Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genomic Alterations Converge on Lipid Metabolism and ERBB2 in Breast Cancer.

Shantanu Chowdhury

shantanuc@igib.in

CSIR-Institute of Genomics and Integrative Biology https://orcid.org/0000-0001-7185-8408

Divya Khanna

CSIR-Institute of Genomics and Integrative Biology

Arnab Ghosh

Biotechnology Research and Innovation Council National Institute of Biomedical Genomics.

Priyanka Bhadwal

CSIR Institute of Genomics and Integrative Biology.

Ramakant Venkata

CSIR Centre for Cellular & Molecular Biology.

Leepakshi Dhingra

CSIR Institute of Genomics and Integrative Biology.

Subrata Das

Biotechnology Research and Innovation Council National Institute of Biomedical Genomics.

Isha Choubey

CSIR Centre for Cellular & Molecular Biology.

Ankita Singh

Integrative and Functional Biology Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110026, India

Geeta Jadaun

CSIR Institute of Genomics and Integrative Biology.

Hriday Chowdhury

CSIR Institute of Genomics and Integrative Biology.

Suryanshi Tiwari

CSIR Institute of Genomics and Integrative Biology.

Vibha Mattoo

CSIR Institute of Genomics and Integrative Biology.

Pooja Dixit

CSIR Institute of Genomics and Integrative Biology.

Safeer Khan

CSIR Institute of Genomics and Integrative Biology.

Vinit Gupta
CSIR Institute of Genomics and Integrative Biology.
Tuneer Mallick
Biotechnology Research and Innovation Council National Institute of Biomedical Genomics.
Surya Mishra
CSIR Centre for Cellular & Molecular Biology.
Tulasi Nagabandi
CSIR Centre for Cellular & Molecular Biology.
Venkata Krishna Vanamamalai
CSIR Centre for Cellular & Molecular Biology. https://orcid.org/0000-0002-9388-521X
Malini Nemalikanti
CSIR Centre for Cellular & Molecular Biology. https://orcid.org/0009-0004-3088-9346
Renu Kumari
CSIR Institute of Genomics and Integrative Biology.
Diksha Singhal
CSIR Institute of Genomics and Integrative Biology.
Laxmi Yadav
CSIR Institute of Genomics and Integrative Biology.
Vaishali Pandey
CSIR Institute of Genomics and Integrative Biology.
Shailya Verma
CSIR Institute of Genomics and Integrative Biology.
Rashmi Gudur
Krishna Vishwa Vidyapeeth
Senthilkumar Ramasamy
BALCO Medical centre, Vedanta Medical Research Foundation https://orcid.org/0000-0002-4896-4576
Susanta Roychoudhury
Saroj Gupta Cancer Centre & Research Institute
Rohan Chaubal
tata memorial Centre https://orcid.org/0000-0002-7226-8757
Sonam Dhamija
South Asian University https://orcid.org/0000-0002-9569-2729
Koushik Mondal
Saroj Gupta Cancer Centre & Research Institute
Lipi Thukral
CSIR-Institute of Genomics and Integrative Biology https://orcid.org/0000-0002-1961-039X
Aruna Korlimarla
Sri Shankara Cancer Hospital & Research Centre, Bangalore, India

Rekha Kumar
Sri Shankara Cancer Hospital and Research Centre
Ashutosh Mishra
All India Institute of Medical Sciences Delhi
Sandeep Mathur
All India Institute of Medical Sciences Delhi
Bhawna Sirohi
BALCO Medical centre, Vedanta Medical Research Foundation
SVS Deo
All India Institute of Medical Sciences Delhi , Apollo Hospitals, New Delhi
Samir Bhattacharya
Saroj Gupta Cancer Centre & Research Institute
B Srinath
Sri Shankara Cancer Hospital and Research Centre
Juhi Tayal
Rajiv Gandhi Cancer Institute and Research Center
Anurag Mehta
Rajiv Gandhi Cancer Institute and Research Center
IBCGA Consortium
IBCGA Consortium
Shilpak Chatterjee
CSIR Indian Institute of Chemical Biology
Sanjeev Khosla
CSIR Institute of Microbial Technology
Divya Tej Sowpati
CSIR Centre for Cellular and Molecular Biology https://orcid.org/0000-0003-4340-6011
Kumardeep Chaudhary
Institute of Genomics and Integrative Biology is a scientific research
Karthik Tallapaka
CSIR - Centre for Cellular and Molecular Biology (CSIR-CCMB)
Sherry Bhalla
CSIR Institute of Genomics and Integrative Biology.
Nidhan Biswas
Biotechnology Research and Innovation Council-National Institute of Biomedical Genomics (BRIC-NIBMG)
Sudeep Gupta
Tata Memorial Centre https://orcid.org/0000-0002-6742-6378

Biological Sciences - Article

Keywords:

Posted Date: September 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10980563/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Genomic alterations converge on lipid metabolism and ERBB2 in breast cancer

Divya Khanna[1#], Arnab Ghosh[2#], Priyanka Bhadwal[1$], Leepakshi Dhingra[1$], Ramakanth Chirravuri-Venkata[3$], Subrata Das[2$], Isha Choubey[3], Ankita Singh[1], Geeta Jadaun[1], Hriday Home Chowdhury[1], Suryanshi Tiwari[1], Vibha Mattoo[1], Pooja Dixit[1], Safeer Khan[1], Vinit Gupta[1], Tuneer R. Mallick[2], Surya Narayan Mishra[3], Tulasi Nagabandi[3], Venkata Krishna Vanamamalai[3], Malini Nemalikanti[3], Renu Kumari[1], Diksha Singhal[1], Laxmi Yadav[1],Vaishali Pandey[1], Shailya Verma[1], Rashmi Gudur[4] , Senthil Kumar[9], Susanta Roychoudhury[5], Rohan Chaubal[6], Sonam Dhamija[1], Koushik Mondal[5], Lipi Thukral[1], Aruna Korlimarla[7], Rekha V Kumar[7], Ashutosh Mishra[8], Sandeep Mathur[8] , Bhawna Sirohi[10], SVS Deo[8,9], Samir Bhattacharya[5], B S Srinath[7], Juhi Tayal[11], Anurag Mehta[11], IBCGA Consortium, Shilpak Chatterjee[12] , Sanjeev Khosla[13], Divya Tej Sowpati[3], Kumardeep Chaudhary[1], Karthik Bharadwaj Tallapaka[3✉], Sherry Bhalla[1], Nidhan K. Biswas[2✉], Sudeep Gupta[6,14✉], Shantanu Chowdhury[1,15,16✉].

[1]CSIR Institute of Genomics and Integrative Biology, South Campus, Opp. Bus Depot, Sukhdev Vihar, Mathura Road, New Delhi – 110025

[2]Biotechnology Research and Innovation Council-National Institute of Biomedical Genomics (BRIC-NIBMG), P.O. NSS, Kalyani, West Bengal, India-741251

[3]CSIR- Centre for Cellular & Molecular Biology, Habsiguda, Uppal Road, Hyderabad - 500 007, Telangana.

[4]Krishna Vishwa Vidyapeeth, Dhebewadi Road, Malkapur, Karad, Taluka Karad, District Satara, Maharashtra, India Karad 415539

[5]Saroj Gupta Cancer Centre & Research Institute, Mahatma Gandhi Road, Thakurpukur, Kolkata - 700 063.

[6]Tata Memorial Hospital (TMH), is Dr. Ernest Borges Marg, Parel, Mumbai-400012, Maharashtra

[7]Sri Shankara Cancer Hospital and Research Centre, 1st Cross, Shankara Math Premises, Shankarapuram, Basavanagudi, Bengaluru, Karnataka 560 004.

[8]All India Institute of Medical Sciences, Ansari Nagar, New Delhi – 110029

[9]Oncology Lead & Senior Consultant Surgical Oncology, Apollo Hospitals, New Delhi

[10] BALCO Medical centre, Vedanta Medical Research Foundation, Raipur, Sector-36, Atal Nagar, PO: Uparwara, Raipur, Chhattisgarh 493661.

[11]Rajiv Gandhi Cancer Institute and Research Center, D-18, Sector-5, Rohini, Delhi-110085

[12]CSIR-Indian Institute of Chemical Biology, Raja S.C. Mullick Road, Kolkata, India

[13]CSIR-Institute of Microbial Technology, Sector 39-A, Chandigarh – 160036

[14]Department of Medical Oncology, Homi Bhabha National Institute, Mumbai, India.

[15]Integrative and Functional Biology Unit, CSIR Institute of Genomics and Integrative Biology, New Delhi 110025, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India.

[16]Current Address: Bharat Charitable Cancer Hospital & Institute (Trust), Mysuru, 570024, India

✉ Corresponding author
shantanuc@igib.in
sudeep.gupta@actrec.gov.in
nkb1@nibmg.ac.in

karthikt@ccmb.res.in

# First Authors: equal contribution
$ Shared second author

Abstract

Breast cancer genomes harbour diverse alterations[1,2], but how these converge on tumour states and clinical behaviour remains unclear, particularly in underrepresented populations[3]. Here, whole-genome sequencing of 501 treatment-naïve breast tumour–normal pairs from the Indian Breast Cancer Genome Atlas was integrated with transcriptomic, copy-number, structural-variation (SV) and clinical data. We identify candidate positively selected genes not previously reported in major cohorts, and frequent alterations in lipid metabolic regulators associated with PI3K pathway activation and *ERBB2* amplification. SVs at the *ERBB2* locus revealed a distinct IKZF3-associated rearrangement activating *ERBB2* independently of canonical amplification. Transcriptomic integration identified four molecular states transcending receptor classification; the HER2–androgen-receptor state, marked by androgen signalling, lipid reprogramming and reduced immune signalling, had the poorest recurrence-free survival. We further detect a HER2-like programme within HER2 IHC-negative triple-negative tumours with reduced immune infiltration. Together, these define a lipid–*ERBB2* axis linking genomic diversity to tumour state and outcome.

Introduction

Large-scale genomic studies, including TCGA and METABRIC, have catalogued recurrent driver events, established molecular subtypes and revealed substantial genomic and transcriptomic heterogeneity in breast cancer[1,2]. Yet an important question remains: how does extensive genomic diversity integrate on the cellular programmes that determine tumour states, disease progression and clinical outcome?

Metabolic adaptation is a fundamental component of malignant transformation[4]. Altered lipid synthesis, storage and utilization support membrane biogenesis, energy homeostasis and signalling, and can promote proliferation and metastatic competence[5,6]. Oncogenic pathways like PI3K–AKT–mTOR and *ERBB2* signalling are known to regulate lipid metabolism, underlining their biological and therapeutic importance in breast cancer[7–9] . What remains unclear is whether lipid metabolism is solely a downstream consequence of oncogenic signalling, or whether recurrent genomic alterations target lipid-metabolic regulators and thereby contribute to specific tumour-states.

Tumour phenotypes can also be altered without changing gene dosage. Structural rearrangements can reconfigure enhancer–promoter interactions and activate oncogenes through regulatory rewiring,

including enhancer or promoter hijacking[10]. However, whether such regulatory alterations cooperate with metabolic programmes to shape breast-cancer states remains poorly understood.

Large-scale breast-cancer genomic studies have predominantly profiled European-ancestry and, more recently, East Asian populations[11,12], leaving South Asian breast cancer comparatively underrepresented. Population diversity, including where breast cancer screening is not prevalent[13], provides an opportunity to identify population-enriched alterations and to examine whether genomic heterogeneity is associated with distinct biological programmes.

Here we present the Indian Breast Cancer Genome Atlas (IBCGA) (Supplementary Table S1a), a multi-institutional cohort integrating whole-genome sequencing of 501 treatment-naïve breast tumour–normal pairs with transcriptomic, copy-number, structural-variation and clinical data. We use this integrated framework to examine whether and how diverse genomic alterations align across signalling, metabolic and regulatory layers, and whether such convergence identifies biologically and clinically distinct tumour states.

Results with Figure Legends

Somatic alteration landscape reveals population-enriched drivers and evolutionary selection

Whole-genome sequencing (WGS) of 501 tumour–normal pairs (40x median depth for tumour and normal) (Supplementary Figures S1, S2, S3; Supplementary Tables S1a, S1b, S2, S3) identified 4.6 million somatic mutations, including 12.53% insertions and deletions. Mutation burden was highest in triple-negative breast cancers (TNBCs; median 2.13/Mb) and lowest in hormone receptor (HR)–positive HER2-negative tumours (1.42/Mb), with intermediate levels in HER2+ subtypes. TNBCs also exhibited substantially greater copy-number alteration burden, with a median 42.51% of the genome altered compared with 26.62% in other subtypes ($p = 1.58 \times 10^{-5}$), driven by higher amplification ($p = 0.0003$) and deletion ($p = 4.76 \times 10^{-6}$) burdens (Supplementary Figure S4).

Using MutSig2CV[14], we identified 22 significantly mutated genes (SMGs; $q < 0.05$), considering nonsynonymous somatic variants present in $\geq$2% of tumours overall and within at least one IHC-defined subtype (Figure 1A, Supplementary Figure S5, S6, S7, S8). These included canonical breast cancer drivers such as *TP53, PIK3CA, GATA3* and *MAP3K1*, as well as *ARID1A* and *ERBB2* (Supplementary Figure S9), which are less commonly identified as significantly mutated genes in other breast cancer cohorts (Figure 1A).

Although the overall driver landscape was broadly consistent with TCGA and other published cohorts, several alterations differed in frequency, particularly in TNBCs: KMT2C (17% vs 6%), ARID1A (9% vs 2%) and MAP3K1 (8% vs 1%) were significantly more frequent in IBCGA than TCGA, while AKT1

mutations occurred in 2% of IBCGA TNBCs but were absent in TCGA (Figure 1B, Supplementary Tables S4a, S4b).

Copy-number alterations also showed population-specific patterns: *AKR1C3, NCEH1, PLD1, IL10, ADIPOR2* and *TERT* amplifications and *BTNL3* deletion were observed in Indian TNBCs but not TCGA TNBCs, while several amplifications were enriched in Indian HER2+ tumours; notably, *MYC* amplification was significantly more frequent in Indian HR+HER2− tumours than in TCGA (23% vs 9%; Supplementary Tables S9, S10).

Although at relatively low frequency, we found support for four genes *ISM2, QRICH2, CBX3* and *OTOP1* as SMGs, and candidate driver genes, in the IBCGA cohort (Figure 1A-C). In thirty-five population-level breast cancer cohorts comprising 17,708 samples including TCGA these four genes were not reported as SMGs (Figure 1B and Supplementary Tables S4a and S4b). All the four SMGs were supported by evidence of significant positive selection. Nonsynonymous mutation rates and dN/dS calculated independently for missense mutations, truncating mutations and insertion/deletions compared strongly with established breast cancer drivers after correction for GC content and background mutation rate (shown in log scale in Figure 1C) – supporting these as candidate drivers not previously identified as significantly mutated in the reference breast-cancer cohorts examined (Figure 1D). Further, majority of the identified nonsynonymous variations within *ISM2, QRICH2, CBX3* and *OTOP1* were IBCGA-specific, mapped to conserved and functional protein domains, and were predicted to be deleterious by multiple algorithms (Figure 1E–G; Supplementary Table S5).

Next, we considered all the identified SMGs. A substantial fraction of nonsynonymous variants (median, 86% of SNVs and 36% of indels), which were not reported earlier in thirty-five population-level cohorts including TCGA and METABRIC, mapped to conserved protein domains, and independent pathogenicity prediction algorithms indicated their deleterious role (Figure 1E-G), supporting functional consequence of somatic alterations detected in this cohort. Driver alterations also varied across IHC-defined subtypes (Figure 1H & I).

Consistent with this, protein structure-based analysis localised variants to constrained functional surfaces. Fifteen of 19 *PIK3CA* variants unique in IBCGA, that is, absent from TCGA-BRCA, mapped to the interface with regulatory subunits PIK3R1/2/3, and docking predicted loss of interface salt bridges and hydrogen bonds (V344G, N345S, E418K, E453Q, Q546E; Supplementary Results, Supplementary Figure S10, S11). Release of p85-mediated inhibition was reported to activate canonical *PIK3CA* hotspots[15], suggesting association of these variants with PI3K activation.

Several driver mutations, including recurrent hotspot mutations in *PIK3CA* and *TP53*, were predicted to generate neoantigens in the context of patient-specific germline HLA alleles (Supplementary Table S6). Notably, the prevalence of specific hotspot mutations showed an inverse correlation with their predicted HLA presentation (e.g., neoantigenic) potential, suggesting that immune-mediated selection may influence the observed driver-mutation spectrum (Supplementary Figures S12, S13).

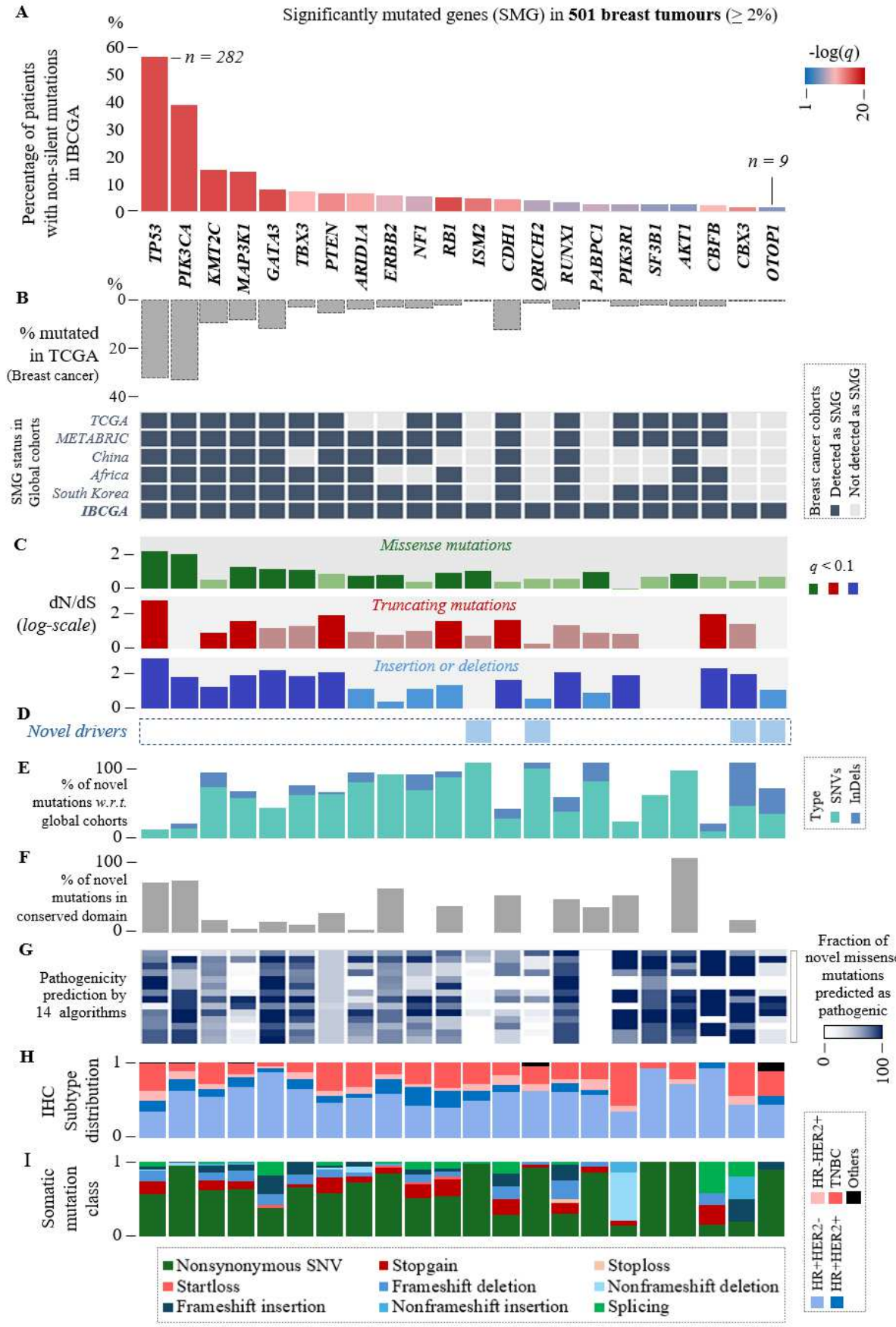

A
Significantly mutated genes (SMG) in 501 breast tumours (≥ 2%)
Percentage of patients with non-silent mutations in IBCGA
n = 282
n = 9
-log(q)
TP53
PIK3CA
KMT2C
MAP3K1
GATA3
TBX3
PTEN
ARID1A
ERBB2
NF1
RB1
ISM2
CDH1
QRICH2
RUNX1
PABPC1
PIK3R1
SF3B1
AKT1
CBFB
CBX3
OTOP1
B
% mutated in TCGA (Breast cancer)
SMG status in Global cohorts
TCGA
METABRIC
China
Africa
South Korea
IBCGA
Breast cancer cohorts
Detected as SMG
Not detected as SMG
C
dN/dS (log-scale)
Missense mutations
Truncating mutations
Insertion or deletions
q < 0.1
D
Novel drivers
E
% of novel mutations w.r.t. global cohorts
Type
SNVs
InDels
F
% of novel mutations in conserved domain
G
Pathogenicity prediction by 14 algorithms
Fraction of novel missense mutations predicted as pathogenic
H
IHC Subtype distribution
I
Somatic mutation class
HR-HER2-
TNBC
Others
HR+HER2-
HR+HER2+
Nonsynonymous SNV
Stopgain
Stoploss
Startloss
Frameshift deletion
Nonframeshift deletion
Frameshift insertion
Nonframeshift insertion
Splicing

Figure 1: Population-enriched driver landscape identifies positively selected breast cancer genes in the IBCGA cohort. (A-B) Twenty-two significantly mutated genes (SMGs) identified by MutSig2CV analysis of 501 treatment-naïve breast tumours ($q < 0.05$; genes present in ≥2% of tumours). Colour intensity denotes statistical significance (-log10(q)) (A); and, mutation frequencies of the identified SMGs in IBCGA compared with representative global breast cancer cohorts (TCGA and 4 others). Comparison with all 35 external cohorts is provided in Supplementary Tables S4a and S4b. (C) Gene-specific dN/dS ratios for missense, truncating and insertion/deletion mutations following correction for background mutation rate and GC content, consistent with positive selection. **(D)** Four genes (*ISM2, QRICH2, CBX3* and *OTOP1*) identified as candidate driver genes in the IBCGA cohort. **(E)** Fraction of novel somatic SNVs and InDels compared to major breast cancer cohorts. **(F)** Proportion of novel mutations mapping to conserved functional protein domains. **(G)** Pathogenicity prediction of novel missense variants using fourteen independent computational algorithms (Supplementary Table S5). **(H)** Distribution of SMGs across immunohistochemistry-defined breast cancer subtypes. **(I)** Spectrum of nonsynonymous mutation classes observed for each SMG. Number of patients corresponding to each subtype: HR+ (ER+/PR+) HER2-: 260; HR- (ER-/PR-) HER2+: 45; HR+ HER2+: 64; TNBC: 128; Others (IHC inconclusive): 4.

Genomic alterations converge on a coordinated lipid metabolic program

Across subtypes, genomic alterations are consistent with a coordinated lipid metabolic programme characterized by PI3K–AKT pathway activation and *ERBB2* signalling (Figure 2A; Supplementary Table S7). Surprisingly, 73.9% (362 of 490; in 11 cases CNV could not be resolved unambiguously) of tumours harboured either activating mutations in known PI3K pathway genes[16] or focal copy-number alterations in lipid metabolic regulators[17]. Gene-specific occurrences within specific tumour subtypes: *PIK3CA* oncogenic mutations or *PTEN* deletions, alongside somatic mutations/copy number amplifications in genes that contribute directly to impaired lipid metabolism are shown in Figure 2A (*PIK3CA* or *PTEN* oncogenic mutations in 18.98% of the tumours; *PIK3CA* amplifications or *PTEN* deletions along with amplification in other lipid related genes in 36.35%; and, both of these events were found in 18.55% cases).

Recurrent copy-number gains affected genes involved in lipid synthesis, lipolysis, and trafficking (including *NCEH1, PLD1, ADIPOR2, ESYT2*), alongside amplification of genes directly affecting the PI3K-AKT pathway (*PI4KB, GATA3, PIK3CA*), and, additionally, deletion of the PI3K-AKT suppressor *PTEN* (Figure 2A, Supplementary Figure S14). Moreover, these alterations were associated

with significantly increased expression of lipid metabolism genes: including fatty acid metabolism, cholesterol homeostasis and adipogenesis (Figure 2B).

PIK3CA mutations were associated with increased metabolic pathway activity in both HR+ and TNBC tumours, with subtype-specific differences in pathways (Figure 2B, Supplementary Figure S15). Similarly, in *ERBB2* amplification associated HER2+ tumours, enhanced expression of lipid- associated genes and pathways, including oestrogen response, PI3K-AKT-mTOR signalling, xenobiotic metabolism and adipogenesis was notable (Figure 2B-C). These alterations co-occur across lipid synthesis, storage and utilization (Figure 2D) indicating subtype-dependent lipid metabolic reprogramming that links oncogenic signalling to the metabolic state.

To understand if this was cohort-specific, we independently tested the lipid-alteration genomic changes in data from other large breast cancer cohorts TCGA-BRCA and CPTAC. In TCGA-BRCA, 551 out of 857 (64.3%) patients with invasive/ductal carcinoma had at least one of the genomic changes within the lipid-related genes (Supplementary Figure S16A). Systematic comparisons of ssGSEA scores between the two groups (lipid-altered vs unaltered at the somatic mutational or copy number level) identified lipid-related hallmark pathways including, cholesterol homeostasis, glycolysis, MTORC1, and PI3K-Akt-mTOR signalling to be concordantly upregulated in 'lipid-altered' tumours (($p < 0.05$, one-tailed Wilcox; Supplementary Figure S16B).

In the CPTAC data comprising 122 breast cancers, 78.7% of the tumours had lipid-related genomic alterations (Supplementary Figure S17A); and, ssGSEA analysis identified four lipid-related hallmark pathways (cholesterol homeostasis, glycolysis, MTORC1 and PI3K-Akt-mTOR signalling) had increased expression in the 'lipid-altered' tumours relative to the unaltered group (Supplementary Figure S17B). Although underpowered in CPTAC, these findings were directionally consistent and support recurrent genomic selection affecting lipid metabolism.

Overall, arm-level amplifications at chr6, deletions at chr21, focal amplifications in genes (including *PI4KB, PIK3CA, CCND1, MYC* and *IL10*), and deletions (including *CDKN2A, PTEN, RB1* and *ARID3A*) were prevalent. 79% HER2+ tumours had *ERBB2* amplification, and TNBC tumours had significant copy number amplifications (*PIK3CA, PI4KB, TERT, TERC, SOX11, FGF14, GATA3* and *EGFR*), and deletions (*RB1, ARID3A, BTNL3* and *CREBBP*) genes.

Individual metabolic genes have previously been shown to undergo copy-number-driven selection in breast cancer — notably *FLAD1* amplification at 1q21.3, which drives lipogenesis[18], and *ENSA* amplification, which promotes cholesterol biosynthesis[19]. Our observation is distinct in scope: rather than a single amplified enzyme acting through one downstream axis, we find that alterations distributed

227 across multiple lipid-regulatory loci co-occur with PI3K–AKT and *ERBB2* events in the majority of
228 tumours and are reproduced in TCGA-BRCA and CPTAC, indicating class-level rather than gene-level
229 selection on the lipid programme.

230
231

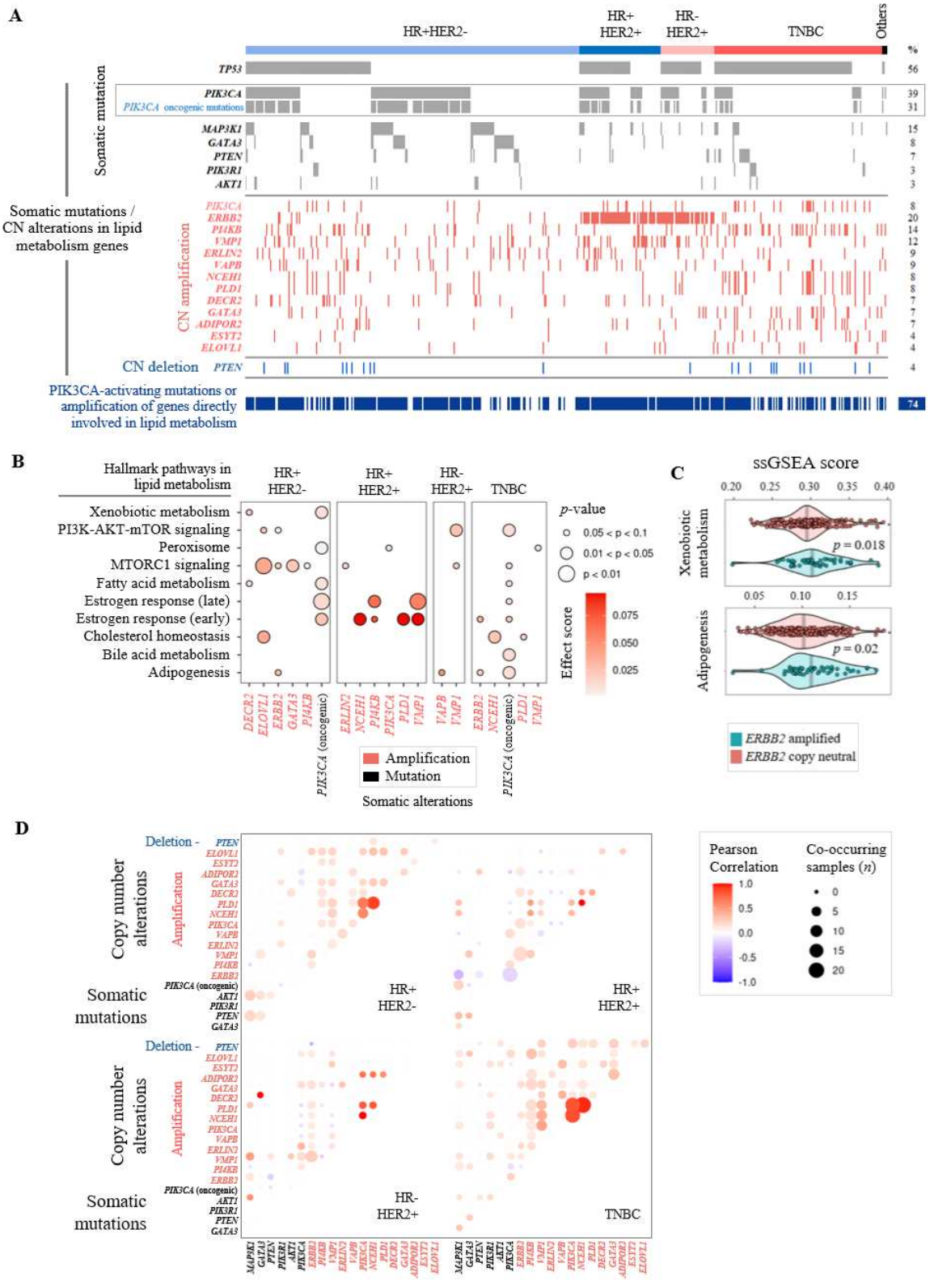

A
HR+HER2-
HR+ HER2+
HR- HER2+
TNBC
Others
%
Somatic mutation
TP53
PIK3CA
PIK3CA oncogenic mutations
MAP3K1
GATA3
PTEN
PIK3R1
AKT1
Somatic mutations / CN alterations in lipid metabolism genes
CN amplification
PIK3CA
ERBB2
PI4KB
VMP1
ERLIN2
VAPB
NCEH1
PLD1
DECR2
GATA3
ADIPOR2
ESYT2
ELOVL1
CN deletion
PTEN
PIK3CA-activating mutations or amplification of genes directly involved in lipid metabolism
B
Hallmark pathways in lipid metabolism
Xenobiotic metabolism
PI3K-AKT-mTOR signaling
Peroxisome
MTORC1 signaling
Fatty acid metabolism
Estrogen response (late)
Estrogen response (early)
Cholesterol homeostasis
Bile acid metabolism
Adipogenesis
p-value
0.05 < p < 0.1
0.01 < p < 0.05
p < 0.01
Effect score
0.075
0.050
0.025
Amplification
Mutation
Somatic alterations
C
ssGSEA score
0.20
0.25
0.30
0.35
0.40
Xenobiotic metabolism
p = 0.018
0.05
0.10
0.15
Adipogenesis
p = 0.02
ERBB2 amplified
ERBB2 copy neutral
D
Deletion -
Copy number alterations
Amplification
Somatic mutations
PIK3CA (oncogenic)
AKT1
PIK3R1
PTEN
GATA3
HR+ HER2-
HR+ HER2+
HR- HER2+
TNBC
Pearson Correlation
1.0
0.5
0.0
-0.5
-1.0
Co-occurring samples (n)
0
5
10
15
20

Figure 2 | Genomic alterations converge on coordinated lipid metabolic reprogramming across breast cancer subtypes. (A) Distribution of activating mutations in PI3K pathway genes together with recurrent copy-number amplifications or deletions (GISTIC, q < 0.1) affecting lipid metabolic regulators across individual tumours. Homozygous focal copy-number events (<2 Mb) are shown. TP53 mutation status is included for reference. Overall, 73.9% (362/490) of tumours harboured at least one genomic alteration predicted to perturb lipid metabolism or PI3K signalling. **(B)** Hallmark pathway enrichment (ssGSEA) associated with individual genomic alterations across breast cancer subtypes. Circle size represents statistical significance and colour denotes effect size. **(C)** ERBB2 copy-number amplification is associated with significantly increased expression of adipogenesis and xenobiotic metabolism pathways. **(D)** Correlation matrix illustrating co-occurrence between somatic mutations and recurrent copy-number alterations affecting lipid metabolic pathways. Circle size indicates frequency of co-occurrence and colour indicates Pearson correlation coefficient.

Structural variation rewires regulatory landscapes *ERBB2* activation and genome instability

We identified 90,553 high-confidence structural variants (SV; 439 of 501 samples with consensus SVs from at least two independent algorithms were considered (Methods)) with subtype-specific distributions: highest in hormone-positive HER2+, followed by TNBC and hormone-negative HER2+, while hormone-positive HER2- had lowest SV burden in the cohort (Figure 3A, Supplementary Figures S18-S22). Recurrent intra- and inter-chromosomal rearrangements (>5% patients) affected key genes (Figure 3B); increased SV density and gain in super-enhancers (SE) in tumours compared to normal breast tissue (analysed using previous breast cancer data SEdb3.0[20]versus ENCODE[21]) and supported specific copy number amplification/deletions (middle and bottom panels; Figure 3B). These clustered at key oncogenic, chromatin-regulatory and lipid-metabolic loci, including *ERBB2, CCND1, ESR1, PTEN, ARID1B, PLD1* and *GATA3* (Figure 3B). Chromosomal rearrangements further revealed subtype-specificity – for instance in *TTC28, IKZF3* and *RUNX1* genes supporting context-specific influence of the rearrangements (Figure 3C).

HER2-positive breast cancer is characterized by substantial biological and therapeutic heterogeneity despite the central role of *ERBB2* amplification and/or overexpression[22]. At the *ERBB2* locus, we identified recurrent promoter hijacking (31.2%, 34 of 109 HER2+ tumours) where *IKZF3* regulatory elements are juxtaposed to *ERBB2* (illustrated in Figure 3D) consistent with an alternative mechanism *ERBB2* activation beyond focal amplification (Figure 3D-E). In support, from published data, increase in SE density at the *ERBB2* locus in breast tumours relative to normal breast tissue was evident (SEdb3.0 /ENCODE; Figure 3D lower panels) providing broader structural basis for increased *ERBB2* expression. In addition, increased expression of *ERBB2* in presence of *IKZF3* local translocation, and not otherwise, was clear (Figure 3F). These together implicate a mechanistic model consistent with structural

variation-mediated rewiring of the *ERBB2* locus, resulting in *ERBB2* upregulation in a subset of breast cancers.

Transcriptionally, lipid metabolism pathways were significantly affected by SVs within *MACROD2, RUNX1, ARID1B, RAD51B,* and *ASIC2,* particularly adipogenesis, cholesterol homeostasis, fatty acid metabolism, and PI3K-AKT-mTOR signalling across tumour subtypes (Figure 3G). *TTC28* translocation, noted in about 12% patients and across subtypes (Figure 3G), were significantly associated with increased ploidy (Figure 3H) implicating a role for these SVs in genome instability.

Recent CUBRICS [12] data identified recurrent extragenic SVs surrounding ERBB2, whereas our analysis identifies a distinct promoter-hijacking rearrangement involving IKZF3 regulatory elements in primary tumours. Elevated *ERBB2* expression at comparable copy number supports a copy-number-independent route to *ERBB2* activation.

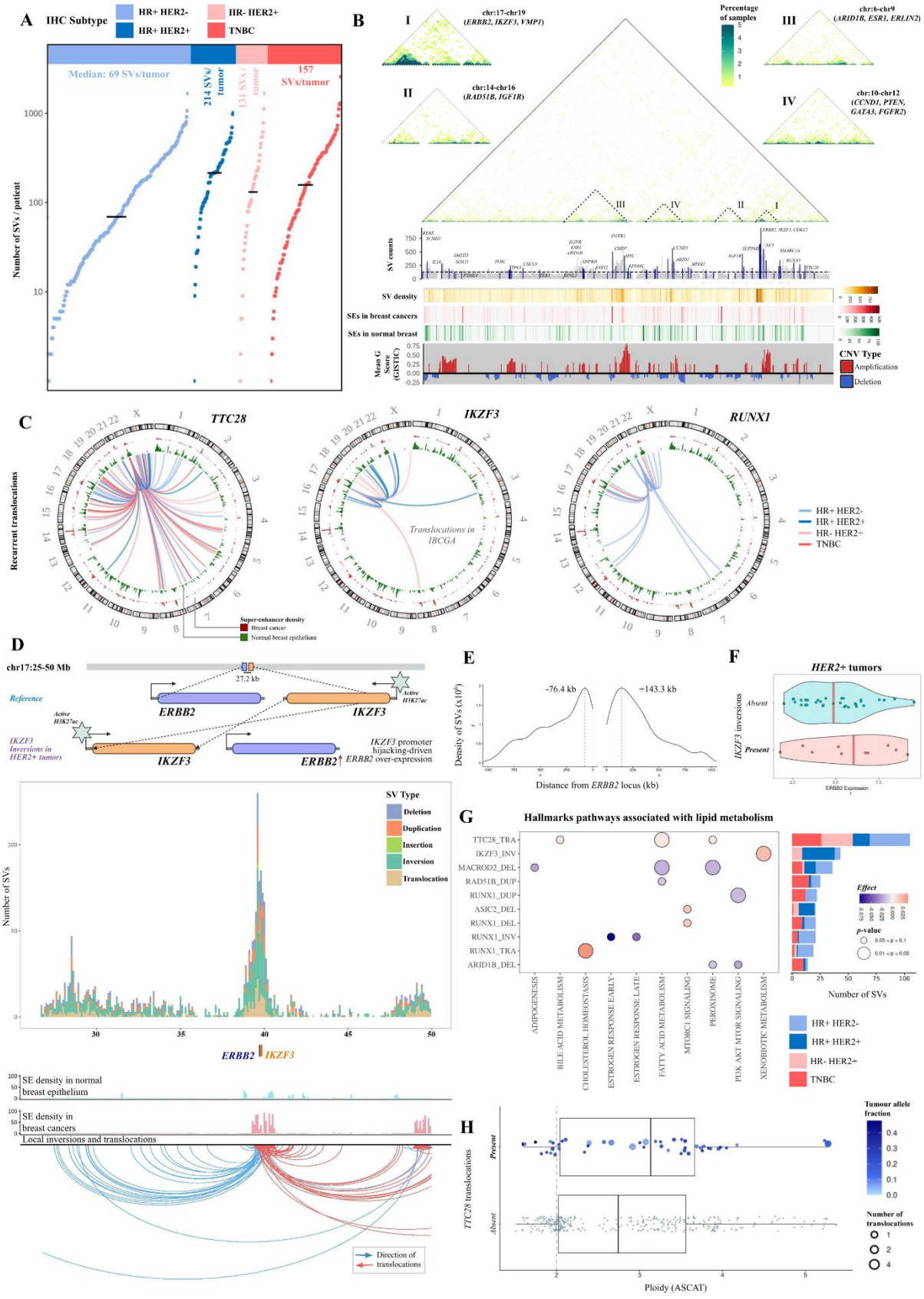

A
IHC Subtype
HR+ HER2-
HR+ HER2+
HR- HER2+
TNBC
Median: 69 SVs/tumor
214 SVs/ tumor
131 SVs/ tumor
157 SVs/tumor
Number of SVs / patient
B
chr:17-chr19 (ERBB2, IKZF3, VMP1)
chr:14-chr16 (RAD51B, IGF1R)
chr:6-chr9 (ARID1B, ESR1, ERLIN2)
chr:10-chr12 (CCND1, PTEN, GATA3, FGFR2)
Percentage of samples
SV counts
SV density
SEs in breast cancers
SEs in normal breast
Mean G Score (GISTIC)
CNV Type
Amplification
Deletion
C
TTC28
IKZF3
RUNX1
Recurrent translocations
Translocations in IBCGA
Super-enhancer density
Breast cancer
Normal breast epithelium
HR+ HER2-
HR+ HER2+
HR- HER2+
TNBC
D
chr17:25-50 Mb
27.2 kb
Reference
ERBB2
IKZF3
Active H3K27ac
IKZF3 Inversions in HER2+ tumors
IKZF3 promoter hijacking-driven ERBB2 over-expression
SV Type
Deletion
Duplication
Insertion
Inversion
Translocation
Number of SVs
SE density in normal breast epithelium
SE density in breast cancers
Local inversions and translocations
Direction of translocations
E
-76.4 kb
+143.3 kb
Density of SVs (x10⁶)
Distance from ERBB2 locus (kb)
F
HER2+ tumors
IKZF3 inversions
Absent
Present
ERBB2 Expression
G
Hallmarks pathways associated with lipid metabolism
TTC28_TRA
IKZF3_INV
MACROD2_DEL
RAD51B_DUP
RUNX1_DUP
ASIC2_DEL
RUNX1_DEL
RUNX1_INV
RUNX1_TRA
ARID1B_DEL
ADIPOGENESIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
MTORC1 SIGNALING
PEROXISOME
PI3K AKT MTOR SIGNALING
XENOBIOTIC METABOLISM
Effect
p-value
Number of SVs
HR+ HER2-
HR+ HER2+
HR- HER2+
TNBC
H
TTC28 translocations
Present
Absent
Tumour allele fraction
Number of translocations
Ploidy (ASCAT)

Figure 3: Structural variation remodels regulatory architecture and promotes ERBB2 activation. (A) Structural variant (SV) burden across breast cancer subtypes demonstrating highest burden in HR+HER2-positive tumours. (B) Genome-wide interaction map of recurrent (>5% patients) inter- and intra-chromosomal SVs (5 Mb window), illustrating genomic regions undergoing recurrent structural remodelling: Frequent events I – IV including key genes highlighted (top panel); SV density and super-enhancers (SEs) in tumour/normal breast (middle panels); recurrent CNVs identified by GISTIC (bottom panel). (C) Recurrent translocations originating from TTC28, IKZF3, and RUNX1 across breast cancer subtypes; colour of individual translocation events represents affected tumour subtype. (D) IKZF3-associated rearrangements affecting the *ERBB2* locus in 31.2% (34/109) of HER2+ tumours. Scheme showing the proposed promoter-hijacking configuration (top), distribution of local SV (middle panel), super-enhancer in normal/tumours, and local rearrangement profiles (lower panels). Chromosomal window indicated in top. (E) Density of SVs around the *ERBB2* locus (1 Mb) showing recurrent rearrangements involving super-enhancer regions. (F) ERBB2 expression according to presence or absence of IKZF3-associated rearrangements. (G) Association between recurrent structural variants and altered lipid metabolic pathway activity; two-tailed Wilcoxon rank-sum test; number of SV events across IHC subtypes (right). (H) TTC28 rearrangements associate with increased tumour ploidy, implicating structural variation in genomic instability; median tumour allele fraction and number of events in each patient with TTC28 translocation (top panel); ploidy inferred from ASCAT.

Mutational processes stratify tumours along the lipid metabolic axis

Recent breast-cancer studies demonstrate that APOBEC3-associated mutagenesis contributes to genomic alterations and treatment response[23]. In IBCGA, mutational signature analysis identified homologous recombination deficiency (HRD: SBS3) and APOBEC hyperactivity (SBS2, SBS13) as dominant processes (Figure 4A, Supplementary Figure S23, Supplementary Table S8-S10), consistent with earlier breast cancer WGS studies[1,24]. Further, HRD was relatively enriched in TNBCs, as noted before[25,26], while HR+ tumours showed more heterogeneity in HRD/APOBEC hyperactivity. Tumours with high HRD had more frequent: (i) rare germline alterations in *BRCA1* and *BRCA2*, as expected[26]; (ii) somatic mutations in *TP53*; and (iii) copy number deletions of *PTEN, CDKN2A* and *RB1* and amplifications of *MYC, CCND2, BIRC5* (Figure 4B).

Further, HRD stratified TNBCs into transcriptionally distinct states, with high-HRD tumours showing increased *KRAS* signalling, and reduced lipid- and hormone-related metabolic programmes, including adipogenesis, cholesterol homeostasis and oestrogen response (Figure 4C). While this supports HRD-associated transcriptional layers in TNBC, their direction and deployment were context-dependent. We also found five predominant SV signatures, including HRD (SV3; particularly in TNBC) consistent with mutational signatures, alongside SV2/SV4/SV5/SV7 of unknown aetiology (Figure 4D).

Synthetic-lethal relationship between *BRCA1/2* deficiency and PARP inhibition establish HRD as a genomic phenotype associated to therapeutic vulnerability[27]. In TNBC particularly, *BRCA1/2* alterations and broader 'BRCAness' phenotypes are associated with sensitivity to DNA-damaging therapies. Consistent with this, here we find HRD-associated mutagenesis was enriched in TNBC and associated with *BRCA1/2* alterations. In addition, our findings extend this biological framework suggesting the role of lipid metabolism in HRD-related genomic and transcriptional change.

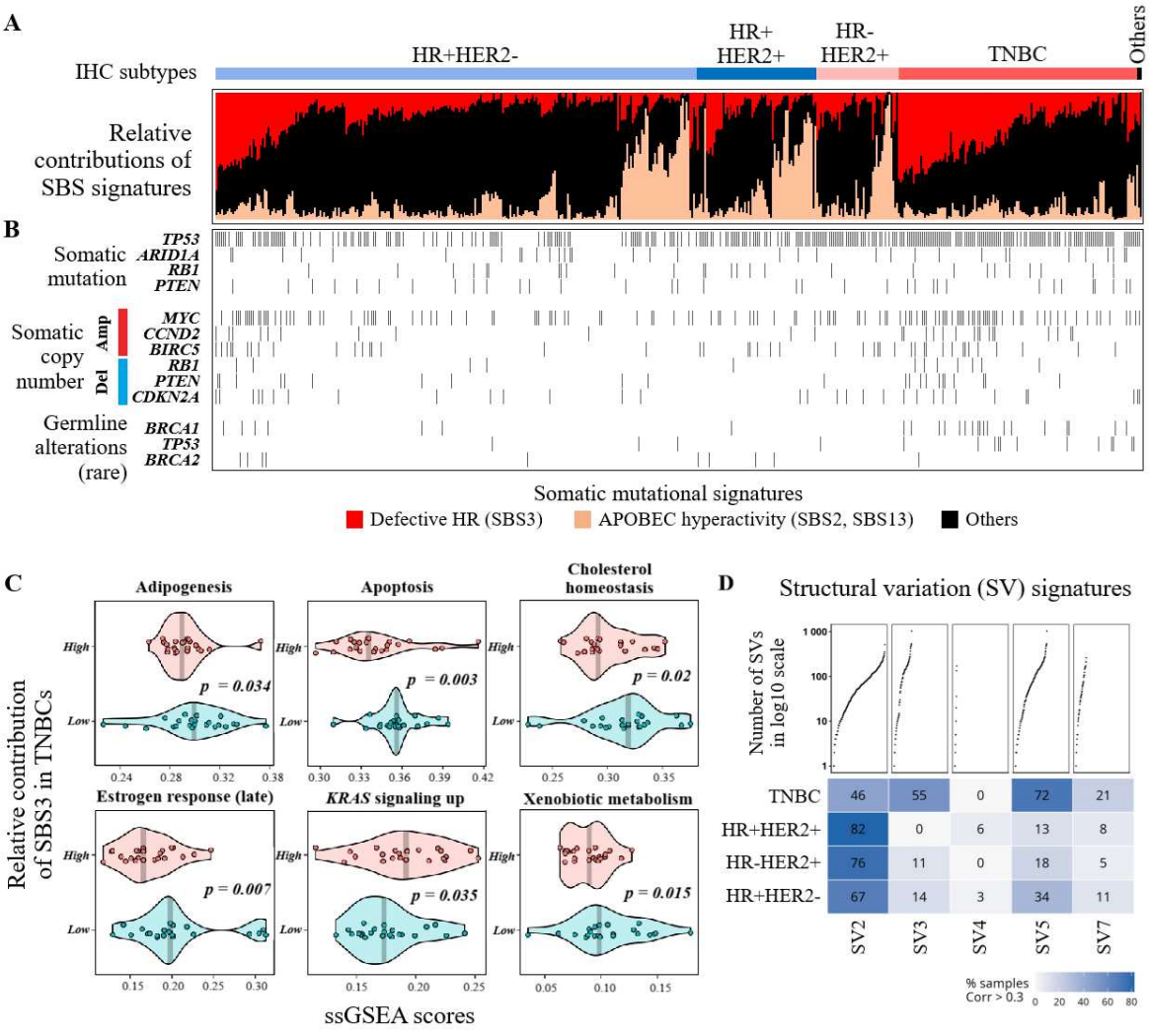


Figure 4: **Mutational processes identify HRD-associated lipid metabolic states.** (A) Somatic mutational signatures revealed two dominant mutational processes: APOBEC hyperactivity and homologous recombination deficiency (HRD), with HRD significantly enriched in TNBCs (median, 28.3%; $P < 5.043 \times 10^{-14}$, *t*-test). (B) HRD and APOBEC activity in relation to mutational and copy-number alterations in key genes; HRD was significantly associated with somatic *TP53* mutations ($P = 1.494 \times 10^{-10}$, *t*-test). HRD/APOBEC associated with germline alterations in *TP53* and *BRCA1/2* shown in bottom panels. (C) Tumours with genomic HRD showed altered expression of lipid- and hormone-

related pathways and *KRAS* signalling. (D) Five predominant somatic structural-variant signatures identified across breast cancer subtypes (COSMIC v104): SV3, non-clustered tandem duplications (>1 kb and <100 kb); SV2/SV4, non-clustered and clustered translocations; and SV5/SV7, deletions <1 kb and 10 kb–1 Mb.

Genomic alterations associated with recurrence align on shared programs

Hotspot mutations in *PIK3CA* or *TP53* (HR = 3.6; 95%CI: 1.16-11.5; p = 0.027; Figure 5A) or mutations in SF3B1 (HR = 4.6; 95%CI: 1.19-18.0; p = 0.027; Figure 5B) significantly associated with reduced recurrence-free survival of HR+ HER2- patients. The *PIK3CA* or *TP53* hotspot mutations significantly decreased recurrence-free survival probability in HER2+ patients (HR = 3.93; 95%CI: 1.11-13.89; p = 0.0289; Figure 5C). In TNBC, recurrence-free survival was low in presence of *PIK3CA* mutations (including non-hotspots) or *GATA3* copy number amplification (HR = 3.1; 95%CI: 1.23-7.8; p = 0.009; Figure 5D). The associations remained significant when analysed with clinical covariates (age and tumour stage) using multi-variable Cox-PH model.

On integrating transcriptome data, we noted recurrence-associated genomic alterations affected lipid metabolism pathways namely, adipogenesis, estrogen response, and fatty acid and xenobiotic metabolism, alongside other pathways including DNA repair, glycolysis, myogenesis, and angiogenesis, with somewhat distinct effect sizes across subtypes (Figure 5E). Among 36/501 (7%) patients with distant metastasis, *MYC* amplification (47% vs 23% in non-metastatic)**,** *ERBB2* amplification 42% vs 18%**,** *CDH1* mutation 14% vs 4%**,** and *TP53* hotspot mutation (R175H, 8% vs 2%) was significant ($p < 0.05$, Fisher's exact test; Supplementary Results and Supplementary Table S11).

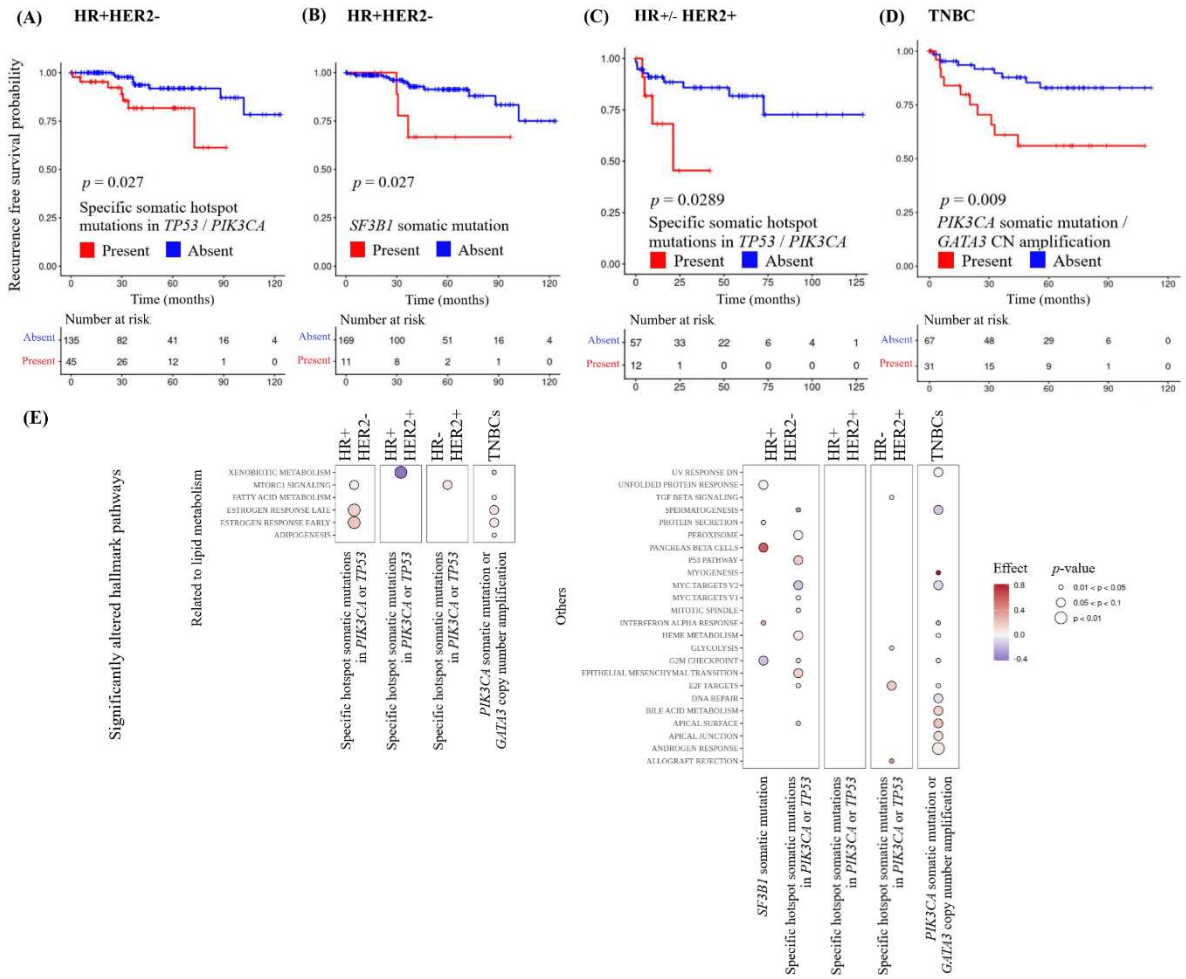


Figure 5. Genomic alterations associated with poor recurrence-free survival. (A–D) Associations between recurrent genomic alterations and recurrence-free survival: TP53 or PIK3CA hotspot mutations in HR+HER2− tumours (A), SF3B1 mutations in HR+HER2− tumours (B), TP53 or PIK3CA hotspot mutations in HER2+ tumours (C), and PIK3CA mutations or GATA3 copy-number amplification in TNBCs (D). Hazard ratios were estimated using Cox proportional-hazards models. (E) Pathway enrichment associated with recurrence-linked genomic alterations across breast cancer subtypes. Clinical follow-up: disease recurrence for this study was defined as local recurrence, distant recurrence or breast-cancer-related death; multivariable models were adjusted for age and tumour stage.

Transcriptome-based clusters reveal tumour states with metabolic lipid rewiring linked to clinical outcome

We next performed unsupervised clustering of 438 transcriptome profiles (following QC which excluded 17 out of 455 tumour profiles, Methods and Supplementary Table S12). This identified four tumour states: Basal-proliferative (BaPro), Luminal- proliferative (LumPro), Luminal-differentiated (LumDiff), and HER2-androgenic (HER2-AR) (Supplementary Figure S24; Supplementary Figure S25D, Supplementary Table S17). These clusters partially overlapped with PAM50[28] and IHC

classifications but identified biologically distinct tumour populations (Figure 6A; Supplementary Figure S25). For example, the HER2-AR transcriptomic state comprised tumours from both IHC HER2-positive and TNBC groups, while LumDiff was enriched for HR+ tumours. Supporting *de novo* transcriptome-based clustering further refined tumour stratification vis-à-vis reported states (Figure 6A). We next sought to understand if the newly identified cluster-based transcriptional states impact clinical outcomes. This revealed that the molecularly classified transcriptional groups independently show significant differences in recurrence-free survival (Figure 6B), with molecular clusters remaining prognostic after multivariable adjustment. (Supplementary Figure S25E)**.**

Further, centroid-based analysis of TCGA and METABRIC data (Methods) independently recapitulated all four states (BaPro, LumPro, LumDiff and HER2-AR) with distinct survival stratification, supporting their broader biological relevance (Supplementary Results; Supplementary Figure S26). Notably, HER2-AR consistently showed the poorest outcome in both METABRIC and TCGA, with the steepest decline in overall survival in TCGA (log-rank p = 0.017) and significantly worse disease-free and overall survival in METABRIC (log-rank $p < 0.001$ for both disease-free and overall survival), remaining an independent adverse prognostic factor on multivariable Cox regression after adjustment for age, T-stage, and nodal status (METABRIC OS: HR 1.78, 95% CI 1.42–2.23, $p < 0.001$).

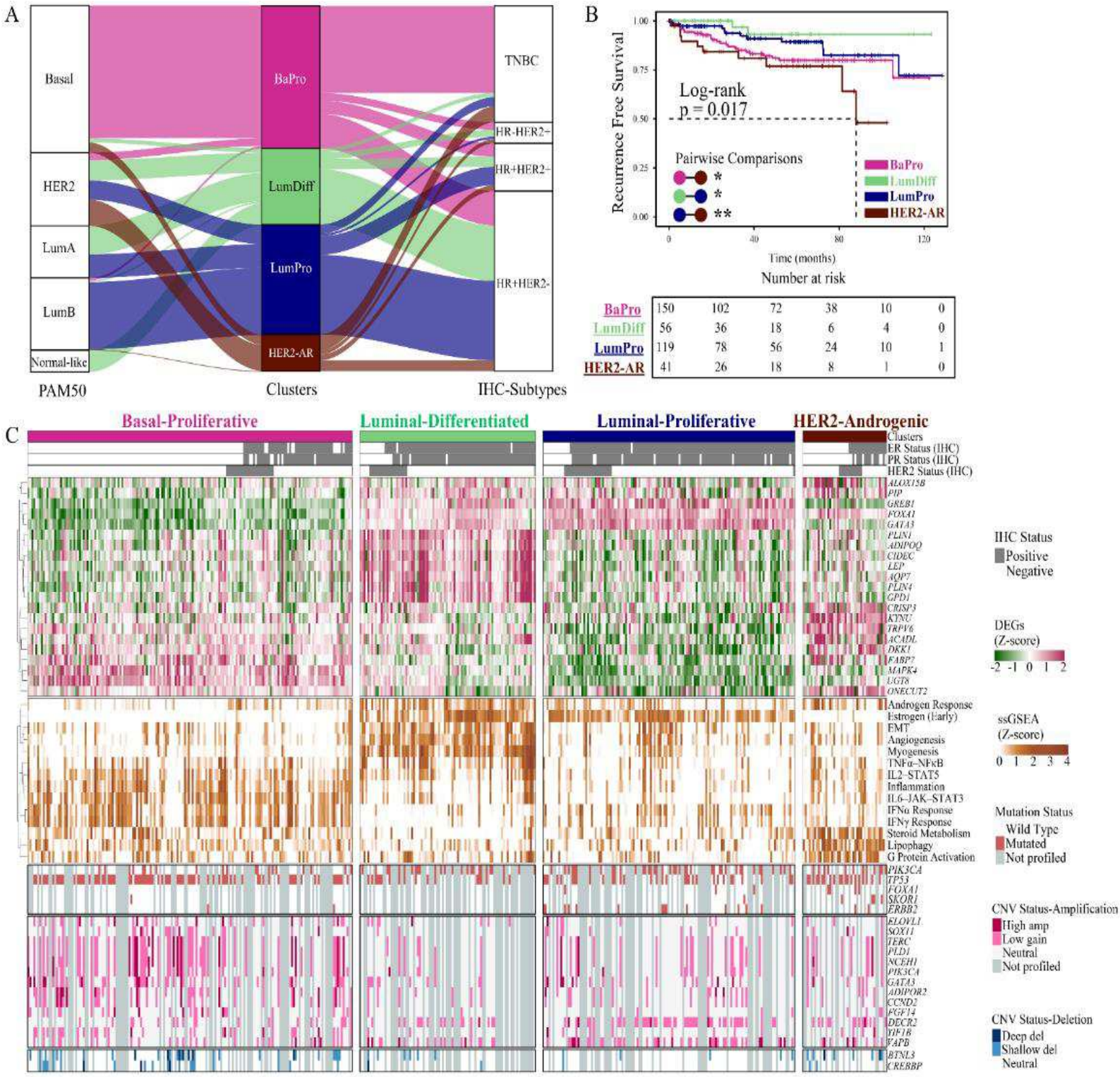
To understand the basis of these outcome differences, we examined the underlying transcriptomic programmes of each cluster (Figure 6C, Supplementary Table S19, S20). HER2-AR tumours were characterised by activation of androgen signalling (*PIP*, *CRISP3*), lipid metabolism (*ALOX15B*, *ACADL*), and WNT signalling (*DKK1*), with pathway enrichment confirming activation of steroid hormone metabolism, lipophagy, and G-protein-coupled receptor signalling (Supplementary Table S18) - pointing to a tumour state characterized by the association of androgen receptor activity and lipid metabolic reprogramming.

Consistent with this, HER2-AR tumours showed markedly suppressed IFNα and IFNγ response pathways, supporting attenuated innate immune signalling, which aligned with their comparatively low immune and leukocyte infiltration scores (Figure 6C) - together pointing to an immune-cold phenotype in this cluster. On the contrary, the BaPro tumours were marked by inflammatory and interferon signalling alongside higher enrichment scores for immune cells, reflecting an immune-active state. Both luminal clusters were supported by estrogen signalling. LumDiff additionally showed enhanced angiogenesis and EMT signatures alongside a strong adipogenic programme and was distinguished by the lowest leukocyte and macrophage infiltration across all four clusters, consistent with an immune-cold microenvironment (Figure 6C). The four clusters also carried distinct genomic signatures, including differential distribution of *TP53* and *PIK3CA* mutations and cluster-specific copy-number alterations (Figure 6C, bottom panels; Supplementary Table S19, S20). A comprehensive

characterisation of immune signalling, immune-cell composition, cytokine expression, and transcription factor activity across breast cancer subtypes and transcriptomic clusters is discussed in the supplementary results**.**

We noted that TNBCs were distributed across all the four clusters (top panels showing ER/PR/HER2 states, Figure 6C). This prompted us to do a TNBC-restricted clustering, which revealed a metabolically distinct fatty acid-enriched, LAR-like subgroup (n=39), of which 18 (46.2%) overlapped with the HER2-AR state identified in the whole-cohort consensus clustering. The LAR-like subgroup showed significantly worse prognosis than the Basal-like cluster in the TNBC cohort (log-rank p = 0.039; Supplementary Figure S27A-B, Supplementary Table S23). Together, these transcriptional states provide phenotypic resolution beyond existing clinical classifiers underscoring observations from earlier studies showing that IHC-defined categories and expression-derived intrinsic states are related but not identical, and their relationship to clinical classification is complex[29,30]. Findings herein, in addition, suggest lipid and metabolic rewiring as an important axis that influence outcome heterogeneity in breast cancer.

456



457

458

459

460

461

Figure 6 | Transcriptome-derived molecular states identify biologically distinct breast cancer subgroups associated with prognosis. Consensus clustering of transcriptomic profiles identified four molecular states: BaPro, LumPro, LumDiff and HER2-AR. (A) Alluvial plot showing relationships between PAM50 intrinsic subtypes, transcriptome-derived molecular clusters and immunohistochemistry-defined breast cancer subtypes (Supplementary Table S13). Band width is proportional to the number of tumours assigned to each category. (B) Kaplan–Meier analysis of recurrence-free survival according to transcriptomic cluster. Survival differences were assessed using the log-rank test; the HER2-AR cluster showed the poorest clinical outcome (log-rank $p = 0.017$). Pairwise comparisons are indicated. (C) Integrated heatmap showing receptor status, cluster-associated genes, pathway enrichment, recurrent somatic mutations and copy-number alterations. Gene-expression and pathway scores are displayed as z-scores. Samples lacking whole-genome sequencing data are indicated as "Not profiled".

HER2 (*ERBB2*)-associated transcriptional programs extend beyond IHC-defined HER2 status

Patients with HER2-low, including HER2-ultralow, breast cancers derive clinically meaningful benefit from HER2-targeted ADCs[31,32]. This has largely been attributed to efficient target engagement together with ADC-mediated bystander killing[31]. However, whether an underlying HER2-associated transcriptional program is active, and captures biologically distinct tumour states with clinical relevance, remains unclear.

To identify tumours with transcriptionally active HER2 (*ERBB2*), we analysed differentially expressed genes in tumours with high HER2 (IHC ≥3) against those with low or no HER2 protein (category-wise differential gene expression followed by concordance analysis, see Methods). This revealed a distinct 144-gene HER2-associated expression signature ($p = 0.05$; Supplementary Figure S28, Supplementary Table S21); including ten genes from the *ERBB2* locus on chromosome 17q12 genes, such as *GRB7* and *CDK12*, supporting relevance of the signature. As oestrogen receptor (ER) signalling directly influences *ERBB2* transcription[33], further analysis was restricted to 161 hormone-negative (HR−) transcriptomes.

The HER2-associated signature resolved HR- tumours into HER2-Enriched (HER2-E; n = 43) and Basal-Like (n = 118) transcriptional states (Figure 7A; Supplementary Table S22). Notably, 19.7% of TNBCs (27/137) were in the HER2-E group, supporting an *ERBB2*-associated transcriptional programme despite lacking HER2 protein expression by IHC. Additionally, expression analysis supported activation of an *ERBB2*-associated transcriptional signature in the HER2-E tumours - including GRB7-mediated ERBB2 signalling, constitutive *ERBB2* overexpression, and *ERBB2* ECD

mutant signalling. Together, pointing to a transcriptionally active HER2 axis that operates independently of IHC status (Figure 7A, bottom panel). Beyond *ERBB2* activity, HER2-E tumours were enriched for the LAR subtype compared with Basal-Like tumours (as per Lehmann classifications; TNBCtype-6: 32.6% vs. 4.2%; $p = 6.46 \times 10^{-6}$ and TNBCtype-4: 41.9% vs. 5.9%; $p = 2.61 \times 10^{-7}$; Figure 7B; Methods and Supplementary Tables S14-S15). To note, the HER2-AR cluster identified using all transcriptomes (Figure 6A) and the HER2-E LAR subtype (Figure 7B) from the HER2-associated signature represent related but independently derived transcriptional classifications.

At the tumour microenvironment level, HER2-E TNBCs had significantly lower immune scores than Basal-Like TNBCs ($p = 4.5\times10^{-6}$; Figure 7C), suggesting an immune cold tumour microenvironment characterised by reduced immune infiltration. This is in line with the complex and heterogenous relationship between tumour transcriptomes, molecular subtypes and the immune microenvironment in TNBCs, noted by others[30,34].

That a subset of IHC HER2-negative tumours harbour HER2-associated transcription is consistent with long-standing observations: molecular apocrine tumours were described two decades ago[35], a substantial proportion of LAR-subtype TNBCs classify as HER2-enriched by PAM50[36,37] and HER2-low TNBC has been reported as a biologically distinct entity within the LAR subtype[38]. Our finding is not the existence of this subset but its derivation — a 144-gene signature defined from HER2 IHC3+ tumours, independent of PAM50 and Lehmann classifiers — and its coupling to the lipid–*ERBB2* axis and to reduced immune infiltration, which together frame these tumours as a candidate stratum for HER2-directed ADC evaluation rather than a further transcriptional label.

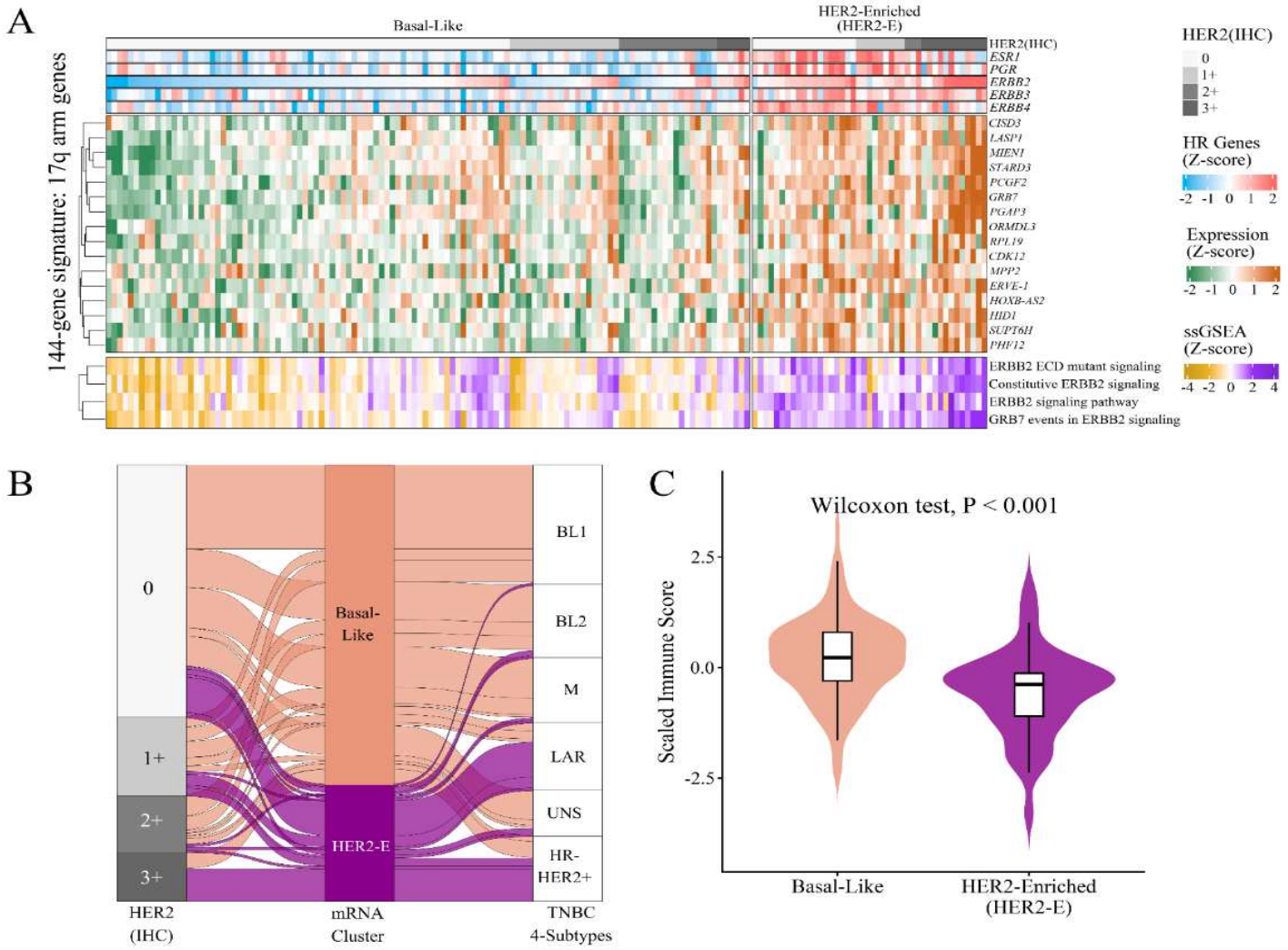


Figure 7. A HER2-associated transcriptional programme identifies an immune-cold, androgen receptor-enriched subset of HER2 IHC-negative breast cancers. A 144-gene HER2-associated expression signature derived from HER2 IHC3+ tumours identifies transcriptionally HER2-active tumours independently of HER2 protein expression. **(A)** Heatmap of HR-negative tumours clustered using the 144-gene HER2-expression signature, resolving two transcriptional states (HER2-Enriched and Basal-Like). Lower panels show ssGSEA enrichment scores for HER2-associated transcriptional programmes, including GRB7-mediated ERBB2 signalling, constitutive ERBB2 overexpression and ERBB2 extracellular-domain signalling. Annotation tracks on top panels display IHC HER2 score (0–3+) and HR-family receptor gene expression (*ESR1, PGR*, *ERBB2*, ERBB3, ERBB4). **(B)** Association of HER2-Enriched and Basal-Like transcriptional states with Lehmann TNBCtype-4 molecular subtypes, with unassigned samples and HR-HER2+ tumours shown separately. Correspondence with

IHC HER2 score shown in left. Statistical significance was assessed using Fisher's exact test. Band width is proportional to sample count. **(C)** Comparison of tumour immune scores between HER2-Enriched and Basal-Like tumours. HER2-Enriched tumours exhibit significantly reduced immune infiltration ($p = 4.5\times10^{-6}$, Wilcoxon rank-sum test).

Discussion

Breast-cancer classification has progressively evolved from receptor-defined categories to intrinsic expression states and, more recently, to integrated views of genomic, transcriptomic and whole-genome variation[2,11,12]. Here, we find that tumour behaviour cannot be understood solely from individual genomic alterations, but instead emerges from their integration with oncogenic signalling, metabolic and transcriptional states, immune context and clinical phenotype.

PI3K–AKT and *ERBB*-family signalling are established determinants of breast cancer biology and therapeutic response, including resistance to HER2-directed treatment[39]. Mutational activation of PI3K–AKT has been implicated in trastuzumab resistance, while HER2-driven tumours can remain dependent on HER2–PI3K signalling after therapeutic perturbation[40]. Our findings suggest that genomic alteration of lipid-metabolic regulators is coupled to this oncogenic signalling, a relationship also observed in independent TCGA-BRCA and CPTAC cohorts. Rather than viewing lipid metabolism as a downstream consequence[41] , these observations raise the possibility of reciprocal coupling between metabolic state and oncogenic signalling.

APOBEC-associated mutagenesis and HRD were prominent mutational processes in IBCGA, with HRD particularly enriched in TNBC. These findings are consistent with established evidence linking APOBEC3 activity to breast-cancer evolution and treatment resistance and BRCA1/2-associated HRD and broader 'BRCAness' phenotypes to TNBC biology and therapeutic vulnerability[23,27]. Notably, HRD was also associated with distinct lipid- and hormone-related transcriptional programmes, suggesting that DNA-repair phenotypes may intersect with lipid metabolism.

Structural variation provided an additional route to oncogenic activation. HER2-positive breast cancer is characterized by substantial biological and therapeutic heterogeneity despite the central role of *ERBB2* amplification and/or overexpression[22,42]. In IBCGA, recurrent *IKZF3*-associated rearrangements were linked to increased *ERBB2* expression. This complements recent evidence for enhancer-mediated *ERBB2* activation[12], while defining a distinct promoter-hijacking architecture involving *IKZF3*.

Transcriptomic analysis further indicated that genomic diversity is organized into biologically meaningful tumour states that partially correspond to conventional classifications, consistent with previous studies[2,43]. The HER2-AR transcriptional state, from whole-cohort unsupervised clustering,

showed the poorest outcome in IBCGA (Figure 6A-B), and was reproducible independently across TCGA and METABRIC supporting its broader biological relevance. The resulting states therefore represent a progression from classification by receptor or intrinsic expression phenotype towards integrated tumour states defined by coordinated metabolic, signalling and immune programmes. In particular, the HER2-E associated state within IHC-negative TNBC (Figure 7) highlights the limitation of defining HER2 biology solely by protein expression and raises the possibility that transcriptional state may capture clinically relevant heterogeneity not resolved by conventional receptor classification. The immune associations are consistent with previous studies of TNBC tumour–immune heterogeneity and recent spatial analyses linking immune microenvironments to prognosis[44,45].

The population context of IBCGA adds an important dimension. WGS in East Asian breast cancer has established that ancestry-associated cohorts reveal distinct mutational and structural spectra[12]. The four candidate drivers reported here (*ISM2*, *QRICH2*, *CBX3*, *OTOP1*) were absent as significantly mutated genes across 35 cohorts totalling 17,708 tumours, indicating that the South Asian genomic spectrum contributes non-overlapping candidates rather than replicating an existing catalogue, in addition to linking genomic diversity to metabolic and transcriptional states.

Several limitations should be considered. The genomic and transcriptomic associations do not establish causality, and functional studies are needed to test the effects of lipid-metabolic alterations, *IKZF3–ERBB2* rearrangement and candidate drivers. The prognostic states require prospective validation, and immune infiltration was inferred from bulk transcriptomic rather than spatial or cellular measurements.

Together, these findings identify convergent lipid–ERBB2 regulatory states as biologically meaningful features of breast-cancer heterogeneity. Their association with oncogenic signalling, immune context and clinical outcome provides a rationale for testing their functional and therapeutic consequences.

Methods

Patient recruitment and sample collection

This ambispective study included both prospectively and retrospectively collected breast tumour samples. Eligible participants were treatment-naïve women aged ≥18 years with a confirmed diagnosis of breast carcinoma. Retrospective cases were selected from hospital biorepositories with ≥3 years of clinical follow-up (IRB approvals: RGCIRC/IRB-BHR/41/2022; KVV/IEC/01/2025). Prospective recruitment was conducted across multiple hospitals in India (Supplementary Figure S1). All participants provided written informed consent, and the study was approved by the Institutional Human Ethics Committee of the CSIR–Institute of Genomics and Integrative Biology (CSIR-IGIB/IHEC/2020-21/02).

Comprehensive clinico-pathological data, including AJCC stage, ER/PR/HER2 status, nodal involvement, menopausal status, Ki-67 index, and family history were recorded (Supplementary Figure S2; Supplementary Table S1a, S24). HER2 IHC-equivocal cases were resolved by fluorescent in situ hybridisation. Tumours were classified into HR+HER2−, HR+HER2+, HR−HER2+, and triple-negative breast cancer (TNBC). All patients had a minimum of 3 years of follow-up documenting treatment, recurrence, and survival. Disease recurrence was defined as local recurrence, distant recurrence or breast-cancer-related death. Tumour biopsies were snap-frozen or preserved in RNAlater at 2–8 °C and stored at −80 °C. Matched normal samples consisted of adjacent normal tissue (retrospective) or peripheral blood (prospective). All procedures adhered to institutional and national ethical guidelines.

DNA and RNA extraction and sequencing

High-quality DNA and RNA were co-extracted from fresh-frozen tumour and matched normal tissues using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Nucleic acid quality was assessed by spectrophotometry (A260/A280), fluorimetry, and RNA Integrity Number (RIN). Libraries for whole-genome sequencing (WGS) were prepared using the TruSeq DNA PCR-Free Kit (Illumina) and sequenced on the NovaSeq 6000 platform (Illumina) to a mean depth of ~40× for tumour and normal samples (S4 reagent kit).

Whole-transcriptome sequencing (WTS) libraries were generated following rRNA depletion (NEBNext rRNA Depletion Kit v2) and directional library preparation (NEBNext Ultra II Directional RNA Library Prep Kit). Libraries were validated using the Agilent TapeStation (DNA 1000 ScreenTape) and Qubit dsDNA HS Kit, and sequenced (2×150 bp) on the NovaSeq 6000 (S1 reagent kit).

Processing and analysis of WGS data to identify germline and somatic mutations and copy number alterations

Whole-genome sequencing (40× tumour–normal pairs) underwent standard Illumina demultiplexing, quality control, and alignment to GRCh38, followed by duplicate marking, recalibration, and contamination assessment. Germline variants were jointly genotyped and clinically annotated, and pathogenic germline alterations were identified using ACMG criteria. Somatic variants were called

using paired tumour–normal analysis, stringently filtered, and annotated with multiple pathogenicity and domain-level predictors, incorporating population-specific controls. Significantly mutated genes and mutation rates were identified using MutSig2CV[14], with manual IGV review of all SMG variants. Copy-number alterations were inferred using ASCAT[46]and GISTIC[47] , and mutual exclusivity/co-occurrence patterns were assessed with maftools. Selection pressures on driver genes were evaluated using dndscv.

Characterization of Somatic mutational and copy-number signatures

We have utilized accelerated SignatureAnalyzer[48] with COSMICv3-reference to identify single base substitution (SBS) signatures from SNVs through the extraction of substitutions based on tri-nucleotide content, followed by cosine similarity calculation. Similarly, somatic insertions and deletions were used for the detection of indel (ID) signatures. The results from signature fitting through SignatureAnalyzer were verified through an orthogonal method using SigProfilerExtractor[49] .From the genome-wide somatic copy number profiles (generated through AscatNGS: 4.5.0), we have detected copy-number signatures as described by Steele et al. 2022[50].

Protein Structure-based mutation mapping

Mutation data were converted to MAF format and analyzed using maftools to identify high-mutation-load genes. Corresponding TCGA-BRCA mutation data (from GDC) were compared against IBCGA, and unique mutations were retained for the top four driver genes (TP53, PIK3CA, KMT2C, MAP3K1). These were annotated using ProtVar for residue conservation, protein stability (FoldX v5.0), pathogenicity (AlphaMissense, CADD Phred, ESM-1b), structural domain localization (Pfam), and disease associations (dbSNP, ClinVar, Ensembl, UniProt).

For PIK3CA specifically, five mutations (V344G, N345S, E418K, E453Q, Q546E) were identified at the PIK3R1-interacting interface. Using the crystal structure of the p110α/niSH2–p85α complex (PDB: 5XGH), mutant complexes were modeled and docked with HADDOCK (100+ structures generated), then compared to wildtype to assess changes in hydrogen bonding and salt bridges via an in-house PyMOL script. Interface residues were identified from modeled PIK3CA–PIK3R1/2/3 complexes, with pDockQ used as a model-confidence metric; since PDB 5XGH represents the p110α/niSH2–p85α heterodimer, the PIK3R1 interface is directly captured in the experimental structure.

Prediction of neoantigens from somatic mutation data

Germline HLA class I alleles for each patient were inferred computationally from whole-genome sequencing data of matched normal tissue. Details in the supplementary methods section.

Structural variant detection

Structural variants (SVs) were called from WGS data of 501 tumour-normal pairs using three tools—GRIDSS (v2.13.2), Manta (v1.6.0), and SvABA (v1.1.0)—each leveraging distinct signals (discordant/split reads, soft-clipping, or local assembly). Manta and SvABA were run in chromosome-wise, chunked mode for efficiency, while GRIDSS was run genome-wide to ensure accurate translocation detection. Calls were merged into a consensus set using Jasmine v1.1.5 (500 bp breakpoint proximity threshold), retaining variants supported by ≥2–3 callers; GRIDSS-reported translocations were manually validated in the merged set. This yielded high-confidence consensus SVs for 439 samples, which were annotated with AnnotSV v3.4.4 and restricted to protein-coding regions (±2kb) using bedtools. Custom in-house scripts handled merging and downstream processing of caller outputs.

RNA-seq processing and quantification

Adapter trimming and quality filtering were performed using Trimmomatic v0.39. rRNA contamination was assessed using bbsplit.sh (BBMap v38.18) and removed using bbduk.sh. Clean reads were aligned to GRCh38.p14 using STAR v2.7.10b. QC metrics (≥75% uniquely mapped reads; >10 million reads per sample) were summarised using MultiQC v1.27. Gene-level quantification was performed using featureCounts (Subread v2.0.6), and transcript-level quantification using RSEM v1.2.28 with GENCODE annotations. Expression values were reported as TPM and FPKM.

Normalization and batch correction

The initial dataset comprised RNAseq data from 486 subjects across 28 sequencing batches (64 tumour-normal paired, 402 tumours only and 20 normals without matched counterparts, Supplementary Figure S1 and Supplementary Table S12). Batches with <3 samples and genes expressed in <10% of samples, yielded 455 samples. Library size and tumour purity (tidyestimate v1.1.1) were computed. VariancePartition v1.32.5 was used to quantify technical and biological variance. Unwanted variation was removed using the RUVIII-PRPS framework, with negative control genes selected based on low biological signal and high technical variance. Normalization performance was assessed using PCA and variance decomposition with RUVIII parameters and negative control genes summarized in Supplementary Table S13 (Supplementary Figure S29).

Integration of gene expression and genomic data

Single-sample geneset enrichment analysis (ssGSEA) scores for MSigDB hallmark pathways were calculated with the TPM values as input using the package 'GSVA' [51]. Pathways that were significantly associated with *ERBB2* amplifications were identified by Wilcoxon's rank sum test between samples with and without *ERBB2* amplifications for each hallmark pathway. Subtype-specific enrichment of hallmark pathways for lipid-related alterations, as well as recurrence signature were identified by Wilcoxon's rank sum test of ssGSEA scores between patients with and without genomic features within each IHC subtype.

Validation of genomic lipid alteration features in external datasets

Mutation, CNV, and PAM50 subtype information for TCGA-BRCA and CPTAC samples were obtained from 'cBioPortal'. Unstranded TPM values for gene expression were obtained from GDC.

Samples-wise ssGSEA scores for hallmark pathways were estimated using 'GSVA'. Genomic lipid-alteration associated pathways were identified by Wilcoxon's rank sum tests.

Unsupervised Consensus Clustering and Downstream Analyses

To stratify tumours based on transcriptional profiles, consensus clustering was performed on 454 samples (1 sample was removed from 455 tumour profiles because of inconclusive IHC (ER/PR/HER2) classification) using the top 5,000 most variable genes, implemented in the ConsensusClusterPlus R package (v1.66.0). Clustering used hierarchical agglomeration with Euclidean distance and Ward.D2 linkage, 1,000 resamplings, and 80% item resampling. The optimal number of clusters (k) was selected using the delta area plot, CDF curve, and average silhouette width. Samples with negative silhouette scores (16 tumour profiles) were excluded, and the clustering procedure was repeated until all remaining samples, that is 438 transcriptome profiles exhibited positive silhouette widths.

We assessed pathway activity by applying single-sample Gene Set Enrichment Analysis (ssGSEA) to MSigDB Hallmark and C2–Reactome gene sets using the GSVA R package (v1.50.5). Lymphocyte infiltration and macrophage/CSF1 response were additionally scored per sample using the Thorsson et al. immune signature gene sets. Cluster-specific genomic alterations, including somatic mutations and copy number variations (CNVs), were analysed using Chi-square (χ^2) tests implemented in the stats R package (v4.3.3). To explore transcriptional heterogeneity within TNBCs, consensus clustering was independently performed on the clinically defined cohort of 137 TNBC samples using the top 3500 variable genes. Clustering was conducted using k-means hierarchical agglomeration, Euclidean distance, Ward.D2 linkage, applying 1,000 resampling iterations with 80% item resampling. Downstream pathway analysis was done using the Hallmark gene set, and survival analyses were then carried out as described above.

Development of a Specificity-Based Centroid Classifier for the Four Identified Transcriptional Clusters in the IBCGA Cohort

The transcriptomic clusters were originally identified using unsupervised consensus clustering of the top 5,000 most variable genes. To develop a robust classifier for assigning new samples to these clusters, a specificity-based centroid classifier was subsequently constructed using the normalized expression matrix of all protein-coding genes rather than being restricted to the 5,000 variable genes. Gene expression values were standardized by calculating gene-wise z-scores across all samples. For each cluster, the mean expression of every gene was computed, and a specificity score was defined as the difference between the mean expression in the target cluster and the highest mean expression observed across the remaining clusters. Genes with a mean z-score >0.5 in the target cluster were ranked according to this specificity score, and the top 25 genes from each of the four clusters were selected, yielding a 100-gene classifier signature. Cluster centroids were then generated as the average z-score expression profile of these signature genes within each cluster. During classification, samples were assigned to the cluster whose centroid exhibited the highest Pearson correlation with the sample's expression profile. The genes displayed in Figure 6C represent a subset of differentially expressed

marker genes selected to illustrate the biological characteristics of each cluster and are intended for visualization only; they are distinct from the complete 100-gene classifier signature used for centroid construction.

Independent validation of the IBCGA-derived four-cluster centroid classifier was performed in two external, publicly available cohorts. For TCGA-BRCA[52], raw RNA-seq counts and associated clinical/PAM50 annotations were retrieved using TCGA biolinks (v2.36.0), collapsed to unique gene symbols by maximum-variance selection, and normalized using TMM followed by voom transformation (edgeR/limma); for METABRIC[53], pre-processed log2-normalized microarray expression values and clinical annotations were obtained from cBioPortal. In both cohorts, expression was restricted to the 100 signature genes shared with the IBCGA centroid matrix, z-scored per gene across samples, and each sample was assigned to the cluster (BaPro, HER2-AR, LumDiff, or LumPro) whose centroid yielded the maximum Pearson correlation with the sample's expression profile. Cluster assignments were evaluated against PAM50 subtype (TCGA) and IHC-defined clinical subtype (both cohorts, derived from ER/PR/HER2 status) and against clinical outcome by Kaplan–Meier analysis (RFS/OS in METABRIC; DFS/OS in TCGA) multivariable in TCGA; multivariable in METABRIC, adjusting for age, tumour stage, and lymph node status, with LumDiff as reference), with hazard ratios summarized via forest plots.

HER2-associated transcriptional program: stratification and transcriptomic characterization

Hormone receptor-negative (HR−) tumours were stratified into HER2 IHC categories (0, 1+, 2+, and 3+) according to routine clinical pathology scoring. *ERBB2* expression was quantified from RUVIII-normalized RNA-seq data. To identify a HER2-associated transcriptional program independent of IHC status, a stepwise differential expression analysis between groups was first performed as follows. First, HER2- IHC 0, IHC 1+ and IHC 2+ tumour categories were independently divided into *ERBB2*-high or *ERBB2*-low using the within-category median *ERBB2* expression, followed by differential gene expression (DEG) analysis within each IHC category. For HER2-IHC 0 cases, high *ERBB2*-expressing (n = 46) was compared with low *ERBB2*-expressing (n= 47) tumours. This identified 2,669/2,990 up/downregulated genes ($p < 0.05$). Similar comparison of high vs low *ERBB2*-expressing tumours with HER2-IHC 1+ (n = 14 vs. 15 cases) or IHC 2+ (n = 10 vs. 11 cases) identified 896/757 and 451/581 up/downregulated genes, respectively ($p < 0.05$). Second, DEG analysis between HER2-IHC 3+ cases (n = 18) vs IHC 0 (n = 93) tumours identified 1,076 upregulated and 2,396 downregulated genes ($p < 0.05$). DEGs between IHC 3+ and IHC 0 were retained if they were concordant with at least two of the three (IHC 0, IHC 1+ or IHC 2+) *ERBB2*-high/low DEGs. This gave a 144-gene DEG set which was designated as the HER2-associated signature (57 upregulated, 87 downregulated; Supplementary Table S21). This HER2-associated signature was used for consensus clustering of hormone-negative (HR−) tumours using ConsensusClusterPlus (v1.66.0; hierarchical agglomeration, Euclidean distance, Ward.D2 linkage, 1,000 resamplings, 80% item resampling).

Biological differences between the resulting transcriptional states were assessed by ssGSEA using MSigDB Hallmark and C2 gene sets with GSVA (v1.50.5) (Figure 7A). Lymphocyte infiltration and macrophage/CSF1-response programmes were scored using Thorsson et al. immune signatures, and tumour purity-adjusted immune infiltration was estimated using ESTIMATE on FPKM expression matrices. Associations with the previously identified HER2-androgenic (HER2-AR) state and TNBCtype-4/TNBCtype-6 classifications were assessed using Fisher's exact test (Figure 7B).

Transcriptomic subtyping

Transcriptomic subtyping was performed using log2-transformed, normalized FPKM expression data. Intrinsic molecular subtypes were assigned using the PAM50 gene-expression signature and the AIMS algorithm (AIMS v1.32.0). TNBC samples were further classified using the Lehmann TNBCtype framework. The original TNBCtype-6 classifier identified six transcriptional subtypes: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR) subtypes[37] (Supplementary Figure S30). Following subsequent refinement of the classifier, the TNBCtype-4 framework was applied to assign samples to the four tumour-intrinsic subtypes BL1, BL2, M, and LAR, with IM and MSL excluded because these subtypes were found to predominantly capture immune and stromal signals, respectively[54].

In parallel, the Burstein classification was applied to the same log2-transformed, normalized FPKM expression profiles to derive four molecular subtypes: luminal androgen receptor (LAR), mesenchymal (MES), basal-like immune-suppressed (BLIS), and basal-like immune-activated (BLIA)[55]. These complementary classification frameworks were used to characterize distinct dimensions of transcriptomic heterogeneity within TNBC.

Data availability

Processed and raw sequencing data from the Indian Breast Cancer Genome Atlas (IBCGA) will be accessible through the IBCGA data portal (https://csir-ibcga.igib.res.in/). The portal provides public access to processed datasets with gene-wise query capabilities and interactive visualisation tools. Raw sequence data are available under controlled access via the same portal. Researchers are required to submit a data access request outlining the proposed research to the Data Management Group. Raw data access is possible following review of the proposal keeping with institutional and regulatory guidelines.

Acknowledgement

We acknowledge the Council of Scientific and Industrial Research (CSIR), India, for funding the CSIR–Indian Breast Cancer Genome Atlas (IBCGA) project (HCP0043). We thank all collaborating

institutions and partner hospitals for their continued support. We especially thank all collaborating clinicians for their critical inputs and support staff for their assistance in administering the project across multiple sites. We remain especially grateful to the patients and their families for participating in the project. We also acknowledge the guidance received from the Project Monitoring Committee and CSIR for project management.

Funding: Council of Scientific and Industrial Research, GOI

Conflict of Interest

The authors declare that they have no conflicts of interest that could have influenced the work reported in this manuscript.

Author Contribution – IBCGA Consortium

References

1. **Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54 (2016).**
2. **Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70 (2012).**
3. **Rony, R., Deng, S., Yang, S., Doig, K. & Goode, D. L. Putative breast cancer risk variants from populations of South Asian ancestry are under-represented in public variant classification databases. *Breast Cancer Res.* 28, (2026).**
4. **Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 12, 31–46 (2022).**
5. **Broadfield, L. A., Pane, A. A., Talebi, A., Swinnen, J. V. & Fendt, S. M. Lipid metabolism in cancer: New perspectives and emerging mechanisms. *Dev. Cell* 56, 1363–1393 (2021).**
6. **Wan, M. *et al.* Lipid metabolic reprograming: the unsung hero in breast cancer progression and tumour microenvironment. *Mol. Cancer* 24, (2025).**
7. **Vasan, N. & Cantley, L. C. At a crossroads: how to translate the roles of PI3K in oncogenic and metabolic signalling into improvements in cancer therapy. *Nature Reviews Clinical Oncology 2022 19:7* 19, 471–485 (2022).**
8. **Hoxhaj, G. & Manning, B. D. The PI3K–AKT network at the interface of oncogenic signalling and cancer metabolism. *Nature Reviews Cancer 2019 20:2* 20, 74–88 (2019).**
9. **Ligorio, F. *et al.* Targeting lipid metabolism is an emerging strategy to enhance the efficacy of anti-HER2 therapies in HER2-positive breast cancer. *Cancer Lett.* 511, 77–87 (2021).**
10. **Sreenivasan, V. K. A., Yumiceba, V. & Spielmann, M. Structural variants in the 3D genome as drivers of disease. *Nat. Rev. Genet.* 26, 742–760 (2025).**
11. **Jiang, Y. Z. *et al.* Integrated multiomic profiling of breast cancer in the Chinese population reveals patient stratification and therapeutic vulnerabilities. *Nat. Cancer* 5, 673–690 (2024).**
12. **Kim, R. *et al.* Whole-genome landscapes of 1,364 breast cancers. *Nature 2025 649:8099* 649, 1282–1291 (2025).**
13. **Thorn, G. J. *et al.* The clinical and molecular landscape of breast cancer in women of African and South Asian ancestry. *Nature Communications 2025 16:1* 16, 4237- (2025).**
14. **Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501 (2014).**
15. **Huang, C. H. *et al.* The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. *Science* 318, 1744–1748 (2007).**
16. **Koundouros, N. *et al.* Metabolic Fingerprinting Links Oncogenic PIK3CA with Enhanced Arachidonic Acid-Derived Eicosanoids. *Cell* 181, 1596-1611.e27 (2020).**
17. **Garcia-Cao, I. *et al.* Systemic elevation of PTEN induces a tumour-suppressive metabolic state. *Cell* 149, 49–62 (2012).**
18. **Song, X. Q. *et al.* Copy number amplification of FLAD1 promotes the progression of triple-negative breast cancer through lipid metabolism. *Nat. Commun.* 16, (2025).**
19. **Chen, Y. Y., Ge, J. Y., Zhu, S. Y., Shao, Z. M. & Yu, K. Da. Copy number amplification of ENSA promotes the progression of triple-negative breast cancer via cholesterol biosynthesis. *Nat. Commun.* 13, (2022).**
20. **Song, S. *et al.* SEdb 3.0: a comprehensive super-enhancer database across multiple species. *Nucleic Acids Res.* 54, D322–D333 (2026).**
21. **Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46, D794–D801 (2018).**
22. **Marra, A., Chandarlapaty, S. & Modi, S. Management of patients with advanced-stage HER2-positive breast cancer: current evidence and future perspectives. *Nat. Rev. Clin. Oncol.* 21, 185–202 (2024).**
23. **Gupta, A. *et al.* APOBEC3 mutagenesis drives therapy resistance in breast cancer. *Nat. Genet.* 57, 1452–1462 (2025).**
24. **Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* 500, 415–421 (2013).**

25. **Watkins, J. *et al.* Genomic Complexity Profiling Reveals That HORMAD1 Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. *Cancer Discov.* 5, 488–505 (2015).**
26. **Tutt, A. *et al.* A randomised phase III trial of carboplatin compared with docetaxel in BRCA1/2 mutated and pre-specified triple negative breast cancer "BRCAness" subgroups: the TNT Trial. *Nat. Med.* 24, 628 (2018).**
27. **Lord, C. J., Tutt, A. N. J. & Ashworth, A. Two decades of PARP inhibitor synthetic lethality in cancer. *Nature* 653, 41–51 (2026).**
28. **Bernard, P. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27, 1160–1167 (2009).**
29. **Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929–944 (2014).**
30. **Fox, N. S., Haider, S., Harris, A. L. & Boutros, P. C. Landscape of transcriptomic interactions between breast cancer and its microenvironment. *Nat. Commun.* 10, (2019).**
31. **Bardia, A., Schwaederle, M. & Curigliano, G. Trastuzumab Deruxtecan in Low or Ultralow HER2 Metastatic Breast Cancer. Reply. *N. Engl. J. Med.* 392, 830–831 (2025).**
32. **Modi, S. *et al.* Trastuzumab deruxtecan in HER2-low metastatic breast cancer: long-term survival analysis of the randomized, phase 3 DESTINY-Breast04 trial. *Nature Medicine 2025 31:12* 31, 4205–4213 (2025).**
33. **Hurtado, A. *et al.* Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* 456, 663–666 (2008).**
34. **Tovey, H. *et al.* Integrated Multimodal Analyses of DNA Damage Response and Immune Markers as Predictors of Response in Metastatic Triple-Negative Breast Cancer in the TNT Trial (NCT00532727). *Clin. Cancer Res.* 29, 3691–3705 (2023).**
35. **Farmer, P. *et al.* Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 24, 4660–4671 (2005).**
36. **Bareche, Y. *et al.* Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Annals of Oncology* 29, 895–902 (2018).**
37. **Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* 121, 2750–2767 (2011).**
38. **Yang, Y., Pan, L., Zhou, W. & Shao, Z. Clinical and molecular characteristics of HER2-low triple-negative breast cancer: insights from multi-omics analysis. *Breast* 84, (2025).**
39. **Chandarlapaty, S. *et al.* Frequent mutational activation of the PI3K-AKT pathway in trastuzumab-resistant breast cancer. *Clin. Cancer Res.* 18, 6784–6791 (2012).**
40. **Arteaga, C. L. & Engelman, J. A. ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* 25, 282–303 (2014).**
41. **Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* 7, 763–777 (2007).**
42. **Ng, C. K. Y. *et al.* Intra-tumour genetic heterogeneity and alternative driver genetic alterations in breast cancers with heterogeneous HER2 gene amplification. *Genome Biol.* 16, (2015).**
43. **Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* 53, 1334–1347 (2021).**
44. **Marra, P. *et al.* IL15RA drives antagonistic mechanisms of cancer development and immune control in lymphocyte-enriched triple-negative breast cancers. *Cancer Res.* 74, 4908–4921 (2014).**
45. **Mukund, K. *et al.* Spatially distinct cellular and molecular landscapes define prognosis in triple-negative breast cancer. *Cell Rep.* 45, (2026).**
46. **Raine, K. M. *et al.* ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr. Protoc. Bioinformatics* 56, 15.9.1-15.9.17 (2016).**
47. **Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, (2011).**
48. **Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* 20, (2019).**

49. **Islam, S. M. A. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* 2, (2022).**

50. **Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature 2022 606:7916* 606, 984–991 (2022).**

51. **Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14, (2013).**

52. **Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120 (2013).**

53. **Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun.* 7, (2016).**

54. **Lehmann, B. D., Pietenpol, J. A. & Tan, A. R. Triple-negative breast cancer: molecular subtypes and new targets for therapy. *Am. Soc. Clin. Oncol. Educ. Book* e31–e39 (2015) doi:10.14694/EDBOOK_AM.2015.35.E31.**

55. **Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.* 21, 1688–1698 (2015).**

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SCFinalIBCGASupplementarytextSept92026.pdf
- ScFinalIBCGASupplementaryTablesSept92026.xlsx