


RESEARCH

Open Access



# Whole-genome sequencing of phenotypically distinct inflammatory breast cancers reveals similar genomic alterations to non-inflammatory breast cancers

Xiaotong Li<sup>1,2</sup>, Sushant Kumar<sup>1,3</sup>, Arif Harmanci<sup>4</sup>, Shantao Li<sup>1</sup>, Robert R. Kitchen<sup>1,3,5</sup>, Yan Zhang<sup>1,6,7</sup>, Vikram B. Wali<sup>2</sup>, Sangeetha M. Reddy<sup>8,9</sup>, Wendy A. Woodward<sup>10,11</sup>, James M. Reuben<sup>10,12</sup>, Joel Rozowsky<sup>1,3</sup>, Christos Hatzis<sup>2</sup>, Naoto T. Ueno<sup>9,10</sup>, Savitri Krishnamurthy<sup>10\*</sup>, Lajos Pusztai<sup>2\*</sup> and Mark Gerstein<sup>1,3,13,14\*</sup> 

## Abstract

**Background:** Inflammatory breast cancer (IBC) has a highly invasive and metastatic phenotype. However, little is known about its genetic drivers. To address this, we report the largest cohort of whole-genome sequencing (WGS) of IBC cases.

**Methods:** We performed WGS of 20 IBC samples and paired normal blood DNA to identify genomic alterations. For comparison, we used 23 matched non-IBC samples from the Cancer Genome Atlas Program (TCGA). We also validated our findings using WGS data from the International Cancer Genome Consortium (ICGC) and the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium. We examined a wide selection of genomic features to search for differences between IBC and conventional breast cancer. These include (i) somatic and germline single-nucleotide variants (SNVs), in both coding and non-coding regions; (ii) the mutational signature and the clonal architecture derived from these SNVs; (iii) copy number and structural variants (CNVs and SVs); and (iv) non-human sequence in the tumors (i.e., exogenous sequences of bacterial origin).

(Continued on next page)

\* Correspondence: [skrishna@mdanderson.org](mailto:skrishna@mdanderson.org); [lajos.pusztai@yale.edu](mailto:lajos.pusztai@yale.edu); [mark@gersteinlab.org](mailto:mark@gersteinlab.org)

<sup>10</sup>Morgan Welch Inflammatory Breast Cancer Research Program and Clinic, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>2</sup>Yale Cancer Center, Breast Medical Oncology, Yale School of Medicine, 300 George Street, Suite 120, Rm133, New Haven, CT 06511, USA

<sup>1</sup>Program in Computational Biology and Bioinformatics, Yale University, 266 Whitney Ave., Bass 432A, New Haven, CT 06520, USA

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Results:** Overall, IBC has similar genomic characteristics to non-IBC, including specific alterations, overall mutational load and signature, and tumor heterogeneity. In particular, we observed similar mutation frequencies between IBC and non-IBC, for each gene and most cancer-related pathways. Moreover, we found no exogenous sequences of infectious agents specific to IBC samples. Even though we could not find any strongly statistically distinguishing genomic features between the two groups, we did find some suggestive differences in IBC: (i) The *MAST2* gene was more frequently mutated (20% IBC vs. 0% non-IBC). (ii) The TGF  $\beta$  pathway was more frequently disrupted by germline SNVs (50% vs. 13%). (iii) Different copy number profiles were observed in several genomic regions harboring cancer genes. (iv) Complex SVs were more frequent. (v) The clonal architecture was simpler, suggesting more homogenous tumor-evolutionary lineages.

**Conclusions:** Whole-genome sequencing of IBC manifests a similar genomic architecture to non-IBC. We found no unique genomic alterations shared in just IBCs; however, subtle genomic differences were observed including germline alterations in TGF $\beta$  pathway genes and somatic mutations in the *MAST2* kinase that could represent potential therapeutic targets.

**Keywords:** Inflammatory breast cancer, Whole-genome sequencing, Single nucleotide variant, Copy number variant, Structural variant

## Background

Inflammatory breast cancer (IBC) is a rare form of breast cancer with very little known about its molecular etiology that is responsible for its aggressive clinical course. IBC accounts for 2–4% of all breast cancers in the USA [1] and causes 7–10% of breast cancer-related deaths in Western countries [2, 3]. IBC includes all known molecular subtypes of breast cancer, but they are considerably more aggressive than in non-IBC, with poorer disease-free survival and overall survival [2, 4]. The disease often presents with rapidly progressing symptoms of swelling of the breast, redness, and thickening of the skin of the breast which resembles an active inflammatory process, which led to the name of the disease. However, the symptoms are not caused by inflammatory cells, but by cancer cells blocking lymph vessels in the skin and breast parenchyma [5]. IBC also has a propensity for rapid dissemination and distant metastatic spread. Gene expression profiling studies have not revealed any consistent IBC-specific gene expression patterns; consequently, there is no molecular diagnostic test to define this disease [4, 6]. The diagnosis is based on the unique and rapidly progressive clinical features of the cancer. Targeted sequencing of ~ 200 cancer-related genes in IBC showed that the most frequently altered gene was TP53, with reported frequencies between 43 and 75% [7–9]. Currently, there is no whole-exome or whole-genome sequence data available for IBC and its DNA level alterations have not been characterized. We hypothesize that specific DNA sequence changes in the coding or non-coding regions of the genome may be responsible for the unique phenotype of IBC. The goal of this project was to perform deep characterization of the complete genomic features of IBC specimens to identify IBC-specific sequence alterations that could potentially explain its etiology and provide new diagnostic markers.

## Methods

### Tissues

Twenty IBC tissues and paired normal DNA from blood were obtained from the Morgan Welch Inflammatory Breast Cancer Research Program and Clinic at MD Anderson Cancer Center under an IRB-approved study. All IBC tissues were individually reviewed by a breast pathologist (Savitri K.) and a clinical investigator (N.U) for accuracy of diagnosis and to ensure tumor cellularity > 60%. All patients provided informed consent for genomic analysis of their cancer and germline DNA. Characteristics for 20 IBC patients are shown in Additional file 1: Table S1. Twenty-three non-IBC samples were selected from the Cancer Genome Atlas (TCGA) study of breast cancer cohort that were proportionally matched by molecular subtype, clinical stage, age, and race. This was done to ensure that various covariate distributions were similar between the IBC and the non-IBC samples studied in this project. Characteristics for those selected non-IBC samples are shown in Additional file 1: Table S1.

### DNA extraction

DNA was extracted from the snap frozen core needle biopsy of the breast tumor and peripheral blood using the QiAamp DNA Mini kit (Qiagen). The tissue was disrupted in buffer ATL, homogenized, and then lysed using Proteinase K. Buffer AL and ethanol was then added to the lysate creating conditions that promoted selective binding of the DNA to the QIAamp spin columns. The sample was then applied to the mini spin columns. The DNA bound to the membrane was eluted in buffer TE at pH 8.0.

### Whole-genome sequencing

One-microgram germline and tumor DNA were used for WGS that was performed under a Yale IRB-approved protocol (HIC #1406014226). It was performed on the Illumina HiSeq 2500 sequencing platform at Macrogen. The samples were prepared according to the Illumina TruSeq DNA library preparation guide. The 150-base pair (bp) paired-end libraries were sequenced with median coverage of 60X for the tumor samples and 40X for the matched normal samples. Detailed sequencing information, including sequencing depth, raw and mapped read numbers, and mapping rates, is summarized in Additional file 2: Table S2.

### Sequence alignment and quality control

We mapped raw FASTQ files for tumor and matched normal samples for 20 IBC samples to the hg19 reference genome using BWA-MEM [10] algorithm with default parameters. Subsequently, reads were sorted and duplicate reads were marked using Samtools [11] and Picard tools (<http://broadinstitute.github.io/picard>) to obtain the final set of BAM files for variant calling. The BAM files have been deposited in the European Genome-Phenome Archive (EGA) under EGA accession EGAS00001004117 (<https://wwwdev.ebi.ac.uk/ega/studies/EGAS00001004117>). For the non-IBC samples, we followed the same procedure. Details of sequencing for each sample are summarized in Additional file 2: Table S2.

### Germline SNV and INDELS calling

For both IBC and non-IBC cohorts from TCGA, we generated the germline SNVs and INDEL call set using the GATK tool [12]. Briefly, we followed the GATK best practice to call germline variants. We realigned the original bam using IndelRealigner and base recalibrator module in the GTAK. Subsequently, variants were called using GATK HaplotypeCaller algorithm. Raw variants were filtered using the variant recalibration module in the GATK. Briefly, the variant recalibration method uses a continuous adaptive error model, while taking into account of the relationship between variant and the probability of it being a true positive instead of a sequencing artifact.

### Somatic SNV and INDELS calling

We called somatic variants for IBC and non-IBC samples from TCGA using MuTect [13] and Strelka [14] tools. Briefly, these tools take tumor and matched normal bam files as input to identify somatic variants supported by minimum number of reads. Somatic SNVs in this study were based on MuTect and Strelka, whereas somatic INDELS were called using Strelka. The initial PASS only call set obtained from both MuTect and

Strelka were further filtered for potential germline contaminated call by removing common variants as defined in the 1000 Genomes Project [15]. Furthermore, we also removed somatic SNVs and INDELS falling outside the high mappability regions of the genome as defined by the Genome in a Bottle Consortium (GIAB) [16]. Finally, we took the intersection of MuTect and Strelka call sets and removed those somatic SNVs and INDELS that appeared in germline call set.

### Detection of loss of heterozygosity (LOH) in tumor DNA

For each site of germline SNV identified from a normal blood sample, we determined the corresponding somatic genotype in tumor DNA by using Samtools [11] and Bcftools [17], which reported information for reference allele, alternative allele, allele count, and allele frequency. An LOH event was identified when the site met these two criteria: (i) it was called as a heterozygous variant (alternative allele frequency = 0.5) in normal blood DNA and (ii) it was shown as homozygous (alternative allele frequency = 0 or 1) in tumor DNA.

### Somatic SV calling

We applied Meerkat [18] to identify somatic structural variants in the IBC and the non-IBC cohorts from TCGA. Briefly, Meerkat extract soft-clipped and unmapped reads from the bam files. These reads are subsequently remapped to the reference genome using BLAT [19] to identify discordant read pairs for SV discovery. Meerkat also characterizes breakpoint around the SVs to assign the underlying mechanism generating SVs. Meerkat generated SVs were further filtered based on mappability criterion and supporting read pairs > 2.

### Identification of somatic CNVs

We implemented BIC-Seq2 [20] to call somatic CNVs using default parameters. In the SeqNorm step, we set read length to 151 bp and bin size to 1000 bp. The fragment size was calculated using the first one million properly mapped reads with mapping quality at least 20 in the BAM files.

We also used a signal processing approach for filtering the somatic copy number segments (sCNSs) identified by BIC-Seq2 in the last step. In this analysis, we focused on large-scale events affecting > 100 kB in length. Below are the specific procedures:

1. For each sample, compute the read depth (RD) signal levels using the mapped reads. This is done by counting the number of reads that overlap with each base. For each patient, we computed RD signal for tumor and the matching normal tissue.
2. Next we normalized normal tissue profiles using reads per million normalization. Given  $i^{\text{th}}$  sample's

tumor and normal signal profiles, we multiplied the normal signal profile with the ratio of total RD signal in tumor and total RD signal in normal.

$$\widehat{RD}_n(i) = RD_n(i) \times \frac{\sum RD_t(j)}{\sum RD_n(j)}$$

$\widehat{RD}_n(i)$  denotes the normalized normal RD signal at  $i^{\text{th}}$  base position in the genome.

3. We next divided the genome into 3000 bp bins and computed the total tumor signal and normal sample's normalized RD signal in each bin.
4. We next computed the log ratio (LR) profiles by dividing the total tumor RD signal by normal sample RD signal in each bin and computing the  $\log_2$  of this ratio. This profile represents a measure of the deletions (LR < 0) and amplifications (LR > 0)

$$LR(b) = \log_2 \frac{\sum_i RD_t(i)}{\sum_{i \in c_b} \widehat{RD}_n(i)}$$

where  $c_b = [(b-1) \cdot l_{bin}, b \cdot l_{bin}]$  represents the base positions for  $b^{\text{th}}$  bin.

5. The LR profile is generally extremely noisy. We use median-based smoothing to smooth the signal. We use a sliding window approach where window size is set to 1000 bins and replace the LR value at each bin with the median of the LR values within the 1000 bins' vicinity. The smoothing operation removes substantial amount of noise from the LR signal

$$LR(b) = \text{median}(LR(b-l_{win}), \dots, LR(b+l_{win}))$$

6. Next, we identify sCNS by evaluating the regions where smoothed LR is constant. On each sCNS, we assign the tumor-to-normal log ratio signal by computing the ratio of total tumor to total normal RD signal. The segments with LR < 0 are assigned as deletions and segments with LR > 0 amplifications. We denote the LR value for segment  $s$  on sample  $k$  with  $LR_s^k$ . Finally, we only took the strongest calls from BIC-Seq2, after filtering by the signal processing approach introduced above, as the final call set. Copy number gain was defined as  $\log_2$  (tumor/expected) ratio > 0.2. Copy number loss was defined as  $\log_2$  (tumor/expected) ratio < -0.2.

### Functional annotation and impact prediction

Both somatic and germline SNVs were annotated by FunSeq2 [21]. Since non-coding variants in regulatory elements (promoter, enhancer, etc.) can be associated with potential target genes, this pipeline helps to identify both coding and non-coding variants of a given gene. Additionally, functional impact of each variant was predicted by PredictSNP2 [22], which could be neutral, deleterious, or unknown. Only deleterious (high functional impact) variants were selected for gene and pathway-level analysis.

### Identification of candidate driver genes

Candidate driver genes in IBC cohort were detected by ActiveDriverWGS [23] with default parameters. Final call sets for all somatic SNVs in IBC samples were used as input. Coordinates of genes were extracted from Ensemble database under hg19 reference genome, using biomaRt package [24]. Genes with FDR < 0.05 were identified as candidate drivers.

### Mutation spectra and mutational signatures

Somatic SNVs across the whole genome were analyzed in single nucleotide and tri-nucleotide context, respectively. DeconstructSig [25] was used to deconstruct the mutation spectrum (96 possible tri-nucleotide combinations) of each sample into 30 reference mutational signatures in the COSMIC database [26], in order to calculate the weight of each reference signature.

### Estimation of number of clones

We implemented SciClone [27] to estimate the number of clones for each IBC and non-IBC sample. First, all somatic SNVs with allele frequency higher than 0.6 were removed from the input file as they were likely affected by copy number loss events. Next, function "sciClone" was called with "minimumDepth" set as 14 and "clusterMethod" set as "binomial.bmm". Finally, the output of the function reported the predicted number of clones detected in the given sample. After repeating the above procedures for all IBC and non-IBC samples, the predicted numbers of clones were compared between two cohorts by Fisher's exact test.

### Evolutionary trees build-up using PhyloWGS

We used PhyloWGS [28] to infer the evolutionary trees for each individual sample. We followed a similar workflow as previously described [29]. Somatic SNVs from the consensus calls of Strelka and Mutect were used. The observed alternative allele and reference allele counts were from Strelka. To remove copy number effects, we removed SNVs in the regions with an absolute "log2.copyRatio" ( $\log_2$  tumor to normal copy number ratio, reported by BIC-Seq2) higher than 0.2 and  $p$ -value

lower than 0.01. Then, we ran PhyloWGS [28] using default parameters and set genders all to female. We only plotted and analyzed the tree with the highest likelihood reported by PhyloWGS.

#### Estimation of tumor purity

In order to estimate the tumor purity for each IBC tumor sample, a computational pipeline called PurBayes [30] was implemented with default parameters. Estimation results were summarized in Additional file 3: Table S3.

#### Microorganism sequences' detection and enrichment

In order to identify sequences that are potentially of exogenous origin and not arising of the host genome, we modified a portion of the *exceRpt* pipeline that was developed for the identification of endogenous and exogenous extracellular RNAs [31]. After reads are aligned to the host genome, we performed a second pass alignment against the host genome in order to remove sequences that might potentially come from the host human genome. We then removed reads that align with a high number of mismatches (5 mismatches per 100 bp). We also filter out reads that align against repetitive sequences in the human genome and reads that multi-map up 200 locations in the human genome. While we cannot confidently assign these reads to the human genome, the goal is to filter them out in order to obtain a set of reads that we are confident that do not come from the host human genome.

These reads are then aligned against indices for a set of full genomes for all sequenced bacteria, viruses, plants, fungi, protist, metazoa, and the following 12 vertebrate genomes: chicken, cod, cow, dog, duck, frog, horse, rabbit, pig, sheep, tilapia, and turkey. Since many exogenous genomes have a high degree of sequence similarity based on evolution, we find that many reads that align to an exogenous genome align to multiple genomes. By default, the pipeline allows for no mismatches during this step (in order to be as conservative as possible in identifying possible exogenous sequences). We assign reads that align to exogenous genomes to the position in the phylogenetic taxonomy tree based on the node that is most parsimonious with the different genomes that the read aligns.

#### Validation cohorts

In order to validate key genomic findings identified from IBC cohort, we expanded our analysis to multiple other cohorts, including breast cancer and other types of cancers, as well as general population. More specifically, we investigated high-functional impact mutation frequencies of genes and pathways in (1) PCAWG breast cancer cohort [32], (2) twenty-three types of primary cancers from ICGC (<https://dcc.icgc.org/>), and (3) general

population data from The Genome Aggregation Database [33].

#### Statistical analysis

There are two types of statistical testing methods used in this study: (1) Wilcoxon rank-sum test and (2) Fisher's exact test. Wilcoxon test was implemented when comparing median of IBC and non-IBC samples. Fisher exact test was used when comparing fraction of IBC and non-IBC samples in each category. All original *p*-values from above tests were adjusted by the very conservative Bonferroni correction. Adjusted *p*-values < 0.05 were considered statistically significant. Additionally, we implemented randomization test to validate the statistical significances by three steps: (1) mix IBC and non-IBC samples and randomly assign them to two groups, (2) test the significance under new sample labeling, and (3) Repeat the analysis for 1000 times and summarize the statistics. All statistical analysis was performed by R software (<https://www.r-project.org/>). R packages *ggplot2* (<http://ggplot2.org>), *ComplexHeatmap* [34], and *RCircos* [35] were used to visualize the results.

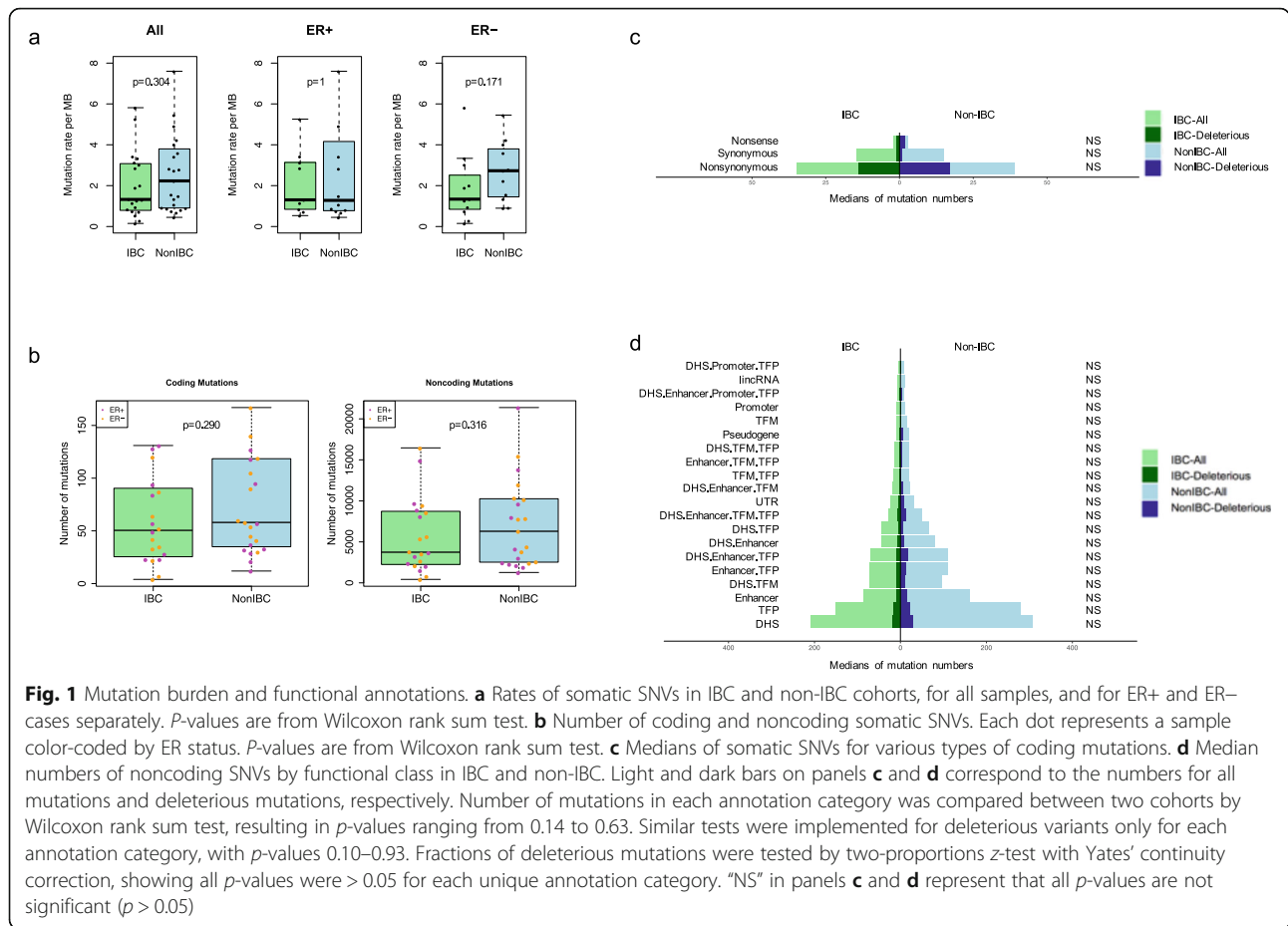
## Results

#### Somatic mutation burden and functional annotations

WGS identified 114,563 somatic SNVs in 20 IBC samples (range 424–16,662 per tumor; median 3789), among which 1282 variants (1.12%) were in coding regions. IBC and non-IBC showed similar mutation rate per megabase (MB) (Fig. 1a). The number of somatic coding and noncoding SNVs were similar between the IBC and non-IBC cohorts (Fig. 1b). Noncoding somatic SNVs were annotated with FunSeq2 [21] into 20 different, non-overlapping functional categories. The number of somatic SNVs within each annotation category was similar between the two cohorts (Fig. 1c, d).

#### Mutation spectra and mutational signatures

IBC has similar proportions of base changes as non-IBC, for all single-nucleotide mutation contexts (C>A, C>G, C>T, T>A, T>C and T>G) (Additional file 4: Fig. S1a) (Wilcoxon test, adjusted *p*-values > 0.05 by Bonferroni method), as well as tri-nucleotide mutation contexts (Additional file 4: Fig. S1b) (Wilcoxon test, adjusted *p*-values > 0.05 by Bonferroni method). The mutation spectrum of each sample was deconstructed using DeconstructSig [25] into 30 reference mutational signatures in the COSMIC database [26]. IBC and non-IBC samples showed no difference in mutational signature distribution (Additional file 4: Fig. S1c) (Wilcoxon test, adjusted *p*-values > 0.05 by Bonferroni method). In particular, there was no difference in signature 3, which has been associated with homologous recombination defect (HRD) in breast cancer [36]. Statistical comparison of



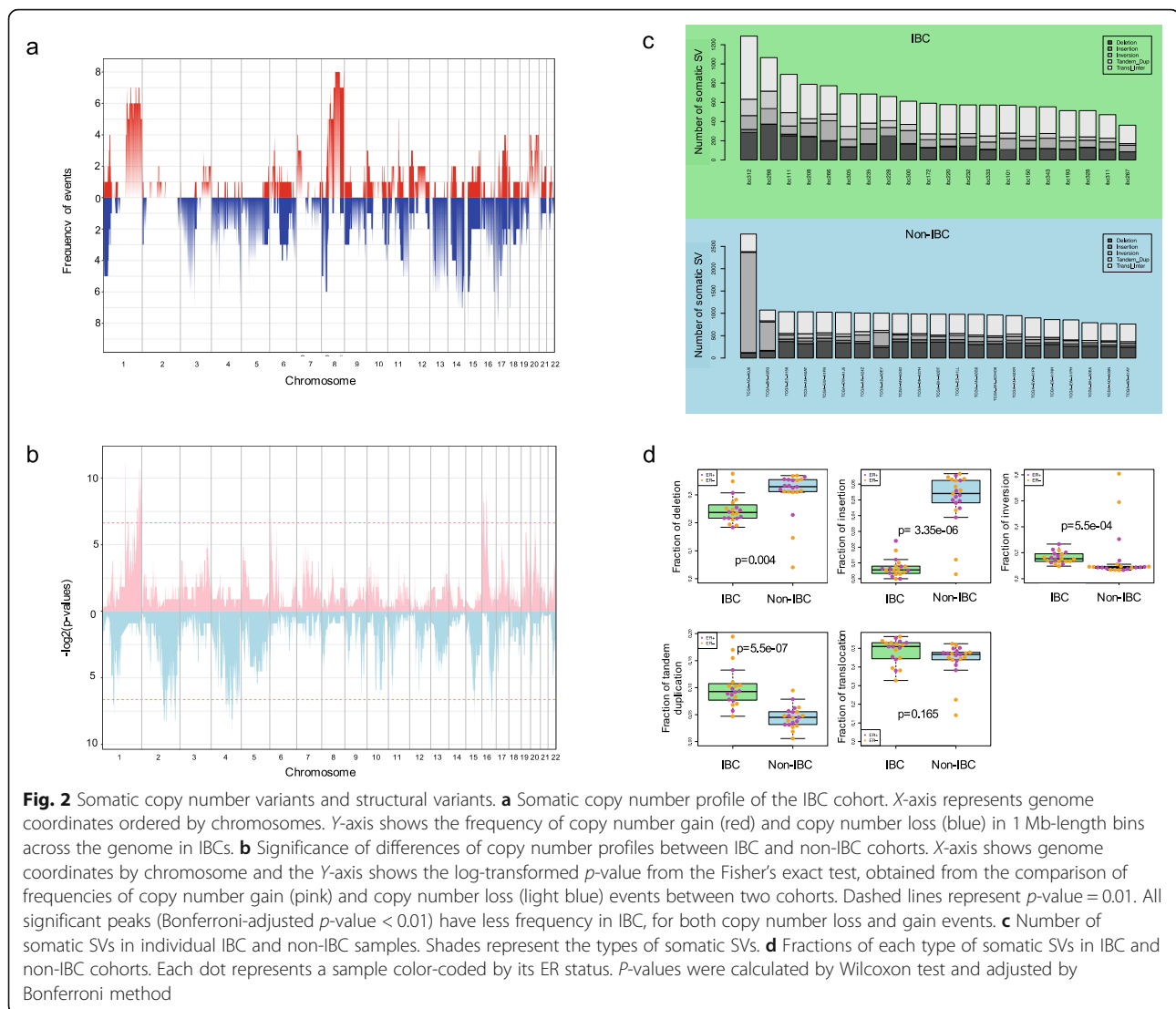
weights of signature 3 indicates that IBC samples have similar degree of HRD to non-IBC ones (Additional file 4: Fig. S1d) (Wilcoxon test, *p* = 0.85).

**Copy number variants and structural variants**

Copy number loss or gain events were mapped into 1-MB-sized bins across the entire genome (Fig. 2a, Additional file 4: Fig. S2). For each bin, frequencies of copy number loss or gain events were summarized separately and then compared between IBC and non-IBC cohorts. For copy number gain events (defined as log<sub>2</sub> (observed tumor/expected) ratio > 0.2), 108 peaks showed significantly different frequencies between two cohorts, locating at chromosome 1, 3, 6, 16, 17, 19, and 20 (Additional file 5: Table S4) (Fisher's exact test, *p*-values < 0.05), in contrast to 34 significant peaks reported by the randomization test (median, 34; minimum, 16; maximum, 62). On the other hand, for copy number loss events (defined as log<sub>2</sub> (observed tumor/expected) ratio < - 0.2), 221 peaks showed significantly different frequencies between two cohorts, locating at chromosome 1, 2, 4, 5, 9, 10, 11, 12, 15, 16, and 17 (Additional file 5: Table S4), in contrast to 57 significant peaks reported by the randomization test (median: 57; minimum: 38;

maximum: 80). There were 26 cancer-related genes involved in these differentially affected genomic regions (Additional file 6: Table S5), including *LRP1B* as a putative tumor suppressor gene, and *ERBB4* as a member in the EGFR subfamily of receptor tyrosine kinases.

Large structural variants were classified into five categories: deletion, insertion, inversion, tandem duplication, and inter-chromosomal translocation (Fig. 2c). The fraction of large somatic SVs in each category was compared between IBC and non-IBC cohorts. IBC showed significantly higher fraction of complex events than non-IBC, including tandem duplications (median 0.093 vs. 0.045) (Wilcoxon test, Bonferroni adjusted *p* = 5.5e-07) and inversions (median 0.154 vs. 0.088) (Wilcoxon test, Bonferroni adjusted *p* = 5.5e-04) (Fig. 2d). On the contrary, IBC showed significantly lower fraction of large deletions (Wilcoxon test, Bonferroni adjusted *p* = 0.004) and insertions (Wilcoxon test, *p* = 3.4e-06), compared with non-IBC samples. Additionally, comparison of the absolute numbers of somatic SVs in each category also presented significant differences (Additional file 4: Fig. S3). For small insertions and deletions (INDELs), both categories of mutations reported similar numbers between IBC and non-IBC cohorts (Additional file 4: Fig.



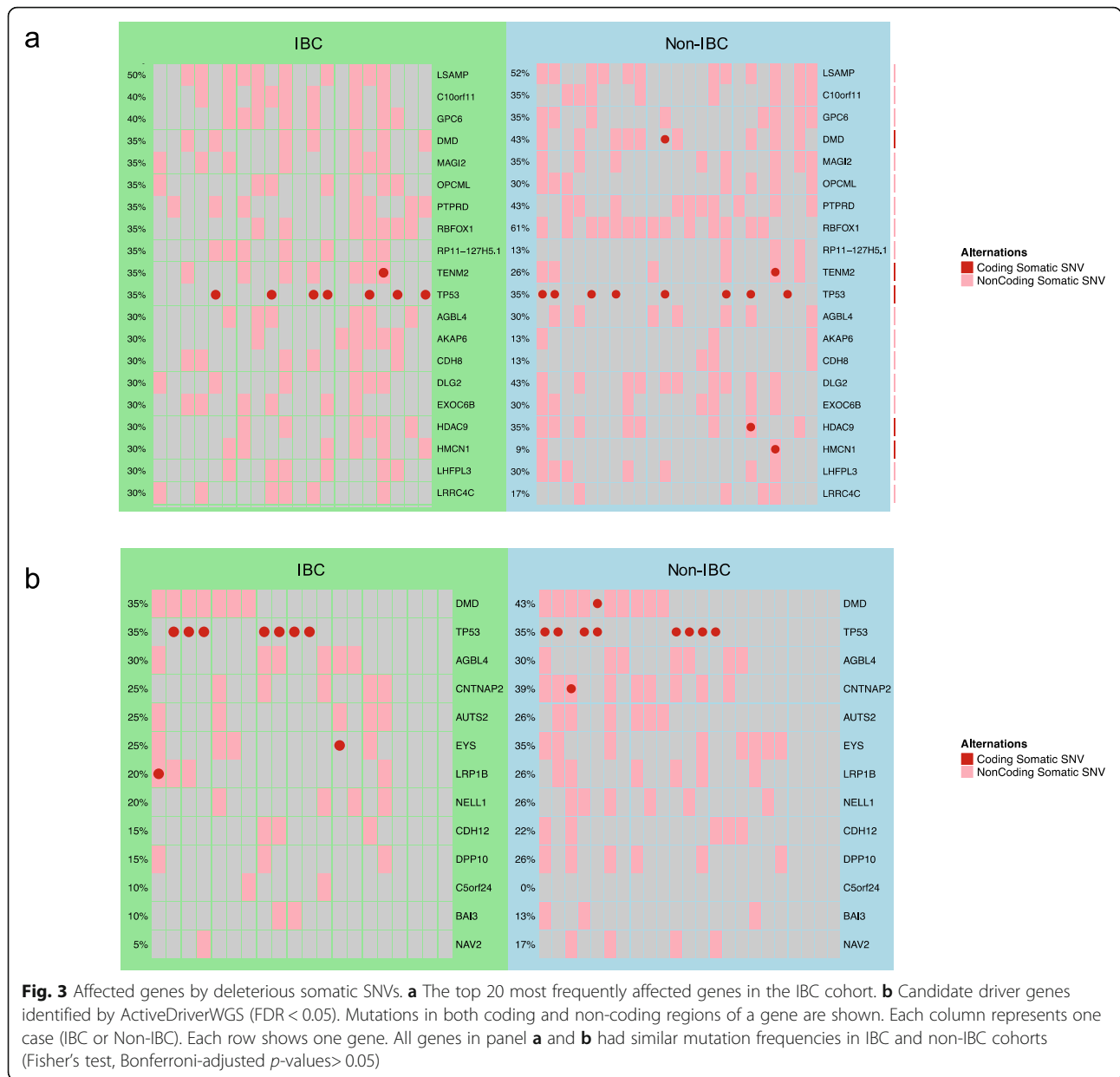
S4). Circos plots summarizing the combined germline and somatic genetic aberrations detected from WGS for each individual IBC are shown in Additional file 4: Fig. S5.

#### High functional impact mutations and affected genes

High-functional impact (HFI) somatic SNVs were selected based on the deleteriousness predictions reported by PredictSNP2 [22], including both coding and noncoding SNVs. Affected genes were then extracted for each sample, and their mutation frequencies were compared between two cohorts. Our analysis showed that all genes with at least one high-functional impact somatic SNVs in the IBC cohort were similarly affected in the IBC and non-IBC cohorts (Fisher's test, Bonferroni adjusted  $p$ -values > 0.05). For IBC, the top 20 genes most frequently affected by deleterious somatic SNVs included *LSAMP*, *GPC6*, and *TP53* among others (Fig. 3a). The top 20 most frequently affected genes by coding and non-

coding deleterious somatic SNVs were summarized in Additional file 4: Fig. S6a and Additional file 4: Fig. S6b, respectively. Additionally, thirteen candidate driver genes were detected with ActiveDriverWGS in the IBC cohort (FDR < 0.05) (Fig. 3b). However, all of them showed similar mutation frequencies between the IBC and the non-IBC cohorts (Fisher's test, Bonferroni adjusted  $p$ -values > 0.05) (Fig. 3b).

Notably, we identified four of 20 (20%) IBCs had unique predicted deleterious mutations in the non-coding (promoter and intron) region of *MAST2* (Microtubule-Associated Serine/Threonine-Protein Kinase 2), while no deleterious mutation was detected in any of the 23 non-IBC cases in our cohort. In the PCAWG [32] breast cancer cohort, we could find only 1 out of 198 samples (0.5%) having a mutation in this gene. In the largest WGS study for breast cancer (BRCA-EU from The ICGC Breast Cancer Project), the mutation



frequency of *MAST2* was 1/569 (0.18%) [37], which was significantly lower than our IBC cohort (Fisher's test, Bonferroni adjusted  $p = 0.024$ ). We also determined the frequency of high-functional impact mutations in the *MAST2* gene in the ICGC Data Portal (<https://dcc.icgc.org/>), for 22 different primary cancer sites. We found that the two highest mutation frequencies were in thyroid cancer (3/50 = 6%) and nasopharyngeal cancer (1/21 = 4.76%), while the frequencies were < 2% in all other cancer types.

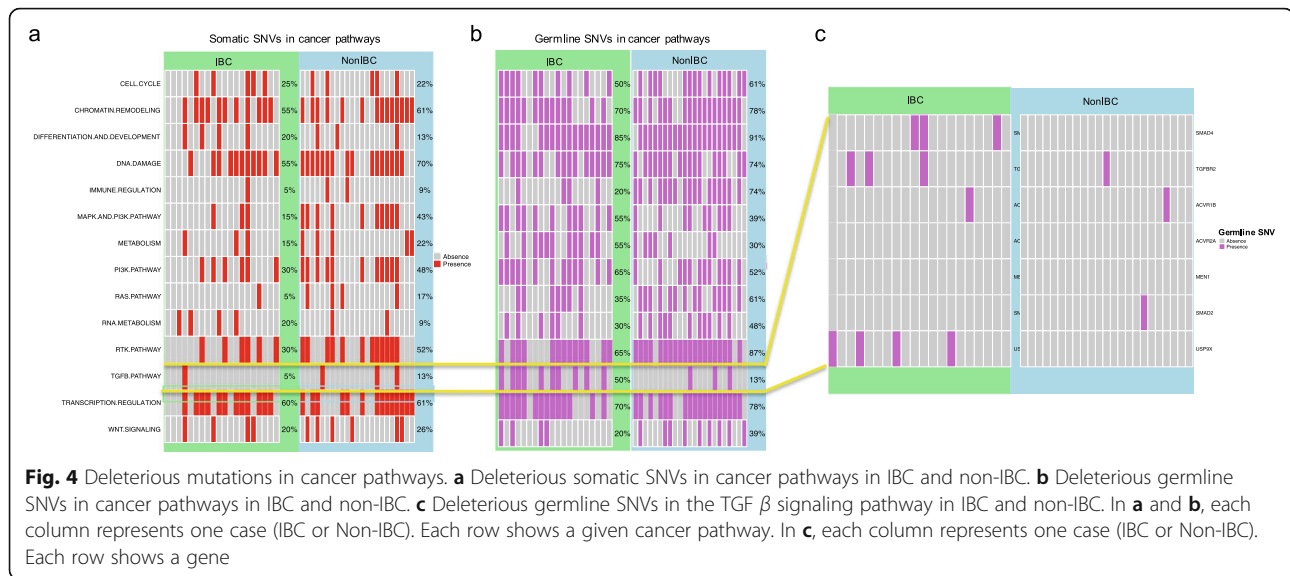
**Alterations of cancer-related signaling pathways**

We investigated pathway-level aberrations in 14 cancer-related biological pathways [38]. For somatic SNVs, none

of these pathways had significantly different mutation frequencies between the IBC and non-IBC cohorts (Fisher's test, Bonferroni adjusted  $p > 0.05$ ) (Fig. 4 (a)). For germline SNVs, the IBC cohort showed a significantly lower frequency of aberrations in the immune regulation pathway than non-IBC (Fisher's test, Bonferroni adjusted  $p = 0.009$ ) (Fig. 4 (b)). Randomization test ( $N = 10,000$ ) showed that the probability of observing a significant difference in the immune regulation pathway was 21/10,000 = 0.0021, which is significant.

Previous studies have identified the TGF  $\beta$  pathway as a potential therapeutic target in IBC [39, 40]. In this study, we observed a numerically higher (but not statistically significant after correction for multiple testing)





mutation frequency of predicted deleterious germline SNVs in the TGF  $\beta$  pathway in IBCs (50% vs. 13%, Fisher's test, Bonferroni adjusted  $p = 0.25$ ) (Fig. 4 (b)). Seven IBC cases (35%) had deleterious germline SNVs in either *SMAD4* or *USP9X*, both involved in TGF $\beta$  signaling; one of these cases had variants in both the coding and non-coding regions and the rest of cases only had non-coding variants (Additional file 7: Table S6). However, none of the non-IBC cases had any deleterious germline SNVs in these two genes in the coding or non-coding regions (Fig. 4 (c)). We did not observe any LOH event at the corresponding genomic locations in the tumor DNA (Additional file 7: Table S6). Notably, none of these deleterious germline SNVs was detected in the ICGC breast cancer cohort ( $n = 1970$ ), and they are also very rare in the general population, with variant allele frequency (VAF)  $< 0.007$  in The Genome Aggregation Database ( $n = 141,456$ ) [33] (Additional file 7: Table S6).

#### Clonal architecture and evolutionary trees

IBC had similar mutant-allele tumor heterogeneity (MATH) [41] as non-IBC cases (Fig. 5a). For each sample, the number of clones was estimated by SciClone [27], based on the model fitting procedures on the distribution of variant allele frequencies (Additional file 4: Fig. S7 and Additional file 4: Fig. S8). The results revealed that 6/20 (30%) IBC cancers were clonal (consisting of only one clone), whereas all non-IBC cases had at least two clones (Fisher's test,  $p = 0.006$ ) (Fig. 5b). We then constructed evolutionary trees for each case to further explore the mutational process heterogeneity as previously described [29] (Additional file 4: Fig. S9). These trees were derived from the whole-genome mutation calls, with their topology suggesting a temporal ordering to the mutations. We could classify the trees into

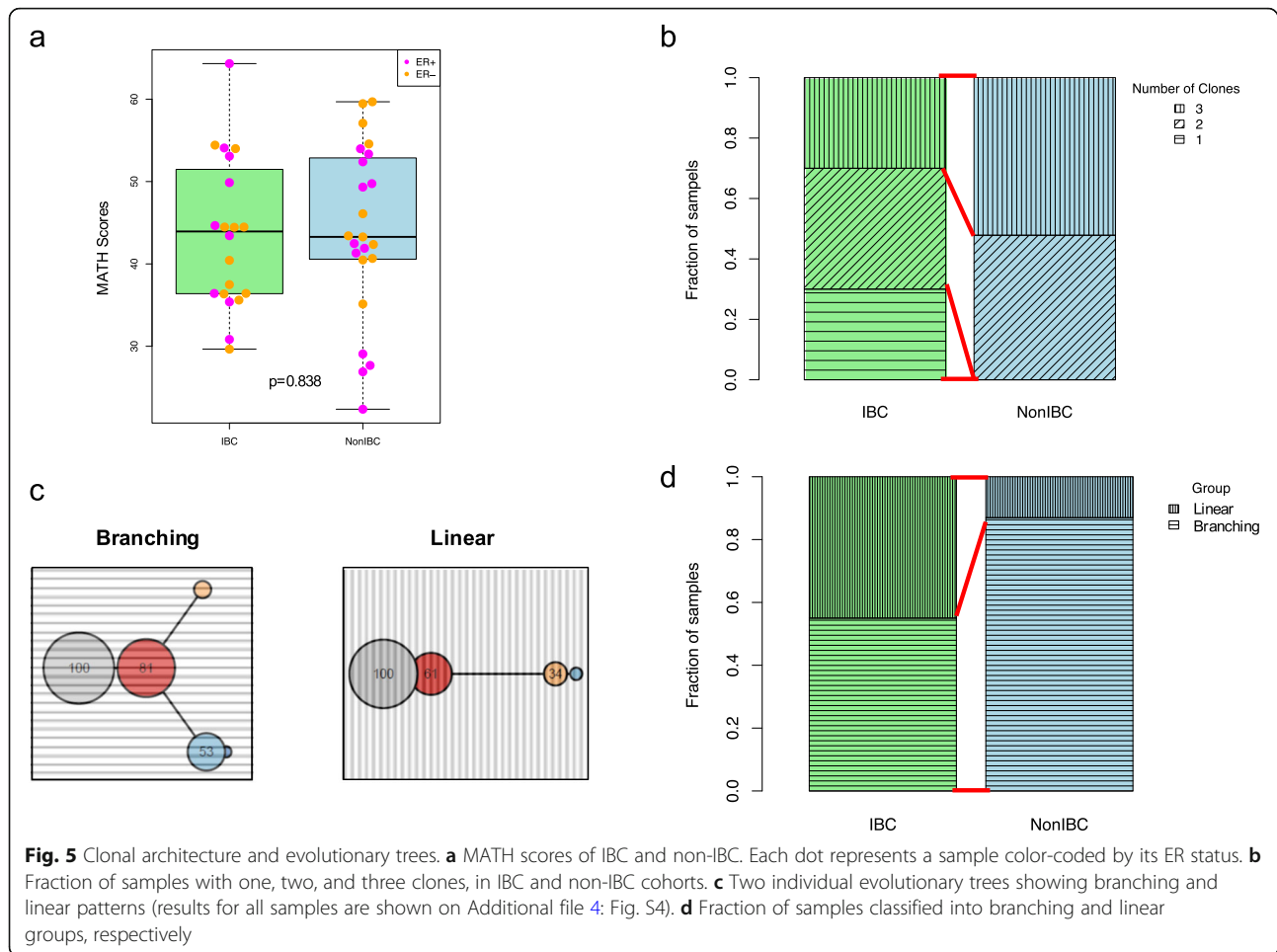
two groups based on their topology: branching or linear (Fig. 5c). Nine out of 20 (45%) of IBC cancers were linear, which is significantly more than non-IBC cases (3/23, 13%) (Fisher's test,  $p = 0.039$ ) (Fig. 5d and Additional file 4: Fig. S9). Our results illustrated that IBC is evolutionarily more homogeneous than non-IBC, with less clonality and less complex evolutionary features. These findings may result from the faster growth of IBC tumors, compared with non-IBC ones.

#### Microorganism sequences' detection and enrichment

As IBC clinically mimics bacterial infection of the breast and/or skin [42], therefore, we also looked for infectious agents in the IBC tumor tissues. We applied a modified exceRpt pipeline [31] to examine the sequence reads that did not map to the human reference genome, in order to detect microorganism sequences in the DNA of IBC and matching normal samples. The top 100 most frequent microorganism sequences in all samples including cancer and normal were highly enriched in *Propionibacterium acnes* (Additional file 4: Fig. S10). *P. acnes* is a ubiquitous skin bacterium that represents a common source of contamination in sequencing studies that can originate from the patient or acquired during tissue handling [43]. We found no infectious agent DNA specific to the IBC cancer samples, which diminished the possibility that IBC was caused by bacterial infection.

#### Discussion

Our study is the first WGS analysis of IBC. We could not identify a single genomic abnormality that is shared by all samples and therefore could molecularly define IBC. IBC tissues showed similar mutation load, mutational spectra, and mutation signatures as non-IBC, and most somatic mutations occurred at similar frequencies



in both cohorts. We did not detect any cancer-specific infectious agents in the DNA extracted from IBC tissues. However, we did identify several subtle genomic differences that distinguished IBC from non-IBC in our cohort. The non-coding region of the *MAST2* gene was mutated at a higher frequency than what reported in any previous WGS analysis of breast cancer. In our IBC cohort, 20% of the cases had a mutation, while the mutation frequency of this gene in coding or non-coding regulatory regions is between 0.18 and 0.5% in non-IBC cases in the PCAWG and ICGC breast cancer WGS datasets. *MAST2* is a microtubule-associated serine/threonine kinase that interacts with the Protocadherin-LKC, a recently proposed tumor suppressor gene for colon and liver cancers, which mediates contact inhibition of cell proliferation [44]. *MAST2* also regulates lipopolysaccharide-induced IL-12 synthesis in macrophages by forming a complex with TRAF6 and inhibiting NF-kappa-B activation [45]. *MAST2* gene rearrangements were previously noted in some breast cancers, and overexpression of *MAST2* (or *MAST1*) gene fusions in breast epithelial cells led to increased proliferation

in vitro and in vivo [46]. In our study, we found deleterious mutations in the non-coding regions of *MAST2*; however, the functional impact of these variants has not yet been investigated. Since *MAST2* has not been included in any of the previous targeted sequencing studies of IBC, future datasets of IBC will be needed to validate this finding.

Complex structural variants also appeared to be more common in IBC, including tandem duplication and inversion, suggesting greater genome complexity than in non-IBC. Several genomic regions showed significantly different copy number profiles harboring genes involved in cancer biology (Additional file 6: Table S5). However, sequencing platforms with different coverages and depths could introduce bias when calling large structural variants. As our IBC and non-IBC cohorts were sequenced separately, some observed differences may arise from the different sources of sequencing data.

A surprising finding of our study was the low clonality of IBC at the time of diagnosis. A substantial minority of IBC had only one detectable clone whereas all non-IBC cases had more than one clones. When we examined the

evolutionary trees of the tumor cell populations, we observed two distinct groups that we describe as branching and linear evolution. IBCs showed significantly more of the linear evolutionary pattern than non-IBC (45% vs. 13%,  $p = 0.039$ ). These results suggest that IBC cells are evolutionarily more homogeneous and exhibit lower clonality than non-IBC cancer cells, leading to the hypothesis that a high proliferation rate and rapid expansion of a single aggressive clone could be responsible for the rapid initial clinical course of the disease, which often unfolds in a few weeks. In comparison, non-IBC often grow for years before becoming detectable, which may allow for the development of greater clonal heterogeneity at the time of diagnosis. Future work will be needed to validate these observations, via high-depth targeted sequencing and subsequent characterization of subclonal entities.

We also examined the host genome for germline variants that might be associated with IBC. Currently no genetic predisposing factors are known for IBC, but some familiar occurrences have been reported and IBC is more prevalent in certain geographic regions that suggest genetic contribution to its etiology [47, 48]. We identified heterozygous germline alterations in the TGF $\beta$  pathway that appear to be more frequent in IBC than in non-IBC (50% vs. 13%). Due to the rarity of IBC (0.5–2% of all breast cancers), our sample size is very small and this observation will need to be confirmed in larger independent IBC datasets. However, TGF $\beta$  has been implicated in the biology of IBC. *USP9X*, affected in 4 out of 20 IBC cases by a germline variant, is a deubiquitinating enzyme that controls *SMAD4* mono-ubiquitination and therefore affects TGF $\beta$  signaling [49]. One previous study showed that the expression of TGF $\beta$  signaling pathway components are lower in IBC compared to non-IBC, and this may contribute to tumor emboli formation and facilitate lymphatic invasion of IBC cells [40]. Another study on head and neck cancer reported that loss of *SMAD4* was associated with increased TGF $\beta$  1 activity [50]. Overall, these results suggest the possibility of aberrant host TGF $\beta$  signaling contributing to IBC biology and predisposition.

We recognize that our results are descriptive and hypothesis generating in terms of biological importance of the findings. However, it is clear from our analysis that there is no shared DNA level pathognomonic alteration in IBC. The sample size of our study is small; nevertheless, it is the largest study so far to examine the whole genome of IBC. Previous genomic analyses included only a few hundreds of genes that were sequenced using targeted sequencing platforms (Additional file 8: Table S7) [7–9]. We observed lower mutation frequencies in *PIK3CA* in our IBC cohort than previous ones, which may be due to sampling bias arising from the small

sample size, as *PIK3CA* was more frequently mutated in hormone receptor-positive (HR+) cancers, and our study had a lower fraction of HR+ cases [51].

Besides various genetic features discussed in this study, it has been shown that IBC is significantly different with non-IBC in several non-genetic factors, including lower prevalence of parous women, higher oral contraceptive use, and higher frequency of regular alcohol consumption [52].

Overall, our results suggest that IBC falls within the continuum of breast cancer in terms of its molecular make up. Its particularly aggressive phenotype may result from unique co-occurrence of heterozygous host germline polymorphisms with subtle effects on TGF $\beta$  signaling and somatic mutations that together enable rapid growth and expansion of a malignant cell clone.

## Conclusions

Here we present the first complete genomic landscape of IBC by whole-genome sequencing of tumor and their matched normal samples. Even though there was no unique, shared genomic alteration in IBCs, we identified several subtle but intriguing genomic differences between IBC and non-IBC which could potentially explain its etiology and result in new diagnostic markers, but will require validation in independent datasets in future studies.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00879-x>.

**Additional file 1: Table S1.** Characteristics of IBC and non-IBC patients.

**Additional file 2: Table S2.** Technical details for sequencing.

**Additional file 3: Table S3.** Tumor purity of IBC samples.

**Additional file 4: Fig. S1-S10.** Supplementary Fig. S1-S10.

**Additional file 5: Table S4.** Significant peaks with somatic CNVs.

**Additional file 6: Table S5.** Cancer genes located in significant peaks with somatic CNVs.

**Additional file 7: Table S6.** Germline SNVs in TGF $\beta$  pathway.

**Additional file 8: Table S7.** Summary of previously published IBC studies.

## Acknowledgements

Not applicable

## Authors' contributions

MG, LP, Savitri K., NU, and CH contributed to the conception and design of the study. XL, SK, AH, SL, RK, YZ, VW, SR, WW, James R., and Joel R. contributed to the acquisition, analysis, or interpretation of data. All authors wrote, reviewed, and/or revised the manuscript. All authors read and approved the final manuscript.

## Funding

Research reported in this publication was supported by a SWOG HOPE Foundation Pilot Grant (L.P. and M.G.), NIH grant R01 HG 008126 (M.G.), grants from the Breast Cancer Research Foundation (L.P., C.H., and S.R.), and a State of Texas Rare and Aggressive Breast Cancer Research Program Grant

(S.R., W.A.W., J.M.R., N.T.U., and Savitri K.). S.R. also received supports from a National Institutes of Health T32 Training Grant (T32 CA 009666), Conquer Cancer Foundation American Society of Clinical Oncology Young Investigator Award, and Cancer Prevention and Research Institute of Texas Award (RR190020). W.A.W. also received supports from Genomic Health, Inc.

#### Availability of data and materials

The datasets generated and analyzed during the current study have been submitted to the European Genome-Phenome Archive (EGA) under accession number EGAS00001004117 (<https://www.ebi.ac.uk/ega/studies/EGAS00001004117>) [53].

#### Declarations

##### Ethics approval and consent to participate

The study was approved by the Yale Cancer Center Human Investigations Committee (HIC #1406014226). The research conformed to the principles of the Helsinki Declaration. All patients provided written informed consent for genomic analysis of their cancer and germline DNA as well as participation in the study.

##### Consent for publication

All participants provided consent for publication.

##### Competing interests

L.P. has received consulting fees and honoraria from Astra Zeneca, Merck, Novartis, Bristol-Myers Squibb Genentech, Eisai, Pieris, Immunomedics, Seattle Genetics, Clovis, Syndax, H3Bio, and Daiichi. C.H. is currently a full-time employee of Bristol-Myers Squibb. V.W. is currently a full-time employee of Janssen Pharmaceutical Companies of Johnson & Johnson. The remaining authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Program in Computational Biology and Bioinformatics, Yale University, 266 Whitney Ave., Bass 432A, New Haven, CT 06520, USA. <sup>2</sup>Yale Cancer Center, Breast Medical Oncology, Yale School of Medicine, 300 George Street, Suite 120, Rm133, New Haven, CT 06511, USA. <sup>3</sup>Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave., Bass 432A, New Haven, CT 06520, USA. <sup>4</sup>Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center Houston, Houston, TX, USA. <sup>5</sup>Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. <sup>6</sup>Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, USA. <sup>7</sup>The Ohio State University Comprehensive Cancer Center (OSUCCC – James), Columbus, OH, USA. <sup>8</sup>Division of Hematology/Oncology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>9</sup>Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>10</sup>Morgan Welch Inflammatory Breast Cancer Research Program and Clinic, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>11</sup>Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>12</sup>Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>13</sup>Department of Computer Science, Yale University, 266 Whitney Ave., Bass 432A, New Haven, CT 06520, USA. <sup>14</sup>Department of Statistics and Data Science, Yale University, 266 Whitney Ave., Bass 432A, New Haven, CT 06520, USA.

Received: 9 September 2020 Accepted: 25 March 2021

Published online: 26 April 2021

#### References

- Chang S, Parker SL, Pham T, Buzdar AU, Hursting SD. Inflammatory breast carcinoma incidence and survival: the surveillance, epidemiology, and end results program of the National Cancer Institute, 1975–1992. *Cancer*. 1998; 82(12):2366–72. [https://doi.org/10.1002/\(SICI\)1097-0142\(19980615\)82:12<2366::AID-CNCR10>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0142(19980615)82:12<2366::AID-CNCR10>3.0.CO;2-N).
- Lim B, Woodward WA, Wang XP, Reuben JM, Ueno NT. Inflammatory breast cancer biology: the tumour microenvironment is key. *Nat Rev Cancer*. 2018; 18(8):485–99. <https://doi.org/10.1038/s41568-018-0010-y>.
- Hance KW, Anderson WF, Devesa SS, Young HA, Levine PH. Trends in inflammatory breast carcinoma incidence and survival: the Surveillance, Epidemiology, and End Results Program at the National Cancer Institute. *Jnci-J Natl Cancer I*. 2005;97(13):966–75. <https://doi.org/10.1093/jnci/dji172>.
- Masuda H, Baggerly KA, Wang Y, Iwamoto T, Brewer T, Pusztai L, Kai K, Kogawa T, Finetti P, Birnbaum D, Dirix L, Woodward WA, Reuben JM, Krishnamurthy S, Symmans WF, van Laere SJ, Bertucci F, Hortobagyi GN, Ueno NT. Comparison of molecular subtype distribution in triple-negative inflammatory and non-inflammatory breast cancers. *Breast Cancer Res*. 2013; 15(6):R112. <https://doi.org/10.1186/bcr3579>.
- Fouad TM, Barrera AMG, Reuben JM, Lucci A, Woodward WA, Stauder MC, Lim B, DeSnyder SM, Arun B, Gildy B, Valero V, Hortobagyi GN, Ueno NT. Inflammatory breast cancer: a proposed conceptual shift in the UICC-AJCC TNM staging system. *Lancet Oncol*. 2017;18(4):E228–E32. [https://doi.org/10.1016/S1470-2045\(17\)30192-4](https://doi.org/10.1016/S1470-2045(17)30192-4).
- Iwamoto T, Bianchini G, Qi Y, Cristofanilli M, Lucci A, Woodward WA, Reuben JM, Matsuoka J, Gong Y, Krishnamurthy S, Valero V, Hortobagyi GN, Robertson F, Symmans WF, Pusztai L, Ueno NT. Different gene expressions are associated with the different molecular subtypes of inflammatory breast cancer. *Breast Cancer Res Treat*. 2011;125(3):785–95. <https://doi.org/10.1007/s10549-010-1280-6>.
- Liang X, Vacher S, Boulai A, Bernard V, Baulande S, Bohec M, Bièche I, Lerebours F, Callens C. Targeted next-generation sequencing identifies clinically relevant somatic mutations in a large cohort of inflammatory breast cancer. *Breast Cancer Res*. 2018;20(1):88. <https://doi.org/10.1186/s13058-018-1007-x>.
- Matsuda N, Lim B, Wang Y, Krishnamurthy S, Woodward W, Alvarez RH, Lucci A, Valero V, Reuben JM, Meric-Bernstam F, Ueno NT. Identification of frequent somatic mutations in inflammatory breast cancer. *Breast Cancer Res Treat*. 2017;163(2):263–72. <https://doi.org/10.1007/s10549-017-4165-0>.
- Ross JS, Ali SM, Wang K, Khaira D, Palma NA, Chmielecki J, Palmer GA, Morosini D, Elvin JA, Fernandez SV, Miller VA, Stephens PJ, Cristofanilli M. Comprehensive genomic profiling of inflammatory breast cancer cases reveals a high frequency of clinically relevant genomic alterations. *Breast Cancer Res Treat*. 2015;154(1):155–62. <https://doi.org/10.1007/s10549-015-3592-z>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–9. <https://doi.org/10.1038/nbt.2514>.
- Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK, Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811–7. <https://doi.org/10.1093/bioinformatics/bts271>.
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74.
- Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol*. 2014;32(3):246–51. <https://doi.org/10.1038/nbt.2835>.
- Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93. <https://doi.org/10.1093/bioinformatics/btq509>.
- Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, Zhang C, Ren X, Protopopov A, Chin L, Kucherlapati R, Lee C, Park PJ. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell*. 2013;153(4):919–29. <https://doi.org/10.1016/j.cell.2013.04.010>.
- Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002;12(4):656–64. <https://doi.org/10.1101/gr.229202>.
- Xi R, Lee S, Xia Y, Kim TM, Park PJ. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility

- variants. *Nucleic Acids Res.* 2016;44(13):6274–86. <https://doi.org/10.1093/nar/gkw491>.
21. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, Khurana E, Gerstein M. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* 2014;15(10):480. <https://doi.org/10.1186/s13059-014-0480-5>.
  22. Bendl J, Musil M, Stourac J, Zendlulka J, Damborsky J, Brezovsky J. PredictSNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS Comput Biol.* 2016;12(5):e1004962. <https://doi.org/10.1371/journal.pcbi.1004962>.
  23. Zhu H, Uuskula-Reimand L, Isaev K, Wadi L, Alizada A, Shuai S, et al. Candidate cancer driver mutations in distal regulatory elements and long-range chromatin interaction networks. *Mol Cell.* 2020;77(6):1307–21 e10. <https://doi.org/10.1016/j.molcel.2019.12.027>.
  24. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nat Protoc.* 2009;4(8):1184–91. <https://doi.org/10.1038/nprot.2009.97>.
  25. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 2016;17(1):31. <https://doi.org/10.1186/s13059-016-0893-4>.
  26. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, de T, Teague JW, Stratton MR, McDermott U, Campbell PJ. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 2015;43(Database issue):D805–11. <https://doi.org/10.1093/nar/gku1075>.
  27. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, Vij R, Tomasson MH, Graubert TA, Walter MJ, Ellis MJ, Schierding W, DiPersio JF, Ley TJ, Mardis ER, Wilson RK, Ding L. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol.* 2014;10(8):e1003665. <https://doi.org/10.1371/journal.pcbi.1003665>.
  28. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* 2015;16(1):35. <https://doi.org/10.1186/s13059-015-0602-8>.
  29. Li S, Shuch BM, Gerstein MB. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. *PLoS Genet.* 2017;13(3):e1006685. <https://doi.org/10.1371/journal.pgen.1006685>.
  30. Larson NB, Fridley BL. PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. *Bioinformatics.* 2013;29(15):1888–9. <https://doi.org/10.1093/bioinformatics/btt293>.
  31. Rozowsky J, Kitchen RR, Park JJ, Galeev TR, Diao J, Warrell J, Thistlethwaite W, Subramanian SL, Milosavljevic A, Gerstein M. exceRpt: a comprehensive analytic platform for extracellular RNA profiling. *Cell Syst.* 2019;8(4):352–7 e3. <https://doi.org/10.1016/j.cels.2019.03.004>.
  32. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature.* 2020;578(7793):82–93. <https://doi.org/10.1038/s41586-020-1969-6>.
  33. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016; 536(7616):285–91. <https://doi.org/10.1038/nature19057>.
  34. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–9. <https://doi.org/10.1093/bioinformatics/btw313>.
  35. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics.* 2013;14(1):244. <https://doi.org/10.1186/1471-2105-14-244>.
  36. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, Rosebrock D, Litvitz D, Kübler K, Mouw KW, Kamburov A, Maruvka YE, Leshchiner I, Lander ES, Golub TR, Zick A, Orthwein A, Lawrence MS, Batra RN, Caldas C, Haber DA, Laird PW, Shen H, Ellisen LW, D'Andrea AD, Chanock SJ, Foulkes WD, Getz G. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet.* 2017;49(10):1476–86. <https://doi.org/10.1038/ng.3934>.
  37. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GJK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivrot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, Veer L, Tutt A, Knappskog S, Tan BKT, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, van Laere S, Lakhani SR, Eyfjord CE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* 2016;534(7605):47–54. <https://doi.org/10.1038/nature17676>.
  38. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* 2017;355(6322).
  39. Cohen EN, Gao H, Anfossi S, Mego M, Reddy NG, Debeb B, Giordano A, Tin S, Wu Q, Garza RJ, Cristofanilli M, Mani SA, Croix DA, Ueno NT, Woodward WA, Luthra R, Krishnamurthy S, Reuben JM. Inflammation mediated metastasis: immune induced epithelial-to-mesenchymal transition in inflammatory breast cancer cells. *PLoS One.* 2015;10(7):e0132710. <https://doi.org/10.1371/journal.pone.0132710>.
  40. Van Laere SJ, Ueno NT, Finetti P, Vermeulen P, Lucci A, Robertson FM, et al. Uncovering the molecular secrets of inflammatory breast cancer biology: an integrated analysis of three distinct affymetrix gene expression datasets. *Clin Cancer Res.* 2013;19(17):4685–96. <https://doi.org/10.1158/1078-0432.CCR-12-2549>.
  41. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol.* 2013;49(3):211–5. <https://doi.org/10.1016/j.oraloncology.2012.09.007>.
  42. Yaghoobi R, Talaizade A, Lal K, Ranjbari N, Sohrabian N, Feily A. Inflammatory breast carcinoma presenting with two different patterns of cutaneous metastases: carcinoma telangiectaticum and carcinoma erysipeloïdes. *J Clin Aesthet Dermatol.* 2015;8(8):47–51.
  43. Møllerup S, Friis-Nielsen J, Vinner L, Hansen TA, Richter SR, Fridholm H, Herrera JAR, Lund O, Brunak S, Izzarugaza JMG, Mourier T, Nielsen LP, Hansen AJ. Propionibacterium acnes: disease-causing agent or common contaminant? Detection in diverse patient samples by next-generation sequencing. *J Clin Microbiol.* 2016;54(4):980–7. <https://doi.org/10.1128/JCM.02723-15>.
  44. Okazaki N, Takahashi N, Kojima S, Masuho Y, Koga H. Protocadherin LKC, a new candidate for a tumor suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. *Carcinogenesis.* 2002; 23(7):1139–48. <https://doi.org/10.1093/carcin/23.7.1139>.
  45. Xiong H, Li H, Chen Y, Zhao J, Unkeless JC. Interaction of TRAF6 with MAST205 regulates NF-kappaB activation and MAST205 stability. *J Biol Chem.* 2004;279(42):43675–83. <https://doi.org/10.1074/jbc.M404328200>.
  46. Robinson DR, Kalyana-Sundaram S, Wu YM, Shankar S, Cao X, Ateeq B, Asangani IA, Iyer M, Maher CA, Grasso CS, Lonigro RJ, Quist M, Siddiqui J, Mehra R, Jing X, Giordano TJ, Sabel MS, Kleer CG, Palanisamy N, Natrajan R, Lambros MB, Reis-Filho JS, Kumar-Sinha C, Chinnaiyan AM. Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med.* 2011;17(12):1646–51. <https://doi.org/10.1038/nm.2580>.
  47. Aziz SA, Pervez S, Khan S, Kayani N, Azam SI, Rahbar MH. Case control study of prognostic markers and disease outcome in inflammatory carcinoma breast: a unique clinical experience. *Breast J.* 2001;7(6):398–404. <https://doi.org/10.1046/j.1524-4741.2001.07604.x>.
  48. Zucali R, Uslenghi C, Kenda R, Bonadonna G. Natural history and survival of inoperable breast cancer treated with radiotherapy and radiotherapy followed by radical mastectomy. *Cancer.* 1976;37(3):1422–31. [https://doi.org/10.1002/1097-0142\(197603\)37:3<1422::AID-CNCR2820370325>3.0.CO;2-Y](https://doi.org/10.1002/1097-0142(197603)37:3<1422::AID-CNCR2820370325>3.0.CO;2-Y).
  49. Zhang J, Zhang X, Xie F, Zhang Z, van Dam H, Zhang L, Zhou F. The regulation of TGF-beta/SMAD signaling by protein deubiquitination. *Protein Cell.* 2014;5(7):503–17. <https://doi.org/10.1007/s13238-014-0058-8>.
  50. Bornstein S, White R, Malkoski S, Oka M, Han GW, Cleaver T, Reh D, Andersen P, Gross N, Olson S, Deng C, Lu SL, Wang XJ. Smad4 loss in mice causes spontaneous head and neck cancer with increased genomic instability and inflammation. *J Clin Invest.* 2009;119(11):3408–19. <https://doi.org/10.1172/JCI38854>.
  51. Stemke-Hale K, Gonzalez-Angulo AM, Lluh A, Neve RM, Kuo WL, Davies M, Carey M, Hu Z, Guan Y, Sahin A, Symmans WF, Pusztai L, Nolden LK, Horlings H, Berns K, Hung MC, van de Vijver MJ, Valero V, Gray JW, Bernards R, Mills GB, Hennessey BT. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* 2008;68(15):6084–91. <https://doi.org/10.1158/0008-5472.CAN-07-6854>.

52. Moslehi R, Freedman E, Zeinomar N, Veneroso C, Levine PH. Importance of hereditary and selected environmental risk factors in the etiology of inflammatory breast cancer: a case-comparison study. *BMC Cancer*. 2016; 16(1):334. <https://doi.org/10.1186/s12885-016-2369-z>.
53. Li X, Kumar S, Harmanci A, Li S, Kitchen RR, Zhang Y, et al. Whole-genome sequencing of phenotypically distinct inflammatory breast cancers reveals similar genomic alterations to the more commonplace non-inflammatory breast cancers. *Eur Genome-Phenome Archive*. EGAS00001004117. Available from: <https://wwwdev.ebi.ac.uk/ega/studies/EGAS00001004117>.

### Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

