**Supplementary Information**

**Germline genomic landscapes of breast and lung cancer patients significantly predict clinical outcomes**

**Supplementary Table 1 List of genes of network operational signatures derived from breast cancer germline mutations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Apop1** | **Apop2** | **Apop3** | **CCycle1** | **CCycle2** | **Ccycle3** | **Cell Adh1** | **Cell Adh2** | **CellAdh3** | **Cytosk1** | **Cytosk2** | **Cytosk3** | **Imm Res1** | **Imm Res2** | **ImmRes3** | **Prolif1** | **Prolif2** | **Prolif3** |
| BRCA1 | APOE | ANXA1 | ACVR1 | BRSK1 | BRCA1 | ADAM17 | ADAM17 | ADAM17 | ACTA2 | ACTRT1 | ACTRT1 | CD19 | BCAP31 | B2M | AGT | ADRA1B | ANXA1 |
| CASP8 | AVEN | BCAP31 | ANXA1 | CCNB1 | C13orf34 | AMBN | BRCA1 | CAV1 | ACTR2 | APOE | AKAP9 | CD4 | CCL4 | C3 | AMBN | BIRC2 | BOK |
| CDK11A | BIRC2 | BFAR | ATR | CCNB3 | CCNB1 | BAI1 | CCL4 | CD209 | ACTRT1 | BBS4 | ALMS1 | CD81 | CCR5 | CCL23 | ASCL1 | BMPR2 | CD81 |
| DOCK1 | BRCA1 | BRCA1 | BRD7 | CDK11A | CCNB3 | CD209 | CD209 | COL11A2 | ASAP1 | CASP8 | CASP8 | CHIA | CD160 | CCL3 | CD81 | BRCA1 | CITED1 |
| EP300 | CDK11A | CDK11A | BRSK1 | CHFR | CDC25B | CD4 | COL6A3 | COL6A3 | BRSK1 | CDK5RAP2 | CYTH2 | CIITA | CD19 | CD19 | CITED1 | CD160 | CITED2 |
| ESPL1 | DAPK1 | CITED2 | CASP8AP2 | CLASP1 | CDCA5 | CD58 | CTGF | CTNND1 | CASP8 | CENPJ | DAPK3 | HCST | CD81 | CD81 | CSF1R | CD81 | CKS1B |
| FGF2 | ESPL1 | ESPL1 | CCNB1 | DCTN2 | CDK11A | CDH17 | DDR1 | EPHA3 | CCNB1 | CEP72 | DIAPH2 | HLA-DMB | CIITA | CIITA | DBH | CDK9 | CTF1 |
| GPX1 | GSN | FAIM3 | CCNB3 | DDIT3 | CHFR | CDH4 | EZR | FAT1 | CDC25B | CEP78 | GSG2 | HLA-DOA | FASLG | ENPP2 | GHRL | CITED1 | DAB2 |
| GSN | GULP1 | FBXO7 | CDK11A | E2F6 | FBXO31 | COL5A1 | F11R | FLOT2 | DAPK3 | CKAP5 | GSN | HLA-DOB | HCST | HCST | GNB1 | DCTN2 | DBH |
| HIPK1 | GZMB | GSN | CENPJ | GSG2 | KIF11 | COL6A3 | FAT1 | FN1 | GSN | CLASP1 | LSP1 | HLA-DPA1 | ICAM1 | ICAM2 | IGSF8 | EDNRA | FAS |
| HYAL1 | GZMH | HIPK1 | CHAF1B | KIF11 | NASP | DPP4 | FLOT2 | IBSP | KIF11 | DAPK1 | MAP2K1 | HLA-DQA2 | ICAM2 | ICAM3 | LTK | ENG | GHRL |
| IKBKB | HIPK1 | IER3 | CHFR | MAPK7 | NCAPD2 | ENG | FN1 | ITGAM | LSP1 | DAPK3 | NCAPG | HLA-DRB3 | ICAM3 | IFITM1 | MDM4 | GHRL | GNB1 |
| IL17A | IKBKB | IKBKB | EP300 | MRE11A | NCAPG | FAT1 | ICAM1 | LAYN | NCAPG | DCTN2 | NF1 | HLA-DRB4 | IFITM1 | IL17A | MMP7 | GNB1 | HOXB4 |
| IL7 | IL17A | IL17A | FBXO31 | NASP | NOTCH1 | FLOT2 | IGFALS | LEF1 | NCAPH | DCTN3 | NPM1 | HLA-DRB5 | IL15 | IL21 | NASP | IL2RA | LTK |
| MAP2K7 | IL7 | IL7 | FOXN2 | NCAPD3 | OFD1 | FN1 | ITGA4 | LIMS1 | NF1 | DIAPH1 | PCM1 | ICAM2 | IL17A | IL7 | NCF1 | INHA | MPL |
| NEFH | LTK | LTBR | KIF11 | NCAPG | PIN1 | ITGAL | ITGAV | MAP2K1 | NPM1 | EZR | PCNT | ICAM3 | IL7 | ITGA4 | PIM2 | MRE11A | PAK7 |
| PIM2 | PLEC | MDM4 | LZTS1 | NOTCH1 | PINX1 | LEF1 | JAM3 | MAP2K2 | NUCB1 | GSN | PDPK1 | ICAM4 | IL9 | ITGAL | PLAU | PLAU | PIN1 |
| PSMD5 | PSME1 | MUC1 | MAP2K1 | OFD1 | PLK1 | LY9 | LEF1 | MLLT4 | OFD1 | GTF2F2 | PFN1 | IFITM1 | ITGAL | KLRC1 | PLK1 | PLK1 | PLAU |
| PSME1 | PSME4 | PCNT | MDC1 | P2RY2 | RAD21 | MAP2K2 | MAP2K2 | MOG | PLK1 | LSP1 | PLEKHG6 | IL17A | KLRD1 | NOTCH1 | PPP1R9B | PPP1R9B | PPP1R9B |
| PSME4 | RASSF5 | PIM2 | MTBP | PINX1 | RBBP8 | MOG | MDC1 | MUC4 | PTPN13 | NCAPG | PLK1 | IL18 | MAP3K14 | PDCD1LG2 | RAP1B | PTEN | SAT1 |
| PTK2B | SEMA3A | PSME1 | NASP | PPP1CA | REC8 | MUC4 | MOG | NFASC | RASSF5 | NF1 | PTK2B | IL2RG | MICB | RAET1G | SAT1 | SAT1 | SGK1 |
| RAD21 | SERPINB9 | PSME4 | NCAPG | RAD50 | RNF8 | MYH6 | MUC4 | NLGN4X | RPA1 | NPM1 | RASSF1 | IL7 | NCF4 | RAET1L | SCRIB | SRA1 | SOX9 |
| SEMA3A | SRA1 | SLK | NOTCH1 | RNF8 | RPA1 | NLGN4X | NLGN4X | NRP1 | S100A9 | PLK1 | RASSF5 | ITGAL | NOTCH1 | SPN | SPHK2 | TACC3 | TCF7L2 |
| SMNDC1 | TGFBR2 | TGFB2 | OFD1 | RPA1 | RPS6KA1 | NPNT | NRP1 | OMD | SCNN1A | PSMD10 | RPA1 | NOTCH1 | RAET1E | TLR10 | TCF7L2 | TCF7L2 | TGFB2 |
| TGFBR2 | TNFRSF10B | TGFBR2 | PIM2 | RPS6KA1 | SMC1A | NRP1 | OMD | PAK1 | SORBS1 | RASSF5 | S100A9 | RAET1G | RAET1G | TNFRSF4 | TGFB2 | TGFB2 | TGFBR2 |
| TMEM173 | TNFSF10 | TNFRSF10B | PINX1 | SMC1B | SMC2 | OMD | PDPK1 | PKN2 | TESK2 | RPA1 | SORBS1 | RGS1 | SERPINB9 | TNFSF10 | TGFBR2 | TGFBR2 | THPO |
| TNFRSF10B | TNFSF9 | TNFRSF25 | RNF8 | SMC3 | SMC3 | PTK2B | SORBS1 | PLEC | TPM1 | S100A9 | TPM1 | TLR10 | TLR10 | TNFSF15 | TNFSF8 | TNFSF9 | TNFSF9 |
| TNFSF10 | TP53BP2 | TNFSF10 | RPA1 | TGFB2 | SMC4 | S1PR1 | TENC1 | SORBS1 | TSSK2 | SORBS1 | TUBA4B | TNFSF10 | TNFSF10 | ULBP1 | TNFSF9 | UBE2L3 | UBE2L3 |
| TREX1 | TP63 | TREX1 | RPS6KA1 | TGFB3 | STAG1 | SORBS1 | TPPP | TNS1 | UBXN6 | TPM1 | UNC119 | ULBP2 | TNFSF9 | ULBP3 | UBE2L3 | UCHL1 | UCHL1 |
| YARS | TREX1 | VHL | SMC3 | USP2 | STAG2 | TPPP | VCAM1 | TPPP | UNC119 | UNC119 | WIPF1 | ULBP3 | ULBP3 | VCAM1 | UCHL1 | DBH | ZEB1 |

Abbreviations: Apop, Apoptosis; CCycle, Cell Cycle; CellAdh, Cell Adhesion; Cytosk, Cytoskeleton, ImmRes, Immune Response; Prolif, Cell Proliferation.

**Supplementary Table 2 Prediction accuracy and recall rate for the gene signatures derived from germline mutations of breast cancer tumors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Number of samples** | **Cancer Hallmark** | **Low-risk** | **High-risk** |
| **Accuracy (%)\*** | **Recall (%)†** | **Accuracy (%)\*\*** | **Recall (%)††** |
| Training | 200 | Apoptosis 1 | 94.8 | 83.3 | 24.0 | 53.5 |
| Apoptosis 2 | 94.9 | 83.3 | 24.5 | 54.7 |
| Apoptosis 3 | 91.7 | 73.3 | 21.2 | 51.8 |
| Cell Cycle 1 | 92.3 | 76.7 | 21.1 | 49.4 |
| Cell Cycle 2 | 91.4 | 70.0 | 21.9 | 55.9 |
| Cell Cycle 3 | 91.1 | 70.0 | 21.2 | 54.1 |
| Cell Adhesion 1 | 90.1 | 66.7 | 20.2 | 53.5 |
| Cell Adhesion 2 | 84.4 | 53.3 | 14.6 | 44.7 |
| Cell Adhesion 3 | 93.7 | 80.0 | 22.9 | 52.4 |
| Cytoskeleton 1 | 86.7 | 56.7 | 16.7 | 50.0 |
| Cytoskeleton 2 | 77.0 | 23.3 | 7.0 | 45.3 |
| Cytoskeleton 3 | 78.7 | 36.7 | 9.9 | 41.2 |
| Immune Response 1 | 90.5 | 70.0 | 20.0 | 50.6 |
| Immune Response 2 | 87.9 | 60.0 | 17.8 | 51.2 |
| Immune Response 3 | 86.9 | 56.7 | 16.8 | 50.6 |
| Cell Proliferation 1 | 85.6 | 50.0 | 15.6 | 52.4 |
| Cell Proliferation 2 | 86.4 | 53.3 | 16.5 | 52.4 |
|  |  | Cell Proliferation 3 | 93.6 | 80.0 | 22.6 | 51.8 |
| TCGA-Nature | 200 | Apoptosis 1 | 91.9 | 50.6 | 11.9 | 60.0 |
| Apoptosis 2 | 91.3 | 52.2 | 11.3 | 55.0 |
| Apoptosis 3 | 93.1 | 52.8 | 13.3 | 65.0 |
| Cell Cycle 1 | 89.3 | 51.1 | 9.3 | 45.0 |
| Cell Cycle 2 | 93.8 | 50.0 | 13.5 | 70.0 |
| Cell Cycle 3 | 92.9 | 51.1 | 12.9 | 65.0 |
| Cell Adhesion 1 | 88.6 | 51.7 | 8.4 | 40.0 |
| Cell Adhesion 2 | 93.7 | 57.8 | 14.6 | 65.0 |
| Cell Adhesion 3 | 87.4 | 50.0 | 7.2 | 35.0 |
| Cytoskeleton 1 | 88.2 | 50.0 | 8.2 | 40.0 |
| Cytoskeleton 2 | 89.6 | 52.8 | 9.6 | 45.0 |
| Cytoskeleton 3 | 90.4 | 47.2 | 10.4 | 55.0 |
| Immune Response 1 | 87.8 | 47.8 | 7.8 | 40.0 |
| Immune Response 2 | 90.2 | 46.1 | 10.2 | 55.0 |
| Immune Response 3 | 90.3 | 51.7 | 10.3 | 50.0 |
| Cell Proliferation 1 | 92.2 | 52.2 | 12.2 | 60.0 |
| Cell Proliferation 2 | 93.6 | 57.2 | 14.4 | 65.0 |
|  |  | Cell Proliferation 3 | 93.3 | 54.4 | 13.7 | 65.0 |
| TCGA-CPTAC | 295 | Apoptosis 1 | 92.8 | 59.4 | 17.2 | 64.7 |
| Apoptosis 2 | 95.1 | 59.4 | 19.7 | 76.5 |
| Apoptosis 3 | 92.8 | 59.0 | 17.1 | 64.7 |
| Cell Cycle 1 | 93.4 | 59.4 | 17.8 | 67.7 |
| Cell Cycle 2 | 94.2 | 56.3 | 18.0 | 73.5 |
| Cell Cycle 3 | 91.5 | 57.5 | 15.3 | 58.8 |
| Cell Adhesion 1 | 91.6 | 62.5 | 16.2 | 55.9 |
| Cell Adhesion 2 | 93.0 | 61.3 | 17.9 | 64.7 |
| Cell Adhesion 3 | 91.8 | 64.4 | 17.0 | 55.9 |
| Cytoskeleton 1 | 91.9 | 69.7 | 18.6 | 52.9 |
| Cytoskeleton 2 | 93.2 | 57.5 | 17.2 | 67.7 |
| Cytoskeleton 3 | 94.6 | 67.4 | 22.0 | 70.6 |
| Immune Response 1 | 92.2 | 59.0 | 16.4 | 61.8 |
| Immune Response 2 | 91.4 | 65.5 | 16.7 | 52.9 |
| Immune Response 3 | 92.5 | 70.9 | 20.0 | 55.9 |
| Cell Proliferation 1 | 92.1 | 62.5 | 17.0 | 58.8 |
| Cell Proliferation 2 | 93.1 | 56.7 | 16.9 | 67.7 |
|  |  | Cell Proliferation 3 | 93.5 | 65.9 | 19.8 | 64.7 |

Notes:

\*Percentage of non-recurred (i.e., non-metastatic) samples in the predicted low-risk group.

†Percentage of the predicted low-risk samples from the non-recurred group.

\*\*Percentage of recurred (i.e., metastatic) samples in the predicted high-risk group.

††Percentage of the predicted high-risk samples from the recurred group.

**Supplementary Table 3 Cox proportional hazards regression model of uni- and multiple-factor for breast cancer**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **P-Value** | **HR** | **95% CI** |
| Age  | 0.13 | 1.006762 | 0.998-1.016 |
| Subtype, Luminal A v Luminal B | 0.04 | 1.32708 | 1.0194-1.728 |
| Subtype, Luminal A v Unknown | 0.82 | 0.95363 | 0.6335-1.436 |
| Subtype, Luminal B v Unknown | 0.15 | 1.3874 | 0.8933-2.155 |
| Localization, Left v Right | 0.03 | 0.7659 | 0.6014-0.9755 |
| Stage, I v II | 0.31 | 1.1703 | 0.8612-1.590 |
| Stage, I v III | 0.28 | 1.2240 | 0.8475-1.768 |
| Stage, I v IV | 0.06 | 0.2250 | 0.0622-1.045 |
| Stage, I v X | 0.31 | 0.5482 | 0.1715-1.752 |
| Subtype + Localization + Stage, Luminal A v Luminal B | 0.02 | 1.38377 | 1.05771-1.810 |
| Subtype + Localization + Stage, I v IV | 0.04 | 0.23098 | 0.05586-0.9551 |

Abbreviations: HR, hazard ratio; CI, Confidence Interval

**Supplementary Table 4 Sample filtering steps in breast cancer dataset**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Clinical Information** | **Sequencing** | **Training Set** | **Testing Set** | **Validation Set 1 (TCGA-CPTAC)** | **Validation Set 2 (TCGA-Nature)** |
| Breast | 1067 | 755 | 200 | 60 | 295 | 200 |

**Supplementary Table 5 List of genes of network operational signatures derived from lung cancer germline mutations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Apop1** | **Apop2** | **Apop3** | **Apop4** | **CCycle** | **CellAdh1** | **CellAdh2** | **Cytosk1** | **Cytosk2** | **Cytosk3** | **Cytosk4** | **ImmRes1** | **ImmRes2** | **Prolif** |
| AEN | ADAM17 | ADAM17 | AEN | AHR | APBA1 | ACTN1 | ACTR2 | ACTN4 | ACTN1 | ACTN1 | AIRE | AIRE | AURKB |
| AHR | AEN | AHR | AHR | BIRC5 | CASK | ADAM17 | APC2 | ACTR2 | ACTR2 | AIRE | BMPR1A | CCL11 | BIRC7 |
| APOE | AHR | BCL2L2 | ATP7A | CDKN2A | CCL2 | AMIGO3 | CAMK2B | ADAM17 | AURKB | AURKB | CCL11 | CCL19 | BST2 |
| CASP3 | ANGPT4 | E2F1 | BCAP31 | CKS1B | CDH17 | APBA1 | CDK7 | AURKB | CAMK2B | CDK7 | CCL17 | CCL7 | CD81 |
| DAD1 | ANXA1 | FAIM3 | BNIP1 | DCTN1 | CLDN11 | CASK | DCTN1 | CAMK2B | CDC42 | CTNNA1 | CD200 | CD96 | CDK11B |
| DCTN1 | ARHGEF4 | FGF2 | CASP3 | EHMT1 | CNTNAP1 | CDH17 | DES | CDK7 | CDK7 | DCTN1 | CTSG | CHIA | CTNNA1 |
| E2F1 | ATF5 | HIPK3 | DAD1 | ENSA | COL19A1 | CDH3 | EPB42 | CTNNA1 | CTNNA1 | DES | CTSW | CXCL13 | DBP |
| FAIM3 | BIRC7 | IFI6 | E2F1 | FOXO4 | COL4A3 | CNTNAP1 | FOXJ1 | CTNNA2 | DCTN1 | DNAJC7 | CXCL11 | ENPP1 | DLX5 |
| FGF2 | BNIP1 | IFT57 | FGF2 | FZD9 | COL4A6 | COL11A2 | FRMD4B | DES | DES | FOXJ1 | CXCL12 | HLA-B | ESRRA |
| FOXO3 | CXCR4 | LY86 | FOXO3 | FZR1 | COL8A1 | COL19A1 | IFT57 | FHL2 | EPB42 | FRMD4B | ETS1 | HLA-DQA1 | FOXO3 |
| HIPK1 | DAD1 | MAPK1 | HIPK1 | GGCX | CTNNA1 | COL8A1 | IPP | FRMD4B | FHL3 | HAUS2 | HLA-B | HLA-DQB1 | GLUL |
| HIPK2 | DIABLO | MCF2L | HIPK2 | GPS1 | DPP4 | CTNNA1 | KLHL41 | HAUS2 | FOXJ1 | IFT57 | HLA-DQA1 | HLA-DQB2 | IL1A |
| HSPA1B | E2F1 | NBN | HSPA1B | HAUS2 | ENAH | CXCL12 | LIMD1 | IFT57 | FRMD4B | IPP | HLA-DQB1 | IFITM2 | KAT2A |
| IFT57 | FAIM3 | NOP56 | HTT | JAG2 | EVL | DPP4 | MAPK1 | IPP | IPP | KIF5B | ICAM2 | IK | KIF15 |
| MAPK1 | FOXO3 | OGT | IFT57 | LATS1 | FEZ1 | FEZ1 | MYOT | KIF5B | KIF11 | KLHL41 | IGHA2 | IL18R1 | LIPA |
| OGT | GAS2 | PRUNE2 | IL2 | LIG1 | ITGA3 | HAS1 | NCAPH | KLHL41 | KLHL41 | LIMD1 | IGHM | IL18RAP | LRP1 |
| PRKCQ | HIPK2 | PSMB5 | MAPT | LIG4 | LAMB1 | LRFN3 | NEB | LIMD1 | MAPK1 | MAPK1 | IGHV3-23 | ITGA4 | MET |
| PSMA4 | IFT57 | PSMD1 | NBN | MAPK1 | LY9 | LY9 | NEFM | MACF1 | MYOT | NCAPH | IL17A | KIR2DL2 | MORF4L1 |
| PSMA6 | MAPK1 | RAC1 | OGT | MAPK6 | MCAM | MAPK1 | NR1I3 | MAPK1 | NCAPH | NCK1 | IL7 | KIR2DS1 | MXD1 |
| PSMB10 | NOD1 | ROCK1 | PRKCQ | MEN1 | MYH7 | MCAM | NUP62 | NCK1 | NEB | NR1I3 | IL7R | KIR2DS5 | PCNA |
| PSMB5 | PAK1 | RRAGA | PSMA6 | NEK9 | NPTN | MMRN1 | PCNA | NEFM | NR1I3 | PCNA | KIR2DS2 | KLRD1 | PES1 |
| RAC1 | PGAP2 | SIAH1 | PSMB10 | NOTCH1 | OPCML | MYBPH | PFN2 | NOS3 | PCNA | PFN2 | KLRD1 | LILRB2 | PPP1R8 |
| ROCK1 | PLP1 | SLC9A3R1 | RAC1 | PPM1D | PAK1 | MYH7 | RPGR | NR1I3 | ROCK1 | ROCK1 | LILRB2 | OPRK1 | REG1B |
| SIAH1 | PRKCQ | SOD2 | ROCK1 | SGSM3 | PCDHGB2 | NEO1 | RPS7 | PARVA | RPGR | RPGR | LST1 | PVR | RPS27 |
| SLC9A3R1 | PRUNE2 | STK3 | SLC9A3R1 | SPIN1 | RAC1 | NLGN3 | SLC16A3 | PFN2 | SLC16A3 | SLC16A3 | NOTCH1 | PVRL2 | TACC1 |
| STK3 | PSMA6 | TARDBP | SOD2 | TACC1 | SCARB1 | PCDHAC2 | TBCA | ROCK1 | SNCA | TCTN2 | PDCD1LG2 | TAPBP | TACSTD2 |
| TCTN3 | ROCK1 | TBP | SPIN2B | TNFAIP1 | SELPLG | SELPLG | TCTN2 | SLC16A3 | TACC1 | TRIP4 | PVR | TLR6 | TBX1 |
| TP53 | SOD2 | TCTN3 | STK3 | TUBB2A | SEMA5A | SEMA5A | TRIP4 | TCTN2 | TBCA | TUBB2A | SELL | TNFAIP1 | TGFBI |
| VPS35 | TICAM1 | TP53 | TP53 | UBE2I | TPBG | SRC | TUBA1A | TMSB4Y | TCTN2 | TUBGCP4 | VCAM1 | VIPR1 | TNFSF9 |
| YWHAE | TRIB3 | YWHAE | VPS35 | VASH1 | ZYX | THEMIS2 | TUBGCP4 | TTLL1 | TUBGCP4 | WDR1 | XBP1 | XBP1 | UBE2V2 |

Abbreviations: Apop, Apoptosis; CCycle, Cell Cycle; CellAdh, Cell Adhesion; Cytosk, Cytoskeleton, ImmRes, Immune Response; Prolif, Cell Proliferation.

**Supplementary Table 6 Prediction accuracy and recall rate for the gene signatures derived from germline mutations of lung cancer tumors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Number of samples** | **Cancer Hallmark** | **Low-risk** | **High-risk** |
| **Accuracy (%)\*** | **Recall (%)†** | **Accuracy (%)\*\*** | **Recall (%)††** |
| Training | 200 | Apoptosis 1 | 51.89 | 55.00 | 52.13 | 49.00 |
| Apoptosis 2 | 51.11 | 69.00 | 52.31 | 34.00 |
| Apoptosis 3 | 63.20 | 79.00 | 72.00 | 54.00 |
| Apoptosis 4 | 46.88 | 45.00 | 47.12 | 49.00 |
| Cell Cycle  | 54.92 | 67.00 | 57.69 | 45.00 |
| Cell Adhesion 1 | 34.75 | 41.00 | 28.05 | 23.00 |
| Cell Adhesion 2 | 46.79 | 51.00 | 46.15 | 42.00 |
| Cytoskeleton 1 | 64.96 | 76.00 | 71.08 | 59.00 |
| Cytoskeleton 2 | 63.28 | 81.00 | 73.61 | 53.00 |
| Cytoskeleton 3 | 60.32 | 76.00 | 67.57 | 50.00 |
| Cytoskeleton 4 | 57.69 | 75.00 | 64.29 | 45.00 |
| Immune Response 1 | 52.17 | 60.00 | 52.94 | 45.00 |
| Immune Response 2 | 33.98 | 35.00 | 32.99 | 32.00 |
| Cell Proliferation | 32.04 | 33.00 | 30.93 | 30.00 |
| Validation Set | 176 | Apoptosis 1 | 54.00 | 63.53 | 59.21 | 49.45 |
| Apoptosis 2 | 48.94 | 54.12 | 52.44 | 47.25 |
| Apoptosis 3 | 42.86 | 56.47 | 42.19 | 29.67 |
| Apoptosis 4 | 52.87 | 54.12 | 56.18 | 54.95 |
| Cell Cycle | 56.57 | 65.88 | 62.34 | 52.75 |
| Cell Adhesion 1 | 60.26 | 55.29 | 61.22 | 65.93 |
| Cell Adhesion 2 | 55.79 | 62.35 | 60.49 | 53.85 |
| Cytoskeleton 1 | 49.51 | 60.00 | 53.42 | 42.86 |
| Cytoskeleton 2 | 49.04 | 60.00 | 52.78 | 41.76 |
| Cytoskeleton 3 | 49.07 | 62.35 | 52.94 | 39.56 |
| Cytoskeleton 4 | 47.62 | 58.82 | 50.70 | 39.56 |
| Immune Response 1 | 50.00 | 51.76 | 53.41 | 51.65 |
| Immune Response 2 | 56.99 | 62.35 | 61.45 | 56.04 |
| Cell Proliferation | 65.38 | 60.00 | 65.31 | 70.33 |

Notes:

\*Percentage of non-recurred (i.e., non-metastatic) samples in the predicted low-risk group.

†Percentage of the predicted low-risk samples from the non-recurred group.

\*\*Percentage of recurred (i.e., metastatic) samples in the predicted high-risk group.

††Percentage of the predicted high-risk samples from the recurred group.

**Supplementary Table 7 Cox proportional hazards regression model of uni- and multiple-factor for lung cancer**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **P-Value** | **HR** | **95% CI** |
| Age  | 0.87 | 0.9985 | 0.9811-1.016 |
| Gender, Male v Female | 0.31 | 0.8411 | 0.6007-1.178 |
| Localization, Left v Right | 0.34 | 1.175 | 0.8414-1.641 |
| Stage, I v II | 0.07 | 0.6636 | 0.4271-1.031 |
| Stage, I v III | 0.07 | 0.5987 | 0.3469-1.033 |
| Stage, I v IV | 0.46 | 0.7102 | 0.2880-1.751 |
| Gender + Localization + Stage, I v II | 0.08 | 0.6695 | 0.4302-1.042 |
| Gender + Localization + Stage, I v III | 0.08 | 0.6112 | 0.3537-1.056 |

Abbreviations: HR, hazard ratio; CI, Confidence Interval

**Supplementary Table 8 Sample filtering steps in lung cancer dataset**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dataset** | **Clinical Information** | **Sequencing** | **Training Set** | **Testing Set** | **Validation Set** |
| Lung | 517 | 436 | 200 | 60 | 176 |

**Supplementary Table 9 Pathway enrichment analysis of network operational signature genes from germlines’ mutations of breast cancer patients**

|  |  |  |
| --- | --- | --- |
| **Category** | **Term** | **FDR P\_Value** |
| GOTERM | Antigen processing and presentation | 8.38E-09 |
| KEYWORDS | Mitosis | 2.12E-08 |
| PATHWAY | Cytokine-cytokine receptor interaction | 3.01E-08 |
| GOTERM | Cell division | 1.99E-06 |
| GOTERM | Natural killer cell lectin-like receptor binding | 2.14E-06 |
| PATHWAY | Leishmaniasis | 7.55E-06 |
| GOTERM | T cell co-stimulation | 1.10E-05 |
| PATHWAY | Graft-versus-host disease | 1.43E-05 |
| GOTERM | Natural killer cell mediated cytotoxicity | 1.64E-05 |
| PATHWAY | Rheumatoid arthritis | 1.74E-05 |
| PATHWAY | Inflammatory bowel disease (IBD) | 1.84E-05 |
| GOTERM | MHC class II protein complex | 4.02E-05 |
| PATHWAY | Viral myocarditis | 4.40E-05 |
| PATHWAY | Allograft rejection | 4.87E-05 |
| PATHWAY | Intestinal immune network for IgA production | 5.15E-05 |
| PATHWAY | Negative regulation of extrinsic apoptotic signaling pathway via death domain receptors | 9.50E-05 |
| PATHWAY | Antigen processing and presentation | 1.60E-04 |
| GOTERM | Mitotic chromosome condensation | 1.62E-04 |
| PATHWAY | Type I diabetes mellitus | 1.82E-04 |
| PATHWAY | Staphylococcus aureus infection | 2.39E-04 |
| PATHWAY | Influenza A | 3.50E-04 |
| PATHWAY | Toxoplasmosis | 9.54E-04 |
| PATHWAY | Asthma | 0.00131511 |
| PATHWAY | Autoimmune thyroid disease | 0.00153788 |
| PATHWAY | Tuberculosis | 0.00210254 |
| PATHWAY | HTLV-I infection | 0.00252675 |
| PATHWAY | FoxO signaling pathway | 0.00487651 |
| PATHWAY | Cell cycle | 0.00936092 |
| PATHWAY | Stimulatory C-type lectin receptor signaling pathway | 0.01846333 |
| PATHWAY | Negative regulation of canonical Wnt signaling pathway | 0.03468057 |
| PATHWAY | Interferon-gamma-mediated signaling pathway | 0.04199796 |
| GOTERM | Positive regulation of inflammatory response | 0.05155117 |

**Supplementary Table 10 Pathway enrichment analysis of network operational signature genes from germlines’ mutations of lung cancer patients**

|  |  |  |
| --- | --- | --- |
| **Category** | **Term** | **FDR P\_Value** |
| GOTERM | Inflammatory response | 3.58E-04 |
| PATHWAY | T cell receptor signaling pathway | 0.00339697 |
| GOTERM | T cell co-stimulation | 0.00364395 |
| GOTERM | Chemokine activity | 0.00602057 |
| PATHWAY | Cytokine-cytokine receptor interaction | 0.01423008 |
| PATHWAY | Chemokine signaling pathway | 0.01709647 |
| PATHWAY | Antigen processing and presentation | 0.050449 |
| PATHWAY | Negative regulation of canonical Wnt signaling pathway | 0.05194507 |
| PATHWAY | Leukocyte transendothelial migration | 0.06788154 |
| GOTERM | Cell division | 0.11210005 |
| PATHWAY | Adherens junction | 0.19492581 |

**Supplementary Methods**

**Sequencing data pre-processing and variant calling**

The GATK1 pipeline-based whole exome-sequence data pre-processing was described previously2 . Briefly, duplicate reads were marked and removed using GATK’s markDuplicates removed using BamTools. Reads with a low mapping quality (n=60) were also removed using BamTools3. Local realignment around indels was made using GATK’s IndelRealign/RealignTargetCreator and finally, base recalibration was conducted using GATK’s BaseRecalibrator. All the variants were obtained using the Varscan24 somatic option by analyzing normal/tumor matched sequencing files. Variants with strand-specific bias, coverage less than 30 reads and variant frequency for heterozygous calls of less than 0.08 were removed.

**Determining tumor purity**

Tumor purity was obtained using absCNseq5. To run absCNseq, for a given tumor, we generated a segmentation file and a SNV (Single Nucleotide Variants) file. The segmentation file was generated by running Varscan2 using the standard protocol. Briefly, we ran VarScan2’s copyNumber on normal and tumor BAM files, and VarScan2’s copyCaller to adjust for GC content and finally applied circular binary segmentation. The SNV file was then transformed from the VCF file by running VarScan2. For some samples, absCNseq could give a few purity solutions. In this situation, the consensus purity was selected. The samples with purity greater than 70% were retained for downstream analyses. Ultimately, 755 ER+ breast tumor samples were available for further analysis. For lung cancer, 436 samples were used for downstream analysis.

**Training and validation set for ER+ breast and lung cancer datasets**

To identify gene signatures of ER+ breast cancer, we randomly selected 200 samples, which have follow-up time, as the training set (30 and 170 for recurred and non-recurred samples, respectively). By default, ~15% of ER+ breast tumors get recurred within 10 years6. Clinically, at present almost all of the ER+/luminal breast cancer patients receive tamoxifen treatment. However, tamoxifen treatment for the ‘real’ low-risk patients does not affect patients’ survival. To develop gene signatures for prognosis, we controlled the training set such that we tried to make sure that the selected ‘low-risk’ patients are ‘real low-risk’ patients by applying these rules: (1) the low-risk patients who have relatively longer survival in the cohort, (2) we further confirmed them by predicting them using the gene-expression-based prognostic signatures we developed previously7. This signature was developed using a cohort where the patients have not been treated with any chemotherapy. The predicting accuracy for low-risk ER+ breast cancer reached 95%. Except these training samples, the rest of the ER+ breast tumor samples in the GDC was used for validation. Sixty non-recurred samples were retained for obtaining optimal signature cutoffs (Supplementary Table 5). Finally, all the remaining ER+ samples were separated into 2 validation sets (TCGA-Nature, TCGA-CPTAC) composed of 200 and 295 samples, respectively. For TCGA-Nature set, we used a ratio of 10% of recurred samples (20 recurred and 180 non-recurred samples). For TCGA-CPTAC, we used a ratio of 11.5% of recurred samples (34 recurred and 261 non-recurred samples).

For lung cancer, we randomly selected 200 samples, which also have follow-up time, as the training set (100 and 100 for recurred and non-recurred samples, respectively). The predicting accuracy for low-risk lung cancer reached 73%. Except the training samples, the rest of the lung tumor samples in the GDC can be used for validation. As for the ER+ breast cancer dataset, sixty non-recurred samples were retained for obtaining optimal signature cutoffs (Supplementary Table 10). Finally, all the remaining lung samples were taken as a validation set of 176 samples. We used a ratio of 43.8% of recurred samples (77 recurred and 99 non-recurred samples).

**Identifying germline mutations**

Based on tumor purity, sequencing reads from each variant were adjusted accordingly, and then the VAF (Variant Allele Frequency) was recalculated. Only mutations in 2n regions of chromosomes (i.e., not in the amplification and deletion regions), which were obtained from the segmental files of tumors from TCGA, were considered. Germline mutations included (1) a homozygous mutation whose VAF is greater than 90 in both normal and tumor samples; (2) a heterozygous mutation whose VAF is 55 >= N >= 45 in normal samples. Using these data, we re-ran eTumorMetastasis and obtained similar results.

**Construction of lung cancer-specific metastasis network**

To construct a lung cancer specific metastasis network, we modified the procedure for constructing ER+ specific cancer survival and proliferation networks2. Briefly, we extracted a subnetwork by mapping the lung cancer specific metastasis-associated genes onto the global literature-curated human signaling network. To build the global literature-curated network, we combined genes and their relations from a variety of existing databases and networks: BioGRID8, BIND9, CRG-all network10, I2D11, IntAct12, PID13, Mint14, MIPS15, HumanNet16 and our own literature-curated human signaling network8. Then, we first identified metastasis-associated genes using lung cancer cell lines and tumor samples. We obtained the gene expression data of 22 lung cancer cell lines from the *Cancer Cell Line Encyclopedia* (CCLE, http://www.broadinstitute.org/ccle). Gene expression data normalization was conducted with the median centering and z-score normalization method described previously7. Using the ratio of two genes’ expression values (i.e., CDH1/VIM for determining epithelial–mesenchymal transition), we classified these lines into epithelial (n=13, CDH1/VIM > 1.2) and mesenchymal (n=9, CDH1/VIM <= 1.2) lines. Modulated genes (called Set 1) were identified by conducting t-test (P < 0.05) comparison of the two groups of cell lines with 10 re-samplings (i.e., each re-sampling randomly took 60% of the original samples). We also obtained the gene expression data of 442 lung tumor samples from the GEO Series (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68465). These samples have information about cancer recurrence and clinical follow-up. Using this set, we used a t-test (P < 0.05) to identify the modulated genes between the recurred and non-recurred samples. Next, we performed Kaplan-Meier survival tests on the modulated genes to identify survival-associated genes (called Set 2) using 500 times of re-samplings. We mapped all the genes from Sets 1 and 2 onto the global signaling network (i.e., we kept the network genes which are common between the genes in Sets 1 or 2 and their links in the network, and then removed other genes and their links from the network) to obtain a lung cancer specific metastasis network which contains 11,166 genes and 442,617 interactions.

**References**

1 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).

2 Zaman, N. *et al.* Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep* **5**, 216-223, doi:10.1016/j.celrep.2013.08.028 (2013).

3 Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P. & Marth, G. T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691-1692, doi:10.1093/bioinformatics/btr174 (2011).

4 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).

5 Bao, L., Pu, M. & Messer, K. AbsCN-seq: a statistical method to estimate tumor purity, ploidy and absolute copy numbers from next-generation sequencing data. *Bioinformatics* **30**, 1056-1063, doi:10.1093/bioinformatics/btt759 (2014).

6 Voduc, K. D. *et al.* Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* **28**, 1684-1691, doi:10.1200/JCO.2009.24.9284 (2010).

7 Li, J. *et al.* Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* **1**, 34, doi:10.1038/ncomms1033 (2010).

8 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34(D1)**, D535–D539, doi:10.1093/nar/gkj109 (2006).

9 Bader, G. D. *et al.* BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31(1)**, 248–250 (2003).

10 Bossi, A. *et al.*Tissue specificity and the human protein interaction network. *Mol Syst Biol.* **5**, 260, doi:10.1038/msb.2009.17 (2009).

11 Kotlyar, M. *et al.* Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res*. **44(D1)**, D536-41. doi: 10.1093/nar/gkv1115 (2016).

12 Kerrien, S. *et al*. IntAct--open source resource for molecular interaction data. *Nucleic Acids Res.* **35(Database issue)**, D561-5, doi:10.1093/nar/gkl958 (2007).

13 Schaefer, C. F. *et al*. PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37(Database issue)**, D674–D679, doi:10.1093/nar/gkn653 (2009).

14 Persico, M. *et al.* HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* **6**, S4-S21, doi:10.1186/1471-2105-6-S4-S21 (2005).

15 Pagel, P. *et al*. The MIPS mammalian protein-protein interaction database.

 *Bioinformatics* **21(6)**, 832-4, doi:10.1093/bioinformatics/bti115 (2005).

16 Lee, I. *et al*. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res* **21(7)**, 1109-21, doi: 10.1101/gr.118992.110 (2011).