

VIEWPOINT

Regulation of mRNA expression in breast cancer - a *cis*-tematic *trans*-action

Roel GW Verhaak^{1*} and Gordon B Mills²

Abstract

Large research consortia such as the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), The Cancer Genome Atlas and International Cancer Genomics Consortium are systematically interrogating large sets of tumor samples through integrated analysis of genome-wide DNA copy number and promoter methylation, transcriptome-wide RNA expression, protein expression and exome-wide sequencing. A recent METABRIC study explored the effects of *cis*-acting and *trans*-acting factors of gene expression regulation in breast cancer. By making their data sets publicly available, these large consortia are inviting new types of analysis that have the potential to drive breast cancer research into previously unexplored avenues.

The central dogma of molecular biology dictates that DNA is transcribed to mRNA, which is translated to protein. DNA dosage is frequently altered in cancer and is an important determinant of mRNA expression. Transcription is organized by *trans*-acting transcription factors or transcription factor complexes that associate with binding sites of a specific sequence. The number and binding affinity of such *cis*-binding elements provide a mechanism of transcription regulation [1]. DNA copy number changes affect the cell by altering the amount of DNA for transcription factors to act upon (*cis*) or by altering the production of transcription factors that would alter the expression of genes elsewhere (*trans*). Chromosomal gains and losses thus lead to increased or decreased numbers of mRNA molecules that are transcribed from the altered locus, thereby providing a proliferative advantage to the cell. Elucidating the

mechanisms underlying the effects of DNA copy number aberrations on expression regulation in cancer would aid in identifying master regulators and the design of therapeutic modalities that specifically block key elements in a regulatory network.

Transcript levels are regulated through multiple processes, including chromatin organization and modification, and effects of microRNAs and long non-coding RNAs. The multitude of regulatory mechanisms complicates the effective unraveling of *cis*- and *trans*-acting factors. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) circumvented these complicating factors through a *tour-de-force* approach [2]. In a collaborative effort of British and Canadian institutes, METABRIC collected a very large number of fresh frozen breast cancer tissues ($n = 2,136$) with long-term follow-up and generated gene expression, genotype and DNA copy number profiles for all cases. This effort resulted in a data set of previously unmatched proportions, adding significantly to the number of breast cancer genomic profiles available in the public domain (The European Genome-phenome Archive [3] accession number EGAS00000000083).

After separating the sample cohort into a training set of approximately 1,000 samples and a similarly sized validation data set, which contained those profiles of lower cellularity or with missing matching normal sample data, the investigators performed integrated analysis of DNA copy number and transcript levels in order to identify target genes of DNA copy number alterations. The large sample size of the data set also allowed investigation of expression quantitative trait loci, which are chromosomal segments whose genotypes or copy number levels show an association with the expression levels of distal genes (>3 Mb). Combined, the two types of analysis were aimed at exposing the *cis*- and *trans*- circuitry of breast cancer and effectively showed that whereas *trans*-acting loci influence expression of a larger number of transcripts, *cis*-acting loci had a more pronounced effect on transcript levels. The availability of a large validation data set allowed the investigators to provide convincing results. Importantly, recent studies of breast cancer sample purity have shown that the average tumor cellularity of breast

*Correspondence: rverhaak@mdanderson.org

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA
Full list of author information is available at the end of the article

cancer tissues is about 49% [4]. The large number of stromal cells, immune cells and tumor-adjacent normal tissue contributes significantly to gene expression levels, which is reflected by the prominent *trans*-acting association of a T-cell receptor loci gene expression module identified by the investigators. Samples highly expressing the T-cell receptor gene set represent one of the 10 patient clusters that were generated through clustering of expression profiles based on the expression levels of the 1,000 genes, which were most strongly regulated in *cis*. Interestingly, the *HER2* and basal gene expression subtypes that are covered by the PAM50 gene expression classification of breast cancer [5] are also identified by *cis*-element clustering. This suggests that the dominant effects driving the *HER2* and basal gene expression subtypes are due to alterations in copy number of the protein targets directly rather than of transcription factors that act in *trans*.

As in publications from The Cancer Genome Atlas Research Network [6,7], the analysis described in the METABRIC publication only scratches the surface of what is possible with a data set of this magnitude. The large number of samples makes the data set particularly useful for identification of mutually exclusive copy number alterations, as well as co-occurring abnormalities [8], but also for further exploration of emergent questions, such as the contributions of non-tumor cells to breast cancer homeostasis.

The METABRIC study has effectively characterized the DNA copy number alteration landscape of breast cancer. These studies are most powerful when integrated with a series of recent studies that have deciphered the mutational landscape of breast cancer through whole-genome and whole-exome sequencing of four independent cohorts containing 80 to 100 breast cancer tissues [9-12]. These efforts uncovered frequent mutations of genes in the mitogen-activated protein kinase (MAPK) pathway and further highlighted the important role of the phosphoinositide 3-kinase (PI3K) pathway as rational therapeutic targets. Integrated analysis of point mutations, methylation profiles, DNA copy number alterations, gene expression, and functional proteomics, as is underway by The Cancer Genome Atlas (TCGA) Research Network, further improves our understanding of breast cancer tumorigenesis [13]. Using genome sequencing, these and other recent studies showed that breast tumors harbor a clonal hierarchy in which genomic abnormalities may only be present in a subset of tumor cells [14,15]. Such analytical approaches may distinguish tumor-initiating changes from abnormalities that lead to tumor progression and provide insights into the temporal order of genomic alteration events during breast cancer progression [11,14-16]. This may be particularly important, given marked intratumor heterogeneity, by

identifying aberrations present in all tumor cells that, if druggable, would represent optimal targets.

Large consortia such as METABRIC, TCGA and the International Cancer Genomics Consortium are generating high quality data sets that invite creative research questions and build a foundation for data analysis for years to come, providing insights into breast cancer tumorigenesis that have never before been possible. Whereas this makes the current era of breast cancer research more exciting than ever before, the key to translating these findings into palpable clinical improvements has not yet been determined. With the emergence of these high quality data sets, the emphasis of genomic studies is rapidly shifting away from data generation towards meta-analysis, integrated analysis, data mining and computational analysis of existing data. With petabytes of high content data available, asking the right research question has become a critical factor towards finding a curative strategy for this complex disease.

Abbreviations

METABRIC, Molecular Taxonomy of Breast Cancer International Consortium; TCGA, The Cancer Genome Atlas.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA. ²Department of Systems Biology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA.

Published: 29 October 2012

References

1. Wittkopp PJ, Kalay G: **Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.** *Nat Rev Genet* 2012, **13**:59-69.
2. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.** *Nature* 2012, **486**:346-352.
3. **The European Genome-phenome Archive** [<http://www.ebi.ac.uk/ega/>]
4. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Børresen-Dale AL, Kristensen VN: **Allele-specific copy number analysis of tumors.** *Proc Natl Acad Sci U S A* 2010, **107**:16910-16915.
5. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**:1160-1167.
6. Network TCGAR: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-1068.
7. Network TCGAR: **Integrated genomic analyses of ovarian carcinoma.** *Nature* 2011, **474**:609-615.
8. Cirillo G, Cerami E, Sander C, Schultz N: **Mutual exclusivity analysis identifies oncogenic network modules.** *Genome Res* 2012, **22**:398-406.
9. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Piña V, Duke F, Francis J, Jung J,

- Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, *et al*: **Sequence analysis of mutations and translocations across breast cancer subtypes.** *Nature* 2012, **486**:405-409.
10. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, *et al*: **Whole-genome analysis informs breast cancer response to aromatase inhibition.** *Nature* 2012, **486**:353-360.
 11. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, *et al*: **The clonal and mutational evolution spectrum of primary triple-negative breast cancers.** *Nature* 2012, **486**:395-399.
 12. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, *et al*: **The landscape of cancer genes and mutational processes in breast cancer.** *Nature* 2012, **486**:400-404.
 13. Network TCGAR: **Comprehensive molecular portraits of human breast tumors.** *Nature* 2012, in press.
 14. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M: **Tumour evolution inferred by single-cell sequencing.** *Nature* 2011, **472**:90-94.
 15. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, *et al*: **The life history of 21 breast cancers.** *Cell* 2012, **149**:994-1007.
 16. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, Pons J, Chan SW, Lam ET, Chu C, Park K, Hong SW, Hur JS, Huh N, Neuhaus IM, Yu SS, Grekin RC, Mauro TM, Cleaver JE, Kwok PY, LeBoit PE, Getz G, Cibulskis K, Aster JC, Huang H, Purdom E, Li J, Bolund L, Arron ST, Gray JW, *et al*: **Temporal dissection of tumorigenesis in primary cancers.** *Cancer Discov* 2011, **1**:137-143.

doi:10.1186/bcr3227

Cite this article as: Verhaak RGW, Mills GB: **Regulation of mRNA expression in breast cancer - a cis-tematic trans-action.** *Breast Cancer Research* 2012, **14**:322.