nature research

Corresponding author(s): Marta Łuksza, Benjamin Greenbaum

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

n/a

2. Data exclusions

Describe any data exclusions.

We exclude neoantigens with mutations on position 2 and 9 and a non hydrophobic wildtype residue at this positions from contributing to the fitness model calculation.
We exclude 10 non-metastatic melanoma samples from the Van Allen et al. cohort. with lacking subclass information.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The analysis did not involve experiments.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

n/a

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

n/a

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Data preparation:

Raw sequence data reads were aligned to the reference human genome (hg19) using the Burrows-Wheeler Alignment tool. Local realignment and quality score recalibration were conducted using the Genome Analysis Toolkit (GATK). For sequence alignment and mutation identification, the FASTQ files were processed to remove any adapter sequences at the end of the reads using cutadapt (v1.6) The files were then mapped using the BWA mapper (bwa mem v0.7.1246, the SAM files sorted, and read group tags added using the PICARD tools. After sorting in coordinate order, the BAM's were processed with PICARD MarkDuplicates. First realignment was carried out using the InDel realigner followed by base quality value recalibration with the Base-QRecalibrator.

A combination of 4 different mutation callers (Mutect 1.1.4, Somatic Sniper 1.0.4, Varscan 2.3.7, and Strelka 1.013) were used to identify single nucleotide variants (SNVs).

The assignment of a somatic mutation to a neoantigen was estimated using a previously described bioinformatics tool called NASeek (ref: Snyder. et al. N. Engl. J. Med. 371, 2189-2199 (2014).) NASeek evaluates putative MHC Class I binding for both wild type and mutant nonamers using a sliding window method using NetMHC3.4.

Tumor clones are reconstructed using the PhyloWGS software package (https://github.com/morrislab/phylowgs).

Neoantigen fitness model:
Model and formula are presented in Methods. Sequence alignments were performed with blastp and alignment scores were evaluated with Biopython Bio.paiwise2 package. Custom script examples for computation of neoantigen fitness cost are included as Supplementary Data 7. Additional custom code available upon reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

n/a

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

n/a

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

n/a

b. Describe the method of cell line authentication used.

n/a

c. Report whether the cell lines were tested for mycoplasma contamination.

n/a

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	n/a

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	n/a