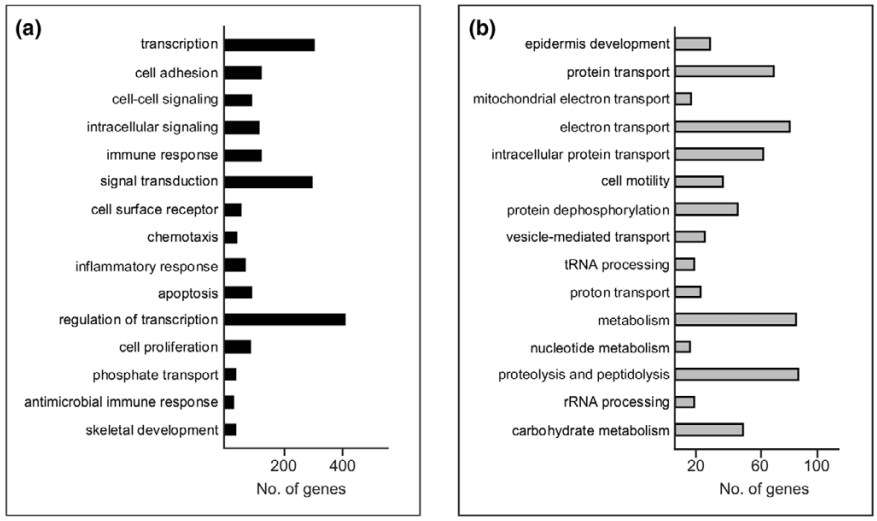
**Figure 2**




Functional classification of the differentially expressed epithelial tumour transcriptome. The top 15 biological processes showing overall **(a)** up-regulation and **(b)** down-regulation are shown. The biological processes are ranked from top to bottom according to their ascending *P* value as described in the Materials and methods. The numbers of genes within each process that are up-regulated or down-regulated for each category are also shown as black and grey bars, respectively.

Having established a common denominator in terms of gene annotation, those genes reported as differential between the normal and malignant tumour sample by microarrays were defined and then compared with the MPSS data. The criteria for differential expression used were that expression measurements between the normal and the malignant sample reported had to be both statistically significant ($P \leq 0.05$) and in the same direction (up or down). Out of the four microarray platforms, the two single colour oligonucleotide platforms (Affymetrix and CodeLink™) validated as differential 3,206 (48.9%) and 3,004 (45.8%) of all MPSS transcripts present on their platforms, respectively, whereas the two-colour microarray technologies confirmed only 1,257 (19.1%) and 1,379 (21%), for Agilent and 20 k brk, respectively (Figure 1a). Overall, a total of 3,902 genes were obtained in which at least one microarray confirmed the MPSS data without any other platform reporting an opposite result (Figure 1a; 1 platform). Expression measurements for 2,440 MPSS differential transcripts could not be confirmed using any of these microarray platforms (Figure 1b, "MPSS-only"). The microarray data were also used to identify any genes reported as differential by at least two platforms, but which did not appear as such in the MPSS analysis. This comprised a total of 4,149 transcripts (Figure 1b, "Array-only"). To establish which of those sets could be most relied on to constitute the validated differential tumour epithelial transcriptome (DTET), examples of each group were analysed by semi-quantitative RT-PCR (Figure 1b). This showed that only 30% (6/20) of the "MPSS-only" identified differentials could be validated, while 78% (78/100) and 92% (37/40) of the "MPSS and array" and "Array-only" differentially expressed transcripts were reported as differential by RT-PCR (Additional file 6). The comparison of RT-PCR results was not given any statistical treatment and is simply presented to illustrate that the array confirmed differentials have a much lower false positive rate (20% to 70%). Consequently, the latter two groups were combined and comprised 8,051 up- and down-regulated genes that constitute the DTET and were subjected to further analysis (Additional file 7).

## Functional classification of differentially expressed genes

GO classification of the 8,051 genes of the DTET revealed that, as might be expected, multiple cellular processes, such as transcription, signal transduction, cell adhesion, cell cycle, metabolism, transport and development, are different in normal luminal epithelium and their malignant counterparts (the full list of perturbed biological processes is provided as Additional file 8). In terms of overall differences, the largest functional group of up-regulated transcripts (Figure 2a) corresponded to genes associated with transcription and regulation in transcription, in agreement with several other profiling studies. The second largest functional group comprised genes involved in signal transduction. These consisted, amongst others, of genes encoding proteins involved in mitogen-activated protein kinases (MAPK) signalling (*FGF4*, -7, -13, *IL1A*, *IL1B*, *NGFB*, *TGFB1* and *TGFB3*) and the JAK-STAT signalling pathway (*IL6*, *IL10*, *OSM*, *SPRY2*), as well as ligands and receptors involved in cytokine-cytokine interaction, including members of the CXC and CC chemokines, platelet-derived growth factor, gp130, tumour necrosis factor and transforming growth factor-β subfamilies. Many of these genes have already been correlated with breast cancer growth and invasion, and their epithelial expression has been demonstrated. In contrast to previously published SAGE data, comparing purified normal