nature research

Corresponding author(s): Valsamo Anagnostou, Victor Velculescu

Last updated by author(s): Nov 4, 2019

# 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<br>*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<br>*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. Previously published genomic data, re-analyzed here, were obtained from dbGaP under accession code phs000980.v1.p1 (cohort 2) and from the CBioPortal for Cancer Genomics https://www.cbioportal.org/study/summary?id=nsclc_pd1_msk_2018 (MSKCC, J Clin Oncol 2018; cohort 3) and https://www.cbioportal.org/study/summary?id=tmb_mskcc_2018 (MSKCC, Nat Genet 2019; cohort 4). |
| Data analysis | Somatic mutations were identified using the VariantDx software (Jones et al, Science Translational Medicine, 7:283, 2015). Missense mutations were evaluated for their potential as cancer drivers by CHASMplus (v1.0.0) with consequence prediction using CRAVAT (v4.3). HLA genotyping was performed using OptiType (v1.2). MHC class I prediction of neoantigens utilized netMHCpan (Nielsen et al, Genome Med, 8:33, 2016) and netCTLpan (Stranzl et al, Immunogenetics, 62, 2010). Mutational signatures were quantified using the deconstructSigs R package (v1.8.0). Somatic copy number profiles were estimated using CNV-kit (Talevich et al, PLoS Comput Biol 12, 2016), FACETS (v0.5.0) and as described in Anagnostou et al, Cancer Discovery, 2017. POLYSOLVER was applied to detect somatic mutations in class I HLA genes (Shukla et al, Nat Biotechnol, 33:11, 2015). LOHHLA was applied to determine loss of HLA genes (McGranahan et al, Cell, 171, 2017). Further statistical analyses were performed using the SPSS software program (version 25.0.0) and R version 3.2 and higher).Statistical analyses were done using the SPSS software program (version 25.0.0 for Windows, IBM) and R version 3.2 and higher, http://www.R-project.org/). |

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

Whole exome sequencing data that support the findings of this study have been deposited in the database of Genotypes and Phenotypes (dbGaP) and the European Genome-phenome Archive (EGA) under accession codes 36406 and EGAS00001003892 respectively (cohort 1). Source data for the TCGA tumor samples were retrieved from http://cancergenome.nih.gov. WES-derived somatic mutation calls from the TCGA PanCancer Atlas MC3 project were retrieved from the NCI Genomic Data Commons (https://gdc.cancer.gov/about-data/publications/ mc3-2017). Previously published genomic data, re-analyzed here, were obtained from dbGaP under accession code phs000980.v1.p1 (cohort 2) and from the CBioPortal for Cancer Genomics https://www.cbioportal.org/study/summary?id=nsclc_pd1_msk_2018 (MSKCC, J Clin Oncol 2018; cohort 3) and https://www.cbioportal.org/study/summary?id=tmb_mskcc_2018 (MSKCC, Nat Genet 2019; cohort 4). All other data supporting the findings of this study are available from the corresponding author on reasonable request.

# 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 determined based on the the number of patients available retrospectively that were treated with immune checkpoint blockade. No statistical methods were used to predetermine sample size. |
| Data exclusions | Cases were not included in the final analyses when tumor purity <10% or absence of matched normal samples, which represent pre-established criteria. |
| Replication | The improved predictive efficacy of corrected TMB was validated in cohort 1 (WES) and cohort 4 (targeted NGS). The enrichment of RTK activating mutations in non-responding tumors was noted in cohorts 1, 2 and 3. The integrated multiparameter predictive model of outcome (consisting of corrected TMB, molecular smoking signature, RTK mutations and HLA germline variation) was generated in cohort 1 and validated in cohort 2. Complete sequencing data to validate the multivarable model (i.e. raw sequence data to perform copy number analyses) were not available for cohorts 3 and 4. |
| Randomization | No randomization was performed as this was a retrospective study. |
| Blinding | Non applicable, as this is a retrospective study. |

# 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 | Primary mouse anti-human CD8 antibody, (Dako, catalog number m7103, clone C8/144B, 1:100 dilution) and anti-PD-L1 primary antibody (Cell Signaling Technologies, catalog number 13684, E1L3N clone, 1:100 dilution). |

| | |
|---|---|
| Validation | SDS-PAGE analysis of immunoprecipitates formed between lysates of 125I-labeled human T lymphoblasts and the antibody shows reaction primarily with a 32 kDa polypeptide corresponding to CD8a (Mason et al., J Clin Pathol 1992:45:1084-8). For the PDL1 antibody Western blot analysis of extracts from KARPAS-299, SUP-M2, and PC-3 cells using PD-L1 (E1L3N) Rabbit mAb detects the PDL1 protein as per the manufacturer's product insert. |

## Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | Cohort 1 consisted of 104 NSCLC patients treated with immune checkpoint blockade at Johns Hopkins Sidney Kimmel Cancer Center and the Nederlands Kanker Instituut. Clinical characteristics for all patients are summarized in Extended Data Table 1. Briefly cohort 1 consisted of 46 males and 43 females with an average age of 65 years. Eighty patients were treated with anti-PD1 therapy, 7 patients received combination anti-PD1 and anti-CTLA4 therapy and 2 patients were treated with chemotherapy and anti-PD1 therapy. Exome data from a published cohort of NSCLC patients treated with PD1 blockade (cohort 2) were obtained and analyzed to validate key findings from cohort 1 (Rizvi et al., Science, 2015). A publicly available cohort of 240 NSCLC patients treated with ICB2 was obtained through CBioPortal for Cancer Genomics (MSK, JCO 2018; http://www.cbioportal.org/study?id=nsclc_pd1_msk_2018) and used to validate the association of RTK mutations with outcome (cohort 3). A publicly available cohort of 1,661 tumors analyzed by targeted next-generation sequencing was obtained through CBioportal for Cancer Genomics (MSKCC, Nat Genet 2019) to validate the correlation between TMB and tumor purity in the setting of higher sequencing depth (cohort 4). |
| Recruitment | There was no patient recruitment given the retrospective nature of this study. Patients in cohort 1 were consecutive patients receiving immunotherapy in our institutions, for which tissue was available for genomic analyses. |
| Ethics oversight | The study was approved by the Institutional Review Board (IRB) and patients provided written informed consent for sample acquisition for research purposes. |

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

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| | |
|---|---|
| Clinical trial registration | Non applicable |
| Study protocol | Non applicable |
| Data collection | Non applicable |
| Outcomes | Non applicable |