

Recurrent and pathological gene fusions in breast cancer: current advances in genomic discovery and clinical implications

Jamunarani Veeraraghavan^{1,2,3} · Jiacheng Ma^{1,2,3} · Yiheng Hu^{1,2,3} ·
Xiao-Song Wang^{1,2,3,4,5,6,7}

Received: 9 March 2016 / Accepted: 18 June 2016 / Published online: 2 July 2016
© Springer Science+Business Media New York 2016

Abstract Gene fusions have long been considered principally as the oncogenic events of hematologic malignancies, but have recently gained wide attention in solid tumors due to several milestone discoveries and the advancement of deep sequencing technologies. With the progress in deep sequencing studies of breast cancer transcriptomes and genomes, the discovery of recurrent and pathological gene fusions in breast cancer is on the focus. Recently, driven by new deep sequencing studies, several recurrent or pathological gene fusions have been identified in breast cancer, including *ESR1-CCDC170*, *SEC16A-NOTCH1*, *SEC22B-NOTCH2*, and *ESR1-YAP1* etc. More important, most of these gene fusions are preferentially identified in the more aggressive breast cancers, such as luminal B, basal-like, or endocrine-resistant breast cancer, suggesting recurrent gene fusions as additional key driver events in these tumors other

than the known drivers such as the estrogen receptor. In this paper, we have comprehensively summarized the newly identified recurrent or pathological gene fusion events in breast cancer, reviewed the contributions of new genomic and deep sequencing technologies to new fusion discovery and the integrative bioinformatics tools to analyze these data, highlighted the biological relevance and clinical implications of these fusion discoveries, and discussed future directions of gene fusion research in breast cancer.

Keywords Breast cancer · Recurrent gene fusions · Next-generation sequencing · Genomics · Bioinformatics · Tumorigenesis

Abbreviations

CFS	Congenital fibrosarcoma
CGH	Comparative genomic hybridization
ConSig	Concept signature
ER	Estrogen receptor
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescent in situ hybridization
GSI	Gamma-secretase inhibitor
GWAS	Genome-wide association studies
Her2	Human epidermal growth factor receptor 2
NGS	Next-generation sequencing
NICD	NOTCH intracellular domain
NSCLC	Non-small cell lung cancer
PDX	Patient-derived xenograft
PR	Progesterone Receptor
RNA-seq	RNA sequencing
RPPA	Reverse phase protein array
SBC	Secretory breast carcinoma
TCGA	The cancer genome atlas
TGS	Third-generation sequencing
TIC	Transcription-induced chimera

✉ Xiao-Song Wang
wangx13@pitt.edu

¹ Lester & Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA

² Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

³ Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA

⁴ Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

⁵ University of Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, PA 15232, USA

⁶ Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15232, USA

⁷ Hillman Cancer Center, Research Pavilion, University of Pittsburgh Cancer Institute, 5117 Centre Avenue, Room G.5a, Pittsburgh, PA 15213, USA

WES	Whole exome sequencing
WGS	Whole genome sequencing
YAC	Yeast artificial chromosome

Introduction

Gene fusions resulting from chromosomal rearrangements have been well recognized as an important class of genomic aberrations that are key drivers of oncogenesis [1]. While many of the cancer-associated mutations have a variable and heterogeneous nature of occurrence across different tumor types, gene fusions for the most part are typically disease specific. Gene fusions have an important role in the initial steps of tumor development [1, 2]. This is exemplified by the archetype fusion, *BCR-ABL* in chronic myeloid leukemia, also known as Philadelphia chromosome [3–5]. More important, many of the recurrent gene fusions identified in cancer have been proven as robust therapeutic targets and their discoveries have been the driving force of modern precision therapeutics [6]. The significance of gene fusions in the development of epithelial cancers is on the rise in recent years since the discovery of *TMPRSS2-ERG* fusion in about 50 % of prostate cancer [7] and *EML4-ALK* fusion in 6.7 % of non-small cell lung cancer (NSCLC) [8]. The *EML4-ALK* gene fusion has been matched with an effective ALK-targeted therapy that has generated enormous patient impact [9]. Recent genome-wide studies provide evidence that gene fusions are not only present but also abundant in solid tumors including breast cancer [10, 11]. An interesting recent review by Mertens et al. recapitulates the evolution of gene fusion research over the years, including the technological advancements, the breakthrough fusion discoveries made, and the resultant therapeutic developments [12]. A more detailed account of these developments, as discussed in the recent book by Rowley et al., is also a valuable resource in this context [13].

Breast cancer is a highly heterogeneous disease comprising of distinct molecular subtypes, variety of genetic aberrations including mutations, and a varied clinical outcome [14, 15]. While endocrine therapy and Her-2-targeted therapy have been very successful in treating estrogen receptor (ER)-positive or Her-2-positive breast cancers, about half of the tumors in advanced cases will exhibit de novo or acquired resistance to these therapies [16–20]. In addition, treatment options are limited for breast cancers that are negative for both receptors. The genetic aberrations driving these more aggressive breast cancers are ill-understood. Recently, driven by new deep sequencing studies of breast cancer, several recurrent or pathological gene fusions have been identified, and interestingly, most of which are preferentially found in the more aggressive

breast cancers such as luminal B, triple-negative, or endocrine-resistant breast cancer. This review summarizes the current achievements in identification and characterization of recurrent or pathological gene fusions in breast cancer. We also highlight the biological relevance and clinical implications of these discoveries, and discuss the possible applications and future directions of gene fusion study in breast cancer.

Next-generation sequencing-aided fusion detection

Although important gene fusion discoveries transpired before the advent of second-generation sequencing or next-generation sequencing (NGS) through conventional methods like comparative genomic hybridization (CGH) and Fluorescent in situ hybridization (FISH), the invention of NGS technologies revolutionized the field of genomics and made large-scale and high-throughput fusion discovery possible. NGS generates high-throughput data and can comprehensively catalog chimerical transcripts and genomic rearrangements in the cancer transcriptome/genome. Since the introduction of first NGS method in 2005, several different and improved NGS platforms have been developed, including whole genome sequencing (WGS), whole exome sequencing (WES), and whole transcriptome sequencing (RNA-Seq). To date, more than 90 % of the total number of gene fusions reported in all types of tumors has been discovered using various NGS approaches in the past 5 years [12].

RNA-seq

RNA-seq is the most commonly selected method for fusion transcript detection since it focuses only on the expressed regions of the genome, making the discovered fusions more relevant. In addition to being relatively less expensive with a quick turnaround time, it can also quantify expression levels and facilitate the detection of multiple fusion variants generated during the fusion event, making this a more ideal technology for gene fusion detection. However, RNA-seq routinely generates a daunting quantity of chimeric sequences, most of which are artifactual fusion sequences as a result of library artifacts and mapping errors [21, 22], transcription-induced chimeras (TIC) resulting from intergenic splicing of adjacent genes, or passenger gene fusions. Bioinformatics methods that use stringent parameters to filter out artifactual chimeras may become less sensitive in detecting authentic fusions or underestimate their incidence, while filtering out the chimeras from adjacent genes will dismiss the close-range gene translocations.

Whole genome sequencing

WGS-based gene fusion detection works through identifying structural variations, a subset of which generates expressed gene fusions. Despite being unbiased and providing a comprehensive and integrative characterization of novel genomic alterations, WGS falls short owing to the short read length, technical artifacts, poor coverage, and higher false-positive calls due to sequencing errors. In addition, the gene fusions discovered through WGS are at the genomic level covering both coding and noncoding regions, and so further evaluation of their ability to produce fusion transcripts is needed to determine their significance. More recently, a third-generation sequencing technology based on single-molecule sequencing was developed by Pacific Biosciences (PacBio) in 2010 [23]. Although the technology is still relatively immature and not widely available, it promises longer read lengths (>40,000 bp with average length around 10,000–15,000 bp), higher accuracy, small amount of starting material, and low cost [24, 25], and thus may overcome the limitations of the current WGS technology.

Bioinformatics approaches and tools in gene fusion discovery

A comprehensive list of the gene fusion detection tools and other NGS data analysis tools [26] is available in the OMICtools portal (www.omictools.com). Currently, there are over 20 different fusion detection tools that utilize WGS, RNA-seq, or both and follow different discovery algorithms [27]. Each tool has its fair share of intrinsic pros and cons and also exhibits diverse behavior based on the dataset used. Achieving an adequate balance between fusion detection sensitivity and accuracy, while critical, is still in the making. The study by Carrara et al. comparing eight most used RNA-seq-based fusion detection tools revealed that the tools vary greatly in fusion detection sensitivity and false fusion discovery rate [21], suggesting the need for improvement in their specificity and sensitivity. More recently, Kumar et al. compared the performance of 12 fusion detection tools on the basis of fusions detected, sensitivity, and positive prediction values in four different types of RNA-seq datasets [22]. Among these tools, they found that EricScript [28] had 100 % positive predictive value with a sensitivity of 78 %, while requiring the least amount of time and memory utilization. Moreover, the performance of each tool differed based on the quality of dataset, such as read length and number of reads, suggesting that the attributes of RNA-seq data on hand should be one of the key deciding factors while choosing

the fusion detection tool. Another latest study by Liu et al. [29] comprehensively evaluated 15 fusion transcript detection algorithms and found that no single method is dominantly the best, but SOAPfuse [30] has overall best performance, followed by FusionCatcher [31] and JAFFA [32]. EricScript performed well in synthetic data but not in the three real data sets they evaluated. Since WGS and RNA-seq have its own limitations when employed separately, integrative tools such as INTEGRATE, BreakTrans, and Comrad that utilize both WGS and RNA-seq were developed to increase the sensitivity and specificity of fusion detection [33–35].

While these approaches aim to detect authentic fusion sequences, identifying the recurrent gene fusions generated by genomic rearrangements and distinguishing the pathological fusions from passenger fusions is challenging. In this regard, deploying additional dimensions of genomic datasets such as copy number, exon expression, and molecular concepts will be critical to provide additional filters. As copy number alterations frequently accompany genomic fusion events, although not necessarily at all times (such as in copy-neutral balanced rearrangements), the presence of intragenic copy number breakpoints within one or both of the fusion partner genes are indications of genomic rearrangements [36, 37]. In our recent discovery of *ESR1-CCDC170* fusion [38], we utilized a fusion detection pipeline to capture authentic fusion sequences from RNA-seq data, nominate tumor-specific recurrent fusion candidates based on their expression profile, and deploy copy number profiling datasets to reveal the fusions generated by unbalanced genomic rearrangements (Fig. 1a). These candidates were then ranked through Concept Signature (ConSig) analysis (<http://consig.cagenome.org/>), which prioritizes biologically important fusion genes by assessing their association with molecular concepts characteristic of cancer genes [39]. This discovery indicated the importance of data integration and use of appropriate additional filters during fusion detection process, which will provide multiple levels of genomic evidence to lock in the authentic recurrent gene fusions.

The landscape of gene fusions in breast cancer

A considerable diversity in the incidence of gene fusions has been observed across different tumor types, with breast cancer falling more toward the higher end of this spectrum [40]. In a study by Robinson et al., RNA-seq on a panel of 89 breast cancer cell lines and tumors showed that gene fusions are not rare events in breast cancer, with an average of 5.5 fusions per breast cancer cell line and 4.2 fusions per primary tumor [10]. Sequencing analyses of breast cancer genome have broadly revealed the presence of complex

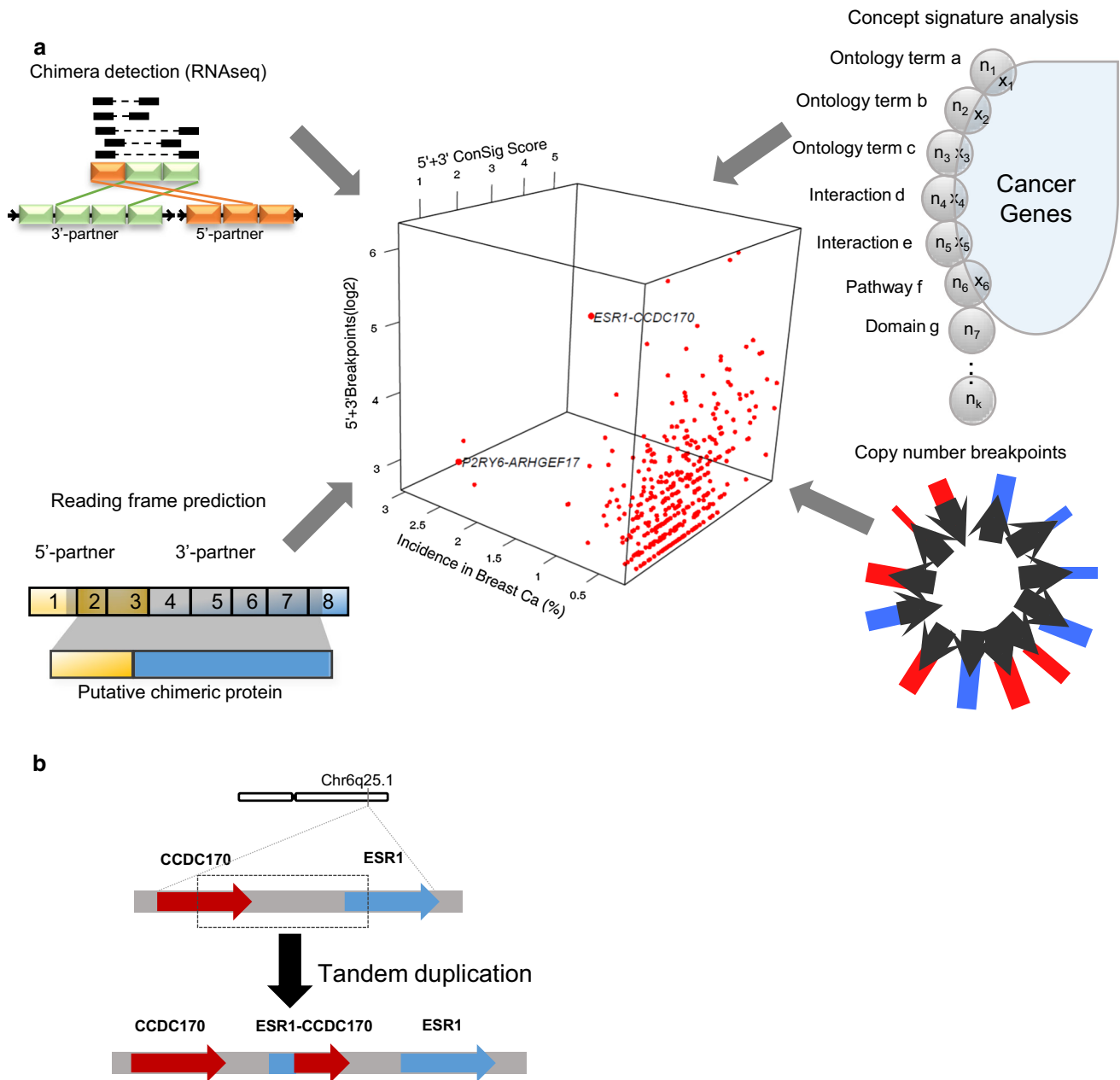


Fig. 1 Schematic representation of bioinformatics pipeline to discover pathological recurrent gene fusions in the cancer genome **a** The FusionZoom discovery pipeline detects recurrent pathological fusion candidates from the RNA-seq data, catalogs the unbalanced breakpoints at the genomic loci of fusion partner genes from copy number

data, and prioritizes pathological gene fusions through the ConSig analysis **b** Recurrent *ESR1-CCDC170* gene fusion identified in aggressive luminal breast cancers. The schematic shows the tandem duplication in chr6q25.1 as possible major genetic mechanism generating this fusion

somatic rearrangements in both breast cancer cells and primary tumor samples [41, 42]. Most gene fusions found in breast cancer involve intrachromosomal rather than interchromosomal rearrangements [10, 42], and tandem duplications are especially common in some cancers [42]. Further, a study by Kalyana-Sundaram et al. reported the presence of “amplicon-associated gene fusions”, a distinct group of genetic alterations involving genes in the chromosomal amplifications loci, to be largely prevalent in

breast cancers [43]. However, this class of genomic aberrations is predominantly a by-product of chromosomal amplifications, and constitutes a subset of “pseudo” or passenger aberrations that should be carefully considered during the prioritization of gene fusion candidates.

To date, only a handful of the discovered fusions have been found to be recurrent or pathological genomic events in breast cancer. In spite of the slow progress in breast cancer gene fusion discovery, we speculate that recurrent

and pathological genomic fusions are still to be discovered considering the limitations of current deep sequencing technologies and bioinformatics approaches in detecting recurrent genomic gene fusions (as discussed above and in perspectives), the complex and diverse protein-coding structures and oncogenic properties of authentic recurrent gene fusions, and the fact that the gene fusion pool inhabiting a breast tumor is comprised of a daunting quantity of miscellaneous passenger rearrangements and recurrently expressed TICs. In contrast, driver genetic aberrations including driver gene fusions are usually mutually exclusive such that they rarely coexist with one another in the same tumor [44–46]. Thus, looking for driver genomic fusions within the large quantity of passenger rearrangements and chimerical transcripts in a tumor is like finding the needle in a haystack. All these perspectives will complicate the discovery and identification process of true recurrent gene fusions resulting from genomic rearrangements. Herein, we review the gene fusions currently reported in breast cancer, categorized by their functionally distinctive molecular classes to summarize the current advances in gene fusion identification. Table 1 summarizes the currently known recurrent and pathologically important gene fusions in breast cancer, together with TICs.

Estrogen receptor gene fusions

Estrogen receptor (ER) is a key regulator of cell growth and survival in a large fraction of breast cancers, especially the ER+ subtype. Recently, mutations in ER have been linked to acquired endocrine resistance in breast cancer [47–50]. Our recent discovery of recurrent genomic rearrangements between *ESR1* and its neighbor gene *CCDC170* in 6–8 % of luminal B tumors revealed the presence of recurrent gene fusions in this more aggressive and endocrine-resistant form of ER+ breast cancer [38]. In addition, a majority of *ESR1-CCDC170*-positive tumors harbor tandem duplications between *ESR1* and *CCDC170* genes, which is consistent with the previous finding that tandem duplications are particularly common in certain breast tumors [42]. Additional clinicopathological studies will be required to further evaluate the incidence of this fusion in different breast cancer cohorts and elucidate its association with breast cancer endocrine resistance. In addition to our study, several other gene fusions involving the *ESR1* gene have been identified recently, including *ESR1-YAP1*, *ESR1-POLH*, and *ESR1-AKAP12* in ER+ patient-derived xenograft (PDX) samples [51]. Although these fusions appear to happen in individual cases among the samples tested, they share a common structure in that the ligand-binding domain of *ESR1* is absent, but the hormone-independent transactivation domain and DNA-binding domain are retained suggesting their

pathological significance in ER+ breast cancer. The Cancer Genome Atlas (TCGA) gene fusion study using RNA-seq data from 16 tumor types (<http://www.tumorfusions.org>) including 1019 breast tumors reported the identification of 16 *ESR1*-associated fusions in breast cancer [40]. However, no molecular or functional validation of these fusions was performed in that study.

Kinase gene fusions

Gene fusions involving kinases as functional fusion partner are crucial in cancer, since the chimeric protein generated by this class of fusion often signify ideal and specific targets for drug development. The *ETV6-NTRK3* fusion generated by balanced chromosomal translocation between *ETV6* gene on chromosome 12p13 and *NTRK3* gene on chromosome 15q25 comprises the helix-loop-helix (HLH) dimerization domain of *ETV6* transcription factor linked to the protein tyrosine kinase (PTK) domain of *NTRK3* and encodes a chimeric protein functioning as constitutive-active PTK. In addition to being a pathologically significant and functionally well-characterized fusion, the systematic approaches that were employed during the initial discovery of *ETV6-NTRK3* served as a prospect for further gene fusion discoveries. *ETV6-NTRK3* was initially detected in pediatric mesenchymal tumors in 1998, using conventional cytogenetic approaches [52, 53]. Karyotype analysis of fibroblasts from congenital fibrosarcoma (CFS) cases revealed abnormal clones with rearrangements of chromosome 15q25-26 and abnormalities of 12p13. To map the breakpoints, FISH analysis of the 12p13 and 15q25-26 alterations was performed using a series of non-chimeric yeast artificial chromosomes (YACs). The YACs yielding split FISH signals were identified and used together in dual-colored FISH and the resultant fusion signal detected represented a der(15)t(12;15). Subsequent cloning and sequencing of the fragments revealed that the *ETV6* gene is fused in-frame to *NTRK3* gene. Interestingly, this rearrangement gained importance in the context of breast cancer when it was later discovered as the only cytogenetic abnormality in human secretory breast carcinoma (SBC) in 2002 [54]. In that study, *ETV6-NTRK3* was observed in 12 of the 13 (92 %) formalin-fixed paraffin-embedded (FFPE) SBC tissues analyzed but not in other ductal carcinomas, suggesting that SBC is characterized by *ETV6-NTRK3*. Later, Makretsov et al. developed a FISH assay for the detection of *ETV6-NTRK3* in FFPE TMAs from SBC tissues, which showed 80 % sensitivity and 100 % specificity in detecting the fusion. Using this assay, the frequency of *ETV6-NTRK3* was assessed in 202 invasive breast carcinoma cases including 1 SBC case [55]. *ETV6-NTRK3* was detected only in that SBC tissue but not in non-SBC

Table 1 List of the currently known recurrent gene fusions and TICs in breast cancer

Gene fusion	Clinicopathological association	Frequency in BC subtypes	Fusion type	RNA assay	DNA assay	Biologically validated	Study
<i>ETV6-NTRK3</i>	Secretory BC	12/13 (92 %) SBC	Interchr	qRT-PCR	FISH	Yes	Tognon et al. [54]
<i>ESR1-CCDC170</i>	ER ⁺ BC (luminal B)	8/200 (4 %) ER ⁺	Adjacent	qRT-PCR	Genomic PCR	Yes	Veeraraghavan et al. [38]
<i>SEC16A-NOTCH1</i>	Triple-negative BC	2/89 (2 %) mixed BC subtypes	Adjacent	qRT-PCR	Genomic PCR	Yes	Robinson et al. [10] Stoeck et al. [70]
<i>SEC22B-NOTCH2</i>	Triple-negative BC	1/89 mixed BC subtypes	Intrachr	qRT-PCR	Genomic PCR	Yes	Robinson et al. [10]
<i>ARID1A-MAST2</i>	No	1/89 mixed BC subtypes	Intrachr	qRT-PCR	Genomic PCR	Yes	Robinson et al. [10]
<i>ZNF700-MAST1</i>		1/89 mixed BC subtypes	Adjacent				
<i>NFIX-MAST1</i>		1/89 mixed BC subtypes	Adjacent				
<i>TADA2A-MAST1</i>		1/89 mixed BC subtypes	Interchr				
<i>GPBP1L1-MAST2</i>		1/89 mixed BC subtypes	Adjacent				
<i>MAGI3-AKT3</i>	Triple-negative BC	1/235 mixed BC subtypes	Intrachr	qRT-PCR	Genomic PCR	Yes	Banerji et al. [41] Mosquera et al. [60] Wu et al. [63]
<i>FGFR2-AFF3</i>	Metastatic BC	1/unknown	Interchr	qRT-PCR	No	Yes	
<i>FGFR2-CASP7</i>		1/unknown	Intrachr				
<i>FGFR2-CCDC6</i>		1/unknown	Intrachr				
<i>ERLIN2-FGFR1</i>		1/unknown	Adjacent				
<i>ESR1-YAP1</i>	ER ⁺ BC	1/22 mixed BC subtypes	Interchr	qRT-PCR	No	Yes	Li et al. [51]
<i>RPS6KB1-VMP1</i>	No	22/70 (31 %) mixed BC subtypes	Adjacent	qRT-PCR	No	No	Inaki et al. [62]
<i>SCNN1A-TNFRSF1A</i>	Triple-negative and ER ⁺ BC	3/42 (7 %) TN	TIC	qRT-PCR	No	No	Varley et al. [67]
<i>CTSD-IFITM10</i>	Triple-negative and ER ⁺ BC	5/42 (12 %) ER ⁺	TIC	qRT-PCR	No	Yes	Varley et al. [67]
<i>SLC2A1-FAF1</i>	Micropapillary Carcinomas	9/42 (21 %) TN	Intrachr	qRT-PCR	No	Yes	Natrajan et al. [81]
<i>BCAS4-AURKA</i>		5/42 (12 %) ER ⁺					
<i>EEF1D3-FRY</i>	No	1/5 MPC	TIC	qRT-PCR	No	No	Kim et al. [82]

TIC Transcription-induced chimera. BC Breast cancer. TN Triple-negative. SBC Secretory breast carcinoma. MPC Micropapillary carcinoma

tissues. Further, four additional SBC cases were examined and three of them expressed *ETV6-NTRK3* fusion. These results suggest *ETV6-NTRK3* as the characteristic genetic alteration of SBC. Although restricted to a rare subset of breast cancers, this fusion represents a dominantly acting oncogene in breast cancer.

Further, the *EML4-ALK* fusion initially detected in NSCLC [8, 56] was later also detected in breast cancer. However, the incidence of *EML4-ALK* in breast cancer remains uncertain as a result of inconsistent results from different studies. While the study by Lin et al. in 2009 reported the expression of this fusion in 2.4 % of breast cancers [57], the study by Fukuyoshi et al. did not find any *EML4-ALK* rearrangement during their analysis of 90 breast cancer cases in 2008 [58], and the sequence analysis of 65 triple-negative breast cancers by Grob et al. failed to detect any *ALK* rearrangements [59]. The results of all these studies using different breast cancer tissue cohorts, though inconsistent, could indicate rare *EML4-ALK* rearrangements in breast cancer as opposed to NSCLC. Another recent discovery of *MAGI3-AKT3* fusion in breast cancer started off as a promising recurrent rearrangement with therapeutic implications in triple-negative breast cancer [41] but was subsequently amended. Mosquera et al. examined 236 triple-negative breast cancer samples and failed to detect these rearrangements in any of the samples [60]; Pugh et al. analyzed the 3 positive TNBC cases from original screen [41] as well as additional tumors by exome plus hybrid capture and illumina sequencing and observed that *MAGI3-AKT3* is expressed in only one of the index breast cancer cases, suggesting that this fusion may be a private event [61].

Another recently reported recurrent *RPS6KB1-VMP1* gene fusion expressed in 30 % of breast cancers involves adjacent genes and is generated by tandem duplication in 17q23 [62]. However, this fusion was also detected in low levels in the normal breast tissue, and the chimeric protein does not contain a functional protein domain. This fusion is proposed to serve as an indicator of genomic instability at the 17q23 locus which leads to gene amplification and/or overexpression of crucial oncogenic elements such as *MIR21* and *RPS6KB1*. The authors concluded that *RPS6KB1-VMP1* fusion is not a driver in tumor development. Several gene fusions involving FGFR family member as 5' or 3' fusion partner such as *ERLIN2-FGFR1*, *FGFR2-AFF3*, *FGFR2-CASP7*, and *FGFR2-CCDC6* are identified in breast cancer [63]. Interestingly, all fusions retain the intact kinase domain of FGFRs, suggesting their potential functionality and the *ERLIN2-FGFR1* and *FGFR2-CCDC6* fusions produce an active FGFR kinase. Moreover, the FGFR fusion partners in these fusions are proposed to mediate oligomerization, thereby triggering the activation of respective FGFR kinase. Similarly, another

transcriptome sequencing study by Robinson et al. in a panel of 89 breast cancer cell lines and tumors identified several MAST kinase gene fusions—*ARID1A-MAST2*, *GPBP1L1-MAST2*, *ZNF700-MAST1*, *NFIX-MAST1*, and *TADA2A-MAST1* involving MAST family members as the 5' or 3' fusion partner in 3–5 % of breast cancers [10]. All five fusions retain the PDZ domain and the 3' kinase-like domain of MAST kinase. Additional sporadic kinase fusions identified in breast cancer include *ERC1-RET* and *PDGFRA-KIT* fusions identified in one case and, the *TBLIXR1-PIK3CA* fusion detected in 2 cases [46].

Gene fusions involving transcription factors

Transcription factors are the master regulators of the expression of multiple downstream target genes, and so their involvement in any gene fusion event could have substantial consequences not only on itself but also on its downstream transcriptional targets. The transcriptome sequencing study by Robinson et al. in a panel of 89 breast cancer cell lines and tumors identified 8 rearrangements involving transcription factors, *NOTCH1* or *NOTCH2*—*SEC16A-NOTCH1*, and *SEC22B-NOTCH2* among which *SEC16A-NOTCH1* fusion showed recurrence in 2 out of 89 samples [10]. NOTCH proteins function as receptors for membrane-bound ligands Jagged1/2 and Delta1 to regulate cell differentiation, proliferation, and apoptotic programs. Upon ligand activation, the protein fragment containing NOTCH intracellular domain (NICD) will be released and form transcriptional activator complex. Interestingly, all the *NOTCH* translocations were observed in ER-negative cases, and the fusion open reading frames retain the NICD domain, which might lead to constitutive activation of notch receptor. Subsequently, Clay et al. analyzed 501 breast cancer tissues and observed 5 cases with *NOTCH1* rearrangements [64]. However, the *NOTCH1* fusions in this study were all observed in ER-positive cases as opposed to previous observations by Robinson et al. in ER-negative cases. This could be in part attributed to the differences in the detection techniques and breast cancer tissue cohorts utilized in these studies. Further studies are necessary before drawing definite conclusions about the demographic distribution of these fusions in breast cancer.

Transcription-induced chimeras

In addition to the largely known genomic events generating gene fusions, RNA-seq is unraveling yet another class of gene fusions generated by intergenic splicing, termed as the transcription-induced chimeras (TICs). Such chimeras are most frequently generated by cis-splicing between

collinearly positioned neighboring genes (also known as read-through events), and also occasionally by trans-splicing between noncollinearly positioned genes or distant genes [65]. Interestingly, at least 4–5 % of the tandem gene pairs in human genome are estimated to generate read-through events [66]. It has been suggested that most of the frequently expressed gene fusions are the result of TIC events [67]. Although many TICs are observed to be considerably expressed in normal human tissues as well, some are selectively overexpressed in human cancers, suggesting their functional significance in cancer development [68], as exemplified by the *SLC45A3-ELK4* fusion identified in prostate cancer [69]. Some TICs such as *SCNN1A-TNFRSF1A*, and *CTSD-IFITM10* [67] are identified to be more frequently expressed in breast cancers compared to normal breast tissues. Both the fusions involve membrane proteins suggesting that these fusions could be breast cancer-specific cell surface markers. Despite the lack of tumor specificity or tissue specificity as opposed to genomic gene fusions, the functional significance of TIC in breast cancer cannot be excluded. Further studies are required to determine whether these events are pathological, or merely passenger events.

The pathological role of gene fusions in breast cancer

Gene fusions are powerful drivers of cancer, and can generate novel chimeric proteins, change gene expression levels, alter protein activities, force oligomerization, or change the subcellular localization of a protein [11]. Functional characterization of fusion genes in breast cancer primarily focuses on elucidating their role in tumorigenesis and therapeutic resistance. To examine the biological effects of fusion genes, fusion transcripts are either ectopically overexpressed in normal breast epithelial cells and breast cancer cells, or silenced in endogenous fusion-positive breast cancer cell lines. Some studies also explore the *in vivo* phenotypes endowed by these fusion genes using breast cancer xenograft/orthotopic models [38, 70]. Various recurrent gene fusions have been shown to possess oncogenic activity through promoting cancer cell proliferation and/or migration [10, 38, 67, 70]. Fusion genes might drive cancer progression through a) generating constitutively active kinases or transcription factors or amplify the growth factor signaling, b) inactivating apoptotic factors and promote uncontrollable cell growth.

One of the well-studied gene fusions in breast cancer is the *ETV6-NTRK3* fusion discovered in SBC. This fusion encodes a chimeric protein comprising the oligomerization domain of ETV6 and PTK domain of NTRK3 [52]. *ETV6-NTRK3* expression in mammary tissues results in the

development of a fully penetrant, multifocal malignant breast cancer with short latency. Moreover, studies have established that the *ETV6-NTRK3* fusion protein is sufficient to initiate mammary tumorigenesis. Interestingly, upregulation and activation of the c-Jun/Fos1 AP1 complex and several of the AP1 target genes including cyclin D1 have been observed in *ETV6-NTRK3*-expressing tumors, which stimulate cell proliferation. Moreover, activation of c-Jun/Fos1 complex and establishment of ETV6-NTRK3/AP1 invasiveness program might be early events in breast cell transformation [71]. These results suggest the critical role of AP1 complex in ETV6-NTRK3-mediated breast cancer development and progression.

Gene fusions involving the estrogen receptor gene (*ESR1*) is preferentially found in luminal B or endocrine-resistant tumors. In our previous study, we have identified recurrent genomic rearrangements between *ESR1* and its neighbor gene *CCDC170* in the more aggressive and endocrine-resistant luminal B tumors (6–8 %) [38]. The observed fusion joins the 5'-untranslated region of *ESR1* to the coding region of *CCDC170*, leading to the expression of truncated CCDC170 protein variants under *ESR1* promoter (Fig. 1b). Thus, the *ESR1* gene does not contribute to the coding sequences. The expression of *ESR1-CCDC170* was associated with significant upregulation of Gab1, a key docking protein that enhances the signaling of many receptor tyrosine kinases and a key scaffold protein involved in the formation of invadopodia [72], along with activation of its downstream signaling molecules, AKT and ERK [38]. Our data show that Gab1 appears to play a role in the increased invasiveness driven by this fusion. However, further studies are required to determine the role of this fusion in breast cancer endocrine resistance and how the fusion proteins could engage Gab1 and interfere with signaling pathways, as there is very little knowledge regarding the role of *CCDC170* in either normal or cancer cells. Interestingly, genome-wide association studies (GWAS) have revealed that *CCDC170* locus is associated with breast cancer susceptibility in women [73, 74]. Another latest fine-scale mapping study revealed the existence of causal genetic variants regulating *CCDC170* expression with a direct effect on breast cancer risk [75]. Together, these studies strengthen the significance of *CCDC170* locus in breast cancer development.

The discovery of recurrent *ESR1-CCDC170* fusion in luminal B subtype is important as it sheds light on the significance of gene fusions as oncogenic events that may promote the aggressive form of ER+ breast cancer. As one of the lead studies in the identification of recurrent and pathological gene fusions in breast cancer, it also highlights the necessity and significance of further elucidating the pathological role of recurrent gene fusions in breast cancer. The other *ESR1* fusions reported thus far, such as *ESR1-*

YAP1, *ESR1-POLH*, and *ESR1-AKAP12*, share distinct coding structures from *ESR1-CCDC170* whereby the fusions result in truncated ER fragments that lack the hormone-dependent transactivation domain (AF2) and ligand-binding domain, but retain the hormone-independent transactivation domain (AF1) and DNA-binding domain. Among these, the function of *ESR1-YAP1* fusion identified in an endocrine therapy-resistant PDX model is best documented. The *ESR1-YAP1* fusion has been shown to possess constitutive ER transcriptional activity and estrogen-independent signaling and thereby induce estradiol-independent cell growth and promote endocrine resistance [51, 76]. Therefore, the *ESR1* gene fusions generating chimeric ER proteins may represent a recurrent mechanism for development of endocrine resistance via constitutively active ER.

For gene fusions involving partners whose functions have been well documented, such as the NOTCH gene family, studies about the underlying mechanism of how these fusions might lead to a malignant phenotype are extensively guided by existing knowledge about the fusion partners. The functional recurrent NOTCH fusion transcripts retain the exons encoding NICD that is responsible for NOTCH-induced transcriptional activities, and the fusion-positive cells exhibit substantially enhanced activation of NOTCH pathway [10]. The endogenous NOTCH fusion-positive breast cancer cell lines as well as the TERT-HME1 cells ectopically expressing NOTCH fusion exhibited decreased cell–matrix adhesion and grow in suspension or as weakly attached clusters. In addition, the NOTCH fusion index cell lines displayed a dependence on NOTCH signaling for proliferation and survival. Deregulated NOTCH signaling has been reported to be oncogenic in several tumor types, for example somatic activating mutations of NOTCH1 are present in more than 50 % of T-cell acute lymphoblastic leukemia [77]. Therefore, these previous studies laid important foundation for studying the role of activated NOTCH1 in the molecular pathogenesis of breast tumors harboring NOTCH1-activating fusions and also provide a rationale for targeted therapies that interfere with the NOTCH signaling pathway.

On the contrary, some gene fusions discovered in breast cancer involve fusion partners that are not well-characterized functionally or biologically. Although structurally well-characterized, the biological/functional role of MAST kinase family genes in cancer has not been well studied. Ectopic expression of the MAST kinase family fusions in benign breast epithelial and breast cancer cells induced cell proliferation and endowed growth advantage [10]. Moreover, following MAST2 knockdown, the endogenous *ARID1A-MAST2* fusion-positive breast cancer cell line failed to develop tumors in vivo. This suggests that

ARID1A-MAST2 could function as a key oncogenic driver at least in this breast cancer cell line. A more detailed investigation using additional breast cancer cell lines is necessary to establish the oncogenic potential of these fusions in breast cancer as well as to provide mechanistic insights. On the other hand, in case of FGFR family fusions, despite involving biologically significant fusion partners, their implication in breast cancer development and progression is uncertain due to the rarity of their occurrence. Overexpression of these fusions was observed to induce the proliferation of 293T cells and TERT-HME cells, while knockdown of FGFR-BAIAP2L1 in the index bladder cancer cell line significantly suppressed cell proliferation [63]. Moreover, FGFR small-molecule inhibitor treatment inhibited the growth of xenografts developed from FGFR fusion expressing bladder cancer cell lines. Although these results suggest the significance of *FGFR* fusions in cancer, the impact of these fusions specifically in breast cancer cells has not yet been studied.

Together, all the recurrent gene fusions currently identified in breast cancer invariably have a key role in promoting tumorigenesis and endocrine resistance. Despite our increased understanding toward the transforming power of fusion genes, further studies are needed to comprehensively understand the mechanisms utilized by these fusions to promote breast cancer development/progression. These studies will not only enrich our knowledge about the disease, but also provide valuable concepts about how we might treat breast cancers harboring these fusions.

Clinical implications and future directions

Despite the high occurrence of gene fusion events, the biological significance and clinical implications of gene fusions in breast cancer have been largely elusive. From a therapeutic perspective, gene fusions that involve well-characterized and targetable fusion partners are of particular interest. This is represented by the *ETV6-NTRK3* gene fusion. Clinical trials are currently recruiting breast cancer patients harboring *NTRK* rearrangements to test the efficacy of Entrectinib, an orally available inhibitor of the tyrosine kinases, NTRKs, ROS, and ALK [<https://clinicaltrials.gov/show/NCT02097810>, <https://clinicaltrials.gov/show/NCT02568267>]. In addition, Crizotinib, the ALK inhibitor currently in clinical trials, has been shown to inhibit ETV6-NTRK3 fusion kinase, block proliferation of ETV6-NTRK3-dependent tumor cells, and induce the regression of tumor xenografts in mice [78]. Further, the breast cancer cells harboring ETV6-NTRK3 fusion show upregulation and activation of AP1 complex. TAM67, a dominant negative inhibitor of AP1 shown to inhibit the growth of malignant breast cells [79], suggests that

targeting AP1 complex in the *ETV6-NTRK3*-positive breast cancer cases might also be beneficial.

Further, the study of functional recurrent NOTCH gene family fusions indicated that NOTCH1 rearrangements were associated with high levels of activated NOTCH1 (N1-ICD) and were sensitive to the gamma-secretase inhibitor (GSI), MRK-003 [70]. This study not only provided a feasible rationale for the treatment of breast cancers harboring NOTCH1 rearrangements, but also offers valuable insights about the functional recurrent fusion gene families which were largely overlooked previously. For example, the *FGFR2-AFF3*, *FGFR2-CASP7*, *FGFR2-CCDC6*, and *ERLIN2-FGFR1* fusions that join the FGFR family members to various 3' or 5' gene partners have been identified in metastatic breast cancers [63]. Although none of these fusion genes showed recurrence in breast tumors, all the fusion partners exhibited oligomerization capability, and thereby might contribute to upregulated FGFR signaling through forced oligomerization of the growth factor receptor [63]. Inhibition of FGFR might offer substantial therapeutic potential in treating metastatic breast cancers. Taken together, although activating gene fusions involving druggable gene families such as *NOTCH*, *MAST*, *FGFR* are mostly private events, these families of gene fusions in accrual will affect a significant subset of patients that could be potentially manageable through targeting the common targetable partners.

In addition to serving as viable therapeutic targets, gene fusions also bear potential prognostic and diagnostic significance in breast cancer. For example, luminal B breast tumors are notoriously known to develop de novo resistance to endocrine therapy, and clinically it is even difficult to clearly define this more aggressive form of ER+ breast cancer due to the lack of reliable and accurate genetic biomarkers. Current available classification methods like PAM50 gene expression profile or Ki67 index are limited by controversial performance, and the patients, especially those who are on the borderline, may be misclassified and thus may not receive the appropriate treatment. Our discovery of the *ESR1-CCDC170* fusions in luminal B breast cancers [38] provided a potential genetic biomarker for defining and subtyping luminal B breast cancers. Detection of the *ESR1-CCDC170* gene fusion may be used as an independent or companion diagnostic to screen for patients who may harbor this more aggressive form of breast cancer that may require advanced treatment. In addition, several studies have suggested activating mutations in *ESR1* as a key mechanism in acquired endocrine resistance in breast cancer therapy [47, 49, 51]. Given that some of the *ESR1* rearrangements have been shown to be associated with endocrine resistance, such as the *ESR1-YAP1*, future discovery of such rearrangements will have critical clinical implications in predicting endocrine-resistant breast

cancer. In addition, evaluating the druggability of these *ESR1* gene fusions including *ESR1-CCDC170* will be of great interest for the development of new targeted therapies.

Conclusion and perspectives

In summary, integrative genomic research methodologies play an indispensable role in gene fusion discoveries in breast cancer. In the era of precision medicine, molecular subtyping of breast tumors is of utmost importance for genetic characterization of breast cancer subtypes and identification of effective treatment strategies. Current studies have identified diverse recurrent or pathological fusion genes in breast cancer that may drive the development and progression of more aggressive tumors. Despite the diversity of most fusion events in breast cancer, our finding of the recurrent *ESR1-CCDC170* gene fusion suggests the presence of true recurrent genomic fusions in aggressive breast cancer subtypes, and more such pathological recurrent events are yet to be discovered.

While several deep sequencing studies of breast cancer transcriptome or genome have concluded that breast cancer genome exhibits complex rearrangement pattern, therefore recurrent gene fusions are rare, we believe that more significant recurrent gene fusions are still to be uncovered in breast cancer for the following reasons. First, the sensitivity of current NGS technology in detecting gene fusions may not be as good as expected. One of the reasons is that in gene fusion detection, only paired-reads spanning or encompassing the fusion junction are identifiable as chimeric reads, while reads from rest of the fusion sequences will be mapped to wild-type genes. This will result in substantially lower sensitivity of fusion detection compared to wild-type genes. In addition, deep sequencing technologies are not sensitive to the high-GC or high-AT sequences [80], which may be present in fusion junctions. Second, RNA-seq routinely generates a daunting quantity of chimerical sequences, most of which are artifactual fusion sequences, TICs, or nonfunctional passenger gene fusions. Third, the prohibiting costs of WGS technology have prevented large-scale deep sequencing of breast cancer genomes and the short read lengths of current WGS technologies impede the de novo assembly of fusion sequences. Fourth, although the primary tumors of breast cancer have been intensively sequenced with the current deep sequencing technologies, it will be interesting to further explore whether pathological recurrent gene fusions exist in the special breast tumor entities such as metastatic breast cancer, endocrine-resistant breast cancer, or Her2 therapy-resistant breast cancer. Therefore, it is reasonable to believe that with more advances in the sequencing

technologies, extended sequencing of special breast tumor entities, as well as improvements in bioinformatics analyses, more recurrent and pathologically important gene fusions in breast cancers could be discovered in the near future.

Acknowledgments JV is supported by the Congressionally Directed Medical Research Programs fellowship, W81XWH-13-1-0431. X-SW is supported by the NIH grants 1R01CA181368 and 1R01CA183976, and Commonwealth of PA Tobacco Phase 15 Formula Fund.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals.

References

- Mitelman F, Johansson B, Mertens F (2007) The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 7(4):233–245. doi:10.1038/nrc2091
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR (2004) A census of human cancer genes. *Nat Rev Cancer* 4(3):177–183. doi:10.1038/nrc1299
- Ren R (2005) Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* 5(3):172–183. doi:10.1038/nrc1567
- Goldman JM, Melo JV (2003) Chronic myeloid leukemia—advances in biology and new approaches to treatment. *New Engl J Med* 349(15):1451–1464. doi:10.1056/NEJMra020777
- de Klein A, van Kessel AG, Grosveld G, Bartram CR, Hagemeyer A, Bootsma D, Spurr NK, Heisterkamp N, Groffen J, Stephenson JR (1982) A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature* 300(5894):765–767
- Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM (2015) Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Gen Med* 7:129. doi:10.1186/s13073-015-0252-1
- Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310(5748):644–648. doi:10.1126/science.1117679
- Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448(7153):561–566. doi:10.1038/nature05945
- Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, Choi HG, Kim J, Chiang D, Thomas R, Lee J, Richards WG, Sugarbaker DJ, Ducco C, Lindeman N, Marcoux JP, Engelman JA, Gray NS, Lee C, Meyerson M, Janne PA (2008) EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin Cancer Res* 14(13):4275–4283. doi:10.1158/1078-0432.CCR-08-0168
- 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 (2011) Functionally recurrent rearrangements of the MAST kinase and notch gene families in breast cancer. *Nat Med* 17(12):1646–1651. doi:10.1038/nm.2580
- Edwards PA (2010) Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol* 220(2):244–254. doi:10.1002/path.2632
- Mertens F, Johansson B, Fioretos T, Mitelman F (2015) The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 15(6):371–381. doi:10.1038/nrc3947
- Rowley JD, Beau MML, Rabbitts TH (2015) Chromosomal Translocations and Genome Rearrangements in Cancer. Springer International Publishing
- Malhotra GK, Zhao X, Band H, Band V (2010) Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther* 10(10):955–960
- Cancer Genome Atlas N (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61–70. doi:10.1038/nature11412
- Chang M (2012) Tamoxifen resistance in breast cancer. *Biomol Ther* 20(3):256–267. doi:10.4062/biomolther.2012.20.3.256
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New Engl J Med* 344(11):783–792. doi:10.1056/NEJM200103153441101
- Montemurro F, Valabrega G, Aglietta M (2004) Trastuzumab-based combination therapy for breast cancer. *Expert Opin Pharmacother* 5(1):81–96. doi:10.1517/14656566.5.1.81
- Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 20(3):719–726
- Osborne CK, Schiff R (2011) Mechanisms of endocrine resistance in breast cancer. *Annu Rev Med* 62:233–247. doi:10.1146/annurev-med-070909-182917
- Carrara M, Beccuti M, Lazzarato F, Cavallo F, Cordero F, Donatelli S, Calogero RA (2013) State-of-the-art fusion-finder algorithms sensitivity and specificity. *BioMed Res Int* 2013:340620. doi:10.1155/2013/340620
- Kumar S, Vo AD, Qin F, Li H (2016) Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci Rep* 6:21597. doi:10.1038/srep21597
- McCarthy A (2010) Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol* 17(7):675–676. doi:10.1016/j.chembiol.2010.07.004
- English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7(11):e47768. doi:10.1371/journal.pone.0047768
- Kim KS, Peluso P, Haw E, Sisneros N, Rank D, Badgett MR, Gu JG (2015) Best practices for whole genome de novo sequencing with long-read smrt sequencing. Poster Session Presented at: Genome Technologies—Large Insert Libraries 23rd International Plant and Animal Genome Conference San Diego, CA
- Leprohon P, Fernandez-Prada C, Gazanion E, Monte-Neto R, Ouellette M (2015) Drug resistance analysis by next generation sequencing in Leishmania. *Int J Parasitol Drugs and Drug Resistance* 5(1):26–35. doi:10.1016/j.ijpddr.2014.09.005

27. Latysheva NS, Babu MM (2016) Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res*. doi:[10.1093/nar/gkw282](https://doi.org/10.1093/nar/gkw282)
28. Benelli M, Pescucci C, Marseglia G, Severgnini M, Torricelli F, Magi A (2012) Discovering chimeric transcripts in paired-end RNA-seq data by using *ericScript*. *Bioinformatics* 28(24):3232–3239. doi:[10.1093/bioinformatics/bts617](https://doi.org/10.1093/bioinformatics/bts617)
29. Liu S, Tsai WH, Ding Y, Chen R, Fang Z, Huo Z, Kim S, Ma T, Chang TY, Priedigkeit NM, Lee AV, Luo J, Wang HW, Chung IF, Tseng GC (2016) Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res* 44(5):e47. doi:[10.1093/nar/gkv1234](https://doi.org/10.1093/nar/gkv1234)
30. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, Yu Y, Zhu D, Nickerson ML, Wan S, Liao X, Zhu X, Peng S, Li Y, Wang J, Guo G (2013) SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Gen Biol* 14(2):R12. doi:[10.1186/gb-2013-14-2-r12](https://doi.org/10.1186/gb-2013-14-2-r12)
31. Nicorici D, Satalan M, Edgren H, Kangaspeska S, Murumagi A, Kallioniemi O, Virtanen S, Kilkku O (2014) Fusioncatcher—a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv*. doi:[10.1101/011650](https://doi.org/10.1101/011650)
32. Davidson NM, Majewski IJ, Oshlack A (2015) JAFFA: high sensitivity transcriptome-focused fusion gene detection. *Gen Med* 7(1):43. doi:[10.1186/s13073-015-0167-x](https://doi.org/10.1186/s13073-015-0167-x)
33. Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, Wilson RK, Maher CA (2016) INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Gen Res* 26(1):108–118. doi:[10.1101/gr.186114.114](https://doi.org/10.1101/gr.186114.114)
34. McPherson A, Wu C, Hajirasouliha I, Hormozdiari F, Hach F, Lapuk A, Volik S, Shah S, Collins C, Sahinalp SC (2011) Comrad: detection of expressed rearrangements by integrated analysis of RNA-Seq and low coverage genome sequence data. *Bioinformatics* 27(11):1481–1488. doi:[10.1093/bioinformatics/btr184](https://doi.org/10.1093/bioinformatics/btr184)
35. Chen K, Navin NE, Wang Y, Schmidt HK, Wallis JW, Niu B, Fan X, Zhao H, McLellan MD, Hoadley KA, Mardis ER, Ley TJ, Perou CM, Wilson RK, Ding L (2013) BreakTrans: uncovering the genomic architecture of gene fusions. *Gen Biol* 14(8):R87. doi:[10.1186/gb-2013-14-8-r87](https://doi.org/10.1186/gb-2013-14-8-r87)
36. Li H, Wang J, Mor G, Sklar J (2008) A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* 321(5894):1357–1361. doi:[10.1126/science.1156725](https://doi.org/10.1126/science.1156725)
37. Fernandez-Cuesta L, Sun R, Menon R, George J, Lorenz S, Meza-Zepeda LA, Peifer M, Plenker D, Heuckmann JM, Leenders F, Zander T, Dahmen I, Koker M, Schottle J, Ullrich RT, Altmuller J, Becker C, Nurnberg P, Seidel H, Bohm D, Goke F, Ansen S, Russell PA, Wright GM, Wainer Z, Solomon B, Petersen I, Clement JH, Sanger J, Brustugun OT, Helland A, Solberg S, Lund-Iversen M, Buettner R, Wolf J, Brambilla E, Vingron M, Perner S, Haas SA, Thomas RK (2015) Identification of novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data. *Gen Biol* 16:7. doi:[10.1186/s13059-014-0558-0](https://doi.org/10.1186/s13059-014-0558-0)
38. Veeraghavan J, Tan Y, Cao XX, Kim JA, Wang X, Chamness GC, Maiti SN, Cooper LJ, Edwards DP, Contreras A, Hilsenbeck SG, Chang EC, Schiff R, Wang XS (2014) Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun* 5:4577. doi:[10.1038/ncomms5577](https://doi.org/10.1038/ncomms5577)
39. Wang XS, Prensner JR, Chen G, Cao Q, Han B, Dhanasekaran SM, Ponnala R, Cao X, Varambally S, Thomas DG, Giordano TJ, Beer DG, Palanisamy N, Sartor MA, Omenn GS, Chinnaiyan AM (2009) An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nature Biotechnol* 27(11):1005–1011. doi:[10.1038/nbt.1584](https://doi.org/10.1038/nbt.1584)
40. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, Verhaak RG (2015) The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* 34(37):4845–4854. doi:[10.1038/onc.2014.406](https://doi.org/10.1038/onc.2014.406)
41. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Pina V, Duke F, Francis J, Jung J, Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, Thompson KM, Uribe-Figueroa L, Baselga J, Beroukheim R, Polyak K, Sgroi DC, Richardson AL, Jimenez-Sanchez G, Lander ES, Gabriel SB, Garraway LA, Golub TR, Melendez-Zajgla J, Tokar A, Getz G, Hidalgo-Miranda A, Meyerson M (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486(7403):405–409. doi:[10.1038/nature11154](https://doi.org/10.1038/nature11154)
42. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerod A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Borresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462(7276):1005–1010. doi:[10.1038/nature08645](https://doi.org/10.1038/nature08645)
43. Kalyana-Sundaram S, Shankar S, Deroo S, Iyer MK, Palanisamy N, Chinnaiyan AM, Kumar-Sinha C (2012) Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer. *Neoplasia* 14(8):702–708
44. Sequist LV, Heist RS, Shaw AT, Fidias P, Rosovsky R, Temel JS, Lennes IT, Digumarthy S, Waltman BA, Bast E, Tammireddy S, Morrissey L, Muzikansky A, Goldberg SB, Gainer J, Channick CL, Wain JC, Gaissert H, Donahue DM, Muniappan A, Wright C, Willers H, Mathisen DJ, Choi NC, Baselga J, Lynch TJ, Ellisen LW, Mino-Kenudson M, Lanuti M, Berger DR, Iafrate AJ, Engelman JA, Dias-Santagata D (2011) Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. *Annals Oncol* 22(12):2616–2624. doi:[10.1093/annonc/mdr489](https://doi.org/10.1093/annonc/mdr489)
45. Kris M, Johnson B, Kwiatkowski D, Iafrate AJ, Wistuba II, Aronson SL, et al. (2011) Identification of driver mutations in tumor specimens from 1,000 patients with lung adenocarcinoma: The NCI's Lung Cancer Mutation Consortium (LCMC). *J Clin Oncol* 29: 2011 (suppl); abstr CRA7506)
46. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C (2014) The landscape of kinase fusions in cancer. *Nat Commun* 5:4846. doi:[10.1038/ncomms5846](https://doi.org/10.1038/ncomms5846)
47. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, Li Z, Gala K, Fanning S, King TA, Hudis C, Chen D, Taran T, Hortobagyi G, Greene G, Berger M, Baselga J, Chandarlapaty S (2013) ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nature Gen* 45(12):1439–1445. doi:[10.1038/ng.2822](https://doi.org/10.1038/ng.2822)
48. Merenbakh-Lamin K, Ben-Baruch N, Yeheskel A, Dvir A, Soussan-Gutman L, Jeselsohn R, Yelensky R, Brown M, Miller VA, Sarid D, Rizel S, Klein B, Rubinek T, Wolf I (2013) D538G mutation in estrogen receptor-alpha: a novel mechanism for acquired endocrine resistance in breast cancer. *Cancer research* 73(23):6856–6864. doi:[10.1158/0008-5472.CAN-13-1197](https://doi.org/10.1158/0008-5472.CAN-13-1197)
49. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, Kalyana-Sundaram S, Wang R, Ning Y, Hodges L, Gursky A, Siddiqui J, Tomlins SA, Roychowdhury S, Pienta KJ, Kim SY, Roberts JS, Rae JM, Van Poznak CH, Hayes DF, Chugh R, Kunju LP, Talpaz M, Schott AF, Chinnaiyan AM (2013) Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nature Gen* 45(12):1446–1451. doi:[10.1038/ng.2823](https://doi.org/10.1038/ng.2823)

50. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, Ferrer-Lozano J, Perez-Fidalgo JA, Cristofanilli M, Gomez H, Arteaga CL, Giltane J, Balko JM, Cronin MT, Jarosz M, Sun J, Hawryluk M, Lipson D, Otto G, Ross JS, Dvir A, Soussan-Gutman L, Wolf I, Rubinek T, Gilmore L, Schnitt S, Come SE, Pusztai L, Stephens P, Brown M, Miller VA (2014) Emergence of constitutively active estrogen receptor- α mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* 20(7):1757–1767. doi:[10.1158/1078-0432.CCR-13-2332](https://doi.org/10.1158/1078-0432.CCR-13-2332)
51. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, He X, Liu S, Hoog J, Lu C, Ding L, Griffith OL, Miller C, Larson D, Fulton RS, Harrison M, Mooney T, McMichael JF, Luo J, Tao Y, Goncalves R, Schlosberg C, Hiken JF, Saied L, Sanchez C, Giuntoli T, Bumb C, Cooper C, Kitchens RT, Lin A, Phommaly C, Davies SR, Zhang J, Kavuri MS, McEachern D, Dong YY, Ma C, Pluard T, Naughton M, Bose R, Suresh R, McDowell R, Michel L, Aft R, Gillanders W, DeSchryver K, Wilson RK, Wang S, Mills GB, Gonzalez-Angulo A, Edwards JR, Maher C, Perou CM, Mardis ER, Ellis MJ (2013) Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* 4(6):1116–1130. doi:[10.1016/j.celrep.2013.08.022](https://doi.org/10.1016/j.celrep.2013.08.022)
52. Knezevich SR, McFadden DE, Tao W, Lim JF, Sorensen PH (1998) A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nature Gen* 18(2):184–187. doi:[10.1038/ng0298-184](https://doi.org/10.1038/ng0298-184)
53. Knezevich SR, Garnett MJ, Pysher TJ, Beckwith JB, Grundy PE, Sorensen PH (1998) ETV6-NTRK3 gene fusions and trisomy 11 establish a histogenetic link between mesoblastic nephroma and congenital fibrosarcoma. *Cancer Res* 58(22):5046–5048
54. Tognon C, Knezevich SR, Huntsman D, Roskelley CD, Melnyk N, Mathers JA, Becker L, Carneiro F, MacPherson N, Horsman D, Poremba C, Sorensen PH (2002) Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* 2(5):367–376
55. Makretsov N, He M, Hayes M, Chia S, Horsman DE, Sorensen PH, Huntsman DG (2004) A fluorescence in situ hybridization study of ETV6-NTRK3 fusion gene in secretory breast carcinoma. *Genes Chromosomes Cancer* 40(2):152–157. doi:[10.1002/gcc.20028](https://doi.org/10.1002/gcc.20028)
56. Mano H (2008) Non-solid oncogenes in solid tumors: eML4-ALK fusion genes in lung cancer. *Cancer Sci* 99(12):2349–2355. doi:[10.1111/j.1349-7006.2008.00972.x](https://doi.org/10.1111/j.1349-7006.2008.00972.x)
57. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, Pandita A, Tang J, Modrusan Z (2009) Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. *Mol Cancer Res* 7(9):1466–1476. doi:[10.1158/1541-7786.MCR-08-0522](https://doi.org/10.1158/1541-7786.MCR-08-0522)
58. Fukuyoshi Y, Inoue H, Kita Y, Utsunomiya T, Ishida T, Mori M (2008) EML4-ALK fusion transcript is not found in gastrointestinal and breast cancers. *Br J cancer* 98(9):1536–1539. doi:[10.1038/sj.bjc.6604341](https://doi.org/10.1038/sj.bjc.6604341)
59. Grob TJ, Heilenkotter U, Geist S, Paluchowski P, Wilke C, Jaenicke F, Quaas A, Wilczak W, Choschzick M, Sauter G, Lebeau A (2012) Rare oncogenic mutations of predictive markers for targeted therapy in triple-negative breast cancer. *Breast Cancer Res Treat* 134(2):561–567. doi:[10.1007/s10549-012-2092-7](https://doi.org/10.1007/s10549-012-2092-7)
60. Mosquera JM, Varma S, Pauli C, MacDonald TY, Yashinskii JJ, Varga Z, Sboner A, Moch H, Rubin MA, Shin SJ (2015) MAGI3-AKT3 fusion in breast cancer amended. *Nature* 520(7547):E11–E12. doi:[10.1038/nature14265](https://doi.org/10.1038/nature14265)
61. Pugh TJ, Banerji S, Meyerson M (2015) Pugh et al. reply. *Nature* 520(7547):E12–14. doi:[10.1038/nature14266](https://doi.org/10.1038/nature14266)
62. Inaki K, Hillmer AM, Ukil L, Yao F, Woo XY, Vardy LA, Zawack KF, Lee CW, Ariyaratne PN, Chan YS, Desai KV, Bergh J, Hall P, Putti TC, Ong WL, Shahab A, Cacheux-Rataboul V, Karuturi RK, Sung WK, Ruan X, Bourque G, Ruan Y, Liu ET (2011) Transcriptional consequences of genomic structural aberrations in breast cancer. *Gen Res* 21(5):676–687. doi:[10.1101/gr.113225.110](https://doi.org/10.1101/gr.113225.110)
63. Wu YM, Su F, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao X, Lonigro RJ, Vats P, Wang R, Lin SF, Cheng AJ, Kunju LP, Siddiqui J, Tomlins SA, Wyngaard P, Sadis S, Roychowdhury S, Hussain MH, Feng FY, Zalupski MM, Talpaz M, Pienta KJ, Rhodes DR, Robinson DR, Chinnaiyan AM (2013) Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov* 3(6):636–647. doi:[10.1158/2159-8290.CD-13-0050](https://doi.org/10.1158/2159-8290.CD-13-0050)
64. Clay MR, Varma S, West RB (2013) MAST2 and NOTCH1 translocations in breast carcinoma and associated pre-invasive lesions. *Hum Pathol* 44(12):2837–2844. doi:[10.1016/j.humpath.2013.08.001](https://doi.org/10.1016/j.humpath.2013.08.001)
65. Ren G, Zhang Y, Mao X, Liu X, Mercer E, Marzec J, Ding D, Jiao Y, Qiu Q, Sun Y, Zhang B, Yeste-Velasco M, Chelala C, Berney D, Lu YJ (2014) Transcription-mediated chimeric RNAs in prostate cancer: time to revisit old hypothesis? *Omic: J Integr Biol* 18(10):615–624. doi:[10.1089/omi.2014.0042](https://doi.org/10.1089/omi.2014.0042)
66. Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, Thomson TM, Antonarakis SE, Guigo R (2006) Tandem chimerism as a means to increase protein complexity in the human genome. *Gen Res* 16(1):37–44. doi:[10.1101/gr.4145906](https://doi.org/10.1101/gr.4145906)
67. Varley KE, Gertz J, Roberts BS, Davis NS, Bowling KM, Kirby MK, Nesmith AS, Oliver PG, Grizzle WE, Forero A, Buchsbaum DJ, LoBuglio AF, Myers RM (2014) Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat* 146(2):287–297. doi:[10.1007/s10549-014-3019-2](https://doi.org/10.1007/s10549-014-3019-2)
68. Nacu S, Yuan W, Kan Z, Bhatt D, Rivers CS, Stinson J, Peters BA, Modrusan Z, Jung K, Seshagiri S, Wu TD (2011) Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med Gen* 4:11. doi:[10.1186/1755-8794-4-11](https://doi.org/10.1186/1755-8794-4-11)
69. Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, Rubin MA (2009) SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer. *Cancer Res* 69(7):2734–2738
70. Stoeck A, Lejnine S, Truong A, Pan L, Wang H, Zang C, Yuan J, Ware C, MacLean J, Garrett-Engle PW, Kluk M, Laskey J, Haines BB, Moskaluk C, Zawel L, Fawell S, Gilliland G, Zhang T, Kremer BE, Knoechel B, Bernstein BE, Pear WS, Liu XS, Aster JC, Sathyanarayanan S (2014) Discovery of biomarkers predictive of GSI response in triple-negative breast cancer and adenoid cystic carcinoma. *Cancer Discover* 4(10):1154–1167. doi:[10.1158/2159-8290.CD-13-0830](https://doi.org/10.1158/2159-8290.CD-13-0830)
71. Li Z, Tognon CE, Godinho FJ, Yasaitis L, Hock H, Herschkowitz JI, Lannon CL, Cho E, Kim SJ, Bronson RT, Perou CM, Sorensen PH, Orkin SH (2007) ETV6-NTRK3 fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex. *Cancer Cell* 12(6):542–558. doi:[10.1016/j.ccr.2007.11.012](https://doi.org/10.1016/j.ccr.2007.11.012)
72. Rajadurai CV, Havrylov S, Zaoui K, Vaillancourt R, Stuble M, Naujokas M, Zuo D, Tremblay ML, Park M (2012) Met receptor tyrosine kinase signals through a cortactin-Gab1 scaffold complex, to mediate invadopodia. *J Cell Sci* 125(Pt 12):2940–2953. doi:[10.1242/jcs.100834](https://doi.org/10.1242/jcs.100834)
73. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Breast Cancer Susceptibility C, Hoening M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PD, Stratton MR, Dunning AM, Rahman N, Easton DF (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Gen* 42(6):504–507. doi:[10.1038/ng.586](https://doi.org/10.1038/ng.586)

74. Wang Y, He Y, Qin Z, Jiang Y, Jin G, Ma H, Dai J, Chen J, Hu Z, Guan X, Shen H (2014) Evaluation of functional genetic variants at 6q25.1 and risk of breast cancer in a Chinese population. *Breast Cancer Res* 16(4):422. doi:[10.1186/s13058-014-0422-x](https://doi.org/10.1186/s13058-014-0422-x)
75. Dunning AM, Michailidou K, Kuchenbaecker KB, Thompson D, French JD, Beesley J, Healey CS, Kar S, Pooley KA, Lopez-Knowles E, Dicks E, Barrowdale D, Sinnott-Armstrong NA, Sallari RC, Hillman KM, Kaufmann S, Sivakumaran H, Moradi Marjaneh M, Lee JS, Hills M, Jarosz M, Drury S, Canisius S, Bolla MK, Dennis J, Wang Q, Hopper JL, Southey MC, Broeks A, Schmidt MK, Lophatananon A, Muir K, Beckmann MW, Fasching PA, Dos-Santos-Silva I, Peto J, Sawyer EJ, Tomlinson I, Burwinkel B, Marne F, Guenel P, Truong T, Bojesen SE, Flyger H, Gonzalez-Neira A, Perez JI, Anton-Culver H, Eunjung L, Arndt V, Brenner H, Meindl A, Schmutzler RK, Brauch H, Hamann U, Aittomaki K, Blomqvist C, Ito H, Matsuo K, Bogdanova N, Dork T, Lindblom A, Margolin S, Kosma VM, Mannermaa A, Tseng CC, Wu AH, Lambrechts D, Wildiers H, Chang-Claude J, Rudolph A, Peterlongo P, Radice P, Olson JE, Giles GG, Milne RL, Haiman CA, Henderson BE, Goldberg MS, Teo SH, Yip CH, Nord S, Borresen-Dale AL, Kristensen V, Long J, Zheng W, Pytkas K, Winqvist R, Andrulis IL, Knight JA, Devilee P, Seynaeve C, Figueroa J, Sherman ME, Czene K, Darabi H, Hollestelle A, van den Ouweland AM, Humphreys K, Gao YT, Shu XO, Cox A, Cross SS, Blot W, Cai Q, Ghousaini M, Perkins BJ, Shah M, Choi JY, Kang D, Lee SC, Hartman M, Kabisch M, Torres D, Jakubowska A, Lubinski J, Brennan P, Sangrajrang S, Ambrosone CB, Toland AE, Shen CY, Wu PE, Orr N, Swerdlow A, McGuffog L, Healey S, Lee A, Kapuscinski M, John EM, Terry MB, Daly MB, Goldgar DE, Buys SS, Janavicius R, Tihomirova L, Tung N, Dorfling CM, van Rensburg EJ, Neuhausen SL, Ejlertsen B, Hansen TV, Osorio A, Benitez J, Rando R, Weitzel JN, Bonanni B, Peissel B, Manoukian S, Papi L, Ottini L, Konstantopoulou I, Apostolou P, Garber J, Rashid MU, Frost D, Embrace Izatt L, Ellis S, Godwin AK, Arnold N, Niederacher D, Rhiem K, Bogdanova-Markov N, Sagne C, Stoppa-Lyonnet D, Damiola F, Collaborators GS, Sinilnikova OM, Mazoyer S, Isaacs C, Claes KB, De Leeneer K, de la Hoya M, Caldes T, Nevanlinna H, Khan S, Mensenkamp AR, Hebon Hooning MJ, Rookus MA, Kwong A, Olah E, Diez O, Brunet J, Pujana MA, Gronwald J, Huzarski T, Barkardottir RB, Laframboise R, Soucy P, Montagna M, Agata S, Teixeira MR, kConFab I, Park SK, Lindor N, Couch FJ, Tischkowitz M, Foretova L, Vijai J, Offit K, Singer CF, Rappaport C, Phelan CM, Greene MH, Mai PL, Rennert G, Imyanitov EN, Hulick PJ, Phillips KA, Piedmonte M, Mulligan AM, Glendon G, Bojesen A, Thomassen M, Caligo MA, Yoon SY, Friedman E, Laitman Y, Borg A, von Wachenfeldt A, Ehrencrona H, Rantala J, Olopade OI, Ganz PA, Nussbaum RL, Gayther SA, Nathanson KL, Domchek SM, Arun BK, Mitchell G, Karlan BY, Lester J, Maskarinec G, Woolcott C, Scott C, Stone J, Apicella C, Tamimi R, Luben R, Khaw KT, Helland A, Haakensen V, Dowsett M, Pharoah PD, Simard J, Hall P, Garcia-Closas M, Vachon C, Chenevix-Trench G, Antoniou AC, Easton DF, Edwards SL (2016) Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nature genetics* 48(4):374–386. doi:[10.1038/ng.3521](https://doi.org/10.1038/ng.3521)
76. Ma CX, Reinert T, Chmielewska I, Ellis MJ (2015) Mechanisms of aromatase inhibitor resistance. *Nature Rev Cancer* 15(5):261–275. doi:[10.1038/nrc3920](https://doi.org/10.1038/nrc3920)
77. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT, Aster JC (2004) Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* 306(5694):269–271. doi:[10.1126/science.1102160](https://doi.org/10.1126/science.1102160)
78. Taipale M, Krykbaeva I, Whitesell L, Santagata S, Zhang J, Liu Q, Gray NS, Lindquist S (2013) Chaperones as thermodynamic sensors of drug-target interactions reveal kinase inhibitor specificities in living cells. *Nature Biotechnol* 31(7):630–637. doi:[10.1038/nbt.2620](https://doi.org/10.1038/nbt.2620)
79. Ludes-Meyers JH, Liu Y, Munoz-Medellin D, Hilsenbeck SG, Brown PH (2001) AP-1 blockade inhibits the growth of normal and malignant breast cells. *Oncogene* 20(22):2771–2780. doi:[10.1038/sj.onc.1204377](https://doi.org/10.1038/sj.onc.1204377)
80. Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40(10):e72. doi:[10.1093/nar/gks001](https://doi.org/10.1093/nar/gks001)
81. Natrajan R, Wilkerson PM, Marchio C, Piscuoglio S, Ng CK, Wai P, Lambros MB, Samartzis EP, Dedes KJ, Frankum J, Bajrami I, Kopec A, Mackay A, A'Hern R, Fenwick K, Kozarewa I, Hakas J, Mitsopoulos C, Hardisson D, Lord CJ, Kumar-Sinha C, Ashworth A, Weigelt B, Sapino A, Chinnaiyan AM, Maher CA, Reis-Filho JS (2014) Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. *J Pathol* 232(5):553–565. doi:[10.1002/path.4325](https://doi.org/10.1002/path.4325)
82. Kim J, Kim S, Ko S, In YH, Moon HG, Ahn SK, Kim MK, Lee M, Hwang JH, Ju YS, Kim JI, Noh DY, Kim S, Park JH, Rhee H, Kim S, Han W (2015) Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. *Genes Chromosomes Cancer* 54(11):681–691. doi:[10.1002/gcc.22279](https://doi.org/10.1002/gcc.22279)