

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. [For final submission](#): please carefully check your responses for accuracy; you will not be able to make changes later.

► Experimental design

1. Sample size

Describe how sample size was determined.

We did not predetermine a samples size as and a power calculation could not be performed, as this is the first study of its kind. Sample size was determined based on availability of samples with sufficient cellular material and annotated clinical metadata. For relapse prediction models, samples included in this analysis required at least 3 years of follow up from initial diagnosis.

2. Data exclusions

Describe any data exclusions.

Patient samples that did not have at least 3 years of follow up data from the time of diagnosis were excluded from the DDPN relapse prediction modeling. These comprised UPN 60-69, as shown in Table S1 . These samples were included in the developmental classification.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

The patient samples utilized in this work were analyzed in a series of batches. Each sample was divided into 5 treatment conditions as described in Table S3. Thus, each sample was replicated 5 times and responses across patients were reproducible as well as across batches of samples to enable generalization regarding per patient responses and across patient responses.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

During the cross-validation construction of the DDPN model, samples were randomly assigned to training and validation cohorts using a random assignment in R that preserves proportions of cases vs. controls.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded to group allocation.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ A statement indicating how many times each experiment was replicated
- ☐ ☒ 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 any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ Test values indicating whether an effect is present
*Provide confidence intervals or give results of significance tests (e.g. *P* values) as exact values whenever appropriate and with effect sizes noted.*
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

All analyses were performed using Cytobank (cytobank.org) and R. Single-cell protein expression data were extracted using Bioconductor software (www.bioconductor.org). DREMI is available for download at www.c2b2.columbia.edu/danapeerlab/html/dremi.html. The developmental classifier and DDPR code is available to editors and reviewers and will be made available upon request. FCS files are available on Github (<https://github.com/kara-davis-lab/DDPR/releases>).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

Materials, outside of patient samples, utilized in this study are not unique and are available for distribution by a third party. All antibodies are commercially available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

All antibodies were validated in human cells (cell lines or primary cells) known to our laboratory to be positive or negative controls for a given antibody target. Mass cytometry antibody reagents listed below in the following order:

Protein/ Clone/ Lot Number/ Manufacturer/ Metal Isotope/ Surface (S) or Intracellular(I) Stain

CD10/ HI10a/ 6155527/ Biolegend/ Gd156/ S
 CD123/ 6H6/ B199259/ Biolegend/ Eu151/ S
 CD127/ HCD127/ B173990/ Biolegend/ Dy162/ S
 CD16/ 3G8/ B175991/ Biolegend/ Yb171/ S
 CD179a/ HSL96/ B129864/ Biolegend/ Sm149/ I
 CD179b/ HSL11/ B179047/ Biolegend/ Gd158/ I
 CD19/ H1B19/ B157781/ Biolegend/ Nd142/ S
 CD20/ 2H7/ B164952/ Biolegend/ Sm147/ S
 CD22/ HIB22/ B165323/ Biolegend/ Nd143/ S
 CD235/ HIR2/ B132247/ Biolegend/ In113/ S
 CD24/ ML5/ B167884/ Biolegend/ Gd160/ S
 CD3/ HIT3a/ B151232/ Biolegend/ Er170/ S
 CD33/ HIM3-4/ B183522/ Biolegend/ Yb171/ S
 CD34/ 8G12/ B163230/ Biolegend/ Nd148/ S
 CD38/ HIT2/ B170151/ Biolegend/Er168/ S
 CD43/ CD43-10G7/ B149905/ Biolegend/ Er167/ S
 CD45/ HI30/ B159992/ Biolegend/ In115/ S
 CD58/ TS2-9/ B145718/ Biolegend/ Tm169/ S
 CD61/ VI-PL2/ B176028/ BD Biosciences/ In113/ S
 CD79b/ CB3-1/ 4203934/ Biolegend/ Nd146/ S
 HLA-DR/ L243 B161762/ Biolegend/ Yb174/ S
 IgHi/ polyclonal/ 10689/ Invitrogen/ Eu153/ I
 IgHs/ polyclonal/ 10689/ Invitrogen/ Lu175/ S
 IKAROS/ D10E5 2/ Cell Signaling Technology/ Nd145/ I
 IgL kappa/ MHK-49/ B162243/ Biolegend/ Sm154/ I
 IgL lambda/ MHL-38/ B171739/ Biolegend/ SM154/ I
 PAX5/ 1H9/ B178991/ eBioscience/ Ho165/ I
 RAG1/ D36B3/ 3968BF/ Cell Signaling Technology/ Dy163/ I
 TdT/ E17-1519/ 21361/ BD Biosciences/ Dy164/ I
 CRLF2/ 1B4/ E028811/ eBioscience/ Dy161/ S
 AKT (pS473)/ 193H12/ 4060BF/ Cell Signaling Technology/ Tb159/ I
 4EBP1(pT37/T46)/ 236B4/ 18/ Cell Signaling Technology/ Nd144/ I
 cPARP/ F21-852/ 5089576/ BD Biosciences/ La139/ I
 CREB (pS133)/ 87G3/ 9198BF/ Cell Signaling Technology/ Yb176/ I
 ERK1/2 (pT202/pY204)/ D13/ 4370BF/ Cell Signaling Technology/ Yb173/ I
 Ki-67/ B56/ 3305519/ BD Biosciences/ Sm152/ I
 IKAROS (pS63)/ STA9/ 1/ Epitomics (made for Nolan lab)/ Gd155/ I
 PLCg2 (pY759)/ K86-689.37/ 26057/ BD Biosciences/ Pr141/ I
 rpS6 (pS235/pS236)/ N7-548/ 4044686/ BD Biosciences/ Yb172/ I
 STAT5 (pY694)/ 47/ 4044688/ BD Biosciences/ Nd150/ I
 ZAP70/SYK (pY319/pY352)/ 17a/ 85582/ BD Biosciences/ Er166/ I

10. Eukaryotic cell lines

- State the source of each eukaryotic cell line used.
- Describe the method of cell line authentication used.
- Report whether the cell lines were tested for mycoplasma contamination.
- If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines were used in the study.

No eukaryotic cell lines were used in the study.

No eukaryotic cell lines were used in the study.

No eukaryotic cell lines were used in the study.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

No animals were used in the study.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Clinical data were available for the samples used in this study, including minimal residual disease (MRD) risk group and final risk assignment as per AIEOP-BFM ALL 2000 protocol (ClinicalTrials.gov identifier: NCT00613457), diagnostic cytogenetics, age at diagnosis, gender, white blood cell count at diagnosis, date of diagnosis, date of relapse, and date of the last follow-up.