

## VIEWPOINT

# Prognostic gene network modules in breast cancer hold promise

Andrew E Teschendorff<sup>\*1</sup>, Yan Jiao<sup>1</sup> and Carlos Caldas<sup>\*2,3</sup>

### Abstract

A substantial proportion of lymph node-negative patients who receive adjuvant chemotherapy do not derive any benefit from this aggressive and potentially toxic treatment. However, standard histopathological indices cannot reliably detect patients at low risk of relapse or distant metastasis. In the past few years several prognostic gene expression signatures have been developed and shown to potentially outperform histopathological factors in identifying low-risk patients in specific breast cancer subgroups with predictive values of around 90%, and therefore hold promise for clinical application. We envisage that further improvements and insights may come from integrative expression pathway analyses that dissect prognostic signatures into modules related to cancer hallmarks.

### Background

An outstanding problem in the clinical management of breast cancer is overtreatment. It is estimated that approximately 55 to 75% of breast cancer patients who receive adjuvant chemotherapy would do equally well without it [1], but identifying this low-risk population with a high enough predictive value ( $\geq 90\%$ ) is not possible using standard prognostic factors such as lymph node status or tumour size. Several recently developed gene expression classifiers have shown promise of achieving the required predictive values.

One such classifier is Oncotype DX, a prognostic test based on the expression levels of 21 genes, which has

been shown to identify low-risk patients with an accuracy of at least 90%, but is restricted to lymph node-negative oestrogen receptor-positive (ER+) breast cancer [2]. Another classifier is the 7 gene immune response (IR) module, which allows identification of low-risk patients in oestrogen receptor-negative (ER-) breast cancer [3]. Both of these signatures appear to be robust, demonstrating a high predictive value across many different breast cancer cohorts [2,3]. Gene Ontology (GO) analyses of prognostic signatures [2-6] have shown that specific biological processes play particularly important roles and that this is subgroup-specific. Thus, while cell-proliferation is strongly prognostic in ER+ breast cancer [6], the clinical heterogeneity of ER- breast cancers appears to be explained mainly by differential expression of genes related to immune response pathways, highlighting the need to conduct survival analysis within specific breast cancer subgroups [7-10].

### The article

In line with this, Li and colleagues [11] have recently conducted a novel bioinformatic analysis of existing breast cancer expression data sets in order to identify gene expression modules that may predict patients at low risk of distant metastasis in specific breast cancer subgroups. A common difficulty in identifying robust prognostic gene signatures is the presence of noise and spurious signals, which often render the resulting gene signatures unstable to perturbations. To overcome these limitations, Li and colleagues used a novel multiple survival screening (MSS) algorithm, designed to extract the most robust signals from the data [11]. Using this algorithm, Li and colleagues report novel prognostic gene modules of high negative predictive values (87 to 100%) and that are related to cancer hallmarks, notably cell-cycle, apoptosis and cell-adhesion [11]. The modules were derived using only one data set [12], and while individual modules were not prognostic across all eight validation sets examined, specific non-linear combinations of these modules were. This suggests that classifiers built from non-linear combinations of modules related to distinct cancer hallmarks may yield more powerful predictors than those based on individual modules or

\*Correspondence: a.teschendorff@ucl.ac.uk; cc234@cam.ac.uk

<sup>1</sup>Medical Genomics Group, Paul O'Gorman Building, UCL Cancer Institute, University College London, 72 Huntley Street, London WC1E 6BT, UK

<sup>2</sup>Breast Cancer Functional Genomics Laboratory, Cancer Research UK Cambridge Research Institute and Department of Oncology University of Cambridge, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK

<sup>3</sup>Cambridge Breast Unit and NIHR Biomedical Research Centre, Addenbrookes Hospital, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 2QQ, UK

linear combinations thereof. The modules were shown to be specific for either ER+ or ER- breast cancer, further supporting the view that cancer subtype analyses are necessary [6,7].

Li and colleagues' work is also significant for two other reasons. First, by dissecting prognostic gene lists into various cancer hallmarks they identified prognostic modules (for example, apoptosis) that are normally only associated with the primary tumour and not with the subsequent risk of relapse or distant metastasis. A similar finding implicating apoptosis as a molecular determinant of distant metastasis was made in a different study [13]. Second, genes making up the prognostic modules were mapped to a protein interaction network and many were shown to directly interact with genes known to be mutated in breast cancer (COSMIC) [14]. Interestingly, the prognostic power of the modules could be related to these driver genes, as other sets of genes with the same GO annotations and also direct neighbours of the drivers were observed to be as prognostic as the original modules. This may explain the redundancy in prognostic gene sets, which is often observed in gene expression studies.

### The viewpoint

The insights provided by Li and colleagues are important, yet there are some cautionary remarks. First, a predictive value of 90% is only of clinical relevance if this can be achieved using a genuine single-sample classifier. However, it would appear that the algorithm of Li and colleagues [11] is not a single-sample predictor, since the specific weights in the centroids are retrained in each of the validation sets. Thus, the validation presented is only of the actual gene sets, and not of a single classifier. Therefore, the predictive values quoted should be interpreted with caution. In contrast, the Oncotype DX and IR module classifiers are each defined by a single centroid and therefore constitute (modulo a trivial gene-wise recentering and rescaling) *bona-fide* single-sample predictors. In other words, the Oncotype and IR module centroids used to classify test samples are unchanged and unique to the training set. Thus, it remains to be seen if a single-sample classifier derived with the MSS algorithm [11] could be applied in a clinical setting.

Another important limitation of this study is the sole reliance on one data set [12] to infer prognostic gene sets. It is very likely that specific prognostic modules may have been missed, specially for the minority ER- subgroup, for which one data set would typically not provide the necessary sample numbers and power [7]. This may explain why Li and colleagues do not find a prognostic immune response module in ER- breast cancer, when in fact many other studies have reported such a signature [3,8,9].

Probably, the most important insight from Li and colleagues' work is the observation that non-linear

combinations of prognostic modules related to different cancer hallmark GO terms may significantly improve the prognostic power of gene expression signatures. This makes sense because more often than not gene expression signatures are biased towards those genes and GO terms with the largest effect sizes, thereby diluting the smaller yet equally important predictive effects of other genes or GO terms. Alternatively, the interactions between cancer hallmarks are likely to be non-linear and therefore non-linear combinatorial classifiers may be required to capture these effects. Thus, development of further algorithmic tools that incorporate such non-linearities may be a fruitful endeavour.

The observation by Li and colleagues that underlying these robust prognostic modules may be a common set of mutated driver genes is an important one and supports similar observations made by at least two other groups [15,16]. We envisage that as the quality and coverage of protein interaction networks improves, and as we approach the completion of a final catalogue of mutated breast cancer genes, methods that interpret gene expression, protein expression and protein phosphorylation data in the context of these structural signalling networks and breast cancer genes are likely to provide us with the biological insights needed to drive this field forward.

### Abbreviations

ER, oestrogen receptor; GO, Gene Ontology; IR, immune response; MSS, multiple survival screening.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgements

AET is supported by a Heller Research Fellowship.

Published: 8 December 2010

### References

1. van't Veer LJ, Bernardis R: **Enabling personalized cancer medicine through analysis of gene-expression patterns.** *Nature* 2008, **452**:564-570.
2. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**:2817-2826.
3. Teschendorff AE, Caldas C: **A robust classifier of high predictive value to identify good prognosis patients in er-negative breast cancer.** *Breast Cancer Res* 2008, **10**:R73.
4. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernardis R: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347**:1999-2009.
5. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M: **Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.** *J Natl Cancer Inst* 2006, **98**:262-272.
6. Teschendorff AE, Naderi A, Barbosa-Morais NL, Pinder SE, Ellis IO, Aparicio S, Brenton JD, Caldas C: **A consensus prognostic gene expression classifier for er positive breast cancer.** *Genome Biol* 2006, **7**:R101.

7. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C: **An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer.** *Genome Biol* 2007, **8**:R157.
8. Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrman M: **The humoral immune system has a key prognostic impact in node-negative breast cancer.** *Cancer Res* 2008, **68**:5405-5413.
9. Calabrò A, Beissbarth T, Kuner R, Stojanov M, Benner A, Asslaber M, Ploner F, Zatloukal K, Samonigg H, Poustka A, Sültmann H: **Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer.** *Breast Cancer Res Treat* 2009, **116**:69-77.
10. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, Delorenzi M, Piccart M, Sotiriou C: **Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes.** *Clin Cancer Res* 2008, **14**:5158-5165.
11. Li J, Lenferink AE, Deng Y, Collins C, Cui Q, Purisima EO, O'Connor-McCourt MD, Wang E: **Identification of high-quality cancer prognostic markers and metastasis network modules.** *Nat Commun* 2010, **1**:1-8.
12. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoa T, Berns EM, Atkins D, Foekens JA: **Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.** *Lancet* 2005, **365**:671-679.
13. Teschendorff AE, Severini S: **Increased entropy of signal transduction in the cancer metastasis phenotype.** *BMC Syst Biol* 2010, **4**:104.
14. COSMIC [<http://www.sanger.ac.uk/genetics/CGP/cosmic/>]
15. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
16. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: **Dynamic modularity in protein interaction networks predicts breast cancer outcome.** *Nat Biotechnol* 2009, **27**:199-204.

doi:10.1186/bcr2774

**Cite this article as:** Teschendorff AE, *et al.*: Prognostic gene network modules in breast cancer hold promise. *Breast Cancer Research* 2010, **12**:317.