nature portfolio

Corresponding author(s): Francesco Maura, MD

Last updated by author(s): 09/17/2023

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	Newly generated data: - WGS: 60 CD138+ bone marrow mononuclear cell samples were sequenced using WGS. Samples were collected from 49 patients with newly diagnosed multiple myeloma enrolled in two phase clinical trials (NCT03290950). - Single cell RNA sequencing data: 37 CD138- bone marrow mononuclear cells were sequenced. Samples were collected from 22 patients enrolled in the phase II clinical trial NCT03290950. - Clinical data from the trial can be found in Landgren et al. JAMA Onc 2021: PMID: 33856405. All patients gave written informed consent to the use of their data and to enroll in the trial. Validation cohorts: - The Kydar scRNA-seq data are publicly available at the National Center for Biotechnology Information's Gene Expression Omnibus (accession no. GSE161195). - The CoMMpass data was downloaded from the MMRF researcher gateway portal (https://research.themmrf.org). CoMMpass raw data are accessible on dbgap: phs000748.v1.p1.
Data analysis	WGS: Short insert paired-end reads were aligned to the reference genome (GRCh37) using the Burrows–Wheeler Aligner (v0.5.9). All samples were uniformly analyzed by the following bioinformatic tools: somatic mutations were identified by integrating CaVEman v1.7.5, Mutect2 4.0.1.2, Strelka v2.9.1 and Pindel v1.5.4; copy number analysis and tumor purity (i.e., cancer cell fraction) were evaluated using Battenberg v1.4.0 (https://github.com/Wedge-Oxford/battenberg); structural variants were defined by BRASS v4.0.5 (https://github.com/cancerit/BRASS) via discordant mapping of paired-end reads, passed through additional quality filters, and were manually curated to define complex events (i.e., templated insertions, chromothripsis, and chromoplexy). Mutational signatures were analyzed across all whole genomes. To estimate the activity of mutational signatures, we first employed a three-

step process of de novo extraction, assignment, and fitting . For the first step, we ran SigProfiler for SBS signatures (https://github.com/AlexandrovLab/SigProfilerExtractor). All extracted signatures were then compared with the latest Catalogue of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic/signatures/SBS) to identify the known mutational processes active in the cohort. Finally, we applied mmsig (https://github.com/UM-Myeloma-Genomics/mmsig), a fitting algorithm, to confirm the presence and estimate the contribution of each mutational signature in each sample guided by the catalog of signatures extracted for each individual sample.

scRNA: Sequencing results were demultiplexed and converted to FASTQ format using Illumina bcl2fastq software. The Cell Ranger Single-Cell Software Suite (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger) was used to perform sample demultiplexing, barcode processing, and single-cell 5 gene counting. Data were then analyzed using Seurat R package

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

WGS and scRNA seq data have been uploaded on EGA. Study: EGAS00001007404, Dataset: EGAD00001011132. The trial enrolled newly diagnosed multiple myeloma 18 year or older. Inclusion and exclusion criteria can be found on https://clinicaltrials.gov/study/NCT03290950
The Kydar scRNA-seq data are publicly available at the National Center for Biotechnology Information's Gene Expression Omnibus (accession no. GSE161195). The CoMMpass data was downloaded from the MMRF researcher gateway portal (https://research.themmrf.org). CoMMpass raw data are accessible on dbgap: phs000748.v1.p1.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	60% of the patients were female.
Population characteristics	The key clinical and demographic features are summarized in Supplemental Table 1-2. Overall, 32% of patients were >65 years old and 11% were self-defined African Americans. Because race doesn't affect clinical outcomes in multiple myeloma once corrected for treatment accessibility, the limit linked to race self-identification did not have any impact in our results. Inclusion and exclusion criteria can be found on https://clinicaltrials.gov/study/NCT03290950. Clinical data from the trial can be found in Landgren et al. JAMA Onc 2021: PMID: 33856405. All patients gave written informed consent to the use of their data and to enroll in the trial.
Recruitment	This study included all patients enrolled in two phase clinical trials (NCT03290950 and NCT02937571) with available bone marrow mononuclear cells collected before treatment. This was the only inclusion criteria for this study.
Ethics oversight	This study was approved by the Institutional Review Board (IRB) at Memorial Sloan Kettering Cancer Center (#14-276 and 18-143) with an assurance filed with and approved by the U.S. Department of Health and Human Services.

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	49 patients. This was established according to the samples availability and quality.
Data exclusions	None of the patients enrolled in the phase clinical trial NCT03290950 were removed unless for sample and sequencing quality issue. Five patients failed WGS for low sample purity. Five baseline samples were excluded from the scRNA analysis as they had received prior treatment.

Replication	Phase II clinical trial usually don;t require replications. To validate our findings we leveraged extrenal data set (i.e. CoMMpass and Kydar trial; see above)
Randomization	Phase 2 clinical trials don't require randomization.
Blinding	Not applicable. Phase 2 clinical was not blind in its first place and analysis.

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

Methods

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

Antibodies

Antibodies used

Hashtag antibodies
-TotalSeq™-C0251 anti-human Hashtag 1 Antibody
- TotalSeq™-C0252 anti-human Hashtag 2 Antibody
-TotalSeq™-C0253 anti-human Hashtag 3Antibody
-TotalSeq™-C0254 anti-human Hashtag 4 Antibody

Flow sorting antibodies
- PE anti-human CD138 Antibody-DL-101 (1:100) Biolegend 352305
- PE anti human CD38 antibody IB6 (1:100) Miltenyi 130-113-427
- APC anti-human CD45 HI30 Antibody (1:200) Biolegend 304011

Validation

Hashtag antibodies: Each lot of this antibody is quality control tested by immunofluorescent staining with flow cytometric analysis and the oligomer sequence is confirmed by sequencing. TotalSeq™-C antibodies are compatible with 10x Genomics Chromium Single Cell Immune Profiling Solution. β2-microglobulin (β2M) is a 12 kD nonpolymorphic Ig like protein. It is a non-membraneanchored glycoprotein and is noncovalently associated with 39-44 kD polymorphic heavy chains of MHC class I molecules to form HLA class I antigen complex. In association with HLA class I, β2M is expressed on all leukocytes, platelets, endothelial cells, and epithelial cells. β2M plays an essential role both in governing MHC class I molecules stability and in promoting antigen binding and presenting the antigen to CD3/TCR complex of CD8 T cells. CD298 or the β3 Na/K ATPase, is a 42 kD type II transmembrane protein, also known as ATP1B3. An integral plasma membrane protein, Na/K ATPase is composed of one α and one β subunits. Four isoforms of the α and three isoforms of the β subunits have been reported. Na/K ATPase couples ATP hydrolysis to the development of an ionic gradient by pumping Na and K ions in opposite directions across the cell plasma membrane. It has broad tissue distribution, including all leukocytes and many other tissues
1. Liu C, et al. 2021. Cell. 184(7):1836-1857.e22. PubMed 2. Witkowski M, et al. 2021. Nature. 600:295. PubMed 3. Wagner KI, et al. 2022. Cell Rep. 38:110214. PubMed

Flow sorting antibodies
- PE anti-human CD138 Antibody-DL-101 (1:100) Biolegend 352305 The epitope recognized by MI15 is found within the ectodomain of the CD138 core protein. It has been reported that MI15 blocks the binding of clone B-B4 but not clone DL-101 by flow cytometric analysis. Clones DL-101 and MI15 exhibit differential staining patterns on in vitro generated plasma cells and some CD138 cell lines. 1. Osama MA. 2010. Int. J. Clin. Exp. Pathol. 3:280. (IHC) 2. Itoua MR, et al. 2014. Biomed. Res. Int. 2014:536482. PubMed
- PE anti human CD38 antibody IB6 (1:100) Miltenyi 130-113-427 Clone IB6 recognizes the human CD38 antigen, a single-chain type II trans-membrane glycoprotein with enzymatic activity. It is present on the majority of hematopoietic cells, prevalent during early differentiation and activation processes. Terminally differentiated B cells (plasma cells) express CD38 brightly. Furthermore, CD38 is constitutively expressed in several tissues, for example brain, muscle, and kidney.
Yoon-Sang, K. et al. (2010) Transduction of human primitive repopulating hematopoietic cells with lentiviral vectors pseudotyped with various envelope proteins. Mol. Ther. (7) 18: 1310 - 1317
- APC anti-human CD45 HI30 Antibody (1:200) Biolegend 304011 CD45 is a 180-240 kD single chain type I membrane glycoprotein also known as leukocyte common antigen (LCA) and T200. It is a tyrosine phosphatase expressed on the plasma membrane of all hematopoietic cells, except erythrocytes and platelets. CD45 is a signaling molecule that regulates a variety of cellular processes including cell growth, differentiation, cell cycle, and oncogenic transformation. CD45 plays a critical role in T and B cell antigen receptor-mediated activation by dephosphorylating substrates including p56Lck, p59Fyn, and other Src family kinases. The antibody was purified by affinity chromatography, and conjugated with APC under optimal conditions. Knapp W, et al. 1989. Leucocyte Typing IV. Oxford University Press. New York. 2 Kishihara K, et al. 1993. Cell 74:143. 3. Esser M, et al. 2001. J. Virol. 75:6173. (WB)

4. Yamada T, et al. 2002. J. Biol. Chem. 277:28830.
Nagano M, et al. 2007. Blood 110:151.

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	NCT03290950
Study protocol	study protocol can be found here: https://clinicaltrials.gov/ct2/show/NCT03290950
Data collection	The clinical and correlative study was conducted under the leadership of the study PI Dr. Landgren at Memorial Sloan Kettering Cancer Center in New York, NY. Patients with newly diagnosed multiple myeloma were enrolled between October 1, 2018 and November 15, 2019. As stated above, genomic profiling was based on samples obtained at the time of progression of disease or during an ongoing response, and compared to the pre-treatment baseline samples. The aim was to explore whether pathways leading to emergence of resistance to the drug regimen can be identified. All patients gave written informed consent to the use of their data and to enroll in the trial. Clinical data from the trial can be found in Landgren et al. JAMA Onc 2021: PMID: 33856405.
Outcomes	Primary Objective: The primary objective of the study is to assess the rate of MRD negativity after completion of the combination therapy using multiparametric flow cytometry (sensitivity of 10^-5) Secondary Objectives: 1. To determine safety and tolerability. 2. To evaluate the rates of overall response (partial response (PR) or better), very good partial response (VGPR) or better, and complete response (CR) or stringent CR (sCR). 3. To estimate overall survival (OS) and progression-free survival(PFS). 4. To compare MRD techniques of multi-parametric flow cytometry with next-generation sequencing and mass spectrometry. 5. To create a bone marrow, urine, stool, and peripheral blood sample bank. These samples may be used to later-evaluate biological activity of daratumumab, carfilzomib, lenalidomide, and dexamethasone. Potential analyses include sequencing and gene expression profiling on pre and post therapy bone marrow samples, identification of potential biomarkers (blood, urine, stool, bone marrow aspirates) associated with clinical outcomes. Exploratory Objectives: 1. Genomic profiling (of tumor cells and host immune cells) to explore whether any mutations appear to be associated with response to therapy or toxicity. 2. Genomic profiling will be evaluated using samples at the time of progression of disease or during an ongoing response and will be compared to the pre-treatment baseline samples to explore whether pathways leading to emergence of resistance to the drug regimen can be identified.

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	Frozen bone marrow cells from paired diagnosis and follow-up samples were thawed in warm Roswell Park Memorial Institute 1640 Medium (RPMI, Gibco, USA). They were washed twice in FACS buffer (FACS buffer: PBS+10%FBS+DNAse I), filtered through a 100uL strainer, and incubated with Fc blocker for 10 min, and then with CD38 (PE anti-human CD38 Antibody, Miltenyi, USA, 1:100), CD138 (APC anti-human CD138 Antibody, Biolegend, USA, 1:100), and CD45 (FITC Antibody, Miltenyi, USA, 1:200), and Live/Dead exclusion (DAPI, Invitrogen, USA). Samples were washed twice in FACS buffer and analysed
Instrument	BD FACS-Aria III cell sorter
Software	Diva
Cell population abundance	As this was the negative fraction, the purity was not of interest. Residual plasmacell were present in the single cell analysis. For the positive fraction we checked the purity using WGS data.
Gating strategy	Cells were selected on the forward and side scatter plots and debris excluded. A forward scatter height (FSC-H) vs. forward

Gating strategy

scatter area (FSC-A) density plot was used to exclude doublets. Only double positive CD38+CD138 were considered plasmacells. The rest fell in to the negative fraction.

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