CLAIMS

1. A computer-implemented method, comprising:

using at least one computer hardware processor to perform:

obtaining expression data obtained at least in part by sequencing a biological sample of a subject having, suspected of having or at risk of having cancer, the expression data comprising expression levels for a plurality of genes, the plurality of genes comprising a set of genes;

ranking two or more genes in the set of genes, based on their expression levels in the expression data to obtain a gene ranking, wherein the gene ranking includes values identifying the relative ranks for genes in the gene ranking; and

determining, using the gene ranking and a statistical model trained using training data indicating a plurality of gene rankings of at least some of the genes in the set of genes obtained, at least one characteristic of the biological sample, wherein each of the plurality of gene rankings is obtained based on respective expression levels for the at least some genes in the set of genes, and wherein determining the at least one characteristic comprises providing the gene ranking as input to the statistical model and obtaining an output indicating the at least one characteristic.

2. The method of claim 1, wherein the at least one characteristic of the biological sample is a physiological characteristic of cells in the biological sample or a tissue from which the cells originate.

3. The method of claim 1 or any other preceding claim, wherein the at least one characteristic is selected from cancer grade for cells in the biological sample, tissue of origin for cells in the biological sample, tissue type for cells in the biological sample, and cancer subtype for cells in the biological sample.

4. The method of claim 1 or any other preceding claim, wherein the at least one characteristic includes cancer grade for cells in the biological sample.

5. The method of claim 1 or any other preceding claim, wherein the at least one characteristic includes tissue of origin for cells in the biological sample.

6. The method of claim 1 or any other preceding claim, wherein the set of genes is selected from the following group of genes: UBE2C, MYBL2, PRAME, LMNB1, CXCL9, KPNA2, TPX2, PLCH1, CCL18, CDK1, MELK, CCNB2, RRM2, CCNB1, NUSAP1, SLC7A5, TYMS, GZMK, SQLE, C1orf106, CDC25B, ATAD2, QPRT, CCNA2, NEK2, IDO1, NDC80, ZWINT, ABCA12, TOP2A, TDO2, S100A8, LAMP3, MMP1, GZMB, BIRC5, TRIP13, RACGAP1, ASPM, ESRP1, MAD2L1, CENPF, CDC20, MCM4, MKI67, PBK, CKS2, KIF2C, MRPL13, TTK, BUB1, TK1, FOXM1, CEP55, EZH2, ECT2, PRC1, CENPU, CCNE2, AURKA, HMGB3, APOBEC3B, LAGE3, CDKN3, DTL, ATP6V1C1, KIAA0101, CD2, KIF11, KIF20A, CDCA8, NCAPG, CENPN, MTFR1, MCM2, DSCC1, WDR19, SEMA3G, KCND3, SETBP1, KIF13B, NR4A2, NAV3, PDZRN3, MAGI2, CACNA1D, STC2, CHAD, PDGFD, ARMCX2, FRY, AGTR1, MARCH8, ANG, ABAT, THBD, RAI2, HSPA2, ERBB4, ECHDC2, FST, EPHX2, FOSB, STARD13, ID4, FAM129A, FCGBP, LAMA2, FGFR2, PTGER3, NME5, LRRC17, OSBPL1A, ADRA2A, LRP2, C1orf115, COL4A5, DIXDC1, KIAA1324, HPN, KLF4, SCUBE2, FMO5, SORBS2, CARD10, CITED2, MUC1, BCL2, RGS5, CYBRD1, OMD, IGFBP4, LAMB2, DUSP4, PDLIM5, IRS2 and CX3CR1.

7. The method of claim 1, wherein the set of genes comprises at least 10 genes selected from the following group of genes: UBE2C, MYBL2, PRAME, LMNB1, CXCL9, KPNA2, TPX2, PLCH1, CCL18, CDK1, MELK, CCNB2, RRM2, CCNB1, NUSAP1, SLC7A5, TYMS, GZMK, SQLE, C1orf106, CDC25B, ATAD2, QPRT, CCNA2, NEK2, IDO1, NDC80, ZWINT, ABCA12, TOP2A, TDO2, S100A8, LAMP3, MMP1, GZMB, BIRC5, TRIP13, RACGAP1, ASPM, ESRP1, MAD2L1, CENPF, CDC20, MCM4, MKI67, PBK, CKS2, KIF2C, MRPL13, TTK, BUB1, TK1, FOXM1, CEP55, EZH2, ECT2, PRC1, CENPU, CCNE2, AURKA, HMGB3, APOBEC3B, LAGE3, CDKN3, DTL, ATP6V1C1, KIAA0101, CD2, KIF11, KIF20A, CDCA8, NCAPG, CENPN, MTFR1, MCM2, DSCC1, WDR19, SEMA3G, KCND3, SETBP1, KIF13B, NR4A2, NAV3, PDZRN3, MAGI2, CACNA1D, STC2, CHAD, PDGFD, ARMCX2, FRY, AGTR1, MARCH8, ANG, ABAT, THBD, RAI2, HSPA2, ERBB4, ECHDC2, FST, EPHX2, FOSB, STARD13, ID4, FAM129A, FCGBP, LAMA2, FGFR2, PTGER3, NME5, LRRC17, OSBPL1A, ADRA2A, LRP2, C1orf115, COL4A5, DIXDC1, KIAA1324, HPN, KLF4, SCUBE2, FMO5, SORBS2, CARD10, CITED2, MUC1, BCL2, RGS5, CYBRD1, OMD, IGFBP4, LAMB2, DUSP4, PDLIM5, IRS2 and CX3CR1.

8. The method of claim 1 or any other preceding claim, wherein the set of genes is selected from the following group of genes: PLTP, C1S, LY96, TSKU, TPST2, SERPINF1, SRPX2, SAA1, CTHRC1, GFPT2, CKAP4, SERPINA3, CFH, PLAU, BASP1, PTTG1, MOCOS, LEF1, SLPI, PRAME, STEAP3, LGALS2, CD44, FLNC, UBE2C, CTSK, SULF2, TMEM45A, FCGR1A, PLOD2, C19orf80, PDGFRL, IGF2BP3, SLC7A5, PRRX1, RARRES1, LHFPL2, KDELR3, TRIB3, IL20RB, FBLN1, KMO, C1R, CYP1B1, KIF2A, PLAUR, CKS2, CDCP1, SFRP4, HAMP, MMP9, SLC3A1, NAT8, FRMD3, NPR3, NAT8B, BBOX1, SLC5A1, GBA3, EMCN, SLC47A1, AQP1, PCK1, UGT2A3, BHMT, FMO1, ACAA2, SLC5A8, SLC16A9, TSPAN18, SLC17A3, STK32B, MAP7, MYLIP, SLC22A12, LRP2, CD34, PODXL, ZBTB42, TEK, FBP1 and BCL2.

9. The method of claim 1, wherein the set of genes comprises at least 10 genes selected from the following group of genes: PLTP, C1S, LY96, TSKU, TPST2, SERPINF1, SRPX2, SAA1, CTHRC1, GFPT2, CKAP4, SERPINA3, CFH, PLAU, BASP1, PTTG1, MOCOS, LEF1, SLPI, PRAME, STEAP3, LGALS2, CD44, FLNC, UBE2C, CTSK, SULF2, TMEM45A, FCGR1A, PLOD2, C19orf80, PDGFRL, IGF2BP3, SLC7A5, PRRX1, RARRES1, LHFPL2, KDELR3, TRIB3, IL20RB, FBLN1, KMO, C1R, CYP1B1, KIF2A, PLAUR, CKS2, CDCP1, SFRP4, HAMP, MMP9, SLC3A1, NAT8, FRMD3, NPR3, NAT8B, BBOX1, SLC5A1, GBA3, EMCN, SLC47A1, AQP1, PCK1, UGT2A3, BHMT, FMO1, ACAA2, SLC5A8, SLC16A9, TSPAN18, SLC17A3, STK32B, MAP7, MYLIP, SLC22A12, LRP2, CD34, PODXL, ZBTB42, TEK, FBP1 and BCL2.

10. The method of claim 1, wherein the set of genes is selected from the following group of genes: ITPKB, MYBL1, LMO2, BATF, IRF4, LRMP, CCND2, SLA, SP140, PIM1, CSTB, BCL2, TCF4, P2RX5, SPINK2, VCL, PTPN1, REL, FUT8, RPL21, PRKCB1, CSNK1E, GPR18, IGHM, ACP1, SPIB, HLA-DQA1, KRT8, FAM3C, HLA-DMB, GOT2, PIM2 and PLEK.

11. The method of claim 1, wherein the set of genes comprises at least 10 genes selected from the following group of genes: ITPKB, MYBL1, LMO2, BATF, IRF4, LRMP, CCND2, SLA, SP140, PIM1, CSTB, BCL2, TCF4, P2RX5, SPINK2, VCL, PTPN1, REL, FUT8, RPL21, PRKCB1, CSNK1E, GPR18, IGHM, ACP1, SPIB, HLA-DQA1, KRT8, FAM3C, HLA-DMB, GOT2, PIM2 and PLEK.

12. The method of claim 1 or any other preceding claim, wherein the set of genes is selected from the following group of genes: APOBEC3B, ATAD2, BIRC5, CCL20, CCND1, CDC45, CDC7, CDK1, CDKN2A, CDKN2C, CDKN3, CENPF, CENPN, CXCL14, DCN, DHFR, DKK3, DLGAP5, EPCAM, FANCI, FEN1, GMNN, GPX3, ID4, IGLC1, IL18, IL1R2, KIF18B, KIF20A, KIF4A, KLK13, KLK7, KLK8, KNTC1, KRT19, LAMP3, LMNB1, MCM2, MCM4, MCM5, ME1, MELK, MKI67, MLF1, MMP12, MTHFD2, NDN, NEFH, NEK2, NUP155, NUP210, NUSAP1, PDGFD, PLAGL1, PLOD2, PPP1R3C, PRIM1, PRKDC, PSIP1, RAD51AP1, RASIP1, RFC5, RNASEH2A, RPA2, RPL39L, RSRC1, RYR1, SLC35G2, SMC2, SPARCL1, STMN1, SYCP2, SYNGR3, TIMELESS, TMPO, TPX2, TRIP13, TYMS, UCP2, UPF3B, USP1, ZSCAN18, APOBEC3B, ATAD2, BIRC5, CCL20, CCND1, CDC45, CDC7, CDK1, CDKN2A, CDKN2C, CDKN3, CENPF, CENPN, CXCL14, DCN, DHFR, DKK3, DLGAP5, EPCAM, FANCI, FEN1, GMNN, GPX3, ID4, IGLC1, IL18, IL1R2, KIF18B, KIF20A, KIF4A, KLK13, KLK7, KLK8, KNTC1, KRT19, LAMP3, LMNB1, MCM2, MCM4, MCM5, ME1, MELK, MKI67, MLF1, MMP12, MTHFD2, NDN, NEFH, NEK2, NUP155, NUP210, NUSAP1, PDGFD, PLAGL1 and PLOD2.

13. The method of claim 1, wherein the set of genes comprises at least 10 genes selected from the following group of genes: APOBEC3B, ATAD2, BIRC5, CCL20, CCND1, CDC45, CDC7, CDK1, CDKN2A, CDKN2C, CDKN3, CENPF, CENPN, CXCL14, DCN, DHFR, DKK3, DLGAP5, EPCAM, FANCI, FEN1, GMNN, GPX3, ID4, IGLC1, IL18, IL1R2, KIF18B, KIF20A, KIF4A, KLK13, KLK7, KLK8, KNTC1, KRT19, LAMP3, LMNB1, MCM2, MCM4, MCM5, ME1, MELK, MKI67, MLF1, MMP12, MTHFD2, NDN, NEFH, NEK2, NUP155, NUP210, NUSAP1, PDGFD, PLAGL1, PLOD2, PPP1R3C, PRIM1, PRKDC, PSIP1, RAD51AP1, RASIP1, RFC5, RNASEH2A, RPA2, RPL39L, RSRC1, RYR1, SLC35G2, SMC2, SPARCL1, STMN1, SYCP2, SYNGR3, TIMELESS, TMPO, TPX2, TRIP13, TYMS, UCP2, UPF3B, USP1, ZSCAN18, APOBEC3B, ATAD2, BIRC5, CCL20, CCND1, CDC45, CDC7, CDK1, CDKN2A, CDKN2C, CDKN3, CENPF, CENPN, CXCL14, DCN, DHFR, DKK3, DLGAP5, EPCAM, FANCI, FEN1, GMNN, GPX3, ID4, IGLC1, IL18, IL1R2, KIF18B, KIF20A, KIF4A, KLK13, KLK7, KLK8, KNTC1, KRT19, LAMP3, LMNB1, MCM2, MCM4, MCM5, ME1, MELK, MKI67, MLF1, MMP12, MTHFD2, NDN, NEFH, NEK2, NUP155, NUP210, NUSAP1, PDGFD, PLAGL1 and PLOD2.

14. The method of claim 1 or any other preceding claim, further comprising:

ranking two or more genes in a second set of genes based on their expression levels in the expression data to obtain a second gene ranking; and

determining, using the second gene ranking and a second statistical model trained using second training data indicating a plurality of rankings for at least some of the genes in the second set of genes, at least one second characteristic of the biological sample.

15. The method of claim 1 or any other preceding claim, wherein determining the gene ranking comprises determining a relative rank for each gene in the set of genes based on the expression levels.

16. The method of claim 1 or any other preceding claim, wherein the statistical model comprises a gradient boosted decision tree classifier.

17. The method of claim 1 or any other preceding claim, wherein the statistical model comprises a classifier selected from the group consisting of: a gradient boosted decision tree classifier, a decision tree classifier, a gradient boosted classifier, a random forest classifier, a clustering-based classifier, a Bayesian classifier, a Bayesian network classifier, a neural network classifier, a kernel-based classifier, and a support vector machine classifier.

18. The method of claim 1, further comprising presenting, to a user, an indication of the at least one characteristic.

19. A system comprising:

at least one computer hardware processor; and

at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one hardware processor, cause the at least one hardware processor to perform the method of any one of claims 1-18.

20. At least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by at least one computer hardware processor, cause the at least one computer hardware processor to perform the method of any one of claims 1-18.