TGFB1*6A as a modifier of breast cancer risk and progression: advances and future prospects

Kojo Agyemang, Allan M. Johansen, Grayson W. Barker, Michael J. Pennison, Kimberly Sheffield, Hugo Jimenez, Carl Blackman, Sambad Sharma, Patrick A. Fordjour, Ravi Singh, Katherine L. Cook, Hui-Kuan Lin, Wei Zhang, Hui-Wen Lo, Kounosuke Watabe, Peiqing Sun, Carl D. Langefeld, and Boris Pasche

There is growing evidence that germline mutations in certain genes influence cancer susceptibility, tumor evolution, as well as clinical outcomes. Identification of a disease-causing genetic variant enables testing and diagnosis of at-risk individuals. For breast cancer, several genes such as BRCA1, BRCA2, PALB2, ATM, and CHEK2 act as high- to moderate-penetrance cancer susceptibility genes. Genotyping of these genes informs genetic risk assessment and counseling, as well as treatment and management decisions in the case of high-penetrance genes. TGFB1*6A (rs11466445) is a common variant of the TGF-β receptor type I (TGFB1) that has a global minor allelic frequency (MAF) of 0.051 according to the 1000 Genomes Project Consortium. It is emerging as a high frequency, low penetrance tumor susceptibility allele associated with increased cancer risk among several cancer types. The TGFB1*6A allele has been associated with increased breast cancer risk in women, OR 1.15 (95% CI 1.01–1.31). Functionally, TGFB1*6A promotes breast cancer cell proliferation, migration, and invasion through the regulation of the ERK pathway and Rho-GTP activation. This review discusses recent findings on the genetic, functional, and mechanistic associations between TGFB1*6A and breast cancer risk and proposes future directions as it relates to genetic association studies and mechanisms of action for tumor growth, metastasis, and immune suppression.

Published in partnership with the Breast Cancer Research Foundation

HERITABLE BREAST CANCER GENES

Heritable predisposition genes are important risk factors for breast cancer susceptibility, accounting for 5.03% of all breast cancer cases. Pathogenic variants of high-risk predisposition genes such as BRCA1 and BRCA2 are the most widely known and are used in genetic testing and counseling to predict breast cancer risk and clinical outcomes. However, these variants are uncommon (<1.3% allelic frequency), and account for less than 1.5% of all breast cancer cases. Like other complex genetic traits such as diabetes, obesity, and autoimmune diseases, recent advances in genome-wide association studies reveal that the vast majority of hereditary breast cancer cases are genetically multifactorial, involving numerous other polymorphisms of varying penetrance acting as tumor modifier genes. Thus, several lower penetrance risk variants are now included in breast cancer susceptibility gene screening panels for testing and counseling. Of these polymorphisms into screening panels results in a 40 to 50% increase in breast cancer risk detection among women, and a 5 to 15% increase in detection among BRCA1/2-negative females.

The Transforming Growth Factor-Beta (TGF-β) signaling pathway plays a critical role during cancer development and progression. Variants of the TGF-β pathway genes, particularly TGFB1*6A, have been studied in a large number of females, and have been associated with low penetrance risk for breast cancer (Table 1). This review details our current clinical and preclinical knowledge on TGFB1*6A as a high frequency, low penetrance tumor susceptibility allele, and provides further rationale to assess its role as a modifier gene for breast cancer predisposition and tumor progression.

TGFB1*6A AND BREAST CANCER

Identification and characterization

The TGF-β pathway functions as a tumor suppressor during cancer development but enhances tumor growth, immune evasion, and metastasis in advanced cancers. During TGF-β signaling, the TGF-β receptor 1 (TGFB1) plays a critical role of binding and activating downstream receptor-regulated (R-) SMADs (SMAD2 and SMAD3), co-SMADs (SMAD4), and non-SMAD (MAPK-ERK, RAS, AKT, JNK, and RHOA) intermediary proteins. In 1998, Pasche et al. identified TGFB1*6A (rs11466445) as a polymorphic variant of TGFB1 with an in-frame deletion of three GCG codons encoding alanine within exon 1 of the human TGFB1 signal peptide sequence (Fig. 1). Signal peptides are responsible for intracellular transport, targeting of nascent proteins (secretory and membrane proteins) to the endoplasmic reticulum, and integration of newly translated proteins into their respective compartments. N-terminal signal peptides are usually cleaved off and degraded after insertion and targeting of the nascent protein to the endoplasmic reticulum.

Investigations into TGFB1*6A signal peptide cleavage showed that the 3-alanine deletion from the 9-alanine repeat within the hydrophobic core of the signal peptide does not affect the posttranslational cleavage of a signal peptide. Studies showed that TGFB1*6A signal peptide is cleaved between Ala30 and Leu31, whereas the wild-type TGFB1 is cleaved...
between Ala33 and Leu34. Importantly, the mature forms of the TGFBR1*6A and TGFBR1 receptors are identical and the TGF-β ligand binds to each receptor with the same affinity.23,25. Also, the differences in cleavage sites do not influence TGFBR1*6A protein targeting and translocation functions25 nor its sensitivity, and half-life23. Despite TGFBR1*6A and TGFBR1 physicochemical similarities, TGFBR1*6A is intriguingly associated with breast cancer risk25,27 and promotion of cell growth, migration, and invasion25,28. Only the released signal peptide separates TGFBR1*6A from TGFBR1, which strongly suggests that TGFBR1*6A signal peptide contributes to oncogenesis.

### Table 1. Case-control studies showing TGFBR1*6A genotypic and allelic distribution and frequency.

| Study                  | Country of study participants | Self-reported race/ethnicity | Genotype distribution and frequency (%) | Allelic frequency |
|------------------------|------------------------------|-----------------------------|----------------------------------------|-------------------|
|                        |                              |                             | Cases                                  | Controls          |
|                        |                              |                             | 9/9A                                   | 9A/6A            |
|                        |                              |                             | 6A/6A                                  | 9A/9A            |
|                        |                              |                             | 9A/6A                                  | 6A/6A            |
| Pasche et al. 1999 (a) | US                            | Mixed                       | 128 (84.2)                             | 24 (15.8)        |
| Pasche et al. 1999 (b) | Northern Italy                | Caucasian                   | 39 (81.3)                              | 8 (16.7)         |
| Baxter et al. 2002     | United Kingdom                | Caucasian                   | 268 (75.5)                             | 83 (23.4)        |
| Reiss, 2004            | US                            | Mixed                       | 87 (88.8)                              | 11 (12.2)        |
| Caldes, 2004           | Spain                         | Caucasian                   | 214 (79.0)                             | 56 (20.7)        |
| Offit, 2004            | US                            | NS                          | 391 (84.6)                             | 67 (14.5)        |
| Northwestern, 2004     | US                            | NS                          | 74 (86.1)                              | 12 (13.9)        |
| Jin et al. 2004 (a)    | Finland                       | Caucasian                   | 117 (80.1)                             | 38 (27.2)        |
| Jin et al. 2004 (b)    | Poland                        | Caucasian                   | 140 (82.4)                             | 28 (16.5)        |
| Kaklamani et al. 2005  | US                            | Mixed                       | 515 (84.3)                             | 92 (15.1)        |
| Chen et al. 2006       | US                            | Mixed                       | 92 (80.0)                              | 23 (20.2)        |
| Feigelson et al. 2006  | US                            | Mixed                       | 387 (80.5)                             | 94 (19.5)        |
| Cox et al. 2007        | US                            | NS                          | 968 (81.6)                             | 207 (17.4)       |
| Song et al. 2007       | Sweden                        | Caucasian                   | 598 (78.4)                             | 152 (19.9)       |
| Jakubowska et al. 2009 | Poland                        | Caucasian                   | 282 (88.7)                             | 33 (10.4)        |
| Colleran et al. 2009   | Ireland                       | Caucasian                   | 796 (82.9)                             | 154 (16.5)       |
| Joshi 2011 (a)         | India                         | Asian                       | 163 (97.6)                             | 4 (2.4)          |
| Joshi 2011 (b)         | India                         | Asian                       | 33 (78.6)                              | 8 (19)           |
| Kamali et al. 2015     | Iran                          | Middle Eastern             | 251 (89.6)                             | 25 (8.9)         |

Case-control studies from 1999 to date have included a total of 14,837 participants (6787 cases /8050 controls); of which 13,312 (6026 cases/7286 controls) were in Hardy–Weinberg (HW) equilibrium. The data shows genotype distribution and allelic frequency (%) in the order TGFBR1*9A/9A > TGFBR1*9A/6A > TGFBR1*6A/6A.

NS not stated.

*Study population not in Hardy–Weinberg equilibrium.

### GENETIC ASSOCIATION WITH BREAST CANCER RISK

#### Case-control studies

**Allelic frequency.** TGFBR1*6A (rs11466445) is a common variant of the TGF-β receptor type I (TGFBR1) with a study-wide minor allelic frequency (MAF) of 0.051 in the 1000 Genomes Project Consortium, ranging from 0.0079 among East Asians to 0.0975 among Europeans.29-31 Nineteen case-control studies that included 14,837 participants (6787 cases /8050 controls) have investigated the association of TGFBR1*6A with breast cancer risk (Table 1). Seventeen of these studies, which included 13,312 individuals (6026 cases/7286 controls), show the TGFBR1*6A genotype frequencies were in Hardy–Weinberg equilibrium (HWE). HWE is a useful genotyping quality control metric that relates allele frequencies to genotype frequencies with an expected assumption that genotype frequencies will remain constant in a randomly mating population. The aforementioned TGFBR1*6A case/control studies consisted predominantly of Caucasian individuals of European ancestry (>80%), but also included modest numbers of individuals from other self-reported races/ethnicities (e.g., African American, Hispanic, Asian, and Middle Eastern). Early studies examined a mixed sample of Caucasians, African Americans, Hispanics, and Asians from the US, and Northern Italy, and found an allelic frequency of 7.9% in breast cancer patients compared to 5.3% in healthy controls.32. Several other studies from the US consisting of >80% Caucasians and <20% non-Caucasians reported similar TGFBR1*6A allelic frequency of 5.6–10.0% among breast cancer patients.33-35 The TGFBR1*6A allelic frequency also ranged from 6.1 to 12.8% among breast cancer patients from northern and southwestern Europe.
including samples from the UK, Spain, Sweden, Finland, Poland, and Ireland. In all, Asian and Middle Eastern women, particularly those from Western India had the lowest TGFBR1*6A allelic frequency, 1.2–11.9% among cases. In all, TGFBR1*6A genotype frequency is in the order 9A/9A > 9A/6A > 6A/6A in all populations studied to date (Table 1)."}

**Risk association.** Published case-control studies examining the association of TGFBR1*6A with breast cancer risk among individuals from different geographical locations and ethnicities show both significant and non-significant risk associations. It remains unclear if the differences in TGFBR1*6A risk association correlate with geographical location, ethnicity, age, tumor stage, and other confounding factors such as other polymorphism frequencies, lifestyle, and environment. In 2002, Baxter et al. reported the first association between TGFBR1*6A and breast cancer risk (OR 1.6, 95% CI 1.1–2.5). Cases and controls were residents of Southampton, UK. Controls were healthy females. Breast cancer cases were selected based on age at onset under 40 years, family history of breast cancer irrespective of age at onset, or bilateral breast cancer irrespective of family history or age at onset. The study noted that TGFBR1*6A allelic frequency among Caucasians from the UK does not differ by age of onset (<40 years), bilateral breast cancer, family history, and germline mutations of BRCA1 and BRCA2. In 2004, Caldes investigated breast cancer patients of Caucasian descent only from Madrid, Spain. The study found a significant association between TGFBR1*6A and breast cancer risk (OR 1.55, 95% CI 1.02–2.34). In the Kaklamani 2005 study that included >80% Caucasians from New York, NY, TGFBR1*6A association with breast cancer risk was significant under both dominant (OR 1.50, 95% CI 1.07–2.11) and additive (OR 1.46, 95% CI 1.04–2.06) models. Breast cancer risk for women aged >50 years was higher (OR 2.20, 95% CI 1.25–3.87), than for women aged ≤50 years (OR 1.18, 95% CI 0.75–1.84). However, further subgroup analyses show no association between TGFBR1*6A and ER/PR status or cancer stage at diagnosis. In this study, both breast cancer patients and healthy controls were matched for age, gender, and location. Other studies that included participants of similarly mixed ethnicities but from other states in the United States did not confirm an association between TGFBR1*6A and breast cancer risk (Fig. 2a). This indicates that TGFBR1*6A association with breast cancer risk may have a more modest effect size than in the original reports (regression toward the mean). The above studies were not powered enough to account for the effects of other confounding-modifying factors apart from ethnicity and geographical location that modify the magnitude of the risk attributable to the TGFBR1*6A polymorphism.

Intriguingly, other studies have investigated participants of similar Caucasian backgrounds from other European countries (Sweden, Ireland, Finland, and Poland) and found no significant associations (Fig. 2a). In a case-control study among Swedish Caucasians, the study further enriched its population sampling for genetic susceptibility by recruiting patients with family history of bilateral breast cancer cases. Overall, TGFBR1*6A carrier status was not associated with breast cancer risk (OR 1.12, 95% CI 0.90–1.39). However, subgroup analyses showed that TGFBR1*6A was associated with increased risk among low-risk familial breast cancer patients with one first- or second-degree breast cancer relative (OR 1.3, 95% CI 1.0–1.9). In that study, TGFBR1*6A was also correlated with higher tumor grade (OR 2.27 (95% CI 1.01–5.11) but had no association with tumor stage or ER/PR status. For women of Asian and Middle Eastern descent, TGFBR1*6A association with breast cancer risk was neither significant nor protective. Among Asians from India, Joshi 2011 found no association between Western Indians (OR 0.69, 95% CI 0.18–1.92) and those from the Parsi community (OR 1.85, 95% CI 0.84–4.05). Although these are important studies, the sample sizes are modest, and the lower allelic frequency in this population results in very low a priori statistical power to detect an association of the size previously reported. To date, TGFBR1*6A appears to be associated with decreased breast cancer risk only in Middle Eastern women, particularly in Iran (OR 0.62, 95% CI 0.39–0.98). However, it is worth noting that the genotype frequencies in this sample of 280 cases and 280 controls show strong deviations from Hardy–Weinberg equilibrium expectations ($X^2$ p value = 2.81E-12), an important measure of genotyping quality. Specifically, there were an excess of both TGFBR1*6A and TGFBR1 homozygotes, with a corresponding deficit of heterozygotes, when compared to HWE expectations.

In summary, it appears that TGFBR1*6A is a common breast cancer susceptibility allele in diverse ethnic backgrounds and its influence may vary by ancestry, geographical location, and other risk modifying factors. Large case/control studies, with important related phenotypes, paralleling the magnitude of other complex genetic traits, and meta-analyses that incorporate strict genotype quality control assessment will be necessary to clarify its population-specific subgroup risk associations.

**META-ANALYSES OF TGFBR1*6A ASSOCIATION WITH BREAST CANCER RISK**

To derive a more precise estimate of TGFBR1*6A association with breast cancer risk, several meta-analyses have investigated its risk association in up to 14,837 participants (6787 cases/8050 controls) from 19 case-control studies. In all, eight meta-analyses have studied the association of TGFBR1*6A with breast cancer risk (Fig. 2b). Five of the studies found evidence that supports the association between TGFBR1*6A and breast cancer risk. Early meta-analyses by ref. and ref. that analyzed up to 12 case-control studies (4871 subjects) found an association of TGFBR1*6A with breast cancer risk with odds ratios of 1.48 (95% CI 1.11–1.96) and 1.38 (95% CI 1.14–1.67), respectively. Subsequent studies examined ten case/control studies and did not find a significant risk association (OR 1.10, 95% CI 0.89–1.38), albeit in a concordant direction. The heterogeneity in the reported results may be partly due to meaningful differences in the genotyping methods employed across these studies for this GCG repeat polymorphism, and increased variation in estimates due to lower sample sizes included in the case/control studies.

Three of the most recent meta-analyses that investigated up to 17 case/control studies (14,068 subjects) found an association between TGFBR1*6A and breast cancer risk. Liao et al. 2010 reported an odds ratio of 1.16 (95% CI 1.01–1.34) among 11,220 case controls. Wang et al. 2012 investigated 14,068 participants and showed a nearly identical OR of 1.15 (95% CI 1.01–1.31) and 1.38 (95% CI 1.14–1.67), respectively. Additionally, Ou et al. 2015 found 14.8 and 9.6% TGFBR1*6A allele distribution among 6275 cases and controls, respectively. There was a significant association of the TGFBR1*6A allele with breast cancer risk (OR 1.33 95% CI 1.02–1.73). The study also noted that the homozygous TGFBR1*6A/6A is not significantly associated with breast cancer risk. Colleran 2010 and Krishna 2020 found no significant associations between TGFBR1*6A and breast cancer risk (Fig. 2b). It is interesting to note that the latter two meta-analyses selectively excluded some data from the Pasche 2004, and other case/control studies in their analyses. For example, case/control studies by refs. 32,39 all of which are in Hardy–Weinberg equilibrium were selectively omitted from the ref. meta-analyses. Colleran et al. 2010 explained that some case/control studies in their analyses. For example, case/control studies by refs. 32,39 all of which are in Hardy–Weinberg equilibrium were selectively omitted from the ref. meta-analyses.
**Fig. 2**  **Studies investigating TGFBR1*6A association with breast cancer risk.** Forest plot showing the number of subjects and odds ratios of **a** Case-control studies, and **b** Meta-analyses associating TGFBR1*6A to breast cancer risk. Plot **a** includes only case/control studies that are in Hardy–Weinberg equilibrium. CI confidence interval, † dominant association ($p \leq 0.01$), ‡ additive association ($p \leq 0.05$), † allelic association ($p \leq 0.05$).
heterogeneity among different case/controls as another contributing factor. Overall, most meta-analyses published to date show an association between TGFBR1*6A and breast cancer risk. Further studies are warranted to establish the risk association and its relatedness to different population sub-groups.

**TGFBR1*6A FUNCTIONAL EFFECTS**

**Promotion of cell proliferation, migration, and invasion**

Physicochemical studies using R18/L17 and HEK 293 cells revealed that the mature TGFBR1*6A and the wild-type TGFBR1 receptors have similar half-life, receptor turnover, and binding affinity to TGF-β ligand. In MCF-7 breast cancer cells stably transfected with either TGFBR1 or TGFBR1*6A, TGF-β/SMAD signaling was comparable. However, the stably transfected TGFBR1*6A MCF-7 breast cancer cells exhibited enhanced cell growth, migration, and invasion. Importantly, TGFBR1*6A switched TGF-β anti-proliferative effects into growth stimulatory effects in the MCF-7 breast cancer cells. The TGFBR1*6A-mediated switch to growth stimulation was independent of TGFBR1*6A kinase domain, indicating a mechanism likely due to TGFBR1*6A signal peptide. A similar conclusion was also reached following the investigation of TGFBR1*6A induction of migration and invasion. There was a 1.2-fold and 1.7-fold increase in migration and invasion, respectively, in TGFBR1*6A MCF-7 cells when compared to TGFBR1 transfected cells. In the MCF-7 cells expressing low TGFBR1*6A levels, there were 1.3 and 1.9 times increase in migration and invasion respectively, when compared to the vector controls. Cells expressing higher and intermediate TGFBR1*6A levels showed 1.8 and 2.2 times increase in migration and invasion, respectively. The induction of migration and invasion observed in the TGFBR1*6A cells were not affected by TGF-β stimulation, suggesting an underlying migration mechanism that is TGF-β independent. Similar observations of increased growth and invasion were reported in colorectal cancer cells that endogenously harbor TGFBR1*6A. Using SW48 (TGFBR1/TGFBR1) and DLD1 (TGFBR1/ TGFBR1*6A) cells, TGF-β treatment resulted in growth inhibition in the TGFBR1/TGFBR1 SW48 cells while it resulted in growth stimulation in the TGFBR1/TGFBR1*6A DLD1 cells. Stable transfection of the colorectal cancer cells with TGFBR1*6A also exhibited enhanced proliferation when compared with the vector-transfected cells, indicating that TGFBR1*6A has similar functional outcomes in breast and colorectal cancer.

**MECHANISM OF ACTION**

The functional and mechanistic understanding of tumor susceptibility alleles are important in establishing their clinical significance. TGFBR1*6A function in breast cancer cells is postulated to be mediated by its cleaved signal peptide, which is 3-alanine shorter than the wild-type TGFBR1 signal peptide (Figs. 1, 3a). Thus far, available evidence using SBE4-lux and 3TP-lux luciferase reporters suggest a mechanism that is not mediated by the mature TGFBR1*6A receptor. The SBE4-lux luciferase reporter assesses the binding of activated SMAD2/3/4 with SMAD binding elements in the nucleus as a measure of TGF-β/SMAD activation. Likewise, 3TP-lux transcription reporter measures the binding of activated SMAD2/3/4 to three consecutive 12-O-tetradecanoylphorbol-13-acetate (TPA) response elements (TREs) and a portion of the plasminogen inhibitor-1 (PAI-1) promoter region. In a study that assessed the proliferation of MCF-7 cells expressing intermediate and high levels of TGFBR1*6A with kinase-inactivated domains, TGFBR1*6A enhanced cell proliferation in both clones similar to the kinase-activated TGFBR1*6A MCF-7 clones. Inactivation of the kinase domain did not influence TGFBR1*6A growth stimulation, indicating a cell proliferation enhancing mechanism that is not affected by the TGFBR1*6A kinase domain. Analyses of the TGF-β/SMAD signal transduction system using SBE4-lux and 3TP-lux transcription reporters, and phosphorylated-SMAD2 and SMAD3 levels showed comparable TGF-β/SMAD signaling levels for TGFBR1*6A and TGFBR1 cells. The absence of TGFBR1*6A influence on TGF-β signaling is consistent with the observed TGF-β independent effect on migration and invasion in MCF-7 cells. Investigations into TGF-β/ non-SMAD pathways revealed that TGFBR1*6A increases ERK1/2 phosphorylation but shows no significant influence on p38 and JNK activation. The upregulation of ERK1/2 phosphorylation is suggestive of MAP-kinase-mediated migration and invasion (Fig. 3b). A similar increase in MAP-kinase activation has also been correlated with increased invasion of TGFBR1*6A-transfected colorectal cancer cells. Zhou et al. showed that TGFBR1*6A expression enhances proliferation and invasion in stably transfected SW48 and DLD1 cells accompanied by increased ERK1/2 phosphorylation. In contrast to breast cancer cells, the study noted an increase in p38 activation in the colorectal cancer cells. In all, the identified functional and mechanistic responses in colorectal cancer cells appears to be TGF-β-dependent, indicating cancer-type differences in response mechanisms.

Further investigations into the signaling network of TGFBR1*6A-induced MCF-7 cell migration and invasion using Affymetrix GeneChip Human Genome U133 Plus 2.0 Array identified ARHGAP5 and FN1 as the key differentially expressed genes. The levels of ARHGAP5 and FN1 were downregulated in TGFBR1*6A MCF-7 cells when assessed with RT-qPCR and western blot assays. ARHGAP5 encodes the Rho GTPase activating protein 5 that negatively regulates the Rho-family of small GTPases. The Rho-family of small GTPases are small guanosine triphosphatases (GTPases) of the rat sarcoma (Ras) superfamily that function as molecular switches for cytoskeletal remodeling during cell division, cell-cell adhesion, cell contractility, migration, and invasion. ARHGAP5 (P190B) and ARHGAP35 (P190A) are the main regulators of the Ras homolog (Rho) family of actin-based regulators and are implicated in cellular adhesion, migration, and invasion. ARHGAP5 silencing inhibits migration and invasion of AGS and MGC-803 gastric adenocarcinoma cells. In colorectal cancer, however, ARHGAP5 is markedly overexpressed in the liver of metastatic tissues compared to matched primary tumor tissues. ARHGAP5 suppression leads to decreased wound healing, migration, and invasion in DLD1 and SW480 colorectal cancer cells. Similarly, amplification of ARHGAP5 on the 14q12 chromosomal locus promotes the spreading and migration of Huh-7 hepatocellular carcinoma cells. FN1 encodes fibronectin, a soluble glycoprotein that binds to the cell surface and extracellular matrix. It is involved in cell-cell adhesion, cell motility, and maintenance of cell shape. Positive stromal fibronectin expression is significantly associated with low metastatic spread among patients with invasive breast carcinoma. On the other hand, relatively low expression of stromal fibronectin correlates with lymph node metastasis, TNM stage, recurrence, and mortality. Studies have also shown that tamoxifen-induced TGF-β expression regulates fibronectin levels in a feedback loop. TGF-β treatment reduced the levels of tamoxifen-induced FN1 in MCF-7 cells. Overall, the unbiased gene expression analysis revealed that TGFBR1*6A alters the expression of known pro- and anti-metastatic effectors, which may contribute to its effect on breast cancer progression.

**PERSPECTIVES AND FUTURE DIRECTIONS**

**Genetic association with breast cancer risk and progression**

The case/control studies and meta-analyses assessing TGFBR1*6A association with breast cancer risk suggest an association between TGFBR1*6A and risk for breast cancer, albeit with demonstrable heterogeneity in odds ratios. They also suggest that
TGFBR1*6A-associated breast cancer risk may depend on ancestry, geographical location, and other confounding factors. In genetic association studies, there are primarily four possible explanations that account for a statistical association between a genetic polymorphism (e.g., TGFBR1*6A) and risk for a disease (e.g., breast cancer). They include (i) the allele directly affects the actual disease phenotype/expression, (ii) the allele is in linkage disequilibrium (correlated) with the true disease allelic locus, (iii) there is a spurious association due to population stratification or other confounders, and (iv) type II error. To assert a direct allelic effect on disease phenotype/expression, the study design must control for confounding factors such as population stratification that could lead to false-positive/spurious associations. In general, ancestral population substructure is the main confounding concern contributing to spurious genetic associations. Other sources include age, other mutations, and disease-specific contributory factors. By design, population-based case/control studies such as those used in most of the TGFBR1*6A case/control studies are susceptible to spurious risk associations due to their random or convenient selection of cases and controls from participants that may not belong to the same ancestral population. On the other hand, family-based case/control studies can help control for confounding population stratification because of shared family pedigrees; albeit these designs can be more difficult to recruit. Thus, in a bid to conclusively establish the risk association between TGFBR1*6A and breast cancer, future case/control studies should consider multicenter family-based population studies.

For multicenter population-based case/control studies that draw subjects from different self-reported races/ethnicities, careful consideration should be given to methods such as (i) structured association tests, (ii) principal components analyses (PCA), and (iii) multidimensional scaling (MDS) from GWAS ancestry to control genetic admixture and confounding effects. Additionally, the effect sizes reported for nearly all complex genetic traits require large sample sizes for robust power and precise estimation. Thus, future studies should be designed with the range of magnitude of the effect, and allele frequencies observed in the published studies, and recruit cases and controls accordingly to enable well-designed and well-powered subgroup analyses of the contribution of the homozygous and heterozygous TGFBR1*6A genotypes. This will allow assessment of TGFBR1*6A

Fig. 3 Schematic showing the role of TGFBR1*6A signal peptide. a TGFBR1*6A protein translation and processing, the TGFBR1*6A signal peptide is cleaved between Ala30 and Leu31, whereas the wild-type TGFBR1 is cleaved between Ala33 and Leu34. Both TGFBR1*6A and TGFBR1 wild-type exhibit similar binding affinity to TGFβ ligand and similar protein targeting and translocation functions as the wild-type. b TGFBR1*6A intracellular signaling, TGFBR1*6A maintains intact TGF-β signaling to induce growth and migration in breast cancer cells. It shows similar TGF-β signaling as wild-type TGFBR1 but enhances phosphorylation of ERK1/2 to induce its tumor-promoting effects.
risk association by age, ethnicity, family history, tumor subtypes, tumor grade/histology, other mutation (BRCA1/2) status, response to cancer treatment, and survival/mortality outcomes. When compared to high (BRCA1 and BRCA2; odds ratios ranging from 5.0 to 10.6) and moderate (CHEK2 and ATM; odds ratios ranging from 2.1 to 2.5) penetrance gene variants, TGFBR1*6A has been described as low penetrance (<odds ratios <2.0) tumor susceptibility allele. Although low penetrance gene variants and SNPs are common, their genetic risk to complex diseases such as cancer are comparatively small; carriers of low penetrance breast cancer gene variants are estimated to have less than 20% lifetime risk of developing breast cancers. As such, low penetrance gene variants are postulated to exert their oncogenic influence through additive or multiplicative covariance interactions with other residual gene variants, whose presence or absence can determine disease trait. Thus, the coupling of the TGFBR1*6A family-based case/control studies with the participants’ GWAS data will allow an additional estimation of the genome-wide composite association of TGFBR1*6A with breast cancer risk, as it relates to other gene variants and gene loci that are in linkage disequilibrium, polygenic association or epistatic interaction. Altogether, the estimation of the genetic risk and polygenic risk score of TGFBR1*6A from a robust family-based study will inform clinical guidelines for early detection and adoption of preventive measures.

**TGFBR1*6A MECHANISM OF ACTION AND INFLUENCE ON TGF-β BIOMARKER AND DRUG DEVELOPMENT**

Over the last decade, efforts aimed at developing anti-TGF-β drugs has for the most part led to inconclusive pre-clinical and clinical results and serious adverse events. This is at least partly due to imprecise biomarkers and drug targets identified so far. For example, in a recent phase I study of LY3022859 (an anti-TGFBR2 kinase inactivating monoclonal antibody) in patients with advanced solid tumors including breast carcinomas, there were associated infusion-related reactions such as cytokine release syndrome. Also, the TGF-β small molecule inhibitor galunisertib (LY2157299) and several others acting as neutralizing monoclonal antibodies for TGFBR1 and TGFBR2, and protein traps for TGFBR1 and TGFBR2 invariably yielded promising but minimal clinical benefits. In all, a key weakness is a difficulty to identify precise biomarkers that would tailor TGF-β inhibitors to a specific patient population. Recent studies have proposed and developed a TGF-β gene response signature (TBR5) to assess TGF-β signaling response as a biomarker for cancer predisposition, clinical outcome, and therapeutic response. It utilizes up to 153 genes as probes, which include mainly the TGF-β superfamily ligands (TGFβ1, BMP2), TGF-β receptors (TGFBR1 and TGFBR2), transcription factors (BACH1, TXNIP, and CREB1), and TGF-β responsive genes (ID1, HMOX1, MPM2, and ZEB1). Padua, 2008 and Wahdan-Alasawd 2016, used the TGF-β gene response signature (TBR5) to classify breast tumors as TGF-β gene responsive (TBR5+) and TGF-β gene unresponsive (TBR5-) breast tumor subtypes and found a higher correlation between TBR5+, and ER- and TNBC breast tumor subtypes. Additionally, recent studies relate TGF-β signaling with several cross-talk pathways including microRNA synthesis, stromal fibrosis, and endoplasmic reticulum (ER) stress that regulate immune checkpoint inhibitors, stem cell formation, and metastasis.

The question of whether TGFBR1*6A functions singularly or in combination with other pro-oncogenic pathways to signal modifications in oncogenic traits needs to be answered to provide clues for tailored drug and biomarker development for the affected TGFBR1*6A individuals. First, it is imperative to firmly establish whether TGFBR1*6A signaling is a signal peptide or receptor-mediated in various breast cancer subtypes and cell lines. Subsequent interactions and networking with identified cross-talk pathways and mediators can be traced using recent investigational tools such as Whole Transcriptome RNA Sequencing (RNA-seq) and proteomic profiling techniques coupled with computational network analyses such as comparative gene ontology (GO) enrichment analyses, ingenuity pathway analyses (IPA), and protein–protein interaction (PPI) network analyses. In all, the development of a unique TGF-β gene response signature (TBR5) for TGFBR1*6A as well as the elucidation of a druggable target from the TGFBR1*6A transcriptomic and proteomic studies will be valuable for the prognosis of at-risk individuals and clinical evaluation for precision medicine.

**CONCLUDING REMARKS**

Genetic association studies have identified TGFBR1*6A as a high frequency, low penetrance breast cancer susceptibility allele in breast cancer patients of varied ethnic backgrounds and geographical locations. Its oncogenic influence is attributed to the promotion of cell proliferation, migration, and invasion. Further investigations into the risk associated with individual TGFBR1*6A genotypes as well as their correlation with breast cancer subtypes, disease progression (tumor grade), metastasis, and survival are needed to clarify the population-specific susceptibilities of TGFBR1*6A carriers.

**DATA AVAILABILITY**

Any additional dataset other than cited published data were available upon request to the corresponding author.

Received: 18 January 2022; Accepted: 13 June 2022; Published online: 19 July 2022

**REFERENCES**

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. CA Cancer J. Clin. 69, 7–34 (2019).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. CA Cancer J. Clin. 70, 7–30 (2020).
3. DeSantis, C. E., Miller, K. D., Goding, S. A., Jemal, A. & Siegel, R. L. Cancer statistics for African Americans, 2019. CA Cancer J. Clin. 69, 211–233 (2019).
4. Hu, C. et al. A population-based study of genes previously implicated in breast cancer. N. Engl. J. Med. 384, 440–451 (2021).
5. Metsalfe, K. et al. International trends in the uptake of cancer risk reduction strategies in women with a BRCA1 or BRCA2 mutation. Br. J. Cancer 121, 15–21 (2019).
6. Couch, F. J. et al. Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol. 3, 1190–1196 (2017).
7. Zhang, H., Ahearn, T. J., Lecarpentier, J. & Barnes, D. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat. Genet. 52, 572–581 (2020).
8. Ferreira, M. A. et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat. Commun. 10, 1741 (2019).
9. Barnes, D. R. et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. Genet. Med. 22, 1653–1666 (2020).
10. Ho, W. K. & Tan, M. M. European polygenic risk score for prediction of breast cancer shows similar performance in Asian women. Nat. Commun. 11, 3833 (2020).
11. Desmond, A. et al. Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. JAMA Oncol. 1, 943–951 (2015).
12. Hauke, J. et al. Gene panel testing of 5589 BRCA1/2-negative index patients with breast and ovarian cancer. JAMA Oncol. 3, 1190–1196 (2017).
13. K. Agyemang et al. Concluding remarks (TBR5) and proteomic pro.
18. Moore-Smith, L. & Pasche, B. TGFBR1 signaling and breast cancer. J. Mammary Gland Biol. Neoplasia 16, 89–95 (2011).
19. Feng, X. H. & Derynick, R. Specificity and versatility in tgf-beta signaling through Smads. Annu. Rev. Cell Dev. Biol. 21, 659–693 (2005).
20. Tzavlati, K. & Moustakas, A. TGF-β Signaling. Biomolecules 10 (2020).
21. Moustakas, A. & Heldin, C. H. Non-Smad TGF-beta signals. J. Cell Sci. 118, 3537–3546 (2005).
22. Pasche, B. et al. Type I transforming growth factor beta receptor maps to 9q22 and exhibits a polymorphism and a rare variant within a polyaniline tract. Cancer Res. 58, 2727–2732 (1998).
23. Pasche, B. et al. TbetaR-I(6A) is a candidate tumor susceptibility allele. Cancer Res. 59, 5678–5682 (1999).
24. Buus, R. W. Progress in molecular biology and translational science. Prec. Prog. Mol. Biol. Transl. Sci. 95, xi (2010).
25. Pasche, B. et al. Somatic acquisition and signaling of TGFBR1*6A in cancer. JAMA 294, 1634–1646 (2005).
26. Kaklamani, V. G. et al. Combined genetic assessment of transforming growth factor-beta signaling pathway variants may predict breast cancer risk. Cancer Res. 65, 3454–3461 (2005).
27. Wang, Y. Q., Qi, X. W., Wang, F., Jiang, J. & Guo, Q. N. Association between TGFBR1 polymorphisms and cancer risk: a meta-analysis of 35 case-control studies. PLoS ONE 7, e42899 (2012).
28. Rosman, D. S., Phukan, S., Huang, C. C. & Pasche, B. TGFBR1*6A enhances the migration and invasion of MCF-7 breast cancer cells through RhoA activation. Cancer Res. 68, 1319–1328 (2008).
29. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).
30. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. Nature 526, 68–74 (2015).
31. Pasche, B. et al. TGFBR1*6A and cancer: a meta-analysis of 12 case-control studies. J. Clin. Oncol. 22, 756–758 (2004).
32. Chen, T. et al. Int2G24A variant of transforming growth factor-beta receptor type I is associated with invasive breast cancer. Clin. Cancer Res. 12, 392–397 (2006).
33. Feigelson, H. S. et al. Transforming growth factor beta receptor type I and transforming growth factor beta1 polymorphisms are not associated with postmenopausal breast cancer. Cancer Epidemiol. Biomark. Prev. 15, 1236–1237 (2006).
34. Cox, D. G., Penney, K., Guo, Q., Hankinson, S. E. & Hunter, D. J. TGF BR1 and TGFBR1 polymorphisms and breast cancer risk in the Nurses’ Health Study. BMC Cancer 7, 175 (2007).
35. Baxter, S. W., Choong, D. Y., Eccles, D. M. & Campbell, I. G. Transforming growth factor beta receptor 1 polyalanine polymorphism and exon 5 mutation analysis in breast and ovarian cancer. Cancer Epidemiol. Biomark. Prev. 11, 211–214 (2002).
36. Song, B. et al. TGFBR1*6A and Int2G24A variants of transforming growth factor-beta receptor 1 in Swedish familial and sporadic breast cancer. J. Clin. Oncol. 22, 4275–4277 (2004).
37. Song, B. et al. A novel amplification target, TGFBR2, promotes cell spreading and migration by negatively regulating RhoA in Huh-7 hepatocellular carcinoma cells. Cancer Lett. 275, 27–34 (2009).
38. Christensen, L., Nielsen, M., Andersson, J. & Cleemputts, I. Stromal fibronectin staining pattern and metastasizing ability of human breast carcinoma. Cancer Res. 48, 6227–6233 (1988).
39. Takei, H. et al. Angiogenesis and stromal fibronectin expression in invasive breast carcinoma. Int. J. Oncol. 12, 517–523 (1998).
40. Hori, Y. et al. The regulatory effect of tamoxifen on fibronectin expression in estrogen-dependent MCF-7 breast cancer cells. Oncol. Rep. 15, 1191–1195 (2006).
41. Bhandari, V. Genetic Inflammation in Lung Development and Injury. In The Newborn Lung: Neonatology Questions and Controversies (ed. Bancalari, E.), Chapter 2, 29–55 (Saunders, Philadelphia, PA, 2012).
42. Tasha E., F. & David A., S. Genetics of Lung Disease. In Annual Rev. Cell Dev. Biol. (ed. Bancalari, E.), Chapter 2, 101–168 (Elsevier, Philadelphia, PA, 2016).
43. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. Cell 177, 1080 (2019).
44. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 570, 514–518 (2019).
45. Coingard, J., Lush, M., Beesley, J. & O’Mara, T.A. A case-only study to identify genetic modifiers of breast cancer risk for BRCA1/BRCA2 mutation carriers. Nat. Commun. 12, 1078 (2021).
46. Sekhar, D., Pooja, S., Kumar, S. & Rajender, S. RAD51 35G>C substitution increases breast cancer risk in an ethnic-specific manner: a meta-analysis on 21,236 cases and 19,407 controls. Sci. Rep. 5, 11588 (2015).
47. Chen, X. et al. Associations between RAD51D germline mutations and breast cancer risk and survival in BRCA1/2-negative breast cancers. Ann. Oncol. 29, 2046–2051 (2018).
48. Kadouri, L. et al. A single-nucleotide polymorphism in the RAD51 gene modifies breast cancer risk in BRCA1 carriers, but not in BRCA1 carriers or noncarriers. Br. J. Cancer 90, 2002–2005 (2004).
49. Antoniou, A. C. et al. RAD51 135G>C modifies breast cancer risk among BRCA2 mutation carriers: results from a combined analysis of 19 studies. Am. J. Hum. Genet. 81, 1186–1200 (2007).
50. Wang, W. W. et al. A single nucleotide polymorphism in the 5′ untranslated region of RAD51 and risk of cancer among BRCA1/2 mutation carriers. Cancer Epidemiol. Biomark. Prev. 10, 955–960 (2001).
51. Levy-Lahad, E. et al. A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers. Proc. Natl Acad. Sci. USA 98, 3232–3236 (2001).
52. Zondervan, K. T. & Cardon, L. R. Designing candidate gene and genome-wide case-control association studies. Nat. Protoc. 4, 2492–2501 (2009).
53. Tjian, C., Gregersen, P. K. & Seldin, M. F. Accounting for ancestry: population substructure and genome-wide association studies. Hum. Mol. Genet. 17, R143–R150 (2008).
54. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
55. Helgason, A., N迎接, D. & Stafansson, K. An Icelandic example of the impact of population structure on association studies. Nat. Genet. 40, 90–95 (2008).
56. Narod, S. A. Which genes for hereditary breast cancer? N. Engl. J. Med. 384, 471–473 (2021).
57. Dorrin, L. et al. Breast cancer risk genes - association analysis in more than 113,000 women. N. Engl. J. Med. 384, 428–439 (2021).
58. Zeldov, N. et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am. J. Hum. Genet. 104, 21–34 (2019).
59. Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9, 855–867 (2008).
60. Tolcher, A. W. et al. Phase 1 study of anti-TGFβ receptor type II monoclonal antibody LY3022859 in patients with advanced solid tumors. Cancer Chemother. Pharm. 79, 673–680 (2017).
79. de Gramont, A., Fainvre, S. & Raymond, E. Novel TGF-β inhibitors ready for prime time in onco-immunology. *Oncoimmunology* **6**, e1257453 (2017).

80. Morris, J. C. et al. Phase I study of GC1008 (fresolimumab): a human anti-transforming growth factor-beta (TGFβ) monoclonal antibody in patients with advanced malignant melanoma or renal cell carcinoma. *PloS ONE* **9**, e90353 (2014).

81. Zhu, H. et al. A novel TGFβ trap blocks chemotherapeutics-induced TGFβ signaling and enhances their anticancer activity in gynecologic cancers. *Clin. Cancer Res.* **24**, 2780–2793 (2018).

82. Murao, K. et al. Blockade of TGF-beta inhibits mammary tumor cell viability, migration, and metastases. *J. Clin. Invest.* **109**, 1551–1559 (2002).

83. Padua, D. et al. TGFbeta primes breast tumors for lung metastasis seeding through angioptatin-like 4. *Cell* **133**, 66–77 (2008).

84. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).

85. Wahdan-Alaswad, R. et al. Metformin attenuates transforming growth factor beta (TGF-β) mediated oncogenesis in mesenchymal stem-like/claudin-low triple negative breast cancer. *Cell Cycle* **15**, 1046–1059 (2016).

86. Vanpouille-Box, C. et al. TGFβ is a master regulator of radiation therapy-induced antitumor immunity. *Cancer Res.* **75**, 2232–2242 (2015).

87. Zhang, D. et al. Proteomic characterization of differentially expressed proteins in *Proteomics: Clin. Appl.* **14**, 1332–1340 (2020).

88. Bai, W. D. et al. MiR-200c suppresses TGF-beta-1 induced epithelial-mesenchymal transition and metastasis by suppressing the TGF-beta receptor I to inhibit TGF-beta-induced epithelial-to-mesenchymal transition and metastasis by suppressing the TGF-β/SMAD pathway in breast cancer. *Int J. Oncol.* **55**, 1097–1109 (2019).

89. Wang, S. et al. MicroRNA-133b targets TGFβ receptor I to inhibit TGF-β-induced epithelial-to-mesenchymal transition and metastasis by suppressing the TGF-β/SMAD pathway in breast cancer. *Int J. Oncol.* **55**, 1097–1109 (2019).

90. Xu, X. et al. Association of germline variants in natural killer cells with tumor immune microenvironment subtypes, tumor-infiltrating lymphocytes, immune-therapy response, clinical outcomes, and cancer risk. *JAMA Netw. Open* **2**, e1199292 (2019).

91. Foroutan, M., Cursons, J., Hediyeh-Zadeh, S., Thompson, E. W. & Davis, M. J. A transcriptional program for detecting TGFβ-induced EMT in cancer. *Mol. Cancer Res.* **15**, 619–631 (2017).

92. Wang, X. et al. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* **14**, 2676–2687 (2014).

93. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *JAMA* **284**, 742–752 (2005).

94. Milanese, J. S. et al. Germline variants associated with leukocyte genes predict tumor recurrence in breast cancer patients. *NPJ Precis. Oncol.* **3**, 28 (2019).

95. Wang, X., Liu, Q. & Zhang, B. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* **14**, 2676–2687 (2014).

96. Foroutan, M., Cursons, J., Hediyeh-Zadeh, S., Thompson, E. W. & Davis, M. J. A transcriptional program for detecting TGFβ-induced EMT in cancer. *Mol. Cancer Res.* **15**, 619–631 (2017).

97. Wang, X., Liu, Q. & Zhang, B. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* **14**, 2676–2687 (2014).

98. Li, Y. et al. RNA-Seq and network analysis revealed interacting pathways in TGF-β-treated lung cancer cell lines. *Cancer Inf.* **13**, 129–140 (2014).

**ACKNOWLEDGEMENTS**

The authors acknowledge Michael Olivier, Ph.D. (Center for Precision Medicine, Department of Internal Medicine, Wake Forest Baptist Health, Winston-Salem, USA) for his contribution to discussing parts of the manuscript. All images in Figs. 1 and 3 were created with Biorender (https://biorender.com/) online visual drawing tool.

**AUTHOR CONTRIBUTIONS**

All authors contributed in parts to the conception, writing, and revision of the manuscript. The first author K.A and corresponding author B.P conceived and wrote the manuscript. The co-authors A.M.J, G.W.B, M.J.P, K.S., H.J., and C.B. contributed to discussing and revising the cancer genetics sections, S.S., P.A.F., R.S., K.L.C., H.-K.L., W.Z., H.-W.L., K.W., and P.S. contributed to discussing and revising the molecular signaling sections, and C.D.L. contributed to discussing and revising the cancer genetics and genome-wide association (GWAS) implications of the article.

**COMPETING INTERESTS**

The authors declare no competing interests.

**ADDITIONAL INFORMATION**

Correspondence and requests for materials should be addressed to Boris Pasche.

Reprints and permission information is available at http://www.nature.com/reprints

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