A transcriptional sketch of a human breast cancer by 454 deep sequencing

Guffanti A[(1)*], Iacono M[(1)*], Pelucchi P[(1)*], Kim N[(2,3)], Soldà G[(4)], Croft LJ[(5)], Taft RJ[(5)], Rizzi E[(1)], Askarian-Amiri M[(5)], Bonnal RJ[(1)], Callari M[(11)], Mignone F[(6)], Pesole G[(1,7)], Bertalot G[(8,9)], Bernardi LR[(10)], Albertini A[(1)], Lee C[(2)], Mattick JS[(5)], Zucchi I[(1)] & de Bellis G[(1)]

[(1)] Institute of Biomedical Technologies, National Research Council, Milan, Italy
[(2)] Department of Biochemistry and Molecular Biology, University of California Los Angeles, USA
[(3)] Current address: Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, 52 Eoeun-dong, Yuseong-gu, Daejeon, 305-806, South Korea
[(4)] Department of Biology and Genetics for Medical Sciences, University of Milan, Italy.
[(5)] ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia QLD 4072, Australia
[(6)] Faculty of Pharmacological Sciences, University of Milan, Italy.
[(7)] Department of Biochemistry and Molecular Biology, University of Bari, Italy
[(8)] European Institute of Oncology, Milano, Italy
[(9)] Current address: Department of Pathology, Desenzano sul Garda Hospital, Leno, Italy
[(10)] Science and Technology Pole, Istituto di Ricovero e Cura a Carattere Scientifico MultiMedica, Milan, Italy
[(11)] Translational Research Unit, Department of Experimental Oncology, Istituto Nazionale Tumori, Milan.

Motivation

The cancer transcriptome is a complex object to explore, given the heterogeneity of quantitative and qualitative transcriptional events linked to the disease status. An increasing number of “unconventional” transcripts, such as novel isoforms, noncoding RNAs, somatic gene fusions and deletions have been associated with the tumoural state. In addition, there is growing evidence that antisense transcription may fine-tune gene expression in cis and in trans, including cancer genes. Massively parallel sequencing techniques make full-transcriptome sequencing feasible with a limited laboratory and financial effort. Hence, we developed a 454 deep sequencing and bioinformatics analysis protocol to investigate the molecular composition of a breast cancer poly(A)+ transcriptome. This method combines the direct pyrosequencing of a breast cancer cDNA library with a normalization step and biology-oriented bioinformatic analyses to increase the detection of rare transcripts of possible interest in the aetiology of cancer.

Methods

We classified 132.113 breast cancer partial cDNA sequences, mapping to the human genome, in well-defined categories (intragenic, extragenic, novel transcripts, known and novel exons and isoforms) and detected a range of unusual transcriptional events, some of which could be related to the disease. We subsequently validated by RT-PCR on RNA from a total of nine breast cancer samples and by direct sequencing on the RT-PCR products from the original sample one deletion, two novel ncRNAs (one intergenic and one intragenic) and a number of new isoforms. A gene fusion, instead, could be identified clearly only in the original RNA. We also explored the non-protein-coding portion of the breast cancer transcriptome, identifying hundreds of sequences corresponding to the non-coding RNA MALAT1, known to be overexpressed in many human carcinomas. Analysis of cDNA array expression data from a large panel of breast tumours weakly supports the hypothesis of a link between MALAT1 overexpression and breast cancer recurrence. We also identified many novel non-coding transcripts supported by EST evidence and/or conservation analysis. Finally, we investigated the reciprocal expression pattern of senseantisense transcripts in our sample and detected an enrichment of the coexpression of the transcript pairs overlapping at their 3’ end.

Results

Our results demonstrate how 454 deep sequencing of a cDNA library, combined with a normalization step and careful bioinformatic analysis to increase the detection of rare transcripts, can be used as a qualitative tool to characterize transcriptome complexity, revealing many interesting and hitherto unknown transcripts as candidates for further wet lab functional investigation, even at relatively low sequencing depth.

Contact : alessandro.guffanti@itb.cnr.it