Table 2

Multivariant proportional-hazard analysis

Parameter	Hazard ratio (95% confidence interval)	*P* (Cox)
Grade	*2.6434 (1.3935–5.0144)*	*0.0029*
LN status	*3.6408 (1.5264–8.6840)*	*0.0036*
POSTN	*1.8099 (0.9247–3.5422)*	*0.0833*
LN status	**5.6924 (1.3331–24.3076)**	**0.0189**
POSTN	**2.8151 (1.2048–6.5775)**	**0.0168**

The tissue microarray cohort was analysed using the Cox proportional hazards model for disease-free survival (italic) and overall survival (bold). Only those statistically significant independent prognostic factors as determined by the model are shown. LN, lymph node status at diagnosis.

Enrichment of luminal and myoepithelial genes and gene sets in the differential tumour epithelial transcriptome

To identify functionally related gene sets of luminal or myoepithelial phenotype within the DTET, GSEA was carried out on the perturbed biological processes that were statistically significant ($P \leq 0.05$) and composed of at least 10 genes [31]. This resulted in a total of 72 gene sets, 53 and 19 for the up- and down-regulated modules, respectively (Additional file 8). In the top 20, four categories showed enrichment of genes belonging to our luminal transcriptome, including protein modification (GO:0006464), cell motility (GO:0006928) and protein dephosphorylation (GO:0006470), as down-regulated modules, as well as antimicrobial immune response (GO:0019735) as an up-regulated one (Figure 4a). Overall, GSEA analysis showed marked enrichment for the expression of myoepithelial genes in the functional groups of the tumour overexpressed transcripts compared to the luminal epithelial transcriptome (Figure 4a). The gene set with the most statistically significant representation of myoepithelial type genes consisted of members of the collagenase family (GO:0006817), with *COL3A1*, *COL6A1*, *COL1A1*, *COL5A12*, *COL15A2*, *COL1A1* and *COL12A1* representing the discriminator genes. The second most statistically significant enrichment of expression in myoepithelial type genes with higher abundance in the malignant breast epithelium was found in the functional category of skeletal development (GO:0001501; Figure 4a,b). This set of bone related genes included *COL1A2*, *COL1A1*, *GHR*, *COL12A1*, *PAPSS2*, *TBX3*, *FRZB*, *EXT1*, *MSX1*, *EN1*, *TWIST1* and *AEBP1*, with *POSTN* being the most prominent discriminator of this gene set (Figure 4b).

Clinical significance of POSTN using tissue microarray analysis

To evaluate whether the luminal and myoepithelial annotations of our epithelial deregulated transcriptome identify genes with any correlation with clinical outcome in breast cancer, we performed immunohistochemical analysis POSTN on a tissue microarray consisting of 245 primary breast tumours. POSTN, usually expressed in mesenchymal cells, was chosen, not only because it was one of the most highly differentially expressed genes in normal myoepithelial cells over all microarray platforms (Figure 3), but also because it belongs to the functional group of skeletal development that showed overall myoepithelial-specificity and up-regulation in the malignant breast epithelium (Figure 4b). When POSTN expression was examined at the protein level, no detectable expression was observed in the normal breast epithelium, but only in the stroma, in concordance with its known mesenchymal expression (not shown). However, 42/224 (18.75%) invasive breast carcinomas clearly showed epithelial expression (Figure 5a), whereas the remainder showed the expected expression pattern only in the stroma (Figure 5b). POSTN expression in neoplastic cells was significantly correlated with positivity for progesterone receptor (PR) ($P < 0.05$) and low proliferation rates as defined by Ki67 (MIB1) staining ($P < 0.05$) (Additional file 10). When the whole cohort was analysed, POSTN-positive breast cancers showed a trend towards a poorer outcome, although this did not reach statistical significance (Additional file 11a,b). Since the estrogen receptor (ER) status is the most important marker in defining the prognosis and treatment of breast cancer, the correlation of POSTN expression with overall survival and disease free survival was analysed in ER-positive and ER-negative subgroups. No significant correlation was observed in the ER-negative cohort. However, within the ER-positive subgroup, 20.8% (37/178) of breast tumours were positive and there was a significant correlation with both overall survival ($P = 0.0083$) and disease-free survival ($P = 0.0136$) (Figure 6a,b, respectively). In this cohort, modified Bloom-Richardson grade ($P < 0.01$), lymph node status at diagnosis ($P < 0.005$) and POSTN expression ($P < 0.05$) were statistically significant predictors of disease-free survival in univariate analysis, whereas only lymph node status at diagnosis ($P < 0.001$) and POSTN expression ($P < 0.01$) were associated with overall survival in univariate analysis. By multivariate analysis of disease-free survival in the ER-positive cohort, POSTN did not reach formal statistical significance as an independent factor ($P = 0.0833$) (Table 2, italics), although it did constitute an independent prognostic factor for overall survival ($P = 0.0168$) (Table 2, bold). Two other genes that showed up-regulation in the malignant breast epithelium were also analysed on the protein level by tissue microarray, namely those encoding COMP [38], a skeletal developmental protein that was not