Published in final edited form as:
J Pathol. 2012 May ; 227(1): 29–41. doi:10.1002/path.4003.

A whole-genome massively parallel sequencing analysis of *BRCA1* mutant oestrogen receptor negative and positive breast cancers

Rachael Natrajan[#1], **Alan Mackay**[#1], **Maryou B Lambros**[#1], **Britta Weigelt**[2], **Paul M Wilkerson**[1], **Elodie Manie**[3], **Anita Grigoriadis**[4], **Roger A'Hern**[5], **Petra van der Groep**[6], **Iwanka Kozarewa**[1], **Tatiana Popova**[3], **Odette Mariani**[7], **Samra Turaljic**[8], **Simon J Furney**[8], **Richard Marais**[8], **Daniel-Nava Rodruigues**[1], **Adriana C Flora**[1], **Patty Wai**[1], **Vidya Pawar**[1], **Simon McDade**[9], **Jason Carroll**[10], **Dominique Stoppa-Lyonnet**[3,7], **Andrew R Green**[11], **Ian O Ellis**[11], **Charles Swanton**[12,13], **Paul van Diest**[6], **Olivier Delattre**[3], **Christopher J Lord**[1], **William D Foulkes**[14], **Anne Vincent-Salomon**[3,7], **Alan Ashworth**[1], **Marc Henri Stern**[3,7], and **Jorge S Reis-Filho**[1]

[1]The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, SW3 6JB, UK [2]Signal Transduction Laboratory, Cancer Research UK London Research Institute, WC2A 3LY, UK [3]Institut Curie, INSERM U830, 75248 Paris, France [4]Breakthrough Research Unit, Bermondsey Wing, Guy's Hospital, London, SE1 9RT, UK [5]CRUK Clinical Trials Unit, The Institute of Cancer Research, Sutton, SM2 5NG, UK [6]University Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands [7]Institut Curie, Department of Tumour Biology, 75248 Paris, France [8]Signal Transduction Team, Division of Cell and Molecular Biology, The Institute of Cancer Research, London, SW3 6JB, UK [9]Centre for Cancer Research and Cell Biology, Queen's University, Belfast, BT9 7BL, Northern Ireland, UK [10]Nuclear Receptor Transcription Laboratory, Cancer Research UK Cambridge Research Institute, Cambridge, CB2 0RE, UK [11]Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham and Nottingham University Hospitals Trust, Nottingham, NG7 2UH, UK [12]Translational Cancer Therapeutics Laboratory, Cancer Research UK London Research Institute, WC2A 3LY, UK [13]UCL Cancer Institute, Huntley Street, London WC1E 6DD, UK [14]Program in Cancer Genetics, Departments of Human Genetics and Oncology, McGill University, Montreal, QC, H2W 1S6, Canada

[#] These authors contributed equally to this work.

Abstract

Corresponding author: Professor Jorge S Reis-Filho, The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, SW3 6JB, UK, Jorge.Reis-Filho@icr.ac.uk.

Author contributions: BW, AV-S, AA and JSR-F supervised the research. RN, AM, CJL, AV-S, AA, MHS and JSR-F conceived and designed the experiments. RN, MBL, BW, PMW, DNR, ACF, PW and VP performed the experiments. AM, AG, TP and RA performed statistical analyses. RN, AM, MBL, BW, EM, AG, TP, ST, SF, SM, JC, PvD, WDF, AV-S and JSR-F analysed the data. PvdG, IK, DM, RM, JC, DS-L, ARG, IOE, PvD, OD, CJL, AV-S, MHS and JSR-F contributed reagents, materials and/or analysis tools. The manuscript was written by RN, BW, PMW, WDF, CS, AV-S and JSR-F. All authors read and approved the manuscript.

Conflict of interest statement: The authors have no conflict of interest to declare.

Microarray comparative genomic hybridisation data, the analysis history, script and code are available at http://rock.icr.ac.uk/collaborations/Mackay/BRCA1.Complete.Genomics

BRCA1 encodes a tumour suppressor protein that plays pivotal roles in homologous recombination (HR) DNA repair, cell-cycle checkpoints, and transcriptional regulation. *BRCA1* germline mutations confer a high risk of early-onset breast and ovarian cancer. In >80% of cases, tumours arising in *BRCA1* germline mutation carriers are oestrogen receptor (ER)-negative, however up to 15% are ER-positive. It has been suggested that *BRCA1* ER-positive breast cancers constitute sporadic cancers arising in the context of a *BRCA1* germline mutation rather than being causally related to BRCA1 loss-of-function. Whole-genome massively parallel sequencing of ER-positive and ER-negative *BRCA1* breast cancers, and their respective germline DNAs, was used to characterise the genetic landscape of *BRCA1* cancers at base-pair resolution. Only *BRCA1* germline mutations and somatic loss of the wild-type allele, and *TP53* somatic mutations were recurrently found in the index cases. *BRCA1* breast cancers displayed a mutational signature consistent with that caused by lack of HR DNA repair in both ER-positive and ER-negative cases. Sequencing analysis of independent cohorts of hereditary *BRCA1* and sporadic non-*BRCA1* breast cancers for the presence of recurrent pathogenic mutations and/or homozygous deletions found in the index cases revealed that *DAPK3*, *TMEM135*, *KIAA1797*, *PDE4D* and *GATA4* are potential additional drivers of breast cancers. This study demonstrates that *BRCA1* pathogenic germline mutations coupled with somatic loss of the wild-type allele are not sufficient for hereditary breast cancers to display an ER-negative phenotype, and has led to the identification of three potential novel breast cancer genes (i.e. *DAPK3*, *TMEM135* and *GATA4*).

Keywords

BRCA1; breast cancer; next generation sequencing; mutation; tumour suppressor genes

Introduction

BRCA1 germline mutations are one of the main causes of hereditary early-onset breast and ovarian cancer syndrome[1]. *BRCA1* (17q21) encodes a protein with diverse biological functions in homologous recombination (HR) DNA repair, cell cycle checkpoints, DNA decatenation, and transcriptional regulation[1,2].

Breast cancers arising in patients with germline *BRCA1* mutations (*BRCA1* breast cancers) have characteristic histological and molecular features. Contrary to sporadic breast cancers, which predominantly express oestrogen receptor (ER), >80% of cases of *BRCA1* breast cancers lack ER expression. Furthermore, the majority of *BRCA1* breast cancers display a triple-negative phenotype (i.e. lack of ER, progesterone receptor (PR) and HER2), a basal-like 'intrinsic' subtype[3,4] and *TP53* somatic mutations[4]. Recent studies have provided circumstantial evidence to suggest that BRCA1 directly modulates ER expression in breast cancers and that BRCA1 loss-of-function would result in ER-negative breast cancers[5,6]. Furthermore, there is evidence to suggest that BRCA1 may play a role in the differentiation status of breast stem cells and that BRCA1 breast cancers originate from ER-negative luminal progenitor cells[7,8].

The enrichment for tumours with a triple-negative phenotype in *BRCA1* germline mutation carriers have led to the generalisation that BRCA1 tumours are of ER-negative status. At

least 15% of all breast cancers arising in *BRCA1* germline mutation carriers, however, do express ER. Some argue that these cancers may constitute sporadic ER-positive cancers developing in patients with *BRCA1* germline mutations and not causally linked to BRCA1 loss-of-function[9], or that they might be pathological and molecular "intermediates" between ER-negative *BRCA1* cancers and ER-positive sporadic breast cancers, developing through distinct evolutionary pathways[10].

The predisposition to breast and ovarian cancer development caused by *BRCA1* mutations has been linked to BRCA1's role in HR DNA repair[1,11]. Deficiency in HR leads to the use of error-prone DNA repair mechanisms (e.g. non-homologous end-joining) to correct DNA double-strand breaks, resulting in the accumulation of genetic aberrations. These error-prone DNA repair mechanisms leave mutation signatures in the genome of cancer cells, including sequences of micro-homology around structural rearrangement breakpoints[11–14].

Massively parallel sequencing (MPS) has revealed the complexity of sporadic breast cancer genomes[15–17]. Although ER-positive and ER-negative breast cancers had previously been shown to differ in terms of gene copy number alterations[18,19], the repertoire of somatic mutations, structural and copy number aberrations of sporadic breast cancers of the same phenotype appears to be rather diverse[15,18–20]. *BRCA1* breast cancers offer an alternative approach to study the genetic heterogeneity of breast cancers in the presence of known driver events, given that these tumours by definition harbour a *BRCA1* germline mutation and in the vast majority of cases, display somatic loss of the *BRCA1* wild-type allele and *TP53* somatic mutations[21].

The aim of this study was to characterise the genomic landscape of ER-positive and ER-negative *BRCA1* cancers, where the *BRCA1* wild-type allele was somatically lost and the cancer cells harboured *TP53* somatic mutations. These data were subsequently used i) to address whether ER-positive *BRCA1* breast cancers harbour a pattern of genetic aberrations consistent with the mutation signature found in cells with defects in HR, and ii) as a basis for the identification of novel breast cancer genes.

Material and Methods

Samples

Representative frozen samples and matched peripheral blood lymphocytes from one ER-negative (BRCA1/ER-BC) and one ER-positive (BRCA1/ER+BC), lymph node negative, invasive ductal carcinomas developing in patients with *BRCA1* germline mutations, in which *TP53* somatic mutations had already been identified, were retrieved from the tissue bank of Institut Curie, Paris. This study was approved by the local research ethics committees and specific patient consent was obtained from both patients. The histological features of each case were reviewed by two pathologists (AV-S, JSR-F). Expression of ER, PR, and HER2 were defined as previously described[22] and reviewed by two pathologists (AV-S, JSR-F). DNA and RNA were extracted from frozen sections after gross dissection of the frozen blocks to ensure that the samples contained >60% of tumour cells (Supplementary Methods S1). Germline and somatic *BRCA1* and *TP53* mutation profiles were defined as

previously described[22] (Supplementary Methods S1). Patient characteristics and tumour phenotypes are summarised in Table 1.

Massively parallel sequencing (MPS)

BRCA1/ER-BC and its matched germline DNA, and BRCA1/ER+BC and its matched germline DNA were subjected to MPS using the unchained combinatorial probe anchor ligation chemistry on arrays of self-assembling DNA nanoballs method (Complete Genomics, http://www.completegenomics.com/) as previously described[23,24]. Analysis of the data generated by MPS was performed essentially as described by Lee et al.[24]. The methods for the characterisation of mutation rates, ratios of transitions to transversions, the frequency of base-pair alterations surrounding somatic single nucleotide variations (SNVs), and the identification of somatic mutations in protein coding genes are described in the Supplementary Methods S1 [23–25] (Supplementary Table S1). *BRCA1* allelotyping and Sanger sequencing validation of the aberrations identified by MPS are described in the Supplementary Methods S1.

To address the issue of differences in tumour cell content between BRCA1/ER-BC and BRCA1/ER+BC, sequencing coverage and sequencing errors, somatic score thresholds (SSTs) (Supplementary Methods S1) were chosen to identify mutations with a true-positive rate of 90% and a false-positive rate of 20% for somatic SNVs (Supplementary Figure S1). Identification of in-dels was shown to be less accurate than the identification of SNVs and SSTs were chosen to minimise the false positive rate to 20% for both tumours.

The presence of regions of sequences of microhomology surrounding deletions and structural variants (SVs) was investigated as described by Stephens et al.[26] (Supplementary Methods S1).

SNP6 and microarray-based gene expression profiling

Tumour DNA was subjected to SNP (single nucleotide polymorphism) analysis using Affymetrix SNP6 arrays, and RNA samples were subjected to gene expression profiling analysis using the Affymetrix U133Plus2 chip as previously described[27,28] (Supplementary Methods S1). Genome Alteration Print (GAP) was employed to determine the tumour cell content as described previously[27]. Nearest centroid classification of tumours into the ‘intrinsic’ molecular subtypes was performed using the unfiltered MAS5 normalised data and based upon single sample predictors as previously described[29–31]. Microarray data, the analysis history, script and code are available at http://rock.icr.ac.uk/collaborations/Mackay/BRCA1.Complete.Genomics (Supplementary Methods S1).

Additional *BRCA1* and sporadic breast cancers

Sixteen additional *BRCA1* breast cancers (12 ER-negative and 4 ER-positive) were retrieved from the pathology files of the University Medical Center Utrecht, Utrecht, The Netherlands, and Institut Curie, Paris, France. Sixteen non-*BRCA1* breast cancers (12 ER-negative and 4 ER-positive) were retrieved from The University Hospital La Paz, Madrid, Spain (Supplementary Table S2) [18]. DNA was extracted from samples with >60% of tumour cells as defined by histological assessment carried out by two pathologists (JSR-F and AV-

Europe PMC Funders Author Manuscripts

S). To define whether any potentially disease-causing/pathogenic mutations identified in the two index cases would constitute recurrent genetic events in breast cancer, the coding region of 11 genes was subjected to Sanger sequencing as previously described[32] (Supplementary Methods S1) in these 16 *BRCA1* and 16 ER-matched sporadic breast cancers. Primers are detailed in Supplementary Table S3.

Analysis of a publicly available microarray-based comparative genomic hybridisation (aCGH) dataset

An aCGH dataset was retrieved from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus archive (GEO), comprising *BRCA1* mutant, non-*BRCA1* familial and sporadic breast cancers (GSE22133)[20]. Details of the bioinformatic analysis of this dataset is described in the Supplementary Methods S1 [20,33–36]. Amplifications identified in BRCA1/ER-BC and BRCA1/ER+BC were assessed in the *BRCA1* mutant breast cancer datasets. Non-synonymous coding variants, homozygous deletions (HODs) and genes disrupted by SVs in BRCA1/ER-BC and BRCA1/ER+BC were interrogated for the presence of HODs.

Identification of potential novel cancer genes operative in breast cancer

To determine whether the recurrent alterations identified to be deleterious by one or more prediction algorithms (i.e. SIFT[37], polyPhen[38], Condel[39] and CanPredict[40]), disrupted by HODs or by structural rearrangements would be potential novel cancer genes operative in breast cancer, we first established if they are known cancer genes according to the Cancer Gene Consensus (www.sanger.ac.uk/genetics/CGP/Census/). Next, we defined if these genes are described as mutated in breast cancer or other cancer types as registered in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic/), or in other high-throughput breast cancer sequencing data published previously[16,17,26,41–43]. If a gene was identified as recurrently altered in this study, not described as a known cancer gene and not previously described in breast cancer, it was defined as a potential novel breast cancer gene.

Results

BRCA1 breast cancers have complex landscapes of genetic aberrations regardless of the ER status

MPS using the unchained combinatorial probe-anchor ligation chemistry on arrays of self-assembling DNA nanoballs method[23,24] of tumour and germline DNA from BRCA1/ER-BC and BRCA1/ER+BC revealed multiple somatic alterations, including single nucleotide variants (SNVs), small insertions or deletions (in-dels), structural variations (SVs), and copy number alterations (Table 1, Figure 1, Supplementary Figure S2).

By employing SSTs to allow detection of SNVs with similar accuracy in both tumours, 12,184 and 5,513 somatic SNVs were identified in BRCA1/ER-BC and BRCA1/ER+BC, respectively (Supplementary Table S4). A direct comparison of the in-dels between the two samples is not possible, given the different accuracies for the detection of these aberrations in the BRCA1/ER-BC and BRCA1/ER+BC (Supplementary Methods S1).

Using the SSTs determined above for the SNVs, we calculated a genome-wide point mutational rate per Mb of 4.26 in BRCA1/ER-BC and 1.93 in BRCA1/ER+BC (Figure 2). Both tumours expressed a similar number of genes (12,499 and 12,845 in BRCA1/ER-BC and BRCA1/ER+BC, respectively; Supplementary Methods S1). BRCA1/ER-BC and BRCA1/ER+BC harboured a higher number of mutations in intergenic compared to genic regions (1.29 in BRCA1/ER-BC, 1.17 in BRCA1/ER+BC), and in non-expressed compared to expressed genes (1.43 in BRCA1/ER-BC, 1.16 in BRCA1/ER+BC), consistent with active transcription-coupled DNA repair in both cancers[44]. The ratios of mutations in intergenic:genic and non-expressed:expressed genes were, however, significantly higher in BRCA1/ER-BC than in BRCA1/ER+BC ($p=0.00431$ and <0.00001, respectively, Fisher's exact test, Figure 2), which may result from an increased level of genomic instability, mutation rate or high proliferation found in BRCA1/ER-BC compared to BRCA1/ER+BC, as expected in an ER-negative breast cancer[18–20,45]. Theoretically, in a scenario where most stochastic mutations affecting genes are deleterious or neutral, mutations affecting intergenic regions or genes not expressed would have a higher probability of being maintained in the cancer cell population than mutations affecting genic regions or expressed genes that may be deleterious[46]. Therefore, relatively higher number of mutations in intergenic regions and non-expressed genes than in genic regions and expressed genes would be expected in tumours with relatively higher levels of genetic instability.

Based on the analysis of high-confidence SNVs found in each *BRCA1* mutant genome, an increase in the number of transversions relative to the number of transitions was observed (Figures 3A and 3B), akin to previous cancer genomes subjected to whole-genome sequencing analysis[24,47]. In addition, C>G/G>C transversions, as defined on the basis of the analysis of the positive DNA strand, were significantly more frequently preceded and/or succeeded by a thymine base than expected by chance, an observation not previously described in *BRCA1* breast cancers (Figures 3C and 3D, Supplementary Figure S3, Chi-square LOD score >20).

Ninety-three and 51 somatic SNVs were identified above the defined SSTs in BRCA1/ER-BC and BRCA1/ER+BC, respectively. Of these, 62 (67%) and 28 (55%) were confirmed by Sanger sequencing of tumour and germline samples as somatic and not present in the germline DNA in BRCA1/ER-BC and BRCA1/ER+BC, respectively. Forty-one and 20 somatic in-dels were identified above the defined SSTs in BRCA1/ER-BC and BRCA1/ER +BC, respectively. Of these, 9 (22%) and 1/14 (5%) were confirmed as somatic and not detectable in the germline DNA by Sanger sequencing in BRCA1/ER-BC and BRCA1/ER +BC, respectively. These analyses led to the identification and validation of 71 and 29 coding somatic mutations in BRCA1/ER-BC and BRCA1/ER+BC, respectively. These figures are in agreement to those reported for a sporadic triple-negative breast cancer[16] and an ER-positive invasive lobular carcinoma of the breast[17], which harboured 50 and 32 validated non-synonymous coding mutations, respectively, but are higher than the figures previously observed in medulloblastoma, pancreatic cancers, glioblastoma and colorectal cancers[48].

BRCA1/ER-BC and BRCA1/ER+BC harboured similar numbers of gene copy number aberrations (Supplementary Figure S4, Supplementary Table S5), with a large number of

gains and losses affecting multiple chromosomal regions. In addition, both tumours displayed a pattern of alterations consistent with a complex 'sawtooth' genomic profile, known to be significantly associated with high grade ER-negative tumours[18,19,45]. BRCA1/ER-BC displayed a higher number of tandem duplications, which is consistent with the so-called 'mutator' phenotype described by Stephens et al.[26] (Figure 4A). 16q loss, a hallmark feature of ER-positive breast cancers[49], was not observed in either tumour.

A similar number of amplified genes in both tumours were observed (281 in BRCA1/ER-BC and 284 in BRCA1/ER+BC (Supplementary Table S5), which is higher than the median number of amplified genes (i.e. 33) seen in re-analysis of a series of 71 grade 2 ER-positive HER2-negative sporadic breast cancers[20]. BRCA1/ER-BC, however, harboured a higher number of genes disrupted by HODs (n=30) than BRCA1/ER+BC (n=0) (Supplementary Table S5). The copy number changes inferred by MPS were confirmed by analysis of the copy number states defined by SNP6 data analysis (data not shown).

To identify putative large-scale structural variations i.e. chromosomal-rearrangements and gene-fusions events (SVs), *de novo* assembly around SV breakpoints was conducted as previously described[23], and SVs present in the patient's germline were excluded. BRCA1/ER-BC harboured a higher number of SVs than BRCA1/ER+BC (Supplementary Tables S6, S7, Figures 1 and 4A). Although this could potentially stem from differences in tumour cell percentage between BRCA1/ER-BC and BRCA1/ER+BC, qualitative differences could also be observed. BRCA1/ER-BC displayed a significantly higher intra-chromosomal:inter-chromosomal SV ratio than BRCA1/ER+BC (Fisher's exact test p=0.0042, Figure 4B, Table 2). BRCA1/ER+BC showed multiple complex structural rearrangements between chromosomes 6 and X, associated with copy number aberrations of the affected regions, indicative of chromothripsis (Figure 1)[50]. The presence of a mutator phenotype (i.e. tandem duplications) in BRCA1/ER-BC and of chromothripsis in BRCA1/ER+BC may also account for the differences in number of intra- and inter-chromosomal rearrangements observed. Nominated fusion genes were tested by Sanger sequencing (index cases), and 32/40 and 13/19 were validated in BRCA1/ER-BC and BRCA1/ER+BC, respectively (Supplementary Table S6). Out of the validated fusion genes, BRCA1/ER-BC displayed a significantly lower proportion of in-frame fusion genes than BRCA1/ER+BC (p=0.0202, Fisher's exact test, Figure 4C).

Next, the expression of the 45 validated fusion genes was tested at the RNA level by reverse transcription PCR (RT-PCR) and Sanger sequencing of tumour cDNA (Figure 4D). It should be noted that primers for cDNA sequencing of 9 of the 45 fusion gene events validated at the DNA level could not be reliably designed due to the complexity of the rearrangements. Out of the 36 fusion genes tested, 11 were expressed, of which two were found to be in-frame in BRCA1/ER+BC (*SUPT6H-ARHGAP12* and *EXT2-TTC17*) (Supplementary Figure S5). *EXT2* and *TTC17* were shown to contain intra-genic oestrogen responsive elements[51], suggesting that the expression of some in-frame fusion genes in breast cancer may be driven by ER (Supplementary Methods S1 and Results S1). To define whether these fusion genes may be recurrent events, we interrogated RNA-sequencing data from five *BRCA1* mutant ER-negative breast cancer cell lines and three *BRCA1* ER-negative mutant primary breast cancers[52] and the massively parallel DNA sequencing data of 24 breast cancers (including

4 *BRCA1* mutant ER-negative samples)[26]. None of the partners involved in these two chimaeric transcripts were rearranged and expressed in additional samples, suggesting that they may constitute private events consistent with previous observations for the majority of fusion genes in breast cancer[26,52,53].

These results demonstrate that BRCA1/ER+BC differs from the majority of sporadic ER-positive breast cancers in its patterns of gene copy number changes, however it also shows important quantitative and qualitative differences in SNVs in comparison with BRCA1/ER-BC.

Both *BRCA1* ER-positive and ER-negative breast cancers display signatures of HR DNA repair defects

Defects in HR DNA repair, such as those resulting from loss of BRCA1 function, contribute to an increase in the number of small deletions flanked by homologous sequences if other mechanisms, such as non-homologous end joining, are used by the cell for the repair of DNA double strand breaks[11,14]. Therefore we sought to determine whether the BRCA1/ER+BC would display a pattern of deletions and structural rearrangements flanked by sequences of homology or micro-homology as previously described in BRCA1 ER-negative breast cancer[26]. Given that the accuracy for the detection of in-dels was suboptimal using the MPS technology employed in this study[24], we concentrated on homology surrounding structural rearrangement breakpoints. Using a previously described approach[26], we observed a significantly higher number of large deletions and rearrangements flanked by homology and micro-homology in BRCA1/ER+BC than expected by chance (Supplementary Figure S6). These results are similar to those observed in BRCA1/ER-BC (Figures 4E and 4F) and in *BRCA1* mutant samples analysed by Stephens et al.[26], suggesting that BRCA1 dysfunction also contributes to the genomic landscape of this ER-positive *BRCA1* germline mutant breast cancer, similar to ER-negative disease. Although some samples devoid of *BRCA1* mutations were shown to display a similar pattern in Stephens et al.[26], it has been demonstrated that a substantial proportion of breast cancers lack competent HR DNA repair[54].

Prevalence of somatic mutations, copy number aberrations and structural aberrations in independent datasets

Despite BRCA1/ER-BC and BRCA1/ER+BC harbouring the same driving genetic events (i.e. BRCA1 and p53 loss-of-function), no additional gene was affected by somatic mutations or HODs in both cancers. Hence, we sought to use the mutational landscape of these tumours as the basis for the identification of novel breast cancer genes. To determine whether validated somatic mutations affecting genes found in BRCA1/ER-BC or BRCA1/ER+BC predicted to be deleterious/disease-causing by functional prediction algorithms[37–40] would be recurrent in *BRCA1* breast cancers, exons and intron-exon boundaries of 11 genes were sequenced in the validation series of *BRCA1* (n=16) and ER-matched non-*BRCA1* (n=16) breast cancers (Supplementary Table S2). The 11 genes were chosen based on i) the functional effect of the mutation in the index case on the protein as determined by SIFT[37], polyPhen[38], Condel[39], CanPredict[40]; ii) had previously been reported to be mutated in the COSMIC database (http://www.sanger.ac.uk/genetics/CGP/

cosmic/); or iii) had previously been implicated in cancer. The probability of one of these 11 genes being recurrently mutated by chance in ≥2 of the 16 *BRCA1* samples tested is <0.05 (Supplementary Methods S1). Recurrent and potentially deleterious/disease-causing somatic mutations in *DAPK3* were present in BRCA1/ER+BC and in 2/16 ER-negative *BRCA1* breast cancers, and in none of the 16 non-*BRCA1* breast cancers (Table 3, Supplementary Table S2, Supplementary Figure S7), suggesting that this gene may constitute a novel breast cancer gene independent of ER status.

We next performed a re-analysis of a publicly available aCGH dataset[20] to search for HODs of genes targeted by inactivating mutations, HODs or disrupted by SVs in either BRCA1/ER-BC or BRCA1/ER+BC. The probability of a specific gene being recurrently homozygously deleted by chance in ≥2 of the 20 *BRCA1* breast cancers investigated is <0.0001 and in ≥3 of the 343 breast cancers tested is <0.0001 (Supplementary Methods S1) [55]. This analysis confirmed that the *CDKN2A/2B* locus is recurrently homozygously deleted in breast cancer (6/343) and has led to the identification of recurrently mutated or homozygously deleted genes in independent *BRCA1* breast cancers (n=20) and/or sporadic breast cancers (n=323). These included *TMEM135* (2/20), which was recurrently homozygously deleted in additional *BRCA1* breast cancers, *GATA4* (3/343), *PDE4D* (3/343) and *KIAA1797* (3/343), which were homozygously deleted in either *BRCA1* or sporadic breast cancers (Table 3, Supplementary Results S1, Supplementary Figures S8 and S9, Supplementary Table S8). Of these, *TMEM135* and *GATA4*, have not been previously implicated in breast cancer (Table 3).

Discussion

Functional preclinical studies have demonstrated that BRCA1 can transcriptionally induce ER expression, and that BRCA1 loss-of-function leads to loss of ER expression[5,6,56]. Animal model studies have recently demonstrated that *BRCA1* breast cancers are likely to originate from breast luminal progenitor cells, which are of ER-negative phenotype; however, even in animals where concurrent inactivation of *Brca1* and *Tp53* was restricted to these cells, up to 20% of tumours were of ER-positive phenotype[8]. Our observations that the BRCA1/ER+BC displayed loss of *BRCA1* wild-type allele, as previously reported in 50-90% of all *BRCA1* breast cancers irrespective of ER-status[22,57], and a pattern of genetic aberrations consistent with those found in tumours with HR DNA repair defects suggest that the BRCA1/ER+BC, in a way akin to BRCA1/ER-BC, did not have a functional BRCA1. These data provide evidence to suggest that BRCA1 loss-of-function may be, by itself, insufficient to condition the development of breast cancers that display an ER-negative or basal-like phenotype.

In this study, we were able to address the hypothesis that ER-positive *BRCA1* cancers are merely sporadic tumours arising in a *BRCA1* mutant background. This would have implications for the treatment of patients who present with similar tumours, given that drugs targeting cells with defects in HR DNA repair (such as PARP inhibitors and platinum salts) would be ineffective in these patients if this hypothesis were true. We identified a number of molecular features in BRCA1/ER+BC suggesting that at least a subset of ER-positive *BRCA1* mutant tumours are not sporadic, but are likely to be causally linked to the BRCA1

loss-of-function. Not only was somatic loss of the wild-type *BRCA1* allele observed but also somatic *TP53* mutations and a higher ratio of transversions to transitions were identified, which has been suggested to stem from DNA repair defects[58]. In addition, enrichment for regions of microhomology flanking large deletions was identified in both tumours, a signature previously shown to be associated with defects in HR DNA repair[11–14,26]. Moreover, BRCA1/ER-BC displayed the patterns of genetic aberrations expected in a grade III, ER-negative breast cancer, whereas BRCA1/ER+BC lacked deletion of 16q, one of the hallmark features of sporadic ER-positive disease[18,45,49], and showed a 'sawtooth' pattern[45], uncommon in grade II ER-positive breast cancers.

The constellation of deleterious somatic mutations, HODs and SVs found in the *BRCA1* breast cancers subjected to whole-genome MPS was employed as the basis for the identification of novel breast cancer genes. We confirmed the existence of recurrent HODs of *CDKN2A/2B* and *PDE4D*[59–61], and identified *DAPK3*, *TMEM135* and *GATA4* as potential novel breast cancer genes. TMEM135 is a transmembrane protein that appears to integrate biological processes involving fat metabolism and energy expenditure[62]; although a SNP near to *TMEM135* has been recently reported[3], no links between this gene and breast cancer have previously been established. *DAPK3* is of particular interest, given that it encodes a cancer-associated kinase reported to act as a *bona fide* tumour suppressor in other cancer types[63]. *GATA4*, a transcription factor recently shown to have tumour suppressor functions and to play a role in tumourigenesis[64–66], was also recurrently mutated or homozygously deleted in ER-positive breast cancers, suggesting a potential tumour suppressor role for this gene in ER-positive disease. Although no recurrent mutations in *GATA4* were identified in our four additional ER-positive *BRCA1*-mutant tumours, further studies of the role of *GATA4* in ER-positive breast cancer are warranted.

One limitation of this study is the small number of cases subjected to whole-genome sequencing. The method used in this study for whole-genome sequencing requires large quantities of DNA from fresh frozen samples and matched germline DNA. Despite extensive searches in the tissue banks of the authors' institutions, only two samples yielded sufficient DNA and fitted the inclusion criteria. Second, we could not test the HR signature in additional sporadic breast cancers for which data could be retrieved through academic collaborations or direct access to public repositories, given that the sequencing methodologies employed were distinct and differences between the results obtained with the analysis of the *BRCA1* cancers and sporadic breast cancers could be attributed to differences in sequencing platforms. Third, the current algorithms to define pathogenic/disease-causing mutations are more effective for the identification of loss-of-function rather than gain-of-function mutations. Therefore, the list of 11 genes sequenced in additional *BRCA1* and sporadic breast cancers is enriched for genes whose loss-of-function is advantageous for cancer cells.

In conclusion, whole-genome sequencing of *BRCA1* breast cancers driven by *BRCA1* germline mutations and *TP53* somatic mutations revealed that despite the common drivers, the phenotypic characteristics and repertoire of somatic mutations and HODs in each cancer were dramatically different. Hence, our results suggest that the heterogeneity of the genomic landscapes of breast cancers reported so far cannot be solely attributed to the existence of

Europe PMC Funders Author Manuscripts

multiple early driver genes. BRCA1/ER+BC appears to have patterns of genetic aberrations intermediate between non-hereditary ER-positive and BRCA1/ER-BC, given that BRCA1/ER+BC lacked some of the hallmark genetic aberrations of ER-positive breast cancers, displayed somatic loss of *BRCA1* wild-type allele and a mutational signature consistent with that of BRCA1/ER-BC. Finally, our study has led to the identification of three putative novel breast cancer genes. Further studies to characterise the functional consequences of pathogenic/disease-causing mutations of these genes and of *KIAA1797* and *PDE4D* in breast cancer and their potential therapeutic implications are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding: This study was funded by Breakthrough Breast Cancer. BW and CS are funded in part by CRUK. PMW is in receipt of a Wellcome Trust Clinical Research Fellowship grant. The authors acknowledge NHS funding to the NIHR Biomedical Research Centre.

References

1. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. Nat Rev Cancer. 2004; 4:665–676. [PubMed: 15343273]
2. Mullan PB, Quinn JE, Harkin DP. The role of BRCA1 in transcriptional regulation and cell cycle control. Oncogene. 2006; 25:5854–5863. [PubMed: 16998500]
3. Teraoka SN, Bernstein JL, Reiner AS, et al. Single nucleotide polymorphisms associated with risk for contralateral breast cancer in the Women's Environment, Cancer, and Radiation Epidemiology (WECARE) Study. Breast Cancer Res. 2011; 13:R114. [PubMed: 22087758]
4. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010; 363:1938–1948. [PubMed: 21067385]
5. Hosey AM, Gorski JJ, Murray MM, et al. Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. J Natl Cancer Inst. 2007; 99:1683–1694. [PubMed: 18000219]
6. Gorski JJ, Kennedy RD, Hosey AM, et al. The complex relationship between BRCA1 and ERalpha in hereditary breast cancer. Clin Cancer Res. 2009; 15:1514–1518. [PubMed: 19223511]
7. Lim E, Vaillant F, Wu D, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009; 15:907–913. [PubMed: 19648928]
8. Molyneux G, Geyer FC, Magnay FA, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell. 2010; 7:403–417. [PubMed: 20804975]
9. Vaziri SA, Krumroy LM, Elson P, et al. Breast tumor immunophenotype of BRCA1-mutation carriers is influenced by age at diagnosis. Clin Cancer Res. 2001; 7:1937–1945. [PubMed: 11448907]
10. Tung N, Wang Y, Collins LC, et al. Estrogen receptor positive breast cancers in BRCA1 mutation carriers: clinical risk factors and pathologic features. Breast Cancer Res. 2010; 12:R12. [PubMed: 20149218]
11. Gudmundsdottir K, Ashworth A. The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. Oncogene. 2006; 25:5864–5874. [PubMed: 16998501]
12. Guirouilh-Barbat J, Huck S, Bertrand P, et al. Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol Cell. 2004; 14:611–623. [PubMed: 15175156]
13. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008; 24:529–538. [PubMed: 18809224]

14. Barber LJ, Rosa Rosa JM, Kozarewa I, et al. Comprehensive genomic analysis of a BRCA2 deficient human pancreatic cancer. PLoS One. 2011; 6:e21639. [PubMed: 21750719]
15. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010; 466:869–873. [PubMed: 20668451]
16. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464:999–1005. [PubMed: 20393555]
17. Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009; 461:809–813. [PubMed: 19812674]
18. Natrajan R, Lambros MB, Rodriguez-Pinilla SM, et al. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. Clin Cancer Res. 2009; 15:2711–2722. [PubMed: 19318498]
19. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006; 10:529–541. [PubMed: 17157792]
20. Jonsson G, Staaf J, Vallon-Christersson J, et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res. 2010; 12:R42. [PubMed: 20576095]
21. Kay GF, Penny GD, Patel D, et al. Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell. 1993; 72:171–182. [PubMed: 8425217]
22. Manie E, Vincent-Salomon A, Lehmann-Che J, et al. High frequency of TP53 mutation in BRCA1 and sporadic basal-like carcinomas but not in BRCA1 luminal breast tumors. Cancer Res. 2009; 69:663–671. [PubMed: 19147582]
23. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010; 327:78–81. [PubMed: 19892942]
24. Lee W, Jiang Z, Liu J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature. 2010; 465:473–477. [PubMed: 20505728]
25. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
26. Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009; 462:1005–1010. [PubMed: 20033038]
27. Popova T, Manie E, Stoppa-Lyonnet D, et al. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome Biol. 2009; 10:R128. [PubMed: 19903341]
28. Natrajan R, Weigelt B, Mackay A, et al. An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. Breast Cancer Res Treat. 2009 epub ahead of print.
29. Weigelt B, Mackay A, A'Hern R, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol. 2010; 11:339–349. [PubMed: 20181526]
30. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. 2006; 7:96. [PubMed: 16643655]
31. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009; 27:1160–1167. [PubMed: 19204204]
32. Geyer FC, Lacroix-Triki M, Savage K, et al. beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. Mod Pathol. 2011; 24:209–231. [PubMed: 21076461]
33. Didraga MA, van Beers EH, Joosse SA, et al. A non-BRCA1/2 hereditary breast cancer sub-group defined by aCGH profiling of genetically related patients. Breast Cancer Res Treat. 2011; 130:425–436. [PubMed: 21286804]
34. Joosse SA, Brandwijk KI, Mulder L, et al. Genomic signature of BRCA1 deficiency in sporadic basal-like breast tumors. Genes Chromosomes Cancer. 2011; 50:71–81. [PubMed: 21104783]
35. Joosse SA, van Beers EH, Tielen IH, et al. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. Breast Cancer Res Treat. 2009; 116:479–489. [PubMed: 18704682]
36. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007; 23:657–663. [PubMed: 17234643]

37. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31:3812–3814. [PubMed: 12824425]
38. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002; 30:3894–3900. [PubMed: 12202775]
39. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011; 88:440–449. [PubMed: 21457909]
40. Kaminker JS, Zhang Y, Watanabe C, et al. CanPredict: a computational tool for predicting cancer-associated missense mutations. Nucleic Acids Res. 2007; 35:W595–598. [PubMed: 17537827]
41. Stephens P, Edkins S, Davies H, et al. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. Nat Genet. 2005; 37:590–592. [PubMed: 15908952]
42. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314:268–274. [PubMed: 16959974]
43. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. Science. 2007; 318:1108–1113. [PubMed: 17932254]
44. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463:191–196. [PubMed: 20016485]
45. Hicks J, Krasnitz A, Lakshmi B, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res. 2006; 16:1465–1479. [PubMed: 17142309]
46. Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. Cell. 2011; 145:30–38. [PubMed: 21458666]
47. Pleasance ED, Stephens PJ, O'Meara S, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]
48. Parsons DW, Li M, Zhang X, et al. The genetic landscape of the childhood cancer medulloblastoma. Science. 2011; 331:435–439. [PubMed: 21163964]
49. Natrajan R, Lambros MB, Geyer FC, et al. Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype? Genes Chromosomes Cancer. 2009; 48:351–365. [PubMed: 19156836]
50. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011; 144:27–40. [PubMed: 21215367]
51. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 481:389–393. [PubMed: 22217937]
52. Ellis MJ, Suman VJ, Hoog J, et al. Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. J Clin Oncol. 2011; 29:2342–2349. [PubMed: 21555689]
53. Robinson DR, Kalyana-Sundaram S, Wu YM, et al. Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. Nat Med. 2011; 17:1646–1651. [PubMed: 22101766]
54. Graeser M, McCarthy A, Lord CJ, et al. A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary breast cancer. Clin Cancer Res. 2010; 16:6159–6168. [PubMed: 20802015]
55. Bignell GR, Greenman CD, Davies H, et al. Signatures of mutation and selection in the cancer genome. Nature. 2010; 463:893–898. [PubMed: 20164919]
56. Harte MT, O'Brien GJ, Ryan NM, et al. BRD7, a subunit of SWI/SNF complexes, binds directly to BRCA1 and regulates BRCA1-dependent transcription. Cancer Res. 2010; 70:2538–2547. [PubMed: 20215511]
57. Tung N, Miron A, Schnitt SJ, et al. Prevalence and predictors of loss of wild type BRCA1 in estrogen receptor positive and negative BRCA1-associated breast cancers. Breast Cancer Res. 2010; 12:R95. [PubMed: 21080930]

58. Biggs PJ, Warren W, Venitt S, et al. Does a genotoxic carcinogen contribute to human breast cancer? The value of mutational spectra in unravelling the aetiology of cancer. Mutagenesis. 1993; 8:275–283. [PubMed: 8377645]

59. Jonsson G, Naylor TL, Vallon-Christersson J, et al. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res. 2005; 65:7612–7621. [PubMed: 16140926]

60. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007; 450:893–898. [PubMed: 17982442]

61. Nancarrow DJ, Handoko HY, Smithers BM, et al. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays. Cancer Res. 2008; 68:4163–4172. [PubMed: 18519675]

62. Exil VJ, Silva Avila D, Benedetto A, et al. Stressed-induced TMEM135 protein is part of a conserved genetic network involved in fat storage and longevity regulation in Caenorhabditis elegans. PLoS One. 2010; 5:e14228. [PubMed: 21151927]

63. Brognard J, Zhang YW, Puto LA, et al. Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. Cancer Res. 2011; 71:3152–3161. [PubMed: 21487036]

64. Agnihotri S, Wolf A, Munoz DM, et al. A GATA4-regulated tumor suppressor network represses formation of malignant human astrocytomas. J Exp Med. 2011; 208:689–702. [PubMed: 21464220]

65. Hellebrekers DM, Lentjes MH, van den Bosch SM, et al. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. Clin Cancer Res. 2009; 15:3990–3997. [PubMed: 19509152]

66. Wen XZ, Akiyama Y, Pan KF, et al. Methylation of GATA-4 and GATA-5 and development of sporadic gastric carcinomas. World J Gastroenterol. 2010; 16:1201–1208. [PubMed: 20222162]

Europe PMC Funders Author Manuscripts

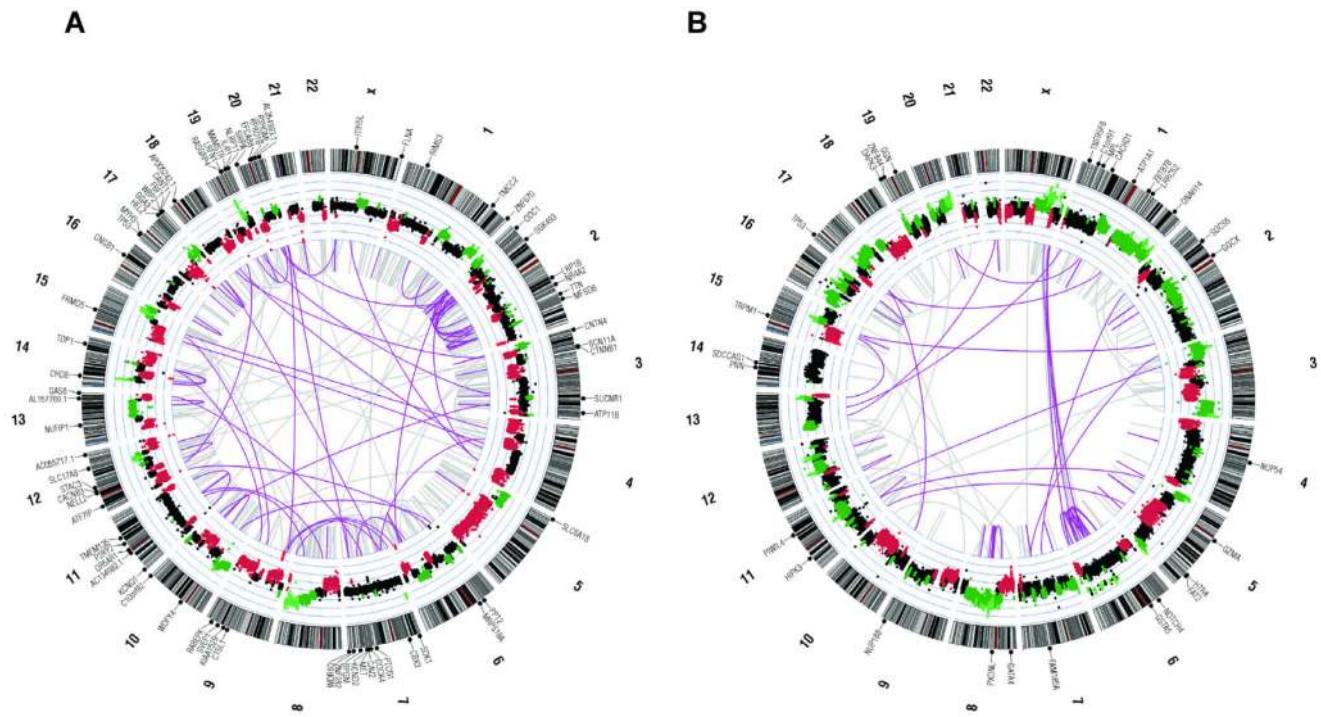

Figure 1. The genomic landscape of somatic alterations in BRCA1 ER-negative and ER-positive breast cancers.

Circos plots of (**A**) *BRCA1* mutant ER-negative (BRCA1/ER-BC) and (**B**) *BRCA1* mutant ER-positive (BRCA1/ER+BC) primary breast cancers, derived from massively parallel sequencing data at a coverage of >30x of the genome (Supplementary Table S1). Circles from outside to inside depict the following: i) chromosomes and validated mutations above the somatic score threshold (Supplementary Methods S1); ii) copy number derived from GC normalised sequence coverage (2kb window). Red indicates copy number loss, green copy number gain and bright green amplifications; iii) high-confidence structural rearrangements (grey), with rearrangements validated by Sanger sequencing highlighted in purple.

Europe PMC Funders Author Manuscripts

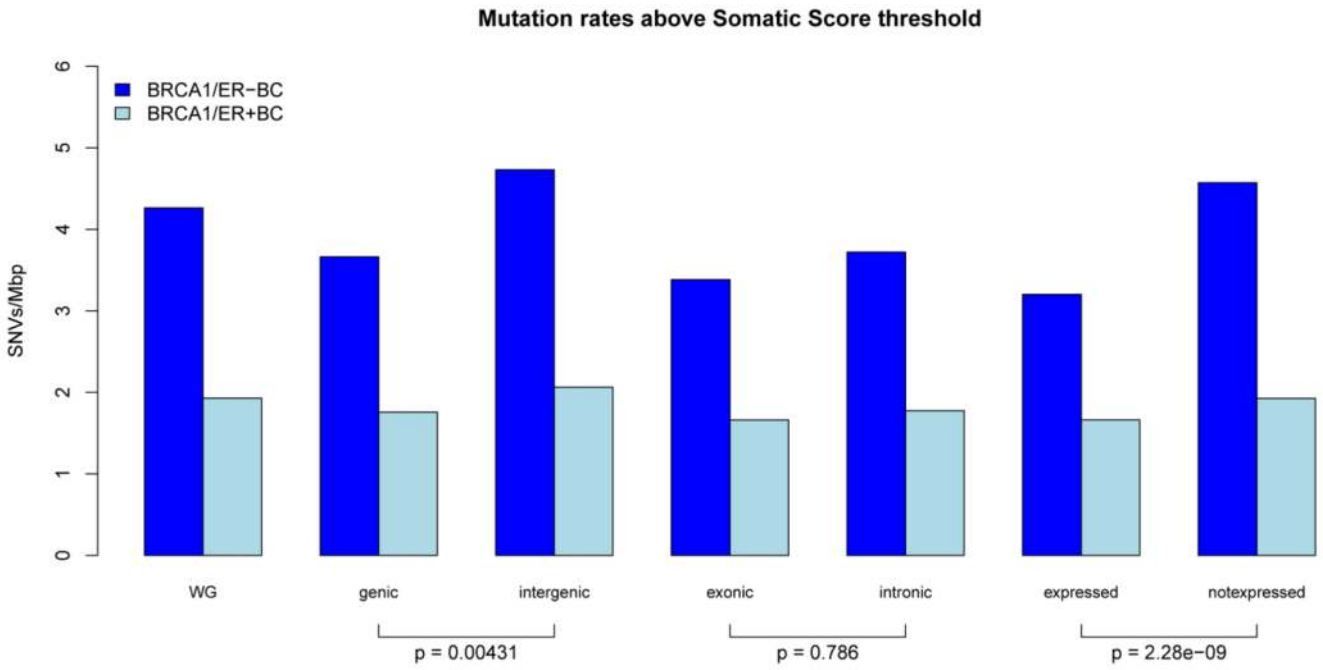

Figure 2. Mutation rates in *BRCA1* mutant ER-negative and ER-positive breast cancers.

Mutation rates as defined by single nucleotide variants (SNVs) per mega-base pair (Mbp), above the somatic score threshold (i.e. BRCA1/ER-BC-0.057 and BRCA1/ER+BC-0.034), in BRCA1/ER-BC (dark blue) and BRCA1/ER+BC (light blue) for whole-genome (WG), genic regions, intergenic regions, exonic regions, intronic regions, expressed genes, and non-expressed genes. P values: two-tailed Fisher's exact test. SNV: single nucleotide variation.

Europe PMC Funders Author Manuscripts

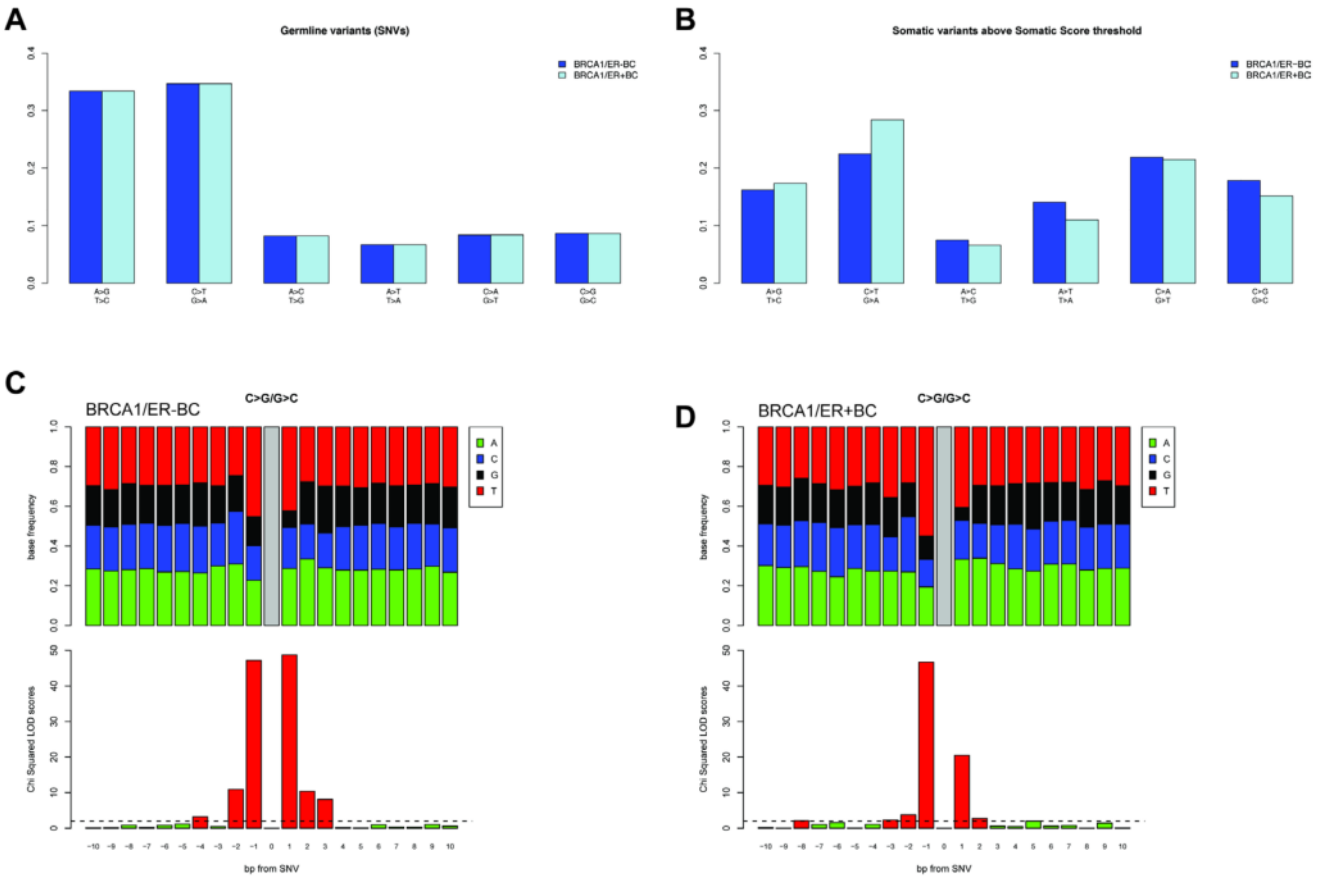

Figure 3. Mutational trends and patterns in *BRCA1* mutant ER-negative and ER-positive breast cancers.

Frequency of transitions to transversions in BRCA1/ER-BC (dark blue) and BRCA1/ER +BC (light blue) in the germline (A) and tumour (B) DNA samples of BRCA1/ER-BC (dark blue) and BRCA1/ER+BC (light blue). The y-axis depicts the frequency of base changes. In both tumours, the ratio of transition:transversion decreases in comparison to the germline ratio of 2:1. The sequence context of somatic substitutions of C>G/G>C transversions in BRCA1/ER-BC (C) and in BRCA1/ER+BC (D). The bars at position zero indicate the variation itself. The remaining bars indicate the fractions of different nucleotides at positions 5' (minus) and 3' (plus) to the variation. C>G/G>C transversions were more frequently preceded and/or succeeded by T. Chi-square logarithm (base 10) of odds (LOD) scores are depicted below each panel. Chi-square LOD scores of > 2 were considered significant.

Europe PMC Funders Author Manuscripts

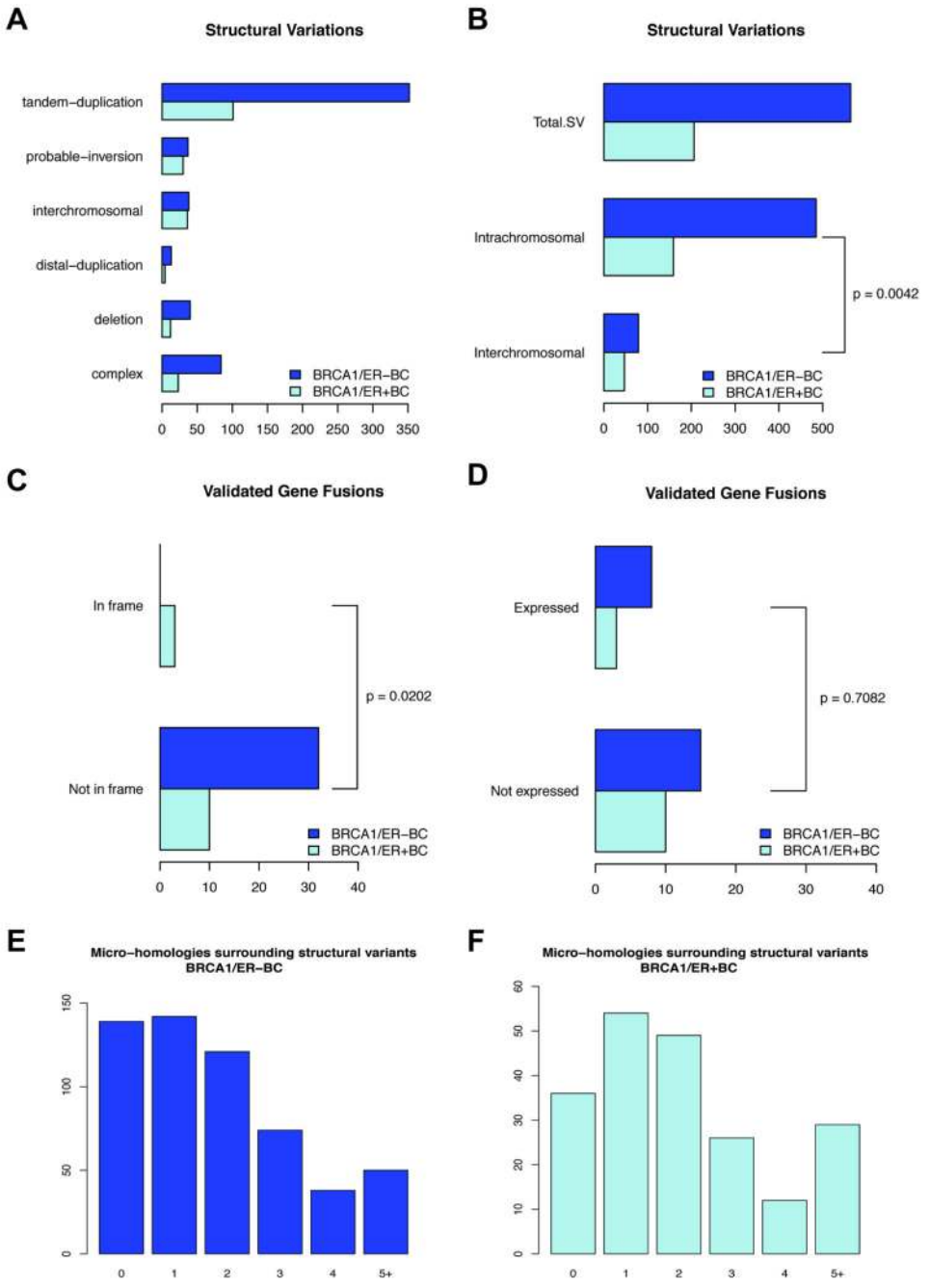

Figure 4. Landscape of structural rearrangements in *BRCA1* breast cancers.

Summary of the landscape of high-confidence structural rearrangements in BRCA1/ER-BC (dark blue) and BRCA1/ER+BC (light blue) according to individual classes of structural rearrangements (A). Total number of structural rearrangements in BRCA1/ER-BC (dark blue) and BRCA1/ER+BC (light blue) that are intra-chromosomal and inter-chromosomal (B). Number of validated structural variants predicted to be in-frame and not in-frame (C). Number of structural variants validated by Sanger sequencing that were expressed and not expressed at the RNA level (D). The number of structural rearrangements is plotted in the y-

axis. P values: two-tailed Fisher's exact test. Extent of overlapping micro-homology sequences at structural rearrangement breakpoints in BRCA1/ER-BC (E) and in BRCA1/ER +BC (F). The number of structural rearrangements (y-axis) is plotted according to the number of base pairs of micro-homology (x-axis).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Summary of the clinicopathological and molecular characteristics of the index cases.

Characteristic	BRCA1/ER-BC	BRCA1/ER+BC
Age	54	52
Sex	Female	Female
Histological grade	3	2
Mitoses/ 10 HPF	81	15
ER	Negative	Positive
PR	Negative	Negative
HER2	Negative	Negative
***BRCA1* mutation**	c.124delA/p.Ile42TyrfsX8	c.4485-?_4986+?del/p.Ser1496CysfsX14
Somatic *BRCA1* wild-type allele *	Loss	Loss
***TP53* mutation**	c.927_928delTA	c.951C>T
Molecular subtype	Basal-like	Luminal B
Tumour cell content GAP* (H&E) +	71% (70%)	42% (60%)

Mitoses 10/HPF (per 10 high power field). Molecular subtype determined by single sample predictors [29] according to Hu et al. [30] and Parker et al. [31].

*Defined by Genome Alteration Print (GAP) based on single nucleotide polymorphism array data (SNP6, Affymetrix)[27].

+Haemotoxylin and Eosin staining of frozen tumour samples.

Table 2

Summary of genomic features of BRCA1/ER-BC and BRCA1/ER+BC

Criteria	BRCA1/ER-BC	BRCA1/ER+BC
Whole genome mutation rate	4.26	1.93
Intronic:exonic mutation rate	1.1	1.07
Intergenic:genic mutation rate ratio	1.29	1.17
Non-expressed:expressed mutation rate ratio	1.43	1.16
Number of expressed genes *	12499	12845
Validated SNVs (% validated)	62 (67)	28 (55)
Validated in-dels (% validated)	9 (22)	1 (5)
Copy number profile	sawtooth	sawtooth
Number of amplified genes	281	284
Number of homozygously deleted genes	30	0
Number of total SVs	564	206
Number of tandem duplications	352	101
Number of inter-chromosomal SVs	79	47
Number of intra-chromosomal SVs	485	159
Validated gene fusions	32	13
In-frame validated gene fusions	0	3
Expressed fusion genes	8	3
Expressed in-frame fusion genes	0	2
Micro-homologies surrounding SVs	yes	yes
Evidence of chromothripsis[50]	no	yes
Mutator phenotype[26]	yes	no

* Determined from Affymetrix U133Plus2 arrays; in-del: insertion and deletion; SNV: single nucleotide variant; SV: structural variant

Europe PMC Funders Author Manuscripts

Table 3

Identification of recurrent alterations in *BRCA1* mutant and *BRCA1* non-mutant breast cancers.

	Index cases				Recurrence screen							
Gene	**BRCA1/ER-BC**	**BRCA1/ER+BC**	**Prediction***	**Cosmic**	**Cancer Gene Census**	**Novel breast cancer gene****	***BRCA1* mutant ER-negative (n=12)**	***BRCA1* mutant ER-positive (n=4)**	**non-*BRCA1* ER-negative sporadic (n=12)**	**non-*BRCA1* ER-positive sporadic (n=4)**	**Mutation**	**Prediction***
DAPK3	0	c.703C>T	deleterious	c.646C>T, lung cancer, NCI-H1770; c.481G>A, ovarian mucinous carcinoma; c.335C>T, large_intestinal adenocarcinoma; c.27G>A, Grade IV astrocytoma;	N	Y	1) c.1013_1023 del_GCCTGCGC GAG; 2) c.2T>G	0	0	0	1) frameshift 2) misssense	1) trunctating 2) damaging
Gene	**BRCA1/ER-BC**	**BRCA1/ER+BC**	**Prediction**	**Cosmic**	**Cancer Gene Census**	**Novel breast cancer gene****	***BRCA1* mutant ER-negative (n=18)**	***BRCA1* mutant ER-positive (n=2)**	**non-*BRCA1* ER-negative sporadic (n=98)**	**non-*BRCA1* ER-positive sporadic (n=225)**	**Mutation**	**Prediction***
GATA4	0	c.185A>T	neutral, deleterious	c.1312G>A, Grade IV astrocytoma	N	Y	0	0	0	3 homozygous deletions (1 *BRCA2* mutant)	homozyggous deletion	NA
PDE4D	homozygous deletion	0	NA	c.1865A>C, breast ductal carcinoma cell line HCC2157; c.1045-1G>A, Grade IV astrocytoma	N	N	1 homozygous deletion		1 homozygous deletion-familial	1 homozygous deletion		
TMEM135	c.360C>T	0	deleterious	c.1023G>C, upper aerodigestive tract; c.199G>C, ovarian serous carcinoma	N	Y	2 homozygous deletions	0	0	0	homozyggous deletion	NA
CDKN2A/2B	homozygous deletion	0	NA	c.?,c.151_457del307,c.151_457del307,c.151_457del307, c.247C>T,c.155T>A,c.?_?del?,c.?_?del?, c.?_?del?,c.1_471del471,c.1_471del471,c.1_471del471, c.1_471del471,c.1_471del471,c.1_471del471,c.1_471del471, c.1_471del471,c.1_471del471,c.1_471del471,c.1_471del471, c.151_457del307,c.1_471del471,c.1_471del471,c.1_471del471, c.1_457del457,c.1_471del471,c.1_471del471,c.1_471del471, c.151_457del307,c.1_471del471,c.1_471del471,c.156G>C, c.156G>C,c.156G>C,c.456_457+25AGGTGAGGACTGATGATCTGAGAATTT>C, c.150+2T>C,c.225delC,c.224C>T,c.318G>A, c.151_457del307,c.1_471del471,c.1_471del471,c.1_471del471, c.1_471del471,c.1_471del471,c.203C>T, breast cancer	Y	N	0	0	1 homozygous deletion	5 homozygous deletions(3 *BRCA2*)	homozyggous deletion	NA
KIAA1797	structural variation-interchromosomal fusion	0	NA	c.2862G>A, invasive ductal breast cancer; c.5148T>A, clear cell renal carcinoma; c.2374G>T,c.2212C>G,c.4106G>A,c.2946T>A, c.2516T>A,c.2678G>T,c.378T>C,c.15C>A, serous ovarian cancer	N	N	0	0	0	3 homozygous deletions (1 *BRCA2* mutant)	homozyggous deletion	NA

Identification of recurrent somatic genetic alterations identified in the index cases BRCA1/ER-BC and BRCA1/ER+BC in an independent validation series of 16 *BRCA1* mutant breast cancers and 16 non-*BRCA1* mutant sporadic breast cancers. The complete list of mutations is available in Supplementary Table S2. Recurrent homozygous deletions were identified through re-analysis of publicly available microarray comparative genomic hybridisation datasets of *BRCA1* mutant (n=20) and non-*BRCA1* mutant (n=323) breast cancers from Jonsson et al. [20].

* Predictions of mutations on protein structure determined with SIFT[37], polyPhen[38], Condel[39] and CANpredict[40].