Stem Cell Reports
Article

PLAG1 and USF2 Co-regulate Expression of Musashi-2 in Human Hematopoietic Stem and Progenitor Cells

Muluken S. Belew,[1] Sonam Bhatia,[1] Ava Keyvani Chahi,[1] Stefan Rentas,[1] Jonathan S. Draper,[1] and Kristin J. Hope[1,*]
[1]Department of Biochemistry and Biomedical Sciences, Stem Cell and Cancer Research Institute, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada
*Correspondence: kristin@mcmaster.ca
https://doi.org/10.1016/j.stemcr.2018.03.006

SUMMARY

MSI2, which is expressed predominantly in hematopoietic stem and progenitor cells (HSPCs), enforces HSPC expansion when overexpressed and is upregulated in myeloid leukemias, indicating its regulated transcription is critical to balanced self-renewal and leukemia restraint. Despite this, little is understood of the factors that enforce appropriate physiological levels of MSI2 in the blood system. Here, we define a promoter region that reports on endogenous expression of *MSI2* and identify USF2 and PLAG1 as transcription factors whose promoter binding drives reporter activity. We show that these factors co-regulate, and are required for, efficient transactivation of endogenous *MSI2*. Coincident overexpression of USF2 and PLAG1 in primitive cord blood cells enhanced *MSI2* transcription and yielded cellular phenotypes, including expansion of CD34+ cells *in vitro*, consistent with that achieved by direct MSI2 overexpression. Global chromatin immunoprecipitation sequencing analyses confirm a preferential co-binding of PLAG1 and USF2 at the promoter of *MSI2*, as well as regulatory regions corresponding to genes with roles in HSPC homeostasis. PLAG1 and USF2 cooperation is thus an important contributor to stem cell-specific expression of MSI2 and HSPC-specific transcriptional circuitry.

INTRODUCTION

The unique ability of hematopoietic stem cells (HSCs) to self-renew while giving rise to mature cells by differentiation enables their regenerative capacity. As such, identification and characterization of key regulators of this property are of both fundamental and clinical interest. The Musashi-2 (MSI2) RNA-binding protein is now recognized as one such key HSC regulator. Its expression is highest in the primitive HSC compartment and progressively decreases upon commitment (de Andres-Aguayo et al., 2011; Hope et al., 2010; Park et al., 2014; Rentas et al., 2016). MSI2 loss of function results in a significant depletion in the reconstitution capacity of primitive murine hematopoietic cells (de Andres-Aguayo et al., 2011; Hope et al., 2010; Ito et al., 2010; Park et al., 2014). By contrast, when moderately overexpressed, it imparts enhanced HSC self-renewal activity as measured by increased competitiveness in *in vivo* reconstitution assays (Hope et al., 2010). In the human system we have shown an analogous detrimental effect on cord blood (CB) HSC-mediated reconstitution when MSI2 is repressed. These same stem cells undergo significant *ex vivo* expansion when MSI2 is overexpressed (Rentas et al., 2016). MSI2 has also been implicated in aspects of leukemia pathogenesis (Kharas et al., 2010; Park et al., 2015; Ito et al., 2010). For instance, in mouse models of chronic myeloid leukemia (CML) and myelodysplastic syndrome (MDS), ectopic expression of MSI2 encourages promotion of the disease to acute phases (Kharas et al., 2010; Taggart et al., 2016). In the human

context, aberrantly high expression of MSI2 correlates with more aggressive CML disease states and is associated with poor prognosis in acute myeloid leukemia and MDS (Ito et al., 2010; Kharas et al., 2010; Taggart et al., 2016). Taken together, these studies suggest that the precise molecular regulation of MSI2 gene expression may be among the critical mechanisms underlying balanced HSC self-renewal/differentiation and the restraint of leukemia progression. Despite the importance of MSI2 in stem cell behavior, it remains poorly understood how *MSI2* expression is maintained at appropriate levels, and very little is known of the promoter elements or transcription factors (TFs) that mediate this. Here, we report an approach to address HSC cell fate control through the systematic dissection of the *MSI2* promoter functional in hematopoietic cells. Through this strategy, we have identified two TFs that function as cooperative regulators of *MSI2* and that together play a key role in HSPC function.

RESULTS

Dissection of the *MSI2* Minimal Promoter

MSI2 expression is evolutionarily conserved in both mouse and human HSPCs. Therefore, as an initial step in mapping its promoter we concentrated on the region directly upstream of the translational start site sharing extensive sequence similarity between the two species. This corresponded to a region extending to ~3.2 kb upstream wherein homology peaks were detected throughout as
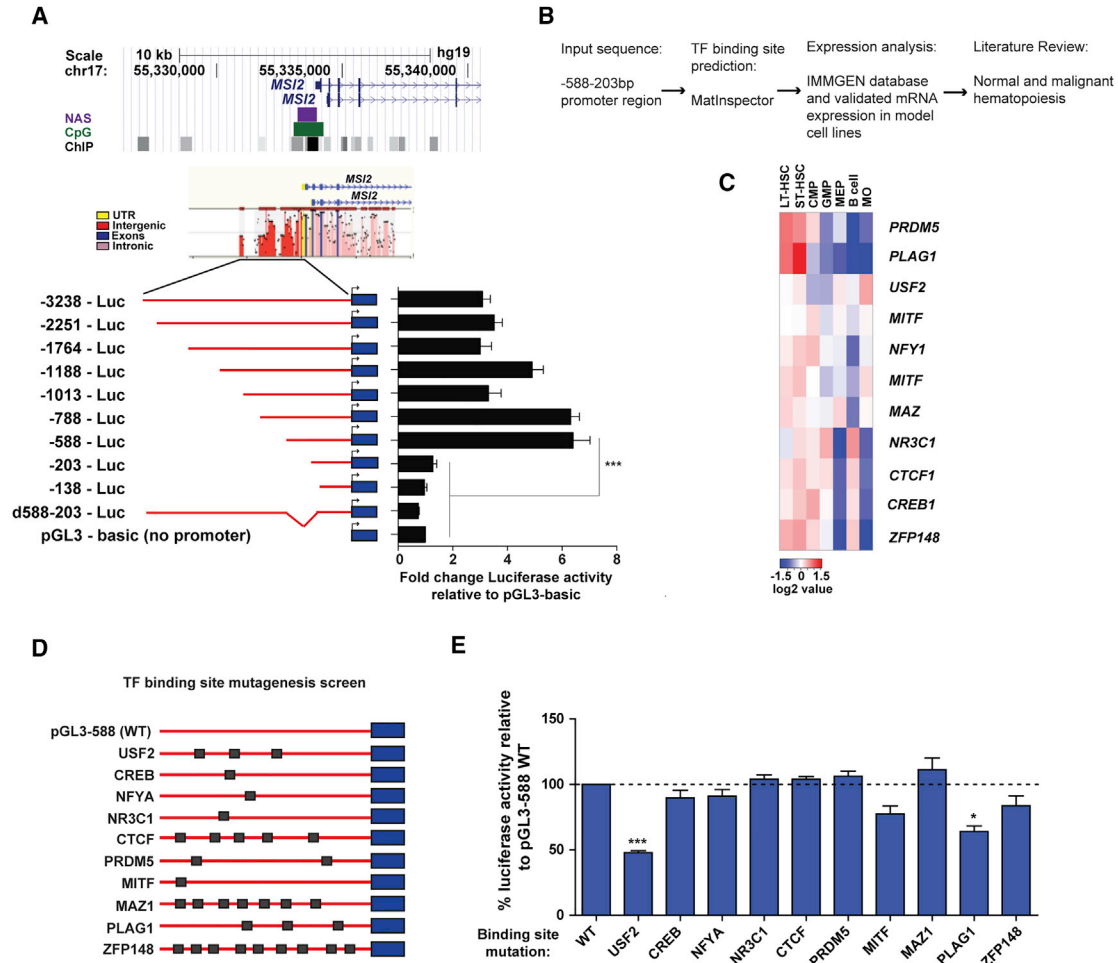
Figure 1. *In Silico* Mapping and Mutagenesis Screening Identifies the *MSI2* Promoter in Hematopoietic Cells with Dependence on USF2 and PLAG1 Binding Sites for Activity

(A) UCSC genome browser annotation of features within the region directly 5′ upstream of *MSI2* (top panel) including ChIP-validated transcription factor (TF) binding sites, a CpG island, and nuclease accessible site (NAS). Middle panel depicts *MSI2* genomic sequence alignment and homology between mouse and human species as analyzed by MULAN. Bottom panel shows a schematic representation of the serial 5′- promoter truncations (red) placed upstream of the firefly luciferase (Luc) reporter gene (blue) and their corresponding luciferase reporter activity.

(B) Workflow of TF selections for binding site mutagenesis screen.

(C) Heatmap demonstrating the relative expression across the hematopoietic hierarchy of a prioritized subset of TFs predicted *in silico* to bind the *MSI2* promoter.

(D) Schematic depicting the binding sites mutated for each of the ten candidate *MSI2*-regulating TFs in independent reporter constructs.

(E) Percentage changes in luciferase activity after specific TF binding site mutagenesis within the minimal promoter sequence compared with the wild-type (WT) promoter.

Data for all reporter assays was generated from n = 3–6 independent experiments. Data are presented as means ± SEM. *p < 0.05, ***p < 0.001. See also Tables S1 and S2.

identified by the multiple sequence local alignment and visualization tool (MULAN) (Ovcharenko et al., 2004) (Figure 1A, middle panel). Multiple sequence features including a nuclease accessible site (NAS), CpG island, and TF binding sites as identified by chromatin immuno-precipitation sequencing (ChIP-seq) within a conserved re-

gion ~1 kb upstream of the translational start site further suggested the potential for this region to function in a promoter capacity (Figure 1A). Introduction of this 3.2 kb region upstream of firefly luciferase in pGL3-basic yielded significantly greater reporter activity compared with the promoterless construct in MSI2-expressing K562 and
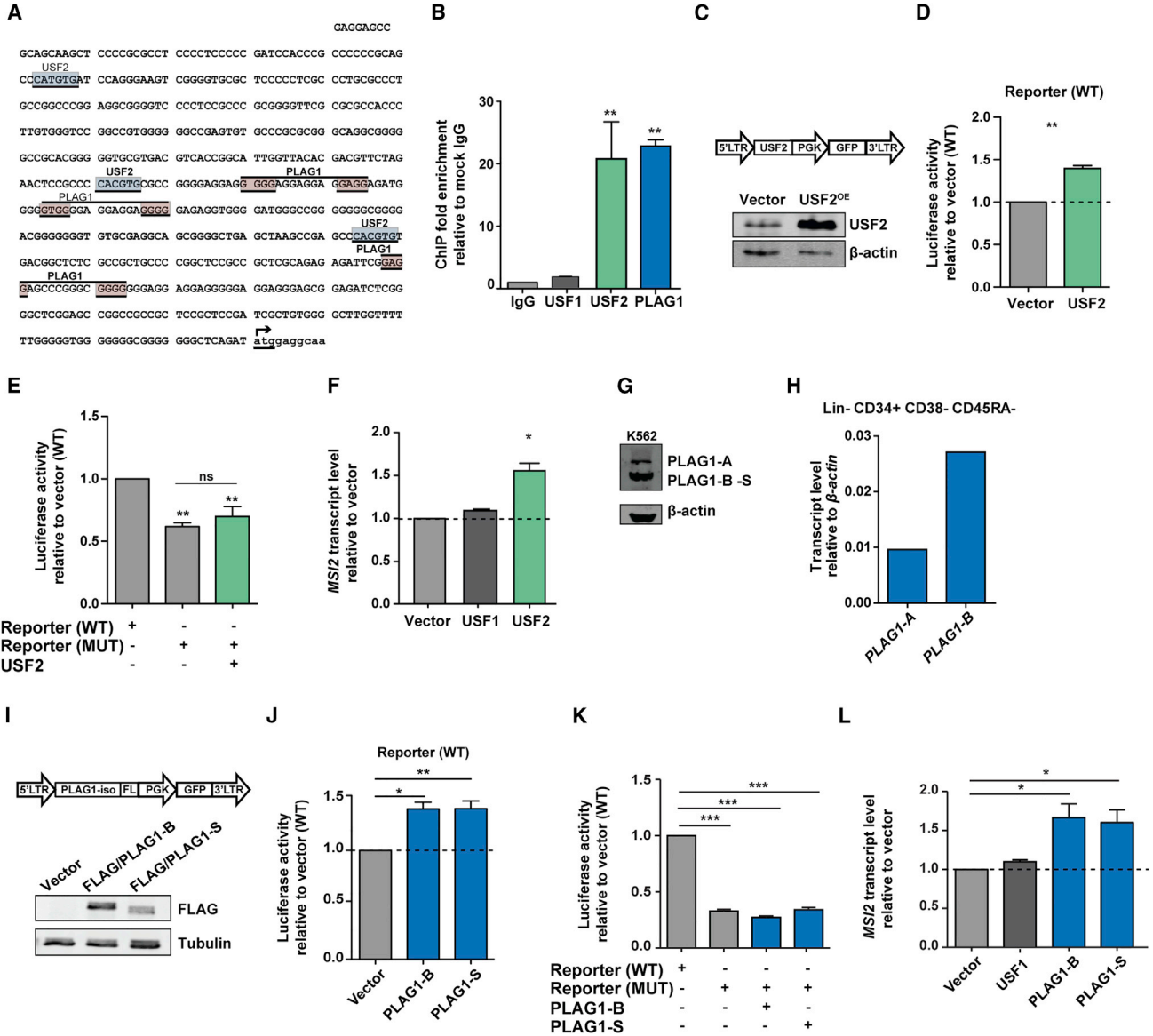
Figure 2. Ectopic PLAG1 and USF2 Bind the *MSI2* Promoter and Enhance Its Transcription in K562 Cells

(A) DNA sequence of the minimal promoter for *MSI2* with predicted (MatInspector) USF2 (blue) and PLAG1 (red) binding sites shown. The translation start site of *MSI2* is identified by the forward arrow from lower case nucleotides and the initiation codon is underlined.

(B) ChIP-qPCR for PLAG1 and USF2 binding of the minimal *MSI2* promoter (n = 4 experiments).

(C) USF2-overexpressing lentiviral construct schematic (top) and western blot quantification of its enhanced protein levels in K562 cells (bottom).

(D) Wild-type (WT) *MSI2* minimal promoter reporter construct luciferase activity upon USF2 overexpression (n = 3 experiments).

(E) Luciferase activity of USF2 binding site-mutated (MUT) reporter with and without USF2 overexpression compared with WT (n = 4 experiments).

(F) Endogenous *MSI2* transcript levels following USF2 overexpression in K562 cells (n = 3 experiments).

(G) Characterization of PLAG1 isoform expression by western blot in K562 cells.

(H) Comparison of PLAG1-A and PLAG1-B expression in flow-sorted CB HSPCs (performed once from pooled CB samples).

(I) FLAG-tagged PLAG1 short-isoform overexpressing lentiviral construct schematic and western blot quantification of enhanced protein levels in K562 cells (bottom).

(J) WT reporter construct luciferase activity upon PLAG1-B or PLAG1-S overexpression (n = 4 experiments).

(legend continued on next page)

HEK293 cell lines (3-fold and 7.5-fold respectively) (Figure 1A, data not shown). Using variations in the extent of homology peaks as endpoints, we generated a set of luciferase reporter constructs with serial 5′-truncations of the ~3.2 kb sequence. A significant drop in reporter activity resulted only when the upstream sequence driving reporter expression was reduced from −588 to −203 bp (Figure 1A). In confirmation that a minimal promoter region containing essential *cis* elements governing *MSI2* expression is contained within this 385 bp region we found its deletion from the full-length 3.2 kb fragment was sufficient to repress luciferase activity to the level of the promoterless reporter (Figure 1A).

USF2 and PLAG1 Binding Sites Are Required for *MSI2* Promoter Activity

We next implemented a mutagenesis screen to systematically test the functionality of TF consensus sites within the minimal promoter region in order to pinpoint key *MSI2*-regulators. First we used *in silico* TF binding site prediction (MatInspector) to identify a total of 107 TF candidates representing 65 different TF families. From within this set we removed those TFs not appreciably expressed in both murine and human HSPCs, where *MSI2* expression is known to be concentrated. A final refinement based on an extensive literature search to identify novel TFs, or those implicated in the regulation of stem cell function and/or leukemia, yielded a list of ten candidate regulators of primitive hematopoietic cell expression of *MSI2* (Figures 1B and 1C). Using the −588-Luc reporter construct as a template we generated a series of constructs possessing mutations of the consensus binding sites of each of these TFs and screened for loss of luciferase reporter activity (Figures 1D and 1E). The only TFs whose binding site mutagenesis resulted in a significant reduction in reporter activity were the E-box motif-binding USF2 and the C_2H_2 zinc finger protein PLAG1 (Figure 1E). The results of our rationalized screen thus indicate that intact PLAG1 and USF2 binding sites are required for full *MSI2* promoter activity.

USF2 and PLAG1 Bind the Promoter of *MSI2* and Promote Its Transcription

The minimal *MSI2* promoter sequence contains a total of three consensus E-box motifs for USF2 (Bendall and Molloy, 1994) and three G-rich bipartite consensus sequences for PLAG1 (Hensen et al., 2002) (Figure 2A). In support of the functionality of these binding sites, ChIP-qPCR

demonstrated that PLAG1 and USF2 yield significant fold enrichment at the minimal promoter region of *MSI2* in K562 cells (Figure 2B). Combined with the mutagenesis data, these results are consistent with the direct binding of USF2 and PLAG1 to their consensus recognition elements as requirements for full activity of the *MSI2* minimal promoter. To quantify the transactivation potential of USF2 and/or PLAG1 on the promoter region we independently overexpressed each gene. USF2-overexpressing (USF2) cells exhibited a significant 40% (1.4-fold) enhanced reporter activity compared with vector controls (Figures 2C and 2D). In addition, USF2-overexpressing cells did not exhibit any elevation of luciferase reporter levels when transfected with a promoter construct that had mutations in all three USF2 binding sites, further confirming the specificity of USF2 binding to the predicted sites (Figure 2E). Lastly, USF2 overexpression resulted in a 1.4-fold increased *MSI2* transcript level compared with negative controls. Binding to the E-box motif within the *MSI2* promoter and transactivation of *MSI2* expression is specific to USF2 as its closely related family member, USF1, does not bind the *MSI2* promoter (Figure 2B) and yielded no increase in *MSI2* expression when overexpressed (Figure 2F). Together, these results validate the reporter results and confirm the direct connection of USF2 levels with endogenous *MSI2* expression.

There are three human PLAG1 transcript variants encoding two protein isoforms. The larger isoform (PLAG1-A) contains all annotated functional domains, whereas the shorter form (PLAG1-B) is missing Zn fingers F1 and F2. A third reported slightly smaller PLAG1 translational isoform exists through translation from methionine 100 (PLAG1-S), is missing the first 17 N-terminal residues of PLAG1-B, and thus co-migrates with it on SDS-PAGE (Debiec-Rychter et al., 2001) (Figures 2G and S1). In both K562 and primary primitive human CB cells, PLAG1 appears to be contributed largely by the short-form(s) (Figures 2G and 2H). Thus to characterize their respective contributions to *MSI2* transactivation we overexpressed PLAG1-B and PLAG1-S or the negative control USF1 in K562 cells (Figure 2I). Both PLAG1 forms independently enhanced reporter activity ~1.4 fold in a PLAG1-binding-site-dependent manner (Figures 2J and 2K) and, most importantly, resulted in a significant 1.6-fold increase in endogenous *MSI2* transcription, suggesting they behave interchangeably with respect to *MSI2* regulation (Figure 2L). Together, these findings indicate that both USF2 and PLAG1 directly

(K) Luciferase activity of PLAG1-B and PLAG1-S MUT reporters with PLAG1-B or PLAG1-S overexpression compared with WT (n = 4 experiments).

(L) Endogenous *MSI2* transcript levels following PLAG1-B or PLAG1-S overexpression in K562 cells (n = 3 experiments).

Data are presented as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001. See also Figure S1 and Table S3.
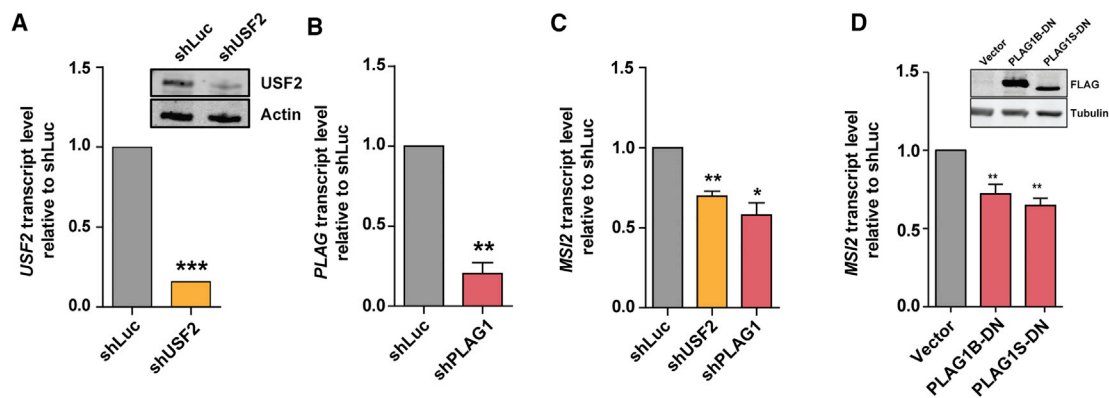
Figure 3. PLAG1 and USF2 Are Required for *MSI2* Promoter Activity and Expression
(A and B). Knockdown efficiencies of shUSF2 (E) and shPLAG1 (F) as measured post-infection by q-RT-PCR and western blot where sufficient cells were available (inset, A) in K562 cells (n = 3 experiments for each hairpin).
(C) *MSI2* transcript level reductions following *USF2* and *PLAG1* knockdown (n = 3 experiments).
(D) Western blot (inset) validation of overexpression of the dominant-negative forms of PLAG1-B (PLAG1B-DN) and PLAG1-S (PLAG1S-DN), and graph of *MSI2* transcript levels following introduction of PLAG1B-DN or PLAG1S-DN into K562 cells. (n = 3 experiments).
Data are presented as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001. See also Table S3.
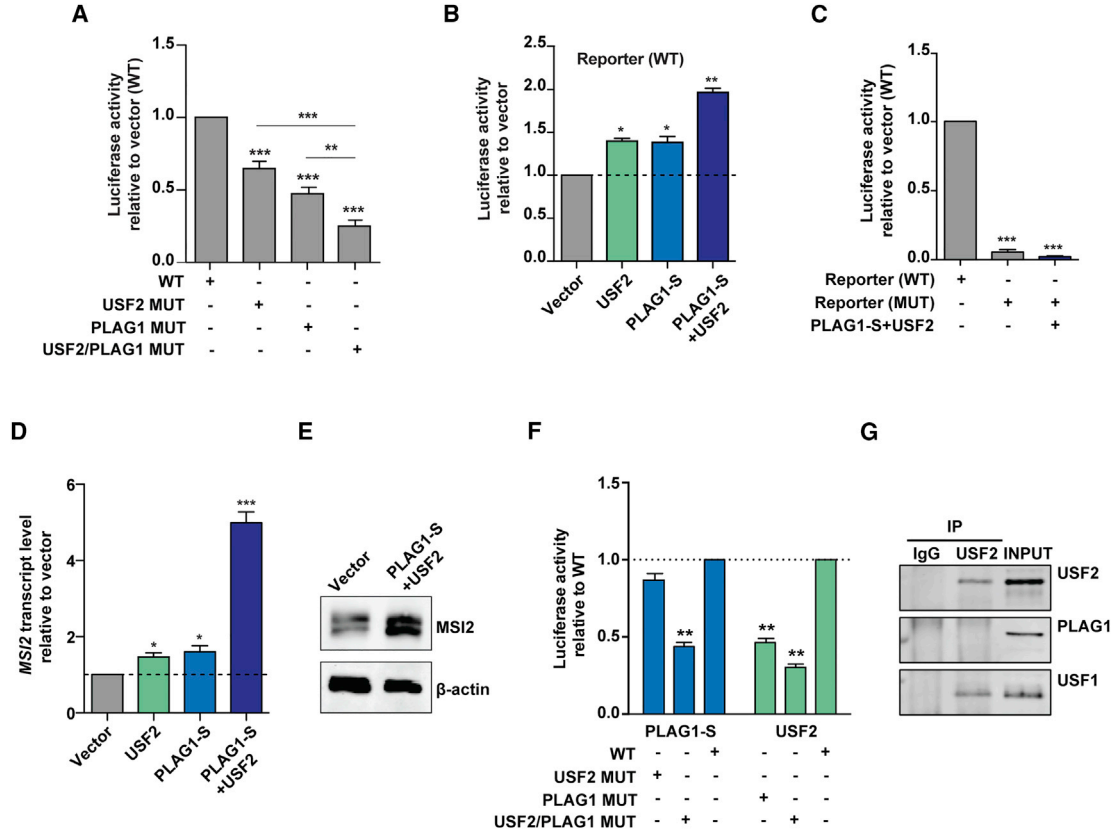
interact with the *MSI2* minimal promoter and are capable of ectopically transactivating its expression.

Reciprocal knockdown experiments with short hairpin RNAs (shRNAs) targeting *USF2* and *PLAG1* resulted in a 40% and 50% decrease in endogenous *MSI2* transcription, respectively, in comparison with control cells expressing non-targeting shRNA (shLuc) (Figures 3A–3C). In addition, we measured the effects of PLAG1 loss of function using a dominant-negative approach wherein we introduced short PLAG1 forms missing their transactivation domains (Kas et al., 1998; Voz et al., 2000) (PLAG1-B DN and PLAG1-S-DN, respectively) (Figure 3D). The introduction of either form led to the attenuation of *MSI2* transcription by approximately 1.4-fold compared with controls (Figure 3D). These results confirm that PLAG1 transactivation activity is a requirement for maintenance of physiological levels of *MSI2* transcription. In summary, these gain- and loss-of-function assays demonstrate that USF2 and PLAG1 constitute critical components of the regulatory circuitry controlling *MSI2* transcription and are essential for maintaining physiological MSI2 expression levels.

USF2 and PLAG1 Function Collaboratively to Promote *MSI2* Transactivation

We next tested whether Plag1 and USF2 cooperate to transactivate *MSI2* expression. First we generated a minimal promoter reporter assay containing mutations in both the PLAG1 and USF2 consensus sites. Combinatorial binding site mutagenesis yielded a more significant reduction in minimal promoter activity than when PLAG1 or USF2 binding sites were mutated in isolation (Figure 4A). As this suggested that USF2 and PLAG1 potentially cooperate

to regulate *MSI2* promoter activity we sought to further address this possibility using PLAG1-S and USF2 co-overexpression in K562 cells. Luciferase activity following co-overexpression was approximately 2-fold greater than in negative controls and over 1.5-fold greater than in cells overexpressing either USF2 or PLAG1-S alone (Figure 4B). In addition, the same assay performed in singly and doubly USF2- and PLAG1-S-overexpressing cells transfected with promoter reporter constructs with co-mutated USF2/PLAG1 binding sites demonstrated background level reporter activity, confirming the specificity of the predicted binding sites (data not shown and Figure 4C). Strikingly, co-overexpression increased endogenous *MSI2* transcription more than 4.5-fold in comparison with vector transfected control cells, an increase that was also manifested at the protein level (Figures 4D and 4E). This robust increase in *MSI2* transcription is more than 2.5-fold higher than that observed in cells overexpressing USF2 or PLAG1-S independently, indicating a potential synergistic response on wholesale *MSI2* transactivation when the two factors are coordinately elevated. To assess the nature of their collaboration on the minimal promoter of *MSI2* we performed reporter assays in single factor overexpressing cells carrying the promoter sequence containing mutations of the alternate factor's consensus sites. We find that when USF2 binding sites are mutated, transactivation by PLAG1 is unaffected, whereas when PLAG1 binding sites are mutated transactivation of the promoter by USF2 is attenuated by more than 2-fold (Figure 4F). This is intriguingly suggestive of a requisite role for PLAG1 in the recruitment of cooperative TFs. To further characterize the potential cooperation between PLAG1 and USF2 we

Figure 4. PLAG1-S and USF2 Collaboratively Regulate *MSI2* Promoter Activity and *MSI2* Expression

(A) Reporter activity upon combinatorial binding site mutagenesis for USF2 and PLAG1 within the *MSI2* minimal promoter region. n = 11 independent experiments.

(B) Reporter construct luciferase activity upon independent or coincident overexpression of USF2 and PLAG1-S (n = 3 experiments).

(C) Luciferase activity of PLAG1 and USF2 MUT reporters upon PLAG1-S and USF2 co-overexpression (n = 3 experiments).

(D and E) Endogenous *MSI2* transcript (D) (n = 3 experiments) and protein levels (E) following overexpression of PLAG1-S and USF2 in K562 cells compared with vector control.

(F) Luciferase reporter activity of the *MSI2* minimal promoter region upon PLAG1-S or USF2 overexpression when the binding site for the reciprocal factor or for both factors is mutated (MUT) (n = 3 experiments).

(G) Co-immunoprecipitation analysis performed on K562 cells using anti-USF2 for immunoprecipitation and immunoblotting with antibodies to USF2, the known USF2 interacting partner USF1 or PLAG1.

Data are presented as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001. See also Table S3.

used co-immunoprecipitation (coIP) to test whether the two TFs are involved in a protein-protein interaction and found that they are not engaged in an interaction detectable by coIP (Figure 4G). Thus, although PLAG1 and USF2 exhibit an epistatic relationship, these data suggest that PLAG1 and USF2 do so in the absence of forming a direct, stable physical interaction.

If the cooperative effect of USF2 and PLAG1 is indeed a key component in driving *MSI2* expression in primary hematopoietic cells, it would be expected that maximal *MSI2* expression would be observed in cells where the combined levels of these two factors peak. Indeed, this is the case, as *USF2* is ubiquitously expressed across the human hemato-

poietic hierarchy, while *PLAG1* is at its highest levels in the most primitive HSC subset and declines rapidly in expression with differentiation (Figure 5A), a profile that is also consistent with *MSI2* expression (Rentas et al., 2016). Moreover, we confirmed PLAG1-B isoform expression follows the same pattern of being more highly expressed in HSC enriched CB populations (Figure 5B). To support that this maximal presence of PLAG1 in combination with USF2 in more primitive hematopoietic cells translates to an increase in actual *MSI2* promoter binding we performed ChIP to examine PLAG1 and USF2 binding to the −588 bp region of Lin-CD34+ HSPCs and Lin-CD34-restricted progenitors. ChIP-qPCR demonstrated that
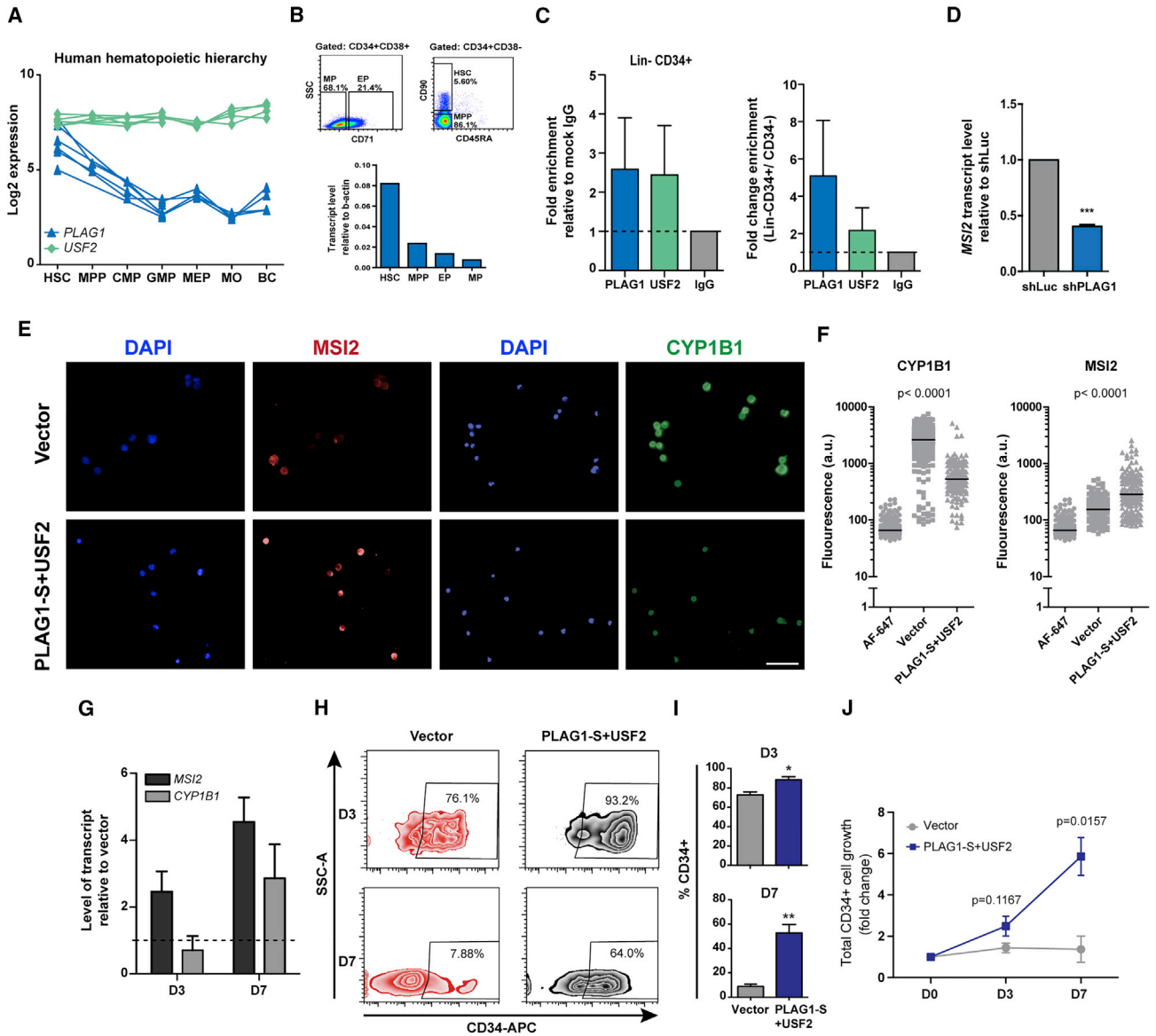
Figure 5. PLAG1-S and USF2 Co-regulate *MSI2* Expression and Its Downstream Functions in Primitive Hematopoietic Cells

(A) Expression profile of PLAG1 and USF2 in the human hematopoietic hierarchy. BC, band cell; CMP, common myeloid progenitor; GMP, granulocyte monocyte progenitor; MEP, megakaryocyte erythroid progenitor; MO, monocytes; MPP, multipotent progenitors.

(B) Flow cytometry gating strategy to isolate erythroid progenitors (EP), myeloid progenitors (MP), MPP and HSCs from cord blood (top) and the q-RT-PCR validated expression of PLAG1-B within these populations (bottom, performed once from pooled CB samples).

(C) ChIP-qPCR-defined fold enrichment of PLAG1 and USF2 at the *MSI2* promoter in purified cord blood Lin-CD34+ (left) and fold change ChIP enrichment in Lin-CD34+ compared with Lin-CD34- (right) (n = 3 experiments).

(D) qPCR determination of *MSI2* transcript levels following *PLAG1* knockdown in Lin-CD34+ cells (n = 3 experiments).

(E) Representative immunofluorescence images of MSI2 and CYP1B1 proteins in Lin-CD34+ CB cells expressing control vector or co-overexpressing PLAG1-S and USF2 and cultured for 7 days. Scale bar, 25 μm.

(F) Quantification of MSI2 and CYP1B1 expression on a per cell basis. Each point in the graph represents an individual stained cell and the solid line denotes the mean fluorescence intensity of each dataset. Cells stained with Alexa Fluor 647 (AF-647) conjugated secondary antibody alone were considered as background fluorescence controls (>100 cells analyzed per condition for each experiment, plots summarize n = 3 experiments).

(G) Fold change in the transcript level of *MSI2* and *CYP1B1* upon co-expression of PLAG1-S and USF2 at day 3 (D3) and day 7 (D7) of *in vitro* culture relative to vector control.

PLAG1 and USF2 do indeed directly bind to the *MSI2* promoter with more than 4.5- and 2.5-fold enrichment, respectively, in Lin-CD34+ compared with Lin-CD34-cells (Figure 5C). Furthermore, selective knockdown of PLAG1 in Lin-CD34+ cells results in a 60% down-regulation of *MSI2* transcript validating the PLAG1-dependent nature of MSI2 expression in primitive hematopoietic cells (Figure 5D).

Co-overexpression of USF2 and PLAG1 Phenocopies Expression of Ectopic MSI2 and Enhances Primitive Cell Output

We have previously reported that MSI2 enforces HSPC expansion *ex vivo* in part through the post-transcriptional repression of the aryl hydrocarbon receptor (AHR) effector CYP1B1 (Rentas et al., 2016). As would be expected if a major effect of their cooperative function is MSI2 upregulation, co-ordinate overexpression of USF2 and PLAG1-S in Lin-CD34+ cells resulted not only in a significant 4.5-fold elevation in *MSI2* transcription at day 7 of *ex vivo* culture (D7) but also in the selective repression of CYP1B1 protein (and not its mRNA levels) (Figures 5E–5G). Most importantly, relative to controls, significant 5- and 6-fold enhancements in the proportion and total numbers of primitive CD34+ cells, respectively, were observed following a 7-day culture of CB cells jointly overexpressing USF2 and PLAG1-S (Figures 5H–5J). Together, these results are consistent with USF2 and PLAG1-S acting in combination to enforce MSI2 expression, thereby promoting primitive hematopoietic cell maintenance.

Genome-wide Identification of USF2 and PLAG1 Co-occupancy Identifies HSC Regulators as a Subset of their Common Targets

To explore the potential for cooperativity of USF2 and PLAG1 in binding additional regulatory elements across the genome we performed cross-linking immunoprecipitation followed by massively parallel next-generation sequencing (ChIP-seq) on two different clones (#4 and #5) of K562 cells co-overexpressing USF2 and FLAG-PLAG1-S (Figure S2). The chromatin was sheared to <500 bp fragments, the efficiency of pull-down was validated by ChIP followed by a western blot experiment and subsequent ChIP-seq fingerprint plot analysis (Figures S2A–S2D) (Ramirez et al., 2016). ChIP-seq peak calling revealed that USF2 and PLAG1 occupied a total of 2,409 and 21,469 sites, respectively (Figure S2E). To evaluate the likely biological impact, we assayed for co-binding within a stringent 100 bp window and found a statistically significant association of USF2 binding with PLAG1 sites as the majority (56%) of USF2 regions (1,386 sites) were accompanied by PLAG1 binding (Figure 6A, p value 2.2×10^{-16} by Fisher's exact test; Table S1A). We noted that USF2-PLAG1 (PLAG1 + USF2) co-occupancy typically occurred at promoter regions, with ~70% of peaks detected within 3 kb of a transcription start site (TSS) (Figure 6B). These peaks also show a strong distribution centered at the TSS supporting the role of USF2 and PLAG1 as TFs at these locations (Figure 6C). We next performed *de novo* motif discovery on sites bound by USF2 or PLAG1. Analysis of the USF2 bound sites uncovered the expected E-box motif for USF2 as the most significantly enriched, as well as a second, less frequent G-rich motif. The motif discovery for PLAG1 revealed not only the G-rich core consensus binding motif recognized by PLAG family members (Voz et al., 1998) but also identified an E-box containing consensus sequence bound with an equivalent frequency to the PLAG1 G-rich core consensus (Figure 6D). The motif distribution profile of the USF2 consensus at the center of PLAG1 peaks (Figure 6E) provide further support for the concept that PLAG1 can associate with regions of chromatin to which USF2 would be expected to show an affinity and echoes the preferential co-localization of these two factors at selected genomic loci as quantified above.

We next sought to explore whether USF2 and PLAG1 regulate HSC-specific genes and/or pathways. Using gene set enrichment analysis (GSEA) we created ranked lists with publicly available gene expression profiles assembled from sub-fractions of the human hematopoietic hierarchy ranging from primitive (Lin-CD34+) to mature (Lin+) (Beck et al., 2013) and then used the 621 genes associated with the most high-confidence USF2-PLAG1 co-bound promoters (peak scores >200) as the gene set. Interestingly, we found no significant enrichment of HSC-specific gene promoters being bound by the two factors (Figure S2F). Furthermore, the USF2-PLAG1 co-bound sites did not show a significant enrichment of a separate HSC stem/progenitor gene set (Ivanova et al., 2002) (Figure S2F). Despite this, however, we did observe promoter co-occupancy for a number of recognized HSC regulators, including *DNMT3A*, *SOCS2*, *HOXA1*, and *MSI2* (Figure 6F, Table S1A). Strikingly, MSI2 is the only factor in this list of promoters co-bound by USF2 and PLAG1-S that is

(H) Representative flow cytometry plots showing the frequency of CD34+ cells at the D3 and D7 *in vitro* culture points with the co-overexpression of PLAG1-S and USF2.
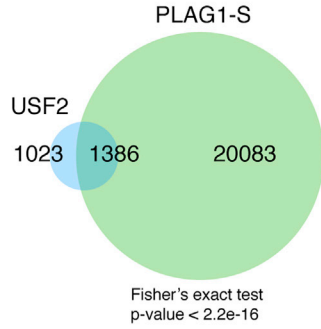(I) Quantification of CD34+ cell frequency (n = 3 experiments).
(J) Fold change in total CD34+ cells over the 7-day culture period.
Data are presented as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001. See also Table S3.

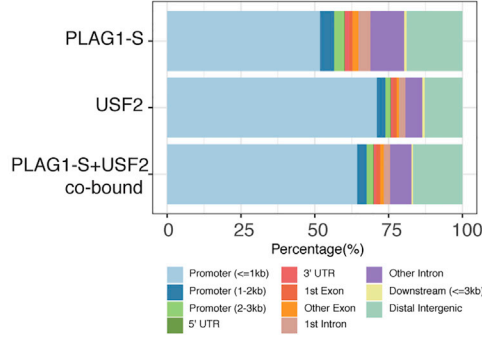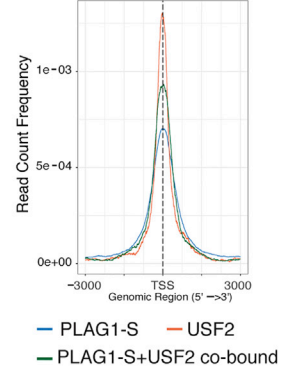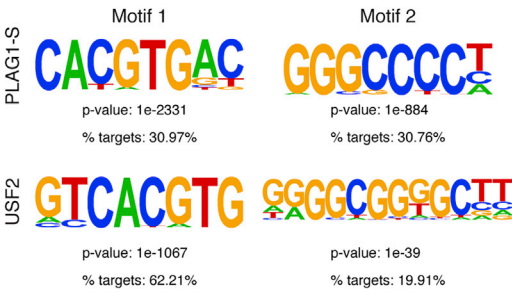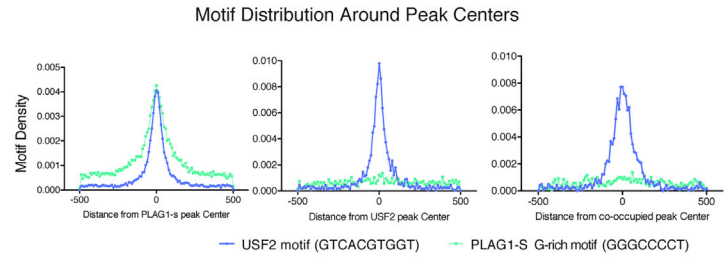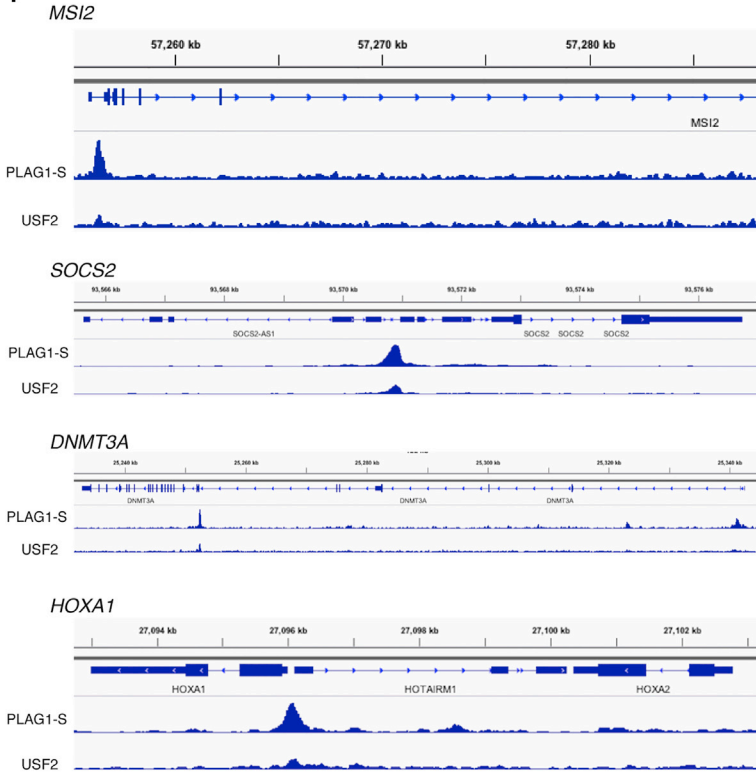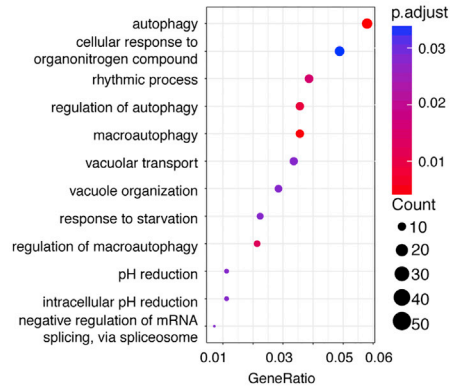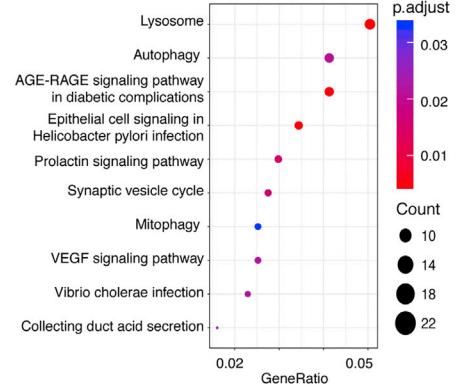(legend on next page)

documented as capable of expanding human HSC, and the identification of *MSI2* in this independent ChIP-seq experiment provides key support of our earlier findings in this study. Interestingly, upon Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of those genes that show co-occupancy, there was an abundance of genes involved in autophagy, a pathway important for stem cell control and homeostasis in the blood system (Ho et al., 2017; Mortensen et al., 2011) (Figures 6G and 6H, Tables S1B and S1C). Finally, performing unbiased GSEA we also find that USF2-PLAG1 co-occupied genes are positively enriched for those involved in cell cycle, membrane transport, RNA processing, and infectious diseases (Table S1D). Together these results suggest that a subset of PLAG1-USF2 bound sites are in fact localized to promoter regions with relevance to stem cell regulation as well as other genes whose involvement in the function of these two factors will now be of interest to explore further. Thus, we show that in the HSC context the co-ordinate regulation by USF2 and PLAG1 of these genes, in conjunction with MSI2, provides complementary mechanisms that contribute to the maintenance of stemness features.

DISCUSSION

To date, we have little knowledge of the endogenous regulators of *MSI2* transcription in the human HSPC compartment. Given the powerful role MSI2 plays in promoting HSPC self-renewal, identifying *MSI2* activators and their mechanism of action is required before their regenerative potential can be fully harnessed. Here, we have identified USF2 and PLAG1 as being both necessary, and, together, sufficient for significant transcriptional activation of *MSI2*. This discovery advances the understanding of how stem cell fate is propagated via the master regulator MSI2. In addition to describing a role for USF2 and PLAG1, our study provides key insights into the nature of their regula-

tory action. We observed that PLAG1 and USF2 appear to function additively on the *MSI2* promoter, but synergistically on actual *MSI2* transcription, which implies additional regulatory mechanisms act on the state of native chromatin through which these factors associate to further elevate transactivation. Indeed, our findings indicate that the competency of PLAG1 binding is dominant over that of USF2 at the *MSI2* promoter. Of note, our ChIP-seq analysis identified the G-rich consensus sequence previously described for PLAG1 (Voz et al., 1998) but also reveals that PLAG1 occupancy is frequently found at sites containing or surrounding an E-box motif that is highly similar to that bound by USF2. These data offer the possibility that PLAG1 could be an important nucleator for subsequent USF2 binding, the precise mechanism of which will now require more in-depth exploration. To this point, the specific sequence, occupancy status, and, by extension, shape of genomic regions proximal to E-boxes are known to modulate USF binding specificity (Gordan et al., 2013). Thus, the possibility that PLAG1 binding at or proximal to E-box motifs establishes a local structural configuration that is favorable to USF2 binding presents an attractive subject for subsequent exploration. The feasibility of PLAG1 acting to facilitate a permissive local environment for USF2 binding is supported by recent work on the related PLAG family member PLAGL1, which has highlighted the potential for PLAGL1 to display aspects of "pioneer factor" behavior (Varrault et al., 2017).

Our attempts to measure a direct interaction between PLAG1 and USF2 by coIP did not detect such a relationship, but it is important to point out that a negative coIP result does not preclude PLAG1 and USF2 having direct physical interaction or indirect cooperativity. Indeed, transcriptional control mechanisms are often complex, involving myriad co-factor interactions, and occur in the context of highly structured chromatin and often with dynamic kinetics (Morgunova and Taipale, 2017). As such, dissection of the physical and mechanistic relationship between PLAG1 and USF2 at regions of co-regulation is beyond

Figure 6. ChIP-Seq Mapping of USF2 and PLAG1-S Binding Identifies Co-occupancy at Promoters of *MSI2* and Other HSC Regulators
(A) Venn diagram showing the overlap of USF2 and PLAG1-S ChIP peaks within a 100 bp distance.
(B) Peak distribution profiles of all PLAG1-S and USF2 peaks, along with the co-occupied sites (PLAG1-S + USF2).
(C) Read distribution profiles of peaks found within ±3 kb of the TSS.
(D) *De novo* motif discovery analysis of the PLAG1-S and USF2 peaks. Top two most significantly enriched motifs identified by HOMER are shown for each factor.
(E) Motif density distribution profiles around peak centers for the USF2 consensus and the G-rich motif identified as the PLAG1-S consensus (motif 2) during *de novo* motif discovery.
(F) Integrative genome browser tracks showing the ChIP signal at regulatory regions of *MSI2*, *SOCS2*, *DNMT3A*, and *HOXA1*.
(G and H) Enrichment plots showing the biological process (G) and KEGG (H) pathways enriched in PLAG1-S + USF2 co-occupied sites using ChIPseeker R package. Counts represent the number of genes identified in each cluster and p.adjust are the adjusted p values using BH (Benjamini-Hochberg) method. GO, Gene Ontology.
See also Figure S2 and Tables S4, S5, S6, and S7.

the scope of our study but presents a priority topic for future study. However, our identification of PLAG1 and USF2 as key mediators of *MSI2* regulation improves our understanding of the upstream mechanism that controls HSPC maintenance and expansion.

Our work also supplements the current understanding on the nature of USF2 action in HSPCs. USF2 is known to regulate a number of diverse biological processes but has also previously been shown to regulate HOXB4 expression in human hematopoietic cells, via association with the primitive cell enriched-TF NF-Ya (Giannola et al., 2000; Zhu et al., 2003). We found that USF2, when co-expressed with PLAG1, is bound to the promoter of *HOXA1*, which has been shown to promote self-renewal of murine HSPCs (Bach et al., 2010; Fabb et al., 2004). Thus, in addition to hinting that USF2 may have a larger potential to function in combination with other, more HSC-restricted factors in the regulation of stemness genes, these findings also suggest that USF2 may play a larger role in HOX gene regulation.

PLAG1 was initially discovered in human pleomorphic adenomas of the salivary gland as a proto-oncogene (Debiec-Rychter et al., 2001; Kas et al., 1997; Voz et al., 1998). Here, we have expanded the repertoire of PLAG1 functionality into the maintenance of HSPCs. We have shown that it is the selective expression of PLAG1 within the most primitive cells of the blood system (Gazit et al., 2013) that enables it to function in co-ordination with the ubiquitously expressed USF2, ensuring appropriate elevated expression of *MSI2* in these cells. Our findings position PLAG1 as a key transcriptional regulator of HSPC in its own right. We have shown that USF2 and PLAG1 work cooperatively, but the large pool of binding sites for PLAG1 that do not overlap with USF2 may regulate genomic targets beyond which we can study here. In combination with the markedly enriched expression of PLAG1 in HSCs, this may hint at a larger USF2-independent role for PLAG1 in the control of HSC function, a topic that will now be important to explore in future studies.

EXPERIMENTAL PROCEDURES

MSI2 Promoter Cloning and Screening for Associated TFs

A 3,238 bp 5′-flanking sequence upstream of the translation start site of human *MSI2* (transcript variants 1 and 4) was PCR amplified from BAC clone RP11-784M23 (Roswell Park Cancer Institute) and cloned upstream of firefly luciferase in pGL3-basic (Promega). The transcriptional start site is predicted to be 173 nucleotides upstream of the translational start site. 5′-Serial truncations were generated using primers listed in Table S2A. TFs predicted by Mat-Inspector (Quandt et al., 1995) to bind the human 385 bp minimal

MSI2 promoter were retained for comparison against publicly available databases of global gene expression profiles of diverse hematopoietic subpopulations to facilitate a final selection of genes that are expressed in primitive mouse and human HSCs (Bagger et al., 2016; Heng and Painter, 2008; Jojic et al., 2013; Novershtern et al., 2011). For subsequent promoter mutagenesis screening we prioritized TFs that (1) had commercially available antibodies, (2) are expressed in K562s, (3) are known to be expressed in other stem cell types and/or to regulate their homeostasis, and (4) exhibit binding sites that are conserved in the *MSI2* promoter region shared between human and mouse. Predicted TF binding sites were mutated by rolling circle site-directed mutagenesis (Zheng et al., 2004). Primers and nucleotide substitutions are provided in Table S2B. Heatmap depicting putative *MSI2* regulating TFs and their expression in the mouse hematopoietic hierarchy was generated using IMMGEN's (Immunological Genome Project) on-line My GeneSet tool.

Promoter Clone Transfections and Reporter Assays

Five-hundred nanograms of each clone were transiently co-transfected with 25 ng of pRL-TK (internal transfection control expressing *Renilla* luciferase) in K562 cells in triplicate. Luminescence was measured 24–36 hr after transfection using Dual-Luciferase-Assay Kit (Promega) in a FLUOstar Omega luminometer (BMG Labtech). Firefly luciferase activity was normalized to *Renilla* luciferase activity to adjust for transfection efficiency.

Western Blots

Total protein was extracted from K562 cells using lysis buffer (50 mM Tris-HCl [pH 8], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, EDTA 2 mM) supplemented with protease inhibitor cocktail (Roche diagnostics). Lysate was spun at 14,000 rpm for 20 min on a refrigerated tabletop centrifuge. Total protein concentration was determined by Bradford assay using BSA as standard (Bio-Rad). Ten micrograms of total protein was resolved on 10% Bis-Tris PAGE. Resolved protein was transferred onto an immobilon-P polyvinylidene fluoride (PVDF) membrane (EMD Millipore) on transfer buffer (25 mM Tris, 250 mM glycine, and 15% methanol) using Trans-blot Turbo apparatus (Bio-Rad). Membrane was blocked for 1 hr using LI-COR blocking buffer (LI-COR Biosciences). Primary antibodies against MSI2 (#ab7614, Abcam) β-actin (catalog no. A5441 clone AC15, Sigma-Aldrich), were diluted in LI-COR blocking buffer (1:1000) and incubated for 1 hr at room temperature, washed three times at 15 min intervals by using 1× TBST buffer (50 mM Tris, 150 mM NaCl, 0.05% Tween 20 [pH 7.6]). IRDye 680 Goat Anti-Rabbit (catalog no. 926–32221, LI-COR Biosciences) and IRDye 800CW Goat anti-Mouse (catalog no. 926–32210, LI-COR Biosciences) secondary antibodies were diluted (1:15,000) in LI-COR blocking buffer and incubated for 1 hr at room temperature. The membrane was washed three times at 15 min interval using 1× TBST buffer. Blots were imaged using LI-COR Odyssey imaging station (LI-COR Biosciences).

qRT-PCR

Total RNA was isolated using Trizol-LS reagent (Life Technologies). cDNA synthesis and qPCR were done using the same protocol as Rentas et al. (2016) using Roche UPL primer and probe sets specific

to each of *PLAG1*, *β-ACTIN*, *MSI2*, and *CYP1B1* (Table S2C). *PLAG1* isoform-specific qPCR assays were purchased from IDT (PrimeTime Mini qPCR Assay, PLAG1-A assay ID, Hs.PT.58.26984634; PLAG1-B assay ID, Hs.PT.58.22864822). Fold change in transcript level was calculated according to the $2^{-\Delta\Delta Ct}$ method.

Generation of PLAG1 and USF2 Independent Overexpression

Open reading frame cDNA clones for *PLAG1* (Clone Id: 30915361) and *USF2* (Clone Id: 5483523) were obtained from Mammalian Gene Collections (GE Healthcare). cDNAs were PCR amplified using PrimeSTAR Max enzyme and subcloned into EcoRI/XhoI (NEB) digested MSCV-PGK/GFP vector. Primers used were: USF2-sense, CAT GCA TGG AAT TCG CCA CCA TGG ACA TGC TGG ACC, USF2-antisense, CAT GCA TGC TCG AGT CAC TGC CGG GTG CC (EcoRI and XhoI sites are underlined). To generate 3xFLAG epitope-tagged versions of the predominantly expressed variants of PLAG1-B and PLAG1-S, the respective PCR products were generated and cloned into the shuttle vector p3xFLAG-CMV10 (Sigma-Aldrich) at EcoRI/KpnI (NEB) sites. FLAG-tagged clones were PCR amplified and subcloned into MSCV-PGK/GFP vector at EcoRI/XhoI sites using MfeI/SalI (NEB) compatible end ligation. Primers used were: FLAG-PLAG1-sense, GAT CGA TCC AAT TGA TGG ACT ACA AAG ACC, FLAG-PLAG1-antisense, CAT GCA TGG TCG ACC TAC TGA AAA GCT TGA (MfeI and SalI sites are underlined).

Generation of PLAG1 and USF2 Overexpression and Knockdown Constructs

PLAG1 isoforms and *USF2* were cloned into the bidirectional promoter pSMALB vector where the transgene is driven in one direction and mtagBFP (blue fluorescent protein [BFP], as the transduction marker) in the other (van Galen et al., 2014). For co-overexpression, 3xFLAG-*PLAG1-100* and *USF2* were cloned consecutively into pSMALB and separated by a P2A site so that BFP positivity marks cells co-overexpressing both transgenes. shRNAs against *PLAG1* and *USF2* were designed using the RNAi Central tool and ligated downstream of the H1 promoter in the modified cppt-PGK-EGFP-IRES-PAC-WPRE lentiviral expression vector (Hope et al., 2010). Lentivirus production and transduction of K562 or primary Lin- CB cells was done as previously reported (Rentas et al., 2016). Stable lines were propagated with biological replicate lines created for every condition from three or more distinct transductions.

For details of ChIP and all other procedures, see Supplemental Experimental Procedures.

Isolation of Primitive Human Hematopoietic Cells and Flow Cytometry

All patient samples were obtained with informed consent and with the approval of the local human subject research ethics board at McMaster University.

ACCESSION NUMBERS

The accession number for the ChIP-seq data reported in this paper is GEO: GSE111469.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and seven tables and can be found with this article online at https://doi.org/10.1016/j.stemcr.2018.03.006.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Bach, C., Buhl, S., Mueller, D., Garcia-Cuellar, M.P., Maethner, E., and Slany, R.K. (2010). Leukemogenic transformation by HOXA cluster genes. Blood *115*, 2910–2918.

Bagger, F.O., Sasivarevic, D., Sohi, S.H., Laursen, L.G., Pundhir, S., Sonderby, C.K., Winther, O., Rapin, N., and Porse, B.T. (2016). BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. Nucleic Acids Res. *44*, D917–D924.

Beck, D., Thoms, J.A., Perera, D., Schutte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Gottgens, B., et al. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood *122*, e12–e22.

Bendall, A.J., and Molloy, P.L. (1994). Base preferences for DNA binding by the bHLH-Zip protein USF: effects of MgCl2 on specificity and comparison with binding of Myc family members. Nucleic Acids Res. *22*, 2801–2810.

de Andres-Aguayo, L., Varas, F., Kallin, E.M., Infante, J.F., Wurst, W., Floss, T., and Graf, T. (2011). Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood *118*, 554–564.

Debiec-Rychter, M., Van Valckenborgh, I., Van den Broeck, C., Hagemeijer, A., Van de Ven, W.J., Kas, K., Van Damme, B., and Voz, M.L. (2001). Histologic localization of PLAG1 (pleomorphic adenoma gene 1) in pleomorphic adenoma of the salivary gland: cytogenetic evidence of common origin of phenotypically diverse cells. Lab. Invest. *81*, 1289–1297.

Fabb, S.A., Haines, G., Dworkin, S., Simmons, P.J., Gudas, L.J., and Purton, L.E. (2004). Novel roles for the homeobox gene, HoxA1, and its truncated form, HoxA1-T, in the regulation of hematopoietic stem cell self-renewal. Blood *104*, 1706.

Gazit, R., Garrison, B.S., Rao, T.N., Shay, T., Costello, J., Ericson, J., Kim, F., Collins, J.J., Regev, A., Wagers, A.J., et al. (2013). Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. Stem Cell Reports *1*, 266–280.

Giannola, D.M., Shlomchik, W.D., Jegathesan, M., Liebowitz, D., Abrams, C.S., Kadesch, T., Dancis, A., and Emerson, S.G. (2000). Hematopoietic expression of HOXB4 is regulated in normal and leukemic stem cells through transcriptional activation of the HOXB4 promoter by upstream stimulating factor (USF)-1 and USF-2. J. Exp. Med. *192*, 1479–1490.

Gordan, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. Cell Rep. *3*, 1093–1104.

Heng, T.S., and Painter, M.W. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. *9*, 1091–1094.

Hensen, K., Van Valckenborgh, I.C., Kas, K., Van de Ven, W.J., and Voz, M.L. (2002). The tumorigenic diversity of the three PLAG family members is associated with different DNA binding capacities. Cancer Res. *62*, 1510–1517.

Ho, T.T., Warr, M.R., Adelman, E.R., Lansinger, O.M., Flach, J., Verovskaya, E.V., Figueroa, M.E., and Passegue, E. (2017). Autophagy maintains the metabolism and function of young and old stem cells. Nature *543*, 205–210.

Hope, K.J., Cellot, S., Ting, S.B., MacRae, T., Mayotte, N., Iscove, N.N., and Sauvageau, G. (2010). An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell *7*, 101–113.

Ito, T., Kwon, H.Y., Zimdahl, B., Congdon, K.L., Blum, J., Lento, W.E., Zhao, C., Lagoo, A., Gerrard, G., Foroni, L., et al. (2010). Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature *466*, 765–768.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., and Lemischka, I.R. (2002). A stem cell molecular signature. Science *298*, 601–604.

Jojic, V., Shay, T., Sylvia, K., Zuk, O., Sun, X., Kang, J., Regev, A., Koller, D., Immunological Genome Project, C., Best, A.J., et al. (2013). Identification of transcriptional regulators in the mouse immune system. Nat. Immunol. *14*, 633–643.

Kas, K., Voz, M.L., Hensen, K., Meyen, E., and Van de Ven, W.J. (1998). Transcriptional activation capacity of the novel PLAG family of zinc finger proteins. J. Biol. Chem. *273*, 23026–23032.

Kas, K., Voz, M.L., Roijer, E., Astrom, A.K., Meyen, E., Stenman, G., and Van de Ven, W.J. (1997). Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. Nat. Genet. *15*, 170–174.

Kharas, M.G., Lengner, C.J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., Morgan, K., Tam, W., Paktinat, M., Okabe, R., et al. (2010). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nat. Med. *16*, 903–908.

Morgunova, E., and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. Curr. Opin. Struct. Biol. *47*, 1–8.

Mortensen, M., Soilleux, E.J., Djordjevic, G., Tripp, R., Lutteropp, M., Sadighi-Akha, E., Stranks, A.J., Glanville, J., Knight, S., Jacobsen, S.E., et al. (2011). The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. J. Exp. Med. *208*, 455–467.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

Ovcharenko, I., Nobrega, M.A., Loots, G.G., and Stubbs, L. (2004). ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res. *32*, W280–W286.

Park, S.M., Deering, R.P., Lu, Y., Tivnan, P., Lianoglou, S., Al-Shahrour, F., Ebert, B.L., Hacohen, N., Leslie, C., Daley, G.Q., et al. (2014). Musashi-2 controls cell fate, lineage bias, and TGF-beta signaling in HSCs. J. Exp. Med. *211*, 71–87.

Park, S.M., Gonen, M., Vu, L., Minuesa, G., Tivnan, P., Barlowe, T.S., Taggart, J., Lu, Y., Deering, R.P., Hacohen, N., et al. (2015). Musashi2 sustains the mixed-lineage leukemia-driven stem cell regulatory program. J. Clin. Invest. *125*, 1286–1298.

Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. *23*, 4878–4884.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Rentas, S., Holzapfel, N.T., Belew, M.S., Pratt, G.A., Voisin, V., Wilhelm, B.T., Bader, G.D., Yeo, G.W., and Hope, K.J. (2016). Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. Nature *532*, 508–511.

Taggart, J., Ho, T.C., Amin, E., Xu, H., Barlowe, T.S., Perez, A.R., Durham, B.H., Tivnan, P., Okabe, R., Chow, A., et al. (2016). MSI2 is required for maintaining activated myelodysplastic syndrome stem cells. Nat. Commun. *7*, 10739.

van Galen, P., Kreso, A., Mbong, N., Kent, D.G., Fitzmaurice, T., Chambers, J.E., Xie, S., Laurenti, E., Hermans, K., Eppert, K., et al. (2014). The unfolded protein response governs integrity of

the haematopoietic stem-cell pool during stress. Nature *510*, 268–272.

Varrault, A., Dantec, C., Le Digarcher, A., Chotard, L., Bilanges, B., Parrinello, H., Dubois, E., Rialle, S., Severac, D., Bouschet, T., and Journot, L. (2017). Identification of Plagl1/Zac1 binding sites and target genes establishes its role in the regulation of extracellular matrix genes and the imprinted gene network. Nucleic Acids Res. *45*, 10466–10480.

Voz, M.L., Agten, N.S., Van de Ven, W.J., and Kas, K. (2000). PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. Cancer Res. *60*, 106–113.

Voz, M.L., Astrom, A.K., Kas, K., Mark, J., Stenman, G., and Van de Ven, W.J. (1998). The recurrent translocation t(5;8)(p13;q12) in pleomorphic adenomas results in upregulation of PLAG1 gene expression under control of the LIFR promoter. Oncogene *16*, 1409–1416.

Zheng, L., Baumann, U., and Reymond, J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. *32*, e115.

Zhu, J., Giannola, D.M., Zhang, Y., Rivera, A.J., and Emerson, S.G. (2003). NF-Y cooperates with USF1/2 to induce the hematopoietic expression of HOXB4. Blood *102*, 2420–2427.

Supplemental Information

PLAG1 and USF2 Co-regulate Expression of Musashi-2 in Human Hematopoietic Stem and Progenitor Cells

Muluken S. Belew, Sonam Bhatia, Ava Keyvani Chahi, Stefan Rentas, Jonathan S. Draper, and Kristin J. Hope

Figure S1

Figure S1. Sequence alignment of human PLAG1 isoforms. Related to Figure 2.
Shown is the alignment of the three PLAG1 isoforms and the location and number of Zn finger domains (blue boxes), transactivation domain (green box) and alternative start sites for PLAG1 isoforms (red).

Figure S2
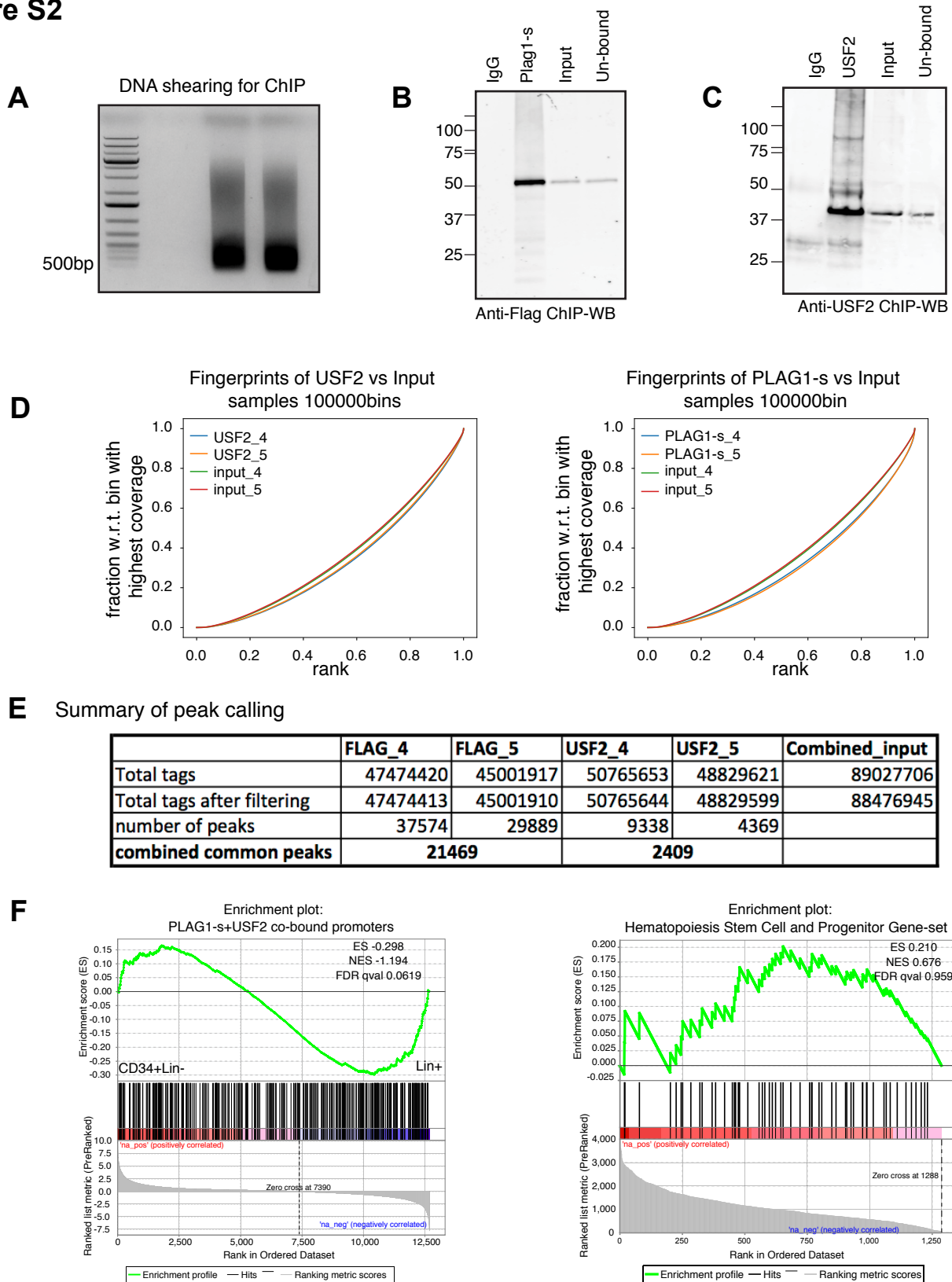
Figure S2. ChIP-seq experiments for PLAG1-s and USF2. Related to Figure 6.
(A.) Shearing profiles of the chromatin that was immunoprecipitated for ChIP-seq experiments. (B. and C.) ChIP-western blot (ChIP-WB) profiles to assess the efficiency of chromatin pull down prior to sequencing by (B.) anti-Flag antibody for PLAG1-s ChIP and (C.) anti-USF2 antibody for USF2 ChIP. (D.) Fingerprint plots showing the sequencing read coverage distribution across 100,000 randomly selected genomic bins for inputs as well as USF2, and PLAG1-s ChIP samples suggests that PLAG1-s ChIP was more efficient compared to USF2 ChIP (E.) Summary of peak calling for both individual replicates from K562 co-overexpression clones #4 and #5 for PLAG1-s and USF2 ChIP experiments. Peaks that were common to the two replicates (combined common peaks) were used for all downstream analysis. (F.) GSEA for PLAG1-s+USF2 co-occupied promoters against a ranked list of genes based on expression from Lin- to Lin+ populations (left) GSEA for PLAG1-s+USF2 co-occupied sites using the hematopoietic stem/progenitor gene set (right).

Table S1: Promoter fragment truncation and deletion primer sequences. Related to Figure 1. Same reverse primer was used in combination with the respective forward primer sequence to generate full length and truncation clones. Underlined are the XhoI and HindIII restriction sites in the forward and reverse sequences, respectively.

Promoter clone	Forward primer sequence	Reverse primer sequence
-3238-Luc	CTAGCTAG<u>CTCGAG</u>ATGGAGAGCAGAAAGGCACG	CTAGCTAG<u>AAGCTT</u>ATCTGAGCCCCCCCGCCCCC
-2251-Luc	CTAGCTAG<u>CTCGAG</u>GTGAGTTTGGCACCTCTGAG	
-1764-Luc	CTAGCTAG<u>CTCGAG</u>GTGAGTTTGGCACCTCTGAG	
-1188-Luc	CTAGCTAG<u>CTCGAG</u>AGGGGTTTCCTGAAATAACC	
-1013-Luc	CTAGCTAG<u>CTCGAG</u>CTTTCTCGCCACTGGAATCG	
-788-Luc	CTAGCTAG<u>CTCGAG</u>CGGGGGCCGCCAAAGGAG	
-588-Luc	CTAGCTAG<u>CTCGAG</u>GAGGAGCCGCAGCAAGC	
-203-Luc	CTAGCTAG<u>CTCGAG</u>TGAGCTAAGCCGAGCCC	
-138-Luc	CTAGCTAG<u>CTCGAG</u>ATTCGGAGGAGCCCGG	
d588-203-Luc	TGAGCTAAGCCGAGCCC	ACCCGGACCTGGGAGAAG

Table S2: Transcription factor binding site mutagenesis primers. Related to Figure 1.

Transcription Factor	Predicted Number of Binding Sites	Consensus Binding Sequence	Substitution	Forward Primer	Reverse Primer
Upstream stimulatory factor-2 (USF2)	3	CANNTG	GTNNGA	CCCGCCCCCGCAGCCTCTGGAATCCAGGGAAGTCGGG TCTAGAACTCCGCCCTCCGGAGCGCGGGAGGAGGG GAGCTAAGCCGAGCCTCCGGATGACGGGCTCTCGCCG	CCCGACTTCCCTGGATTCCAGAGGCTGCGGGGGCGGG CCCTCCTCCCGGCGTCCGGAGGGCGGGGAGTTCTAGA CGGCGAGCAGCCGTCATCCGGAGGCTCGGCTTAGCTC
Nuclear factor Y-A (NFYA)	1	GGCATTGGTT	TCAGCTGTAT	GGTGCGTGACGTCACCTCAGCTGTATACAGACGTTCTAGAAC	GTTCTAGAACGTCTGTATACAGCTGAGGTGACGTCACGCACC
cAMP responsive element binding protein (CREB)	1	TGACGT	TTAAGT	CGCACGGGGGTGCGTTAATTTACCGGCATTGGTTACA	TGTAACCAATGCCGGTAAATTAACGCACCCCCGTTGCG
Glucocorticoid Nuclear Receptor (NR3C1/GRE)	1	GTTC	GCCC	CATTGGTTACACGACGCCCTAGAAACTCCGCCCCA	TGGGGCGGAGTTCTAGGGCGTCGTGTAACCAATG
Microphthalmia-Associated Transcription Factor (MITF)	1	CATGTG	GATATA	CCCGCCCCCGCAGCCGGATATAATCCAGGGAAGTCGGG	CCCGACTTCCCTGGATTATATCGGCTGCGGGGGCGGG
CCCTC-binding factor (CTCF)	5	GGGG	ATATT	CCCCGGATCCACCCGAATATCGCAGCCCATGTGAT GTGGGTCCGGCCGTGATATTCCGAGTGTGCGCGCG GTGCGCGCGGGCAGGCATATTCCGCACATATTTGCGTGACGTCACCGGC GCCGGGGGGCGGGGAATATTGGGTGTGCGAGGCAGC	ATCACATGGGCTGCGATATTCGGGTGGATCGGGGGG CGCGCGCACACTCGGAATATCACGGCCGGACCCAC GCCGGTGACGTCACGCAAATATGTGCGGAATATGCCTGCCCGCGCGCGCCAC GCTGCCTCGCACACCAATATTCCCGCGCCCCCCGGC
PR Domain Containing 5 (PRDM5)	2	AGGG	TCTT	CGTGCGCCGGGAGGTCTTTGGAGGAGGAGGA GCGGGGGGAGGAGGAGGTCTTTGGAGGAGGAGGCGGA	TCCTCCTCCTCCTCCAAGACCTCCCGGCGCACG TCCGCTCCCTCCTCCAAGACCTCCTCCCCCCGC
Zinc fingerDNA binding protein 148/89 (ZFP148/ZBP89)	10	C-rich: CCCC G-rich: GGGG	C-rich: CGGT G-rich: GCCA	CGCCTCCCCTCCGGTCGATCCACCCGCGGTCCGCAGCCCAT GTCGGGGTGCCGCTCGGTCTCGCCCTGCGCCC GGAGGAGATGGCCATGGGAGGAGGAGGCCAAGAGGTGGGGAT GGGATGGGCCGGGGCCCACGGGGACGGGGG CGGGGAGCGGGGCCATGTGCGAGGCAGC CCCGGGGGGCGGGGCCAAGGAGGAGGCCAAGGAGGGGAGGCG	ATGGGCTGCGGACCGCGGGTGGATCGACCGAGGGGAGGCG GGGCCAGGCGGAGACGAGCGCACCCCGAC CCTCCTCCTCCTTGGCCTCCTCCCCGG ATCCCACCTCTTGGCCTCCTCCTCCCATGGCCATCTCCTCC CCCCGTCCCCGTGCCCCGGCCCCATCCC GCTGCCTCCCACATGGCCCCCGTCCCCG CGCTCCCTCCCTTGGCCTCCTCCTTGGCCCCGCCCGG
Pleomorphic adenoma gene 1 (PLAG1)	3	GGGG	TGAT	GTGCGCCGGGAGGATGATGAGTAATAGGAGGAGATGG GGAGATGGGGGTATGGAGGAGGAATGAGAGAGGTGGGG CAGAGAGATTCGTAATAGCCCGGGCGTGATGGAGGAGGAGGGGGAGGGGAGGGAGCG	CCATCTCCTCCTATTACTCATCATCCTCCCCGGCGCAC CCCCACCTCTCTCATTCCTCCTCCATACCCCCATCTCC CGCTCCCTCCCTCATCCTCCTCATCACGCCCGGGCTATTACGAAATCTCTG
Myc associated zinc finger protein1 (MAZ1)	7	GGGAGGGNNNN CCCTCCCNNNN	TGTATGTNTNT ACATACANANA	AAGCTCCCCGCACTCCACTACACAGATCCACCCGCC GGCAGGCGGGTCACATAGACATGCCCCCTGCC GAACCCCGGGGTGTATGTGACCCCGCGCTCC CCACGTGCGCCGTGTATGATGGGGAGGAGGAG GGGAGGAGGAGGTGTATATGTGGGGATGGGC GGGAGGAGGAGGTGTATGATGGGCGGAGATC TTTTGGGGTGGGGTGTCTGTGGGGCTCAGAT	GGCGGGTGGATCTGTGTATGTGAGTCGCGGGGAGCTT GGCAGGCGGGCCATGTCTATGTGGAGGCGCACCCC GAACCCCGGGGTGTATGTGACCCCGCGCTCC CTCCTCCTCCCACATACACGGCCGCACGTCG GCCCATCCCACCATATACACCTCCTCCTCCC GATCTCGCTCCATCAATACACCTCCTCCTCCC ATCTGAGCCCCACAGACACCCCACCCCCAAAA

Table S3. qPR-PCR Primers and Universal Probe Library Number for MSI2 and CYP1B1. Related to Figure 2-5.

Gene	Accession	Primer sequences	Length	Position	Tm	%GC	UPL#
MSI2 isoform 1	NM_138962.2	ggcagcaagaggatcagg ccgtagagatcggcgaca	18 18	947 - 964 1016 - 1033	59 60	61 61	26
CYP1B1	NM_000104.3	acgtaccggccactatcact ctcgagtctgcacatcagga	20 20	1359 - 1378 1450 - 1469	59 60	55 55	20

Table S5: KEGG Pathway Analysis of genes co-bound by PLAG1-S and USF2. Related to Figure 6.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjusted	qvalue	geneID	Count
hsa04142	Lysosome	22/436	123/7317	2.78E-06	7.79E-04	6.85E-04	3916/9114/57192/4758/2548/285362/2799/9179/1509/1201/1508/51606/23062/1200/1497/4074/8722/2629/5538/8905/3425/427	22
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	15/436	68/7317	7.94E-06	1.11E-03	9.78E-04	9114/528/5336/8992/9550/529/155066/998/51606/6300/5599/102/4792/6714/3725	15
hsa04933	AGE-RAGE signaling pathway in diabetic complications	18/436	99/7317	1.78E-05	1.66E-03	1.46E-03	581/5581/5336/998/6776/5296/1729/596/6300/5599/4893/5054/4846/4088/5594/595/7423/3725	18
hsa04917	Prolactin signaling pathway	13/436	70/7317	2.09E-04	1.46E-02	1.28E-02	8835/2592/6776/5296/6300/5599/25759/4893/5594/2309/595/6714/894	13
hsa04721	Synaptic vesicle cycle	12/436	63/7317	2.84E-04	1.59E-02	1.40E-02	9114/528/8992/8775/9550/529/155066/51606/2054/774/5864/594855	12
hsa04140	Autophagy - animal	18/436	128/7317	5.39E-04	2.15E-02	1.89E-02	3916/64121/64422/1509/1508/29982/11337/5296/596/5599/4893/10325/55102/5594/55062/22863/9474/54541	18
hsa05110	Vibrio cholerae infection	10/436	50/7317	6.06E-04	2.15E-02	1.89E-02	9114/528/5336/8992/9550/529/155066/51606/60/375	10
hsa04370	VEGF signaling pathway	11/436	59/7317	6.14E-04	2.15E-02	1.89E-02	5336/998/5296/6300/25759/4893/4846/8877/5594/7867/6714	11
hsa04966	Collecting duct acid secretion	7/436	27/7317	7.95E-04	2.47E-02	2.18E-02	9114/528/8992/9550/529/155066/10723	7
hsa04137	Mitophagy - animal	11/436	65/7317	1.42E-03	3.62E-02	3.18E-02	79735/11337/5599/4893/2309/665/9474/55288/6714/51024/3725	11
hsa05131	Shigellosis	11/436	65/7317	1.42E-03	3.62E-02	3.18E-02	998/1729/6300/5599/3688/5594/9474/4792/6714/60/7414	11
hsa04931	Insulin resistance	15/436	107/7317	1.59E-03	3.70E-02	3.26E-02	6945/5581/92579/133522/5296/5599/4846/5106/10891/5770/1374/6720/4792/79660/10999	15

Table S6: GO Pathway Analysis of genes co-bound by PLAG1-S and USF2. Related to Figure 6.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjusted	qvalue	geneID	Count
GO:0006914	autophagy	57/983	428/16672	6.24E-09	2.96E-05	2.84E-05	LAMP1/ATP6V0D1/MCOLN1/ULK3/RRAGC/ATG3/MTDH/TRAPPC8/CTSD/ATP6V1C1/VPS26A/ATP6V0E1/ATP6V1G1/RAB3GAP2/TBC1D17/ATP6V1E1/ATP6V0E2/LAMTOR1/CLN3/WDR6/ATP6V1H/NRBF2/VPS37B/VPS11/U2AF2/PIK3R2/BCL2/FBXL2/MTM1/PIP4K2C/MAPK8/GAPDH/VPS33A/VPS16/RNF185/TICAM1/GBA/TPCN1/RRAGB/ATG2B/EXOC1/CDK5R1/LAMTOR4/PIM2/WIPI1/PPARGC1A/BNIP3L/ATG14/FUNDC2/ATG5/MAP1S/KIAA1324/SREBF1/SPTLC2/NEDD4/FIS1/UBC	57
GO:0016236	macroautophagy	35/983	234/16672	3.46E-07	8.20E-04	7.87E-04	ATP6V0D1/MCOLN1/RRAGC/ATG3/TRAPPC8/ATP6V1C1/VPS26A/ATP6V0E1/ATP6V1G1/RAB3GAP2/ATP6V1E1/ATP6V0E2/LAMTOR1/CLN3/ATP6V1H/MTM1/PIP4K2C/MAPK8/GAPDH/VPS33A/VPS16/GBA/RRAGB/ATG2B/EXOC1/CDK5R1/LAMTOR4/WIPI1/BNIP3L/ATG14/ATG5/KIAA1324/SPTLC2/NEDD4/UBC	35
GO:0010506	regulation of autophagy	35/983	261/16672	4.55E-06	7.19E-03	6.90E-03	ATP6V0D1/RRAGC/MTDH/ATP6V1C1/VPS26A/ATP6V0E1/ATP6V1G1/RAB3GAP2/ATP6V1E1/ATP6V0E2/CLN3/WDR6/ATP6V1H/NRBF2/U2AF2/PIK3R2/BCL2/FBXL2/MTM1/PIP4K2C/MAPK8/GAPDH/TICAM1/GBA/TPCN1/RRAGB/EXOC1/CDK5R1/PIM2/BNIP3L/ATG14/KIAA1324/SREBF1/SPTLC2/NEDD4	35
GO:0016241	regulation of macroautophagy	21/983	123/16672	9.67E-06	1.15E-02	1.10E-02	ATP6V0D1/ATP6V1C1/VPS26A/ATP6V0E1/ATP6V1G1/RAB3GAP2/ATP6V1E1/ATP6V0E2/CLN3/ATP6V1H/MTM1/PIP4K2C/MAPK8/GAPDH/GBA/EXOC1/CDK5R1/BNIP3L/KIAA1324/SPTLC2/NEDD4	21
GO:0048511	rhythmic process	38/983	310/16672	1.51E-05	1.43E-02	1.38E-02	BAX/NR1D2/DYRK1A/VGF/TEF/BHLHE41/HNRNPD/DBP/NR1D1/SRD5A1/BCL2/MAPK8/CRY1/RELB/ADNP/HLF/SERPINE1/NOS3/KLF10/CDK5R1/ANG/BMPR1B/CSF2/TIMELESS/DDX5/PER3/FOXO3/PPARGC1A/CPT1A/RORA/TOP1/SREBF1/CHRNA7/NR5A1/JUND/MSH4/SOD1/JUN	38
GO:0007033	vacuole organization	28/983	209/16672	4.06E-05	2.95E-02	2.83E-02	MCOLN1/PDCD6IP/GAA/ATG3/TRAPPC8/ARFGEF2/RAB3GAP2/LAMTOR1/CLN3/HPS1/VPS37B/VPS11/MTM1/PIP4K2C/TPP1/VPS33A/VPS16/HOOK2/PPT1/ATG2B/WIPI1/STX6/P2RX7/ATG14/ATG5/KIAA1324/CLN6/ARF1	28
GO:0051452	intracellular pH reduction	11/983	45/16672	4.50E-05	2.95E-02	2.83E-02	ATP6V0D1/ATP6V1C1/ATP6V0E1/ATP6V1G1/ATP6V1E1/ATP6V0E2/CLN3/ATP6V1H/BCL2/PPT1/CLN6	11
GO:0045851	pH reduction	11/983	46/16672	5.60E-05	2.95E-02	2.83E-02	ATP6V0D1/ATP6V1C1/ATP6V0E1/ATP6V1G1/ATP6V1E1/ATP6V0E2/CLN3/ATP6V1H/BCL2/PPT1/CLN6	11
GO:0048025	negative regulation of mRNA splicing, via spliceosome	7/983	19/16672	6.53E-05	2.95E-02	2.83E-02	DYRK1A/U2AF2/PTBP1/SFSWAP/SRSF6/SRSF7/HNRNPA2B1	7
GO:0007034	vacuolar transport	33/983	272/16672	6.62E-05	2.95E-02	2.83E-02	AP5Z1/LAMP1/VPS50/AP4M1/MICALL2/TRAPPC8/VPS26A/RAB29/TBC1D17/SPAG9/SNX8/CLN3/SNX2/RAB21/VPS37B/VPS11/MTM1/GPRASP1/VPS33A/M6PR/VPS16/HOOK2/MAPK1/SNX9/WIPI1/STX6/ATG14/SRC/RHOB/NEDD4/ARF1/UBC	33
GO:0042594	response to starvation	22/983	150/16672	6.84E-05	2.95E-02	2.83E-02	RRAGC/FNIP1/SLC38A2/HMGCL/UCP2/SRD5A1/BCL2/CAD/COMT/BCAS3/GBA/RRAGB/KLF10/PPARGC1A/SEH1L/ATG14/ATG5/	22

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
							KIAA1324/NUAK2/SREBF1/AACS/JUN	
GO:0071417	cellular response to organonitrogen compound	48/983	462/16672	9.61E-05	3.71E-02	3.56E-02	ATP6V0D1/SOCS2/RRAGC/MGARP/PRKAR2B/ATP6V1C1/ATP6V0E1/ATP6V1G1/APEX1/ATP6V1E1/ATP6V0E2/LAMTOR1/OSBPL8/HNRNPD/ATP6V1H/PLA2G1B/STAT5A/UCP2/SRD5A1/PIK3R2/RAB31/COL6A1/DIAPH1/DNMT3A/GCLC/AGTRAP/PDK2/KLF15/RRAGB/KLF10/MAPK1/LAMTOR4/KLF4/FER/HDAC5/FOXO3/FGF21/PPARGC1A/BAIAP2L1/CAPN10/PTPN1/COL16A1/SLC34A1/PRKAR1A/SREBF1/PDE2A/SRC/SOD1	48
GO:1901652	response to peptide	45/983	425/16672	1.02E-04	3.71E-02	3.56E-02	ATP6V0D1/SOCS2/MGARP/PRKAR2B/ATP6V1C1/VGF/BSG/ATP6V0E1/ATP6V1G1/APEX1/ATP6V1E1/ATP6V0E2/OSBPL8/ATP6V1H/PLA2G1B/STAT5A/UCP2/SRD5A1/PIK3R2/RAB31/GCLC/CAD/AGTRAP/KHK/CRY1/PDK2/KLF15/KLF10/MAPK1/KLF4/FER/CACYBP/HDAC5/FOXO3/FGF21/BAIAP2L1/CAPN10/PTPN1/SLC34A1/PRKAR1A/SREBF1/NFKBIA/SRC/SRSF6/JUND	45
GO:1901653	cellular response to peptide	35/983	304/16672	1.17E-04	3.97E-02	3.81E-02	ATP6V0D1/SOCS2/MGARP/PRKAR2B/ATP6V1C1/ATP6V0E1/ATP6V1G1/APEX1/ATP6V1E1/ATP6V0E2/OSBPL8/ATP6V1H/PLA2G1B/STAT5A/UCP2/SRD5A1/PIK3R2/RAB31/GCLC/AGTRAP/PDK2/KLF15/KLF10/MAPK1/KLF4/FER/HDAC5/FOXO3/FGF21/BAIAP2L1/CAPN10/PTPN1/PRKAR1A/SREBF1/SRC	35
GO:0008333	endosome to lysosome transport	10/983	42/16672	1.26E-04	4.00E-02	3.84E-02	VPS11/MTM1/GPRASP1/VPS33A/M6PR/VPS16/SNX16/HOOK2/ATG14/RHOB	10
GO:1903312	negative regulation of mRNA metabolic process	9/983	35/16672	1.46E-04	4.32E-02	4.14E-02	DYRK1A/U2AF2/PTBP1/CCNT1/SFSWAP/SRSF6/SRSF7/TOB1/HNRNPA2B1	9
GO:0000082	G1/S transition of mitotic cell cycle	28/983	227/16672	1.74E-04	4.53E-02	4.35E-02	BAX/CACUL1/HINFP/CDK7/APEX1/EIF4E/BCL2/NACC2/GSPT1/PRMT1/RPA1/ITGB1/MEPCE/JADE1/KANK2/PIM2/PLK5/CNOT4/CCNE1/CCND1/CUL4B/CDC25C/MCM5/MCM2/CTDSP2/CCND2/PLAGL1/UBC	28
GO:0007040	lysosome organization	11/983	52/16672	1.83E-04	4.53E-02	4.35E-02	GAA/LAMTOR1/CLN3/HPS1/VPS11/TPP1/HOOK2/PPT1/P2RX7/CLN6/ARF1	11
GO:0080171	lytic vacuole organization	11/983	52/16672	1.83E-04	4.53E-02	4.35E-02	GAA/LAMTOR1/CLN3/HPS1/VPS11/TPP1/HOOK2/PPT1/P2RX7/CLN6/ARF1	11
GO:0007035	vacuolar acidification	6/983	16/16672	1.99E-04	4.53E-02	4.35E-02	ATP6V0E1/ATP6V0E2/CLN3/ATP6V1H/PPT1/CLN6	6
GO:0050686	negative regulation of mRNA processing	8/983	29/16672	2.00E-04	4.53E-02	4.35E-02	ATP6V0E1/ATP6V0E2/CLN3/ATP6V1H/PPT1/CLN6	8

Table S7: Reactome GSEA of genes co-bound by PLAG1-S and USF2. Related to Figure 6.

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
REACTOME_CELL_CYCLE	25	0.69933075	1.444864	0	0.092173874	0.088	257	tags=36%, list=20%, signal=44%
REACTOME_TRANSMEMBRANE_TRANSPORT_OF_SMALL_MOLECULES	37	0.6706475	1.4014298	0.002	0.11639127	0.196	176	tags=32%, list=14%, signal=36%
REACTOME_CELL_CYCLE_MITOTIC	20	0.6407614	1.2703049	0.07214429	0.503188	0.78	257	tags=35%, list=20%, signal=43%
REACTOME_MRNA_PROCESSING	22	0.6112415	1.2238253	0.118236475	0.6260868	0.924	319	tags=41%, list=25%, signal=53%
REACTOME_HIV_INFECTION	19	0.605965	1.2200736	0.125	0.51733893	0.934	309	tags=47%, list=24%, signal=61%
REACTOME_METABOLISM_OF_PROTEINS	37	0.5639318	1.1745437	0.12825651	0.6692175	0.99	343	tags=41%, list=27%, signal=54%
REACTOME_METABOLISM_OF_CARBOHYDRATES	17	0.56408656	1.122007	0.27108434	0.8701489	1	180	tags=29%, list=14%, signal=34%
REACTOME_SLC_MEDIATED_TRANSMEMBRANE_TRANSPORT	24	0.5438985	1.106202	0.284	0.8493914	1	144	tags=21%, list=11%, signal=23%
REACTOME_PHOSPHOLIPID_METABOLISM	18	0.5415519	1.0580957	0.4084507	1	1	390	tags=44%, list=30%, signal=63%
REACTOME_GENERIC_TRANSCRIPTION_PATHWAY	16	0.5298262	1.0539083	0.4248497	0.95460856	1	305	tags=31%, list=24%, signal=40%
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS	39	0.49715915	1.0450431	0.388	0.91145444	1	390	tags=41%, list=30%, signal=57%
REACTOME_TRANSCRIPTION	18	0.52930707	1.0423437	0.40480962	0.8432317	1	394	tags=33%, list=31%, signal=47%

REACTOME_PLATELET_ACTIVATION_SIGNALING_AND_AGGREGATION	15	0.5285818	1.0418788	0.42566192	0.77960515	1	330	tags=27%, list=26%, signal=35%
REACTOME_ADAPTIVE_IMMUNE_SYSTEM	26	0.513643	1.0310123	0.452	0.76632285	1	344	tags=38%, list=27%, signal=51%
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	19	0.5146221	1.0180796	0.49	0.75990134	1	319	tags=32%, list=25%, signal=41%
REACTOME_HEMOSTASIS	31	0.43617868	0.9102693	0.726	1	1	353	tags=26%, list=27%, signal=35%
REACTOME_SIGNALING_BY_GPCR	23	0.4300659	0.8785485	0.73547095	1	1	330	tags=26%, list=26%, signal=34%
REACTOME_IMMUNE_SYSTEM	53	0.399912	0.85586494	0.888	1	1	570	tags=47%, list=44%, signal=81%
REACTOME_SIGNALLING_BY_NGF	20	0.42523456	0.85515	0.7755511	1	1	570	tags=55%, list=44%, signal=97%
REACTOME_MRNA_SPLICING	16	0.4288823	0.84179723	0.77911645	0.9739479	1	319	tags=25%, list=25%, signal=33%
REACTOME_GPCR_DOWNSTREAM_SIGNALING	18	0.41564652	0.81756485	0.8266129	0.9627115	1	330	tags=28%, list=26%, signal=37%
REACTOME_INNATE_IMMUNE_SYSTEM	21	0.30153662	0.60955846	0.9839357	1	1	747	tags=57%, list=58%, signal=134%
REACTOME_DEVELOPMENTAL_BIOLOGY	24	0.27081522	0.54898703	0.986	1	1	869	tags=71%, list=67%, signal=214%
REACTOME_NEURONAL_SYSTEM	15	0.26906207	0.526231	0.9919355	0.993899	1	712	tags=47%, list=55%, signal=103%

Supplemental Experimental Procedures

Co-Immunoprecipitation

1×10^7 Cells were lysed in RIPA buffer containing protease inhibitors and incubated with rotation overnight at 4°C with 2µg anti-USF2 (SantaCruz, #sc-293443) or anti-mouse IgG (SantaCruz) and Protein G dynabeads. The precipitated complex was washed with RIPA buffer and coimmunoprecipitated proteins were eluted in 2x SDS Loading buffer and analyzed by Western Blot. Similar experiments were performed with 1% PFA cross-linked cells and anti-Flag (Sigma, F1804).

Chromatin Immunoprecipitation (ChIP)

ChIP-qPCR was done as per Carey et al with slight modifications (Carey et al., 2009). Briefly, cells were cross-linked, lysed and nuclei subjected to M220 focused-ultra sonicator (Covaris) to generate sheared chromatin fragments of ~500 bp. 5µg anti-USF2 (SantaCruz, #sc-862), anti-PLAG1 (GenTex, # GTX124217) or anti-rabbit IgG (SantaCruz) were added per 1000µL of IP volume followed by precipitation of chromatin-antibody complex by Protein G dynabeads (Invitrogen). Immunoprecipitated DNA was extracted using phenol/chloroform, quantified and subjected to qPCR. Primers and probes were designed by the Universal Probe Library (UPL) probe finder tool (Roche) using input sequence comprising 588 bp upstream of the MSI2 TSS. For the promoter region we used UPL probe #63 and forward and reverse primers as follows: CGCTCGCAGAGAGATTCG; GAGATCTCCGCTCCCTCCT. ChIP fold enrichment was calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008).

For ChIP-seq, 1×10^8 cells K562 cells co-overexpressing Flag-PLAG1-S and USF2 were cross-linked in 1% PFA and excess PFA was washed with glycine, cells were lysed in

RIPA buffer (50mM Tris-Cl pH7.45, 150mM NaCl, 0.1% SDS, 2% NP-40, 1% Sodium deoxycholate) containing protease inhibitors and nuclei were subjected to probe sonication. Two percent of each sample was used for Input sequencing. Chromatin was diluted in dilution buffer (50mM Tris-Cl pH7.45, 150mM NaCl) to get a final SDS concentration of 0.025% to assist in immunoprecipitation (IP). Twenty µg anti-USF2 (SantaCruz, #sc-293443) or anti-Flag (Sigma, F1804) and Protein G dynabeads were used for IP. The precipitated complex was washed in low (50mM Tris-Cl, pH 7.45, 250mM NaCl, 2mM EDTA, 0.1%SDS w/v, 1% Triton X-100 v/v) and high salt buffer (same as low salt buffer except use 500mM NaCl) and chromatin was eluted in TE buffer at 65°C before reverse cross-linking with Proteinase K (NEB, P8107S). Sequencing libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB, E7645S) and sequencing as 50 bps single ended reads was performed on Illumina HiSeq 1500 at a depth of 55 million reads per input and 42.5 million reads per IP.

ChIP-seq analysis

Sequencing reads were quality controlled using fastQC, and the adaptors were removed using cutadapt software using the illumina universal primer sequence AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC. The reads were aligned to human genome assembly GRCh38 using hisat2 aligner (version hisat2-2.0.4) (Kim et al., 2015) and duplicates were removed using samtools (Li et al., 2009). Fingerprint plots were generated using deepTools (Ramirez et al., 2016) to check efficiency of the ChIP experiments. Samtools merge was used to merge the input files to create a

master background files, and MACS2 (Feng et al., 2011) was used to called peaks from the individual ChIP replicates using the merged input files. The following peak calling parameters were used:

macs2 callpeak -t ChIP_file1.bam -c Input_merged.bam -f BAM -g hs –n ChIP_file1_peaks -B -s 50 --bw 150 -q 0.01 --outdir

Since the IP efficiency was lower for USF2 the peaks were called using the following less stringent parameters:

macs2 callpeak –t USF2_ChIP_file1.bam -c Input_merged.bam -f BAM -g hs -n ChIP_file1_peaks -B -s 50 --bw 150 -q 0.05 –m 2 50 --outdir

Bedtools intersect (Quinlan and Hall, 2010) was used to identify peaks common to the two replicates and were used for all downstream analysis. Peak files were annotated using HOMER (Heinz et al., 2010) annotatepeaks.pl script using hg38 as the background. HOMER's mergepeaks function was used to identify PLAG1-S and USF2 co-bound targets within 100bps from each other. findMotifsGenome.pl was used for motif analysis with the following parameters -size 200 –mask. Motif distribution profiles were generated using the HOMER function annotatePeaks.pl with the USF2 consensus motif file and the second G-rich region identified as being the PLAG1-S consensus. The R-packages ChIPSeeker (Yu et al., 2015), and clusterProfiler   (Yu et al., 2012) were used to perform peak distribution and pathway analysis. Significance of any enriched pathways and GO processes was addressed by setting adjusted pvalueCutoff to 0.05, and by adjusting the p-values using Benjamini & Hochberg (BH) method that is built into the clusterProfiler package. Gene-set enrichment analyses were performed using the GSEA tool with 500-1500 permutations.

Cord blood CD34+ cell isolation, lentiviral transduction and *in vitro* culture

FACS isolated Lin⁻CD34⁺ 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). Cells were transduced at an MOI of 50-100 with pSMALB lentiviral particles encoding 3xFLAG-PLAG1-p2A-USF2 (PLAG1+USF2) or firefly luciferase (Vector). BFP+ cells were purified using a MoFlo XDP (Beckman Coulter) 72hrs post-transduction. For *in vitro* culture, cells were seeded at a density of 1×10^5 cells/ml in IMDM 10% FBS supplemented with IL-6 (10 ng/ml), SCF (50 ng/ml), FLT3-L (50 ng/ml), and TPO (20 ng/ml) and counted and analyzed by flow cytometry for CD34 positivity at day 3 and 7. Flow cytometry data was analyzed by Flowjo (Tree star). Immunocytochemistry staining, imaging and analysis for MSI2 and CYP1B1 protein levels was done as previously reported (Rentas et al., 2016).

Statistical analysis

All statistical analyses were performed using PRISM (Graphpad software, Inc). Where appropriate statistical significance was estimated using non-parametric two-tailed t-test for pairwise comparisons or 1-way ANOVA multiple comparisons of control and test transfections/transductions. Data shown as mean ±SEM. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Supplemental References

Carey, M.F., C.L. Peterson, and S.T. Smale. 2009. Chromatin immunoprecipitation (ChIP). *Cold Spring Harbor protocols* 2009:pdb prot5279.

Feng, J., T. Liu, and Y. Zhang. 2011. Using MACS to identify peaks from ChIP-Seq data. *Curr Protoc Bioinformatics* Chapter 2:Unit 2 14.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576-589.

Kim, D., B. Langmead, and S.L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357-360.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and S. Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.

Ramirez, F., D.P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dundar, and T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* 44:W160-165.

Rentas, S., N.T. Holzapfel, M.S. Belew, G.A. Pratt, V. Voisin, B.T. Wilhelm, G.D. Bader, G.W. Yeo, and K.J. Hope. 2016. Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. *Nature* 532:508-511.

Schmittgen, T.D., and K.J. Livak. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* 3:1101-1108.

Yu, G., L.G. Wang, Y. Han, and Q.Y. He. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284-287.

Yu, G., L.G. Wang, and Q.Y. He. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31:2382-2383.