nature research

Corresponding author(s): Christian Steidl

Last updated by author(s): Dec 16, 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 *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	Mutascope (v1.02), snpEff (v3.6), OncoSNP (v1.3) , GISTIC (v2.0.12) , STAR aligner (STAR_2.5.1b_modified), DESEQ2 package (v1.10.1) , Metafer (v3.11.3), MetaMorph (Molecular Devices), Icy (http://icy.bioimageanalys), ReactomePA (v1.14.4) , R software v3.2.2, CytoScape v3.2.1, TCGA biolink v2.2.10, FlowJo v10.2, GraphPad Prism Version 7
Data analysis	Mutation and copy number were analyzed by Mutascope (v1.02), snpEff (v3.6), OncoSNP (v1.3) , GISTIC (v2.0.12), glmnet in R software v3.2.2, PyClone. RNAseq data were processed by STAR aligner (STAR_2.5.1b_modified), DESEQ2 package (v1.10.1) , ReactomePA (v1.14.4) , CytoScape v3.2.1, Samtools (0.1.19), bwa (0.7.5a). Raw count data from the TCGA project was retrieved using the TCGA biolinks package (v2.2.10) in Bioconductor. Flow cytometry data was analyzed using FlowJo v10.2. Proteomics data was analyzed by PECA in R software v3.2.2 and Proteome Discoverer (2.1). All graphs were generated and analyzed using GraphPad Prism 8, and R software v3.2.2

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

A full data including patient characteristics, mutation data and experimental data were provided in the manuscript as supplementary tables (Supplementary

Tables 2, 3, 8 and 9) and our previous paper (Ennishi et al. J Clin Oncol. 2019 Jan 20;37(3):190-201). The raw sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAS00001002657.

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	The 5% threshold for incidence was based on a calculation of power to detect survival differences in a cohort size of 347 patients at varying hazard ratios and mutational incidences (Supplementary Table 1). Sample sizes for experimental works were determined based on the Authors' experience of what is necessary to generate a convincing and compelling result
Data exclusions	No data were excluded from the analyses.
Replication	We validated the mutation rate of TMEM30A using publicly available dataset. All in vitro experiments were confirmed across multiple cell lines using technical and biological replicates, and all attempts were successful. In vitro and vivo experiments were confirmed in different model systems such as cell line derived xenografts and patient tumor biopsies. Some results were confirmed by different investigators. All data points are represented in the Figures and Extended Figures.
Randomization	Experimental samples, both in vitro and in vivo, were randomly selected for analysis and reported on in their entirety. For mouse model analysis, mice were randomly and without bias selected for treatment by a support technician with no previous knowledge of this project or the expected results.
Blinding	Our study is a retrospective cohort study without blinding

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

1. Antibodies used for immunohistochemstry
Antigen, Antibody Clone, Manufacturer, Catalog number, working dilution, lot number.
CD10, 56C6, Novocastra, NCL-CD10-270, 1:25, 6020664
BCL6, LN22, Novocastra, NCL-L-BCL-6-564, 1:100, 6024089
IRF4, MUM1p, Dako, M7259, 1:1000, 20000312
MYC, Y69, Ventana, 790-4628, neat, E00211
BCL2, 124, Dako, M088701, 1:20, 00095784
2. Antibodies used for immunoblotting and flow cytometry
Antigen, Antibody Clone, Manufacturer, Catalog number, working dilution, lot number.
pan Akt, C67E7, Cell Signaling, 4691S, 1:1000, 20
pan Akt, C67E7, Cell Signaling, 5186S, 1:50, 6
phospho-Akt (Ser473), 193H12, Cell Signaling, 4058S, 1:1000, 30
phospho-Akt (Ser473), 193H12, Cell Signaling, 2337S, 1:50, 24
ERK1/2, unknown, Cell Signaling, 9102S, 1:1000, 26
phospho-ERK1/2, D13.14.4E, Cell Signaling, 9101S, 1:1000, 29
Beta-actin, AC15 Sigma-Aldrich, A5441-100UL, 1:10000, 064M4789V
cleaved Caspase 3, 617S, New England Biolabs, 9661S,1:1000, 45

CD14-APC, 61D3, eBioscience, 17-0149-42, 1:50, 4344410
CD11b-PE, 1CRI44, eBioscience, 12-0118-41, 1:50, 4329537
CD69-PB, FN50, Biolegend, 310920, 1:50, B243908
CD25-APC, BC96, Biolegend, 302610, 1:50, B199705
CD23-PE, EBVCS-5, Biolegend, 338507, 1:50, B199705
CD19-PE-Cy5, HIB19, Biolegend, 302210, 1:50, B142829
CD20-PE, 2H7, Biolegend, 302305, 1:50, unknown
kappa-APC-Cy7, MHK-49, Biolegend, 316522, 1:50, B159771
lambda-FITC, MHL-38, Biolegend, 316606, 1:50, B136011

Validation

All antibodies used in this study are commercially available. Antibodies used in a specific species or application have been appropriately validated by manufacturers for that application and this information is provided on their website and product information datasheets. All antibodies described here have been further optimized for an appropriate concentration by testing several dilutions, and antigen/heat retrieval in citrate and/or EDTA-based buffers (IHC). For those antibodies used in immunoblotting, specificity was validated by detection of the correct molecular weight and non-specific detection was minimized through optimizing concentration and blocking reagent (milk or BSA in 0.5% TBST).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Cell line, source
DOHH-2, DSMZ
Jurkat, by donation from A. Weng, BC Cancer, Vancouver
Karpas422, DSMZ
NU-DUL-1, by donation from M. Dyer, Leicester, UK
A20, ATCC
HEK293-T, Open Biosystems

Authentication

HEK293-T, DOHH-2, NU-DUL-1 and Karpas422 were validated using STR profiling, described below. Jurkat and A20 were not validated by our group during this study.
STR profiling was performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada to authenticate the cell lines.
Cell line authentication was performed using the GenePrint®10 System (Promega Corporation, Madison, WI, USA) using the manufacturer's instructions. Briefly, the kit employs a short tandem repeat (STR) multiplex assay that amplifies nine (9) loci (eight loci recommended by the ANSI Standard (ASN-0002)) and the Amelogenin gender-determining marker in a single PCR amplification. For each reaction, 10 ng of genomic DNA was added to 5 ul GenePrint® 10 5X Master Mix, 5 ul GenePrint® 10 5X Primer Pair Mix and water to a total volume of 25 ul. The cycling conditions were 96 °C for 1 min; followed by 30 cycles of (94 °C for 10 sec, 59 °C for 1 min, 72 °C for 30 sec); 1 cycle of 60 °C for 10 min and 4 °C soak. Samples were run on an AB 3130 Genetic Analyzer using POP7, Promega's internal lane standard 600, dye set F and analyzed in GeneMapper v3.7.

Mycoplasma contamination

All cell lines reported herein have been tested for mycoplasma contamination in-house using the VenorGeM Mycoplasma Detection kit (PCR based), Sigma-Aldrich, and have consistently tested negative.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Species, strain, sex, age
1. Mouse, NSG (NOD scid gamma), female, 6-8 weeks
2. Mouse, Balb/c, female, 8-9 weeks

Wild animals

The study did not involve wild animals.

Field-collected samples

The study did not involve samples collected from the field.

Ethics oversight

All mouse studies were approved by the Institutional Animal Care Committee (IACC) at the University of British Columbia, in accordance with the Canadian Council on Animal Care Guidelines. This study was approved by the Research Ethics Board of the University of British Columbia and the BCCA, conducted in accordance with the Declaration of Helsinki.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics

Patients with de novo DLBCL were included if they met the following criteria: patients had to be 16 years of age or older, treated uniformly with R-CHOP with curative intent at BC Cancer, had complete clinical, laboratory and outcome data available, and had a fresh frozen diagnostic biopsy.

Recruitment	Patients were selected based on the BCCA Lymphoid Cancer database with diffuse large B-cell lymphoma diagnosed between 1985 and 2011. The database cover the all lymphoma patients treated or diagnosed within British Columbia - population based lymphoma registry. One of selection criteria with final cohort of 347 cases was the availability of fresh frozen samples, which may be associated with lower proportion of patients with two or more extranodal sites in the study cohort (Ennishi D et al. Blood 2017). However, the outcomes in the study cohort were not significantly different from the registry-based population (Ennishi D et al. Blood 2017), showing that the nature of the study cohort appears representative of the DLBCL population in BC as a whole.
Ethics oversight	This study was reviewed and approved by the University of British Columbia-BC Cancer Agency Research Ethics Board (H14-02304), in accordance with the Declaration of Helsinki.

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	This was not a clinical trial.
Study protocol	For this retrospective study, inclusion criteria were defined as stated in the manuscript and in this form. Briefly, patients were required to suffer from DLBCL and to have received R-CHOP therapy with curative intent, and were available for fresh frozen samples obtained prior to R-CHOP therapy.
Data collection	Clinical and demographic data, and samples were obtained from the BCCA Lymphoid Cancer database and tissue bank between 2011 and 2016. A CONSORT diagram is included in the manuscript.
Outcomes	We estimated the time to progression (TTP; progression/relapse or death from lymphoma or acute treatment toxicity), progression-free survival (PFS; progression/relapse or death from any cause), disease-specific survival (DSS; death from lymphoma or acute treatment toxicity) and overall survival (OS; death from any cause) for outcome analyses.

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	Flow cytometry was utilized in several unique procedures during this study. Unless indicated, 0.25 to 1 million cells from DLBCL or T cell lymphoma cell cultures were tested in polystyrene flow tubes. For surface marker detection, cells from were collected in polystyrene flow tubes, washed and resuspended in 1% fetal bovine serum (FBS) in PBS, (flow cytometry buffer) and the antibody of interest was added and incubated, in the dark and on ice, for 15 minutes. Unbound antibody was then washed away with flow cytometry buffer, and the cells were resuspended and analyzed in flow cytometry buffer. To detect phosphatidylserine exposure, cells were resuspended in a calcium rich annexin V binding buffer and incubated at room temperature for 10 minutes with annexin V-APC. Following this incubation, annexin V binding buffer was added and the labeled cells were immediately tested by flow. For drug uptake experiments using the naturally fluorescent doxorubicin, cells were similarly washed and resuspended in PBS, and incubated with 1 ug/ml doxorubicin at various times. Primary tumor cells from DLBCL patients, which were stored at -80C, were rapidly thawed at 37C, immediately washed in RPMI1640 plus 20% FBS and DNAseI, then similarly treated with doxorubicin. These cells were then labeled with various surface markers and stains to gate for live B cells and analyzed by flow. Cells were evaluated by mitotic indexing with the following preparation. Cells were treated overnight with the microtubule polymerization inhibitor vincristine. Following this incubation, cells were washed in flow cytometry buffer, and resuspended in 500 uL PBS. Cold 70% ethanol was gradually added to a final volume of 5 mL with vortexing on a low setting. Cells were held on ice for 2 hours, resulting in permeabilization, then washed once in PBS. Cells were stained with propidium staining solution for one hour (0.1% Triton X-100 , 10 ug/ml propidium iodide, 100 ug/ml RNase A in PBS). Singlets were then detected by flow cytometry. Finally, flow cytometry was utilized for an in vitro macrophage engulfment assay, and samples were prepared as follows. In brief, human monocyte-derived macrophages, isolated from fresh, commercial human normal peripheral blood Bcells and primed with human recombinant M-CSF and IFNg were co-cultured with DLBCL cell lines fluorescently labeled on the cell surface with the a pH-tolerant Violet Proliferation Dye, in a 96-cell untreated culture plate. Following a 2-hour period of coculture, wherein phagocytosis was expected to occur, the plate was washed with cold FACS buffer (2 mM EDTA, 2% heat inactivated fetal bovine serum, 0.1% sodium azide in PBS). Human FcR binding inhibitor was added and incubated for 10 minutes at room temperature. Cells were then labeled for 30 minutes on ice with macrophage detection antibodies, targeting CD11b and CD14, and including a live/dead fixable near-IR cell stain. Singly stained macrophage-only controls were included to generate a compensation matrix for this assay. The plate was washed four times with FACS buffer, and the cells were resuspended in a stabilizing fixative solution. Live CD11b/CD14+ macrophages were gated, and tested for the presence of engulfed DLBCL cells by high Violet Proliferation Dye.

Instrument	Make, model number BD FSRFortessa, 64779426 or 64779425 FACSCalibur E1507, E6303 (serial no.)
Software	Flow analysis on the FACSCalibur was conducted using BD CellQuestPro V5.2. Flow analysis on the FSRFortessa was conducted using BD FACS Diva software v6.2. Flow analysis was conducted using FlowJo Version V10.2 software (TreeStar).
Cell population abundance	In this study, flow cytometry was generally used for analysis, by collecting 20,000 events, and not for live cell sorting to enrich a specified population. Those procedures that used live cell sorting, namely enrichment of GFP positive cells following nucleofection of DLBCL cell lines using the CRISPR vector, as least 10,000 cells were collected using this method. Purity was determined visually, by inspection with fluorescent microscopy, and enrichment of the GFP-expressing CRISPR vector was measured by a Surveyor assay, which measures the degree of insertion-deletions in a cell population, informing on the degree of accurate sorting.
Gating strategy	In general, the gating strategy used for all relevant experiments starts with a preliminary FSC (linear)/SSC (log) gate around the live cell population, excluding low FSC/SSC events which are expected to be dead cells or particulates. In the co-culture mixing monocyte-derived macrophages and DLBCL cells, this gate also excluded the DLBCL cell population, which presented with a lower SSC population. Single cells are selected by gating at SSC (height) vs SSC area. Positive and negative boundaries for all stains are gated through the use of positive and negative controls, namely unlabeled cells as negative controls, and singly labeled cells as positive controls. Stain concentrations were selected that show the greatest separation between negative and positive stain, with the lowest detectable non-specific staining (in instances of a mixed culture, such as the macrophage engulfment assay).

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