Supplementary Table 1. Curated list of 51 DNA repair genes by pathway.

|  |  |
| --- | --- |
| DNA Repair Pathway | Genes |
| HR | BRCA1, BRCA2, RAD51, RNF168 |
| FA | FANCD2, FANCI, FANCL, FANCA |
| MMR | MLK1, MLH3, MSH2, MSH6, PMS2, MSH3 |
| NHEJ/AEJ | XRCC6, XRCC5, POLQ, RNF168, RNF8, MRE11, NBN, PARP1, RAD50 |
| TLS | RAD18, POLH, POLI, POLK, REV1, UBE2A, UBE2B |
| NER | XPA, XPC, LIG3, LIG1, XRCC1 |
| BER | NEIL1, NEIL2, XRCC1, LIG3, FEN1, POLB |
| APOBEC | APOBEC1, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H |
| CTA | HORMAD1, MAGEA4, (MAGEA1, MAGEA10, MAGEA6) |
| Genes Selected For Classifier | XPA, HORMAD1, PMS2, MLH3, BRCA1, APOBEC3B, FANCA, POLK, MLH1, RAD50, XRCC1, FANCD2, NEIL2, LIG3, MSH6, UBE2B, MAGEA4, XPC, MSH3, XRCC5, POLI, FANCI, FEN1, MAGEA6, NEIL1, REV1, MSH2, APOBEC3A, RAD51, POLQ |

Supplemental Table 2) Table of Silhouette Width and SigClust p-value for unsupervised DNA Repair (U-HRD) groups in TCGA and CBCS. Silhouette widths indicate the degree to which samples within a cluster are more similar to one another than to samples in other clusters, with higher values showing better concordance, while SigClust tests whether observed clusters explain a greater proportion of the variance in expression than the null hypothesis of no clusters.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TCGA |  |  |  |  |  |
|  | High Repair | APOBEC | Heterogeneous | HR/FA |  |
| High Repair | -- | 0.041 | < 0.001 | <0.001 |  |
| APOBEC | -- | -- | 0.176 | 0.019 |  |
| Heterogeneous | -- | -- | -- | < 0.001 |  |
| Silhouette Width | 0.12 | 0.19 | 0.19 | 0.34 |  |
|  |  |  |  |  |  |
| CBCS |  |  |  |  |  |
|  | High Repair | APOBEC | Heterogeneous | HR/FA |  |
| High Repair | -- | < 0.001 | 0.025 | > 0.99 |  |
| APOBEC | -- | -- | < 0.001 | 0.576 |  |
| Heterogeneous | -- | -- | -- | < 0.001 |  |
| Silhouette Width | 0.17 | 0.13 | 0.19 | 0.12 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Supplemental Table 3)

Table of clinical features in U-HRD groups in TCGA and CBCS. Age, PAM50 subtype, ER status, race, stage, and TP53 RNA subtype distributions are listed. P-values determined by Chi-Square tests.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | TCGA |   |   |   |   | CBCS |   |   |   |   |
|   | APOBEC-High | APOBEC-Low | HRD-High | HRD-Low | p | APOBEC-High | APOBEC-Low | HRD-High | HRD-Low | p |
| N | 137 | 421 | 233 | 303 |   | 252 | 130 | 581 | 501 |   |
| **ER Status** |   |   |   |   |   |   |   |   |   |   |
| Negative | 27 (20.6) | 26 ( 6.5) | 150 (66.7) | 34 (11.9) | <0.001 | 79 (31.6) | 8 ( 6.2) | 365 (63.0) | 34 ( 6.8) | <0.001 |
| Positive | 104 (79.4) | 377 (93.5) | 75 (33.3) | 251 (88.1) |   | 171 (68.4) | 122 (93.8) | 214 (37.0) | 467 (93.2) |   |
| **P53 Status (RNA)** |   |   |   |   |   |   |   |   |   |   |
| Mut | 91 (68.4) | 35 ( 9.2) | 187 (85.8) | 158 (55.2) | <0.001 | 109 (74.7) | 37 (30.3) | 261 (57.0) | 56 (12.5) | <0.001 |
| Wt | 42 (31.6) | 345 (90.8) | 31 (14.2) | 128 (44.8) |   | 37 (25.3) | 85 (69.7) | 197 (43.0) | 393 (87.5) |   |
| **Race** |  |  |  |  |  |  |  |  |  |  |
| AA | 27 (22.7) | 50 (12.6) | 57 (25.9) | 49 (18.7) | <0.001 | 124 (49.2) | 72 (55.4) | 372 (64.0) | 215 (42.9) | <0.001 |
| Non-AA | 92 (77.3) | 348 (87.4) | 163 (74.1) | 213 (81.3) |   | 128 (50.8) | 58 (44.6) | 209 (36.0) | 286 (57.1) |   |
| **Age** |  |  |  |  |  |  |  |  |  |  |
| <=50 | 38 (27.7) | 127 (30.2) | 79 (33.9) | 85 (28.1) | 0.462 | 146 (57.9) | 81 (62.3) | 358 (61.6) | 214 (42.7) | <0.001 |
| >50 | 99 (72.3) | 293 (69.8) | 154 (66.1) | 218 (71.9) |   | 106 (42.1) | 49 (37.7) | 223 (38.4) | 287 (57.3) |   |
| **Stage** |   |   |   |   |   |   |   |   |   |   |
| Stage\_I | 21 (15.7) | 88 (21.4) | 34 (14.8) | 39 (13.2) | 0.001 | 73 (30.0) | 38 (29.2) | 170 (29.4) | 232 (46.5) | <0.001 |
| Stage\_II | 74 (55.2) | 216 (52.4) | 155 (67.7) | 174 (59.0) |   | 130 (53.5) | 69 (53.1) | 314 (54.3) | 203 (40.7) |   |
| Stage\_III | 38 (28.4) | 102 (24.8) | 33 (14.4) | 76 (25.8) |   | 34 (14.0) | 15 (11.5) | 75 (13.0) | 54 (10.8) |   |
| Stage\_IV | 1 ( 0.7) | 6 ( 1.5) | 7 ( 3.1) | 6 ( 2.0) |   | 6 ( 2.5) | 8 ( 6.2) | 19 ( 3.3) | 10 ( 2.0) |   |
| **PAM50 Subytpe** |   |   |   |   |   |   |   |   |   |   |
| Basal | 12 ( 8.8) | 6 ( 1.4) | 154 (66.1) | 18 ( 5.9) | <0.001 | 14 ( 9.6) | 7 ( 5.7) | 273 (59.6) | 7 ( 1.6) | <0.001 |
| Her2 | 22 (16.1) | 8 ( 1.9) | 22 ( 9.4) | 30 ( 9.9) |   | 23 (15.8) | 4 ( 3.3) | 44 ( 9.6) | 14 ( 3.1) |   |
| LumA | 54 (39.4) | 364 (86.5) | 29 (12.4) | 120 (39.6) |   | 73 (50.0) | 45 (36.9) | 70 (15.3) | 376 (83.7) |   |
| LumB | 40 (29.2) | 21 ( 5.0) | 20 ( 8.6) | 134 (44.2) |   | 22 (15.1) | 66 (54.1) | 61 (13.3) | 46 (10.2) |   |
| Normal | 9 ( 6.6) | 22 ( 5.2) | 8 ( 3.4) | 1 ( 0.3) |   | 14 ( 9.6) | 0 ( 0.0) | 10 ( 2.2) | 6 ( 1.3) |   |

Supplemental Table 4) Demographic characteristics in TCGA influencing RNA-based S-HRD classifier accuracy. True HRD status was determined using gold-standard DNA-based HRD score, and compared to ClaNC estimated HRD status. The classifier was more sensitive for traditionally aggressive indicators of breast cancer biology, but more specific for less aggressive indicators.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accurate | Inaccurate | False Negative | False Positive | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| N | 720 | 238 | 29 | 209 |   |   |   |
| Age Category |  |  |  |  |   |  |  |
| <=50 Years | 209 (29.1)  |  79 (33.2)  |  9 (31.0)  |  70 (33.5)  | 86 | 71 | 69 |
| >50 Years | 510 (70.9)  | 159 (66.8)  | 20 (69.0)  | 139 (66.5)  | 83 | 75 | 75 |
|  |  |  |  |  |   |  |  |
| PAM50 |  |  |  |  |   |  |  |
| LumA | 444 (61.7)  |  49 (20.6)  | 16 (55.2)  |  33 (15.8)  | 30 | 93 | 89 |
| LumB | 128 (17.8)  |  66 (27.7)  | 10 (34.5)  |  56 (26.8)  | 76 | 63 | 65 |
| Basal | 103 (14.3)  |  61 (25.6)  |  0 ( 0.0)  |  61 (29.2)  | 100 | 2 | 60 |
| Her2 |  28 (3.9)  |  49 (20.6)  |  1 ( 3.4)  |  48 (23.0)  | 91 | 27 | 36 |
| Normal |  17 (2.4)  |  13 (5.5)  |  2 ( 6.9)  |  11 (5.3)  | 50 | 58 | 47 |
|  |  |  |  |  |   |  |  |
| Race |  |  |  |  |   |  |  |
| Black | 113 (18.1)  |  40 (20.8)  |  4 (15.4)  |  36 (21.7)  | 92 | 66 | 71 |
| White | 511 (81.9)  | 152 (79.2)  | 22 (84.6)  | 130 (78.3)  | 80 | 77 | 75 |
|  |  |  |  |  |   |  |  |

Supplemental Table 5)

Table of clinical features in low S-HRD and high S-HRD classifier samples in CBCS dataset. Age, PAM50 subtype, tumor size group, grade, ER status, PR status, ROR-P classification, and lymph node status distributions are listed.

|  |  |  |  |
| --- | --- | --- | --- |
|  | High HRD | Low HRD | RFD (95% CI) |
| N | 648 | 813 |   |
| Age (<=50 years) | 413 (63.7) | 385 (47.4) | 16% (11% - 21%) |
| African American | 422 (65.1)  | 359 (44.2)  | 21% (16% - 26%) |
| Premenopausal | 351 (54.2)  | 340 (41.8)  | 12% (7% - 17%) |
| PAM50 |  |  |  |
|  LumA |  56 (11.6)  | 507 (73.5)  | -62% (-66% - -57%) |
|  LumB |  77 (16.0)  | 118 (17.1)  | -1% (-5% - 3%) |
|  Basal | 287 (59.5)  |  13 ( 1.9)  | 58% (53% - 62%) |
|  Her2 |  56 (11.6)  |  29 ( 4.2)  | 7% (4% - 11%) |
|  Normal |  6 ( 1.2)  |  23 ( 3.3)  | -2% (-4% - 0%) |
|  Missing | 166 | 123 |  |
| Tumor Size |  |  |  |
|  <=2 cm | 238 (37.2)  | 454 (56.3)  | -19% (-24% - -14%) |
|  >2-5 cm | 330 (51.6)  | 298 (37.0)  | 15% (10% - 20%) |
|  >5 cm |  71 (11.1)  |  54 ( 6.7)  | 4% (1% - 7%) |
| P53 (RNA) |  |  |  |
|  Wild-Type | 534 (63.3)  | 177 (22.6)  | -41% (-46% - -35%) |
|  Mutant-Like | 177 (36.7)  | 534 (77.4)  | 41% (35% - 46%) |
|  Missing | 166 | 123 |  |
| Stage |  |  |  |
|  Stage I/II | 529 (82.7) | 698 (86.5) | -4% (-7% - 0%) |
|  Stage III/IV | 111 (17.3) | 109 (13.5) | 4% (0% - 7%) |
|  Missing | 8 | 6 |  |
| Grade |  |  |  |
|  I |  14 (2.2)  | 177 (21.8)  | -24% (-28% - -21%) |
|  II |  68 (10.5)  | 319 (39.2)  | -35% (-40% - -30%) |
|  III | 411 (63.4)  | 159 (19.6)  | 59% (54% - 64%) |
|  Missing | 155 | 158 |  |
| Node Status |  |  |  |
|  Negative | 375 (58.4)  | 475 (58.6)  | 0% (-5% - 5%) |
|  Positive | 267 (41.6)  | 336 (41.4)  | 0% (-5% - 5%) |
|  Missing | 6 | 2 |  |
| ER Status |  |  |  |
|  Positive | 214 (33.2)  | 723 (89.0)  | -56% (-60% - - 52%) |
|  Borderline/Indeterminate |  29 ( 4.5)  |  6 ( 0.7)  | 4% (2% - 6%) |
|  Negative | 402 (62.3)  |  83 (10.2)  | 52% (48% - 56%) |
| PR Status |  |  |  |
|  Positive | 169 (26.3)  | 618 (76.5)  | -50% (-55% - -46%) |
|  Borderline/Indeterminate |  35 ( 5.5)  |  49 ( 6.1)  | -1% (-3% - 2%) |
|  Negative | 438 (68.2)  | 141 (17.5)  | 51% (46% - 55%) |
| ROR-P Classification |  |  |  |
|  Low |  12 ( 2.5)  | 253 (36.7)  | -34% (-38% - -30%) |
|  Intermediate | 212 (44.0)  | 398 (57.7)  | -14% (-19% - -8%) |
|  High | 258 (53.5)  |  39 ( 5.7)  | 48% (43% - 53%) |
|  Missing | 166 | 123 |  |