Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome

David Venet[1], Jacques E. Dumont[2], Vincent Detours[2,3]*

1 IRIDIA-CoDE, Université Libre de Bruxelles (U.L.B.), Brussels, Belgium, 2 IRIBHM, Université Libre de Bruxelles (U.L.B.), Campus Erasme, Brussels, Belgium, 3 WELBIO, Université Libre de Bruxelles (U.L.B.), Campus Erasme, Brussels, Belgium

Abstract

Bridging the gap between animal or *in vitro* models and human disease is essential in medical research. Researchers often suggest that a biological mechanism is relevant to human cancer from the statistical association of a gene expression marker (a signature) of this mechanism, that was discovered in an experimental system, with disease outcome in humans. We examined this argument for breast cancer. Surprisingly, we found that gene expression signatures—unrelated to cancer—of the effect of postprandial laughter, of mice social defeat and of skin fibroblast localization were all significantly associated with breast cancer outcome. We next compared 47 published breast cancer outcome signatures to signatures made of random genes. Twenty-eight of them (60%) were not significantly better outcome predictors than random signatures of identical size and 11 (23%) were worst predictors than the median random signature. More than 90% of random signatures >100 genes were significant outcome predictors. We next derived a metagene, called meta-PCNA, by selecting the 1% genes most positively correlated with proliferation marker PCNA in a compendium of normal tissues expression. Adjusting breast cancer expression data for meta-PCNA abrogated almost entirely the outcome association of published and random signatures. We also found that, in the absence of adjustment, the hazard ratio of outcome association of a signature strongly correlated with meta-PCNA ($R^2 = 0.9$). This relation also applied to single-gene expression markers. Moreover, >50% of the breast cancer transcriptome was correlated with meta-PCNA. A corollary was that purging cell cycle genes out of a signature failed to rule out the confounding effect of proliferation. Hence, it is questionable to suggest that a mechanism is relevant to human breast cancer from the finding that a gene expression marker for this mechanism predicts human breast cancer outcome, because most markers do. The methods we present help to overcome this problem.