Introduction

Breast cancer is a clinically heterogeneous disease and consists of many different cell types, including normal and reactive stromal components in addition to the malignant neoplastic compartment. Moreover, it comprises a series of distinct malignant tumours that present diverse cellular features with varying differentiation status, distinct genetic changes, responses to therapy and outcome [1]. Likewise, the normal breast is also composed of different parenchymal and stromal cell types, with the terminal ductal-lobular unit being the most important feature with regard to neoplasia. The latter is composed of two morphologically recognisable cell types, epithelial cells on the luminal surface and basally located myoepithelial cells. While typical breast cancers have been traditionally regarded as exhibiting characteristics akin to luminal epithelial cells, recent data have shown that some also exhibit, in part or whole, myoepithelial/basal features [2-4]. Based on the restricted expression of genes representing the phenotypes of luminal epithelial and basal cells [4], major subtypes of breast cancer have been defined and linked to both long term survival [5] and their response to therapy [6]. Therefore, detailed characterisation of the normal luminal and myoepithelial/basal phenotypes is a prerequisite for understanding the genetic alterations that occur in breast cancers and how they may impact on disease progression and outcome.

The use of solid tissues, as in most previous breast cancer gene expression analyses, results in greatly enhanced complexity of data because of the widely varying degrees of stromal responses (desmoplasia) and inflammatory infiltrates in individual tumours. Laser capture microdissection partially alleviates this problem in respect to tumour samples, but is unsuited to the large-scale separation of the normal epithelial cell types in breast because of the close contact between these cells. Immunomagnetic separation of individual cell types from normal human breast tissue [7,8] and primary breast cancers [9] has enabled direct comparisons of normal epithelial and malignant epithelial cells to be made. Previous proteomic [9,10] and gene expression analyses of such samples [10-13] have established a partial molecular characterisation of the epithelial compartment in the normal breast and breast cancer [2], but, due to the limitations of technology available at the time of these studies, did not provide a comprehensive comparison of all proteins or transcripts.

Multiple large-scale analytical techniques now make it possible to capture entire transcriptomes of defined cell populations. Breast cancers have been extensively analysed with both expression arrays [14] and with direct sequencing techniques such as serial analysis of gene expression (SAGE) [15]. Although several studies have correlated expression data based on microarray and SAGE [16,17], a comprehensive genome-wide expression profile using a combination of complementary technologies has not yet been achieved for purified malignant epithelial breast cells in comparison with purified normal breast epithelial cells. In this study, massively parallel signature sequencing (MPSS) [18,19] and multiple genome-wide microarrays have been used to analyse immunomagnetically separated normal luminal epithelial cells and primary breast cancers substantially enriched for the neoplastic epithelial component. The aim of this study was to establish a virtually complete coverage of transcripts deregulated in the neoplastic cells of human breast cancer. In addition, expression profiles from normal luminal and myoepithelial cells have been used to identify cell-type specific transcripts and ontologically related gene sets in the differentially expressed tumour epithelial transcriptome. The use of highly enriched cell preparations in combination with a multiplatform approach to their expression analysis has revealed novel markers and potential targets, the clinical significance of some of which has also been examined, using tissue microarrays.

Materials and methods

Sample preparation

Ten primary cultures (approximately 10^7) of normal human breast luminal and myoepithelial cells were prepared from reduction mammoplasty samples by double immunomagnetic sorting methods [7,8,10]. In brief, breast epithelial cells were immunomagnetically purified using combined positive magnetic activated cell sorting (MACS; Miltenyi Biotec, Auburn, CA) selection with antibodies against the luminal epithelial marker EMA (rat monoclonal ICR-2, Seralab, Leicestershire, UK) and the myoepithelial membrane antigen CD10 (mouse monoclonal CALLA clone SS2/36, DAKO Corporation, Glostrup, Denmark), followed by negative Dynabead (Dynal, UK) selection using mouse monoclonal antibodies against anti-β-4-integrin clone A9, a myoepithelial cell-surface antigen (Santa Cruz Biotechnology, CA, USA) and BerEp-4 Epithelial Antigen, a luminal antigen (DAKO Corporation, Glostrup, Denmark). Immunostaining with myoepithelial and luminal-specific lineage markers showed the final sort of epithelial cells used in this study to be >95% pure. Full details of these procedures are not only contained in previous publications [10,11], but are also appended, as required, to the Minimum information about a microarray experiment (MIAME) protocol that accompanies submission E-TABM-66 [20].

Malignant breast epithelial cells of 50 freshly isolated primary infiltrating ductal carcinomas of histological grade 2 and 3 were enriched from disaggregated tumour tissue as described previously [9]. In brief, fresh tumour biopsies (1 to 2 g) were comminuted to approximately 1 mm^3, using scalpel blades, and subjected to a controlled disaggregation using 0.25% collagenase Type1 (Sigma-Aldrich, Dorset, UK) in L-15