nature research

Corresponding author(s): Christine Mayr

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	For 3'-seq and RNA-seq analyses, read counts from new samples and publicly available samples (see Data availability statement) were obtained. Information on DNA mutations of cancers other than CLL were obtained from the cbioportal. Information on DNA mutations of CLL samples were obtained from published data and anonymized identities were obtained from Dan A. Landau.
Data analysis	3'-seq and RNA-seq: Peak-calling and quantification were performed using in-house R packages biosignals and TagSeq. These packages depend on the following Bioconductor and CRAN packages: GenomicRanges, Rsamtools, GenomicAlignments, data.table, foreach. The source code of these packages is available in a public repository on BitBucket (https://bitbucket.org/leslielab/apa_2018/). All the GLM based differential analysis was performed using DEXSeq and DESeq. The remaining analyses were done using R version 3.1.2 (2014-10-31). Western blot: Odyssey CLx imaging system (Li-Cor). Northern blot: Fuji phosphorimager. FACS: Flow Jo. Statistics: R and SPSS. DNA sequencing: Vector NTI (Invitrogen) and Chromas (Technelysium Pty Ltd).

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 upon request. 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

All 3'-seq and RNA-seq data generated and analyzed for this study have been deposited in the Gene Expression Omnibus database under accession numbers GSE111310 and GSE111793.
The code to analyze the data is available under https://bitbucket.org/leslielab/apa_2018/ and the processed data are available in Supplementary Table S1 (for Fig. 1b-d, 2a, 4a, Extended Data Fig. 3, and 4) and Supplementary Table S2 (for Extended Data Fig. 8a), and in the Source data files (for Fig. 1e, 2c, 2e, 3a, 3d, 4b-d, 4g, Extended Data Fig. 2c, 6j, 7c, and 8a). Data on DNA mutations from CLL patients were provided by Dan A. Landau (Weill-Cornell Medical College) and need to be requested from him. The mutation data on solid cancers was obtained through the MSK cbio portal. The data can be accessed through www.cbioportal.org.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. CLL samples investigated by 3'-seq were chosen based on availability. The samples needed to be untreated and contain a minimum of 75,000 WBC/ul. We only used fresh samples. We sequenced samples from N = 13 different CLL patients and obtained significantly different IPA isoforms. This indicates that the sample number was sufficient. As controls, we used N = 4 CD5+ normal B cells from lymphatic tissues obtained from different healthy donors. The 3'-seq was confirmed by obtaining concordant results from RNA-seq data from the same samples. To increase the sample size, we included 46 previously published RNA-seq samples for which exome sequencing data and IGVH mutational data were available.
Data exclusions	We did not exclude samples from the analysis.
Replication	3'-seq was performed in a single experiment, but we used normal B cells from N = 4 and CLL samples from N = 13 different donors. The data was reproducible among normal and malignant B cells obtained from different individuals. Furthermore, 3'-seq data were validated by several different methods. Genome-wide IPA isoforms were validated by RNA-seq and other genome-wide methods and 80% (4,456/5,587) were validated. In addition to genome-wide validation, selected truncated mRNAs generated from CLL-IPAs were validated by RT-PCRs (N = 18) and by western blots (N = 13). One CLL-IPA isoform was present, but did not show a difference in expression to normal B cells investigated by RT-PCR and another CLL-IPA isoform could not be validated by western blot analysis. All other tested CLL-IPA isoforms were successfully validated. Several shRNAs per knock-down were used. All experiments were performed as several biological replicates.
Randomization	The experimental groups are defined as cancer vs normal samples.
Blinding	There was no blinding during data collection as the normal and cancer cells are derived from different sources. 3'-seq and RNA-seq libraries from cancer and normal samples were generated at the same time. As the analysis aimed to identify differences between normal and cancer B cells, the data analysis was not performed in a blinded fashion.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

Fresh primary CLL B cells were obtained from peripheral blood of untreated CLL patients. The material is no longer available. The BLCL cell line was made by us and is available upon request.

Antibodies

Antibodies used

For sorting of primary B cell subsets and FACS analysis of B cells and CLL cells, the following antibodies were used:
anti-CD3-PE (mouse, BD Biosciences, 555333, Lot#2317603, 1:100 dilution), anti-CD5-FITC (mouse, BD Biosciences, 555352, Lot#3046601, 1:100 dilution), anti-CD14-PECy7 (mouse, ebioscience, 25-0149-42, Lot#E10278-1635, 1:300 dilution), anti-CD19-APC (mouse, BD Biosciences, 555415, Lot#2347818, 1:100 dilution), anti-CD27-PE (mouse BD Biosciences, 555441, Lot#3051680, 1:50 dilution), anti-CD38-APC (mouse, BD Biosciences, 555462, Lot#3057748, 1:100 dilution), anti-CD38-FITC (mouse, BD Biosciences, 555459, Lot#2289722, 1:50 dilution).

For FACS analysis:
anti-phospho-NF-κB p65 (rabbit, Cell Signaling 3033, Lot#16, used 1:1,500 dilution, validated by citations and western blots in HeLa cells); secondary antibody: Alexa Fluor 647 (goat, Invitrogen, A27040, Lot#1834794, 1:10,000 dilution). Data in this paper: Figure 2c, Extended Data Figure 6d.

For Western blot analysis:
anti-ACTIN (mouse, clone AC-40, Sigma, A4700 , used 1:2,000 dilution; rabbit, Sigma, A2066, Lot#106M4770V, 1:7,000 dilution, validated by protein size prediction and citations, data in this paper: Figure 2a, 4f, Extended Data Figure 4, 5b, 5d, 6b, 6e-g, 8f), anti-GAPDH (goat, V-18, Santa Cruz Biotechnology, Lot#A1316, 1:500 dilution, validated by protein size prediction and citations, data in this paper: Extended Data Figure 6c), anti-AKAP10 (mouse, clone 51, Santa Cruz Biotechnology, sc-136512, Lot#F0410, 1:500 dilution, validated by protein size prediction and citations, data in this paper: Extended Data Figure 4), anti-CARD11 (rabbit, clone 93H1, Cell Signaling, 4440S, Lot#1, 1:1,000 dilution, validated by RT-PCR, protein size prediction, exclusive knockdown of endogenous isoforms in human TMD8 cells and citations, data in this paper: Figure 2a, Extended Data Figure 5d, 6b, c), anti-DICER1 (rabbit, gift from Dr. Witold Filipowicz (FMI Basel), 1:7,000 dilution, validated by the Filipowicz lab, citations, RT-PCR, protein size prediction and endogenously induced isoform expression in HeLa cells, data in this paper: Figure 2a, Extended Data Figure 5d, 6f), anti-DNM1L (mouse, Abcam, ab56788, Lot# GR237898-1, 1:1,000 dilution, validated by protein size prediction and citations, data in this paper: Extended Data Figure 4), anti-MGA (rabbit, H-286, Santa Cruz Biotechnology, sc-382569, Lot#A1516, 1:200 dilution, validated by protein size prediction and ectopic expression of isoforms in HEK293T cells, data in this paper: Figure 2a, Extended Data Figure 6g), anti-SFRS15 (SCAF4; mouse, Abnova, H00057466-B01, Lot#08163WULZ, 1:1,500 dilution, validated by protein size prediction, data in this paper: Extended Data Figure 4, 5d), anti-WSTF (BAZ1B; mouse, clone G-5, Santa Cruz Biotechnology, sc-514287, Lot#B0415, 1:500 dilution, validated by protein size prediction and citations, data in this paper: Extended Data Figure 4), anti-NUP98 (rabbit, Novus Biologicals, NB100-93325, Lot#A1, 1:2,000 dilution, validated by RT-PCR, protein size prediction, endogenously induced isoform expression in A549 cells (confirmed by RT-PCR and Western blot) and citations, data in this paper: Extended Data Figure 4, 5b), anti-SGK223 (mouse, clone A-6, Santa Cruz, sc-398164, Lot#G1715, 1:100 dilution, validated by RT-PCR, protein size prediction and endogenously induced isoform expression in HEK293 cells (confirmed by RTPCR), data in this paper: Extended Data Figure 4), anti-SENP1 (rabbit, Bethyl Labs, A302-927A-T, Lot#A302-927A-T-1, 1:1,000 dilution, validated by RT-PCR and protein size prediction, data in this paper: Extended Data Figure 4), anti-CUL3 (rabbit, Bethyl Labs, A301-108A-T, Lot#A301-108A-T-1, 1:1,000 dilution, validated by RT-PCR and protein size prediction, data in this paper: Extended Data Figure 4), anti-PAWR (rabbit, Abcam ab92590, clone EPR3991, 1:5,000 dilution, validated by RT-PCR and protein size prediction, data in this paper: Extended Data Figure 4), anti-RIPK1 (rabbit, Cell Signaling 4926, Lot#2, 1:1,000 dilution, validated by RT-PCR, protein size prediction and citations, data in this paper: Extended Data Figure 4), anti-V5 tag antibody (mouse, Invitrogen R960-25, Lot#1949337, 1:2,500 dilution, validated by citations, data in this paper: Extended Data Figure 6e), anti-GFP (chicken, Abcam ab13970, Lot#GR236651-17, 1:2,000 dilution, validated by citations, data in this paper: Extended Data Figure 6g), anti-WNT5a/b (rabbit, clone C27E8, Cell Signaling 2530, Lot#4, 1:1,000 dilution, validated by citations, data in this paper: Figure 4f, Extended Data Figure 8f).

The secondary antibodies used included anti-mouse IRDye 700 (donkey, Rockland Immunochemicals, 610-730-002), anti-rabbit IRDye 680 (donkey, Li-Cor Biosciences, 926-68073), anti-rabbit IRDye 800 (donkey, Li-Cor Biosciences, 926-32213), and anti-mouse IRDye 800 (donkey, Li-Cor Biosciences, 926-32212).

For protein immunoprecipitation:
anti-V5 tag antibody (mouse, Invitrogen R960-25, Lot#1949337, validated by citations, data in this paper: Extended Data Figure

	6e).
Validation	Validations of the antibodies are indicated above.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	B lymphoblastoid cells (BLCL) are Epstein Barr virus-immortalized human blood B cells. They were immortalized by us and were described before (Lianoglou et al., Genes Dev 2013). MEC1 cells are malignant B cells from B-Prolymphocytic leukemia and were provided by Dr. Abdel-Wahab (MSKCC). They are also available at ATCC. Raji and TMD8 cells are malignant B cells from lymphomas and were a gift from Dr. Hans-Guido Wendel (MSKCC) and are also available from ATCC. HEK293 and HEK293T cells (embryonic kidney), HeLa cells (cervical cancer) and A549 cells (lung adenocarcinoma) were purchased from ATCC. The parental and DICER KO HCT116 cells were published and were provided by V. Narry Kim (Seoul National University).
Authentication	None of the cell lines used have been authenticated.
Mycoplasma contamination	All cell lines were tested for mycoplasma contamination and found to be negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

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	Human TMD8 cells were fixed with 4% formaldehyde at room temperature for 15 mins. After two washes with excess PBS, fixed cells were resuspended with ice-cold PBS and permeablized with 90% methanol for 20 mins on ice. Cells were then washed with cold PBS twice and resuspended with the incubation buffer (PBS + 0.5% BSA). Cells were aliquoted and incubated with anti-phospho-NF-κB p65 (1:1500 dilution, Cell signaling #3033) for 1.5 hrs at room temperature. Cells were washed with the incubation buffer twice and incubated with the fluorochrome-conjugated secondary antibody solution (1:10,000 Alexa Fluor 647 A27040, Invitrogen) for 15 mins at room temperature. After two washes with the incubation buffer, cells were analyzed using a FACS Calibur.
Instrument	Becton Dickinson FACSCalibur Flow Cytometer
Software	FlowJo
Cell population abundance	GFP-positive cells were used for the analysis. The cell numbers of analyzed cells were at least 2,500 up to 7,000 cells, with the percentage range of 16.3-44% of total live cells.
Gating strategy	FSC-H (>200) and SSC-H (50-400) were set to gate live cells. By using the unstained cell control, GFP (FL1-H)-positive cells were gated at FL1-H > 60. All the GFP-positive cells were analyzed for FL4-H intensity.

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