

Targeting Germline BRCA1-Associated Breast Cancer Through Transcriptome-Guided Immunoprevention

Ying Ni

niy@ccf.org

Cleveland Clinic <https://orcid.org/0000-0002-0338-5239>

Article

Keywords:

Posted Date: March 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8502955/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.

Abstract

Women with germline *BRCA1* variants face a high lifetime risk of breast cancer, necessitating novel immunoprevention strategies. While the *BRCA1*-mutant tumor immune microenvironment (TME) is known to be immunologically active, the specific, targetable alterations that distinguish malignant tumors from the *BRCA1*-mutant normal tissue of the same carrier remain undefined.

We performed integrated transcriptomic and immunologic profiling on a unique cohort of trio *BRCA1*-mutant breast tumors, their matched adjacent and distant normal tissues, and separate normal breast tissues from tumor-free women harboring pathogenic germline *BRCA1* variants. This within-carrier design enables precise dissection of tumor-specific transcriptional and immune microenvironmental changes while controlling for both germline *BRCA1* background and host immune context.

Our analysis revealed a "hot but suppressed" TME, with high CD8⁺ T-cell infiltration counteracted by PD-L1/CTLA4 upregulation. We uncovered that *BRCA2* is upregulated in tumors, where its high expression paradoxically correlates with a "cold," T-cell-excluded immune phenotype. Concurrently, we identified an 80-gene "public" antigen signature, prioritizing five lead candidates (*LEMD1*, *EPYC*, *H3C13*, *H3C14*, and *H4C15*), that are, critically, most highly expressed in these "cold," high-*BRCA2* tumors. The upregulation of *Lemd1* was subsequently confirmed in the *Brca1*-deficient mouse model.

These findings identify *BRCA2* upregulation as a marker of an immune-suppressed state in *BRCA1*-mutant tumors and highlight candidate targets for *BRCA*-mutant patient-specific preventative vaccine development.

INTRODUCTION

Over 12 million individuals in the United States have a genetic predisposition to cancer, a number likely underestimated, as a recent large-scale study suggests the prevalence of pathogenic variants in key cancer susceptibility genes may be higher than previously expected, affecting up to 5% of the population¹. Heritable variants in *BRCA1* (0.22% prevalence) and *BRCA2* (0.42% prevalence) represent the most common cause of hereditary breast cancer². Women carrying germline *BRCA1/2* variants face a lifetime risk of up to 85% for developing breast cancer, often developing in their early 30s, compared to a 13% risk in the general population³. For the past two decades, this understanding has guided clinical decision-making, primarily involving aggressive management strategies like annual radiological surveillance and risk-reducing surgeries such as prophylactic mastectomy⁴. While effective, these interventions are highly invasive, highlighting a critical need for non-surgical preventive strategies.

The success of PD-1/PD-L1 inhibitor immunotherapies has firmly established the principle that the immune system can be harnessed to eradicate tumors⁵⁻⁷. This success raises an intriguing question for cancer interception: what is the baseline immune landscape in *BRCA1*-mutant tissues before and during malignant transformation? Recent studies have begun to characterize this landscape, revealing that

BRCA1-deficient tumors are often immunologically "hot," characterized by T-cell infiltration, genomic instability, and activation of the cGAS-STING pathway. However, this pro-inflammatory state is frequently coupled with potent immune resistance mechanisms, such as PD-L1 upregulation and the recruitment of suppressive immune cells. This suggests a complex TME that must be navigated^{8–12}. A key barrier to developing effective prophylactic cancer vaccines remains the identification of truly tumor-specific antigens that can be safely and selectively targeted to generate a productive immune response in an often-hostile tumor microenvironment.

The search for tumor-specific antigens has primarily focused on putative targets, often overlooking the broader tumor immune microenvironment (TME) in which these antigens are presented. A tumor's intrinsic biology can create an immune-suppressed or "cold" state, rendering even highly immunogenic antigens ineffective^{13–15}. The extent to which key driver genes, such as *BRCA1* and its homologous recombination partner *BRCA2*, shape this immune landscape is poorly understood. Compounding this challenge is the traditional, narrow definition of a "tumor-specific" antigen, which has prioritized "private," mutated neoantigens arising from DNA sequence alterations. This focus overlooks a broader spectrum of "public" antigens that are equally tumor-specific and potentially suitable for an off-the-shelf vaccine. Work from our group and others has demonstrated that potent anti-tumor immunity can be directed against non-mutated self-proteins that are aberrantly expressed in cancer but absent from normal adult tissues. Notable examples include the re-expression of placental NY-ESO-1 in sarcomas and melanomas¹⁶, and the tumor-specific expression of α -lactalbumin, normally restricted to lactation^{17,18}. To capture this complexity, an optimal immunoprevention discovery strategy must build a comprehensive atlas encompassing both public and private antigens and situate these targets within their immunological context. We addressed this by integrating transcriptomic and immunologic analyses of matched *BRCA1*-mutant breast tumor/normal trios to map the immune landscape, identify tumor-specific antigens, and establish a framework for future immunoprevention strategies.

METHODS

Patient Selection and Study Design

This study was approved by the Cleveland Clinic Institutional Review Board. Patients were identified from current physician schedules in the Breast Center and Plastic Surgery Center (Cleveland Clinic, Cleveland, Ohio, USA). Adult female patients with qualifying variants and those who had undergone surgical elections were recruited. Inclusion criteria were as follows: (1) Adult women > 18 years of age; (2) with a germline pathogenic *BRCA1* variant documented in the medical record; (3) planning to undergo a standard of care treatment mastectomy, a treatment mastectomy with contralateral risk-reducing mastectomy, or a bilateral risk-reducing mastectomy; and (4) able to sign informed consent. Exclusion criteria were: (1) Men; Women < 18 years of age; (2) absence of *BRCA1* variant or variant documented in medical record; (3) not planning a standard of care treatment mastectomy, treatment mastectomy with

contralateral risk-reducing mastectomy, or a bilateral risk-reducing mastectomy; and (4) refusal or unable to sign informed consent.

The following tissue samples were collected by our team at the Multidisciplinary Breast Center at the Cleveland Clinic: Matched trio breast tissue samples from individuals with germline *BRCA1* variants, including (1) breast tumors - Tumor; (2) matched normal breast tissues adjacent to the tumors – adjacent normal (Adj_Normal); (3) matched normal breast tissues distant from the tumor via risk-reducing contralateral breast surgeries in patients with *BRCA1* tumors – distant normal (Dis_Normal); and (4) additional normal breast tissues (from unrelated women) via prophylactic risk-reducing double mastectomies in *BRCA1* variant carriers with no history of breast cancer –prophylactic normal breast tissue (PNT).

Tissue Processing

Post-surgery tissue samples were obtained from the Department of Pathology, Cleveland Clinic (Cleveland, Ohio, USA). A qualified breast pathologist performed an initial gross dissection to separate tissue into tumor, tumor-adjacent normal, and distant normal cohorts. Each specimen was then processed in parallel: (1) for molecular analysis, a tissue portion (approx. 1 cm³) was snap-frozen in a cryotube and stored at -80°C for subsequent RNA isolation; and (2) for histological quality assessment (QC), an adjacent piece of tissue was fixed in 10% neutral buffered formalin and embedded in paraffin to create a corresponding FFPE block.

Prior to RNA isolation from the frozen tissue, a definitive microscopic quality control was performed. Sections from the matched FFPE block were stained with Hematoxylin and Eosin (H&E). A board-certified breast pathologist reviewed these H&E slides to confirm the histological classification and assess sample purity, annotating any contaminating cell types (e.g., fat in tumor samples; tumor cells in normal samples). Based on this H&E review, only snap-frozen tissues corresponding to histologically pure blocks were used directly for RNA isolation. In cases where the H&E slide revealed significant tissue heterogeneity, the corresponding frozen tissue was embedded in OCT compound, sectioned, and specific regions of interest were isolated via microdissection to ensure sample purity before proceeding with RNA extraction.

RNA Extraction and Library Preparation For RNA-Sequencing

Flash-frozen tissue sections were homogenized in 1 ml Trizol (Life Technologies, Carlsbad, CA) using 5 mm stainless steel beads for 2 min at 25 Hz with a TissueLyser II (Qiagen, Valencia, CA). We utilized the Ambion PureLink RNA Mini Kit (Life Technologies, Carlsbad, CA) for RNA extraction, DNase treatment, and cleanup. Briefly, we applied 15 µg of total RNA, mixed with lysis buffer and ethanol (1:3), to the column and rinsed it with Wash Buffer II (70% ethanol). DNA was digested on-column for 15 min with PureLink DNase (Life Technologies, Carlsbad, CA), and the RNA was rinsed twice with Wash Buffer II

(70% ethanol). We assessed RNA quality using a Bioanalyzer, and 5 µg DNA-free RNA was shipped to the Case Western Reserve University Genomics Core (Cleveland, Ohio, USA) for library preparation and sequencing. RNA-sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA with Ribo Zero Kit (Illumina, San Diego, CA). The libraries were multiplexed three to a lane, grouped in biological triplicates. We performed 100 base pair, paired-end sequencing using the Illumina HiSeq 2500 Rapid Run v2 flow cell platform.

RNA-sequencing Data Analysis

Sequencing reads were aligned to the human reference genome hg38 using STAR¹⁹ (version 2.54b). Sequences that remained unaligned to hg38 were included in a separate pool for BLAST comparison of sequences between normal and tumor samples to maximize the identification of potential neoantigens. Aligned sequence reads were processed for differential expression using an in-house pipeline that utilizes HTSeq and NOISeq. HTSeq was used to count the raw reads associated with a given gene, and NOISeq was used for normalization and differential expression analysis. Expressed non-coding sequences identified via alignment to hg38 were compared between tumor and normal tissues for potential neoantigens²⁰.

Differential Expression Analysis

Following predetermined contrasts, we calculated differential expression by comparing wild-type Transcripts Per Million (TPM) values of *BRCA1* mutant breast tumors (tissue type 1) to matched normal adjacent (tissue type 2) and to normal distant (tissue type 3), as well as to normal breast tissue obtained from prophylactic risk reduction surgery in *BRCA1* variant carriers (tissue type 4). The differential gene expression was modeled with the “dream” method from variantPartition package in R. Genes were considered differentially expressed at $q < 0.05$ following Benjamini-Hochberg multiple testing correction of the original p -values.

Gene-signature Score and Immune Cell Abundance Estimation

The signature-based scoring was calculated using the rank-based single-sample gene set scoring method (simpleScore) provided by the R package IOBR^{21,22}. Singscore method implements a simple single-sample geneset (gene-signature) scoring method, which scores individual samples independently without relying on other samples in gene expression datasets.

Bulk immune cell abundance estimation was profiled using MCP-counter²³. More detailed immune cell with activation states was profiled using CIBERSORT²⁴, which is a signature matrix file consisting of 547 genes that accurately distinguish 22 mature human hematopoietic cell populations isolated from

peripheral blood or in vitro culture conditions, including seven T-cell types, naïve and memory B cells, plasma cells, NK cells, and myeloid subsets.

Fusion Neoantigen Analysis

We performed gene fusion analysis on our RNA-seq data using the Arriba 12²⁵ pipeline, a tool developed for clinical research settings based on the ultrafast STAR aligner, as described above.

Putative Neoantigens Analysis from TCGA

Somatic whole-exome sequencing variant data from The Cancer Genome Atlas (TCGA) breast cancer study were downloaded from dbGaP. The matched HLA typing data were generated by the TCGA pan-cancer immune landscape effort from 885 individuals (965 samples) to incorporate individualized neoantigens specific to the individual HLA alleles²⁶. The germline *BRCA1* variant status was downloaded from dbGaP (Proj 21943). Only pathogenic or likely pathogenic variants as reported in ClinVar were considered as *BRCA1* germline variant-positive.

All discovered protein-coding somatic variants are translated into 17-mer peptides with the mutant amino acid positioned in the center. The 9-mer amino acid substrings within the mutant peptides with a projected MHC Class I binding affinity to each of the six most common patient-specific HLA alleles were identified using a sliding window approach. The binding affinity for mutant and matched wild-type peptides was reported to discover neoantigens using the NetMHCpan4.1²⁷ software.

Statistical Analysis

For reporting, continuous variables are expressed as mean $\pm$ standard deviation (SD) and were compared using the Student's t-test or Paired Wilcoxon signed-rank tests. Categorical variables are presented as count (%) and were compared using the chi-squared or Fisher's exact test. The Spearman correlation coefficient was used to analyze the correlation between different variables. Overall differences were evaluated by Kruskal–Wallis test. Paired Student's t-test or Wilcoxon signed-rank tests were used for post-hoc pairwise comparisons. All statistical tests were two-sided, and a p-value of less than 0.05 was considered significant across all analyses performed, or a False Discovery Rate (FDR) of less than 0.05 after multiple testing correction. Statistical analyses were performed using R (version 4.0.3) and RStudio (version 2024.12.1).

RESULTS

The Transcriptomic Landscape of BRCA1-Mutant Breast Cancer

To comprehensively characterize the transcriptomic landscape associated with germline *BRCA1* variants, we performed RNA-sequencing on four distinct breast tissue cohorts. Our primary goal was to identify molecular differences that could inform targeted prevention strategies for *BRCA1* variant carriers. Specifically, we analyzed (1) tumor tissues, (2) matched adjacent normal breast tissues (Adj_Norm), and (3) matched distant normal breast tissues (Dis_Norm). As an additional control group, we analyzed (4) normal breast tissues from unrelated individuals with a germline *BRCA1* variant but no history of breast cancer (prophylactic normal breast tissue; PNT).

Principal component analysis (PCA) revealed clear segregation of global transcriptomic profiles among the four cohorts (Fig. 1A). Tumor samples clustered distinctly from all three normal tissue groups, underscoring a robust molecular divergence driven by malignant transformation. Notably, the matched normal from individuals with tumor (Adj_Norm and Dis_Norm) also clustered separately from non-tumor PNT, suggesting unique transcriptional profiles exist even within normal mammary tissue environments with or without developed tumor. Pairwise PCA comparisons confirmed this separation across all contrasts (tumor vs. Adj_Norm (Fig. 1B), tumor vs Dis_Norm (Fig. 1C), and tumor vs PNT (Fig. 1D). Consistent with these findings, pairwise differential gene expression analyses identified several significantly differentially expressed genes between the tumor and each normal tissue group (Fig. 1E-G).

We next asked how *BRCA1* is expressed in breast tissue from individuals carrying germline *BRCA1* variants, and whether its homolog, *BRCA2*, shows compensatory changes. While *BRCA1* transcript levels were similar across all cohorts (Fig. 2A), *BRCA2* expression was significantly upregulated in tumors compared to all normal tissue groups (Fig. 2B). Together, these data reveal that while *BRCA1* mRNA levels remain unchanged in variant carriers, *BRCA2* expression rises specifically in tumors, suggesting a dynamic interplay between the two homologous recombination genes during malignant transformation. Given that such profound transcriptional alterations, including the specific upregulation of *BRCA2*, and the underlying genomic instability they imply, are known to be primary drivers of immune cell infiltration and tumor immune microenvironment remodeling²⁸⁻³⁰, we next sought to characterize the resulting TME.

The Tumor Immune Microenvironment is Pro-Inflammatory but Immunosuppressed

We used deconvolution algorithms to infer immune cell populations from our bulk RNA-seq data to understand the tumor immune microenvironment (TIME). Our analysis revealed strong evidence of an active, pro-inflammatory TIME. Tumors were significantly more infiltrated by key anti-cancer immune cells compared to normal tissues, including total T-cells ($p = 0.015$) and, more specifically, cytotoxic CD8 + T-cells ($p = 0.010$) (Fig. 3A,B). This inflamed state was further supported by a definitive shift in macrophage polarization. The TIME showed a higher abundance of pro-inflammatory M1 macrophages ($p < 0.001$) and a concurrent lower abundance of immune-suppressive M2 macrophage signatures ($p = 0.048$) relative to normal tissues (Fig. 3C-E). However, this pro-inflammatory environment was coupled with potent signals of adaptive immune resistance. We observed significantly higher expression of the

critical immune checkpoint genes *CTLA4* ($p = 0.0095$) and *CD274* (PD-L1) ($p = 0.014$) in tumors compared to all normal cohorts (Fig. 3F,G).

Collectively, these data indicate that while *BRCA1*-mutant tumors are immunologically "hot"—recognized and infiltrated by the immune system—they simultaneously upregulate powerful inhibitory pathways to suppress an effective anti-tumor response. This inflamed but immunosuppressed TME highlights the need for therapeutic strategies that can tip the balance in favor of productive immunity, providing a strong rationale for the identification of novel vaccine targets.

BRCA2 Upregulation Correlates with an Immunologically Cold Tumor Microenvironment

While our analysis revealed that tumors as a group are more inflamed than normal tissues, we hypothesized that substantial immune heterogeneity exists within the tumor cohort itself. To investigate the potential drivers of this variability, we focused on the relationship between key homologous recombination genes, *BRCA1* and *BRCA2*, and the TME.

First, we examined the state of the normal tissues. In all three non-malignant cohorts, *BRCA2* expression showed a consistent positive correlation with the abundance of pro-inflammatory immune cells, indicating a state of healthy immune coordination^{31,32} (Fig. 4A-C). Specifically, *BRCA2* expression showed a consistent positive correlation with the abundance of pro-inflammatory immune cells, including T-cells and cytotoxic lymphocytes. This indicates that in a normal physiological state, higher *BRCA2* expression is associated with a more active immune environment, likely reflecting normal immune surveillance^{33–36}. In contrast, when we analyzed tumor tissues, we observed a striking, paradoxical reversal of this relationship (Fig. 4D). The positive correlations seen in normal tissues were completely inverted. In the tumor microenvironment, high *BRCA2* expression was instead significantly negatively correlated with the presence of anti-cancer immune cells. For example, *BRCA2* expression showed negative correlations with the abundance of T-cells ($r = -0.24$), CD8 + T-cells ($r = -0.21$), and cytotoxic lymphocytes ($r = -0.22$).

This critical finding clarifies the heterogeneity of the TME. While tumors, on average, are inflamed, those with high *BRCA2* expression create an immunologically "cold," or T-cell-excluded, microenvironment. The detailed correlation patterns for both *BRCA1* and *BRCA2* against a wider panel of immune markers are provided in Supplementary Fig. 1, further confirming this phenomenon. This discovery of a *BRCA2*-associated immune suppression mechanism provides a crucial insight into how these tumors evade immune destruction.

Identification of a Repertoire of Immunotherapeutic Targets

Having established that *BRCA1*-mutant tumors are immunologically active yet remain functionally restrained, we next identified tumor-specific antigens that could serve as suitable targets for therapeutic vaccine development. Our approach combined comparative transcriptomic profiling with neoantigen

discovery, integrating both shared antigens overexpressed across tumors for an "off-the-shelf" vaccine, and private neoantigen sources.

A Core Signature of Shared, Overexpressed Antigens

To identify shared targets, our analysis of tumors versus both adjacent and distant normal tissues identified over 1500 genes with a twofold or greater overexpression (P -adjusted < 0.05). To prioritize robust, tumor-associated targets, this list was filtered to include only coding genes with minimal expression in normal tissues ($< 15\%$) but prevalent expression in tumors ($> 40\%$).

This stringent filtering process yielded a final signature of 80 tumor-associated genes (Supplementary Table 1). These genes were consistently overexpressed in tumor samples compared to the minimal expression observed in all normal tissue cohorts (Fig. 5A). Comparing tumors to adjacent and distant normal tissues identified 44 and 64 differentially expressed genes (DEGs), respectively, with 28 genes common to both comparisons (Fig. 5B). Analysis of the prevalence of these 80 genes across the tumor cohort revealed that 42 genes (52.5%) were overexpressed in more than 50% of tumors, and two genes (S100P and a novel transcript C5orf46) were overexpressed in over 80% of tumors (Fig. 5C). This suggested that breast carcinomas arising in *BRCA1* variant-positive women exhibit common signaling alterations driven by their underlying genetic predisposition, which is further reflected in a tumor-specific differential transcriptome and protein expression profile. Notably, none of these DEGs were expressed differently when comparing adjacent normal, distant normal, and normal breast tissues from *BRCA1* variant carriers without a history of breast cancer. Finally, to confirm tumor-specificity, the expression of these 80 candidates was evaluated across a panel of normal human tissues in the GTEx database, allowing for the prioritization of genes with the lowest expression in non-mammary tissues (Fig. 5D).

The Landscape of Fusion-derived Neoantigens

In addition to shared antigens, we characterized the landscape of private neoantigens arising from two classes of tumor-specific somatic variants. First, we analyzed gene fusions, a key source of novel open reading frames^{37–39}. Across the 62 samples, we detected 973 (30%) fusions predicted with medium or high confidence to generate novel protein sequences (Supplementary Fig. 2, Supplementary Table 2 - HighConfidenceFusion), yielding a median of 14 predicted fusion peptides per sample (range: 1–50).

Analysis of this high-confidence fusion set revealed a significant difference in their abundance across the four tissue cohorts (Kruskal-Wallis $p = 0.0372$) (Supplementary Fig. 2). Specifically, tumor tissues exhibited a higher median number of these high-confidence, neoantigen-producing fusions compared to adjacent normal tissues ($p = 0.03662$), and showed a strong trend toward higher numbers when compared to distant normal tissues ($p = 0.06152$). No significant differences were observed between the various normal tissue cohorts (Supplementary Fig. 2).

Further in silico analysis predicted that 618 of the fusion events could produce neo-peptides. These neo-peptide-generating fusions were detected in 14 tumor samples and 40 normal samples (Supplementary Table 2 - NeoPeptide). Interestingly, the repertoire of predicted fusion peptides was largely distinct

between tumors and normal tissues, with minimal overlap observed, with predicted fusion peptides being significantly different in tumor tissues compared to normal tissues, both within matched trios and between all cancers versus any normal tissues (despite tumors having a higher load of fusion events). This indicates that the fusion events are highly heterogeneous and mostly private, which represent a rich source of tumor-specific peptides that may serve as personalized vaccine targets.

From this comprehensive repertoire of shared and private antigens, we prioritized a shortlist of five lead candidates from the shared overexpression signature for deeper investigation: *EPYC*, *LEMD1*, *H3C13* (*HIST2H3D*), *H3C14* (*HIST2H3C*), and *H4C15* (*HIST2H4B*). These specific genes were selected based on a combination of factors, including the highest magnitude and prevalence of overexpression in our tumor cohort (Supplementary Fig. 3), minimal to no expression across vital normal tissues in the GTEx database (Fig. 5D), and high sequence conservation in murine models, a prerequisite for future preclinical validation studies.

Validation of Human Gene Targets in a Brca1-Deficient Mouse Model

Following the prioritization of our five lead candidates as potential targets based on human tumor data, we next assessed which of our top human vaccine candidates are preserved in a preclinical setting suitable for preclinical vaccine studies. To this end, we utilized a genetically engineered mouse model that faithfully recapitulates human *BRCA1*-mutant breast cancer⁴⁰. We analyzed bulk RNA-seq data from spontaneous tumors arising in a K14-Cre; p53^{f/f}; Brca1^{f/f} model in which both *Brca1* and *p53* are deleted (n = 21). As a stringent control, we used tumors that were also p53-deficient but retained a functional *Brca1* gene (K14-Cre p53^{f/f} model, n = 8), allowing us to isolate the specific effects of *Brca1* loss.

We focused our analysis on the mouse homologs of our top five human candidates, which all showed sequence conservation (Supplementary Fig. 4A). The expression patterns of these homologs varied significantly (Supplementary Fig. 4B). Expression was not detected for the mouse homologs of the histone cluster genes *H3C13*, *H3C14*, or *H4C15*, while the mouse homologs for Epiphykan (*Epyc*) and LEMD1 (*Lemd1*) were detected in the murine tumors, suggesting potential species-specific regulation or developmental restriction.

Critically, *Lemd1* was the only candidate to show a positive expression trend associated with *Brca1*-deficiency, exhibiting a log2 fold change of 1.64 (padj = 0.2), while *Epyc* showed a log2 fold change of 0.18 (padj = 0.89) (Supplementary Fig. 4B). While this trend did not reach statistical significance after correction for multiple testing, the magnitude of the effect clearly isolates *Lemd1* as the most promising and relevant candidate for further preclinical development. These results provide important direct evidence supporting the advancement of an LEMD1-targeted vaccine.

The Landscape of Private, Variant-Derived Neoantigens in TCGA Breast Cancer Samples with Germline BRCA1

Variants

To complement our cohort-based findings, we investigated neoantigens originating from single-nucleotide variations (SNVs) using data from the TCGA breast cancer cohort. Neoantigens that originate from non-synonymous somatic variants, such as SNVs and insertions or deletions (Indels), are the most extensively studied class of cancer-specific antigens. We focused on 18 individuals with a germline *BRCA1* variant from the TCGA breast cancer cohort and analyzed their somatic variant data. Compared to non-*BRCA1* mutated breast tumors, tumors carrying *BRCA1* germline variant trended toward a higher burden of somatic, protein-coding variants compared to other breast cancer subtypes (mean = 134 vs. 54, $p = 0.08$). Correspondingly, this resulted in a higher number of predicted neoantigens, with a trend toward more peptides predicted to bind the patient's specific HLA types (mean = 350 vs. 136, $p = 0.09$) (Supplementary Fig. 5). To identify high-confidence neoantigen candidates, we applied stringent filtering to an initial set of 135,261 peptide-MHC-I bindings. We excluded predictions where the mutant peptide had a weak binding score (rank > 10%) or where the corresponding wild-type peptide was also a strong binder (ranking < 2%). This process yielded a final set of 8710 high-confidence, variant-specific bindings. Notably, this analysis identified a neoantigen originating from a variant in the *H3C13* (H3 clustered histone 13) gene in one of the TCGA *BRCA1* variant carriers. This is a particularly relevant finding, as *H3C13* was also identified as one of the overexpressed genes in our primary analysis of *BRCA1*-mutant tumors, providing a direct link between the two independent datasets.

Vaccine Candidates Are Preferentially Expressed in Tumors with the BRCA2-Associated Immune-Cold Phenotype

The discovery that high *BRCA2* expression delineates an immunologically "cold" T-cell-excluded subset of tumors presented a critical test for our therapeutic strategy. For a vaccine to be effective, particularly one designed to "prime" a non-responsive TME, its target antigen must be highly expressed on the tumor cells that have most successfully evaded the immune system⁴¹. We therefore hypothesized that our lead candidate antigens would be co-expressed with *BRCA2*, our established proxy for the "cold" tumor microenvironment.

We performed a correlation analysis between the expression of our five lead candidates (*EPYC*, *LEMD1*, *H3C13*, *H3C14*, and *H4C15*) and *BRCA2* transcript levels specifically within our tumor cohort. Remarkably, the results showed a clear and consistent pattern: all five candidates demonstrated a strong and statistically significant positive correlation with *BRCA2* expression (Fig. 6). This finding confirms the ideal therapeutic scenario: the tumors with the highest levels of *BRCA2* and therefore the most immune-suppressed TME are the very same tumors that most highly express our target antigens. This coordinated expression pattern establishes a direct link between the immune-suppressed tumor state and the molecular presence of our vaccine targets. It suggests that these antigens are the "hallmark" of the exact tumor cells that require immune-mediated clearance.

DISCUSSION

This study was designed to address a challenge in hereditary breast cancer research: identifying tumor-specific targets that can enable safe and effective immunoprevention for women carrying germline *BRCA1* variants. Building on our group's experience translating tumor-specific mRNA vaccine technology into first-in-human clinical trials for solid tumors^{42,43}, we applied a precision transcriptomic approach to discover antigens uniquely expressed in *BRCA1*-mutant tumors. A defining strength of our experimental approach lies in the use of matched trios of breast tissues—tumor, adjacent normal, and distant normal—from the same *BRCA1* variant carriers, complemented by cancer-free prophylactic samples from additional carriers. This design minimizes inter-individual variability and allows precise delineation of transcriptional changes that are specific to malignant transformation, rather than effects of *BRCA1* haploinsufficiency or genetic background. By controlling both patient genotype and tissue context, our study provides an exceptionally robust framework to identify tumor-restricted molecular signatures that represent genuine targets for immunoprevention.

Leveraging this unique sample architecture, we integrated transcriptomic profiling, immune deconvolution, and neoantigen prediction to construct a comprehensive antigen landscape in *BRCA1*-associated breast cancer. Our analysis first revealed that these tumors exist in a paradoxically "hot but suppressed" immune microenvironment, characterized by significant T-cell infiltration that is counteracted by the upregulation of immune checkpoints. Within this immune context, we identified a core signature of 80 shared, overexpressed genes with tumor-specific expression, providing a rich pool of candidates for an "off-the-shelf" vaccine. Critically, we discovered that these candidate antigens are most highly expressed in tumors with high levels of *BRCA2*, a feature we identified as a marker for an immunologically "cold" or T-cell-excluded state, making them ideal targets for a vaccine designed to awaken a dormant immune response. A top candidate from this signature, *LEMD1*, was subsequently validated as overexpressed in a relevant *Brca1*-deficient murine model, so that this target may be tested in preclinical vaccine studies. In parallel, our characterization of neoantigens confirmed that while gene fusions yield patient-private targets, SNV-derived neoantigens are more abundant in these tumors, and our discovery of *H3C13* as a "dual-hit" candidate highlights a novel link between the two antigen classes. Collectively, these findings provide distinct, data-driven paths toward the development of both public and personalized immunoprevention strategies.

A particularly interesting finding of this study is the paradoxical association of *BRCA2* expression with the TME. It is well established that *BRCA1*-mutant tumors possess a distinct and complex immune landscape. Consistent with our observations and prior reports^{28,29,44–47}, these tumors display marked genomic instability, a high mutation burden, and dense infiltration by immune cells—particularly CD8⁺ T-cells—yet simultaneously exhibit elevated PD-L1 expression and expansion of myeloid-derived suppressor cells. This "hot but suppressed" phenotype is a defining feature of *BRCA1* deficiency.

In contrast, the specific contribution of *BRCA2* expression to this immune context has remained unclear. Although both *BRCA1* and *BRCA2* are classically defined by their roles in DNA repair⁴⁸, our data uncover an interesting correlation between *BRCA2* expression and the TME. Loss of *BRCA2* function has been linked to genomic instability, increased neoantigen load, and an inflamed ("hot") TME responsive to

checkpoint blockade⁴⁹. Consistent with this, we found that in normal tissues, *BRCA2* expression correlated positively with T-cell populations, reflecting effective immune surveillance. However, within tumors, this relationship was inverted; high *BRCA2* expression was instead associated with a T-cell-excluded ("cold") phenotype. This shift suggests that, while *BRCA1* deficiency may create a baseline inflamed state, the tumor-specific upregulation of wild-type *BRCA2* is a feature of established tumors that have successfully evaded immune clearance.

Our data demonstrate that *BRCA1*-mutant tumors are not immunologically barren; rather, they exist in a paradoxically "hot but suppressed" state. We found the TME to be significantly infiltrated by anti-cancer immune cells, including cytotoxic CD8 + T-cells and pro-inflammatory M1 macrophages. These findings are well-supported by existing evidence, which has consistently shown that *BRCA1*-deficient tumors are enriched with tumor-infiltrating lymphocytes (TILs), including CD8 + and CD4 + T-cells, as well as tumor-associated macrophages (TAMs)^{29,44,50}. However, our results also show that this productive immune infiltration is actively held in check by potent co-expressed inhibitory signals, including the significant upregulation of the *CTLA4* and *CD274* (PD-L1) checkpoint genes. This confirms that a pre-existing but stalled anti-tumor immune response is a hallmark of these tumors, a finding that aligns with reports of increased PD-L1/PD-1 expression^{28,51} and the expansion of other suppressive populations like MDSCs in *BRCA1*-mutant models⁴⁷.

A major strength of our study is that we focused on RNA-based vaccine targets, rather than just looking for a single type of antigen. By integrating our cohort's transcriptomic data with fusion analysis and TCGA variant data, we constructed a comprehensive antigen atlas for *BRCA1*-mutant cancer. This "multi-angle" approach revealed a complementary landscape of potential targets. Our analysis of gene fusions, for instance, confirmed they are a source of highly immunogenic "private" neoantigens, but their patient-specific nature necessitates a costly and time-intensive personalized vaccine approach. Similarly, our TCGA analysis confirmed a trend of higher neoantigen burden from SNVs, further supporting the rationale for individualized therapies. Our most clinically actionable discovery is the 80-gene signature of "public," or shared, overexpressed antigens. This provides a clear path toward a more feasible "off-the-shelf" vaccine, which is far better suited for the preventive or adjuvant setting in a large high-risk population. This dual discovery underscores a critical point: a one-size-fits-all approach is insufficient for this complex disease, and our work provides a data-driven rationale for pursuing both public and personalized vaccine strategies in parallel.

The significance of *LEMD1* as a lead target is strongly supported in the literature. *LEMD1* (LEM Domain Containing 1) is a predicted integral membrane protein whose expression is normally restricted to the testis, classifying it as a cancer-testis antigen (CTA)^{52,53}, a premier category of vaccine targets valued for their high tumor specificity and minimal expression in normal adult tissues. Our identification of *LEMD1* in *BRCA1*-mutant tumors, which are predominantly triple-negative, is consistent with recent reports independently recognizing *LEMD1* as a novel oncogene and therapeutic target in triple-negative breast cancer (TNBC)⁵⁴. Notably, its overexpression extends beyond breast cancer and has been

observed in several other aggressive malignancies, including gastric, non-small cell lung (NSCLC), colorectal, and thyroid cancers, where it similarly correlates with poor prognosis.^{55–57} *LEMD1* functions as an active oncogenic driver rather than a passive bystander. Elevated expression correlates with poor prognosis, increased proliferation, invasion, metastasis, and cancer stemness across multiple malignancies^{54,56–59}. Mechanistically, *LEMD1* promotes tumor aggressiveness by activating PI3K/AKT and ERK signaling pathways^{55,56,60}. These findings underscore its biological relevance and therapeutic potential, as *LEMD1* knockdown suppresses tumor growth and enhances chemosensitivity^{54–56}. Our data support *LEMD1* as a compelling candidate for prophylactic vaccine development to prevent tumor initiation or progression in *BRCA1*-mutant settings for at-risk individuals.

Additional lead candidates—*EPYC*, *H3C13*, *H3C14*, and *H4C15*—represent intriguing avenues for future research. *EPYC* (epiphykan) is a compelling target that has recently been identified as an oncogenic driver and prognostic biomarker in pancreatic and ovarian cancers, where it promotes metastasis via the PI3K-AKT pathway^{61,62}. Although its specific role as an immunotherapeutic target is currently uncharacterized, its link to a core survival pathway in other aggressive cancers warrants further investigation. The histone cluster genes—*H3C13*, *H3C14*, and *H4C15*—represent a particularly novel finding. These genes, which encode H3.2 and H4 histone variants, are overexpressed across multiple malignancies, including breast and gastric cancers⁶³. Their variants have been described as "oncohistones" that drive tumorigenesis through aberrant chromatin regulation^{64–66}. Our identification of *H3C13* as a "dual-hit" antigen—both highly overexpressed in our cohort and a source of neoantigens in the TCGA dataset—highlights its strong biological relevance. However, given the essential and ubiquitous roles of core histones in chromatin organization and gene regulation,^{67,68} these proteins are unlikely to be viable candidates for therapeutic development. Systemic targeting would carry a high risk of toxicity and off-target effects in normal proliferating cells.

Finally, the implications of our findings may extend significantly beyond the germline *BRCA1* variant carrier population. A large subset of sporadic triple-negative breast cancers (TNBCs), which lack a germline variant, nevertheless exhibit a "BRCAness" phenotype characterized by defects in homologous recombination and a similar molecular signature to *BRCA1*-mutant tumors^{69–73}. Our lead candidate, *LEMD1*, has recently been identified by others as a key oncogene and therapeutic target specifically in sporadic TNBC, where it drives proliferation and is associated with poor prognosis^{54–56,60}. This convergence suggests that our 80-gene signature, and *LEMD1* in particular, may represent a shared antigenic landscape common to the entire "BRCAness" subtype. This considerably broadens the potential clinical impact of our work, positioning these antigens as viable "off-the-shelf" vaccine targets for a much larger population of patients with sporadic TNBC who currently lack effective targeted therapies.

While the transcriptomic foundation of this study provides a powerful and comprehensive atlas of potential targets, several considerations are essential for their future clinical translation. Our discovery pipeline, by design, identified candidates at the RNA level. The necessary next steps will involve

confirming the robust expression of these 80 candidates at the protein level and, more critically, validating their processing and presentation on MHC Class I molecules on the tumor cell surface, which is the ultimate prerequisite for T-cell recognition. Similarly, our characterization of the TME, based on computational deconvolution, has provided a critical blueprint of the immune landscape. This *in silico* framework should be further validated using spatial technologies, such as multiplex immunofluorescence, to confirm the physical co-localization and spatial relationships between immune subsets (like CD8 + T-cells) and cells expressing high levels of *BRCA2* or our vaccine targets. This finding does not diminish the potential of these antigens in humans but rather reinforces the critical importance of our human-tissue-first discovery approach. These considerations provide a clear and actionable roadmap for advancing our lead candidates from *in silico* discovery to a clinically viable immunoprevention strategy.

In conclusion, this study provides a comprehensive immunogenomic atlas of *BRCA1*-mutant breast cancer, extending beyond antigen discovery to define the immune context in which these tumors emerge. We identified a paradoxical “hot but suppressed” tumor immune microenvironment, characterized by robust T-cell infiltration constrained by checkpoint upregulation. Our analyses further revealed that upregulation of wild-type *BRCA2*, rather than its loss, correlates with a T-cell-excluded (“cold”) phenotype, suggesting an active immune evasion mechanism. Critically, we uncovered a core set of 80 shared, overexpressed antigens, including the high-priority target *LEMD1*, which are preferentially expressed in the most immunologically cold tumors. Validation of *LEMD1* in a *Brca1*-deficient murine model establishes a clear translational foundation for vaccine development. Together, these findings provide a data-driven rationale and preclinical framework for a prophylactic vaccine approach, designed for primary disease prevention in high-risk carriers or to prevent disease recurrence in the adjuvant setting.

Declarations

ACKNOWLEDGMENTS

We are grateful to all our patients and families who have participated in research studies for more than two decades. We appreciate Tammy Sadler, Matthew Wolf, and Dr. Ritika Jaini for their contributions to the initial cohort identification and sample collection that laid the partial groundwork for this project. We also thank Dr. Erinn Downs-Kelly (Pathology) for her expert histopathologic review and assistance with specimen triage and processing, and Dr. Stephen R. Grobmyer for patient recruiting.

We dedicate this work to the memory of Dr. Charis Eng, who conceived and led this project. Dr. Eng was the founding Chair of the Genomic Medicine Institute and the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic. Her vision, mentorship, and lifelong commitment to advancing the understanding of hereditary cancer syndromes made this study possible. Her scientific rigor, compassion for patients, and unwavering dedication continue to inspire our work.

We are also grateful for our funding sources, including the American Cancer Society (RPG-02-151-01-CCE and Clinical Research Professorship) and Cancer Preventive Initiative HeritX (CE). CE was an American Cancer Society Clinical Research Professor. NLS is funded by NDCR, NIHK08, and the American Cancer Society. DJM is funded by NCI R00CA240689. YN is funded by the Breast Cancer Research Foundation. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

DATA AVAILABILITY

Aggregate data are provided in the associated tables. To protect individual patient information, the Cleveland Clinic Foundation IRB (IRB #16-520) and Legal Department do not permit clinical information to be deposited in a publicly accessible database (by policy). RNAseq data will be deposited to the National Institute of Health (NIH) Sequence Read Archive (SRA).

COMPETING INTERESTS

All authors declare no competing interests.

AUTHOR CONTRIBUTIONS

CE and YN conceptualized and designed the project. PB, YN, LL, LY, VM, HD, and TA performed the data analyses. YN, LY, VSC, NLS, and DJM interpreted the data. LL, YN, and VSC performed the statistical analyses. CE and YN supervised the project. VSC, LY, YN, NS, DM, and CE drafted the original version of the manuscript. LY, VSC, DJM, NLS, JA, and YN made critical inputs for revision. All authors drafted, critically revised, and gave final approval of the paper. CE passed away on 13 August 2024.

References

1. Idumah, G. *et al.* Pathogenic Germline Variants in Cancer Susceptibility Genes. *JAMA* (2025) doi:10.1001/JAMA.2025.16372.
2. Petrucelli, N., Daly, M. B. & Pal, T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. *GeneReviews®* (2025).
3. Antoniou, A. C. & Easton, D. F. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* **25**, 190–202 (2003).
4. Daly, M. B. *et al.* Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc. Netw.* **19**, 77–102 (2021).
5. Jönsson, J. M. *et al.* Molecular subtyping of serous ovarian tumors reveals multiple connections to intrinsic breast cancer subtypes. *PLoS One* **9**, (2014).

6. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
7. Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**, 2443–2454 (2012).
8. van Vugt, M. A. T. M. & Parkes, E. E. When breaks get hot: inflammatory signaling in BRCA1/2-mutant cancers. *Trends in Cancer* **8**, 174–189 (2022).
9. Bruand, M. *et al.* Cell-autonomous inflammation of BRCA1-deficient ovarian cancers drives both tumor-intrinsic immunoreactivity and immune resistance via STING. *Cell Rep.* **36**, 109412 (2021).
10. Bhat-Nakshatri, P. *et al.* Signaling Pathway Alterations Driven by BRCA1 and BRCA2 Germline Mutations are Sufficient to Initiate Breast Tumorigenesis by the PIK3CAH1047R Oncogene. *Cancer Res. Commun.* **4**, 38–54 (2024).
11. Miklikova, S. *et al.* The Role of BRCA1/2-Mutated Tumor Microenvironment in Breast Cancer. *Cancers (Basel)*. **13**, 575 (2021).
12. Pal, B. *et al.* A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* **40**, 1–23 (2021).
13. Sayed, R. El, Haibe, Y., Amhaz, G., Bouferraa, Y. & Shamseddine, A. Metabolic Factors Affecting Tumor Immunogenicity: What Is Happening at the Cellular Level? *Int. J. Mol. Sci.* **22**, 1–28 (2021).
14. Khosravi, G. R. *et al.* Immunologic tumor microenvironment modulators for turning cold tumors hot. *Cancer Commun. (London, England)* **44**, 521–553 (2024).
15. Liu, Y. T. *et al.* Turning cold tumors into hot tumors to ignite immunotherapy. *Mol. Cancer* **24**, 254 (2025).
16. Thomas, R. *et al.* NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. *Front. Immunol.* **9**, (2018).
17. Jaini, R. *et al.* An autoimmune-mediated strategy for prophylactic breast cancer vaccination. *Nat. Med.* **16**, 799–803 (2010).
18. Tuohy, V. K. & Jaini, R. Prophylactic cancer vaccination by targeting functional non-self. *Ann. Med.* **43**, 356–365 (2011).
19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
20. Laumont, C. M. *et al.* Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, (2018).
21. Zeng, D. *et al.* Enhancing immuno-oncology investigations through multidimensional decoding of tumor microenvironment with IOBR 2.0. *Cell Reports Methods* **4**, (2024).
22. Zeng, D. *et al.* IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* **12**, (2021).
23. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, (2016).

24. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
25. Uhrig, S. *et al.* Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, 448–460 (2021).
26. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812–830.e14 (2018).
27. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
28. Wen, W. X. & Leong, C. O. Association of BRCA1- and BRCA2-deficiency with mutation burden, expression of PD-L1/PD-1, immune infiltrates, and T cell-inflamed signature in breast cancer. *PLoS One* **14**, (2019).
29. Kelly, R. L. *et al.* The Breast Tumor Immune Microenvironment of DNA Double-Strand Break Repair Pathogenic Variant Carriers Is Enriched with Tumor-Associated Macrophages. *Cancer Epidemiol. Biomarkers Prev.* **34**, 868–874 (2025).
30. van Vugt, M. A. T. M. & Parkes, E. E. When breaks get hot: inflammatory signaling in BRCA1/2-mutant cancers. *Trends in Cancer* **8**, 174–189 (2022).
31. Cicchese, J. M. *et al.* Dynamic balance of pro- and anti-inflammatory signals controls disease and limits pathology. *Immunol. Rev.* **285**, 147–167 (2018).
32. Warrick, K. A., Vallez, C. N., Meibers, H. E. & Pasare, C. Bidirectional Communication Between the Innate and Adaptive Immune Systems. *Annu. Rev. Immunol.* **43**, 489–514 (2025).
33. Gebhardt, T., Palendira, U., Tschärke, D. C. & Bedoui, S. Tissue-resident memory T cells in tissue homeostasis, persistent infection, and cancer surveillance. *Immunol. Rev.* **283**, 54–76 (2018).
34. Beura, L. K. *et al.* CD4⁺ resident memory T cells dominate immunosurveillance and orchestrate local recall responses. *J. Exp. Med.* **216**, 1214–1229 (2019).
35. Robert, C. & Kupper, T. S. Inflammatory skin diseases, T cells, and immune surveillance. *N. Engl. J. Med.* **341**, 1817–1828 (1999).
36. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, (2019).
37. Martin, M. V. *et al.* The Neo-Open Reading Frame Peptides That Comprise the Tumor Framome Are a Rich Source of Neoantigens for Cancer Immunotherapy. *Cancer Immunol. Res.* **12**, 759–778 (2024).
38. Skorecki, K. L., Verkman, A. S. & Ausiello, D. A. Cross talk between stimulatory and inhibitory guanosine 5'-triphosphate binding proteins: role in activation and desensitization of the adenylate cyclase response to vasopressin. *Biochemistry* **26**, 639–645 (1987).
39. Nelde, A. *et al.* Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. *Cell. Mol. Life Sci.* **79**, (2022).
40. Pulver, E. M. *et al.* A BRCA1 Coiled-Coil Domain Variant Disrupting PALB2 Interaction Promotes the Development of Mammary Tumors and Confers a Targetable Defect in Homologous Recombination

- Repair. *Cancer Res.* **81**, 6171–6182 (2021).
41. Huber, F. & Bassani-Sternberg, M. Defects in antigen processing and presentation: mechanisms, immune evasion and implications for cancer vaccine development. *Nat. Rev. Immunol.* (2025) doi:10.1038/S41577-025-01208-8.
 42. Qdaisat, S. *et al.* Sensitization of tumours to immunotherapy by boosting early type-I interferon responses enables epitope spreading. *Nat. Biomed. Eng.* **2025 99 9**, 1437–1452 (2025).
 43. Mendez-Gomez, H. R. *et al.* RNA aggregates harness the danger response for potent cancer immunotherapy. *Cell* **187**, 2521–2535.e21 (2024).
 44. Launonen, I. M. *et al.* Single-cell tumor-immune microenvironment of BRCA1/2 mutated high-grade serous ovarian cancer. *Nat. Commun.* **13**, (2022).
 45. Yu, X. *et al.* Dissection of transcriptome dysregulation and immune characterization in women with germline BRCA1 mutation at single-cell resolution. *BMC Med.* **20**, (2022).
 46. Glodzik, D. *et al.* Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers. *Nat. Commun.* **11**, (2020).
 47. Li, J. *et al.* S100A9-CXCL12 activation in BRCA1-mutant breast cancer promotes an immunosuppressive microenvironment associated with resistance to immunotherapy. *Nat. Commun.* **13**, (2022).
 48. Bertwistle, D. *et al.* Nuclear location and cell cycle regulation of The BRCA2 protein. *Cancer Res.* **57**, 5485–5488 (1997).
 49. Samstein, R. M. *et al.* Mutations in BRCA1 and BRCA2 differentially affect the tumor microenvironment and response to checkpoint blockade immunotherapy. *Nat. cancer* **1**, 1188–1203 (2021).
 50. Jørgensen, N. *et al.* Tumour-infiltrating CD4-, CD8- and FOXP3-positive immune cells as predictive markers of mortality in BRCA1- and BRCA2-associated breast cancer. *Br. J. Cancer* **125**, 1388–1398 (2021).
 51. Corvigno, S. *et al.* Immune microenvironment composition in high-grade serous ovarian cancers based on BRCA mutational status. *J. Cancer Res. Clin. Oncol.* **147**, 3545–3555 (2021).
 52. Yuki, D., Lin, Y.-M., Fujii, Y., Nakamura, Y. & Furukawa, Y. Isolation of LEM domain-containing 1, a novel testis-specific gene expressed in colorectal cancers. *Oncol. Rep.* **12**, 275–80 (2004).
 53. Zhang, H., Xu, C., Jiang, F. & Feng, J. A Three-Genes Signature Predicting Colorectal Cancer Relapse Reveals LEMD1 Promoting CRC Cells Migration by RhoA/ROCK1 Signaling Pathway. *Front. Oncol.* **12**, (2022).
 54. Li, X. *et al.* LEM Domain Containing 1 Acts as a Novel Oncogene and Therapeutic Target for Triple-Negative Breast Cancer. *Cancers (Basel)*. **15**, (2023).
 55. Li, Q. *et al.* LEM domain containing 1 promotes proliferation via activating the PI3K/Akt signaling pathway in gastric cancer. *J. Cell. Biochem.* **120**, 15190–15201 (2019).

56. Li, L. & Zhang, P. Elevation of LEM Domain Containing 1 Predicts Poor Prognosis of NSCLC Patients and Triggers Malignant Stemness and Invasion of NSCLC Cells by Stimulating PI3K/AKT Pathway. *Curr. Mol. Med.* **24**, 366–378 (2024).
57. Takeda, R. *et al.* Identification and functional analysis of variants of a cancer/testis antigen LEMD1 in colorectal cancer stem-like cells. *Biochem. Biophys. Res. Commun.* **485**, 651–657 (2017).
58. Xu, M. *et al.* LEM domain containing 1 promotes thyroid cancer cell proliferation and migration by activating the Wnt/ β -catenin signaling pathway and epithelial-mesenchymal transition. *Oncol. Lett.* **21**, (2021).
59. Li, D., Wang, D., Liu, H. & Jiang, X. LEM domain containing 1 (LEMD1) transcriptionally activated by SRY-related high-mobility-group box 4 (SOX4) accelerates the progression of colon cancer by upregulating phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt) signaling pathway. *Bioengineered* **13**, 8087–8100 (2022).
60. Li, Z., Wang, J., Wu, J., Li, N. & Jiang, C. Long Noncoding RNA LEMD1-AS1 Increases LEMD1 Expression and Activates PI3K-AKT Pathway to Promote Metastasis in Oral Squamous Cell Carcinoma. *Biomed Res. Int.* **2022**, (2022).
61. Yang, Z. *et al.* EPYC functions as a novel prognostic biomarker for pancreatic cancer. *Sci. Rep.* **14**, (2024).
62. Deng, L., Wang, D., Chen, S., Hu, W. & Zhang, R. Epiphykan Predicts Poor Outcomes and Promotes Metastasis in Ovarian Cancer. *Front. Oncol.* **11**, (2021).
63. Vardabasso, C. *et al.* Histone variants: emerging players in cancer biology. *Cell. Mol. Life Sci.* **71**, 379–404 (2014).
64. Espinoza Pereira, K. N., Shan, J., Licht, J. D. & Bennett, R. L. Histone mutations in cancer. *Biochem. Soc. Trans.* **51**, (2023).
65. Lai, P. M., Gong, X. & Chan, K. M. Roles of Histone H2B, H3 and H4 Variants in Cancer Development and Prognosis. *Int. J. Mol. Sci.* **25**, (2024).
66. Espiritu, D. *et al.* Cancer histone mutations impact protein binding and DNA repair with possible links to genomic instability. *Nucleic Acids Res.* **53**, (2025).
67. Martire, S. & Banaszynski, L. A. The roles of histone variants in fine-tuning chromatin organization and function. *Nat. Rev. Mol. Cell Biol.* **21**, 522–541 (2020).
68. Vijayalakshmi, P. *et al.* Unveiling the guardians of the genome: The dynamic role of histones in DNA organization and disease. *Adv. Protein Chem. Struct. Biol.* **143**, 39–68 (2025).
69. Stok, C., Kok, Y. P., Van Den Tempel, N. & Van Vugt, M. A. T. M. Shaping the BRCAness mutational landscape by alternative double-strand break repair, replication stress and mitotic aberrancies. *Nucleic Acids Res.* **49**, 4239–4257 (2021).
70. Byrum, A. K., Vindigni, A. & Mosammaparast, N. Defining and Modulating 'BRCAness'. *Trends Cell Biol.* **29**, 740–751 (2019).

71. Liu, L. *et al.* BRCAness as a prognostic indicator in patients with early breast cancer. *Sci. Rep.* **10**, (2020).
72. Guo, M. & Wang, S. M. The BRCAness Landscape of Cancer. *Cells* **11**, (2022).
73. Tsyganov, M. M. *et al.* Determination of BRCAness Phenotype in Breast Tumors for the Appointment of Neoadjuvant Chemotherapy Based on Platinum and Taxanes. *Int. J. Mol. Sci.* **24**, (2022).

Figures

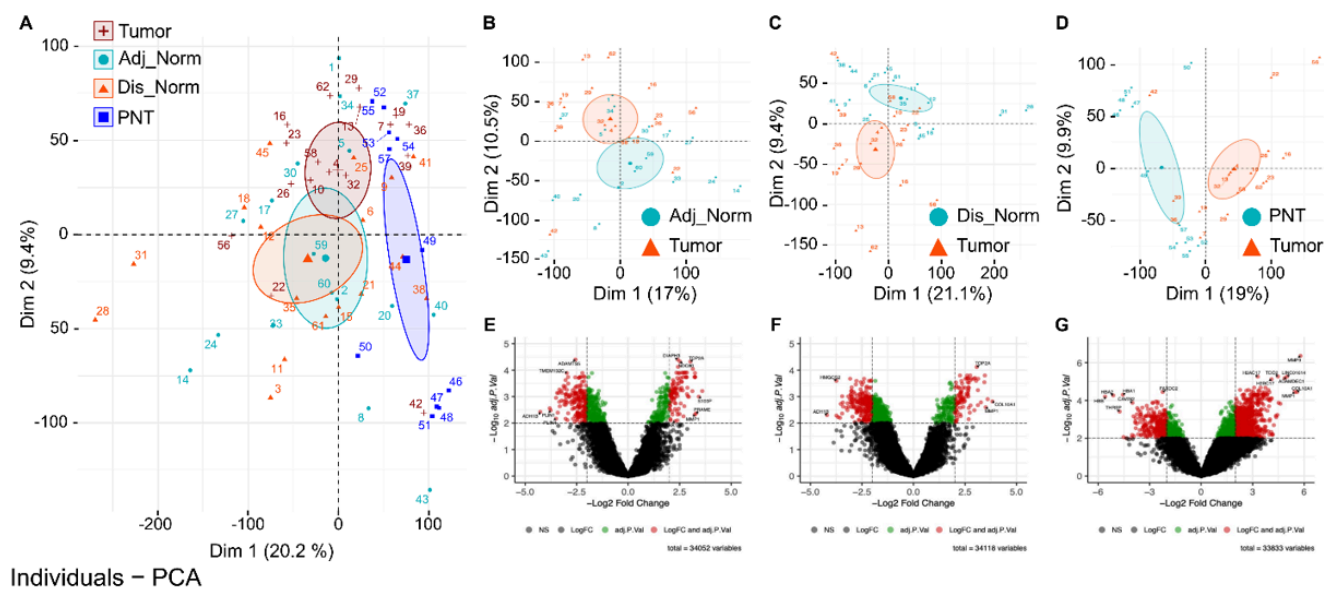


Figure 1

Global Transcriptomic Profiles and Differential Gene Expression in BRCA1-Mutant Breast Tissues. (A) Principal Component Analysis (PCA) plot depicting the global transcriptomic relationships among four study cohorts: tumor, adjacent normal (Adj_Norm), distant normal (Dis_Norm), and prophylactic normal (PNT) breast tissue samples. Each point represents a sample, with ellipses indicating 95% confidence intervals. Pairwise PCA plots illustrating the distinct separation of sample groups: (B) Tumor versus Adj_Norm; (C) Tumor versus Dis_Norm; and (D) Tumor versus prophylactic normal tissue (PNT). Volcano plots displaying differential gene expression for pairwise comparisons: (E) Tumor versus Adj_Norm; (F) Tumor versus Dis_Norm; and (G) Tumor versus PNT.

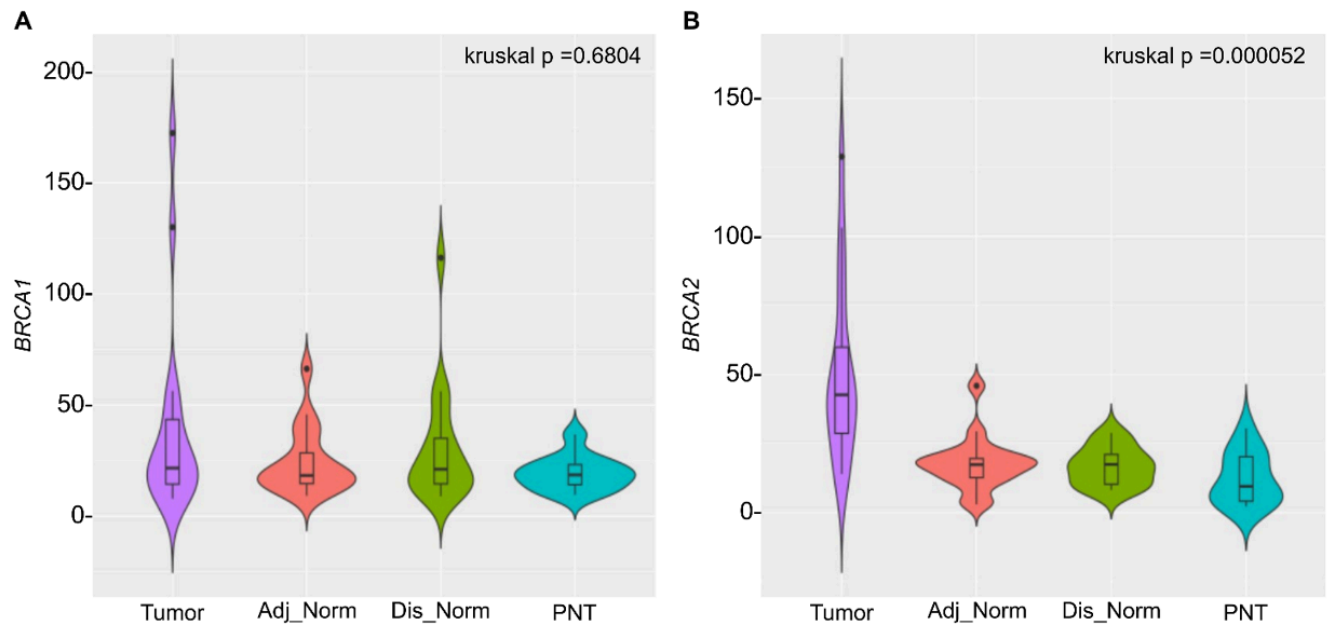


Figure 2

Expression of *BRCA1* and *BRCA2* in *BRCA1*-Mutant Breast Tissues. Violin plots showing the distribution of transcript levels for **(A) *BRCA1*** and **(B) *BRCA2*** across tumor, adjacent normal (Adj_Norm), distant normal (Dis_Norm), and prophylactic normal breast tissue (PNT). Statistical significance across all four groups was assessed using the Kruskal-Wallis test, with p-values shown for each gene.

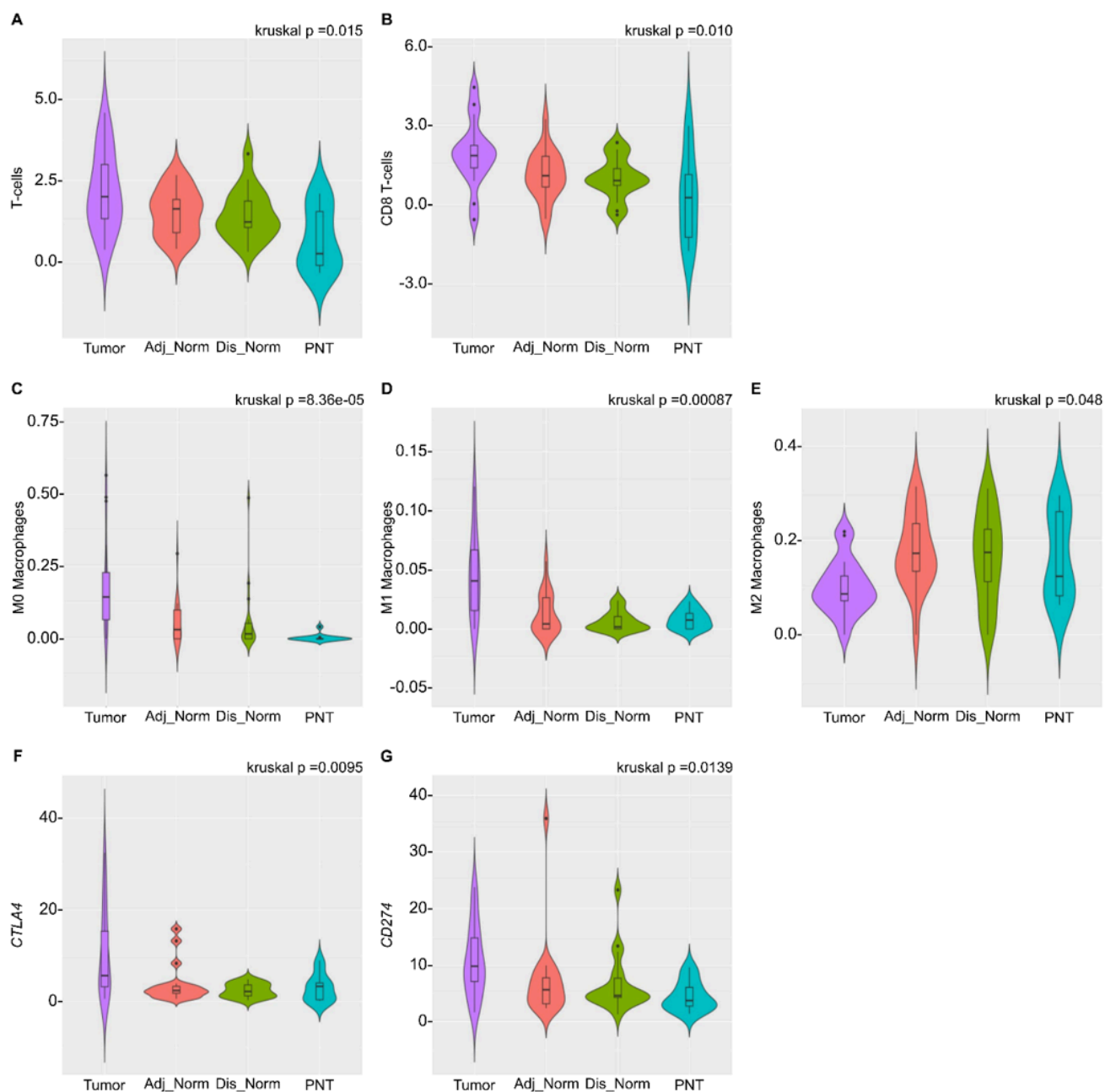


Figure 3

The Tumor Immune Microenvironment is Pro-Inflammatory but Suppressed. Violin plots showing (A) T-cell, (B) CD8 T-cell, (C) M0 macrophage, (D) M1 macrophage, and (E) M2 macrophage counts and (F) CTLA4 and (G) CD274 (PD-L1) transcript levels in BRCA1-mutant tumor, adjacent normal (Adj_Norm), distant normal (Dis_Norm), and prophylactic normal breast tissue (PNT). Statistical significance across all four groups was assessed using the Kruskal-Wallis test, with p-values shown for each panel.

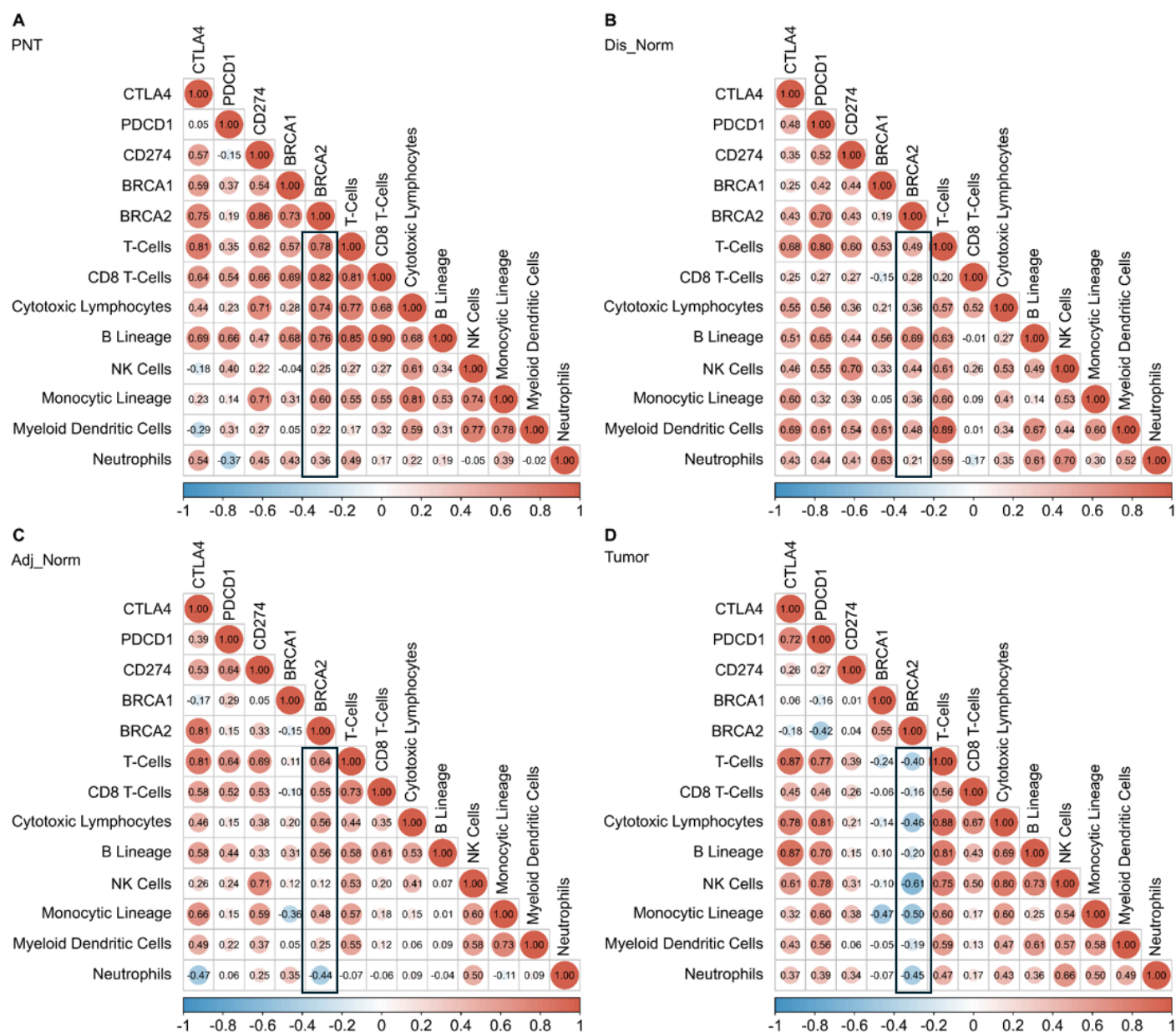


Figure 4

***BRCA2* Upregulation Correlates with an Immunologically Cold Tumor Microenvironment.** Correlation matrices showing the correlation of *BRCA1* and *BRCA2* transcript levels with immune cells estimated using the MCPcounter abundance scoring method for (A) Prophylactic normal breast tissue (PNT), (B) distant normal tissue (Dis_Norm), (C) adjacent normal tissue (Adj_Norm), and the tumor.

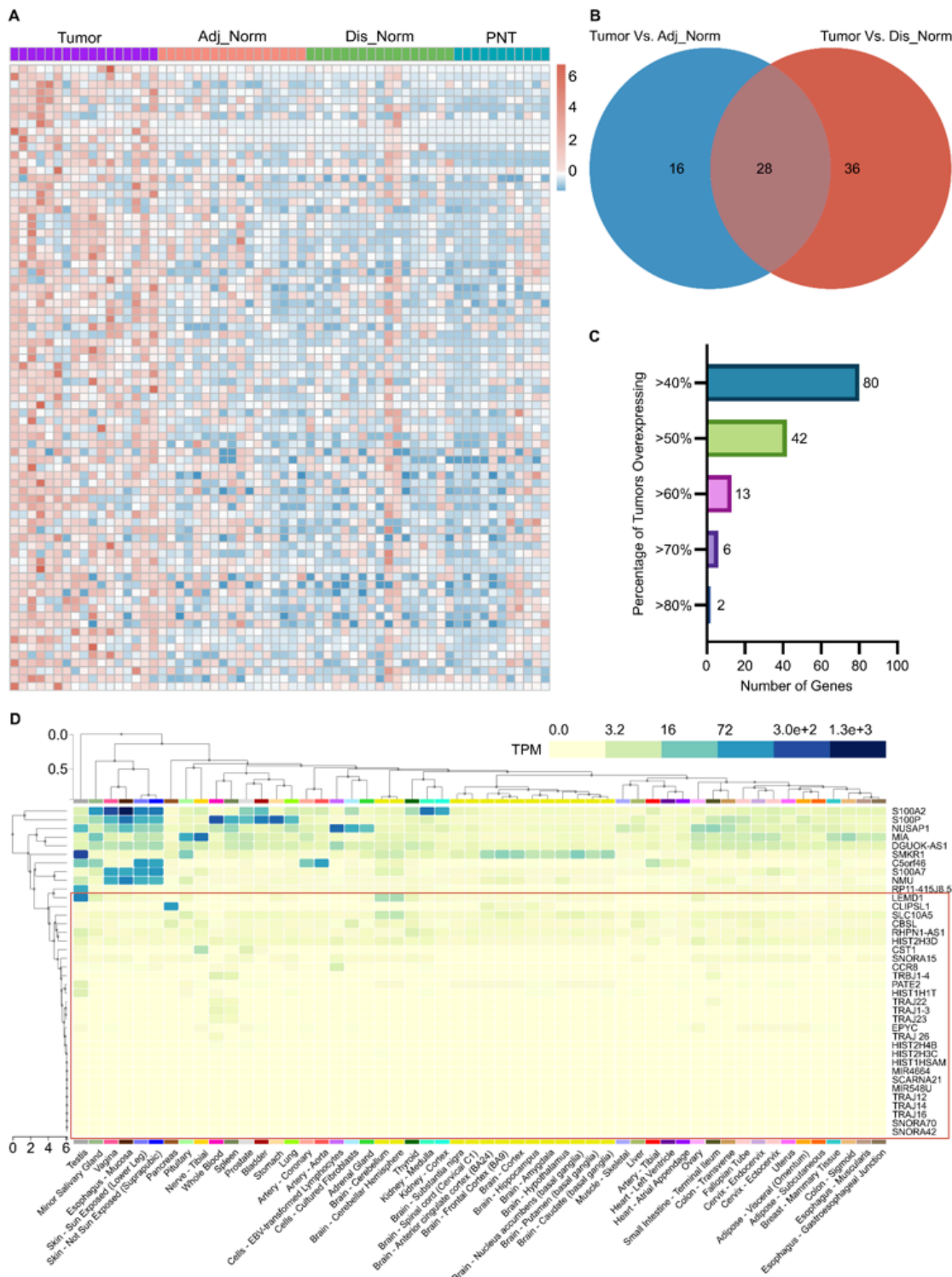


Figure 5

Identification and Characterization of an 80-Genes Tumor-Specific Signature in *BRCA1*-Mutant Breast Cancer. (A) Heatmap displaying the relative expression (row Z-score) of the 80 signature genes across the four sample cohorts: Tumor, Adj_Norm, Dis_Norm, and Prophylactic Normal Breast Tissue (PNT). (B) Venn diagram illustrating the overlap of differentially expressed genes (DEGs) identified from the comparison of tumor vs. adjacent normal tissues (Adj_Norm) and tumor vs. distant normal tissues

(Dis_Norm). **(C)** Bar chart quantifying the prevalence of the 80 signature genes. The bars show the number of genes overexpressed in greater than 40%, 50%, 60%, 70%, or 80% of the tumors analyzed. **(D)** Heatmap showing the expression levels in Transcripts Per Million (TPM) of the 80 signature genes across a panel of normal human tissues from the GTEx database. The red box highlights candidate genes with minimal expression in most non-mammary tissues, indicating high tumor specificity. The genes listed in the heatmap include alternative gene name for following: *H3C13* (*HIST2H3D*), *H3C14* (*HIST2H3C*), and *H4C15* (*HIST2H4B*).

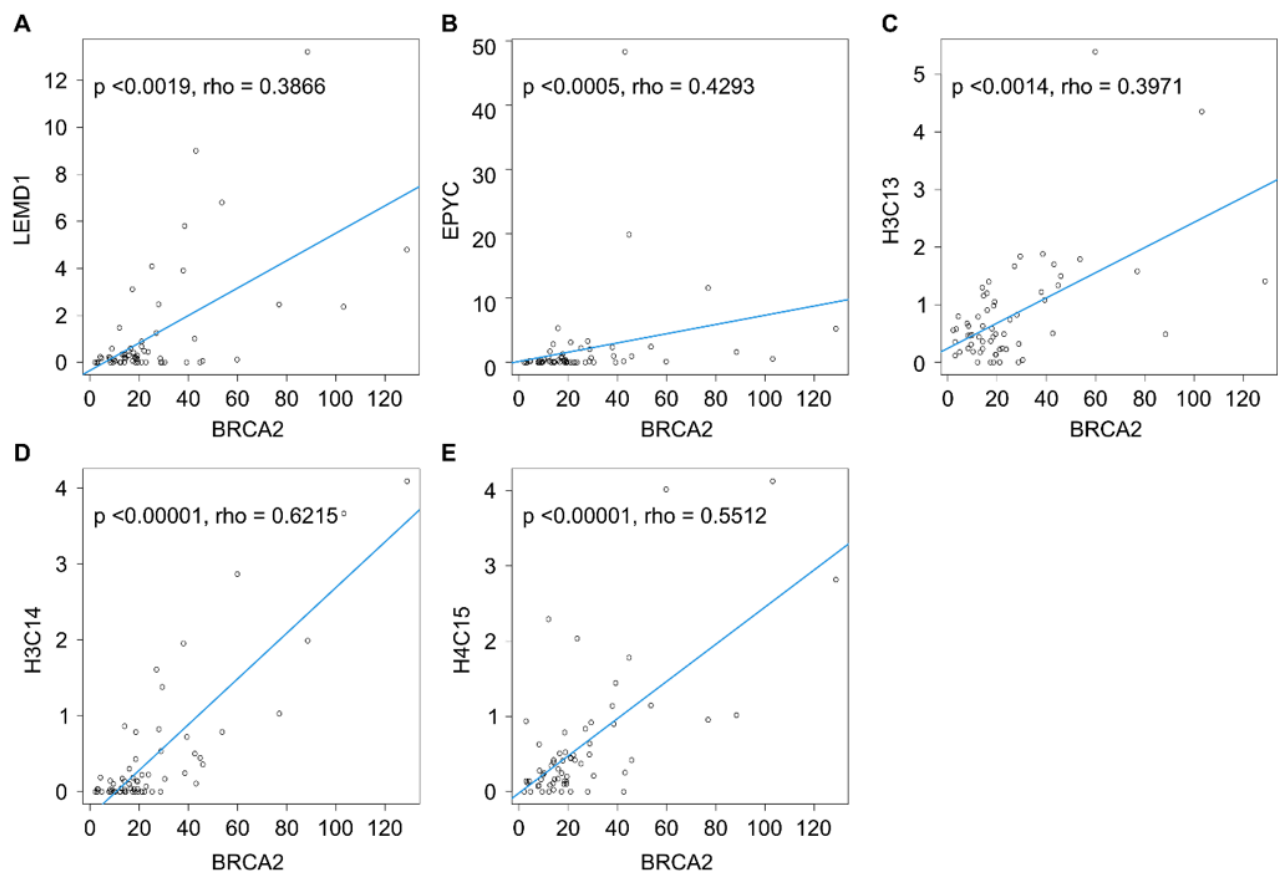


Figure 6

Vaccine Candidates Are Highly Expressed in Immunologically Cold Tumors. Scatter plots showing the correlation between *BRCA2* transcript level and the expression of individual lead candidate antigens **(A)** *LEMD1*, **(B)** *EPYC*, **(C)** *H3C13*, **(D)** *H3C14*, and **(E)** *H4C15* in *BRCA1*-mutant breast tumor samples. Rho values represent Spearman's rank correlation coefficient.

Supplementary Files

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

- [SupplementaryTable1.xlsx](#)

- [SupplementaryTable2.xlsx](#)
- [SupplementaryFiguresandFigurelegendsFinal260102.docx](#)