nature research

Corresponding author(s): Prof. Yardena Samuels

Last updated by author(s): February 4th 2021

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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	SMURF package in R, V1 ggplot2 package in R, v3.3.0 ggtree package in R, v2.2.4 BWA-MEM software Kraken v.1.0 phyloseq package in R, v1.3.2.0 vegan version 2.5-4 package in R CD-HIT software, v4.6.6 MaxQuant version 1.5.0.25 NetMHCpan version 4.0 NetMHCIIpan version 3.2 GibbsCluster 2.0 server Seq2Logo 2.0 Peptides package in R, v2.4.1 R package OrgMassSpecR, v0.5.3 Trimmomatic v0.38-WGS data for bacteria (Bolger, A. M. et al., 2014) , SPAdes v3.14 -Quality filter of WGS paired-end de novo assembled (Bankevich A. et al., 2012), QUAST v5.0 -Genome assembly quality(Gurevich, A. et al., 2013), CheckM v1.0.18 -Genome assembly quality(Parks, D. H. et al., 2015), BWA-MEM software v0.7(Li, H.et al,2013), metabat2 v2.14 -Contig grouping (Kang, D. et al, 2015), GTDB release 04-RS89-Taxonomic classifications (Parks, D. H. et al., 2020), GTDB-tk v0.3.2- Taxonomic classification(Chaumeil, P. et al, 2019), fastANI v1.1-Taxonomic classification(Jain, C. et al,2018), dRep- clustering (Olm, M. et al,2017), Prokka v1.140- protein sequence annotations(Seemann, T. et al, 2014). FlowJo software v7

ImageJ 1.53C
Imaris 9.1 (Bitplane) software

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. 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

All raw MS files as well as human and bacteria proteomes and the MaxQuant version used for analyzing the data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022150. Raw sequence data of isolated bacteria WGS were deposited in the NCBI Sequence Read Archive (SRA), under BioProject accession number PRJNA669827. Source Data 1 is related to Figure 2a, and includes all raw files of identified peptides for HLA peptidomics experiments. Source Data 2 is related to Figure 3e, and includes all the data recived from bacteria WGS (16S rRNA sequences, genome assembly and Protein sequences).
Other publicly available detests used:
The Greengenes database (May 2013 version) for 16S sequencing reference
High-throughput DNA sequencing of 108 pairs of tumor and blood samples (Hayward, N. K. et al., 2017)

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample size was determine according to biological material availability
Data exclusions	Data was not excluded
Replication	All experiments required were done in triplicates and repreduced
Randomization	For all high throughput data analysis-randomization for samples into experiment groups is not relevant, as the design of the study requires explorative analysis for identifying discriminative features between already established experimental groups.
Blinding	For all high throughput data analysis-blinding to investigators is not relevant to this study, as the design of the study requires explorative analysis for identifying discriminative features.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-Lipoteichoic acid antibody, Cat #HM5018, Hycult Biotech, Clone 55 , LOT 25823M0119-A anti-LPS, Cat# HM6011, Clone WN1 222-5

anti-mouse secondary antibody labeled with Alexa flour 488, Cat #115-545-146, Jackson Immunoresearch, Clone , LOT 117790
anti-mouse secondary antibody labeled with Alexa flour 546, Cat #A-11030, Invitrogen, AB_141370
AF647-conjugated CD45, Cat# 7368537, lot B239311s, clone 2D1,BioLegend
APC/Cy7-conjugated CD19 , Cat# 302218, lot B279663, clone HIB19,BioLegend
APC-conjugated CD19, Cat# 302212, lot B182157, clone HIB19,BioLegend
APC-conjugated HLA-A,B,C ,Cat# 311409, lot B213470, clone W6/32,BioLegend
APC-conjugated HLA-DR,DP,DQ ,Cat# 361714, lot B289409, clone Tu39,BioLegend
BV421-conjugated CD69,Cat# 310930, lot B284312, clone FN50,BioLegend
anti-GAPDH, Cat# MAB374, lot 3043558, clone 6C5,Millipore
anti-CD45, Cat# 13917, lot 6, clone D9M8I,Cell Signaling
anti-HLA-DP1, Cat# ab157210, lot GR3173627-19, clone EPR11226,Abcam
APC/Cy7-conjugated CD14 Cat# 301820, lot B280758, clone M5F2,BioLegend
Anti-pan-HLA–I antibody, Cat #H1650 ,Sigma, Clone W6/32, LOT 22130603
IFNg capture antibody, as part of kit, Cat #130-054-202, Miltenyi Biotec, LOT 5190207203

Validation

All antibodies used in the study were validated by the manufacturer, data is available at the manufacturer's website, as indicated bellow:
anti-Lipoteichoic acid antibody, Cat #HM5018, Hycult Biotech, Clone 55 , LOT 25823M0119-A: https://www.hycultbiotech.com/hm5018-1
anti-LPS, Cat# HM6011, Clone WN1 222-5, https://www.hycultbiotech.com/hm6011
anti-mouse secondary antibody labeled with Alexa flour 488, Cat #115-545-146, Jackson Immunoresearch, Clone , LOT 117790, https://www.jacksonimmuno.com/catalog/products/115-545-146
anti-mouse secondary antibody labeled with Alexa flour 546, Cat #A-11030, Invitrogen, AB_141370, https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11030
AF647-conjugated CD45, Cat# 7368537, lot B239311s, clone 2D1,BioLegend, https://www.biolegend.com/en-us/products/apc-anti-human-cd45-antibody-12397
APC/Cy7-conjugated CD19 , Cat# 302218, lot B279663, clone HIB19,BioLegend, https://www.biolegend.com/en-us/products/apc-cyanine7-anti-human-cd19-antibody-1910
APC-conjugated CD19, Cat# 302212, lot B182157, clone HIB19,BioLegend, https://www.biolegend.com/en-us/products/apc-anti-human-cd19-antibody-715
APC-conjugated HLA-A,B,C ,Cat# 311409, lot B213470, clone W6/32,BioLegend, https://www.biolegend.com/en-us/products/apc-anti-human-hla-a-b-c-antibody-1870
APC-conjugated HLA-DR,DP,DQ ,Cat# 361714, lot B289409, clone Tu39,BioLegend, https://www.biolegend.com/en-us/products/apc-anti-human-hla-dr-dp-dq-antibody16319
BV421-conjugated CD69,Cat# 310930, lot B284312, clone FN50,BioLegend, https://www.biolegend.com/en-us/products/brilliant-violet-421-anti-human-cd69-antibody-7141
anti-GAPDH, Cat# MAB374, lot 3043558, clone 6C5,Millipore, https://www.merckmillipore.com/INTL/en/product/Anti-Glyceraldehyde-3-Phosphate-Dehydrogenase-Antibody-clone-6C5,MM_NF-MAB374
anti-CD45, Cat# 13917, lot 6, clone D9M8I,Cell Signaling, https://www.cellsignal.com/products/primary-antibodies/cd45-intracellular-domain-d9m8i-xp-rabbit-mab/13917
anti-HLA-DP1, Cat# ab157210, lot GR3173627-19, clone EPR11226,Abcam, https://www.abcam.com/hla-class-i-antibody-w632-low-endotoxin-azide-free-ab23755.html

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

51AL,55A3,58A were produced by MD Anderson
EBV-transformed B-cells were purchased from the IHWG Cell and DNA Bank
Hybridoma cells HB95 and HB145 were purchased from the ATCC
LCL 721.221 cell line was purchased from the ATCC
THP-1 cells were a kind gift from Prof. Steffen Jung (originally from ATCC)

Authentication

Commercial cells were authenticate by the supplier
51AL, 55A3, 58A- The Cell line authentication test was performed at the Genomics Center of Biomedical Core Facility, Technion. The test was performed using the Promega GenePrint 24 System in order to determine short tandem repeat (STR) profile of 23 loci plus Amelogenin for gender determination (X or XY). In addition, the male-specific DYS391 locus is included to identify null Y allele results for Amelogenin. DNA sample from the kit (2800M Control DNA) was included in the analysis and served as positive control for the PCR step. No DNA template was also included as negative control. The results were analyzed using the 3500xl Genetic Analyzer (Life Technologies) and GeneMapper IDX software. Allelic ladder was included in the run.

Mycoplasma contamination

All cell lines were tested regularly and were found negative for mycoplasma contamination (EZ-PCR Mycoplasma Kit, Biological Industries).

Commonly misidentified lines
(See ICLAC register)

no commonly misidentified cell lines were used in the study

Human research participants

Policy information about studies involving human research participants

Population characteristics	All patients included in the analysis were diagnosed with metastatic melanoma. Tumor samples from patients 19, 42, 51, 55, 58, 70, 86, 92, 112, 152 and 422 were received from eleven different patients treated at the University of Texas MD Anderson Cancer Center, all of who had signed an informed consent for the collection and analysis of their tumor samples. The protocol for tumor samples was approved by the MD Anderson Institutional Review Board (protocol numbers 2012-0846, LAB00-063 and 2004-0069, NCT00338377). Synchronous metastatic tumors were resected via surgery at the same time. Sample and patient information are provided in Extended Data Table 1. From each patient, we collected between 1-3 different metastases. Tumor sample HG38 was received from patient biopsies removed at the University of Zurich Hospital. The patient had signed a patient release form, which has been approved by an ethics committee and assigned the numbers EK647 and EK800. Tumor 261MS was received from patient biopsies removed at the Ella Lemelbaum Institute for Immuno-Oncology (Sheba medical center, Israel), under Institutional Review Board number 6786-20-SMC.
Recruitment	Patient were recruited by University of Texas MD Anderson Cancer Center, all of who had signed an informed consent for the collection and analysis of their tumor samples. Tumor sample HG38 was received from patient biopsies removed at the University of Zurich Hospital. The patient had signed a patient release form, which has been approved by an ethics committee and assigned the numbers EK647 and EK800. Tumor 261MS was received from patient biopsies removed at the Ella Lemelbaum Institute for Immuno-Oncology (Sheba medical center, Israel), under Institutional Review Board number 6786-20-SMC.
Ethics oversight	The protocol for tumor samples was approved by the MD Anderson Institutional Review Board (protocol numbers 2012-0846, LAB00-063 and 2004-0069, NCT00338377). Tumor sample HG38 was received from University of Zurich Hospital under an ethics committee and assigned the numbers EK647 and EK800. Tumor 261MS was received from the Ella Lemelbaum Institute for Immuno-Oncology (Sheba medical center, Israel), under Institutional Review Board number 6786-20-SMC.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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	TILs after co-culture with EBV transformed B-cells loaded with peptides, were labeled with an IFNg capture antibody (130-054-202, Miltenyi Biotec) and then incubated for 45 minutes at 37°C. Then cells were stained with the detection antibody. For HLA-I and HLA-II staning all cells were washed X2 and staind for 30 minutes on ice with the antibodies.
Instrument	BD LSR II flow cytometer (BD Biosciences)
Software	FlowJo software
Cell population abundance	Known cell amounts, of the different cell types, in known ratio and was constant during experiments
Gating strategy	In all assays, cells were first gated for live, single cells, and then gated according to the requested staining.

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