Genetic barcoding systematically comparing genes in del(5q) MDS reveals a central role for CSNK1A1 in clonal expansion

Supplemental Methods

Animal Studies

Csnk1a1 conditional knockout mice in which exon 3 is flanked by loxP sites (floxed Csnk1a1 or Csnk1a1fl/fl) were crossed to Mx1-Cre transgenic mouse to create Csnk1a1fl/+Mx1-Cre+ as described before. Mx1-Cre mouse strain was ordered from Jackson laboratory (JAX stock #002527). C57Bl6 ApcloxP/B6;129-Egr1tm1jmi/J were a kind gift from Tyler Jacks lab. In this conditional knockout, LoxP sites are flanking all 15 protein coding exons in the Apc locus to generate a Cre regulated null allele. Egr1tm1jmi/J were ordered from Jackson laboratories. In these mice, a neomycin cassette is inserted upstream of the DNA binding domain in the Egr1 locus, thereby disrupting Egr1 protein expression. B6.129-Ctnnb1tm2kem/KnwJ mice were a kind gift from Scott Armstrong’s lab. These mice carry two loxP sites flanking a region of the Ctnnb1 gene sequence from exon 2 through exon 6. Compound haploinsufficiency was achieved by crossing the above mentioned conditional knockout mice to each other for the resulting combinations: Csnk1a1fl/+ Egr1-/+ Mx1-Cre+, Csnk1a1fl/+ Apcfl/+ Mx1-Cre+, Csnk1a1fl/+ Ctnnb1fl/+ Mx1-Cre+.

Mouse experiments were performed according to an Institutional Animal Care and Use Committee (IACUC)-approved protocol at Children’s Hospital Boston and protocols approved by the Central Animal Committee (Centrale Commissie Dierproeven; CCD; the Netherlands under approval number AVD1010020173387).

Competitive Transplants

In competitive bone marrow transplantation studies, 2x10^6 freshly isolated bone marrow cells were harvested, equally mixed with 2x10^6 freshly isolated CD45.1+ bone marrow competitor cells and transplanted through tail vein injection into lethally irradiated (10.5 Gy) CD45.1+ recipient mice. Donor cell chimerism was first determined four weeks after the transplant by flow cytometry analysis from peripheral blood samples. Mice were then subjected to three rounds of poly(I:C) in one week to induce the excision of the floxed alleles (5µg/g body weight). We confirmed that the gene expression of these respective genes was reduced to 50% (haploinsufficient) only after the injection of Poly(I:C) but not after transplantation (Figure 1).
Figure S1: Quantitative qRT-PCR of peripheral blood cells taken before transplant, 4 weeks after transplant (before polyIC) and 4 weeks after poly(IC).

Donor blood cell chimerism was determined every four weeks.

In the competitive transplant with low-dose inflammatory stimulation, we let all mice recover for two weeks after the first three poly(I:C) injections to induce the excision (see above). We split mice into two groups per genotype and started injections with biweekly 5µg/g body weight poly(I:C) or PBS for one block of two weeks and two blocks of one week, leaving two weeks of recovery in between all blocks (Figure 7A).

Barcoding Library and Barcoding Transplant
A 33-bp DNA barcode was generated consisting of a defined 6-bp library sequence at the 5’ followed by a large, random 27-bp (single cell-specific) sequence. In the competitive assay of transplantation of different genotypes, each genotype was labeled with a specific library ID. The 33-bp DNA barcode sequence was cloned into the non-expressing region of the lentiviral pLKO (GFP)-vector. The U6-promoter was truncated to avoid any residual expression. Linkers complementary to the primers required for the Illumina GA II high-throughput sequencing were appended at the end of the barcode sequence in the lentiviral construct.

Ckit+ hematopoietic stem and progenitor cells were isolated from the bone marrow of the different haploinsufficient or compound haploinsufficient mouse lines described above using the MACS ckit-micro bead-separation kit and cultured in SFEM (StemSpan) supplemented with 50ng/ml SCF and 50ng/ml murine thrombopoietin (both Peprotech) for 16 16 hours and then transduced with concentrated and titered lentiviral supernatant in the presence of 2 µg/ml Polybrene. Multiplicity of infection (MOI <0.5) was set such that less than half of the donor ckit enriched cells would be GFP expressing after infection, to prevent multiple barcodes per cell. 24 hours after infection, cultured and transduced cells were sorted for GFP+ viable cells (separately for each genotype). Sort-purified cells from the different genotypes were counted using a Neubauer chamber and then mixed in in equal ratios. From this mixture, 500,000
cells were transplanted in equal volumes (150µl/mouse) into the lateral tail veins of five lethally irradiated recipients (10.5 Gy). For the secondary transplantation, 3 x 10^6 whole bone marrow cells from each mouse were transplanted into lethally irradiated secondary recipients.

Hematopoietic lineages, such as granulocytes (CD11b+, Gr1+), B-cells (CD11b-, CD19+), T-cells (CD11b-, CD3+) and hematopoietic progenitors (lin-, ckit+, sca1+ LSK cells) were analyzed and purified (Aria, Becton-Dickinson) from bone marrow (26 weeks) and peripheral blood (week 4, 8, 12, 26). Defined cell populations were sorted into DNA lysis buffer. Initial experiments showed that 2500 cells are sufficient for genomic DNA isolation used for sequencing.

Genomic DNA was isolated from purified cell populations using the DNA micro blood and tissue kit (Qiagen). A PCR step (Illumina Sequencing Primers) was used to amplify the DNA barcode and in the same step add linkers necessary for the high-throughput sequencing. The PCR product at the right size was cut out from 3% agar gels. Sequencing was performed using the Illumina GAII sequencer of the RNAi consortium at the Broad Institute.

First, sequences were combined allowing for mismatches up to 1bp in total. Next, the 6 first sequences were used to detect the genotypes and the remaining 27 bases were used to count the barcodes. See Table S1 and S2 for barcode information and barcode count matrix. Clustering of barcoding libraries was based on the Morisita Horn index.

**Single-Cell Sequencing**

Experimental mice were transplanted with 3x10^6 Mx1Cre+ or Csnk1a1-/-Mx1Cre+ bone marrow cells. Excision of floxed allele was induced four weeks after transplantation and full excision was tested by PCR. At 24 weeks after transplantation mice were sacrificed, bone marrow was isolated and lineage depleted using magnetic bead separation with indirect lineage depletion kit (Milteny Biotec). Lin negative bone marrow was washed and labeled with antibodies for 15 min at 4°C. Cells were analyzed and viable, lin-, Sca1+, ckit+ (LSK) hematopoietic stem and progenitor cells were sorted into PBS/10% FCS using a BD FACS Aria II cell sorter.

Sorted cells were immediately processed using 10x Genomics Chromium Controller to encapsulate each cell into emulsion droplets. ScRNA-seq libraries were constructed according to the manufacturer's protocol using Chromium Single Cell 3' Reagent Kits (v2, v3.1). Quality was evaluated using the Agilent TapeStation System. Pooled Libraries were sequenced on a Novaseq 6000 platform (Illumina). Reads were aligned to the mouse genome mm10 and sample demultiplexing, barcode and gene count processing was performed using CellRanger V.2.1.1. Filtering of doublets and low quality cells was performed using R package Seurat v3.1.4 (>5% mitochondrial counts, unique feature counts over 5000 or less than 800, or total UMI
over 20000 and less than 1000). This left 76,963 high quality cells. We removed ribosomal genes, i.e. genes matching the regular expression “Rps” or “Rpl” to control sample specific bias. We regressed out cell cycle S-phase/G2M-Phase differences, the proportion of mitochondrial gene counts and UMI counts and performed a log-normalization of read counts using Seurat. Samples were integrated with Seurat (v3.1.4) anchors based method. We used the first 20 principal components (PC) for anchors identification.

Unsupervised clustering was performed using the first 30 PCs with a shared nearest neighbor graph with resolution 0.3 and error bound of 0.5 when performing nearest neighbor search. UMAP representations were generated using the parameter min.dist equal to 0.75. We used FindMarkers gene function using a Wilcoxon Rank Sum test with logFC cut-off of 0.25 and adjusted p value < 0.05 to find only positive cluster specific markers expressed in at least 25% of cells. Gene set enrichment was tested for each cluster using CAMERA from edgeR (Version 3.26.7; 7) Hallmark gene sets to be tested were obtained from Molecular Signatures Database (MSigDB) gene sets using R package msigdb (Version 7.2.1). PROGENy analysis was conducted as described using the Progeny package (R, V1.11.0) for murine datasets 8,9. For each interrogated cluster the full generated list of genes differentially expressed between Csnk1a1 haploinsufficient cells and wt control without fold-change cutoff was used. The scala products of obtained log2-fold changes and PROGENy weights were calculated for respective pathway scores. To estimate significance label permutation was performed 10000 times to generate a Null distribution for each pathway. DoRoThea analysis was conducted using the DoRoThea R package (V1.0.0) for murine datasets 8,9. For each interrogated cluster the full generated list of genes differentially expressed between Csnk1a1 haploinsufficient cells and wt control without fold change cutoff was used. We estimated TF activities as a proxy of the expression levels of the targeted genes using the aREA method from the VIPER R package (V.1.20.0) 10 based on the average ranks of the targets. The NES was used as a measure of relative TF activity and only TF regulons with at least five targets were tested. Datasets are deposited in gene expression omnibus (GEO accession number: GSE165395).

**Mathematical Model**

**Model structure**

The quantitative mathematical model used here is an extension of the ordinary differential equation model from Stiehl et al. 11,12, which has been validated based on data and is in good agreement with clinical observations. Here, we calibrate the model to murine hematopoiesis and extend it to account for the excision of floxed alleles. Due to the high number of cells, the use of differential equations is justified 11–14.
For each genotype the model considers mitotic cells in the bone marrow and mature granulocytes in the blood stream. The two-compartmental design is supported by the following considerations: (i) Mathematical \(^1\) and computational \(^2\) results demonstrate that mature cell dynamics observed during clonal evolution are insensitive to the number and the parameters of the considered progenitor stages. (ii) The two-compartmental structure can faithfully reproduce clinical and experimental observations \(^1\). \(^2\)–\(^4\). (iii) We intend to keep the number of free parameters in a reasonable balance to the available data to avoid overfitting.

The number of mitotic wild type cells per gram of body weight at time \(t\) is denoted by \(c_{WT,1}(t)\) and that of mature wildtype neutrophils by \(c_{WT,2}(t)\). For the considered genotypes, the respective cell numbers are denoted by \(c_{APC,1}(t)\), \(c_{EGR,1}(t)\), \(c_{CSNK,1}(t)\), \(c_{APC,2}(t)\), \(c_{EGR,2}(t)\) and \(c_{CSNK,2}(t)\) before excision of the floxed alleles and by \(c_{APC,2}(t)\), \(c_{EGR,2}(t)\), \(c_{CSNK,2}(t)\) afterwards. Before allele excision all cells are assumed to have the same kinetic properties as the respective wildtype cells.

The division rates of mitotic cells at time \(t\) are denoted by \(p_{WT}(t)\), \(p_{APC}(t)\), \(p_{EGR}(t)\), \(p_{CSNK}(t)\) for the respective genotypes. Consequently, the flux to mitosis for wild type cells at time \(t\) is given by \(p_{WT}(t)\cdot c_{WT,1}(t)\), and analogously for all other genotypes. During division these cells give rise to \(2\cdot p_{WT}(t)\cdot c_{WT,1}(t)\) cells. The fraction \(a_{WT}(t)\) of these cells remain mitotic cells in the bone marrow and the fraction \((1-a_{WT}(t))\) differentiate into mature granulocytes. The parameter \(a_{WT}\) is referred to as fraction of self-renewal \(^1\), \(^2\)–\(^4\) and is equivalent to the probability that offspring originating from division belong to the same cell type as their parent cell. The fractions of self-renewal for the other genotypes are denoted by \(a_{APC}\), \(a_{EGR}\) and \(a_{CSNK}\). They quantify how often cells divide in average before further differentiation \(^5\). Mature granulocytes die at a constant rate \(d\).

Feedback regulations

Based on experimental data \(^2\)–\(^4\) we assume that proliferation rates and fractions of self-renewal depend on non-linear feedbacks. For the proliferation rates we use a feedback that is based on the physiology of G-CSF, the major cytokine of granulopoiesis \(^5\), \(^6\). The feedback signal at time \(t\) depends on the total number \(N_2(t)\) of mature granulocytes. We set \(s_p(t) = 1/(1 + k_p N_2(t))\) and \(p_{WT}(t) = p_{WT,max}s_p(t)\), \(p_{APC}(t) = p_{APC,max}s_p(t)\), \(p_{EGR}(t) = p_{EGR,max}s_p(t)\) and \(p_{CSNK}(t) = p_{CSNK,max}s_p(t)\). These expressions are derived from a quasi-steady state approximation of G-CSF production and receptor-mediated endocytosis \(^5\), \(^6\). The use of such Hill-type functions is very common and leads to realistic dynamics \(^1\), \(^2\)–\(^4\), \(^5\)–\(^8\)–\(^10\), \(^19\)–\(^21\), \(^28\)–\(^31\).

For the regulation of self-renewal, we set \(s_a(t) = 1/(1 + k_a N_1(t))\) and \(a_{WT}(t) = a_{WT,max}s_a(t), \ a_{APC}(t) = a_{APC,max}s_a(t), \ a_{EGR}(t) = a_{EGR,max}s_a(t)\) and \(a_{CSNK}(t) = a_{CSNK,max}s_a(t)\). \(N_1(t)\) denotes the total number of mitotic cells at time \(t\). This takes into account that self-renewal is regulated by local signals in the bone-marrow \(^2\)–\(^4\).

Full model
After Poly(I:C) induction at day 28 cells transit to the new phenotype at a rate \(\alpha_0\). We obtain:

\[
\begin{aligned}
\text{WT:} & \quad \frac{d}{dt} c_{WT,1}(t) = (2a_{WT,max}s_a(t) - 1)p_{WT,max}sp(t)c_{WT,1}(t) \\
& \quad \frac{d}{dt} c_{WT,2}(t) = 2\left(1 - a_{WT,max}s_a(t)\right)p_{WT,max}^2sp(t)c_{WT,1}(t) - dc_{WT,2}(t) \\
\text{APC non-induced:} & \quad \frac{d}{dt} \tilde{c}_{APC,1}(t) = (2a_{APC,max}s_a(t) - 1)p_{APC,max}^2sp(t)c_{APC,1}(t) \\
& \quad \frac{d}{dt} \tilde{c}_{APC,2}(t) = 2\left(1 - a_{APC,max}s_a(t)\right)p_{APC,max}^2sp(t)c_{APC,1}(t) - dc_{APC,2}(t) \\
\text{APC induced:} & \quad \frac{d}{dt} \tilde{c}_{APC,1}(t) = \alpha(t)\tilde{c}_{APC,1}(t) + (2a_{APC,max}s_a(t) - 1)p_{APC,max}sp(t)c_{APC,1}(t) \\
& \quad \frac{d}{dt} \tilde{c}_{APC,2}(t) = \alpha(t)\tilde{c}_{APC,2}(t) + 2\left(1 - a_{APC,max}s_a(t)\right)p_{APC,max}sp(t)c_{APC,1}(t) - dc_{APC,2}(t) \\
\text{EGR non-induced:} & \quad \frac{d}{dt} \tilde{c}_{EGR,1}(t) = (2a_{EGR,max}s_a(t) - 1)p_{EGR,max}sp(t)c_{EGR,1}(t) - \alpha(t)\tilde{c}_{EGR,1}(t) \\
& \quad \frac{d}{dt} \tilde{c}_{EGR,2}(t) = 2\left(1 - a_{EGR,max}s_a(t)\right)p_{EGR,max}sp(t)c_{EGR,1}(t) - dc_{EGR,2}(t) \\
\text{EGR induced:} & \quad \frac{d}{dt} \tilde{c}_{EGR,1}(t) = \alpha(t)\tilde{c}_{EGR,1}(t) + (2a_{EGR,max}s_a(t) - 1)p_{EGR,max}sp(t)c_{EGR,1}(t) \\
& \quad \frac{d}{dt} \tilde{c}_{EGR,2}(t) = \alpha(t)\tilde{c}_{EGR,2}(t) + 2\left(1 - a_{EGR,max}s_a(t)\right)p_{EGR,max}sp(t)c_{EGR,1}(t) - dc_{EGR,2}(t) \\
\text{CSNK non-induced:} & \quad \frac{d}{dt} \tilde{c}_{CSNK,1}(t) = (2a_{CSNK,max}s_a(t) - 1)p_{CSNK,max}sp(t)c_{CSNK,1}(t) - \alpha(t)\tilde{c}_{CSNK,1}(t) \\
& \quad \frac{d}{dt} \tilde{c}_{CSNK,2}(t) = 2\left(1 - a_{CSNK,max}s_a(t)\right)p_{CSNK,max}sp(t)c_{CSNK,1}(t) - dc_{CSNK,2}(t) \\
\text{CSNK induced:} & \quad \frac{d}{dt} c_{CSNK,1}(t) = \alpha(t)c_{CSNK,1}(t) + (2a_{CSNK,max}s_a(t) - 1)p_{CSNK,max}sp(t)c_{CSNK,1}(t) \\
& \quad \frac{d}{dt} c_{CSNK,2}(t) = \alpha(t)c_{CSNK,2}(t) + 2\left(1 - a_{CSNK,max}s_a(t)\right)p_{CSNK,max}sp(t)c_{CSNK,1}(t) - dc_{CSNK,2}(t) \\
\text{feedbacks:} & \\
& \quad sp(t) = \frac{1}{1 + k_p(c_{WT,2}(t) + c_{APC,2}(t) + \tilde{c}_{EGR,2}(t) + c_{CSNK,2}(t) + \tilde{c}_{APC,2}(t) + \tilde{c}_{EGR,2}(t) + \tilde{c}_{CSNK,2}(t))} \\
& \quad sa(t) = \frac{1}{1 + k_a(c_{WT,1}(t) + c_{APC,1}(t) + \tilde{c}_{EGR,1}(t) + c_{CSNK,1}(t) + \tilde{c}_{APC,1}(t) + \tilde{c}_{EGR,1}(t) + \tilde{c}_{CSNK,1}(t))} \\
& \quad c_{WT,1}(0) = \tilde{c}_{APC,1}(0) = \tilde{c}_{EGR,1}(0) = \tilde{c}_{CSNK,1}(0) = c_0; \quad c_{WT,2}(0) = \tilde{c}_{APC,2}(0) = \tilde{c}_{EGR,2}(0) = \tilde{c}_{CSNK,2}(0) = 0 \\
& \quad c_{APC,1}(0) = c_{EGR,1}(0) = c_{CSNK,1}(0) = c_{APC,2}(0) = c_{EGR,2}(0) = c_{CSNK,2}(0) = 0 \\
& \quad \alpha(t) = \begin{cases} 
\alpha_0, & \text{for } 28 \text{ days } \leq t \leq 35 \text{ days} \\
0, & \text{for } t > 35 \text{ days}
\end{cases}

\text{Model calibration}

Cell counts are quantified as cells per gram of body weight. Parameters of wild type cells and feedbacks are calibrated such that the model leads to realistic steady state cell counts and describes granulocyte engraftment after transplantation. Proliferation rates and fractions of self-renewal of Apc, Egr1 and Csnk1a1 haplo-insufficient cells are fitted based on the experimental data as described in the next section.
In our mice the peripheral leukocyte count is 5000-12000 per µl of blood, one third of which are neutrophils (approx. 1600-4000 per µl of blood). Assuming a blood volume of 90 ml per kg of body weight, this corresponds to cWT,2=25.2x10^4 neutrophils per gram of body weight. A body weight of 20 gram and a total nucleated cell count of 5x10^8, correspond to 2.5x10^7 nucleated cells per gram of body weight. Of these 8.5% are LT-HSC, ST-HSC, GMP, myeloblasts, promyelocytes and myelocytes, which implies cWT,1=2.1x10^6 cells per gram of body weight. The clearance rate of peripheral neutrophils is set to 1.33/day, corresponding to a half-life of 0.52 days.

Denote steady state cell counts as \( \bar{c}_{WT,1} \) and \( \bar{c}_{WT,2} \) and the feedback signals as \( \bar{s}_p \) and \( \bar{s}_m \). Then it holds: 
\[
0 = (2a_{WT}\bar{s}_a - 1)p_{WT}\bar{s}_p\bar{c}_{WT,1} \quad \text{and} \quad 0 = 2(1 - a_{WT}\bar{s}_a)p_{WT}\bar{s}_p\bar{c}_{WT,1} - d_2\bar{c}_{WT,2},
\]
which implies \( \bar{s}_a = 1/(2a_{WT}) \), \( k_a = (2a_{WT} - 1)/\bar{c}_{WT,1} \) and \( p_{WT}\bar{s}_p = d_2\bar{c}_{WT,2}/\bar{c}_{WT,1} \approx 0.16/day \). The latter corresponds to a steady state division rate between HSC and myeloblasts, which is plausible for an averaged mitotic compartment. Taking into account that under G-CSF stimulation cell proliferation rates increase by about 50% [33], we set \( k_p = 1.984 \times 10^{-6} \) and \( p_{WT,max} = 0.24/day \).

We set \( a_{WT} \) such that the model recapitulates neutrophil engraftment after transplantation. Transplantation of 3x10^6 whole bone marrow cells, 8.5% of which contribute to granulopoiesis, imply a cell dose of 1.3x10^4 per gram of body weight. For \( a_{WT,max} = 0.82 \) we observe 600 neutrophils per µl after 4 weeks, as in experiments. This implies \( k_a = 3.048 \times 10^{-7} \). The transition rate \( \alpha_0 \) describing the change of phenotype after Poly(I:C) is set to 0.5/day, implying that 95% of the floxed alleles are excised 2 days after the last Poly(I:C) dose, i.e., 6 days after the first dose. We assume that the poly(I:C) activity lasts for approximately one week, which is in agreement with the observations from reference 44.

**Parameter estimation**

We use *fmincon* from MATLAB (Version R2021a) and a multi-start approach (5000 starts) with latin hypercube sampling to fit \( a_{APC}, a_{APC}, a_{EGR}, a_{CSNK}, p_{APC}, p_{EGR}, p_{CSNK} \) (least squares). For the fitting the transplanted 500000 cKit+ cells are equally distributed among the genotypes. To calculate \( c_0 \) we assume that 55% of these cells contribute to neutrophil production. We fit the proportion of the genotypes in the experiments to the proportion of the genotypes among peripheral neutrophils in the model. We use blood samples at 4, 8, 12, 26 weeks and the LSK sample at 26 weeks. To detect and quantify inter-individual differences in cell parameters we fit the model to each mouse separately. Confidence intervals were calculated with *lsqcurvefit* and *nlparci* using all available data of the respective mouse. The obtained best fits and the 95% confidence intervals are depicted in Figure 2 B-C. The calculation of the confidence intervals is based on the residual of the fit and properties of the model.

**Simulations**

We use the solver *ode45* with adaptive time steps. To study the impact of changes in cell parameters we repeat the simulations of the fitted models with 1000 random
perturbations of the cell properties in a range of ±5% and in a range of ±20%, sampled from a uniform distribution. To study the impact of unequal transplanted/engraftment cell doses of the different genotypes, we run simulations, where we randomly perturb the cell numbers of each genotype by ±10%, ±20% and ±60% sampled from a uniform distribution.

**Quantification and Statistical Analysis**

Statistical analysis - excluding that for single cell RNAseq data - was conducted using GraphPad Prism v8.0 and v9.0. Unless otherwise specified, data are presented as mean ± SEM. Significance is depicted throughout the manuscript as follows: p ≤ 0.05: *; p ≤ 0.01: **; p ≤ 0.0001: ***

**References**

1. Schneider RK, Ademà V, Heckl D, et al. Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell*. 2014;26(4):509-520.
2. Cheung AF, Carter AM, Kostova KK, et al. Complete deletion of Apc results in severe polyposis in mice. *Oncogene*. 2010;29(12):1857-1864.
3. Lee SL, Tourtellotte LC, Wesselschmidt RL, Milbrandt J. Growth and differentiation proceeds normally in cells deficient in the immediate early gene NGFI-A. *J Biol Chem*. 1995;270(17):9971-9977.
4. Heidel FH, Bullinger L, Feng Z, et al. Genetic and Pharmacologic Inhibition of β-Catenin Targets Imatinib-Resistant Leukemia Stem Cells in CML. *Cell Stem Cell*. 2012;10(4):412-424.
5. Horn HS. Measurement of “Overlap” in Comparative Ecological Studies. *The American Naturalist*. 1966;100(914):419-424. doi:10.1086/282436
6. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21. doi:10.1016/j.cell.2019.05.031
7. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*. 2012;40(10):4288-4297. doi:10.1093/nar/gks042
8. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim Biophys Acta Gene Regul Mech*. 2020;1863(6):194431.
9. Holland CH, Tanevski J, Perales-Patón J, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol*. 2020;21(1):36.
10. Alvarez MJ, Shen Y, Giorgi FM, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*. 2016;48(8):838-847.
11. Stiehl T, Baran N, Ho AD, Marciniak-Czochra A. Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse. *J R Soc Interface*. 2014 Mar 12;11(94):20140079. doi: 10.1098/rsif.2014.0079.
12. Haurie C, Dale DC, Mackey MC. Cyclical neutropenia and other periodic hematological disorders: a review of mechanisms and mathematical models. Blood. 1998 Oct 15;92(8):2629-40. PMID: 9763544.

13. Rubinow SI, Lebowitz JL. A mathematical model of the acute myeloblastic leukemic state in man. Biophys J. 1976 Aug;16(8):897-910. doi: 10.1016/S0006-3495(76)85740-2. PMID: 1065387.

14. Manesso E, Teles J, Bryder D, Peterson C. Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation. J R Soc Interface. 2013 Mar 6;10(80):20120817. doi: 10.1098/rsif.2012.0817. PMID: 23256190.

15. Lorenzi T, Marciniak-Czochra A, Stiehl T. A structured population model of clonal selection in acute leukemias with multiple maturation stages. J Math Biol. 2019 Oct;79(5):1587-1621. doi: 10.1007/s00285-019-01404-w. PMID: 31350582.

16. Stiehl T, Baran N, Ho AD, Marciniak-Czochra A. Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: a model to predict patient survival. Cancer Res. 2015 Mar 15;75(6):940-9. doi: 10.1158/0008-5472.CAN-14-2508. PMID: 25614516.

17. Stiehl T, Lutz C, Marciniak-Czochra A. Emergence of heterogeneity in acute leukemias. Biol Direct. 2016 Oct 12;11(1):51. doi: 10.1186/s13062-016-0154-1. PMID: 27733173.

18. Stiehl T, Ho AD, Marciniak-Czochra A. Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis. Sci Rep. 2018 Feb 12;8(1):2809. doi: 10.1038/s41598-018-1115-4. PMID: 29434256.

19. Banck JC, Görlich D. In-silico comparison of two induction regimens (7 + 3 vs 7 + 3 plus additional bone marrow evaluation) in acute myeloid leukemia treatment. BMC Syst Biol. 2019 Jan 31;13(1):18. doi: 10.1186/s12918-019-0684-0. PMID: 30704476.

20. Stiehl T, Marciniak-Czochra A. Stem cell self-renewal in regeneration and cancer: Insights from mathematical modeling. Current Opinion in Systems Biology 2017; 5: 112-120.

21. Stiehl T, Ho AD, Marciniak-Czochra A. The impact of CD34+ cell dose on engraftment after SCTs: personalized estimates based on mathematical modeling. Bone Marrow Transplant. 2014 Jan;49(1):30-7. doi: 10.1038/bmt.2013.138. PMID: 24056742.

22. Lord BI, Molineux G, Pojda Z, Souza LM, Mermod JJ, Dexter TM. Myeloid cell kinetics in mice treated with recombinant interleukin-3, granulocyte colony-stimulating factor (CSF), or granulocyte-macrophage CSF in vivo. Blood. 1991 May 15;77(10):2154-9. PMID: 1709372.

23. Lord BI, Woolford LB, Molineux G. Kinetics of neutrophil production in normal and neutropenic animals during the response to filgrastim (r-metHu G-CSF) or filgrastim SD/01 (PEG-r-metHu G-CSF). Clin Cancer Res. 2001 Jul;7(7):2085-90. PMID: 11448927.

24. Morrison SJ, Wright DE, Weissman IL. Cyclophosphamide/granulocyte colony-stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. Proc Natl Acad Sci U S A. 1997 Mar 4;94(5):1908-13. doi: 10.1073/pnas.94.5.1908. PMID: 9050878.

25. Metcalf D. Hematopoietic cytokines. Blood. 2008 Jan 15;111(2):485-91. doi: 10.1182/blood-2007-03-079681. PMID: 18182579.
26. Layton JE, Hockman H, Sheridan WP, Mostyn G. Evidence for a novel in vivo control mechanism of granulopoiesis: mature cell-related control of a regulatory growth factor. Blood. 1989 Sep;74(4):1303-7. PMID: 2475185.

27. Marciniak-Czochra A, Mikelić A, Stiehl T. Renormalization group second order approximation for singularly perturbed nonlinear ordinary differential equations. Math. Methods in Applied Sciences. 2018; 41: 5691-5710.

28. Colijn C, Mackey MC. A mathematical model of hematopoiesis: II. Cyclic neutropenia. J Theor Biol. 2005 Nov 21;237(2):133-46. doi: 10.1016/j.jtbi.2005.03.034. PMID: 15975606.

29. De Souza C, Humphries AR. Dynamics of a Mathematical Hematopoietic Stem-Cell Population Model. SIAM Journal on Applied Dynamical Systems 2019 18:2, 808-852

30. Craig M, Humphries AR, Mackey MC. A Mathematical Model of Granulopoiesis Incorporating the Negative Feedback Dynamics and Kinetics of G-CSF/Neutrophil Binding and Internalization. Bull Math Biol. 2016 Dec;78(12):2304-2357. doi: 10.1007/s11538-016-0179-8. PMID: 27324993.

31. Walenda T, Stiehl T, Braun H, Fröbel J, Ho AD, Schroeder T, Goecke TW, Rath B, Germain U, Marciniak-Czochra A, Wagner W. Feedback signals in myelodysplastic syndromes: increased self-renewal of the malignant clone suppresses normal hematopoiesis. PLoS Comput Biol. 2014 Apr 24;10(4):e1003599. doi: 10.1371/journal.pcbi.1003599. PMID: 24763223.

32. Zhang J, Niu C, Ye L, Huang H, He X, Tong WG, Ross J, Haug J, Johnson T, Feng Q, Harris S, Wiedemann LM, Mishina Y, Li L. Identification of the haematopoietic stem cell niche and control of the niche size. Nature. 2003 Oct 23;425(6960):836-41. doi: 10.1038/nature02041. PMID: 14574412.

33. Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC, Martin RP, Schipani E, Divieti P, Brinthurst FR, Milner LA, Kronenberg HM, Scadden DT. Osteoblastic cells regulate the haematopoietic stem cell niche. Nature. 2003 Oct 23;425(6960):841-6. doi: 10.1038/nature02040. PMID: 14574413.

34. da Silva CL, Gonçalves R, dos Santos F, Andrade PZ, Almeida-Porada G, Cabral JM. Dynamic cell-cell interactions between cord blood haematopoietic progenitors and the cellular niche are essential for the expansion of CD34+, CD34+CD38- and early lymphoid CD7+ cells. J Tissue Eng Regen Med. 2010 Feb;4(2):149-58. doi: 10.1002/term.226. PMID: 19937912.

35. Nombela-Arrieta C, Manz MG. Quantification and three-dimensional microanatomical organization of the bone marrow. Blood Adv. 2017 Feb 14;1(6):407-416. doi: 10.1182/bloodadvances.2016003194. PMID: 29296956.

36. Endicott KM, Gump H. Hemograms and myelograms of healthy female mice of C-57 brown and CFW strains. Blood. 1947 Jul;(1):60-3. PMID: 20266787.

37. Pegg DE. A Quantitative Study of Bone Marrow Grafting: Implications for Human Bone Marrow Infusion. Br J Cancer. 1962 Sep;16(3):400-16. doi: 10.1038/bjc.1962.46. PMID: 21772461.

38. Yang M, Büsche G, Ganser A, Li Z. Morphology and quantitative composition of hematopoietic cells in murine bone marrow and spleen of healthy subjects. Ann Hematol. 2013 May;92(5):587-94. doi: 10.1007/s00277-012-1653-5. PMID: 23307597.

39. Passegù E, Wagers AJ, Giuriato S, Anderson WC, Weissman IL. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J Exp Med. 2005 Dec 5;202(11):1599-611. doi: 10.1084/jem.20050967. PMID: 16330818.
40. Pillay J, den Braber I, Vrisekoop N, Kwast LM, de Boer RJ, Borghans JA, Tesselaar K, Koenderman L. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. Blood. 2010 Jul 29;116(4):625-7. doi: 10.1182/blood-2010-01-259028. PMID: 20410504.

41. Cheng H, Zheng Z, Cheng T. New paradigms on hematopoietic stem cell differentiation. Protein Cell. 2020 Jan;11(1):34-44. doi: 10.1007/s13238-019-0633-0. PMID: 31201709.

42. Lord BI, Molineux G, Pojda Z, Souza LM, Mermod JJ, Dexter TM. Myeloid cell kinetics in mice treated with recombinant interleukin-3, granulocyte colony-stimulating factor (CSF), or granulocyte-macrophage CSF in vivo. Blood. 1991 May 15;77(10):2154-9. PMID: 1709372.

43. Lord BI, Woolford LB, Molineux G. Kinetics of neutrophil production in normal and neutropenic animals during the response to filgrastim (r-metHu G-CSF) or filgrastim SD/01 (PEG-r-metHu G-CSF). Clin Cancer Res. 2001 Jul;7(7):2085-90. PMID: 11448927.

44. Kühn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. Science. 1995 Sep 8;269(5229):1427-9. doi: 10.1126/science.7660125. PMID: 7660125.
Figure S2. Sorting of barcoded cells, gene marking during transplantation and clonal complexity. (A) Sorting strategy of ckit+ cells transduced with barcode vector by live/dead marker and GFP expression. (B) Barcode marking within the stem cell compartment of the bone marrow at the end of 1st and 2nd transplant. (C) Number of individual barcodes (observed richness) recovered from peripheral blood per mouse over time. Only the larger clones summing up to 99% cumulative abundance are considered. (D) Population measures of clonal evolution over time per genotype for all mice. Richness is the number of unique observed barcodes, Evenness is the population diversity normalized by
richness and Diversity shows the Inverse Simpson measure, describing heterogeneity in composition of clone sizes. Only the larger clones summing up to 99% cumulative abundance are considered. (E) Representative example of clonal proportions and tracking of 20 individual clones (20 largest clones in 1st transplant LSK) within Granulocytes over time in Mouse 1 and 4.

Figure S3. Hierarchical clustering and dynamical systems model. The panel depicts the confidence intervals of fitted proliferation rates and self-renewal probabilities for Apc\textsuperscript{+/-}, Egr1\textsuperscript{+/-} and Csnk\textsuperscript{+/-} per mouse in comparison to the wild type cell properties. Hierarchical agglomerative average-linkage clustering based on Morisita-Horn similarity index groups samples from the same mouse together.
Figure S4. Gene marking of genetic barcoding, genotype abundances over time and barcode clustering. (A) Barcode marking within the stem cell compartment of the bone marrow and within peripheral blood at the end of 1st

and 2nd transplant. (B) Number of individual barcodes (observed richness) recovered from peripheral blood per mouse over time. Only the larger clones summing up to 99% cumulative abundance are considered. (C) Population measures of clonal evolution over time per genotype for all mice. Richness is the number of unique observed barcodes, Evenness is the population diversity normalized by richness and Diversity shows the Inverse Simpson measure, describing heterogeneity in composition of clone sizes. Only the larger clones summing up to 99% cumulative abundance are considered. (D) Hierarchical agglomerative average-linkage clustering based on Morisita-Horn index of all samples. (E) Representative example of 85% dominant clonal proportions within Granulocytes and LSK over time in Mouse 53 and 55. Lines represent the tracking of 20 largest clones in 1st transplant LSK.

Figure S5. Engraftment of haploinsufficient mouse models and intracellular β-catenin flow. (A) Engraftment of genotypes as single cell CD45.2 chimerism in competitive transplant against CD45.1+ wt bone marrow at 4 weeks after transplantation before Poly(I:C) treatment and after Poly(I:C) at 8 weeks after transplantation. (B) Mean Fluorescence intensity of β-catenin-FITC antibody in LSK of WT, Csnk1a1 Δ/Δ, Csnk1a1 Δ/Δ Apc Δ/Δ, Csnk1a1 Δ/Δ Ctnnb1 Δ/Δ. Statistical significance was tested by one-way Anova with Dunnett’s multiple comparison test. Data represents mean
± SEM. (C) Percentage of LK and LSK within CD45.2+ and CD45.1+ cells in the bone marrow LSK (lin−;Sca1++;kit+) after 2nd transplant. Statistical significance was tested by unpaired t-test. Data represents mean ± SEM.
Figure S6. Sorting strategy, quality metrics of scRNAseq and cluster idenfication&proportions. (A) Sorting strategy to isolate lin-c-kit+ sca1+ cells at the time of sacrifice (24 weeks after transplantation of Csnk1a1+/+ or Mx1Cre+ bone marrow into WT recipients) (B) Quality metrics of all samples: Percentage of mitochondrial genes, Log10 of gene counts per cell, Log10 of unique features per cell (C) Bar plot of cells per cluster (Csnk1a1+/+ versus WT). Normalization to overall number of input per condition is shown. Fisher's exact test with Benjamini-Hochberg correction was used. (D) Relative cell proportions per experimental sample color-coded by identity cluster as identified by unsupervised clustering (E) Ridge Plots of Csnk1a1 expression in cell identity clusters split by genotype (F) Heatmap of average expression of top cluster markers as identified per cluster of cells. (G) Heatmap of pathway response signature scores for 14 pathways inferred from differentially expressed genes between Csnk1a1+/+ versus WT per cell identity cluster using the method PROGENy (Pathway responsive genes for activity inference).

Figure S7. Blood counts and cell cycle analysis of competitive transplant under inflammatory conditions. (A) Blood counts taken during competitive transplant (B) Intracellular ki67 and DAPI staining for non-cycling(G0) and cycling cells (G1-S-G2M) within CD45.2+ LSK, MPP and LT-HSC. Data represents mean ± SEM. Statistical significance was tested by one-way Anova with Dunnett’s multiple comparison test and by one-tailed t-test comparing PBS and PIC condition per genotype. Differences are non-significant.