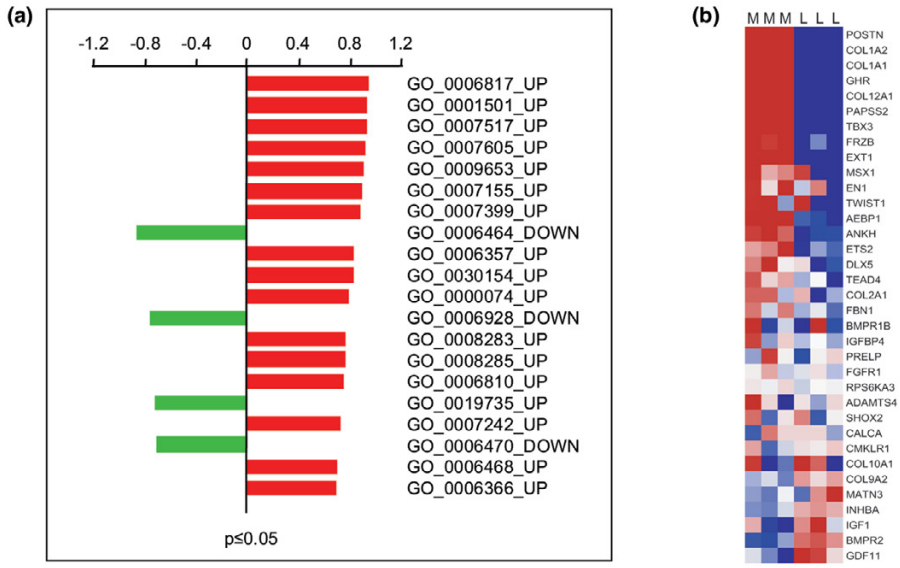
breast epithelial tissue with solid tumour breast tissue [12,13] in which reduced expression of cytokines such as IL6 and IL8 was observed, higher abundance of these genes was detected in our malignant breast epithelial sample in comparison with the normal luminal sample. Ninety genes belonging to the GO category of 'apoptosis', including members of the BAG family (*BAG1*, *BAG2*, *BAG3*), as well as members of the breast cancer 'proliferation signatures' (*BUB1*, *PLK1*, *CCNE1*, *CCND1* and *CCNB1*) were also identified as up-regulated in our DTET [35,36].

The most significantly perturbed functional gene set identified in the down-regulated tumour epithelial transcriptome (Figure 2b) was epidermis development, including members of the kallikrein family (*KLK5*, *KLK7*, *KLK8*, *KLK10*) and the keratin family (*CK10*, *CK14*), as well as the family of extracellular matrix glycoproteins, such as *LAMC2*, *LAMB3* and *LAMA3*. The second most perturbed subset of down-regulated genes included several members of the RAS-related proteins, *RAP1A*, *RALB*, *RAB5B*, *RAB4A*, *RAB3B*, *RAB2* and *RAB25* (protein transport; Figure 2b), some of which counteract the mitogenic function of RAS-MAPK signalling pathways [37].

Differentially expressed transcripts in normal breast epithelial cells

Whether tumours exhibit a luminal or myoepithelial/basal phenotype has been correlated with prediction and prognosis in breast cancer [2-4]. Global transcriptomes of normal myoepithelial and luminal epithelial cells were, therefore, compared to identify all transcripts that were differentially expressed in these normal cell types. The purpose was to further define breast epithelial specificity within the tumour transcriptome by annotating the DTET with respect to their expression in these normal epithelial cell types. Differential gene expression profiles of immunomagnetically purified luminal and myoepithelial cell samples were established using the criterion of differential detection by at least two of the four genome-wide microarray platforms, as used previously when comparing the normal luminal with the malignant sample. We identified 907 transcripts with higher abundance in the normal luminal cells and 955 transcripts were higher in the normal myoepithelial cells. These collectively comprised the differential normal epithelial transcriptome. The top 50 discriminator genes over all four microarray platforms are shown in Figure 3 (complete list is given as Additional file 9). These genome-wide gene signatures agreed with previous data from individual luminal and myoepithelial sample analyses [11]. All the main classifiers for the myoepithelial cell type, such as *LGALS7*, *S100A2*, *SFN*, *SPARC* and *CAV1* (and *CD24*, *LCN2*, *CLDN4*, *MUC1* and *SEMA3B* for the luminal epithelial cell type) were identified as differential in the present study. However, as expected from the enhanced coverage provided by the methods used here, many other genes that may play an important role in the biology of these two cell types were also identified (for example, *PADI2*, *TSPAN2*, *DACT1* for the luminal, and *POSTN*, *DCN*, *ADAMTS5* for the myoepithelial cell type).

Figure 4



Enrichment of luminal and myoepithelial transcripts in the differentially expressed epithelial tumour transcriptome. **(a)** The top 20 deregulated biological processes identified by gene set enrichment analysis that are enriched in luminal (green) and myoepithelial (red) expression are shown. The definition of each Gene Ontology (GO) category is given in Additional file 8. **(b)** Heatmap of the skeletal developmental gene subset (GO:0001501) based on the Affymetrix expression data. L (luminal) and M (myoepithelial) show results from individual arrays. Genes are ranked according to their significance of enrichment as described in the Materials and methods.