Supplementary methods

Read alignment and quantification

FASTQ files were aligned with *STAR* (version 2.6.1d) to the human genome (version GRCh37) to generate bam files[1]. Gene expression quantification was performed with *Salmon* (version 0.12)[2]. Raw genewise counts to be used in downstream analyses were extracted as length scaled TPM values from *Salmon* outputs using the *Tximport* R package (version 1.12.0)[3].

Gene expression analysis

DGE analysis was performed with the *Limma* R package (version 3.40.2)[4]. GSEA was performed with *CAMERA*[5] using the hallmark and gene ontology (GO) gene sets from the Molecular Signatures Database (MSigDB)[6].

Confirmation of tumor viral status

Tumor viral status was confirmed by running *Xenomapper* against the MCPyV (NCBI Reference Sequence: NC_010277.2) and human genome (GRCh37).

Variant calling from RNA-Seq

Variants were called using a modified version of the GATK best practices for variant calling from RNA-Seq (https://gatkforums.broadinstitute.org/gatk/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail), with the exception that GATK version 4 rather than 3 was used (GATK version 4.1.0.0)[7 8]. Variants were also called using *VarScan* (version 2.4.3)[9] with a minimum of 8 bases of coverage and 3 bases supporting each variant call. To be considered a true call, the variant had to be called by both the GATK

HaploypeCaller and *VarScan*. To filter out germline variants, all variants present in the dbSNP (version 151) were removed and additionally any variant present in more than 20% of the cohort was removed. To remove mutations caused by RNA-editing, any sites present in the RADAR RNA-editing database[10] were also removed. The method was originally developed using 167 lung squamous cell carcinoma (LUSC) RNA-Seq samples from the TCGA with matched whole exome sequencing (WES) data. The method was able to call variants with only 6% recall, but 74% precision. The TMB from RNA-Seq variant calling was also validated against an external cohort made up of 11 WGS samples comprised of 9 cancer of unknown primary (CUP) and 2 MCC samples with a Pearson's correlation of 0.98. Despite the low recall rate, UV mutational signatures were still able to be identified and were dominant in the viral negative samples.

Calculation of TMB

TMB was calculated for a given sample by counting the total number of exon bases that had a coverage of 8 or more using *Mosdepth* (version 0.2.5)[11], dividing this by the total number of detected coding mutations and multiplying by 1 million (mutations/megabase).

Mutational signature generation

Mutation signatures were generated for each sample in the cohort using *MutationalPatterns*[12]. The generated signatures were then matched to the single bases substitution (SBS) signatures from COSMIC v3[13 14].

Table S1. Genes used in determining cell-type-specific gene expression. These gene sets were identified by Danaher et al[15] by using co-expression patterns from large tumor gene expression datasets to evaluate previously reported marker and gene lists.

Gene	Cell Type
BLK	B-cells
CD19	B-cells
MS4A1	B-cells
TNFRSF17	B-cells
FCRL2	B-cells
KIAA0125	B-cells
PNOC	B-cells
SPIB	B-cells
TCL1A	B-cells
PTPRC	CD45
CD8A	CD8 T cells
CD8B	CD8 T cells
CTSW	Cytotoxic cells
GNLY	Cytotoxic cells
GZMA	Cytotoxic cells
GZMB	Cytotoxic cells
GZMH	Cytotoxic cells
KLRB1	Cytotoxic cells
KLRD1	Cytotoxic cells

KLRK1	Cytotoxic cells
PRF1	Cytotoxic cells
NKG7	Cytotoxic cells
CCL13	DC
CD209	DC
HSD11B1	DC
CD244	Exhausted CD8
EOMES	Exhausted CD8
LAG3	Exhausted CD8
PTGER4	Exhausted CD8
CD163	Macrophages
CD68	Macrophages
CD84	Macrophages
MS4A4A	Macrophages
MS4A2	Mast cells
TPSAB1	Mast cells
CPA3	Mast cells
HDC	Mast cells
TPSB2	Mast cells
CSF3R	Neutrophils
S100A12	Neutrophils
CEACAM3	Neutrophils
FCAR	Neutrophils

FCGR3B	Neutrophils
FPR1	Neutrophils
SIGLEC5	Neutrophils
IL21R	NK CD56dim cells
KIR2DL3	NK CD56dim cells
KIR3DL1	NK CD56dim cells
KIR3DL2	NK CD56dim cells
NCR1	NK cells
XCL2	NK cells
XCL1	NK cells
CD3D	T-cells
CD3E	T-cells
CD3G	T-cells
CD6	T-cells
SH2D1A	T-cells
TRAT1	T-cells
TBX21	Th1 cells
FOXP3	Treg

Table S2. Hallmark gene sets. Hallmark gene sets called as significantly enriched in patients with an observable response to ICI (FDR < 0.05)**.**

Gene set	Number of Genes	Direction	*P* Value	False Discovery Rate
HALLMARK_E2F_TARGETS	199	Down	1.10E-06	5.51E-05
HALLMARK_MYC_TARGETS_V1	199	Down	4.32E-06	1.08E-04
HALLMARK_G2M_CHECKPOINT	198	Down	1.66E-04	0.00210671
HALLMARK_MTORC1_SIGNALING	198	Down	2.10E-04	0.00210671
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	110	Down	2.11E-04	0.00210671
HALLMARK_OXIDATIVE_PHOSPHORYLATION	200	Down	3.71E-04	0.00309176
HALLMARK_UV_RESPONSE_UP	155	Down	0.00373543	0.02668164

Table S3. GO gene sets. Gene ontology (GO) gene sets called as significantly enriched in patients with an observable response to ICI (FDR < 0.05)**.**

Gene set	Number of Genes	Direction	*P* Value	False Discovery Rate
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	103	Down	1.18E-19	6.99E-16
GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	122	Down	3.63E-18	1.07E-14
GO_CYTOSOLIC_RIBOSOME	107	Down	6.17E-18	1.22E-14

GO_TRANSLATIONAL_INITIATION	141	Down	3.48E-15	4.77E-12
GO_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT	58	Down	4.04E-15	4.77E-12
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	115	Down	1.47E-14	1.45E-11
GO_PROTEIN_TARGETING_TO_MEMBRANE	152	Down	8.95E-14	7.56E-11
GO_RIBOSOMAL_SUBUNIT	157	Down	7.54E-13	5.58E-10
GO_MULTI_ORGANISM_METABOLIC_PROCESS	135	Down	3.36E-12	2.21E-09
GO_CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT	41	Down	8.88E-11	5.25E-08
GO_LARGE_RIBOSOMAL_SUBUNIT	92	Down	1.29E-10	6.91E-08
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	198	Down	4.89E-09	2.41E-06
GO_RIBOSOME	216	Down	5.49E-09	2.50E-06
GO_IMMUNOGLOBULIN_COMPLEX	21	Down	1.35E-08	5.64E-06
GO_RNA_CATABOLIC_PROCESS	214	Down	1.43E-08	5.64E-06
GO_SMALL_RIBOSOMAL_SUBUNIT	65	Down	1.87E-08	6.92E-06
GO_CYTOSOLIC_PART	200	Down	1.77E-07	6.14E-05
GO_IMMUNOGLOBULIN_RECEPTOR_BINDING	22	Down	1.34E-06	4.41E-04
GO_KERATINIZATION	40	Up	1.73E-06	5.37E-04
GO_CORNIFIED_ENVELOPE	33	Up	3.17E-06	9.38E-04
GO_VIRAL_LIFE_CYCLE	278	Down	3.46E-06	9.68E-04
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_MEMBRANE	257	Down	3.60E-06	9.68E-04
GO_CHAPERONE_MEDIATED_PROTEIN_FOLDING	47	Down	4.13E-06	0.001028 65
GO_KERATIN_FILAMENT	27	Up	4.17E-06	0.001028

				65
GO_POSITIVE_REGULATION_OF_NUCLEASE_ACTIVITY	15	Down	8.69E-06	0.00205641
GO_BITTER_TASTE_RECEPTOR_ACTIVITY	20	Down	1.29E-05	0.00293593
GO_RRNA_METABOLIC_PROCESS	248	Down	1.96E-05	0.00428899
GO_PERK_MEDIATED_UNFOLDED_PROTEIN_RESPONSE	11	Down	2.57E-05	0.00542086
GO_DNA_PACKAGING_COMPLEX	57	Down	3.71E-05	0.00757515
GO_UNFOLDED_PROTEIN_BINDING	93	Down	4.23E-05	0.00834014
GO_PROTEIN_LOCALIZATION_TO_MEMBRANE	366	Down	5.14E-05	0.00952335
GO_DNA_PACKAGING	141	Down	5.15E-05	0.00952335
GO_DETECTION_OF_CHEMICAL_STIMULUS_INVOLVED_IN_SENSORY_PERCEPTION_OF_TASTE	31	Down	5.89E-05	0.0105592
GO_CYTOPLASMIC_TRANSLATION	40	Down	6.73E-05	0.01154811
GO_CELL_CYCLE_PHASE_TRANSITION	249	Down	6.83E-05	0.01154811
GO_KERATINOCYTE_DIFFERENTIATION	88	Up	8.41E-05	0.01316369

GO_INNER_MITOCHONDRIAL_MEMBRANE_PROTEIN_COMPLEX	99	Down	8.63E-05	0.013163 69
GO_PEPTIDE_CROSS_LINKING	44	Up	8.73E-05	0.013163 69
GO_PROTEIN_FOLDING_IN_ENDOPLASMIC_RETICULUM	8	Down	8.78E-05	0.013163 69
GO_ANTIGEN_BINDING	92	Down	8.90E-05	0.013163 69
GO_CENTROMERE_COMPLEX_ASSEMBLY	36	Down	9.32E-05	0.013344 39
GO_CHROMATIN_ASSEMBLY_OR_DISASSEMBLY	129	Down	9.48E-05	0.013344 39
GO_PROTEIN_DNA_COMPLEX_SUBUNIT_ORGANIZATION	181	Down	1.11E-04	0.015303 04
GO_PHAGOCYTOSIS_RECOGNITION	28	Down	1.17E-04	0.015707 39
GO_PROTEIN_FOLDING	208	Down	1.30E-04	0.017079 88
GO_PROTEIN_TARGETING	390	Down	1.47E-04	0.018889 63
GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN	52	Down	1.51E-04	0.018962 48
GO_CELL_CYCLE_G2_M_PHASE_TRANSITION	136	Down	1.68E-04	0.020646 39
GO_ORGANIC_CYCLIC_COMPOUND_CATABOLIC	396	Down	1.71E-04	0.020647

_PROCESS				68
GO_STRUCTURAL_CONSTITUENT_OF_MUSCLE	36	Up	1.90E-04	0.0223471
GO_PROTEIN_REFOLDING	19	Down	1.93E-04	0.0223471
GO_DNA_REPLICATION_INDEPENDENT_NUCLEOSOME_ORGANIZATION	41	Down	1.97E-04	0.0224258 2
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ORGANELLE	348	Down	2.03E-04	0.0226745 7
GO_RESPIRATORY_CHAIN	75	Down	2.19E-04	0.0236326 1
GO_OXIDATIVE_PHOSPHORYLATION	80	Down	2.22E-04	0.0236326 1
GO_HISTONE_EXCHANGE	40	Down	2.24E-04	0.0236326 1
GO_ATP_DEPENDENT_CHROMATIN_REMODELING	61	Down	2.52E-04	0.0261378 6
GO_B_CELL_MEDIATED_IMMUNITY	81	Down	2.91E-04	0.0296602 3
GO_REGULATION_OF_NUCLEASE_ACTIVITY	23	Down	3.08E-04	0.0302431 9
GO_POLYSOME	37	Down	3.12E-04	0.0302431 9
GO_TASTE_RECEPTOR_ACTIVITY	22	Down	3.16E-04	0.0302431 9

GO_MRNA_METABOLIC_PROCESS	572	Down	3.17E-04	0.030243 19
GO_MITOCHONDRIAL_PROTEIN_COMPLEX	128	Down	3.30E-04	0.030949 74
GO_AMIDE_BIOSYNTHETIC_PROCESS	483	Down	3.36E-04	0.030949 74
GO_EPIDERMAL_CELL_DIFFERENTIATION	129	Up	3.40E-04	0.030949 74
GO_REGULATION_OF_CELL_CYCLE_G2_M_PHA SE_TRANSITION	57	Down	3.55E-04	0.031785 29
GO_NUCLEAR_ENVELOPE_REASSEMBLY	17	Down	3.64E-04	0.032123 51
GO_RIBOSOME_BIOGENESIS	295	Down	4.38E-04	0.037661 36
GO_INTERMEDIATE_FILAMENT	85	Up	4.39E-04	0.037661 36
GO_LYMPHOCYTE_MEDIATED_IMMUNITY	124	Down	4.48E-04	0.037875 36
GO_PROTEIN_LOCALIZATION_TO_ORGANELLE	533	Down	4.97E-04	0.041368 47
GO_RIBONUCLEOPROTEIN_COMPLEX	685	Down	5.24E-04	0.043009 03
GO_INTERSPECIES_INTERACTION_BETWEEN_O RGANISMS	627	Down	5.37E-04	0.043505 89
GO_MITOCHONDRIAL_ELECTRON_TRANSPORT_	15	Down	5.59E-04	0.044657

CYTOCHROME_C_TO_OXYGEN				5

References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29(1):15-21.
2. Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* 2017;14(4):417.
3. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 2015;4
4. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015;43(7):e47-e47.
5. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic acids research* 2012;40(17):e133-e33.
6. Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739-40.
7. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010;20(9):1297-303.
8. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011;43(5):491.
9. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* 2012;22(3):568-76.
10. Ramaswami G, Li JB. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic acids research* 2013;42(D1):D109-D13.
11. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 2017;34(5):867-68.
12. Blokzijl F, Janssen R, Van Boxtel R, et al. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome medicine* 2018;10(1):33.
13. Alexandrov L, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *BioRxiv* 2018:322859.
14. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic acids research* 2018;47(D1):D941-D47.
15. Danaher P, Warren S, Dennis L, et al. Gene expression markers of tumor infiltrating leukocytes. *Journal for immunotherapy of cancer* 2017;5(1):18.