Research article

Open Access

Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data

Anita Grigoriadis[1], Alan Mackay[2], Jorge S Reis-Filho[2], Dawn Steele[2], Christian Iseli[3], Brian J Stevenson[3], C Victor Jongeneel[3], Haukur Valgeirsson[2], Kerry Fenwick[2], Marjan Iravani[2], Maria Leao[1], Andrew JG Simpson[4], Robert L Strausberg[5], Parmjit S Jat[6], Alan Ashworth[2], A Munro Neville[1] and Michael J O'Hare[1]

[1]Ludwig Institute for Cancer Research/University College London Breast Cancer Laboratory, 91 Riding House Street, London, W1W 7BS, UK
[2]The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK
[3]Office of Information Technology, Ludwig Institute for Cancer Research and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[4]Ludwig Institute for Cancer Research, New York Branch at Memorial Sloan-Kettering Cancer Centre, New York, NY 10021, USA
[5]The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
[6]Department of Neurodegenerative Disease, Institute of Neurology, London, WC1N 3BG, UK

Corresponding author: A Munro Neville, munroneville@f2s.com

Received: 17 Jul 2006 Revisions requested: 7 Aug 2006 Revisions received: 7 Sep 2006 Accepted: 2 Oct 2006 Published: 2 Oct 2006

Breast Cancer Research 2006, **8**:R56 (doi:10.1186/bcr1604)
This article is online at: http://breast-cancer-research.com/content/8/5/R56

Abstract

Introduction Diverse microarray and sequencing technologies have been widely used to characterise the molecular changes in malignant epithelial cells in breast cancers. Such gene expression studies to identify markers and targets in tumour cells are, however, compromised by the cellular heterogeneity of solid breast tumours and by the lack of appropriate counterparts representing normal breast epithelial cells.

Methods Malignant neoplastic epithelial cells from primary breast cancers and luminal and myoepithelial cells isolated from normal human breast tissue were isolated by immunomagnetic separation methods. Pools of RNA from highly enriched preparations of these cell types were subjected to expression profiling using massively parallel signature sequencing (MPSS) and four different genome wide microarray platforms. Functional related transcripts of the differential tumour epithelial transcriptome were used for gene set enrichment analysis to identify enrichment of luminal and myoepithelial type genes. Clinical pathological validation of a small number of genes was performed on tissue microarrays.

Results MPSS identified 6,553 differentially expressed genes between the pool of normal luminal cells and that of primary tumours substantially enriched for epithelial cells, of which 98% were represented and 60% were confirmed by microarray profiling. Significant expression level changes between these two samples detected only by microarray technology were shown by 4,149 transcripts, resulting in a combined differential tumour epithelial transcriptome of 8,051 genes. Microarray gene signatures identified a comprehensive list of 907 and 955 transcripts whose expression differed between luminal epithelial cells and myoepithelial cells, respectively. Functional annotation and gene set enrichment analysis highlighted a group of genes related to skeletal development that were associated with the myoepithelial/basal cells and upregulated in the tumour sample. One of the most highly overexpressed genes in this category, that encoding periostin, was analysed immunohistochemically on breast cancer tissue microarrays and its expression in neoplastic cells correlated with poor outcome in a cohort of poor prognosis estrogen receptor-positive tumours.

Conclusion Using highly enriched cell populations in combination with multiplatform gene expression profiling studies, a comprehensive analysis of molecular changes between the normal and malignant breast tissue was established. This study provides a basis for the identification of novel and potentially important targets for diagnosis, prognosis and therapy in breast cancer.

COMP = cartilage oligomeric matrix protein; DTET = differential tumour epithelial transcriptome; ER = estrogen receptor; GO = Gene Ontology; GSEA = gene set enrichment analysis; HTR = human transcriptome database; IL = interleukin; MAPK = mitogen-activated protein kinases; MIAME = minimum information about a microarray experiment; MPSS = massively parallel signature sequencing; POSTN = periostin; PR = progesterone receptor; RT-PCR = reverse transcription PCR; SAGE = serial analysis of gene expression; TMA = tissue microarray; tpm = transcripts per million; VEGF = vascular endothelial growth factor; VEGFR = vascular endothelial growth factor receptor.