

Original Article

Clinically actionable alterations in Indian breast cancer patients derived through whole transcriptome sequencing

Nilesh Gardi,^{1,2,#} Rohan Chaubal,^{1,2,3,#} Khushboo A. Gandhi,^{1,2} Anushree Kadam,^{1,2} Ankita Singh,^{1,2} Aishwarya Sabari Raja,^{1,2,4} Vaibhav Vanmali,³ Rohini Hawaldar,³ Suhani Sale,^{1,2} Shalaka Joshi,^{2,3} Rajendra Badwe^{2,3} & Sudeep Gupta^{1,2,4}

¹Clinician Scientist Laboratory, Advanced Centre for Treatment, Research, and Education in Cancer, Tata Memorial Centre, ²Training School Complex, Homi Bhabha National Institute, Departments of ³Surgical Oncology, and ⁴Medical Oncology, Tata Memorial Hospital, Tata Memorial Centre, Mumbai, Maharashtra, India

Received October 16, 2024; Accepted May 23, 2025; Published February 28, 2026

Background and objectives: Genomic studies are essential for identifying mutations that may influence key aspects of breast tumours, such as susceptibility, aggressiveness, and response to treatment. There are deficient molecular and genomic data from Indian breast cancer patients.

Methods: mRNA from primary breast cancer samples were subjected to next-generation transcriptome (mRNA) sequencing on an Illumina platform, in duplicates and triplicates to generate 30–60 M reads/sample. PAM50, and absolute intrinsic molecular subtyping (AIMS) gene expression-based classifiers were used for intrinsic subtyping. Variants were called using, GATK, MuTect2, VarScan2, and VarDict, followed by filtering for somatic and non-synonymous changes. Germline variants were excluded using public databases. ClinVar annotations prioritised pathogenic variants, and the STRING algorithm was used for network analysis.

Results: A total of 207 RNA-Seq datasets from 97 breast cancer patients were analysed. There was good concordance between the immunohistochemical receptor and AIMS classification for all subtypes, but there was discordance between immunohistochemical and PAM50 subtypes within the ER-positive/HER2-positive subgroup, wherein only 38.5% (n= 5) were classified as HER2-like by gene expression classification. Variant analysis identified 145 high-confidence somatic mutations, with *TP53* (n=46, 47%) and *PIK3CA* (n=33, 34%) being the most frequent. Additional actionable mutations in *BRCA1*, *BRCA2*, *FGFR2*, *PTEN*, *AKT1*, and *mTOR* pathways were identified. At least one actionable mutation was found in 52% of patients. Fusion transcript analysis identified 91 recurrent fusions, including novel partners with *ERBB2*, *MED1*, and *CDK12*, suggesting the possibility of unique molecular events.

Interpretation and conclusions: This study demonstrates that Indian breast cancer patients exhibit molecular subtypes and actionable mutations comparable to Caucasian cohorts.

Keywords Breast cancer; Intrinsic subtype; PIK3CA; TP53; Transcriptome

Identifying and characterising somatic mutations and gene fusions have provided important insights into intricate molecular mechanisms underlying breast cancer. These genetic alterations can lead to aberrant signalling pathways, disrupted cellular processes, and dysregulation of critical genes involved in tumour suppression or promotion. Understanding the underlying genetic aberrations in this disease has

been shown to provide important input into patients' prognoses and suggest treatment strategies.¹

There is a paucity of genome-wide molecular characterisation in Indian breast cancer patients. However, a few molecular studies have been reported. Thakkar *et al*¹ identified 108 differentially expressed genes (DEG) in 31 ER-positive breast tumours, implicating mRNA transcription and cellular

[#]Equal contribution

How to cite this article: Gardi N, Chaubal R, Gandhi KA, Kadam A, Singh A, Raja AS, *et al*. Clinically actionable alterations in Indian breast cancer patients derived through whole transcriptome sequencing. *Indian J Med Res*. 2026;163:95-103. DOI: 10.25259/IJMR_1642_2024.

differentiation pathways. In another study, microarray profiling of 29 breast tumours revealed 2,413 DEGs with perturbed cell-cycle, extracellular matrix (ECM), and lipid-metabolism pathways, while PAM50 confirmed canonical subtypes.² A study involving targeted sequencing of 56 genes in 275 breast tumours found somatic variants in 71% of cases, predominated by *TP53* and *PIK3CA* alterations, with 46% actionable, implicating PI3K/AKT/PTEN pathway activation and *PIK3CA*-driven trastuzumab resistance.³ Another study using transcriptomic analysis revealed subtype-specific mRNA and lncRNA signatures, identifying a combined 25 mRNA-27 lncRNA panel that segregated subtypes and showed potential prognostic value.⁴ However, there have been only a few studies evaluating a genome-wide sweep of mutations in Indian breast cancer patients.

Our group has recently reported the clonal evolution of Triple Negative Breast Cancer (TNBC) using multi-omic analysis of tumour samples biopsied longitudinally during a patient's life history.⁵ We also recently reported the therapeutic implications of a three-gene signature identified through whole-genome sequencing of endocrine therapy-sensitive and resistant breast cancer samples.⁶ We have also previously reported immunohistochemical characterisation and outcomes in Indian cohorts of breast cancer.^{7,8,9} Additionally, we have also reported the transcriptomic changes occurring in breast tumours after progesterone administration,¹⁰ and following to surgical resection.¹¹ Our previous studies generated large-scale RNA-Seq data wherein we reported gene-expression changes in context-dependent experiments.^{5,10,11}

In this analysis, we subjected this RNA-Seq data to a robust bioinformatics analysis to identify mutations in key genes reported as actionable alterations by others.¹² Actionable genomic alterations comprise those in which a potential drug treatment is available either as an approved therapy or within clinical trials. We performed a comprehensive investigation of mutations and fusion events in our RNA-Seq data comprising breast cancer tumours from Indian patients with the aim of cataloguing their frequency in our population. Additionally, we investigated molecular subtypes based on transcriptomic data and compared them with the Caucasian population.

Methods

This study was conducted at the department of surgical oncology, medical oncology and pathology at the Tata Memorial Hospital (TMH), Mumbai and the Advanced Centre for Treatment, Research and Education in

Cancer (ACTREC), Tata Memorial Centre, Navi Mumbai, India between 2013-2022.

The analysis was conducted in accordance with the Declaration of Helsinki. Patients were recruited after obtaining informed consent prior to the start of the study. The study also received clearance from the Institutional Ethics Committee of TMH and ACTREC and was also registered with the Clinical Trials Registry of India (CTRI/2017/11/010553, CTRI/2016/11/007430, CTRI/2017/11/010553) and National ClinicalTrials.gov (NCT03797482).

Study design, sample biobanking, and next generation sequencing: Patient recruitment, sample biobanking, RNA extraction, and whole transcriptome sequencing (RNA-Seq) were carried out at the Clinician Scientist laboratory, ACTREC, as previously described.^{5,10,11}

RNA-Seq was conducted to generate 30-60M paired-end reads per sample, as described earlier.^{5,10,11} RNA-Seq data were re-analysed to identify genomic (DNA) alterations and classify the tumours using gene expression-based classifiers. This RNA-Seq data was mined for genomic alterations using a bioinformatics pipeline specific for variant calling from transcriptomic data. When RNA-Seq data from multiple samples from the same patient tumour were available, the analysis was independently performed in all samples available for that patient, treating each sample as a replicate. Accordingly, the analysis used triplicate samples for some patients, duplicate samples for others, and single samples for the remaining patients.

Bioinformatics analysis:

Transcriptomic characterization for PAM50 subtypes: Molecular subtyping was performed using the PAM50 intrinsic gene signature and the Absolute Intrinsic Molecular Subtyping (AIMS) algorithm¹³ with the default settings. Transcript-level quantification was performed using Salmon¹⁴ (version 0.8.1, RRID: SCR_017036) on the RNA-Seq data (FASTQ files) with default settings. A transcriptome index was built using the University of California Santa Cruz (UCSC) *Homo sapiens* reference genome (build hg19) GTF file followed by transcript quantification. Transcript-abundance files were imported from Salmon and converted to gene-level information using the tximport (v1.0.3) R Bioconductor package.¹⁵

Alignment and variant calling pre-processing: The paired-end raw data available in fastq format were aligned to the hg38 reference genome using the STAR aligner.¹⁶ A STAR index reference genome was created

using the Gencode (Version 34) GTF file. BAM files were further processed using the Picard tool (v.2.10.0) (<https://broadinstitute.github.io/picard/>) for sorting and duplicate removal steps. SplitNCigarReads, BaseRecalibrator and ApplyBQSR utilities from the GATK¹⁷ bundle (Version 4.1.2.0) were used for post-processing of the data.

Variant calling and filtering: Variants were detected in the genomic regions corresponding to the gene bodies of known cancer genes.¹² Three state-of-the-art variant callers, MuTect2,¹⁸ VarScan2,¹⁹ and VarDict²⁰ were used to call variants in tumour-only mode. Variants were considered true positive if they were annotated as “PASS”, were present in at least five sequencing read pairs, and were detected by at least two variant callers. These putative true positive variants were functionally annotated using ANNOVAR.²¹ Variants annotated as indels were discarded due to documented problems with misalignment on account of alternate-splicing and splice-site read-through sequence reads. Variants annotated as synonymous genetic change, *i.e.*, a genetic change leading to the same amino acid with no change in protein coding, were also excluded from further study. We thus restricted our analysis to variants that lead to a change in amino acids and subsequently in the protein, which were annotated as non-synonymous, stop-gain, or splice-site.

Germline filtering: Further filtering was carried out based on prior reporting of variants in the dbSNPv156 (database of single-nucleotide polymorphisms)²² and COSMIC (Catalogue of Somatic Mutations in Cancer) database (release 98).²³⁻²⁵ dbSNP v151 onwards has included TMC-SNPdb2.0 (https://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=TMC_SNPDB2), a database of Indian ethnicity-specific variants from normal tissues which is maintained and curated at Tata Memorial Centre. Of note, TMC-SNPdb 2.0 (<https://academic.oup.com/database/article/doi/10.1093/database/baac029/6583650>) also includes variants from the indigen project at Institute of Genomics and Integrative Biology (IGIB, <https://indigen.igib.in/>).²⁴ Variants with a dbSNP^{26,27} id were considered as germline variants and excluded from further analysis, while those with a COSMIC²³ and/or ICGC²⁸ (International Cancer Genome Consortium) id were considered putative somatic variants and carried forward. Any variant with a dbSNP id and an ICGC and/or COSMIC id, was considered putative somatic and carried forward for further analysis. Additionally, germline status for each variant was inferred using the flags ExAC_all, ExAC_SAS, ExAC_nontcga_all, ExAC_nontcga_SAS, gnomAD_exome_ALL,

andgnomAD_exome_SAS. If the variant population prevalence in any of the above flags was more than 1%, it was classified as a germline variant and excluded from further analysis. Based on the above filters, we restricted our analysis to only putative somatic variants.

Variant prioritisation: Candidate somatic variants were filtered by ClinVar annotation flags: variants labelled ‘Pathogenic,’ ‘Likely pathogenic,’ ‘Conflicting interpretations of pathogenicity,’ or ‘Uncertain significance’ were retained for downstream analyses, whereas those marked ‘Benign’ or ‘Likely benign’ were discarded.

Network analysis for mutations: Mutations present in COSMIC or ICGC leading to a protein change and predicted to be deleterious were selected for this analysis. Genes harbouring these mutations were identified and grouped according to the molecular subtype of the underlying samples. These genes were then subjected to a network pathways analysis using STRINGdb Version 2.22.0db²⁹ with *P* values set at a stringency of 0.005 and a False Discovery Rate (Benjamini-Hochberg) of 0.05. The list of proteins for each subtype were searched against Homo Sapiens Validated STRING networks.

Fusion transcript analysis: RNA-Seq data was aligned to the hg38 reference genome to identify potential fusion events using Star-fusion³⁰ with default settings and analysed as described earlier.³¹ Fusion events supported by at least two sequence reads spanning the fusion breakpoint were considered for further analysis.

qPCR validation: DNA was extracted from FFPE blocks using Qiagen (Hilden, Germany), QIAamp DNA FFPE Tissue Kit for DNA extraction as per the manufacturer's protocol. The extracted DNA was subjected to integrity analysis on the TapeStation using HSDNA assays. DNA samples that satisfied the DNA Integrity Number (DIN) quality threshold were analysed by qPCR on an ARIAMx system with the Easy PGx Ready PIK3CA kit (Diatech Pharmacogenetics), following the manufacturer's instructions.

Data availability statement: The human sequence data generated in this analysis are not publicly available due to patient privacy requirements but are available upon reasonable request to the corresponding author. Other data generated in this study are available within the article and its supplementary data files.

Results

Patient population: Across the pooled studies, 97 patients contributed tumour samples with whole-

Table I. Clinical characteristics

Characteristic	ER+ and/or PR+ and HER2-neg (n=46)	Triple Negative TNBC (n=24)	ER+ and/or PR+ and HER2-pos (n=13)	ER- and/or PR- and HER2-pos (n=8)	HER2 Equivocal (n=6)	Total (N=97)
Median age in yr (Range)	56 (27-81)	54 (30-75)	50 (37-67)	59.5(45-71)	57 (50-68)	55 (27-81)
Nodes (%)	Positive	24 (52.17)	11 (45.83)	8 (61.53)	3 (50)	49 (50.51)
	Negative	22 (47.82%)	13 (54.16)	5 (38.46)	5 (62.5)	48 (49.48)
Grade (%)	I	1 (2.17)	0 (0)	0 (0)	0 (0)	1 (1.03)
	II	12 (26.08)	0 (0)	1 (7.69%)	0 (0)	14 (14.43)
	III	33 (71.73)	24 (100)	12 (92.30)	8 (100)	82 (84.53)

transcriptome (RNA-seq) data and were therefore included. RNA-seq data were available from three samples (in triplicate) for 34 patients, from two samples (in duplicate) for 42 patients, and as a single sample for 21 patients, yielding 207 tumour RNA-seq datasets for analysis. Patients were stratified based on oestrogen receptor (ER), progesterone receptor (PR), and HER2 status (**Table I**).

Molecular subtyping by gene expression analysis reveals canonical intrinsic breast cancer subtypes in Indian patients: We identified a median of 11,650 (range 100-33,100) transcripts across 207 samples from 97 patients. Transcript distribution across paired samples revealed a median of 11,300 (range 100–28,600) transcripts in at least three samples, 11,100 (range 100-33,100) transcripts in at least two samples, and 11,900 (range 1,100-21,400) transcripts in at least 1 sample. Transcript distribution and overlap have been shown in **Supplementary Figure 1**. The 97 patients were classified according to both AIMS and PAM50 intrinsic subtyping classifiers. AIMS identified five subtypes (Luminal A, n=27; Basal-like, n=25; HER2-enriched, n=23; Luminal B, n=20; Normal-like, n=2; **Supplementary Table I**), whereas PAM50 resulted in four subtypes (Luminal B, n=36; Basal-like, n=29; Luminal A, n=19; HER2-enriched, n=13; **Fig. 1, Table II**). A detailed comparison of these subtype distributions and their concordance is presented in **Supplementary Table II**. This comparative analysis (**Supplementary Table II**) reveals significant discordance in the classification of Luminal A and Luminal B subtypes between the two methodologies. In contrast, Basal-like and HER2-enriched tumours demonstrated high concordance across both classifiers. Patients classified as Luminal A by the PAM50 gene signature were found to be distributed among several distinct AIMS subtypes.

Molecular architecture: Following stringent filtering, which mandated a minimum of five sequence reads for mutant allele support and excluded synonymous nucleotide changes, a total of 1,803 unique genetic variants were identified across the patient cohort. On a per-patient basis, a median of 302 such variants were detected. Of these 1,803 unique variants, 1,322 were absent from the dbSNP, COSMIC, and ICGC databases. These variants, potentially novel and specific to the Indian ethnic background (as they were not catalogued in the Indigenomes or TMC-SNPdb2.0 databases), underscore the genetic diversity within this population; they were subsequently excluded from further analysis.

The remaining 481 variants showed varied database representation: 179 were listed in dbSNP, COSMIC, and ICGC; 190 in dbSNP and COSMIC; 23 in dbSNP and ICGC; 40 in COSMIC and ICGC; 47 exclusively in COSMIC; and 2 solely in ICGC. A subsequent clinical significance assessment using ClinVar led to the exclusion of an additional 336 variants, comprising 239 with no ClinVar annotations and 97 classified as benign or likely benign.

From the 145 variants that proceeded past these filters, 60 were categorised as variants of unknown significance (VUS), 48 exhibited conflicting interpretations of pathogenicity, and 37 were determined to be pathogenic or likely pathogenic. Ultimately, 85 variants—those with conflicting significance or classified as pathogenic/likely pathogenic were selected for downstream analysis. As anticipated, mutations in *TP53* (observed in 47% of the cohort) and *PIK3CA* (34%) were the most prevalent genetic alterations identified (**Fig. 2, Supplementary Table III**).

Actionable alterations: The 481 variants identified in at least one of the dbSNP, COSMIC, or ICGC databases

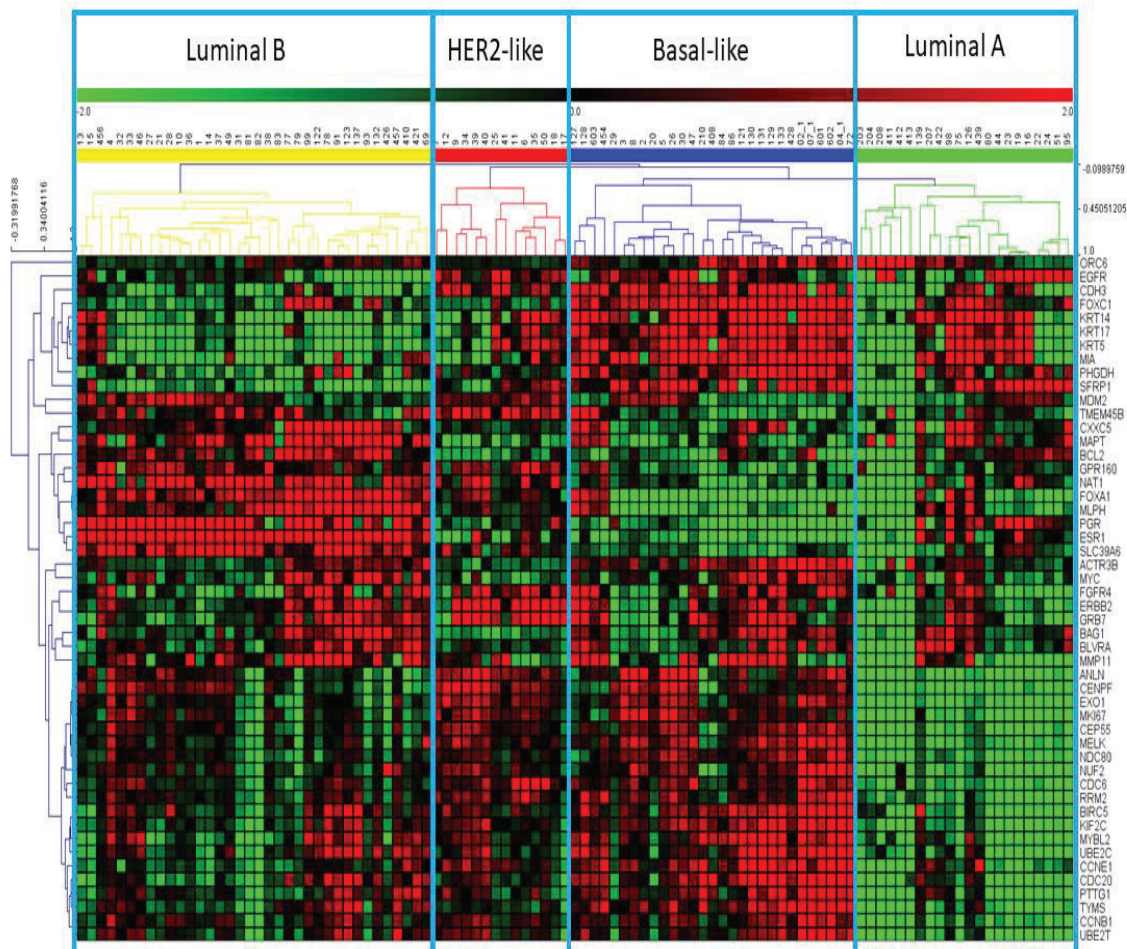


Fig. 1. PAM50 clustering and gene expression for 97 patients. The heatmap was generated using R software.

Table II. Comparison of receptor-based classes with intrinsic subtypes identified using PAM50 gene signature

PAM50 subtype	ER+ and/or PR+ and HER2-neg (n=46), %	Triple Negative Breast Cancer (n=24), %	ER+ and/or PR+ and HER2-pos (n=13), %	ER- and/or PR- and HER2-pos (n=8), %	HER2 Equivocal (n=6), %	Total (N=97), %
Luminal A	12 (26.08)	2 (8.33)	4 (30.76)	1 (12.5)	0 (0)	19 (19.58)
Luminal B	29 (63.04)	1 (4.16)	1 (7.69)	0 (0)	5 (83.33)	36 (37.11)
Basal-like	5 (10.86)	18 (75)	3 (23.07)	2 (25)	1 (16.6)	29 (29.89)
HER2-like	0 (0)	3 (12.5)	5 (38.46)	5 (62.5)	0 (0)	13 (13.40)

underwent further assessment for clinical actionability. This analysis was performed using the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT) framework, facilitated by a local academic installation of the Cancer Genome Interpreter (CGI).^{32,33} This evaluation revealed that a significant portion of the cohort, 51 out of 97 patients (52%), possessed at least one actionable mutation. The evidence supporting these actionable findings was categorised as follows: 20 mutations were supported by Level A evidence, 10 by

Level B evidence, and 15 by Level C evidence. Among the 51 patients with actionable mutations, alterations in *PIK3CA* and *TP53* were prominent. Specifically, 24 of these 51 patients had a *PIK3CA* mutation, and 17 had a *TP53* mutation. Furthermore, 9 patients within this group harboured concurrent mutations in both *PIK3CA* and *TP53*. These identified mutations were predominantly well-characterised, previously documented hotspot mutations (**Supplementary Fig. 2**).

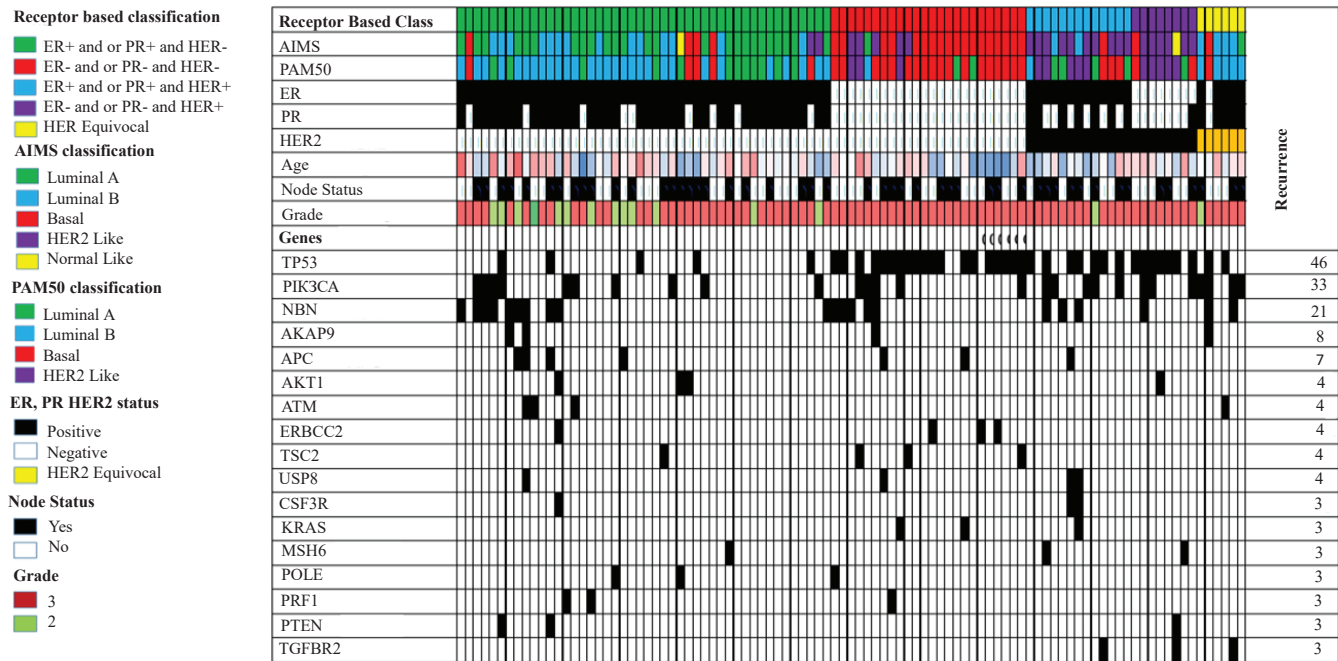


Fig. 2. Heatmap of genes mutated in at least 3% in cohort.

Validation of mutations: An attempt was made to validate the mutations initially identified through RNA-sequencing using DNA extracted from formalin-fixed paraffin-embedded (FFPE) tumour tissues. DNA was successfully extracted from 10 available tissue samples. However, the inherent degradation caused by the paraffinization process resulted in poor-quality DNA for most samples, rendering them unsuitable for subsequent assays. Of the evaluated samples, only two yielded DNA of acceptable quality, defined by a DIN greater than 2.5 and a concentration of at least 100 ng/ μ l. From these two samples, one mutation (specifically, *PIK3CA* H1047R in sample 14T) was successfully validated. The attempted validation of a second mutation (*PIK3CA* E545K in sample 69T) was unsuccessful. It is likely that this could be due to the limit of detection of the qPCR assay used (1% for this particular mutation). The tumour allele fraction in the FFPE sample might have been reduced due to the presence of normal tissue or fat infiltration, falling below the qPCR detection threshold, though it was identifiable by the more sensitive high-depth sequencing.

Network analysis in each intrinsic subtype: Our analysis identified 227 unique genes harbouring mutations that met our defined criteria. Twenty-one (9.3% of the 227) genes were found to be altered across all molecular subtypes. Further analysis revealed distinct sets of

uniquely mutated genes in each subtype. Specifically, 30 genes (13.2%) were mutated exclusively in the Luminal A subtype, 32 (14.1%) were unique to Luminal B, 28 (12.3%) were specific to the HER2-enriched subtype, and 33 (14.5%) were uniquely mutated in the Basal-like subtype (**Supplementary Fig. 3**). Pathway analyses were performed on these subtype-specific gene pools. This revealed that pathways related to immune response, DDD (Disease, Drug, and Development), haematopoiesis, and growth/development were significantly impaired in both Luminal A (**Supplementary Fig. 4A**) and Luminal B subtypes (**Supplementary Fig. 4B**). In contrast, the HER2-enriched subtype showed enrichment in pathways associated with Epidermal Growth Factor Receptor Family (ERBB) signalling, PI3K signalling, and T-cell regulation (**Supplementary Fig. 4C**). The Basal-like subtype was characterised by an enrichment of pathways involved in cell cycle regulation, including those governing microtubule and cytoplasmic regulation, nuclear excision repair, and developmental growth processes (**Supplementary Fig. 4D**).

Fusion transcript analysis: Across 207 patient samples (derived from 97 patients), a total of 225 potential fusion events were initially identified. Predicting fusion transcripts from RNA-sequencing data is challenging due to misalignment errors, library preparation artefacts, and the absence of germline control data. To

address these limitations, the analysis focused only on fusion transcripts detected in multiple tumour samples from the same patient, as true fusions are more likely to recur across independent experiments.

Applying this stringent criterion, 91 of the initial 225 fusion events were classified as recurrent, *i.e.*, present in more than one sample from the patient. Notably, this filtered set included novel fusion transcripts involving partner genes previously implicated in fusions, such as *ERBB2*, *CBX1*, *MED1*, *CDK12*, *ELF2*, and *KANSL1*. Despite these findings, the analysis suggests that commonly known fusion transcripts reported in other breast cancer populations do not appear to be prevalent in this Indian breast cancer cohort.

Discussion

In this analysis of 97 Indian breast cancer patients profiled by multi-sample RNA-Seq, we confirmed the presence of canonical intrinsic subtypes but noted discordance between AIMS and PAM50 in luminal cancers; uncovered a catalogue of 1,803 expressed variants, three-quarters of which were absent from global and Indian reference databases; identified clinically actionable alterations in over half the cohort, driven chiefly by hotspot *PIK3CA* and *TP53* mutations, and delineated subtype-specific pathway perturbations ranging from immune and haematopoietic signalling in luminal tumours to *ERBB/PI3K* activation in *HER2*-enriched disease and cell-cycle/DNA-repair dysregulation in basal-like cancers.

Both AIMS and PAM50 classifiers reproduced the four classical intrinsic subtypes, confirming that the molecular architecture of Indian tumours is broadly similar to that reported in Western cohorts. The discordance between Luminal A and Luminal B assignments by the two algorithms could be due to the different gene lists, cut-offs, and normalisation procedures in each algorithm, and cautions against the interchangeable use of signatures without population-specific calibration.

RNA-Seq-based variant calling revealed a median of 302 alterations per patient. However, 73% of the 1,803 unique variants were absent from dbSNP, COSMIC, and ICGC, even after cross-referencing two Indian germline catalogues. These putatively population-specific changes highlight the genetic heterogeneity of the Indian subcontinent and the need to expand reference panels that currently under-represent South Asian genomes. Although excluded from downstream analyses to minimise noise, these variants constitute

a resource for future germline and somatic discovery efforts.

Applying the ESCAT framework showed that 52% of patients harboured at least one clinically actionable mutation, led by hotspot *PIK3CA* and *TP53* events with Level A to Level C evidence. Importantly, nine patients carried co-occurring *PIK3CA* and *TP53* lesions, a combination associated elsewhere with endocrine resistance and poor prognosis.

Our study revealed the extensive molecular heterogeneity within breast cancer, corroborating previous reports.¹² We identified mutations in known tumour suppressor genes (*e.g.*, *TP53*, *PTEN*) and oncogenes (*e.g.*, *PIK3CA*, *AKT1*, *ERBB3*), providing insights into their prevalence and clinical relevance. We identified distinct mutational profiles and fusion events across different breast cancer subtypes, emphasising the importance of molecular subtyping in guiding treatment strategies. Integrating genomic alterations with traditional immunohistochemical receptor-based classification allowed for a refined classification.

Our study has notable strengths. Firstly, 76 of the 97 patients in our cohort had RNA-Seq data from at least two different tissue samples from the same tumour, each with an independent library preparation and analysis. This allows us to filter noise from this RNA-Seq data, excluding variants that are present only in a single sample, and thus report variants with high confidence (**Fig. 2**). Second, our sample size ($n=97$) constitutes a reasonably sized cohort in which we have conducted an unbiased, genome-wide sequencing. Most studies from India have been performed on targeted gene panels and, therefore, were restricted in their ability to make discovery-level findings.

Our study has some limitations. RNA-based variant detection captures only expressed alleles, may miss truncal mutations in lowly expressed genes and is susceptible to artefacts from RNA editing or reverse-transcription errors. The lack of matched normal tissue limits discrimination of somatic versus germline events, which was partly mitigated by stringent database filtering, but may have excluded true somatic variants unique to South Asians. We also did not elucidate epigenetic alterations such as *BRCA1* promoter methylation, which can have therapeutic implications.³⁴

In conclusion, our study demonstrates that Indian breast cancers display the recognised intrinsic subtypes and a high prevalence of therapeutically actionable

शोध-संदेश

यह अध्ययन भारतीय स्तन कैंसर रोगियों में सूक्ष्म स्तरों और जीनोमिक परिवर्तनों के विश्लेषण पर केंद्रित है, क्योंकि इन रोगियों से संबंधित आंकड़े सीमित हैं। शोध का उद्देश्य यह समझना था कि भारतीय रोगियों में पाए जाने वाले जीनोमिक म्यूटेशन (mutation) ट्यूमर की सर्वेदनशीलता, आक्रामकता, और उपचार प्रतिक्रिया को किस प्रकार प्रभावित करते हैं। अध्ययन से यह स्पष्ट हुआ कि भारतीय स्तन कैंसर रोगियों में पाए जाने वाले उपप्रकार और उपचार योग्य म्यूटेशन (actionable mutation) काकेशियन आबादी में रिपोर्ट किये गए निष्कर्षों के समान हैं। ये परिणाम भारतीय रोगियों में भी जीनोमिक-आधारित निदान और व्यक्तिगत उपचार रणनीतियों की उपयोगिता को समर्थन प्रदान करते हैं।

somatic genetic lesions. Expansion to larger, prospectively accrued cohorts with matched germline DNA, fresh-frozen tissue for orthogonal validation, and longitudinal clinical data will be essential to translate these genomic insights into precision-oncology interventions tailored to Indian patients.

Author contributions: SG: Conceptualized and designed this study; RB, SG: Acquired the funding for this study; NG, RC, VV, RH, SJ, RB, SG: Screened, consented, and recruited the patients for this study; NG, RC, KG, AK, AS, SS: Bio-banked and processed the patient samples for all assays; NG, RC, KG, AK, AS, AR, SS: Followed-up patients and obtained clinical data; NG, RC, SG: Analysed the data, manuscript writing. All authors have read and approved the final printed version of the manuscript.

Acknowledgment: Authors acknowledge Dr Omshree Shetty for performing the qPCR analysis of the PIK3CA hotspot mutation for validation.

Financial support and sponsorship: This study was funded by the Department of Atomic Energy, Government of India. This study was also funded by the Department of Biotechnology (DBT), GOI, through the DBT-Virtual National Cancer Institute (VNCI) Breast Cancer 2015 Grant (BT/MED/30/VNCI-Hr-BRCA/2015) awarded to SG. The study also received funding from Department of Science and Technology (DST) - Scientific Engineering and Research Board (SERB) and Prime Minister's Fellowship awarded to NG. We acknowledge funding from Mizuho Bank Limited for research infrastructure to the research laboratory. We thank Mr. Akhil Gupta for funding laboratory infrastructure. We acknowledge part research funding for this study from the Women's Cancer Initiative (WCI) – Tata Memorial Hospital. RC and NG were funded by a fellowship from HBNI, Mumbai, and TMC, Mumbai.

Conflicts of Interest: None.

Use of Artificial Intelligence (AI)-Assisted Technology for manuscript preparation: The authors confirm that there was no use of AI-assisted technology for assisting in the writing of the manuscript and no images were manipulated using AI.

References

- Thakkar AD, Raj H, Chakrabarti D, Ravishankar N, Saravanan N, Muthuvelan B, *et al.* Identification of gene expression signature in estrogen receptor positive breast carcinoma. *Biomark Cancer.* 2010;2:1–15.
- Malvia S, Bagadi SAR, Pradhan D, Chintamani C, Bhatnagar A, Arora D, *et al.* Study of gene expression profiles of breast cancers in Indian women. *Sci Rep.* 2019;9:10018.
- Ghosh M, Naik R, Lingaraju SM, Susheela SP, Patil S, Srinivasachar GK, *et al.* Landscape of clinically actionable mutations in breast cancer: A cohort study. *Transl Oncol Neoplasia.* 2020;14:100877.
- Manjunath M, Nirgude S, Mhatre A, Vemuri SG, Nataraj M, Thumsi J, *et al.* Transcriptomic profiling of Indian breast cancer patients revealed subtype-specific mRNA and lncRNA signatures. *Front Genet.* 2022;13:932060.
- Gardi N, Chaubal R, Parab P, Pachakar S, Kulkarni S, Shet T, *et al.* natural history of germlineBRCA1 mutated and BRCA wild-type triple-negative breast cancer. *Cancer Res Commun.* 2024;4:404–17.
- Ghosh A, Chaubal R, Das C, Parab P, Das S, Maitra A, *et al.* Genomic hallmarks of endocrine therapy resistance in ER/PR+HER2- breast tumours. *Commun Biol.* 2025;8:207.
- Ghosh J, Gupta S, Desai S, Shet T, Radhakrishnan S, Suryavanshi P, *et al.* Estrogen, progesterone and HER2 receptor expression in breast tumors of patients, and their usage of HER2-targeted therapy, in a tertiary care centre in India. *Indian J Cancer.* 2011;48:391–6.
- Nair N, Shet T, Parmar V, Havaldar R, Gupta S, Budrukkar A, *et al.* Breast cancer in a tertiary cancer center in India – An audit, with outcome analysis. *Indian J Cancer.* 2018;55:16–22.
- Bhargava P, Rathnasamy N, Shenoy R, Gulia S, Bajpai J, Ghosh J, *et al.* Clinical profile and outcome of patients with human epidermal growth factor receptor 2-positive breast cancer with brain metastases: Real-world experience. *JCO Glob Oncol.* 2022;8:e2200126.
- Chatterjee S, Chaubal R, Maitra A, Gardi N, Dutt A, Gupta S, *et al.* Pre-operative progesterone benefits operable breast cancer patients by modulating surgical stress. *Breast Cancer Res Treat.* 2018;170:431–8.
- Chaubal R, Gardi N, Joshi S, Pantvaidya G, Kadam R, Vanmali V, *et al.* Surgical tumor resection deregulates hallmarks of cancer in resected tissue and the surrounding microenvironment. *Mol Cancer Res.* 2024;22:572–84.
- De Mattos-Arruda L, Sammut SJ, Ross EM, Bashford-Rogers R, Greenstein E, Markus H, *et al.* The genomic and immune landscapes of lethal metastatic breast cancer. *Cell Rep.* 2019;27:2690–708.
- Paquet ER, Hallett MT. Absolute assignment of breast cancer intrinsic molecular subtype. *J Natl Cancer Inst.* 2014;107:357.

14. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417–9.
15. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521.
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al*. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, *et al*. The genome analysis toolkit: A map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
18. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213–9.
19. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, *et al*. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–76.
20. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, *et al*. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44: e108.
21. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
22. Phan L, Zhang H, Wang Q, Villamarin R, Hefferon T, Ramanathan A, *et al*. The evolution of dbSNP: 25 years of impact in genomic research. *Nucleic Acids Res*. 2025;53:D925–31.
23. Sondka Z, Dhir NB, Carvalho-Silva D, Jupe S, Madhumita N, McLaren K, *et al*. COSMIC: A curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res*. 2024;52:D1210–7.
24. Desai S, Mishra R, Ahmad S, Hait S, Joshi A, Dutt A. TMC-SNPdb 2.0: An ethnic-specific database of Indian germline variants. *Database (Oxford)*. 2022;2022:baac029.
25. Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, *et al*. IndiGenomes: A comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res*. 2021;49:D1225–32.
26. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, *et al*. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308–11.
27. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Res*. 2000;28:352–5.
28. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, *et al*. The International cancer genome consortium data portal. *Nat Biotechnol*. 2019;37:367–9.
29. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, *et al*. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51:D638–46.
30. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol*. 2019;20:213.
31. Dharavath B, Butle A, Chaudhary A, Pal A, Desai S, Chowdhury A, *et al*. Recurrent UBE3C-LRP5 translocations in head and neck cancer with therapeutic implications. *NPJ Precis Oncol*. 2024;8:63.
32. Mateo J, Chakravarty D, Dienstmann R, Jezdic S, Gonzalez-Perez A, Lopez-Bigas N, *et al*. A framework to rank genomic alterations as targets for cancer precision medicine: The ESMO scale for clinical actionability of molecular targets (ESCAT). *Ann Oncol*. 2018;29:1895–902.
33. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, *et al*. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*. 2018;10:25.
34. Jagtap SV, Jagtap SS. Methylation of BRCA1 promoter in sporadic breast cancer. *Indian J Med Res*. 2023;158:85–7.

For correspondence: Dr Sudeep Gupta, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai 400012, Maharashtra, India
e-mail: sudeepgupta04@yahoo.com

Supplementary file(s) available at: https://doi.org/10.25259/IJMR_1642_2024