nature portfolio

Corresponding author(s): Anna S. Nam

Last updated by author(s): Jul 9, 2026

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	No commercial, open source, or custom software or code was used for data collection.
Data analysis	Pooled scRNA-seq, CITE-seq and hashing libraries were processed with Cell Ranger (v6.1.1 and v6.1.2) using cellranger-multi pipeline (v1). Multi-omic nuclear data for snATAC-seq and snRNA-seq were processed together with Cell Ranger Arc (v2.0.0). The downstream analysis of the processed data was performed using Seurat (v4.1.0) and Signac (v1.5.0) packages. GoT analysis was carried out via the IronThrone pipeline V2.1. Pseudotime analysis was performed by constructing cell lineage trajectory with Monocle 3. The following packages were also used in downstream analysis in R: lme4 package (v.1.2-1), msigdbr (v7.2.1), fgsea (v1.12.0), FIMO (v5.4.1), Gotcha (v1). RNA velocity was assessed from the spliced and unspliced transcript variants using scVelo (v0.2.5). For a full list of software and packages used, please see methods.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The processed sequencing data is available via Gene Expression Omnibus (GEO) under the accession number GSE215140.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Cryopreserved bone marrow mononuclear cells were obtained from patients with CALR-mutated essential thrombocythemia (ET) and JAK2-mutated ET and polycythemia vera (PV) treated with weekly pegylated IFN-alfa2a during clinical trials MPN-RC-111 (NCT01259817)19 and MPN-RC-112 (NCT01258856). Data from three samples were obtained from a previous study (Nam, et al. Nature Genetics, 2022). Samples were de-identified, and only age and sex were available. Samples included in this study were from both male and female patients across a representative age range for MPN patients. All information is available in Supplementary Table 1.
Reporting on race, ethnicity, or other socially relevant groupings	Clinical samples were de-identified so data on race, ethnicity, or other socially relevant groupings was not available.
Population characteristics	Samples from patients with CALR-mutated or JAK2-mutated myeloproliferative neoplasms were included. Baseline and treated samples from the same patients were included in the study when available.
Recruitment	Samples were banked during clinical trials MPN-RC-111 (NCT01259817) and MPN-RC-112 (NCT01258856) so recruitment did not apply to our study.
Ethics oversight	The study was approved by the local ethics committee and by the Institutional Review Board (IRB) of Weill Cornell Medicine.

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

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	For the GoT-IM study on CD34+ cells, 8 baseline and 13 IFNa-treated samples (3 of which were collected post-IFN discontinuation) were collected from 10 individuals. Three baseline samples from previous work were included as age and sex-matched controls. n = 65,452 cells were profiled across all samples. For the GoT-IM study on CD34- cells, 5 baseline and 10 IFNa-treated samples (2 of which were collected following IFNa discontinuation) were collected from 7 individuals. n = 59,912 cells were profiled across all samples. For scRNA-sequencing of hydroxyurea-treated peripheral blood, 2 samples were collected from 2 individuals. For GoT-ATAC, 4 baseline and 3 treated samples were included. For GoT-ChA, 4 samples from 2 individuals were included int he study.
Data exclusions	We used standard practices set by experts in the single-cell sequencing field for filtering. For individual scRNA-seq datasets, cells with UMI > or < 3 standard deviations from the mean UMI and mitochondrial gene percentage >10% were filtered (Extended Data Fig. 1a, 4a). For genotyping, only UMIs with at least 2 or more supporting reads were retained for final genotyping assignments. For single-cell ATAC-data, cells with blacklist ratio >0.02, TSS enrichment <2 and nucleosome signal >4 were filtered out (Extended Data Fig. 7a-d).
Replication	In total, 24 independent human replicate samples were used for the study of GoT-IM on CD34+ cells, across from 10 individual patients. Data was analyzed separately within each sample and results were replicated across all (data from individual samples available in Extended Data Figures and Supplementary Tables). All in vitro experiments were performed with 2-4 biological replicates, and all replicate values are noted in figures and corresponding legends.
Randomization	Each patient sample contained both mutant and wild-type cells, which are not randomizable. In vitro experiments were completed with several technical and biological replicates and randomized.

Blinding	Blinding was not done in this study.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Supplementary Table 34 contains all antibodies used, manufacturer information, catalog numbers, applications, and dilutions.
Validation	Negative and positive controls were used in every flow cytometry experiment to gate positive cells. CITE-seq antibodies were validated by checking expression in expected cell types as determined by scRNA-seq.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	UT7 cell lines expressing MPL (thrombopoietin receptor) and either the mutant CALR (type 1, L367Tfs*46) or wildtype CALR transgene were acquired from the Mullally Lab. The K562 cell line was acquired from the Kleaveland Lab.
Authentication	Cell lines were authenticated upon purchase from source labs.
Mycoplasma contamination	Cell lines were mycoplasma tested upon arrival to the lab and once during use.
Commonly misidentified lines (See ICLAC register)	N/A

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Data is available via Gene Expression Omnibus (GEO) under the accession number GSE215140.

Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	Data includes 2 independent biological replicates.
Sequencing depth	We acquired 5-10mil reads per sample library. Reads were subsequently down-sampled within each replicate and condition to account for differences between samples.
Antibodies	Anti-human rabbit monoclonal PU.1 antibody (ab76543) and IgG antibody (ab172730) were used in the assay.
Peak calling parameters	Aligned reads were then used for peak calling with MACS3136 with a q-value threshold of 0.01 and reads from IgG antibody as the control.
Data quality	SPI1 over-enrichment in the called peaks was confirmed with Simple Enrichment Analysis (SEA, v5.5.5).
Software	Low-quality reads were filtered out using Trimmomatic, resulting reads were aligned with hg38 genome with Bowtie2, and PCR duplicates were removed with SAMtools. Overlapping peaks were processed with multiBamSummary function from deepTools to obtain counts matrix per sample for downstream analysis. Differential peak enrichment analysis was run between conditions (MUT vs WT and IFN?-treated vs control) using DESEQ2 for distal PU.1 peaks. Differential peaks in each condition were centered around SPI1 motif and motif enrichment was performed for regions +/- 250 bp from SPI1 motif using HOMER (v.4.1.1).

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	For sequencing studies, cryopreserved bone marrow mononuclear cells were thawed and stained using standard procedures with the surface antibody CD34-PE (clone AC136, dilution 1:50, Miltenyi Biotec) and DAPI (Sigma-Aldrich), according to manufacturer's protocol. To isolate CD34+ populations, cells were sorted for DAPI– and CD34+ cells with negative and positive controls used for gating and compensation. To isolate CD34- populations, cells were sorted for DAPI– and CD34– cells. CD34–, high side-scatter (SSC) was used to sort for granulocytes, medium SSC to enrich for monocytes/DCs, and low SSC to enrich for lymphocytes. FACS Buffer was used in all experiments. For additional flow cytometry experiments, see methods for details on sample preparation.
Instrument	BD Influx was used for sorting, and BD LSRFortessa Analyzer was used for in vitro colony analysis studies.
Software	Flow cytometric analysis was performed using FlowJo and OMIQ.
Cell population abundance	CD34+ cells made up ~1% of total bone marrow mononuclear cells.
Gating strategy	First gate: identify cell population and remove debris based on FSC-A and SSC-A. Second and Third gate: remove doublets with FSC-A + FSC-H and SSC-A and SSC-H. Fourth gate: Remove dead cells that are DAPI+ Fifth gate: Assess CD34+ population based on unstained negative control. CD34- populations can be identified by SSC-A due to size differences in morphology of cell types. Sorting strategy was also experimentally verified by scRNA-seq and snATAC-seq data.

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