

Claims:

1. A method for detecting a prostate cancer in a subject in need thereof, the method comprising:

determining the expression level of at least 38 biomarkers from a test sample obtained from the subject by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the at least 38 biomarkers comprise *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TPRSS2-ERG T1E4*, *TPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, *XPC*, and a housekeeping gene, wherein the housekeeping gene is selected from the group consisting of *ALG9*, *SEPN*, *YWHAQ*, *VPS37A*, *PRRC2B*, *DOPEY2*, *NDUFB11*, *ND4*, *MRPL19*, *PSMC4*, *SF3A1*, *PUM1*, *ACTB*, *GAPD*, *GUSB*, *RPLP0*, *TFRC*, *MORF4L1*, *18S*, *PPIA*, *PGK1*, *RPL13A*, *B2M*, *YWHAZ*, *SDHA*, *HPRT1*, *TOX4* and *TPT1*, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

normalizing the expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TPRSS2-ERG T1E4*, *TPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, and *XPC* to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TPRSS2-ERG T1E4*, *TPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, and *XPC*;

inputting each normalized expression level into an algorithm to generate a score wherein the algorithm is a prediction model generated using one or more machine-learning algorithms, wherein the one or more machine-learning algorithms are selected from XGB, RF, glmnet, cforest, CART, treebag, knn, nnet, SVM-radial, SVM-linear, NB, and mlp;

comparing the score with a first predetermined cutoff value; and

identifying the presence of a prostate cancer in the subject when the score is equal to or greater than the first predetermined cutoff value or identifying the absence of a prostate cancer in the subject when the score is less than the first predetermined cutoff value, wherein the first predetermined cutoff value is at least 33% on a scale of 0-100%.

2. A method for determining whether a prostate cancer in a subject is stable or progressive, the method comprising:

determining the expression level of at least 38 biomarkers from a test sample obtained from the subject by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the at least 38 biomarkers comprise *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TMPRSS2-ERG T1E4*, *TMPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, *XPC*, and a housekeeping gene, wherein the housekeeping gene is selected from the group consisting of *ALG9*, *SEPN*, *YWHAQ*, *VPS37A*, *PRRC2B*, *DOPEY2*, *NDUFB11*, *ND4*, *MRPL19*, *PSMC4*, *SF3A1*, *PUM1*, *ACTB*, *GAPD*, *GUSB*, *RPLP0*, *TFRC*, *MORF4L1*, *18S*, *PPIA*, *PGK1*, *RPL13A*, *B2M*, *YWHAZ*, *SDHA*, *HPRT1*, *TOX4* and *TPT1*, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

normalizing the expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TMPRSS2-ERG T1E4*, *TMPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, and *XPC* to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYP7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIP1*, *STX12*, *TMPRSS2-ERG T1E4*, *TMPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, and *XPC*;

inputting each normalized expression level into an algorithm to generate a score, wherein the algorithm is a prediction model generated using one or more machine-learning algorithms, wherein the one or more machine-learning algorithms are selected from XGB, RF, glmnet, cforest, CART, treebag, knn, nnet, SVM-radial, SVM-linear, NB, and mlp;

comparing the score with a first predetermined cutoff value; and identifying that the prostate cancer is progressive when the score is equal to or greater than the first predetermined cutoff value or identifying that the prostate cancer is stable when the score is less than the first predetermined cutoff value, wherein the first predetermined cutoff value is at least 33% on a scale of 0-100%.

3. A method for determining the completeness of surgery in a subject having a prostate cancer, the method comprising:

determining the expression level of at least 38 biomarkers from a test sample obtained from the subject after the surgery by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the 38 biomarkers comprise *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYD7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIPI*, *STX12*, *TMPRSS2-ERG T1E4*, *TMPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, *XPC*, and a housekeeping gene, wherein the housekeeping gene is selected from the group consisting of *ALG9*, *SEPN*, *YWHAQ*, *VPS37A*, *PRRC2B*, *DOPEY2*, *NDUFB11*, *ND4*, *MRPL19*, *PSMC4*, *SF3A1*, *PUM1*, *ACTB*, *GAPD*, *GUSB*, *RPLP0*, *TFRC*, *MORF4L1*, *18S*, *PPIA*, *PGK1*, *RPL13A*, *B2M*, *YWHAZ*, *SDHA*, *HPRT1*, *TOX4* and *TPT1*, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

normalizing the expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*, *EDC4*, *FGFR2*, *FXYD7*, *FYCO1*, *HNRNPU*, *HPN*, *KRT15*, *KRT23*, *MAN2B2*, *MAX*, *MRPS25*, *NDUFS2*, *PPARGC1A*, *PPRC1*, *RAD23A*, *REPIN1*, *SDR39U1*, *SETBP1*, *SLC14A1*, *SLC18A2*, *SMC4*, *SPARC*, *SQLE*, *STRIPI*, *STX12*, *TMPRSS2-ERG T1E4*, *TMPRSS2-ERG T2E4*, *TRIM29*, *UNC45A*, and *XPC* to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP*, *ANO7*, *AR*, *AR-V7*, *C16orf89*, *CHTOP*, *COL1A1*,

EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A, and XPC;

inputting each normalized expression level into an algorithm to generate a score, wherein the algorithm is a prediction model generated using one or more machine-learning algorithms, wherein the one or more machine-learning algorithms are selected from XGB, RF, glmnet, cforest, CART, treebag, knn, nnet, SVM-radial, SVM-linear, NB, and mlp;

comparing the score with a first predetermined cutoff value; and

identifying that the prostate cancer is not completely removed when the score is equal to or greater than the first predetermined cutoff value or identifying that the prostate cancer is completely removed when the score is less than the first predetermined cutoff value, wherein the first predetermined cutoff value is at least 33% on a scale of 0-100%.

4. A method for differentiating benign prostate hyperplasia from a prostate cancer in a subject having an enlarged prostate, the method comprising:

determining the expression level of at least 38 biomarkers from a test sample obtained from the subject by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the at least 38 biomarkers comprise *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A, XPC*, and a housekeeping gene, wherein the housekeeping gene is selected from the group consisting of *ALG9, SEPN, YWHAQ, VPS37A, PRRC2B, DOPEY2, NDUFB11, ND4, MRPL19, PSMC4, SF3A1, PUM1, ACTB, GAPD, GUSB, RPLP0, TFRC, MORF4L1, 18S, PPIA, PGK1, RPL13A, B2M, YWHAZ, SDHA, HPRT1, TOX4 and TPT1*, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

normalizing the expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN,*

KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A, and XPC to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYP7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A, and XPC*;

inputting each normalized expression level into an algorithm to generate a score, wherein the algorithm is a prediction model generated using one or more machine-learning algorithms, wherein the one or more machine-learning algorithms are selected from XGB, RF, glmnet, cforest, CART, treebag, knn, nnet, SVM-radial, SVM-linear, NB, and mlp;

comparing the score with a first predetermined cutoff value; and

identifying the presence of a prostate cancer in the subject when the score is equal to or greater than the first predetermined cutoff value or identifying the presence of benign prostate hyperplasia in the subject when the score is less than the first predetermined cutoff value, wherein the first predetermined cutoff value is at least 33% on a scale of 0-100%.

5. A method for evaluating the response of a subject having a prostate cancer to a first therapy, wherein the therapy is selected from hormone therapy and chemotherapy, the method comprising:

(1) at a first time point:

(a) determining the expression level of at least 38 biomarkers from a test sample obtained from the subject by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the at least 38 biomarkers comprise *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYP7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4,*

TRIM29, UNC45A, XPC, and a housekeeping gene, wherein the housekeeping gene is selected from the group consisting of *ALG9, SEPN, YWHAQ, VPS37A, PRRC2B, DOPEY2, NDUFB11, ND4, MRPL19, PSMC4, SF3A1, PUM1, ACTB, GAPD, GUSB, RPLP0, TFRC, MORF4L1, 18S, PPIA, PGK1, RPL13A, B2M, YWHAZ, SDHA, HPRT1, TOX4 and TPT1*, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

(b) normalizing the expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A*, and *XPC* to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A*, and *XPC*;

(c) inputting each normalized expression level into an algorithm to generate a first score, wherein the algorithm is a prediction model generated using one or more machine-learning algorithms, wherein the one or more machine-learning algorithms are selected from XGB, RF, glmnet, cforest, CART, treebag, knn, nnet, SVM-radial, SVM-linear, NB, and mlp;

(2) at a second time point, wherein the second time point is after the first time point and after the administration of the first therapy to the subject:

(d) determining the expression level of the at least 38 biomarkers from a test sample obtained from the subject by contacting the test sample with a plurality of agents specific to detect the expression of the at least 38 biomarkers, wherein the test sample is blood, serum, plasma, or neoplastic tissue;

(e) normalizing the expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4*,

TRIM29, UNC45A, and XPC to the expression level of the housekeeping gene, thereby obtaining a normalized expression level of each of *AAMP, ANO7, AR, AR-V7, C16orf89, CHTOP, COL1A1, EDC4, FGFR2, FXYD7, FYCO1, HNRNPU, HPN, KRT15, KRT23, MAN2B2, MAX, MRPS25, NDUFS2, PPARGC1A, PPRC1, RAD23A, REPIN1, SDR39U1, SETBP1, SLC14A1, SLC18A2, SMC4, SPARC, SQLE, STRIP1, STX12, TMPRSS2-ERG T1E4, TMPRSS2-ERG T2E4, TRIM29, UNC45A, and XPC*;

(f) inputting each normalized expression level into the algorithm to generate a second score;

(3) comparing the first score with the second score; and

(4) identifying the subject as responsive to the first therapy when the second score is significantly decreased as compared to the first score or identifying the subject as not responsive to the first therapy when the second score is not significantly decreased as compared to the first score, wherein the second score is significantly decreased as compared to the first score when the second score is at least 25% less than the first score.

6. The method of any one of claims 1-5, wherein the housekeeping gene is *TOX4*.
7. The method of any one of claims 1-4, wherein the first predetermined cutoff value is at least 50% on a scale of 0 to 100%
8. The method of claim 1 or claim 4, wherein the predetermined cutoff value has a sensitivity of identifying the subject as having prostate cancer of at least 92%.
9. The method of claim 2, wherein the predetermined cutoff value has a sensitivity of identifying that the prostate cancer is progressive of at least 92%.
10. The method of claim 3, wherein the predetermined cutoff value has a sensitivity of identifying that the prostate cancer is not completely removed of at least 92%.

11. The method of claim 1 or claim 4, wherein the predetermined cutoff value has a specificity of identifying the subject as having prostate cancer of at least 95%.
12. The method of claim 2, wherein the predetermined cutoff value has a specificity of identifying that the prostate cancer is progressive of at least 95%.
13. The method of claim 3, wherein the predetermined cutoff value has a specificity of identifying that the prostate cancer is not completely removed of at least 95%.
14. The method any of claims 1-13, wherein at least one of the at least 38 biomarkers is RNA, cDNA, or protein.
15. The method of claim 14, wherein when the biomarker is RNA, the RNA is reverse transcribed to produce cDNA, and the produced cDNA expression level is detected.
16. The method of claim 14, wherein the expression level of the biomarker is detected by forming a complex between the biomarker and a labelled probe or primer.
17. The method of any of claims 1-4, wherein the first predetermined cutoff value is derived from a plurality of reference samples obtained from subjects not having or not diagnosed with a neoplastic disease.
18. The method of claim 17, wherein the neoplastic disease is prostate cancer.
19. The method of claim 17, wherein each reference sample is blood, serum, plasma, or non-neoplastic tissue.

20. The method of any one of claims 1-19, wherein the one or more machine-learning algorithms is XGB.

21. The method of claim 5, wherein the first time point is prior to the administration of the first therapy to the subject.