nature research

Corresponding author(s): Luca Vago

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	SNP array and gene expression data were generated by the Illumina iScan Reader (Illumina, San Diego, CA, USA). The RNA-seq libraries were sequenced using the Illumina Next-Seq 500 high platform (Illumina, San Diego, CA, USA). Leukemia samples were FACS-purified using a MoFloTM XDP (Beckman Coulter, Brea, CA, USA) or a FACS Diva (BD Biosciences, Franklin Lakes, NJ, USA) cell sorter. Total RNA and DNA fragments were measured and checked for quality on a Agilent Bioanalyser chip or an D1000 Screen Tape System (both from Agilent, Santa Clara, CA, USA) Spots of the ELIspot assays were acquired by the KS Elispot Reader (Zeiss, Oberkochen, Germany). Quantitative PCR data were generated and analyzed by the ABI7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Analysis of human samples was performed using a Canto II flow cytometer equipped with 405nm, 488nm, and 633nm lasers, or using a LSR Fortessa flow cytometer equipped with 355nm, 405nm, 488nm, 561nm and 640nm lasers (both from BD Biosciences, Franklin Lakes, NJ, USA), both using the DB FacsDIVA Software version 7 (BD Biosciences, Franklin Lakes, NJ, USA). Analysis of mouse samples was performed using a Gallios flow cytometer equipped with 488nm, 638nm and 405nm lasers (Beckman Coulter, Brea, CA, USA) using the Summit software (Beckman Coulter, Brea, CA, USA). Cytofluorimetric data were next processed using FCS 4 Express (De Novo Software, Los Angeles, CA), FlowJo version 9.8.5 (Tree Star, Ashland, OR, USA), or Kaluza (Beckman Coulter, Brea, CA, USA).
Data analysis	For SNP array analysis GenomeStudio version 1.9.4 (Illumina, San Diego, CA, USA), Nexus Copy Number 5.0 (BioDiscovery, El Segundo, CA, USA) and DNAcopy (Olshen, A.B et al., Biostatistics, 2004) were used. - Gene expression array data were analyzed using the Illumina GenomeStudio software (Illumina, San Diego, CA, USA), the R software taking advantage of different Bioconductor packages (https://www.bioconductor.org/): in particular LIMMA (Smyth, G.K., at al., Bioinformatics, 2005) and GoSEMSim v2.12 (Yu, G., Bioinfermatics,

2010), the online interface DAVID version 6.7 (Huang D., Nat Protoc., 2009) and the Cluego Cytoscape package (Bindea, G., Bioinformatics, 2009). - The RNA-seq libraries were pseudoaligned to gencode v28 transcripts (Harrow, J., Genome Res., 2012) using kallisto v 0.44.0 (Bray, N., Nat. Biotechnol., 2016). The abundancies were summarized to genes using txImport package (Soneson, C., F1000Res, 2015) and analyzed using edgeR (Robinson, M.D., Bioinformatics, 2010). - Cytokine quantification was performed using LEGENDplex v7.0 (BioLegend, San Diego, CA, USA). All the statistical analysis were performed using GraphPad Prism Version 5.0 or 7.0a (La Jolla, CA, USA). - High-dimensional analysis of flow cytometry was performed using the BH-SNE dimensionality reduction algorithm, running on the MatLab plug-in "cyt" developed by Dana Pe'er's lab (Amir, E.D., Nat. Biotechnol., 2013)

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

SNP array, microarray and RNA-Seq data generated and analysed during the current study are available through ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) with accession numbers E-MTAB-7631, E-MTAB-7628, E-MTAB-7630 and E-MTAB-7456.
All the other relevant data generated and/or analysed in the current study are included in the manuscript or in supplementary information.

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

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	All patient samples coming from the Biobank of the different centres had been collected upon specific written consent in agreement with the Declaration of Helsinki. Selection of samples of interest was operated on the basis of availability of viable samples containing at least 5% leukemic blasts pairwise collected before and after allogeneic HSCT and no evidence of genomic HLA loss at relapse. For the high-throughput analyses also the expected yield of an adequate amount of purified leukemic blasts upon sorting was considered a criterion for sample choice. All samples meeting these requirements were tested, and sample size was determined by their availabilityand considered adequate upon analysis of the interindividual variability in the variables of interest. The study was approved by the scientific and ethic committee of the San Raffaele Hospital Scientific Institute (Milan, IT). For in vivo experiments, number of animals was selected based on variability observed in pilot experiments and on primary leukemia cells availability. All in vivo experiments performed were approved by the San Raffaele Animal Care and Use Committee (IACUC), by the San Raffaele Ethic Committe and by the Italian Ministry of Health.
Data exclusions	No data were excluded.
Replication	For gene expression profiling, leukemic cell purification and microarray profiling was for each sample performed in triplicate whenever possible. For all other experiments performed using patient-derived samples, each sample was tested in a single experiment, with an appropriate number of experimental replicates within the experiments. All attempts at replication of presented data were successful.
Randomization	Not applicable to our study design (retrospective observational).
Blinding	Analysis of biological features was blinded to the clinical data, which were linked for correlation analysis only in a later stage.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Research animals
☐	☒ Human research participants

Unique materials

Obtaining unique materials

For primary leukemia and PDX samples used in this study material restrictions depends on sample availability and are ruled by the San Raffaele Scientific Institute. All the samples used had been collected upon specific written consent under a specific protocol approved be the scientific and ethic committee of the San Raffaele Hospital Scientific Institute (Milan, IT)

Antibodies

Antibodies used

Here below the list of anti-human fluorochrome-conjugated antibodies:
CD34 PerCP-Cy5.5 (BD Biosciences, Clone 8G12, Catalog. N° 347222, Lot N°4092951, 1:30)
CD34 PE (BD Biosciences, Clone 8G12, Catalog. N° 345802, Lot N°77166, 1:50)
CD34 APC (BD Biosciences, Clone 8G12, Catalog. N° 345804, Lot N°6229872, 1:50)
CD34 FITC (BD Biosciences, Clone 8G12, Catalog. N° 345801, Lot N°42362, 1:50)
CD33 PerCP-Cy5.5 (BD Biosciences, Clone P67.6, Catalog. N° 333146, Lot N°37722)
CD117 PerCP-Cy5.5 (BD Biosciences, Clone YB5.B8, Catalog. N° 562094, Lot N°4002885, 1:30)
CD86 FITC (BD Biosciences, Clone 2331, Catalog. N° 555657, Lot N°3338536, 1:50)
HLA-DR FITC (BD Biosciences, Clone L243, Catalog. N° 347400, Lot N°6180996, 1:100)
HLA-DR PE-CF594 (BD Biosciences, Clone G46-6, Catalog. N°562304, Lot N°7191553, 1:100)
DNAM-1 FITC (BD Biosciences, Clone DX11, Catalog. N° 559788, Lot N°4287752, 1:50)
CD137 PE (BD Biosciences, Clone 4B4-1, Catalog. N° 555956, Lot N°33587, 1:100)
CD45 BUV395 (BD Biosciences, Clone HI30, Catalog. N°563792, Lot N°7298798, 1:100)
CD38 BUV737 (BD Biosciences, Clone HB7, Catalog. N°564686, Lot N°7352565, 1:100)
CD14 BV510 (BD Biosciences, Clone M□P9, Catalog. N°563079, Lot N°7341719, 1:30)
CD10 BV786 (BD Biosciences, Clone HI10a, Catalog. N°564960, Lot N°8045564, 1:50)
CD200 PE-Cy7 (BD Biosciences, Clone MRC OX-104, Catalog N°562125, Lot N°5023724, 1:50)
CTLA-4 APC (BD Biosciences, Clone BNI3, Catalog. N°555855, Lot N°4066549, 1:50)
CD45 PC7 (BioLegend, Clone HI30, Catalog. N° 304016, Lot N°B229089, 1:100)
CD45 Pacific Blue (BioLegend, Clone HI30, Catalog. N° 304029, Lot N°B180440, 1:100)
CD3 FITC (BioLegend, Clone SK7, Catalog. N° 344804, Lot N°B231398, 1:100)
CD3 APC-H7 (BioLegend, Clone SK7, Catalog. N° 344818, Lot N°B206985, 1:100)
CD3 BV605 (BioLegend, Clone OKT3, Catalog N°3173322, Lot B24061, 1:100)
CD11a APC (BioLegend, Clone HI111, Catalog. N° 301212, Lot N°B170015, 1:100)
HLA-ABC Pacific Blue (BioLegend, Clone W6/32, Catalog. N° 311418, Lot N°B191431, 1:200)
GAL-9 APC (BioLegend, Clone 9M1-3, Catalog. N° 348908, Lot N°B188210, 1:50)
PVRL2 PE (BioLegend, Clone TX31, Catalog. N° 337410, Lot N°B177553, 1:50)
OX40 APC (BioLegend, Clone ACT35, Catalog. N° 350008, Lot N°B161270, 1:50)
PD-1 FITC (BioLegend, Clone EH12.2H7, Catalog. N° 329904, Lot N°B169871, 1:100)
CD28 Pacific Blue (BioLegend, Clone CD28.2, Catalog. N° 302928, Lot N°B147378, 1:100)
ICOS PB (BioLegend, Clone C398.4A, Catalog. N° 313522, Lot N°B170453, 1:50)
CD4 APC-Cy7 (BioLegend, Clone OKT4, Catalog. N° 317418, Lot N°B171483, 1:100)
CD4 Pacific Blue (BioLegend, Clone OKT4, Catalog. N° 317429, Lot N°B195687, 1:100)
CD4 FITC (BioLegend, Clone SK3, Catalog. N° 344604, Lot N°B205589, 1:100)
CD8 APC-Cy7 (BioLegend, Clone SK1, Catalog. N° 344714, Lot N°B233288, 1:100)
CD14 FITC (BioLegend, Clone M5E2, Catalog. N° 301804, Lot N°B163436, 1:100)
CD33 PerCP-Cy5.5 (BioLegend, Clone WM53, Catalog. N°303414, Lot B261779, 1:30)
PD-L1 PE-Cy7 (BioLegend, Clone 29E.2A3, Catalog. N°329718, Lot N°B227182, 1:50)
CD45RA Alexa700 (BioLegend, Clone HI100, Catalog. N°304120, Lot N°B227350, 1:50)
CD16 APC-Cy7 (BioLegend, Clone 3G8, Catalog. N°302018, Lot N°B260177, 1:50)
ICAM-1 PE (BioLegend, Clone HA58, Catalog. N°353106, Lot N°B169127, 1:50)
ICOSL PE (Thermo Fisher Scientific, Clone MIH12, Catalog. N° 12-5889-42, Lot N°E14114-103, 1:50)
PD-L1 PE (Thermo Fisher Scientific, Clone MIH1, Catalog N°12-5983-42, Lot N°E13991-104, 1:50)
PD-L2 APC (Thermo Fisher Scientific, Clone MIH18, Catalog. N° 17-5888-42, Lot N°E17310-101, 1:50)
TIGIT PE-Cy7 (Thermo Fisher Scientific, Clone MBSA43, Catalog. N° 25-9500-42, Lot N°E25046-101, 1:50)
B7-H3 APC (R&D Systems, Clone 85504, Catalog. N° FAB1027A, Lot N°AAPJ0213061, 1:50)
OX40L PE (R&D Systems, Clone 159403, Catalog. N° FAB10541P, Lot N°ABJJ0313111, 1:50)
VISTA FITC (R&D Systems, Clone 730804, Catalog. N° FAB71261G, Lot N°ADJM0113091, 1:50)
HLA-DP PE (Leinco Technologies, Clone B7/21, Catalog. N° H130, Lot N°0313L280, 1:200)
PVR FITC (Lifespan Biosciences, Clone TX21, Catalog. N° LS-C179432, Lot N°66802, 1:50)
CD80 PE (Invitrogen, Clone MEM-233, Catalog. N° MHCD8004, Lot N°1369609B, 1:50)
CD34 Vioblue (Miltenyi Biotech, Clone AC136, Catalog. N°130-095-393, Lot N°5150415145, 1:50)

The only anti-mouse antibody used is the pan CD45 PerCp5.5 (Clone 30-F11, Catalog. N° 103132, Lot N° B199699, 1:200) from BioLegend, (San Diego, CA, USA), utilized only for the in vivo experiments.

Validation

All the antibodies used in the flow cytometry experiments were from commercial vendors and they were validated for specificity to original targets by the manufacturers. The Certificate of Analysis is available from the manufacturers. All these commercially-available antibodies were also tested on cell lines reported to be positive for the markers of interest and titrated for optimal on target/off target activity.

Research animals

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

Animals/animal-derived materials

For in vivo experiments four week-old male non-irradiated immunodeficient NOD-SCID interleukin (IL)-2Rgamma(null) (NSG) mice were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics

In this retrospective study, we selected based on material availability a discovery cohort (n=40) and a validation cohort (n=36) of patients with diagnosis of de novo or secondary AML who experienced non-HLA loss disease relapse after allo-HCT, and for whom paired pre- and post-transplant viable leukemic samples had been collected upon specific written informed consent, in agreement with the Declaration of Helsinki. Discovery cohort samples came from the San Raffaele Leukemia Biobank (Protocol "ALLO-RELAPSE" approved by the San Raffaele Ethic Committee on 03/11/17 and Protocol "Banca Neoplasie Ematologiche" approved by the San Raffaele Ethic Committee on 10/05/10, latest amendment on 06/14/12) and the Northern Italy Leukemia Group (NILG) BioBank (protocol NILG AML 02/06, NCT00495287). Validation cohort samples came from the San Raffaele Leukemia Biobank, the Universitätsklinikum Essen Biobank, the Universitätsklinikum Freiburg Biobank, the Hokkaido University Biobank, the Leiden University Medical Center Biobank, the Johns Hopkins University Biobank and the Institut Paoli-Calmettes Biobank.
All the covariate-relevant population characteristics of the human cohorts here analyzed are listed in Supplementary Table 1. Supplementary Table 2 and 5 provide a summary of the analyses conducted for each sample.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

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

FACS-Sorting: Samples containing leukemia cells were thawed in and kept overnight at 37 °C in X-Vivo-15 medium supplemented with 5% human serum (complete medium) at a concentration of 1-2x10e6 cells/ml. Leukemia blasts were labeled using the appropriate mixture of antibodies in 100 µl of complete medium and kept for 20 min at 4 °C. Before sorting, cells were washed with 10 ml of complete medium and filtered with round bottom tube with cell a strainer cap.

Immunophenotipic analysis: Samples were thawed in complete medium. A total of 0.2-1 x 10e6 cells were washed in 1XPBS, 2% FBS and then labeled with the appropriate antibodies mix in 50 µl 1XPBS, 2% FBS in a FACS tube. Staining was performed at 4 °C for 15-20 min and washed with 1 ml of 1XPBS, 2% FBS.

Mice samples: In total 50 µl mice peripheral blood, previously treated with heparin, were stained with the relevant mixture of antibodies with the addition of 100 µl of 1X PBS, 2% FBS, for 20 min at 4 °C. The erythrocytes were eliminated from the samples by incubating with 3-5 ml ACK lysis buffer for 4 min at RT. Cells were pelleted by centrifugation (300g for 10 min) and washed with 3 ml of 1X PBS and then re-suspended in 50 µl of 1X PBS and 25 µl of count beads.

Instrument

Sorting of the relevant population was performed using MoFloTM XDP (Beckman Coulter, Brea, CA, USA) or a FACS Diva (BD Biosciences, Franklin Lakes, NJ, USA) cell sorters.

	Analysis of human samples was performed using a Canto II flow cytometer equipped with 405nm, 488nm, and 633nm lasers, or a LSR Fortessa flow cytometer equipped with 355nm, 405nm, 488nm, 561nm and 640nm lasers (both from BD Biosciences, Franklin Lakes, NJ, USA). Analysis of mouse samples was performed using a Gallios flow cytometer equipped with 488nm, 638nm and 405nm lasers (Beckman Coulter, Brea, CA, USA).
Software	Data were processed using FCS 4 Express (De Novo Software, Los Angeles, CA), FlowJo version 9.8.5 (Tree Star, Ashland, OR, USA), or Caluza (Beckman Coulter, Brea, CA, USA).
Cell population abundance	The relevant leukemic cell populations were sorted using MoFloTM XDP (Beckman Coulter, Brea, CA, USA) or a FACS Diva (BD Biosciences, Franklin Lakes, NJ, USA) cell sorters. The cell purity was analyzed by fow cytometry by comparing the cell populations before and after sorting. All the post-sort fractions were at least 90% pure. For in vivo experiments the T cells were depleted from the leukemia samples by column selection (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) after staining the cells with human CD3 microbeads (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) according to manufacturer's instructions.
Gating strategy	The gating strategy for analysis of markers of interest on leukemia and T cells is provided in Supplementary Fig. 1, and the gating strategy for analysis of markers of interest on bone marrow myeloid progenitors and peripheral blood monocytes is provided in Supplementary Fig. 2.

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