

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

WGS data from Patient MSK-HN1 normal and tumor DNA were aligned to GRCh37 human reference genome using Burrows-Wheeler Aligner (www.maq.sourceforge.net), duplicate reads marked by NovoSort (www.novocraft.com), realignment around indels and base calibration with Genome Analysis Toolkit (GATK) (www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit). Structural variants were determined by NBIC-seq (www.compbio.med.harvard.edu/BIC-seq), Crest (www.stjude.com/research/site/lab/zhang), Delly (www.korbel.embl.de/software.html), and BreakDancer (<http://breakdancer.sourceforge.net/>). For Patient MSK-HN1 RNA-seq, reads were aligned to the GRCh37.p13 genome using STAR 2.4.2a. Genes were quantified vs. Gencode v18 (www.gencodegenes.org) annotation using featureCounts from the Subread package (v1.4.3-p1). FusionCatcher (v0.99.5a) was used to detect fusions (<http://code.google.com/p/fusioncatcher/>). The DEK-AFF2 fusion was identified in both WGS and RNA-seq and confirmed by inspection in IGV. All high throughput sequencing results were high quality, and all normal and tumor tissue pairs were confirmed as concordant. Normal and tumor DNA from Patient MSK-HN1 had 145 million and 282 million reads, respectively, over 99% of which correctly aligned to the reference genome. Mean target coverage for MSK-HN1 was 60X for normal DNA and 90X for tumor DNA, using Picard CollectWgsMetrics. Patient MSK-HN1 tumor RNA-seq yielded over 50 million read pairs (125 bp in length per read), over 97% mapped reads, over 85% gene assignment rate (uniquely mapped reads assigned to exons), and 1.56% rRNA reads. GEM mapper was used to detect viral sequences from unmapped or partially mapped reads in WGS and RNA sequences (<http://gemlibrary.sourceforge.net/>). Patient human leukocyte antigens (HLAs) were determined by OptiType (<https://github.com/FRED-2/OptiType>) and Polysolver23 calling from normal DNA exome sequencing data. SNV calling was performed using Somatic Sniper version 1.0.5.0 (<http://gmt.genome.wustl.edu/packages/somatic-sniper/>), VarScan v2.4.0 (<http://varscan.sourceforge.net/>), Strelka v2 (<https://github.com/Illumina/strelka>), MuTect v 1.1.7 (<http://www.broadinstitute.org/cancer/cga/mutect>), and Platypus (<http://www.well.ox.ac.uk/platypus>). Mutations were annotated using SnpEffect and SnpSift version 4.3 (<http://snpeff.sourceforge.net/>). SNVs with an allele read count of less

than 4 or with corresponding normal coverage of less than 7 reads were filtered out. Copy number alterations were assessed by FACETS 0.5.6 (<https://sites.google.com/site/mskfacets/>) with *cval* of 300. For alternative splicing (AS) event calling, junction coordinates corresponding to non-annotated transcripts were extracted from the STAR alignment results and categorized according to the one of the following alternative splicing events: exon-skipping, alternative 5', alternative 3', mutually exclusive event. Translational effect (in-frame, or out-of-frame) were inferred based on the annotation. A database of novel junctions identified in GTEx and TCGA was used to estimate the prevalence of these junctions. AS events absent in GTEx or TCGA data, and with high PSI (percentage spliced inclusion; indicating the percentage of transcripts with the event) of 10% or greater were included.

For Patient MSK-HN1, all possible neoantigen peptides (9-amino acids in length) encompassing the mutated residues (for SNVs), fusion breakpoint (for gene fusions) or alternative splicing events were queried for patient-specific HLA binding by NetMHCpan 4.0 (<http://www.cbs.dtu.dk/services/NetMHCpan/>). Predicted peptide binding to HLA with < 2 %Rank is considered significant. All SNV nonamers predicted to bind patient HLA and all fusion peptides regardless of predicted HLA binding were tested for ability to stimulate patient T cell secretion of IFN- γ .

Neoantigen-HLA binding predictions for MYB-NFIB, NFIB-MYB, and MYBL1-NFIB fusions in ACC samples, and for melanoma samples in the study of Riaz, et al25, were generated from the INTEGRATE-Neo algorithm, which identifies transcripts spanning two genes (gene fusion transcripts) from RNA-seq data and queries fusion-derived peptides for HLA binding by using NetMHC 4.0 (<https://github.com/ChrisMaherLab/INTEGRATE-Neo>). HLAs were identified from RNA-seq data by the HLaminer algorithm (<http://www.bcgsc.ca/platform/bioinfo/software/hlaminer>), which can call more than 6 HLAs per sample due to sequence similarity among HLA subtypes, and with seq2HLA. Predicted peptide binding to HLA with < 500 nM affinity is considered significant. For the samples from Riaz, et al25, a total of 43 tumors had complete pre-therapy and on-therapy matched pairs with RNA-seq data available, of which 38 had a fusion neoantigen predicted in at least one sample.

RNA-seq data were aligned to the hg19 genome using STAR aligner, counted with Rsamtools v1.28 (<https://bioconductor.org/packages/release/bioc/html/Rsamtools.html>), and annotated with UCSC hg19.knownGene (<http://genome.ucsc.edu>). The Union counting mode was used for mapped reads only, and Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) were obtained with DESeq2 (<http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>). Read count capture was performed with Rsamtools and GenomicAlignments package35. Several orthogonal tools for the deconvolution of immune infiltration from RNA-seq data were implemented: Immune Infiltration Score (IIS) and T cell Infiltration Score (TIS) using single sample gene set enrichment analysis (ssGSEA)36, ESTIMATE Immune Score (ImmuneScore)37, Cytolytic Score (CYT)38, Cibersort Absolute Score (<http://cibersort.stanford.edu/>), and Reactome Interferon gamma39, (Reactome.org). The complete TCGA dataset of head and neck squamous cell carcinoma (HNSC) were analyzed by immune deconvolution methods and were ranked along with normalized PD-1 and PD-L1 expression levels to generate a composite z-score.

For analysis of TCGA cases with gene fusions, fusion data was downloaded from Gao, et al12 and immune metrics were downloaded from Thorsson, et al40. Analysis was limited to 5,825 samples with a fusion gene identified, representing 30 cancer types. We excluded hematologic malignancies (DLBC, LAML) and non-malignant paragangliomas (PCPG). Loss of heterozygosity of HLA class I genes (LOH-HLA) was determined by using FACETS of TCGA whole exome sequencing data with default settings (cbioportal.org). Briefly, segments within the chromosome 6p locus containing the HLA-A, -B, and -C loci were identified, and loss of heterozygosity (LOH) was defined as a minor allele copy number estimate of 0 for any of the HLA loci using the expectation-maximization model41,42. Cancer type and somatic mutational burden were each significantly associated with fusion neoantigen status in univariate analysis, and these were included as covariates in multivariable analysis. Multivariable logistic regression was used to evaluate the association of the presence of a fusion neoantigen with leukocyte fraction (from methylation data), lymphocyte infiltration (from RNA-seq data), the proportions of infiltrating CD8+ T cells, activated NK cells, macrophages, Th1 cells, and Th2 cells40, measures of immune cell activation such as the CYT score38 and IFN γ response pathway, and LOH-HLA, in all cases including tumor histology and tumor mutational burden as covariates.

Data analysis

GraphPad 7.04 and SPSS 24 were used to perform the statistical tests indicated in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Whole exome and RNA sequencing data have been deposited in SRA and are available under project number PRJNA527992.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. All experiments were performed in biological replicates with sample sizes of 3. Sample sizes were limited by patient peripheral blood. All sample sizes are described in the text.
Data exclusions	In the analysis of TCGA data, all samples within all cancer types with available data were included, with the exception of non-solid tumors (AML and DLBCL) and the largely non-malignant paraganglioma cohort, as is described in the Methods.
Replication	All attempts at replication were successful.
Randomization	Randomization is not relevant to this study, as no animals or human research participants are involved in this study.
Blinding	Investigators were not blinded to negative control and experimental samples when performing the experiments, but readouts such as ELISpots and flow cytometry were quantified by a plate reader or flow cytometer, and analyzed as such.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials	With written patient consent and under an IRB approved biospecimen protocol, peripheral blood was obtained post-immunotherapy treatment for Patient MSK-HN1, and post-standard-of-care treatment for Patient ACC_M9, during follow-up visits. All constructs, primers and peptides made are freely available without restrictions for use to investigators.
----------------------------	---

Antibodies

Antibodies used	<p>For immunohistochemical staining of MSK-HN1 tumor samples, the antibodies are listed with the supplier, catalogue number, and dilution: anti-CD3 (Leica cat# NCL-L-CD3-565, 1:100), anti-CD8 (Ventana, cat#790-4460, no dilution), anti-CD45 (Ventana cat# 760-2505, no dilution), and anti-PD-L1 (Cell Signaling ca# 13684, 1:800).</p> <p>For cell-based assays, the antibodies used are listed with the supplier, catalogue number, and dilution: PE-anti-HLA-A2 clone BB7.2 (BD cat # 343306, 1:500), Alexa Fluor 594 anti-HLA-A clone EP1395Y (Abcam cat# 207872, 1:1,000), polyclonal anti-HLA-B (Invitrogen cat# PA5-35345, 1:1,000), anti-HLA-C clone DT9 (Biolegend cat# 373302, 1:1,000), PE-Cy5.5-F(ab')₂-goat anti-mouse-IgG H+L (Invitrogen cat# M35018, 1:200), PE-F(ab')₂-goat anti-rabbit IgG H+L (Invitrogen cat#31864, 1:200), Alexa Fluor 405 anti-CD3e Clone UCHT1 (Thermo cat# CD0326, 1:100), PerCP/Cy5.5 anti-CD4 Clone A161A1 (Biolegend cat# 357414, 1:100), APC-H7 anti-CD8 Clone SKI (BD cat# 560179, 1:100), APC anti-PD-1 Clone M1H4 (eBioscience cat# 17-9969-42, 1:100), PE/Dazzle 594 anti-CD137 Clone 4B4-1 (Biolegend cat# 309825, 1:100), PE anti-CD40L Clone 24-31 (Biolegend cat# 310805, 1:100), Aqua Live/Dead (cat# L34966, 1:1,000), PE anti-active caspase-3 (BD cat# 550821, 1:50), anti-c-Myb Clone D2R4Y (Cell Signaling cat# 12319, 1:1,000), anti-DEK Clone E1L3V (Cell Signaling cat# 13962, 1:1,000), anti-HLA-ABC Monoclonal Antibody Clone W6/32 (Memorial Sloan Kettering Cancer Center antibody and bioresource core facility). Anti-CD3/CD28 beads were purchased from Gibco (cat# 11131D).</p>
Validation	<p>For immunohistochemical staining:</p> <p>anti-CD3 (Leica cat# NCL-L-CD3-565, 1:100): Normal human tonsil: immunohistochemical staining for CD3 antigen using NCL-L-CD3-565. Note in tense membrane staining of T lymphocytes. Paraffin section.</p> <p>anti-CD8 (Ventana, cat#790-4460, no dilution): Tonsil, T Cell Lymphoma</p> <p>anti-CD45 (Ventana cat# 760-2505, no dilution): Tonsil</p> <p>anti-PD-L1 (Cell Signaling ca# 13684, 1:800): Immunohistochemical analysis of paraffin-embedded human placenta.</p>

For cell-based assays:

PE-anti-HLA-A2 clone BB7.2 (BD cat # 343306, 1:500): HLA-A2 positive and HLA-A2 negative donors peripheral blood lymphocytes stained with BB7.2 PE.

Alexa Fluor 594 anti-HLA-A clone EP1395Y: Raji cells

polyclonal anti-HLA-B (Invitrogen cat# PA5-35345, 1:1,000): HepG2 cells compared to a negative control

anti-HLA-C clone DT9 (Biolegend cat# 373302, 1:1,000): Human peripheral blood lymphocytes

PE-Cy5.5-F(ab')₂-goat anti- mouse-IgG H+L (Invitrogen cat# M35018, 1:200): Anti-Mouse secondary antibodies are affinity-purified antibodies with well-characterized specificity for mouse immunoglobulins and are useful in the detection, sorting or purification of its specified target.

PE-F(ab')₂-goat anti-rabbit IgG H+L (Invitrogen cat#31864, 1:200): Product # 31864 has been successfully used in Western blot, IF, ICC, IHC, IP and FACS applications.

Alexa Fluor 405 anti-CD3e Clone UCHT1 (Thermo cat# CD0326, 1:100): Human peripheral blood lymphocytes

PerCP/Cy5.5 anti-CD4 Clone A161A1 (Biolegend cat# 357414, 1:100): Human peripheral blood lymphocytes

APC-H7 anti-CD8 Clone SKI (BD cat# 560179, 1:100): Human whole blood

APC anti-PD-1 Clone M1H4 (eBioscience cat# 17-9969-42, 1:100): Normal human peripheral blood lymphocytes

PE/Dazzle 594 anti-CD137 Clone 4B4-1 (Biolegend cat# 309825, 1:100): Human peripheral blood lymphocytes

PE anti-CD40L Clone 24-31 (Biolegend cat# 310805, 1:100): Human peripheral blood lymphocytes

Aqua Live/Dead (cat# L34966, 1:1,000): Jurkat cells

PE anti-active caspase-3 (BD cat# 550821, 1:50): Camptothecin treated cells

anti-c-Myb Clone D2R4Y (Cell Signaling cat# 12319, 1:1,000): Jurkat and MOLT-4 cells

anti-DEK Clone E1L3V (Cell Signaling cat# 13962, 1:1,000): HCT116 and COS-7 cells

anti-HLA-ABC Monoclonal Antibody Clone W6/32 (Memorial Sloan Kettering Cancer Center antibody and biosource core facility): Human peripheral blood lymphocytes

Anti-CD3/CD28 beads: Trickett A et al. (2003) T cell stimulation and expansion using anti-CD3/CD28 beads. J Imm Methods 275:251-255.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

T2 cells, COS-7 cells, HEK293 cells and SCC-9 cells were all obtained directly from ATCC.

Authentication

Cell lines were purchased directly from ATCC and cells had the morphology of published or manufacturer-provided images, thus not further authenticated.

Mycoplasma contamination

Cell lines were not tested for mycoplasma.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Patient MSK-HN1 is a female patient. ACC patients had a mean age of 63 years old: 8 male and 10 female. All of these patients provided written informed consent for the collection and analysis of biospecimens under MSKCC IRB approved biospecimen protocol #11-195.

Recruitment

Patient MSK-HN1 was an exceptional responder to anti-PD-1 after progressing on standard of care and was under the clinical care of the senior author. All ACC patient samples collected at MSKCC were included. The ACC patients were also all under the care of the authors.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were resuspended in FACS buffer (PBS + 2% FBS) and incubated with antibodies using standard techniques described in the methods section.

Instrument

BD Fortessa II, LSR II, BD FACSAria

Software

Sample data were collected with FACSDIVA and analyzed with FlowJo v10.

Cell population abundance

All cells recovered from co-cultures were assessed (10,000 cells or more).

Gating strategy

Cells were gated by FSC and SSC for singlets. Fixable Aqua Dead Cell Stain Kit was used for live/dead cell staining. Fluorescence signal was above the level of that on unstained controls. FMO was used to gate for positive signals.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.