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 at least some genes in the set of genes, based on their expression levels in the expression data to obtain a 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.

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, wherein the method further comprises performing sequencing of the biological sample using a gene expression microarray prior to obtaining the expression data.~~

~~5. The method of claim 1 or any other preceding claim, wherein the method further comprises performing next generation sequencing of the biological sample prior to obtaining the expression data.~~

46. 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.

57. 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.

8. The method of claim 1 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having breast cancer.

69. The method of claim 18 or any other preceding claim, wherein the set of genes is selected from the following group of genes listed in Table 1: 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.

10. The method of claim 8, wherein the set of genes comprises at least 3 genes selected from the group of genes listed in Table 1.

11. The method of claim 8, wherein the set of genes comprises at least 5 genes selected from the group of genes listed in Table 1.

712. The method of claim 18, wherein the set of genes comprises at least 10 genes selected from the following group of genes listed in Table 1: 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.

~~13. The method of claim 8, wherein the set of genes comprises at least 20 genes selected from the group of genes listed in Table 1.~~

~~14. The method of claim 1, wherein the subject has, is suspected of having, or is at risk of having kidney cancer.~~

~~15. The method of claim 1 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having clear cell kidney cancer.~~

8~~16~~. The method of claim 1~~15~~ or any other preceding claim, wherein the set of genes is selected from the following group of genes ~~listed in: Table 2~~ 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.

~~17. The method of claim 15, wherein the set of genes comprises at least 3 genes selected from the group of genes listed in Table 2.~~

~~18. The method of claim 15, wherein the set of genes comprises at least 5 genes selected from the group of genes listed in Table 2.~~

9~~19~~. The method of claim 1~~15~~, wherein the set of genes comprises at least 10 genes selected from the following group of genes ~~listed in: Table 2~~ 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.

~~20. The method of claim 15, wherein the set of genes comprises at least 20 genes selected from the group of genes listed in Table 2.~~

~~21. The method of claim 1 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having lymphoma.~~

10~~22~~. The method of claim 1~~21 or any other preceding claim~~, wherein the set of genes is selected from the following group of genes ~~listed in: Table 3~~ 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.

~~23. The method of claim 21, wherein the set of genes comprises at least 3 genes selected from the group of genes listed in Table 3.~~

~~24. The method of claim 21, wherein the set of genes comprises at least 5 genes selected from the group of genes listed in Table 3.~~

11~~25~~. The method of claim 1~~21~~, wherein the set of genes comprises at least 10 genes selected from the following group of genes ~~listed in: Table 3~~ 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.

~~26. The method of claim 21, wherein the set of genes comprises at least 20 genes selected from the group of genes listed in Table 3.~~

~~27. The method of claim 1 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having head and neck squamous cell carcinoma.~~

~~28. The method of claim 27 or any other preceding claim, wherein the set of genes is selected from the group of genes listed in Table 6.~~

~~29. The method of claim 27, wherein the set of genes comprises at least 10 genes selected from the group of genes listed in Table 6.~~

~~30. The method of claim 1 or any other preceding claim, wherein the at least one characteristic includes human papillomavirus status for cells in a biological sample.~~

12~~31~~. The method of claim 1~~30~~ or any other preceding claim, wherein the set of genes is selected from the following group of genes ~~listed in: Table 8~~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.

1332. The method of claim 130, wherein the set of genes comprises at least 10 genes selected from the following group of genes ~~listed in: Table 8~~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.

1433. The method of claim 1 or any other preceding claim, further comprising:

ranking at least some 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 the at least some of the genes in the second set of genes, at least one second characteristic of the biological sample.

~~34. The method of claim 33 or any other preceding claim, wherein the at least one second characteristic includes cancer grade for cells in the biological sample.~~

~~35. The method of claim 33 or any other preceding claim, wherein the at least one second characteristic includes tissue of origin for cells in the biological sample.~~

15~~36~~. 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~~37~~. The method of claim 1 or any other preceding claim, wherein determining the at least one characteristic further comprises providing the gene ranking as input to the statistical model and obtaining an output indicating the at least one characteristic.

17~~38~~. The method of claim 1 or any other preceding claim, wherein the statistical model comprises a gradient boosted decision tree classifier.

18~~39~~. 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.

~~40. The method of claim 1 or any other preceding claim, wherein the set of genes includes at least 5 genes.~~

~~41. The method of claim 1 or any other preceding claim, wherein the set of genes consists of 5-50 genes.~~

~~42. The method of claim 1 or any other preceding claim, wherein the set of genes consists of 5-300 genes.~~

19~~43~~. The method of claim 1, further comprising presenting, to a user, an indication of the at least one characteristic.

~~44. The method of claim 43, wherein presenting the indication of the at least one characteristic further comprises displaying the at least one characteristic to the user in a graphical user interface (GUI).~~

2045. A system comprising:

the 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 1-1944.

2146. At least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one computer hardware processor, cause the at least one computer hardware processor to perform the method of any one of claims 1-1944.

2247. A method, comprising:

using at least one computer hardware processor to perform:

obtaining expression data for cells in a biological sample of a subject having, suspected of having, or at risk of having cancer;

ranking at least some genes in at least one set of genes based on their expression levels in the expression data to obtain at least one gene ranking; and

determining, using the at least one gene ranking and at least one statistical model trained using training data indicating a plurality of rankings for at least some genes in the at least one set of genes, tissue of origin for at least some of the cells in 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 at least one set of genes.

~~48. The method of claim 47, wherein the expression data was obtained using a gene expression microarray.~~

~~49. The method of claim 47 or any other preceding claim, wherein the expression data was obtained by performing next generation sequencing.~~

~~50. The method of claim 47 or any other preceding claim, wherein the tissue of origin is selected from the group consisting of lung tissue, pancreas tissue, stomach tissue, colon tissue, liver tissue, bladder tissue, kidney tissue, thyroid tissue, lymph node tissue, adrenal gland tissue, skin tissue, breast tissue, ovary tissue, prostate tissue, urothelial tissue, cervical tissue, esophagus tissue, brain tissue, soft tissue, connective tissue, head tissue, and neck tissue.~~

~~51. The method of claim 47 or any other preceding claim, further comprising:~~

~~determining, using the at least one gene ranking and the at least one statistical model, tissue type for at least some of the cells in the biological sample.~~

~~52. The method of claim 51 or any other preceding claim, wherein the tissue type is selected from the group consisting of adenocarcinoma, squamous cell carcinoma, carcinoma, cystadenocarcinoma, sarcoma, and glioma.~~

~~53. The method of claim 51 or any other preceding claim, wherein a combination of the tissue of origin and the tissue type is selected from the group consisting of lung adenocarcinoma, lung squamous cell carcinoma, melanoma, breast carcinoma, colorectal adenocarcinoma, ovarian serous cystadenocarcinoma, phenochromocytoma, bladder urothelial carcinoma, cervical squamous cell carcinoma, glioblastoma multiforme, head squamous cell carcinoma, neck squamous cell carcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, pancreatic adenocarcinoma, paraganglioma, prostate adenocarcinoma, sarcoma, stomach adenocarcinoma, thyroid carcinoma, and uterine corpus endometrial carcinoma.~~

2354. The method of claim 2247 ~~or any other preceding claim~~, wherein the subject has, is suspected of having, or is at risk of having lymphoma.

2455. The method of claim 2354, wherein the subject has, is suspected of having, or is at risk of having Diffuse Large B-Cell Lymphoma (DLBCL).

2556. The method of claim 2455, wherein the tissue of origin is a cell of origin selected from the group consisting of germinal center B-cell (GCB) and activated B-cell (ABC).

2657. The method of any one of claims 23 to 25~~claim 54 or any other preceding claim~~, wherein a set of genes of the at least one set of genes is selected from the following group of genes~~ listed in~~: ~~Table 3~~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.

~~58. The method of claim 54, wherein a set of genes of the at least one set of genes comprises at least 3 genes selected from the group of genes listed in Table 3.~~

~~59. The method of claim 54, wherein a set of genes of the at least one set of genes comprises at least 5 genes selected from the group of genes listed in Table 3.~~

2760. The method of claim 2354, wherein a set of genes of the at least one set of genes comprises at least 10 genes selected from the following group of genes~~ listed in~~: ~~Table 3~~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.

~~61. The method of claim 47 or any other preceding claim, wherein a set of genes of the at least one set of genes includes at least 5 genes.~~

~~62. The method of claim 47 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 5-100 genes.~~

~~63. The method of claim 47 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 10-200 genes.~~

~~64. The method of claim 47 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 20-100 genes.~~

~~65. The method of claim 47 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 50-100 genes.~~

2866. The method of any one of claims 22 to 27 ~~claim 47 or any other preceding claim~~, wherein the expression data includes values, each representing an expression level for a gene in the at least one set of genes, and determining a gene ranking of the at least one gene ranking comprises determining a relative rank for each gene in one of the at least one set of genes based on the values.

2967. The method of any one of claims 22 to 28~~claim 47 or any other preceding claim~~, wherein determining the at least one characteristic further comprises using the at least one gene ranking as an input to the at least one statistical model and obtaining an output indicating the tissue of origin.

3068. The method of any one of claims 22 to 29~~claim 47 or any other preceding claim~~, wherein the at least one statistical model comprises a gradient boosted decision tree classifier.

~~69. The method of claim 47 or any other preceding claim, wherein the at least one statistical model comprises at least one 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.~~

~~70. The method of claim 47, wherein the at least one set of genes comprises a first set of genes associated with predicting a first type of tissue and a second set of genes associated with predicting a second type of tissue.~~

3171. 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 computer hardware processor, cause the at least one computer hardware processor to perform the method of any one of claims 22-3047-70.

3272. 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 22-3047-70.

3373. A method, comprising:

using at least one computer hardware processor to perform:

obtaining expression data for cells in a biological sample of a subject having, suspected of having, or at risk of having cancer;

ranking at least some genes in a set of genes based on their expression levels in the expression data to obtain a gene ranking; and

determining, using the gene ranking and a statistical model trained using training data indicating a plurality of rankings for at least some genes in the set of genes, cancer grade for at least some of the cells in 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.

~~74. The method of claim 73, wherein the expression data was obtained using a gene expression microarray.~~

~~75. The method of claim 73 or any other preceding claim, wherein the expression data was obtained by performing next generation sequencing.~~

3476. The method of claim 3373 ~~or any other preceding claim~~, wherein the cancer grade is selected from the group consisting of at least Grade 1, Grade 2, and Grade 3.

~~77. The method of claim 73 or any other preceding claim, wherein the cancer grade is selected from the group consisting of Grade 1, Grade 2, Grade 3, Grade 4, and Grade 5.~~

~~78. The method of claim 73 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having breast cancer.~~

~~79. The method of claim 78 or any other preceding claim, wherein the set of genes is selected from the group of genes listed in Table 1.~~

~~80. The method of claim 78, wherein the set of genes comprises at least 3 genes selected from the group of genes listed in Table 1.~~

~~81. The method of claim 78, wherein the set of genes comprises at least 5 genes selected from the group of genes listed in Table 1.~~

3582. The method of claim 3378, wherein the set of genes comprises at least 10 genes selected from the following group of genes ~~listed in: Table 1~~ 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.

~~83. The method of claim 73 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having kidney cancer.~~

~~84. The method of claim 73 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having clear cell kidney cancer.~~

~~85. The method of claim 84 or any other preceding claim, wherein the set of genes is selected from the group of genes listed in Table 2.~~

~~86. The method of claim 84, wherein the set of genes comprises at least 3 genes selected from the group of genes listed in Table 2.~~

~~87. The method of claim 84, wherein the set of genes comprises at least 5 genes selected from the group of genes listed in Table 2.~~

~~88. The method of claim 84, wherein the set of genes comprises at least 10 genes selected from the group of genes listed in Table 2.~~

~~89. The method of claim 73 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having lung adenocarcinoma.~~

3690. The method of any one of claims 33 to 35~~claim 89 or any other preceding claim~~, wherein the set of genes is selected from the following group of genes~~ listed in~~: ~~Table 6~~AADAC, ALDOB, ANXA10, ASPM, BTNL8, CEACAM8, CENPA, CHGB, CHRNA9, COL11A1, CRABP1, F11, GGTLC1, HJURP, IGF2BP3, IHH, KCNE2, KIF14, LRRC31, MYBL2, MYOZ1, PCSK2, PI15, SCTR, SHH, SLC22A3, SLC7A5, SPOCK1, TM4SF4, TRPM8 and YBX2.

3791. The method of claim 3389, wherein the set of genes comprises at least 10 genes selected from the following group of genes listed in: ~~Table 6~~AADAC, ALDOB, ANXA10, ASPM, BTNL8, CEACAM8, CENPA, CHGB, CHRNA9, COL11A1, CRABP1, F11, GGTLC1, HJURP, IGF2BP3, IHH, KCNE2, KIF14, LRRC31, MYBL2, MYOZ1, PCSK2, PI15, SCTR, SHH, SLC22A3, SLC7A5, SPOCK1, TM4SF4, TRPM8 and YBX2.

~~92. The method of claim 73 or any other preceding claim, wherein the set of genes includes at least 50 genes.~~

~~93. The method of claim 73 or any other preceding claim, wherein the set of genes consists of 10-100 genes.~~

~~94. The method of claim 73 or any other preceding claim, wherein the set of genes consists of 10-30 genes.~~

~~95. The method of claim 73 or any other preceding claim, wherein the expression data includes values, each representing an expression level for a gene in the set of genes, and~~

~~determining the gene ranking comprises determining a relative rank for each gene in the set of genes based on the values.~~

3896. The method of claims 33 to 37~~claim 73 or any other preceding claim~~, wherein determining that at least one characteristic further comprises using the gene ranking as an input to the statistical model and obtaining an output indicating the cancer grade.

3997. The method of claims 33 to 38~~claim 73 or any other preceding claim~~, wherein the statistical model comprises a gradient boosted decision tree classifier.

~~98. The method of claim 73 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.~~

4099. 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 computer hardware processor, cause the at least one computer hardware processor to perform the method of any one of claims 33-39~~73-98~~.

41~~100~~. 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 33-39~~73-98~~.

42~~101~~. A method, comprising:

using at least one computer hardware processor to perform:

obtaining expression data for cells in a biological sample of a subject having, suspected of having, or at risk of having cancer;

ranking at least some genes in at least one set of genes based on their expression levels in the expression data to obtain at least one gene ranking; and

determining, using the at least one gene ranking and at least one statistical model, a subtype of peripheral T-cell lymphoma (PTCL) for at least some of the cells in the biological sample.

~~102. The method of claim 101, wherein the at least one statistical model was trained using training data indicating a plurality of rankings of expression levels for at least some genes in the at least one set of genes.~~

~~103. The method of claim 101 or any other preceding claim, wherein each of the plurality of gene rankings is obtained based on respective expression levels for the at least some genes in the at least one set of genes.~~

~~104. The method of claim 101 or any other preceding claim, wherein the expression data was obtained using a gene expression microarray.~~

~~105. The method of claim 101 or any other preceding claim, wherein the expression data was obtained by performing next generation sequencing.~~

~~106. The method of claim 101 or any other preceding claim, wherein the expression data was obtained using a hybridization-based expression assay.~~

~~107. The method of claim 101 or any other preceding claim, wherein the subtype of PTCL is selected from the group consisting of: anaplastic large cell lymphoma (ALCL), angioimmunoblastic T-cell lymphoma (AITL), natural killer/T-cell lymphoma (NKTCL), and adult T-cell leukemia/lymphoma (ATLL).~~

~~108. The method of claim 101 or any other preceding claim, wherein the subtype of PTCL is selected from the group consisting of: Peripheral T-Cell Lymphoma, Not Otherwise Specified (PTCL-NOS), anaplastic large cell lymphoma (ALCL), angioimmunoblastic T-cell lymphoma (AITL), cutaneous T-cell lymphoma (CTCL), Natural killer/T-cell lymphoma (NKTCL), Sezary syndrome, adult T-cell leukemia/lymphoma (ATLL), enteropathy-type T-cell lymphoma, nasal NK/T-cell lymphoma, hepatosplenic gamma-delta T-cell lymphoma, T-cell lymphomas of Follicular T-cell (TFH) origin, T-cell lymphomas of the gastrointestinal tract, and cutaneous T-cell lymphomas.~~

43109. The method of claim 42101 or any other preceding claim, wherein a set of genes of the at least one set of genes is selected from the following group of genes listed in: Table 10EFNB2, ROBO1, S1PR3, ANK2, LPAR1, SNAP91, SOX8, RAMP3, TUBB2B, ARHGEF10, NOTCH1, ZBTB17, CCNE1, FGF18, MYCN, PTHLH, SMARCA2, WNK1, NKX2-1, CYP26A1, HPSE, CTLA4, PELI1, PRKCB, SPAST, ALS2, KIF3B, ZFYVE27, FGF18, FNTB, REL, DMRT1, SLC19A2, STK3, PERP, TNFRSF8, TMOD1, BATF3, CDC14B, WDFEY3, AGT, ALK, ANXA3, BTBD11, CCNA1, DNER, GAS1, HS6ST2, IL1RAP, PCOLCE2, PDE4DIP, SLC16A3, TIAM2, TUBB6, WNT7B, SMOX, TMEM158, NLRP7, ADRB2, GALNT2, HRASLS, CD244, FASLG, KIR2DL4, LOC100287534, KLRD1, SH2D1B, KLRC2, NCAM1, CXCR5, IL6, ICOS, CD40LG, CD84, IL21, BCL6, MAF, SH2D1A, IL4, PTPN1, PIM1, ENTPD1, IRF4, CCND2, IL16, ETV6, BLNK, SH3BP5, FUT8, CCR4, GATA3, IL5, IL10, IL13, MMEITPKB, MYBL1, LRMP, KIAA0870, LMO2, CR1, LTBR, PDPN, TNFRSF1A, FCER2, ICAM1, FCGR2B, IKZF2, CCR8, TNFRSF18, IKZF4, FOXP3, IL2, TBX21, IFNG, GZMH, GNLY, EOMES, NCR1, GZMB, NKG7, FGFBP2, KLRF1, CD160, KLRK1, CD226, NCR3, TNFRSF8, BATF3, TMOD1, TMEM158, MSC and POPDC3.

~~110. The method of claim 101, wherein a set of genes of the at least one set of genes comprises at least 3 genes selected from the group of genes listed in Table 10.~~

~~111. The method of claim 101, wherein a set of genes of the at least one set of genes comprises at least 5 genes selected from the group of genes listed in Table 10.~~

44112. The method of claim 42101, wherein a set of genes of the at least one set of genes comprises at least 10 genes selected from the following group of genes listed in: Table 10EFNB2, ROBO1, S1PR3, ANK2, LPAR1, SNAP91, SOX8, RAMP3, TUBB2B, ARHGEF10, NOTCH1, ZBTB17, CCNE1, FGF18, MYCN, PTHLH, SMARCA2, WNK1, NKX2-1, CYP26A1, HPSE, CTLA4, PELI1, PRKCB, SPAST, ALS2, KIF3B, ZFYVE27, FGF18, FNTB, REL, DMRT1, SLC19A2, STK3, PERP, TNFRSF8, TMOD1, BATF3, CDC14B, WDFEY3, AGT, ALK, ANXA3, BTBD11, CCNA1, DNER, GAS1, HS6ST2, IL1RAP, PCOLCE2, PDE4DIP, SLC16A3, TIAM2, TUBB6, WNT7B, SMOX, TMEM158, NLRP7, ADRB2, GALNT2, HRASLS, CD244, FASLG, KIR2DL4, LOC100287534, KLRD1, SH2D1B, KLRC2, NCAM1, CXCR5, IL6, ICOS, CD40LG, CD84, IL21, BCL6,

MAF, SH2D1A, IL4, PTPN1, PIM1, ENTPD1, IRF4, CCND2, IL16, ETV6, BLNK, SH3BP5, FUT8, CCR4, GATA3, IL5, IL10, IL13, MMEITPKB, MYBL1, LRMP, KIAA0870, LMO2, CR1, LTBR, PDPN, TNFRSF1A, FCER2, ICAM1, FCGR2B, IKZF2, CCR8, TNFRSF18, IKZF4, FOXP3, IL2, TBX21, IFNG, GZMH, GNLY, EOMES, NCR1, GZMB, NKG7, FGFBP2, KLRF1, CD160, KLRK1, CD226, NCR3, TNFRSF8, BATF3, TMOD1, TMEM158, MSC and POPDC3.

~~113. The method of claim 101, wherein a set of genes of the at least one set of genes comprises at least 50 genes selected from the group of genes listed in Table 10.~~

~~114. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes includes at least one up-regulated in AITL gene.~~

~~115. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes includes at least one down-regulated in AITL gene.~~

~~116. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes includes at least one MF profile gene.~~

~~117. The method of claim 101 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having lymphoma.~~

~~118. The method of claim 117, wherein the subject has, is suspected of having, or is at risk of having peripheral T-cell lymphoma (PTCL).~~

~~119. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes includes at least 5 genes.~~

~~120. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 5-100 genes.~~

~~121. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 10-200 genes.~~

~~122. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 20-100 genes.~~

~~123. The method of claim 101 or any other preceding claim, wherein a set of genes of the at least one set of genes consists of 50-100 genes.~~

~~124. The method of claim 101 or any other preceding claim, wherein the expression data includes values, each representing an expression level for a gene in the at least one set of genes, and determining a gene ranking of the at least one gene ranking comprises determining a relative rank for each gene in one of the at least one set of genes based on the values.~~

~~125. The method of claim 101 or any other preceding claim, wherein determining the subtype of PTCL further comprises using the at least one gene ranking as an input to the at least one statistical model and obtaining an output indicating the subtype of PTCL.~~

~~126. The method of claim 101 or any other preceding claim, wherein the at least one statistical model comprises a gradient boosted decision tree classifier.~~

~~127. The method of claim 101 or any other preceding claim, wherein the at least one statistical model comprises at least one 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.~~

~~128. The method of claim 101 or any other preceding claim, wherein the at least one statistical model includes a multi-class classifier.~~

~~129. The method of claim 128 or any other preceding claim, wherein the multi-class classifier has at least four outputs each corresponding to a different subtype of PTCL.~~

~~130. The method of claim 129, wherein the at least four outputs include a first output corresponding to anaplastic large cell lymphoma (ALCL), a second output corresponding to angioimmunoblastic T-cell lymphoma (AITL), a third output corresponding to natural~~

~~killer/T-cell lymphoma (NKTCL), and a fourth output corresponding to adult T-cell leukemia/lymphoma (ATLL).~~

45~~131~~. The method of any one of claims 42 to 44~~claim 101 or any other preceding claim~~, wherein the at least one statistical model comprises a plurality of classifiers corresponding to different subtypes of PTCL.

46~~132~~. The method of claim 45~~131~~, wherein the plurality of classifiers includes a first classifier, a second classifier, a third classifier, and a fourth classifier, wherein the first classifier corresponds anaplastic large cell lymphoma (ALCL), a second classifier corresponds to angioimmunoblastic T-cell lymphoma (AITL), a third classifier corresponds to natural killer/T-cell lymphoma (NKTCL), and a fourth classifier corresponds to adult T-cell leukemia/lymphoma (ATLL).

~~133. The method of claim 131, wherein the at least one set of genes includes a first set of genes associated with a first classifier of the plurality of classifiers and a second set of genes associated with a second classifier of the plurality of classifiers.~~

~~134. The method of claim 101 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having lymphoma.~~

~~135. The method of claim 101 or any other preceding claim, wherein the subject has, is suspected of having, or is at risk of having PTCL.~~

~~136. The method of claim 101, further comprising presenting, to a user, an indication of the subtype of PTCL.~~

~~137. The method of claim 136, wherein presenting the indication of the subtype of PTCL further comprises displaying the subtype of PTCL to the user in a graphical user interface (GUI).~~

47~~138~~. 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 computer hardware processor, cause the at least one computer hardware processor to perform the method of any one of claims 42-46~~101-137~~.

48~~139~~. 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 42-46~~101-137~~.