Combined Cohesin-Runx1 Deficiency Synergistically Perturbs Chromatin Looping and Causes Myelodysplastic Syndromes

Yotaro Ochi[1,2], **Ayana Kon**[1], **Toyonori Sakata**[3], **Masahiro M Nakagawa**[1], **Naotaka Nakazawa**[4], **Masanori Kakuta**[5], **Keisuke Kataoka**[1], **Haruhiko Koseki**[6], **Manabu Nakayama**[7], **Daisuke Morishita**[8], **Tatsuaki Tsuruyama**[9], **Ryunosuke Saiki**[1], **Akinori Yoda**[1], **Rurika Okuda**[1], **Tetsuichi Yoshizato**[1], **Kenichi Yoshida**[1], **Yusuke Shiozawa**[1], **Yasuhito Nannya**[1], **Shinichi Kotani**[1,2], **Yasunori Kogure**[1], **Nobuyuki Kakiuchi**[1], **Tomomi Nishimura**[1], **Hideki Makishima**[1], **Luca Malcovati**[10,11], **Akihiko Yokoyama**[12], **Kengo Takeuchi**[13], **Eiji Sugihara**[14], **Taka-aki Sato**[14], **Masashi Sanada**[15], **Akifumi Takaori-Kondo**[2], **Mario Cazzola**[10,11], **Mineko Kengaku**[4,16], **Satoru Miyano**[5], **Katsuhiko Shirahige**[3], **Hiroshi I Suzuki**[17,*], **Seishi Ogawa**[1,18,19,*]

[1]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3]Laboratory of Genome Structure and Function, Research Division for Quantitative Life Sciences, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan

[4]Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto, Japan

[5]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

[6]Laboratory for Developmental Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[7]Laboratory of Medical Omics Research, Department of Frontier Research and Development, Kazusa DNA Research Institute, Kisarazu, Japan

[8]Chordia Therapeutics Inc., Kanagawa, Japan

[9]Department of Drug and Discovery Medicine, Pathology Division, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[10]Department of Molecular Medicine, University of Pavia, Pavia, Italy

[11]Department of Hematology Oncology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

[12]Tsuruoka Metabolomics Laboratory, National Cancer Center, Yamagata, Japan

[13]Pathology Project for Molecular Targets, Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan

[14]Research and Development Center for Precision Medicine, University of Tsukuba, Ibaraki, Japan

[15]Department of Advanced Diagnosis, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan

[16]Graduate School of Biostudies, Kyoto University, Kyoto, Japan

[17]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

[18]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan

[19]Department of Medicine, Centre for Haematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden

Abstract

STAG2 encodes a cohesin component and is frequently mutated in myeloid neoplasms, showing highly significant co-mutation patterns with other drivers, including *RUNX1*. However, the molecular basis of cohesin-mutated leukemogenesis remains poorly understood. Here we show a critical role of an interplay between Stag2 and Runx1 in the regulation of enhancer-promoter looping and transcription in hematopoiesis. Combined loss of Stag2 and Runx1, which co-localize at enhancer-rich, Ctcf-deficient sites, synergistically attenuates enhancer-promoter loops, particularly at sites enriched for RNA polymerase II and Mediator, and deregulates gene expression, leading to myeloid-skewed expansion of hematopoietic stem/progenitor cells (HSPCs) and myelodysplastic syndromes (MDS). Attenuated enhancer-promoter loops in Stag2/Runx1-deficient cells are associated with downregulation of genes with high basal transcriptional pausing, which are important for regulation of HSPCs. Down-regulation of high-pausing genes is also confirmed in STAG2/cohesin-mutated primary leukemia samples. Our results highlight a unique STAG2/RUNX1 interplay in gene regulation and provide insights into cohesin-mutated leukemogenesis.

Keywords

myelodysplastic syndrome; cohesin; chromatin loop; transcriptional pausing; leukemogenesis

Introduction

Myelodysplastic syndromes (MDS) and related disorders are heterogeneous groups of myeloid neoplasms showing varying degrees of cytopenia due to ineffective hematopoiesis and a high propensity to progression to acute myeloid leukemia (AML) (1). During the past two decades, a complete registry of recurrent mutational targets, or driver genes, has been identified using advanced genomics (2-4). However, the functional basis of mutations has not fully been elucidated in many of those drivers. Among these are *STAG2* and other members of the cohesin complex, including *SMC1*, *SMC3*, and *RAD21*, which are a new class of driver genes mutated in ~10% of MDS and other myeloid neoplasms with *STAG2* being most frequently affected (5-7). Most mutations in *STAG2* are nonsense or frameshift, leading to protein truncation and loss-of-function (5). Involved in different cellular processes, such as sister chromatid cohesion during cell division and DNA repair (8), cohesin is also implicated in transcriptional control (9-12), possibly through regulating high-order chromatin structures (13). However, it is largely unknown how mutated cohesin contributes to myeloid leukemogenesis. In this study, through the analysis of interactions of gene mutations in a large cohort of MDS followed by the analysis of relevant mouse models, we show a strong functional interplay between Stag2 and Runx1 in the regulation of chromatin structures and gene expression in the hematopoietic compartment, providing novel insight into the leukemogenic mechanism of a unique subset of myeloid neoplasms characterized by *STAG2* and mutually highly correlated mutations.

Results

Genetic interaction of mutations in human MDS/AML

In MDS/AML, *STAG2* mutations are rarely seen as a solitary mutation, but almost always accompanied by other mutations, frequently involving *SRSF2, RUNX1, ASXL1, CEBPA, BCOR, EZH2, IDH2,* and *NRAS* (2-4). To see this in more detail, we investigated significant mutational correlations in MDS and related myeloid neoplasms, using in-house or publicly available mutation data sets from 3,047 cases with MDS (n=2,498) and related myeloid neoplasms (n=549) (2,3,14-18). After exhaustively evaluating correlations across all pair-wise combinations among 42 major drivers commonly mutated in MDS/AML, we detected a number of significant positive and negative correlations (Fig. 1A, and Supplementary Tables S1-3). Remarkably, the top-ranked 6 correlations were exhausted by all possible pair-wise combinations among four genes, *STAG2, RUNX1, SRSF2*, and *ASXL1* ('*SRSA*' genes), which are involved in gene regulation, and were co-mutated at significantly higher frequencies than expected only by chance (Fig. 1A-B). One or more of these genes accounted for 31.8% (n=970) of MDS/AML, of which 346 (35.7%) had mutations in 2 of these genes and 75 (7.7%) and 33 (3.4%) carried mutations in three and all four genes, respectively (Fig. 1B and Supplementary Fig. S1A-B). Patients with 2 *SRSA* mutations had a significantly poor overall survival, compared with those with just one, which still negatively affected the survival (Fig. 1C). Numbers of other driver mutations did not differ according to the number of *SRSA* mutations, suggesting that *SRSA* combination is not just a consequence of increased total mutations (Supplementary Fig.

S2A). No significant difference was observed in the frequency of missense vs non-sense or frameshift mutations in the *RUNX1* gene between *STAG2*-WT and mutated cases (Supplementary Fig. S2B). Analysis of variant allele frequency suggests that *SRSF2* mutations are acquired earlier than other 3 mutations, followed by *RUNX1* mutations and then *STAG2* and *ASXL1* mutations (Fig. 1D and Supplementary Fig. S2C). We also observed high frequency of converging evolution by way of 'parallel' *STAG2* mutations; multiple, as many as four, independent *STAG2*-mutated subclones were detected in 17 (22%) of 76 evaluable cases with *STAG2* mutations, of which 16 carried *RUNX1, SRSF2,* and/or *ASXL1* mutations in the major tumor population, indicating that *STAG2* mutations should confer a strong selective advantage in these mutational contexts (Fig. 1E). Combined, these findings suggest strong functional interactions among *SRSA* mutations in positive selection that underlie the development/progression of MDS.

Expanded HSPC pools and differentiation block in *Stag2* knockout mice

To understand the leukemogenic mechanism of *SRSA*-mutated MDS, particularly focusing on *STAG2* mutation, which showed a unique converging evolution pattern (Fig. 1E) and has been less studied in terms of functional consequence compared to other *SRSA* genes, we first generated a mouse model having a conditional *Stag2* knockout allele with an *Mx1-Cre* transgene ($Mx1\text{-}Cre^+$ $Stag2^{fl/-}$; SKO) (Supplementary Fig. S3A-C; see Methods). After polyinosinic-polycytidylic acid (pIpC) injection, SKO mice exhibited a slightly decreased white blood cell (WBC) with a large reduction in B-lymphocytes ($B220^+$), compared with littermate wild-type (WT) mice ($Mx1\text{-}Cre^-$ $Stag2^{fl/-}$) (Fig. 2A). While no significant changes were observed in hemoglobin level and platelet count between SKO and WT mice, SKO mice showed significantly increased red cell distribution width (RDW), suggestive of dyserythropoiesis (19) and pathological examination revealed mild tri-lineage bone marrow (BM) dysplasia and a slightly enlarged spleen with evidence of extramedullary hematopoiesis in SKO mice (Fig. 2A and Supplementary Fig. S3D-E). Collectively these findings support the presence of a mild MDS-like phenotype in SKO mice, although overall survival did not differ between SKO and WT mice (Supplementary Fig. S3F). In BM, SKO mice had a higher frequency of $Lin^-/Sca1^+/c\text{-}Kit^+$ (LSK) cells, compared with WT littermate, where all major subfractions of LSK cells were increased (Fig. 2B and Supplementary Fig. S4A), indicating expanded hematopoietic stem/progenitor cell (HSPC) pools. The increase was most prominent in myeloid-biased progenitors, including multipotent progenitor (MPP)-2 ($Lin^-/c\text{-}Kit^+/Sca\text{-}1^+/CD135^-/CD150^+/CD48^+$) and MPP-3 ($Lin^-/c\text{-}Kit^+/Sca\text{-}1^+/CD135^-/CD150^-/CD48^+$), followed by LT-HSC ($Lin^-/c\text{-}Kit^+/Sca\text{-}1^+/CD135^-/CD150^+/CD48^-$) and ST-HSC ($Lin^-/c\text{-}Kit^+/Sca\text{-}1^+/CD135^-/CD150^-/CD48^-$), suggesting the presence of myeloid skewing (20). The myeloid skewing was also evident in more differentiated progenitors as evident from increased common myeloid progenitors (CMPs; $Lin^-/c\text{-}Kit^+/Sca\text{-}1^-/CD34^+/Fc\gamma R^{med}$) and granulocyte-macrophage progenitors (GMPs; $Lin^-/c\text{-}Kit^+/Sca\text{-}1^-/CD34^+/Fc\gamma R^{high}$), and decreased common lymphoid progenitors (CLPs; $Lin^-/c\text{-}Kit^{med}/Sca\text{-}1^{med}/IL\text{-}7R\alpha^+/Flt3^+$) (Fig. 2C and Supplementary Fig. S4B). Of note, SKO mice showed decreased frequencies of megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs; $Lin^-/c\text{-}Kit^+/Sca\text{-}1^-/CD34^-/Fc\gamma R^{low}$) and erythroid progenitors ($Ter119^+/CD71^+$), suggesting a blocked differentiation into erythromegakaryocyte lineages (Fig. 2C and Supplementary Fig. S4B-D). Mature cells were

also skewed to myeloid lineages with increased granulocytes/monocytes (CD11b$^+$) and decreased B-lymphocytes in the BM and spleen (Fig. 2D and Supplementary Fig. S4E-F). Extramedullary hematopoiesis was evident from increased frequencies of erythroid progenitors in the spleen (Supplementary Fig. S4C-D). In agreement with expanded HSPC pools, SKO-derived BM cells showed an enhanced clonogenicity in replating assay (Fig. 2E). HSCs (CD150$^+$/CD48$^-$ LSK cells) from SKO mice showed a decreased frequency of apoptotic cells (Annexin$^+$/7-AAD$^-$) and an enhanced cell cycling (S/G2/M; Ki-67$^+$/Hoechst $^+$) with decreased quiescent (G0; Ki-67$^-$/Hoechst$^-$) cells, compared with those from WT mice (Fig. 2F-G and Supplementary Fig. S4G-H). In competitive repopulation assay, SKO-derived cells showed enhanced chimerism within the LSK fraction, although the chimerism of SKO-derived cells was not significantly changed in total BM and even reduced in peripheral blood, particularly within lymphocytes (Fig. 2H). These results suggest an enhanced self-renewal and repopulation capacity of SKO-derived progenitors with a block in lymphoid differentiation.

As expected from the myeloid skewing in SKO mice, transcriptome analysis demonstrated up- and downregulation of genes implicated in myeloid and lymphoid programs in SKO, respectively, including a number of transcription factors (TFs) (Fig. 2I-J and Supplementary Fig. S4I). Among these, we noted the elevated expression of *Runx1* in SKO-derived LSK and CMP cells (Fig. 2K) (21). The following ATAC sequencing analysis (22) showed enrichment of the binding motifs of Runx1 and Gata2 in enhanced ATAC peaks and enrichment of the binding motifs of interferon regulatory factors (IRFs) in reduced ATAC peaks in SKO-derived LSK and CMP cells (Fig. 2L-M and Supplementary Fig. S5A-C). Furthermore, we confirmed an increased Runx1 binding to its consensus motifs in SKO cells by ChIP-seq (Supplementary Fig. S5D). These findings suggest upregulated expression of Runx1-regulated genes. However, interestingly, genes down-regulated in SKO-derived LSK cells showed a highly significant enrichment in genes down-regulated in *Runx1*-knockout mice (*Mx1-Cre*$^+$ *Runx1*$^{fl/fl}$; RKO) and vice versa (Fig. 2N), suggesting a functional interplay between Stag2 and Runx1, a typical *SRSA* combination.

Stag2/Runx1 codeficiency induces MDS in mice

To see the interplay between both proteins, we investigated the effects of *Stag2/Runx1* double knockout (*Mx1-Cre*$^+$ *Stag2*$^{fl/-}$*Runx1*$^{fl/fl}$; DKO) on hematological phenotype in transplantation setting, in which DKO- as well as single KO-derived BM cells were transplanted into lethally irradiated mice, followed by pIpC injection. Compared with single KO-transplanted mice, in which cytopenia was confined to lymphocytes (SKO and RKO) and platelets (RKO), DKO-transplanted mice exhibited more profound cytopenia (Fig. 3A). WBC was markedly reduced and the reduction was seen not only in lymphocytes but also in granulocytes and monocytes. Although not apparent in single-KO-transplanted mice, anemia was evident with markedly increased MCV and RDW (Fig. 3A and Supplementary Fig. S6A). In contrast to severely reduced peripheral blood counts, the frequency of immature BM progenitors (LSK cells) was almost doubled compared with that in single-KO-transplanted mice, where the increase was mostly explained by MPP2 and MPP3, although LT- and ST-HSC were significantly reduced compared with those in SKO-transplanted mice (Fig. 3B-C and Supplementary Fig. S6B). While more mature progenitors exhibited variable

profiles depending on genotype, they were largely maintained in DKO-transplanted mice, except for severely reduced frequency of colony-forming unit erythroid cells (CFUe) and Ter119$^+$CD71$^+$ erythroid progenitors (Fig. 3D-F and Supplementary Fig. S6C-F). Lineage-committed mature cells were synergistically skewed to myeloid lineages by *Stag2*/*Runx1* deletion (Fig. 3G and Supplementary Fig. S6E). Prolonged clonogenicity was also observed in replating assay (Supplementary Fig. S6G). In competitive repopulation assay, the peripheral blood chimerism was severely impaired, even though the chimerism in the progenitor (LSK) fraction was not significantly affected, which was compatible with severe peripheral blood cytopenias (Supplementary Fig. S6H). Conspicuously, all mice transplanted with DKO-derived BM cells developed overt MDS mostly within a half year after pIpC injection, with a severe cytopenia and marked trilineage dysplasia, while none of the animals transplanted with SKO- or RKO-cells developed MDS, although a number of deaths were also observed in RKO-transplanted mice with no exacerbation of blood cell counts (Fig. 3H-I and Supplementary Fig. S6I-L). In transcriptome analysis and ATAC-seq of LSK cells, DKO-derived LSK cells showed more extensive alterations in gene expression and chromatin accessibility, compared with single-KO-derived cells; DKO cells showed a much higher number of differentially expressed genes (DEGs) and differentially enhanced or attenuated chromatin accessibility sites, compared with single-KO (Fig. 3J-K and Supplementary Fig. S6M). These findings in the phenotype of DKO mice further support the functional interplay between Stag2 and Runx1.

Unique binding of Stag2-cohesin and Runx1 to enhancers

To understand the molecular basis of the interplay between Stag2 and Runx1, we performed an extensive ChIP-sequencing analysis, in which genomic localization of Stag2 and other cohesin components (Stag1 and Smc1), Runx1, Ctcf, and major histone marks was investigated in WT-derived HSPCs (c-Kit$^+$ BM cells). We identified a total of 27,997 cohesin binding sites that showed signal peaks in Stag1 and/or Stag2. On the basis of relative ChIP signals for Ctcf and H3K27ac, indicative for insulator and enhancer sites, respectively, these cohesin binding sites were separated into two discrete groups showing high relative Ctcf signals (cohesin binding cluster-I (CC-I) sites) and high relative H3K27ac signals (cohesin binding cluster-II (CC-II) sites) (Fig. 4A and Supplementary Fig. S7A-B). Explaining most of the cohesin-binding sites (n=24,364, 87%), CC-I sites generally had comparable Stag1 and Stag2 signal intensities with discrete signal peaks and were characterized by the paucity of active histone marks (Fig. 4A). By contrast, accounting for only 13% (n=3,633) of all cohesin binding sites detected, CC-II sites had stronger Stag2 than Stag1 signals showing broader chromatin binding peaks (Fig. 4A-B). As expected from their association with high H3K27ac signals, most CC-II sites also had abundant signals of other active histone marks, including H3K4me1, H3K4me3, with scarcity of H3K27me3 signals, suggesting their enrichment in active promoters and enhancers (23). Conspicuously, CC-II sites were highly enriched for Runx1-binding (Fig. 4A). In accordance with this, a subset (~10%) of Stag2/STAG2 signals were colocalized with Runx1/RUNX1 signals in immunofluorescence of mouse HSPCs and human K562 leukemia cell lines using super-resolution microscopy (Fig. 4C). Moreover, Runx1/RUNX1 and Stag2/STAG2 were shown to be co-immunoprecipitated with Smc1/SMC1 and Smc3/SMC3 in mouse (32Dcl3) and human (K562) myeloid leukemia cell lines (Supplementary Fig. S7C-D). Additionally,

according to the published ChIP-seq data from a murine HSPC cell line (HPC-7) (24), CC-II sites were highly enriched not only for Runx1 binding but also for binding of other TFs implicated in hematopoiesis as well as Asxl1 (25), RNA polymerase II (Pol II) and Med12 (26) (Supplementary Fig. S7E-F), which was also supported by an enrichment of consensus motifs of many hematopoietic TFs in CC-II sites compared with CC-I sites (Supplementary Fig. S7G). The distinct role of Stag2 at CC-II sites, as compared with at CC-I sites was further supported by ChIP-seq analysis in SKO-derived cells, in which only slight changes in Stag1 binding were observed at CC-II sites in contrast to remarkably increased binding of Stag1 at CC-I sites. This suggests that Stag1-cohesin does not replace Stag2-cohesin on CC-II sites, even in the face of Stag2 deficiency (Fig. 4D). Ctcf signals slightly decreased at CC-I sites but not at CC-II sites, consistent with specific binding of Ctcf at CC-I sites (Fig. 4D). These findings in ChIP-sequencing suggest a distinct role of Stag2 and Runx1 bindings at CC-II sites, in which active enhancers are enriched (Fig. 4E).

Stag2/Runx1 codeficiency disrupts enhancer-promoter loops

We next investigated the effects of Stag2 and/or Runx1 deletion on chromatin structures, using deep in situ Hi-C analysis of c-Kit$^+$ HSPCs from mice with different genotypes (27), which yielded 1.79 billion valid interactions per genotype on average. Overall, *Stag2/Runx1* deletions did not substantially affect boundaries of the large genomic structures that are known as compartment A/B, which largely correspond to transcriptionally active/inactive genomic regions, respectively (Supplementary Fig. S8A-C). However, insulations at boundaries of all topologically associating domains (TADs) were slightly enhanced in SKO and DKO cells, compared with those in WT cells (Supplementary Fig. S8D). Cohesin binding sites, CC-I and CC-II, were highly enriched at boundaries and inside regions of TADs, especially those located in compartment A ('A-TADs'), respectively (Fig. 5A and Supplementary Fig. S8E) and as expected, most of the DEGs in SKO-, RKO-, or DKO-derived LSK cells, as compared with WT-derived cells, were mapped within A-TADs (Fig. 5B). Thus, we calculated average difference in Hi-C contacts between mutant and WT cells, focusing on A-TADs (n=3,295) after their size was normalized (Fig. 5C). SKO-derived cells showed slightly reduced short-range contacts in the vicinity of the bottom line, while the contacts were slightly enhanced at the TAD corner. By contrast, chromatin contacts in the corresponding A-TADs exhibited a more profound and wide-spread reduction in DKO cells, even though RKO alone minimally influence intra-TAD chromatin contacts. When the analysis was stratified for TAD hierarchy, the contacts tended to be more attenuated in nested TADs than the parental TADs (Fig. 5D).

We also evaluated the effects of *Stag2/Runx1* KO on the number of chromatin loops. Consistent with relative spatial distribution of CC-I and CC-II sites (Supplementary Fig. S8E), CC-II-anchored loops were shorter than CC-I-anchored loops (Fig. 5E). The number of loops originating from CC-II sites was reduced in SKO-derived cells and further decreased in DKO-derived cells, whereas the effects on loops originating from CC-I sites remained minimum across the genotypes (Fig. 5F and Supplementary Fig. S8F-H). When these loops were characterized by the presence or absence of Ctcf (CC-I), promoter, and enhancer at loop anchors, we found that the loops between enhancer-promoter (E-P), promoter-promoter (P-P), and enhancer-enhancer (E-E) were more selectively disrupted in

SKO and DKO mice than other types of loops associated with Ctcf (Fig 5G-H). Moreover, these findings in the mouse model were confirmed in human leukemia cells using a series of isogenic AML cell lines derived from HL-60, which were targeted for *STAG2* and *RUNX1* using a CRISPR-Cas9 system, in which synergistic disruption of CC-II loops were recapitulated in *STAG2/RUNX1*-double knockout cells (Supplementary Fig. S9A-F).

To further understand the combinatorial effects on chromatin looping, we evaluated how the CC-II-anchored loops identified in WT cells are altered in SKO and DKO cells (Fig. 5I). As shown in Fig. 5J, the CC-II sites anchoring loops identified in WT cells (group (1)) were divided into four groups (groups (2)-(5)) depending on whether the loop at each site was lost or preserved in SKO and DKO cells, which were further investigated for the binding patterns of cohesin components, Pol II, Mediator, and ten hematopoietic transcription factors (TFs) at each group. Interestingly, we found relative enrichment of Stag2 at groups (2) and (4), strong enrichment of Pol II and Mediator at group (4), and relative enrichment of certain TFs at group (4) (Fig. 5K). In contrast, Stag1 enrichment was observed at group (5), i.e., CC-II sites with DKO-resistant loops. In addition, group (4) is bound by multiple TFs more frequently than other groups (Fig. 5L). These results suggest that, in the absence of Stag2, Runx1 deficiency additionally disrupts E-P loops anchored at Stag2 binding sites enriched for Pol II, Mediator, and TFs, leading to profound disruption of short-range chromatin interactions in DKO (Fig. 5M).

Transcriptional pausing underlies transcriptional vulnerability to *Stag2/Runx1* codeficiency

To see the effects of SKO and/or RKO on gene expression, we next investigated gene expression in LSK cells from these mice in greater details. On the basis of unsupervised clustering, we identified five discrete groups of genes (groups I-V) differentially expressed across four genotypes (Fig. 6A-B and Supplementary Fig. S10A). These groups were associated with distinct gene ontology terms and tissue- or hematopoietic lineage-specific gene expression profiles (Fig. 6C and Supplementary Fig. S10B-C), and group II and IV genes were synergistically upregulated and downregulated in DKO mice, respectively. These indicate that despite globally seen across the entire genome, loop attenuation does not necessarily result in a uniform change in overall gene expression. This is in agreement with the report that alteration in gene expression remained moderate and were seen only in a subset of genes, even when all loop structures were disrupted by complete loss of cohesin (28), suggesting that in general, the effect of chromatin looping on gene expression is modest and may be influenced by other contexts.

In this regard, it has been reported that super-enhancer (SE)-associated genes are more prone to downregulated expression in cohesin-deficient cells (28). In our mouse models, SE-associated genes also showed a trend of being downregulated in SKO and DKO cells, compared with RKO and WT cells, on average (Fig. 6D-E and Supplementary Fig. S10D). However, the effect of knockout was highly variable depending on SE site, compared with the case with typical enhancer (TE), particularly in DKO. This was exemplified for four SE-associated genes, *Hoxa9* (group IV), *Gata2* (group II), and *Fos/Fosb* (group II) (Fig. 6D). *Hoxa9* and other Hoxa cluster genes, which have long been implicated in the regulation of

normal hematopoiesis and leukemogenesis (29), were downregulated in SKO/DKO cells, where attenuated loops were observed (Fig. 6F and Supplementary Fig. S10E). Thus we further investigated the roles of Hoxa9, which was also down-regulated in shRNA mouse models targeting cohesin components, *Stag2* and *Smc1* (10) (Supplementary Fig. S11A). In serial replating assay, Hoxa9-overexpression did not prolong the replating for SKO/DKO cells, although the number of colonies increased substantially in Hoxa9-expressed cells. (Supplementary Fig. S11B). On the other hand, in in vitro single-cell differentiation assay, DKO-derived HSPCs showed a reduced frequency of erythroid-containing colonies compared to WT-derived HSPCs, which, however, was substantially rescued by Hoxa9-overexpression (Supplementary Fig. S11C-E). This suggests that the enhanced self-renewal in DKO is not attributed to *Hoxa9* down-regulation but mediated by other mechanisms, while down-regulated *Hoxa9* might contribute to differentiation block in these mice.

In contrast to Hoxa9 downregulation, even being associated with attenuated E-P loops, *Gata2* and *Fos/Fosb*, as well as other AP-1 components, showed upregulated expression (Supplementary Fig. S12A). Of note, in ATAC-seq analysis of DKO-derived LSK cells, consensus binding motifs of AP-1, Gata2, Runx1, IRFs, and Hoxa9 were enriched in enhanced ATAC-peaks frequently associated with group-II,V genes (AP-1 and Gata2) and attenuated ATAC-peaks frequently associated with group-IV genes (Runx1, IRFs, and Hoxa9), respectively (Fig. 6G-H and Supplementary Fig. S12B-C). Moreover, AP-1 and Gata2 motifs were highly enriched in promoter regions in group-II genes (Supplementary Fig. S12D). Thus, modulation of promoter activities by AP-1 and Gata2 may correspond to gene upregulation in group II genes by overriding detrimental effects of global loop suppression.

To further investigate the effect of E-P loop attenuation on gene expression, we next evaluated the link with transcriptional pausing, which has been known to correlate with enhancer activity and be implicated in the regulation of differentiation potential of stem cells (30-32). Of particular interest, in this regard, group-IV genes showed a high basal pausing level, as assessed by pausing index calculated by total Pol II ChIP-seq (Fig. 6I and Supplementary Fig. S13A). We separately confirmed that high-pausing genes were consistently and significantly downregulated in SKO and more profoundly decreased in DKO (Supplementary Fig. S13B). Degrees of pausing did not influence the expression specificity across diverse hematopoietic lineages (Supplementary Fig. S13C). These findings suggest a higher vulnerability of genes with high basal pausing to attenuated chromatin interactions caused by *Stag2/Runx1* deficiency. In addition, in coordination with reduced numbers of E-P loops in SKO and DKO cells across all DEG groups (Fig. 6J), we observed a substantial decrease in ChIP-seq intensities of Ser5-phosphorylated (Ser5-P) Pol II at the promoter proximal regions in SKO and DKO in all DEG groups, suggesting the global impact of E-P loop attenuation on promoter-proximal Pol II dynamics (Fig. 6K).

Explaining partial similarities of transcriptomes in SKO and RKO (Fig. 2N), group-IV genes were downregulated in SKO/RKO in a synergistic manner (Fig. 6A-B and Supplementary Fig. S10A), and were most strongly enriched for genes specifically expressed in HSCs across diverse hematopoietic lineages (Fig. 6C), suggesting a role of downregulated group-IV genes in the hematological phenotypes of Stag2/Runx1 deficiency. We also performed

RNA-seq analysis in HL-60 cell lines with *STAG2/RUNX1* deficiency. Although there were substantial differences in altered pathways in mouse LSK cells and HL-60 cells, we observed considerable overlaps of downregulated pathways in two experimental systems (Supplementary Fig. S14A-B).

Downregulation of high-pausing genes in *STAG2*/cohesin-mutated human AML/MDS

Finally we evaluated to what extent above findings in mice could be extended in human MDS/AML samples carrying *STAG2* and/or other *SRSA* mutations, as well as other cohesin mutations, using transcriptome data from three published cohorts of MDS/AML (6,33,34). An integrated pathway analysis of altered gene expression revealed a number of pathways changing in the same directions among SKO mouse model, cohesin-mutated MDS/AML cohorts, and HL-60 cell line model, including interferon response, regulation of leukocyte, adaptive immune responses, inflammatory response, ribosomal translation, and regulation of DNA/expression (Fig. 7A). Recapitulating the case with *Stag2/Runx1* deficient mouse models, expression of *HOXA9* and other HOXA cluster genes showed a uniform trend of downregulation with an increasing number of *SRSA* mutations in human samples (Fig. 7B). SE-associated genes identified in human normal HSPCs (35) were marginally downregulated in cohesin/*STAG2*-mutated MDS samples, as compared with TE-associated genes; however, this trend was less clear in AML samples (Supplementary Fig. S15A-B). We also evaluated the effect of basal level of transcriptional pausing on gene expression in *SRSA*-mutated samples, where the basal pausing level was calculated using total Pol II ChIP-seq of human normal HSPCs (36). High-pausing genes identified in mouse and human HSPCs were associated with several molecular pathways including interferon response and DNA repair response (Fig. 7C), consistent with downregulation of interferon and inflammatory responses in pathway network analysis (Fig. 7A). Conspicuously, high-pausing genes are consistently downregulated in cohesin- or *STAG2*-mutated samples in all three cohorts studied (Fig. 7D and Supplementary Fig. S15C-D). Moreover, high-pausing genes were preferentially downregulated in samples with 2 *SRSA* mutations, and *STAG2*-mutated cases with other *SRSA* mutations (Fig. 7E-F and Supplementary Fig. S15E). Taken together, these results support the relevance of transcriptional pausing and downregulation of high basal pausing genes in leukemogenesis not only of *Stag2/Runx1* DKO mice but also of human MDS/AML carrying multiple *SRSA* mutations.

Discussion

We have revealed a unique interplay between Stag2 and Runx1 in the maintenance of short-range chromatin interactions and transcriptional regulation, disruption of which leads to the development of MDS in mice. The interplay seems to depend on their localization to Ctcf-deficient, enhancer-proficient sites, in which the role of Stag2 cannot be replaced by Stag1. The distinct role of STAG2 in enhancer-rich regions is also suggested from the analysis of cultured cell lines in a recent report (37). Given the minimum changes with RKO alone, it is rather unexpected that Stag2/Runx1 DKO causes more profound chromatin alterations across the entire genome, compared with SKO alone. In the absence of Stag2, Runx1 loss additionally disrupts E-P loops anchored at cohesin bindings sites highly enriched for Pol II and Mediator. This suggests that such E-P loops vulnerable to combined loss of Stag2 and

Runx1 may be associated with formation of transcriptional condensates through phase separation capacities of Pol II, Mediator, and TFs, as emerged from several recent studies (38-40).

The effects of cohesin loss on whole transcriptomes are modest despite of the essential roles in regulation of looping (28). Of interest, genes showing high basal transcriptional pausing are more prone to downregulated expression, compared with genes with low transcriptional pausing. Such genes may require the assistance of intact E-P loop formation and relevant enhancer-bound TFs, such as Runx1, for their transcription/expression, as group-IV genes are more profoundly downregulated in DKO cells. The link between gene expression and transcriptional pausing is in contrast to a previous study (28) which highlighted the effect of cohesin-mediated looping on the expression of super-enhancer-associated genes as typically seen for Hoxa genes in DKO mice. Given that super-enhancer-associated genes show lower degree of pausing than typical enhancer-associated genes (Supplementary Fig. S13A) (32), transcriptional pausing should be considered as a novel gate keeper to determine selective expression changes, distinct from association with super-enhancers. High-pausing genes in HSPCs are associated with several molecular pathways including interferon response and DNA repair response. Thus, downregulation of high-pausing genes is consistent with downregulation of interferon and inflammatory responses at whole transcriptome levels and enrichment of IRFs in reduced ATAC peaks in SKO and DKO. These findings suggest that transcriptional pausing provides the molecular basis of a previously reported downregulation of interferon and inflammatory responses, which control self-renewal and differentiation of HSPCs, in cohesin-mutated AML and hematopoietic cells (41). Taken together, perturbation of both several super-enhancer-associated genes including *Hoxa9* and highly paused genes may contribute to multiple differentiation abnormalities and cell transformation in MDS/AML.

Distinct roles of Stag1 and Stag2 in HSPC self-renewal and differentiation in mouse models has been reported in a recent study (42). While Stag2-deficient mice display similar phenotypes in two studies, our study illustrates several novel findings: (1) association of mutations in *SRSA* genes; (2) mechanistic insights of combined effects of Stag2 and Runx1 loss on chromatin looping; and (3) the link between transcriptional pausing and selective gene dysregulation. Explaining the top significant associations between driver mutations, *SRSA* genes showed a conspicuous co-mutation pattern among MDS and related disorders, and also significantly co-mutated in a subset of primary AML, reported as AML with chromatin spliceosome mutation (43) or secondary-type primary AML (15), suggesting that these mutations define a unique subset of myeloid neoplasms, where the deregulated interplay of these genes in the maintenance of enhancer activity might explain the common pathogenesis. The significant downregulation of high-pausing genes in samples with cohesin/*STAG2* and other *SRSA* mutations in three independent cohort of MDS/AML strongly supports this idea. A previous study reporting the interaction between Asxl1 and cohesin (25) and our finding of highly enriched Asxl1 binding at CC-II sites in mice (Supplementary Fig. S7F) are also suggestive of this, as well as a recent report demonstrating the enhanced transcriptional pausing induced by mutated *SRSF2* (44). We did not observe no significant increase in abnormal splicing in *Stag2*/*Runx1* DKO cells, suggesting that myelodysplasia in DKO mice is not likely due to altered splicing

(Supplementary Fig. S15F). It is warranted in the future studies to better clarify the molecular basis of this unique subset of myeloid neoplasms.

Methods

Correlations between driver mutations in human MDS/AML

We analyzed in-house or publicly available data on large-scale genetic profiling of MDS/AML (2,3,14-18) for the investigation of correlations among major driver mutations commonly mutated in myeloid neoplasms. A total of 3,047 cases with MDS/AML were analyzed for correlations across 42 frequently mutated genes by Fisher's exact test as previously described (45). Patient survival was analyzed in 831 patients, for which data on survival was available (2). We analyzed number of driver mutations, variant allele frequencies of mutations, type of *RUNX1* mutations, and status of *STAG2* mutations in 76 cases which harbored more than one *STAG2* mutations, in a MDS cohort (2). The tumor cell fraction was calculated as previously described (46) with minor modifications. The tumor content was regarded as the maximum adjusted VAF among driver mutations in each case. Patient cohort is summarized in Supplementary Table S1. Clinical characteristics of MDS patients (2) with *STAG2*, *RUNX1*, *SRSF2*, and/or *ASXL1* mutations is summarized in Supplementary Table S2. Correlations between mutations in patients with MDS/AML are described in Supplementary Table S3.

Mice

Animal care was in accordance with institutional guidelines and approved by the Animal Research Committee, Graduate School of Medicine, Kyoto University (Kyoto, Japan). To generate the *Stag2* conditional knockout mouse model, the *Stag2*-targeting vector was constructed based on the *Cre-LoxP* system, in which two *LoxP* sites were inserted flanking exon 7, which encodes a STAG domain, and *Frt*-flanked neomycin selection cassette in a downstream intron. The linearized targeting vector was electroporated into murine C57BL/6 embryonic stem cells (RENKA strain) and were selected in G418. Homologous recombination was confirmed by direct sequencing of both outer regions of the neomycin cassette, and also by Southern blotting using neomycin-specific probes. Chimeric mice were produced by microinjection of targeted embryonic stem cells into blastocysts and were bred to C57BL/6 mice to establish germ line transmission (Trans Genic Inc., Kobe, Japan). The generated mice were initially crossed to a germline *Flp*-deleter (The Jackson Laboratory), to eliminate the neomycin cassette, and subsequently to the interferon-inducible *Mx1-Cre* transgenic mice (47). *Mx1-Cre* expression was induced by intraperitoneally injecting 500 μg of pIpC on six alternate days. Genotyping of *Stag2* floxed alleles were performed as previously described (48) using the following primers: 5'-GACCACTAAGCTCATAATCGC-3' and 5'- ATTTCTGGCTACTACGCTTGC-3'. *Runx1* conditional knockout mice were described previously (49). CD57BL/6 CD45.1$^+$ mice and CD57BL/6 F1-CD45.1$^+$/CD45.2$^+$ mice were purchased from Sankyo-Lab Service (Tsukuba, Japan).

Retroviral transduction

For Hoxa9 overexpression, we used FLAG-tagged Hoxa9 (50) in the pMSCV-neo retroviral vector in serial replating assay and pGCDNsam-IRES-EGFP retroviral vector in single-cell differentiation assay. Retroviruses were produced by transient transfection of Plat-E packaging cells with retroviral constructs. Purified c-Kit+ HSPCs were cultured in Iscove's modified Dulbecco's media (IMDM) containing 20% FBS and 50 ng/ml mouse SCF, mouse TPO, and human FLT-3L for 24 h, and then were rerto-transduced using Retronectin (Takara).

Serial replating assay

Freshly isolated 20,000 BM or 2,000 c-Kit+ cells were seeded into cytokine-supplemented methylcellulose medium (Methocult, M3434; STEMCELL Technologies) according to the manufacturer's protocol. For Hoxa9-overexpression assay, c-Kit+ cells transduced by pMSCV-neo vector were selected by adding G418 to the medium at the final concentration of 1 mg/ml. Colonies propagated in culture were scored at day 14 in duplicate. For replating, cells were resuspended in PBS, counted, and 10,000 cells were replated in duplicate once a week.

Flow cytometry

Single-cell suspensions were stained with monoclonal antibodies as previously described (48). Stained cells were analyzed with FACS Aria III or LSRFortessa X-10 flow cytometers (BD Bioscience, Franklin Lakes, NJ, USA). Cell sorting was performed with FACS Aria III. Cell cycle and apoptosis analysis were performed as previously described (48). Data were analyzed by FlowJo software (Tree Star, Ashland, CA, USA). The antibodies used in FACS experiments are described in Supplementary Table S4.

Isolation of mouse hematopoietic progenitors

Purification of HSPCs was performed as previously described (48). Briefly, freshly isolated BM cells were stained with APC-conjugated anti-c-Kit antibody and were enriched using anti-APC magnetic beads and MACS LS columns (Miltenyi Biotec) according to the manufacturer's instruction.

Bone marrow transplantation

In non-competitive BMT assay, unfractionated BM cells (2×10^6 cells) from CD45.2 donor mice were transplanted into 8-12 weeks-old female CD45.1 recipient mice which were lethally irradiated at 9.5 Gy dose. In competitive BMT assay, unfractionated BM cells (1×10^6 cells) from CD45.2 donor mice were transplanted into 8-12 weeks-old female CD45.1 recipient mice which were lethally irradiated at 9.5 Gy dose, together with the equal number of BM cells from CD45.1/CD45.2 competitor mice. In the BMT experiments, pIpC was injected at 4 weeks after BMT. The donor chimerism in PB was evaluated at 4, 8, 12, 16, and 20 weeks after transplantation by flow cytometry. Mice were sacrificed at 20 weeks after transplantation, and the chimerism in each BM fraction was assessed by flow cytometry.

Generation of CRISPR knockout cell lines

Human AML HL-60 cell lines were obtained from American Type Culture Collection and not tested for Mycoplasma contamination. sgRNAs targeting *STAG2*/*RUNX1* were designed, and knockout efficiency was confirmed by Guide-it Mutation Detection Kit (Takara). To express SpCas9, HL-60 cell line was transduced by lentiCas9-Blast vector (Addgene #52962) and selected by blasticidin at the concentration of 5 μg/ml. To generate WT, SKO, RKO, and DKO cells, cells were transduced by pLKO5-sgRNA-EFS-GFP vector (Addgene #57822) containing sgRNA targeting STAG2 or non-targeting control sgRNA and pLKO5-sgRNA-EFS-tRFP vector (Addgene #57823) containing sgRNA targeting RUNX1 or non-targeting control sgRNA, and GFP$^+$/RFP$^+$ cells were sorted at single cell per well into a 96-well plate. Generated colonies were genotyped by amplicon sequencing using iSeq 100 (Illumina). All clones for SKO, RKO, and DKO were subjected to Western blotting to confirm the knockout of STAG2 and/or RUNX1. PCR primers and sgRNA sequences are provided in Supplementary Table 5, and genotypes of generated clones are described in Supplementary Table 6.

RNA-sequencing

RNA was extracted using RNeasy Mini Kit (QIAGEN) or NucleoSpin RNA XS (Macherey-Nagel). Libraries for RNA-seq were prepared using the NEBNext Ultra RNA Library Prep kit for Illumina (New England BioLabs) and were subjected to sequencing using HiSeq 2500 or NovaSeq 6000 instrument (Illumina) with a standard 100-150-bp paired-end protocol as previously described (51). RNA-seq experiments were performed in two or more biological replicates. The sequencing reads were aligned to the reference genome (hg19 or mm9) using STAR (v2.5.3) (52). Reads on each refSeq gene were counted with featureCounts (v1.5.3) (53) from Subread package, and edgeR package in R (54) was used to identify the differentially expressed genes with FDR threshold of 0.05 and to generate the multidimensional scaling (MDS) plot. The analysis was performed in genes expressed at >1 count per million (CPM) in two or more samples, and generalized linear models were used to compare gene expression data. MSigDB overlap analysis was performed using MSigDB database and hallmark gene sets (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). RNA expression analysis in the hematopoietic system was carried out using Haemopedia RNA-seq datasets (55), and averages of log2 (TPM + 1) values of each gene set were calculated for each cell type. GSEA (v2.2.4) (56) was used to determine the sets of genes that are significantly different between groups. Enriched pathways were visualized using Enrichment Map (57) with the q-value < 0.05 and overlap similarity coefficient parameters > 0.5. RNA-seq datasets used are described in Supplementary Table S7. Differentially expressed genes shown in Fig. 2I are described in Supplementary Tables S8-9. Differentially expressed genes shown in Fig. 3J are described in Supplementary Tables S10-15.

More detailed methods are described in the Supplementary Methods.

ATAC-sequencing

ATAC-seq experiments were performed using LSK or CMP cells obtained from WT-, SKO-, RKO-, or DKO-transplanted mice using Fast-ATAC protocol as previously described (58) with minor modifications. Briefly, freshly isolated 10,000 cells were pelleted and fifty

microliters of transposase mixture (25 ul of 2 x TD buffer, 2.5 ul of TDE1, 0.5 ul of 1% digitonin, and 22 ul of nuclease-free water) (FC-121-1030, Illumina; G9441) was added to the cells. After transposition reactions at 37°C for 30 min, transposed DNA was purified using QIAGEN MinElute Reaction Cleanup kit. Transposed fragments were PCR-amplified, and the resulting library was sequenced on Hiseq 2500. ATAC-seq experiments were performed in biological duplicates. Reads were aligned to the mouse mm9 reference genome using bowtie2 (v2.3.3) (59) with -X 2000 –no-mixed –very-sensitive parameters following adapter trimming using cutadapt (v1.14) (60). Duplicates were removed by Picard (v2.6.0) (https://broadinstitute.github.io/picard/), and reads on mitochondria genome or blacklisted regions (ENCODE) were removed by bedtools (v2.27.1) (61). Peaks were called with MACS (v2.1.1) (62) with –nomodel –broad parameters with a q-value threshold of 1×10^{-5} for individual replicate as well as merged data of all replicates. Read counts on peaks for merged data were counted with the multicov function in bedtools, and edgeR was used to identify the peaks with differential accessibility with FDR threshold of 0.05. Transcription factor motifs were discovered with HOMER (63) in differentially accessible sites (up or down, compared with WT) using stable peaks or random genome as backgrounds. Peak annotation was performed with HOMER. Differentially accessible ATAC regions in SKO-derived LSK and CMP cells (vs WT) shown in Fig. 2 are described in Supplementary Tables S16-19.

ChIP-sequencing

ChIP-seq experiments were performed using c-Kit[+] HSPCs or HL-60 cell lines. Cells were fixed in PBS with 1% formaldehyde (Thermo Fisher Scientific) for 10 min at room temperature with gentle mixing. The reaction was stopped by adding glycine solution (10x) (Cell Signaling Technology) and incubating for 5 minutes at room temperature, and the cells were washed in cold PBS twice. The cells were then processed with SimpleChIP Plus Sonication Chromatin IP Kit (Cell Signaling Technology) and Covaris E220 (Covaris) according to the manufacturer's protocol. After purification of ChIPed DNA, ChIP-seq libraries were constructed using ThruPLEX DNA-seq kit (Takara) according to the manufacturer's protocol, and then subjected to sequencing using HiSeq 2500 or NoveSeq 6000 (Illumina). ChIP-seq experiments were performed in two or more biological replicates with input controls. The sequencing reads were aligned to the reference genome (hg19 or mm9) using bowtie (v1.2.2) (64) following trimming of adapters and read tails to a total length of 50 base pairs using cutadapt. Duplicates and reads on blacklisted regions (ENCODE) were removed by Picard and bedtools, respectively. Peaks were called using MACS (v2.1.1) for each replicate individually with a *P*-value threshold of 1×10^{-3} unless otherwise specified, and overlapped peaks among replicates were regarded as consensus peak sets. Motif analysis and peak annotation were performed with HOMER. Super-enhancers were identified with H3K27ac ChIP-seq data in WT HSPCs using ROSE (65) with default parameters. Identified super-enhancers are described in Supplementary Table S20. Super-enhancers in human HSCs were previously described (35), and we used the dataset of BI_CD34_Primary_RO01536 for assignment of super-enhancer-associated genes. Calculation of ChIP signal intensities around peaks and generation of read density profile plots and heatmaps were performed using deeptools (66). For clustering of cohesin binding sites, we calculated logarithm of H3K27ac and Ctcf ChIP signal intensities summed up

around ± 200 bp from centers of cohesin binding sites (Stag1 and/or Stag2 peaks) by deeptools, performed clustering using flowPeaks (67), and regarded the H3K27ac high clusters as cohesin cluster II and the others as cluster I. Binding profiles of cohesin components, Pol II, Mediator, and ten hematopoietic TFs were similarly calculated around ± 200 bp from centers of cohesin binding sites (Fig. 5K). We also used ChIP-seq datasets (10 TFs: Asxl1, Fli1, Gata2, Gfi1b, Lmo2, Lyl1, Meis1, Pu1, Runx1, Scl; Pol II, Med12 in mouse, Pol II in human (24-26,36)) in previous studies. More detailed methods are described in the Supplementary Methods.

Pol II pausing analysis

For total Pol II ChIP-seq, two replicates were merged and ChIP-seq peaks were called using MACS (v1.4.2) with a P-value threshold of 1×10^{-3}. For analysis of pausing indices in mice, expressed genes, whose transcription start sites (TSSs) overlapped with Pol II ChIP-seq peaks in WT HSPCs, were subjected to downstream analyses. Pausing indices were calculated as the input-subtracted read density in the promoter-proximal region (−50 bp to +300 bp around TSS) divided by that of the gene body (from +300 bp to 2 kb downstream of the end of gene). Genes with positive signals in both the promoter-proximal region and gene body were further considered. ChIP-seq intensities of Ser5-P Pol II in the promoter proximal region were similarly calculated. Pol II ChIP-seq in human CD133-positive cells was previously described (36), and ChIP-seq peaks were called using MACS (v1.4.2) with a P-value threshold of 1×10^{-5}. For pausing analysis in human, input subtraction was not performed and genes, whose TSSs overlapped with Pol II ChIP-seq peaks, were considered. Statistical significance was assessed with one-sided Wilcoxon rank-sum test.

Hi-C

Hi-C experiments were performed using MboI restriction enzyme as previously described (27). Briefly, two million cells were crosslinked with 1% formaldehyde for 10 min at room temperature. Cells were permeabilized and chromatin was digested with MboI restriction enzyme, and the ends of restriction fragments were labeled with biotinylated nucleotides and ligated. After crosslink reversal, DNA was purified and sheared with Covaris M220 (Covaris). Then point ligation junctions were pulled down with streptavidin beads. Then libraries were constructed with Nextera Mate Pair Sample Preparation Kit (Illumina) according to the manufacturer's protocol, and subject to sequencing using NovaSeq 6000 (Illumina) with a standard 100- or 150-bp paired-end protocol. Hi-C experiments were performed in biological duplicates. The sequencing reads were processed using Juicer (27) and hg19 or mm10 reference genome. After filtering of reads, the average valid interactions per genotype resulted in 1.79 billion for mouse HSPCs and 1.66 billion for HL-60 cells. For comparative analysis, the valid interactions after filtering were randomly resampled and arranged in the number of the lowest sample. Contact matrices used for further analysis were created for each replicate as well as merged one by genotype and Knight-Ruiz (KR)-normalized with Juicer. Loops were called at 5kb and 10kb resolutions using HICCUPS (27) and then merged to construct loop sets. Loops were classified into CC-I loops (whose anchors overlapped with at least one CC-I sites but not with CC-II) and CC-II loops (whose anchors overlapped with at least one CC-II but not CC-I sites). Topologically associating domains (TADs) were called at 5kb resolution using Arrowhead (27). TAD boundaries were

defined as +- 5kb from the 5'- or 3'- ends of TADs, and insides were regions insides of both boundaries. More detailed methods are described in the Supplementary Methods.

Statistical analysis

In human MDS/AML, correlations across 42 frequently mutated genes were assessed by Fisher's exact test. For mouse phenotype analysis, we calculated P-values with two-tailed unpaired Student's t test for two group comparison, or ordinary one-way analysis of variance (ANOVA) with Bonferroni analysis for three or more group comparison using GraphPad Prism (v6). For survival analysis, survival was estimated using the Kaplan-Meier method, and groups were compared using the log-rank test, with survival package in R software. In RNA-seq, ATAC-seq, and ChIP-seq analyses, statistical analyses were performed using edgeR, HOMER, DAVID, TSEA, GSEA, or MACS. For metaplot analysis, bin-wise P-values were obtained using one-sided Wilcoxon rank-sum test. For splicing analysis, differential PSI was assessed using two-sided moderated t-test and Benjamini-Hochberg correction. For pausing analysis and RNA expression analysis according to pausing indices, statistical significance was assessed with one-sided Wilcoxon rank-sum test.

Data availability

External ChIP-seq datasets used in this study are summarized in Supplementary Table S21. The sequencing data generated in this study are available in the GEO repository, under accession number GSE131583. All other data will be made available upon request to the corresponding author.

Additional Methods

Detailed methods for quantitative reverse transcription PCR (qRT-PCR), Western blotting, co-immunoprecipitation, histology and cytology, immunostaining, microscopy and data analysis, Single-cell differentiation assay, and splicing analysis are described in the Supplementary Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. Blood 2013;122:4021–34. [PubMed: 24136165]

2. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 2014;28:241–7. [PubMed: 24220272]

3. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122:3616–27; quiz 99. [PubMed: 24030381]

4. Ogawa S Genetics of MDS. Blood 2019;133:1049–59. [PubMed: 30670442]

5. Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y, et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet 2013;45:1232–7. [PubMed: 23955599]

6. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059–74. [PubMed: 23634996]

7. Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat Genet 2013;45:1293–9. [PubMed: 24056718]

8. Remeseiro S, Losada A. Cohesin, a chromatin engagement ring. Curr Opin Cell Biol 2013;25:63–71. [PubMed: 23219370]

9. Viny AD, Ott CJ, Spitzer B, Rivas M, Meydan C, Papalexi E, et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. J Exp Med 2015;212:1819–32. [PubMed: 26438361]

10. Mullenders J, Aranda-Orgilles B, Lhoumaud P, Keller M, Pae J, Wang K, et al. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. J Exp Med 2015;212:1833–50. [PubMed: 26438359]

11. Mazumdar C, Shen Y, Xavy S, Zhao F, Reinisch A, Li R, et al. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. Cell Stem Cell 2015;17:675–88. [PubMed: 26607380]

12. Galeev R, Baudet A, Kumar P, Rundberg Nilsson A, Nilsson B, Soneji S, et al. Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs. Cell Rep 2016;14:2988–3000. [PubMed: 26997282]

13. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet 2018;19:789–800. [PubMed: 30367165]

14. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med 2012;366:1090–8. [PubMed: 22417201]

15. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood 2015;125:1367–76. [PubMed: 25550361]

16. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H, et al. Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet 2017;49:204–12. [PubMed: 27992414]

17. Yoshizato T, Nannya Y, Atsuta Y, Shiozawa Y, Iijima-Yamashita Y, Yoshida K, et al. Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. Blood 2017;129:2347–58. [PubMed: 28223278]

18. Meggendorfer M, de Albuquerque A, Nadarajah N, Alpermann T, Kern W, Steuer K, et al. Karyotype evolution and acquisition of FLT3 or RAS pathway alterations drive progression of myelodysplastic syndrome to acute myeloid leukemia. Haematologica 2015;100:e487–90. [PubMed: 26294738]

19. Hu L, Li M, Ding Y, Pu L, Liu J, Xie J, et al. Prognostic value of RDW in cancers: a systematic review and meta-analysis. Oncotarget 2017;8:16027–35. [PubMed: 27926498]

20. Pietras EM, Reynaud D, Kang YA, Carlin D, Calero-Nieto FJ, Leavitt AD, et al. Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. Cell Stem Cell 2015;17:35–46. [PubMed: 26095048]

21. de Bruijn M, Dzierzak E. Runx transcription factors in the development and function of the definitive hematopoietic system. Blood 2017;129:2061–9. [PubMed: 28179276]

22. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–8. [PubMed: 24097267]

23. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 2014;15:272–86. [PubMed: 24614317]

24. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell 2010;7:532–44. [PubMed: 20887958]

25. Li Z, Zhang P, Yan A, Guo Z, Ban Y, Li J, et al. ASXL1 interacts with the cohesin complex to maintain chromatid separation and gene expression for normal hematopoiesis. Sci Adv 2017;3:e1601602. [PubMed: 28116354]

26. Aranda-Orgilles B, Saldana-Meyer R, Wang E, Trompouki E, Fassl A, Lau S, et al. MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. Cell Stem Cell 2016;19:784–99. [PubMed: 27570068]

27. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80. [PubMed: 25497547]

28. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin Loss Eliminates All Loop Domains. Cell 2017;171:305–20 e24. [PubMed: 28985562]

29. Collins CT, Hess JL. Deregulation of the HOXA9/MEIS1 axis in acute leukemia. Curr Opin Hematol 2016;23:354–61. [PubMed: 27258906]

30. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 2014;512:96–100. [PubMed: 25043061]

31. Williams LH, Fromm G, Gokey NG, Henriques T, Muse GW, Burkholder A, et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol Cell 2015;58:311–22. [PubMed: 25773599]

32. Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, et al. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev 2018;32:26–41. [PubMed: 29378787]

33. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. Nature 2018;562:526–31. [PubMed: 30333627]

34. Shiozawa Y, Malcovati L, Galli A, Pellagatti A, Karimi M, Sato-Otsubo A, et al. Gene expression and risk of leukemic transformation in myelodysplasia. Blood 2017;130:2642–53. [PubMed: 29097382]

35. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell 2013;155:934–47. [PubMed: 24119843]

36. Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, et al. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell 2009;4:80–93. [PubMed: 19128795]

37. Kojic A, Cuadrado A, De Koninck M, Gimenez-Llorente D, Rodriguez-Corsino M, Gomez-Lopez G, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. Nat Struct Mol Biol 2018;25:496–504. [PubMed: 29867216]

38. Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science 2018;361:412–5. [PubMed: 29930094]

39. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, et al. Coactivator condensation at super-enhancers links phase separation and gene control. Science 2018;361.

40. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 2018;175:1842–55 e16. [PubMed: 30449618]

41. Cuartero S, Weiss FD, Dharmalingam G, Guo Y, Ing-Simmons E, Masella S, et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. Nat Immunol 2018;19:932–41. [PubMed: 30127433]

42. Viny AD, Bowman RL, Liu Y, Lavallee VP, Eisman SE, Xiao W, et al. Cohesin Members Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and Differentiation. Cell Stem Cell 2019;25:682–96 e8. [PubMed: 31495782]

43. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374:2209–21. [PubMed: 27276561]

44. Chen L, Chen JY, Huang YJ, Gu Y, Qiu J, Qian H, et al. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. Mol Cell 2018;69:412–25 e6. [PubMed: 29395063]

45. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nat Genet 2015;47:458–68. [PubMed: 25848751]

46. Ochi Y, Hiramoto N, Yoshizato T, Ono Y, Takeda J, Shiozawa Y, et al. Clonally related diffuse large B-cell lymphoma and interdigitating dendritic cell sarcoma sharing MYC translocation. Haematologica 2018;103:e553–e6. [PubMed: 30237263]

47. Kuhn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. Science 1995;269:1427–9. [PubMed: 7660125]

48. Kon A, Yamazaki S, Nannya Y, Kataoka K, Ota Y, Nakagawa MM, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood 2018;131:621–35. [PubMed: 29146882]

49. Ichikawa M, Asai T, Saito T, Seo S, Yamazaki I, Yamagata T, et al. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. Nat Med 2004;10:299–304. [PubMed: 14966519]

50. Yokoyama A, Ficara F, Murphy MJ, Meisel C, Hatanaka C, Kitabayashi I, et al. MLL becomes functional through intra-molecular interaction not by proteolytic processing. PLoS One 2013;8:e73649. [PubMed: 24040009]

51. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478:64–9. [PubMed: 21909114]

52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. [PubMed: 23104886]

53. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30. [PubMed: 24227677]

54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40. [PubMed: 19910308]

55. Choi J, Baldwin TM, Wong M, Bolden JE, Fairfax KA, Lucas EC, et al. Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans. Nucleic Acids Res 2019;47:D780–D5. [PubMed: 30395284]

56. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50. [PubMed: 16199517]

57. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5:e13984. [PubMed: 21085593]

58. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 2016;48:1193–203. [PubMed: 27526324]

59. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9. [PubMed: 22388286]

60. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 2011;17:10–2.

61. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2. [PubMed: 20110278]

62. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137. [PubMed: 18798982]

63. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–89. [PubMed: 20513432]

64. Langmead B Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics 2010;Chapter 11:Unit 11 7.

65. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 2013;153:320–34. [PubMed: 23582323]

66. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–91. [PubMed: 24799436]

67. Ge Y, Sealfon SC. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. Bioinformatics 2012;28:2052–8. [PubMed: 22595209]

Significance

We demonstrate a critical role of an interplay between Stag2 and a master transcription factor of hematopoiesis, Runx1, in MDS development, and further reveal their contribution to regulation of high-order chromatin structures, particularly enhancer-promoter looping, and the link between transcriptional pausing and selective gene dysregulation caused by cohesin deficiency.
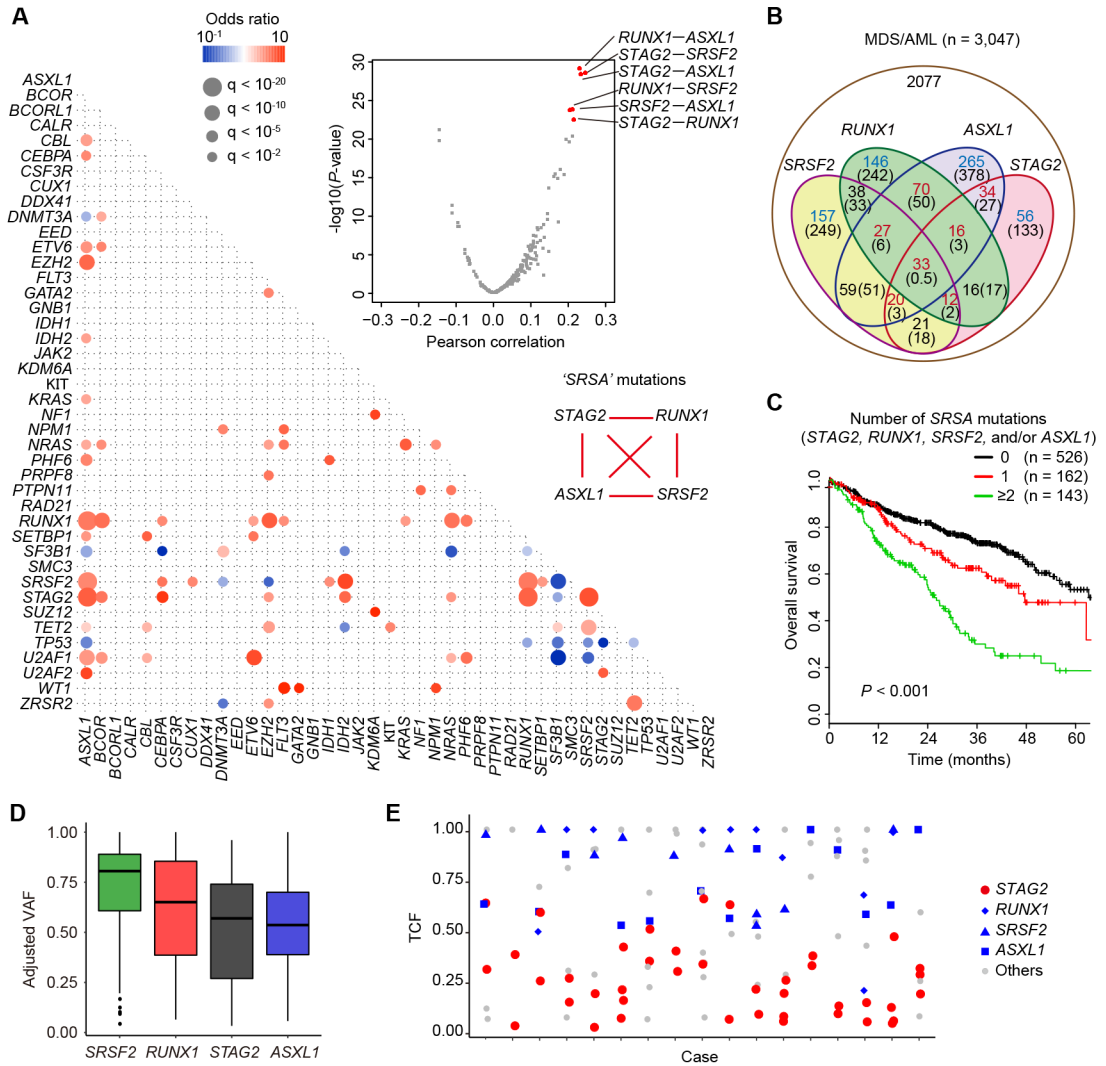
Figure 1. *STAG2* and associated mutations in human MDS/AML.

A, Correlations between driver mutations in MDS/AML. Left panel: Significantly co-occurring and mutually exclusive mutations are shown in red and blue circles, respectively. Odds ratio and associated q-values are indicated by the color gradient and size of circles, respectively. Right upper panel: Volcano plot showing the relationship of Pearson correlation values and corresponding −log10(*P*-value) between any pairs of the co-occurring mutations found in more than five cases. *P*-values were calculated by Fisher's exact test. **B**, Venn diagram showing the overlaps of '*SRSA*' mutations (*STAG2, RUNX1, SRSF2,* and *ASXL1*) in MDS/AML cases. The numbers of cases are indicated in red or blue colors, in which >20% increase or decrease are observed compared with the expected numbers by chance as shown in parenthesis, respectively. **C**, Kaplan–Meier estimates of overall survival according to the number of *SRSA* mutations. *P*-value was calculated by log-rank test. **D**, Adjusted VAF values of *SRSA* mutations. **E**, Tumor cell fractions (TCFs) of indicated driver mutations are shown for the patients harboring two or more different *STAG2* mutations.
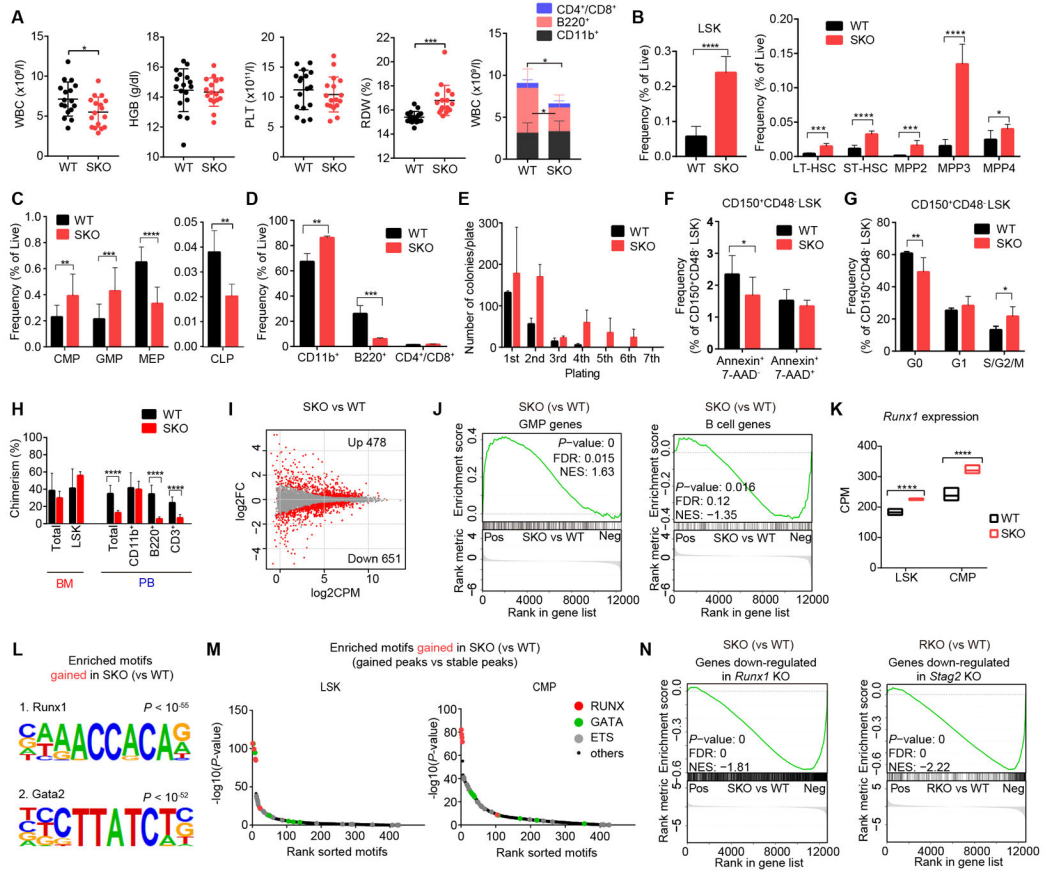
Figure 2. Stag2 depletion alters HSC self-renewal and differentiation in mice.

A, White blood cell (WBC) count, hemoglobin (HGB) level, platelet (PLT) count and red cell distribution width (RDW) in the peripheral blood (PB) of wild-type (WT) and *Stag2* conditional knockout (SKO) littermate male mice are plotted as dots (n = 17), in which the mean ± standard deviation (SD) are indicated as bars (left panels). Number of granulocytes/monocytes (CD11b$^+$), B-lymphocytes (B220$^+$) and T-lymphocytes (CD4$^+$/CD8$^+$) in the PB of WT and SKO mice (mean ± SD, n = 10) are shown in the right panel. **B**, Frequency of lineage (Lin)-negative/Sca1$^+$/c-Kit$^+$ (LSK) cells (left panel), and frequencies of long-term HSC (LT-HSC), short-time HSC (ST-HSC), multipotent progenitor (MPP)-2, MPP-3, and MPP-4 fractions in the BM of WT or SKO mice (mean ± SD, n = 6) (right panel) are shown. **C**, Frequencies of common myeloid progenitors (CMPs), granulocyte-macrophage progenitors (GMPs), megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs) and common lymphoid progenitors (CLPs) in the BM of WT and SKO mice (mean ± SD, n = 6). **D**, Frequencies of each lineage-committed cells in the BM of WT and SKO mice (mean ± SD, n = 4). **E**, Colony counts in methylcellulose replating experiments using nucleated BM cells from WT or SKO mice (mean ± SD, n = 2) are shown. BM cells were plated in duplicate at a density of 20,000 cells/plate for the first plating and 10,000 cells/plate for replating. **F**, Frequency of apoptotic cells (Annexin$^+$/7-AAD$^-$) in CD150$^+$/CD48$^-$ LSK cells (n = 6, mean ± SD). **G**, Frequency of cycling cells (S/G2/M; Ki-67$^+$/Hoechst$^+$), quiescent cells (G0; Ki-67-/Hoechst-), and G1 cells (Ki-67$^+$/Hoechst$^-$) in CD150$^+$/CD48$^-$ LSK cells (n = 5, mean ± SD). **H**, Percentages of CD45.2$^+$ donor cells within each fraction of the BM

or PB after competitive BM transplantation (16 weeks after pIpC injection) are shown (mean ± SD, n = 10 for WT and 6 for SKO). **I**, MA plot showing the transcriptional changes between WT- and SKO-derived LSK cells. Differentially expressed genes (DEGs) (FDR < 0.05) are indicated by red color. FC, fold-change. **J**, Gene set enrichment analysis (GSEA) between WT- and SKO-derived LSK cells, showing a significant enrichment of genes characteristic of GMPs and B-lymphocytes. Nominal *P*-value, false discovery rate (FDR), and normalized enrichment score (NES) are indicated. **K**, Expression levels of *Runx1* in LSK and CMP fractions are indicated by counts per million mapped reads (CPM) (min to max values with mean, n = 3). *P*-values were calculated using edgeR package in R software. **L,** Motifs and corresponding *P*-values identified by de novo motif search in ATAC-seq peaks that gained accessibility in SKO-derived LSK cells. **M**, Enrichment of known transcription factor (TF) motifs in ATAC-seq peaks that gained accessibility in SKO-derived LSK (left panel) and CMP cells (right panel). The sorted motif rank and $-\log10(P\text{-value})$ of a motif enrichment test using stable peaks as backgrounds is indicated in horizontal and vertical axis, respectively. **N**, GSEA analysis between SKO- and WT-derived LSK cells, showing a negative enrichment of genes down-regulated in *Runx1* conditional knockout (RKO)-derived LSK cells compared with WT (left panel), and GSEA analysis between RKO- and WT-derived LSK cells, showing a negative enrichment of genes down-regulated in SKO-derived LSK cells compared with WT (right panel). For panels (**A-G**), mice were analyzed at 12-24 weeks of age. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. Two-tailed unpaired Student's t-test in (**A-H**).
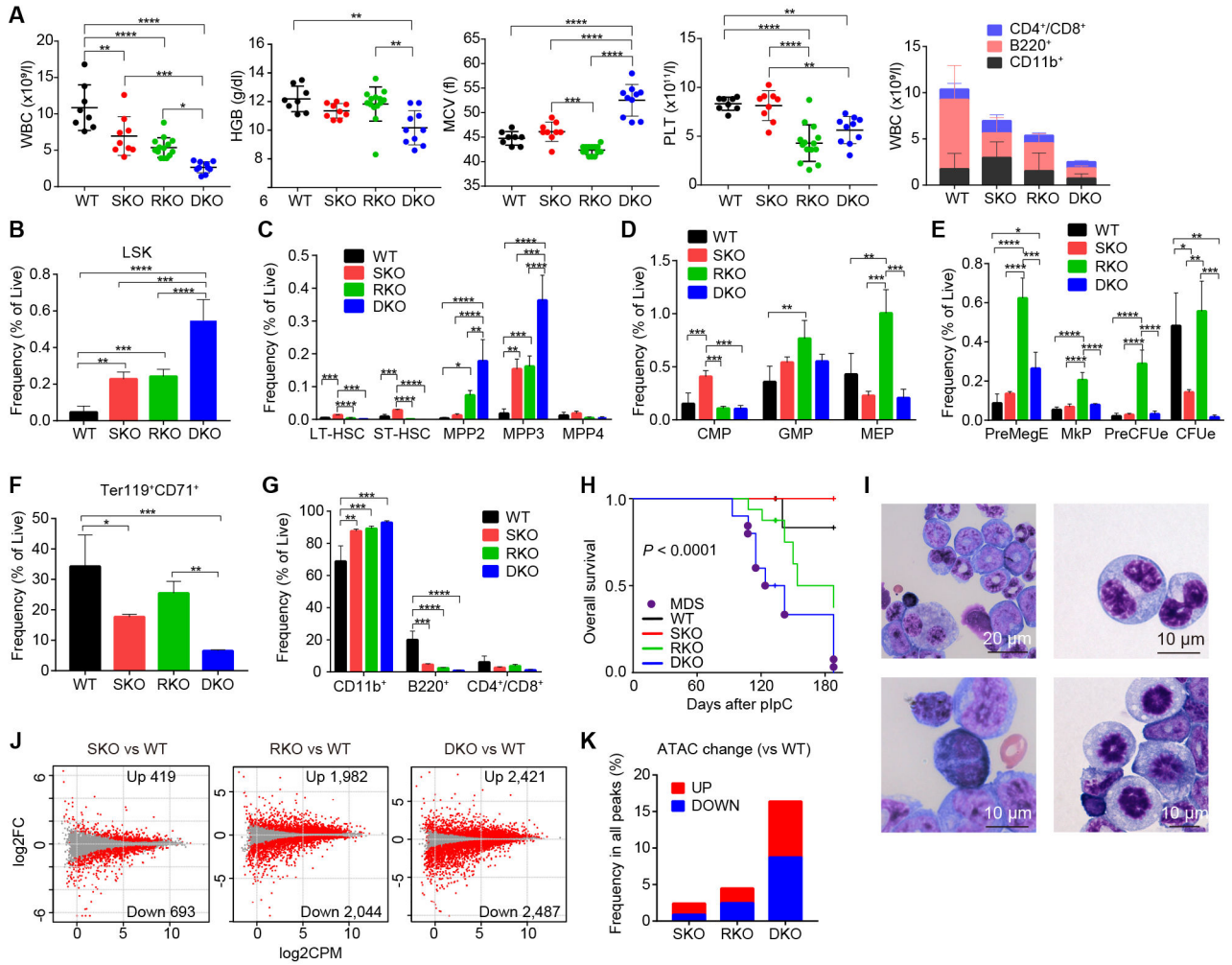
Figure 3. *Stag2/Runx1* double knockouts induce MDS in mice.
A, WBC, HGB, mean corpuscular volume (MCV), and PLT count in the PB of recipient mice transplanted with BM cells of WT, SKO, RKO, or *Stag2/Runx1* double conditional knockout (DKO) mice are plotted as dots (n = 8 for WT, 9 for SKO, 14 for RKO and 10 for DKO), in which the mean ± SD are indicated as bars (left panels). Number of granulocytes/ monocytes (CD11b+), B-lymphocytes (B220+), and T-lymphocytes (CD4+/CD8+) in the PB of WT-, SKO-, RKO-, and DKO- transplanted mice are shown in the right panel (mean ± SD, n = 9 for WT and SKO, 14 for RKO, and 4 for DKO). **B-G**, Frequencies of HSPCs (**B-C**), myeloid progenitors (**D**), megakaryocyte/erythroid progenitors (**E**), erythroblasts (**F**), and lineage-committed cells (**G**) in the BM are shown (mean ± SD, n = 5 for WT and RKO, and 3 for SKO and DKO). PreMegE, pre-megakaryocyte-erythroid progenitors; MkP, megakaryocytic progenitors; PreCFUe, pre-colony-forming unit erythroid cells; CFUe, colony-forming unit erythroid cells. **H**, Kaplan–Meier estimates of overall survival for each genotype (n = 9 for WT and SKO, 16 for RKO, and 10 for DKO). *P*-value was calculated by log-rank test. Death due to MDS is indicated by the purple circle. **I**, Representative May-Grünwald-Giemsa staining of BM cells showing dysplastic features, including pseudo-Pelger–Huët anomalies in neutrophils, binucleated megakaryocytes or erythroblasts, and

abnormal mitosis. **J**, MA plot showing the transcriptional changes in LSK cells derived from SKO, RKO, and DKO mice compared with WT-derived LSK cells. DEGs (FDR < 0.05) are indicated by red color. **K**, Frequency of differentially accessible ATAC peaks for SKO-, RKO- and DKO-derived LSK cells compared with WT. In panels (**A-G**), mice were analyzed 16-20 weeks after pIpC injection. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. *P*-values were calculated by ordinary one-way ANOVA with Bonferroni analysis in (**A-G**).
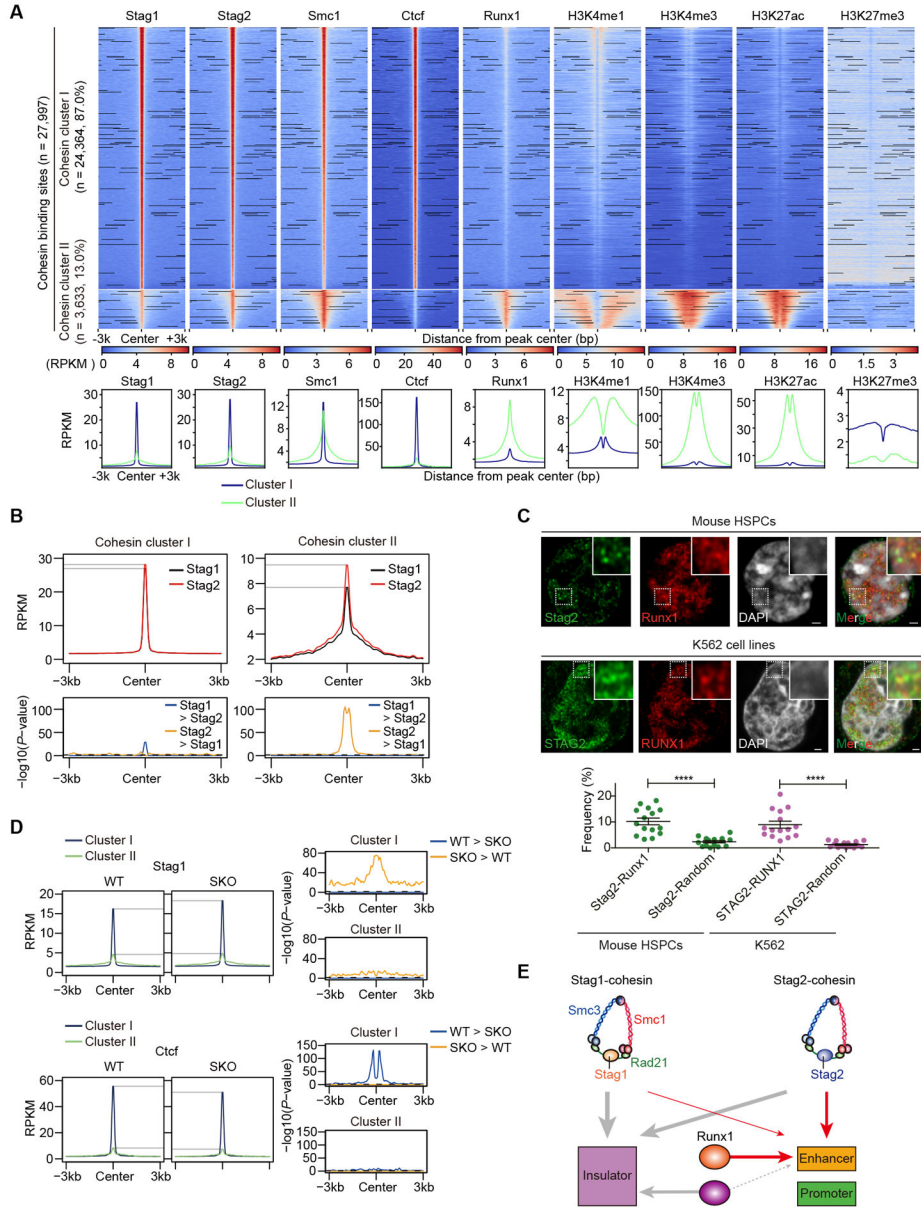
Figure 4. Colocalization of Stag2-cohesin and Runx1 at enhancers.
A, Upper panels: ChIP-seq density heatmap of cohesin components (Stag1, Stag2, and Smc1), Ctcf, Runx1, and histone marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in c-Kit[+] HSPCs of WT mice centered on Stag1- and/or Stag2-cohesin binding sites (n = 27,997) are depicted in descending order of Stag2 peak intensities, in which cohesin binding sites were divided into two clusters (cohesin cluster-I (CC-I) and cohesin cluster-II (CC-II)) according to the ChIP signals for Ctcf and H3K27ac (see also Supplementary Fig. S7A-B). Color scales below the heatmaps indicate ChIP-seq intensities (reads per kilobase per million mapped reads (RPKM)). Lower panels: Average ChIP-seq read intensity plot for CC-I (blue) and CC-II (green) distribution around the cohesin binding sites. **B**, Average ChIP-seq read intensities of Stag1 or Stag2 around CC-I or CC-II sites (upper panels) and *P*-

values for comparison between Stag1 and Stag2 across each bin (lower panels). **C**, Super-resolution images of Stag2/Runx1 localization at the nucleus in a mouse c-Kit$^+$ HSPC (upper panels) and STAG2/RUNX1 localization at the nucleus in a K562 cell line (middle panels). The dotted white box indicates the magnified region shown in the inset (Scale bars: 1μm). The images were obtained using a LSM880 Airy scan super-resolution microscope (Zeiss). Lower panel: Quantification of the colocalization of Stag2-Runx1 in mouse c-Kit$^+$ HSPCs and colocalization of STAG2-RUNX1 in K562 cell lines. The dots indicate the percentages of the areas of Stag2 (STAG2)-Runx1 (RUNX1) double positive spots among total areas of Stag2 (STAG2) positive spots. ****p<0.0001, two-sided Wilcoxon rank-sum test (n = 15 from three biological replicates). **D**, Average ChIP-seq read intensities of Stag1 and Ctcf in WT- and SKO-derived HSPCs around CC-I (blue) and CC-II (green) sites (left panels) and *P*-values for comparison between WT and SKO across each bin (right panels). **E**, Schematic representation representing the preferential binding of Stag2-cohesin to active enhancers together with Runx1. *P*-values were calculated by one-sided Wilcoxon rank-sum test comparing the ChIP-intensities in each bin in (**B**) and (**D**). Horizontal dashed lines indicate *P* = 0.05 in (**B**) and (**D**).
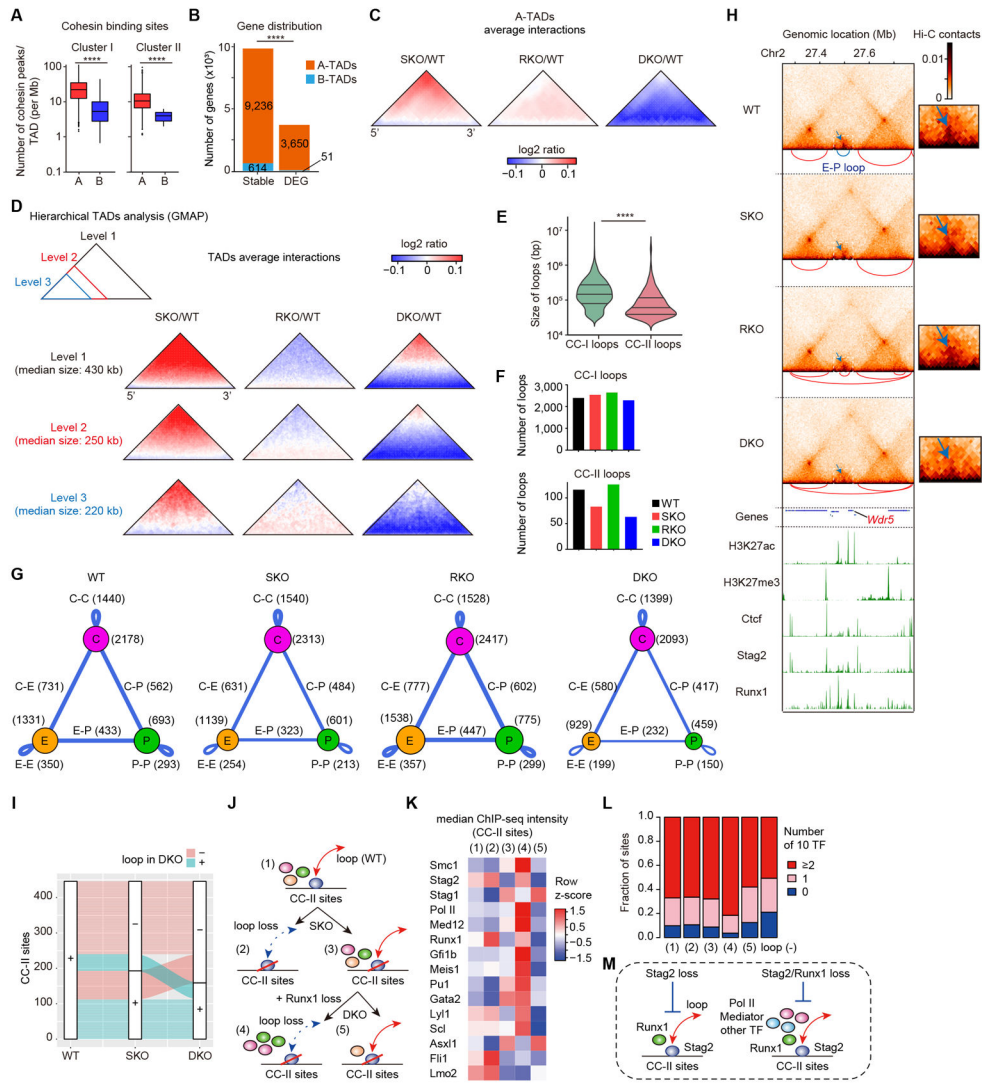
Figure 5. Stag2/Runx1 codeficiency alters chromatin architectures and disrupts enhancer-promoter loops.

A, Number of cohesin peaks (CC-I or CC-II) within topologically-associating domains (TADs) located in genomic compartment A (A-TADs) or B (B-TADs). *P*-values were calculated by two-sided Wilcoxon rank-sum test. **B**, Number of DEGs between WT- and SKO/RKO/DKO-transplanted LSK cells (FDR < 0.05) or other genes (stable) located in A- or B-TADs. *P*-value was calculated by Fisher's exact test. **C**, Average differential changes in Hi-C contacts within a subset of size-normalized A-TADs, visualized as log2 ratio indicated in the color scale. **D**, Average differential changes in Hi-C contacts within each hierarchical level of size-normalized TADs, showing the disruption of short-range interactions particularly within smaller sub-TADs in SKO, and more prominent in DKO. Hierarchical TADs were called using GMAP, and each level of TADs indicated in the upper-left panel was separately analyzed. **E**, Violin plots showing the size distribution of CC-I or CC-II loops with median and quartiles. *P*-value was calculated by two-sided Wilcoxon rank-sum test. Loops were classified by the presence of only one of either CC-I or CC-II sites at their

anchors. **F**, Number of CC-I or CC-II loops independently identified using each Hi-C data. **G**, Summary of the major types of loops identified in each Hi-C data. Ctcf sites (CC-I sites) and active enhancers/promoters in which loops were anchored are displayed as purple, orange, and green circles, respectively. The loops between two sites are displayed as blue lines, and the width of the lines is proportional to the number of loops relative to WT. E, Enhancer; P, Promoter; C, CTCF; C-C, Ctcf-Ctcf; C-E, Ctcf-Enhancer: C-P, Ctcf-Promoter; E-E, Enhancer-Enhancer; E-P, Enhancer-Promoter; P-P, Promoter-Promoter. **H**, Genome browser snapshot demonstrating the Hi-C contacts, chromatin loops (upper panels), and ChIP-seq profiles (lower panels) in WT-/SKO-/RKO-/DKO-transplanted HSPCs at the *Wdr5* gene (a group IV gene in Fig. 6A) locus. The arcs below each Hi-C contact map show the loops identified in the corresponding Hi-C data, and the E-P loop anchored at both promoter of Wdr5 and active enhancer was indicated as blue color. The dotted white box indicates the magnified region shown on the right. Color scale intensities of Hi-C heatmaps are shown in KR-normalized Hi-C contacts. Note that the E-P loop anchored at both promoter of Wdr5 and active enhancer was weakened in SKO, and more prominently in DKO (blue arrows). **I**, An alluvial plot demonstrating the proportion of CC-II sites having loops in WT which retained or lost loops in SKO and DKO. Red sites lost loops in DKO, and green sites retained loops in DKO. **J**, A classification scheme of CC-II sites with loops identified in WT for the analysis in (**K**) and (**L**). **K**, Median ChIP-seq intensities of various factors at each group of CC-II sites shown in (**J**). Color scales are normalized along each row. **L**, Proportions of numbers of co-bound 10 TFs (Asxl1, Fli1, Gata2, Gfi1b, Lmo2, Lyl1, Meis1, Pu1, Runx1, and Scl) at each group of CC-II sites shown in (**J**). **M**, Schematic representation depicting the characteristics of loops susceptible to Stag2/Runx1 loss. **** $P < 0.0001$.
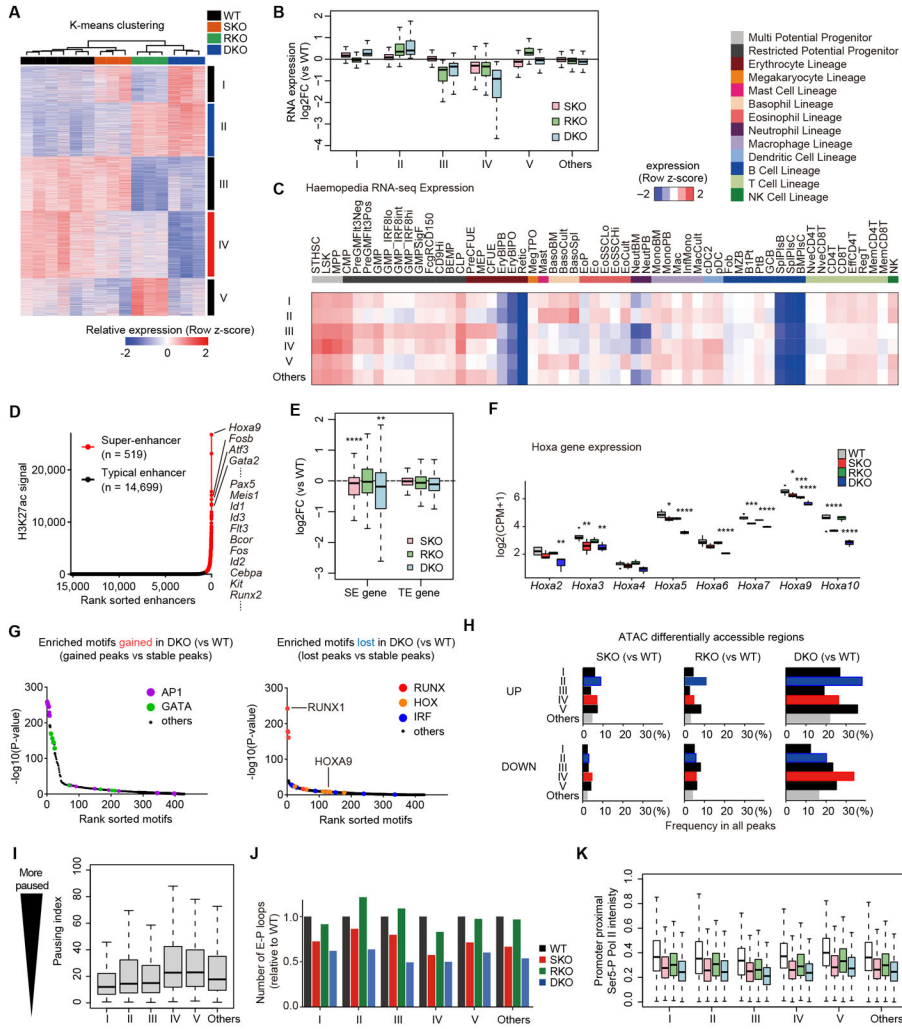
Figure 6. Molecular features of transcriptional vulnerability to Stag2/Runx1 codeficiency.
A, -means clustering analysis of DEGs between WT- and SKO/RKO/DKO-derived LSK
cells in RNA-seq datasets (FDR < 0.05). Color scales are normalized along each row. **B**, Box
plots showing expression changes of each DEG group in SKO/RKO/DKO-derived LSK cells
compared with WT. The vertical axis represents the log2(FC) in the indicated genotype and
DEG group. **C**, Expression specificity of each DEG group across diverse hematopoietic
lineages. Average expression levels of genes in the indicated DEG groups in each
hematopoietic lineage are shown. Mouse expression datasets of diverse hematopoietic
lineages are from Haemopedia RNA-seq datasets. Color scales are normalized along each
row. **D**, Super-enhancers (SEs) and typical enhancers (TEs) identified by the standard ROSE
algorithm using H3K27ac ChIP-seq intensities in HSPCs. **E**, Box plots showing expression
changes of SE- and TE-associated genes in SKO/RKO/DKO-derived LSK cells compared
with WT. *P*-values were calculated by one-sided Wilcoxon rank-sum test comparing SE
genes vs TE genes. **F**, Box plots showing expression levels of Hoxa family genes in
WT/SKO/RKO/DKO-derived LSK cells. *P*-values (vs WT) were calculated with edgeR
package. **G**, Enrichment of known TF motifs in the ATAC-seq peaks that gained (left panel)

or lost (right panel) accessibility in DKO-derived LSK cells compared with WT. The sorted motif rank and −log10(P-value) of a motif enrichment test using stable peaks as backgrounds are indicated in horizontal and vertical axis, respectively. **H**, Frequencies of differentially accessible ATAC-seq peaks in SKO/RKO/DKO-derived LSK cells compared with WT (FDR < 0.05) near genes in the indicated DEG group. **I**, Box plots showing Pol II pausing indices of genes in each DEG group. **J**, Number of E-P loops anchored at the promoters of genes in the indicated DEG groups. The vertical axis represents the relative number of loops in WT/SKO/RKO/DKO-derived HSPCs to WT. **K**, Box plots showing Ser5-P Pol II ChIP-seq intensities in the promoter proximal regions of genes in each DEG group in WT/SKO/RKO/DKO-derived HSPCs. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.
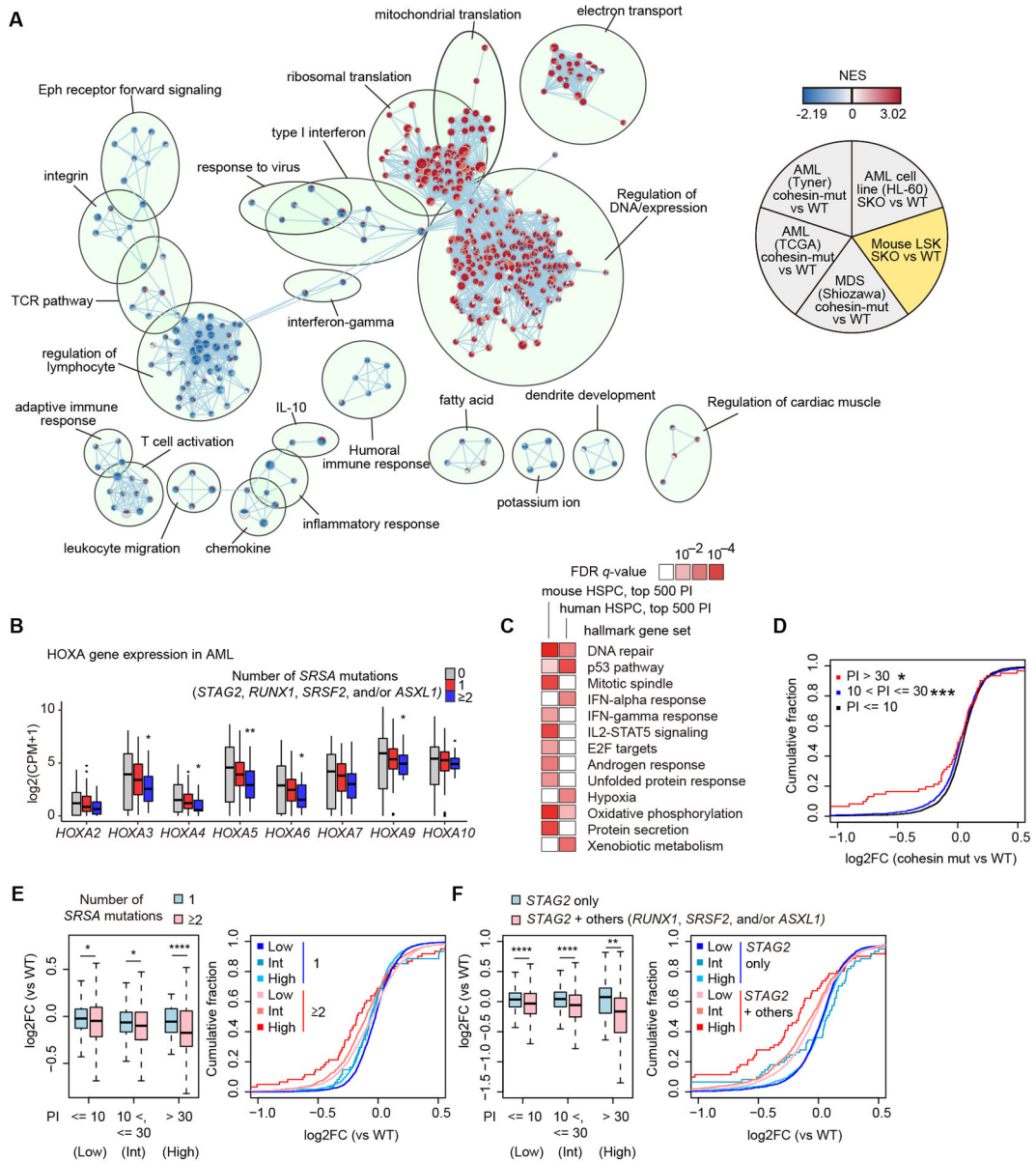
Figure 7. Shared transcriptome changes in human and mice.
A, Comparison of transcriptome changes in mouse model, three human MDS/AML, and HL-60 cell line datasets using enrichment map analysis based on GSEA results. NES values in cohesin-mutated MDS/AML cases compared with cohesin-WT cases in three independent cohort (6,33,34) are indicated in the upper left, lower left, and bottom of each circle, and those in SKO of HL-60 cell lines and LSK cells compared with WT are indicated in the upper right and lower right of each circle, respectively. Each node indicates a gene set of GSEA. The size of each node indicates the number of genes in each gene set, and the color scale indicates the NES value. The width of edge indicates the overlap size of gene sets. **B**, Box plots showing expression levels of HOXA family genes in human AML patients with 0/1/ 2 mutations in *SRSA* genes. *P*-values (vs no mutations in *SRSA* genes) were calculated with edgeR package. **C**, MSigDB overlap analysis between high-pausing genes and

hallmark gene sets in MSigDB. FDR q-values were from MSigDB overlap analysis. Pathways which are significant ($q < 0.01$) in either dataset are shown. **D**, Cumulative probability distributions of expression changes (log2FC) of genes grouped by pausing index (PI) in cohesin-mutated cases (vs WT) in RNA-seq datasets of AML (33). P-values (vs genes with PI no more than 10) were calculated by one-sided Wilcoxon rank-sum test. **E-F**, Left panels: Box plots showing expression changes (log2FC) of genes grouped by PI according to the number of *SRSA* mutations (0/1/ 2) (**E**) or mutations in *STAG2* with/without the other *SRSA* mutations (*RUNX1*, *SRSF2*, and/or *ASXL1*) (**F**) in RNA-seq datasets of AML (33). Right panels show cumulative probability distribution of expression changes (log2FC) shown in left panels. P-values were calculated by one-sided Wilcoxon rank-sum test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.