

APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation

David W. Cescon^{a,b,c,d,1}, Benjamin Haibe-Kains^{a,c,1}, and Tak W. Mak^{a,b,c,2}

^aPrincess Margaret Cancer Centre, University Health Network, M5G 2M9 Toronto, Canada; ^bCampbell Family Institute for Breast Cancer Research, M5G 2M9 Toronto, Canada; ^cDepartment of Medical Biophysics, University of Toronto, M5G 1L7 Toronto, Canada; and ^dDivision of Medical Oncology and Hematology, Department of Medicine, University of Toronto, M5G 2C4 Toronto, Canada

Contributed by Tak W. Mak, January 12, 2015 (sent for review December 3, 2014; reviewed by Carlos Caldas)

Genomic sequencing studies of breast and other cancers have identified patterns of mutations that have been attributed to the endogenous mutator activity of APOBEC3B (A3B), a member of the AID/APOBEC family of cytidine deaminases. A3B gene expression is increased in many cancers, but its upstream drivers remain undefined. Furthermore, there exists a common germ-line deletion polymorphism (A3B^{del}), which has been associated with a paradoxical increase in breast cancer risk. To examine causes and consequences of A3B expression and its constitutive absence in breast cancer, we analyzed two large clinically annotated genomic datasets [The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)]. We confirmed that A3B expression is associated with aggressive clinicopathologic characteristics and adverse outcomes and show that A3B expression is highly correlated with proliferative features (mitosis and cell cycle-related gene expression) in breast and 15 of 16 other solid tumor types. However, breast cancers arising in homozygous A3B^{del} individuals with A3B absent did not differ in these features, indicating that A3B expression is a reflection rather than a direct cause of increased proliferation. Using gene set enrichment analysis (GSEA), we detected a pattern of immune activation in A3B^{del} breast cancers, which seems to be related to hypermutation arising in A3B^{del} carriers. Together, these results provide an explanation for A3B overexpression and its prognostic effect, giving context to additional study of this mutator as a cancer biomarker or putative drug target. In addition, although immune features of A3B^{del} require additional study, these findings nominate the A3B^{del} polymorphism as a potential predictor for cancer immunotherapy.

cancer | mutagenesis | cellular proliferation

The recent application of next generation sequencing technologies to characterize the landscape of somatic alterations in solid tumors has yielded major insights into the genes and pathways operant in various cancers. In addition, these studies have enabled the identification of distinct patterns of DNA base alterations that reflect underlying mutational processes (1–6). One of the major discoveries of this effort was the high prevalence of C > T transitions occurring in a preferred sequence motif (TCW, thymine/cytosine/adenine or thymine). This pattern is consistent with the deaminase activity of the AID/APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of enzymes, and thus, its identification invoked an endogenous mutator as a significant contributor to the somatic mutational burden across several cancer types (2–9). Based on the sequence motifs, expression levels, and subcellular localization of APOBEC family members, APOBEC3B (A3B) has been implicated as the likely mutator (6–8).

Analyses of cell line and tumor datasets have shown that A3B gene expression is up-regulated in malignant vs. normal tissues and epithelial cell lines and have shown correlations between A3B expression and the presence of certain somatic mutations,

particularly in TP53 (7, 10) and PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha) (11). Although the factors that cause up-regulation of A3B in cancer cell lines and tumors remain unknown, these observations form the basis for a model where A3B expression contributes to the accumulation of somatic alterations during the process of carcinogenesis and subsequent evolution, and it has been suggested that inhibition of this activity could represent a strategy for cancer prevention or an adjuvant to other therapies (7, 12). A recent report that A3B expression is associated with adverse outcomes in ER+ breast cancer would be consistent with this hypothesis (13).

In contrast to these findings, a germ-line deletion polymorphism in A3B (A3B^{del}), which is present at allele frequencies approaching 100% in some populations (14), has been associated with an increased risk of breast cancer (15, 16). This seemingly paradoxical finding suggests that, if A3B is contributing cancer-causing mutations, it may also play some other function in restraining tumor development. A more recent analysis of the association between the A3B^{del} polymorphism and the APOBEC somatic mutational signature identified an increased mutational burden among A3B^{del} carriers, the etiology of which remains unknown (17).

We sought to address the biologic and clinical questions raised by these observations using two large clinically annotated breast cancer datasets with gene expression, copy number, and mutational data (1, 18) with the chief aim of resolving the incongruence

Significance

Somatic mutagenesis is fundamental to the development and evolution of cancers. APOBEC3B (A3B) is a cellular deaminase, which is overexpressed in cancers and believed to be an important cause of cancer-associated mutations. The factors responsible for A3B up-regulation are unknown. Interestingly, a germ-line deletion polymorphism exists, such that a significant proportion of the global population does not express A3B protein. Using large human cancer datasets, we show that A3B expression is strongly associated with cellular proliferation. Furthermore, we identify a pattern of immune activation related to hypermutation in tumors arising in A3B deletion carriers suggesting that these patients could respond differently to immune-directed therapies. These results provide important context for the ongoing study of A3B as a therapeutic target or biomarker.

Author contributions: D.W.C., B.H.-K., and T.W.M. designed research; D.W.C. and B.H.-K. performed research; D.W.C. and B.H.-K. analyzed data; and D.W.C., B.H.-K., and T.W.M. wrote the paper.

Reviewers included: C.C., Cancer Research United Kingdom.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹D.W.C. and B.H.-K. contributed equally to this work.

²To whom correspondence should be addressed. Email: tmak@uhnres.utoronto.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424869112/-DCSupplemental.

between gene expression and polymorphism associations. Taking advantage of the natural occurrence of *A3B*-null individuals (homozygous for the *A3B^{del}* allele), we assess the direction of causality for *A3B* gene expression associations and the biological consequences of the germ-line *A3B^{del}* polymorphism. These analyses have identified proliferation as the upstream feature associated with both increased *A3B* expression and adverse outcomes and revealed a distinct pattern of immune-related gene expression in tumors arising in *A3B^{del}* carriers compared with WT individuals.

These findings will enable more focused study of *A3B* as a potential biomarker and suggest the possibility that the *A3B^{del}* polymorphism could be used to identify patients with immune-activated tumors who may have differential benefit from cancer immunotherapies.

Materials and Methods

Datasets. We collected the two largest publicly available breast cancer datasets with both genomic and transcriptomic data, namely the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (18) and The Cancer Genome Atlas (TCGA) (1). For METABRIC, we downloaded the normalized gene expression profiles generated using the Illumina HT 12 IDATS platform and the clinical information from the European genome-phenome archive (EGAS0000000083). The mutation calls for *TP53* generated by Sanger sequencing of coding exons used were from the work by Silwal-Pandit et al. (10). *A3B^{del}* polymorphism genotypes were identified using Nexus Copy Number 7.0 software (BioDiscovery Inc.) to process Affymetrix-generated CEL files (EGAD00010000164). Altogether, we collected genomic, transcriptomic, and clinical information for 1,992 breast cancer patients from METABRIC.

For TCGA, we downloaded the normalized gene expression data generated using the Illumina RNA-seq platform and the clinical information from the cBioPortal (19) through the *cgdsr* R package (version 1.1.30). Gene expression values were transformed as $X' = \log_2(X + 1)$, where X represents the normalized fragments per kilobase transcript per million mapped reads values. The point mutation calls for *TP53*, *PIK3CA*, and other recurrently mutated genes were also downloaded from cBioPortal. *A3B^{del}* polymorphism genotype calls determined by analysis of next generation sequencing data as well as the hypermutation calls used were from the work by Nik-Zainal et al. (17). Altogether, we collected genomic, transcriptomic, and clinical information for 1,006 breast cancer patients from TCGA.

For coexpression analysis, in 59 breast cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) (20) and 16 other solid tumor types from TCGA, we used gene expression data (microarray for CCLE and RNA-seq for TCGA) and the coexpression function implemented on cBioPortal (19).

Molecular Subtyping. We used our robust subtype classification model SCMOD2 (21) implemented in the *genefu* R/Bioconductor package (version 1.16.0) to classify breast tumors into well-established molecular subtypes (21): basal (ER-/HER2-), HER2+, and luminal (ER+/HER2-), which can be further classified into luminal B (ER+/HER2- highly proliferative tumors) and luminal A (ER+/HER2- lowly proliferative tumors).

Statistical Analysis.

Association. We used Spearman's rank-based correlation coefficient (22) to estimate the strength and significance of the association between two continuous variables, such as gene expression. To estimate the significance of the association between a continuous variable and a binary variable, we used the Wilcoxon rank sum test (23); we used the Kruskal-Wallis test (24) for discrete variables with more than two levels.

Prognosis. We used the Cox proportional hazards regression model (25) with disease-free survival to assess the prognostic value of gene expression as a continuous variable. In particular, we used the D index (26) as a robust, unit-free estimate of the hazard ratio (HR). Significance of the D index was computed using the log-rank test (27). We estimated the prognostic value of discrete variables, such as gene expression, divided into binary or ternary groups based on quantiles or *A3B* allele calls using Kaplan-Meier survival curves (28). We used the log-rank test to test for differences between survival curves (27).

Gene Ontology/Pathway Analysis. To identify functional clusters of genes coexpressed with *A3B*, we performed Functional Annotation Clustering using DAVID (29) on the top 100 genes most strongly correlated with *A3B* expression in each dataset/tissue type.

Gene Set Enrichment Analysis. To identify the pathways that are significantly enriched in genes associated with the *A3B^{del}* polymorphism, we

computed the association of each gene with *A3B^{del}* calls using the following logistic regression model: $C = \beta_0 + \beta_1 X$, where C is the variable representing the *A3B^{del}* calls coded as 0 for homozygous and heterozygous deletions (*A3B^{del/del}* and *A3B^{del/WT}*, respectively) and 1 for WT (*A3B^{WT/WT}*) as in ref. 17, and X is the continuous expression value of each gene in the gene expression profiling platform. The strength and significance of association between each gene and *A3B^{del}* calls are given by the coefficient β_1 and its corresponding P value (two-tailed Student t test). Based on the genome-wide ranking of genes based on the significance of their association with *A3B^{del}* calls, we used preranked gene set enrichment analysis (GSEA; version 2-2.1.0) (30) to identify gene sets and pathways significantly enriched in genes highly positively or negatively correlated to *A3B^{del}* calls. Gene sets were defined by all gene ontology (GO) (31) terms extracted from class c5 in MSigDB (32).

We used a similar approach to identify gene sets enriched in genes associated with the hypermutation calls defined by Nik-Zainal et al. (17). More specifically, we used the same logistic regression model but replaced the *A3B^{del}* calls (C) with the hypermutation status (H). To assess the effect of the *A3B^{del}* polymorphism beyond the hypermutation phenotype, we updated the model to explicitly control for hypermutation when ranking the genes based on their association with *A3B^{del}* calls: $C = \beta_0 + \beta_1 X + \beta_2 H$. As before, the genes are ranked based on the significance of β_1 , and preranked GSEA is performed to identify the enriched pathways and gene sets.

We performed the GSEA for *A3B^{del}* in each of TCGA and the METABRIC datasets (separately) and for hypermutation status only in TCGA. In each case, we performed the GSEAs for the global population of breast cancer patients as well as each subtype separately. Additional details are provided in *SI Appendix, Supplementary Methods*.

Reproducibility. To ensure full reproducibility, this work complies with the guidelines proposed by Gentleman (33) in terms of the availability of the code and the reproducibility of results and figures. The procedure to properly set up the software environment and run our analysis pipeline is provided in *SI Appendix*.

All analyses were performed in both the METABRIC and TCGA datasets where corresponding data were available. Results from METABRIC, which include more patients and have more complete clinical annotation, are shown in the figures for analyses involving clinicopathologic or survival end points. Results from TCGA are shown in the figures for coexpression and gene signature analyses, because the RNA-seq data in this dataset provide a more quantitative measurement (34). To show the high consistency of our results across datasets, the corresponding figures using the second study are included in *SI Appendix*.

Results

***A3B* Expression Correlates with Adverse Prognostic Features and Is Associated with Outcome in Breast Cancer.** We evaluated the association between *A3B* gene expression [detected by microarray (METABRIC) or RNA-seq (TCGA)] and established clinicopathologic features, focusing on the METABRIC dataset, where clinical annotation is more complete. We found that presence of lymph node involvement, histologic grade, lack of ER expression, and overexpression/amplification of HER2 are each associated with increased expression of *A3B* (Fig. 1). No significant correlation was observed between *A3B* expression and tumor size at diagnosis. Although *A3B* expression levels do vary by intrinsic subtype (*SI Appendix, Fig. S1*), the association between *A3B* expression and tumor grade is present across all tumors as

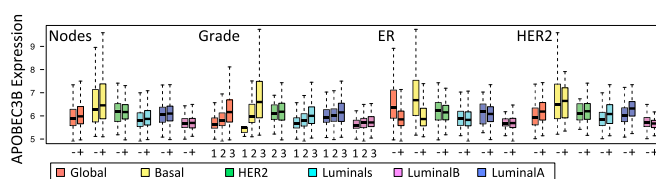


Fig. 1. *A3B* gene expression is associated with clinicopathologic features, including lymph node involvement ($P = 3.2 \times 10^{-4}$), histologic grade ($P = 1.4 \times 10^{-45}$), lack of ER expression ($P = 4.8 \times 10^{-43}$), and HER2 overexpression/amplification ($P = 7.6 \times 10^{-5}$), in METABRIC breast cancers. P values are for the global dataset.

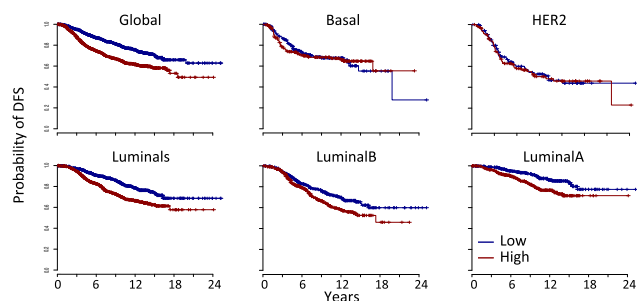


Fig. 2. High *A3B* gene expression is associated with recurrence after treatment for luminal breast cancer in METABRIC. Kaplan–Meier plots are dichotomized by median *A3B* expression for all breast cancers (global) or within each intrinsic subtype. $P < 0.001$ for global, luminals, luminal A, and luminal B. DFS, disease-free survival.

well as within each subtype (Fig. 1). Consistent with these observations, in univariable analysis, *A3B* expression as a continuous variable is associated with recurrence [HR = 1.54; 95% confidence interval (95% CI) = 1.34–1.77; $P = 1.2 \times 10^{-9}$], with its prognostic value being limited to the luminal tumor subtype (HR = 1.60; 95% CI = 1.34–1.93; $P = 4.1 \times 10^{-7}$). To take into account the imprecision of HR estimates within subtype, we included an interaction term between *A3B* expression and luminal status in the Cox model. This interaction term was highly significant ($P = 2.7 \times 10^{-5}$), suggesting that the observed difference of *A3B* prognostic value with respect to luminal status is statistically significant. This adverse prognosis associated with elevated *A3B* expression is similarly evident when patients are binarized into high and low *A3B*-expressing groups. Although this association holds across the entire METABRIC population, it is driven by the luminal subgroups (Fig. 2).

Several gene expression-based prognostic signatures have been developed to predict breast cancer outcomes. We evaluated whether *A3B* expression is associated with scores calculated with these classifiers and observed strong correlation between *A3B* expression and several well-established signatures, including the Genomic Grade Index (35) (Fig. 3 and *SI Appendix*, Fig. S2), Chromosomal Instability (36), MammaPrint (also referred to GENE70) (37), and Oncotype DX (38). These associations were present in both datasets, shared across subtypes, and either specific to or much stronger for *A3B* compared with other *A3* family members. Proliferation has been identified as a common underlying feature driving these prognostic signatures (39). We directly assessed the association between *A3B* expression and key proliferation-associated genes and observed strong correlation [for example, *AURKA* (aurora A kinase) (Fig. 4 and *SI Appendix*, Fig. S3), *MKI67*, and *CCNB1*]. Indeed, coexpression analysis in both METABRIC and TCGA identified strong enrichment of mitosis and cell cycle-related functional ontologies associated with *A3B* expression across all cancers and within subtypes (*SI Appendix*, Table S1). We extended this analysis to 16 other solid tumor types using DAVID with RNA-seq gene expression data from TCGA and confirmed that *A3B* expression is strongly associated with mitosis and cell cycle in 16 of 17 examined, with the exception lung squamous cell carcinoma (*SI Appendix*, File S1). As in the breast cancer datasets, the correlations between *AURKA* and *A3B* gene expression in these other tumor types provide support for the results of the GO analyses (*SI Appendix*, Fig. S4).

To further evaluate the prognostic importance of *A3B* expression, we performed multivariate analyses with a model incorporating clinical features (age, grade, tumor size, nodal status, and ER status) and proliferation-associated gene expression (*AURKA*). When all of these features were considered, *A3B* expression was no longer a significant predictor of outcome (all tumors: HR = 0.98; 95% CI = 0.85–1.12; $P = 0.74$; luminals: HR = 1.12; 95% CI = 0.86–1.47; $P = 0.39$). Borderline significance

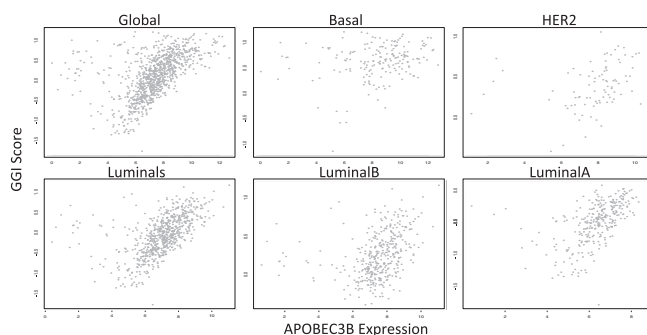


Fig. 3. *A3B* gene expression correlates with the Genomic Grade Index (GGI) score in all breast cancers (global: $\rho = 0.51$, $P < 1 \times 10^{-16}$) and within individual intrinsic subtypes (basal: $\rho = 0.23$, $P = 3.1 \times 10^{-3}$; Her2: $\rho = 0.33$, $P = 1.1 \times 10^{-4}$; luminals: $\rho = 0.55$, $P < 1 \times 10^{-16}$; luminal A: $\rho = 0.39$, $P = 1.4 \times 10^{-14}$; luminal B: $\rho = 0.53$, $P < 1 \times 10^{-16}$) in TCGA.

was observed in luminal A and HER2 subgroups (luminal A: HR = 2.11; 95% CI = 0.99–4.49; $P = 0.053$; HER2: HR = 1.60; 95% CI = 1.07–2.39; $P = 0.02$).

Together, these analyses in two large independent breast cancer datasets affirm that *A3B* expression is associated with adverse prognostic features and outcomes in breast cancer and reveal a link to cellular proliferation.

***A3B* Expression Is Increased in *TP53*- but Not *PIK3CA*-Mutated Breast Cancers.**

Because *A3B* has been implicated as an endogenous mutator in breast and other cancers, we evaluated the association between mutations in the most common recurrently altered genes in breast cancer (*TP53* and *PIK3CA*) (1, 40) and *A3B* expression. As has been recently shown (10), we found that *A3B* expression is increased in tumors with *TP53* mutations in both the METABRIC and TCGA datasets (Fig. 5 and *SI Appendix*, Fig. S5A). For *PIK3CA*, however, the presence of point mutations was, instead, associated with lower *A3B* expression (Fig. 5) in breast cancers. Because the spectrum of base changes ascribed to *A3B* activity occurs in a specific sequence context, we considered the *PIK3CA* hotspot mutations separately: the E542K and E545K alterations conform to the *A3B* motif, whereas H1047R does not (11). Nevertheless, each point mutation exhibited a similar pattern, where the presence of a *PIK3CA* alteration was associated with lower *A3B* expression (Fig. 5 and *SI Appendix*, Fig. S5B). Systematic analysis of genes recurrently mutated in breast cancer (9) using TCGA breast dataset revealed no remarkable pattern of increased *A3B* expression in somatically mutated vs. unmutated tumors, although this analysis was

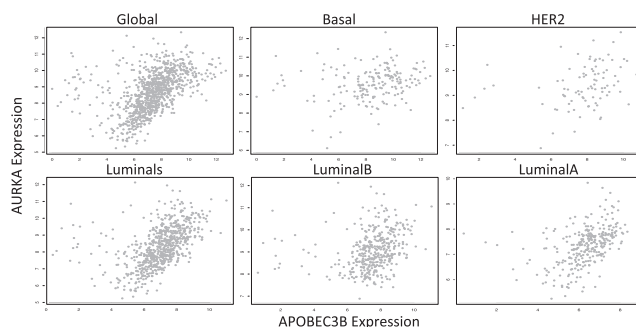


Fig. 4. *A3B* gene expression correlates with expression of the proliferation-associated gene *AURKA* across all breast cancers (global: $\rho = 0.46$, $P < 1 \times 10^{-16}$) and within individual intrinsic subtypes (basal: $\rho = 0.22$, $P = 3.8 \times 10^{-3}$; Her2: $\rho = 0.28$, $P = 9.8 \times 10^{-4}$; luminals: $\rho = 0.50$, $P < 1 \times 10^{-16}$; luminal A: $\rho = 0.31$, $P = 1.1 \times 10^{-9}$; luminal B: $\rho = 0.45$, $P = 8.9 \times 10^{-16}$) in TCGA.

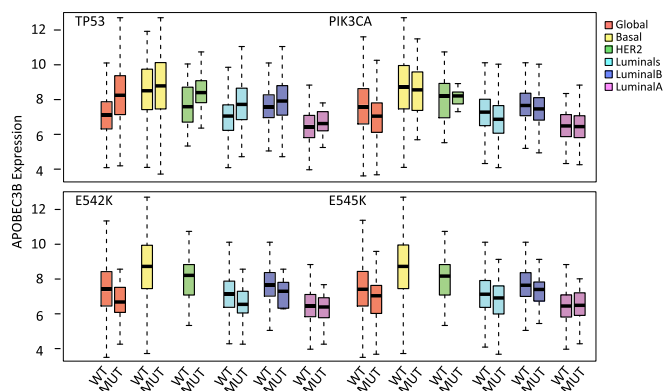


Fig. 5. *A3B* gene expression in TCGA breast cancers is increased in tumors with *TP53* mutation (vs. WT) but decreased in tumors with any *PIK3CA* mutation or the E542K or E545K *A3B* motif mutations (vs. WT). Similar associations are present across all tumors and within individual subtypes. Note that groups with less than three samples were omitted. MUT, mutation.

limited by low mutation frequency (<10%) for all genes with the exceptions of *TP53*, *PIK3CA*, *CDH1*, and *GATA3*.

Association Between *A3B* Expression and Proliferation Is Retained in Breast Cancer Cell Lines. To assess whether the observed associations between *A3B* expression and proliferation are cell-autonomous, we further evaluated the relationship between *A3B* expression and proliferation in the CCLE breast cancer cell lines dataset. As in the METABRIC and TCGA breast tumor datasets, coexpression analysis identified mitosis and cell cycle-related terms as the most strongly associated GOs (*SI Appendix*, Table S2).

Germ-Line *A3B* Deletion Polymorphism Is Associated with *A3B* Expression but Not Clinical Variables. To assess the causality of these *A3B* associations, we evaluated associations between *A3B*^{del} and these same factors, reasoning that, if *A3B* expression contributes to the development of high-grade, proliferative, and clinically aggressive cancers, the constitutional absence of *A3B* protein would result in an altered spectrum of tumor features. *A3B*^{del} allele frequencies were determined for METABRIC, and previously published values (17) were used for TCGA. As expected, *A3B* expression was highly associated with the *A3B*^{del} genotype (Fig. 6 and *SI Appendix*, Fig. S6). Importantly, we did not observe reciprocal up-regulation of other A3 family members in the absence of *A3B*—specifically, expression of *A3A* (the 3' UTR of which is replaced by the *A3B* 3' UTR in the polymorphic *A3B*^{del} allele) was not increased in tumors from *A3B*^{del} carriers (Fig. 6 and *SI Appendix*, Fig. S6).

In contrast to the associations observed between *A3B* expression and clinicopathologic features, we observed no association between the *A3B*^{del} germ-line polymorphism and tumor ER or HER2 status, lymph node involvement, tumor grade, intrinsic subtype distribution (Table 1), or somatic *TP53* or *PIK3CA* mutation status (*SI Appendix*, Table S3). Age at diagnosis was slightly lower in *A3B*^{del} carriers. As previously shown, *A3B*^{del} is associated with the presence of the hypermutation phenotype defined by Nik-Zainal et al. (17).

Similarly, we observed no association (either statistically significant or consistent trend) between *A3B*^{del} status and expression of individual proliferation-associated genes or prognostic signature scores (Fig. 6 and *SI Appendix*, Fig. S7). Finally, we observed no association between the *A3B*^{del} polymorphism genotype and breast cancer outcomes (Fig. 7).

Germ-Line *A3B* Deletion Polymorphism Is Associated with Expression of Immune-Related Genes in Breast Tumors. The surprising observation that *A3B*^{del} is not associated with any standard clinicopathologic characteristics, despite the ascribed role for *A3B* as

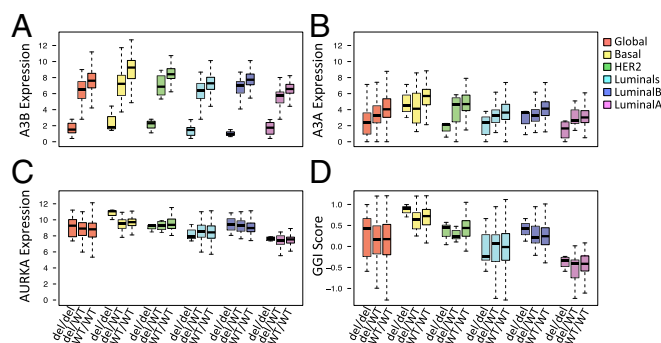


Fig. 6. *A3B* expression is (A) reduced in TCGA breast cancers from patients with germ-line *A3B*^{del} polymorphism (B) without a reciprocal increase in *A3A*. The *A3B*^{del} genotype is not associated with (C) expression of *AURKA* or (D) the Genomic Grade Index (GGI). Similar patterns are present across all tumors and within intrinsic subtypes.

an endogenous mutator and the shown association between *A3B*^{del} and breast cancer risk, prompted us to investigate the METABRIC and TCGA gene expression datasets to identify potential distinguishing features of tumors arising in *A3B*^{del} carriers. To discover biological processes associated with *A3B*^{del}, we applied GSEA (30) in METABRIC and TCGA. Consistent with our observations above, no association between proliferation-associated processes and *A3B*^{del} was observed. Instead, in each dataset, a striking association was identified between *A3B*^{del} status and enrichment of immune response-related gene sets [Table 2 shows the top-ranked gene set for METABRIC (immune response) and TCGA (cytokine binding)].

Immune Activation in *A3B*^{del} Tumors Is Related to a Hypermutation Phenotype. Antitumor immune responses have been attributed to neoepitopes arising from somatic mutations in cancer cells (41),

Table 1. Association of *A3B*^{del} polymorphism with clinicopathologic features of breast cancers in the METABRIC and TCGA datasets

Genotype	<i>A3B</i> ^{del/del}	<i>A3B</i> ^{del/WT}	<i>A3B</i> ^{WT/WT}	P value
METABRIC				
All patients*	18	199	1,771	
Grades 1/2/3	1/7/9	22/71/95	147/695/851	0.7
Age (y) median (SD)	59.8 (10.7)	59.8 (12.8)	62.3 (13.0)	0.04
Size (cm) median (interquartile range)	2.5 (1.0)	2.1 (1.4)	2.3 (1.3)	0.3
ER+	83.3%	77.8%	77.3%	0.8
HER2+	0.0%	18.8%	18.0%	0.4
Node+	50.0%	46.2%	47.8%	0.9
Subtype				0.1
Luminal A/B	4/9	66/57	609/639	
Basal	3	44	343	
Her2	2	32	180	
TCGA				
All patients*	13	86	597	
Age (y) median (SD)	54 (12)	54 (12.4)	58 (13.3)	0.07
ER+	69.2%	77.6%	78.9%	0.7
Node+	62.5%	55.2%	49.8%	0.6
Subtype				0.54
Luminal A/B	3/3	25/34	160/240	
Basal	2	12	95	
Her2	5	12	83	
Hypermutated†	30.8%	19.0%	10.7%	0.02

*For each data type, the total number of cases may differ because of missing or incomplete data.

†From the work by Nik-Zainal et al. (17).

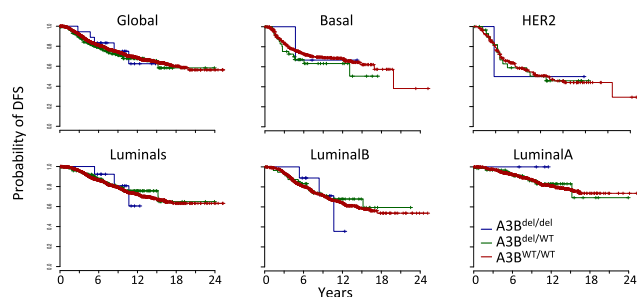


Fig. 7. Germ-line *A3B^{del}* genotype is not associated with recurrence after treatment for early breast cancer in METABRIC. Kaplan-Meier plots show disease-free survival (DFS) by *A3B^{del}* genotype for all breast cancers (global) and by intrinsic subtypes. $P > 0.6$ for all.

and the $A3B^{del}$ polymorphism has been associated with a hypermutation phenotype in breast cancers (17). To assess whether hypermutation could be responsible for the $A3B^{del}$ -associated immune signature that we detected, we compared hypermutated with nonhypermutated breast cancers from TCGA (17) using GSEA. Across all tumors as well as within luminal and basal subtypes, we identified a strong signal of immune activation in the hypermutated cancers reminiscent of the expression patterns in $A3B^{del}$ cancers (Table 2). $A3B^{del}$ is associated with hypermutation status, particularly in luminal tumors (*SI Appendix, Fig. S8*). To determine whether the $A3B^{del}$ polymorphism contributes to immune activation independent of the hypermutation phenotype, we constructed a GSEA linear model comparing tumors based on $A3B^{del}$ status corrected for hypermutation status. In this model, the immune activation pattern was attenuated (Table 2), suggesting that hypermutation contributes to this phenomenon.

Discussion

Since the initial identification of a pattern of mutations fitting the activity of APOBEC3 enzymes in breast cancer sequencing data (3, 7, 8), numerous studies have gone on to identify a similar pattern of mutations across several human cancer types (2, 5, 6, 8, 42), implicating *A3B* in somatic mutagenesis and by inference carcinogenesis. Despite the great interest that these results have attracted, the cause of *A3B* up-regulation that is observed has remained enigmatic. Our results show that *A3B* expression is increased in highly proliferative states, consistent with the increased expression in cancers vs. normal tissues (7, 43) and high- vs. low-grade tumors. Our analysis of *A3B*-null tumors (homozygous for the *A3B*^{del} polymorphism), however, revealed that the constitutive absence of *A3B* has no significant impact on the same clinicopathologic features. These data suggest that *A3B* expression is regulated in a cell cycle-dependent manner, and gene expression profiling of cell cycle states supports this conclusion (44). An association with proliferation is also supported by the observation that APOBEC pattern mutations predominate later in tumor evolution (42, 45). Although increased *A3B* expression might result in accumulation of additional oncogenic mutations, resulting in increased proliferation and a positive feedback loop, we show that *A3B* expression per se is not the primary cause of proliferative or adverse clinical features in breast cancer but rather, a consequence of these features. This finding has implications for the potential clinical use or application of strategies to inhibit *A3B*.

Although these analyses do not directly address the increased risk of breast cancer development that has been associated with the *A3B^{del}* allele (15, 16), we find no evidence of increased *APOBEC3A-B* hybrid mRNA expression in deletion carriers, which has been suggested as a potential cause of increased mutational burden and cancer risk in carriers (46). However, while performing our analyses, we did observe a remarkable association between expression of each of the *APOBEC3* family members with the exception of *A3B* and IFN/

inflammation-associated gene expression (STAT1 signature) (*SI Appendix, Fig. S10*). In light of this finding, it is conceivable that elevated expression of APOBEC3A occurs in *A3B^{del}* carriers during periods of inflammation and results in hit and run mutagenesis but is no longer detectable in the established tumors in the breast cancer datasets.

We found no evidence of systematic differences in traditional clinicopathologic features, subtype distributions, or outcomes between breast cancers based on germ-line $A3B^{del}$ status, although the relatively low $A3B^{del}$ allele frequency (and consequent small numbers of $A3B^{del}$ homozygous individuals) in these predominantly Caucasian patient datasets is a limitation. Future analyses of other ethnic populations with higher $A3B^{del}$ allele frequency (14) will be important to confirm these findings. Nevertheless, we have identified a reproducible pattern of increased immune response—including cytokine and T-cell activation gene expression programs—in tumors arising in $A3B^{del}$ carriers. Although the nature and temporal development of the immune processes associated with the observed expression changes detected by GSEA are of great interest, this technique provides limited resolution to characterize the associated phenomena, and additional study is required. However, given the canonical function of APOBEC enzymes in the immune system (47, 48), several potential mechanisms for this observation could be considered arising from either tumor or immune cell intrinsic effects of the $A3B^{del}$ polymorphism. The previously described association of $A3B^{del}$ with a somatic hypermutation phenotype (17) and recent interest in the relationship between somatic mutational burden and antitumor immune response prompted us to examine the differences in gene expression profiles between the hypermutated and nonhypermutated breast cancers. Our analysis revealed, to our knowledge, previously undescribed and marked immune activation in the hypermutated cancers. By controlling for the hypermutation status, we observed that the original signal that we identified in $A3B^{del}$ tumors can be attributed, at least in part, to increased mutational burden (or the process responsible for it). Because hypermutation is likely a continuum but was binarized in these analyses (17), some or all of the residual immune-associated patterns in $A3B^{del}$ tumors may also be attributable to degrees of hypermutation. Lacking exome- or genome-level sequencing data in METABRIC, confirmatory analyses in our second breast cancer dataset were not possible; however, given the reproducible $A3B^{del}$ -associated immune gene expression signatures, we believe that this phenomenon is likely generalizable.

In light of the immune activation observed in $A3B^{del}$ -hypermutated tumors, the increased breast cancer risk found in $A3B^{del}$ carriers (15, 16) seems, perhaps, counterintuitive. Although our analyses of established tumors are unable to

Table 2. GSEA of $A3B^{del}$ vs. $A3B^{WT}$ and hypermutated vs. nonhypermutated breast cancers

Comparison and gene set	Normalized ES	FDR q value
METABRIC		
<i>A3B^{del}</i> vs. <i>A3B^{WT}</i>		
Immune response	-3.19	$<1 \times 10^{-16}$
Cytokine binding	-2.68	$<1 \times 10^{-16}$
TCGA breast		
<i>A3B^{del}</i> vs. <i>A3B^{WT}</i>		
Immune response	-2.01	0.006
Cytokine binding	-2.18	0.002
Hypermutated vs. not		
Immune response	2.81	$<1 \times 10^{-16}$
Cytokine binding	2.30	$<1 \times 10^{-16}$
<i>A3B^{del}</i> vs. <i>A3B^{WT}</i> hypermutated model		
Immune response	-1.77	0.096
Cytokine binding	-1.93	0.068

ES, enrichment score; FDR, false discovery rate.

ascertain the presence or temporal development of the observed immune phenomena during tumorigenesis, the increased mutational rate/genomic instability presumably overrides the immune activation and results in tumor formation. We did not find any association between *A3B^{del}* and breast cancer outcomes, which one might expect in the setting of an activated immune response. However, recent analyses of tumor-infiltrating lymphocytes in breast cancer have shown this pathologic measure of immune response to have important prognostic or predictive effects only in specific tumor subtypes or clinical contexts (49, 50). Thus, the lack of any detectable effect of *A3B^{del}* on prognosis in this analysis is perhaps unsurprising, especially given the relationship with hypermutation that we identified.

Notwithstanding the lack of prognostic effects of the *A3B^{del}* polymorphism in patients treated with standard therapies in the METABRIC and TCGA cohorts, the recent and dramatic advances in the development of cancer immunotherapies (51) provide an emerging context where the immune effects that we have observed could be of clinical importance. Recent studies examining factors predictive of response to immune checkpoint inhibitors have

identified associations between mutation- and immune-associated features and treatment response (52–54), although they rely on complex tumor analyses, including whole-exome sequencing. It is certainly plausible that the immune features associated with *A3B^{del}* cancers (by way of the hypermutated phenotype) could influence response to immunotherapy-based treatment strategies. Genotyping for *A3B^{del}* (which can be performed quickly and inexpensively from peripheral blood or tissue samples) could provide a simple selection or stratification factor to identify populations of interest for such treatments, and its exploration as a predictive biomarker is warranted.

ACKNOWLEDGMENTS. We thank Dr. Samuel Aparicio [British Columbia Cancer Agency (BCCA)] for his advice and input. This study used data generated by The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium; we thank TCGA, Cancer Research United Kingdom, and the BCCA for sharing these invaluable data with the scientific community. D.W.C. was supported by a Canadian Institutes of Health Research/Canadian Association of Medical Oncologists Fellowship. B.H.-K. was supported by the Gattuso-Slaight Personalized Cancer Medicine Fund at Princess Margaret Cancer Centre.

1. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61–70.
2. Alexandrov LB, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415–421.
3. Nik-Zainal S, et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5):979–993.
4. Stephens PJ, et al.; Oslo Breast Cancer Consortium (OSBREAC) (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486(7403):400–404.
5. Roberts SA, et al. (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45(9):970–976.
6. Burns MB, Temiz NA, Harris RS (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 45(9):977–983.
7. Burns MB, et al. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494(7437):366–370.
8. Taylor BJ, et al. (2013) DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* 2:e00534.
9. Lawrence MS, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.
10. Silwal-Pandit L, et al. (2014) TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clin Cancer Res* 20(13):3569–3580.
11. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR (2014) APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep* 7(6):1833–1841.
12. Harris RS (2013) Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. *Genome Med* 5(9):87.
13. Sieuwerts AM, et al. (2014) Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm Cancer* 5(6):405–413.
14. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE (2007) Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* 3(4):e63.
15. Long J, et al. (2013) A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst* 105(8):573–579.
16. Xuan D, et al. (2013) APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* 34(10):2240–2243.
17. Nik-Zainal S, et al. (2014) Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 46(5):487–491.
18. Curtis C, et al.; METABRIC Group (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346–352.
19. Gao J, et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269):pl1.
20. Barretina J, et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603–607.
21. Haibe-Kains B, et al. (2012) A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst* 104(4):311–325.
22. Spearman C (2010) The proof and measurement of association between two things. *Int J Epidemiol* 39(5):1137–1150.
23. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6):80–83.
24. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 47(260):583–621.
25. Cox DR (1972) Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 34(2):187–220.
26. Royston P, Sauerbrei W (2004) A new measure of prognostic separation in survival data. *Stat Med* 23(5):723–748.
27. Bland JM, Altman DG (2004) The logrank test. *BMJ* 328(7447):1073.
28. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53(282):457–481.
29. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
30. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
31. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
32. Liberzon A, et al. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27(12):1739–1740.
33. Gentleman R (2005) Reproducible research: A bioinformatics case study. *Stat Appl Genet Mol Biol* 4(2005):2.
34. Wang C, et al. (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* 32(9):926–932.
35. Sotiriou C, et al. (2006) Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98(4):262–272.
36. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z (2006) A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38(9):1043–1048.
37. van't Veer LJ, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530–536.
38. Paik S, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27):2817–2826.
39. Desmedt C, Sotiriou C (2006) Proliferation: The most prominent predictor of clinical outcome in breast cancer. *Cell Cycle* 5(19):2198–2202.
40. Lawrence MS, et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495–501.
41. Chen DS, Mellman I (2013) Oncology meets immunology: The cancer-immunity cycle. *Immunity* 39(1):1–10.
42. de Bruin EC, et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346(6206):251–256.
43. Leonard B, et al. (2013) APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res* 73(24):7222–7231.
44. Peña-Díaz J, et al. (2013) Transcription profiling during the cell cycle shows that a subset of Polcomb-targeted genes is upregulated during DNA replication. *Nucleic Acids Res* 41(5):2846–2856.
45. Zhang J, et al. (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346(6206):256–259.
46. Caval V, Suspène R, Shapira M, Vartanian JP, Wain-Hobson S (2014) A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat Commun* 5(2014):5129.
47. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22(2):367–377.
48. Moris A, Murray S, Cardinaud S (2014) AID and APOBECs span the gap between innate and adaptive immunity. *Front Microbiol* 5:534.
49. Loi S, et al. (2013) Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol* 31(7):860–867.
50. Loi S, et al. (2014) Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial. *Ann Oncol* 25(8):1544–1550.
51. Pardoll DM (2012) The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 12(4):252–264.
52. Snyder A, et al. (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 371(23):2189–2199.
53. Herbst RS, et al. (2014) Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515(7528):563–567.
54. Gubin MM, et al. (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 515(7528):577–581.