

## An integrated genomics approach for identifying breast cancer patients with highly aggressive tumors

Breast cancer is the most common cancer in women with over 2 million new cases worldwide in 2018. Three kinds of receptors are usually found on the surface of breast cancer cells: a receptor for female hormone estrogen (ER), a receptor for female hormone progesterone (PR), and a receptor for human epidermal growth factor (HER2). A number of drugs have been developed that block these receptors and kill the cancer cells. However, for a type of breast cancer called triple negative breast cancer (TNBC), the tumor cells lack all three receptors. As a result, tumors in the TNBC category cannot be treated with any of the drugs designed against these receptors. TNBC tumors tend to be not only untreatable, but also highly aggressive. Only 14% of patients with TNBC survive for 5 or more years after being diagnosed. TNBC disproportionately affects young women and women of African and Hispanic descent.

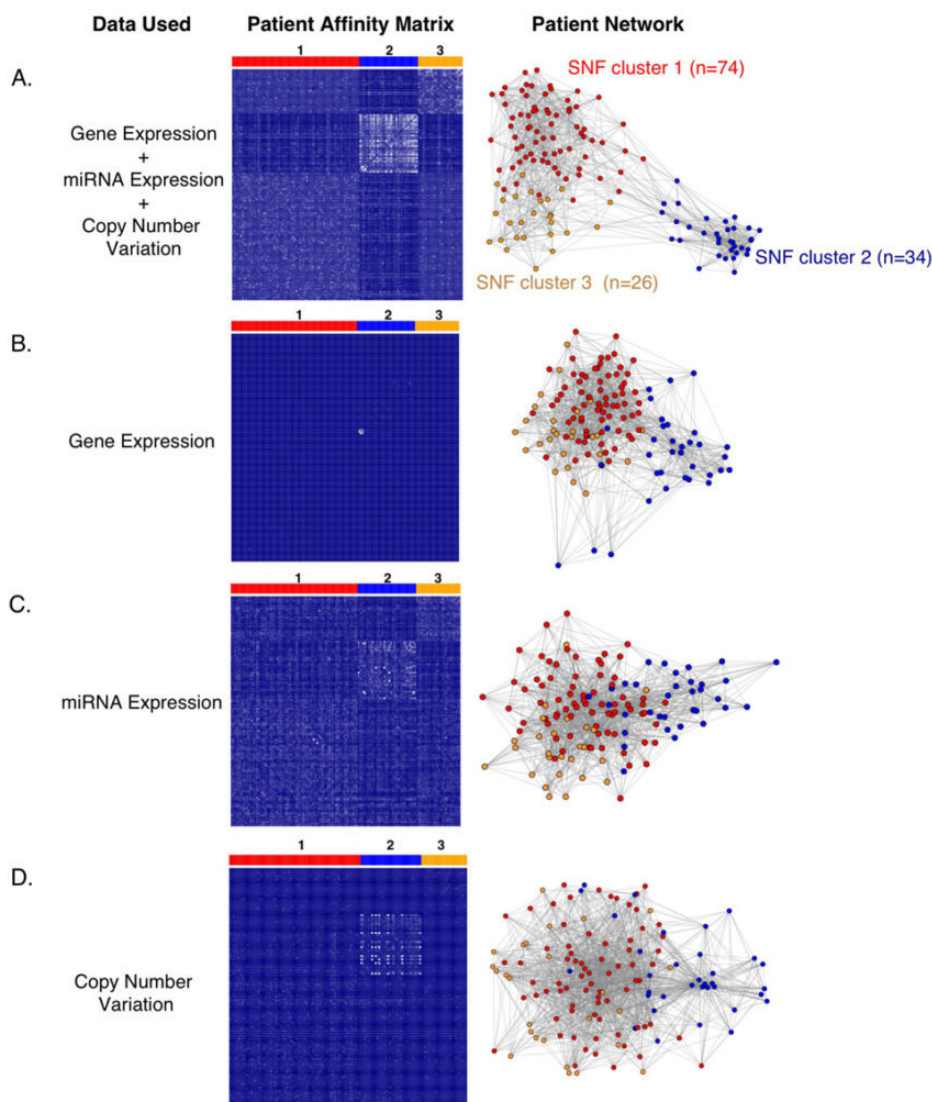


Fig. 1. Integrative clustering of patient tumors (A) Patient tumors were clustered based on a combination of gene expression, miRNA expression and copy number variants. A heatmap (left) was generated by combining gene expression, miRNA expression, and CNV data. The white and blue regions represent high and low similarities between patients. On the right, the patients are represented as a network with patients as nodes (dots) connected by edges (lines). Long edges connect patients with low similarity, while short edges connect patients with high similarity. (B, C, and D) Heatmaps are shown on the left and networks are shown on the right for clusters developed based on a single-data type: gene expression data (B), miRNA data (C), or copy number variant data (D).

Previous studies have shown that the tumors from different TNBC patients can be classified into groups based on the expression of specific genes in the patients' cancer cells. This patient-to-patient variation in TNBC (called inter-tumor heterogeneity) is important to study to understand why some patients with TNBC survive longer than other patients, and for the development of personalized therapies.

We asked whether we could understand the inter-patient differences among TNBCs better if we use multiple different types of genomic data and combine it with gene expression data for classifying the tumors.

We considered two other parameters that varied across the different TNBCs: 1) expression of small RNA molecules called microRNAs that specifically target mRNAs for degradation resulting in no protein expression; and 2) increase in copy number of genes or deletion of genes in DNA (referred to as copy number variation or CNV). Clustering TNBC patients by integrating the three parameters of gene expression, microRNA expression, and CNV lead to three patient clusters that are better separated than clusters obtained from any one parameter by itself. One of the three TNBC clusters we identified based on all of the data available had significantly poorer survival compared to a large group of breast cancer patients, suggesting we had identified a subset of TNBC patients at particularly high risk.

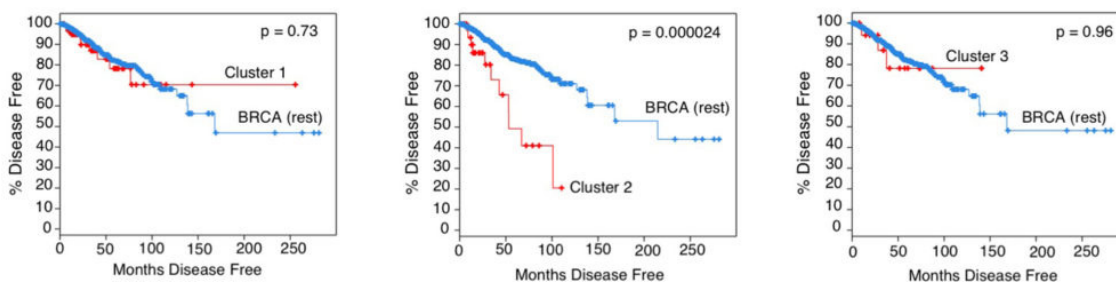


Fig. 2. Comparison of disease-free survival of patients in an integrative cluster with the TCGA breast cancer patient cohort (breast invasive carcinomas, n=1098) excluding the patients associated with this integrative cluster.

We looked further into this group of TNBC patients with particularly poor survival and found it to be associated with certain DNA mutations and CNVs and gene expression changes compared to the other two clusters. Finally, we implemented a program that assigned a new, unknown TNBC tumor to one of the three clusters with high accuracy. Our study provides a framework for understanding the differences among individuals with tumors at the same site, and provides a new classification scheme for TNBC that is predictive of tumor outcome.

**Mithun Mitra**<sup>1,2</sup>, **Alec M. Chiu**<sup>3</sup>, **Hilary A. Collier**<sup>1,2,3,4</sup>

<sup>1</sup>Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, USA

<sup>2</sup>Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, USA

<sup>3</sup>Bioinformatics Interdepartmental Program, University of California, Los Angeles, USA

<sup>4</sup>Molecular Biology Institute, University of California, Los Angeles, USA

## **Publication**

[Integrative analysis of the inter-tumoral heterogeneity of triple-negative breast cancer.](#)

Chiu AM, Mitra M, Boymoushakian L, Coller HA

*Sci Rep. 2018 Aug 7*