nature research

Corresponding author(s): NMED-NT97113
Bendall, Sean C.

Last updated by author(s): Dec 5, 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	Clinical flow cytometry acquisition was performed on BD FACSDiva 8.01. Mass cytometry acquisition was performed on DVS Sciences Instrument Control Software 6.0.626.
Data analysis	Clinical flow cytometry analysis was performed on FCS Express 4.08.0016. Mass cytometry analysis was performed on Cytobank (cytobank.org). Further data analysis was performed with custom code written for the R statistical package 3.5.1.

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

Data availability. Flow cytometry and normalized singlet mass cytometry FCS data are openly available as below. Mass cytometry data without normalization or doublet/debris removal are available upon request.
http://flowrepository.org/id/FR-FCM-Z2DM : 54 initial samples, CyTOF
http://flowrepository.org/id/FR-FCM-Z2D3 : 11 in-depth samples, flow cytometry
http://flowrepository.org/id/FR-FCM-Z2DX : 11 in-depth samples, CyTOF, MM axes
http://flowrepository.org/id/FR-FCM-Z2E6 : 11 in-depth samples, CyTOF, MD axes
http://flowrepository.org/id/FR-FCM-Z2E7 : 17 myeloid blast samples, CyTOF, BD axes

Processed flow and mass cytometry source data plots for the 11 in-depth samples are openly available at the Stanford Digital Repository, https://purl.stanford.edu/dq578cx2728

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	A specific sample size was not chosen, but we tested samples until key antibody reagent was expended. Because we aimed to create a general-purpose diagnostic test where the diagnosis is not known beforehand, it was more important to compare and contrast the major cytometrically-diagnosable classes of the WHO hematopoietic tumor classification than to focus on any specific entity. Furthermore, many of the individual entities are genetic subtypes -- testing for which results are only available after the specimen has degraded too much for cytometry. Some are so rare it would take a decade or more just to obtain five. However, hematopoietic tumors are physically only composed of a handful of cell types. For example, roughly a quarter of the classification is composed of acute leukemias, but for the most part they are morphologically and cytometrically just various flavors of blasts, with some differentiating into granulocytes, monocytes, etc. We could validate the principles of our method by sampling the range of the blast spectrum -- T-ALL, B-ALL, monomorphous CD34+ AML, differentiating AML, MPAL, etc. -- rather than every specific entity. Mature cells, e.g. neoplastic and non-neoplastic mature lymphocytes, showed largely consistent behavior and had less morphometric range than blasts. Furthermore, the main difference between flow and mass cytometry is side scatter (SSC). Across all our samples, we could achieve similar results as SSC using VAMP-7, thus anything done by flow cytometry could also be done by mass cytometry.
Data exclusions	No samples were excluded.
Replication	In order to save reagent and test the broadest range of samples, individual clinical samples in our test set were not re-run. Within the study, individual cell populations have multiple biological replicates run on different days, as described in the main text. For the most common cell populations, neutrophils: n=53, blasts: n=39, T and NK cells: n=54, monocytes: n=44, B cells: n=51. Furthermore, in preliminary/pilot work, the most common technical and biological replicate was normal marrow, tested extensively during titration, which behaved consistently.
Randomization	This study did not involve an intervention with control and test groups. It is an observational study involving comparisons between technologies and samples.
Blinding	This study did not involve an intervention with control and test groups. Flow cytometry and mass cytometry data are visually distinct and not blindable.

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	Mass cytometry antibodies are listed in Supplementary Table 1 and in the attached table. Clinical flow cytometry antibodies and their fluors are listed alongside plots in Supplementary Fig. 5 and the source data plots.
Validation	All antibodies were validated on human cells. Clinical flow cytometry antibodies were purchased as in vitro diagnostic (IVD) or analyte-specific reagents (ASR) from BD and used according to manufacturer's instructions per standard operating procedures in a CAP- and CLIA-certified diagnostic clinical laboratory. Antibodies past the manufacturer's expiration date are discarded per protocol. In addition to validation performed by the manufacturer, at least once a month every lot of each clinical flow cytometry antibody is checked for appropriate positive and negative staining on clinical patient samples.

Mass cytometry antibody control populations are provided in Supplementary Table 1. Scatterbodies (VAMP-7, serpin B1, MPO, lysozyme, lactoferrin, lamin A/C, lamin B1, rRNA, HP1b, beta actin, WGA) were titrated to maximally separate the major cell populations in human bone marrow. Their characterization is the subject of the current study, with their behavior on normal bone marrow populations shown in Figs. 1d and 2, and on clinical sample populations in Figs. 3, 4, and 5c, Supplementary Figs. 2-5 and 7-8, and in the source data plots. Mass cytometry antibodies to standard cluster of differentiation (CD) antigens were titrated and validated on human bone marrow and peripheral blood. As the current study itself includes a validation, extensive and direct comparison of CD2, CD3, CD4, CD5, CD7, CD8, CD10, CD11c, CD13, CD14, CD15, CD16, CD19, CD20, CD22, CD23, CD33, CD34, CD38, CD45, CD56, CD64, CD79a, CD117, HLA-DR, IgK, IgL, and MPO to clinical flow cytometry are provided as source data plots, with full FCS files available online at flowrepository.org, examples in Figure 3g, and statistical analysis performed in Figure 3h. Technical details of panel design are also discussed in the Supplementary text. CD61 was titrated on platelets and used as a clean-up step as shown in the Supplementary text and gating. CD71 and CD235ab were titrated on bone marrow nucleated erythroid precursors, and demonstrated known patterns of reactivity (Supplementary Fig. 4). CD123 was titrated on bone marrow basophils, with a well-known, characteristic CD123+ HLA-DR- pattern. CD298 (Na+/K+ ATPase) and beta-2-microglobulin are present on all human cells and were titrated and characterized in Hartmann et al. (2018).

Human research participants

Policy information about studies involving human research participants

Population characteristics	While some hematopoietic cancers are more common in certain age groups or genders, this study did not involve any diagnoses specific to age, gender, or race, or draw any conclusions based on these patient characteristics. Samples included patients from infants to the elderly, with 34 samples from female patients and 39 from male patients.
Recruitment	Given that we have no control over the patient samples submitted for diagnostic testing, we chose from the available post-diagnostic excess clinical samples to broadly represent the major classes from the WHO hematopoietic tumor diagnostic manual, enriching for more difficult cases. Both common and uncommon entities and phenotypes were selected in order to test the robustness of our platform -- rather than only selecting easy or common samples.
Ethics oversight	Samples were collected under Stanford IRB-30899 and IRB-40765, and only used post-diagnostic excess material prior to disposal, with no change to patient management.

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

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	Sample preparation methods are described in the Methods and Supplementary Text.
Instrument	Flow cytometry was performed on BD FACS Canto II instruments. Mass cytometry was performed on a DVS Sciences CyTOF 2 instrument.
Software	Clinical flow cytometry acquisition was performed on BD FACSDiva 8.01. Mass cytometry acquisition was performed on DVS Sciences Instrument Control Software 6.0.626. Clinical flow cytometry analysis was performed on FCS Express 4.08.0016. Mass cytometry analysis was performed on Cytobank (cytobank.org). Further data analysis was performed with custom code written for the R statistical package 3.5.1.
Cell population abundance	No cell sorting was performed.
Gating strategy	Details of gating and analysis are provided in the Methods and Suppelmentary Text.

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