

CLAIMS

1. A computer-implemented method for determining whether or not a subject is likely to respond to a checkpoint blockade therapy, the method comprising:

accessing expression data for the subject;

using the expression data to determine expression levels, in the subject, for at least three genes; and

determining, using the determined expression levels and a statistical model trained using expression data indicating expression levels for a plurality of genes for a plurality of subjects, whether or not the subject is likely to respond to the checkpoint blockade therapy, wherein the checkpoint blockade therapy is a PD1 inhibitor and/or a CTLA4 inhibitor;

characterized in that the at least three genes are selected from the set of predictor genes consisting of BRAF, ACVR1B, MPRIP, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, SNX6, ACVR1B, MPRIP, COPS3, NLRX1, ELAC2, MON1B, ARF3, ARPIN, SPRYD3, FLI1, TIRAP, GSE1, POLR3K, PIGO, MFHAS1, NPIPA1, DPH6, ERLIN2, CES2, LHFP, NAIF1, ALCAM, SYNE1, SPINT1, SMTN, SLCA46A1, SAP25, WISP2, TSTD1, NLRX1, NPIPA1, HIST1H2AC, FUT8, FABP4, ERBB2, TUBA1A, XAGE1E, SERPINF1, RAI14, SIRPA, MT1X, NEK3, TGFB3, USP13, HLA-DRB4, IGF2, and MICAL1; and the statistical model includes variables representing ratios of gene expression levels for pairs of the predictor genes.

2. The method of claim 1, wherein the expression data:

(a) is used to determine expression levels for at least four genes, at least five genes, at least six genes, at least seven genes, at least eight genes, at least nine genes, or at least ten genes; and/or

(b) is RNA expression data, DNA expression data, or protein expression data.

3. The method of claim 1 or any other preceding claim, wherein the statistical model comprises a first set of dependent variables, each representing a ratio of a pair of genes, wherein the genes are selected from: BRAF, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, SNX6, ACVR1B, MPRIP, COPS3, NLRX1, ELAC2, MON1B, ARF3, ARPIN, SPRYD3, FLI1, TIRAP, GSE1, POLR3K, PIGO, MFHAS1, NPIPA1, DPH6, ERLIN2, CES2, LHFP, NAIF1, ALCAM, SYNE1, SPINT1, SMTN, SLCA46A1, SAP25, WISP2, TSTD1, NLRX1, NPIPA1, HIST1H2AC, FUT8, FABP4,

ERBB2, TUBA1A, XAGE1E, SERPINF1, RAI14, SIRPA, MT1X, NEK3, TGFB3, USP13, HLA-DRB4, IGF2, and MICAL1, preferably wherein the genes are selected from: BRAF, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, SNX6, XAGE1E, SERPINF1, RAI14, SIRPA, MT1X, NEK3, TGFB3, USP13, HLA-DRB4, IGF2, and MICAL1.

4. The method of claim 3 or any other preceding claim, wherein the statistical model comprises a first set of dependent variables, each representing a ratio of a pair of genes, wherein the genes:

(a) comprise: BRAF, RAI14, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, and SNX6; or

(b) consist of: BRAF, RAI14, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, and SNX6.

5. The method of claim 3 or any other preceding claim, wherein the first set of dependent variables comprises at least two ratios selected from:

(a) BRAF:RAI14, ACVR1B:MPRIIP, ACVR1B:COPS3, PRKAG1:STX2, NLRX1:ELAC2, MON1B:STX2, ARF3:MPRIIP, ARPIN:MPRIIP, SPRYD3:FLI1, TIRAP:MPRIIP, GSE1:RAI14, POLR3K:HAUS8, RAB40C:HAUS8, PIGO:MPRIIP, MFHAS1:USP13, GSE1:NPIPA1, DPH6:STX2, ERLIN2:RAI14, CES2:LHFP, and NAIF1:HAUS8; or

(b) MON1B:STX2, FAM234A:LIN37, DPH6:STX2, BRAF:RAI14, ADCK2:C14ORF80, POLR3K:HAUS8, URB1:TMEM181, GCLC:NEK3, RAB40C:HAUS8, NLRX1:ELAC2, CMIP:ROBO4, NXT2:FBXO5, EIF3H:NEK3, EHHADH:SNX6, DMTN:RASD1, SLC46A1:RBM8A, ACVR1B:GTF2H2, NPEPPS:HDAC2, CXCL16:BASP1, MFHAS1:DYRK3, ACVR1B:USP13, DPH6:C14ORF166, CES2:LHFP, ACVR1B:DCP1B, NAIF1:HDGFRP2, ABCC1:TRIO, GPR107:STX2, ZDHHC7:USP13, PRKAG1:TSEN2, PC:HAUS8, LRBA:CEP192, POM121C:CNPY4, KAT14:SETD5, SLC35A5:SNX6, ATP6V1A:GTDC1, TXNL4B:AKAP8L, SLC36A1:HAUS8, PSAP:SH3BP5, CMIP:OLFML2B, TATDN3:NEK3, TASP1:DDX5, SDC4:FUT8, TMEM254:COPS3, ARF3:MPRIIP, SLC46A1:PMF1-BGLAP, ATP6V1A:FBXO30, MFHAS1:C2CD5, ERAP1:SYNE1, F11R:FYN, RCHY1:RNF146, ATP6V1A:PDCD5, ACVR1B:ELAC2, CLN3:HAUS8, NAIF1:HAUS8, PRKAG1:SOCS4, HNRNPH2:USP13, TPD52:MTRF1L,

ACVR1B:C14ORF80, IST1:NPIPA1, DPH6:TCEAL1, CSNK2A1:MRGBP,
 CXCL16:FILIP1L, AGK:USP13, MYO18A:FYN, SIRPA:FLII, C16ORF58:FLII,
 TRIM11:AKAP8L, MFHAS1:DMPK, JMJD8:AKAP8L, DIAPH1:SYNE1,
 BCKDHA:HAUS8, TMEM254:PSMC5, ACVR1B:HTRA2, MON1B:SYNE1,
 DCAKD:PMF1-BGLAP, VWA5A:RASD1, TPD52:TRA2A, ZMIZ1:STX2, NUB1:C2CD5,
 GSE1:RAI14, AGFG1:STX2, NXT2:TRA2A, ACSS1:NPIPA5, FBXW8:USP13,
 CMIP:TRIO, AGPAT3:FYN, PSMF1:PTOV1, CREG1:TARS, SLC46A1:CEP131,
 SIRPA:SERPINF1, DNAJA2:HDAC2, ERLIN2:RAI14, FAM234A:ZNF428,
 CHMP1A:LIN37, FAM110A:TCF7, ACVR1B:COPS3, GSE1:DDX11, CREG1:ARFGAP3,
 BRPF3:USP13, MFHAS1:USP13, LAMP1:MAPK7, ACSS1:PMF1-BGLAP, SUFU:TRIO,
 ARF3:DAD1, NLRX1:TRA2A, NLRX1:SLC39A13, CMIP:SH3BP5, PPIF:HAUS8,
 ANKRD13A:SOCS4, F8:SYNE1, ATP6V1A:USP48, ACVR1B:MPRIP,
 TMEM141:HAUS8, TIRAP:MPRIP, ZDHHC12:HAUS8, SLC46A1:MED9,
 MFHAS1:NCAPD2, ERBB2:CENPL, JMJD7:PQBP1, PHKG2:AKAP8L,
 SLC36A1:SLC26A6, ATP6V1A:SPDL1, DCTN5:CEP89, IPPK:STX2, LAMB3:ADM,
 ARPIN:MPRIP, SLC46A1:FYN, ACVR1B:LTV1, GDE1:ZNF576, DMTN:GFPT2,
 LCMT2:RDH11, ACVR1B:CCDC66, ACVR1B:NEK3, SEC24B:MAP3K7,
 ZNF764:AKAP8L, CHMP1A:HAUS8, PIGO:USP13, ARF3:MAPK7, GSE1:CCDC66,
 ACVR1B:METTL17, C20ORF196:HAUS8, ARF3:CBY1, BRPF3:MPRIP,
 SLC46A1:PSMC5, CMTR2:MAP3K7, TASP1:CCDC66, BRPF3:PIP5K1A, PIGO:FLII,
 MYO18A:ACIN1, PSMF1:AKAP8L, FBXW11:LTV1, CXCL16:ADAMTS2,
 SPRYD3:FLII, DPH6:FRA10AC1, PDXK:HAUS8, ACVR1B:WDR45B,
 MON1B:SMIM10L1, LAT:PRRX1, WDR24:AKAP8L, EHHADH:CCDC174,
 ACVR1B:CEP89, MFHAS1:ODF2, ALDH6A1:GOPC, GSE1:NBPF14, EHHADH:MEX3C,
 BRPF3:STX2, EHHADH:GTF2H2, PIGO:MPRIP, PRKAG1:STX2, EHHADH:SMIM10L1,
 SLC36A1:SH3PXD2A, ADCK2:STX2, SORD:PDCD5, ACVR1B:LLGL1,
 LAMTOR3:ZNF644, PIM1:MICAL2, CREG1:NPIPA1, GSE1:NPIPA1, PRR13:HAUS8,
 WDR55:RAD1, CMIP:NID2, DIAPH1:TARS, SNAP23:SNX6, GSE1:TSPYL2,
 C2ORF68:NPIPA1, MFHAS1:TRIO, DPH6:CGRRF1, KIF13B:STX2, PTK2B:TCF7,
 ATP2A2:STX2, ANKS1A:USP13, JRK:NEK3, LRBA:DDX5, IDH2:HAUS8,
 CCNF:HAUS8, CMIP:CHN1, STAU2:STX2, ACSS1:LHFP, GSR:STX2, IGF2R:FYN,
 CXCL16:ACVRL1.

6. The method of claim 3 or any other preceding claim, wherein the first set of dependent variables:
- (a) consists of the ratios: BRAF:RAI14, PRKAG1:STX2, AGPAT3:FYN, CMIP:ROBO4, RAB40C:HAUS8, SNAP23:SNX6; or
 - (b) comprises at least three ratios, at least four ratios, at least five ratios, or at least six ratios.
7. The method of claim 1, further comprising using the expression data to determine expression levels, in the subject, for at least eight genes from the group consisting of BRAF, ACVR1B, MPRIP, PRKAG1, STX2, AGPAT3, FYN, CMIP, ROBO4, RAB40C, HAUS8, SNAP23, SNX6, ACVR1B, MPRIP, COPS3, NLRX1, ELAC2, MON1B, ARF3, ARPIN, SPRYD3, FLI1, TIRAP, GSE1, POLR3K, PIGO, MFHAS1, NPIPA1, DPH6, ERLIN2, CES2, LHFP, NAIF1, ALCAM, SYNE1, SPINT1, SMTN, SLCA46A1, SAP25, WISP2, TSTD1, NLRX1, NPIPA1, HIST1H2AC, FUT8, FABP4, ERBB2, TUBA1A, XAGE1E, SERPINF1, RAI14, SIRPA, MT1X, NEK3, TGFB3, USP13, HLA-DRB4, IGF2, and MICAL1.
8. The method of claim 1 or any other preceding claim, wherein the statistical model comprises a generalized linear model and/or a logistic regression model.
9. The method of claim 1 or any other preceding claim, further comprising training the statistical model, wherein training the statistical model comprises training a generalized linear model having a plurality of regression variables, each of the plurality of regression variables representing a ratio of a pair of genes for respective pairs of the predictor set of genes, and/or wherein training the statistical model comprises training a logistic regression model having a plurality of regression variables, each of the plurality of regression variables representing a ratio of a pair of genes for respective pairs of the predictor set of genes.
10. The method of claim 9, wherein the logistic regression model comprises a respective plurality of weights for the plurality of regression variables, wherein estimating the parameters of the statistical model comprises:
- estimating the plurality of weights using the expression data for the plurality of subjects and information indicating which of the plurality of subjects responded to the

checkpoint blockade therapy and/or which of the plurality of subjects did not respond to the checkpoint blockade therapy.

11. The method of claim 9 or any other preceding claim, wherein training the statistical model comprises iteratively adding regression variables for respective genes to the statistical model, at least in part by:

- identifying a candidate gene in the subset of genes;
- augmenting a current statistical model with a regression variable for the candidate gene to obtain an augmented statistical model;
- evaluating performance of the augmented statistical model; and
- determining to add the regression variable for the candidate gene to the current statistical model based on results of evaluating the performance.

12. The method of claim 1 or any other preceding claim, wherein the subject has melanoma.

13. The method of claim 1 or any other preceding claim, wherein the method further comprises providing, to a user, an indication of whether or not the subject is likely to respond to the checkpoint blockade therapy.

14. A system for determining whether or not a subject is likely to respond to a checkpoint blockade therapy, the 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, causes the at least one computer hardware processor to perform the method of any of claims 1 to 13.

15. At least one non-transitory computer readable storage medium storing processor-executable instructions that, when executed at least one computer hardware processor, causes the at least one computer hardware processor to perform a method for determining whether or not a subject is likely to respond to a checkpoint blockade therapy according to any of claims 1 to 13.