nature portfolio

Corresponding author(s): Anita K. Gandhi

Last updated by author(s): Jun 21, 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	Sequencing data was aligned using BWA-mem and the Sentieon implementation of Mutect2 (tnhaplotyper). Variants were annotated with SnpEff using the dbnsfp database. For WGS, data copy number aberrations were called using Battenberg and structural variants were called by Manta. For WES data copy number aberrations were called using Sclust. Structural variants were not called for the WES data. RNA-seq data was aligned with STAR aligner and quantified with salmon.
Data analysis	Data analysis was conducted in R 3.14. Prism 9.5.0 was used for graphing and statistical analysis. Classifier code is available at https://github.com/mattstokesBMS/SubLymE

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

Sequencing data for the ROBUST clinical trial may be requested via the European Genome Archive under 674 dataset ID EGAS00001007480. REMoDL-B is available on GEO under GSE117556, and GOYA is available under GSE125966.

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 performed analysis in cohorts that were roughly balanced by sex. We tested the association of sex with our high-risk subgroup, but found no significant association in any cohort.
Reporting on race, ethnicity, or other socially relevant groupings	Information on race and ethnicity was collected for some cohorts. The overwhelming majority of annotated patients were white non-Hispanic, so limited sample size prevented analysis by group. Post-hoc analysis showed no significant association between our high-risk subpopulation and race or ethnicity.
Population characteristics	We use clinical trial and observational cohorts of newly diagnosed DLBCL patients with inclusion/exclusion criteria set per trial.
Recruitment	Refer to clinical trial protocols for patient recruitment details.
Ethics oversight	Clinical trial protocols were approved by the FDA and the MER observational study by Mayo Clinic IRB.

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	Molecular data was collected from cohorts that were powered for other purposes (in the clinical trial datasets, to find a treatment effect between treatment arms), but in most cases are sufficiently large to find moderate-to-large effects (HR>1.5) within treatment arms.
Data exclusions	Raw molecular data was filtered for quality metric such as library size, duplication rate, and alignment rate. A further set of samples was excluded after initial clustering, as they were shown to be a highly distinct set of samples just barely passing the pre-set QC thresholds.
Replication	We replicate our key finding in four independent datasets.
Randomization	Clinical trial datasets randomized patients by treatment group.
Blinding	We performed a novel stratification of ndDBLCL populations. The relevant grouping is this stratification itself, for which we could not be blinded.

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: TCF4 (Proteintech, 22337-1-AP), MYC (aAbcam, ab32072), GAPDH (Cell Signaling Technology, 2118L).

Validation: TCF4 and MYC antibodies were validated by shRNA knockdown

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s): ATCC

Authentication: All cell lines were authenticated by fingerprint

Mycoplasma contamination: Free of Mycoplasma contamination tested by Lonza MycoAlert® Detection Kit

Commonly misidentified lines (See ICLAC register): *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals: Mouse: C57/Bl6 eu-myc/hCRBN cross, 4 weeks.

Wild animals: *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex: *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*

Field-collected samples: *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight: Internal BMS IACUC committee approved animal studies

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: ROBUST: NCT02285062; GOYA: NCT01287741; REMoDL-B:NCT01324596; Reddy: EGAS00001002606

Study protocol: *Note where the full trial protocol can be accessed OR if not available, explain why.*

Data collection: *Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

Outcomes: *Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

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.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.
EGAS00001007480.

Files in database submission — Recently-generated data will be uploaded during review

Genome browser session
(e.g. UCSC)
Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates — *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth — 30,001,486 75-bp single-end mapped reads in treatment; 44,436,693 75-bp single-end mapped reads in input control

Antibodies — Ikaros CST(#14859)

Peak calling parameters — macs2 callpeak --keep-dup all --nomodel --qvalue 0.05 --gsize 2749859687 --format BAM --name 03_ikaros_u2932_dmso_rep1.mLb.clN --treatment 03_ikaros_u2932_dmso_rep1.mLb.clN.sorted.bam --control 14_input_u2932_dmso_rep1.mLb.clN.sorted.bam

Data quality — 3,662 peaks using qvalue cutoff = 5.00e-02

Software — R/Bioconductor package ChIPpeakAnno, MACS version 2.2.7.1

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 — Cells were stained with Alexa Fluor 488 conjugated Annexin V (Thermo Fisher, A13201) according to the manual, followed by being incubated with 0.5 µM of TO-PRO-3 Iodide (Thermo Fisher, T3605)

Instrument — BD FACSCanto

Software — FlowJo V10.8

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

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