

# Somatic copy number alterations modulate tumor suppressor and oncogene expression to drive early onset of breast cancer

**Pamela Derliche Tonouo** (✉ [tonouopamela@gmail.com](mailto:tonouopamela@gmail.com))

University of Douala

**Esther Hortense Murielle Dina Bell**

University of Douala

**Arnol Auvaker Tiofack Zebaze**

Université de Dschang

**Eliane Ndounga**

Centre Hospitalier et Universitaire de Brazzaville

**Sidonie Noa Ananga**

University of Douala

**Etienne Atenguena**

Yaoundé General Hospital

**Gustave Simo**

Université de Dschang

**Abdel Jelil Njouendou**

University of Buea

**Smiths Sengkwawoh Lueong**

University of Douala

---

## Research Article

**Keywords:** CNA, gene expression, breast cancer, early onset, Oncogene, tumor suppressor

**Posted Date:** February 10th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1335891/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

Female breast cancer (BC) is the leading cause of cancer-related deaths among women worldwide and higher mortality rates are observed in developing countries. Data from African Americans women are often used to draw inference in African BC and little is currently known about BC traits in indigenous African (IA) women.

## Methods

This study prospectively enrolled 472 female patients with BC from Cameroon and Congo Brazzaville between 2007 and 2018. Patient demographics and clinical phenotypes were documented. Additional BC molecular data was downloaded from the cancer genome atlas (TCGA) and the gene expression omnibus (GEO). Demographic and clinical data from the different IA patient cohorts were analyzed to identify unique BC traits in African women. Molecular data from the TCGA was then analyzed to understand key BC traits identified in IA women.

## Results

The average age at BC diagnosis in IA women was more than 10 years earlier ( $46.5 \pm 12.9$ ) than within the TCGA samples ( $58.5 \pm 13.2$ ). Early onset of BC was observed more in IA women than in the TCGA (51.7% vs 15.6%). There was no difference in mean age at diagnosis between African American and white women. IA women had about 60% of T4 stage tumors compared with less than 5% in the TCGA cohort. Using CNA, mutation and gene expression data from the TCGA, we identified somatic copy number amplifications and upregulation of oncogenes such as *CDH6*, *FOXM1*, *NAAA* and *CXCL10* in patients with early onset of BC. Deletion of tumor suppressor with corresponding downregulation was equally observed for *TGM3*, *DMBT1* and *MYO18B*. *TP53* mutations were associated with early BC onset (OR: 2.93,  $p < 0.01$ ). Amplification and upregulation of the oncogene *CDH6*, *PPP1R9B* and *SLC1A3* were associated with survival (HR: 0.51, 95% CI: 0.23-1.13,  $p = 0.096$  and HR: 0.36, 95% CI: 0.18-0.68,  $p = 0.002$ , and HR: 0.14, 95% CI: 0.03-0.6,  $p = 0.008$ , respectively).

## Conclusion

Early onset of BC in IA women in this study might reflect yet unidentified genomic alterations. Dedicated studies are required to establish the genetic underpinning of BC in IA women to support tailored management strategies

## Introduction

Breast cancer (BC) is has become the leading contributor to the global cancer burden. In 2020, with some 2.3 million new cases, representing more than 11% of all cancers worldwide [1] were diagnosed. It is the most prevalent female cancer in 159 countries worldwide and the leading cause of cancer-related deaths in women within 110 countries [1]. Female breast cancer incidence largely exceeds the incidence of other malignancies [1], especially in developing countries. Although incidence rates are higher in industrialized countries, mortality rates are 17% in developing countries [1]. BC has therefore become a major public health challenge in these countries, especially in sub-Saharan Africa. In most sub-Saharan African countries, poverty, poor health infrastructure, lack of adequate training and lack of awareness constitute a lethal cocktail for BC patients. Unfortunately, very few studies have attempted to understand female breast cancer traits in these populations. The fight against BC in these low-income countries will inevitably require an in-depth knowledge of the clinicopathological and molecular traits of breast cancer in these populations.

Studies in African American (AA) women have revealed higher (42%) BC-related death rates compared to white Americans, despite advances in the diagnosis and treatment [2]. Additionally, about 33% of AA women were reported to be diagnosed with BC before the age of 50 years, against 22% of white women [3]. Furthermore, AA women are twofold more likely to be diagnosed with BC before the age of 35 with relatively larger tumors when compared with white women [4, 5]. Recent BC updates, however, have revealed similar mean ages at BC diagnosis between AA and white American women was similar, highlighting significant improvement in awareness and management. Most AA diagnosed with BC were reported to be current or former tobacco users, while white American (WA) women with BC reported alcohol use. Time to medical consultation and time to treatment initiation were significantly higher in AA women than in WA women. AA women were predominantly diagnosed with advanced stage and triple negative BC [2]. Whether or not similar clinical phenotypes and risk factors are operational among IA women is not known

Studies in indigenous African (IA) women have revealed increasing incidence with age up to the age of 45 years, after which a decline is observed [6]. The age-adjusted 5-year overall survival rates are reported to be lower in sub-Saharan Africa compared with North Africa [7]. Although differences in development index might explain the poor outcome, the bottom line remains early age at diagnosis of breast cancer (EOBRCA) in black African women. It is less likely, that socioeconomic factors influence age at BC onset to such an extent, and family history of breast cancers has been shown not be associated with early onset of breast cancer [8]. The role of intrinsic molecular traits have so far not been explored, as well as possible exposure to risk factors. Nonetheless, in low-income setting, where prevention is the most achievement combat strategy, understanding the molecular basis of early onset of breast cancer is an unmet need, as it will allow for the development of population screening tools and monitoring of high-risk population.

Studies addressing the etiology of EOBRCA have suggested the involvement of toxic environment, disruption of hormonal internal milieu or genetic susceptibility [9]. The underlying genetic susceptibility loci have however remained unelucidated. Similarly, studies on occupational, diet and environmental risk factors have led to very little insight. Furthermore, although mutations in high penetrance genes such as *BRCA1* and *BRCA2* are associated with very high risk of contralateral breast cancer, these mutations are only seen in a small fraction of patients with EOBRCA [10, 11]. Using BC stem cell population, the role of the osteoclast differentiation factor *RANKL* has been eluded [12]. Dedicated studies addressing the molecular basis of EOBRCA are still missing both in industrialized and developing countries. Most importantly, there is little, if at all any information on EOBRCA in indigenous black African population. Similarly, molecular data from these population is equally lacking, though needed to design tailored screening and monitoring strategies. There is therefore a crucial need for the identification of molecular and lifestyle risk factors driving early onset of breast cancer, especially in black African women

In the light of the aforementioned, we set out to characterize the clinical features of breast cancer in black IA women from Congo and Cameroon and to identify potential molecular traits supporting early onset of breast cancer using publicly available data sets from the TCGA and the gene expression omnibus (GEO). We show that compared with women of other ethnicity, breast cancer development occurs significantly early in black African women, and that T3 and T4 tumors are the most predominant in black IA women. Furthermore, most women were diagnosed with high-grade tumors and molecular analysis reveal somatic copy number amplification of oncogenes and deletions of tumor suppressor genes in patients with breast cancer onset before or at age 45 years

## Methods

### Patient cohorts

Females histologically confirmed with BC and aged >18 years were recruited at the Yaoundé general hospital (Cameroon), the Douala General Hospital (Cameroon) and at the University teaching hospital (CHU, Brazzaville) in Congo. Administrative authorizations were obtained from all local sampling sites and all participants gave written informed consent to the study. Participants were prospectively recruited between 2007 and 2018 during routine medical visits and were monitored during the entire study period or until they were disease free, death or opted out of the study. Lifestyle data was collected using a structured internal questionnaire, while clinical phenotypes were retrieved from the individual patient files from each sampling site. Patient data for the TCGA cohort (>1000 cases) was downloaded from the Genomic data commons (GDC) repository using the TCGAbiolinks Bioconductor package. Additional patient cohort data (> 300 cases) was downloaded from the gene expression omnibus (GSE3494), which had age, subtype as well as *TP53* gene mutation status. All patients with missing age and gender information or who participated for less than six (06) months were excluded from further analysis. Therefore, 115 patients were retained from the Congo cohort, 65 patients in the Yaoundé cohort and the rest were recruited at the Douala general hospital.

### Database mining and *in silico* analyses

Gene expression data was downloaded from the gene expression omnibus (GEO) for breast cancer patients (GSE3494) and from the cancer genome atlas (TCGA). Copy number variation data was equally downloaded from the TCGA for the breast cancer project (TCGA-BRCA). All data were downloaded from the GDC legacy (genome assembly hg19) database. For all TCGA data, only data from primary tumors were used, by specifying the "Sample.type" to "Primary tumor". For copy number variation data, mean copy number segment data from the Affymetrix SNP 6.0 data was used, while illumina Hiseq gene expression quantification data was downloaded for gene expression analysis using the TCGAbiolinks package. Somatic mutation files for TCGA-BRCA were also download as maf files and processes using the maftools Bioconductor package. All non-mutated samples were excluded from the presented oncoplots (but

included in the summary statistics). Both gene expression and copy number variation data were analyzed following the TCGAblinks user guide. Briefly, copy number segment data was downloaded and filtered on a cut-off threshold of 0.3 for gain or -0.3 for loss. The filtered CNV segment file was then used to create a CNV object. A probe metafile serving as marker matrix was obtained from the Broad institute ([ftp://ftp.broadinstitute.org/pub/GISTIC2.0/hg19\\_support/](ftp://ftp.broadinstitute.org/pub/GISTIC2.0/hg19_support/)) and filtered for common CNVs. The filtered marker matrix was the used to create a marker object and recurrent somatic copy number aberrations were identified using the gaia Bioconductor package. CNVs were annotated using the biomaRt and GenomicRanges Bioconductor packages while circus plots were made with the circlize package. TCGA gene expression data was normalized and filtered using the EDASeq package, while edgeR was used for differential gene expression analyses. Genes with a false discovery rate (FDR) of < 0.05 and a log2fold change of at least 1, were considered as differentially expressed. The affy package was used for processing cell files from the GSE 3494 cohort. Genes whose expression was affected by CNV changes were obtained by filtering the list of recurrent somatic CNV from the EOBRCA group for all CNVs that were equally present in the LOBRCA. The resulting list was then intersected with the list of differentially expressed genes

Gene set enrichment was performed using the desktop GSEA application with 1000 permutations on the hallmark gene sets. Only the hallmarks of cancer gene sets were analyzed and gene sets with a FDR < 0.05 and a normalized enrichment score > 1.5 were considered to be significantly enriched.

## Statistical analyses

The TCGA cohort served as a reference to stratify patients into early and late onsets of BC. Patients were considered to have EOBRCA, if they were first diagnosed before the age of 45 years (mean- 1 SD of the TCGA cohort), or LOBRCA for those diagnosed after the age of 50 years. The mean age at diagnosis of BC was determined by computing the column statistics in Graphpad Prism. The same cut-off was also used for all other cohorts and each cohort was dichotomized into two groups: EOBRCA and LOBRCA. The survminer package was used to determine the gene expression cut-off for survival analyses, while the latter was performed using survival package. Genes with consistency in CNA pattern and gene expression profiles were used in a multivariate cox proportional hazards model to find associations with patient survival in the LOBRCA group. The EOBRCA group was not used for the survival analysis, because the relatively young age of these patients would confound the entire analyses. Relationships between molecular and lifestyle factors and breast cancer phenotypes was assessed by logistic regression. The Kaplan-Meier method was used to compare survival differences and significance was tested using the log-rank test. Several groups were compared using the Kruskal-Wallis test and proportions were compared using the Fisher's exact test while setting the significance threshold to a p value < 0.05. All analyses were performed with the R environment, SPSS or using Graphpad prism software 8.0.0 for Windows, (GraphPad Software, San Diego, California USA).

## Results

Demographic and clinical data were analyzed from 472 females with BC from different care centers in Cameroon and Congo (indigenous Africans). Gene expression and copy number variation data were obtained for western populations from the TCGA and the gene expression omnibus (GEO). The mean age at BC diagnosis within the TCGA cohort was  $58.5 \pm 13.2$  years while the mean age at BC diagnosis for the IA cohort was  $46.5 \pm 12.9$  years. Early onset of breast cancer was defined for those diagnosed before the lower limit of the TCGA cohort ( $58.5 - 13.2$  years). Based on this classifications, 244/472 (51.7%) of African with BC were diagnosed with early onset of BC while 228/472 (48.3%) of cases were diagnosed after the age of 45 years. Within the TCGA cases, early onset of BC was found in 159/1018 (15.6%) of all BC cases. Among IA women, as shown in table 1, there were only very slight differences in the tumor size and subtype of BC between patients diagnosed before 45 years compared with those diagnosed later. Similarly, there was no difference in the SBR grade between both groups. However, women who developed BC after the age of 45 years were more likely to have had menarche after the age of 12 years (21% vs 10%, respectively) (OR: 0.41, 95% CI: 0.10-1.42). Furthermore, most of these women had their first pregnancies before the age of 20 years (47.37% vs 58.42%) and most of them did not use contraceptive pills (30.25% vs 41.41%) and breastfed their children for more than 12 months (64.52% vs 43%).

## Early onset of large and high-grade tumors in African patients

Within the TCGA cohort, there was a slight but non-significant difference in the mean age at diagnosis for Asia, and Black American women compared with white American women. However, the mean age at diagnosis within the IA cohort was significantly lower than any other ethnic origin within the TCGA cohort, and even when compared with African women (figure 1a). We observed an inverse

pattern in the tumor T stage within the two cohorts. Meanwhile more than 80% of tumors from the TCGA cohort were T1 and T2 tumors, 80% of tumors in African women were T3 and T4 (UICC 7<sup>th</sup> edition) (figure 1b). There was no tumor grade information for the TCGA cohort. In the African cohort however, we observed about 60% of all tumors being grade III (figure 1c), meanwhile there was an equal distribution (about 30%) of luminal A and triple negative breast cancer within the African cohort. Her2+ and luminal b tumors accounted for about 15% each (figure 1d). To identify lifestyle and molecular features associated with early onset of BC, we performed a multivariate logistic regression on data from the African cohort. Early onset of menarche ( $\leq 12$  years as well as the use of contraceptive pills were associated (but statistically not significant) with higher odds of developing BC before the age of 45 years (EOBRCA). There was a significant association between EOBRCA and late age at first full pregnancy, (OR: 0.12, 95% CI: 0.04-0.34, p value < 0.001) (first pregnancy at age  $\geq 20$  years). Family history of breast cancer was significantly associated with the development of BC after the age of 45 years (LOBRCA), (OR:4.08, 95% CI: 1.34-13.51, p = 0.016). The use of alcohol, as well as obesity were not significantly correlated with development of LOBRCA (figure 2A). Using molecular data from publicly available Breast cancer studies (GSE3494), we analyzed the relationship between different molecular features and early onset of BC. Of all features analyzed, only *TP53* mutational status was significantly associated with age-dependent breast cancer development. Wild type *TP53* status was significantly associated with late onset of breast cancer (figure 2b). Both low grade tumors and estrogen receptor (ER) positive tumors did associated with late onset of breast cancer but this associated did not reach statistical significance. Similarly, Progesterone receptor positivity did correlate with early onset of breast cancer, although not significantly.

## Higher rates of *TP53* mutations are present in EOBRCA

In order to further dissect the genetic basis of early breast cancer onset, we used molecular data from the TCGA. The samples were dichotomized into EOBRCA and LOBRCA based on the criteria described above. Analyses of somatic mutations revealed that both *TP53* and *PIK3CA* were the most predominantly mutated genes in EOBRCA and LOBRCA. However, a higher rate of *TP53* mutations were observed in patients with EOBRCA, compared with LOBRCA patients (29% vs 26%, respectively figure 3a). *PIK3CA* mutations were observed at higher rates in LOBRCA compared with EOBRCA (30% vs 25%, respectively figure 3b). Given that these mutations accounted for less than 50% of all EOBRCA cases, we further analyzed copy number alterations in both patient cohorts. As shown in the circus plots in figure 3c and 3d, both EOBRCA and LOBRCA showed similar CNA patterns and there were more CNA events in LOBRCA compared with EOBRCA. However, remarkable differences were observed in certain chromosomes, especially chromosomes 3, 5 and 14 in EOBRCA. In this group, there was a remarkable genomic amplification, which was almost absent in LOBRCA. Similarly, on chromosomes 5 and 14 in EOBRCA, there was an obvious deletion, which was not observable in LOBRCA.

## EOBRCA is associated with CNA in oncogenes and tumor suppressors

We investigate if genomic CNA affected gene expression and if the affected genes were associated with cancer development. We specifically focused on genes with CNA exclusively in EOBRCA. To this end, we filtered out all genes with CNA in both EOBRCA and LOBRCA and generated a list of EOBRCA CNAs (supplementary table 1). We performed differential gene expression analysis for all patients in both groups and selected all genes with a FDR < 0.05 and an absolute log<sub>2</sub> fold change greater than 1, to constitute a list of differentially expressed genes (figure 4a, and supplementary table 2). Our list of differentially expressed genes was then intersected with the EOBRCA CNA list and resulted in 61 genes (supplementary table 3). From this table, 18 genes showed similar trends in CNA type (amplification & deletion, supplementary table 4) and gene expression pattern (upregulation and down regulation). Of these 34 genes, 17 genes with copy number amplifications were upregulated, while 17 genes with copy number deletions were down regulated (figure 4b). Of note, several cancer associated genes such as *CDH6*, *FOXM1*, *NAAA* and *CXCL10* [13-17] were amplified and upregulated. Candidate tumor suppressor genes such as *TGM3*, *MYO18B*, *SH3GL2*, *DMBT1*, *SEZ6L* and *SLIT1* were equally deleted [18-25]. Further investigation in CNA events in EOBRCA and LOBRCA, indeed indicated, that between 30-40 Mb on chromosome 5, in the region where *CDH6* is located, there was a strong copy number amplification in EOBRCA (figure 4c), compared with LOBRCA (figure 4d). There was equally a very pronounced genetic deletion between 60-100 Mb on chromosome 5 of EOBRCA, which was barely seen in LOBRCA (figures 4c & 4d). These observations, indeed confirmed, that the observed upregulation of *CDH6* in EOBRCA was not epigenetically regulated but associated with gene amplification. In order to understand the hallmarks of EOBRCA, gene set enrichment analysis was performed on gene expression data of both groups. Hallmarks of *MYC* targets as well as early estrogen response were highly enriched in EOBRCA (figure 4e). Other well know hallmarks of cancer such as TGF $\beta$  and mTORC1 signaling were equally highly enriched in EOBRCA. In LOBRCA, *KRAS*, WNT/ $\beta$ -catenin and hallmarks of inflammatory response were enriched (figure 4f). Comparing the enriched hallmark gene sets in both groups, we observed that there were more cancer hallmark gene sets enriched in the EOBRCA compared with

only 4 genes sets (with FDR < 0.05) enriched in LOBRCA. Most importantly, gene sets associated with highly aggressive tumors such as the hallmarks of *MYC* targets, *TGFβ* as well as mTORC1 were strongly enriched in EOBRCA [26, 27].

## Altered genes are associated with stemness and tumor suppression

Gene expression analyses revealed that some amplified genes were upregulated in EOBRCA, while some deleted genes were downregulated in EOBRCA. We then wanted to investigate if the amplified genes had any tumorigenic properties and if the deleted genes had tumor suppressor activities, at least in solid tumors. As seen in figure 5, the cancer hallmark gene, *FOXM1* was upregulated in almost all EOBRCA and not in LOBRCA. *FOXM1* is a master transcription factor, regulating tumor cell proliferation, self-renewal and tumorigenesis in several human cancers [28]. Similarly, *CDH6*, another EMT promoting gene was upregulated in most of the EOBRCA cases. Another amplified upregulated gene, *PPP1R9B*, a gene that has been associated with tumor progression and stemness [29] in human tumors was amplified and upregulated. Among the deleted and downregulated genes, several of these genes have been reported to have tumor suppressor properties in some solid tumors. In lung cancer, the deleted gene *MYO18B* was reported as a tumor suppressor [24]. Another deleted gene, *TGM3*, is also known to play tumor suppressor roles in colorectal cancer by repressing EMT and PIK3/AKT signaling [30]. In urothelial carcinoma, deletion of the gene *SH3GL2* is known to promote malignant behavior [22], while *DMBT1* has been proposed as a tumor suppressor in brain cancers [23].

## EOBRCA CNA gene expression is prognostic

We then investigated possible associations between patient survival and EOBRCA CNA gene expression in the LOBRCA group. As shown in figure 6a, in a multivariate cox regression, low expression of *CDH6* was significantly associated with better survival, (HR: 0.51, 95% CI: 0.23-1.13, p = 0.096). Similarly, low expression of *PPP1R9B* and *SLC1A3* were associated with better patient survival (HR: 0.36, 95% CI: 0.18-0.68, p value = 0.002 & 0.14, 95% CI: 0.033-0.6, p value = 0.008, respectively), but did not reach statistical significance. Low expression of *CXCL10* was contrarily significantly associated with poor survival (HR: 2.88, 95% CI: 1.38-6.1, p value = 0.005) meanwhile low expression of the deleted tumor suppressor genes *DMBT1* was not significantly associated with poor survival (HR: 1.29, 95% CI: 0.44-3.74, p = 0.64). Low expression of *GPR26*, another EOBRCA deleted gene was associated with poor survival. The expression of other EOBRCA deleted and downregulated genes were not significantly associated with survival. Kaplan-Meier survival analysis revealed, that patients with low expression of *CDH6* lived significantly longer than patients with higher expression log-rank p = 0.01, (figure 6b). Similarly, patients with low expression of *PPP1R9B* lived significantly longer than those with higher expression log-rank p = 0.00035, (figure 6c), while low expression of *SLC1A3* was also associated with better overall survival, log-rank p = 0.015 (figure 6d).

## Discussion

We investigated unique breast cancer traits in IA women by comparing their demographic as well as clinical data with data from the cancer genome atlas. The mean age at diagnosis of BC in African women was more than 10 years earlier (46 years vs 58 years) than that of the TCGA cases. Only about 15% of BC cases were diagnosed before the age of 45 years within the TCGA cases, while more than 50% of African women with BC in our study were diagnosed before the age of 45 years. Meanwhile poor health infrastructure and low socioeconomic standard in African might explain poor outcome, none of these factors are likely to explain the relatively higher rates of early disease onset. Analysis of BC risk factors revealed that even family history of BC was not related with early onset of BC, as postulated in previous findings [8]. About 60% of BC in African women were T4 stage tumors compared with less than 5% within the TCGA. The absence of comprehensive health systems and poverty might deter African women from seeking medical help at early stages of the disease, leading to larger tumors at disease diagnosis. Such large tumors might also be associated with its highly aggressive nature, as reflected by high number of grade III tumors in African women, thereby implicating molecular determinants. The latter is highly plausible, as we observed higher rates of *TP53* mutation in BC patients with early disease onset. In effect, higher rates of *TP53* [31] [32] mutations have been reported in highly aggressive breast tumors, and several studies have revealed highly aggressive tumors in African women [33]. Analysis of hallmarks of cancer in patients with early disease onset within the TCGA cases, revealed the enrichment of gene sets associated with tumor aggressiveness such as the hallmarks of *MYC* targets. In effect, higher *MYC* expression has been reported in aggressive breast tumors and may explain the molecular features of aggressive breast cancer subtypes [34, 35]. Interestingly, meanwhile the aggressive TNBC subtype represents a minor subtype in other populations, with between 10-15% of all BC tumors being TNBC [36], we observed about 2-fold increase rates in TNBC among African women. These observations, further support

the implication of some underlying genetic traits governing disease clinical phenotypes. Among African women with breast cancer diagnosed before 45 years of age, lifestyle risk factors such as lower breastfeeding duration, higher rates of contraceptive use, late age at childbirth and early onset of menarche, which are all BC risk factors might in part explain cancer onset [37]. These factors, however, cannot account for low rates of T2 tumors observed within this group. Oncogene amplifications and inactivation of tumor suppressors are hallmarks of cancer development [38, 39]. The transcription factor *FOXM1* is an established master regulator of tumorigenesis across several human cancers [40]. Meanwhile the role of this gene in different aspect of tumor development and progression is undoubtable, it was until now not clear, of this gene is ubiquitously expressed in all tumors of the same cancer entity. Our results here have painted a clear picture of the age-associated expression of *FOXM1* in more than 350 primary breast tumors, with almost exclusive expressing in younger patients. These findings are of key relevance to the field, as *FOXM1* has for so long been a molecular target of interest for the development of antineoplastic drugs [41]. It is therefore compelling, to analyze age-associated discrepancies in cancer development and to adapt clinical trials to take into account such details in cases where it might be relevant. Similarly, the oncogene *CDH6* is a well-known oncogene and has been linked with poor outcome in other cancer entities [14]. *CDH6* is also known to promote EMT and metastasis in cancer [13]. *CDH6* is also responsible for cellular adhesion and invasion in renal and ovarian cancers [42]. It is therefore very likely, that *CDH6* amplification contributes to early onset on BC. Additionally, *DMBT1*, a tumor suppressor involved in immune defense and epithelial differentiation as deleted and downregulated in EOBRCAs. In effect, the gene was previously shown to be down regulated in BC, although the underlying mechanism remained unclear [43] meanwhile. We now show, that downregulation of this gene is predominantly in EOBRCAs and is mediated by copy number deletion. *TGM3*, a gene that functions as a tumor suppressor and is repressed in several cancer entities [18, 30] was also deleted and downregulated. In effect, this gene has been proposed to be a tumor suppressor by repressing EMT and PIK3/AKT pathway in colorectal cancer [30]. Other gene such as *SLIT1*, *SEZ6L* and *MYO18B*, all of which have been reported to be downregulated in other solid cancers were deleted in EOBRCAs and consequently downregulated at the gene expression level [21, 25, 44]. The deletions of these genes in patients with early onset of BC might indeed render patients more susceptible to cancer development upon exposure to other BC risk factors. Meanwhile much still needs to be done to understand the molecular underpinnings of early breast cancer onset in African women, the present study provides a firm background for potential areas that might be exploited and reveals genomic alterations that might be explored for the development of biomarkers for early detection of BC in African women. Furthermore, our data has revealed the possible implication of age-associated molecular differences in treatment outcome and open new horizons that might help in fine-tuning the design of clinical trials.

## Conclusion

Early onset of breast cancer is predominant in black indigenous African women and associated with late age at childbearing, early onset of menarche and reduced duration of breastfeeding. Molecular analyses of early onset of breast cancer reveal higher rates of *TP53* gene mutation as well as copy number amplification of key oncogenes such as *FOXM1* and *CDH6* as well as deletion of candidate tumor suppressor such as *MYO18B*, *DMBT1* and *TGM3* in EOBRCAs. Dedicated molecular analysis of tumor material from IA women will further improve our understanding EOBRCAs in these populations and fine tune combat strategies. The observed age-specific molecular alterations of key tumor suppressor and oncogenes plaid in favor of age-related considerations in subsequent studies and clinical trials, as this might significantly affect both patient outcome and study endpoints.

## Abbreviations

BC Breast cancer

CI confidence interval

CNA copy number alteration

EMT epithelial-to-mesenchymal transition

EOBRCA early onset of breast cancer

ER estrogen receptor

FDR false discovery rate

GDC genomics data commons

GEO gene expression omnibus

GSEA gene set enrichment analysis

HR hazard ratio

IA Indigenous African

LOBRCA late onset of breast cancer

SD standard deviation

TCGA the cancer genome atlas

TNBC triple negative breast cancer

## Declarations

## Ethics approval

This study was approved at all sampling sites. In Cameroon, the study was approved by the ethical committee of the faculty of medicines and pharmaceutical sciences at the University of Douala (N° 1231 DEI-Udo/11/2017/M). The same ethical clearance was applicable at the Yaoundé general hospital after approval by the by the local institutional ethics committee (N° CBI/395/ERCC/CAMBIN, protocol N° 1086). At the CHU Brazzaville, the study was approved by the CHU Brazzaville internal administrative board.

All participants provided informed consent for the collection and use of clinical data and where necessary molecular data for research purposes. When patients were not able to provide consent directly, their legal representatives (caretakers) provided consent. All informed consent forms included consent for publication.

## Consent for publication

Not applicable

## Availability of data and materials

Molecular data used in this study is publicly available from the gene expression omnibus (GEO) with the accession number GSE3494 or from the TCGA under the TGCA-BRCA project. The data supporting this manuscript is provided in supplementary tables 1-4

## Competing interests

The authors declare that they have no competing interests

## Funding

Not applicable

## Authors' contributions

PDTW: Data collection, curation and analysis and manuscript drafting, EHMDB: Patient recruitment and data curation, AATZ: data collection and curation, EN: patient recruitment and data curation, SNA: patient recruitment and data collection, EA: patient recruitment and data collection, GS: provided resources for study, AJN: Data curation and study organization, SSL: Study design, data curation and analysis, bioinformatics analysis and manuscript writing

# Acknowledgements

Not applicable

## References

1. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. CA Cancer J Clin, 2021. **71**(3): p. 209-249.
2. Foy, K.C., et al., *Disparities in breast cancer tumor characteristics, treatment, time to treatment, and survival probability among African American and white women*. NPJ Breast Cancer, 2018. **4**: p. 7.
3. Clarke, C.A., et al., *Existing data on breast cancer in African-American women: what we know and what we need to know*. Cancer, 2003. **97**(1 Suppl): p. 211-21.
4. Kurian, A.W., et al., *Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups*. Breast Cancer Res, 2010. **12**(6): p. R99.
5. Daly, B. and O.I. Olopade, *A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change*. CA Cancer J Clin, 2015. **65**(3): p. 221-38.
6. Joko-Fru, W.Y., et al., *The evolving epidemic of breast cancer in sub-Saharan Africa: Results from the African Cancer Registry Network*. Int J Cancer, 2020. **147**(8): p. 2131-2141.
7. Ssentongo, P., et al., *Regional, racial, gender, and tumor biology disparities in breast cancer survival rates in Africa: A systematic review and meta-analysis*. PLoS One, 2019. **14**(11): p. e0225039.
8. de Bock, G.H., et al., *A family history of breast cancer will not predict female early onset breast cancer in a population-based setting*. BMC Cancer, 2008. **8**: p. 203.
9. Huo, D., et al., *Concordance in histological and biological parameters between first and second primary breast cancers*. Cancer, 2011. **117**(5): p. 907-15.
10. Metcalfe, K., et al., *Predictors of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers*. Br J Cancer, 2011. **104**(9): p. 1384-92.
11. FitzGerald, M.G., et al., *Germ-line BRCA1 mutations in Jewish and non-Jewish women with early-onset breast cancer*. N Engl J Med, 1996. **334**(3): p. 143-9.
12. Schramek, D., et al., *Osteoclast differentiation factor RANKL controls development of progesterin-driven mammary cancer*. Nature, 2010. **468**(7320): p. 98-102.
13. Gugnoni, M., et al., *Cadherin-6 promotes EMT and cancer metastasis by restraining autophagy*. Oncogene, 2017. **36**(5): p. 667-677.
14. Zhao, Z., et al., *High expression of oncogene cadherin-6 correlates with tumor progression and a poor prognosis in gastric cancer*. Cancer Cell Int, 2021. **21**(1): p. 493.
15. Chawla, A.T., et al., *CtBP-a targetable dependency for tumor-initiating cell activity and metastasis in pancreatic adenocarcinoma*. Oncogenesis, 2019. **8**(10): p. 55.
16. Ueda, K., H. Arakawa, and Y. Nakamura, *Dual-specificity phosphatase 5 (DUSP5) as a direct transcriptional target of tumor suppressor p53*. Oncogene, 2003. **22**(36): p. 5586-91.
17. Wightman, S.C., et al., *Oncogenic CXCL10 signalling drives metastasis development and poor clinical outcome*. Br J Cancer, 2015. **113**(2): p. 327-35.
18. Wu, X., et al., *TGM3, a candidate tumor suppressor gene, contributes to human head and neck cancer*. Mol Cancer, 2013. **12**(1): p. 151.
19. Pleshkan, V.V., et al., *[Transcription of the KLRB1 gene is suppressed in human cancer tissues]*. Mol Gen Mikrobiol Virusol, 2007(4): p. 3-7.
20. Liu, Z.H., et al., *ZIC2 is downregulated and represses tumor growth via the regulation of STAT3 in breast cancer*. Int J Cancer, 2020. **147**(2): p. 505-518.
21. Dickinson, R.E., et al., *Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers*. Br J Cancer, 2004. **91**(12): p. 2071-8.
22. Majumdar, S., et al., *Loss of Sh3gl2/endophilin A1 is a common event in urothelial carcinoma that promotes malignant behavior*. Neoplasia, 2013. **15**(7): p. 749-60.

23. Mollenhauer, J., et al., *DMBT1 encodes a protein involved in the immune defense and in epithelial differentiation and is highly unstable in cancer*. *Cancer Res*, 2000. **60**(6): p. 1704-10.
24. Nishioka, M., et al., *MYO18B, a candidate tumor suppressor gene at chromosome 22q12.1, deleted, mutated, and methylated in human lung cancer*. *Proc Natl Acad Sci U S A*, 2002. **99**(19): p. 12269-74.
25. Nishioka, M., et al., *Identification of a 428-kb homozygously deleted region disrupting the SEZ6L gene at 22q12.1 in a lung cancer cell line*. *Oncogene*, 2000. **19**(54): p. 6251-60.
26. Slack, G.W. and R.D. Gascoyne, *MYC and aggressive B-cell lymphomas*. *Adv Anat Pathol*, 2011. **18**(3): p. 219-28.
27. Zou, J., et al., *Secreted TGF-beta-induced protein promotes aggressive progression in bladder cancer cells*. *Cancer Manag Res*, 2019. **11**: p. 6995-7006.
28. Gartel, A.L., *FOXM1 in Cancer: Interactions and Vulnerabilities*. *Cancer Res*, 2017. **77**(12): p. 3135-3139.
29. Verdugo-Sivianes, E.M., et al., *Mutation of SPINOPHILIN (PPP1R9B) found in human tumors promotes the tumorigenic and stemness properties of cells*. *Theranostics*, 2021. **11**(7): p. 3452-3471.
30. Feng, Y., et al., *TGM3 functions as a tumor suppressor by repressing epithelial to mesenchymal transition and the PI3K/AKT signaling pathway in colorectal cancer*. *Oncol Rep*, 2020. **43**(3): p. 864-876.
31. Kaur, R.P., et al., *Role of p53 Gene in Breast Cancer: Focus on Mutation Spectrum and Therapeutic Strategies*. *Curr Pharm Des*, 2018. **24**(30): p. 3566-3575.
32. Zhang, Y., et al., *Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases*. *Nat Commun*, 2018. **9**(1): p. 3953.
33. Dietze, E.C., et al., *Triple-negative breast cancer in African-American women: disparities versus biology*. *Nat Rev Cancer*, 2015. **15**(4): p. 248-54.
34. Xu, J., Y. Chen, and O.I. Olopade, *MYC and Breast Cancer*. *Genes Cancer*, 2010. **1**(6): p. 629-40.
35. Fallah, Y., et al., *MYC-Driven Pathways in Breast Cancer Subtypes*. *Biomolecules*, 2017. **7**(3).
36. Boyle, P., *Triple-negative breast cancer: epidemiological considerations and recommendations*. *Ann Oncol*, 2012. **23 Suppl 6**: p. vi7-12.
37. Kaminska, M., et al., *Breast cancer risk factors*. *Prz Menopauzalny*, 2015. **14**(3): p. 196-202.
38. Osada, H. and T. Takahashi, *Genetic alterations of multiple tumor suppressors and oncogenes in the carcinogenesis and progression of lung cancer*. *Oncogene*, 2002. **21**(48): p. 7421-34.
39. Zhou, W., et al., *Identification of driver copy number alterations in diverse cancer types and application in drug repositioning*. *Mol Oncol*, 2017. **11**(10): p. 1459-1474.
40. Radhakrishnan, S.K. and A.L. Gartel, *FOXM1: the Achilles' heel of cancer?* *Nat Rev Cancer*, 2008. **8**(3): p. c1; author reply c2.
41. Shukla, S., et al., *The FOXM1 Inhibitor RCM-1 Decreases Carcinogenesis and Nuclear beta-Catenin*. *Mol Cancer Ther*, 2019. **18**(7): p. 1217-1229.
42. Bartolome, R.A., et al., *CDH6-activated alpha1beta3 crosstalks with alpha2beta1 to trigger cellular adhesion and invasion in metastatic ovarian and renal cancers*. *Mol Oncol*, 2021. **15**(7): p. 1849-1865.
43. Braidotti, P., et al., *DMBT1 expression is down-regulated in breast cancer*. *BMC Cancer*, 2004. **4**: p. 46.
44. Yanaihara, N., et al., *Reduced expression of MYO18B, a candidate tumor-suppressor gene on chromosome arm 22q, in ovarian cancer*. *Int J Cancer*, 2004. **112**(1): p. 150-4.

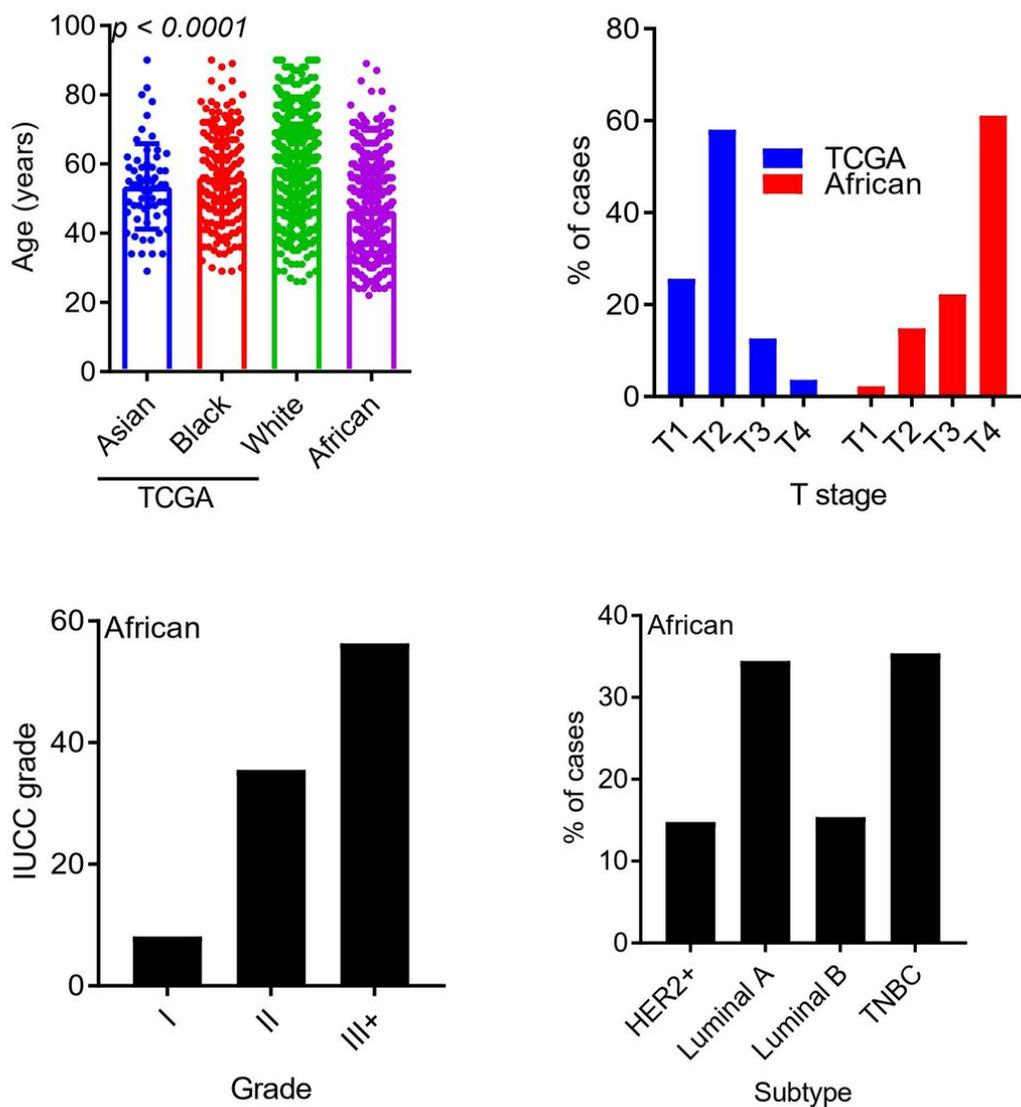
## Table

Table 1

Patient baseline characteristics

Variables	Sample size 472										p
	A					B					
	≤45 (244) 51.70%					>45 (228) 48.30%					
Tumor Size	T1	T2	T3	T4	Tx	T1	T2	T3	T4	Tx	0.015
	6 (4.19)	29 (20.28)	32 (22.38)	73 (51.05)	3 (2.1)	8 (5)	42 (26.25)	32 (20)	72 (45)	6 (37.75)	
Grade	SBR I		SBR II	SBR III		SBR I		SBR II	SBR III		0.17
	31 (20.26)		70 (45.75)	52 (33.99)		31 (18.79)		80 (48.48)	54 (32.73)		
UICC Stage	I	II	III		IV	I	II	III		IV	0.02
	5 (3.70)	25 (18.52)	77 (57.03)		28 (20.75)	4 (3.98)	31 (30.69)	53 (52.48)		13 (12.85)	
Subtype	Luminal A	Luminal B	HER2 Like	TNBC		Luminal A	Luminal B	HER2 Like	TNBC		0.018
	58 (36.25)	25 (15.63)	18 (11.25)	59 (36.87)		61 (33.33)	30 (16.39)	28 (15.30)	64 (34.98)		
Menarche	≤12		>12			≤12		>12			0.0001
	31 (21.53)		113 (78.47)			9 (10)		81 (90)			
Parity	≤6		>6			≤6		>6			0.0001
	117 (80.69)		28 (19.31)			107 (60.80)		69 (39.20)			
Age at first pregnancy	≤20		>20			≤20		>20			0.022
	42 (41.58)		59 (58.42)			60 (52.63)		54 (47.37)			
Contraception	Yes		No			Yes		No			0.0001
	59 (40.41)		87 (59.59)			36 (30.25)		83 (69.75)			
Menopause	Yes		No			Yes		No			0.0001
	12 (7.84)		141 (92.16)			94 (83.93)		18 (16.07)			
Breastfeeding duration	≤12 months		>12 months			≤12 months		>12 months			0.0001
	80 (64.52)		44 (35.48)			36 (43.00)		46 (57.00)			

## Figures



**Figure 1**

**Early onset of highly aggressive breast tumors is prominent among indigenous Africans (IA) with female breast cancer.** A) A scatter plot showing the distribution of age at first diagnosis of breast cancer in patients of different ethnicity in the TCGA cohort (N = 1018) and in female breast cancer patients from Africa (Cameroon and Congo, N = 472). B) A bar plot showing the distribution of the different tumor stages in the TCGA cohort and in the IA cohort. C) A bar plot demonstrating the proportions (as percentages) of different tumor grades observed among IA women with breast cancer. C) A bar plot show the distribution of the different breast cancer molecular subtypes among IA women diagnosed with breast cancer.

A)

Predictors	OR	95%CI	p-value
Menarche age [≤ 12 years]	0.41	0.10 – 1.42	0.178
Contraceptive use [yes]	0.57	0.20 – 1.56	0.283
Alcohol [yes]	1.74	0.57 – 5.60	0.337
BMI [Obese]	2.4	0.64 – 9.55	0.197
BMI [Overweight]	3.23	0.99 – 11.58	0.059
Preganancy age [≥ 20 years]	0.12	0.04 – 0.34	<0.001
Exccercise [yes]	0.93	0.27 – 3.12	0.906
Familial cancer [yes]	4.08	1.34 – 13.51	0.016

B) Molecular factors

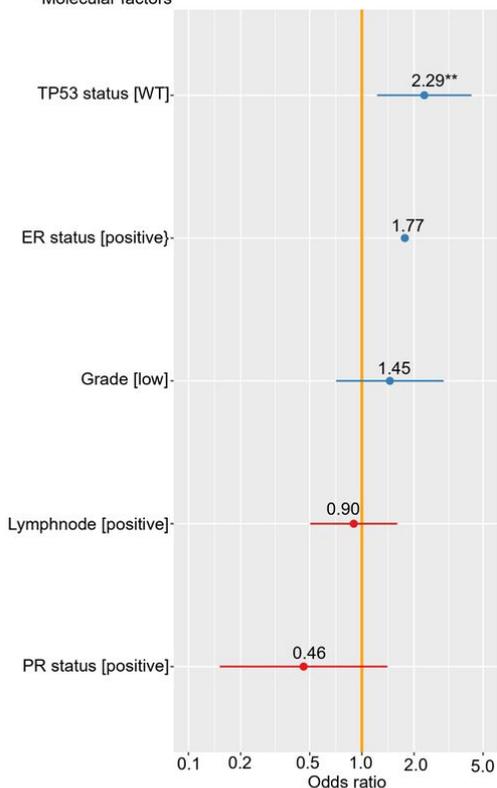
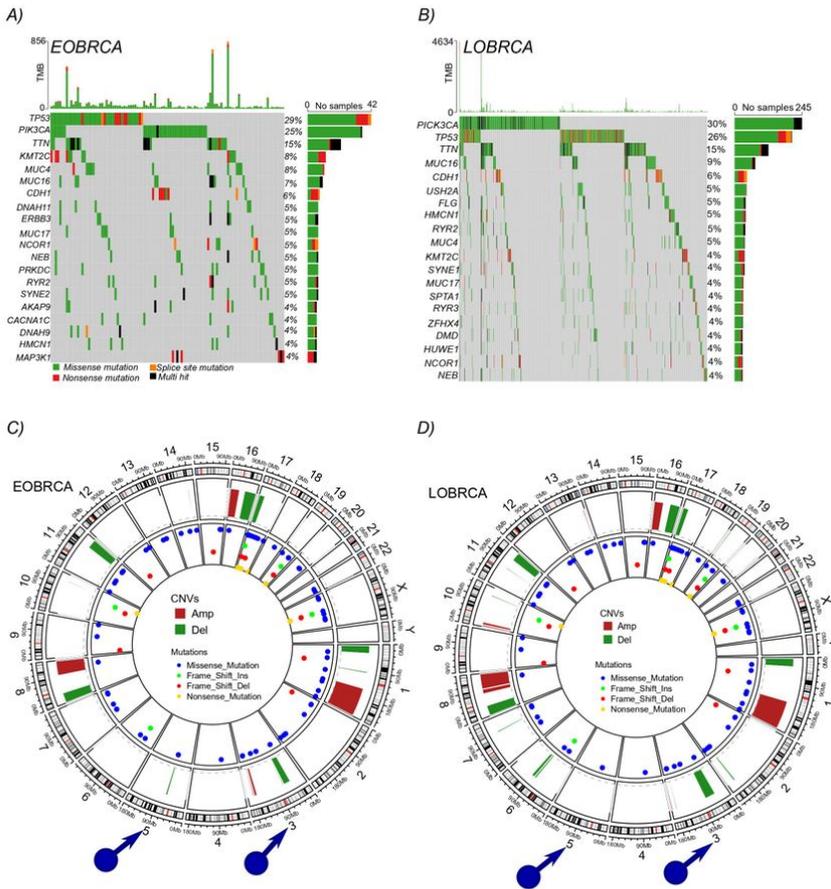


Figure 2

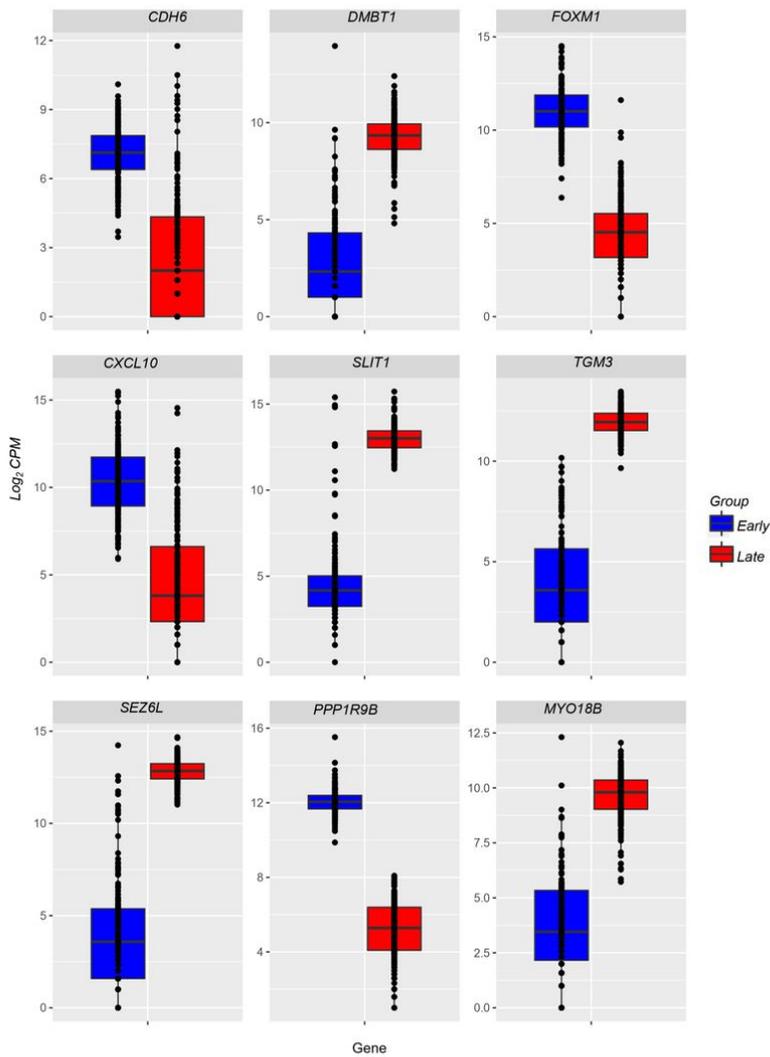
**Molecular and lifestyle risk factors are associated with early onset of breast cancer (EOBRCA) in IA women.** A) A table showing odds ratio for the development of EOBRCA (as reference) from multivariate analyses of lifestyle risk factors collected from IA women with breast cancer. EOBRCA is defined as development of BRCA before the age of 45 years, corresponding to the mean – 1SD derived from the TCGA cohort. B) A forest plot showing molecular factors associated with EOBRCA (as reference). Molecular data was obtained from a publicly available breast cancer cohort (GSE 3494) and patients were stratified to EOBRCA and LOBRCA as described above.



**Figure 3**

**Genetic characterization of EOBRCAs reveals specific patterns in PIK3CA and TP53 mutations and differences in somatic copy number alterations.** A) An oncoprint showing the top 20 most mutated genes in EOBRCAs breast cancer patients from the TCGA cohort (samples without any mutation are excluded from the plot but included in the statistics). B) An oncoprint showing the top 20 most mutated genes in LOBRCA breast cancer patients from the TCGA cohort (samples without any mutation are excluded from the plot but included in the statistics). C) A circos plot, showing a summary of all somatic copy number alterations (green = deletion and red = amplifications) and somatic mutations in the different chromosomes in patients with EOBRCAs from the TCGA cohort. D) A circos plot, showing a summary of all somatic copy number alterations (green = deletion and red = amplifications) and somatic mutations in the different chromosomes in patients with LOBRCA from the TCGA cohort. For computational reasons, the first 200 patients in the LOBRCA were analyzed, leading to comparable sizes in the EOBRCAs (159 cases) and LOBRCA (200 cases)





**Figure 5**

**Copy number alterations in key oncogenes and tumor suppressor genes alter their expression in EOBRCa and LOBRCA.** Boxplots, showing the expression of the amplified and upregulated oncogenes *FOXM1*, *CDH6*, *CXCL10* and *PPP1R9B* as well as deleted and downregulated candidate tumor suppressor genes *TGM3*, *SEZ6L*, *SLIT1* and *DMBT1* in EOBRCa and LOBRCA.

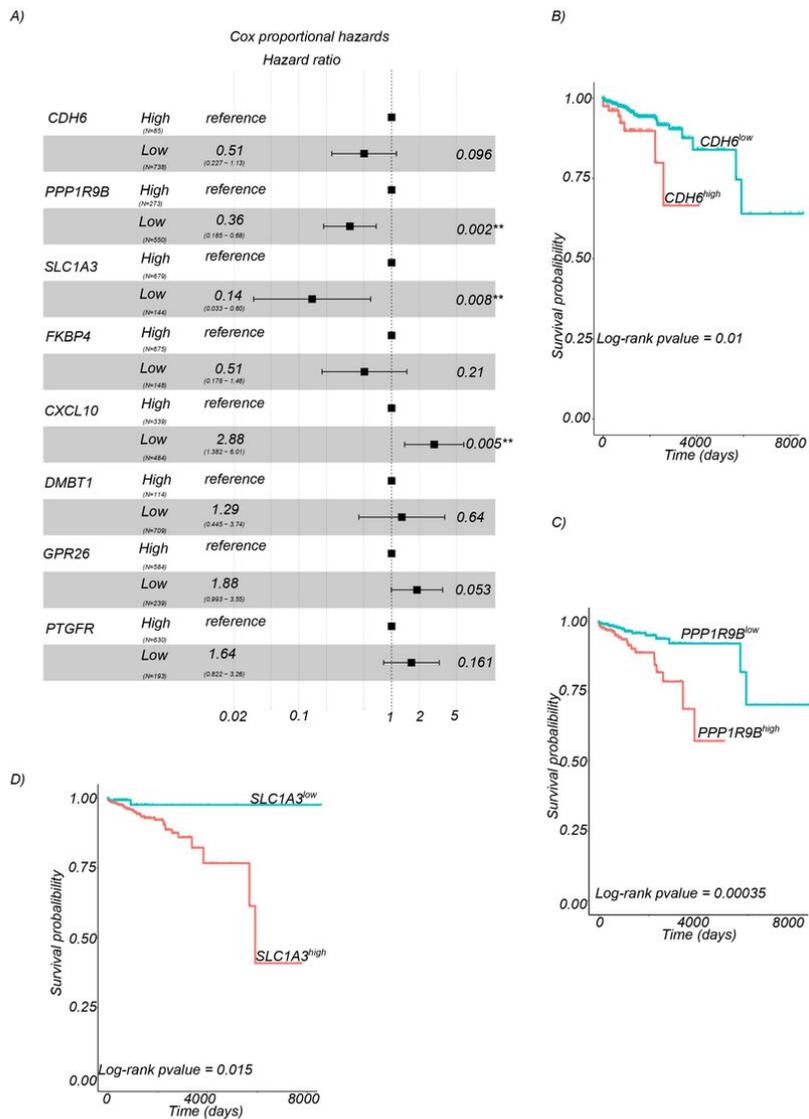


Figure 6

The expression of oncogenes and tumor suppressors with somatic copy number alterations in EOBRCA have prognostic value in BRCA.

A) A forest plot showing multivariate cox proportional hazards ratios for the association of genes with somatic copy number alteration and breast cancer patient survival. Data is shown for all patients with LOBRCA from the TCGA. B) Kaplan-Meier overall survival curve for patients with LOBRCA from the TCGA cohort. Patients were stratified for the expression of the EOBRCA amplified gene *CDH6*. The stratification cut-off was statistically determined using the survminer package. C) Kaplan-Meier overall survival curve for patients with LOBRCA from the TCGA cohort. Patients were stratified for the expression of the EOBRCA deleted gene *KLRB1*. The stratification cut-off was statistically determined using the survminer package.

## Supplementary Files

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

- [Supplementarytable1.xlsx](#)
- [Supplementarytable2.xlsx](#)
- [Supplementarytable3.xlsx](#)

- [Supplementarytable4.xlsx](#)