nature portfolio

Corresponding author(s): Thomas Kuilman & Gavin M. Bendle

Last updated by author(s): Nov 16, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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

Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.

Data analysis

Mutation identification was performed as previously described33 (experiments in Fig. 5). Alternatively, mutation calling was performed using GATK (version 4.1.3.0) according to their best practices for somatic short variant discovery (https://gatk.broadinstitute.org/hc/en-us/articles/360035894731-Somatic-short-variant-discovery-SNVs-Indels-). STAR (version 2.7.3a) was used for aligning RNA reads, and primary transcript counts were used as a proxy for expression of a mutation. An adjusted version of the Neoepiscope tool63 was used for subsequent MG identification (https://github.com/ThijsMaas/neoepiscope) as follows. The first 12 amino acids (AA) of the MG consisted of wild-type sequence. This sequence was followed by the first altered AA in position 13. The MG sequence was completed with additional wild-type sequence (in case of point mutations) or additional out-of-frame sequence until the first STOP codon (in case of out-of-frame mutations). MGs had a length of 25 AA or more (in case of long out-of-frame sequences). In cases where a minimum of 25 AA could not be obtained using the design guidelines described above sequence frames were moved away from start/stop codons to reach a 25 AA MG sequence. Mutations were further selected for screening using a combined, equal weighing of i) variant-allele frequency and ii) expression of the mutation-containing transcript. A variable number of MGs was distributed over TMGs with each TMG flanked by LAMP1 signaling domain (N-terminal), and transmembrane and cytosolic domains (C-terminal) to allow targeting to the HLA class II presentation machinery24,25. TMGs were codon optimized and cloned into a pMX P2A Puro T2A Ly6G retroviral transfer vector to allow coexpression of the TMG, a puromycin resistance gene and murine Ly6G from the same transcript. TMGs were combinatorially encoded in TMG pools by ensuring that a given TMG is present in exactly one combination of TMG pools, and that every combination of TMG pools encodes exactly one TMG. As an example of how to achieve this, the combinatorial TMG design in Extended Data Fig. 4 was created by diagonal mirroring of the 7 x 7 block of TMGs from the R1-R7 design to obtain the C1-C7 design. As a positive control, autologous B cells expressing a construct encoding B2M-linker-HLA-A*02:01 and a minigene encoding a mutated CDK4 epitope were included. These cells can be recognized by the positive control TCR CDK4-17.

Screens were executed in triplicate when B cell pools were not designed using combinatorial TMG encoding (screens in Fig. 3 and Extended Data Fig. 5d (pt3/pt15)), while screens were executed once for each B cell pool (i.e., all TMGs were screened twice given their representation in two independent B cell pools) when applying combinatorial TMG encoding (all other screens, e.g. including the screen in Fig. 4 according to the combinatorial TMG encoding design in Extended Data Fig. 4). Neoantigen-specific TCR candidates were identified using differential count analysis with the DESeq2 Bioconductor package (version 1.26.0). An interaction model was applied to identify TCRs that were enriched in activated samples relative to non-activated samples, specifically in those samples that were derived from cocultures in the presence, but not the absence, of a given TMG (or pool of TMGs) in autologous B cells. Statistical testing was performed in a one-sided fashion. For screens in which combinatorial TMG encoding in pools was applied (Fig. 4 and pt4 (CRC4) in Fig. 5), this analysis was repeated for every combination of row and column pools. All results were combined, and a Bonferroni correction was applied to adjust p-values for multiple testing. For visualization purposes, 'TCR reactivity' (Fig. 3f,5b and Extended Data Fig. 5j,k) was calculated for each TCR as the difference between the sum of all DESeq2 rlog values of relevant activated samples, and the sum of all rlog values of relevant non-activated samples. These values were scaled according to the number of samples in the presence and absence of relevant antigen prior to graphical representation. Bonferroni-corrected p-values were used to identify candidate TCRs, with cut-offs of $p<0.01$ for screens using combinatorial TMG encoding (melanoma patient and pt4 screens) and $p<1$ for other screens (pt3 and pt15 screens).
TCR clustering analysis was performed using GIANA64 with default settings or by calculating the minimal Levenshtein edit distance between the alpha and beta chain CDR3s. A minimal edit distance threshold of 2 or less between any pair of TCRs was used to predict recognition of the same antigen. Network graphs were made using the igraph R package (version 2.0.2). Only fully validated and deconvolved TCRs were included in the analysis.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Read counts from functional genetic screens are available in Supplementary Table 1. Additional data are available upon request unless restricted by consent of study subjects.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex and/or gender-based analyses were not performed as they are not relevant for this proof of concept study.
Reporting on race, ethnicity, or other socially relevant groupings	Race and/or ethnicity-based analyses were not performed as they are not relevant for this proof of concept study.
Population characteristics	Population characteristics were not collected for this study as they are not relevant to this proof of concept study.
Recruitment	Patients whose tumor specimens and matched PBMCs were collected were recruited via collaborations in an anonymzed fashion. Material availability and tumor indication were factors that determined suitability for inclusion in this proof of concept study.
Ethics oversight	Tumor tissue was collected from patients treated at the NKI-AVL (Amsterdam, the Netherlands), LUMC (Leiden, the Netherlands) and Moffitt Cancer Center (Tampa, Unites States) with informed consent and in accordance with guidelines of the Institutional Review Board of the Moffitt Cancer Center and the Medical Ethical Committees of the NKI-AVL and of the Leiden University Medical Centre.

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

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	17 screens are reported in this proof of concept work to demonstrate the broad applicability of TCR discovery using the described platform.

Data exclusions	No data were excluded from the presented analyses.
Replication	TCR discovery screens were performed in duplicate or triplicate. Candidate TCRs from the screens were validated in independent experiments as presented in this manuscript. All key readouts were performed in duplicate or triplicate. All attempts at replication were successful.
Randomization	No experimental groups were employed in this work, and sample allocation was therefore not possible nor relevant.
Blinding	No experimental groups were employed in this work, and blinding was therefore not possible nor relevant.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Biolegend MojoSort Human Pan B cell Isolation kit (480081); as per manufacturer's instructions Miltenyi human CD20 MicroBeads (130-091-104); as per manufacturer's instructions BD Biosciences human CD20 FITC (555622, Clone 2H7); 1:20 dilution BD Biosciences human CD69 PE (557050, Clone FN50): 1:100 dilution Miltenyi human CD62L Microbeads (130-091-758); 20 µL per 10e6 cells BD Biosciences human CD8 PerCPCy5.5 (565310; Clone SK1); 1:100 dilution BD Biosciences mouse TCRbeta chain-PE (553172; Clone H57-597); 1:150 dilution BD Biosciences human CD20 PE-Cy7 (335828; Clone L27); 1:200 dilution BD Biosciences human CD5 BV650 (740564; Clone UCHT2); 1:200 dilution BD Biosciences human CD8 APC (557834; Clone SK1); 1:300 dilution Antibodies Online HLA class II blocking antibody (ABIN5067814; Clone IVA12); 120 µg/mL
Validation	BD: The antibody development process includes testing on a combination of primary cells, cell lines and/or transfectant cell models with relevant controls using multiple immunoassays to ensure biological accuracy. They perform multiplexing with additional antibodies to interrogate antibody staining in multiple cell populations. BioLegend: All newly developed clones at BioLegend undergo validation testing for multiple applications. This serves as a cross-check for specificity and provides clarity for research uses. Typically, antibodies are tested by two or more of the following methods: endogenous positive and negative expressing cells, KO/KD cell lines, isotype controls, biological testing (ie. treatments, activation etc.), multiple donor samples for human cells, side-by-side comparisons with currently available clones. This is followed by Western-Blotting, Flow Cytometry, ChIP, Immunofluorescence imaging, Immunohistochemistry. Miltenyi: All our antibodies are rigorously tested and validated before release. In the application section on the product page, you can find examples of typical performance data. In addition, we provide extended validation data highlighting details of antibody performance, specificity, and fixation compatibility. All antibodies for which any of these datasets are already available will be indicated with the extended validation stamp. That list will be continuously expanded, so be sure to check back soon if you don't see what you need yet. The IVA12 blocking antibody was validated by titrating the antibody into cocultures of Jurkat cells expressing either of 2 HLA class II restricted TCRs and CD4 with B cells expressing the relevant antigens.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Jurkat E6-1 (ATCC) B cells isolated from various patient PBMC specimens
Authentication	Jurkat cells were authenticated via STR profiling. B cells were authenticated by HLA-typing

Mycoplasma contamination	All cell lines were tested negative for mycoplasma
Commonly misidentified lines (See ICLAC register)	The Jurkat cells used throughout this manuscript were obtained from the American Type Culture Collection (ATCC), and were all derivatives from the standardized and validated clone E6-1. Clonal derivatives were confirmed using STR profiling.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

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	All samples were prepared as described in the Methods section. Samples were washed in FACS buffer (PBS + 2% bovine serum albumin) and stained with fluorescent antibodies in FACS buffer, washed in FACS buffer and resuspended in FACS buffer for flow cytometry.
Instrument	All samples were measured on a BD Biosciences LSRFortessa machine, and cell sorting was performed on a BD Biosciences FACSAria Fusion.
Software	BD FACSDiva (version 8.0.2) was used for the acquisition of flow cytometry samples; FlowJo (version 10.7.1) was used for analysis.
Cell population abundance	For cell sorting, post-sorting T cell fraction purity was not determined but the PCR to retrieve TCR inserts was specific for the TCR library transduced T cellls
Gating strategy	For measuring or sorting coculture experiments with TCR or TCR-library transduced T cells, cell populations were gated on SSC-A/FSC-A for to determine lymphocytes, FSC-H/FSC-W or FSC-H/FSC-A or SSC-W/SSC-H for determining single cells, CD69-PE/CD20-FITC or CD69-PE/CD8-PerCPCy5.5 or SSC-A/CD20-PECy7 or CD8-PerCPCy5.5/BCL6-BCLXL-GFP-FITC or CD5-BV650/ BCL6-BCLXL-GFP-FITC for determining T cells and SSC-A/CD69-PE for determining T cell activation or CD8-APC/mTCRbeta-PE for determining TCR transduction.

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