

RESEARCH ARTICLE

Open Access



# Germline *APOBEC3B* deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation

Wei Xiong Wen<sup>1,2†</sup>, Jaslyn Sian-Siu Soo<sup>1,2†</sup>, Pui Yoke Kwan<sup>1</sup>, Elaine Hong<sup>1</sup>, Tsung Fei Khang<sup>3</sup>, Shivaani Mariapun<sup>1</sup>, Christine Shu-Mei Lee<sup>1</sup>, Siti Norhidayu Hasan<sup>1</sup>, Pathmanathan Rajadurai<sup>4</sup>, Cheng Har Yip<sup>1,4</sup>, Nur Aishah Mohd Taib<sup>2,5</sup> and Soo Hwang Teo<sup>1,2,5\*</sup>

## Abstract

**Background:** *APOBEC3B* is a cytosine deaminase implicated in immune response to viral infection, cancer predisposition and carcinogenesis. Germline *APOBEC3B* deletion is more common in East Asian women and confers a modest risk to breast cancer in both East Asian and Caucasian women. Analysis of tumour samples from women of European descent has shown that germline *APOBEC3B* deletion is associated with an increased propensity to develop somatic mutations and with an enrichment for immune response-related gene sets. However, this has not been examined in Asian tumour samples, where population differences in genetic and dietary factors may have an impact on the immune system.

**Methods:** In this study, we determined the prevalence of germline *APOBEC3B* deletion and its association with breast cancer risk in a cross-sectional hospital-based Asian multi-ethnic cohort of 1451 cases and 1442 controls from Malaysia. We compared gene expression profiles of breast cancers arising from *APOBEC3B* deletion carriers and non-carriers using microarray analyses. Finally, we characterised the overall abundance of tumour-infiltrating immune cells in breast cancers from TCGA and METABRIC using ESTIMATE and relative frequency of 22 immune cell subsets in breast cancers from METABRIC using CIBERSORT.

**Results:** The minor allelic frequency of *APOBEC3B* deletion was estimated to be 0.35, 0.42 and 0.16 in female populations of Chinese, Malay and Indian descent, respectively, and that germline *APOBEC3B* deletion was associated with breast cancer risk with odds ratios of 1.23 (95 % CI: [1.05, 1.44]) for one-copy deletion and 1.38 (95 % CI: [1.10, 1.74]) for two-copy deletion compared to women with no deletion. Germline *APOBEC3B* deletion was not associated with any clinicopathologic features or the expression of any *APOBEC* family members but was associated with immune response-related gene sets (FDR  $q$  values < 0.05). Analysis of breast cancers from METABRIC revealed breast cancers from *APOBEC3B* deletion carriers to have significantly higher abundance of tumour-infiltrating immune cells ( $P < 0.001$ ).

(Continued on next page)

\* Correspondence: soohwang.teo@cancerresearch.my

†Equal contributors

<sup>1</sup>Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia

<sup>2</sup>Breast Cancer Research Unit, University Malaya Cancer Research Institute, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia

Full list of author information is available at the end of the article



(Continued from previous page)

**Conclusions:** Taken together, our data suggests that tumour-infiltrating immune cells may be an important feature of breast cancers arising in women with *APOBEC3B* germline deletion, and that this may be of particular interest in Asian women where the germline deletion is more common.

**Keywords:** APOBEC3B, Asian, Breast cancer, Breast cancer risk, Immune response

## Background

Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3B (*APOBEC3B*) is a member of the APOBEC family of cytidine deaminases whose canonical function is in innate immune defence [1]. *APOBEC3B* plays role in retrovirus and endogenous retrotransposon restriction by hyperediting complementary DNA (cDNA) intermediates [1]. Recent studies on whole genome- and exome-sequenced cancers have implicated the enzymatic activity of the APOBEC family in two of the 21 mutational signatures, suggesting that APOBEC mutational processes may be a key driver to carcinogenesis [2–4]. Signature 2 consists of predominantly C to T transitions in the TCX sequence context while Signature 13 consists of predominantly C to G transversions in the TCX sequence context [2]. Given that *APOBEC3B* expression is upregulated in breast tumours, and that its upregulation is associated with APOBEC-associated mutational signatures [3–5], it has been suggested that *APOBEC3B* may be an endogenous source of mutations in cancers, especially breast cancer.

The *APOBEC3B* deletion allele frequency is estimated to be 37 % in East Asian populations, and 6 % in Europeans [6]. The deletion involves the removal of a sequence of length 29.5 kb between apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3A (*APOBEC3A*) and *APOBEC3B*, resulting in a hybrid transcript consisting of the coding region from *APOBEC3A* and the 3' UTR from *APOBEC3B*. *APOBEC3B* deletion polymorphism is associated with increased risk to breast cancer, with odds ratios (ORs) of 1.31 (95 % confidence interval [CI]: [1.21, 1.42]) for one-copy deletion and 1.76 (95 % CI: [1.57, 1.97]) for two-copy deletion in East Asian women [7], and 1.21 (95 % CI: [1.02, 1.43]) for one-copy deletion and 2.29 (95 % CI: [1.04, 5.06]) for two-copy deletion in women of European descent [8]. The prevalence of germline *APOBEC3B* deletion and its association with breast cancer remain unknown in other populations.

Given that *APOBEC3B* has been proposed to be an endogenous source of mutations, it is paradoxical that germline loss of the *APOBEC3B* results in increased propensity to develop tumours with hypermutation. Analysis of breast cancers from The Cancer Genome Atlas (TCGA) showed that germline *APOBEC3B* deletion is associated with an OR of 2.68 for one-copy deletion, and an OR of 3.82 for two-copy deletion with tumours that have a hypermutator phenotype [9]. One plausible explanation is

that germline loss of *APOBEC3B* results in hypermutation driven by other endogenous sources of mutation, for example, by other members of the APOBEC cytosine deaminases. Furthermore, it has been proposed that hypermutation in the absence of *APOBEC3B* generates tumour-specific antigens which activate the immune system. Indeed, immune response-related gene sets were found to be enriched in breast cancers arising from *APOBEC3B* deletion carriers [10].

The purpose of this study is to determine the prevalence of germline *APOBEC3B* deletion and its association with breast cancer risk in a multi-ethnic Asian cohort of Chinese, Malay and Indian women in Malaysia, to compare the gene expression profiles of breast cancers arising from Asian *APOBEC3B* deletion carriers and non-carriers, and lastly, to characterise the presentation of tumour-infiltrating immune cells in breast cancers.

## Methods

### Study subjects

All breast cancer patients who participated in the Malaysian Breast Cancer Genetic Study between October 2002 and March 2014 were included in a hospital-based cohort of breast cancer cases. The cases were drawn from two hospitals, namely University Malaya Medical Centre and Sime Darby Medical Centre [11]. Of the 2452 breast cancer patients, 378 individuals were excluded from the study if they were either (a) male breast cancer patients ( $n = 7$ ); (b) diagnosed with non-invasive breast cancer ( $n = 133$ ); (c) had missing histopathology data ( $n = 151$ ); (d) had insufficient or low-quality genomic DNA samples ( $n = 46$ ) or (e) were of ethnicity other than Chinese, Malay or Indian, or unknown ethnicity ( $n = 58$ ). Control subjects included women who attended an opportunistic mammography screening programme at Sime Darby Medical Centre between October 2011 and September 2014 [11]. Of the 1819 control subjects, 271 individuals were excluded from the study if they were either (a) diagnosed with invasive breast cancer ( $n = 12$ ); (b) were younger relatives of study participants ( $n = 209$ ) or (c) were of ethnicity other than Chinese, Malay or Indian, or unknown ethnicity ( $n = 64$ ). The remaining cases and controls were frequency-matched on 5-year age group within each ethnic group. Thus, a total of 1468 breast cancer cases consisting of 991 Chinese, 253 Indians, and 224 Malays and a total of 1451 healthy controls consisting of 986 Chinese, 246 Indians and 219 Malays

were included in the study. All study participants provided written informed consent. The study was approved by the Medical Ethics Committee of University Malaya Medical Centre (application number: 842.9) and the Independent Ethics Committee of Sime Darby Medical Centre (application numbers: 201109.4 and 201208.1).

#### Copy number analyses

Primers and probes highly specific to the target gene, *APOBEC3B* (assay ID: Hs04504055\_cn), and the reference gene, *RNase P*, were purchased from Applied Biosystems (Foster City, CA, USA). Duplex real-time quantitative PCR assays were performed on 384-well plates using Applied Biosystems ViiA 7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Each plate included one negative control (water) and one calibrator. The calibrator is a sample previously confirmed to carry two copies of *APOBEC3B* using droplet digital PCR. All samples were tested in triplicate, and fluorescence signals were normalised to the passive reference dye ROX. The quantitative PCR amplification curves were analysed using ViiA 7 software on a plate-by-plate basis. CopyCaller software version 2.0 (Applied Biosystems, Foster City, CA, USA) was used to assign the copy number from the raw quantification cycle (Cq). Samples with confidence values below 95 % or absolute  $z$  scores of 2.65 or greater were excluded from further analysis. A total of 1451 breast cancer cases consisting of 984 Chinese, 245 Indians, and 222 Malays and a total of 1442 healthy controls consisting of 985 Chinese, 239 Indians and 218 Malays were included in the analysis.

#### Microarray experiment

Twenty-eight invasive breast cancers which were positive for expression of estrogen receptor (ER) and negative for expression of human epithelial receptor 2 (HER2) (by immunohistochemistry) were selected for gene expression analysis. The clinicopathologic features of the tumour samples are shown in Additional file 1: Table S1. The samples were frequency- and ethnicity-matched between each genotype group. Hematoxylin and eosin-stained frozen sections were obtained and tumour content was confirmed by a pathologist (PR). RNA from each sample (50 ng) was used to generate biotinylated sense-strand cDNA using the GeneChip WT Plus Reagent Kit according to the manufacturer's instruction (Affymetrix Inc., Santa Clara, CA, USA). Microarray experiments were performed using GeneChip Human Transcriptome Array 2.0 (Affymetrix Inc., Santa Clara, CA, USA) according to the manufacturer's instructions.

#### Gene expression analysis

The raw data was normalised in the Affymetrix Expression Console Software using the robust multi-array

average (RMA) method [12, 13]. Subsequent analysis of gene expression data was performed in the statistical computing language R (version 3.2.0) [14]. Tumour samples were categorised into  $A3B^{\text{del}}$  (samples with one- and two-copy *APOBEC3B* deletion) and  $A3B^{\text{wt}}$  (samples with no *APOBEC3B* deletion) breast cancers as previously described [9, 10]. For gene set enrichment analysis (GSEA; [http://www.broadinstitute.org/gsea/]), we used the list of all genes that were pre-ranked based on their moderated  $t$  statistic value using the 'limma' R package [15, 16]. Gene sets were defined by all gene ontology (GO) [17] terms extracted from class c5 in MsigDB [18]. A gene set was considered significantly enriched if its false discovery rate (FDR)  $q$  value was below 0.05.

#### Data accession code

Microarray data are available from the Gene Expression Omnibus (GEO) database (accession number: GSE73540). The submission conformed to requirements for minimum information about a microarray experiment (MIAME) [19].

#### Statistical analyses

The association between odds for breast cancer and *APOBEC3B* copy number was modelled using logistic regression (implemented in the Statistical Package for the Social Sciences (SPSS) version 16.0, SPSS Inc., Chicago, IL, USA). In our logistic regression model, we controlled for the effects of age, number of live births, age of first live birth, number of first-degree relatives with breast cancer, and oral contraceptive use. Odds ratios were estimated for one-copy and two-copy deletion genotypes compared with no deletion genotypes in the three ethnic groups. Odds ratio trend across genotypes was tested using the likelihood ratio test. The differences in the subject characteristics were determined using  $t$  test and chi-squared test while the differences in clinicopathologic features across  $A3B^{\text{del/del}}$ ,  $A3B^{\text{del/wt}}$  and  $A3B^{\text{wt/wt}}$  patients were tested using analysis of variance (ANOVA) and chi-squared test. The APOBEC family member gene expression across  $A3B^{\text{del/del}}$ ,  $A3B^{\text{del/wt}}$  and  $A3B^{\text{wt/wt}}$  breast cancers were compared using the Kruskal-Wallis test while immune scores and the relative frequency of tumour-infiltrating immune cell subsets across  $A3B^{\text{del/del}}$ ,  $A3B^{\text{del/wt}}$  and  $A3B^{\text{wt/wt}}$  breast cancers were compared using the Kruskal-Wallis and Mann-Whitney  $U$  test for three- and two-group comparison, respectively.

#### TCGA and the METABRIC dataset

The two largest publicly available breast cancer datasets with both genomic and transcriptomic data, namely TCGA [20] and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) [21] were analysed. For the TCGA breast cancer dataset, clinical information was retrieved from the cBio Cancer

Genomics Portal [22] and germline *APOBEC3B* copy number status was obtained from Nik-Zainal et al. [9] while normalised RNASeqV2 gene expression profiles were retrieved from TCGA data portal. Gene expression values were transformed as  $X = \log_2(X + 1)$ , where X represents the normalised fragments per kilobase transcript per million mapped reads values. Altogether, 706 breast cancers from TCGA with normalised gene expression and germline *APOBEC3B* copy number status were collected and analysed. For the METABRIC dataset, clinical information was retrieved from the Synapse Commons archive (syn2133322; <https://www.synapse.org/>) and germline *APOBEC3B* copy number status was obtained from Cescon et al. [10] while normalised gene expression profiles were retrieved from the European genome-phenome archive (EGAS0000000083). Altogether, 1988 breast cancers from METABRIC with normalised gene expression profile and germline *APOBEC3B* copy number status were collected and analysed.

#### Characterisation of tumour-infiltrating immune cells

Immune scores for breast cancers from TCGA and METABRIC datasets were defined using the Estimation of STromal and Immune cell in Malignant Tumour (ESTIMATE) tissues using Expression data algorithm [23]. Cell type Identification By Estimating Relative Subsets Of known RNA Transcripts (CIBERSORT), a systematic and comprehensive tool to enumerate leukocyte abundance in bulk tumours using expression data from microarray experiment, was used to characterise the relative frequency of 22 tumour-infiltrating immune cell subtypes in breast cancers from METABRIC [24].

## Results

#### Germline *APOBEC3B* deletion is associated with breast cancer risk but not clinicopathologic features of breast cancer in Malaysian women

The characteristics of the study cohort are shown in Table 1 and Additional file 1: Table S2. Both cases and controls had similar distribution of age, age at menarche, age at menopause and proportion receiving hormone replacement therapy. Breast cancer patients had significantly fewer live births ( $P < 0.01$ ), and were younger at age of first live birth ( $P < 0.01$ ). The fewer number of live births in cases was consistent with other epidemiological studies on breast cancer risk but the younger age of first live birth in cases was not consistent with other studies. One possible explanation is that our opportunistic mammography screening programme attracted women with higher socioeconomic status compared to our cases, which was a hospital-based cohort of patients. As this variable is unlikely to impact significantly on the genetic association, we proceeded with the genotype analysis [25]. Indeed, the genotype distributions of *APOBEC3B*

**Table 1** Demographic characteristics and known breast cancer risk factors of study participants<sup>a</sup>

Category	Cases (N = 1451)	Controls (N = 1442)	P value
Demographic factors			
Age (year)	51.6 ± 7.5	51.4 ± 7.3	0.49 <sup>b</sup>
Reproductive risk factors			
Age at menarche (year)	13.0 ± 1.5	12.9 ± 1.4	0.05 <sup>b</sup>
Age at menopause (year) <sup>d</sup>	49.4 ± 4.5	49.3 ± 4.4	0.54 <sup>b</sup>
Number of live births <sup>e</sup>	2.9 ± 1.3	3.2 ± 1.5	<0.01 <sup>b</sup>
Age at first live birth (year) <sup>e</sup>	26.4 ± 5.0	27.5 ± 4.9	<0.01 <sup>b</sup>
Other risk factors			
First-degree relatives with breast cancer (%)	13.2	11.0	0.08 <sup>c</sup>
First- or second-degree relatives with breast cancer (%)	19.6	17.8	0.23 <sup>c</sup>
Oral contraceptive (%) <sup>f</sup>	28.2	31.5	0.06 <sup>c</sup>
Hormone replacement therapy (%) <sup>f</sup>	9.5	9.8	0.80 <sup>c</sup>

<sup>a</sup>Unless otherwise specified, data are presented in mean ± standard deviation  
<sup>b</sup>t test

<sup>c</sup>Chi-squared test

<sup>d</sup>Among postmenopausal women

<sup>e</sup>Among parous women

<sup>f</sup>Ever user

in all three ethnic groups were in good agreement with expectations under Hardy-Weinberg equilibrium (Additional file 1: Table S3).

The estimated minor allelic frequency (± standard error) of *APOBEC3B* in females of Chinese, Malay and Indian descent were  $0.35 \pm 0.01$ ,  $0.42 \pm 0.02$  and  $0.16 \pm 0.02$ , respectively. We found germline *APOBEC3B* deletion was associated with breast cancer risk, with ORs of 1.23 [95 % CI: 1.05, 1.44] for one-copy deletion and 1.38 [95 % CI: 1.10, 1.74] for two-copy deletion ( $P = 0.005$ ; Table 2) compared to women with no deletion. After adjusting for age, number of live births and age of first live birth, there was negligible change in our results although number of live births and age of first live birth were significant in the logistic regression model

**Table 2** Association between germline *APOBEC3B* deletion and breast cancer risk in Malaysian women

Genotype	Cases (N = 1451) n (%)	Controls (N = 1442) n (%)	OR (95 % CI) <sup>a</sup>
<i>A3B</i> <sup>wt/wt</sup>	591 (40.7)	670 (46.5)	1.00 (reference)
<i>A3B</i> <sup>del/wt</sup>	649 (44.7)	599 (41.5)	1.23 [1.05, 1.44]
<i>A3B</i> <sup>del/del</sup>	211 (14.5)	173 (12.0)	1.38 [1.10, 1.74]
<i>P</i> <sup>b</sup> trend			0.005

*A3B*<sup>wt/wt</sup> no deletion, *A3B*<sup>del/wt</sup> one-copy deletion, *A3B*<sup>del/del</sup> two-copy deletion

<sup>a</sup>Adjusted for age

<sup>b</sup>Test for trends of odds were two-sided and based on likelihood ratio tests



( $P < 0.05$ ; data not shown). The adjusted odds ratio were 1.26 (95 % CI: 1.05, 1.52) for one-copy deletion and 1.36 (95 % CI: [1.05, 1.78]) for two-copy deletion compared with women with no deletion ( $P = 0.014$ ; data not shown). The association was also observed after stratifying the study group by ethnicity, but it was only statistically significant in Chinese women (Additional file 1: Table S4).

We observed no significant association between germline *APOBEC3B* deletion with any clinicopathologic features of breast cancers such as age of diagnosis, tumour grade, tumour size, lymph node involvement, estrogen receptor (ER) status, progesterone receptor (PR) status, human epithelial receptor 2 (HER2) status, and triple-negative breast cancer (TNBC) status in the overall cohort as well as in the Chinese and Indians (Table 3, Additional file 1: Table S5). However, we did observe HER2 status to be significantly associated with non-carriers of *APOBEC3B* deletion in the Malays with frequency of 61.9 % compared to one- and two-copy *APOBEC3B* deletion carriers with frequency of 39.1 % and 36.2 %, respectively ( $P = 0.01$ ; Additional file 1: Table S5).

#### Germline *APOBEC3B* deletion is associated with *APOBEC3B* expression but not with other *APOBEC* family members

We compared the gene expression of the *APOBEC* family members in individuals with different *APOBEC3B* copy number. We found *APOBEC3B* expression to be significantly associated with *APOBEC3B* copy number ( $P < 0.001$ ), but *APOBEC3A* expression was not observed to be significantly higher in *APOBEC3B* deletion carriers (Fig. 1). There was also no association between *APOBEC3B* copy number and expression of other *APOBEC* family members (Additional file 1: Figure S1). The expression of apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 4 (*APOBEC4*) was observed to

be only marginally significant across *APOBEC3B* copy number ( $P = 0.04$ ), and its expression was not found to be associated with the *APOBEC3B* gene copy number.

#### Breast cancers arising from *APOBEC3B* deletion carriers are enriched for immune response-related gene sets

We performed gene expression analysis to investigate the biological processes underlying the loss of *APOBEC3B* in our tumour samples. The result of the GSEA showed that *A3B<sup>del</sup>* breast cancers were significantly enriched for immune response-related gene sets (FDR  $q$  value  $< 0.05$ ; Table 4).

#### Loss of *APOBEC3B* deletion is associated with tumour-infiltrating immune cells

To investigate whether the enrichment of immune response-related gene sets in *A3B<sup>del</sup>* breast cancers is associated with the presentation of tumour-infiltrating immune cells, we characterised the immune scores of *A3B<sup>del/del</sup>*, *A3B<sup>del/wt</sup>* and *A3B<sup>wt/wt</sup>* breast cancers from TCGA and METABRIC datasets using the ESTIMATE algorithm. We observed the immune scores to be significantly different across *APOBEC3B* copy number with *A3B<sup>del/del</sup>*, *A3B<sup>del/wt</sup>*, and *A3B<sup>wt/wt</sup>* breast cancers with median immune scores of 1354, 1321 and 1093 ( $P < 0.001$ ), respectively in breast cancers from METABRIC. However, we did not observe the immune scores to be significantly different across *APOBEC3B* copy number in the smaller sample set of breast cancers from TCGA with *A3B<sup>del/del</sup>*, *A3B<sup>del/wt</sup>*, and *A3B<sup>wt/wt</sup>* breast cancers with median immune scores of -45, 273 and -13, respectively (Fig. 2). We extended our analysis by stratifying the breast cancers using PAM50 classification and found immune scores in luminal A and basal-like breast cancers from METABRIC to be significantly different across *APOBEC3B* copy number ( $P = 0.02$  and  $P = 0.01$ , respectively) but not in luminal B and HER2-enriched

**Table 3** Association between *APOBEC3B* copy number and clinicopathologic features of breast cancers<sup>a</sup>

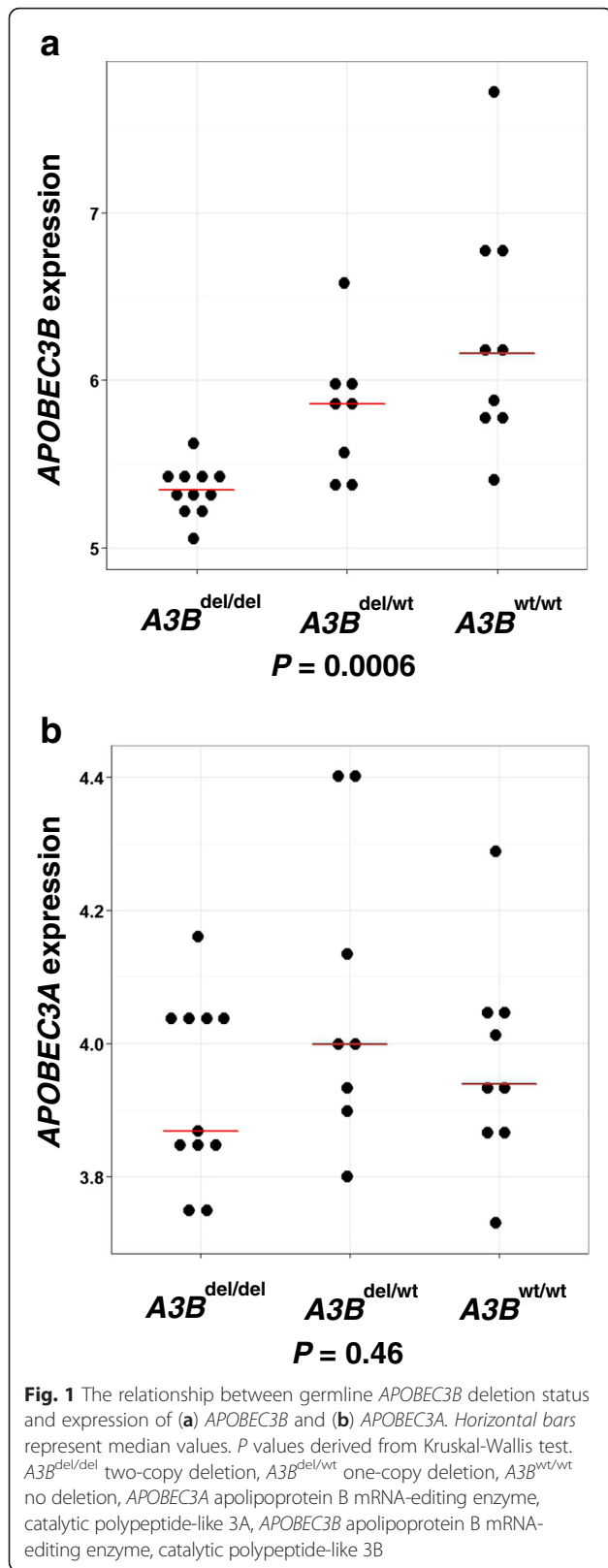
Clinical variables	Data available <i>N</i> (%)	<i>A3B<sup>wt/wt</sup></i>	<i>A3B<sup>del/wt</sup></i>	<i>A3B<sup>del/del</sup></i>	<i>P</i> value
Age (year)	1451 (100)	50.0 ± 7.4	51.0 ± 7.8	51.0 ± 7.1	0.23 <sup>b</sup>
Grade 1/2/3 (%)	1140 (78.6)	14/51/36	11/54/36	14/51/36	0.76 <sup>c</sup>
Size (cm)	1220 (84.1)	2.5 ± 2.5	2.5 ± 2.2	2.5 ± 1.7	0.93 <sup>b</sup>
Node + (%)	1308 (90.1)	47.7	44.4	39.7	0.17 <sup>c</sup>
ER+ (%)	1370 (94.4)	65.8	68.7	65.2	0.49 <sup>c</sup>
PR+ (%)	1346 (92.8)	51.1	51.0	47.5	0.65 <sup>c</sup>
HER2 (%)	1328 (91.5)	49.0	44.6	40.4	0.09 <sup>c</sup>
TNBC (%)	1240 (87.7)	12.5	12.2	17.5	0.17 <sup>c</sup>

*A3B<sup>wt/wt</sup>* no deletion, *A3B<sup>del/wt</sup>* one-copy deletion, *A3B<sup>del/del</sup>* two-copy deletion, ER+ estrogen receptor positive, PR+ progesterone receptor positive, HER2 human epithelial receptor 2, TNBC triple-negative breast cancer

<sup>a</sup>Unless otherwise specified, data are presented in median ± interquartile range

<sup>b</sup>ANOVA

<sup>c</sup>Chi-squared test



**Table 4** Gene set enrichment analysis of  $A3B^{del}$  vs.  $A3B^{wt}$  breast cancers

Gene set	Normalised ES	FDR q value
Chemokine receptor binding	-1.97	0.02
Chemokine activity	-1.93	0.03
Inflammatory response	-1.86	0.04

ES enrichment score, FDR false discovery rate

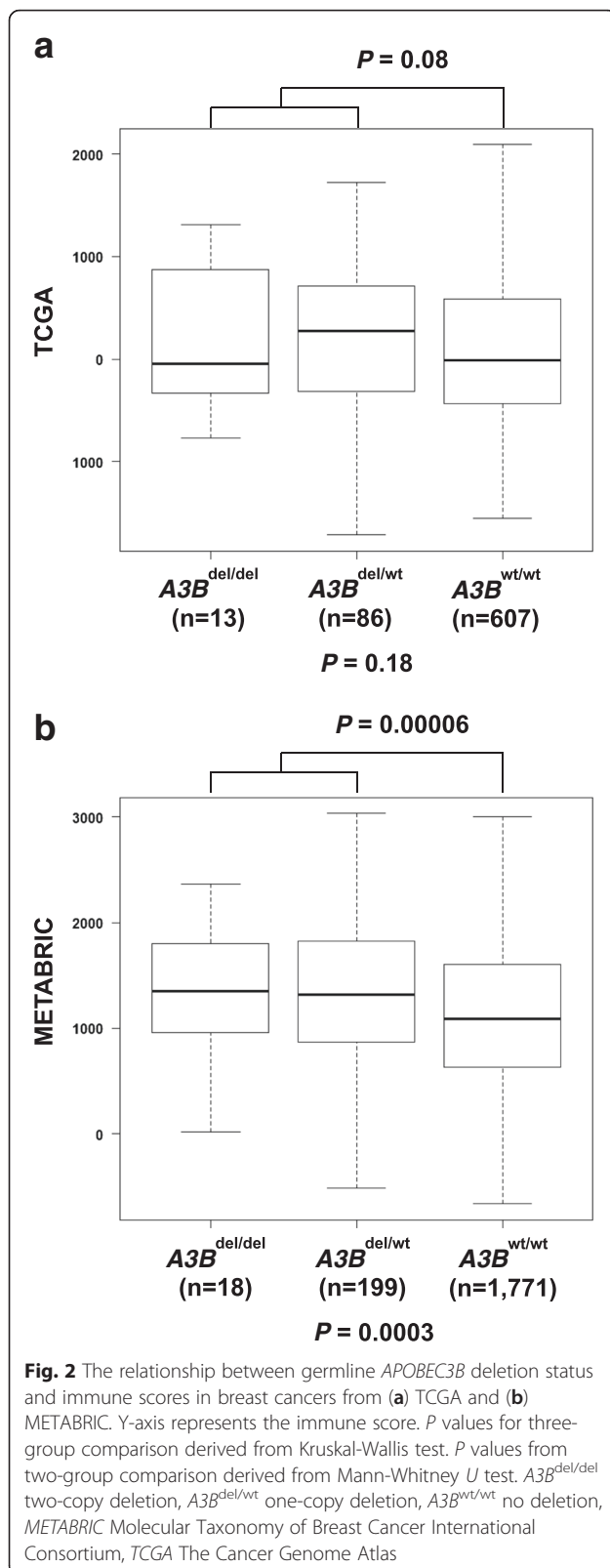
breast cancers nor in any of the subtypes of breast cancers from TCGA (Additional file 1: Figures S2 and S3).

Using CIBERSORT, a bioinformatics tool used to infer immune cell composition from microarray datasets, we computed the relative frequency of 22 tumour-infiltrating immune cell subsets in breast cancers from METABRIC using the gene expression signatures. We observed monocytes and macrophages, CD8 T cells, mast cells, plasma cells, and CD4 T cells to be the most common immune cell subsets with mean fractions of 0.107, 0.080, 0.065, 0.047 and 0.039 respectively. Notably, breast cancers with at least one-copy *APOBEC3B* deletion had lower fraction of macrophages M2 compared to  $A3B^{wt/wt}$  breast cancers with mean fractions of 0.196 and 0.207, respectively, but this difference was only marginally significant ( $P = 0.04$ ; Additional file 1: Table S6). Three-group comparison across *APOBEC3B* copy number for macrophages M2 fractions was not significant with  $A3B^{del/del}$ ,  $A3B^{del/wt}$ , and  $A3B^{wt/wt}$  breast cancers with mean fractions of 0.179, 0.197 and 0.207, respectively (Additional file 1: Table S6).

### Discussion

In this study, we showed that germline *APOBEC3B* deletion is associated with an increased risk to breast cancer among Chinese, Malay and Indian women in Malaysia. Furthermore, we report that germline *APOBEC3B* deletion appears to be associated with an enrichment of immune response genes and that this may arise from the enrichment and activation of tumour-infiltrating immune cells.

Our breast cancer association results are consistent with previous findings in East Asian [7] and Caucasian women [8], which implicate germline *APOBEC3B* deletion as a susceptibility factor for breast cancer by conferring carriers a modest risk to the disease. Given the role of *APOBEC3B* as an endogenous mutator [5], it has been proposed that loss of *APOBEC3B* may be associated with less aggressive tumour phenotype [10]. However, we did not observe any association between germline *APOBEC3B* status and the clinicopathologic features of breast cancer patients. Our findings are consistent with that of a previous study analysing TCGA and the METABRIC datasets, which predominantly comprise patients of European descent [10].



Our gene expression analysis of fresh frozen tumour samples showed that in germline *APOBEC3B* deletion carriers, there is corresponding loss of gene expression of the *APOBEC3B* and this is consistent with previous studies [9, 10]. It has previously been proposed using transient transfection assays in 293T lines, that the deletion of *APOBEC3B* generates a novel *APOBEC3A* transcript, which is fused to the *APOBEC3B* 3' UTR, and could result in a stabilized *APOBEC3A* mRNA transcript that may be more active than in non-*APOBEC3B* deletion carriers [26]. In our study, we did not find any upregulation of *APOBEC3A* or any of the other *APOBEC* cytosine deaminase family members, suggesting that there is no detectable feedback loop at the mRNA level that results in co-regulation of any of the members of this family. These gene expression results are consistent with data from TCGA, which similarly examined fresh frozen breast tumour samples but using a different analysis platform (transcriptomic sequencing) [9]. These data do not exclude the possibility that *APOBEC3A* may nonetheless be the endogenous source of mutations in *APOBEC3B* deletion carriers [27].

We showed that  $A3B^{del}$  breast cancers from our Asian cohort were enriched for immune response-related genes, indicating a possible involvement of immune response in carcinogenesis. This is consistent with previous results using TCGA and METABRIC datasets, where breast cancers arising from predominantly *APOBEC3B* deletion carriers of European descent were reported to be enriched for immune response-related gene sets [10]. Further characterisation of immune scores in breast cancers from the METABRIC dataset demonstrated that  $A3B^{del/del}$  and  $A3B^{del/wt}$  breast cancers present with significantly higher immune scores compared to  $A3B^{wt/wt}$ , reflecting higher abundance of tumour-infiltrating immune cells in  $A3B^{del}$  breast cancers. Taken together, these data suggest that germline *APOBEC3B* deletion is associated with the enrichment of immune response-related gene sets arising from tumour-infiltrating immune cells. While it is anticipated that immune activation reflected by an increase in tumour-infiltrating immune cells would lead to reduced breast cancer risk in *APOBEC3B* deletion carriers, the conundrum between increased tumour-infiltrating immune cells and increased breast cancer risk in these individuals could be attributed to the increased propensity to develop higher mutation load in *APOBEC3B* deletion carriers [9] and this may override immune activation as previously suggested [10]. Although a number of factors are hitherto known to modulate immune cells infiltration, including history of pregnancy and lactation, commensal and pathobionts, and dietary and hormonal factors [28], to the best of our knowledge, this is the first

time that germline genetic status has been suggested to modulate tumour-associated immune infiltrates.

## Conclusions

Taken together, we propose that tumour-infiltrating immune cells may be an important feature of breast cancers arising in women with *APOBEC3B* germline deletions and that this may be of particular interest in Asian women where the germline deletion is more common. Further research into the potential role of *APOBEC3B* genotype as a predictive biomarker for cancer prevention or anticancer immunotherapy may be warranted [10], particularly for Asian women.

## Additional file

**Additional file 1:** Supplementary tables and figures. (DOCX 755 kb)

## Abbreviations

AID, activation-induced cytidine deaminase; APOBEC1, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide 1; APOBEC2, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 2; APOBEC3A, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3A; APOBEC3B, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3B; APOBEC3C, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3C; APOBEC3D, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3D; APOBEC3F, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 F; APOBEC3G, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G; APOBEC4, apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 4; CIBERSORT, Cell type Identification By Estimating Relative Subsets Of known RNA Transcripts; CI, confidence interval; ER, estrogen receptor; ESTIMATE, Estimation of Stromal and Immune cell in Malignant Tumour tissues using Expression data; FDR, false discovery rate; GO, gene ontology; GSEA, gene set enrichment analysis; HER2, human epithelial receptor 2; METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; OR, odds ratio; TCGA, The Cancer Genome Atlas.

## Acknowledgements

We thank participants and their families for taking part in this study: Yoon Sook-Yee, Lee Sheau Yee, Daphne Lee, Phuah Sze Yee, Patsy Ng, Maheswari Jagannathan and Dr. Ho Weang Kee for assistance with recruitment of patients, tissue collection, DNA preparation, and helpful discussions. We thank Prof. Dr. Woo Yin Ling for permitting us to use her laboratory's qPCR machine for our copy number analyses. We thank TCGA and METABRIC for access to the datasets. This study was funded by research grants from the Ministry of Higher Education, University Malaya (UM.C/HIR/MOHE/06), Estee Lauder Group of Companies and Cancer Research Malaysia.

## Authors' contributions

WXW carried out the copy number experiment, performed the statistical and bioinformatics analyses, conceptualised the study and drafted the manuscript. JSSS carried out the copy number and microarray experiments, performed the statistical analysis, conceptualised the study and drafted the manuscript. PYK, TFK and CSML performed the statistical and bioinformatics analyses and revised the manuscript. EH carried out the microarray experiment and revised the manuscript. SM performed the statistical analysis, recruited the study participants, collected and processed the blood samples and revised the manuscript. SNH recruited the study participants, processed the blood and tumour samples and revised the manuscript. PR reviewed the pathological status of the samples and revised the manuscript. CHY recruited the study participants and clinical data, collected the blood and tumour samples and revised the manuscript. NAMT recruited the study participants and clinical data, collected the blood and tumour samples and revised the manuscript. SHT conceptualised the study and drafted the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Cancer Research Malaysia, Subang Jaya, Selangor, Malaysia. <sup>2</sup>Breast Cancer Research Unit, University Malaya Cancer Research Institute, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia. <sup>3</sup>Institute of Mathematical Sciences, Faculty of Science, University Malaya, Kuala Lumpur, Malaysia. <sup>4</sup>Sime Darby Medical Centre, Subang Jaya, Selangor, Malaysia. <sup>5</sup>Department of Surgery, Faculty of Medicine, University Malaya Medical Centre, Kuala Lumpur, Malaysia.

Received: 28 October 2015 Accepted: 8 May 2016

Published online: 27 May 2016

## References

1. Refsland EW, Harris RS. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol*. 2013;371:1–27.
2. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415–21.
3. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*. 2013;45(9):977–83.
4. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013;45(9):970–6.
5. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013;494(7437):366–70.
6. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet*. 2007;3(4):e63.
7. Long J, Delahanty RJ, Li G, Gao YT, Lu W, Cai Q, et al. A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst*. 2013;105(8):573–9.
8. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, et al. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*. 2013;34(10):2240–3.
9. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet*. 2014;46(5):487–91.
10. Cescon DW, Haibe-Kains B, Mak TW. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc Natl Acad Sci U S A*. 2015;112(9):2841–6.
11. Phuah SY, Lee SY, Kang P, Kang IN, Yoon SY, Thong MK, et al. Prevalence of PALB2 mutations in breast cancer patients in multi-ethnic Asian population in Malaysia and Singapore. *PLoS One*. 2013;8(8):e73638.
12. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*. 2001; 98(1):31–6.
13. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249–64.
14. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. <http://www.R-project.org/>. (Accessed on 7 July 2015).
15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
16. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3:Article3.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9.
18. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–40.
19. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001;29(4):365–71.



20. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
21. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346–52.
22. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):11.
23. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
24. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
25. Mariapun S, Li J, Yip CH, Taib NA, Teo SH. Ethnic differences in mammographic densities: an Asian cross-sectional study. *PLoS One*. 2015; 10(2):e0117568.
26. Caval V, Suspene R, Shapira M, Vartanian JP, Wain-Hobson S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat Commun*. 2014;5:5129.
27. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet*. 2015; 47(9):1067–72.
28. Kroemer G, Senovilla L, Galluzzi L, Andre F, Zitvogel L. Natural and therapy-induced immunosurveillance in breast cancer. *Nat Med*. 2015;21(10):1128–38.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

