Cohesin-dependent regulation of gene expression during differentiation is lost in Cohesin-mutated myeloid malignancies.

Authors

Daniel Sasca,[1-3] Haiyang Yun,[1,2] George Giotopoulos,[1,2] Jakub Szybinski,[3] Theo Evan,[1,2] Nicola K. Wilson,[1,2] Moritz Gerstung,[4] Paolo Gallipoli,[1,2] Anthony R Green,[1,2] Robert Hills,[5] Nigel Russell,[6] Cameron S. Osborne,[7] Elli Papaemmanuil,[8] Berthold Göttgens,[1,2] Peter Campbell,[9] Brian JP Huntly[1,2]

Affiliations

1. Wellcome Trust-MRC Cambridge Stem Cell Institute, Cambridge, UK

2. Department of Haematology, University of Cambridge, Cambridge, UK

3. Department of Hematology, Oncology and Pneumology, University Medical Center Mainz, Mainz, Germany

4. European Bioinformatic Institute (EBI), Genome Campus, Hinxton, UK

5. Nuffield Department of Population Health, University of Oxford, Oxford, UK

6. University of Nottingham, Nottingham, UK

7. Department of Medical and Molecular Genetics, Kings College London

8. Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Centre, NY, USA

9. Cancer, Ageing and Somatic Mutation Programme, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK

Key points

- Cohesin depletion abrogates induction of dynamic erythroid transcriptional programmes
- Cohesin-dependent dynamic gene expression during erythroid differentiation is pre-specified and repressed by Etv6 in stem and progenitor cells

Abstract

Cohesin complex disruption alters gene expression and Cohesin mutations are common in myeloid neoplasia, suggesting a critical role in hematopoiesis. Here, we explore Cohesin dynamics and regulation of hematopoietic stem cell homeostasis and differentiation. Cohesin binding increases at active regulatory elements only during erythroid differentiation. Prior binding of the repressive Ets transcription factor Etv6 predicts Cohesin binding at these elements and Etv6 interacts with Cohesin at chromatin. Depletion of Cohesin severely impairs erythroid differentiation, particularly at Etv6-pre-bound loci, but augments self-renewal programmes. Together with corroborative findings in acute myeloid leukemia and myelodysplastic syndrome patient samples, these data suggest Cohesin-mediated alleviation of Etv6 repression is required for dynamic expression at critical erythroid genes during differentiation and how this may be perturbed in myeloid malignancies.

Introduction

Cohesin is an evolutionary conserved multiprotein complex that topologically entraps DNA, thereby establishing interactions of more than one DNA fragment[1]. This Cohesin-mediated DNA tethering occurs across multiple genomic layers and regulates critical cellular functions. The Cohesin complex is an absolute requirement for replication fork stability[2], DNA-damage repair via homologous recombination[3] and, critically, for coordinated sister chromatid cohesion to ensure orderly chromosomal segregation[4]. Recently however, a role for Cohesin has been described in coordinating contact between non-contiguous regions of the same DNA strand, such as in mediating interactions between proximal and distal *cis*-regulatory elements (e.g. promoter-enhancer interactions) and also in insulation of topologically-associating domains (TAD)[5–7]. Structurally, Cohesin forms a distorted ring, with SMC1A/B and SMC3 as its arms, while RAD21 reinforces the interaction of the two SMC subunits[8]. The role of STAG2/1 is less well appreciated but these may function during the sensing and recruitment of the complex to DNA[9].

Alterations of Cohesin function have been described in both inherited developmental disorders such as Cornelia de Lange and Roberts syndromes and via somatically acquired mutations in malignancy. Mutations in members of the Cohesin complex associate with an unusual mixture of cancer types - Bladder Cancer, Glioblastoma, Ewing-Sarcoma and Myeloid Neoplasia[10–12]. Specifically for Myeloid Neoplasia, *STAG2* mutations occur at a frequency of ~5-10% in Myelodysplastic Syndromes (MDS), de-novo and secondary Acute Myeloid Leukemia (AML), while mutations in *RAD21, SMC1A* and *SMC3* altogether display a prevalence of another 5%[13–16]. In general Cohesin mutations are heterozygous, although *STAG2* and *SMC1A* are X chromosome-linked. The mutations are predicted to confer loss or dominant negative functions, and usually lead to a decrement of protein levels of Cohesin members[11,17].

In Cohesin mutated tumors, and specifically in myeloid malignancies with Cohesin mutations, there is no evidence to associate Cohesin mutations with aneuploidy or abrogation of proper chromosome segregation[17–20]. This implies that residual

Cohesin function in mutated cells is sufficient to coordinate sister chromatid cohesion. Furthermore, several groups have demonstrated in different models that loss, down-regulation or over-expression of mutated forms of Cohesin genes cause alterations in the balance between hematopoietic stem and progenitor cells (HSPC) and differentiated cells that are associated with altered gene expression programmes and chromatin accessibility[17–19,21,22]. These observations are therefore compatible with the disease-associated/-causative events occurring through Cohesin-associated defects at cis-regulatory loops and/or TAD-boundaries. Indirect evidence for altered hematopoiesis in Cohesin-perturbed cells also emerges from recent investigations of clonal dynamics of MDS, where patients with STAG2 mutations significantly associated with high-risk disease, increased progression to AML and a shorter overall survival[23].

In this study we comprehensively address the function of the Cohesin complex during normal hematopoiesis and how its abrogation biases differentiation and predisposes towards myeloid malignancy.

Materials and methods

Cell Culture Conditions

HPC-7 cells (kindly provided by Leif Carlsson, Umeå University Sweden) were cultured in Iscove's Modified Dulbecco's Medium (IMDM, Gibco) supplemented with 10% Fetal Bovine Serum (FBS, Sigma-Aldrich), 100 ng/µl murine recombinant Stem Cell Factor (SCF, Peprothech) and 74.8 mM Monothioglycerol (Sigma-Aldrich). For all experiments, cells were used after 1-3 weeks of expansion in liquid culture. For the erythroid differentiation, SCF was decreased to 20 ng/µl and human recombinant Erythropoietin (EPO, Peprotech) was added at 4 U/l. For the myeloid differentiation, SCF was decreased to 20 ng/µl and murine recombinant Interleukin-3 (IL-3, Peprotech) was added at 50 ng/µl.

Flow cytometry/FACS

For the antibodies please refer to the specific section in the Supplemental materials and methods. With the exception of the GFP experiments (Figure S6D), which were performed on a BD Canto II, flow cytometry was performed on a BD Fortessa. Sorting of GFP+ cells was performed on a BD FACS Aria II flow cytometer.

Western blotting

For the preparation of whole cellular extracts and western blot please refer to our previous work[24].

Chromatin Immunoprecipitation (ChIP), library preparation, sequencing

Approximately 15x10e6 cells were suspended in PBS and cross-linked for 15 minutes at room temperature by the addition of 1% formaldehyde (Sigma-Aldrich), followed by quenching with 125 mM glycine for 5 minutes. Cells were washed 2x in cold PBS and stored at −80°C. All ChIP preparations were performed with previously frozen cell stocks. Cross-linked cells were thawed on ice, suspended in lysis buffer (50 mM Tris-HCl ph 8.0, 1% SDS, 10 mM EDTA, 1 × Complete® protease inhibitor) and sonicated on a Diagenode Bioruptor plus® sonicator for 15 cycles and 30 seconds with 30 seconds between cycles. Next, lysates were cleared by centrifugation at 16,000 rcf. for 10 minutes at 4°C, 25 µL/sample were reserved for input, while the remaining lysates were diluted 10X in a modified RIPA buffer (10 mM Tris-HCl ph 8.0, 1% Triton X-100, 0.1% Na-Deoxycholate, 90 mM NaCl, 1 × Complete® protease inhibitor) and incubated for 4 hours with antibodies. Chromatin-antibody conjugates were afterwards supplemented with 17.5 ul of each Protein A and G Dynabeads® (Thermo Fisher Scientific) and further incubated over night. Next, beads were washed three times with wash buffer A (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS) and two times with wash buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS). DNA was eluted off the beads by incubation on a Thermo-mixer (Eppendorf) at 30°C for 15 minutes and 900 rpm in 125 µL elution buffer (1% SDS, 100 mM NaHCO3). Cross-link reversal and RNA degradation (with 2 ul of 10 mg/ml RNAse A (Thermo Fisher Scientific)) were performed simultaneously for 4 hours at 65°C. The DNA was finally purified with the QIAquick PCR Purification Kit® from Qiagen.

For ChIP-seq experiments, ChIP DNA was used to prepare multiplexed libraries following the Illumina TruSeq DNA Sample Preparation v2® protocol and kit. Amplified libraries were size-selected for fragments between 250 and 450 bp. using a 2% freshly prepared Low-Range-Agarose Gel. Libraries were quantified by qPCR using the NEBNext® Library Quant Kit for Illumina (New England Biolabs) as recommended by protocol. Finally, library sizes, purity and free adapters were quantified on an Agilent 2100 Bioanalyzer® using the Agilent DNA 1000® Kit as per the manufacturers instructions. Libraries were sequenced on an Illumina HiSeq 4000 for 50 base pairs in single read mode. To avoid adapter hopping, free adapters were removed (if present) through a second size-selection in gel, while samples were stored separated at -20°C and only multiplexed using unique dual indexing pooling combinations of up to 6 indexes/lane.

Promoter Capture Hi-C (pCHiC)

pCHiC was performed in close analogy to Schoenfelder et al[25] with several differences, as here highlighted: 1. Cross-linking was performed with a final concentration of 1% formaldehyde. 2. Nuclei were isolated during the lysis using 15 ml Dounce homogenizers (10 strokes with the loose grinding pestle, 10 strokes with the tight grinding pestle). 3. Library HiC amplification was performed with 6 cycles.
Final pCHiC were quantified with Qubit® and quality checked on an Agilent 2100 Bioanalyzer® using the Agilent DNA High Sensitivity® Kit as per the manufacturers instructions. Libraries were sequenced on an Illumina HiSeq 4000 (each library in one lane).

Detailed methods for Cloning, virus production, transfection and transduction, Chromatin fractionation, Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME), Co-Immunoprecipitation, RNAseq, ChIP-Seq, pCHiC, RIME and human data analysis are provided in the Supplemental Materials and methods section.

Data availability

All ChipSeq, promoter Capture HiC and RNASeq data have been deposited in the GEO database under the accession numbers GSE129478. Raw RIME-MS will be made available upon request.

Results

Cohesin dosage at active cis-regulatory elements varies between hematopoietic populations during differentiation

To determine the role of Cohesin in hematopoiesis, we first assessed the dynamic expression patterns of Cohesin members in single murine cells during hematopoietic differentiation[26]. Expression of Cohesin members, as exemplified by Rad21 (Figure 1A, diffusion plot), differed significantly between HSPC and committed stages, with increased expression demonstrated upon differentiation. However, this was not uniform across lineages; when we further subdivided progenitors into myeloid (defined as Spi1^{high} single cells) and erythroid precursors (Klf1^{high} single cells), we observed a marked increase of Rad21, Stag2 and Smc3 expression in erythroid progenitors only, with no change evident in myeloid cells (Figures 1A, box plot and S1A). Of note, the expression of Smc1a and Stag1 did not significantly differ between stem and progenitor populations.

To address if this variable expression of Cohesin member transcripts translated into differing global dosages of Cohesin complex protein binding at chromatin between earlier HSC, erythroid and myeloid precursors during differentiation, we utilized a hematopoietic stem and progenitor cell line, HPC7[27] (hereafter HPC), that recapitulates the epigenetic landscape of murine HSC[28]. We performed chromatin immunoprecipitation followed by parallel sequencing (ChIP-Seq) for Cohesin proteins Rad21 (hereafter as a proxy for the whole Cohesin complex) and Smc1a, the major Cohesin interaction partner Ctcf, RNA Polymerase II (PolII) and chromatin marks for promoters and enhancers: H3K4me3, H3K4me1, and H3K27ac. Cohesin-bound chromatin regions were classified into active and primed enhancers, active and inactive promoters and insulator/other regions (as defined in Figure S1B). In agreement with previous studies[7], approximately 60% of the Cohesin peaks were

enriched at insulators, with the remaining 40% bound to active, H3K27ac-associated, regulatory elements (Figure S1C).

To determine dynamic changes in Cohesin binding during hematopoietic differentiation, we repeated ChIP-Seq analysis for Rad21, Smc1a and Ctcf in cells that were induced for 48 hours with EPO (hereafter Erythroid cells) or 72 hours with IL-3 (Myeloid cells) (Figure 1B). These time points were chosen as cells still demonstrated an HPC surface phenotype, however longitudinal flow analysis demonstrated that they were transitioning to more differentiated cells (Figure 1C). A discrete, although significant decrease in Rad21 binding at insulators and other elements was noted between the HPC and Myeloid cells (Figures 1D-F). This decrease was not shared by Smc1a binding (Figure S1D-E). Strikingly, during erythroid differentiation a marked increase in both Rad21 and Smc1a binding was demonstrated at active *cis*-regulatory, but not insulator, elements (Figures 1D-F and S1D-E). Increased Smc1a binding was evident, although Smc1a gene expression did not differ between earlier hematopoietic stages and erythroid precursors, possibly due to high base line protein expression and, thus, high protein availability. Of note, Erythroid cells also displayed increased total protein levels of Rad21 and Stag2 compared to HPC and Myeloid cells, highlighting the necessity of increased cohesin protein availability in these cells (Figure S1F). In contrast to dynamic pattern of the Cohesin complex, Ctcf showed uniformly increased binding across all elements in both differentiated cell types (Figures S1G-H).

Taken together, these observations imply different and context-specific requirements for Cohesin during erythropoiesis and myelopoiesis, with the pattern of expression and binding of Cohesin suggesting it as critical for erythroid differentiation.

Global dynamic Cohesin binding at active promoters correlates with H3K27 acetylation, but not gene expression, during erythroid differentiation.

We next wished to correlate Cohesin dosage at chromatin with dynamic histone modifications by ChIP-Seq, 3-D promoter-enhancer interaction by promoter based Capture Hi-C (pCHiC) and alterations in gene expression by RNA-Seq during erythroid differentiation (Figure 2A). Although Rad21 binding was almost globally increased upon erythroid differentiation, the degree was noted to be highly variable across the genome. Based on this observation we divided all active promoters into 4 notional erythroid "differentiation" tiers. Tier1 promoters manifested the strongest increase of Rad21 dosage, with this dosage gradually decreasing across the remaining tiers, such that Tier4 promoters demonstrated no increase (Figures 2B-C).

We next integrated the dynamics of H3K27ac, pCHiC interactions and gene expression changes across the 4 tiers. On a global scale, H3K27ac decreased in Erythroid cells compared to HPC (Figure 2D). Of note, the degree of H3K27ac alteration significantly correlated with the erythroid differentiation tiers and with Rad21/Cohesin binding; the most evident reduction of H3K27 acetylation was present in Tier4 promoters and the change in the modification was preserved greatest in Tier1 promoters (Figure 2D). However, promoter interaction dynamics did not mirror this correlation and demonstrated no coordinated difference between the tiers (Figure 2E). To compare gene expression changes between the HPC and the Erythroid cell states, genes whose differential expression increased after EPO stimulation and erythroid differentiation were termed 'Erythroid' genes, and those whose expression decreased, and were thus higher in HPC, were called 'Immature' genes (Figure 2F). As expected, Erythroid genes were enriched for erythroid differentiation genesets (Figure S2A). However, when we overlapped our Immature and Erythroid genes with the 4 graded tiers of Rad21 binding, we found no significant correlation between Rad21 binding level and alterations in gene expression (Figure 2G).

We next compared dynamic Rad21 and H3K27ac binding, and 3D interaction frequencies at the promoters of Erythroid and Immature genes. Although Rad21 dosage increased equally at both Erythroid and Immature gene groups upon differentiation, a relative increase in H3K27ac, and interaction frequency was noted

only at Erythroid genes (Figures S2B-C). For example, the *Epo-receptor* gene (*Epor*) is upregulated following EPO stimulation, where activation is associated with increased Rad21 binding, H3K27ac modification and interaction frequency. In contrast, at the *Etv5* gene, increased Rad21 binding associates with decreased H3K27ac, interaction frequency and decreased expression (Figure 2H). Taken together, these studies indicate that binding of Cohesin alone is not sufficient to instruct erythroid differentiation but requires additional factors that allow gene transcription.

Etv6 binding precedes Cohesin binding at cis-regulatory elements of genes upregulated during erythropoiesis

During hematopoietic differentiation, transcription is coordinated by co-activators such as the p300/Crebbp to deposit H3K27ac at promoters and enhancers[29]. These co-activators are classically recruited by specific hematopoietic- and/or lineage-specific transcription factors[30]. We hypothesized that the non-uniform pattern of Cohesin binding during normal erythropoietic differentiation is influenced at an earlier hematopoietic stage by specific transcription factors. To identify proteins that might regulate Cohesin binding, we performed pull-down of endogenous Stag2-containing chromatin complexes, using a technique that enriches for chromatin bound interactions[31] (Figure 3A). We initially compared mock knockdown (Luc-HPC) and *Stag2* knockdown (shS2, Figure S4A) HPC to identify Cohesin interacting proteins. Only proteins that were present in all Luc-HPC replicates and displayed significantly decreased presence upon *Stag2* knockdown (Figure 3B) were considered as putative interacting proteins. To generate a high confidence dataset of Cohesin interacting proteins, we next performed RIME using Rad21 containing chromatin complexes and intersected these with the Stag2-interactors. We could demonstrate a strong overlap between the interacting proteins and further characterized their presumed functions using gene ontology molecular function (GO:MF) overrepresentation (Figure 3C). Most cohesin interactors were nucleic acid-binding proteins and could be assigned into different groups, based on their putative function. One of these groups consisted of 5 hematopoietic transcription factors (hTF) - Etv6, Erg, Runx1, Tal1 and Stat5 (Figure 3C). We then determined the

genome-wide DNA binding patterns of these hTFs in unstimulated HPC cells, including their binding at the erythroid differentiation tiers (Figure 3D). Of the 5 hTFs, only Etv6 displayed globally enhanced co-binding at the Cohesin-associated promoters. Strikingly, Etv6 binding was highest at promoters from Tier 1 loci and decreased through the remaining tiers, a pattern directly proportional to Cohesin binding during erythroid differentiation. Given that Etv6 is predominantly a transcriptional repressor that has been previously associated with maintaining HSC function and megakaryopoiesis[32], we next assessed the dynamic expression patterns of Etv6 during differentiation using published scRNASeq data[33] (Figure S3). In diametric contrast to Cohesin dynamics, Etv6 was most highly expressed in HSPC and strongly decreased during erythroid differentiation. Finally, we confirmed the interaction between Etv6 and cohesin by performing Co-Immunoprecipitation of Rad21 or Stag2 and western blots for Etv6 (Figure 3E). Taken together, we functionally link Etv6 with Cohesin and indirectly suggest that a release of transcriptional repression by Etv6 may be required for proper erythroid differentiation.

Cohesin deficiency severely impairs erythroid differentiation and expands myelopoiesis, but only modestly alters HSPC homeostasis.

The differential landscape of Cohesin binding between HPC and erthyroid cell states predicts that Cohesin is critical for the dynamic remodeling of transcription during erythropoiesis. To test this, we perturbed the function of Cohesin members in HPC (Figure 4A) using inducible knockdown of *Stag2* (shS2_HPC), *Rad21* (shR21_HPC) or *Smc1a* (shS1a_HPC) (Figures S4A-B). No significant differences in the binding of Cohesin proteins or Ctcf were detected upon Cohesin subunit knockdown in HPC (Figures 4B, S4C). Additionally, RNA-Seq revealed only minor changes in gene expression in Cohesin deficient HPC (Figures 4C, S4D). However, these changes consistently decreased signatures of differentiation and, for shR21_HPC and shS1a_HPC, enriched for stem cell signatures (Figure S4E) in GSEA analyses.

In stark contrast, when we determined the role of Cohesin during erythroid differentiation, following knockdown, erythroid differentiation was significantly impaired after *Rad21, Smc1a*, and mostly after *Stag2* down-regulation (Figure 4D and S4F). This was also demonstrated at the level of global gene expression changes using principle component analysis (PCA) of RNA-Seq. Here, a decreased variance in shS2_ery cells in the first component, which coincides with erythroid differentiation (Table S3), and marked differences in gene expression between Luc_ery and shS2_ery cells were demonstrated (Figures 4E-F). GSEA further highlighted an increase of stem cell signatures and lost erythroid and general differentiation signatures following knockdown of *Stag2* (Figure 4G). Of note, Luc_ery genes (genes with significantly higher expression in control cells in comparison to Stag2 knockdown cells following EPO induction) strongly overlapped (196/288 genes, 68%) with the genes upregulated during normal erythroid differentiation (Figure 4H). In line with previous studies,[17–19] IL-3-driven myelopoiesis was significantly expanded in *Stag2* and *Rad21*-perturbed cells (Figures S4G-H).

These data demonstrate that the effects of Cohesin member perturbation are context dependent, and are in consonance with our description of the increased requirement for Cohesin function during erythroid differentiation. They also demonstrate that Cohesin appears to be critical for the expression of dynamic genes during erythroid differentiation, rather than the continued expression of steady-state genes required for the maintenance and homeostasis of the self-renewing HSPC state.

Impaired erythroid differentiation directly relates to decreased Cohesin binding at erythroid-specific genes.

We next tested if the impairment of erythroid differentiation associated with limiting concentrations of Cohesin members relates to a failure of dynamic binding of the intact Cohesin at the required *cis*-regulatory elements that we have previously defined during normal erythroid differentiation. We performed ChIP-Seq for Rad21 in control (Luc_ery) and Cohesin perturbed (shS2_ery) cells and compared signal

intensities at active enhancers and promoters and specifically across the erythroid differentiation tiers described in Figures 2B-C. The global dosage of Rad21 was only mildly, albeit significantly, decreased at the active promoters and enhancers following knockdown of *Stag2* (Figure 5A). However, across the erythroid differentiation tiers, impairment of Rad21 binding was most evident at Tier1 with a graded decline in differential binding across the remaining tiers (Figure 5B). The levels of the H3K27ac modification also positively correlated with this graded alteration of Rad21 binding (Figure 5C). Nevertheless, not all Tier1 promoters displayed an impaired binding of Rad21 in shS2_ery cells (Figure S5A), which prompted us to further define active promoters with significant differential binding of Rad21 following Stag2 knockdown. We identified 782 promoters where Rad21 dosage was significantly impaired in shS2_ery cells ("lost" promoters) and 80 promoters with a significant increase of Rad21 signal ("gained" promoters) (Figure 5D). We next integrated Rad21 and H3K27ac ChIPSeq signal intensities with pCHiC interactions and differential gene expression at these "lost" and "gained" promoters and could demonstrate that loci where Rad21 binding decreased were associated with decreased H3K27ac, impaired interactions and a reduction in gene expression (Figures 5E-G). Importantly, marked decreases in Cohesin binding were observed at crucial erythropoietic master-regulators, including *Klf1* and *Gata1*, thus providing a further explanation for the impaired erythropoietic gene expression programme (Figure 5F). Moreover, the "lost" promoters were consistently enriched for erythropoiesis and differentiation (Figures S5B-C), thus further linking dynamic Cohesin binding requirements to Erythroid and general differentiation genes.

These findings suggest that the impaired erythroid differentiation of Stag2-perturbed HPC following EPO stimulation relates to decreased Cohesin binding at the promoters of a critical subset of Erythroid genes. To corroborate this suggestion, we demonstrated a significantly positive correlation between Rad21 binding and gene expression, H3K27ac and interaction frequency at the corresponding promoters of differentially expressed genes (Figures S5D-F). Examples of critical erythroid genes, *Klf1* and *Gata1*, whose expression is enriched in Luc-ery control cells are shown in Figures 5H and S5G.

Taken together, these data demonstrate that limiting concentrations of functional Cohesin, as occurs in Cohesin-mutated myeloid malignancies, impair the normal binding patterns of the complex at active promoters/enhancers of critical genes during erythroid differentiation. This leads to a correlative decrease in H3K27 acetylation, interaction between promoters and distal *cis*-regulatory elements and reduced gene expression. Notably, the impairment of Cohesin binding is not uniform and is marked at critical regulators of erythropoiesis, explaining the impaired erythroid differentiation observed in Cohesin-mutated myeloid malignancies.

Stag2 perturbation leads to preferential loss of Cohesin binding at differentiation promoters with strong Etv6 pre-binding

As we have demonstrated, the dynamic pattern of Cohesin binding during erythropoiesis appears primed by Etv6. We therefore wondered whether Etv6 binding would also pre-mark those promoters that lose Cohesin following *Stag2* knockdown.

To test this hypothesis, we initially stratified the global alterations in Rad21 binding at active promoters following *Stag2* knockdown; we identified 4 erythroid perturbation tiers (TierA-D), based on the differences of Rad21 binding between shS2_ery and Luc_ery (Figure 6A). As anticipated, there was a degree of overlap between the differentiation and perturbation tiers (Figure S6A). Of note, we observed strong impairment of Rad21 binding at TierA, intermediate dynamics at TierB and TierC and a redistribution/sequestration of Rad21 at the TierD promoters in the shS2_ery cells (Figure 6A). We could again demonstrate significant correlation of Rad21 binding with dynamic H3K27 acetylation and promoter interaction frequencies at these tiers (Figures 6A-C). Moreover, promoters from TierD significantly overlapped with genes upregulated following Stag2 knockdown (shS2_ery genes), while promoters from TierA significantly overlapped with genes whose expression was higher in control cells following Stag2 knockdown (Luc_ery genes) (Figures 6A,D).

We then determined the DNA binding patterns of the 5 previously identified Cohesin co-interacting hTFs at the erythroid perturbation tiers (Figure 6E). Again Etv6 binding was highest at promoters from tier A loci and decreased through the remaining tiers, implying that Etv6 might contribute to altered erythropoiesis following Stag2 knockdown. Moreover, Etv6 binding was even more significantly enriched at differentiation genes that were also impaired after Stag2 knockdown (common peaks between perturbation Tier A and erythroid Tier 1), compared to all Tier1 genes (Figure 6F).

Upon erythroid differentiation in our HPC system, as in primary cells, we could demonstrate a decrease in *Etv6* expression, following EPO stimulation (data not shown). We therefore wondered if Etv6 protein abundance and binding to chromatin might persist following EPO stimulation of shS2 cells. Of note, we noticed a modest and relative increase of Etv6 at the protein level when comparing Stag2 knock down to unstimulated control HPC cells (Figure 6G). However, this did not correlate with an increase in Etv6 transcription, as Luc and shS2 HPC cells displayed similar Etv6 mRNA levels (Figure S6B). Looking at protein abundance in greater detail within individual cellular compartments in unstimulated cells, we could detect an increase in Etv6 in the cytoplasm and nucleoplasm, while the chromatin fraction remained equally saturated (Figure 6H). In contrast, during erythroid differentiation, Etv6 specifically displayed increased chromatin binding in shS2 cells, demonstrating that Etv6 remains active at chromatin following Cohesin loss (Figure 6I). Finally, we investigated the Etv6 binding dynamics by ChIP-qPCR at specifically impaired gene promoters. We chose the promoter regions of *Klf1, Epor* and *Cxcr4*, as these were common among the Tier1 and TierA loci and significantly lost Rad21 binding during Stag2-perturbed erythroid differentiation. No difference in Etv6 binding was observed between Luc_HPC and shS2_HPC cells (Figure 6J). By contrast, Etv6 binding significantly decreased during normal erythroid differentiation, however, in agreement with its higher availability in shS2_Ery cells, Etv6 binding intensity remained significantly higher in shS2_Ery cells compared to Luc_Ery cells (Figure 6J).

These data suggest that Cohesin evicts Etv6 from chromatin, relieving its repressive function, and that deficiency allows for the continued chromatin binding and repressive activity of Etv6 at genes critical for erythroid differentiation. To test this hypothesis, we sought to determine if we could rescue erythroid differentiation in Cohesin perturbed HPC by concomitantly knocking down Etv6. However, Etv6 is a critical regulator of hematopoiesis and its loss leads to impairment of HSC function and blood development[32]. In keeping with this critical function, both single knockdown of *Etv6* (Luc-shEtv6-HPC) and double knockdown of *Stag2* and *Etv6* (shS2-shEtv6-HPC) led to a marked reduction in cellular proliferation, preventing us from testing this hypothesis (Figure S6C-D).

Erythropoiesis is impaired in human Cohesin-mutated MDS and AML

Cohesin mutations are common events in myeloid neoplasia. Among Cohesin genes, mutations of *STAG2* are most frequent[13,15,34]. To determine if our model of impaired erythropoiesis correlates with human disease, we analyzed differential gene expression and matched mutational analysis in a dataset consisting of CD34+ bone marrow cells from 159 MDS patient samples and 17 healthy donors[35]. We initially analyzed all significant differential transcripts that occurred between 8 STAG2-mutated and 151 other MDS samples, noting a total count of 2778 genes that were differentially expressed ($q < 0.05$). Consistent with our Cohesin knockdown datasets, the Hemoglobin genes and the essential erythroid regulator *KLF1* were some of the most highly down regulated genes in *STAG2* mutated samples (Figure 7A). GSEA further highlighted that stem cell signatures were up regulated and erythroid and general differentiation signatures down regulated in *STAG2*-mutated samples (Figure 7B), in consonance with our murine cellular model of altered Cohesin function. Indeed, the majority of the human orthologues of the genes differentially expressed during erythroid commitment following Stag2 knock down also showed decreased gene expression in *STAG2*-mutated MDS samples (Figure S7A), further corroborating the relevance of our model to human disease. Extending our analysis to Cohesin-mutated AML, we repeated this analysis in the TCGA LAML dataset[36], comparing expression changes from 18 Cohesin-mutated patient samples with 133 AML

samples from patients who lacked a Cohesin-mutation. Again, there was a marked overlap between the genes downregulated in Cohesin-mutated AML and our experimental dataset (Figure S7B).

Mutual exclusivity of *ETV6* and *Cohesin* mutations in myeloid malignancies supports their functional interaction

ETV6 and Cohesin members are commonly mutated in MDS and AML. To further interrogate the functional interaction between ETV6 and Cohesin, we examined the pattern of mutations in individual patients from large cohorts of patients with myeloid malignancies. Analyzing mutations in a total of 3612 patient samples with myeloid malignancies from the Cosmic database[37] in which all *STAG2* and *ETV6* coding regions were sequenced, we observed that only 4 out of 338 samples displayed mutations in both *ETV6* and *STAG2*, suggesting a strong ~~degree of~~ tendency towards mutual exclusivity of mutations in these genes (~~p<0.001,~~ Figure 7C). However, perhaps related to the small number of ETV6 mutations, the heterogeneity of the myeloid diseases or the lack of full Cohesin member coverage, we were relatively underpowered to demonstrate this statistically and mutual exclusivity was not significant by Fisher's exact test (p = 0.4).

To further validate these results and underscore the significance of our mechanistic model, we examined cumulative results from an unpublished cohort of 2434 AML samples from the UK NCRI AML trials (MRC), 200 AML samples from the TCGA consortium[36] and 622 AML samples from the Oregon Health & Science University (OHSU)[38], in which all Cohesin members and *ETV6* genes were analyzed for mutations. Among a total of ~~386~~ 472 samples, only one was co-mutated for a Cohesin member and *ETV6*, a comparable frequency to the degree of mutual exclusivity between individual members of Cohesin (~~p<0.001,~~ p=0.028, Figure 7D). Of note, full mutual exclusivity between *ETV6* (22 samples) and cohesin members *STAG2/RAD21* was also evident in another large AML dataset (Papaemmanuil et al[16]). These results ~~almost complete mutual exclusivity~~ further corroborate our experimental data and strongly support a functional interaction between ETV6 and

Cohesin, such that in the absence of physiological levels of Cohesin, intact ETV6 function is required to repress genes critical for erythroid differentiation.

Discussion

Multiple studies[17–19,21] and the high incidence of mutations of Cohesin members in myeloid malignancies[13,15,16] have demonstrated that reduced Cohesin dosage perturbs hematopoiesis. However, the mechanisms whereby Cohesin maintains hematopoietic homeostasis remain obscure. We now present a model that highlights the requirement for Cohesin during the dynamic induction of erythroid differentiation. We link this requirement to the acetylation of proximal and distal *cis*-regulatory elements, to increased interaction between these and to the expression of important differentiation genes. Critically, these changes are not uniform across the genome but appear coordinated by the prior binding of specific, interacting hTFs, such as Etv6. Moreover, we demonstrate the abrogation of dynamic induction of these differentiation processes and genes when Cohesin members are depleted. Finally, we provide evidence of a functional interaction between ETV6 and Cohesin and corroborate its relevance for human disease, correlating our experimental findings with large datasets from patients with MDS and AML.

Our model system was chosen to allow temporal and lineage control over differentiation and, with it, the interrogation of specific cellular states rather than heterogeneous populations of cells along a differentiation continuum. To overcome the limitations of a purely *in-vitro* system, we corroborated our findings at all times to more physiological systems, all the way from the expression of the Cohesin members in murine and human stem and progenitor cells undergoing differentiation, through similarities of our altered gene expression patterns to murine self-renewal signatures, all the way to the physiological correlation of our data with the genotype-specific disease mutational and gene expression signatures in MDS and AML databases. Similarities are seen with phenotypes previously described in murine and human hematopoietic cells following Cohesin depletion or ever expression of mutated forms [17–19,22], where decreased erythroid differentiation

of human or murine haematopoietic stem and progenitor cells have been demonstrated *in vitro*[17], and decreased bone marrow MEP and Ter119+ cells *in vivo*[18], although this same study actually showed an increase of extramedullary erythroid cells in the spleen. Moreover, our identification of the hTF that interact with Cohesin on chromatin may also explain, at least in part, the alterations in chromatin accessibility demonstrated upon Cohesin depletion in these studies. Of note, increased accessibility at chromatin regions enriched for Erg, Gata2, Runx1 and Stat5 binding sites were documented, demonstrating an almost complete overlap with the hTF that we detected to interact with Cohesin at chromatin.

We find that Cohesin member perturbation has little direct influence on the maintenance of existing transcriptional programmes necessary for HSPC homeostasis. This reconciles with previous studies in which only the rapid and complete removal of Cohesin from chromatin affects the 3D maintenance of chromatin[39] and would also explain why malignant cells can tolerate partial loss of Cohesin whilst still maintaining proliferation. Our findings of the requirement for Cohesin function to induce gene expression during the dynamic process of differentiation are in line with a recent study[40], where so-called inducible enhancers were dependent on Cohesin binding to activate gene expression from their target promoters during the dynamic process of macrophage activation following inflammatory signaling. However, our data expands on this observation, mechanistically highlighting that the induction of dynamic processes, such as differentiation, by Cohesin-mediated gene expression programmes is neither random nor uniform, but highly specific and appears regulated by pre-bound hTFs. We speculate that, within HSPCs, critical erythroid genes are repressed by Etv6. However, during normal erythroid differentiation, Cohesin is recruited to these critical promoters. This coincides with the loss of Etv6 binding, deposition of the activating H3K27ac, increased contact between these promoters and distal cis-regulatory enhancers and expression of critical erythroid genes. Moreover, during normal differentiation *Etv6* is downregulated at the transcriptional level to further augment this switch. However, under Cohesin insufficiency, e.g. in Cohesin-mutated myeloid neoplasia, there is no increase in Cohesin binding and Etv6 remains bound

at chromatin, resulting in continued repression of critical erythroid genes and a failure to differentiate properly. Moreover, this model is entirely consistent with the erythroid differentiation defect evident within patients with Cohesin-mutated myeloid neoplasia and is supported by the near mutual exclusivity of *ETV6* and Cohesin member mutations in myeloid malignancies.

In summary, our data shed considerable light on the function of Cohesin during cellular processes that require dynamic gene regulation, such as erythroid differentiation. Further investigation is warranted to demonstrate whether Cohesin is required for other dynamic cellular processes. We propose a mechanism for Cohesin co-ordinated alterations in critical gene expression programmes that relates to interaction with specific hTFs such as Etv6. We speculate that other tissue-specific TF may regulate Cohesin function in other systems and suggest further studies are warranted to test this hypothesis. Finally, we demonstrate that abrogation of Cohesin function alters differentiation through the failure to affect a switch from repression to activation at specific genes and propose these genes and continued Etv6 activity as potential therapeutic targets to induce differentiation in AML and MDS.

Acknowledgements

We are grateful to Robin Antrobus for his help and advice with proteomic experiments and analysis.

The Huntly Lab is funded by the European Research Council (grant 647685 COMAL), CRUK (C18680/A25508), the Kay Kendall Leukaemia Fund, the Medical Research Council (MRC), Bloodwise, the Wellcome Trust. Research in the laboratory is supported by core funding from Wellcome and MRC to the Wellcome-MRC Cambridge Stem Cell Institute. DS was a Postdoctoral Fellow of the Mildred-Scheel Organization, German Cancer Aid (111875). NKW and BG were funded by the Wellcome Trust, Bloodwise and CRUK.

Author contributions

Conceptualization, D.S. and B.J.P.H.; Methodology, D.S., H.H., B.J.P.H.; Software, D.S., M.G., B.G., B.J.P.H.; Formal Analysis, D.S., H.Y., G.G., J.S., T.E.; Resources, D.S., N.W., M.G., A.R.G., R.H., N.R., C.O., E.P., B.G., P.C., B.J.P.H.; Writing - Original Draft, D.S., B.J.P.H.; Writing - Review & Editing, all authors; Visualization, D.S. and B.J.P.H.; Supervision, B.J.P.H.

Conflict-of interest disclosure: The authors declare no competing financial interests.

Correspondence: Brian James Patrick Huntly, Wellcome Trust-MRC Cambridge Stem Cell Institute, Cambridge, UK; e-mail: bjph2@cam.ac.uk

References:

1. Murayama Y, Samora CP, Kurokawa Y, Iwasaki H, Uhlmann F. Establishment of DNA-DNA Interactions by the Cohesin Ring. *Cell*. 2018;172(3):465–477.e15.
2. Lengronne A, McIntyre J, Katou Y, et al. Establishment of Sister Chromatid Cohesion at the S. cerevisiae Replication Fork. *Mol. Cell*. 2006;23(6):787–99.
3. Ström L, Lindroos HB, Shirahige K, Sjögren C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell*. 2004;16(6):1003–15.
4. Michaelis C, Ciosk R, Nasmyth K. Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. *Cell*. 1997;91(1):35–45.
5. Hadjur S, Williams LM, Ryan NK, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. 2009;460(7253):410–3.
6. Kagey MH, Newman JJ, Bilodeau S, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010;467(7314):430–435.
7. Dowen JM, Fan ZP, Hnisz D, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 2014;159(2):374–387.
8. Peters JM, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* 2008;22:3089–3114.
9. Countryman P, Fan Y, Gorthi A, et al. Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates. *J. Biol. Chem.* 2018;293(3):1054–1069.
10. Brohl AS, Solomon DA, Chang W, et al. The Genomic Landscape of the Ewing Sarcoma Family of Tumors Reveals Recurrent STAG2 Mutation. *PLoS Genet.* 2014;10(7):e1004475.
11. Kon A, Shih L-Y, Minamino M, et al. Recurrent mutations in multiple

components of the cohesin complex in myeloid neoplasms. *Nat. Genet.* 2013;45(10):1232–7.
12. Solomon DA, Kim T, Diaz-Martinez LA, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science (80-.).* 2011;333(6045):1039–43.
13. Thota S, Viny AD, Makishima H, et al. Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*. 2014;124(11):1790–1798.
14. Lindsley RC, Mar BG, Mazzola E, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*. 2015;125(9):1367–76.
15. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616–27.
16. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* 2016;374(23):2209–2221.
17. Mazumdar C, Shen Y, Xavy S, et al. Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell*. 2015;17(6):675–688.
18. Mullenders J, Aranda-Orgilles B, Lhoumaud P, et al. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med.* 2015;212(11):1833–50.
19. Viny AD, Ott CJ, Spitzer B, et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J. Exp. Med.* 2015;212(11):1819–32.
20. Balbás-Martínez C, Sagrera A, Carrillo-De-Santa-Pau E, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat. Genet.* 2013;45(12):1464–9.
21. Tothova Z, Krill-Burger JM, Popova KD, et al. Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell*. 2017;21(4):547–555.
22. Galeev R, Baudet A, Kumar P, et al. Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs. *Cell Rep.* 2016;14(12):2988–3000.
23. Makishima H, Yoshizato T, Yoshida K, et al. Clonal evolution in myelodysplastic syndromes. *Nat. Genet.* 2016;57(6):723–729.
24. Sasca D, Hähnel PS, Szybinski J, et al. SIRT1 prevents genotoxic stress-induced p53 activation in acute myeloid leukemia. *Blood*. 2014;124(1):121–33.
25. Schoenfelder S, Javierre B-M, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J. Vis. Exp.* 2018;(136):.
26. Nestorowa S, Hamey FK, Pijuan Sala B, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*. 2016;128(8):e20–e31.
27. Pinto Do Ó P, Kolterud Å, Carlsson L. Expression of the LIM-homeobox gene LH2 generates immortalized Steel factor-dependent multipotent hematopoietic precursors. *EMBO J.* 1998;17(19):5744–5756.
28. Wilson NK, Foster SD, Wang X, et al. Combinatorial transcriptional control in blood stem/progenitor cells: Genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell*. 2010;7(4):532–544.

29. Calo E, Wysocka J. Modification of Enhancer Chromatin: What, How, and Why? *Mol. Cell*. 2013;49(5):825–837.
30. Roe JS, Mercan F, Rivera K, Pappin DJ, Vakoc CR. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol. Cell*. 2015;58(6):1028–1039.
31. Mohammed H, Taylor C, Brown GD, et al. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat. Protoc.* 2016;11(2):316–326.
32. Hock H, Meade E, Medeiros S, et al. Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* 2004;18(19):2336–41.
33. Tusi BK, Wolock SL, Weinreb C, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*. 2018;555(7694):54–60.
34. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(3):241–7.
35. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat. Commun.* 2015;6:5901.
36. TCGA. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
37. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 2017;45(D1):D777–D783.
38. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
39. Rao SSP, Huang SC, Glenn St Hilaire B, et al. Cohesin Loss Eliminates All Loop Domains. *Cell*. 2017;171(2):305–320.e24.
40. Cuartero S, Weiss FD, Dharmalingam G, et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat. Immunol.* 2018;19(9):932–941.

Figure legends

Figure 1. Cohesin dosage at active cis-regulatory elements varies between hematopoietic populations during differentiation

A. Left panel - Diffusion map of scRNASeq expression of Rad21, colored in red with http://blood.stemcells.cam.ac.uk/geneMap.html. The color corresponds to a log2 scale of expression ranging between 0 and the maximum value. LT-HSC, HSPC and Progenitors are highlighted in approximation of the original single cell sorting.
Right panel - box plots with 10-90 percentiles of Rad21 expression in the indicated populations. Erythroid and Myeloid progenitors were empirically considered as all the cells with normalized expression of Klf1 > 10 and Spi1 > 10, respectively.
B. Graphical schema of the experimental design to determine cohesin dynamics during normal erythroid and myeloid differentiation.
C. Representative longitudinal Flow Cytometry plots. HPCs were differentiated towards the erythroid and myeloid lineages. Day 2 of EPO induction was highlighted as the optimal Erythroid (or Ery) transition state, while cells induced for 3 days with IL-3 were called Myeloid (or Myelo) transition cells.
D. and E. Density heat map (D) and average profiles (E) of Rad21 binding at Cohesin-associated regions in the indicated cellular states.
F. Total count of significant differentially bound Rad21 peaks comparing HPC to the indicated differentiation lineage. Left panels show pie charts of genomic location of differential bound peaks. Scatter plots (right) - differential Rad21 binding at active promoters/enhancers in the indicated populations. Blue dots (Lost) show peaks that decrease, while red/purple dots (Gained for Erythroid and Myeloid cells respectively) show peaks that increase during differentiation.

Figure 2. Global dynamic Cohesin binding at active promoters correlates with H3K27 acetylation, but not gene expression, during erythroid differentiation.

A. Experimental system to determine the role of Cohesin during normal erythroid differentiation.

B. Left panel - average profile for Rad21 binding at active promoters in the HPC (blue) and Erythroid (red) cellular states. Right panel - Representation of the erythroid differentiation tiers; density heat maps of Rad21 binding sites from a representative experiment (of 2) at active promoters that were split into 4 equal tiers, based on the incremental change of Rad21 binding during erythroid differentiation.

C. Box plots and 10-90 percentiles of the differences of Rad21 binding in Erythroid cells and HPC (Rad21_Ery minus Rad21_HPC) in the indicated erythroid differentiation tiers. Shown are means from 2 replicates. Box plots to the left of zero indicate increased binding in HPC and to the right indicate increased binding in erythroid differentiating cells.

D. Box plots and 10-90 percentiles of the differences of H3K27ac binding for the same regions as in Fig2C. Shown are means from 2 replicates.

E. Box plots and 10-90 percentiles of the differences of significant interaction frequencies of the baits located at the promoters of the same regions as in Fig2C. Interactions were determined with the CHICAGO pipeline. The threshold for significance was set at 5. Shown are results from 2 replicates.

F. Volcano plot showing gene expression changes during HPC to erythroid differentiation (3 biological replicates, Luc1, Luc2 and Luc3, were used per group for consistency with later experiments). The blue shaded area consists of genes enriched in HPC (immature genes), while the red shaded area shows genes whose expression significantly increased in the Erythroid cell state (erythroid genes).

G. Venn diagrams of overlaps between annotated promoters from the erythroid differentiation tiers (centre circle) and differentially expressed genes from Fig2F (outer circles). Chi-Square analysis was performed for the trend of overlapping events across all tiers.

H. Examples of Rad21 and H3K27ac binding dynamics (upper), as well as interaction frequencies (lower arcs) between HPC and Erythroid cellular states for a representative erythroid gene (Epor) and an immature gene (Etv5).

Figure 3. Etv6 binding precedes Cohesin binding at cis-regulatory elements of genes upregulated during erythropoiesis

A. Schematic of methodology to identify Cohesin complex interacting proteins at chromatin.

B. Volcano plot showing differences in pull-down of Stag2 between Luc1_HPC (4 experimental replicates) and shS2_1_HPC (4 experimental replicates). The brown colored dots show significantly associated proteins (padj <0.1, log2FC 0.33).

C. Upper panel - Venn diagram displaying validation of significant Stag2-associated proteins from Figure 3B after RIME pull down with Rad21 (4 technical replicates, Luc1_HPC). Lower panel right - Putative Cohesin-interacting proteins that bind together at chromatin; Lower panel left - top 5 enrichments of gene ontology (GO) molecular functions of the final full list of cohesin interactors (scale is -log10(qvalue)).

D. Average profiles demonstrating the enrichment of the specified pre-bound hTF at the erythroid differentiation tiers in unstimulated HPC.

E. Co-immunoprecipitation experiments confirm interaction of Etv6 with Stag2 and Rad21.

Figure 4. Cohesin deficiency severely impairs erythroid differentiation but only modestly alters HSPC homeostasis

A. Experimental design to determine Cohesin dynamics following perturbation of Stag2.

B. Average profiles of Rad21 binding at active promoters and active enhancers in Luc_HPC and shS2_HPC cells. To minimize global quantitative biases, the Rad21 signal was CPM normalized and internally calibrated to its signal at the insulator/other sites (see Methods). Shown are means of 2 replicates, while each shRNA was considered an independent replicate (Luc1_HPC, Luc2_HPC, shS2_1 HPC, shS2_2_HPC).

C. Differential gene expression between Luc_HPC and shS2_HPC cells (3 biological replicates representing different knock-down clones per group: Luc1_HPC, Luc2_HPC, Luc3_HPC, shS2_1_HPC, shS2_2_HPC, shS2_3_HPC) demonstrates minimal gene expression changes. The blue shaded area consists of genes enriched in HPC, while the brown shaded area shows genes significantly expressed in

shS2_HPC cells. The numbers at the upper corners show the counts of log2FC >0.5, padj <0.1 and (log2FC >1, padj <0.05) significant differential genes. For comparison, the significance and difference spreads were set intentionally to match those in Figure 4F.

D. Representative flow cytometric analysis of erythroid differentiation following EPO for Luc_Ery and shS2_Ery at the indicated times.

E. Variance by PCA analysis of the shCohesin_HPC, Luc_HPC, shS2_ery and Luc_ery RNASeq datasets. Each dot represents a different shRNA (Luc1-3_HPC, shS2_1-3_HPC, shR21_1-3_HPC, shSmc1a_1-2_HPC, Luc1-2_ery, shS2_1-2_ery).

F. Differential gene expression between Luc_ery and shS2_ery cells (2 biological replicates for each group - Luc1_ery, Luc2_ery, shS2_1_ery, shS2_2_ery). The numbers at the upper corners show the counts of log2FC >0.5, padj <0.1 and (log2FC >1, padj <0.05) significantly differentially expressed genes.

G. NES for significant (FDR > 0.05 datasets) hematopoiesis-related datasets as determined by GSEA. Input was a ranked list of all pajd < 0.1 genes from figure 4F.

H. Venn diagrams of gene expression changes displaying overlaps between Erythroid or Immature genes from Fig. 2F and Luc_ery or shS2_ery genes from Fig. 4F.

Figure 5. Impaired erythroid differentiation directly relates to decreased Cohesin complex binding at erythroid-specific genes

A. Rad21 binding at active promoters or active enhancers in Luc_ery (red) and shS2_ery cells (black). To minimize global quantitative biases, the Rad21 signal was CPM normalized and internally calibrated to its signal at the insulator/other sites. Shown are means of 2 replicates (Luc1_ery, Luc2_ery, shS2_1_ery, shS2_2_ery).

B. Differential Rad21 binding in Luc_ery and shS2_ery cells at the indicated erythroid differentiation tiers from Fig. 2B. Box plots to the left of zero indicate increased binding following Stag2 knockdown and to the right indicate increased binding in control knockdown cells that follow normal erythroid differentiation. Shown are means of 2 replicates as described in Figure 5A.

C. Differential H3K27ac binding as Fig. 2B. Shown are means of 2 replicates as described in Figure 5A.

D. Differential Rad21 binding at active promoters between Luc_ery (red) and shS2_ery cells (blue). Shown are means of 2 replicates as described in Figure 5A.

E. H3K27ac binding at the sites from Fig5D. Shown are means of 2 replicates as described in Figure 5A.

F. Gene expression differences between Luc_ery and shS2-ery cellular states at the indicated sites from Fig5D. Genes that were enriched in Luc_ery were marked with negative values (shades of red), while genes that were enriched in shS2_ery were marked with positive values (shades of blue). N.e. = not expressed. Selected genes of interest are noted.

G. Interaction frequencies at the indicated sites from Fig5D (at gained promoters - blue; at lost promoters - red) are shown as the difference of normalized significant interactions between shS2-Ery and Luc-Ery conditions. Shown are means of 2 replicates as described in Figure 5A.

H. Example of Rad21 and H3K27ac binding dynamics, as well as interaction frequency between Luc_ery and shS2_ery at the *Klf1* promoter (RNASeq log2foldChange -1.8).

Figure 6. *Stag2* perturbation leads to preferential loss of Cohesin binding at differentiation promoters with strong Etv6 pre-binding

A. Representation of the erythroid perturbation tiers; differences of Rad21 binding in Luc-Ery and shS2-Ery cells (Rad21-Ery minus shS2-Ery) in the indicated erythroid perturbation tiers. Box plots to the left of zero indicate increased binding following Stag2 knockdown and to the right indicate increased binding in control knockdown cells with normal erythroid differentiation.

B. Differential H3K27ac binding as in Fig6A.

C. Differential significant interaction frequencies at the promoters of the same regions as in Fig6A.

D. Overlaps between annotated promoters from the indicated perturbation tiers and differentially expressed genes from Fig4F.

E. Enrichment of the specified hTF at the perturbation tiers in wild-type HPC.

F. Enrichment of Etv6 at the indicated regions.

G. Immunoblotting for total protein expression of Etv6, Stag2 and Actin in the indicated conditions.

H. Immunoblotting for protein expression of Etv6, Stag2 H3 and Actin in different cellular fractions in the indicated conditions.

I. Immunoblotting for protein expression of Etv6, Stag2 H2AX and Actin in different cellular fractions in the indicated conditions.

J. ChIP-qPCR of Etv6 binding at *Klf1, Epor* and *Cxcr4* promoter regions in the indicated cellular states. Shown are results from 3 experimental replicates.

Figure 7. Erythropoiesis is impaired in human Cohesin-mutated Myelodysplastic Syndromes and Acute Myeloid Leukemia

A. Differential gene expression in STAG2 mutated MDS vs all other MDS subtypes. Only events with FDRq<0.1 are shown. The red (enriched in other MDS) and black (enriched in shSTAG2mut MDS) segments specify for log2FC >0.5 differential genes. Genes of interest are shown.

B. NES for significant hematopoiesis-related datasets (FDR <0.05) as determined by GSEA. Input was a ranked list of all genes from figure 7A. STAG2 mutant-enriched genesets are to the right of 0 and other MDS subtype genesets to the left.

C. and D. Common and exclusive mutations of the indicated genes among 3612 myeloid neoplasia curated by the Cosmic database (C) and ~~2434~~ 3256 AML samples from the UK NCRI AML trials (2434 samples), BEAT-AML (622) and TCGA LAML (200) datasets (D).

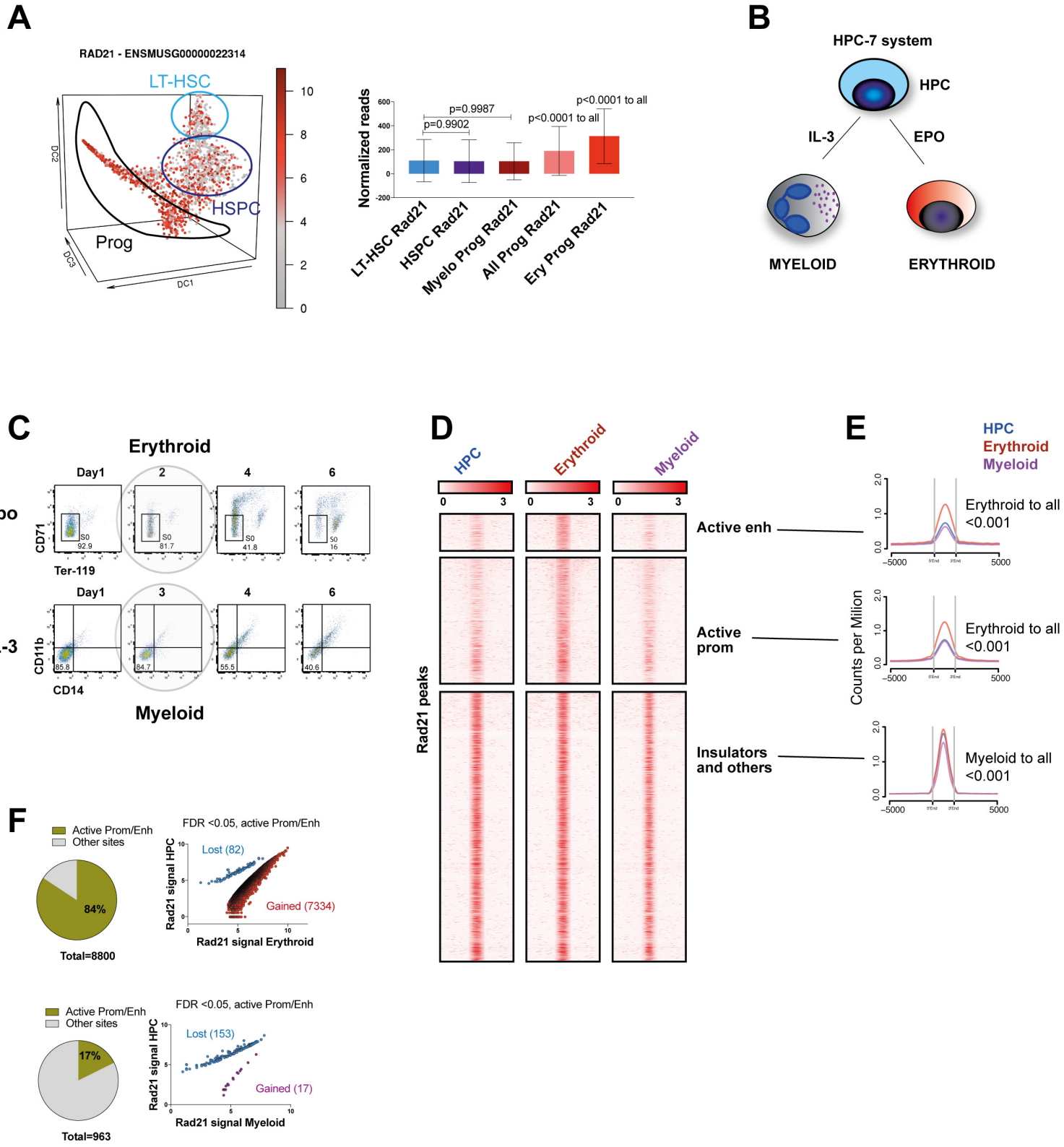

A
RAD21 - ENSMUSG00000022314
LT-HSC
HSPC
Prog
DC2
DC3
DC1
10
8
6
4
2
0
Normalized reads
600
400
200
0
-200
p=0.9987
p=0.9902
p<0.0001 to all
p<0.0001 to all
LT-HSC Rad21
HSPC Rad21
Myelo Prog Rad21
All Prog Rad21
Ery Prog Rad21
B
HPC-7 system
HPC
IL-3
EPO
MYELOID
ERYTHROID
C
Erythroid
Epo
Day1
2
4
6
CD71
Ter-119
S0
92.9
S0
81.7
S0
41.8
S0
16
IL-3
Day1
3
4
6
CD11b
CD14
85.8
84.7
55.5
40.6
Myeloid
D
HPC
Erythroid
Myeloid
0
3
Rad21 peaks
Active enh
Active prom
Insulators and others
E
HPC
Erythroid
Myeloid
Erythroid to all <0.001
Erythroid to all <0.001
Myeloid to all <0.001
Counts per Milion
2.0
1.0
0.0
-5000
5000
F
Active Prom/Enh
Other sites
84%
Total=8800
FDR <0.05, active Prom/Enh
Lost (82)
Gained (7334)
Rad21 signal HPC
Rad21 signal Erythroid
Active Prom/Enh
Other sites
17%
Total=963
FDR <0.05, active Prom/Enh
Lost (153)
Gained (17)
Rad21 signal HPC
Rad21 signal Myeloid

Figure 1

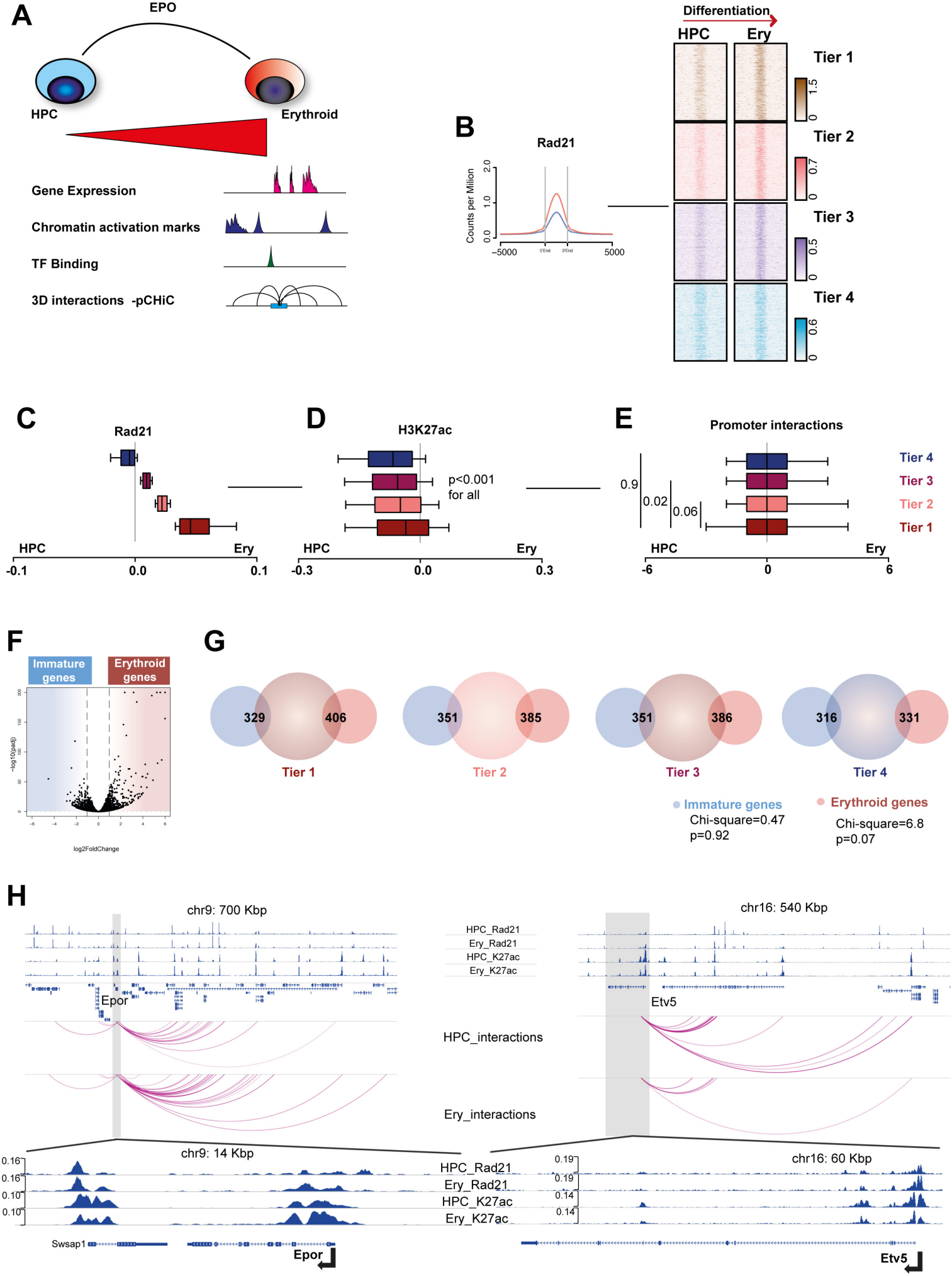

A
EPO
HPC
Erythroid
Gene Expression
Chromatin activation marks
TF Binding
3D interactions -pCHiC
B
Rad21
Counts per Milion
Differentiation
HPC
Ery
Tier 1
Tier 2
Tier 3
Tier 4
C
Rad21
HPC
Ery
D
H3K27ac
p<0.001
for all
E
Promoter interactions
Tier 4
Tier 3
Tier 2
Tier 1
F
Immature genes
Erythroid genes
log2FoldChange
G
329
406
Tier 1
351
385
Tier 2
351
386
Tier 3
316
331
Tier 4
Immature genes
Chi-square=0.47
p=0.92
Erythroid genes
Chi-square=6.8
p=0.07
H
chr9: 700 Kbp
Epor
chr16: 540 Kbp
Etv5
HPC_Rad21
Ery_Rad21
HPC_K27ac
Ery_K27ac
HPC_interactions
Ery_interactions
chr9: 14 Kbp
chr16: 60 Kbp
Swsap1
Epor
Etv5

Figure 2

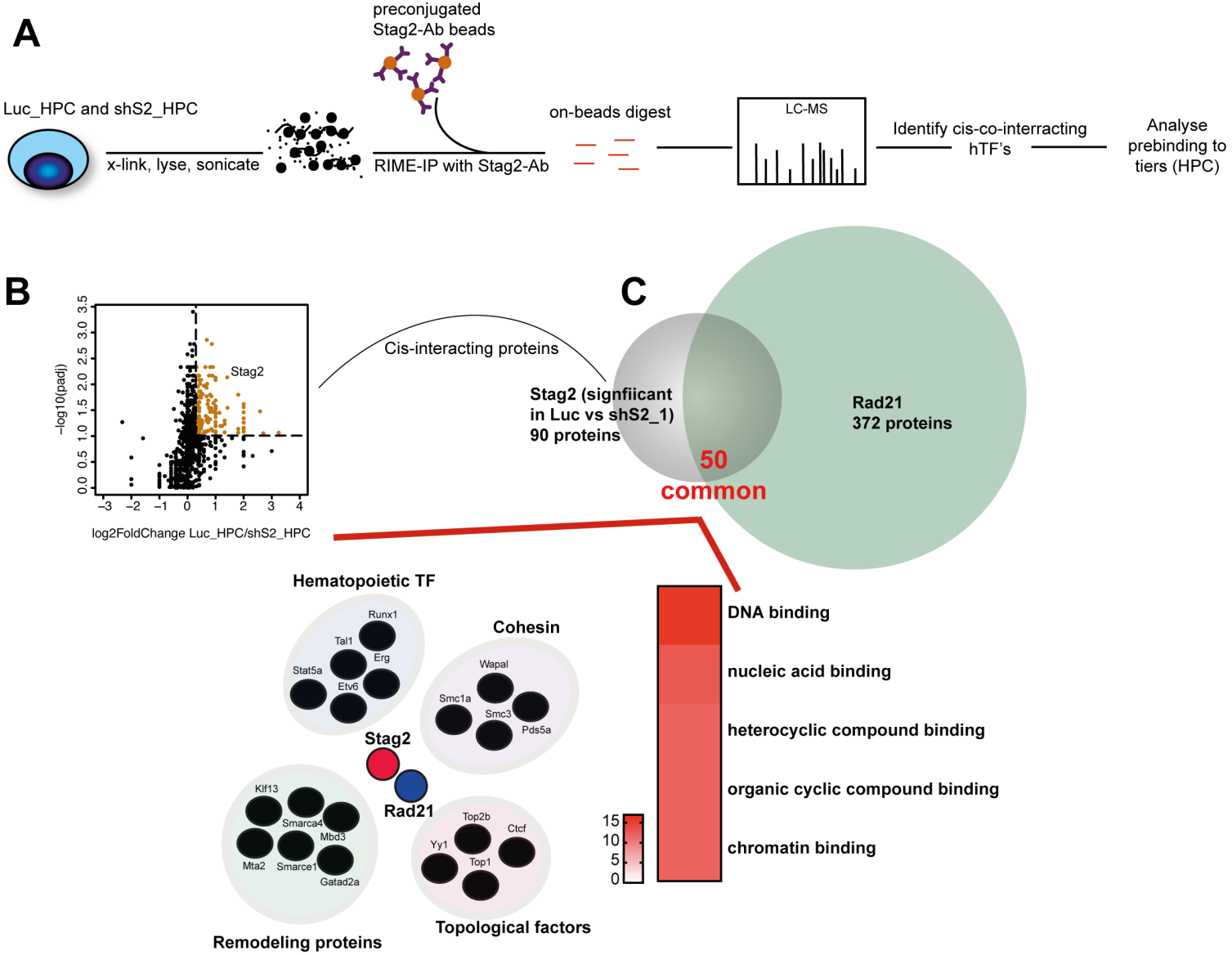

A
preconjugated
Stag2-Ab beads
Luc_HPC and shS2_HPC
x-link, lyse, sonicate
RIME-IP with Stag2-Ab
on-beads digest
LC-MS
Identify cis-co-interracting
hTF's
Analyse
prebinding to
tiers (HPC)
B
Stag2
−log10(padj)
log2FoldChange Luc_HPC/shS2_HPC
Cis-interacting proteins
C
Stag2 (signfiicant
in Luc vs shS2_1)
90 proteins
Rad21
372 proteins
50
common
Hematopoietic TF
Runx1
Tal1
Erg
Stat5a
Etv6
Cohesin
Wapal
Smc1a
Smc3
Pds5a
Stag2
Rad21
Klf13
Smarca4
Mbd3
Mta2
Smarce1
Gatad2a
Top2b
Ctcf
Yy1
Top1
Remodeling proteins
Topological factors
DNA binding
nucleic acid binding
heterocyclic compound binding
organic cyclic compound binding
chromatin binding
D
Etv6
Tal1
uStat5
Runx1
Erg
5'End 3'End
Tier 1
Tier 3
Tier 2
Tier 4
E
IP:IgG
IP:Rad21
IP:Stag2
IB:Etv6
IB:Rad21
IB:Stag2

Figure 3

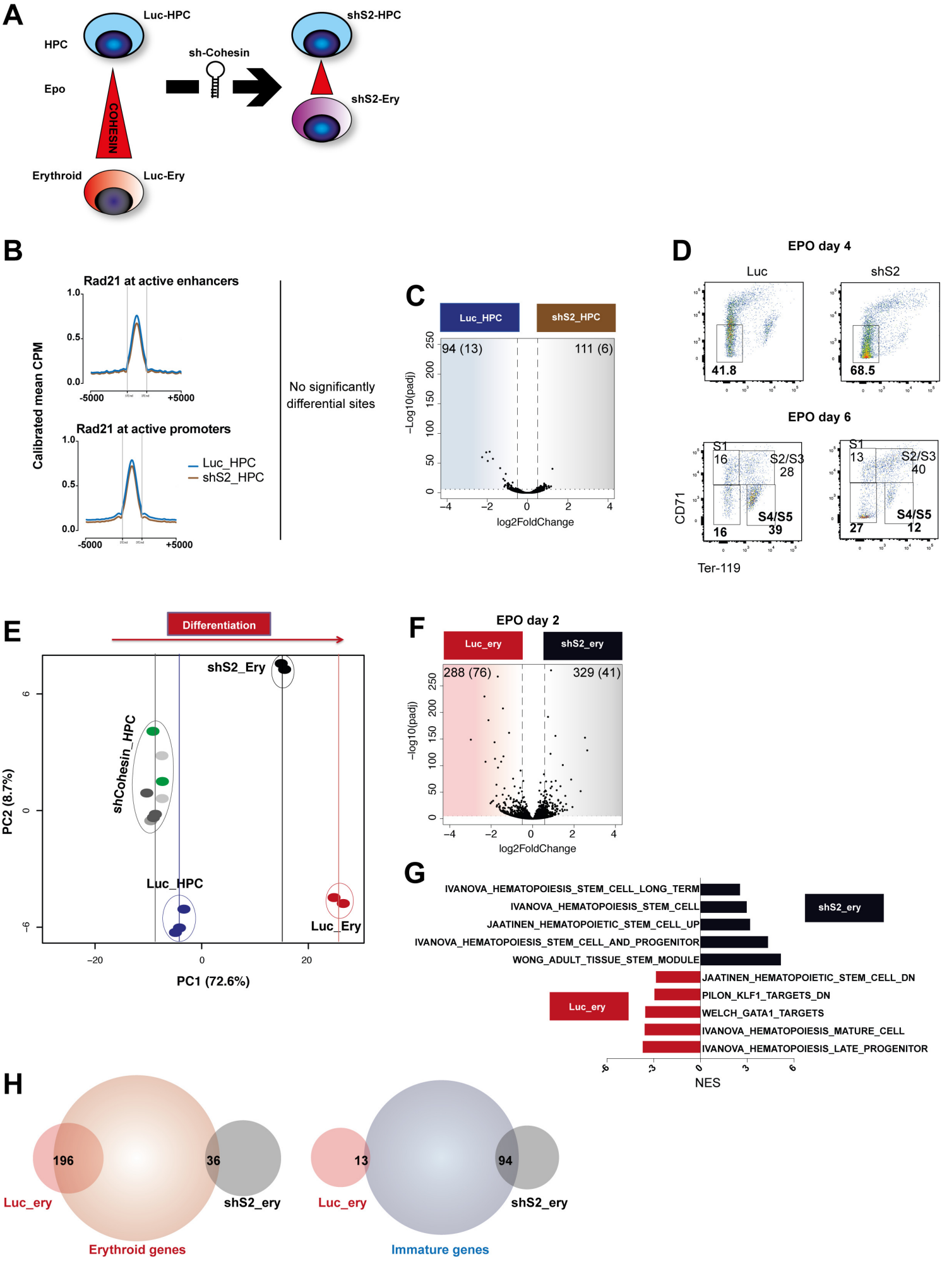

A
Luc-HPC
HPC
sh-Cohesin
shS2-HPC
Epo
COHESIN
shS2-Ery
Erythroid
Luc-Ery
B
Rad21 at active enhancers
Calibrated mean CPM
Rad21 at active promoters
Luc_HPC
shS2_HPC
-5000
+5000
No significantly differential sites
C
Luc_HPC
shS2_HPC
94 (13)
111 (6)
−Log10(padj)
log2FoldChange
D
EPO day 4
Luc
shS2
41.8
68.5
EPO day 6
S1
16
S2/S3
28
S4/S5
39
S1
13
S2/S3
40
S4/S5
12
CD71
Ter-119
E
Differentiation
shS2_Ery
shCohesin_HPC
Luc_HPC
Luc_Ery
PC2 (8.7%)
PC1 (72.6%)
F
EPO day 2
Luc_ery
shS2_ery
288 (76)
329 (41)
−log10(padj)
log2FoldChange
G
IVANOVA_HEMATOPOIESIS_STEM_CELL_LONG_TERM
IVANOVA_HEMATOPOIESIS_STEM_CELL
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP
IVANOVA_HEMATOPOIESIS_STEM_CELL_AND_PROGENITOR
WONG_ADULT_TISSUE_STEM_MODULE
shS2_ery
JAATINEN_HEMATOPOIETIC_STEM_CELL_DN
PILON_KLF1_TARGETS_DN
WELCH_GATA1_TARGETS
IVANOVA_HEMATOPOIESIS_MATURE_CELL
IVANOVA_HEMATOPOIESIS_LATE_PROGENITOR
Luc_ery
NES
H
196
36
Luc_ery
shS2_ery
Erythroid genes
13
94
Luc_ery
shS2_ery
Immature genes

Figure 4

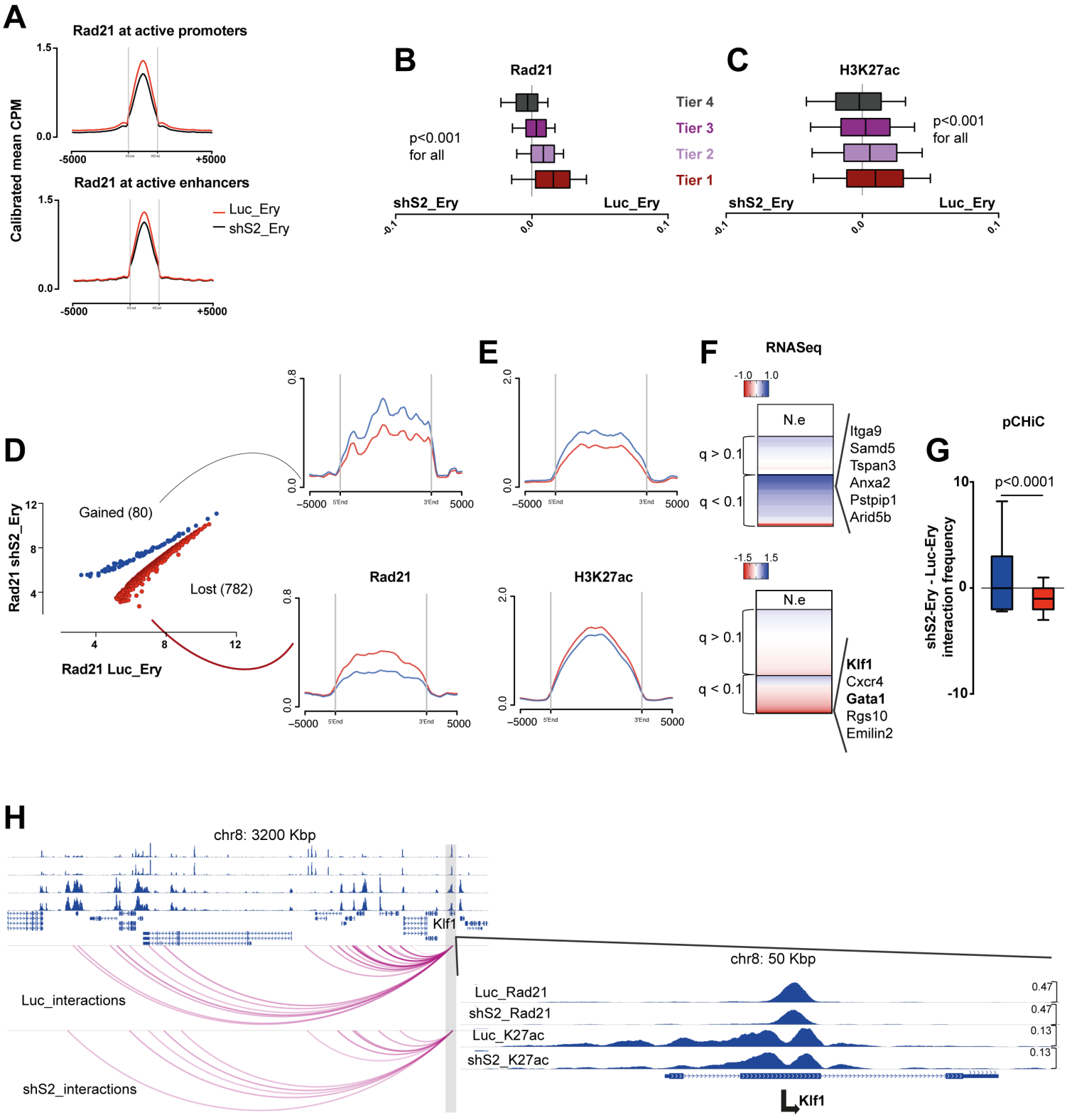

A
Rad21 at active promoters
Rad21 at active enhancers
Calibrated mean CPM
Luc_Ery
shS2_Ery
B
Rad21
p<0.001
for all
shS2_Ery
Luc_Ery
C
H3K27ac
p<0.001
for all
Tier 4
Tier 3
Tier 2
Tier 1
D
Gained (80)
Lost (782)
Rad21 shS2_Ery
Rad21 Luc_Ery
E
Rad21
H3K27ac
F
RNASeq
N.e
q > 0.1
q < 0.1
Itga9
Samd5
Tspan3
Anxa2
Pstpip1
Arid5b
Klf1
Cxcr4
Gata1
Rgs10
Emilin2
G
pCHiC
p<0.0001
shS2-Ery - Luc-Ery
interaction frequency
H
chr8: 3200 Kbp
Klf1
Luc_interactions
shS2_interactions
chr8: 50 Kbp
Luc_Rad21
shS2_Rad21
Luc_K27ac
shS2_K27ac
Klf1

Figure 5

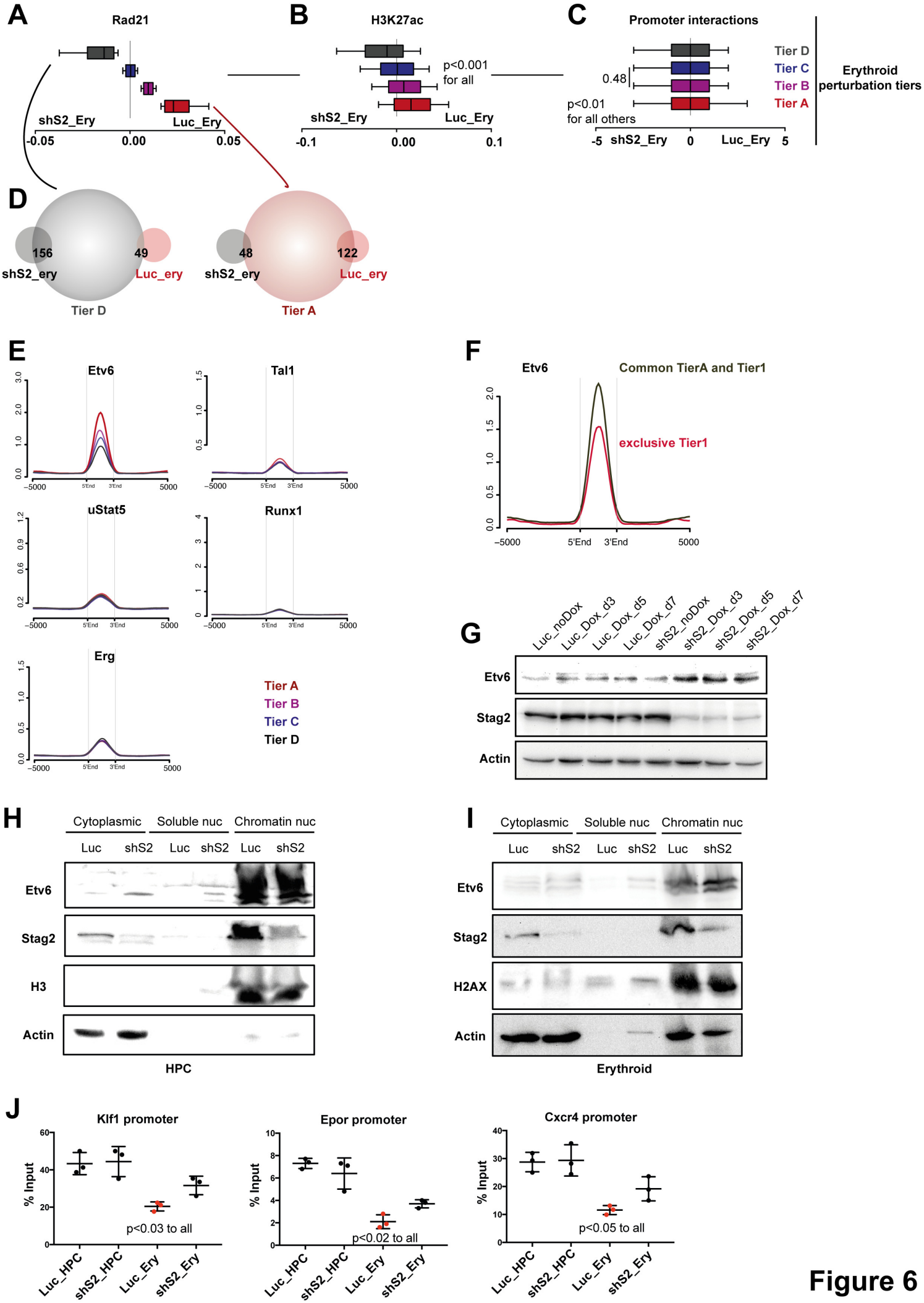
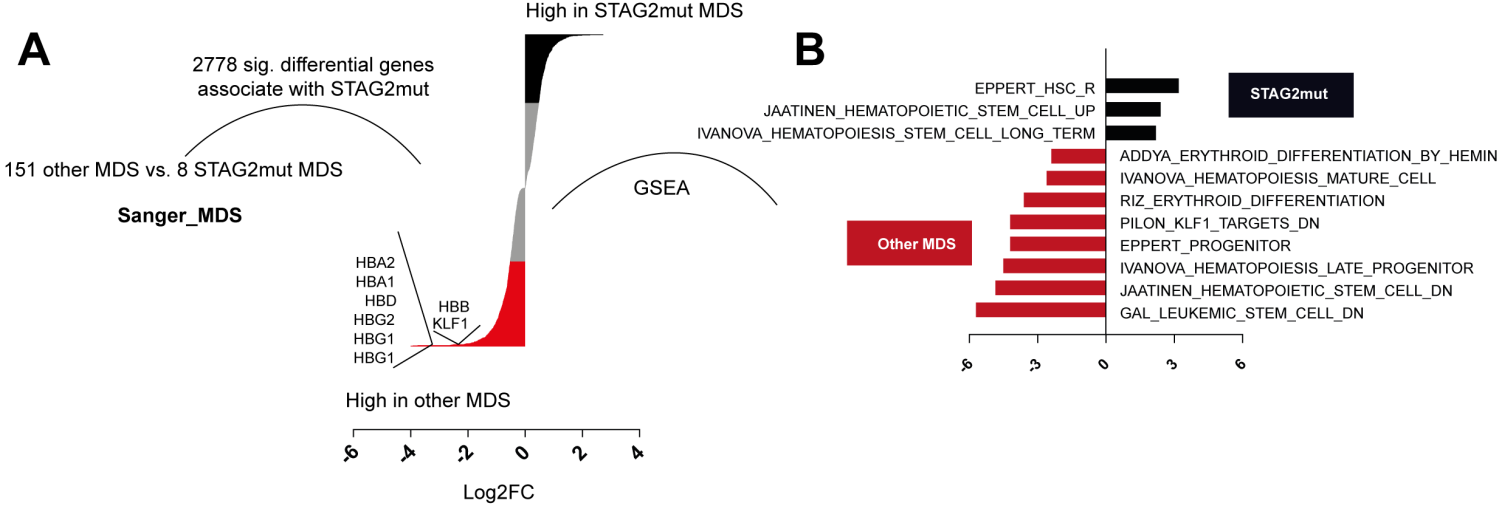
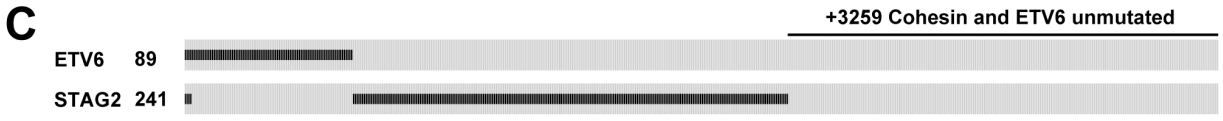

A
Rad21
shS2_Ery
Luc_Ery
-0.05
0.00
0.05
B
H3K27ac
p<0.001
for all
shS2_Ery
Luc_Ery
-0.1
0.00
0.1
C
Promoter interactions
0.48
p<0.01
for all others
-5
shS2_Ery
0
Luc_Ery
5
Tier D
Tier C
Tier B
Tier A
Erythroid
perturbation tiers
D
156
49
shS2_ery
Luc_ery
Tier D
48
122
shS2_ery
Luc_ery
Tier A
E
Etv6
Tal1
uStat5
Runx1
Erg
5'End
3'End
Tier A
Tier B
Tier C
Tier D
F
Etv6
Common TierA and Tier1
exclusive Tier1
5'End
3'End
G
Luc_noDox
Luc_Dox_d3
Luc_Dox_d5
Luc_Dox_d7
shS2_noDox
shS2_Dox_d3
shS2_Dox_d5
shS2_Dox_d7
Etv6
Stag2
Actin
H
Cytoplasmic
Soluble nuc
Chromatin nuc
Luc
shS2
Etv6
Stag2
H3
Actin
HPC
I
Cytoplasmic
Soluble nuc
Chromatin nuc
Luc
shS2
Etv6
Stag2
H2AX
Actin
Erythroid
J
Klf1 promoter
Epor promoter
Cxcr4 promoter
% Input
p<0.03 to all
p<0.02 to all
p<0.05 to all
Luc_HPC
shS2_HPC
Luc_Ery
shS2_Ery

Figure 6





Fisher's exact test p=0.4 for mutual exclusivity

Fisher's exact test p=0.028, Odds ratio 0.17 (ETV6 vs. all Cohesin member mutations) for mutual exclusivity

Figure 7