scientific reports

Check for updates

OPEN

Comprehensive germline and somatic profiling of high-risk Thai breast cancer via next-generation sequencing

Kornyok Kamdee[1,2], Ekkapong Roothumnong[3,4], Wanna Thongnoppakhun[4], Krittiya Korphaisarn[3], Panee Nakthong[4], Peerawat Dungort[4], Chutima Meesamarnpong[4], Supakit Wiboontanasarn[4], Warisara Tansa-nga[4], Kittiporn Punuch[3,4], Khontawan Pongsuktavorn[4], Warunya Tititumjariya[4], Chittapat Lertbussarakam[4], Jantanee Wattanarangsan[4], Jiraporn Sritun[4], Numpueng Ridchuayrod[3,4], Manop Pithukpakorn[3,4]✉ & Bhoom Suktitipat[1,5]✉

Breast cancer genomic landscapes differ across ethnic groups, yet the somatic profile of Thai breast tumours has remained uncharacterised. This study analysed 1676 high-hereditary-risk Thai breast cancer patients, identified according to National Comprehensive Cancer Network (NCCN) guideline. Germline alterations were assessed in 1370 cases using a custom 36-core cancer panel. Somatic mutations were characterised in formalin-fixed, paraffin-embedded tumour tissues from 180 of the 1676 patients using the 501-gene Oncomine Comprehensive Assay Plus panel. Pathogenic or likely pathogenic (P/LP) variants were detected in 13% of the 1370 germline analyses, with *BRCA1* and *BRCA2* being the most frequently altered genes. The prevalence of P/LP variants in *BRCA1*, *BRCA2*, and *PALB2* differed from that observed in other ethnic cohorts. In somatic profiling, *TP53* emerged as the most frequently mutated gene, especially in HER2 and TNBC tumours, whereas *MAP3K1* and *GATA3* were the most frequently mutated genes in the HR+/HER2- tumours. Moreover, hormone-receptor-positive (HR+) tumours showed distinct mutation patterns compared with other ethnicities. Notably, germline carriers exhibited lower *PIK3CA* mutation rates than non-carriers. These findings advance our understanding of Thai breast cancer genomics and underscore the importance of ethnic diversity in cancer research, offering insights into tailored screening and therapeutic approaches.

Keywords Breast cancer, Genomic landscape, Next-generation sequencing, Somatic mutation

Breast cancer is a genetically complex and heterogeneous disease. Female breast cancer is the most common cause of cancer-related death worldwide, accounting for 15.5% of all cancer cases[1]. Incidence and mortality have risen both in Thailand and globally[2]. Germline mutations that increase breast cancer risk occur in 5–10% of the general population[3,4] and in 9–24% of high-risk populations[5–7]. Most breast cancer cases, however, are sporadic, highlighting the crucial role of somatic mutations in tumourigenesis and therapy.

One major research obstacle is the predominance of genomic data derived from Western breast cancer populations, while data on Asian cohorts are comparatively limited[8–14]. Variations in cancer-related gene mutation frequencies have also been reported among different ethnic groups[13–16]. These findings suggest that current genomic information may not accurately reflect the Thai breast cancer population, given regional and ethnic diversity in cancer genomes. Consequently, assumptions of similarity in cancer biology, aetiology, or the selection of treatment and prevention strategies based on genomic data may be incorrect.

Despite these concerns, a comprehensive genomic characterisation of Thai breast cancer, including somatic profiling, remains lacking. The limited data raise crucial clinical questions. Do genes commonly mutated in

[1]Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. [2]Department of Medical Technology, School of Allied Health Sciences, Walailak University, Nakhon Si Thammarat, Thailand. [3]Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. [4]Siriraj Genomics, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. [5]Integrative Computational BioScience (ICBS) Center, Mahidol University, Nakhon Pathom, Thailand. ✉email: manop.pit@mahidol.ac.th; bhoom.suk@mahidol.ac.th

Western populations exhibit similar patterns in Thai patients, or are additional mutations prevalent among Thais? We hypothesised that the Thai breast cancer genomic landscape differs from that of other ethnic groups.

In this study, we characterised the germline and somatic profiles of Thai breast cancer patients using peripheral blood and formalin-fixed, paraffin-embedded (FFPE) tumour tissues. Germline mutations in blood samples were detected with a custom 36-core cancer panel. The 501-gene Oncomine Comprehensive Assay Plus (OCA-Plus; Thermo Fisher Scientific, Waltham, MA, USA) was used for somatic mutation analysis in FFPE specimens.

Results

Spectrum of germline mutations in Thai patients with breast cancer and cross-ethnic comparisons

Blood samples from 1370 patients were subjected to germline mutation sequencing using a custom 36-core cancer panel (Supplementary Table 1). We found that 13% (178/1370) of patients carried pathogenic or likely pathogenic (P/LP) mutations; these included 121 (8.83%) *BRCA1/2* mutations and 60 (4.38%) non-*BRCA1/2* mutations. *BRCA1* ($n=66$) and *BRCA2* ($n=56$) were the most frequently altered genes. Aside from *BRCA1/2*, *ATM* ($n=11$) was the most commonly mutated gene, followed by *TP53* ($n=7$), *RAD50* ($n=6$), *BARD1* ($n=5$), *NF1* ($n=4$), *MUTYH* ($n=3$), *PALB2* ($n=3$), and *PMS2* ($n=3$) (Fig. 1A).

Among the *BRCA1* mutations, c.5511G > T (p.Trp1837Cys) was most frequent, whereas *BRCA2* mutations were dominated by c.22_23del (p.Arg8Alafs*5). The distribution of *BRCA1* and *BRCA2* P/LP variants that caused change in the amino acid sequences appeared random, with no notable clustering in specific domains (Supplementary Fig. 3). However, 8 of the 66 (12%) *BRCA1* variants—7 missense and 1 frameshift—were located in the BRCT domain. Similarly, 9 of the 56 (16%) *BRCA2* variants—2 nonsense, 4 frameshift, and 3 missense—occurred in the DNA-binding domain. The frequencies of P/LP variant types in *BRCA1* and *BRCA2* are summarised in Table 1; additional data on non-*BRCA1/2* genes appear in Supplementary Table 19.

BRCA1 P/LP mutations were most frequent in triple-negative breast cancer (TNBC), whereas *BRCA2* mutations were predominantly observed in the hormone-receptor-positive/human epidermal growth factor receptor 2-negative (HR+/HER2−) subtype (Supplementary Table 3).

To minimise bias arising from different selection criteria, we re-stratified our Thai cohort according to the recruitment guidelines of both a high-hereditary-risk Caucasian cohort[5] (National Comprehensive Cancer Network [NCCN] v2.2017 guidelines) and a Chinese cohort[6]. A summary of differences between the NCCN v2.2017 and v1.2023 guidelines is provided in Supplementary Table 4.

Based on the NCCN v2.2017 guidelines, 545 of our samples met the inclusion criteria. We then compared the 10 most frequently mutated high- and moderate-penetrance breast cancer genes in these 545 Thai samples with those reported in the Caucasian cohort[5]. The P/LP variant frequencies of *BRCA1/2*, *BRCA1*, and *BRCA2* were higher in our Thai patients than in the Caucasian cohort, whereas the frequency of *PALB2* was lower ($P<0.05$; Fig. 1B, Supplementary Table 5).

For comparison with the Chinese cohort[6], we identified 616 Thai samples that met the selection criteria, which included patients aged ≤ 35 years, those with bilateral breast cancer, males with breast cancer, or a family history of cancer. The frequencies of P/LP variants in *BRCA1/2*, *BRCA2*, and *PALB2* were lower among Thai patients than in the Chinese cohort ($P<0.05$; Fig. 1B, Supplementary Table 6).

Characteristics of somatic alterations in Thai patients with breast cancer

To investigate the somatic mutational profile of Thai breast tumours, we performed targeted sequencing of FFPE tumour tissues using the 501-gene OCA-Plus panel. This approach detects single nucleotide variants (SNVs), small insertions/deletions (indels), and copy number variations (CNVs) across more than 1.4 Mb of coding sequence. FFPE tumour samples were sequenced to a mean depth of 2859×. Among a total of 8187 variants, we identified 657 somatic non-synonymous variants (547 SNVs and 110 indels). The median somatic mutation count was three variants per sample.

The most frequently mutated gene in Thai breast tumours was *TP53*, followed by *PIK3CA*, *ARID1A*, *FAT1*, *KMT2C*, *MAP3K1*, *PTEN*, *GATA3*, *CDH1*, and *FOXA1* (Fig. 2A). Variant allele frequency plots revealed that *TP53*, *GATA3*, *MAP3K1*, *PIK3CA*, and *CDH1* had mean allele frequencies exceeding 30% (30–50%; Supplementary Fig. 4).

Genes showing hotspot mutations, defined here as those with mutation frequencies above 2%, included *PIK3CA* p.H1047R (7%) and p.E545K (4%; Supplementary Fig. 5). To understand the roles of these somatic mutations in cancer development and progression, we mapped the protein domains affected by five commonly mutated genes (Supplementary Fig. 6).

Most *TP53* mutations—primarily missense variants—clustered in the P53 DNA-binding domain, impairing its transcriptional activity. Notably, the gain-of-function mutations R248W and R273H were identified[17]. In *PIK3CA*, the gain-of-function mutation H1047R clustered in the kinase domain, while the gain-of-function mutations E545K and E542K were found in the helical binding domain, all of which promote tumour development[18]. Nearly all *FAT1* mutations, again mostly missense, occurred within cadherin repeat domains. Of these, 7 out of 8 (88%) were predicted by SIFT and PolyPhen to have damaging effects on protein function (Supplementary Table 7).

Pairwise comparisons of mutation frequencies by molecular subtype indicated that tumours with *TP53* mutations were more likely to be HER2-positive (HR-/HER2+; OR = 10.14, $P=0.001$) or TNBC (OR = 9.62, $P<0.0001$). Similarly, *MAP3K1* (OR = Inf, $P=0.007$) and *GATA3* (OR = Inf, $P=0.013$) mutations were more likely to be HR+/HER2− subtype. By contrast, tumours harbouring *PIK3CA* mutations were less likely to fall into the HR + subtype (Fig. 2B, Supplementary Table 8).

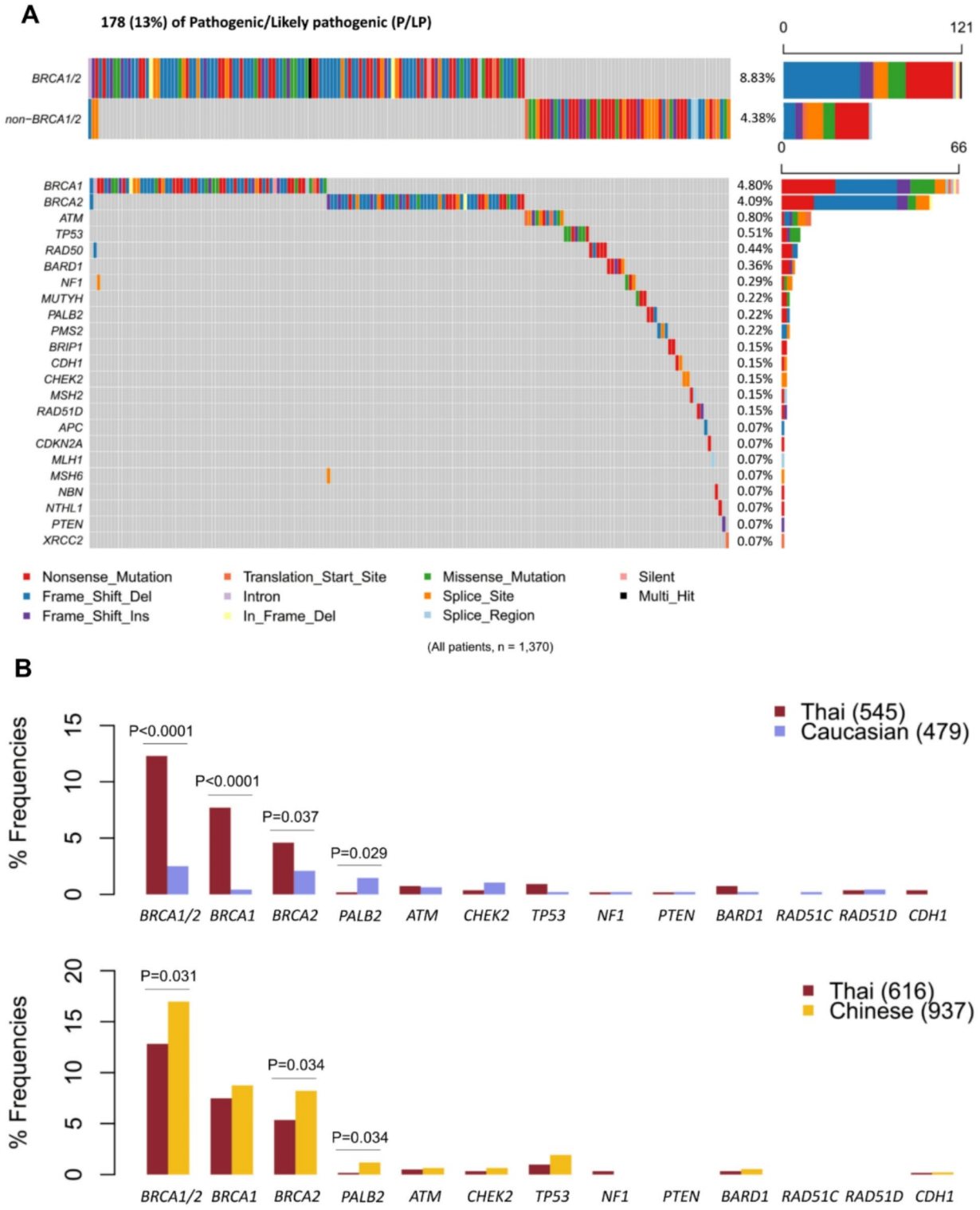

Fig. 1. Germline mutation landscape in Thai breast cancer patients and comparison with other ethnicities. (**A**) Landscape of pathogenic/likely pathogenic variants in 1370 Thai breast cancer patients. (**B**) Comparison of the 10 most frequently mutated genes (high- and moderate-penetrance genes) in Thai patients with those in high-risk Caucasian[5] and Chinese cohorts[6].

After excluding samples that failed quality checks for CNV analysis, 170 breast tumour samples were retained for evaluation. Among these 170 samples, 158 (93%) exhibited altered copy number. The most commonly affected genes were *MCL1*, *PCBP1*, *ERBB2*, *MYC*, and *CCND1* (Fig. 2D). Distinct patterns emerged when stratified by breast cancer subtype. For instance, *ERBB2* amplification was more likely to be HER2-positive

Gene	Frame_ Shift_Del *n* (%)	Frame_ Shift_Ins *n* (%)	In_ Frame_ Del *n* (%)	Intron *n* (%)	Missense *n* (%)	Nonsense *n* (%)	Silent *n* (%)	Splice_ Region *n* (%)	Splice_ Site *n* (%)	Translation_ Start_Site *n* (%)	Total variants	Total samples, *n* (%) (*n* = 1370)
BRCA1	23 (34.8)	5 (7.6)	1 (1.5)	1 (1.5)	9 (13.6)	20 (30.3)	1 (1.5)	1 (1.5)	4 (6.1)	1 (1.5)	66	66 (4.8)
BRCA2	31 (55.4)	4 (7.1)	1 (1.8)	0 (0)	3 (5.4)	12 (21.4)	0 (0)	0 (0)	5 (8.9)	0 (0)	56	56 (4.09)

Table 1. Frequency of pathogenic/likely pathogenic germline variant types in *BRCA1* and *BRCA2*. The table summarises the distribution of pathogenic/likely pathogenic germline variant types identified in *BRCA1* and *BRCA2* in 1370 Thai breast cancer patients.

tumours (OR = Inf, $P < 0.0001$), whereas *CCND1* amplification characterised HR+/HER2 – tumours (OR = 6.95, $P < 0.0001$). In addition, *MPL* amplification was observed in the HER2 subgroup, and *MTAP* deletion was prevalent in the TNBC subtype (Supplementary Fig. 7, Supplementary Table 9).

Co-occurrence and mutual exclusivity of somatically altered genes

We investigated potential associations between somatically altered genes and identified patterns of co-occurrence and mutual exclusivity among several breast cancer driver genes[12,14]. These findings are illustrated in Fig. 2C and Supplementary Table 10. *FLT4* showed the highest incidence of co-occurring events. Significant pairs included *FLT4* and *PTEN* (OR = 66.045, $P < 0.0001$), *FLT4* and *CREBBP* (OR = 38.97, $P = 0.005$), *FLT4* and *FOXA1* (OR = 31.54, $P = 0.008$), *FLT4* and *ZFHX3* (OR = 31.54, $P = 0.008$), and *FLT4* and *KMT2C* (OR = 19.92, $P = 0.02$). We identified four missense variants of *FLT4* (G245E, P422S, R891C, and G914R), all of which were predicted by SIFT and PolyPhen to damage protein function (Supplementary Fig. 11, Supplementary Table 21). In contrast, *TP53* exhibited the highest frequency of mutually exclusive events, including *TP53* and *MAP3K1* (OR = 0, $P = 0.005$), *TP53* and *GATA3* (OR = 0, $P = 0.01$), *TP53* and *SPEN* (OR = 0, $P = 0.04$), and *TP53* and *ARID1A* (OR = 0.22, $P = 0.04$).

The somatic landscape of oncogenic signalling pathways

We investigated the frequent pathway alterations in common oncogenic signalling pathways in breast cancer, including the Cell Cycle, HIPPO, MYC, NOTCH, NRF2, PI3K, RTK-RAS, TGF-β, TP53, and β-catenin/WNT pathways[19]. Using a previously described analytical approach[13], we observed that the TP53 (44%), PI3K (29%), NOTCH (15%), and RTK-RAS (13%) pathways were most frequently altered in Thai breast tumours (Fig. 3A, Supplementary Fig. 8).

Pathway-specific analyses showed that mutations in the TP53 pathway were more common in HER2-positive (OR = 9.63, $P < 0.0001$) and TNBC (OR = 9.02, $P < 0.0001$) subtypes. In addition, the Cell Cycle pathway was altered more often in TNBC tumours (OR = 6.14, $P = 0.008$; Fig. 3B, Supplementary Table 11).

Clinically actionable variants

Personalising cancer therapy can be difficult due to the complexity of tumour genomic data and the need to integrate recommended drugs with treatment guidelines. In this study, we investigated potentially actionable variants in Thai breast tumours and identified corresponding therapeutic options. Using the Oncomine Knowledgebase Reporter software (Thermo Fisher Scientific, Waltham, MA, USA), we classified actionable variants according to the Joint Consensus Recommendation from ASCO, AMP, and CAP (Tier IA to IIC)[20]. We then matched these variants to drugs approved by the US Food and Drug Administration, European Medicines Agency, NCCN, and European Society for Medical Oncology. Overall, 75.56% (136/180) of Thai breast tumours harboured at least one clinically actionable genomic variant. Among these, 50.56% fell into Tier IA (approved for breast cancer), whereas 25% were classified as Tier IIC (approved for non-breast cancer indications; Table 2, Supplementary Fig. 9A–B).

Comparison of the somatic landscape with other ethnic cohorts

To determine whether Thai breast tumours display a distinct biology or whether they align with those of other ethnic groups, we compared the mutation frequencies of commonly mutated breast cancer driver genes in our cohort with data from The Cancer Genome Atlas (TCGA; primarily a US population)[8], METABRIC (mainly Canada-UK)[12], and a Malaysian cohort, where 89% of participants were of Chinese descent[14]. The nine genes analysed—*TP53*, *PIK3CA*, *ARID1A*, *MAP3K1*, *PTEN*, *KMT2C*, *GATA3*, *CDH1*, and *FOXA1*—had been previously identified as putative drivers of breast cancer[12,14].

Previous findings[13,21], supported by our data, suggest that genomic mutations are enriched in different molecular subtypes of breast cancer. Variations in the proportion of each subtype within a cohort can artificially inflate or diminish apparent mutation frequencies across ethnic groups. To reduce this bias, we performed subtype-specific comparisons based on oestrogen receptor (ER) status and molecular classification. Our dataset was stratified into ER+/HER2−, HER2, and TNBC, following the TCGA classification[8]. Additionally, we compared ER-positive versus ER-negative tumours according to the METABRIC[12] and Malaysian[14] cohorts.

Comparison based on molecular subtype or ER status revealed significant differences in *TP53*, *PIK3CA*, *ARID1A*, *GATA3*, *CDH1*, and *FOXA1* mutation frequencies. Notably, *PIK3CA*, *GATA3*, and *CDH1* showed

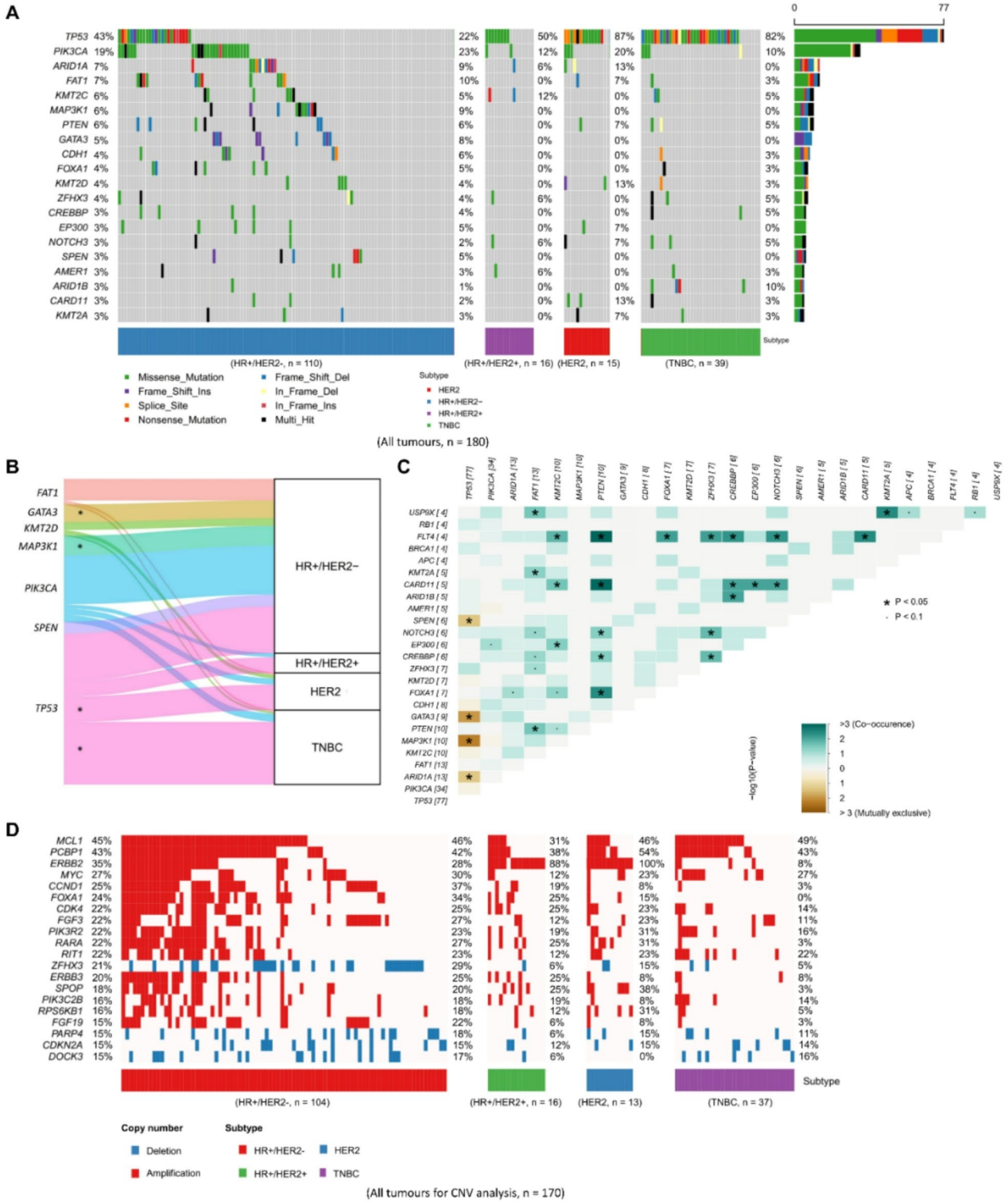

Fig. 2. Somatic landscape and characterisation of Thai breast tumours. (**A**) Somatic mutation landscape of 180 Thai breast tumours, classified by molecular subtype. The first column of percentage frequencies represents the frequency of mutated genes in the overall sample set. (**B**) Clinically relevant subtype-specific mutated genes. Asterisks indicate somatic mutations associated with molecular subtypes. (**C**) Co-occurrence and mutual exclusivity of somatic mutations in 180 Thai breast tumours. Co-occurrence refers to cases in which two genes tend to be mutated together in the same sample. (**D**) Landscape of somatic copy number variations (CNVs) in 170 formalin-fixed, paraffin-embedded (FFPE) breast tumour tissues, classified by molecular subtype.

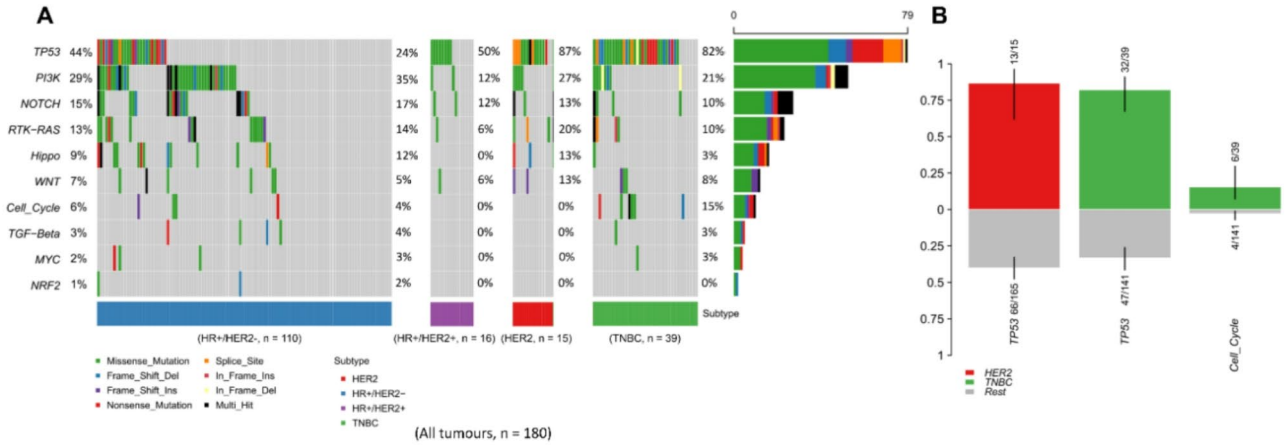

Fig. 3. Somatic alterations in oncogenic signalling pathways. (**A**) Somatic alterations in oncogenic signalling pathways in 180 FFPE breast tumour tissues from Thai breast cancer patients, classified by molecular subtype. The first column of percentage frequencies represents the prevalence of pathway mutations in the overall sample set. (**B**) Significant enrichment of oncogenic pathway mutations in different molecular subtypes ($P < 0.05$; FDR < 0.25). Significant associations with subtype were identified using two-tailed Fisher's exact tests. 'Rest' refers to all other subtypes.

Gene	Variation	Proportion of tumours ($n = 180$) n, (%)		Drugs	Public data sources included in relevant therapies	Other criteria
ERBB2	amplification (CN ≥ 4)	58 (32%)		ado-trastuzumab emtansine[1,2,3,4], irbinitinib + trastuzumab[1], trastuzumab[1,2,3], lapatinib + capecitabine[1,2,3,4], margetuximab[1,4], neratinib[1,2,3,4], pertuzumab + trastuzumab[1,2,3,4], trastuzumab deruxtecan[1,2]	[1]FDA, [2]EMA, [3]NCCN, and [4]ESMO	
PIK3CA	p.E542K, p.E545K, p.H1047R, p.H1047Y	27 (15%)	43 (24%)	alpelisib[1,2,3,4], capivasertib[1]	[1]FDA, [2]EMA, [3]NCCN, and [4]ESMO	*ERBB2* neg, and HR pos
	p.E542Q, p.Q546K			alpelisib[2], capivasertib[1]	[1]FDA and [2]ESMO	*ERBB2* neg, and HR pos
	p.N345K, p.R115L, p.R108H, p.C472_L473insFC, p.R916C, p.G451_L455delinsV, p.R357Q, p.V344G	9 (5%)		capivasertib	FDA	*ERBB2* neg, and HR pos
PTEN	p.P96del, p.A328Qfs*16, p.E43Q, p.M35V, p.D24H, p.T319*, p.L247Cfs*4, p.D301Efs*7, p.Y178H, p.L57S, p.L318Qfs*23, p.D92H	10 (6%)		capivasertib	FDA	*ERBB2* neg, and HR pos
AKT1	p.E17K	1 (1%)		capivasertib	FDA	*ERBB2* neg, and HR pos
ESR1	p.R555C	1 (1%)		elacestrant	FDA and NCCN	*ERBB2* neg, and HR pos

Table 2. Clinically actionable somatic variants (Tier IA). The table presents the identified clinically actionable somatic variants (Tier IA), including the genes affected, mutation types, and the proportion of patients with corresponding actionable mutations. It also lists the approved therapies for each variant, along with public data sources for these therapies. EMA, European Medicines Agency; ESMO, European Society for Medical Oncology; FDA, US Food and Drug Administration; NCCN, National Comprehensive Cancer Network.

lower mutation rates in Thai breast tumour patients ($P < 0.05$; Fig. 4A–B, Supplementary Tables 12 and 13). These differences were most pronounced in HR + tumours.

Among the examined genes, *PIK3CA* mutations were significantly less frequent in tumours with germline mutations than in those without ($P = 0.004$; Fig. 5B, Supplementary Fig. 10). The overall lower *PIK3CA* mutation rate observed in Thai breast tumours, relative to TCGA, METABRIC, and Malaysian cohorts, may partly reflect the high proportion of germline-positive cases (58/180, 32%) in our cohort.

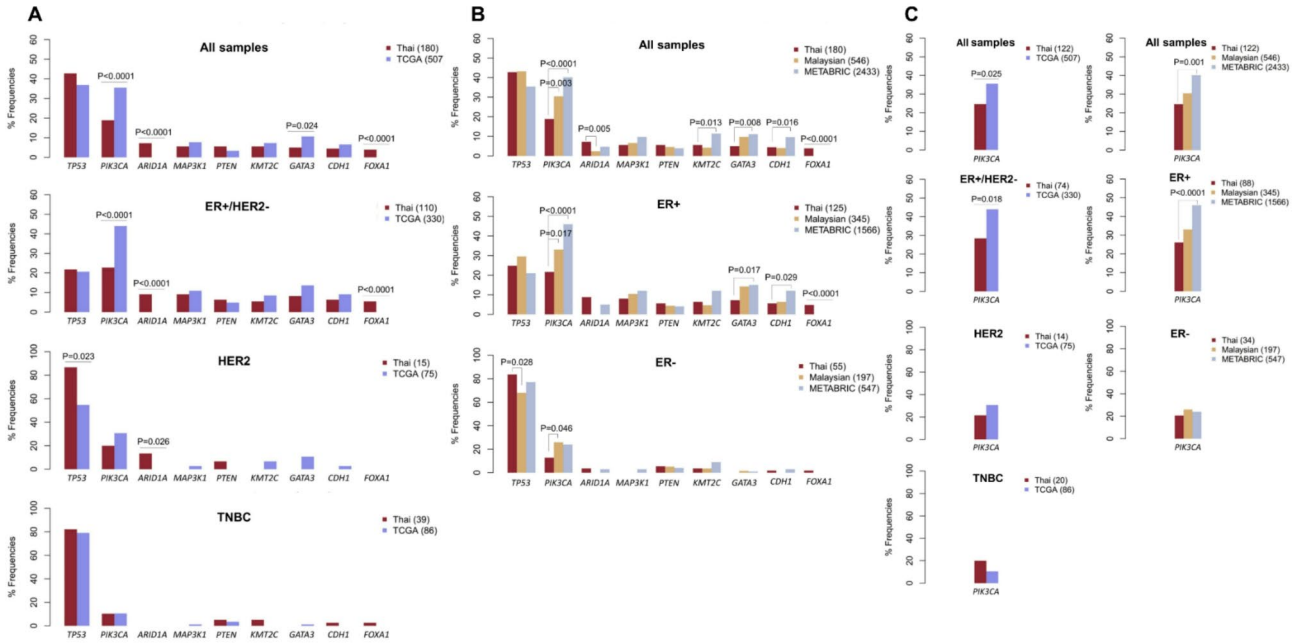

Fig. 4. Population-specific somatic mutations in Thai breast cancer patients compared with other ethnic cohorts. (**A**) Comparison of the prevalence of mutations in major breast cancer driver genes in Thai patients and Caucasian cohorts (TCGA), classified by ER+/HER2−, HER2, and TNBC status[8]. (**B**) Comparison with Caucasian (METABRIC)[12] and Malaysian[14] cohorts, classified by ER + and ER − status. (**C**) Comparison of *PIK3CA* mutation rates in non-germline carriers and other ethnic cohorts. Significance was determined using two-tailed Fisher's exact tests.

To verify these findings, we assessed *PIK3CA* mutations specifically in tumours lacking germline variants. Compared with TCGA and METABRIC cohorts, Thai patients still exhibited fewer *PIK3CA* mutations ($P < 0.05$; Fig. 4C, Supplementary Tables 14 and 15). However, no significant differences emerged when comparing ER + or ER − tumours to the Malaysian cohort (Fig. 4C, Supplementary Table 15).

Clinical impact

We investigated the potential clinical effect of the somatic profile by evaluating its association with overall survival. Patients whose tumours harboured both *TP53* and *PTEN* mutations exhibited worse survival outcomes than those without alterations in these genes (HR = 6.235, 95% CI 0.256–151.7, $P = 0.009$), although the sample size was small ($n = 5$; Fig. 5A).

Differences in the somatic landscape between germline and non-germline carriers

To determine whether germline mutations influence the molecular biology of Thai breast tumours, we compared the frequencies of somatic variants in germline carriers and non-carriers. *PIK3CA* mutations were significantly less common in tumours from germline carriers ($P = 0.004$). By contrast, somatic mutations in *BRCA2* and *HNF1A* were more frequent in germline carriers than in non-carriers ($P = 0.03$; Fig. 5B and C). Given that *BRCA1* and *BRCA2* germline mutations were most prevalent, we further assessed their corresponding tumour somatic profiles. Among tumours carrying germline *BRCA1/2* variants, *TP53* emerged as the most frequently mutated gene (Fig. 5D).

Discussion

The availability of genomic information influences treatment options for patients with breast cancer; however, the somatic profile of Thai breast tumours has not yet been characterised. In this study, we identified both germline and somatic mutational profiles in Thai breast cancer patients at high hereditary risk. According to multiple-gene sequencing panel data, the prevalence of P/LP mutations in high-hereditary-risk breast cancer ranges from 9 to 24% across various ethnicities[5–7,22]. The highest frequencies are reported for *BRCA1/2* (2.5–17%) and for non-*BRCA1/2* (3.9–7.7%)[5–7,22], reflecting differences in selection criteria and in the genes included in each panel. We found that 13% of patients harboured P/LP mutations, which aligns with the 11.8% germline carrier rate reported by the Genomic Thailand Project[23]. This similarity may be explained by the previous study's high proportion of high-risk cases (up to 80.4%). Notably, the frequency of non-*BRCA1/2* mutations was 4.4%, suggesting that multiple-gene panel testing can identify additional hereditary breast cancer patients beyond *BRCA1/2* alone. This finding underscores the clinical utility of panel sequencing in prevention, treatment, and follow-up planning.

Our results are consistent with earlier work showing that variants in *BRCA1* and *BRCA2* are distributed randomly throughout both genes[24]. However, certain mutations cluster in known functional domains, including

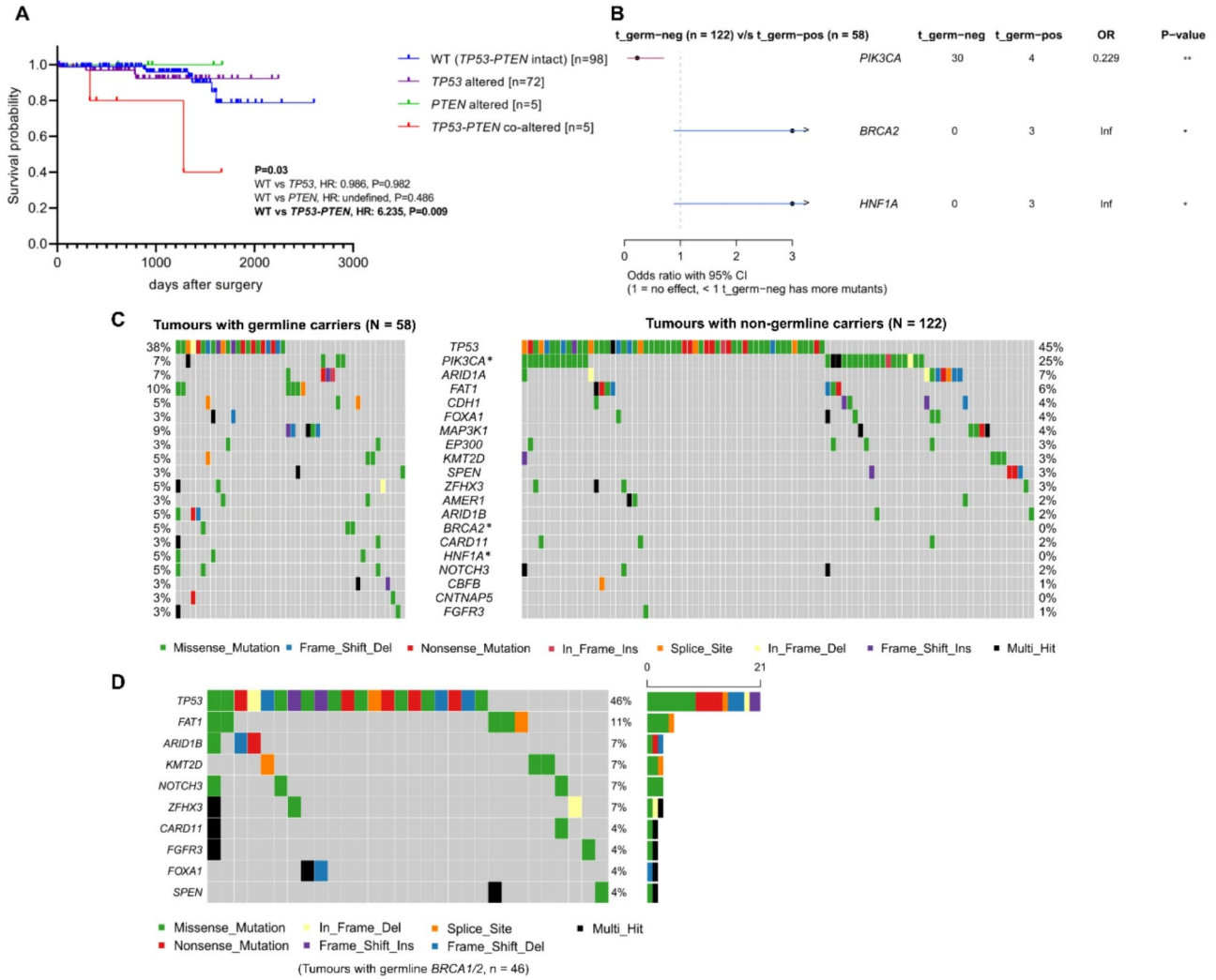

Fig. 5. Clinical impact and germline–somatic mutation interactions in Thai breast tumours. (**A**) Kaplan–Meier survival curve for Thai breast tumours harbouring co-mutations in *TP53* and *PTEN*. (**B**) Comparison of somatic mutation prevalence across all mutated genes in tumours from carriers with and without germline mutations (* $P = 0.03$; ** $P = 0.004$). (**C**) Oncoplot illustrating somatic mutation profiles in tumours with and without germline mutations. Asterisks (*) indicate significant differences in mutation frequency between the two groups. (**D**) Somatic mutation landscape of tumours harbouring germline *BRCA1/2* mutations.

the BRCT domain of *BRCA1* and the DNA-binding domain of *BRCA2*[24,25]. The prevalence of pathogenic *BRCA1* (0.4–8.8%) and *BRCA2* (2.1–8.2%) mutations varies among high-risk breast cancer cohorts in Asian and Caucasian populations[5–7,24]. However, *BRCA1/2* mutations generally occur at higher rates in high-risk Asian populations than in Caucasian populations, with reported prevalence ranging from 2.51%[5] and 9.3%[22] among Caucasians compared with 17% among the Chinese[6]. Our Thai cohort showed an 8.8% prevalence, similar to the 12.1% reported in Koreans[7].These discrepancies likely reflect differences in selection criteria, ethnic backgrounds, and environmental factors. In contrast, the general breast cancer population exhibits no discernible difference in *BRCA1* or *BRCA2* mutation frequencies between Asian and Caucasian groups[14,21]. The higher rate of *BRCA1/2* mutations in high-hereditary-risk Thai breast cancer patients, relative to similar high-risk Caucasian cohorts[5], suggests that hereditary breast cancer may be more common in Thai populations. Indeed, *BRCA1*, *BRCA2*, and *PALB2* encode homologous recombination repair-associated proteins, and mutations in these genes confer homologous recombination deficiency[26,27], thereby increasing an individual's susceptibility to breast cancer[3,4]. Ethnicity-specific patterns of germline *BRCA1*, *BRCA2*, and *PALB2* mutations may thus contribute to inter-ethnic variations in breast cancer risk and to differences in the benefits with PARP inhibitors.

The somatic landscape of the breast tumours examined in this study showed that *TP53* and *PIK3CA* were the most frequently mutated genes, consistent with observations in large breast cancer cohorts[8,11,14–16]. Apart from *FAT1*, the ten most commonly mutated genes in our cohort closely resembled those reported in a Chinese (Hong Kong) cohort[16]. Cancer driver genes play critical roles in cancer initiation and progression, and besides *FAT1*, these genes have been confirmed as breast cancer drivers in previous studies[12,14]. *FAT1* is often mutated in head and neck squamous cell carcinoma (23%)[28]., but it has also been detected in metastatic invasive lobular

carcinoma and metastatic invasive ductal carcinoma breast tumours[29]. Furthermore, *FAT1* mutation correlates with poor outcomes in breast cancer patients treated with CDK4/6 inhibitors by inducing transcription factors that modulate *CDK6* via the HIPPO pathway[30].

TP53, *GATA3*, *MAP3K1*, *PIK3CA*, and *CDH1* had a mean allele frequency of approximately 30–50%, suggesting they are clonally mutated and likely appear during early tumour formation. Nearly all breast tumours (93%) exhibited CNVs, aligning with findings from a previous study[10] and underscoring the role of copy number changes in driving breast cancer. These results indicate that commonly mutated genes may be central to tumorigenesis and progression in Thai breast tumours, highlighting their potential as drug targets.

Different clinical outcomes have been reported among patients with various molecular subtypes of breast cancer[31]. The subtype-specific alterations identified in our study, for both SNVs/indels and CNVs, suggest that genetic changes may drive differences in characteristics and clinical outcomes across subtypes. For instance, patients with the HR + subtype who harbour *MAP3K1* and *GATA3* mutations and receive hormonal therapies generally exhibit a favourable prognosis[32,33]. *PIK3CA* mutations are typically enriched in HR + tumours[8,13,21], with the luminal A subtype demonstrating the highest frequency of approximately 45%[8].

However, in this study, *PIK3CA* was not enriched in HR + tumours. This finding may be explained by the small proportion of luminal A subtype (2.8%) in our cohort (Supplementary Table 2), which contrasts with the reported 28.8% among Thai breast cancer patients[34]. Another contributory factor may be the enrichment of tumours with germline mutations (32%), which likely reduced the observed *PIK3CA* mutation frequency. Consistent with a previous study[21], tumours with germline-positive status exhibited a lower *PIK3CA* mutation rate than those without germline-positive status.

The co-occurrence of mutated driver genes observed in this study may indicate synergistic interactions that promote tumour development. These co-occurrences also suggest potential combination therapeutic targets, whereas mutually exclusive mutations may imply synthetic lethality[35]. Somatic mutations in *FLT4*, including G245E, P422S, R891C, and G914R, exhibited the highest incidence of co-occurring events. *FLT4*, also known as *VEGFR-3*, encodes a receptor for vascular endothelial growth factors that drive angiogenesis and lymphangiogenesis, thereby facilitating cancer cell invasion and metastasis[36]. Mutations in *FLT4* can influence cancer progression and contribute to lymphatic disorders[37,38].

R891C and G914R reside in the tyrosine kinase domain, a phosphorylation site for downstream signalling. These variants may heighten tyrosine kinase activity and disrupt *FLT4* signalling, thereby accelerating lymphatic development and potentially tumour metastasis. G245E arises in the immunoglobulin domain, which mediates ligand binding and receptor dimerisation. This mutation could impair receptor–ligand interactions, thereby affecting *FLT4* activation and lymphatic vessel development. However, the co-occurrence and mutual exclusivity patterns of these mutated driver genes have not yet been reported. Further in vitro and in vivo research is warranted to clarify the functional relevance of these variants.

The differences in mutation frequencies for *TP53* (higher), *ARID1A* (higher), *GATA3* (lower), and *CDH1* (lower) compared with the Caucasian population align with previous reports[13,15]. Breast tumours carrying *TP53*, *PIK3CA*, *ARID1A*, and *FOXA1* mutations are linked to cancer progression and poor prognosis[17,18,39–44]. Thus, the variations in these mutated genes may reflect differences in breast cancer biology and progression across ethnic groups.

However, sequencing coverage and variant-calling pipelines can influence estimates of mutation frequency. In our cohort, the relatively high detection of *ARID1A* and *FOXA1* mutations, which are extremely rare in the TCGA[8] and Malaysian[14] cohorts, may stem from the higher sequencing coverage of our study. This more comprehensive coverage may reveal mutations that lower-coverage WES-TCGA or WES-Malaysian data could miss. Nonetheless, ethnic and environmental factors could also account for these observed differences.

The proportion of breast cancer cases treatable with *ERBB2* amplifications-targeted therapies appears higher in Thai patients (32%) than in Chinese patients (24%). By contrast, *PIK3CA* mutations occur less frequently in Thai patients (15% versus 31%) than in a prior Chinese cohort[45]. These findings may stem from differences in the databases used to define actionable variants and from variations related to ethnicity. The actionable variants uncovered by our somatic profiling underscore the clinical significance of these alterations in Thai breast cancer and facilitate the design of personalised therapies for this patient population.

We explored the potential clinical impact of somatic profiles on overall survival. Several studies have reported that *TP53* and *PTEN* interact and exhibit crosstalk at transcriptional and protein levels, with *PTEN* enhancing *TP53* stability and *TP53* inducing *PTEN* transcription[46]. Our survival analysis showed that patients with *TP53*–*PTEN* co-mutations experienced worse overall survival than those without such co-mutations, in line with a previous study[47]. This outcome suggests that the simultaneous inactivation of *TP53* and *PTEN* through somatic mutations may have a synergistic effect on cancer progression[48,49]. Clinicians should therefore consider both mutated genes when planning treatment. Further studies with larger sample sizes are needed to validate these observations.

We examined the interplay between germline and somatic mutations in breast cancer biology. Of the 58 tumours with germline mutations in our study, 49 harboured mutations in DNA repair-related genes, namely, *BRCA1* ($n = 19$), *BRCA2* ($n = 27$), *BRIP1* ($n = 2$), and *TP53* ($n = 1$). Tumours bearing these germline mutations displayed fewer *PIK3CA* somatic mutations and exhibited more *BRCA2* somatic mutations than non-carrier tumours. This pattern aligns with findings from a Chinese breast cancer cohort[21]. That study reported increased *BRCA2* somatic mutations and decreased *PIK3CA* somatic mutations among all tumours with germline carriers and tumours with germline DNA repair-related gene mutations. These observations suggest potential co-occurrence and mutual exclusivity, as well as differences in tumour biology, between tumours with and without germline carriers.

Previous studies have also shown that germline *BRCA1/2* mutations and *PIK3CA* somatic mutations are mutually exclusive, whereas germline *BRCA1* mutations often co-occur with *TP53* somatic mutations[21,50].

Our results support this finding, as tumours with germline *BRCA1/2* carriers rarely harbour *PIK3CA* somatic mutations yet frequently possess *TP53* somatic mutations, mirroring observations in a Japanese cohort[51]. Moreover, tumours with biallelic inactivation of *TP53* and germline *BRCA1* may undergo loss of the normal chromosome 17. This process can inactivate germline *BRCA1* in carriers with *TP53* mutations, thereby promoting cancer progression[51,52].

This study has several limitations. First, the sample size for tumour sequencing was relatively small, and the follow-up period was short. Second, prior research indicates that early-onset and older patients with breast cancer exhibit distinct somatic mutational differences[53], which should be accounted for when comparing somatic mutation rates across ethnic cohorts. Our somatic profile specifically represents breast tumours with clinical features that prompt germline testing due to high hereditary risk, whereas other ethnic cohorts often encompass more general populations.

Future research should therefore compare our results with findings from other ethnic groups and with a broader Thai population selected through random sampling. Such comparisons would help validate whether the somatic features observed in this cohort are generalisable or primarily unique to high-hereditary-risk breast cancer. This approach may ultimately confirm or refine our conclusions and support the development of more targeted diagnostic and therapeutic strategies for Thai patients.

Conclusion

This is the first study to report both germline and somatic mutational profiles in Thai breast cancer patients. The findings highlight common mutated genes, oncogenic pathways, clinically actionable variants, subtype-specific alterations, and clinical impact. The study also revealed genetic differences between Thai and other ethnic groups, as well as divergent somatic profiles in germline carriers and non-carriers. The identified germline and somatic mutations provide deeper insight into potential therapeutic targets and could guide the development of custom genetic panels in clinical practice. This study underscores the importance of routine genetic sequencing for advancing precision cancer therapies in Thailand.

Materials and methods

Patient sample details

The patients in this study represent a sub-cohort of the Cancer Genomic Thailand Project (January 2016–March 2022, $n = 3170$)[23]. That project, conducted by Siriraj Genomics, investigated germline mutations in a large cohort of Thai breast cancer patients with all cancer types by using next-generation sequencing of blood samples with various panels. In the present study, we included only breast cancer patients who met the NCCN guidelines (Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic: Hereditary Cancer Testing Criteria for Breast Cancer, v1.2023) for high hereditary risk. Each enrolled patient exhibited at least one relevant clinical feature.

The inclusion criteria were as follows:

- diagnosis at ≤ 50 years;
- TNBC at any age;
- bilateral breast cancer;
- lobular breast cancer with a personal or family history of diffuse gastric cancer;
- male breast cancer; or.
- a diagnosis at any age with ≥ 1 close blood relative who had breast cancer at ≤ 50 years, male breast cancer, ovarian cancer, pancreatic cancer, or high-risk/metastatic prostate cancer, or ≥ 3 close blood relatives with breast cancer, or ≥ 2 close blood relatives with either breast cancer or prostate cancer.

Overall, 1676 of the 3170 patients satisfied these guidelines (v1.2023). Their samples were subdivided into germline and somatic cohorts.

For the germline cohort, we retrieved germline data from 1370 of the 1676 high-hereditary-risk patients. Each had undergone blood-sample sequencing using a custom 36-core cancer panel (see Supplementary Table 1). Hence, the germline profile in this study reflects Thai breast cancer patients at high hereditary risk.

For the somatic cohort, we aimed to explore the somatic profile of Thai breast cancer patients and to compare the molecular biology of tumours with and without germline mutations. To accomplish this, we selected 210 archived FFPE tumour tissues from the 1676 high-hereditary-risk cases at Siriraj Hospital for somatic mutation sequencing. Tumours were chosen if they harboured (1) a P/LP germline mutation or (2) no germline mutation. Consequently, the somatic mutational profile in this study reflects Thai breast tumours displaying features indicative of high hereditary risk and enriched in germline-positive cases. All FFPE tumour specimens were derived from primary tumours at diagnosis (biopsy or resection) and had not received neoadjuvant therapy. Each specimen was confirmed by pathologists and then sequenced using the 501-gene OCA-Plus panel. After excluding samples that failed quality checks, 180 FFPE tumours remained for downstream somatic mutational analyses. A flowchart of the study design is provided in Supplementary Fig. 1.

Clinicopathological data and subtype classification

We collected clinicopathological data from Siriraj Hospital. These data included sex, age at diagnosis, family history, tumour site, tumour grading, TNM staging, patient outcome, and oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67 statuses. Pathologists confirmed gene expression status via immunohistochemical analysis and in situ hybridisation, following the 2010 ASCO/CAP guidelines[54]. ER/PR negativity was defined as fewer than 1% positively stained tumour cells by immunohistochemistry. HER2 negativity was indicated by an immunohistochemistry score of 0 or 1 or by a

HER2 amplification ratio below 2.2 on DISH analysis[54]. For Ki67, in line with the St Gallen International Expert Consensus, a value of ≤ 14% positively stained cells was classified as low, and more than 14% was considered high[55].

Based on ER, PR, HER2, and Ki67 statuses, we initially classified the tumours into five molecular subtypes[56]. These were luminal A (ER + and/or PR+, HER2–, Ki67 low), luminal B (ER + and/or PR+, HER2–, Ki67 high), luminal B (HER2) (ER + or PR+, HER2+), HER2 (ER–, PR–, HER2+), and TNBC (ER–, PR–, HER2–). Owing to the small number of luminal A cases, we combined luminal A and luminal B into a single category. We then regrouped the tumours into four subtypes[13]: HR+/HER2– (ER + and/or PR+, HER2–), HR+/HER2+ (ER + and/or PR+, HER2+), HR–/HER2+ (designated as HER2), and TNBC.

Germline sequencing, variant calling, and interpretation

Genomic DNA extraction from peripheral blood, sequencing, and variant interpretation were performed as described in the Cancer Genomics Thailand project[23]. Briefly, genomic DNA from blood samples was sequenced using a custom 36-core cancer panel that encompasses all exons and ± 50 bp of intron–exon boundaries in the target genes (see Supplementary Table 1). Ion Torrent next-generation sequencing technology (Ion S5 XL sequencer; Thermo Fisher Scientific, Waltham, MA, USA) was used for library preparation and sequencing. Raw reads were mapped to the GRCh37 reference genome via Ion Torrent Suite software v5.16.1 (Thermo Fisher Scientific). Germline variants were called and annotated using JSI SeqPilot SeqNext (JSI Medical Systems, New York, NY, USA). The P/LP germline variants were interpreted and classified according to the 2015 ACMG-AMP standards guideline[57] and the ACMG/ClinGen guideline[58], using Golden Helix VS Clinical software. All SNVs and indels were validated by Sanger sequencing, and CNVs were confirmed by digital MLPA.

Somatic sequencing, variant calling, and filtering

Genomic DNA was extracted and purified from FFPE tumour tissue using an in-house protocol with reagents from PerkinElmer Chemagen Technologie GmbH (Baesweiler, Germany). Briefly, FFPE tumour tissues were deparaffinised and lysed at 100 °C in 10% SDS solution and chill-out reagent, followed by protein digestion with Proteinase K. The lysate was then purified using AMPure XP magnetic beads on a semi-automated KingFisher Duo Prime system (Thermo Fisher Scientific, Waltham, MA, USA). The extracted DNA was treated with the NEBNext FFPE DNA Repair Mix to address FFPE-induced DNA damage. We used 20–200 ng of DNA to construct sequencing libraries with the OCA-Plus kit, following the manufacturer's instructions. In the first step, genomic DNA was treated with Uracil DNA Glycosylase to mitigate deamination issues from the FFPE sample. Final sequencing libraries were prepared on the Ion Torrent platform and run on an Ion S5 XL sequencer (Thermo Fisher Scientific).

The raw reads were aligned to the GRCh37 reference genome using Torrent Suite software v5.16.1 (Thermo Fisher Scientific). Samples failing to meet the recommended OCA-Plus quality metrics were excluded. Somatic variant calling and annotation of SNVs, indels, and CNVs were carried out with Ion Reporter software v5.16.0.2 (Thermo Fisher Scientific), using the Oncomine Comprehensive Assay Plus (version w2.1, DNA, single-sample analysis workflow with default parameters). Since there were no matched normal tissues, the default Oncomine Extended v5.16 filter chain was used to filter out potential germline variants, which included a minor allele frequency (MAF) > 10^{-6}, a 5000-exomes global MAF > 10^{-6}, an ExAc global population allele frequency > 10^{-6}, and variants appearing in UCSC common SNPs. We also excluded samples with a deamination score > 60 and a proportion of C:G > T:A mutations > 25%, as these indicated high false-positive rates from FFPE deamination.

VCF files were annotated and converted to MAF files using vcf2maf (Cyriac Kandoth. mskcc/vcf2maf: vcf2maf v1.6.19. [2020]. doi:https://doi.org/10.5281/zenodo.593251) in conjunction with the Ensembl Variant Effect Predictor. We then applied the following filters to remove likely germline variants and artefacts:

- Variants with a variant allele frequency < 5% and read depth < 200× (minimum 10 reads) were discarded, in keeping with the Ion Torrent platform's detection limit.
- Common variants with a population alternate allele frequency ≥ 1% in gnomAD or the 1000 Genomes Project were excluded.
- Variants confirmed as somatic in COSMIC were retained.
- Variants present in dbSNP, and thus likely germline, were removed.
- Germline contamination was excluded by subtracting variants found in each patient's matched blood sample (sequenced with the 36-core cancer panel or other equivalent methods).
- Variants that did not affect protein structure or function, as determined by SIFT and PolyPhen, were excluded.
- Variants detected in five or more samples, but not listed in COSMIC, gnomAD, or the 1000 Genomes Project, were removed.
- All indels in homopolymer stretches of ≥ 4 nucleotides (as visualised in the Integrative Genomics Viewer), or within simple repeats or microsatellite tracts in the UCSC browser, were excluded.
- Only non-silent variants (including those in exons or splice sites) were retained for downstream analyses.

Each somatic variant that passed this pipeline underwent manual review to ensure that no false positives were included.

For somatic CNV filtering, Ion Reporter software was used with default settings. We selected only samples meeting the recommended median absolute pairwise difference (MAPD) threshold of < 0.5, which indicates low noise in copy number data. Ten samples exceeding this threshold were excluded, leaving 170 for CNV analysis. In line with the Oncomine Comprehensive Assay Plus User Guide (Revision C.0), amplification was defined as a gene copy number ≥ 4 with a 95% CI > 4, whereas deletion was defined as a copy number ≤ 1 with a 95% CI < 1.

We retained only CNVs in genomic regions known as gains or losses of function according to the Oncomine panel (OncomineGeneClass parameter).

Identification of clinically actionable variants

We used Oncomine Knowledgebase Reporter software (v5.6) with data version 2023.05(006) (Thermo Fisher Scientific, Waltham, MA, USA) to identify clinically actionable variants. The software classifies variants based on evidence-level Tier IA to IIC, following standards and guidelines for somatic variant interpretation and reporting in cancer, as recommended by AMP, ASCO, and CAP[20]. It also provides the corresponding drug approval statuses from the US Food and Drug Administration, European Medicines Agency, NCCN, and European Society for Medical Oncology. For this study, an *ERBB2* copy number amplification threshold of ≥ 4 was applied.

In 2023, the US Food and Drug Administration approved capivasertib for breast cancers carrying at least one *PIK3CA*, *PTEN*, or *AKT1* mutation in the tumour tissue[59]. Accordingly, we manually reviewed the actionable variants in those three genes to assess their relevance for capivasertib therapy.

Genomic and statistical analysis

We managed BAM and VCF files using *samtools*, *bcftools*, and Visual Studio Code. Analyses and data visualisations were performed in R statistical software (v4.3.1) using the *maftools* package[60] and GraphPad Prism (v8). Statistical tests were performed with R and PASW Statistics (v18). Differences in gene alteration frequencies between Thai patients and other cohorts were assessed by Pearson's chi-squared test or Fisher's exact test. Several functions in the *maftools* package facilitated data interpretation.

We used *lollipopPlot* for visualising the somatic position distribution at the protein level and *plot* for basic graphics. We employed *somaticInteractions* to identify mutually exclusive or co-occurring events and *clinicalEnrichment* to investigate enrichment patterns. We applied *mafCompare* to compare somatically mutated genes in tumours with germline-positive status versus those without. For survival analysis, we used *mafSurvival*, *survGroup*, and *mafSurvGroup* to generate Kaplan–Meier curves. Overall survival was measured from the date of surgery to the time of death or the last follow-up without death (where patients were assumed to be alive).

Statistical significance was defined as a two-tailed P value < 0.05. For somatic interaction analyses, a false discovery rate < 0.25 was considered significant.

Ethics approval

All subjects and study protocols were approved by the Ethics Committee of the Siriraj Hospital Institutional Review Board (protocol number 418/2562, approval number Si 631/2019; and protocol number 140/2566, approval number Si 230/2023). Each participant provided written informed consent. All samples were collected with approval from the Ethics Committee and with patients' consent. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Siriraj Hospital Institutional Review Board for research involving human participants.

Data availability

The datasets generated during this study are available in the Genome Variation Map (GVM) repository under BioProject ID PRJCA033163, accession number GVM000917.

Received: 3 December 2024; Accepted: 24 March 2025
Published online: 03 April 2025

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Huang, J. et al. Global incidence and mortality of breast cancer: a trend analysis. *Aging (Albany NY)*. **13**, 5748–5803 (2021).
3. Hu, C. et al. A population-based study of genes previously implicated in breast cancer. *N Engl. J. Med.* **384**, 440–451 (2021).
4. Couch, F. J. et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncol.* **3**, 1190–1196 (2017).
5. Beitsch, P. D. et al. Underdiagnosis of hereditary breast cancer: Are genetic testing guidelines a tool or an obstacle?? *J. Clin. Oncol.* **37**, 453–460 (2019).
6. Li, J. Y. et al. Germline mutations in 40 cancer susceptibility genes among Chinese patients with high hereditary risk breast cancer. *Int. J. Cancer*. **144**, 281–289 (2019).
7. Shin, H. C. et al. Detection of germline mutations in breast cancer patients with clinical features of hereditary cancer syndrome using a Multi-Gene panel test. *Cancer Res. Treat.* **52**, 697–713 (2020).
8. Koboldt, D. C. et al. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
9. Curtis, C. et al. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
10. Ciriello, G. et al. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
11. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
12. Pereira, B. et al. The somatic mutation profiles of 2433 breast cancers refine their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).
13. Lang, G. T. et al. Characterization of the genomic landscape and actionable mutations in Chinese breast cancers by clinical sequencing. *Nat. Commun.* **11**, 5679 (2020).
14. Pan, J. W. et al. The molecular landscape of Asian breast cancers reveals clinically relevant population-specific differences. *Nat. Commun.* **11**, 6433 (2020).
15. Kan, Z. et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat. Commun.* **9**, 1725 (2018).
16. Zhu, B. et al. Comparison of somatic mutation landscapes in Chinese versus European breast cancer patients. *HGG Adv.* **3**, 100076 (2022).
17. Bargonetti, J. & Prives, C. Gain-of-function mutant p53: history and speculation. *J. Mol. Cell. Biol.* **11**, 605–609 (2019).

18. Zardavas, D., Phillips, W. A. & Loi, S. PIK3CA mutations in breast cancer: reconciling findings from preclinical and clinical data. *Breast Cancer Res.* **16**, 201 (2014).
19. Sanchez-Vega, F. et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell* **173**, 321–337e310 (2018).
20. Li, M. M. et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, American society of clinical oncology, and college of American pathologists. *J. Mol. Diagn.* **19**, 4–23 (2017).
21. Chen, C. et al. Comprehensive genomic profiling of breast cancers characterizes germline-somatic mutation interactions mediating therapeutic vulnerabilities. *Cell. Discov.* **9**, 125 (2023).
22. Tung, N. et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer* **121**, 25–33 (2015).
23. Kansuttiviwat, C. et al. Germline mutations of 4567 patients with hereditary breast-ovarian cancer spectrum in Thailand. *NPJ Genom Med.* **9**, 9 (2024).
24. Shao, D. et al. Prevalence of hereditary breast and ovarian cancer (HBOC) predisposition gene mutations among 882 HBOC high-risk Chinese individuals. *Cancer Sci.* **111**, 647–657 (2020).
25. Heramb, C. et al. BRCA1 and BRCA2 mutation spectrum - an update on mutation distribution in a large cancer genetics clinic in Norway. *Hered Cancer Clin. Pract.* **16**, 3 (2018).
26. den Brok, W. D. et al. Homologous recombination deficiency in breast cancer: A clinical review. *JCO Precis Oncol.* **1**, 1–13 (2017).
27. Stover, E. H., Fuh, K., Konstantinopoulos, P. A., Matulonis, U. A. & Liu, J. F. Clinical assays for assessment of homologous recombination DNA repair deficiency. *Gynecol. Oncol.* **159**, 887–898 (2020).
28. Lawrence, M. S. et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
29. Pareja, F. et al. The genomic landscape of metastatic histologic special types of invasive breast cancer. *NPJ Breast Cancer.* **6**, 53 (2020).
30. Li, Z. et al. Loss of the FAT1 tumor suppressor promotes resistance to CDK4/6 inhibitors via the Hippo pathway. *Cancer Cell.* **34**, 893–905e898 (2018).
31. Yang, S. X. & Polley, E. C. Systemic treatment and radiotherapy, breast cancer subtypes, and survival after long-term clinical follow-up. *Breast Cancer Res. Treat.* **175**, 287–295 (2019).
32. Ellis, M. J. et al. Whole-genome analysis informs breast cancer response to aromatase Inhibition. *Nature* **486**, 353–360 (2012).
33. Griffith, O. L. et al. Author correction: The prognostic effects of somatic mutations in ER-positive breast cancer. *Nat. Commun.* **9**, 4850 (2018).
34. Ditsatham, C. et al. Breast cancer subtypes in Northern Thailand and barriers to satisfactory survival outcomes. *BMC cancer.* **22**, 1147 (2022).
35. El Tekle, G. et al. Co-occurrence and mutual exclusivity: What cross-cancer mutation patterns can tell Us. *Trends Cancer.* **7**, 823–836 (2021).
36. Chen, J. C., Chang, Y. W., Hong, C. C., Yu, Y. H. & Su, J. L. The role of the VEGF-C/VEGFRs axis in tumor progression and therapy. *Int. J. Mol. Sci.* **14**, 88–107 (2013).
37. Liu, N. & Gao, M. FLT4 mutations are associated with segmental lymphatic dysfunction and initial lymphatic aplasia in patients with Milroy disease. *Genes (Basel).* **12**, 1611 (2021).
38. Gu, Y. et al. Case report: unique FLT4 variants associated with differential response to anlotinib in angiosarcoma. *Front. Oncol.* **12**, (2022).
39. Marvalim, C., Datta, A. & Lee, S. C. Role of p53 in breast cancer progression: An insight into p53 targeted therapy. *Theranostics* **13**, 1421–1442 (2023).
40. Zhang, Z., Hao, R., Guo, Q., Zhang, S. & Wang, X. TP53 mutation infers a poor prognosis and is correlated to immunocytes infiltration in breast cancer. *Front. Cell. Dev. Biol.* **9**, 759154 (2021).
41. Reinhardt, K. et al. PIK3CA-mutations in breast cancer. *Breast Cancer Res. Treat.* **196**, 483–493 (2022).
42. Arruabarrena-Aristorena, A. et al. FOXA1 mutations reveal distinct chromatin profiles and influence therapeutic response in breast cancer. *Cancer Cell.* **38**, 534–550e539 (2020).
43. Liu, N. et al. FOXA1 and FOXA2: The regulatory mechanisms and therapeutic implications in cancer. *Cell. Death Discov.* **10**, 172 (2024).
44. Xu, G. et al. ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer. *Nat. Genet.* **52**, 198–207 (2020).
45. Wu, L. et al. Landscape of somatic alterations in large-scale solid tumors from an Asian population. *Nat. Commun.* **13**, 4264 (2022).
46. Nakanishi, A., Kitagishi, Y., Ogura, Y. & Matsuda, S. The tumor suppressor PTEN interacts with p53 in hereditary cancer (Review). *Int. J. Oncol.* **44**, 1813–1819 (2014).
47. Tanic, N. et al. The impact of TP53 and PTEN tumor suppressor genes on response to different breast cancer treatment modalities. *Биомедицинска Истраживања.* **13**, 105–117 (2022).
48. Gasparyan, M., Lo, M. C., Jiang, H., Lin, C. C. & Sun, D. Combined p53- and PTEN-deficiency activates expression of mesenchyme homeobox 1 (MEOX1) required for growth of triple-negative breast cancer. *J. Biol. Chem.* **295**, 12188–12202 (2020).
49. Kim, G. et al. SOCS3-mediated regulation of inflammatory cytokines in PTEN and p53 inactivated triple negative breast cancer model. *Oncogene* **34**, 671–680 (2015).
50. Park, S. et al. Clinical characteristics and exploratory genomic analyses of germline BRCA1 or BRCA2 mutations in breast cancer. *Mol. Cancer Res.* **18**, 1315–1325 (2020).
51. Inagaki-Kawata, Y. et al. Genetic and clinical landscape of breast cancers with germline BRCA1/2 variants. *Commun. Biol.* **3**, 578 (2020).
52. Martins, F. C. et al. Evolutionary pathways in BRCA1-associated breast tumors. *Cancer Discov.* **2**, 503–511 (2012).
53. Waks, A. G. et al. Somatic and germline genomic alterations in very young women with breast cancer. *Clin. Cancer Res.* **28**, 2339–2348 (2022).
54. Hammond, M. E. et al. American society of clinical oncology/college of American pathologists guideline recommendations for immunohistochemical testing of Estrogen and progesterone receptors in breast cancer. *Arch. Pathol. Lab. Med.* **134**, 907–922 (2010).
55. Goldhirsch, A. et al. Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann. Oncol.* **22**, 1736–1747 (2011).
56. Tang, P. & Tse, G. M. Immunohistochemical surrogates for molecular classification of breast carcinoma: A 2015 update. *Arch. Pathol. Lab. Med.* **140**, 806–814 (2016).
57. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* **17**, 405–424 (2015).
58. Riggs, E. R. et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: A joint consensus recommendation of the American college of medical genetics and genomics (ACMG) and the clinical genome resource (ClinGen). *Genet. Med.* **22**, 245–257 (2020).
59. Turner, N. C. et al. Capivasertib in hormone Receptor–Positive advanced breast cancer. *N Engl. J. Med.* **388**, 2058–2070 (2023).
60. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).

Acknowledgements
This work was conducted under the Cancer Genomics Thailand project and was partly funded by the Health Systems Research Institute (HSRI)–Genomics Thailand Initiative Grant awarded to MP and grant number 64-143 to BS. Additional support was provided by the Siriraj Core Research Facility (SiCRF) and Strategic Project Grant to MP, the Siriraj Chalermphrakiat Grant to MP and BS, and the Thanapat Fund (D003752) to MP. Open access funding was provided by Mahidol University. We thank Assistant Professor Norasate Samarnthai, pathologist, for his assistance in confirming FFPE tumour tissue.

Author contributions
KK, BS, and MP designed the experiments and the overall study. KK, ER, and JS optimised the protocols. ER, PN, PD, CM, SW, WTan, KP, KPong, WTit, CL, JW, JS, and NR carried out blood sample pre-sequencing and next-generation germline sequencing, followed by confirmation via Sanger sequencing and MLPA. KK performed pre-sequencing of formalin-fixed, paraffin-embedded tumour tissue and next-generation somatic sequencing. KK, NR, and MP gathered and verified the clinicopathological data. KK, BS, WT, and MP analysed and interpreted the genomic data. KK and BS drafted the manuscript. BS, MP, WT, and KKor supervised the study design and analysis pipeline and provided critical feedback. All authors read and approved the final manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95834-w.

Correspondence and requests for materials should be addressed to M.P. or B.S.