Supplementary Table S1. List of gene expression datasets included in the analysis of the DISC cohort.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Source ID.** | **Total # of Genes in Array** | **Total # of Samples** | **# of Benign** | **# of Primary** | **# of CRPC/Met** |
| GSE6099 | 10137 | 104 | 52 | 32 | 20 |
| GSE6752 | 12418 | 31 | 0 | 10 | 21 |
| GSE6956 | 13020 | 89 | 20 | 69 | 0 |
| GSE8218 | 13020 | 148 | 71 | 78 | 0 |
| GSE32269 | 13020 | 51 | 0 | 22 | 29 |
| GSE2443 | 13020 | 20 | 0 | 20 | 0 |
| GSE25136 | 13020 | 79 | 0 | 79 | 0 |
| GSE7055 | 13020 | 57 | 0 | 57 | 0 |
| E-SMDB-2486 | 13888 | 112 | 41 | 62 | 9 |
| GSE3933 | 15468 | 112 | 41 | 62 | 9 |
| GSE15484 | 16110 | 65 | 13 | 52 | 0 |
| GSE6919 | 16386 | 160 | 72 | 63 | 25 |
| GSE14206 | 16548 | 67 | 14 | 53 | 0 |
| GSE6811 | 16625 | 35 | 0 | 24 | 11 |
| E-MTAB-154 | 16709 | 48 | 0 | 48 | 0 |
| GSE12378 | 17406 | 39 | 3 | 36 | 0 |
| GSE29079 | 17406 | 95 | 48 | 47 | 0 |
| GSE41408 | 17406 | 48 | 0 | 48 | 0 |
| GSE30521 | 17839 | 23 | 5 | 18 | 0 |
| E-TABM-26 | 18804 | 57 | 13 | 44 | 0 |
| GSE8511 | 18848 | 41 | 16 | 12 | 13 |
| GSE11682 | 19075 | 34 | 17 | 17 | 0 |
| GSE41619 | 19497 | 14 | 0 | 0 | 14 |
| GSE35988 | 19596 | 119 | 28 | 59 | 32 |
| GSE27616 | 19751 | 13 | 4 | 5 | 4 |
| GSE38241 | 19751 | 39 | 21 | 0 | 18 |
| TCGA (2013-04-24) | 20437 | 220 | 44 | 176 | 0 |
| GSE3325 | 20678 | 19 | 6 | 7 | 6 |
| GSE26910 | 20678 | 12 | 6 | 6 | 0 |
| GSE17951 | 20678 | 154 | 81 | 73 | 0 |
| GSE32448 | 20678 | 80 | 40 | 40 | 0 |
| GSE2109 | 20678 | 56 | 0 | 56 | 0 |
| GSE16120 | 22153 | 65 | 14 | 51 | 0 |
| GSE21034 | 22261 | 179 | 29 | 131 | 19 |
| GSE40272 | 24013 | 153 | 52 | 101 | 0 |
| GSE32571 | 24319 | 98 | 39 | 59 | 0 |
| GSE29650 | 24384 | 30 | 0 | 0 | 30 |
| GSE28680 | 27317 | 24 | 4 | 20 | 0 |