CHAPTER 4

Discovery of MAST Kinase Fusions in Breast Cancer by Transcriptome Sequencing

INTRODUCTION

Recurrent gene fusions and translocations have long been associated with hematologic malignancies and rare soft tissue tumors as driving genetic lesions (Delattre et al., 1992; Nowell and Hungerford, 1960; Rowley, 1998). Over the last few years, it is becoming apparent that these genetic rearrangements are also found in common solid tumors including a large subset of prostate cancers (Kumar-Sinha et al., 2008; Tomlins et al., 2005) and smaller subsets of lung cancer, gastric cancer, among others (Prensner and Chinnaiyan, 2009). Importantly, a number of these gene fusions are potentially targetable including the *ALK* gene fusions in non-small cell lung cancer (Perner et al., 2008; Soda et al., 2007), *RET* in papillary thyroid cancer (Grieco et al., 1990), and RAF family fusions in prostate cancer and other solid tumors (Palanisamy et al.). Breast cancer is generally classified based on expression of estrogen receptor (*ER*), progesterone receptor (*PR*) and *ERBB2*. Interestingly, two rare forms of breast cancer, secretory breast carcinoma and adenoid cystic carcinoma, are characterized almost exclusively by recurrent *ETV6-NTRK3* and *MYB-NFIB* gene fusions, respectively (Persson et al., 2009; Tognon et al., 2002). In this study, we employed next generation sequencing of
CHAPTER 4

breast cancer transcriptomes to identify functionally recurrent classes of gene fusions with a particular emphasis on classes of molecules that may be therapeutically targetable.

MATERIALS AND METHODS

Cell lines and tissues

Breast cancer cell lines used in this study were obtained from the American Type Culture Collection (ATCC) or individual collections. Cells were grown in appropriate media supplemented with fetal bovine serum and antibiotics (Invitrogen), or supplements designated for the media (Lonza). This study was approved by the University of Michigan Internal Review Board. De-identified breast cancer samples were obtained from the University of Michigan Cancer Center Frozen Biobank and the Breakthrough Breast Cancer Research Centre, Institute of Cancer Research (London, UK). The complete list of cell lines and tissue samples used for this study is shown (Table S4.1).

Paired end transcriptome sequencing and nomination of gene fusions

Total RNA was extracted from normal and cancer breast cell lines and breast tumor tissues using Trizol reagent (Invitrogen) according to the manufacturer’s instructions. Quality assessment of RNA was performed using Agilent Bioanalyzer 2100 (Agilent Technologies). Transcriptome libraries from the mRNA fractions were generated following the RNA-SEQ protocol (Illumina). Each sample was sequenced in a single lane with the Illumina Genome Analyzer II (40-80 nucleotides read length) or with the Illumina HiSeq 2000 (100 nucleotides read length). Number of reads passing filter for
each sample is shown (Table S4.2). Paired-end transcriptome reads passing filter were mapped to the human reference genome (hg18) and UCSC genes, allowing up to two mismatches, with Illumina ELAND software (Efficient Alignment of Nucleotide Databases). Sequence alignments were subsequently processed to nominate gene fusions using the methodology described earlier (Maher et al., 2009a) and further fine-tuned by parallel in-house gene fusion detection algorithm along with manual curation process. Briefly, paired reads were processed to identify any that either ‘encompassed’ or ‘spanned’ candidate fusion junctions. Encompassing paired reads refer to those in which individual reads of a pair aligned to a distinct gene, with the pair encompassing the fusion junction. In contrast, spanning paired reads refer to those in which individual sequence read aligns to a gene and its paired partner spans the fusion junction. Both categories of reads undergo a series of filtering steps to remove false positive before being merged together to generate the final chimera nomination.

Targeted Capture and Sequencing

Following RNA integrity analysis using the Agilent BioAnalyzer 2100 protocol, 92 individual breast carcinomas were placed in two pools. The first pool consisted of 200 ng each of 51 RNAs with RIN values between 3 and 5 and the second pool consisted of 41 RNAs with RIN values between 5.1 and 7.5. The pooled RNAs were depleted of rRNAs using RiboMinus reagents and protocols (Invitrogen). The rRNA depleted pools were converted to paired-end libraries Illumina RNA-SEQ paired end libraries following the standard protocol with the omission of the poly A selection. Following size selection
of 250 to 350 bp fragments on agarose gels, the DNA was recovered using the QIAQuick method (QIAGEN) and amplified for 8 cycles using Illumina PE1.0 and PE 2.0 primers and amplification conditions. After purification by the Ampure XP method (Agencourt) the concentration was determined using a Naondrop spectrophotometer. Capture probes were generated for exons 2-10 of \textit{MAST1} and \textit{MAST2}. Primer pairs generating PCR products between 105 and 140 bp were designed and a sequence encoding the T7 RNA polymerase promoter was added to the 5’ end of the forward primer in each pair. The primers are shown in (Table S4.3). 10 cycles of PCR amplification using 10 ng of cDNA plasmids for each gene was performed using HotStar polymerase reagents (QIAGEN). Biotinylated RNA probes were synthesized by \textit{in vitro} transcription reactions using the T7 Maxi script protocol (Ambion). Reactions were performed using 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.3 mM UTP, and 0.2 mM biotin-16-UTP. After synthesis at 37º C for 1 hr, the reactions were digested with DNase I and RNA was purified using the RNAClean method (Agencourt). Each biotinylated RNA probe was adjusted to a concentration of 100 ng / l and pooled. Pooled probes were hybridized to 2 g of the previously generated paired-end libraries using conditions and reagents of the SureSelect system (Agilent) following hybridization for 48 hr, fragments were captured using Dynal M280 streptavidin magnetic beads, washed and eluted using SureSelect protocols. The captured library was reamplified for 14 cycles using Illumina primers and conditions, purified using Ampure XP reagents and submitted for sequencing.
Array CGH of Breast Cancer Lines

Breast cancer cell line DNAs (ATCC) were labeled and hybridized to Agilent 244K chips using the manufacturer’s protocol. Arrays were scanned with an Agilent Microarray Scanner and data were extracted and analyzed with CGH Analytics software.

Real Time PCR validation

Primers for validation of candidate gene fusions were designed using NCBI Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), with two primer pairs spanning the fusion junction amplifying 70-110 bp products for every chimera tested. Quantitative PCR (QPCR) was performed using SyBr Green Mastermix (Applied Biosystems) on an Applied Biosystems Step One Plus Real Time PCR System. All oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA) and are listed (Table S4.3). GAPDH was used as endogenous control. All assays were performed twice and results were plotted as average fold change relative to GAPDH.

Constructs used for overexpression studies

The ZNF700-MAST1 fusion ORF from BrCa00001 was cloned into pENTR-D-TOPO Entry vector (Invitrogen) following manufacturer’s instructions. Sequence confirmed entry clones in correct orientation were recombined into Gateway pcDNA-DEST40 mammalian expression vector (Invitrogen) by LR Clonase II enzyme reaction following manufacturer’s instructions. Plasmids with C-terminus V5 tags were generated and tested for protein expression by transfection in HEK293 cells. A full-length
expression construct of MAST2 with DDK tag was obtained from Origene (Maryland, USA).

**Immunoblot detection of fusion protein**

Cell pellets were sonicated in NP40 lysis buffer (50 mM Tris-HCl, 1% NP40, pH 7.4, Sigma), complete protease inhibitor mixture (Roche) and phosphatase inhibitor (EMD bioscience). Immunoblot analysis for MAST2 was carried out using MAST2 antibody from Novus Biologicals (Colorado). Human actin antibody (Sigma) was used as a loading control.

**Cell Transfections**

HEK293 cells were transfected with the above mentioned constructs using Fugene 6 reagent (Roche). MAST1 protein overexpression was validated by probing with V5 antibody (Sigma). MAST2 overexpression was validated using DDK antibody (Origene). HMEC-TERT cells were transfected using Fugene 6 and polyclonal populations of cells expressing MAST1, MAST2 or empty vector constructs were selected using geneticin. For siRNA knockdown experiments, Smart-pool siRNAs from Thermo were used (J-004633-06, J-004633-07, and J-004633-08). All siRNA transfections were carried out using oligofectamine reagent (Life Sciences) and three days post transfection the cells were plated for proliferation assays. At the indicated times cell numbers were measured using Coulter Counter. Lentiviral particles expressing the MAST2 shRNA (Cat. TRCN0000001733 Sigma) were transduced using polybrene, according to
manufacturer’s instructions. Polyclonal populations expressing the *MAST2* shRNA sequences were selected using 0.5-1 g/ml puromycin.

**Colony formation assay**

Equal number of MDA-MB-468 cells, transduced with scrambled or *MAST2* shRNA lentivirus particles were plated and selected using puromycin. After 7-8 days the plates were stained with crystal violet to visualize the number of colonies formed. For quantitation of differential staining, the plates were treated with 10% acetic acid and absorbance was read at 750nm.

**Confluence measurements and wound healing assay using Incucyte**

Polyclonal populations of HMEC-TERT overexpressing *MAST1*, *MAST2* or vector control were plated and relative confluence measurements were made at 30 minute intervals using the Incucyte system. Rate of increase in confluence is indicative of increase in cell proliferation. For the wound healing assay vector control or *MAST1* overexpressing cells were plated at high density and 6 hours later, uniform scratch wounds were made using Woundmaker (Incucyte). Relative migration potential of the cells was assessed by confluence measurements at regular time intervals as indicated, over the wound area.

**Chicken chorioallantoic membrane assay**

Chicken chorio-allantoic membrane (CAM) assay for tumor growth was carried out as described earlier (Zijlstra et al., 2002). Briefly, fertilized eggs were incubated in a humidified incubator at 38°C for 10 days, and then CAM was dropped by drilling two
holes: a small hole through the eggshell into the air sac and a second hole near the allantoic vein that penetrates the eggshell membrane but not the CAM. Subsequently, a cutoff wheel (Dremel, Racine, WI) was used to cut a 1 cm² window encompassing the second hole near the allantoic vein to expose the underlying CAM. When ready, CAM was gently abraded with a sterile cotton swab to provide access to the mesenchyme and 2x 10⁶ cells in 50 l volume were implanted on top. The windows were subsequently sealed and the eggs returned to the incubator. After 7 days extra-embryonic tumors were isolated and weighed. 5-10 eggs per group were used in each experiment.

**MDA-MB-468-MAST2 Knockdown Xenograft Model**

Four week-old female SCID C.B17 mice were procured from a breeding colony at University of Michigan, maintained by Dr Kenneth Pienta. MDA-MB-468 cells infected with lentivirus constructs of scrambled or MAST2 shRNA were selected for 3 days using puromycin. Mice were anesthetized using a cocktail of xylazine (80 mg/kg IP) and ketamine (10 mg/kg IP) for chemical restraint. Cells were implanted into right and left abdominal-inguinal mammary fat pads (4 × 10⁶ cells resuspended in 100μl of saline with 20% Matrigel (BD Biosciences). Ten mice were included in each group. Tumor growth was recorded weekly by using digital calipers and tumor volumes were calculated using the formula \((\pi/6) (L\times W^2)\), where \(L\) = length of tumor and \(W\) = width. All procedures involving mice were approved by the University Committee on Use and Care of Animals (UCUCA) at the University of Michigan and conform to their relevant regulatory standards.
RESULTS

Paired end transcriptome sequencing of a panel of 41 breast cancer cell lines, and 14 breast cancer tissues, along with 8 benign breast epithelial cell lines and 2 benign breast tissues (Table S4.1) analyzed through the fusion discovery pipeline described earlier (Maher et al., 2009a; Maher et al., 2009b), lead to the identification of 267 gene fusions, at an average of over four gene fusions per breast cancer sample (Table S4.4). A closer examination of the chromosomal coordinates of the fusion partner genes revealed that a majority of the gene fusions clustered in regions of chromosomal amplifications (Figure 4.1A). Chromosome 17 harbors the ERBB2 amplicon and an adjacent amplicon that includes genes such as BCAS3, RPS6KB1, and TMEM49 among others, accounted for a third of all the gene fusions in samples with CGH data (Table S4.4). Other recurrent loci harboring multiple gene fusions include the BCAS4 amplicon on chr20 and the chr8q amplicon. Interestingly, the breast cancer cell lines BT-474, that harbors both the chr17 as well as chr20 amplicons and MCF7, with prominent amplifications in chr17, chr20 and chr8q, showed the maximum number of gene fusions observed in a sample, with these two samples accounting for as many as 26 gene fusions associated with amplicons compared against only nine in unamplified loci (Figure 4.1A, B).

Remarkably, no single gene fusion from more than 200 identified here was found to be recurrent in our compendium, even as several fusion genes did appear in combination with different fusion partners. For example, three fusions each involving IKZF3 and BCAS3 as 3' partners were found in three different cell lines- all with different 5'
partners; likewise *TRIM37* was a common 5' partner in three distinct gene fusions with different 3' partners. Overall, 24 genes were found to be recurrent fusion partners, almost always associated with amplicons (*Table S4.4*). In order to focus on potentially tumorigenic ‘driver’ fusions, we prioritized the gene fusions based on the known cancer-associated functions of component genes such as if the 3’ partner was a kinase, oncogene, tumor suppressor or known to be fusion partners in the Mitelman Database of chromosomal aberrations in cancer (Mitelman et al., 2010). Following this strategy, many potentially interesting gene fusions were identified, including two different gene fusions involving ribosomal protein S6 kinase (*RPS6KB1*) found in BT-474 and MCF7. Incidentally, both of these cell lines harbor amplifications at the *RPS6KB1* locus (*Figure 4.1B*), express the highest levels of *RPS6KB1* among all the samples examined here (*Figure 4.1C*), but the chimera shows a loss of the kinase domain, thus it is unclear how these aberrations may play a role in cancer progression (*Figure 4.1C inset*). Previously, BT-474 and MCF7 have been shown to express high levels of full length RPS6KB1 protein (Yamnik et al., 2009).

In order to search for functionally relevant, potentially “drug-gable” rearrangements in breast cancer, in a way akin to *EML4-ALK* in lung cancer or *SLC45A3-BRAF* in prostate cancer, we focused on gene fusions with intact reading frames retaining critical functional domains and discovered several gene fusions involving intact kinase domains (*Table S4.5*), of which microtubule associated serine/threonine kinase family (MAST) gene fusions represented functionally recurrent events. MAST
kinase family genes (MAST1-4, and MAST-like) are characterized by the presence of a serine/threonine kinase domain and a PDZ domain, involved in protein scaffolding and interaction with other proteins (Garland et al., 2008). MAST1 and MAST2 are widely expressed in diverse tissues including brain, heart, liver, lung, kidney, and testis, while MAST3 and MAST4 show more restricted expression in several tissues and MAST-like is predominantly expressed in heart and testis (Garland et al., 2008).

Three independent cases of MAST gene fusions were identified by transcriptome sequence analyses—ZNF700-MAST1 in breast cancer tissue BrCa00001, NFIX-MAST1 in breast carcinoma BrCa10017, and ARID1A-MAST2 in a triple negative (ER-/PR-/ERBB2-) breast cancer cell line MDA-MB-468 (Figure 4.2). These gene fusions were among the top scoring fusions observed in their respective index samples, based on the number of unique paired end reads supporting the chimeric transcripts. These index samples ranked among the highest levels of expression of MAST1 (in BrCa00001 and BrCa10017) and MAST2 (in MDA-MB-468) in our compendium of more than 350 cancer samples encompassing more than 17 different tissue types (data not shown). The samples harboring MAST gene fusions are distinct from those with Notch family gene fusion described in our companion paper. Each of the fusions was confirmed by fusion-specific PCR in the respective samples (Figure 4.3A). As a working antibody was available for MAST2, we validated the expression of the fusion protein from the ARID1A-MAST2 gene fusion in the breast cancer cell line MDA-MB-468 (Figure 4.3B). Together, these results suggest that the MAST fusions identified here may
represent key genetic aberrations in the respective breast cancer samples from which they were identified.

FISH-based screening was not feasible for genes that are in close proximity (e.g., ZF700, NFlX, and MAST1 are less than 1Mb apart on Chr 19) or regions of highly repetitive genomic sequences (as is the case with MAST1). As high throughput next generation sequencing now enables the detection of genetic aberrations at a resolution far superior to cytogenetic and FISH based approaches, we devised a targeted sequencing approach to screen additional samples for MAST gene fusions. A transcriptome library of 92 pooled breast carcinoma RNAs was generated and captured in solution with biotinylated baits encompassing the 5' exons 2-10 of MAST1 and MAST2. The captured library was sequenced and analyzed as before. Two new MAST gene fusions were discovered using this strategy. TADA2A-MAST1 and GPBP1L1-MAST2 gene fusions, like the preceding three fusions, encoded contiguous open reading frames, retaining the serine/threonine kinase and PDZ domains of 3’ MAST genes (Figure 4.2B, 4.3C, D). As Figure 4.3 C and D display, the predicted open reading frames of the MAST fusions we identified in this study each retain intact PDZ and serine/threonine kinase domains. Thus overall, we have discovered five novel gene fusions encoding MAST1 and MAST2 in a cohort of a little over 100 breast cancer samples and more than 40 cell lines, suggesting that the novel serine/threonine kinase family gene fusions represent a subset of up to 5% of breast cancers. As these are kinase fusions, they may also be therapeutically tractable in the future.
Next, we focused on the functional aspects of MAST fusion proteins. The
\textit{ZNF700-MAST1} fusion transcript encodes a truncated MAST1 protein that retains the
kinase (as well as PDZ) domain. We cloned the fusion encoded open reading frame from
the index sample, breast cancer tissue BrCa00001, into an expression vector. A
commercially available full-length \textit{MAST2} expression construct was used to mimic the
function of \textit{ARID1A-MAST2} overexpression, as this fusion encodes nearly full length
\textit{MAST2} (along with a 379 amino acid segment from \textit{ARID1A}). To assess the potential
oncogenic functions of MAST genes, we ectopically over-expressed epitope tagged-
truncated \textit{MAST1} and full length \textit{MAST2} in the benign breast cell line, HMEC-
TERT. Expression of the respective constructs was confirmed using anti-V5 and anti-
DDK antibodies (Figure S4.1A, B). Next, polyclonal populations of HMEC-TERT
cells overexpressing \textit{MAST1} and \textit{MAST2} were generated (Figure S4.1C, D). Using the
Incucyte system to measure cell proliferation in real time, both the \textit{MAST1} and \textit{MAST2}
overexpressing cells showed a growth advantage over vector control cells in confluence
measurements (Figure 4.4A). \textit{MAST1} overexpressing HMEC-TERT cells also showed
increased migration potential in a wound healing assay (Figure 4.4B) (MAST2 was not
done). Furthermore, \textit{MAST1} and \textit{MAST2} overexpressing HMEC-TERT cells showed
a significantly increased growth in a chicken chorioallantoic membrane (CAM) assay, as
compared to control cells (Figure 4.4C). Overall, these findings suggest that fusion
encoded truncated \textit{MAST1} and full length \textit{MAST2} overexpression can impart growth
and proliferative advantage thereby promoting an oncogenic phenotype.
To study the role of the endogenous *ARID1A-MAST2* fusion in MDA-MB-468 cells, we used multiple independent *MAST2* siRNAs to achieve a marked knockdown of the *MAST2* fusion (Figure S4.2A). These siRNAs showed significant growth inhibitory effects in cell proliferation assays in MDA-MB-468 cells (Figure 4.4D, left panel). Knockdown of *MAST2* in fusion negative benign breast cells, HMEC-TERT and a breast cancer cell line BT-483 did not have an effect on cell proliferation (Figure 4.4D, right panel), although a significant reduction in the levels of the wild-type *MAST2* transcript was achieved (Figure S4.2B-D). To characterize the effects of the *ARID1A-MAST2* fusion in MDA-MB-468 cells further, we used shRNA targeting *MAST2*, which displayed efficient knockdown of *ARID1A-MAST2* fusion at both the transcript (Figure S4.2E) and protein level (Figure S4.2F). MDA-MB-468 cells treated with *MAST2* shRNA exhibited a dramatic reduction in growth as demonstrated in a colony formation assay (Figure 4E), as well as showed increased apoptosis with S-phase arrest (Figure S4.3A and B). Not surprisingly, *MAST2* shRNA treated MDA-MB-468 cells did not survive long-term culturing, therefore, we carried out in vivo experiments using MDA-MB-468 cells transiently transfected with *MAST2* shRNA. We observed a reduction in tumor burden in the chicken chorioallantoic membrane assay (Figure S4.3C). In the mouse xenograft model, MDA-MB-468 cells transiently transfected with *MAST2*-shRNA, but not the scrambled control, failed to establish palpable tumors over a time course of 4 weeks (Figure 4.4F). Taken together,
the knockdown studies show that the *ARID1A-MAST2* fusion is a critical driver fusion in MDA-MB-468 cells.

**DISCUSSION**

The discovery of *TMPRSS2-ERG* gene fusion in a majority of prostate cancers fuelled expectations that similar high prevalence recurrent gene fusions may pervade other epithelial tumors. The emerging landscape of gene fusions in diverse solid cancers discovered since then however, including in prostate cancers, points to a more complex landscape where small subsets of cancers harbor rare but oncogenic gene fusions that provide novel diagnostic and personalized therapeutic avenues. For example, a number of rare gene fusions in prostate cancer involve ETS member family genes such as *ERG* (*SLC45A3-ERG* (Esgueva et al., 2010; Han et al., 2008)), *ETV1* (*TMPRSS2-ETV1*, C15orf21–ETV1, HERV-K–ETV1, HNRPA2B1–ETV1, *SLC45A3–ETV1* (Tomlins et al., 2007)), *ETV4* (*TMPRSS2-ETV4* (Tomlins et al., 2006), *DDX5-ETV4* (Han et al., 2008)) and *ETV5* (*SLC45A3-ETV5* (Helgeson et al., 2008)), as well as gene fusions involving RAF kinase family genes described recently, including *SLC45A3-BRAF* and *ESRP1-RAF1* (Palanisamy et al., 2010). Similarly, a number of gene fusions recently discovered in lung cancer involve *ALK* (*EML4-ALK* and *TFG-ALK*), and *ROS* (*SLC34A4-ROS* and *CD74-ROS*) (Rikova et al., 2007) kinases. The gene fusions in breast cancer involving MAST kinases (described here) and the Notch family of transcription factors (described in the companion study) (Robinson et al., 2011) represent
novel classes of rare but functionally recurrent gene fusions with therapeutic implications (similar to the ALK fusions in lung cancer). The discovery of recurrent fusions of MAST family serine/threonine kinases in about 5% of breast cancers will likely spur the development of MAST-kinase specific inhibitors or lead to the evaluation of currently available serine/threonine kinase inhibitors such as Sorafenib.

Little is currently known about the role of MAST family genes in normal physiological processes or in cancer, albeit, MAST1 expression has been associated with resistance to the anti-cancer drug 5-fluorouracil (5-FU) (De Angelis et al., 2006). In a recent study of genetic variation in mitotic kinases associated with breast cancer risk, identified common haplotypes of MAST2 to be significantly associated with breast cancer risk (P = 0.04) (Wang et al., 2009). Functionally, MAST2 has been linked with the dystrophin/utrophin network of microtubule filaments via the syntrophins. MAST2 has also been shown to act as a scaffolding protein for TRAF6, regulating its activity, including inhibition of NF-κB, likely regulating cellular inflammatory responses (Xiong et al., 2004). Pertinent to breast cancer, the tumor suppressor phosphatase PTEN has been shown to interact with the PDZ domain of MAST2 and related serine/threonine kinases (Valiente et al., 2005), suggesting potentially regulatory networks impacted by MAST genes.

Identification of MAST and Notch family gene rearrangements in subsets of breast cancer further supports the argument that cancers may be operationally classified by driving molecular events, rather than by organ site or cell of origin, in the context of
rational targeted therapy. Multiple small subclasses of gene fusions revealed through next generation high throughput sequencing underscores the utility of this approach in delineating patient-specific molecular subtypes of cancers, which may facilitate development of personalized therapeutic modalities.

SUMMARY

- Employed paired-end transcriptome sequencing to explore the landscape of gene fusions in a panel of breast cancer cell lines and tissues.
- Identified two classes of recurrent gene rearrangements involving genes encoding microtubule-associated serine-threonine kinase (MAST) and members of the NOTCH family.
- *MAST1* and *MAST2* kinase fusions were identified in breast cancer cell lines and tissues, and in all the index cases, their serine threonine kinase domains were spared.
- Overexpression of *MAST1* or *MAST2* had a proliferative effect on breast epithelial cells *in vitro* and *in vivo*, and knockdown of the *MAST2* fusion in an index breast cancer cell line attenuated growth *in vitro* and *in vivo*.
- While these fusions represent a rare subtype of breast cancer, they potentially may be targetable due to their kinase function.
Figure 4.1. Recurrent loci of amplifications are hotspots of gene fusions in breast cancer. (A) Histograms of number of gene fusions in individual samples with respect to their association with loci of genomic amplifications. The pie chart shows overall distribution of gene fusions associated with amplicons (maroon) or non-amplified loci (blue). (B) Circos plot representation of chromosomal locations of gene fusions in breast cancer cell line BT-474 (left) and MCF7 (right). The outer ring represents genome-wide chromosomal coordinates and the inner ring shows corresponding genomic amplification (in red) and deletion (in green) profiles based on aCGH data. The radial lines represent location of gene fusions, with amplicon associated gene fusions indicated by purple lines and non-amplicon fusions shown in blue. (C) Gene expression levels, expressed in reads per kb per million total reads, RPKM units (y-axis) of ribosomal protein S6 kinase (RPS6KB1), arranged in descending order of expression from left to right, across individual samples in our compendium of breast cancer transcriptome (x-axis). The two highest expressing samples, BT-474 and MCF7 harbor amplifications at the RPS6KB1 locus (1B, above), as well as gene fusions (inset), RPS6KB1-SNF8 in BT-474 and RPS6KB1-TMEM49 in MCF7. The domain architecture of fusion genes shows that the kinase domains are lost in both gene fusions. Red arrows indicate the locations of chromosomal breaks.
Figure 4.2. Discovery of the MAST kinase gene fusions in breast cancer identified by paired-end transcriptome sequencing. (A) List of MAST family gene fusions identified in breast carcinomas with subtype status. (B) Diagram of MAST family gene fusions. \textit{ZNF700-MAST1} in BrCa00001, \textit{TADA2A-MAST1} in BrCa10038, \textit{NFIX-MAST1} in BrCa10017, \textit{ARID2A-MAST2} in the breast cancer cell line MDA-MB-468, and \textit{GPBP1L1-MAST2} in BrCa10039 are shown. Histograms of top ranking gene fusion nomination scores in each of the index samples are shown on the right of the chimera schematics, with MAST fusion gene names shown in red and additional gene fusions in each sample indicated in black. Schematic representation of unique passed purity filter paired-end reads supporting the gene fusions are shown spanning the fusion junctions with respective exon numbers (and nucleotide positions) comprising the chimeric transcripts shown below.
Figure 4.3. Experimental validations of MAST gene fusions in the index breast cancer samples. (A) Expression of \(ZNF700-MAST1\) gene fusion in breast cancer tissue BrCa00001, \(NFIX-MAST1\) in BrCa10017, \(TADA2A-MAST1\) fusion in BrCa10038, and \(ARID1A-MAST2\) fusion in MDA-MB-468 validated by RT-PCR normalized against glyceraldehyde 6-phosphate dehydrogenase (\(GAPDH\)) values in each sample. (B) Western blot showing a higher molecular weight band above MAST2, corresponding to the fusion protein ARID1A-MAST2, specifically observed in the index breast cancer cell line MDA-MB-468. Western blot with the housekeeping protein actin was used as the loading control. (C) Schematic representation of functional domains retained in the putative chimeric proteins involving MAST1 and (D) involving MAST2.
**Figure 4.4. Functional characterization of MAST fusion genes.** (A) Percentage confluency over a time course was measured using the Incucyte system for polyclonal populations of HMEC-TERT cells overexpressing full length MAST2, allelic MAST1 (truncated ORF from ZNF700-MAST1 transcript in BrCa00001) and empty vector control. (B) Wound healing assay using the Incucyte system. Polyclonal populations of HMEC-TERT cells overexpressing MAST1 or empty vector were assessed for their ability to close wound areas over a period of time. (C) Histogram showing growth of HMEC-TERT cells stably overexpressing MAST1, MAST2 or vector control on chicken chorionic allantoic membrane (CAM) assay. (D) Graphical representation of cell proliferation assay showing cell numbers (y-axis) over the indicated time course (x-axis) with MAST2 knockdown using three independent siRNAs and one shRNA construct in MDA-MB-468 cells harboring the ARID1A-MAST2 fusion (left) and in fusion negative HMEC-TERT and BT-483 cells, as indicated (right). (E) Histogram representation of colony formation assay with MDA-MB-468 cells treated with MAST2 specific shRNA or control-scrambled sequence-shRNA. The inset shows crystal violet staining of cells treated with either scrambled or MAST2 shRNA. (F) Tumor growth in immunodeficient mice implanted with MDA-MB-468 cells transfected with MAST2-shRNA or scrambled control shRNA.
Figure S4.1. Immunoblot analysis of HEK293 cells overexpressing (A) fusion allelic MAST1 using anti-V5 antibody and (B) full length MAST2 using anti-DDK antibody. qPCR validation of MAST1 (C) and MAST2 (D) overexpression in polyclonal populations of HMEC-TERT cells overexpressing the respective constructs.
Figure S4.2. MAST2 Knockdown Assays. (A) qPCR validation of MAST2 and ARID1A-MAST2 knockdown using MAST2 siRNAs in MDA-MB-468 cells. qPCR validation of MAST2 knockdown (B) in fusion negative BT-483 cells (C) in H16N2 cells (D) in HMEC-TERT cells. Validation of MAST2 knockdown in MDA-MB-468 cells by (E) qPCR and (F) anti-MAST2 immunoblot.
Figure S4.3. Functional characterization of MAST2 fusion (A) Flow cytometric analysis of MDA-MB-468 cells treated with scrambled shRNA or MAST2 shRNA. (B) Percentage distribution of the MDA-MB-468 cells in different phases of the cell cycle after treatment with either the scrambled shRNA or MAST2 shRNA. (C) Chicken chorioallantoic membrane assay showing tumor weight of MDA-MB-468 cells treated with either scrambled shRNA or MAST2 shRNA.
SUPPLEMENTARY FILES (Enclosed in CD)

Table S4.1. Breast cell lines and tissues analyzed. List of breast samples sequenced along with their ER, PR, ERBB2 status, source and culture media

Table S4.2. Sequencing statistics for samples analyzed from breast cohort. List of breast samples analyzed along with sequencing statistics

Table S4.3. Primer sequences - MAST1 and MAST2 fusion Study. List of primer sequences used to validate MAST1 and MAST2 fusion and the list of primers used for the target capture.

Table S4.4. Gene fusions identified in breast carcinomas. List of gene fusions identified in breast cancer cell lines, and tissues along with their copy number status.

Table S4.5. Fusions involving genes encoding protein kinases in breast carcinomas. List of gene fusions candidates involving kinase genes, which are identified in breast cancer cohort.