Supplementary Information

A STAT5B-CD9 axis determines self-renewal in hematopoietic and leukemic stem cells

This file includes the Supplementary Methods, Supplementary Tables S1 and S2, Supplementary Figures S1-S7, Supplementary Figure Legends, the Supplementary References and an Annex Part I (cropped western blot membranes with all loading controls) and Part II (uncropped western blot membranes). Supplementary Table S3 is provided as separate .xlsx file.

Supplementary Materials and Methods

Graphical abstract

The graphical abstract was created with BioRender.com.

Primary patient-derived samples

Primary patient BM samples (routine examinations at diagnosis) were obtained from 6 patients with chronic phase CML, 8 with *FLT3*-ITD-mutated AML, 4 with *JAK2*^{V617F}-mutated MPN and one with a JAK2^{V617F+} myelodysplastic/myeloproliferative (MDS/MPN) overlap syndrome (listed in Supplementary **Table S2**). BM MNC were isolated using Ficoll and frozen.

Cell culture maintenance

HPC^{LSK} cells were generated, cultured and maintained as described[1]. HPC-7 cells were maintained as described[2]. HEL, SET2 and K562 were cultured in RPMI medium supplemented with 10% heat-inactivated fetal bovine serum (FBS), 50 µM 2-mercaptoethanol, 100 U/ml penicillin, and 100 µg/ml streptomycin. Platinum-E (plat-E) cells were cultured and maintained in DMEM medium supplemented with 10% FBS and 100 U/ml penicillin, and 100 µg/ml streptomycin.

Retroviral transductions

Virus supernatants were produced using plat-E cells that were separately transfected with a pMSCV retroviral vector containing following oncogenes: JAK2^{V617F}-GFP or FLT3-ITD-GFP using Turbofect according to the manufacturer's instruction. BCR/ABL^{p210}–GFP virus was produced by gP^{p210} BCR/ABL^{p210} retroviral producer cell line[3]. One day after transfection, medium was changed to IMDM culture media of target cells. Viral supernatants were harvested, filtered (0.45 µm), supplemented with 4 µg/mL polybrene and 2% stem cell factor (SCF) for HPC-7 cells, 4 µg/mL polybrene, 2% SCF and 12.5 ng/ml IL-6 for HPC^{LSK} cells, 4 µg/mL polybrene, 2% SCF, 12.5 ng/ml IL-6, 10 ng/ml IL-3 and 25 ng/ml TPO for *ex vivo* FACS-sorted LSK cells. Target cells were infected with virus supernatant on each of the following two days and GFP^{+} cells were FACS-sorted.

Flow Cytometry

Murine BM cells were recovered by crushing of femora and tibiae in PBS using a mortar and pestle. The cell suspension was filtered through a 70 µm cell strainer and washed once in PBS. Cells were stained on ice for 45 min in 2% FBS/PBS.

For characterization of the HSC compartment and FACS-sorting of LSK or HSC/MPP1 cells derived from BM of wt, *Stat5a*$^{-/-}$, *Stat5b*$^{-/-}$, cS5A^{hi} or STAT5B^{N642H} mice antibodies directed against Gr-1-APC/Cy7, TER-119-APC/Cy7, CD19-APC/Cy7, CD3-APC/Cy7, Mac1-APC/Cy7, Sca1-PE/Cy7, c-Kit-PE-Cy5, CD150-APC, CD48-PE, CD135-Biotin., CD34-FITC and Streptavidin-PB were used.

To detect CD9 levels on wt, *Stat5a*$^{-/-}$, *Stat5b*$^{-/-}$, cS5A^{hi}, STAT5B^{N642H} and JAK2^{V617F} LSK cells, BM cells were stained with antibodies directed against: Gr-1-APC/Cy7, TER-119-APC/Cy7, CD19-APC/Cy7, CD3-APC/Cy7, Mac1-APC/Cy7, Sca1-PE/Cy7, c-Kit-PE-Cy5 and CD9-PB.

For characterization of aCD9 blocking JAK2^{V617F} BM transplantation assays antibodies directed against Gr-1-APC/Cy7, TER-119-APC/Cy7, CD19-APC/Cy7, CD3-APC/Cy7, Mac1-APC/Cy7, Sca1-PE/Cy7, c-Kit-PE-Cy5, CD150-APC, CD48-PE, CD41-PB, Ly5.1-PE, Ly5.2-FITC, CD93-APC, CD43-Biotin, Streptavidin-BV650, CD127-PB, CD19-BV605, Ly5.1-FITC, Ly5.2-PB were used.

For intracellular stainings the following antibodies were used: HSC/LSK: CD19-PB, CD11b-PB, Gr-1-PB, TER-119-PB, CD3-PB, CD117-FITC, CD150-BV510, CD48-PE, LY-6A/E-Biotin., Streptavidin-PE/Cy7; B-cells: CD45R-PE; T-cells: CD3-FITC; Granulocytes: Gr-1-FITC, CD11b-PE/Cy7; erythroid cells: CD71-FITC; megakaryocytes: CD41-Biotin., Streptavidin-PE/Cy7; human patient BM MNC samples: CD9-FITC; phospho-STAT5-APC, IgG-APC, STAT5.

Human patient samples were surface stained with antibodies against: CD34-PE, CD38-PE/Cy7, CD11b-APC/Cy7, CD45-PB, CD9-FITC.

Samples were acquired using a FACS CantoII (BD), a Cytoflex (Beckman Coulter) or Cytoflex S (Beckman Coulter) flow cytometer. Data were analysed using FACSDiva, CytExpert or FlowJo for data analysis.

All high-purity FACS-sorting experiments were performed at 4 °C on a FACSAriaIII device equipped with a 488 nm, 561 nm, 633 nm and 395 nm laser and analysed using FACSDiva software.

Cell cycle staining

KI67/DAPI staining was performed as previously described[4]. Cells were stained using antibodies directed against: Gr-1-APC/Cy7, TER-119-APC/Cy7, CD19-APC/Cy7, CD3-APC/Cy7, Mac1-APC/Cy7, Sca1-PE/Cy7, c-Kit-PE-Cy5, CD150-APC and CD48-PE, KI67 and 2 µg/ml DAPI.

Apoptosis staining

Cells were harvested and stained using antibodies directed against: Gr-1-APC/Cy7, TER-119-APC/Cy7, CD19-APC/Cy7, CD3-APC/Cy7, Mac1-APC/Cy7, Sca1-PE/Cy7, c-Kit-PE-Cy5, CD150-APC and CD48-PE. All further steps were performed in Annexin-V Binding Buffer (1x). Cells were washed and incubated for 10-15 min at room temperature (RT) with KI67 antibody. After washing, cells were resuspended in Annexin-V Binding Buffer containing 2 µg/ml DAPI.

Intracellular staining

For intracellular pYSTAT5 and STAT5 stainings several settings were used: 3x10^6 BM cells of wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ mice were stimulated with: 100 ng/ml TPO (PeproTech 135-14) for CD150$^+$CD48$^-$ LSKs, 200 U/ml IL-2 and 50 ng/ml IL-7 for B-cells/T-cells, 100 ng/ml GM-CSF for granulocytes/macrophages/neutrophils or 10 U/ml EPO for erythroid cells or 100 ng/ml TPO for megakaryocytes. Cells were washed once in PBS and stained for surface markers CD150$^+$CD48$^-$LSK cells or B-cells/T-cells (B220$^+$/CD3$^+$) or granulocytes/neutrophils (Gr-1$^+$/CD11b$^+$) or erythrocytes (CD71$^+$) or megakaryocytes (CD41$^+$) *OR* 3x10^6 BM cells of wt, cS5A^{hi} or STAT5B^{N642H} mice were analysed *ex vivo* without cytokine stimulation. Cells were washed once in PBS and surface-stained for CD150$^+$CD48$^-$LSK cells. *OR* 2.5x10^5 AML$^{FLT3\text{-}ITD+}$ patient BM MNCs were used without stimulation. Cells were washed once with PBS and then surface-stained for CD9.

Surface-stained cells were washed once with PBS and fixed by 2% paraformaldehyde (Aldrich)/PBS at 37 °C for 10 min. Cells were permeabilized with 90% ice cold methanol in PBS/2% FBS/0.2% Tween-20 for 30 min at 4 °C. Subsequent washing steps were performed with PBS/2% FBS/0.2% Tween-20. Cells were washed twice and incubated with phospho-STAT5-APC or STAT5 and Streptavidin-PE/Cy7 (if a biotinylated surface antibody was used) antibodies on ice for one hour. Cells incubated with the STAT5 antibody were washed twice and then incubated with IgG-APC (H+L) on ice for 20 min. Cells were washed two times before analysis via flow cytometry.

qPCR analyses

RNA was isolated using the RNeasy MiniKit according to the manufacturer's instructions. RNA was transcribed with the iSCRIPT cDNA synthesis kit. Quantitative real-time PCR was performed on a CFX96 Real-Time System C1000Touch Thermal Cycler (Bio-Rad) with SsoAdvanced universal SYBR GreenSupermix. Murine mRNA levels of *Socs2*, *Cish*, *Meg3*, *Ifitm1*, *Ifitm3*, *Stat1*, *Cd9*, *Rplp0* were analysed. Target gene expression was normalized to expression of Ribosomal Protein Lateral Stalk Subunit P0 (Rplp0).

Cytoplasmic and nuclear fractionation

$1x10^7$ SET2 or HEL cells were washed once with PBS *OR* $2x10^7$ HPC^{LSK} cells stimulated with either 10 U/ml EPO, 100 ng/ml GM-CSF, 100 ng/ml IL-3 or 100 ng/ml TPO for 20 min at 37 °C or left unstimulated. Cells were washed once with PBS *OR* $2x10^7$ HPC^{LSK} cells retro-virally expressing FLT3-ITD or BCR/ABL^{p210} and HPC-7 cells retro-virally expressing $JAK2^{V617F}$ were washed once with PBS. All steps were performed at 4 °C. Cells listed above were resuspended in Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 2 mM DTT, 0.4 mM Na-Vanadate, 25 mM Na-Fluoride, 1 mM PMSF, 20 µg/ml Leupeptin and 20 U/ml Aprotinin) and incubated for 15 min on ice. 10% NP-40 was added, vortexed and cytoplasmic supernatant was collected after centrifugation at 13.000 g for 1 min. The pellet was washed 5 times with pre-cooled PBS and resuspended in Buffer B (20 mM HEPES pH 7.9, 25% Glycerol, 400 mM NaCl, 1 mM EDTA, 1 mM EGTA, 2 mM DTT, 0.4 mM Na-Vanadate, 25 mM Na-Fluoride, 1 mM PMSF, 20 µg/ml Leupeptin and 20 U/ml Aprotinin) and incubated for 1h at 4°C while shaking. Nuclear supernatant was collected after centrifugation at 13,000 g for 5 min.

Immunoprecipitation

Whole cell lysates were harvested as described[5]. 300µg protein lysates of HEL, K562, SET2 or 1mg protein lysates of HPC^{LSK} cells were incubated at 4°C o/n while rotating with either STAT5A or STAT5B antibody in radioimmunoprecipitation assay (RIPA) buffer. Next day, SureBeads™ Protein Magnetic Beads were added and incubated for 2h at 4°C while rotating. Beads were washed three times with pre-cooled RIPA buffer and then eluted with 4x Laemmli buffer for 5min at 95°C. Eluates were loaded for immunoblot analyses.

Immunoblotting

Proteins were separated on an 8% SDS polyacrylamide gel and transferred to nitrocellulose membranes. The following antibodies were used for immunoblotting: STAT5A, STAT5B, STAT5A/B, pSTAT5 (Y694/699). α-TUBULIN or RCC1 served as loading control. Anti-mouse and anti-rabbit HRP conjugated secondary antibodies were used. Chemiluminescent visualisation was performed with a ChemiDoc™ Touch Imaging System after incubation of the membranes with Clarity Western ECL reagent.

Chromatin Immunoprecipitation (ChIP) - qPCR

$2.5x10^7$ HPC^{LSK} cells were either stimulated with 100 ng/ml TPO or left unstimulated for 30min. Cells were washed once with PBS and crosslinked with 2 mM DSG in PBS for 30 min at RT, washed twice and incubated with 1% Formaldehyde in PBS for 10 min at RT. Cell lysis and sonication was performed as described in[6]. Sheared chromatin was incubated with 5 µg anti-STAT5B or anti-IgG (as control) antibodies overnight at 4 °C and processed as described in[6]. Selected pulled down genomic regions were quantified by quantitative real-time PCR performed on a CFX96 Real-Time System C1000Touch Thermal Cycler with SsoAdvanced universal SYBR GreenSupermix. The following genomic regions were analysed: *Cd9 Prom, Cd9 neg1, Cd9 neg2, Cish bdg, Cish neg, Bcl2-l1 bdg* and *Bcl2-l1-neg*.

Immunohistochemistry (IHC)

Cytospins of $5*10^4$ HPC^{LSK} wt cells (unstimulated or stimulated for 20min with TPO) were fixed for 5 min in ice-cold Methanol and air-dried. Retrieval step of slides was performed in 0.01 M citric acid (pH 6.95) in a water bath at 95 °C for 10 min. Slides were washed three times with PBS and then permeabilized with 0.1% TritonX-100 in PBS for 10 min. Slides were washed two times and blocked with 10% goat serum for 30 min. STAT5A (for IHC, 1:500) or STAT5B (1:100) antibody was incubated in PBS containing 2% goat serum at 4 °C o/n. Slides were washed two times and incubated with biotinylated secondary antibody (from VECTASTAIN® Elite® ABC HRP Kit) in PBST containing 2% goat serum for 30 min at RT. Slides were washed two times with PBST and then incubated with ABC solution (VECTASTAIN® Elite® ABC HRP Kit) for 45 min at RT. Slides were washed two times and counter-stained with Hemalaun for 90 s. After wash step with tap water slides were mounted with Aquatex. Pictures were taken with an OLYMPUS IX71 inverted microscope using the CellSens Dimension software with an OLYMPUS DP72 camera. For nuclear STAT5A/B staining intensity quantification, high-power field sections (x200 objective) were obtained for each sample and analysed with HistoQuest (TissueGnostics GmbH) quantification software as described[7].

***in vitro* experiments**

Analyses of CD9 levels

$1.5x10^6$ HPC^{LSK} cells were treated with 100 ng/ml TPO in culture conditions for 30 min, 2 h, 24 h or left unstimulated. Cells were washed once with PBS and mRNA levels were analysed by qPCR (see above) *or* $3x10^6$ wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ BM cells were stimulated with 100 ng/ml TPO in IMDM, 5% FBS, 15 mM monotioglycerol (MTG), 2 mM L-Glutamine and cultured for up to 96 h. TPO was added every 24 h. Cells were analysed for CD9 levels on LSK cells via flow cytometry every 24 h *or* HPC^{LSK}

cells retro-virally expressing FLT3-ITD or BCR/ABL^{p210} and HPC-7 cells expressing FLT3-ITD or JAK2^{V617F} were analysed for CD9 levels via flow cytometry.

Single HSC–derived culture

FACS-sorted single HSC/MPP1 (LSK, CD150^{+}, CD48^{-}) cells obtained from wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ mice were cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF and 25 ng/ml TPO. After 10 days, single cell-derived clones were analysed via flow cytometry for clonal outgrowth, numbers of LSK and total cells.

LSK-derived culture

1x10^4 LSK cells obtained from wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ mice were FACS-sorted and cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF and 25 ng/ml TPO. Cells were analysed via flow cytometry for total cell numbers after 4, 7, 10 and 15 days.

Stat5a/b over-expression in LSK culture

5x10^3 LSK cells obtained from wt mice were FACS-sorted and infected with a pMSCV-IRES-GFP-based construct encoding *Stat5a* or *Stat5b* as described[8]. Infections with the empty vector (eV) served as control. Infections were performed in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF, 25 ng/ml TPO, 10 ng/ml IL-3 and 10 ng/ml IL-6. Upon two days of infection, medium was exchanged to IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF and 25 ng/ml TPO. Cells were analysed via flow cytometry for GFP expression and total cell numbers after 5, 7, 10, 15, 20, 25 and 29 days and for LSK markers after 28 days.

LSK Re-plating

1x10^3 FACS-sorted LSK cells obtained from wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ mice were cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF and 25 ng/ml TPO. After 2 days all cells were harvested and seeded in methylcellulose (MethoCult 03231) supplemented with 10 ng/ml IL3, 10 ng/ml IL6, 5 U/ml EPO (3 U/ml), 0.5% SCF, 200 µg/ml transferrin and 10 µg/ml Insulin. Colonies were counted after 8 days, liquefied with PBS and analysed via flow cytometry for total cell numbers. 3x10^4 cells were seeded again in methylcellulose (MethoCult 03231) supplemented with cytokines as described above. Colonies were counted after 8 days.

HPC^{LSK} Re-plating – over-expressing eV, STAT5B, STAT5B^{N642H}

1.25x10^3 HPC^{LSK} cells retro-virally expressing STAT5B, STAT5B^{N642H} or empty vector (eV, as control) were harvested and seeded in methylcellulose (MethoCult 03231) supplemented with 0.3% SCF. Colonies were counted after 10 days, liquefied with PBS and analysed via flow cytometry for total cell numbers. This procedure was repeated for up to 3 serial platings.

CD9 blocking experiments

Colony formation assays (murine)

5x10^5 wt or JAK2^{V617F}, vav-Cre BM cells were treated with 5 µg/ml aCD9 or IgG antibodies and seeded in methylcellulose (MethoCult, 03231) supplemented with 20 ng/ml IL-3, 20 ng/ml IL-6, 5 U/ml EPO

(3 U/ml), 0.5% SCF, 200 µg/ml transferrin and 10 µg/ml insulin. Colonies were counted after 12 days, liquefied with PBS and analysed via flow cytometry for total and LSK cell numbers and for levels of c-kit and CD11b.

Liquid culture (murine)

$2.5x10^5$ wt or $JAK2^{V617F}$, vav-Cre BM cells were treated with 2 µg/ml aCD9 or IgG antibodies and cultured in IMDM with 5% FBS, 15 mM MTG, 2 mM L-Glutamine for 2 days. Cells were analysed for Annexin-V levels, total and LSK cell numbers.

Liquid culture - aCD9: KMC8 vs. MZ3 (murine)

$3x10^6$ $JAK2^{V617F}$, vav-Cre BM cells were treated with either 5 µg/ml aCD9 (KMC8), aCD9 (MZ3) or IgG antibodies and cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF, 25 ng/ml TPO, 10 ng/ml IL-3 and 10 ng/ml IL-6 for 4 days. Cells were analysed by flow cytometry for Annexin-V and CD9 levels, as well as for total and LSK cell numbers.

Liquid culture - aCD9 vs. AC-4-130

$3x10^6$ $JAK2^{V617F}$, vav-Cre BM cells were treated with either 5 µg/ml aCD9 or IgG antibodies or with 5 µM AC4-130[9] or respective amount of DMSO and cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF, 25 ng/ml TPO, 10 ng/ml IL-3 and 10 ng/ml IL-6 for 4 days. Cells were analysed by flow cytometry for Annexin-V, c-kit, and CD9 levels, as well as for total and LSK cell numbers.

Intracellular staining – aCD9 affects pYSTAT5 levels

$3x10^6$ $JAK2^{V617F}$, vav-Cre BM cells were treated with either 5 µg/ml aCD9 or IgG antibodies and cultured in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine. After 1 h treatment, cells were stimulated with 100ng/ml TPO, 100 ng/ml GM-CSF, 10 U/ml EPO or 100 ng/ml IL-3 for 15min. Cells were washed once in PBS and surface-stained for $CD150^+CD48^-$LSK cells following intracellular staining for pYSTAT5.

CD9 blocking experiments – patient samples

Liquid culture

$5x10^5$ $AML^{FLT3\text{-}ITD+}$, $CML^{BCR/ABL1+}$ or $MPN^{JAK2V617F}$ patient-derived BM MNC (listed in Supplementary **Table S2**) were treated with 2 µg/ml aCD9 or IgG antibodies and cultured in IMDM with 5% FBS, 15 mM MTG, 2 mM L-Glutamine for 2 days. $AML^{FLT3\text{-}ITD+}$ and $CML^{BCR/ABL1+}$ were maintained without cytokines, while $MPN^{JAK2V617F}$ BM MNCs were maintained with 100 ng/ml hSCF, 10 ng/ml hIL-3, 25 ng/ml hTPO, 5 U/ml EPO. After 2 days cells were analysed for $CD45^+$ and $CD34^+CD38^-$ cell numbers, Annexin-V, CD11b and CD34 levels.

Serial plating

$6x10^4$ $MPN^{JAK2V617F+}$ BM cells were treated with 10 µg/ml aCD9 or IgG antibodies and seeded in methylcellulose (MethoCult 04431). Colonies were counted after 14 days, liquefied with PBS and analysed via flow cytometry for total and $CD34^+CD38^-$ cell numbers. $3x10^4$ cells were again treated with 10 µg/ml aCD9 or IgG antibodies, and seeded in methylcellulose and as described above. Colonies

were counted after 20 days. $6x10^4$ $CML^{BCR/ABL1+}$ BM cells were treated with 5 µg/ml aCD9 or IgG antibodies and seeded in methylcellulose (MethoCult 04435). Colonies were counted after 11 days. $3x10^4$ cells were again treated with 5 µg/ml aCD9 or IgG antibodies, and seeded in methylcellulose as described above. Colonies were counted after 23 days. Colonies were always liquefied with PBS and analysed via flow cytometry for total and $CD34^+CD38^-$ cell numbers.

Liquid culture - non-toxicity test

$5x10^5$ control human patient BM MNC (listed in Supplementary **Table S2**) were treated with either 2 µg/ml, 5 µg/ml or 10 µg/ml aCD9 or IgG antibodies and cultured in IMDM with 5% FBS, 15 mM MTG, 2 mM L-Glutamine for 2 days. Cells were analysed for $CD45^+$ and $CD34^+CD38^-$ cell numbers.

Serial plating – non-toxicity test

$6x10^4$ control human patient BM MNC (listed in Supplementary **Table S2**) were treated with either 2 µg/ml, 5 µg/ml or 10 µg/ml aCD9 or IgG antibodies and seeded in methylcellulose (MethoCult 04431). After 14 days, colonies were liquefied with PBS and analysed via flow cytometry for $CD45^+$ and $CD34^+CD38^-$ cell numbers.

in vivo experiments

5-FU injection

wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ mice (6-10 weeks of age) were injected i.p. with 150 mg/kg 5-FU (resolved in 200 µl physiological sodium chloride solution). BM cells were analysed for total cell numbers, the presence of individual LSK populations, cell cycle distribution and apoptosis 8 days post treatment.

Serial transplantation assays

5×10^6 BM cells derived from wt (*Ly5.2*$^+$), *Stat5a*$^{-/-}$ (*Ly5.2*$^+$) and *Stat5b*$^{-/-}$ (*Ly5.2*$^+$) mice were injected intravenously (i.v.) into lethally irradiated (9 Gy) C57Bl/6J (*Ly5.1*$^+$) recipient mice. The irradiated control mice had to be sacrificed after 9 days. For the assessment of long-term repopulation capacities of transplanted HSCs, mice were euthanized 16 weeks post-transplantation. BM was analysed by flow cytometry for the contribution of *Ly5.1*$^+$ and *Ly5.2*$^+$ cells to individual HSC populations and total cell numbers. 5×10^6 of collected BM cells were re-injected in lethally irradiated (9 Gy) C57Bl/6J (*Ly5.1*$^+$) mice This procedure was repeated up to four rounds of transplantation.

Serial transplantation of BCR/ABL^{p210+} *LSK cells*

FACS-sorted LSK cells from wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ BM were infected with $BCR/ABL^{p210\text{-}GFP}$ virus-containing supernatant as previously described[10]. Two days post infection, $2x10^4$ GFP^+ LSK cells were injected i.v. in NOG mice. When the first recipient showed signs of disease, four mice per genotype were euthanized, analysed by flow cytometry for total cell numbers, LSK and lineage populations and $2.5x10^4$ GFP^+ cells were re-injected i.v. in NSG mice. In both rounds of transplantation, recipients were sacrificed upon first signs of disease (paralysis of hind legs) and analysed for total GFP^+ and GFP^+ LSK cell numbers.

JAK2^{V617F} *BM Transplantation – aCD9 pre-treatment*

5×10^6 BM cells derived from *JAK2^{V617F}* (*Ly5.2$^+$*) mice were treated *in vitro* with 5 µg/ml aCD9 or IgG in IMDM supplemented with in IMDM, 5% FBS, 15 mM MTG, 2 mM L-Glutamine, 2% SCF, 25 ng/ml TPO, 10 ng/ml IL-3 and 10 ng/ml IL-6 for 24 h. After a viability check using a live/dead dye (SytoxBlue), cells were injected i.v. in NSG mice (*Ly5.1$^+$*, 4 mice per condition) Three weeks after transplantation, spleen size and *Ly5.2$^+$* HSC and total cell numbers (flow cytometry) were analysed.

JAK2^{V617F} BM Transplantation – in vivo aCD9 treatment

5×10^6 BM cells from *JAK2^{V617F}* (*Ly5.2$^+$*) mice were injected i.v. in NSG mice (*Ly5.1$^+$*, n=8). One week thereafter, the recipients (*Ly5.1$^+$*) were treated 4 times (4-day-interval) with aCD9 or IgG i.v. (1.25 mg/kg, 4 mice per condition). Four weeks after the treatment, the recipient mice were analysed for spleen size and by flow cytometry for *Ly5.1$^+$* and *Ly5.2$^+$* HSC and total cell numbers.

Single cell RNA-Seq (10x genomics)

LSK cells were sorted from tibiae and femora of wt, *Stat5a$^{-/-}$* or *Stat5b$^{-/-}$* mice were subjected to single cell RNA-seq profiling using the 10X Genomics Chromium Single Cell Controller and sequenced on an Illumina Hi-Seq3000/4000 system. Further details are listed in the supplementary information. These data can be accessed in the GEO database, accession number GSE178648.

Mouse BM cells of wt, *Stat5a$^{-/-}$* or *Stat5b$^{-/-}$* mice were recovered from femora and tibiae in PBS. The cell suspension was filtered through a 70 µm cell strainer, washed once in PBS and frozen at -150 °C, according to "*10x Genomics Sample Preparation Demonstrated Protocol*" (CG00039). Three samples of wt, *Stat5a$^{-/-}$* or *Stat5b$^{-/-}$* BM were thawed and $3x10^4$ LSK cells per genotype were FACS sorted individually at the core facility flow cytometry of the Medical University of Vienna using the lineage cocktail of antibodies including (anti-mouse CD3; anti-mouse Ly-6G/Ly-6C; anti-mouse CD11b; anti-mouse CD45R/B220; anti-mouse TER-119), anti-mouse Ly6A/E (Sca-1) and anti-mouse CD117 (c-kit). FACS-purified cells were harvested in PBS with 0.04% BSA and subjected to single cell RNA-seq profiling using the 10X Genomics Chromium Single Cell Controller with the Chromium Single cell 3' Kit v1 following the manufacturers' instructions. Successful libraries were quality controlled using the QBit system for quantification and the Agilent Bioanalyzer for fragment length verification and sequenced on an Illumina Hi-Seq3000/4000 system with the 75bp paired end setting.

Single cell RNA-Seq: Sample demultiplexing, barcode processing and UMI counting

The 10x Genomics Cell Ranger pipeline 3.0.2 was used for sample demultiplexing, barcode processing and UMI counting. Reads were aligned to the mouse reference genome (version refdata-cellranger-mm10-3.0.0). Each library was processed separately using the CellRanger count tool to generate gene x barcode matrices.

Single cell RNA-Seq: Data cleaning and normalization

The R package Seurat v3[11,12] was used to analyze the data. Gene-barcode matrices were imported into Seurat and normalized and scaled with default settings. To remove cells with low sequencing quality, cutoffs for a minimal number of detected genes and UMI per cell were defined based on visualizations

of the distributions of the number of detected genes and UMI per cell for each sample. Potential doublets were identified using the following approach: normalized and scaled expression values of detected genes on sex chromosomes were used for dimension reduction by Uniform Manifold Approximation and Projection (UMAP). The first two UMAP dimensions were then used to find k nearest neighbors and construct a Shared Nearest Neighbor (SNN) graph which was then used for clustering. For that purpose the FindNeighbors (default settings) and FindClusters (resolution = 0.01) functions implemented in Seurat were used. In each sample, a cluster with a higher mean number of detected genes per cell as well as a higher median UMI per cell than the other clusters was identified. These clusters contain potential doublets and were removed. After filtering, 8036 *Stat5a*$^{-/-}$, 8347 *Stat5b*$^{-/-}$, and 6564 wt cells were preserved. Raw counts of kept cells were again normalized using the NormalizeData function of Seurat with default settings, and the top 2000 most variable genes were determined using Seurats FindVariableFeatures function (default settings).

Data integration and removing of contaminating cell clusters

Filtered and normalized samples were integrated using Seurat's "FindIntegrationAnchors"- and "IntegrateData"-functions (both with standard parameters). First, we checked for potential contaminating cell populations. For that purpose data was scaled (Seurat's "ScaleData"-function; all genes were used for scaling), dimensions reduced (Seurat's "RunPCA"; default settings), KNN calculated and used for SNN graph construction (Seurats "FindNeighbors"; dims = 1:20), and clusters defined (Seurat's "FindClusters"; resolution = 0.5). Visualization of these clusters on the first two dimensions from UMAP (Seurat's "RunUMAP" on first 20 principal components) identified small clusters localized far away from the bulk of cells representing potential contaminations. Removing of these contaminating clusters (by visual inspection of clusters in the UMAP and filtering accordingly) resulted in 6726 *Stat5a*$^{-/-}$, 7350 *Stat5b*$^{-/-}$, and 4904 wt cells.

Removing cell cycle genes from integration analysis

In addition to contaminating clusters, the UMAP plot showed a distinct group of cells, composed of two clusters, which mapped away from the rest of the cells. Marker gene identification revealed that these cells showed high expression of a G2-M cell cycle signature (data not shown).

In an attempt to identify clusters based on genes related to cell type rather than cell cycle state, cell cycle-related genes were removed from the clustering analysis. Inspired by a similar analysis[13], we categorised genes as cell-cycle related if they were part of the Gene Ontology (GO) set "mitotic cell cycle process" (GO:1903047). First, any gene from the 2500 anchor features that was found in the GO set, was removed from the anchor features list. Second, all genes from the remaining anchor features that were highly correlated with at least one of the cell cycle genes in the GO set were removed, too. To do the latter, Pearson correlation coefficients were calculated between any remaining anchor feature and the genes that were part of the GO set. Features that had a maximum Pearson correlation coefficient of greater than or equal to 0.15 between themselves and any gene of the GO set, were excluded from the

anchor features, too. After this procedure 1963 genes remained in the integration anchors list and were used for re-integration of the data.

Re-integration of datasets, cluster identification and force directed graph representation

The three data sets were re-integrated using the 1963 remaining anchor features and the wildtype sample as a reference. The newly integrated data were again scaled and PCA (default parameters) and UMAP (with dims = 1:30) dimensionality reductions were performed. A separation of subgroups of cells as observed before was not apparent in this analysis. We next calculated the k=7 nearest neighbours (KNN) and from that shared nearest neighbours (SNN) as implemented in the FindNeighbors function in Seurat. The SNN were used to identify clusters (FindClusters function with a resolution parameter of 0.9).

For visualization, the KNN graph was exported and the data were used to generate a force directed graph (using Gephi 0.9.2 with the Force Atlas 2 layout). The force directed graph was imported into R and added to the Seurat analysis object as a dimensional reduction object and used for visualization of the data.

For visual comparison of force directed graphs from the three samples the cell numbers were adjusted such that each plot showed 4904 cells (all cells for the wild-type sample, random samples for the other two). Force directed graphs were also used to visually inspect the expression of cluster marker genes (FeaturePlot function with min.cutoff = "q20", max.cutoff = "q95").

Cluster marker genes, cluster annotation, cluster heatmap

Cluster-specific maker genes (Supplementary **Table S3**) were identified using Seurat's FindConservedMarker function. These marker genes allowed for the manual annotation of clusters as being enriched for individual hematopoietic progenitor cell types. A random sample of 2000 cells was drawn in order to generate a heatmap demonstrating marker gene expression. Genes to be plotted were chosen as follows: The marker gene lists for each cluster were sorted by maximum p-value (low to high) and possible ties were sorted by fold change of the wt sample (high to low). The top ranking genes with a maximum p-value of < 1e-10 were plotted for each cluster, up to a total of 20 genes per cluster.

Single cell RNA-Seq: Comparison of cluster sizes

Relative cluster sizes were calculated for each sample separately as the percentage of the cells within a cluster. Differences between samples were calculated as log2 fold change between *Stat5a*$^{-/-}$ and wt as well as *Stat5b*$^{-/-}$ and wt.

Label-Transfer

We downloaded the data from Gene Expression Omnibus (GEO accession: GSE107727) and processed six samples (the three LSK and the three LK samples) following the methods described in Dahlin *et al*. 2018[13]. We reproduced their analysis and then performed Louvain clustering to identify 16 clusters and annotated them for hematopoietic stem and progenitor cell types following Dahlin *et al*. where possible. Clusters that did not receive a label in the original publication were not labelled by us either - except for one cluster adjacent to the HSC cluster, which we labelled "adjacent_HSC" (**Figure S1I**). We performed the label-transfer onto our dataset, following the standard label-transfer workflow

(https://satijalab.org/seurat/archive/v3.0/integration.html, "Standard Workflow", 09.07.2021) using default parameters.

Single cell RNA-Seq: Cell cycle score

Cell cycle phases were assigned to cells using the cyclone function[14] from the package scran with default parameters. The provided pre-trained classifier for mouse was used.

Single cell RNA-Seq: Identification of a Stat5b specific stem cell signature

Markers for HSCs were downloaded from CellMarker database (http://biocc.hrbmu.edu.cn/CellMarker/download.jsp). Genes from studies PMIDs 27580035, 28479188, 27225119 and 29915358 detected in any cell of our scRNA-seq were included (577 unique genes).

The mean expression of the HSC marker genes scaled within cells of one genotype was calculated for each cell and used to classify cells into HSC signature high (top10% = "HSC-high") and low ("HSC-low") cells. Differential expression analysis between HSC-high and HSC-low cells was performed for each genotype using the FindMarker function (logfc.threshold = 0, min.pct = 0). Only genes detected in at least 10% in any of the analyzed groups and significantly higher expressed in HSC-high cells in all genotypes (adjusted p value < 0.05) were included for further analysis. In order to discriminate between Stat5b and Stat5a effects, we looked for genes with opposing effects in both genotypes. 35 genes were downregulated in *Stat5b*$^{-/-}$ and up-regulated in *Stat5a*$^{-/-}$ compared to wt, and 2 genes were up-regulated in *Stat5b*$^{-/-}$ and down-regulated in *Stat5a*$^{-/-}$.

Gene Set Enrichment Analysis (GSEA)

FindMarkers function from Seurat (logfc.threshold = 0, min.pct = 0, min.cells.feature = 1, min.cells.group = 1) was used to perform differential gene expression analysis (DEA) of clusters HSC_1 and HSC_2 between wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ cells. GSEA was performed using fGSEA version 1.14.0 with ranked gene lists (sign of fold change * -log10(p-value)) of different DEAs (*Stat5a*$^{-/-}$ vs. wt; *Stat5b*$^{-/-}$ vs. WT; *Stat5b*$^{-/-}$ vs. *Stat5a*$^{-/-}$) and gene sets from CellMarker-database (obtained from Clustermole version 1.0.1).

Publicly available gene expression data sets

For survival analysis of STAT5B-specific HSC genes, the gene expression data sets GSE6891, GSE37642, GSE13159, GSE76008 and GSE63270 were obtained from GEO database and data set OHSU was obtained by cBioPortal for Cancer Genomics[15]. Raw Affymetrix gene expression data were processed, normalised and log2 transformed using the frozen robust multiarray (fRMA) algorithm and R 3.4.2 software[16]. Pre-processed Illumina BeadArray data (GSE76008) and RNA-seq data (OHSU) were used as provided by the databases. Differential expression was calculated using the lmFit function of the R package limma and an FDR<0.05 was considered statistically significant. Optimal cut points to stratify samples into CD9-high and CD9-low for survival analyses were calculated using maximally selected rank statistics (R package maxstat) as previously described[17]. Statistical significance between

these groups was assessed through the log-rank test. Survival analyses and visualization were performed using the R packages survival and survminer.

Processing of published RNA-seq data

For expression visualization of genes of interest we analysed a published RNA-seq dataset of FACS sorted HSC subpopulations (E-MTAB-2262)[18]. Briefly, raw FASTQ files were trimmed with Trimmomatic (version 0.36) and quality checked before and after trimming using FastQC. Next reads were mapped to the GENECODE M13 genome using STAR (version 2.5.2b) with default parameters. Counts for union gene models were obtained using featureCounts from the Subread package (version 1.5.1). Regularized, log-transformed counts (DESeq2 (version 1.28.1)) were used for plotting.

Supplementary Table S1: Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Antibodies used for Immunoblotting, IP, ChIP, IHC or blocking experiments		
STAT5A	R&D	MAB2174
STAT5A (for IHC)	Abcam	Ab32043
STAT5B	Abcam	Ab178941
Rabbit IgG	Abcam	Ab17273
pSTAT5 (Y694/699)	BD	611965
STAT5A/B	R&D	AF2168
A-TUBULIN	CST	21255
RCC1	Santa Cruz	Sc-55559
Anti-human CD9 Monoclonal Antibody	eBioscience	SN4 C3-3A2
CD9 Rat anti-Mouse	eBioscience	eBioKMC8
Purified anti-mouse CD9 Antibody	Biozym	(DRAP-27, MRP-1, p-24), clone: MZ3
Rat IgG2a kappa Isotype Control	eBioscience	eBR2a
Mouse IgG1 kappa Isotype Control	eBioscience	P3.6.2.8.1
Murine antibodies for flow cytometry		
TER119-APC/Cy7	Invitrogen	TER-119
Gr1-APC	Biozym	RB6-8C5
CD19-APC/Cy7	eBioscience	eBio1D3
CD3-APC/Cy7	Biozym	17A2
CD11b-APC/Cy7	Invitrogen	M1/70
Sca1-PE/Cy7	eBioscience	D7
c-kit-PE/Cy5	Invitrogen	2B8
CD150-APC	Biozym	TC15-12F12.2
CD48-PE	Biozym	HM48-1

CD135-Biotin	eBioscience	A2F10
CD34-FITC	eBioscience	RAM34
CD9-PB	Invitrogen	KMC8
CD19-PB	eBioscience	eBIO1D3
Gr1-PB	Invitrogen	RB6-8C5
CD11b-PB	eBioscience	M1/70
TER-119	eBioscience	TER-119
CD3-PB	Biozym	17A3
c-kit-FITC	Biozym	2B8
CD150-BV521	Biozym	TC15-12F12.2
Sca1-Biotin.	Biozym	D7
Streptavidin-PE/Cy7	Biozym	Streptavidin
Streptavidin-PB	eBioscience	Streptavidin
CD45R-PE	BD	RA3-6B2
CD3-FITC	BD	17A2
Gr1-FITC	BD	RB6-8C5
CD11b-PE/Cy7	eBioscience	M1/70
CD71-FITC	Invitrogen	R17217
CD41-Biotin.	eBioscience	eBioMWReg30
CD41-PB	Biozym	MWReg30
Ly5.1-PE	Biozym	A20
Ly5.1-FITC	Biozym	A20
Ly5.2-FITC	eBioscience	104
Ly5.2-PB	eBioscience	104
CD93-APC	Invitrogen	AA4.1
CD43-Biotin	BD	S7
Streptavidin BV650	Biozym	Streptavidin
CD19-BV605	Biozym	6D5
CD127-PB	eBioscience	A7R34
Phospho-STAT5-APC	BD	47/Stat5(pY694)
IgG-APC (H+L)	Invitrogen	Polyclonal
STAT5	CST	D2O6Y
AnnexinV-FITC	BD	556420
FITC Mouse Anti-KI67 Set	BD	556026
Human Antibodies for flow cytometry		

CD34-PE	Biozym	581
CD38-PE/Cy7	Biozym	Kappa HB-7
CD11b-APC/Cy7	Biozym	ICRF44
CD45-PB	eBioscience	HI30
CD9-FITC	Invitrogen	eBioSN4
Cytokines		
Murine Cytokines		
IL-2	PeproTech	212-12
IL-3	R&D	403-ML
IL-6	R&D	406-ML
IL-7	R&D	407-ML
GM-CSF	R&D	415-ML
TPO	PeproTech	315-14-100UG
EPO	Janssen	ERYPO® FS
SCF	VetMed	
Human Cytokines		
hIL-3	PeproTech	AF-200-03-100UG
hSCF	PeproTech	AF-300-07-100UG
hTPO	PeproTech	300-18
EPO	Janssen	ERYPO® FS
Chemicals, Peptides, and Recombinant Proteins		
Disuccinimidyl Glutarate	Thermo Scientific	20593
16% Formaldehyde solution (w/v), Methanol-free	Thermo Scientific	28906
DNAse type I	SIGMA	DN25-1G
TurboFect Transfection Reagent	Thermo Scientific	R0532
PenStrep	SIGMA	P4333-100ML
2-Mercaptoethanol	SIGMA	M3148-250ML
RPMI Medium	SIGMA	R8758-500ML
Fetal Bovine Serum (FBS)	Capricorn Scientific	FBS-12A
IMDM Medium	SIGMA	I3390-500ML
DMEM Medium	SIGMA	D6429-500ML
HEPES	SIGMA-ALDRICH	H0887-100ml
Aquatex Aqueous mounting medium	SIGMA-ALDRICH	1085620050
Mayer's hemalum solution	SIGMA-ALDRICH	1092491000
L-Glutamine	SIGMA-ALDRICH	G7513-100ML

Trypsin-EDTA Solution 1x	SIGMA	59417c
DAPI	SIGMA-ALDRICH	D9542
SytoxBlue	Fisher Scientific	S34857
5-Fluorouracil (5-FU)	SANDOZ	L01BC02
Clarity Western ECL Substrate	BIO-RAD	1705061
Tween20	ROTH	9127.2
(Polybrene) Hexadimethrine Bromide	SIGMA	H9268-5g
SsoAdvanced universal SYBR GreenSupermix	BIO-RAD	1725275
AnnexinV Binding Buffer (10x)	Invitrogen	00-0055-56
Critical Commercial Assays		
MethoCult H4435 Enriched	StemCell Technologies	04435
MethoCult H4431 Enriched	StemCell Technologies	04431
MethoCult M3231	StemCell Technologies	03231
VECTASTAIN® Elite® ABC HRP Kit	Szabo Scandic	PK-6101
RNeasy MiniKit	Qiagen	74106
Qiagen MinElute PCR Purification Kit	Qiagen	28006
iSCRIPT cDNA synthesis kit	BIO-RAD	1708891BUN
Cytofix/Cytoperm Plus	BD	555028
Deposited Data		
GEO Single cell RNA-Seq in progress	hscs	
Experimental Models: Cell Lines		
gP^{p210} BCR/ABL^{p210} retroviral producer cell line	VetMed	
K562	Kindly provided by Giulio Superti-Furga	
HEL	Kindly provided by Peter Valent	
SET2	Kindly provided by Peter Valent	
HPC-7	Kindly provided by Anthony Green	
HPC^{LSK}	Generated in house	
Experimental Models: Organisms/Strains		
Stat5a$^{-/-}$	VetMed	C57BL/6
Stat5b$^{-/-}$	VetMed	C57BL/6
$cS5A^{hi}$	VetMed	C57BL/6
$STAT5B^{N642H}$	VetMed	C57BL/6
$JAK2^{V617F}$	VetMed	C57BL/6
Vav1-Cre	VetMed	C57BL/6
C57BL/6N	VetMed	C57BL/6

NOG-F	Taconic Biosciences	NOD.Cg-*Prkdcscid Il2rgtm1Sug*/JicTac
NSG	Charles River	NOD.Cg-*Prkdc*^{scid} *Il2rg*tm^{1Wjl}/SzJ
Ly5.1^{+}	Charles River	B6.SJL-Ptprca

Oligonucleotides

Murine Primer (qPCR)	*forward (5'>3')*	*reverse (5'>3')*	*Origin*
Socs2	CAGTCAAACAGATGGTACTG	GGTAGTCTGAATGCGAACTATCTC	Microsynth
Cish	CCTTCGGGAATCTGGGTGG	GACTGACAGTGTGAACAGGTAGC	Microsynth
Meg3	ACTCTTGCCACATTAGCCC	CACAGGAAGACGCGACAG	Microsynth
Ifitm1	GCCTATGCCTACTCCGTG	AGGATGGTGAAGAACAGGGA	Microsynth
Ifitm3	AGCCTATGCCTACTCCGTGA	GACCAAGGTGCTGATGTTCA	Microsynth
Stat1	CACATTCACATGGGTGGAAC	TCTGGTGCTTCCTTTGGTCT	Microsynth
Cd9	CTGGCATTGCAGTGCTTGCTA	AACCCGAAGAACAATCCCAGC	Microsynth
Rplp0	GCTTTCTGGAGGGTGTCC	GCTTCAGCTTTGGCAGGG	Microsynth
Murine Primer (ChIP-qPCR)	*forward (5'>3')*	*reverse (5'>3')*	*Origin*
Cd9 Prom	GTAGTCCCTAGAGCCTGCCT	TCCTGTGGCTTTGTGGTTGA	Microsynth
Cd9 neg1	GCTACTCCGCACCCTAACTC	GCCTTGCTCACTCACCTTCTC	Microsynth
Cd9 neg2	TCCTGGTGCAAACCTACCC	GACAGCAGTTGAGCAGACCTA	Microsynth
Cish bdg	CAACTCTAGGAGCTCCCGCC	AACACCTTTGACAGATTTCCAAGAAC	Microsynth
Cish neg	TACCCCTTCCAACTCTGACTGAGC	TTCCCTCCAGGATGTGACTGTG	Microsynth
Bcl2-l1 bdg	TGAGCTTCAGGGAATCTTTGGG	CCGCTTCCTGTTCTGAGAAATG	Microsynth
Bcl2-l1 neg	ACATACTGCCACTGAGTACCAC	TGAGAAGAGCCCAGCCTAATTG	Microsynth

Recombinant DNA

JAK2^{V617F}-GFP in pMSCV	Kindly provided by Richard Moriggl	
FLT3-ITD-GFP in pMSCV	Kindly provided by Richard Moriggl	
STAT5A-GFP in pMSCV	Kindly provided by Richard Moriggl	
STAT5B-GFP in pMSCV	Kindly provided by Richard Moriggl	
STAT5B-V5-GFP in pMSCV	Cloned in-house	
STAT5B^{N642H}-V5-GFP in pMSCV	Cloned in-house	

Software and Algorithms

GraphPad Prism (8.4.3.686)	GraphPad Software	
FlowJo (v10.6.1)	DeNovo Software	
Cytexpert (2.4.0.28)	Beckman Coulter	

FACSDiva	BD	
ImageJ	NIH	
ImageLab 5.2.1	BIO-RAD	
CellSens Dimension 1.5	OLYMPUS	
Adobe Photoshop CS5.1	Adobe	
Adobe Illustrator CS5.1	Adobe	
CFX Maestro 1.1 (4.1.2433.1219)	BIO-RAD	
CellRanger pipeline 3.0.2	10X Genomis	
R version 4.0.2		
Seurat version 3	Bioconductor	
scran version 1.16.0	Bioconductor	
gephi 0.9.2	https://github.com/gephi/gephi	
Simpleaffy version 2.62.0	Bioconductor	
Frma version 1.38.0	Bioconductor	
Survival version 3.2-7	CRAN	
Survminer version 0.4.8	CRAN	
Ggpubr version 0.4.0	CRAN	
CellMarker database	http://biocc.hrbmu.edu.cn/CellMarker/download.jsp)	
STAR version 2.5.2b	https://github.com/alexdobin/STAR	
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	
Trimmomatic version 0.36	http://www.usadellab.org/cms/?page=trimmomatic	
Subread package version 1.5.1	http://subread.sourceforge.net/	
DESeq2 version 1.28.1	Bioconductor	
sva version 3.36.0	Bioconductor	
Histoquest 6.0.1.0127	TissueGnostics	
fgsea version 1.14.0	Bioconductor	
clustermole version 1.0.1	Bioconductor	
Other		
SureBeads™ Protein Magnetic Beads	BIO-RAD	161-4023

419

Supplementary Table S2: Information on primary patient samples

Nr	Sex	Date of Birth	Age	Diagnosis	FAB AML	% Blasts BM	Karyotype	Key Mutations / WHO
1	f	06.09.1940	79	CML CP	n.a.	1	46,XX,t(9;22)	BCR/ABL1: 35.1% IS
2	m	09.09.1992	27	CML CP	n.a.	0.5-1	n.a.	BCR/ABL1: 39.2% IS
3	f	01.11.1979	39	CML CP	n.a.	3	46,XX,t(9;22)	BCR/ABL1: 40.5% IS
4	f	29.10.1998	19	CML CP	n.a.	6	46,XX,t(9;22)	BCR/ABL1: 75.4% IS
5	m	22.09.1987	31	CML CP	n.a.	<2	46,XY,t(9;22)	BCR/ABL1: 27.9% IS
6	m	10.04.1948	70	CML CP	n.a.	2	46,XY,t(9;22)	BCR/ABL1: 68% IS
7	f	06.11.1965	53	MPN-PV	n.a.	<1	n.a.	JAK2 V617F+: 31.48%
8	f	29.07.1946	71	MPN/MDS	n.a.	2-3	46,XX, t(3;22)	JAK2 V617F+: 30.97%
9	f	04.04.1961	56	MPN-PV	n.a.	<1	46,XX	JAK2 V617F+: 16,32%
10	f	10.07.1976	42	MPN-ET	n.a.	<1	46,XX	JAK2 V617F+: 12.87%
11	f	19.04.1959	57	MPN-PV	n.a.	<1	46,XX	JAK2 V617F+: 53.56%
13	m	04.09.1962	53	AML (FLT3-ITD)	biphenotypic	90	46,XY	FLT3-ITD+
14	f	03.03.1965	51	AML (FLT3-ITD)	M1	95	46,XX	NPM1+, FLT3-ITD+
15	f	03.08.1972	44	AML (FLT3-ITD)	M1	80-85	46,XX	NPM1+, FLT3-ITD+
16	m	03.07.1967	50	AML (FLT3-ITD)	M4	67	46,XY,t(5;15)	NPM1+, FLT3-ITD+
18	m	10.02.1998	20	AML (FLT3-ITD)	M2/ AML with dysplasia	45	46,XY	NPM1+, FLT3-ITD+, KIT D816V, CEBPA
19	f	09.10.1975	42	AML (FLT3-ITD)	M1	85	46,XX	NPM1+, FLT3-ITD+
20	f	25.10.1964	53	AML (FLT3-ITD)	M4	35	46,XX	NPM1+, FLT3-ITD+
21	m	09.10.1956	62	AML (FLT3-ITD)	M2	85	46,XY	NPM1+, FLT3-ITD+
23	f	29.08.1948	65	Lymphoma	n.a.	1	46,XX	n.a.
24	m	14.06.1981	34	Lymphoma	n.a.	<1	n.a.	n.a.
25	m	19.12.1976	39	Lymphoma	n.a.	1	n.a.	none
26	m	14.07.1940	79	Lymphoma	n.a.	0	n.a.	none

Figure S1

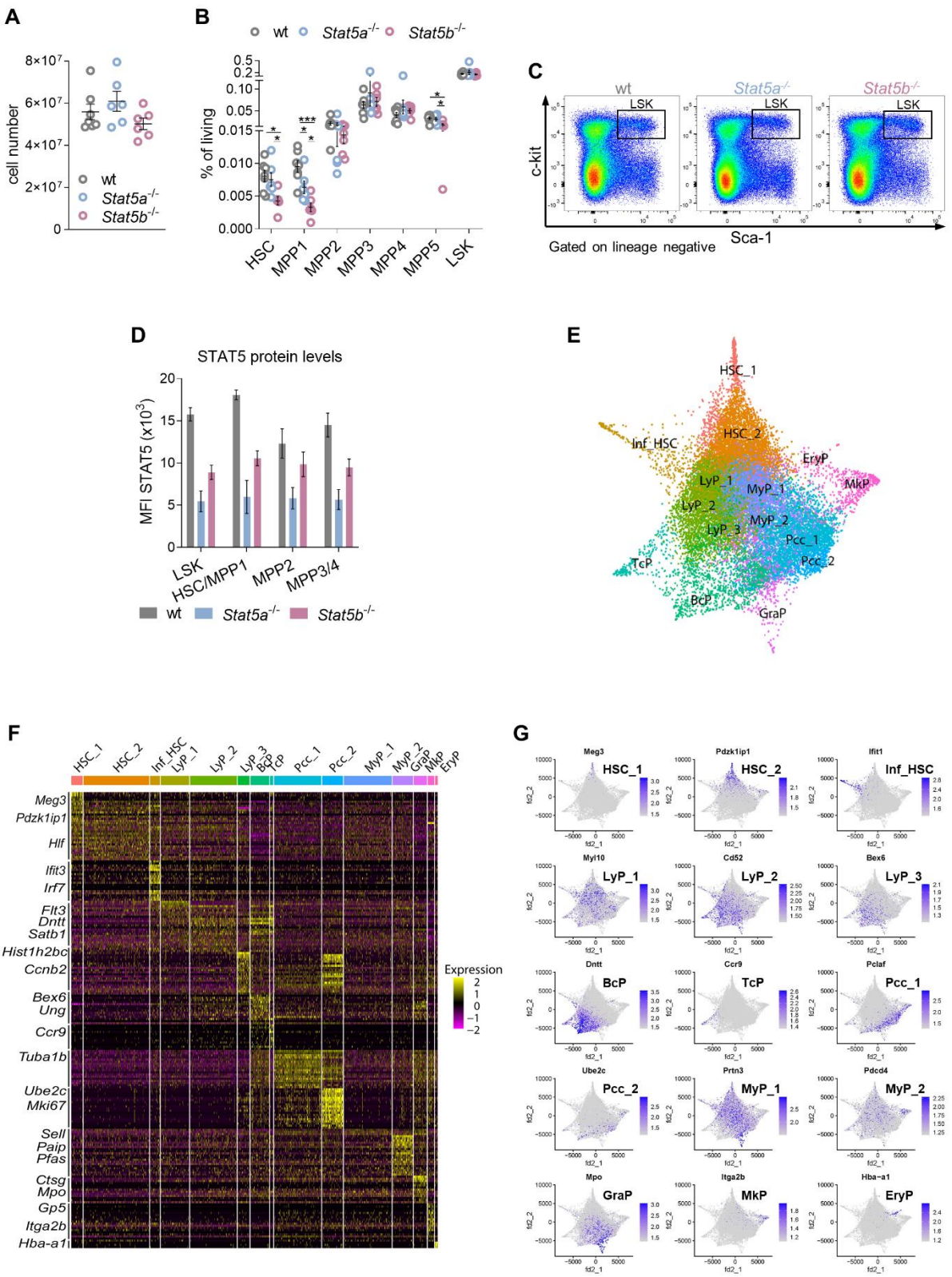
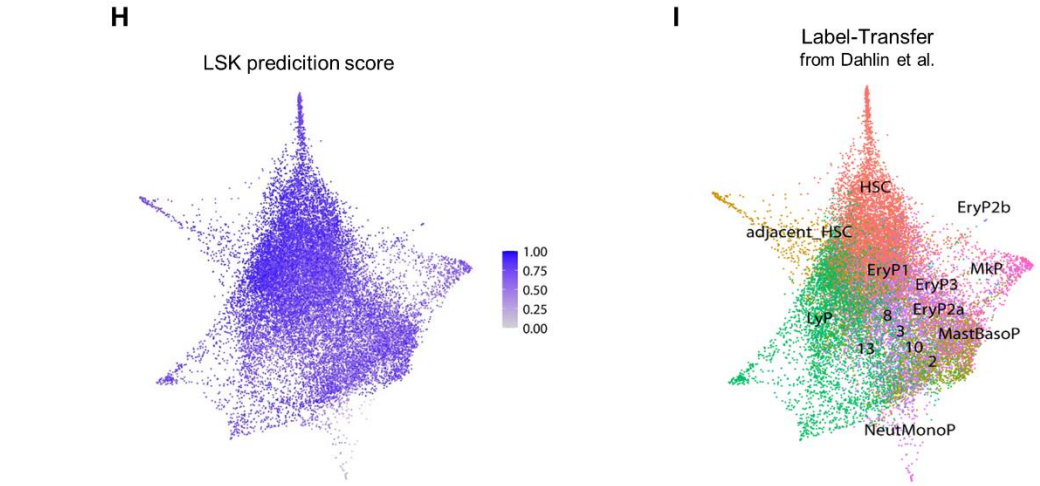

A
cell number
8×10^7
6×10^7
4×10^7
2×10^7
0
wt
Stat5a-/-
Stat5b-/-
B
wt
Stat5a-/-
Stat5b-/-
% of living
0.5
0.2
0.10
0.05
0.015
0.010
0.005
0.000
HSC
MPP1
MPP2
MPP3
MPP4
MPP5
LSK
C
wt
Stat5a-/-
Stat5b-/-
LSK
c-kit
Sca-1
Gated on lineage negative
D
STAT5 protein levels
MFI STAT5 (x10^3)
20
15
10
5
0
LSK
HSC/MPP1
MPP2
MPP3/4
wt
Stat5a-/-
Stat5b-/-
E
HSC_1
HSC_2
Inf_HSC
EryP
LyP_1
MyP_1
MkP
LyP_2
MyP_2
LyP_3
Pcc_1
Pcc_2
TcP
BcP
GraP
F
HSC_1
HSC_2
Inf_HSC
LyP_1
LyP_2
LyP_3
BcP
TcP
Pcc_1
Pcc_2
MyP_1
MyP_2
GraP
MkP
EryP
Meg3
Pdzk1ip1
Hlf
Ifit3
Irf7
Flt3
Dntt
Satb1
Hist1h2bc
Ccnb2
Bex6
Ung
Ccr9
Tuba1b
Ube2c
Mki67
Sell
Paip
Pfas
Ctsg
Mpo
Gp5
Itga2b
Hba-a1
Expression
2
1
0
−1
−2
G
Meg3
HSC_1
Pdzk1ip1
HSC_2
Ifit1
Inf_HSC
Myl10
LyP_1
Cd52
LyP_2
Bex6
LyP_3
Dntt
BcP
Ccr9
TcP
Pclaf
Pcc_1
Ube2c
Pcc_2
Prtn3
MyP_1
Pdcd4
MyP_2
Mpo
GraP
Itga2b
MkP
Hba-a1
EryP
fd2_1
fd2_2

Figure S1


H
LSK predicition score
1.00
0.75
0.50
0.25
0.00
I
Label-Transfer
from Dahlin et al.
HSC
EryP2b
adjacent_HSC
EryP1
MkP
EryP3
LyP
8
EryP2a
3
MastBasoP
13
10
2
NeutMonoP

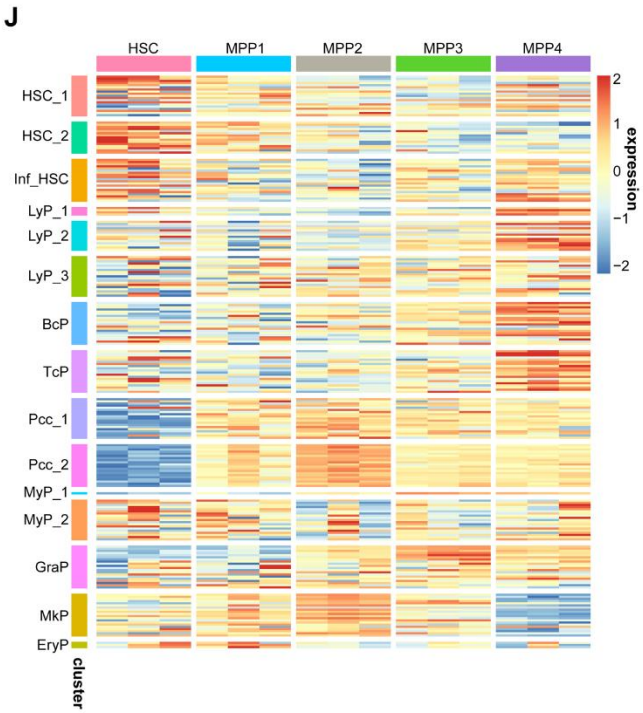
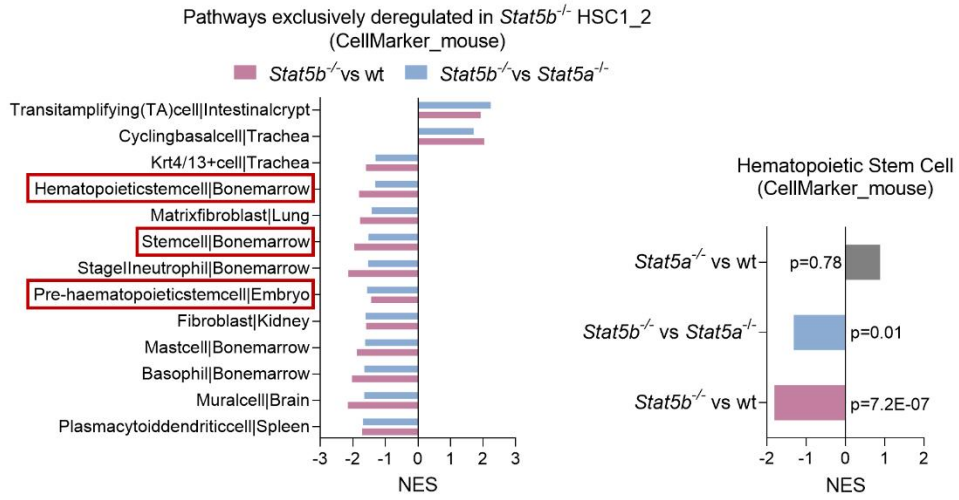

J
HSC
MPP1
MPP2
MPP3
MPP4
HSC_1
HSC_2
Inf_HSC
LyP_1
LyP_2
LyP_3
BcP
TcP
Pcc_1
Pcc_2
MyP_1
MyP_2
GraP
MkP
EryP
cluster
2
1
0
−1
−2
expression

Figure S1

K



L

Molecular cell cycle score of HSC_1 and HSC_2

%	wt	*Stat5a*$^{-/-}$	*Stat5b*$^{-/-}$
G_0/G_1	94.20	93.68	91.06
S	2.90	4.36	3.52
G_2M	2.90	1.96	5.42

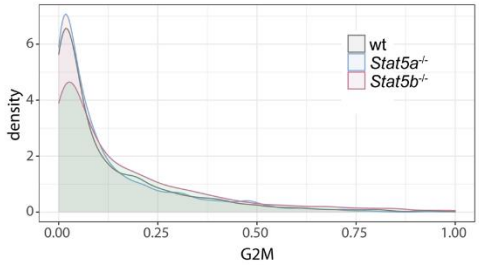


Figure S1: STAT5B, but not STAT5A, has an important role in HSC dormancy.

(A)-(B) FACS analyses of wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ BM. Quantification of (A) total cell numbers (n≥6/genotype, mean±SEM) and (B) relative HSC, MPP1-5 and LSK cell numbers (n≥6/genotype, mean±SEM) and (C) representative gatings of LSK population on lineage negative cells.

(D) STAT5 levels in wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ LSK, HSC/MPP1, MPP2 and MPP3/4 cells determined by flow cytometry (n=5/genotype, mean±SEM).

(E-L) Single cell RNA-seq of LSK cells from wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ mice (n=3 pooled/genotype). (E) Force-directed graph showing all defined and annotated clusters. HSC_1-2: Hematopoietic stem cells_1-2, Inf_HSC: Interferon activated HSCs, LyP_1-3: Lymphoid progenitors_1-3, BcP_B-cell progenitors, TcP: T-cell progenitors, Pcc_1-2: Progenitors in cell cycle_1-2, MyP_1-2: Myeloid progenitors_1-2. GraP: Granulocyte progenitors, MkP: Megakaryocyte progenitors, EryP: Erythroid progenitors. (F) Heatmap illustrating the top 20 upregulated genes of each cluster, color code indicates z-scores. (G) Force-directed graphs with color key indicating log normalised gene expression for one representative marker gene of each defined cluster. Marker genes were chosen based on[18–21]. (H) Force-directed graph showing an LSK prediction score (based on[13]) for each cell in our data. Cells with a score >0.5 are predicted to be LSKs. (I) Force-directed graph showing all the transferred labels from[13]. HSC: Hematopoietic stem cells, LyP: Lymphoid progenitors, NeutMonoP: Neutrophil, granulocyte and monocyte progenitors, MastBasoP: Mast cell and basophil progenitors, EryP: Erythroid progenitors 1-3, MkP: Megakaryocyte progenitors. 92% of our HSC_1 cells and 84% of our HSC_2 cells received the label "HSC", while 90% of our Inf_HSC cells received the label "adjacent_HSC". 71% of our BcPs and 99% of our TcPs were labelled LyP. Cells with the strongest MkP, EryP and GraP expression profiles received MkP, EryP2b, and NeutMonoP labels. (J) Expression levels of the top (p-value < 1E-10) cluster specific genes on an independent RNA-Seq dataset of FACS-sorted HSC, MPP1, MPP2, MPP3 or MPP4 cell populations from Cabezas-Wallscheid et al.[18]. (K) GSEA (Gene Set Enrichment Analysis) of HSC_1 and HSC_2 cells using the CellMarker database[22]. NES: Normalized Enrichment Score (L) Molecular cell cycle score. (*Left*) Percentages of cells predicted to be in G_0/G_1, S and G_2M phase in pooled HSC_1 and HSC_2 clusters and (*right)* density plots of the G_2M score of each genotype.

Levels of significance were calculated using 1-way ANOVA in (B) and unpaired t-test in (K) *p < 0.05, ***p < 0.001.

Figure S2

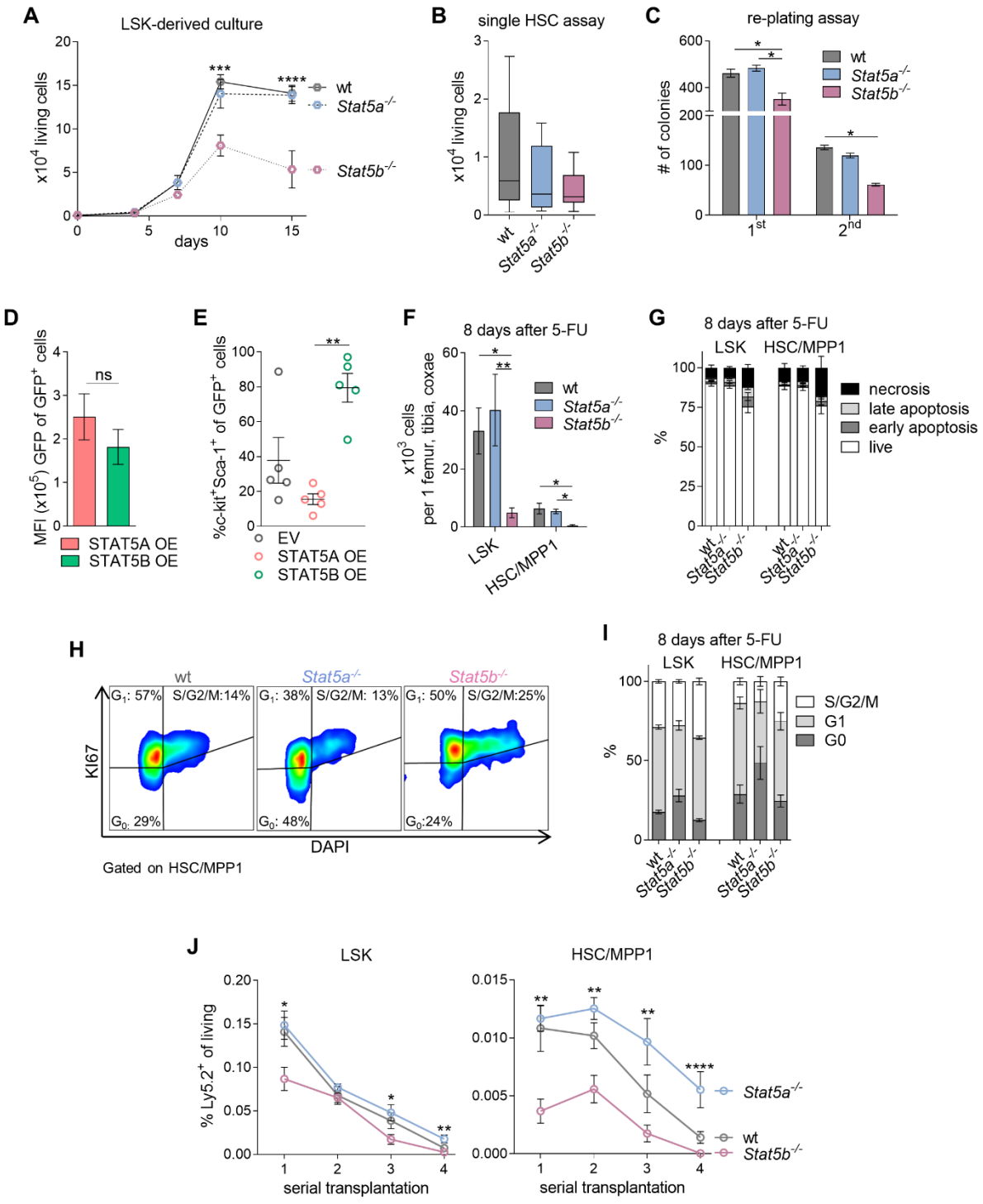

A
LSK-derived culture
x10^4 living cells
days
wt
Stat5a-/-
Stat5b-/-
B
single HSC assay
x10^4 living cells
C
re-plating assay
# of colonies
1st
2nd
D
MFI (x10^5) GFP of GFP+ cells
ns
STAT5A OE
STAT5B OE
E
%c-kit+Sca-1+ of GFP+ cells
EV
STAT5A OE
STAT5B OE
F
8 days after 5-FU
x10^3 cells per 1 femur, tibia, coxae
LSK
HSC/MPP1
G
8 days after 5-FU
LSK
HSC/MPP1
necrosis
late apoptosis
early apoptosis
live
H
wt
Stat5a-/-
Stat5b-/-
G1: 57%
S/G2/M:14%
G1: 38%
S/G2/M: 13%
G1: 50%
S/G2/M:25%
G0: 29%
G0: 48%
G0:24%
KI67
DAPI
Gated on HSC/MPP1
I
8 days after 5-FU
LSK
HSC/MPP1
S/G2/M
G1
G0
J
LSK
HSC/MPP1
% Ly5.2+ of living
serial transplantation
Stat5a-/-
wt
Stat5b-/-

Figure S2: STAT5B drives HSC self-renewal.

(A) *Ex vivo* growth analysis of LSK cells of wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ BM. 5x10^3 LSK cells were FACS-sorted, cultured and counted over 15 days (n=3/genotype, mean±SEM).

(B) Single HSC assay using cell surface markers (Lineage$^-$ (CD3, CD19, CD11b, Gr-1, Ter-119), c-kit$^+$, Sca-1$^+$, CD150$^+$, CD48$^-$). wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ BM: total cell numbers derived from a single HSC/MPP1 cell after 10 days (n=74/genotype, whiskers: Tukey).

(C) Colony numbers of serial plating of FACS-sorted LSK cells of wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ BM (n=3/genotype, mean±SEM).

(D)-(E) STAT5A (STAT5A-GFP) or STAT5B (STAT5B-GFP) over-expression in LSK cells. EV (empty vector-GFP) was used as control. (D) GFP levels of LSK cells expressing STAT5A-GFP or STAT5B-GFP (n=5/genotype, mean±SEM). (E) Quantification of c-kit and Sca-1 expression on GFP$^+$ cells (n=5/genotype, mean±SEM).

(F)-(I) 5-FU recovery assays of wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ mice. Mice were analysed 8-days after 5-FU injection. (F) Quantification of LSK and HSC/MPP1 cell numbers (n=5/genotype, mean±SEM). (G) Annexin-V/DAPI staining of LSK and HSC/MPP1 cells showing distribution of live (Annexin-V$^-$/DAPI$^-$), necrotic (Annexin-V$^-$/DAPI$^+$), early apoptotic (Annexin-V^{-+}/DAPI$^-$) and late apoptotic (Annexin-V$^+$/DAPI$^+$) cells (n=5/genotype, mean±SEM). (H) Representative FACS plots and gating strategy of cell cycle analysis (KI67/DAPI) of HSC/MPP1 cells and (I) the cell cycle phase distribution of LSK and HSC/MPP1 cells (n=5/genotype, mean±SEM).

(J) Serial transplantation assays of wt, *Stat5a*$^{-/-}$ or *Stat5b*$^{-/-}$ BM cells. 5x10^6 BM cells (Ly5.2$^+$) were injected in lethally Ly5.1$^+$ irradiated recipient mice and reinjected 8 weeks thereafter. Quantification of (*left*) Ly5.2$^+$ LSK cells (n=5/genotype, mean±SEM) and (*right*) Ly5.2$^+$ HSC/MPP1 cells in recipient mice in primary, secondary, tertiary and quaternary transplantation (n=5/genotype, mean±SEM).

Levels of significance were calculated using 1-way ANOVA in (C), (E), (F), (J) and using 2-way ANOVA in (A) and using Wilcoxon test in (D). *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.

Figure S3

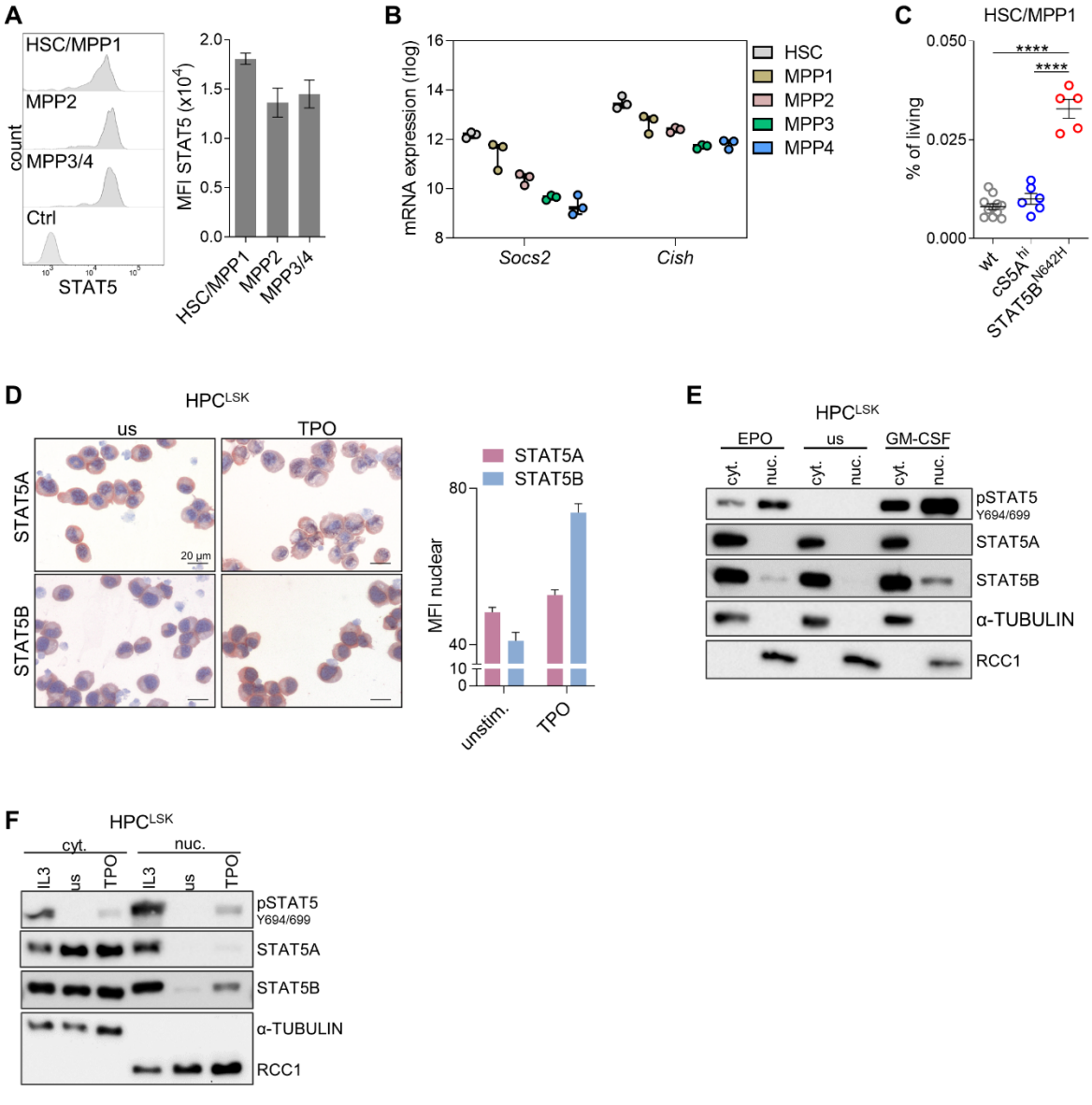

A
HSC/MPP1
MPP2
MPP3/4
Ctrl
count
STAT5
MFI STAT5 (x10^4)
HSC/MPP1
MPP2
MPP3/4
B
mRNA expression (rlog)
HSC
MPP1
MPP2
MPP3
MPP4
Socs2
Cish
C
HSC/MPP1
% of living
wt
cS5A hi
STAT5B N642H
D
HPC LSK
us
TPO
STAT5A
STAT5B
20 µm
MFI nuclear
STAT5A
STAT5B
unstim.
TPO
E
HPC LSK
EPO
us
GM-CSF
cyt.
nuc.
pSTAT5 Y694/699
STAT5A
STAT5B
α-TUBULIN
RCC1
F
HPC LSK
cyt.
nuc.
IL3
us
TPO
pSTAT5 Y694/699
STAT5A
STAT5B
α-TUBULIN
RCC1

Figure S3: Selective STAT5B activation in HSCs drives self-renewal.

(A)-(B) Analysis of STAT5 signaling in HSC subpopulations. (A) (*Left*) Representative histograms of STAT5 MFI levels of wt HSC/MPP1, MPP2 and MPP3/4 cells and isotype control (Ctrl) and (*right*) the quantification of STAT5 levels. (B) *Socs2* and *Cish* rlog transformed mRNA expression of FACS-sorted murine HSC and MPP1-4 subpopulations (analysis of published data[18]).

(C) Relative quantification of HSC/MPP1 cells expressing hyperactive STAT5A ($cS5A^{hi}$) or STAT5B ($STAT5B^{N642H}$) isolated from BM of transgenic mice ($n \geq 5$/genotype, mean±SEM).

(D) (*Left*) STAT5A or STAT5B IHC staining of HPC^{LSK} wt cytospins with or without 20 min TPO stimulation. Scale bar indicates 20 µm. Representative pictures of three independent experiments. (*Right*) Quantification of nuclear STAT5A and STAT5B MFI levels. Quantification of three pictures of one out of four independent HPC^{LSK} cell lines.

(E)-(F) pSTAT5 (Y694/699), STAT5A, and STAT5B immunoblot analysis of cytoplasmic and nuclear fractions of HPC^{LSK} wt cells after 20 min stimulation with (D) EPO or GM-CSF, or (E) IL3 or TPO stimulation or without stimulation (us). RCC1 is used as loading control for the nuclear, α-TUBULIN for the cytoplasmic fraction.

Levels of significance were calculated using 1-way ANOVA in (C). ****$p < 0.0001$.

Figure S4

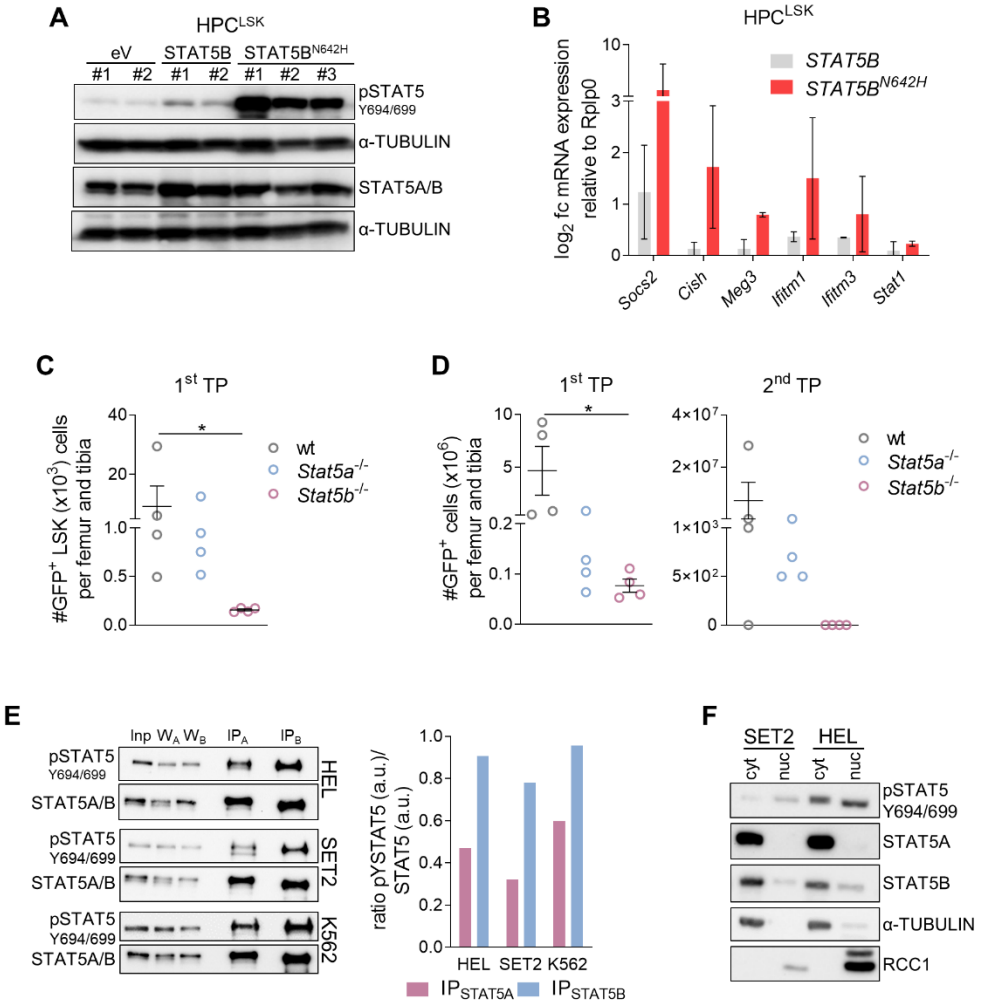

A
HPC^LSK
eV
STAT5B
STAT5B^N642H
#1 #2 #1 #2 #1 #2 #3
pSTAT5 Y694/699
α-TUBULIN
STAT5A/B
α-TUBULIN
B
HPC^LSK
STAT5B
STAT5B^N642H
log2 fc mRNA expression relative to Rplp0
Socs2
Cish
Meg3
Ifitm1
Ifitm3
Stat1
C
1st TP
#GFP+ LSK (x10^3) cells per femur and tibia
wt
Stat5a-/-
Stat5b-/-
D
1st TP
2nd TP
#GFP+ cells (x10^6) per femur and tibia
wt
Stat5a-/-
Stat5b-/-
E
Inp WA WB IPA IPB
pSTAT5 Y694/699
STAT5A/B
HEL
SET2
K562
ratio pYSTAT5 (a.u.)/ STAT5 (a.u.)
HEL SET2 K562
IPSTAT5A
IPSTAT5B
F
SET2
HEL
cyt
nuc
pSTAT5 Y694/699
STAT5A
STAT5B
α-TUBULIN
RCC1

Figure S4: Selective STAT5B activation drives the self-renewal of LSCs.

(A)-(B) Analysis of HPC^{LSK} cells expressing empty Vector (eV), STAT5B or $STAT5B^{N642H}$. (A) pSTAT5 (Y694/699), STAT5A/B and α-TUBULIN (loading control) immunoblot analysis and (B) quantification of mRNA expression (n=3/genotype, mean±SEM).

(C)-(D) $BCR/ABL^{p210\text{-}GFP+}$ LSK re-transplantation assay. (C) $BCR/ABL^{p210\text{-}GFP+}$ LSK cell numbers in recipients' BM 13 weeks after 1st transplantation (TP, n=4/genotype, mean±SEM) and (D) $BCR/ABL^{p210\text{-}GFP+}$ cell numbers after 1st and 2nd TP (n=4/genotype, mean±SEM).

(E)-(F) Analysis of pYSTAT5 levels in the presence of oncoproteins. Input (Inp), W_A (wash fraction of STAT5A IP), W_B (wash fraction of STAT5B IP), IP (immunoprecipitate). (E) *Left*: pSTAT5 (Y694/699) and STAT5A/B immunoblot analysis of IP of either STAT5A or STAT5B of human HEL ($JAK2^{V617F}$), SET2 ($JAK2^{V617F}$) and K562 (BCR/ABL1) leukemic cell lines and (*right*) their quantification. (F) pSTAT5 (Y694/699), STAT5A and STAT5B immunoblots of cytoplasmic and nuclear fractions isolated from HEL and SET2 cell lines. RCC1 is used as loading control for the nuclear, α-TUBULIN for the cytoplasmic fraction. Representative blots of three independent experiments are shown.

Levels of significance were calculated using 1-way ANOVA in (C) and (D). *$p < 0.05$.

Figure S5

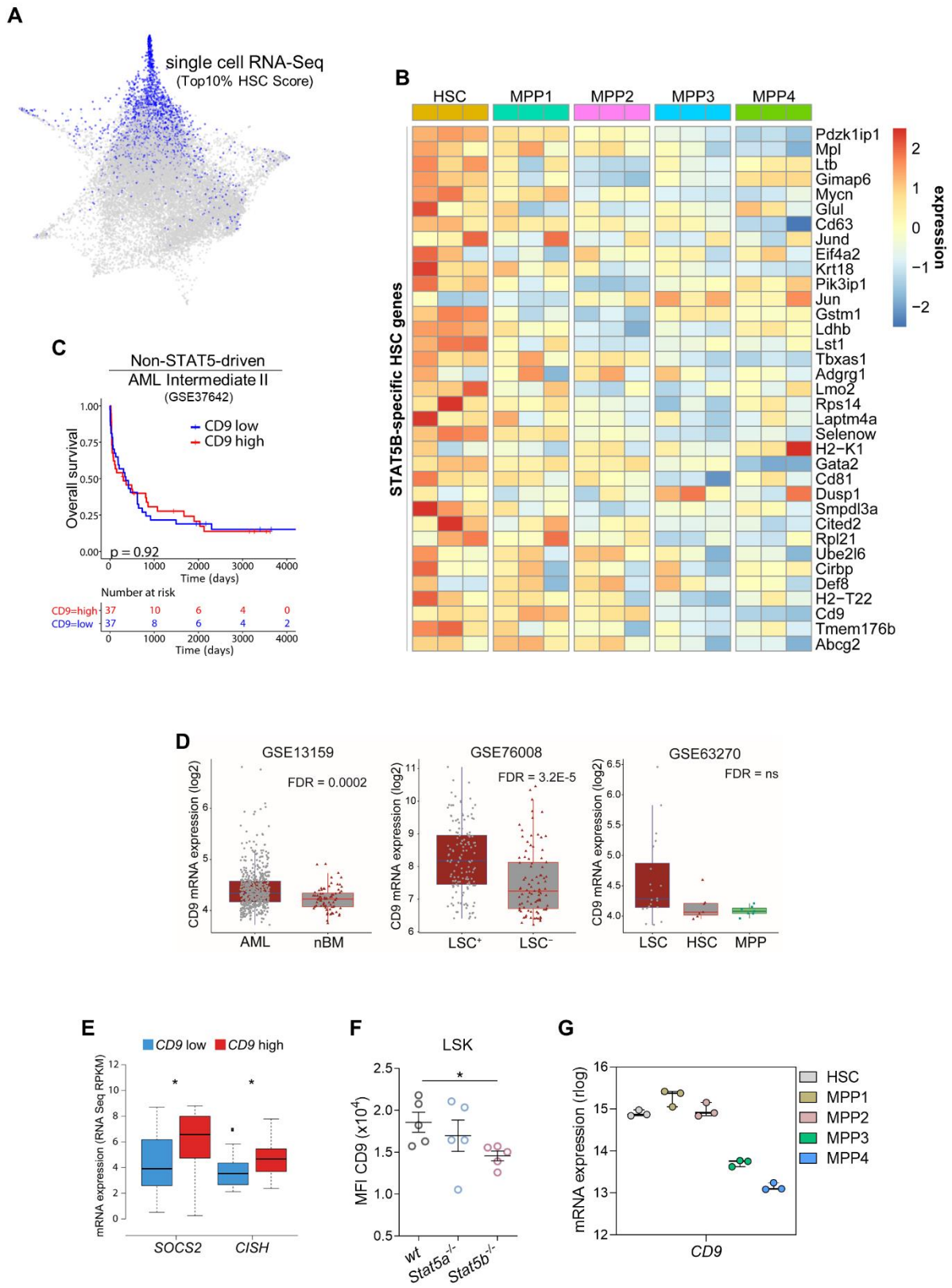

A
single cell RNA-Seq
(Top10% HSC Score)
B
HSC
MPP1
MPP2
MPP3
MPP4
STAT5B-specific HSC genes
Pdzk1ip1
Mpl
Ltb
Gimap6
Mycn
Glul
Cd63
Jund
Eif4a2
Krt18
Pik3ip1
Jun
Gstm1
Ldhb
Lst1
Tbxas1
Adgrg1
Lmo2
Rps14
Laptm4a
Selenow
H2−K1
Gata2
Cd81
Dusp1
Smpdl3a
Cited2
Rpl21
Ube2l6
Cirbp
Def8
H2−T22
Cd9
Tmem176b
Abcg2
expression
2
1
0
−1
−2
C
Non-STAT5-driven
AML Intermediate II
(GSE37642)
CD9 low
CD9 high
Overall survival
p = 0.92
Time (days)
Number at risk
CD9=high 37 10 6 4 0
CD9=low 37 8 6 4 2
D
GSE13159
FDR = 0.0002
CD9 mRNA expression (log2)
AML
nBM
GSE76008
FDR = 3.2E-5
LSC+
LSC−
GSE63270
FDR = ns
LSC
HSC
MPP
E
CD9 low
CD9 high
mRNA expression (RNA Seq RPKM)
SOCS2
CISH
F
LSK
MFI CD9 (x10^4)
wt
Stat5a−/−
Stat5b−/−
G
mRNA expression (rlog)
CD9
HSC
MPP1
MPP2
MPP3
MPP4

Figure S5

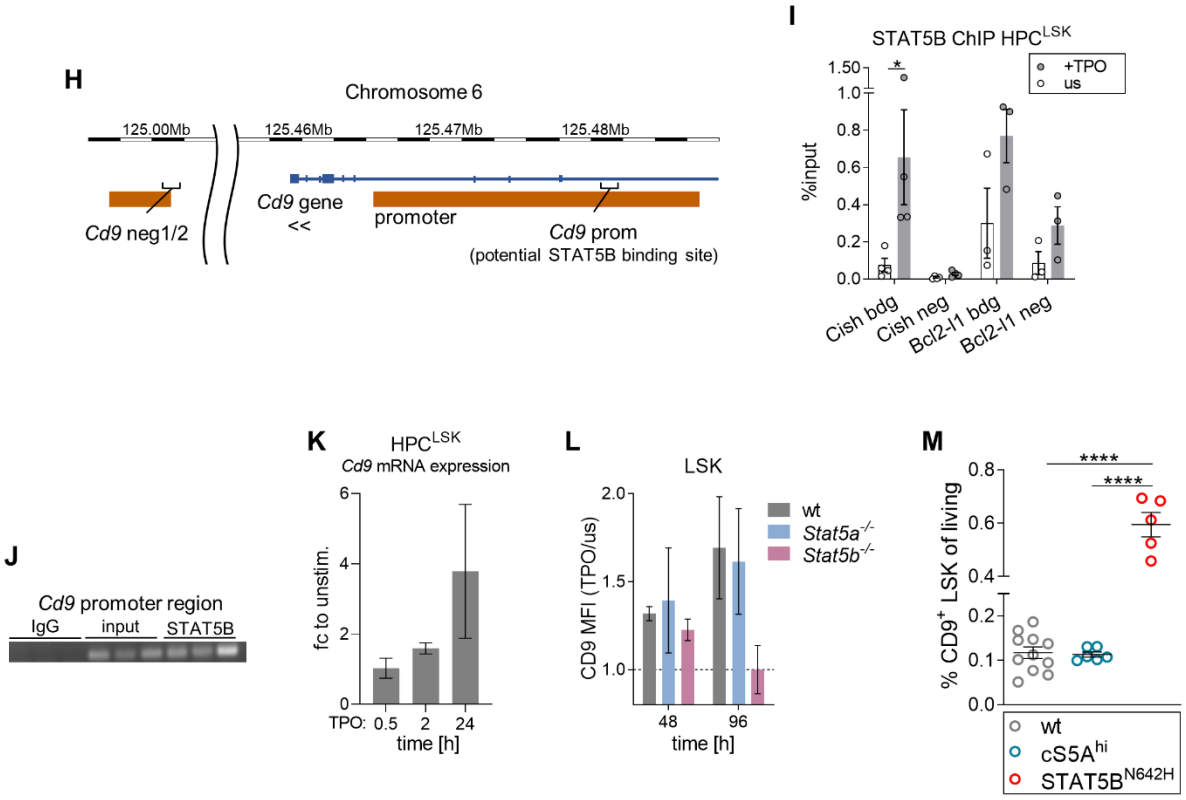

H
Chromosome 6
125.00Mb
125.46Mb
125.47Mb
125.48Mb
Cd9 neg1/2
Cd9 gene
<<
promoter
Cd9 prom
(potential STAT5B binding site)
I
STAT5B ChIP HPC^LSK
+TPO
us
%input
Cish bdg
Cish neg
Bcl2-l1 bdg
Bcl2-l1 neg
J
Cd9 promoter region
IgG
input
STAT5B
K
HPC^LSK
Cd9 mRNA expression
fc to unstim.
TPO: 0.5 2 24
time [h]
L
LSK
wt
Stat5a-/-
Stat5b-/-
CD9 MFI (TPO/us)
48
96
time [h]
M
% CD9+ LSK of living
****
****
wt
cS5A^hi
STAT5B^N642H

Figure S5: CD9 as marker and direct target of pYSTAT5B signaling.

(A) Literature-defined HSC score analysed in wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ single cell RNA-seq data, force-directed graph illustrating the top 10% cells with the highest HSC score.

(B) Expression of the STAT5B-specific HSC genes on an independent RNA-Seq dataset of FACS-sorted HSC, MPP1, MPP2, MPP3 or MPP4 cell populations from Cabezas-Wallscheid et al.[18]

(C) Survival analysis comparing *CD9*-low vs. *CD9*-high expressing patients in non-STAT5-driven cohort GSE37642 (Intermediate II).

(D) Comparison of *CD9* mRNA expression in (*left*) AML (n=542) and nBM (n=74; normal BM, GSE13159), (*middle*) LSC^{-} (n=89) and LSC^{+} (n=138; GSE76008) and (*right*) LSC (n=20), HSC (n=7) and MPP (n=7; GSE63270).

(E) *SOCS2* and *CISH* mRNA expression in *CD9*-low and *CD9*-high expressing FLT3-mut. NPM1-mut. patients (OHSU) (n=45).

(F) Flow cytometry analysis of CD9 surface levels on wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ LSK cells (n=5/genotype, mean±SEM).

(G) *Cd9* rlog transformed mRNA expression ~~levels~~ on an independent RNA-Seq dataset of FACS-sorted HSC, MPP1, MPP2, MPP3 or MPP4 cell populations from Cabezas-Wallscheid et al.[18]

(H)-(J) STAT5B ChIP-qPCR in HPC^{LSK} wt cells with or without 30 min TPO stimulation. (H) Schematic overview of the *Cd9* promoter indicating the primer binding sites designed for *Cd9* promoter (*Cd9* prom) and *Cd9* negative regions 1 and 2 (*Cd9* neg1/2). (I) Quantification of STAT5B binding to the *Cish* and *Bcl2l1* promoter (*bdg*) or negative regions (*neg*) in HPC^{LSK} wt cells with or without TPO stimulation (n≥3/condition, mean±SEM). (J) Gel electrophoresis of the *Cd9* promoter PCR product. IgG: IgG antibody ChIP was performed as control.

(K) *Cd9* mRNA expression of HPC^{LSK} wt cells 0.5, 2 and 24 h after TPO stimulation relative to unstimulated (fc – fold change, n=2/condition, mean±SEM).

(L) CD9 surface levels of wt, *Stat5a*$^{-/-}$ and *Stat5b*$^{-/-}$ *ex vivo* LSK cells cultured for 48 and 96 h with TPO relative to unstimulated (n=2/genotype and condition, mean±SEM).

(M) Relative quantification of CD9 expressing LSK cells of wt, cS5A^{hi} and STAT5B^{N642H} mice (n≥5/genotype, mean±SEM).

Levels of significance were calculated using log rank test in (C), unpaired t-test for *CISH* and Mann-Whitney for *SOCS2* in (E), 1-way ANOVA in (F), (M) and paired t-test in (I). False discovery rates (FDR) in (D) were calculated using the lmfit function of the limma package. *p < 0.05 and ****p < 0.0001.

Figure S6

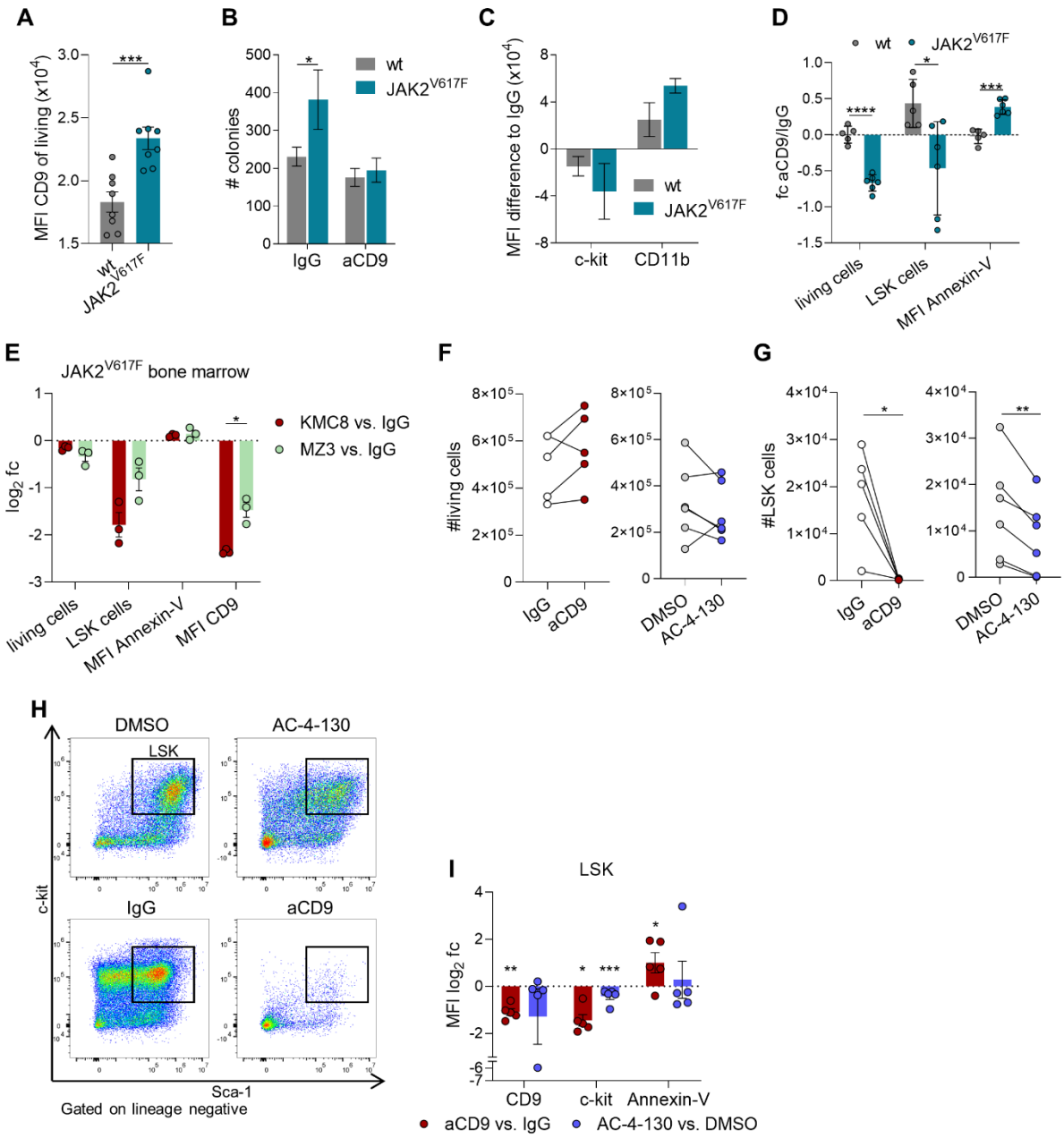

A
MFI CD9 of living (x10^4)
wt
JAK2V617F
B
colonies
wt
JAK2V617F
IgG
aCD9
C
MFI difference to IgG (x10^4)
c-kit
CD11b
D
fc aCD9/IgG
living cells
LSK cells
MFI Annexin-V
E
JAK2V617F bone marrow
log2 fc
KMC8 vs. IgG
MZ3 vs. IgG
living cells
LSK cells
MFI Annexin-V
MFI CD9
F
#living cells
IgG
aCD9
DMSO
AC-4-130
G
#LSK cells
IgG
aCD9
DMSO
AC-4-130
H
DMSO
AC-4-130
LSK
IgG
aCD9
c-kit
Sca-1
Gated on lineage negative
I
LSK
MFI log2 fc
CD9
c-kit
Annexin-V
aCD9 vs. IgG
AC-4-130 vs. DMSO

Figure S6: CD9 blocking affects STAT5 activation and impedes self-renewal.

(A) CD9 surface levels on wt and JAK2^{V617F} BM cells (n=8/genotype, mean±SEM) determined by flow cytometry.

(B)-(C) Colony formation assays of wt and JAK2^{V617F} BM cells either treated with IgG or aCD9 (5 µg/ml). (B) Colony numbers (n≥3/genotype and condition, mean±SEM) and (C) c-kit and CD11b expression levels (n=3/genotype and condition, mean±SEM).
(D) *In vitro* treatment of wt and JAK2^{V617F} BM either with IgG or aCD9. Ratio of aCD9/IgG of Annexin-V levels, living and LSK cell numbers (n≥5/genotype and condition, mean±SEM).

(E) *In vitro* treatment of JAK2^{V617F} BM either with IgG, aCD9 (KMC8) or aCD9 (MZ3) for 96h. Log_2 fc of aCD9/IgG of number of living or LSK cells and MFI of Annexin-V and CD9 levels (n=3/genotype and condition, mean±SEM).

(F)-(I) Comparison of 96 h treatment with AC-4-130 (5 µM, DMSO was used as control) to aCD9 (5 µg/ml, IgG was used as control) of JAK2^{V617F} BM cells.(F) Total cell numbers of treated cells. (G) Total LSK cell numbers of treated cells. (H) Representative flow cytometry dot plots of treated LSK cells. (I) Log_2 fc of CD9, c-kit, or AnnexinV MFI levels of aCD9- or AC-4-130 treated cells compared to the respective control (n≥5/condition, mean±SEM).

Fold change (fc). Levels of significance were calculated using unpaired t-test in (A), (B) and (D) and paired t-test in (E), (G) and (I). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

Figure S7

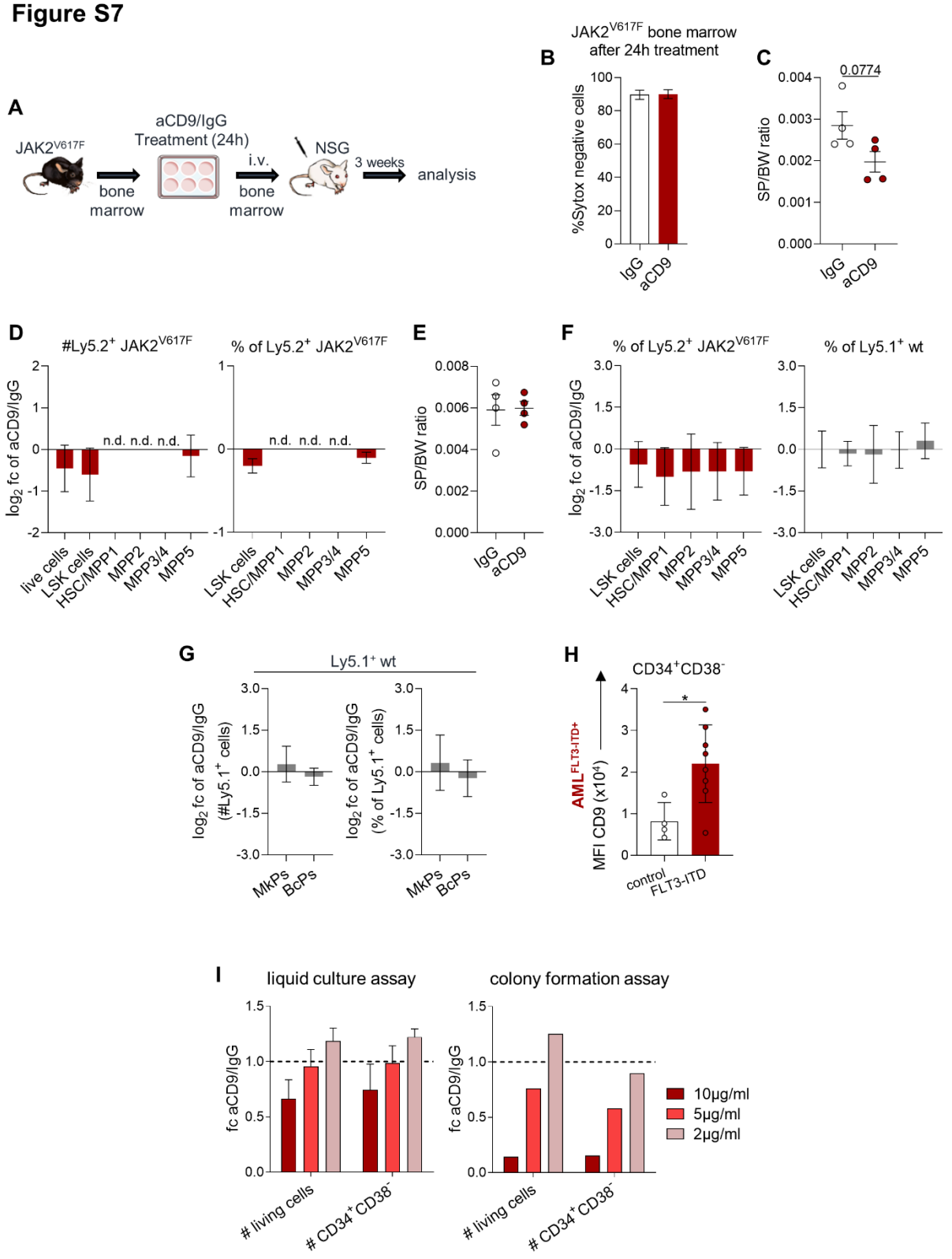

A
JAK2V617F
bone marrow
aCD9/IgG Treatment (24h)
i.v.
bone marrow
NSG
3 weeks
analysis
B
JAK2V617F bone marrow after 24h treatment
%Sytox negative cells
IgG
aCD9
C
0.0774
SP/BW ratio
IgG
aCD9
D
#Ly5.2+ JAK2V617F
% of Ly5.2+ JAK2V617F
log2 fc of aCD9/IgG
n.d. n.d. n.d.
live cells
LSK cells
HSC/MPP1
MPP2
MPP3/4
MPP5
E
SP/BW ratio
IgG
aCD9
F
% of Ly5.2+ JAK2V617F
% of Ly5.1+ wt
log2 fc of aCD9/IgG
G
Ly5.1+ wt
log2 fc of aCD9/IgG (#Ly5.1+ cells)
log2 fc of aCD9/IgG (% of Ly5.1+ cells)
MkPs
BcPs
H
CD34+CD38-
AMLFLT3-ITD+
MFI CD9 (x10⁴)
*
control
FLT3-ITD
I
liquid culture assay
colony formation assay
fc aCD9/IgG
living cells
CD34+CD38-
10µg/ml
5µg/ml
2µg/ml

Figure S7

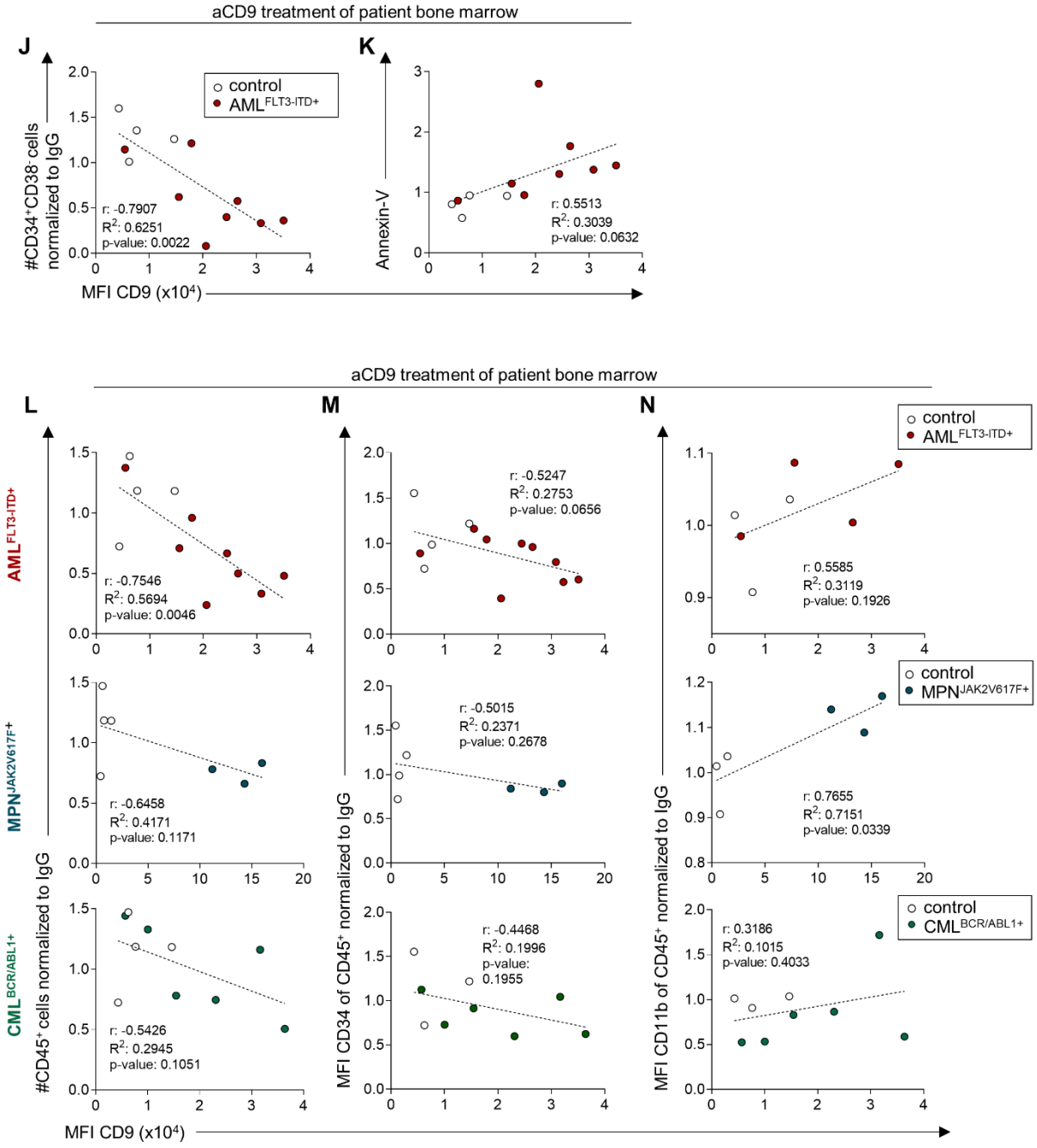

aCD9 treatment of patient bone marrow
J
#CD34+CD38- cells normalized to IgG
control
AML FLT3-ITD+
r: -0.7907
R2: 0.6251
p-value: 0.0022
MFI CD9 (x10⁴)
K
Annexin-V
r: 0.5513
R2: 0.3039
p-value: 0.0632
aCD9 treatment of patient bone marrow
L
M
N
AML FLT3-ITD+
MPN JAK2V617F+
CML BCR/ABL1+
#CD45+ cells normalized to IgG
r: -0.7546
R2: 0.5694
p-value: 0.0046
r: -0.6458
R2: 0.4171
p-value: 0.1171
r: -0.5426
R2: 0.2945
p-value: 0.1051
MFI CD34 of CD45+ normalized to IgG
r: -0.5247
R2: 0.2753
p-value: 0.0656
r: -0.5015
R2: 0.2371
p-value: 0.2678
r: -0.4468
R2: 0.1996
p-value: 0.1955
MFI CD11b of CD45+ normalized to IgG
control
AML FLT3-ITD+
r: 0.5585
R2: 0.3119
p-value: 0.1926
control
MPN JAK2V617F+
r: 0.7655
R2: 0.7151
p-value: 0.0339
control
CML BCR/ABL1+
r: 0.3186
R2: 0.1015
p-value: 0.4033
MFI CD9 (x10⁴)

Figure S7

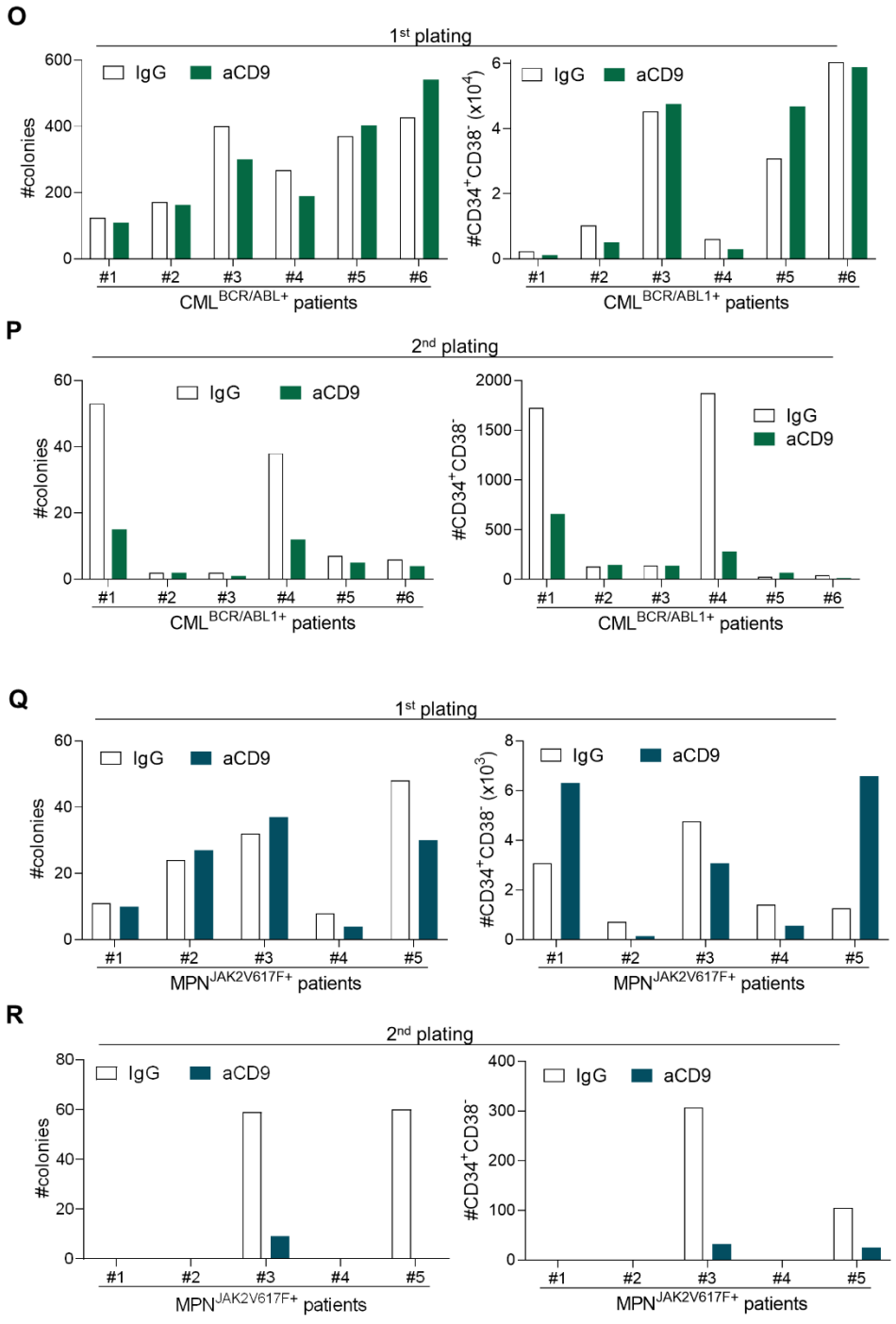

O
1st plating
IgG
aCD9
#colonies
#CD34+CD38- (x10^4)
#1
#2
#3
#4
#5
#6
CML BCR/ABL+ patients
CML BCR/ABL1+ patients
P
2nd plating
#CD34+CD38-
Q
1st plating
#CD34+CD38- (x10^3)
MPN JAK2V617F+ patients
R
2nd plating

Figure S7: CD9 as a therapeutic target for pYSTAT5-driven leukemia.

(A)-(D) $JAK2^{V617F}$ transplantation with aCD9 pre-treatment. (A) Experimental workflow. (B) Viability of IgG or aCD9 treated $JAK2^{V617F}$ BM cells prior transplantation determined by Sytox staining. (C) Spleen (SP) relative to body weight of NSG recipient mice. (D) Fold changes (aCD9 vs IgG) of $Ly5.2^+$ $JAK2^{V617F}$ donor (*left*) HSC subpopulations and total BM cell numbers and (*right*) HSC subpopulations % of $Ly5.2^+$ cells (n=4/treatment, mean±SEM). Not determined (n.d.).

(E)-(G) $JAK2^{V617F}$ transplantation with aCD9 *in vivo* treatment. (E) SP sizes relative to body weight of NSG recipient mice. (F) Log_2 fold changes (aCD9 vs IgG) of HSC/subpopulations % of $Ly5.2^+$ cells of (*left*) $Ly5.2^+$ $JAK2^{V617F}$ donor or (*right*) $Ly5.1^+$ wt NSG recipient cells. (G) Log_2 fold changes (aCD9 vs IgG) of megakaryocyte progenitor (MkPs, Lin^- c-kit^+ $Sca\text{-}1^-$ $CD150^+$ $CD41^+$) and early B-cell progenitor (BcPs, Lin^- $cKit^{mid}$ $CD93^+$ $CD127^+$ $CD19^-$) (*left*) total cell numbers and (*right*) % of $Ly5.2^+$ and $Ly5.1^+$ cells (n=4/treatment, mean±SEM, 4x 1.25 mg/kg i.v. IgG or aCD9 were applied).

(H) Analysis of CD9 levels on $CD34^+CD38^-$ cells of $AML^{FLT3\text{-}ITD}$ (n=8) and control (n=4) BM.

(I) aCD9 dosage to achieve non-toxic condition in human patient BM samples. Fold change (aCD9 relative to IgG) of $CD34^+CD38^-$ and total cell numbers in (*left*) 2-day liquid culture and (*right*) CFA using different aCD9 concentrations.

(J)-(K) *In vitro* treatment of control and patient BM cells either treated with IgG or aCD9 (2 µg/ml). XY plots showing CD9 levels and (J) $CD34^+CD38^-$ cell numbers or (K) Annexin-V levels of control (n=4) and $AML^{FLT3\text{-}ITD+}$ (n=8) patient BM MNCs.

(L)-(N) *In vitro* treatment of control and patient BM either with IgG or aCD9 (2 µg/ml). XY plots showing CD9 levels and (L) $CD45^+$ cell numbers, (M) CD34 levels or (N) CD11b levels of control (n=4) in (*top*) $AML^{FLT3\text{-}ITD+}$ (n=8), (*middle*) $MPN^{JAK2V617F+}$ (n=3) and (*bottom*) $CML^{BCR/ABL1+}$ (n=6) patient BM.

(O)-(P) Serial plating assays of $CML^{BCR/ABL1+}$ (n=6) patient BM either treated with IgG or aCD9. Quantification of (*left*) colony and (*right*) $CD34^+CD38^-$ cell numbers after the (O) primary and (P) secondary plating.

(Q)-(R) Serial plating assays of $MPN^{JAK2V617F+}$ (n=5) patient BM either treated with IgG or aCD9. Quantification of (*left*) colony and (*right*) $CD34^+CD38^-$ cell numbers after the (Q) primary and (R) secondary plating.

Levels of significance were calculated using unpaired t-test in (C) and (H). Levels of significance and correlation were calculated using Pearson in (J)-(N).

Supplementary References

1. Doma E, Mayer IM, Brandstoetter T, et al. A robust approach for the generation of functional hematopoietic progenitor cell lines to model leukemic transformation. *Blood Adv.* 2021;5(1):39–53.
2. Pinto Do Ó P, Kolterud Å, Carlsson L. Expression of the LIM-homeobox gene LH2 generates immortalized Steel factor-dependent multipotent hematopoietic precursors. *EMBO J.* 1998;17(19):5744–5756.
3. Hoelbl A, Schuster C, Kovacic B, et al. Stat5 is indispensable for the maintenance of *bcr/abl* -positive leukaemia. *EMBO Mol. Med.* 2010;2(3):98–110.
4. Galvin A, Weglarz M, Folz-Donahue K, et al. Cell Cycle Analysis of Hematopoietic Stem and Progenitor Cells by Multicolor Flow Cytometry. *Curr. Protoc. Cytom.* 2019;87(1):e50.
5. Berger A, Hoelbl-Kovacic A, Bourgeais J, et al. PAK-dependent STAT5 serine phosphorylation is required for BCR-ABL-induced leukemogenesis. *Leukemia*. 2014;28(3):629–641.
6. Schmidt L, Heyes E, Scheiblecker L, et al. CEBPA-mutated leukemia is sensitive to genetic and pharmacological targeting of the MLL1 complex. *Leukemia*. 2019;33(7):1608–1619.
7. Maurer B, Nivarthi H, Wingelhofer B, et al. High activation of STAT5A drives peripheral T-cell lymphoma and leukemia. *Haematologica*. 2020;105(2):435–447.
8. Warsch W, Grundschober E, Berger A, et al. STAT5 triggers BCR-ABL1 mutation by mediating ROS production in chronic myeloid leukaemia. *Oncotarget*. 2012;3(12):1669–87.
9. Wingelhofer B, Maurer B, Heyes EC, et al. Pharmacologic inhibition of STAT5 in acute myeloid leukemia. *Leukemia*. 2018;32(5):1135–1146.
10. Hoelbl A, Kovacic B, Kerenyi MA, et al. Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. *Blood*. 2006;107(12):4898–906.
11. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21.
12. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 2018;36(5):411–420.
13. Dahlin JS, Hamey FK, Pijuan-Sala B, et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. 2018;131(21):e1–e11.
14. Scialdone A, Natarajan KN, Saraiva LR, et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*. 2015;85:54–61.
15. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 2013;6(269):pl1.
16. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics*. 2010;11(2):242–53.
17. Nguyen CH, Schlerka A, Grandits AM, et al. IL2RA Promotes Aggressiveness and Stem Cell–Related Properties of Acute Myeloid Leukemia. *Cancer Res.* 2020;80(20):4527–4539.
18. Cabezas-Wallscheid N, Klimmeck D, Hansson J, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*. 2014;15(4):507–522.
19. Pei W, Shang F, Wang X, et al. Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. *Cell Stem Cell*. 2020;27(3):383-395.e8.
20. Giladi A, Paul F, Herzog Y, et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat. Cell Biol.* 2018;20(7):836–846.
21. Wilson NK, Kent DG, Buettner F, et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*. 2015;16(6):712–24.
22. Zhang X, Lan Y, Xu J, et al. CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 2019;47(D1):D721–D728.

Annex

Part I: Cropped WB membranes including all loading controls

Figure 3E

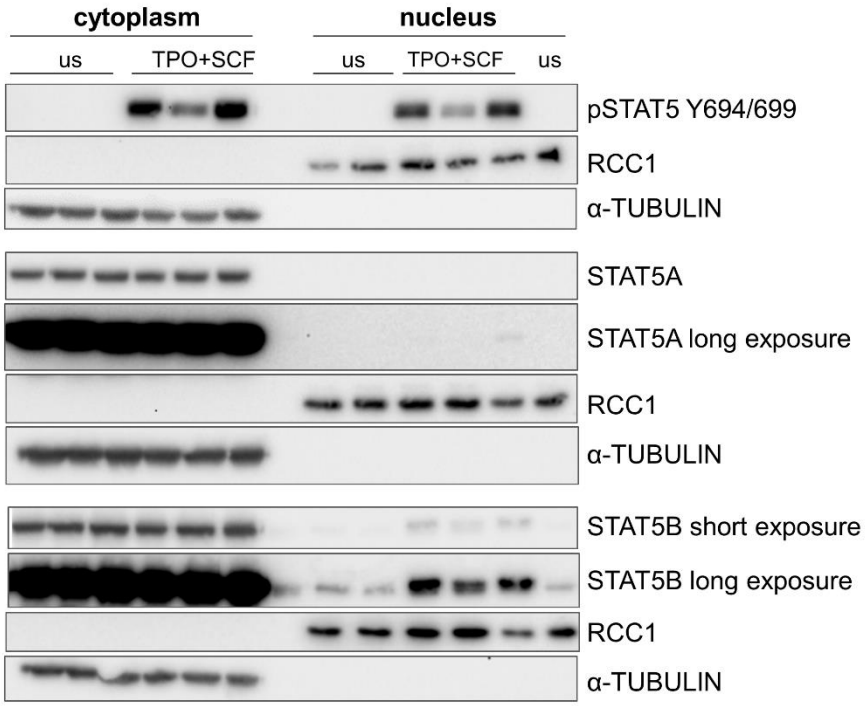


Figure S3E

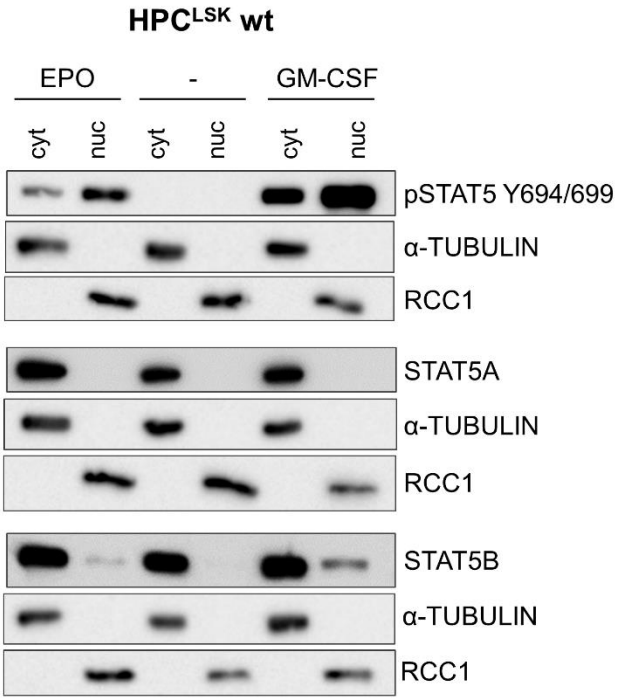
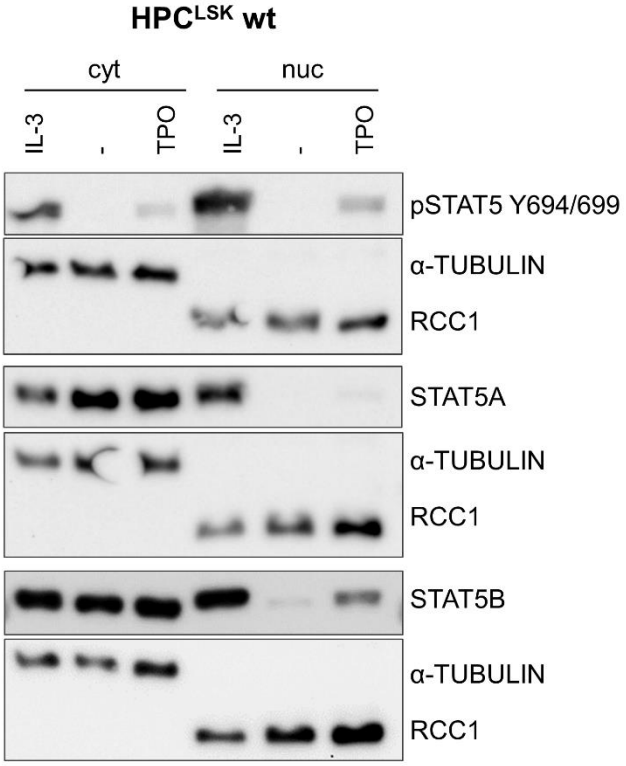
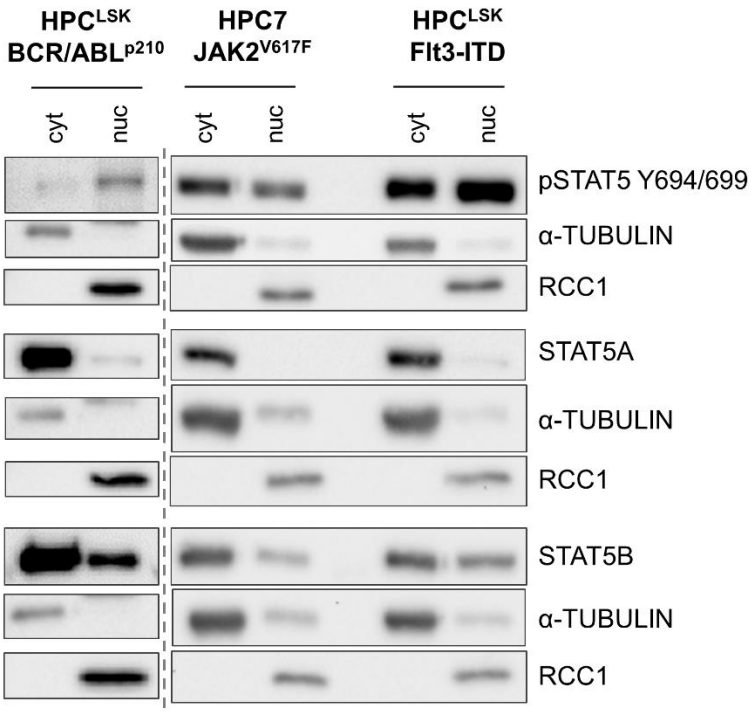


Figure S3F



53

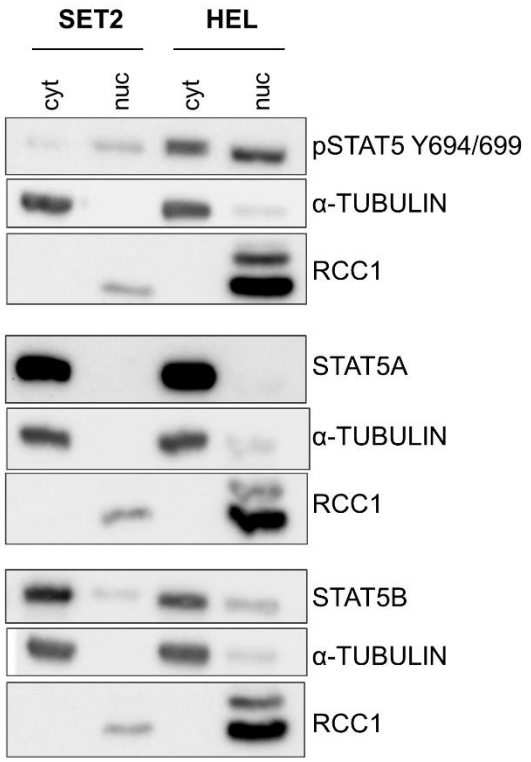
Figure 4C



54

55

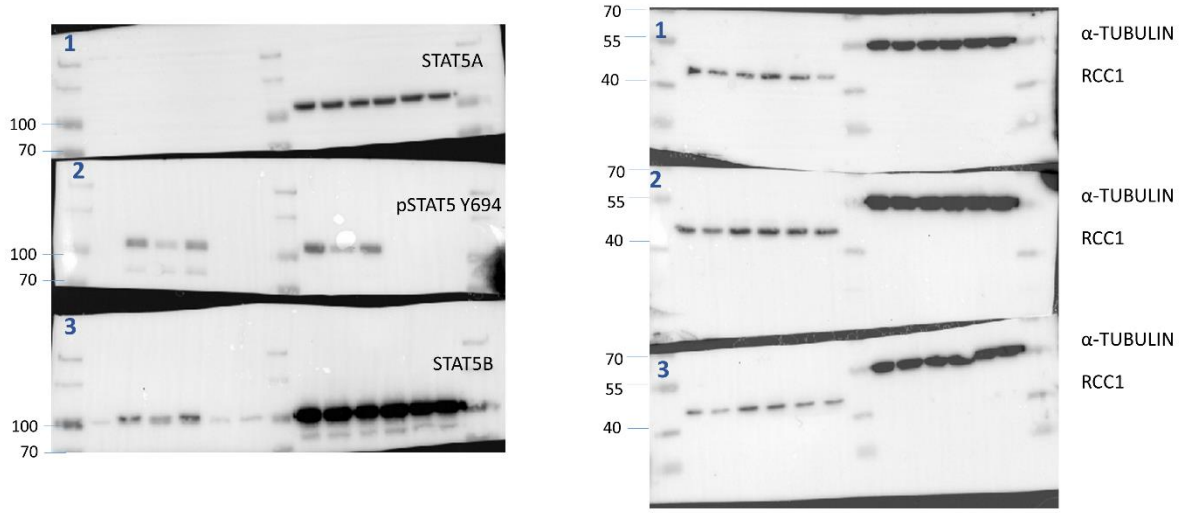
Figure S4F


SET2
HEL
cyt
nuc
cyt
nuc
pSTAT5 Y694/699
α-TUBULIN
RCC1
STAT5A
α-TUBULIN
RCC1
STAT5B
α-TUBULIN
RCC1

Part II: Uncropped WB membranes

Figure 3E

3 8%-polyacrylamide gels have been loaded with the same lysates. The membrane has been cut at the 70 kDa protein ladder signal. The respective loading control for each blot is shown on the right side.



Supplementary Figure 3

3 8%-polyacrylamide gels have been loaded with the same lysates. The membrane has been cut at the 70 kDa protein ladder signal. The respective loading control for each blot is shown on the right side.

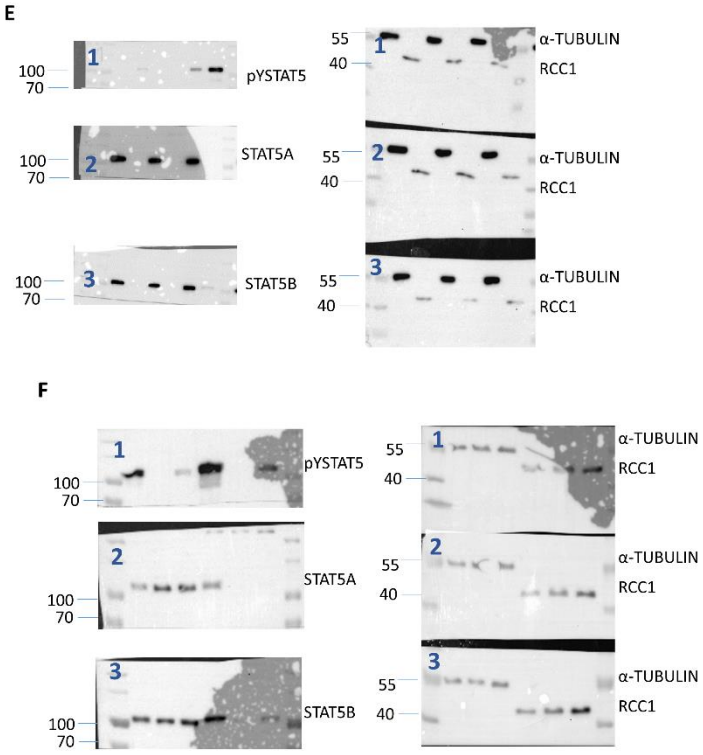
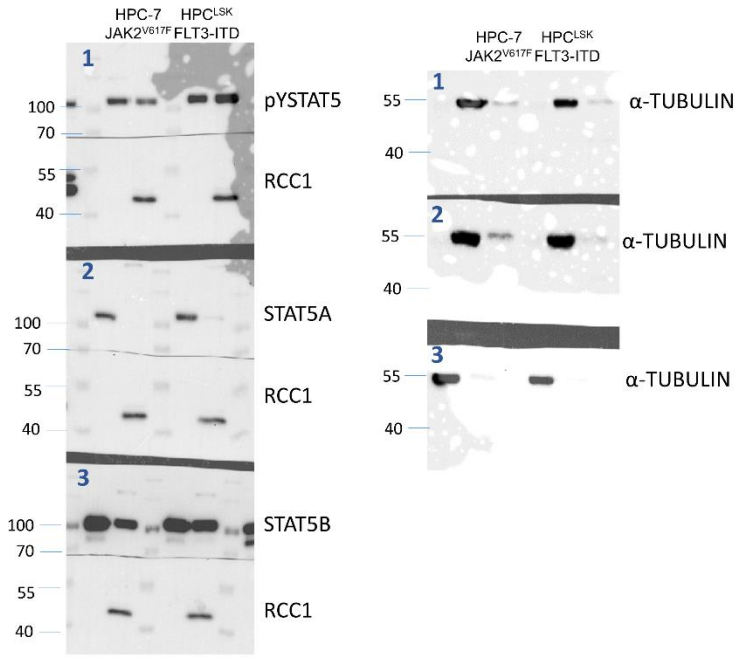
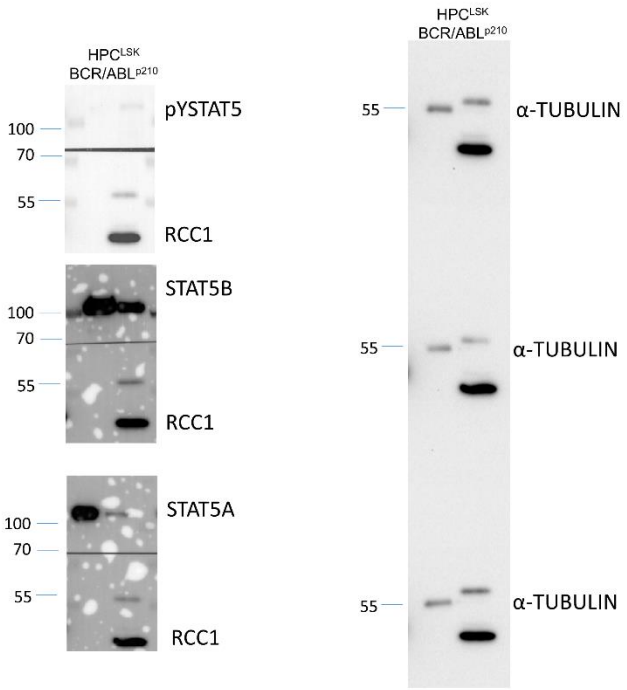


Figure 4C

3 8%-polyacrylamide gels have been loaded with the same lysates. The membrane has been cut at the 70 kDa protein ladder signal. The respective loading controls are shown.

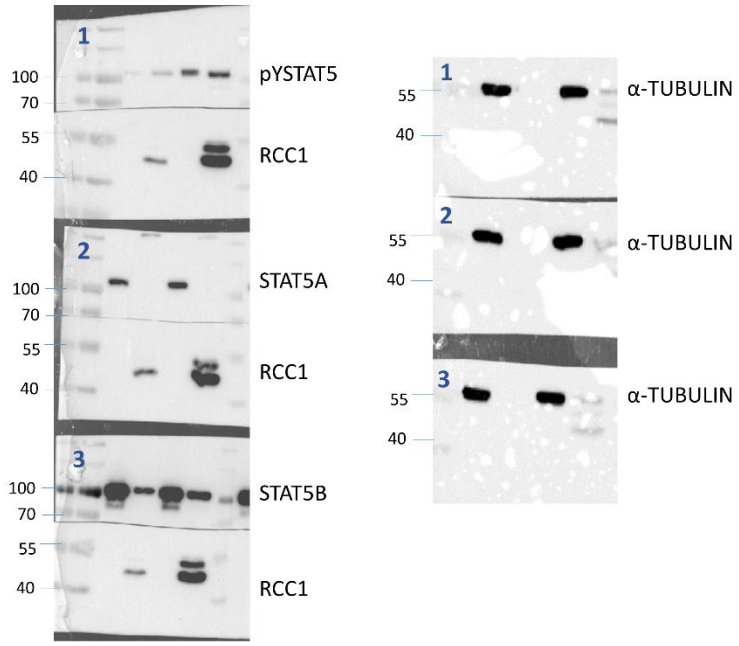




60

Supplementary Figure 4F

3 8%-polyacrylamide gels have been loaded with the same lysates. The membrane has been cut at the 70 kDa protein ladder signal. The respective loading controls are shown.



61