nature portfolio

Corresponding author(s): Jonathan Keats & Sagar Lonial

Last updated by author(s): Feb 2, 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	Clinical data was entered into an eCRF by the individual clinical sites. All patients, visits and sample processing from nucleic acid isolation through sequencing is tracked in a custom LIMS at TGen. All primary analysis is conducted in an automated data analysis pipeline.
Data analysis	BCL2FASTQ (v1.8.4 & 2.17.1), bwa (v0.7.8-r455), SAMtools (v0.1.19-44428cd), GATK (3.1-1-g07a4bf8 & 3.5), Picard (v1.111(1901)), Seurat (v2.6), Strelka (v1.0.13), MuTect (v2.2-25-g2a68eab), snpEFF (v4.2 (build 2015-12-05)), Delly (v0.7.6), STAR (v2.3.1z 01/24/2013), HtSeq (v0.6.0), Salmon (0.7.2), TopHat2 (2.0.8b), SNPsniffer (v5), VEP(v87)/LOFTEE, Matlab2019a R packages: survival (v3.1-8), survminer (v0.4.6), M3C (v3.9), sva (v3.26.0), ConsensusClusterPlus (v1.42.0) https://github.com/tgen/medusaPipe https://github.com/tgen/Post_Medusa_Processing https://github.com/tgen/MMRF_CoMMpass The full reference genome and gene model GTF can be recreated using (https://github.com/tgen/MMRF_CoMMpass/blob/master/fasta_and_gtf_creation/make_ref_genome_and_gtf.sh) GRCh37 fasta:ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz ensembl74 GTF: ftp://ftp.ensembl.org/pub/release-74/gtf/homo_sapiens/Homo_sapiens.GRCh37.74.gtf.gz

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

DNA and RNA sequencing data is available from dbGAP, phs000748, and the Genomic Data Commons. Summarized somatic and clinical data are also available on the MMRF researcher gateway (https://research.themmrf.org/). Clinical data is updated and new molecular data is added with each bi-annual interim release. Supplemental Tables are available on Zenodo (DOI: 10.5281/zenodo.10608273)

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	We only use the term sex in the analysis results and it was only used to confirm the expected over-represenation of males in a normally distributed cohort of myeloma patients and to ensure the molecular data originated from the a person of the expected sex. For certain analysis genes on sex chromosomes were omit to limit sex based clustering.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were self-reported. Neither is used in any analysis presented beyond the cohort description.
Population characteristics	Patients had newly diagnosed, previously untreated multiple myeloma. Patients were recruited from 84 clinical sites located in the United States, Canada, Spain, and Italy. Median age at diagnosis was 63 (range 27-93) with an expected over representation of males (60.4%) versus females (39.6%). This cohort is largely composed of patients from the United States and the distribution of self-reported ancestry reflects US Census Bureau statistics with 80.6% Caucasian, 17.5% Black, and 1.9% Asian.
Recruitment	Patients with a suspected diagnosis or existing diagnosis of multiple myeloma were invited to participate at any of the 84 participating clinical sites. In this observational study enrolled patients must have a confirmed diagnosis of myeloma, had received no previous anti-myeloma therapy, and had to initiate therapy within 30 days with a regimen containing an IMID and/or a proteasome inhibitor. The cohort reflects established age and sex distributions and the race distribution largely reflects the US census.
Ethics oversight	WIRB - Western Institutional Review Board

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	No specific sample size determination was performed in this observational study. For all analysis efforts all patients with data meeting quality control requirements are included
Data exclusions	Samples were excluded from sequencing analysis if they did not meet the specified quality control metrics. Sequencing data were excluded from downstream analyses if they did not meet the specified quality control metrics. Quality control requirements are outlined in detail in the Methods.
Replication	No external data sets were used for replication. For subtype identification all consensus clustering efforts were repeated in triplicate with three independent seeds. The difference and level of concordance is provided in the manuscript.
Randomization	This was an observational clinical study and there was no randomization, therapy was determined by the physician. Patients were required to receive a proteasome inhibitor (PI), and/or immunomodulatory agent (IMiD) as part of the initial treatment regimen. The distribution of patients in all comparisons is expected to be random

Blinding	In this observational trial the site investigators and the analysis team are not blinded to the treatment arms selected. No direct comparison of treatment specific outcomes or response are included in this 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 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	CD38 (BD Biosciences, 340677), CD45/PTPRC (BD Biosciences, 340665), CD138/SDC1 (BD Biosciences, 347205), CD319/SLAMF7 (Invitrogen/eBioscience, 12-2229-42), CD13/ANPEP (BD Biosciences, 340686), CD19 (BD Biosciences, 340720), CD20/MS4A1 (BD Biosciences, 346581), CD27 (BD Biosciences, 654665), CD28 (BD Biosciences, 348047), CD33 (BD Biosciences, 340679), CD52 (Life Technologies, MHCD5204), CD56/NCAM1 (BD Biosciences, 340724), CD117/KIT (BD Biosciences, 340867), FGFR3/CD333 (R&D Systems, FAB766P), Kappa (BD Biosciences, 643774), and Lambda (Life Technologies, MH10614)
Validation	Flow cytometry was conducted in a CAP/CLIA lab as a laboratory developed test at Spectrum Health, Grand Rapids Michigan

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	NCT01454297
Study protocol	Full protocol is available in the related Zenodo repository
Data collection	Patients were enrolled into the study from sites in Canada, Spain, and the United States between July 2011 and June 2015. Sites in Italy collected patients until June 2016
Outcomes	The primary objective was to identify molecular profiles and clinical characteristics that define subsets of multiple myeloma at diagnosis and relapse. These findings of DNA and RNA defined subtypes is presented in this manuscript. Secondary objectives of progression free survival and overall survival are also presented in this manuscript along with the identification of potential targets for myeloma therapeutics.

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	Not available - Conducted in a CAP/CLIA certified laboratory running a laboratory developed test
Instrument	Not available - Conducted in a CAP/CLIA certified laboratory running a laboratory developed test
Software	Not available - Conducted in a CAP/CLIA certified laboratory running a laboratory developed test
Cell population abundance	Not available - Conducted in a CAP/CLIA certified laboratory running a laboratory developed test
Gating strategy	Not available - Conducted in a CAP/CLIA certified laboratory running a laboratory developed test

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