

## Table 2. Selected cell-type specific expressed genes

The table provides a modified “ $t$  statistic” calculated as described in the text (*Results*, Eq. 2) for each cell type [tumor (TSTAT), BPH, (BSTAT), and stroma (SSTAT)] defined and selected differences (T-B) for each gene with  $t > 2.4$ . The modified  $t$  statistics  $t_{ij}$  incorporates goodness of fit and effect size for every gene  $j$  and every tissue type  $i$ , where  $\sigma_{\beta}$  is the standard error of the coefficient,  $\beta_{ij}$ , and  $k$  is a small constant:

$t_{ij} = \beta_{ij} / (k + \sigma_{\beta})$ . The  $\beta$  are the characteristic gene expression level for a particular cell type and gene. The  $\beta$  are determined according to Eq. 1 as described (see text). For  $n = 88$ , a  $t$  statistic of 2.4 sets thresholds corresponding greater than 4-fold expected differences in expression between the respective cell types ( $P < 0.02$ ). 3384 transcripts displayed cell-type associated gene expression patterns according to the threshold and are listed here. 1096 genes have strong “tumor” association, yet the majority (683) of these represent primarily differences in tumor-stroma gene expression (Fig. 3, tumor > stroma). Conversely, a large number of transcripts are predicted to be stroma associated (Fig 3, groups stroma and stroma > tumor). A total of 492 are strongly associated with BPH cell content. A subset of these (196) also benign prostatic hypertrophy (BPH) showed a strong negative association with tumor cell content indicating potential clinically useful markers of BPH. 413 are predicted to be “tumor specific,” being strongly associated with tumor and displaying negative associations with both BPH and stroma (Figure. 3, tumor). “BSTAT,” “SSTAT,” and “TSTAT” are the corresponding  $t$  statistic values for simple regression using percent composition for each tissue type alone (see Eq. 1), i.e., BPH epithelial cells content, stroma cell content, or tumor epithelial cell content, respectively. The  $t$  statistic value are color coded red for  $t$  statistic  $> 2.0$ ; tan for  $2.0 > t$  statistic  $> 1.0$ , green for  $t$  statistic  $< -2.0$ , and blue for  $-2.0 < t$  statistic  $< -1.0$ .