HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2016 October 28.

Published in final edited form as:

Nature. 2016 April 28; 532(7600): 508–511. doi:10.1038/nature17665.

Musashi-2 Attenuates AHR Signaling to Expand Human Hematopoietic Stem Cells

Stefan Rentas[1], **Nicholas Holzapfel**[#1], **Muluken S Belew**[#1], **Gabriel Pratt**[#2,6], **Veronique Voisin**[4], **Brian T Wilhelm**[5], **Gary D Bader**[4], **Gene W Yeo**[2,3,6], and **Kristin J Hope**[1]

[1] Department of Biochemistry and Biomedical Sciences, Stem Cell and Cancer Research Institute, McMaster University, Hamilton, ON, Canada

[2] Department of Cellular and Molecular Medicine, Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA

[3] Department of Physiology, National University of Singapore and Molecular Engineering Laboratory, A*STAR, Singapore

[4] The Donnelly Centre, University of Toronto, Toronto, ON, Canada.

[5] Institute for Research in Immunology and Cancer, University of Montreal, Montreal, QC, Canada

[6] Bioinformatics Graduate Program, University of California, San Diego, La Jolla, CA, USA

[#] These authors contributed equally to this work.

Abstract

Umbilical cord blood (CB)-derived hematopoietic stem cells (HSCs) are essential in many life saving regenerative therapies, but their low number in CB units has significantly restricted their clinical use despite the advantages they provide during transplantation[1]. Select small molecules that enhance hematopoietic stem and progenitor cell (HSPC) expansion in culture have been identified[2,3], however, in many cases their mechanisms of action or the nature of the pathways they impinge on are poorly understood. A greater understanding of the molecular pathways that underpin the unique human HSC self-renewal program will facilitate the development of targeted strategies that expand these critical cell types for regenerative therapies. Whereas transcription factor networks have been shown to influence the self-renewal and lineage decisions of human HSCs[4,5], the post-transcriptional mechanisms guiding HSC fate have not been closely

Correspondence and requests for materials should be addressed to G.W.Y. (; Email: geneyeo@ucsd.edu), K.H. (; Email: kristin@mcmaster.ca).

Author Contributions

S.R. designed and performed experiments, analyzed data and wrote the manuscript. N.H. constructed CLIP-seq libraries. M.S.B. helped perform CB experiments. G.P. and G.W.Y advised CLIP-seq library construction, performed CLIP-seq bioinformatic analyses and wrote the manuscript. B.W. performed RNA-seq analyses. V.V and G.D.B performed RNA-seq bioinformatic analyses. K.H. conceived the project, supervised the study, analyzed data, interpreted results and wrote the manuscript.

The authors declare no competing financial interests.

Supplementary Information

Supplementary Tables 1-7

Supplementary Figure 1

investigated. Here we show that overexpression of the RNA-binding protein (RBP) Musashi-2 (MSI2) induces multiple pro-self-renewal phenotypes, including a 17-fold increase in short-term repopulating cells and a net 23-fold ex vivo expansion of long-term repopulating HSCs. By performing a global analysis of MSI2-RNA interactions, we determined that MSI2 directly attenuates aryl hydrocarbon receptor (AHR) signaling through post-transcriptional downregulation of canonical AHR pathway components in CB HSPCs. Our study provides new mechanistic insight into RBP-controlled RNA networks that underlie the self-renewal process and give evidence that manipulating such networks ex vivo can provide a novel means to enhance the regenerative potential of human HSCs.

RBP-mediated control of translation in human HSCs and its potential to regulate HSC self-renewal remains underexplored. Here we investigated the role of MSI2 in post-transcriptionally controlling human HSPC self-renewal as it is known to regulate mouse HSCs[6-8], and is predicted to impact mRNA translation[9]. *MSI2* was present and elevated in primitive CB HSPCs and decreased during differentiation, whereas its paralog, *MSI1*, was not expressed (Extended Data Fig. 1a-f). Lentiviral overexpression (OE) of *MSI2* resulted in a 1.5-fold increase in colony forming units (CFU) relative to control, principally due to a 3.7-fold increase in the most primitive CFU-Granulocyte Erythrocyte Monocyte Megakaryocyte (GEMM) colony type (Extended Data Fig. 2a, Fig. 1a). Remarkably, 100% of MSI2 OE CFU-GEMMs generated secondary colonies compared to only 40% of controls. In addition, MSI2 OE yielded 3-fold more colonies per re-seeded CFU-GEMM (Fig. 1b, c, Extended Data Fig. 2b). During in vitro culture MSI2 OE resulted in 2.3- and 6-fold more cells relative to control at the 7 and 21-day time points, respectively (Extended Data Fig. 2c, d). Moreover after 7 days in culture MSI2 OE yielded a cumulative 9.3-fold increase in colony forming cells in the absence of changes in cell cycling or death (Extended Data Fig. 2e-h). Altogether, our data demonstrate that enforced expression of MSI2 has potent self-renewal effects on early progenitors and promotes their in vitro expansion.

Short-term repopulating cells (STRC) produce a transient multi-lineage graft in NOD-*scid*-*IL2Rγc*$^{-/-}$ (NSG) mice[10], and in patients reconstitute granulocytes and platelets critical for preventing post-transplant infection and bleeding[1]. STRCs overexpressing MSI2 exhibited 1.8-fold more primitive CD34^{+} cells post-infection and a dramatic 17-fold increase in functional STRCs relative to control as determined by limiting dilution analysis (LDA) of human chimerism at 3 weeks post-transplant (Fig. 1d-f, Extended Data Fig. 3a, b). Furthermore, at a protracted engraftment readout time of 6.5 weeks at non-limiting transplant doses, 100% of MSI2 OE STRC transplanted mice were engrafted compared to only 50% of controls, indicating MSI2 OE extended the duration of STRC-mediated engraftment (Extended Data Fig. 3c).

We next explored the effect of shRNA-induced MSI2 knockdown (KD) on HSPC function. MSI2 KD did not alter clonogenic potential but did decrease CFU replating 3-fold (Extended Data Fig. 4a-c). When effects on more primitive culture-initiating cells were explored we found MSI2 KD significantly decreased cell number over culture (Extended Data Fig. 4d, e) independent of increased death or altered cell cycling (data not shown). Upon transplantation, engrafted MSI2 KD GFP^{+} cells showed no evidence of lineage

skewing, yet were strikingly reduced relative to the percentage of GFP^+ cells initially transplanted (Extended Data Fig. 4f-h). Combined, our in vitro and in vivo data show that MSI2 KD reduces self-renewal in early progenitors and HSCs.

To characterize the earliest transcriptional changes induced by modulating MSI2 expression, we performed RNA-seq on $CD34^+$ MSI2 OE and KD cells immediately post-transduction (Supplementary Tables 1, 2). MSI2 OE-induced transcriptional changes anti-correlated with those of MSI2 KD, suggesting OE and KD have opposite effects (Extended Data Fig. 5a). When compared to transcriptome data from 38 human hematopoietic cell subpopulations[4], we observed that genes significantly upregulated upon MSI2 OE and downregulated upon MSI2 KD were exclusively enriched for those highly expressed in HSC and other primitive $CD34^+$ populations (Extended Data Fig. 5b).

Since MSI2 OE conferred an HSC gene expression program, we hypothesized that it could facilitate HSC expansion ex vivo. Remarkably, MSI2 OE induced a 4-fold increase in $CD34^+CD133^+$ phenotypic HSCs relative to control after 7 days of culture (Fig. 2a). We next performed an LDA to define functional HSC frequency before (day 3 post-transduction, D3) and after 7 days of ex vivo culture (day 10, D10; Extended Data Fig. 6a). D3 recipients displayed no altered engraftment as a result of MSI2 OE, however, recipients of MSI2 OE D10-expanded cells displayed multiple phenotypes of enhanced reconstitution relative to control, including a 2-fold increase in BM GFP^+ levels without changes to lineage output, proportionally more GFP^+ cells within the human graft relative to pre-transplant D10 levels, greater GFP mean fluorescence intensity and enrichment of CD34 expression in GFP^{high} cells (Fig. 2b, c, Extended Data Fig. 6b-h). As the lentiviral construct design ensures GFP levels mirror MSI2's, this indicates high levels of MSI2 impart enhanced competitiveness and are conducive to in vivo HSPC activity. Importantly, D10 MSI2 OE cultures contained more $CD34^+CD133^+$ cells prior to transplant (Extended Data Fig. 6i), and in accordance, the HSC frequency in D10 MSI2 OE cultures was increased 2-fold relative to the D3 culture time point, whereas control cultures displayed a 3-fold decrease. These results demonstrate that MSI2 OE facilitated an ex vivo net 6-fold increase in HSCs relative to control (Fig. 2g, h, Supplementary Tables 3, 4).

Secondary LDA transplants were performed to fully explore the effects of MSI2 OE and culturing has on self-renewal and long-term HSCs (LT-HSC). Robust engraftment with MSI2 OE did not display altered myelo-lymphopoiesis or leukemic development (Fig. 2e). Secondary LDA measurements revealed BM GFP^+ percentage increased 4.6-fold and LT-HSC frequency 3.5-fold with MSI2 OE compared to controls (Fig. 2d, e, f, Supplementary Table 5). This corresponds to MSI2 OE GFP^+ HSCs having expanded in primary mice 2.4-fold over input as compared to a decrease of 1.5-fold for control HSCs (Fig. 2g). The level of MSI2 OE-induced in vivo fold expansion reflects the behaviour of uncultured HSCs, which undergo similarly controlled expansion during passage in mice[3,11,12]. Finally, when accounting for the total change in GFP^+ HSCs upon ex vivo culture, MSI2 OE provided a cumulative 23-fold-expansion of secondary LT-HSC relative to control (Fig. 2g, h) indicating elevated MSI2 provides a significant self-renewal advantage to functional HSCs during ex vivo culture.

To gain mechanistic insights we examined MSI2 OE-induced differentially expressed genes and found Cytochrome P450 1B1 Oxidase (CYP1B1), an effector of AHR signaling[13], amongst the most repressed (Supplementary Table 1). Pathway analysis revealed many predicted AHR targets were enriched in MSI2 OE downregulated (Fig. 3a) and MSI2 KD upregulated gene sets (Extended Data Fig. 7a, b). Binding of the nuclear receptor transcription factor AHR with StemRegenin 1 (SR1) inhibits AHR target gene activation leading to human HSPC expansion in culture[2]. Gene set enrichment analysis (GSEA) revealed MSI2 OE downregulated genes significantly matched the SR1 downregulated gene set in an SR1 dose-dependent manner (Fig. 3b, c), whereas MSI2 KD induced the opposite expression profile (Extended Data Fig. 7c, d). We next examined the overlap of downregulated genes with ChIP-seq-identified AHR targets[14]. This comparison was extended to downregulated genes upon treatment with UM171, another compound that expands HSPCs, but independent of AHR[3]. Direct transcriptional targets of AHR were enriched in the MSI2 OE and SR1 downregulated gene sets by 3.8 and 5.6-fold, respectively, compared to UM171, an overrepresentation maintained for predicted AHR targets and one that suggests MSI2 OE expands HSPCs through AHR signaling attenuation (Fig. 3d, Extended Data Fig. 7e). Furthermore, SR1 treatment increased the percentage of $CD34^+$ cells 8-fold for control cultures compared to only 4-fold with MSI2 OE (Extended Data Fig. 8a, b), a finding that indicates a redundancy in SR1 and MSI2 OE activity on HSPCs, which suggests they act on the same pathway.

To further elucidate the AHR connection, MSI2 OE and control cultures were treated with the AHR agonist 6-Formylindolo(3,2-b)carbazole (FICZ), whose induction of canonical AHR targets in MSI2 OE cells, demonstrates that they remain competent for AHR activation (Extended Data Fig. 8c). FICZ induced a dramatic reversal of the MSI2 OE-mediated increases in primary CFU-GEMMs and their replating capacities (Fig. 3e, f). Furthermore, FICZ-treated MSI2 OE cultures displayed greater losses in phenotypic HSPCs compared to controls which showed no change (Extended Data Fig. 8d, e). Altogether, these results demonstrate agonist-induced restoration of AHR activity reduces MSI2 OE's pro-self-renewal effects and strongly supports downregulation of AHR signaling as the mechanism through which MSI2 OE achieves HSPC expansion.

To identify key RNA targets underlying MSI2 function, we analyzed global MSI2 protein-RNA interactions using CLIP-seq (Extended Data Fig. 9a, b)[15]. Mapped reads identified highly correlated gene RPKMs and enrichment of significantly overlapping clusters within >6000 protein-coding genes from replicate experiments (Extended Data Fig. 9c, Fig. 4a, b). Within the top 40% of reproducible clusters, MSI2 bound predominantly to mature mRNA within their 3’ untranslated regions (3'UTRs) (Fig. 4c). Importantly, 9% of annotated protein-coding genes were reproducible MSI2 targets, compared to 0.2% of long non-coding RNAs (Extended Data Fig. 9d), suggesting MSI2 controls stability or translation of coding mRNAs. Motif analyses identified a consensus pentamer (U/G)UAGU resembling the known mouse Msi1 binding sequence[9,16] within binding sites in all genic regions and significantly more conserved than background (Fig. 4d, Extended Data Fig. 9e-h). MSI2 binding sites within Msi1 targets[16] indicate Musashi proteins may bind the same genes through 3'UTR-embedded motifs (Extended Data Fig. 9i). Finally, target gene ontology analysis revealed 186 biological processes categories (Supplementary Table 6), among the

most significant of which were electron transport, estrogen receptor signaling regulation and metabolism of small molecules, all processes known to be transcriptionally impacted by AHR signaling[17].

Strikingly, among the top 2% of enriched CLIP-seq targets (Supplementary Table 7) were the 3'UTRs of heat shock protein 90 (HSP90) and CYP1B1, two AHR pathway components. Each exhibited multiple MSI2 binding motifs correlating with overlapping clusters of CLIP-seq reads (Fig. 4e, Extended Data Fig. 10a). Querying MSI2's ability to post-transcriptionally regulate these genes during HSPC expansion, we looked for instances of uncoupled transcript and protein expression. HSP90 displayed uncoupling of transcript (1.6-fold up) and protein (1.6-fold down) early in culture, but at 7 days showed further upregulated transcript (2.5-fold) and variable protein levels (Fig. 4f, Extended Data Fig. 10b). Since AHR-HSP90 binding is critical for ligand-dependent transcriptional activity[13], downregulation of HSP90 protein at the outset of HSPC culture would be expected to reduce latent AHR complex formation and attenuate AHR signaling (Fig. 3a). Indeed, CYP1B1 transcript and protein displayed 2-fold reductions early in culture, consistent with decreased AHR pathway activity, however at day 7, CYP1B1 transcripts were upregulated 1.7-fold, uncoupled from 2-fold downregulated protein (Fig. 4g, Extended Data Fig. 10c). To test if MSI2 directly mediates post-transcriptional repression of these targets, *CYP1B1* and *HSP90* 3'UTR regions were coupled to luciferase. MSI2 OE induced significant reductions in luciferase signal from both reporters, an effect mitigated upon mutating the core CLIP-seq-identified UAG motifs (Extended Data Fig. 10d, e). Since MSI2 OE-mediated post-transcriptional downregulation of the AHR pathway converged on CYP1B1 protein repression throughout culture, we explored the effects of inhibiting CYP1B1 independently with (E)-2,3',4,5'-Tetramethoxystilbene (TMS) on HSPCs. During culture, TMS increased the frequency and total numbers of $CD34^+$ cells by 1.5- and 2-fold respectively (Fig. 4h, i), phenocopying the effects of MSI2 OE. Lastly, co-overexpression of 3'UTRless CYP1B1 with MSI2 decreased secondary CFU-GEMM replating efficiency (Extended Data Fig. 10f and g) suggesting that CYP1B1, while typically used to report on AHR signaling, promotes HSPC differentiation.

Our work identifies MSI2 as an important new mediator of human HSPC self-renewal and ex vivo expansion by coordinating the post-transcriptional regulation of proteins belonging to a shared self-renewal regulatory pathway (Extended Data Fig. 10h). We envision that manipulation of the post-transcriptional circuitry controlled by RBPs will provide a novel and powerful means by which to enhance the regenerative potential of not only human HSCs but other stem cell types as well.

Methods

Mice

NOD-*scid*-*IL2Rγc*$^{-/-}$ (Jackson Laboratory) mice were bred and maintained in the Stem Cell Unit animal barrier facility at McMaster University. All procedures received the approval of the Animal Research Ethics Board at McMaster University.

Isolation of primitive human hematopoietic cells and flow cytometry

All patient samples were obtained with informed consent and with the approval of local human subject research ethics boards at McMaster University. Human umbilical cord blood mononuclear cells were collected by centrifugation with Ficoll-Paque Plus (GE), followed red blood cell lysis with ammonium chloride (Stemcell Technologies). Cells were then incubated with a cocktail of lineage specific antibodies (CD2, CD3, CD11b, CD11c, CD14, CD16, CD19, CD24, CD56, CD61, CD66b, and GlyA; Stemcell Technologies) for negative selection of Lin^- cells using an EasySep immunomagnetic column (Stemcell Technologies). Live cells were discriminated on the basis of cell size, granularity and, as needed, absence of viability dye 7-AAD (BD Biosciences) uptake. All flow cytometry analysis was performed using a BD LSR II instrument (BD Biosciences). Data acquisition was conducted using BD FACSDiva software (BD Biosciences) and analysis was performed using FlowJo software (Tree Star).

HSPC sorting and qRT-PCR analysis

To quantify *MSI2* expression in human HSPCs, Lin^- CB cells were stained with the appropriate antibody combinations to resolve HSC ($CD34^+$ $CD38^-$ $CD45RA^-$ $CD90^+$), MPP ($CD34^+$ $CD38^-$ $CD45RA^-$ $CD90^-$), CMP ($CD34^+$ $CD38^+$ $CD71^-$) and EP ($CD34^+$ $CD38^+$ $CD71^+$) fractions as similarly described previously[18,19] with all antibodies from BD Biosciences: CD45RA (HI100), CD90 (5E10), CD34 (581), CD38 (HB7) and CD71 (M-A712). Cell viability was assessed using the viability dye 7AAD (BD Biosciences). All cell subsets were isolated using a BD FACSAria II cell sorter (BD Biosciences) or a MoFlo XDP cell sorter (Beckman Coulter). HemaExplorer[20] analysis was used to confirm *MSI2* expression in human HSPCs and across the hierarchy. For all qRT-PCR determinations total cellular RNA was isolated with TRIzol LS reagent according to the manufacturer's instructions (Invitrogen) and cDNA was synthesized using the qScript cDNA Synthesis Kit (Quanta Biosciences). qRT-PCR was done in triplicate with PerfeCTa qPCR SuperMix Low ROX (Quanta Biosciences) with gene specific probes (Universal Probe Library, UPL, Roche) and primers: MSI2 UPL-26, F- ggcagcaagaggatcagg, R-ccgtagagatcggcgaca; HSP90 UPL-46, F-gggcaacacctctacaagga, R-cttgggtctgggtttcctc; CYP1B1 UPL-20, F-acgtaccggccactatcact, R-ctcgagtctgcacatcagga; GAPDH UPL-60, F-agccacatcgctcagacac, R-gcccaatacgaccaaatcc; ACTB (UPL Set Reference Gene Assays, Roche). The mRNA content of samples compared by qRT-PCR was normalized based on the amplification of GAPDH or ACTB.

Lentivirus production and western blot validation

MSI2 shRNAs were designed with the Dharmacon algorithm (http://www.dharmacon.com). Predicted sequences were synthesized as complimentary oligonucleotides, annealed and cloned downstream of the H1 promoter of the modfied cppt-PGK-EGFP-IRES-PAC-WPRE lentiviral expression vector[18]. Sequences for the MSI2 targeting and control RFP targeting shRNAs were as follows: shMSI2, 5'-gagagatcccactacgaaa-3'; shRFP, 5'-gtgggagcgcgtgatgaac-3'. Human MSI2 cDNA (BC001526; Open Biosystems) was subcloned into the MA bi-directional lentiviral expression vector[21]. Human CYP1B1 cDNA (BC012049; Open Biosystems) was cloned in to psMALB[22]. All lentivirus was prepared by

transient transfection of 293FT (Invitrogen) cells with pMD2.G and psPAX2 packaging plasmids (Addgene) to create VSV-G pseudotyped lentiviral particles. All viral preparations were titrated on HeLa cells before use on cord blood. Standard SDS-PAGE and western blotting procedures were performed to validate the effect of knockdown on transduced NB4 cells (DSMZ) and over expression on 293FT cells. Immunoblotting was performed with anti-MSI2 rabbit monoclonal IgG (EP1305Y, Epitomics) and β-actin mouse monoclonal IgG (ACTBD11B7, Santa Cruz Biotechnology) antibodies. Secondary antibodies used were IRDye 680 goat anti-rabbit IgG and IRDye 800 goat anti-mouse IgG (LI-COR). 293FT and NB4 cell lines tested negative for mycoplasma. NB4 cells were authenticated by ATRA treatment prior to use.

CB transduction

CB transductions were conducted as described previously[18,23]. Briefly, thawed Lin^- CB, flow-sorted Lin^- $CD34^+$ $CD38^-$ or Lin^- $CD34^+$ $CD38^+$ cells were prestimulated for 8-12 hours in StemSpan medium (StemCell Technologies) supplemented with growth factors Interleukin 6 (IL-6; 20 ng/ml, Peprotech), Stem cell factor (SCF; 100 ng/ml, R&D Systems), Flt3 ligand (FLT3-L; 100 ng/ml, R&D Systems) and Thrombopoietin (TPO; 20 ng/ml, Peprotech). Lentivirus was then added in the same medium at a multiplicity of infection of 30–100 for 24 hours. Cells were then given 2 days post transduction before use in in vitro or in vivo assays. For in vitro CB studies biological (experimental) replicates were performed with three independent CB samples.

Clonogenic progenitor assays

Human clonogenic progenitor cell assays were done in semi-solid methylcellulose medium (Methocult H4434; Stem Cell Technologies) with flow-sorted GFP^+ cells post transduction (500 cells/ml) or from day seven cultured transduced cells (12000 cells/ml). Colony counts were carried out after 14 days of incubation. CFU-GEMMs can seed secondary colonies due to their limited self-renewal potential[24]. MSI2 OE and control CFU-GEMM replating for secondary CFU analysis was performed by picking single CFU-GEMMs at day 14 and disassociating colonies by vortexing. Cells were spun and resuspended in fresh methocult, mixed with a blunt end needle and syringe, and then plated into single wells of a 24-well plate. Secondary CFU analysis for shMSI2 and shControl expressing cells was performed by harvesting total colony growth from a single dish (since equivalent numbers of CFU-GEMMs), resuspending cells in fresh methocult by mixing vigorously with a blunt end needle and syringe and then plating into replicate 35 mm tissue culture dishes. In both protocols, secondary colony counts were done following incubation for 10 days. For primary and secondary colony forming assays performed with the AHR agonist 6-Formylindolo(3,2-b)carbazole (FICZ; Santa Cruz Biotechnology), 200 nM FICZ or 0.1% DMSO was added directly to H4434 methocult medium. 2-way ANOVA analysis was performed examining secondary CFU output and FICZ treatment for MSI2 OE or control conditions. Colonies were imaged with a Q-Colour3 digital camera (Olympus) mounted to an Olympus IX5 microscope with a 10X objective lens. Image-Pro Plus imaging software (Media Cybernetics) was used to acquire pictures and subsequent image processing was performed with ImageJ software (NIH).

Lin$^-$ CB and Lin$^-$ CD34$^+$ suspension cultures

Transduced human Lin$^-$ CB cells were sorted for GFP expression and seeded at a density of 1×10^5 cells/ml in IMDM 10% FBS supplemented with human growth factors IL-6 (10 ng/ml), SCF (50 ng/ml), FLT3-L (50 ng/ml), and TPO (20 ng/ml) as previously described[25]. To generate growth curves, every seven days cells were counted, washed, and resuspended in fresh medium with growth factors at a density of 1×10^5 cells/ml. Cells from suspension cultures were also used in clonogenic progenitor assays, cell cycle and apoptosis assays. Experiments performed on transduced Lin$^-$ CD34$^+$ CB cells used serum free conditions as described in the CB transduction subsection of the Methods. For in vitro CB studies biological (experimental) replicates were performed with three independent CB samples.

Cell cycle and apoptosis assays

Monitoring cell cycle progression was performed with the addition of BrdU to day 10 suspension cultures at a final concentration of 10 μM. After three hours of incubation, cells were assayed with the BrdU Flow Kit (BD Biosciences) according to the manufacture's protocol. Cell proliferation and quiescence was measured by Ki67 (BD Bioscience) and Hoechst 33342 (Sigma) on day 4 suspension cultures after fixing and permeabilizing cells with the Cytofix/Cytoperm kit (BD Biosciences). For apoptosis analysis, Annexin V (Invitrogen) and 7-AAD (BD Bioscience) staining on day 7 suspension cultures was performed according to the manufacture's protocol.

Intracellular flow cytometry

Lin$^-$ CB cells were initially stained with anti-CD34 PE (581) and anit-CD38 APC (HB7) antibodies (BD Biosciences) then fixed with the Cytofix/Cytoperm kit (BD Biosciences) according to the manufacturer's instructions. Fixed and permeabilized cells were immunostained with anti-MSI2 rabbit monoclonal IgG antibody (EP1305Y, Abcam) and detected by Alexa-488 goat anti-rabbit IgG antibody (Invitrogen).

RNA-Sequencing data processing

CD34$^+$ cells were transduced with MSI2 OE or KD lentivirus along with their corresponding controls and then sorted for GFP expression 3 days later. Transductions for MSI2 OE and KD were each performed on two independent CB samples. Total RNA from transduced cells ($>1\times10^5$) was isolated using TRIzol LS as recommended by the manufacturer (Invitrogen), and then further purified using RNeasy columns (Qiagen). Sample quality was assessed using Bioanalyzer RNA Nano chips (Agilent). Paired-end, barcoded RNA-seq sequencing libraries were then generated using the TruSeq RNA Sample Prep Kit (v2) (Illumina) following the manufacturer's protocols starting from 1ug of total RNA. Quality of library generation was then assessed using a Bioanalyzer platform (Agilent) and Illumina MiSeq-QC run was performed or quantified by qPCR using KAPA quantification kit (KAPA Biosystems). Sequencing was performed using an Illumina HiSeq2000 using TruSeq SBS v3 chemistry at the Institute for Research in Immunology and Cancer's Genomics Platform (University of Montreal) with cluster density targeted at 750k clusters/mm2 and paired-end 2×100bp read lengths. For each sample, 90-95M reads were produced and mapped to the hg19 (GRCh37) human genome assembly using CASAVA (version 1.8). Read counts

generated by CASAVA were processed in EdgeR (edgeR_3.12.0, R 3.2.2) using TMM normalization, paired design, and estimation of differential expression using a generalized linear model (glmFit). The false discovery rate (FDR) was calculated from the output p-values using the Benjamini-Hochberg method. The fold change of logarithm of base 2 of TMM normalized data (logFC) was used to rank the data from top up-regulated to top down-regulated genes and FDR (0.05) was used to define significantly differentially expressed genes. RNA-seq data has been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE70685.

GSEA and iRegulon AHR target prediction

iRegulon[26] was used to retrieve the top 100 AHR predicted targets with a minimal occurrence count threshold of 5. The data were analyzed using GSEA[27] with ranked data as input with parameters set to 2000 gene-set permutations.

GSEA and StemRegenin 1 (SR1) gene-sets

The Gene Expression Omnibus (GEO) dataset GSE28359 that contains Affymetrix Human Genome U133 Plus 2.0 Array gene expression data of $CD34^+$ cells treated with SR1 at 30 nM, 100nM, 300nM and 1000nM was used to obtain lists of genes differentially expressed in the treated samples compared to the control ones (0 nM)[2]. Data were background corrected using Robust Multi-Array Average (RMA) and quantile normalized using the *expresso()* function of the affy Bioconductor package (affy_1.38.1 , R 3.0.1). Lists of genes were created from the 150 genes top up regulated and top down regulated by the SR1 treated samples at each dose compared to the non-treated samples (0 nM). The data were analyzed using GSEA with ranked data as input with parameters set to 2000 gene-set permutations. Normalized enrichment score (NES) and false discovery rate (FDR) were calculated for each comparison.

DMAP population comparisons

The GEO dataset GSE24759 that contains Affymetrix GeneChip HT-HG_U133A Early Access Array gene expression data of 38 distinct hematopoietic cell states[4] was compared to the OE and KD data. GSE24759 data were background corrected using Robust Multi-Array Average (RMA), quantile normalized using the *expresso()* function of the affy Bioconductor package (affy_1.38.1 , R 3.0.1), batch corrected using the *ComBat()* function of the sva package (sva_3.6.0) and scaled using the standard score. Bar graphs were created by calculating for significantly differentially expressed genes the number of scaled data that were above (>0) or below (<0) the mean for each population. Percentages indicating time the observed value (set of up or downregulated genes) was better represented in that population than random values were calculated from 1000 trials.

AHR ChIP-seq comparison to downregulated gene sets

A unique list of genes closest to AHR bound regions previously identified from TCDD-treated MCF7 ChIP-seq data[14] was used to calculate the overlap with genes showing >1.5-fold down regulation with UM171 (35 nM) and SR1 (500 nM) relative to DMSO treated samples[3] as well as with genes significantly down regulated in MSI2 OE versus control

treated samples (FDR<0.05). Percentage of down regulated genes with AHR bound regions was then graphed for each gene set. P-values were generated with Fisher's exact test between gene lists.

oPOSSUM analysis for promoter AHR binding sites in downregulated gene sets

AHR transcription factor binding sites in downregulated gene sets were identified with oPOSSUM-3[28]. Genes showing >1.5-fold down regulation with UM171 (35 nM) and SR1 (500 nM) relative to DMSO treated samples[3] were used along with significantly downregulated genes (FDR<0.05) with EdgeR analyzed MSI2 OE versus control treated samples. The three gene lists uploaded into oPOSSUM-3 and the AHR:ARNT transcription factor binding site profile was used with the matrix score threshold set at 80% and analyzing the region 1500 bp upstream and 1000 bp downstream of transcription start site. Percentage of downregulated genes with AHR binding sites in the promoter was then graphed for each gene set. Fisher's exact test was used to identify significant overrepresentation of AHR binding sites in gene lists relative to background.

Analysis for human chimerism

NSG mice were sublethally irradiated (315 cGy) one day prior to intrafemoral injection with transduced cells carried in IMDM 1% FBS at 25μl per mouse. Injected mice were analyzed for human hematopoietic engraftment at 12-14 weeks post transplantation or 3 and 6.5 weeks for STRC experiments. Mouse bones (femurs, tibiae and pelvis) and spleen were harvested and bones were crushed with a mortar and pestle then filtered in to single cell suspensions. Bone marrow and spleen cells were blocked with mouse Fc block (BD Biosciences) and human IgG (Sigma) and then stained with fluorochrome-conjugated antibodies specific to human hematopoietic cells. For multilineage engraftment analysis, cells from mice were stained with CD45 (HI30) (Invitrogen), CD33 (P67.6), CD15 (HI98), CD14 (MϕP9), CD19 (HIB19), CD235a/GlyA (GA-R2), CD41a (HIP8) and CD34 (581) (BD Biosciences).

HSC and STRC xenotransplantations

For MSI2 KD in HSCs 5.0×10^4 and 2.5×10^4 sorted Lin^- $CD34^+$ $CD38^-$ cells were used per short-hairpin transduction experiment leading to transplantation of day zero equivalent cell doses of 10×10^3 and 6.25×10^3, respectively per mouse. For STRC LDA transplantation experiments 1×10^5 sorted $CD34^+CD38^+$ cells were used per control and MSI2 OE transduction. After assessing levels of gene transfer, day zero equivalent GFP^+ cell doses were calculated to perform the LDA. Recipients with greater than 0.1% $GFP^+CD45^{+/-}$ cells were considered to be repopulated. For STRC experiments that readout extended engraftment at 6.5 weeks, 2×10^5 $CD34^+$ $CD38^+$ cells were used per OE and control transduction allowing for non-limiting 5×10^4 day zero equivalent cell doses per mouse. For HSC expansion and LDA experiments $CD34^+CD38^-$ cells were sorted and transduced with MSI2 OE or control vectors (50,000 cells per condition) for three days and then analyzed for gene-transfer levels (% $GFP^{+/-}$) and primitive cell marker expression (% CD34 and CD133). To ensure equal numbers of GFP^+ cells were transplanted in both control and MSI2 OE mice, we added identically cultured GFP^- cells to the MSI2 culture to match % GFP^+ of the control culture (necessary due to the differing efficiency of transduction). The adjusted OE

culture was recounted and aliquoted (63,000 cells) to match the output of half the control culture. Three day 0 equivalent GFP^+ cell doses (1000, 300 and 62) were then transplanted per mouse to perform the D3 primary LDA. A second aliquot of the adjusted OE culture was then taken and put in to culture in parallel with the remaining half of the control culture for performing another LDA after 7 days of growth (10 days total growth, D10 primary LDA). Altogether, 4 cell doses were transplanted, if converted back to day 0 equivalents these equaled approximately 1000, 250, 100, and 20 GFP^+ cells per mouse, respectively. Pooled bone marrow from six engrafted primary mice that received D10 cultured control or MSI2 OE cells (from 2 highest doses transplanted) was aliquoted in to 4 cell doses of 15 million, 10 million, 6 million, 2 million and 1 million cells. Numbers of GFP^+ cells within primary mice was estimated from nucleated cell counts obtained from NSG femurs, tibias and pelvis and from Colvin et al.[29]. Actual numbers of GFP^+ cells used for determining numbers of GFP^+ HSCs and the number of mice transplanted for all LDA experiments is found in Supplementary Tables 3-5. The cutoff for HSC engraftment was a demonstration of multilineage reconstitution which was set at BM having >0.1% GFP^+ $CD33^+$ and >0.1% GFP^+ $CD19^+$ cells. Assessment of HSC and STRC frequency was carried out by using ELDA software[30]. For all mouse transplantation experiments, mice were aged (6-12wk) and sex matched. All transplanted mice were included for analysis unless mice died from radiation sickness before the experimental endpoint. No randomization or blinding was performed for animal experiments. Approximately 3-6 mice were used per cell dose for each CB transduction and transplantation experiment.

UV cross-linking immunoprecipitation sequencing (CLIP-seq) library preparation

CLIP-seq was performed as previously described[15]. Briefly, 25 million NB4 cells, a transformed human cell line of hematopoietic origin, were washed in PBS and UV-cross-linked at 400mJ/cm^2 on ice. Cells were pelleted, lysed in wash buffer (PBS, 0.1% SDS, 0.5% Na-Deoxycholate, 0.5% NP-40), DNAse treated, and supernatants from lysates were collected for immunoprecipitation. MSI2 was immunoprecipitated overnight using 5 μg of anti-MSI2 antibody (EP1305Y, Abcam) and Protein A Dynabeads (Invitrogen). Beads containing immunoprecipated RNA were washed twice with wash buffer, high-salt wash buffer (5X PBS, 0.1% SDS, 0.5% Na-Deoxycholate, 0.5% NP-40), and PNK buffer (50 mM Tris-Cl pH 7.4, 10 mM $MgCl_2$, 0.5% NP-40). Samples were then treated with 0.2 U MNase for 5 minutes at 37 degrees with shaking to trim immunopreciptated RNA. MNase inactivation was then carried out with PNK + EGTA buffer (50 mM Tris-Cl pH 7.4, 20 mM EGTA, 0.5% NP-40). The sample was dephosphorylated using alkaline phosphatase (CIP, NEB) at 37 degrees for 10 followed by washing with PNK+EGTA, PNK buffer, and then 0.1 mg/mL BSA in nuclease free water. 3’ RNA linker ligation was performed at 16 degrees overnight with the following adapter: 5' PUGGAAUUCUCGGGUGCCAAGG-puromycin. Samples were then washed with PNK buffer, radiolabelled using P32-y-ATP (Perkin Elmer), run on a 4-12% Bis-Tris gel and then transferred to a nitrocellulose membrane. The nitrocellulose membrane was developed via autoradiography and RNA-protein complexes 15-20 kDa above the molecular weight of MSI2 was extracted with Proteinase K followed by RNA extraction with acid phenol-chloroform. A 5’ RNA linker (5' HO-GUUCAGAGUUCUACAGUCCGACGAUC-OH) was ligated to the extracted RNA using T4 RNA ligase (Fermentas) for two hours and the RNA was again purified using acid

phenol-chloroform. Adapter ligated RNA was re-suspended in nuclease free water and reverse transcribed using Superscript III reverse transcriptase (Invitrogen). 20 cycles of PCR were performed using NEB Phusion Polymerase using a 3' PCR primer that contained a unique Illumina barcode sequence. PCR products were run on an 8% TBE gel. Products ranging between 150-200 bp were extracted using the QIAquick gel extraction kit (Qiagen) and re-suspended in nuclease free water. Two separate libraries were prepared and sent for single-end 50 bp Illumina sequencing at the Institute for Genomic Medicine at the University of California, San Diego. 47,098,127 reads from the first library passed quality filtering of which 73.83% mapped uniquely to the human genome. 57,970,220 reads from the second library passed quality filtering of which 69.53% mapped uniquely to the human genome. CLIP-data reproducibility was verified through high correlation between gene RPKMs and statistically significant overlaps in the clusters and genes within replicates. CLIP-seq data has been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE69583.

CLIP-seq mapping and cluster identification

Before sequence alignment of CLIP-seq reads to the human genome was performed, sequencing reads from libraries were trimmed of polyA tails, adapters, and low quality ends using Cutadapt with parameters --match-read-wildcards --times 2 -e 0 -O 5 --quality-cutoff' 6 -m 18 -b TCGTATGCCGTCTTCTGCTTG -b ATCTCGTATGCCGTCTTCTGCTTG -b CGACAGGTTCAGAGTTCTACAGTCCGACGATC -b TGGAATTCTCGGGTGCCAAGG -b AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA -b TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT. Reads were then mapped against a database of repetitive elements derived from RepBase (version 18.05). Bowtie (version 1.0.0) with parameters -S -q -p 16 -e 100 -l 20 was used to align reads against an index generated from Repbase sequences[31]. Reads not mapped to Repbase sequences were aligned to the hg19 human genome (UCSC assembly) using STAR (version 2.3.0e)[32] with parameters --outSAMunmapped Within –outFilterMultimapNmax 1 –outFilterMultimapScoreRange 1. To identify clusters in the genome of significantly enriched CLIP-seq reads, reads that were PCR replicates were removed from each CLIP-seq library using a custom script of the same method as[33], otherwise reads were kept at each nucleotide position when more than one read's 5' end was mapped. Clusters were then assigned using the CLIPper software with parameters --bonferroni --superlocal --threshold-[34]. The ranked list of significant targets was calculated assuming a Poisson distribution, where the observed value is the number of reads in the cluster, and the background is the number of reads across the entire transcript and or across a window of 1000 bp +/− the predicted cluster.

Gene annotations for CLIP-seq

Transcriptomic regions and gene classes were defined using annotations found in gencode v17. Depending on the analysis clusters were either associated by the Gencode annotated 5' UTR, 3' UTR, CDS or intronic regions. If a cluster overlapped multiple regions, or a single part of a transcript was annotated as multiple regions, clusters were iteratively assigned first as CDS, then 3' UTR, 5' UTR and finally as proximal (<500 bases from an exon) and distal (>500 bases from an exon) introns. Overlapping peaks were calculated using bedtools and pybedtools[35,36].

Gene ontology analysis for CLIP-seq

Significantly enriched gene ontology (GO) terms were identified using a hypergeometric test that compared the number of genes that were MSI2 targets in each GO term to genes expressed in each GO term as the proper background. Expressed genes were identified by using the control samples in SRA study SRP012062. Mapping was performed identically to CLIP-seq mapping, without peak calling and changing the STAR parameter outFilterMultimapNmax to 10. Counts were calculated with featureCounts[37] and RPKMs were then computed. Only genes with a mean RPKM > 1 between the two samples were used in the background expressed set.

De-novo motif and conservation analysis for CLIP-seq

Randomly located clusters within the same genic regions as predicted MSI2 clusters were used to calculate a background distribution for motif and conservation analyses. Motif analysis was performed using the HOMER algorithm as in Lovci et al.[34]. For evolutionary sequence conservation analysis the mean (mammalian) phastCons score for each cluster was used.

Immunofluorescence

$CD34^+$ cells ($>5\times10^4$) were transduced with MSI2 OE or control lentivirus, after 3 days GFP^+ cells were sorted and then put back in to StemSpan medium containing growth factors IL-6 (20 ng/ml), SCF (100 ng/ml), FLT3-L (100 ng/ml) and TPO (20 ng/ml). A minimum of 10,000 cells were used for immunostaining at culture days 3 and 7 post GFP-sort. Cells were fixed in 2% PFA for 10 minutes, washed with PBS and then cytospun on to glass slides. Cytospun cells were then permeabilized (PBS 0.2 % Triton X-100) for 20 minutes, blocked (PBS 0.1% saponin 10% donkey serum) for 30 minutes and stained with primary antibodies (anti-CYP1B1 antibody, EPR14972, Abcam; anti-HSP90 antibody, 68/hsp90, BD Biosciences) in PBS 10% donkey serum for 1 hour. Detection with secondary antibody was performed in PBS 10% donkey serum with Alexa-647 donkey anti-rabbit antibody or Alexa-647 donkey anti-mouse antibody for 45 minutes. Slides were mounted with Prolong Gold Antifade containing DAPI (Invitrogen). Several images (200-1000 cells total) were captured per slide at 20X magnification using an Operetta HCS Reader (Perkin Elmer) with epifluorescence illumination and standard filter sets. Columbus software (Perkin Elmer) was used to automate the identification of nuclei and cytoplasm boundaries in order quantify mean cell fluorescence.

Luciferase reporter gene assay

A 271 bp region of the *CYP1B1* 3'UTR that flanked CLIP-seq identified MSI2 blinding sites was cloned from human HEK293FT genomic DNA using the forward primer GTGACACAACTGTGTGATTAAAAGG and reverse primer TGATTTTTATTATTTTGGT AATGGTG and placed downstream of renilla luciferase in the dual-luciferase reporter vector, pGL4 (Promega). A 271 bp geneblock (IDT) with 6 TAG>TCC mutations was cloned in to pGL4 using XbaI and NotI. The *HSP90* 3'UTR was amplified from HEK293FT genomic DNA with the forward primer TCTCTGGCTGAGGGATGACT and reverse primer TTTTAAGGCCAAGGAATTAAGTGA and cloned in pGL4. A geneblock of the *HSP90*

3'UTR (IDT) with 14 TAG>TCC mutations was cloned in to pGL4 using SfaAI and NotI. Co-transfection of wild type or mutant luciferase reporter (40 ng) and control or MSI2 OE lentiviral expression vector (100 ng) was performed in the non-Musashi1 and 2 expressing NIH-3T3 cell line. 50,000 cells were used per co-transfection. Reporter activity was measured using the Dual-Luciferase Reporter Assay System (Promega) 36-40 hours later.

MSI2 OE suspension cultures with AHR antagonist SR1 and agonist FICZ

For MSI2 OE cultures with the AHR antagonist SR1, Lin^- $CD34^+$ cells were transduced with MSI2 OE or control lentivirus in medium supplemented with SR1 (750 nM; Abcam) or DMSO vehicle (0.1%). GFP^+ cells were isolated (20,000 cells per culture) and allowed to proliferate with or without SR1 for an additional 7 days at which point they were counted and immunophenotyped for CD34 and CD133 expression. For MSI2 OE cultures with the AHR agonist FICZ, Lin^- $CD34^+$ cells were transduced with MSI2 OE or control lentivirus. GFP^+ cells were isolated (20,000 cells per culture) and allowed to proliferate with FICZ (200 nM; Santa Cruz Biotechnology) or DMSO (0.1%) for an additional 3 days at which point they were immunophenotyped for CD34 and CD133 expression.

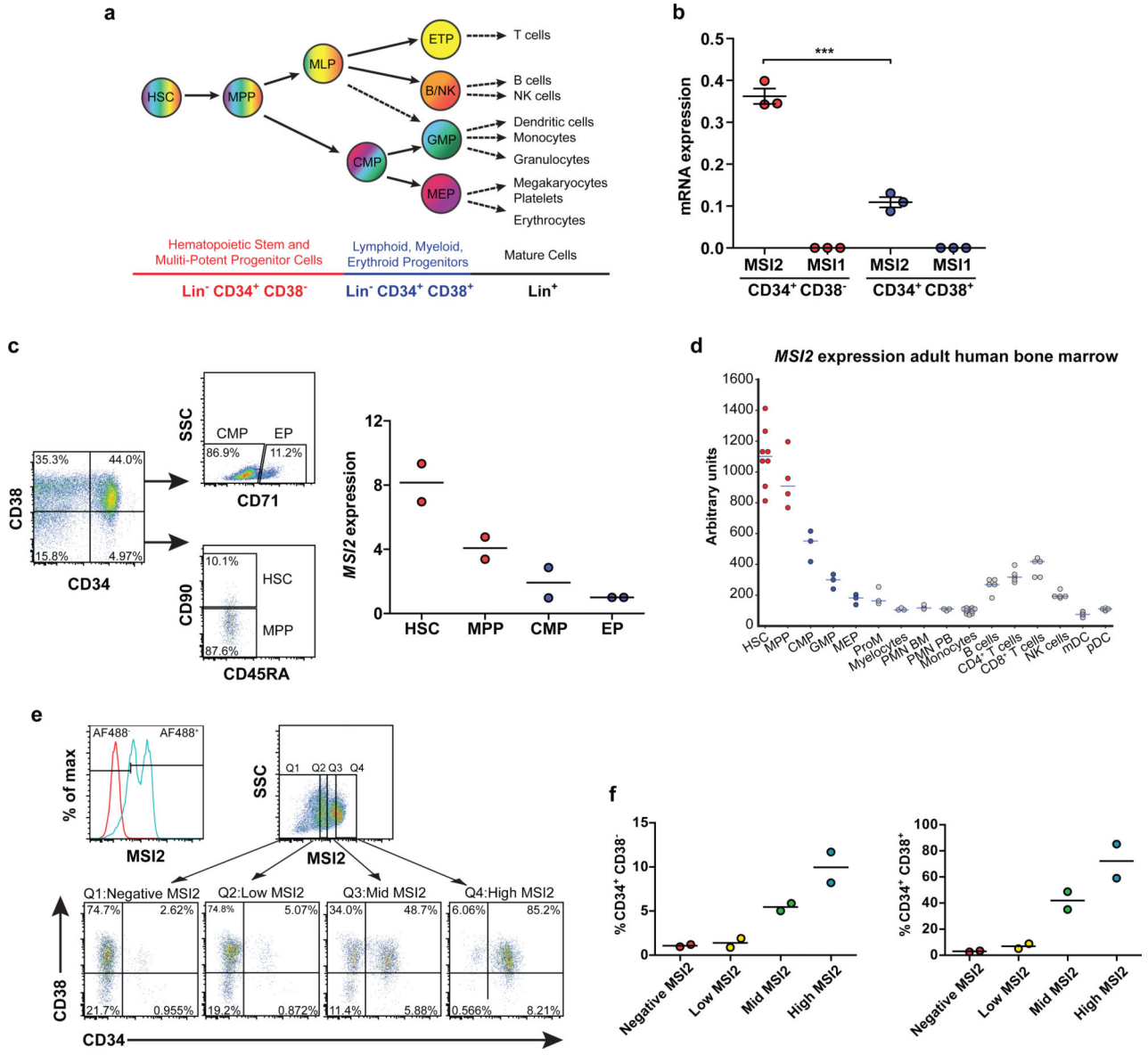
HSPC expansion with (E)-2,3',4,5'-Tetramethoxystilbene (TMS)

Lin^- $CD34^+$ cells were cultured for 72 hours (lentiviral treated but non-transduced flow-sorted GFP^- cells) in StemSpan medium containing growth factors IL-6 (20 ng/ml), SCF (100 ng/ml), FLT3-L (100 ng/ml) and TPO (20 ng/ml) before the addition of the CYP1B1 inhibitor TMS (Abcam) at a concentration of 10 μM or mock treatment with 0.1% DMSO. Equal numbers of cells (12,000 per condition) were then allowed to proliferate for 7 days at which point they were counted and immunophenotyped for CD34 and CD133 expression.

Statistical analysis

Unless stated otherwise (i.e., analysis of RNA-seq and CLIP-seq data sets), all statistical analysis was performed using GraphPad Prism (GraphPad Software version 5.0). Unpaired student t-tests or Mann-Whitney tests were performed with $p < 0.05$ as the cut-off for statistical significance.

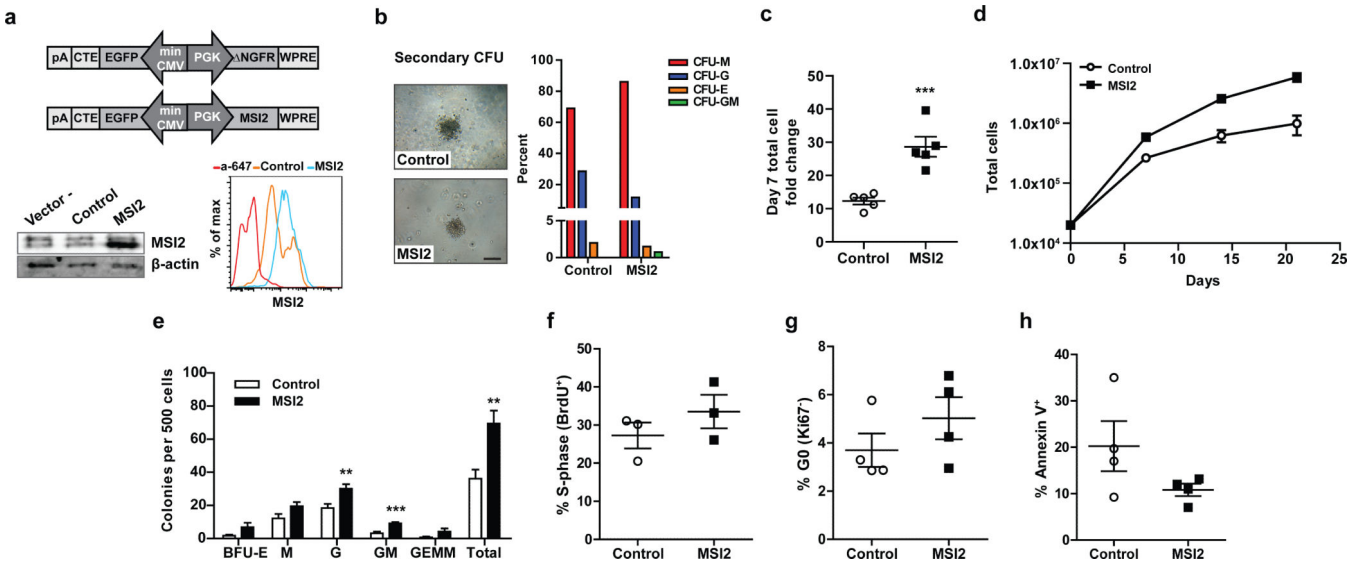
Extended Data



Extended Data Figure 1. MSI2 is highly expressed in human hematopoietic stem and progenitor cell populations

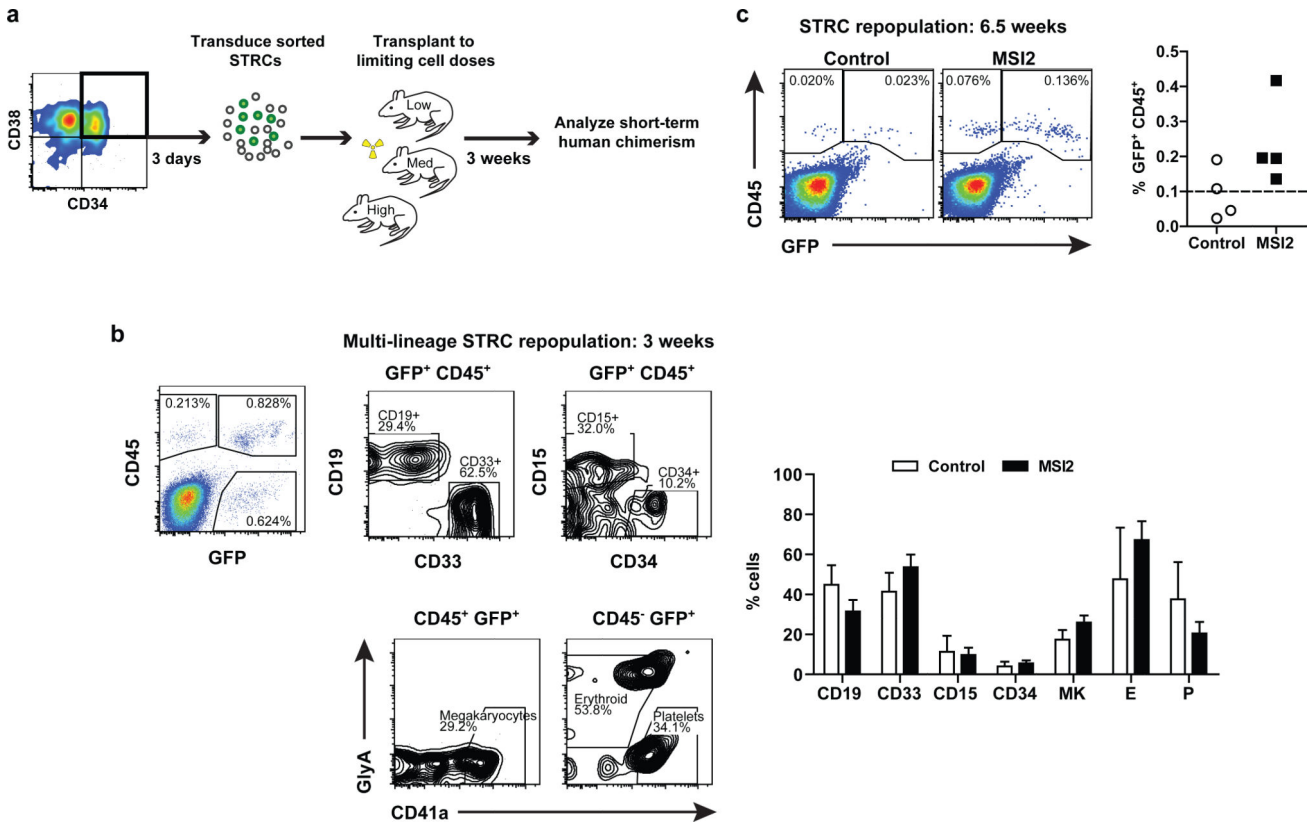
a, Schematic of the human hematopoietic hierarchy showing key primitive cell populations and simplified surface marker expression. **b,** qRT-PCR analysis of *MSI1* and *MSI2* expression in Lin$^-$ CB cell populations (n=3 independent Lin- CB samples). **c,** Gating strategy used to sort sub-fractions of Lin$^-$ CB HSPCs for *MSI2* qRT-PCR expression analysis (n=2 independent pooled Lin$^-$ CB samples). **d,** *MSI2* expression across the human hematopoietic hierarchy. Each circle represents an independent gene expression dataset curated by HemaExplorer. **e,** Intracellular flow cytometry analysis of MSI2 protein levels in Lin$^-$ CB. Histograms show background staining with secondary antibody (red) and positive staining with anti-MSI2 antibody plus secondary in Lin$^-$ CB (blue). MSI2 fluorescence

intensity was divided into quartiles of negative (Q1), low (Q2), mid (Q3) and high (Q4) level expression. **f,** Plots show cell percentage within each quartile from **e** that are $CD34^+$ $CD38^-$ (left) and $CD34^+$ $CD38^+$ (right) (n=2 independent Lin^- CB samples). All data presented as mean ± SEM. *p<0.05; ***p<0.001.



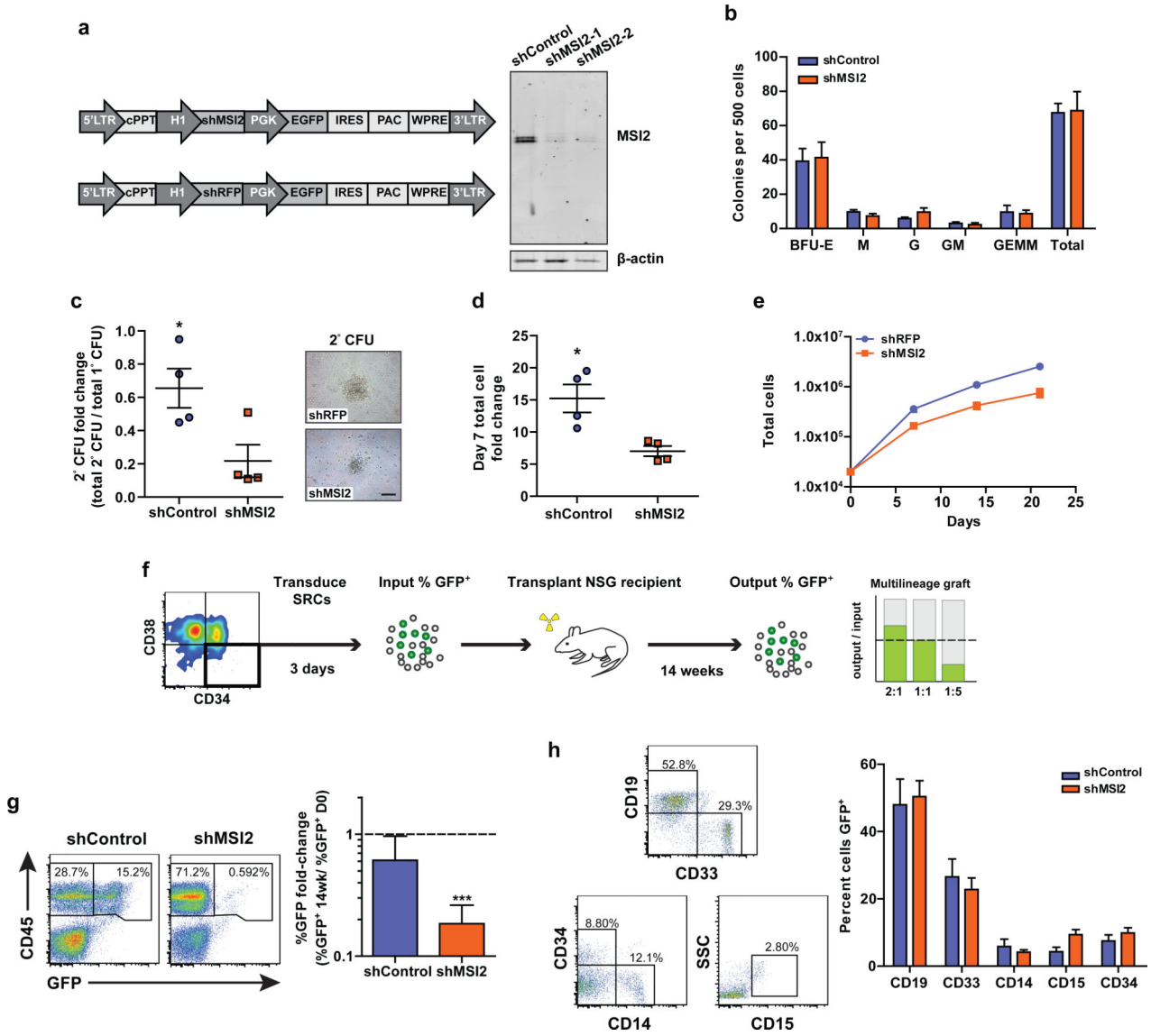
Extended Data Figure 2. MSI2 OE enhances in vitro culture of primitive CB cells

a, Top: Schematic of bi-directional promoter lentivirus used to overexpress MSI2. Bottom: Western blot and histogram showing intracellular flow validation of enforced MSI2 expression in 293FT cells (left) and Lin^- CB (right), respectively. **b,** Representative images of secondary CFU made from replated control and MSI2 OE CFU-GEMMs and types of colonies made, scale bar = 200 μm. **c,** Fold change in Lin^- CB transduced cell number after 7 days in culture following transduction (n=5 experiments). **d,** 21 day growth curve of transduced Lin^- CB cells (n=4 experiments). **e,** Colony output of transduced Lin^- CB from day 7 cultures (n=8 cultures from 4 experiments). **f,** BrdU cell cycle analysis of transduced Lin^- CB cells from day 10 cultures (n=3 experiments). **g,** Ki67 cell cycle analysis of transduced Lin^- CB cells from day 4 cultures (n=4 experiments). **h,** Apoptotic and dead cells in day 7 cultures of transduced Lin^- CB by Annexin V staining (n=3 experiments). Western blot source data in Supplementary Figure 1. All data presented as mean ± SEM. **p<0.01; ***p<0.001.



Extended Data Figure 3. MSI2 OE does not affect STRC lineage output and extends STRC-mediated engraftment

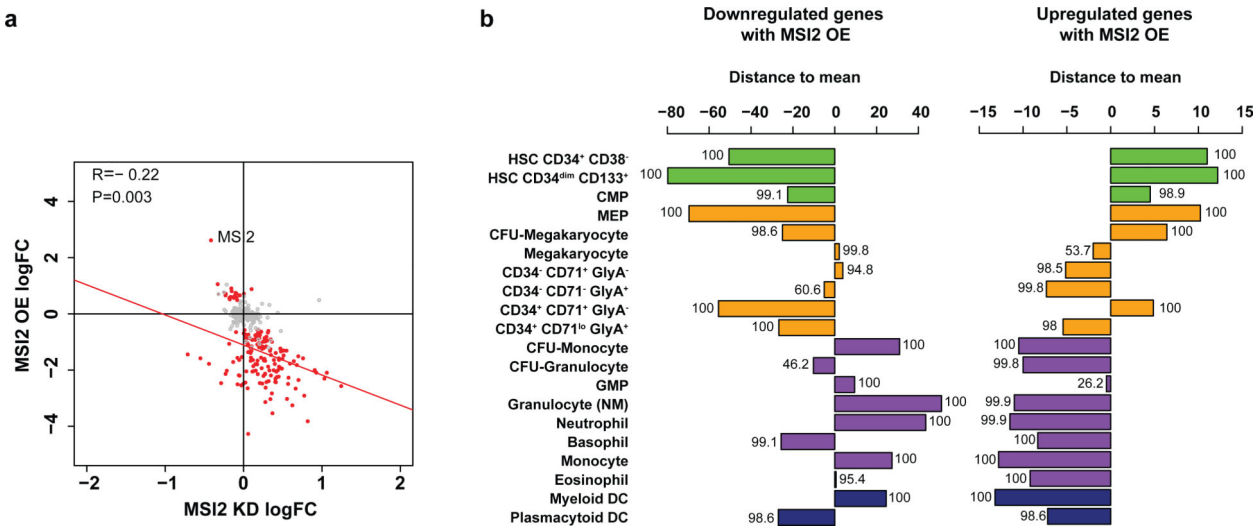
a, Schematic of STRC LDA experimental setup. **b,** Left: Gating strategy to identify engrafted GFP$^+$ CD45$^+$ progenitor and myelo-lympho lineage positive cell types or GFP$^+$ CD45$^-$ erythroid cells and platelets. Right: Summary of lineage output in the injected femur at 3 weeks post-transplant (n=4 control and n=18 MSI2 OE mice). MK = Megakaryocyte, E = Erythroid cells, P = Platelets. **c,** Representative flow plots and summary of transduced STRCs read out for human CD45$^+$ engraftment at 6.5 weeks post-transplant (n=4 mice per condition). All data presented as mean ± SEM.



Extended Data Figure 4. MSI2 KD impairs secondary CFU replating potential and HSC engraftment capacity

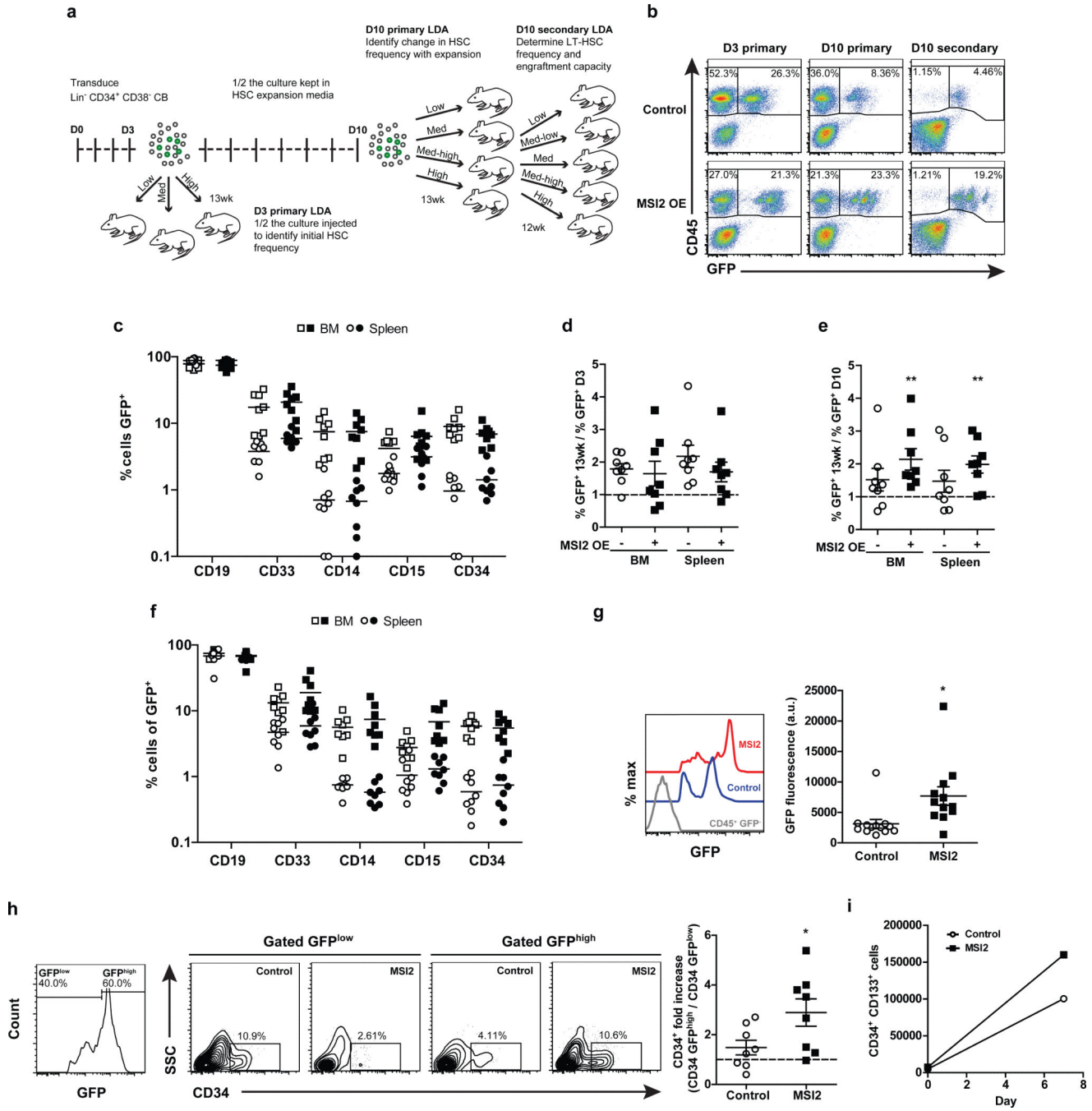
a, Left: Schematic of MSI2- and control RFP-targeted shRNA lentiviruses. Right: Confirmation of MSI2 protein knockdown (both isoforms that can be detected by western blot) in transduced NB4 cells. **b,** CFU production by shMSI2 and shControl transduced Lin^- CB (n=8 cultures from 4 experiments). **c,** Secondary CFU output from shMSI2 and images of representative secondary CFU (scale bar = 200 μm, performed on n=4 cultures from 2 experiments). **d,** Fold change in transduced cell number after 7 days in culture (n=4 experiments). **e,** Growth curves of cultures initiated with transduced Lin^- CB cells (n=4 experiments). **f,** Experimental design to readout changes in HSC capacity with MSI2 KD. **g,** Left: Representative flow analysis of transduced $CD34^+$ $CD38^-$ derived human chimerism in NSG mouse BM. Right: Shown is the ratio of the percentage of GFP^+ cells in the $CD45^+$ population post-transplant to the initial pre-transplant GFP^+ cell percentage. Dotted line indicates the proportion of GFP^+ cells is unchanged relative to input. (One sample t-test, no

change = 1, n = 6 shControl and n = 8 shMSI2 mice pooled from 2 experiments). **h,** Representative flow plots and summary of multilineage engraftment with shControl and shMSI2 (gated on GFP$^+$ cells). Western blot source data is shown in Supplementary Figure 1. Data presented as mean ± SEM. *p<0.05; ***p<0.001.



Extended Data Figure 5. MSI2 OE confers an HSC gene expression signature

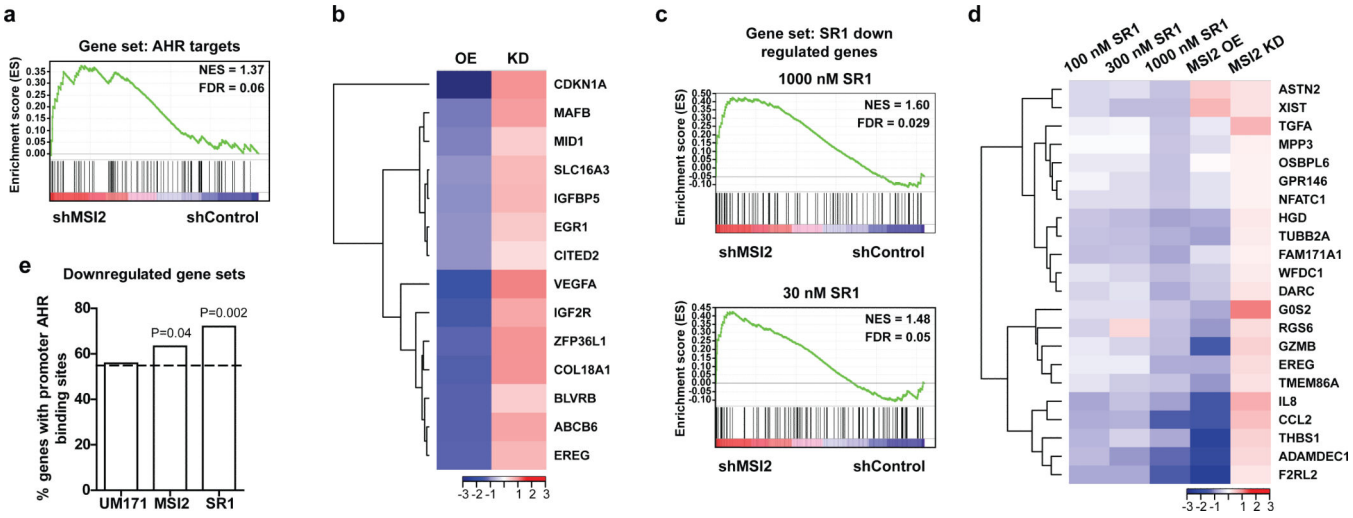
a, Genes that are up in OE (21 genes, logFC>0) and down (156 genes, logFC<0) relative to control with a FDR<0.05 were compared to MSI2 KD normalized to shControl expression data. Red circles represent 177 significantly differentially expressed genes with OE. Gray outlined circles represent random genes (number of gray circles = number of red circles). Only genes that are significantly up or down regulated in OE have anti-correlation with KD. **b,** Differentially expressed genes between MSI2 OE and control cells (FDR <0.05) compared to DMAP populations. Numbers beside each bar indicate the percentage of the time the observed value (set of up or downregulated genes) was better represented in that population than random values (equal number of randomly selected genes based on 1000 trials).



Extended Data Figure 6. MSI2 OE enhances HSC activity after ex vivo culture

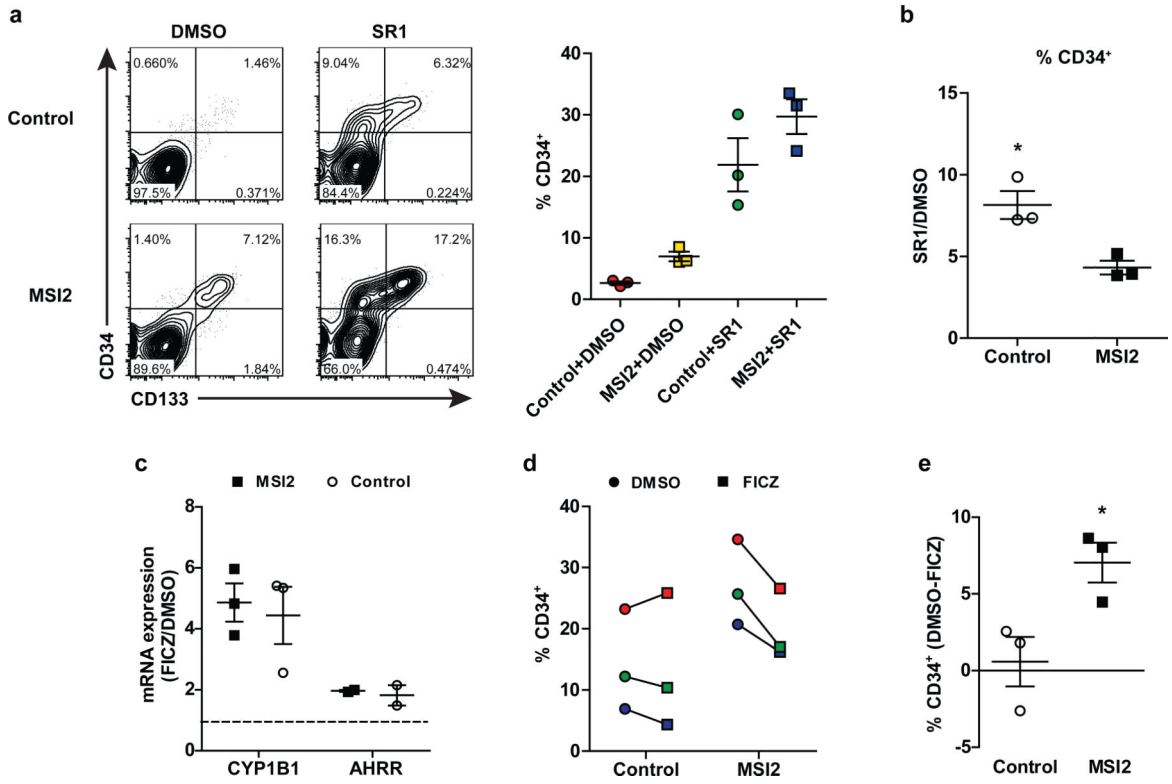
a, Experimental layout to measure changes in HSC engraftment capacity and frequency with ex vivo culture. **b,** Representative flow plots of $CD45^+$ GFP^+ reconstitution from mice receiving the highest cell dose transplanted per time point. **c,** Multilineage engraftment of mice injected with D3 cultures. **d,** Proportion of the human $CD45^+$ graft containing GFP^+ cells from D3 primary mice relative to pre-transplant levels of GFP^+ cells before expansion. Shown are mice transplanted at the two highest doses (n=8 mice for both conditions). **e,** Proportion of the human $CD45^+$ graft containing GFP^+ cells from D10 primary mice relative

to pre-transplant levels of GFP$^+$ cells after expansion. Shown are mice transplanted at the two highest doses (n=8 mice for both conditions, one sample t-test, no change = 1). **f,** Summary of multilineage engraftment from mice in **c**. **g,** GFP mean fluoresence intensity (MFI) in D10 primary engrafted mice. Shown are mice transplanted with the highest three doses, n=11 control and 13 MSI2 mice. **h,** CD34 expression in GFP-high (top 60%) relative to GFP-low (bottom 40%) gated cells (set per mouse) from engrafted recipients in **c**. **i,** Number of transduced phenotyped HSCs after 7 days of culture from HSC expansion experiment described in **a**. Symbols represent individual mice and shaded symbols represent MSI2 OE. All data presented as mean ± SEM.



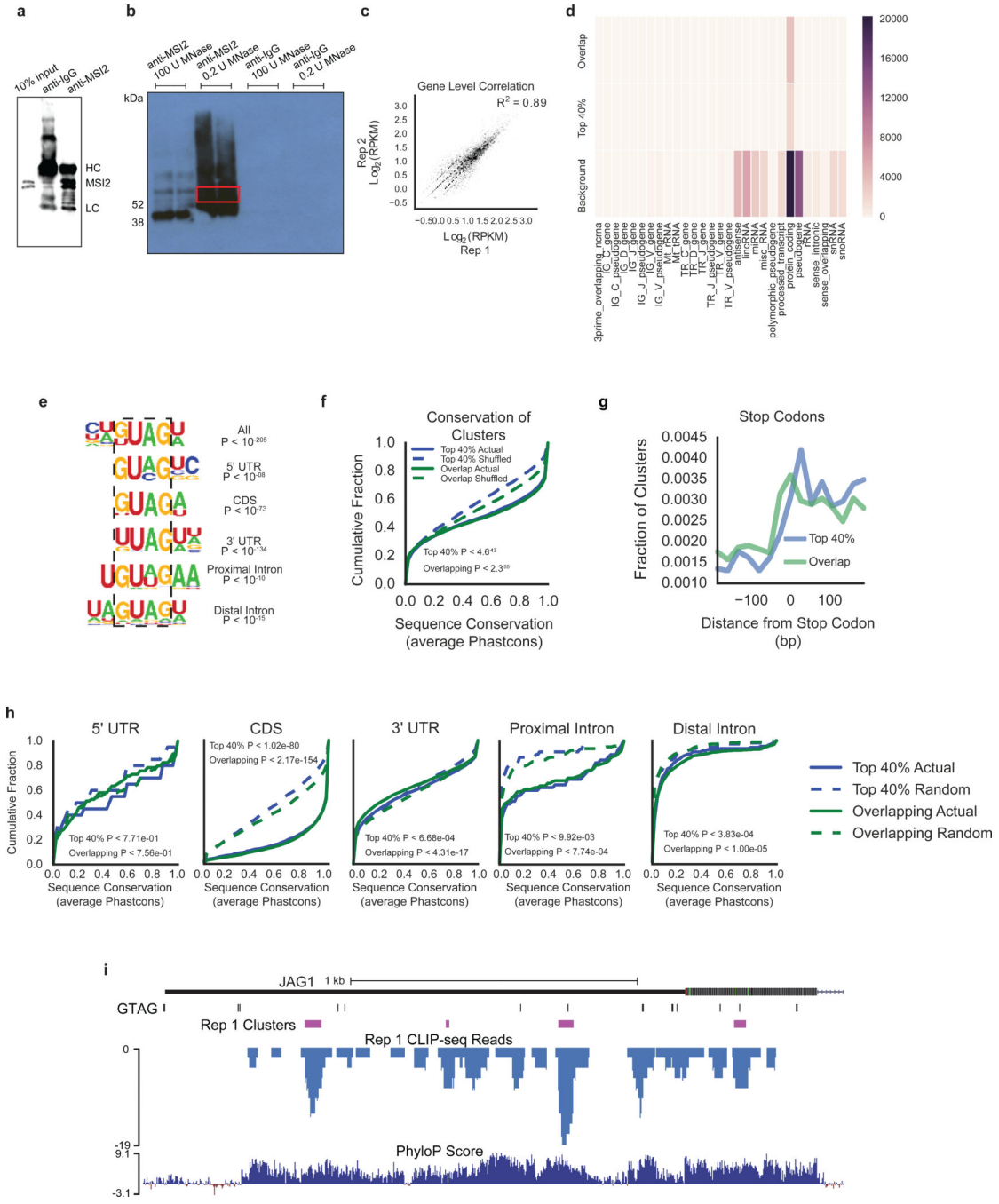
Extended Data Figure 7. Predicted AHR targets and genes downregulated with SR1 and MSI2 OE are upregulated with MSI2 KD

a, Predicted AHR targets were identified with the iRegulon tool and compared with MSI2 KD normalized to shControl upregulated gene signature by GSEA. **b,** Heatmap of MSI2 OE and KD shared leading edge AHR target genes from GSEA. **c,** GSEA comparing SR1 high and low dose downregulated gene sets with the MSI2 KD upregulated gene signature. **d,** Heatmap shows list of shared leading edge genes identified by GSEA from MSI2 OE, KD and SR1 at varying doses. **e,** The percentage of downregulated genes from UM171-, SR1-treated and MSI2 OE data sets containing at least one AHR binding site 1500 bp upstream and downstream of transcription start site. Dotted line indicates the background percentage of genes with AHR binding sites. P-values were generated relative to background with Fisher's exact test.



Extended Data Figure 8. AHR antagonism with SR1 has redundant effects with MSI2 OE, and AHR activation with MSI2 OE results in a loss of HSPCs

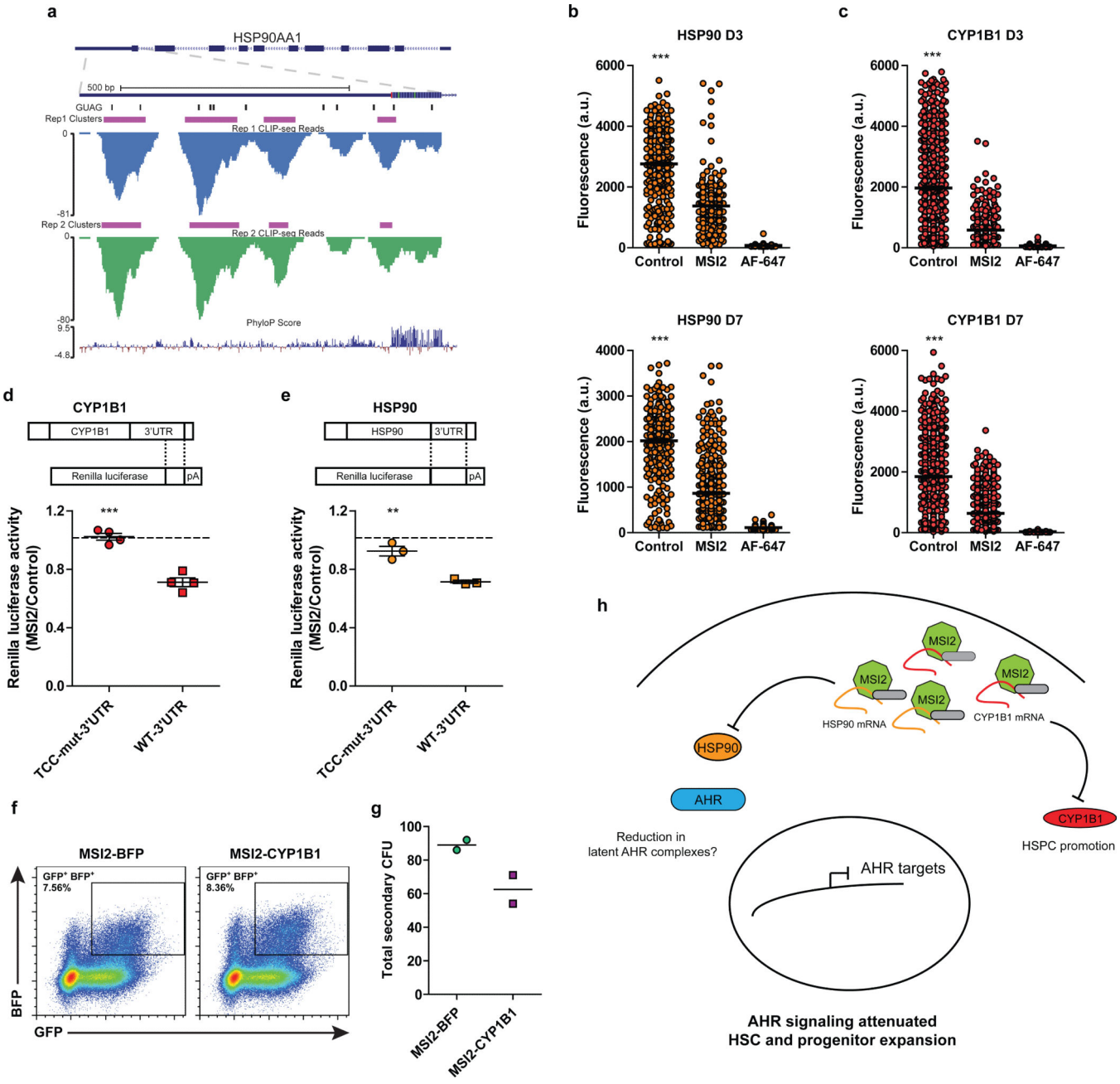
a, Representative flow plots and summary of CD34 expression with MSI2 OE and control transduced CD34$^+$ CB cells grown for 10 days in medium containing SR1 or DMSO vehicle (n=3 experiments). **b,** Fold change in CD34 expression from results shown in **a**. **c,** Fold-increase in *CYP1B1* and *AHRR* transcript levels with FICZ in transduced cultures (n=3 experiments). **d,** Transduced CD34$^+$ CB cells grown for 3 days in medium supplemented with FICZ and the corresponding change in CD34 expression. Each coloured pair (DMSO and FICZ) represents a matched CB sample (n=3 experiments). **e,** Differences in culture CD34 levels from **d**. All data presented as mean ± SEM. * $p<0.05$.



Extended Data Figure 9. MSI2 preferentially binds mature mRNA within the 3'UTR

a, Validation of the capacity of the anti-MSI2 antibody to immunoprecipitate MSI2 compared to IgG control pulldowns. **b,** Autoradiogram showing anti-MSI2 immunoprecipitated, MNase digested and radiolabelled RNA isolated for CLIP library construction and sequencing (red box). Low levels of MNase show a smearing pattern extending upwards from the modal weight of MSI2. **c,** Scatter plot of total number of uniquely mapped CLIP-seq reads for each gene, comparing both replicates. **d,** Heatmap indicating the number of different classes of Gencode annotated genes that contain at least

one predicted MSI2 binding site. **e,** Consensus motifs within MSI2 clusters in the different genic regions. P-values for the most statistically significant enriched motif is presented for all overlapping clusters between replicates. **f,** Cumulative distribution function of mean conservation score (Phastcons) of MSI2 clusters, compared to a shuffled background control, computed for all overlapping clusters and the top 40% of overlapping clusters. P-values were obtained by a Kolgomorov-Smirnov two-tailed test comparing the distributions from actual and shuffled locations. **g,** Number of clusters within 200 bases of the annotated stop codon in known mRNA transcripts for all overlapping clusters between replicates and the top 40% of overlapping clusters. **h,** Cumulative distribution function of mean conservation score (Phastcons) of MSI2 clusters, compared to a shuffled background control, computed for overlapping clusters between the replicates and the top 40% of overlapping clusters found in different genic regions. Similarity in the 3'UTR conservation for the top 40% with the shuffled background control is likely due to MSI2 sites being small and not needing structural contexts for conservation. P-values were obtained by a Kolgomorov-Smirnov two-tailed test comparing the distributions from actual and shuffled locations. **i,** Genome browser views displaying CLIP-seq mapped reads from replicate 1 (blue), predicted clusters (purple), exact matches for the GUAG sequence (black) and mammal conservation scores (PhyloP) in the 3' UTRs for a previously predicted Msi1 target.



Extended Data Figure 10. MSI2 OE represses CYP1B1 and HSP90 3'UTR Renilla Luciferase reporter activity

a, CLIP-seq reads (replicate 1 in blue and replicate 2 in green) and clusters (purple) mapped to the 3'UTR of HSP90. Matches to the GUAG motif are shown in black. Mammal PhyloP score listed in last track. **b and c,** Representative data of mean per cell fluoresence for HSP90 and CYP1B1 protein in transduced CD34^{+} CB. Protein level in cells during in vitro culture was analyzed 3 days (D3) and 7 days (D7) after transduction and sorting for GFP. Corresponding secondary alone antibody staining is shown for each experiment. Each circle represents a cell, and greater than 200 cells were analyzed per condition. **d and e,** Levels of renilla luciferase activity in NIH-3T3 cells co-transfected with control or MSI2 OE vectors

and the CYP1B1 or HSP90 wild type or TCC mutant 3'UTR luciferase reporter (dotted line indicates no change in renilla activity; n=4 CYP1B1 and n=3 HSP90 experiments). **f,** Flow plots of co-transduced $CD34^+$ CB cells with MSI2 (GFP) and CYP1B1 (BFP) lentivirus. **g,** GFP^+ BFP^+ CFU-GEMMs generated from **f** were replated in to secondary CFU assays and enumerated for total number of colonies formed. A total of 24 CFU-GEMMs from MSI2-BFP and MSI2-CYP1B1 were replated (n=2 experiments). Data presented as mean ± SEM. ***p<0.001, **p<0.01. **h,** A model for AHR pathway attenuation through MSI2 post-transcriptional processing. MSI2 mediates the post-transcriptional down regulation of HSP90 at the outset of culture and continuously represses the prominent AHR pathway effector CYP1B1 to facilitate HSPC expansion. The resultant MSI2-mediated repression of AHR signaling enforces a self-renewal program and allows HSPC expansion ex vivo.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Eric Lechman and Peter Van Galen for providing H1 and pSMALB vectors. The MA overexpression vector was provided as a gift by Luigi Naldini. We also thank the SCC-RI core flow cytometry staff, the Obstetrics and Gynecology Unit at McMaster Children's Hospital for cord blood, Brad Doble and Mick Bhatia for critical assessment of this work and all members of the Hope laboratory for experimental support and advice. This work was supported by an Ontario Institute for Cancer Research New Investigator Award (IA-033), an Ontario Institute for Cancer Research Cancer Stem Cell Program Team Grant (P.CSC.005) and a Canadian Institutes of Health Research (MOP-126030) grant to K.H. S.R. is supported by a Canadian Blood Services Graduate Fellowship and Health Canada. The views expressed herein do not necessarily represent the view of the federal government of Canada. This work was partially supported by grants from the National Institute of Health (HG004659 and NS075449) to G.W.Y. and from the California Institute of Regenerative Medicine (RB3-05219) to G.W.Y. G.P. was supported by a National Science Graduate Fellowship. G.W.Y. is an Alfred P. Sloan Research Fellow. We thank the UCSD Institute for Genomic Medicine's Genomics Center for providing access to high-throughput sequencing facilities.

References

1. Miller PH, Knapp DJ, Eaves CJ. Heterogeneity in hematopoietic stem cell populations: implications for transplantation. Curr Opin Hematol. 2013; 20:257–264. [PubMed: 23615054]
2. Boitano AE, et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science. 2010; 329:1345–1348. [PubMed: 20688981]
3. Fares I, et al. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. Science. 2014; 345:1509–1512. [PubMed: 25237102]
4. Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011; 144:296–309. [PubMed: 21241896]
5. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat Immunol. 2013; 14:756–763. [PubMed: 23708252]
6. Hope K, et al. An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell. 2010; 7:101–114. [PubMed: 20621054]
7. de Andrés-Aguayo L, et al. Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood. 2011; 118:554–618. [PubMed: 21613258]
8. Park SM, et al. Musashi-2 controls cell fate, lineage bias, and TGF-beta signaling in HSCs. J Exp Med. 2014; 211:71–87. [PubMed: 24395885]
9. Ohyama T, et al. Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. Nucleic Acids Res. 2012; 40:3218–3231. [PubMed: 22140116]

10. Glimm H, et al. Previously undetected human hematopoietic cell populations with short-term repopulating activity selectively engraft NOD/SCID-beta2 microglobulin-null mice. J Clin Invest. 2001; 107:199–206. [PubMed: 11160136]
11. Cashman JD, Eaves CJ. Human growth factor-enhanced regeneration of transplantable human hematopoietic stem cells in nonobese diabetic/severe combined immunodeficient mice. Blood. 1999; 93:481–487. [PubMed: 9885209]
12. Holyoake TL, Nicolini FE, Eaves CJ. Functional differences between transplantable human hematopoietic stem cells from fetal liver, cord blood, and adult marrow. Exp Hematol. 1999; 27:1418–1427. [PubMed: 10480433]
13. Mimura J, Fujii-Kuriyama Y. Functional role of AhR in the expression of toxic effects by TCDD. Biochim Biophys Acta. 2003; 1619:263–268. [PubMed: 12573486]
14. Lo R, Matthews J. High-Resolution Genome-wide Mapping of AHR and ARNT Binding Sites by ChIP-Seq. Toxicol Sci. 2012; 130:349–361. [PubMed: 22903824]
15. Yeo GW, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol. 2009; 16:130–137. [PubMed: 19136955]
16. Katz Y, et al. Musashi proteins are post-transcriptional regulators of the epithelialluminal cell state. Elife. 2014; 3:e03915. [PubMed: 25380226]
17. Tijet N, et al. Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. Mol Pharmacol. 2006; 69:140–153. [PubMed: 16214954]
18. Doulatov S, et al. PLZF is a regulator of homeostatic and cytokine-induced myeloid development. Genes Dev. 2009; 23:2076–2163. [PubMed: 19723763]
19. Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell Stem Cell. 2007; 1:635–645. [PubMed: 18371405]
20. Bagger FO, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. Nucleic Acids Res. 2013; 41:D1034–1039. [PubMed: 23143109]
21. Amendola M, Venneri MA, Biffi A, Vigna E, Naldini L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. Nat Biotechnol. 2005; 23:108–116. [PubMed: 15619618]
22. van Galen P, et al. The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. Nature. 2014; 510:268–272. [PubMed: 24776803]
23. Lechman ER, et al. Attenuation of miR-126 activity expands HSC in vivo without exhaustion. Cell Stem Cell. 2012; 11:799–811. [PubMed: 23142521]
24. Carow CE, Hangoc G, Broxmeyer HE. Human multipotential progenitor cells (CFU-GEMM) have extensive replating capacity for secondary CFU-GEMM: an effect enhanced by cord blood plasma. Blood. 1993; 81:942–949. [PubMed: 7679010]
25. Milyavsky M, et al. A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. Cell Stem Cell. 2010; 7:186–197. [PubMed: 20619763]
26. Janky R, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput Biol. 2014; 10:e1003731. [PubMed: 25058159]
27. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. [PubMed: 16199517]
28. Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. G3 (Bethesda). 2012; 2:987–1002. [PubMed: 22973536]
29. Colvin GA, et al. Murine marrow cellularity and the concept of stem cell competition: geographic and quantitative determinants in stem cell biology. Leukemia. 2004; 18:575–583. [PubMed: 14749701]
30. Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods. 2009; 347:70–78. [PubMed: 19567251]
31. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

32. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29:15–21. [PubMed: 23104886]
33. Darnell R. CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. Cold Spring Harb Protoc. 2012; 2012:1146–1160. [PubMed: 23118367]
34. Lovci MT, et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat Struct Mol Biol. 2013; 20:1434–1442. [PubMed: 24213538]
35. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
36. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics. 2011; 27:3423–3424. [PubMed: 21949271]
37. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930. [PubMed: 24227677]

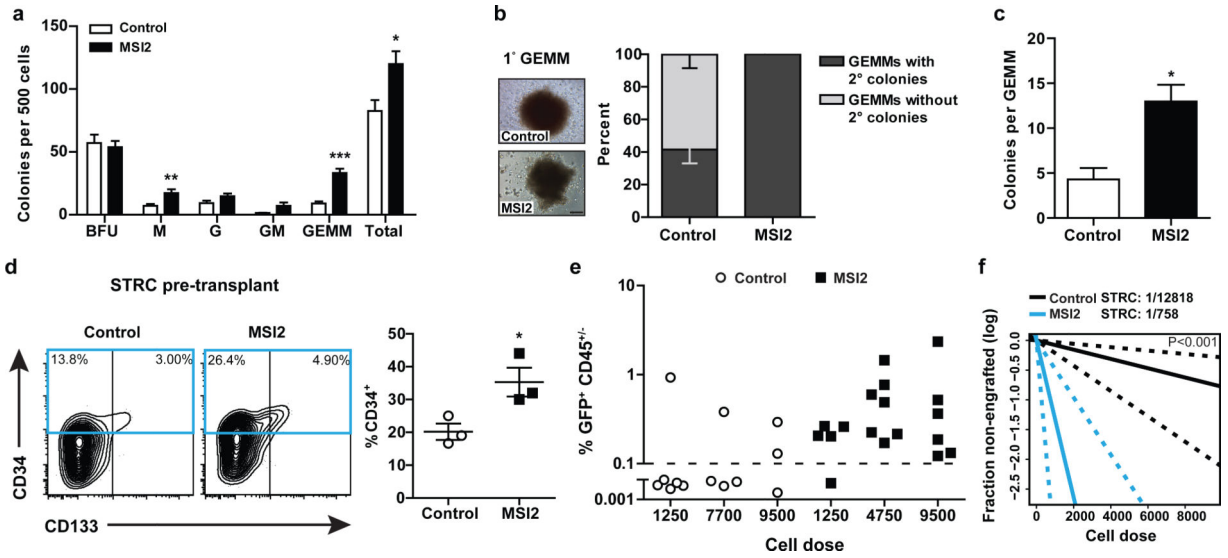


Figure 1. MSI2 OE enhances in vitro CB progenitor activity and increases numbers of STRCs

a, CFU output from transduced Lin$^-$ CB (n=9 control and 10 MSI2 OE cultures from 5 experiments). **b,** CFU-GEMM secondary CFU replating potential (n=24 control and 30 MSI2 OE from 2 experiments) and images of primary GEMMs (scale bar 200 μm). **c,** Number of secondary colonies per replated CFU-GEMM from **b**. **d,** CD34 expression in STRCs prior to transplant (n=3 experiments). **e,** Human chimerism at 3 weeks in mice transplanted with varying doses of transduced STRCs. Dashed line indicates engraftment cutoff (n=3 experiments). **f,** STRC frequency as determined by LDA from **e**. Dashed lines indicate 95% C.I. Data shown as mean ± SEM. *p<0.05; **p<0.01; ***p<0.001.

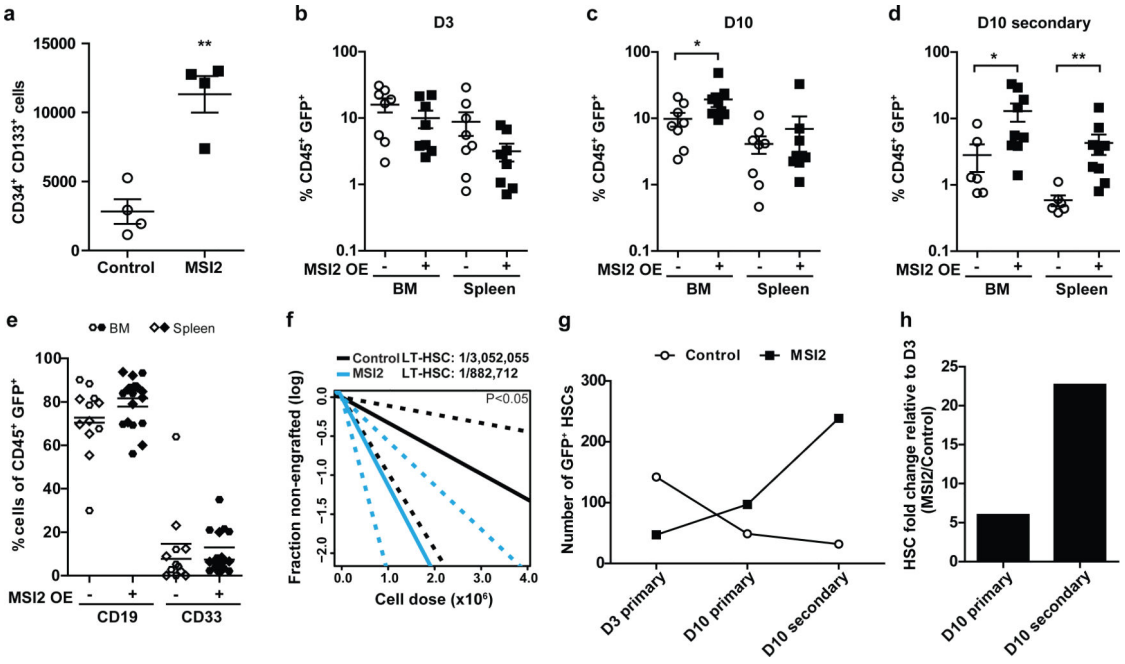


Figure 2. MSI2 OE expands LT-HSCs with ex vivo culture

a, Transduced $CD34^{+}CD133^{+}$ cells after one week of culture (n=4 experiments, unpaired t-test). **b-d,** $CD45^{+}GFP^{+}$ engraftment from mice receiving the highest two cell doses for D3 and D10 (n=8 mice for both conditions) and the highest three doses for D10 secondary mice (n=6 control and 9 MSI2 OE mice, Mann-Whitney test). **e,** Myelo-lymphopoiesis in D10 secondary mice. **f,** Multi-lineage LT-HSC frequency in BM cells of D10 primary mice. Dashed lines indicate 95% C.I. **g,** Numbers of GFP^{+} HSCs as evaluated by LDA. **h,** Cumulative fold change in MSI2 OE-transduced HSCs. Data shown as mean ± SEM. *p<0.05; **p<0.01.

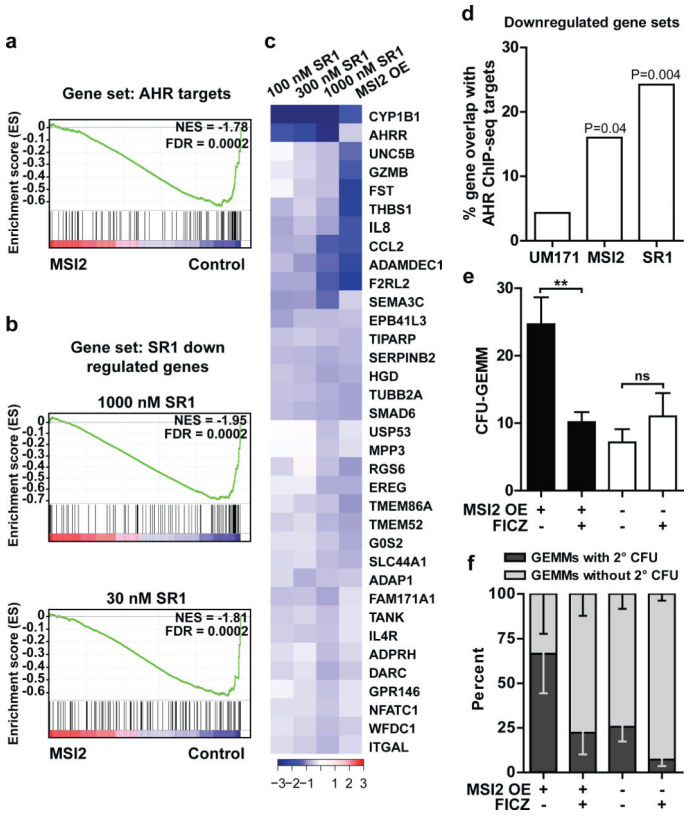


Figure 3. MSI2 OE in human HSPCs attenuates AHR signaling

a, Predicted AHR targets compared by GSEA to genes downregulated with MSI2 OE. **b,** GSEA of SR1 and MSI2 OE downregulated gene sets. **c,** Log fold-change of MSI2 OE and SR1 GSEA leading edge genes. **d,** Percentage of gene overlap between UM171-, SR1-treated and MSI2 OE downregulated gene sets and AHR ChIP-seq-identified targets. **e,** Number of CFU-GEMMs generated from transduced cells grown in CFU medium containing FICZ or DMSO (n=3 experiments). **f,** CFU-GEMMs from **e** replated into CFU assays containing FICZ or DMSO (n=30 control and 29 MSI2 OE per treatment). Data are presented as mean ± SEM. **p<0.01.

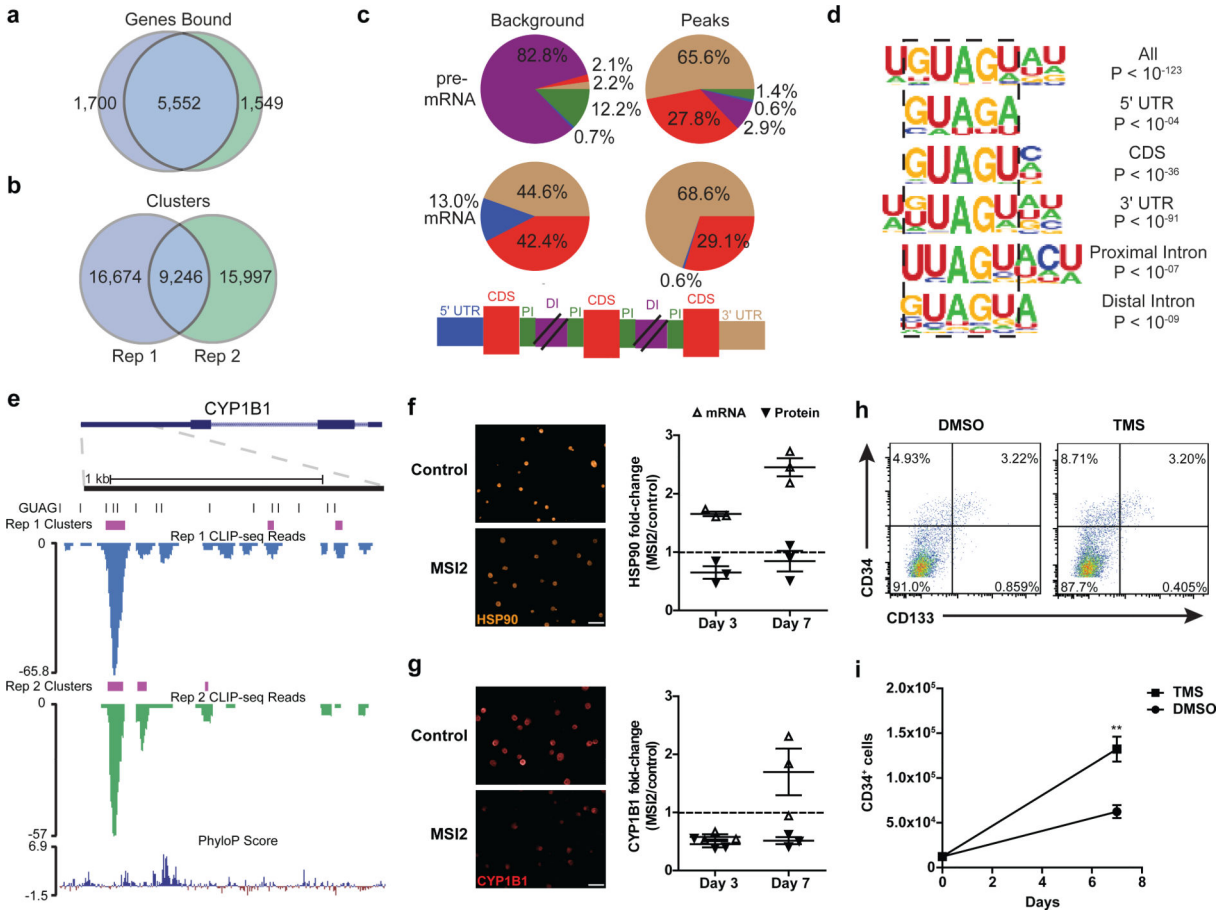


Figure 4. MSI2 OE post-transcriptionally downregulates AHR pathway components

a, Overlap between MSI2 target genes from separate CLIP-seq experiments. **b,** Statistically significant overlap ($p<0.0001$, hypergeometric test) of clusters between the replicates. **c,** Percent of CLIP-seq clusters in different genic regions. **d,** Consensus motifs within MSI2 clusters in different genic regions. P-values presented for the top 40% of clusters. **e,** CLIP-seq reads (replicate 1: blue, replicate 2: green) and clusters (purple) mapped to the 3'UTR of *CYP1B1*. Matches to the GUAG motif shown in black. **f and g,** Immunofluorescent images of HSP90 and CYP1B1 3 days after transduction and summary of fold-change in HSP90 and CYP1B1 protein and transcript levels with MSI2 OE at 3 and 7 days post transduction (scale bar 20 μm, dotted line indicates no change, n=3 experiments). **h,** HSPC marker expression by $CD34^+$ cells treated with TMS for 10 days. **i,** Absolute $CD34^+$ cell number with TMS (n=4 experiments). Data are presented as mean ± SEM. $^{**}p<0.01$.