Germline microRNA-based variants and lymph node burden in breast cancer

J. Weidhaas[1], K. McGreevy[2], J. Le[3], M. Alcaraz[1], E. Rietdorf[1], M. Ensenyat-Mendez[4], J. Baker[5], M. DiNome[4], D. Marzese[4], D. Telesca[6]
[1]Radiation Oncology, UCLA, Los Angeles, CA, [2]Biostatistics, MiraDx, Los Angeles, CA, [3]Surgery, UCSD, San Diego, CA, [4]Surgery, Duke, Durham, NC,
[5]Surgical Oncology, UCLA, Los Angeles, CA, [6]Department of Biostatistics, UCLA, Los Angeles, CA.

Background

Axillary lymph nodal burden is a central prognostic and treatment-defining factor in breast cancer management[1]. In addition, germline variants in microRNAs (mirSNPs) have been found to shape tumor biology in breast cancer [2] and to predict response to therapy across cancer types [3][4].

In this study, we tested whether **mirSNPs can distinguish patients with high versus low axillary nodal burden** in invasive ductal carcinoma (IDC), with the specific goal of identifying germline determinants that can be integrated into a **multi-SNP model** for predicting this high-burden and poor prognosis phenotype.

Cohort Demographics

The study cohort consisted of **73 patients with IDC**, where most patients had HR+/HER2- breast cancer (72.6%), 15.1% were HER2+, and 12.3% had TNBC.

Most patients were clinically node positive at diagnosis (91.8%), and 42 patients (57.5%) had four or more positive lymph nodes (≥pN2).

The average BMI was 27.3 (SD 7.3). cT2 was the most common clinical T stage (57.5%), and the mean tumor size at diagnosis was 28.3 mm (SD 19.7).

Table 1. IDC Breast Cancer Patients' Demographics

	≤3 Positive Nodes (N=31)	≥ 4 Positive Nodes (N=42)
Age at diagnosis		
Mean (SD)	60.9 (10.9)	57.8 (14.1)
Race		
Asian	8 (25.8%)	9 (21.4%)
Black	4 (12.9%)	5 (11.9%)
Hispanic	3 (9.7%)	6 (14.3%)
Middle Eastern	1 (3.2%)	5 (11.9%)
White	15 (48.4%)	17 (40.5%)
Breast Cancer Status		
HR+/HER2-	21 (67.7%)	32 (76.2%)
HER2+	5 (16.1%)	6 (14.3%)
TNBC	5 (16.1%)	4 (9.5%)
Grade		
Mean (SD)	2.42 (0.62)	2.32 (0.72)
1	2 (6.5%)	6 (14.3%)
2	14 (45.2%)	16 (38.1%)
3	15 (48.4%)	19 (45.2%)
Ki67		
Mean (SD)	32.1 (22.1)	26.9 (18.7)

Methods

DNA was isolated from microdissected tumor specimens from IDC patients who underwent axillary lymph node dissection. Nodal burden was the outcome of interest, which was dichotomized into high (**≥pN2, ≥4 positive nodes**) versus low/no burden (**<pN2, 0-3 positive nodes**).

22 mirSNPs were pre-filtered for having a marginal association with nodal burden (Fisher exact p-value < 0.2). mirSNPs to build the final model were further reduced by iteratively removing the bottom five predictors by AUC impact and re-fitting models, resulting in a **final set of 17 mirSNPs**. We performed a Gene Ontology (GO) analysis with these mirSNPs against a universal background.

Clinical variables were included if they had non-zero median relative importance across five leave-one-out cross-validation (LOOCV) rounds. This resulted in retention of age at diagnosis, tumor size, BMI, race, Ki67, clinical T stage, cancer subtype (TNBC, HER2+, HR+/HER2-), tumor grade, lobular features, and use of adjuvant therapies (endocrine, chemotherapy, radiotherapy). These selected clinical features were then used alone and in conjunction with mirSNP genetic data in subsequent predictive modeling to compare predictive model performance.

mirSNP Genetic Model Performance

Elastic net regression using the 17 mirSNPs yielded the strongest predictive accuracy (AUC 0.793) for high nodal burden (**Table 2**). When mirSNPs were combined with clinical variables the resulting performance was lower than for genetics alone (AUC reducing to 0.71). Models using clinical variables only performed the worst, achieving an AUC of only 0.55.

Table 2. LOOCV Performance Metrics Predicting ≥pN2

Variables	Sensitivity	Specificity	PPV	NPV	AUC
Clinical Factors Only	0.610	0.484	0.610	0.484	0.547
17 mirSNPs Only	0.714	0.871	0.882	0.692	0.793
SNPs + Reduced Clin	0.810	0.613	0.739	0.704	0.711

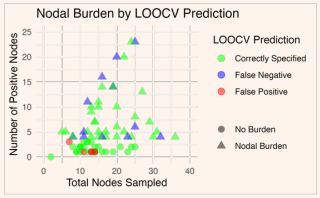


Figure 1. The predictive accuracy did not correlate with the number of nodes sampled, and misclassifications by the model were predominantly concentrated around the clinical threshold of four positive nodes- rather than being biased by the depth of nodal sampling.

Genetic Factors Important for Nodal Burden

The most significant mirSNP was in *P2RX7*, a gene whose expression has been previously associated with the risk of progression in breast cancer [5], and the second most significant mirSNP in *CSMD1*, a gene that is associated with outcome in breast cancer [6] (**Table 3**). Notably, a variant in *KRAS*, which predicts breast cancer risk [2], is also important in this model.

Table 3. Proportion of Patients with High Nodal Burden (≥pN2) by Genotype

mirSNP	Homo-zygous wildtype	Hetero-zygous	Homo-zygous mutant	Fisher p-value
P2RX7 A>C exonic	46.5%	80.0%	100%	0.015
CSMD1 T>C 3' UTR	54.1%	100%	50.0%	0.023

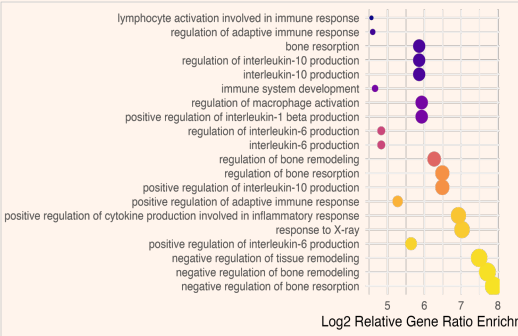


Figure 2. Top Enriched GO Processes from mirSNP genetic model. A total of 148 pathways were enriched. Terms included bone remodeling, somatic recombination of immune receptors, and cytokine regulation. The prominence of bone-related pathways is notable, as breast cancer frequently metastasizes to bone, potentially suggesting that inherited genetic variation may contribute to this metastatic pattern.

Conclusions

Our findings support the potential of mirSNPs as an additional tool to help identify patients at higher risk of high nodal burden in breast cancer. Incorporation of these findings with other biomarkers of lymph node burden and further validation are warranted, as these biomarkers could give prognostic information as well as be a potential path to identify targets for future breast cancer treatments.

Contact

Joanne B. Weidhaas, MD, PhD, MSM
Organizations: University of California, Los Angeles
Email: jweidhaas@mednet.ucla.edu

References

1. Dings et al. The Prognostic Value of Lymph Node Ratio in Node-Positive Breast Cancer: A Dutch Nationwide Population-Based Study. *Breast Oncology, 20, 2013.*
2. Paranjape et al. A 3'-untranslated region *KRAS* variant and triple-negative breast cancer: a case-control and genetic analysis. *The Lancet Oncology, 12, 2011.*
3. Ratner et al. A *KRAS*-variant is a biomarker of poor outcome, platinum chemotherapy resistance and a potential target for therapy in ovarian cancer. *Oncogene, 31, 2012.*
4. Kishan AU, et al. Validation and Derivation of MicroRNA-based Germline Signatures Predicting Radiation Toxicity in Prostate Cancer. *Clin Cancer Res. April 2025.*
5. Du et al. Role of the P2X7 receptor in breast cancer progression. *Purinergic Signaling, 21, 2025.*
6. Kamal et al. Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. *Breast Cancer Research, 121, 2010.*