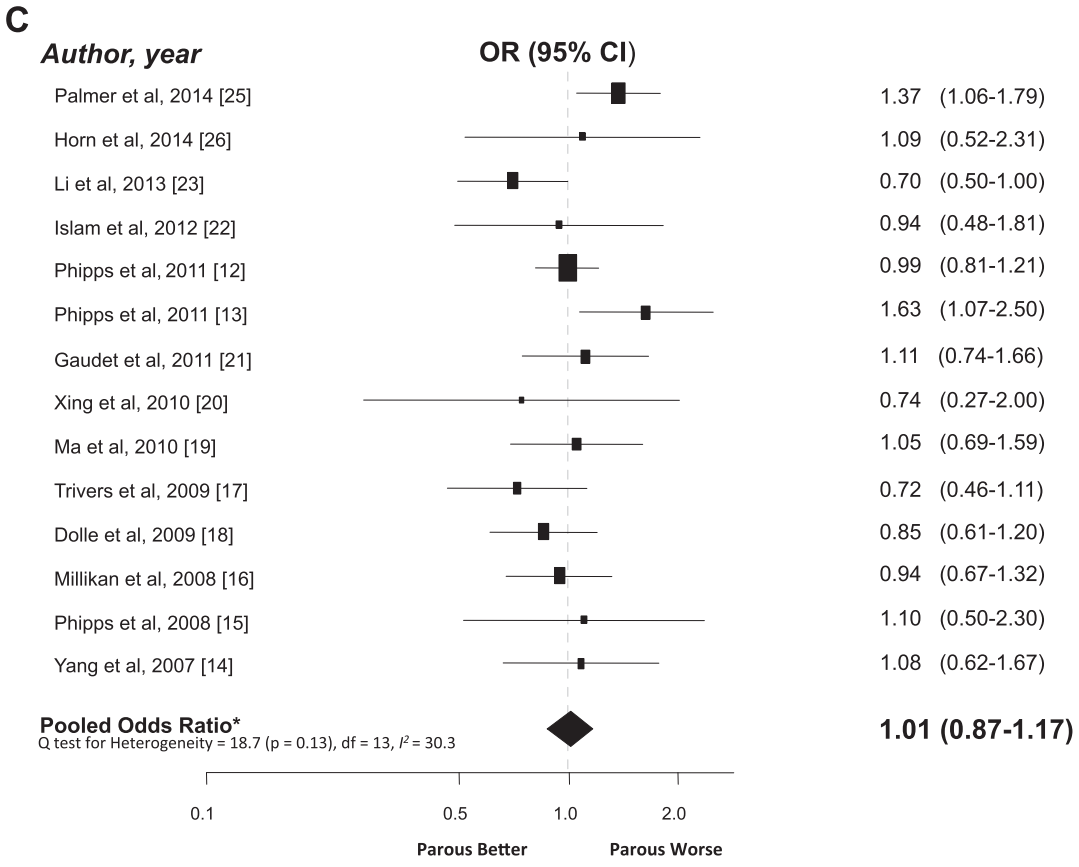


Fig. 2 (*continued*)

heterogeneity. On evaluating the impact of parity, all studies except one [22] showed a reduced risk of luminal breast cancer for parous women. In this case–control Japanese study, parity showed overall no effect on the risk of developing luminal breast cancer but there was a trend toward a reduced risk for women with 3 or more live births as compared to nulliparous women [22]. Similarly, the majority of the studies included in the present meta-analysis showed that advanced age at first birth is associated with an increased risk of luminal breast cancer, except one study [20]. This case–control study conducted in China showed an inverse association with overall reduced risk of luminal breast cancer for women with advanced age at first birth: however, this effect was mainly restricted to women who developed luminal B subtype (HR-positive, HER2-positive) with no impact for those with luminal A subtype (HR-positive, HER2-negative). Thus, heterogeneity seemed to be largely driven by one study in both these analysis, which were outliers and both conducted in Asian populations. On the other hand, the heterogeneity observed in the breastfeeding analysis could be explained by the different definition of "ever" breastfeeding across the evaluated studies, which was relatively broad ranging from less than 2 weeks to more than 12 months.

The results of our meta-analysis suggest a possible explanation on the different incidence of breast cancer subtypes across different regions worldwide. Overall, luminal subtype accounts for the majority of all breast cancers, with HER2-positive and triple-negative breast cancer subtypes representing less than 30% of all breast tumors [29]. However, African-American women have the lowest proportion of luminal breast cancer subtype and the largest proportion of triple-negative breast cancer subtype compared with women of other ethnicities [29]. The higher parity [30] and lower prevalence of lactation [31] in African-American women as compared to white women might represent a contributor to the racial disparity in the incidence of luminal and triple-negative breast cancer subtypes.

The important role of partly modifiable reproductive factors on the risk of developing the different breast cancer subtypes might have important clinical implications. Any lifestyle modification that could minimize cancer risk would have a major impact on public health due to the high incidence of breast cancer. Particularly, any action that might reduce the risk of triple-negative tumors, the breast cancer subtype associated with the worst prognosis, would be even more impactful [28]. A modification in childbearing habit, specifically breastfeeding, might represent a practical way to potentially reduce the risk of breast cancer overall, but also of the most aggressive subtype (i.e. triple-negative breast cancer). Breastfeeding is associated with known benefits for both children and women; it is estimated that the scaling up of breastfeeding to a near universal level could prevent more than 20,000 annual deaths from breast cancer [32]. Supportive measures at many levels (e.g. political support and financial investment) are needed to protect, promote, and support breastfeeding [33]. As outlined in the Healthy People 2020 initiative, a progress toward meeting breastfeeding goals should be pursued particularly among African women, a group that lags in this process [34].

The findings from the present meta-analysis suggest also a possible etiologic heterogeneity among breast cancer subtypes, reflecting possible different mechanisms of carcinogenesis. As already shown decades ago, pregnancy has a dual effect on the risk of developing breast cancer, with transient increased risk after childbirth but reduced in later years [35]. In women with two or more pregnancies, the short-term adverse effect might be masked by the long-term protection given by the first pregnancy [35].

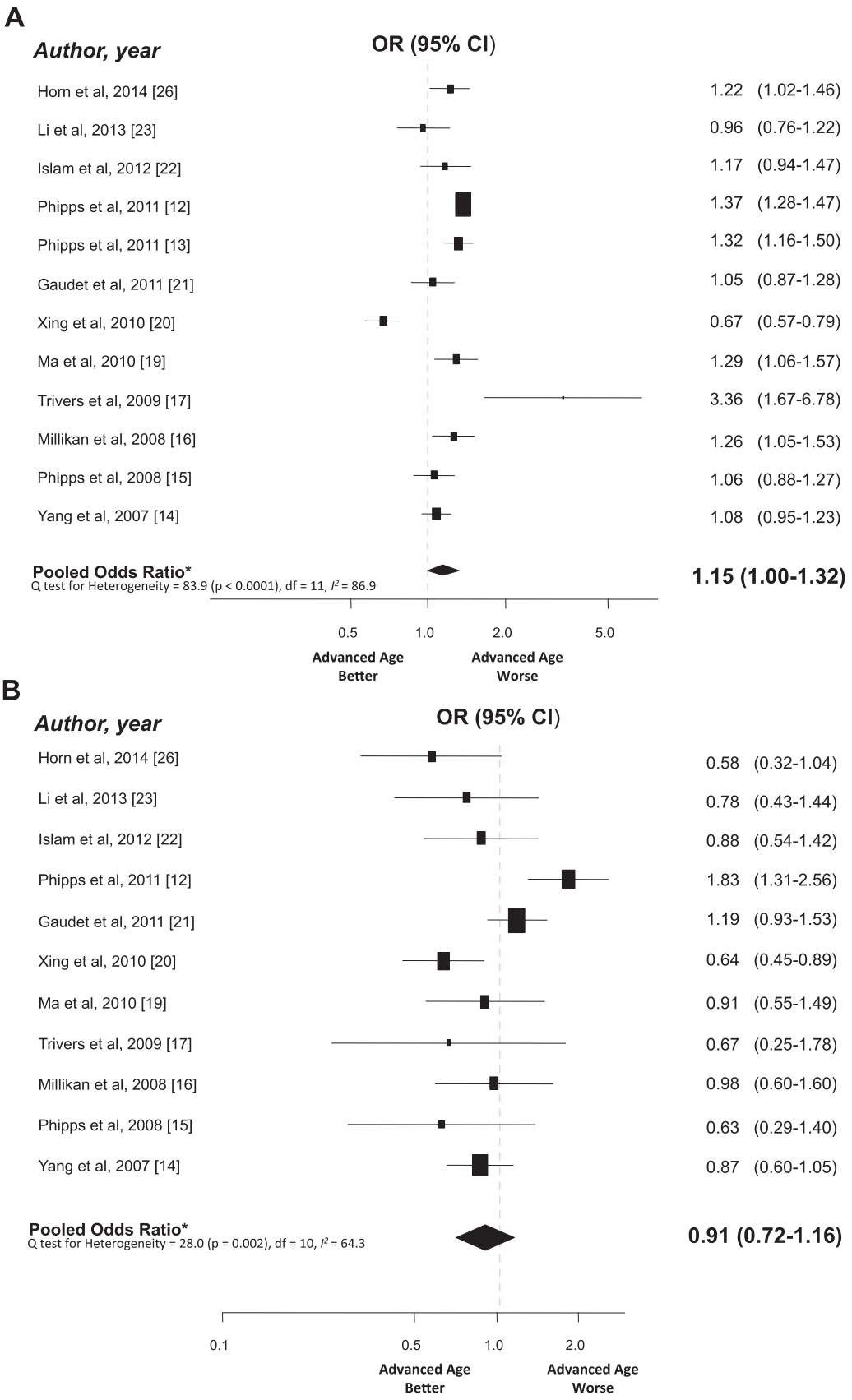


Fig. 3. Age at first birth and risk of breast cancer subtypes. (A) Risk of luminal breast cancer subtype; no publication bias, Macaskill test p value = 0.22. (B) Risk of HER2 positive breast cancer subtype; publication bias, Macaskill test p value = 0.008. (C) Risk of triple-negative breast cancer subtype; no publication bias, Macaskill test p value = 0.27. *Mixed effect model: estimates adjusted for the correlation within studies and heterogeneity between studies. Abbreviations: OR, odds ratio; CI, confidence intervals.

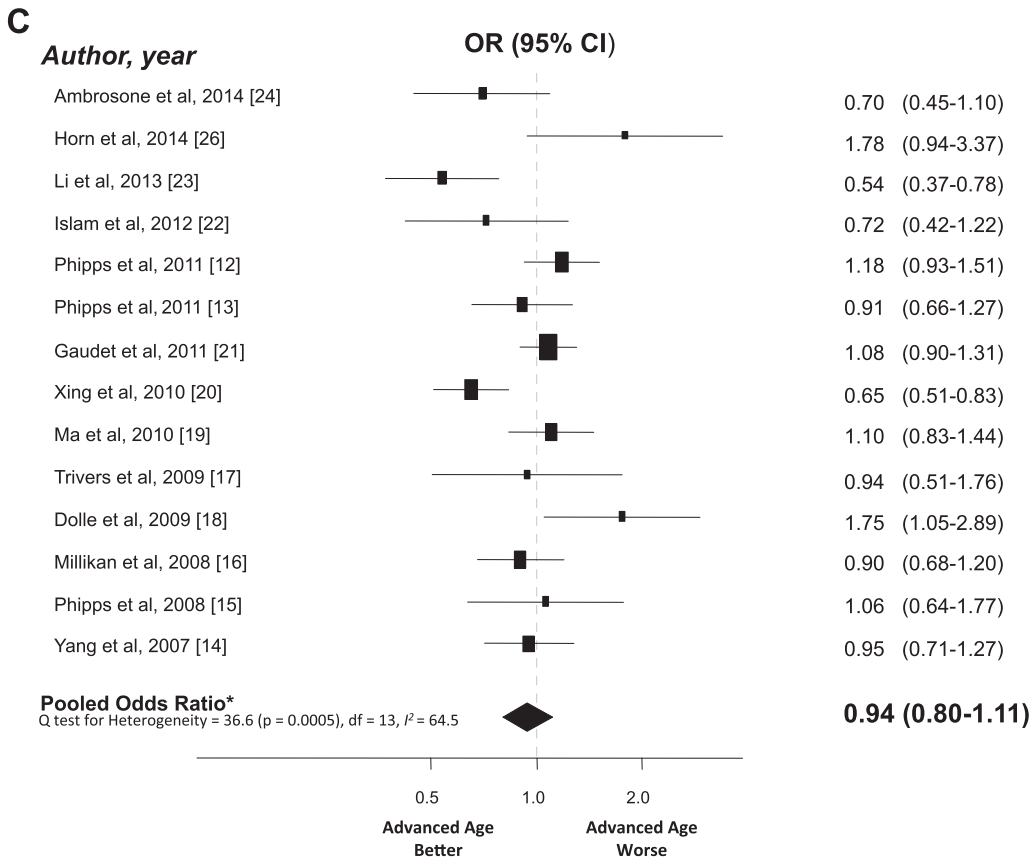


Fig. 3 (*continued*)

Specifically, pregnancy at a young age leads to a strong and lifelong protective effect against the risk of developing breast cancer [36]. The protective effect of parity and young age at first birth against luminal breast cancer but not against HR-negative subtypes (i.e. HER2 and triple-negative breast cancer) suggests that the risk might be influenced predominantly through hormonal mechanisms involving sex hormones. Systemic hormonal changes taking place during pregnancy, particularly in ovarian steroids (i.e. estrogens and progesterone), seem to be major determinants in protecting against mammary cancer [37]. Alterations in systemic hormone levels are generally associated with local cellular processes of the mammary gland that include changes in organ differentiation, hormone responsiveness, proapoptotic pathways, extracellular stroma and mammary stem cell fate [38]. Particularly, a decreased activity of the mammary stem/progenitor cell-related signaling pathway upon pregnancy, renders these cells less proliferative and susceptible to oncogenic transformation [39]. Two recent gene-expression analyses showed unique gene expression patterns in human breast after pregnancy with differential expression of apoptosis-related genes and genes related to cell cycle and cell signaling: this suggests that pregnancy may induce a protective signature against the risk of developing breast cancer [40,41]. Parity, and specifically pregnancy at a young age, showed to be associated with a decrease in the proportion of HR-positive cells and pronounced changes in gene expression (i.e. downregulation of Wnt and transforming growth factor b [TGFb] signaling) leading to a decreased proliferation potential in stem/progenitor cells that showed also to persist into advanced age [42]. Moreover, parous women showed to have a reduction in the number of hormone responsive $p27^{+}$ stem cells with consequent induction of quiescence in mammary hormone-responsive progenitors regulated by the TGFb signaling [43].

On the other hand, breastfeeding appeared to be protective against the risk of developing both luminal breast cancer and triple-negative breast cancer subtypes providing further evidence for possible etiologic heterogeneity in breast tumors. The reduced risk of developing luminal breast cancer subtype in women who ever breastfed suggests the influence of hormonal mechanisms involving sex hormones also for breastfeeding characterized by a shorter exposure to endogenous sex hormones which are reduced during lactation-induced amenorrhea [44]. In the presence of exposure to estrogens, HR-positive progenitor cells produce paracrine signals that stimulate the proliferation of neighboring populations of HR-negative cells [45], suggesting also a possible influence on the risk of developing triple-negative breast cancer subtype. Furthermore, breastfeeding is associated with a permanent alteration in the molecular histology of the breast, characterized by involution of terminal duct lobular units: this is a process known to be associated with a reduced breast cancer risk [46]. Specifically, the lack of involution showed to be associated with risk of developing basal-like (i.e. mainly HR-negative) breast cancers [47]. A molecular "involution signature" regulated by STAT3 signaling seems to be the key mediator of the involution process occurring in the mammary gland during breastfeeding [48].

Some limitations of the present meta-analysis should be acknowledged. This is a meta-analysis of epidemiological studies (cohort studies and case–control studies) with intrinsic methodological limits. All data extracted are not based on individual patient data but were retrieved from published available articles, thus it was not possible to investigate the impact of other important factors (i.e. race/ethnicity, number of children, different ages at first birth, duration of breastfeeding) on the findings of the present meta-analysis. Tumor subtype was defined in the studies included in the present analysis based on the expression of HR

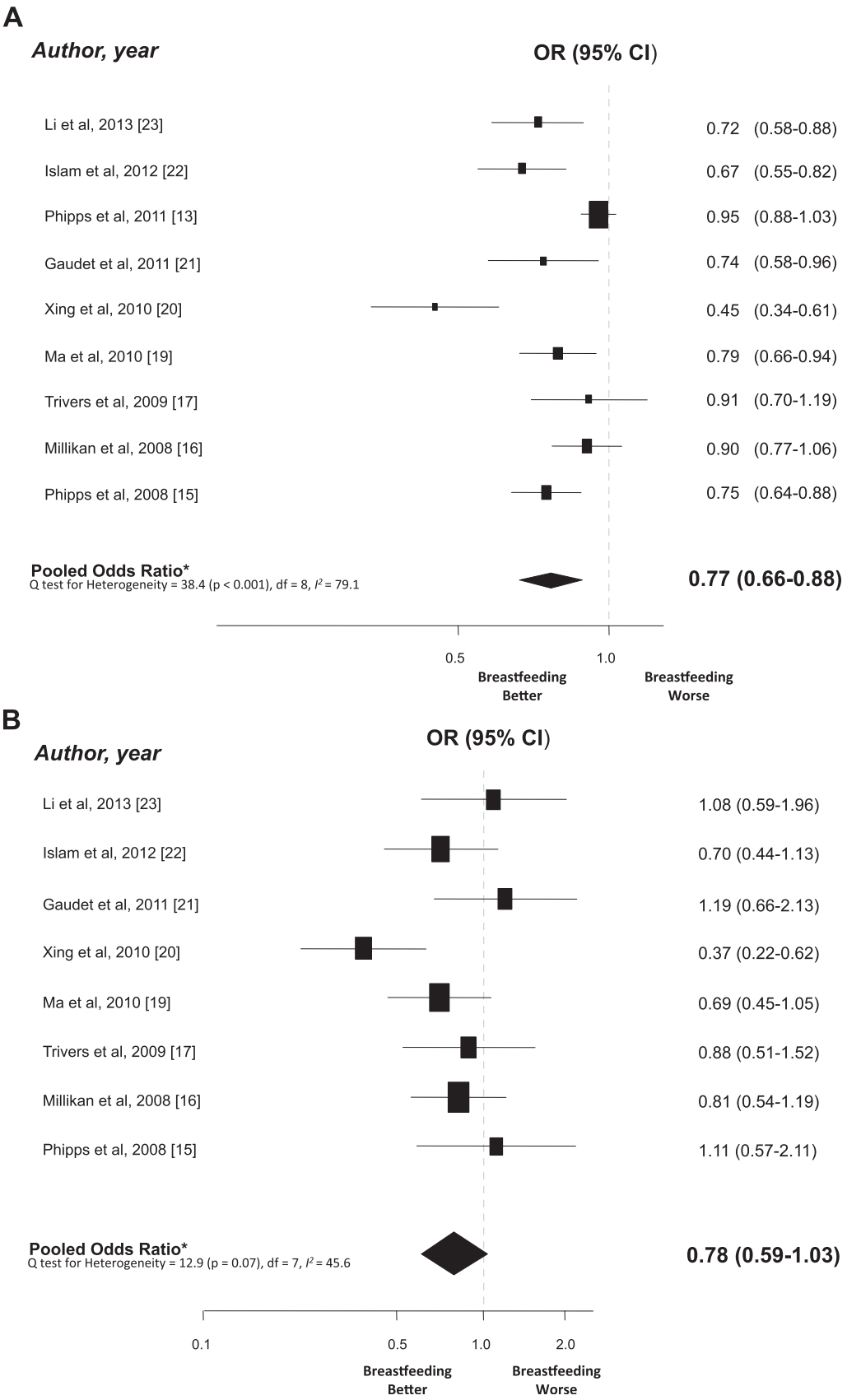


Fig. 4. Breastfeeding and risk of breast cancer subtypes. (A) Risk of luminal breast cancer subtype according to breastfeeding; no publication bias, Macaskill test p value = 0.05. (B) Risk of HER2 positive breast cancer subtype; no publication bias, Macaskill test p value = 0.70. (C) Risk of triple-negative breast cancer subtype; no publication bias, Macaskill test p value = 0.27. *Mixed effect model: estimates adjusted for the correlation within studies and heterogeneity between studies. Abbreviations: OR, odds ratio; CI, confidence intervals.

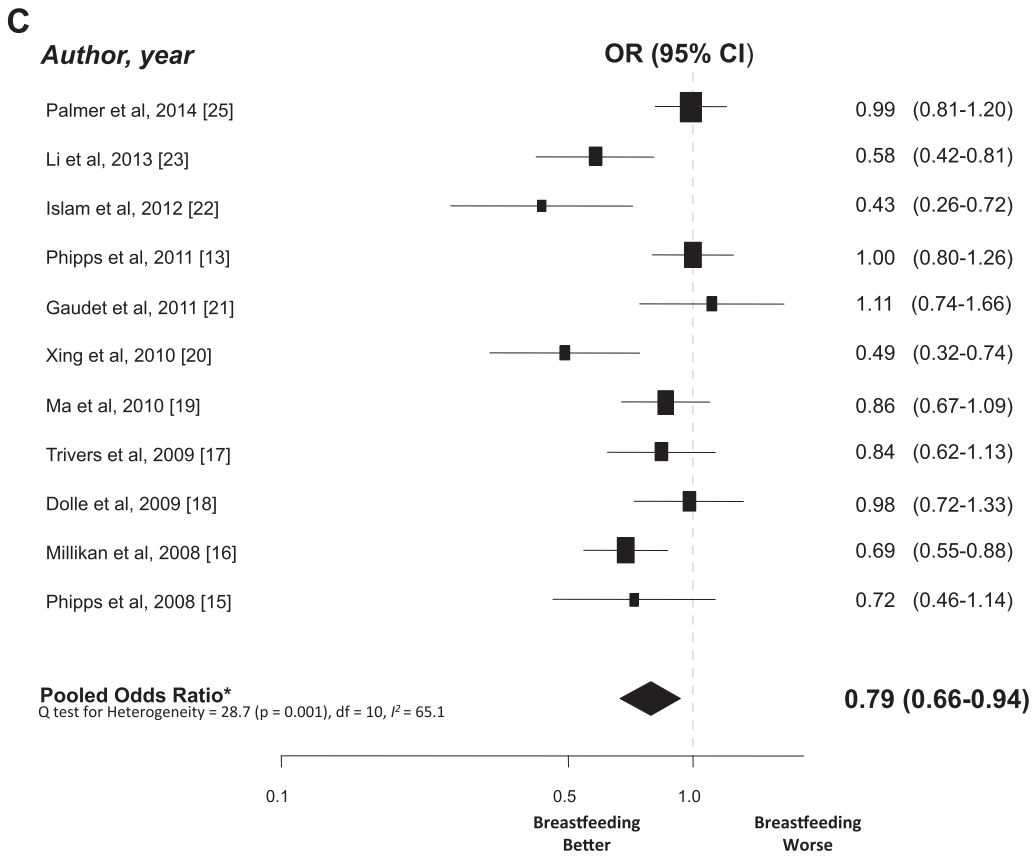


Fig. 4 (*continued*)

and HER2 instead of gene-expression analyses: hence, the molecular subtype of breast cancer was not available. In addition, for the HER2 breast cancer subtype, most of the studies included HER2-positive patients with HR-positive disease in the luminal breast cancer subtype and thus the association among risk factors and HER2-positive/HR-positive breast cancer could not be evaluated. However, these limitations should not significantly influence the overall interpretation of our results especially in light of the rigorous methodology that we applied.

In conclusion, the association between breast cancer risk and parity, age at first birth and breastfeeding varies according to breast cancer subtype. The information from our meta-analysis could partially explain the different incidence of breast cancer subtypes across different regions worldwide and potentially open the door to further study the biological effects of modifiable reproductive behaviors on breast cancer initiation. In addition, our findings could potentially be useful in a more personalized breast cancer risk counseling.

Conflict of interests

The authors declare that they have no competing interests. Matteo Lambertini acknowledges the support from the European Society for Medical Oncology (ESMO) for a Translational Research Fellowship at Institut Jules Bordet. Hatem A. Azim Jr, MD, PhD reports employment at Innate Pharma at the end of this study; this employment is not related in any sort to the subject of the current study.

Funding

This research project was supported by ESMO with the aid of a grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC; investigator Grant No.: 2013-14272). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Any views, opinions, findings, conclusions, or recommendations expressed in this material are those solely of the authors and do not necessarily reflect those of ESMO or AIRC.

Authors' contributions

ML, LS and HAA Jr. conceived the study, had full access to all the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. DU designed the search strategy and discussed with ML and LL. ML and LL performed study selection, data extraction and synthesis. ML, LS and HAA Jr. drafted and led on the writing of the manuscript. All the other authors participated in the analysis and interpretation of the data, revised the manuscript critically for important intellectual content and re-drafted some of its section. All the authors read and approved the final version of the manuscript, and agreed to be accountable for all aspects of the work to ensure its accuracy and integrity.

Acknowledgment

No writing assistance was utilized in the production of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ctrv.2016.07.006.

References

[1] The Globocan Project. Available at: http://www.globocan.iarc.fr. [accessed July 19, 2016].
[2] Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies-improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 2015;26:1533–46.
[3] Bernstein L, Ross RK. Endogenous hormones and breast cancer risk. Epidemiol Rev 1993;15:48–65.
[4] Azim Jr HA, Partridge AH. Biology of breast cancer in young women. Breast Cancer Res 2014;16:427.
[5] Key TJ, Pike MC. The role of oestrogens and progestogens in the epidemiology and prevention of breast cancer. Eur J Cancer Clin Oncol 1988;24:29–43.
[6] Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res 2006;8:R43.
[7] Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast Cancer Res Treat 2014;144:1–10.
[8] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
[9] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
[10] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
[11] Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001;20:641–54.
[12] Phipps AI, Buist DSM, Malone KE, Barlow WE, Porter PL, Kerlikowske K, et al. Reproductive history and risk of three breast cancer subtypes defined by three biomarkers. Cancer Causes Control 2011;22:399–405.
[13] Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Wactawski-Wende J, Kuller LH, et al. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. J Natl Cancer Inst 2011;103:470–7.
[14] Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomark Prev 2007;16:439–43.
[15] Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer. Cancer 2008;113:1521–6.
[16] Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109:123–39.
[17] Trivers KF, Lund MJ, Porter PL, Liff JM, Flagg EW, Coates RJ, et al. The epidemiology of triple-negative breast cancer, including race. Cancer Causes Control 2009;20:1071–82.
[18] Dolle JM, Daling JR, White E, Brinton LA, Doody DR, Porter PL, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. Cancer Epidemiol Biomark Prev 2009;18:1157–66.
[19] Ma H, Wang Y, Sullivan-Halley J, Weiss L, Marchbanks PA, Spirtas R, et al. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. Cancer Res 2010;70:575–87.
[20] Xing P, Li J, Jin F. A case–control study of reproductive factors associated with subtypes of breast cancer in Northeast China. Med Oncol 2010;27:926–31.
[21] Gaudet MM, Press MF, Haile RW, Lynch CF, Glaser SL, Schildkraut J, et al. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. Breast Cancer Res Treat 2011;130:587–97.
[22] Islam T, Matsuo K, Ito H, Hosono S, Watanabe M, Iwata H, et al. Reproductive and hormonal risk factors for luminal, HER2-overexpressing, and triple-negative breast cancer in Japanese women. Ann Oncol 2012;23:2435–41.
[23] Li CI, Beaber EF, Tang M-TC, Porter PL, Daling JR, Malone KE. Reproductive factors and risk of estrogen receptor positive, triple-negative, and HER2-neu overexpressing breast cancer among women 20–44 years of age. Breast Cancer Res Treat 2013;137:579–87.
[24] Ambrosone CB, Zirpoli G, Ruszczyk M, Shankar J, Hong C-C, McIlwain D, 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 2014;25:259–65.
[25] Palmer JR, Viscidi E, Troester MA, Hong C-C, Schedin P, Bethea TN, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. J Natl Cancer Inst 2014;106.
[26] Horn J, Opdahl S, Engstrøm MJ, Romundstad PR, Tretli S, Haugen OA, et al. Reproductive history and the risk of molecular breast cancer subtypes in a prospective study of Norwegian women. Cancer Causes Control 2014;25:881–9.
[27] 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;103:250–63.
[28] Islami F, Liu Y, Jemal A, Zhou J, Weiderpass E, Colditz G, et al. Breastfeeding and breast cancer risk by receptor status-a systematic review and meta-analysis. Ann Oncol 2015;26:2398–407.
[29] DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: convergence of incidence rates between black and white women. CA Cancer J Clin 2016;66:31–42.
[30] Hamilton BE, Martin JA, Ventura SJ. Births: preliminary data for 2012. Natl Vital Stat Rep 2013;62:1–20.
[31] Centers for Disease Control and Prevention (CDC). Racial and ethnic differences in breastfeeding initiation and duration, by state - National Immunization Survey, United States, 2004–2008. MMWR Morb Mortal Wkly Rep 2010;59:327–34.
[32] Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 2016;387:475–90.
[33] Rollins NC, Bhandari N, Hajeebhoy N, Horton S, Lutter CK, Martines JC, et al. Why invest, and what it will take to improve breastfeeding practices? Lancet 2016;387:491–504.
[34] Koh HK, Piotrowski JJ, Kumanyika S, Fielding JE. Healthy people: a 2020 vision for the social determinants approach. Health Educ Behav 2011;38:551–7.
[35] Lambe M, Hsieh C, Trichopoulos D, Ekbom A, Pavia M, Adami HO. Transient increase in the risk of breast cancer after giving birth. N Engl J Med 1994;331:5–9.
[36] Meier-Abt F, Bentires-Alj M. How pregnancy at early age protects against breast cancer. Trends Mol Med 2014;20:143–53.
[37] Rajkumar L, Guzman RC, Yang J, Thordarson G, Talamantes F, Nandi S. Prevention of mammary carcinogenesis by short-term estrogen and progestin treatments. Breast Cancer Res 2004;6:R31–7.
[38] Meier-Abt F, Bentires-Alj M, Rochlitz C. Breast cancer prevention: lessons to be learned from mechanisms of early pregnancy-mediated breast cancer protection. Cancer Res 2015;75:803–7.
[39] Russo IH, Russo J. Pregnancy-induced changes in breast cancer risk. J Mammary Gland Biol Neoplasia 2011;16:221–33.
[40] Russo J, Balogh GA, Russo IH. Full-term pregnancy induces a specific genomic signature in the human breast. Cancer Epidemiol Biomark Prev 2008;17:51–66.
[41] Asztalos S, Gann PH, Hayes MK, Nonn L, Beam CA, Dai Y, et al. Gene expression patterns in the human breast after pregnancy. Cancer Prev Res (Phila) 2010;3:301–11.
[42] Meier-Abt F, Milani E, Roloff T, Brinkhaus H, Duss S, Meyer DS, et al. Parity induces differentiation and reduces Wnt/Notch signaling ratio and proliferation potential of basal stem/progenitor cells isolated from mouse mammary epithelium. Breast Cancer Res 2013;15:R36.
[43] Choudhury S, Almendro V, Merino VF, Wu Z, Maruyama R, Su Y, et al. Molecular profiling of human mammary gland links breast cancer risk to a p27 (+) cell population with progenitor characteristics. Cell Stem Cell 2013;13:117–30.
[44] Turkoz FP, Solak M, Petekkaya I, Keskin O, Kertmen N, Sarici F, et al. Association between common risk factors and molecular subtypes in breast cancer patients. Breast 2013;22:344–50.
[45] Dontu G, El-Ashry D, Wicha MS. Breast cancer, stem/progenitor cells and the estrogen receptor. Trends Endocrinol Metab 2004;15:193–7.
[46] Milanese TR, Hartmann LC, Sellers TA, Frost MH, Vierkant RA, Maloney SD, et al. Age-related lobular involution and risk of breast cancer. J Natl Cancer Inst 2006;98:1600–7.
[47] Yang XR, Figueroa JD, Falk RT, Zhang H, Pfeiffer RM, Hewitt SM, et al. Analysis of terminal duct lobular unit involution in luminal A and basal breast cancers. Breast Cancer Res 2012;14:R64.
[48] Faupel-Badger JM, Arcaro KF, Balkam JJ, Eliassen AH, Hassiotou F, Lebrilla CB, et al. Postpartum remodeling, lactation, and breast cancer risk: summary of a National Cancer Institute-sponsored workshop. J Natl Cancer Inst 2013;105:166–74.