Beyond the tumor: a CAF-Driven Gene Signature for Early Clinical Stratification of Metastasis Risk in Breast Cancer

D. P. Barrera[1], V. Toledo[2], J. S. Contreras-Riquelme[1], J. Sapunar[2], B. Chahuán[2], L. Moyano[3], B. Prieto[1], J. Cerda-Infante[1,4]

Environ[1], Fundación Arturo López Pérez[2], Instituto Nacional del Cáncer[3], Facultad de Medicina Pontificia Universidad Católica de Chile[4]

PS2-09-11

Abstract

Introduction. Despite advances in breast cancer management, tools for early and accurate prediction of metastatic risk remain insufficient. Current models focus mainly on tumor-intrinsic features and overlook the tumor microenvironment. Cancer-associated fibroblasts (CAF), major stromal drivers of ECM remodeling, EMT, angiogenesis, and immune suppression, play a central role in metastatic progression. Given their biological relevance, we hypothesized that CAF from primary tumors harbor a transcriptomic signature with independent prognostic value for distant metastasis. **Methods.** We analyzed 182 FFPE breast tumor biopsies with ≥5 years of follow-up across all molecular subtypes. Reactive stroma was quantified on H&E and Masson's trichrome slides using QuPath, and its association with MFS and OS was evaluated through Kaplan–Meier and Cox models. RNA from FFPE tissue and CAF-enriched explants underwent differential expression analysis, identifying 233 CAF-related genes. A 10-gene prognostic signature was generated using LASSO regression and validated in an independent cohort and external datasets, with performance assessed by AUC, specificity, and Cox modeling. **Results.** A reactive stroma content >53.2% was linked to worse outcomes, including a 3.75-fold higher risk of distant metastasis. The 10-gene CAF signature—enriched strong predictive accuracy (AUC 0.938 in training; 0.956 and 100% specificity in external datasets). Patients positive for the signature had a 19.95-fold increased metastasis risk, independent of receptor status, grade, or tumor size. The signature remained prognostic across molecular subtypes and sequencing platforms. **Conclusion.** This study underscores the prognostic importance of the stromal compartment, particularly CAF, in breast cancer metastasis. The validated 10-gene signature offers a novel TME-based biomarker for early risk stratification, enabling earlier identification of high-risk patients and supporting more personalized therapeutic decisions to improve outcomes. **Acknowledgments**. CORFO 22CVID-206707; ANID VIU24P0138.

Results

Cohort Selection Based on FFPE RNA Quality and Stromal Content

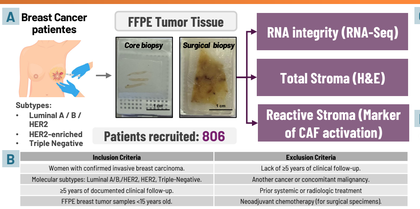


Figure 1. Cohort selection, FFPE pre-analytical evaluation, stromal quantification, and fibroblast marker assessment. A) Workflow for cohort assembly: 806 breast cancer biopsies from different molecular subtypes were collected from core biopsies or surgical specimens. All samples underwent RNA extraction for RNA-Seq and pre-analytical quality evaluation. Histopathological analysis included quantification of total stroma on H&E (as a proxy for tumor stroma enriched in CAF) and reactive stroma on Masson Trichrome (as a marker of fibroblast activation). **B)** Inclusion and exclusion criteria applied to define the analytical cohort. **C)** RNA integrity measured by RIN, showing that only samples with RIN ≥ 1.8 were selected for downstream transcriptomic analysis. **D)** DV200 distribution showing the proportion of RNA fragments >200 nucleotides; samples with DV200 ≥ 30% were retained for RNA-Seq. **E)** Digital pathology workflow for quantifying total stroma on H&E slides using QuPath. (F) Representative H&E images and QuPath predictions illustrating stromal content categories (<25%, ~50%, and >75%). **G)** Pathologist validation of the automated stromal quantification. **H)** Frequency distribution of stromal percentages across the cohort. **I–J)** Relative expression of fibroblast markers VIM and P4H across stroma categories used to identify the stromal range enriched in CAF and suitable for RNA-Seq selection.

Reactive Stroma Prognosticates Metastasis Risk Across BC Subtypes

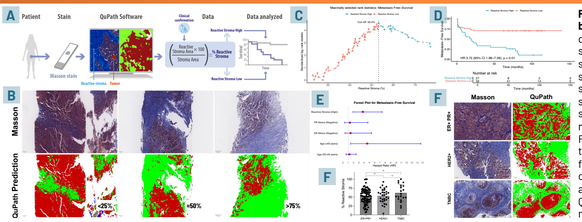


Figure 2. Prognostic value of reactive stroma across breast cancer subtypes. A) Workflow for automated quantification of reactive stroma using Masson's trichrome staining and QuPath. **B)** Representative breast tumor samples, low, intermediate, and high levels of reactive stroma. **C)** Analysis of metastasis-free survival (MFS) in a subset of patients (n = 85). The optimal cutoff for reactive stroma was determined at 53.2%. **D)** Kaplan–Meier curves showing significantly reduced MFS in patients with high reactive stroma (HR = 3.75; 95% CI: 1.98–7.09; p < 0.01). **E)** Forest plot from multivariate Cox regression demonstrating that reactive stroma is an independent prognostic factor for MFS. **F)** Representative Masson's trichrome and corresponding QuPath segmentations for ER+/PR+, HER2+, and triple-negative breast cancer (TNBC) tumors. **G)** Quantitative comparison of reactive stromal content across molecular subtypes revealed no significant differences.

RNA Integrity and Stromal Content Enable High-Quality FFPE Transcriptomics

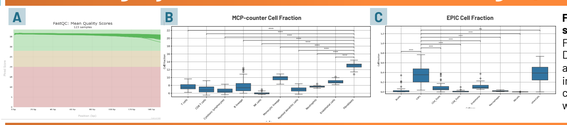


Figure 3. Sequencing quality and cell-type composition in FFPE samples. A) FastQC high and stable read quality across all 123 FFPE samples, confirming that pre-analytical criteria (RIN ≥ 1.8; DV200 ≥ 30%) yield reliable sequencing output. **B)** MCP-counter analysis indicates higher fibroblast abundance compared with immune cell populations. **C)** EPIC deconvolution independently confirms fibroblasts as the predominant cell type, consistent with the high reactive stroma quantified histologically.

CAF-Derived Genes Uncover a Metastasis-Associated Program in FFPE Tumors

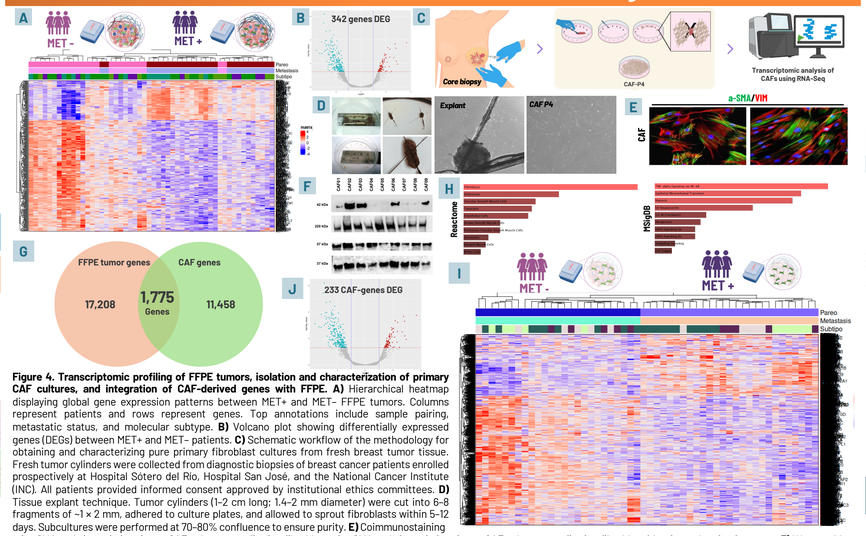


Figure 4. Transcriptomic profiling of FFPE tumors, isolation and characterization of primary CAF cultures, and integration of CAF-derived genes with FFPE. A) Hierarchical heatmap displaying global gene expression patterns between MET+ and MET− FFPE tumors. Columns represent patients and rows represent genes. Top annotations include sample pairing, metastatic status, and molecular subtype. **B)** Volcano plot showing differentially expressed genes (DEGs) between MET+ and MET− patients. **C)** Schematic workflow of the methodology for obtaining and characterizing pure primary fibroblast cultures from fresh breast tumor tissue. Fresh tumor cylinders were collected from diagnostic biopsies of breast cancer patients enrolled prospectively at Hospital Sótero del Río, Hospital San José, and the National Cancer Institute (INC). All patients provided informed consent approved by institutional ethics committees. **D)** Tissue explant technique. Tumor cylinders (1–2 cm long; 1.4–2 mm diameter) were cut into 6–8 fragments of ~1 × 2 mm, adhered to culture plates, and allowed to sprout fibroblasts within 5–12 days. Subcultures were performed at 70–80% confluence to ensure purity. **E)** Coimmunostaining of α-SMA and vimentin in primary CAF cultures, confirming fibroblast of α-SMA and vimentin in primary CAF cultures, confirming fibroblast identity and activation state. **F)** Western blot characterization of CAF cultures showing expression of fibroblast markers FN, α-SMA, and VIM, with GAPDH as loading control. **G)** Venn diagram showing overlap between highly expressed CAF genes and genes expressed in whole FFPE tumor tissue. **H)** Functional enrichment analysis of the 1,775 CAF-associated genes using Reactome pathways and MSigDB gene sets. **I)** Heatmap of the 1,775 CAF-associated genes in FFPE patient samples, highlighting clear separation between MET− and MET+ groups. Top annotations depict patient pairing, metastatic status, and molecular subtype. **J)** Volcano plot of the 233 differentially expressed genes (41 upregulated and 192 downregulated in MET+ patients) based on log2 fold change.

LASSO Modeling Defines a CAF-Driven 10-Gene Prognostic Signature

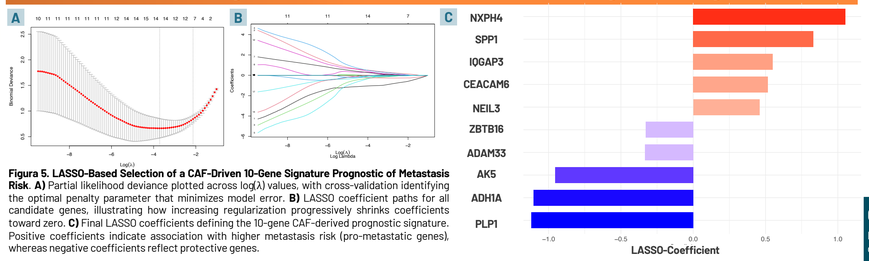


Figure 5. LASSO-Based Selection of a CAF-Driven 10-Gene Signature Prognostic of Metastasis Risk. A) Partial likelihood deviance plotted across log(λ) values, with cross-validation identifying the optimal penalty parameter that minimizes model error. **B)** LASSO coefficient paths for all candidate genes, illustrating how increasing regularization progressively shrinks coefficients toward zero. **C)** Final LASSO coefficients defining the 10-gene CAF-derived prognostic signature. Positive coefficients indicate association with higher metastasis risk (pro-metastatic genes), whereas negative coefficients reflect protective genes.

Validation Confirms the Prognostic Performance of the CAF-Driven

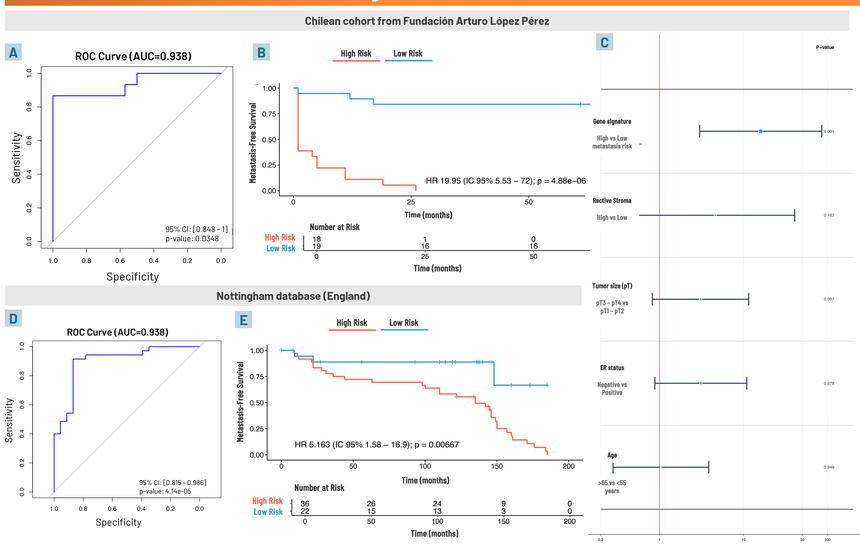


Figure 6. Validation confirms the prognostic performance of the CAF-driven 10-gene signature. A) ROC curve for the 10-gene signature in the Chilean cohort from Fundación Arturo López Pérez, showing its ability to discriminate patients at high versus low risk of metastasis. **B)** Kaplan–Meier curve for metastasis-free survival in the same Chilean cohort, stratified by the 10-gene signature into high- and low-risk groups (with corresponding HR and 95% CI). **C)** Multivariate Cox regression forest plot for metastasis-free survival in the Chilean cohort. **D)** ROC curve for the 10-gene signature in the independent Nottingham dataset, confirming its prognostic performance in an external cohort. **E)** Kaplan–Meier curve for metastasis-free survival in the Nottingham dataset, showing separation between high- and low-risk groups defined by the 10-gene signature.

Conclusions



- **Stromal determinant of metastasis**: Reactive stroma in primary breast tumours is a strong, subtype-independent prognostic indicator of metastasis risk.
- **CAF-driven biology captured in FFPE:** A robust CAF-associated transcriptomic program is detectable in FFPE tumours and differentiates metastatic from non-metastatic patients.
- **10-gene prognostic signature:** LASSO modelling identified a CAF-driven 10-gene signature that accurately stratifies patients into high- and low-risk groups.
- **Cross-cohort validation:** The signature shows high predictive performance in both the Chilean clinical cohort (FALP) and the independent Nottingham dataset.
- **Clinical translation potential:** This CAF-based signature offers a feasible pathway toward early metastasis risk stratification and personalised therapeutic decision-making at the time of diagnosis.

Contact

Daniela P. Barrera Álvarez
daniela@environ.bio

Javier Cerda-Infante
javier@environ.bio

This presentation is the intellectual property of the author/presenter. Contact them at javier@environ.bio for permission to reprint and/or distribute.