**Supplemental Tables**

**Supplemental Table S1**. Literature search for key immunohistochemical markers.

|  |  |
| --- | --- |
|  | Number of publications retrieved in a PubMed Search for protein name and immunohistochemistry |
| Journal | GATA3 | KRT5 | KRT14 | RB | p16 |
| American Journal of Surgical Pathology | 45 | 15 | 14 | 19 | 177 |
| Histopathology | 27 | 38 | 6 | 9 | 69 |
| Archives of Pathology Laboratory Medicine | 9 | 7 | 4 | 1 | 32 |
| total | 81 | 60 | 24 | 29 | 278 |

GATA3, GATA-binding protein 3; KRT5, keratin 5; KRT14, keratin 14; RB, retinoblastoma protein

**Supplemental Table S2.** Ranked performance of individual proteins for identifying subtypes assessed using binary ROC analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Uro vs. Basal |  | GU vs. Basal |  | Uro vs. GU |
| Rank |  | **feature** | **AUC [95% CI]** |  | **feature** | **AUC [95% CI]** |  | **feature** | **AUC [95% CI]** |
| 1 |  | CK5 | 94.39 [91.25-97.52] |  | CK5 | 95.87 [91.81-99.93] |  | RB1 | 92.62 [87.17-98.06] |
| 2 |  | KRT14 | 94.11 [88.87-99.35] |  | KRT14 | 93.93 [88.63-99.22] |  | CCND1 | 90.11 [85.15-95.07] |
| 3 |  | CDH3 | 93.01 [89.05-96.98] |  | CDH3 | 93.74 [89.02-98.47] |  | p16 | 87.15 [80.93-93.37] |
| 4 |  | GATA3 | 91.46 [86.92-96.01] |  | FOXA1 | 89.48 [83.26-95.69] |  | TP63 | 81.68 [73.93-89.44] |
| 5 |  | FOXA1 | 91.32 [87.25-95.39] |  | GATA3 | 89.11 [82.05-96.17] |  | FGFR3 | 78.11 [70.4-85.82] |
| 6 |  | PPARG | 90.69 [87.03-94.35] |  | EPCAM | 88.98 [82.61-95.35] |  | CCNB1 | 71.86 [63.33-80.39] |
| 7 |  | EPCAM | 85.49 [79.75-91.23] |  | TP63 | 88.3 [81.27-95.34] |  | PPARG | 64.16 [54.19-74.13] |
| 8 |  | CDH1 | 77.26 [69.27-85.25] |  | PPARG | 86.45 [79.99-92.91] |  | RXRA | 63.45 [53.94-72.96] |
| 9 |  | KRT20 | 77.2 [72.47-81.93] |  | KRT20 | 79.38 [72.41-86.35] |  | UPK3 | 62.45 [55.04-69.87] |
| 10 |  | VIM | 71.15 [61.72-80.58] |  | CCND1 | 77.17 [67.54-86.8] |  | E2F3 | 62.13 [52.42-71.84] |
| 11 |  | TP63 | 67.77 [59.8-75.75] |  | CDH1 | 72 [61.54-82.45] |  | GATA3 | 60.09 [50.38-69.81] |
| 12 |  | CCNB1 | 65.67 [56.25-75.1] |  | p16 | 71.68 [60.98-82.38] |  | EPCAM | 58.04 [49.01-67.07] |
| 13 |  | UPK3 | 64.58 [59.67-69.48] |  | RB1 | 71.59 [61.86-81.32] |  | VIM | 57.42 [48-66.83] |
| 14 |  | FGFR3 | 62.78 [53.72-71.83] |  | VIM | 67.28 [56.39-78.16] |  | CDH1 | 56.87 [47.84-65.89] |
| 15 |  | CHGA | 56.25 [53.46-59.04] |  | E2F3 | 59.12 [47.54-70.71] |  | FOXA1 | 54.65 [44.8-64.5] |
| 16 |  | RB1 | 56.05 [46.51-65.6] |  | RXRA | 58.54 [47.03-70.05] |  | ZEB2 | 54.32 [51.24-57.4] |
| 17 |  | SYP | 52.21 [50.47-53.94] |  | UPK3 | 57.92 [51.56-64.27] |  | CK5 | 52.69 [43.49-61.89] |
| 18 |  | E2F3 | 51.91 [41.72-62.1] |  | CHGA | 57.41 [52.63-62.19] |  | CDH3 | 52.1 [42.11-62.09] |
| 19 |  | RXRA | 49.98 [40.27-59.7] |  | NCAM1 | 51.78 [42.61-60.96] |  | SYP | 52.07 [48.87-55.27] |
| 20 |  | p16 | 48.65 [38.69-58.62] |  | SYP | 50.93 [49.11-52.74] |  | KRT14 | 51.3 [41.88-60.72] |
| 21 |  | CCND1 | 48.46 [39.17-57.74] |  | CCNB1 | 46.27 [34.41-58.14] |  | TUBB2B | 50.17 [43.66-56.68] |
| 22 |  | NCAM1 | 47.58 [40.05-55.11] |  | FGFR3 | 46 [34.96-57.04] |  | KRT20 | 47.69 [38.19-57.18] |

**Abbreviations:** 95% CI, 95% confidence interval; AUC, area under the receiver operator characteristic curve; CCNB1, cyclin B1; CCND1, cyclin D1; CDH1, cadherin 1; CDH3, cadherin 3; CGHA, chromogranin A; CK5, cytokeratin 5; E2F3, E2F transcription factor 3; EPCAM, epithelial cell adhesion molecule; FGFR3, fibroblast growth factor receptor 3; FOXA1, forkhead box A1; GATA3, GATA-binding protein 3; GU, genomically unstable; KRT5, keratin 5; KRT14, keratin 14; KRT20, keratin 20; NCAM1, neural cell adhesion molecule 1; p16, cyclin-dependent kinase inhibitor 2A; PPARG, peroxisome proliferator-activated receptor gamma; ROC, receiver operator characteristic; RB1, retinoblastoma protein; RXRA, retinoid X receptor alpha; SYP, synaptophysin; TP63, tumor protein 63; TUBB2B, tubulin beta 2B class IIb; UPK3, uroplakin 3; Uro, urothelial-like; VIM, vimentin; ZEB2, zinc finger E-box binding homeobox 2.

**Supplemental Table S3.** Summary table of basal/luminal decision tree: four-fold validation and model accuracies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tree Number | Number of Branches | Markers Used | Marker #1- GATA3-Threshold | Marker #2- CK5 -Threshold | Accuracy |
| Tree 1 | 3 | GATA3 + CK5 | 0.21 | 0.16 | 0.93(0.81-0.99) |
| Tree 2 | 3 | CK5 + GATA3  | 0.35 | 0.51 | 0.86(0.73-0.95) |
| Tree 3 | 3 | GATA3 + CK5 | 0.21 | 0.34 | 0.95(0.85-0.99) |
| Tree 4 | 3 | GATA3 + CK5 | 0.21 | 0.093 | 0.93(0.82-0.99) |

**Abbreviations:** CK5, cytokeratin 5; GATA3, GATA-binding protein 3; IHC, immunohistochemistry

**Supplemental Table S4.** Summary table of basal/GU/Uro decision: tree four-fold validation and model accuracies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Tree Number | Number of Branches | Markers Used | Marker #1- GATA3-Threshold | Marker #2- CK5 -Threshold | Marker #3 P16-Threshold | Accuracy |
| Tree 1 | 3 | GATA3 + p16 | 0.21  |  | 0.42  | 0.74 (0.59-0.87) |
| Tree 2 | 3 | CK5 + p16 |  | 0.51  | 0.42 | 0.72(0.56-0.85) |
| Tree 3 | 3 | GATA3+p16 | 0.21 |  | 0.42 | 0.69(0.53-0.82) |
| Tree 4 | 5 | GATA3 + CK5 + p16 | 0.21 | 0.19 | 0.42 + 0.50 | 0.76(0.61-0.87) |

**Abbreviations:** CK5, cytokeratin 5; GATA3, GATA-binding protein 3; IHC, immunohistochemistry; p16, cyclin-dependent kinase inhibitor 2A