Diverse Heterochromatin-Associated Proteins Repress Distinct Classes of Genes and Repetitive Elements

Ryan L. McCarthy[1,2,3,*], **Kelsey E. Kaeding**[1,2,3,*], **Samuel H. Keller**[5], **Yu Zhong**[6], **Liqin Xu**[6], **Antony Hsieh**[4], **Yong Hou**[5], **Greg Donahue**[1,2,3], **Justin S. Becker**[1,2,3], **Oscar Alberto**[1,2,3], **Bomyi Lim**[5], **Kenneth S. Zaret**[1,2,3]

[1]Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA

[2]Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA

[3]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA

[4]Division of Gastroenterology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA

[5]Department of Chemical and Biomedical Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[6]BGI-Shenzhen, Shenzhen 518083, China

Summary

Heterochromatin, typically marked by histone H3 trimethylation on lysine 9 (H3K9me3) or lysine 27 (H3K27me3), represses different protein-coding genes in different cells, as well as repetitive elements. The basis for locus specificity is unclear. Previously we identified 172 proteins embedded in sonication-resistant heterochromatin (srHC) harboring H3K9me3. Here we investigate in humans how 97 of the H3K9me3-srHC proteins repress heterochromatic genes. We reveal four groups of srHC proteins that each repress many genes and repeat elements in common. Two groups repress H3K9me3-embedded genes with different extents of flanking srHC, one group is specific for srHC genes with H3K9me3 and H3K27me3, and one group is specific

---

Corresponding author: Dr. Ken Zaret zaret@pennmedicine.upenn.edu, telephone: 215-573-5813, fax: 215-573-5844.
*These authors contributed equally

for genes with srHC as the primary feature. We find that enhancer of rudimentary homolog (ERH) is conserved from *S. pombe* in repressing meiotic genes and, in humans, now represses other lineage-specific genes and repeat elements. The study greatly expands our understanding of H3K9me3-based gene repression in vertebrates.

Keywords

heterochromatin; H3K9me3; gene repression; reprogramming; differentiation; hiHep; cell identity

Main

Heterochromatic repression of genes and repetitive elements maintains cell identity and genome integrity[1]. Heterochromatin consists of compacted arrays of nucleosomes that can be mapped as sonication-resistant heterochromatin (srHC), which is transcriptionally silent and contains histone domains enriched for H3K9me3, H3K27me3, both marks, or neither [1–3]. H3K9me3 has classically been associated with silencing of repeat elements, including transposons and centromeric repeats[4], while H3K27me3 has been demonstrated to repress developmentally regulated genes[3]. Recent findings show that H3K9me3 orchestrates repression of genes during development[1,5–7] and impedes binding of transcription factors[2,8]. The orchestration of these changes, maintaining repression at certain genes, while others may lose repression, as well as repressing repetitive elements, suggests that H3K9me3 heterochromatin may be controlled by diverse mechanisms.

H3K9me3 in humans is established by the three lysine methyltransferases (KMTs) SUV39H1, SUV39H2, and SETDB1[9,10]. Recruitment of H3K9 KMTs can be facilitated by transcription factors, such as KRAB-ZNFs[11], or by RNA-dependent mechanisms[12], but how H3K9me3 is established and maintained at different classes of protein coding genes, non-coding genes, and repeat elements remains to be determined. In the fission yeast *S. pombe*, where H3K9 methylation is catalyzed solely by the single histone methyltransferase (HMT) Clr4[13], and in plants, the relevant mechanisms are more fully understood[14]. Facultative heterochromatin formation in *S. pombe* is established by an RNA-dependent mechanism where nascent transcripts, especially from meiotic genes, are bound by the Erh1-Mmi1 complex (EMC) containing Erh1 and Mmi1[15]. The EMC complex targets a transcript for degradation and recruits Clr4 to methylate H3K9 and enforce silencing to repress sporulation[16,17], the primary differentiation capability of fission yeast. In *S. pombe* and plants, constitutive heterochromatin, at regions including repeats and centromeres, occurs via the cooperation of RNA interference (RNAi) and an RNA-dependent RNA polymerase complex[18–20] that is absent from metazoans,[21] leaving open the question of how most repeat elements in humans are targeted for H3K9me3 repression.

Utilizing the property of crosslinking and sonication resistance to separate euchromatin and heterochromatin for sequencing and proteomic analysis, we previously identified 172 H3K9me3 srHC-associated proteins, including known constituents of heterochromatin, such as HP1α[22], but also many proteins with no known role in heterochromatin[2]. By functionally assessing the srHC proteins and identifying the genes that they repress, we sought to better

understand how heterochromatin is maintained and could be manipulated to enable selective gene accessibility.

srHC proteins repress genes and repeats in heterochromatin

To assess the function of srHC associated proteins, we performed two successive cycles of siRNA depletion of 94 srHC proteins, chosen from the 172 srHC proteins[2], to include proteins with known and unknown roles in heterochromatin based upon literature review, as well as of the three H3K9me3 KMTs, SUV39H1, SUV39H2, and SETDB1, in human primary fibroblast cells (Fig. 1a). Knockdown of heterochromatin proteins via the transient nature of siRNA would allow temporary target gene access, e.g., for a cell fate change, and then re-utilization of the proteins to establish new heterochromatin states as needed. We chose the most robust of two independently targeting siRNAs, in duplicate, for each srHC protein target, validating knockdown efficiency by RT-qPCR and a limited number by Western blots, with 75 siRNAs exhibiting greater than 70% mRNA depletion (Extended Data Fig. 1a, b). Global gene induction of srHC genes, as previously annotated[2], was assessed by RNA-seq relative to a non-targeting siControl, in order to identify srHC embedded genes and repeat elements that were upregulated by depletion of srHC associated proteins. We separately analyzed hepatic[23], neural[24], pluripotent[25], spermatogenic[26], and oogenic[27] srHC genes, defined as those transcriptionally silent in fibroblasts but with elevated expression ($p < 0.05$, 2-fold) in the indicated lineage (Fig. 1b). We found that 8 of the 97 depleted srHC proteins significantly activated from 170 to 376 srHC genes, often in more than one lineage category, with many of the activated srHC genes commonly co-repressed by ERH, SUV39H1, RBMX, XRN2 and ZNF207 (Fig. 1b, c). RNA from additional cells treated with a second siRNA was sequenced for siERH, siSUV39H1, siRBMX, siXRN2, and siZNF207 and the results from both siRNAs in duplicate are reported for these targets (Fig. 1a). Knockdown of srHC proteins, including GATAD2A and ERH, impaired cell proliferation (Extended Data Fig. 1c).

Knockdown of ERH, a protein highly enriched in srHC[2] (Extended Data Fig. 2a), and of SUV39H1 significantly activated the greatest total number of srHC genes, 376 and 219 respectively, with 65 genes in common (Fig. 1c). To assess whether ERH controls RNA processing[16] in this context, we measured pre-RNA induction for 3 srHC genes that were highly induced by ERH depletion and found their pre-RNA levels to also be elevated relative to siControl (Extended Data Fig. 2b). Thus, the primary role of ERH here is in transcriptional control. While srHC embedded spermatogenic genes were most profoundly repressed by ERH and SUV39H1 (>5% srHC gene activation), knockdown of 33 of the other 97 targeted srHC proteins, including UBE2I, which represses iPS reprogramming[28], also led to upregulation of at least 3% of the srHC spermatogenic genes[26] (Fig. 1b). Indeed, motifs for spermatogenic transcription factors were enriched in the promoters of induced spermatogenic genes (Extended Data Fig. 2c) and various spermatogenic transcription factors are expressed in the starting fibroblast cells (Extended Data Fig. 2d), compared to the lack of expression of transcription factors for the other lineages tested (Fig. 1d). Among srHC genes that remained repressed in our screen, we observed enrichment of promoter motifs for OCT4, SOX17, and HOXA9 that are not expressed in human fibroblasts, while non-spermatogenic srHC genes exhibiting activation after ERH knockdown were enriched

for motifs bound by factors with high expression in fibroblasts, such as ETV5 and NFYB (Fig. 1e); crucially the expression of these factors was itself not induced by ERH depletion (Extended Data Fig. 2e). The correspondence of activated srHC genes to the presence of activating factors expressed in the starting human fibroblasts suggests that the combination of de-repression of heterochromatin, assessed here through the proxy of gene activation, and an activating transcription factor are required for expression.

We used SalmonTE[29] to assess potential repeat element activation following heterochromatin protein knockdown (Fig. 1b). ERH depletion led to the induction of multiple ERV types and SINEs, while depletion of other factors allowed for activation of other repeat classes, such as for POLDIP3 depletion, which allowed for robust activation of LINE-1 and SINEs (Fig. 1b; Extended Data Fig. 2f, orange and purple arrows). No single knockdown showed robust activation of all ERV, LINE and SINE elements, indicating a diversity in mechanisms for silencing of repeat elements (Fig. 1b).

ERH maintains H3K9me3 heterochromatin in human cells

Orthologs with high sequence similarity to human ERH exist in *S. pombe* and plants such as *A. thaliana*[17], but ERH orthologs are notably absent from *S. cerevisiae,* which lacks H3K9 methylation, and *N. crassa*, where the H3K9 KMT DIM-5 is recruited through a different mechanism[30]. Depletion of ERH in human fibroblasts was sufficient to globally decrease H3K9me3, as measured by immunofluorescence (Fig. 2a; Extended Data Fig. 3a), despite not decreasing the expression of the H3K9me3-methyltransferase genes (Extended Data Fig. 3b). ERH depletion in another cell line, HepG2, also decreased H3K9me3 by immunofluorescence (Extended Data Fig. 3c). We investigated if the change in H3K9me3 may be due to changes of HMT production and found that, upon ERH depletion, the total cellular levels of SUV39H1, SUV39H2, and SetDB1 were unperturbed (Extended Data Fig 3d). Yet strikingly, upon ERH depletion, SUV39H1 abundance decreased markedly within crosslinked and sonicated chromatin (Extended Data Fig. 3e), demonstrating that ERH recruits and/or helps maintain the HMT in heterochromatin. The global role we observe for ERH in humans is in contrast to *S. pombe,* where the RNA-dependent RNA polymerase complex dominates by directing H3K9 methylated heterochromatin at repeats and ERH directs H3K9me3-based repression primarily at meiotic genes[15]. Indeed, the global decrease in H3K9me3 upon *ERH* knockdown in human cells, as well as the induction of srHC-H3K9me3 genes, was greater than that observed for knockdowns of *SUV39H1* or *SETDB1* (Extended Data Fig. 3f). ERH depletion decreased only H3K9me3 and not H3K9me2 (Fig. 2a), as confirmed with multiple antibodies (Extended Data Fig. 3g). In *S. pombe* the ERH complex interacts with MTREC[15] (PAXT in humans), several components of which we previously observed to be enriched on H3K9me3 srHC[2], however depletion of the key PAXT component SKIV2L2 in human fibroblasts did not phenocopy the loss of H3K9me3 (Extended Data Fig. 3h); thus the regulation of H3K9me3 by ERH may function independent of PAXT. Taken together, the results indicate that ERH is integral to the H3K9me3 methylation pathway and has evolved to become a dominant heterochromatin effector in humans.

Knockdown of *ERH* in human fibroblasts significantly upregulated 36 of the 154 human orthologs of meiotic genes observed to be upregulated in an *Erh* knockout *S. pombe*[16] (Fig. 2b). By contrast, knockdowns of *SUV39H1, RBMX,* or *XRN2* upregulated only 18, 14 and 9 meiotic homolog genes, respectively (Fig. 2b). ERH depletion also upregulated more evolutionarily recent spermatogenic genes, including a cluster of *SPANX* family genes (Extended Data Fig. 4a) that are unique to hominids[31]. We conclude that in addition to expanding its targets to repeat elements, human ERH has evolutionarily retained its role in repressing meiotic and gametogenic genes, as seen in fission yeast.

ERH is localized to nascent RNAs in *S. pombe* via binding of Mmi1 to Determinant of Selective Removal (DSR) motifs in RNA, which recruits ERH as part of the EMC complex[15]. However, in contrast to the highly conserved ERH protein[16] no conserved Mmi1 ortholog exists in humans and the closest homologs, YTHDF1–3, show poor conservation of the ERH interacting domain (Extended Data Fig. 4b). Human ERH has been independently detected in nascent RNA proteomics[32], albeit at low levels, and upon examining the R-Deep mass spectrometry database[33], we found that ERH co-fractionates with histones in an RNA-dependent manner (Extended Data Fig. 4c). More significantly, the protein domain by which Mmi1 binds the DSR motif in *S. pombe* is absent from all Mmi1 human orthologs[34]. Thus, the recruitment mechanism for ERH has diverged from that in *S. pombe*[15].

Concordant with the cytological diminution of H3K9me3 after ERH knockdown, H3K9me3 ChIP-seq in *ERH*-knockdown cells, with *Drosophila* chromatin as a spike-in control, revealed loss of H3K9me3 at broad domains containing protein coding genes, non-coding genes, and repeats (e.g., Fig. 2c). Notably, H3K9me3 was significantly decreased at 75% of the 15,154 H3K9me3 domains we previously mapped in human fibroblasts[2] (Fig. 2d). *ERH* knockdown elicited significant H3K9me3 loss for all categories of H3K9me3 domains, annotated based upon the presence of genes or repeat elements, with especially strong H3K9me3 loss at pseudogenes, which are often silenced by H3K9me3 and can encode non-coding RNAs involved in silencing[35] and are often misregulated in cancer[36].

In contrast to ERH function in *S. pombe,* which does not focus on constitutive heterochromatin[15], we observed that nearly all (59 of 60) H3K9me3 domains containing satellite repeats exhibited a strong loss of H3K9me3 in *ERH* knockdowns (Fig. 2e), which may account for much of the global H3K9me3 changes measured by immunofluorescence (Fig. 2a). Further categorization of *ERH* knockdown-induced satellite repeat changes using RepEnrich[37] revealed a loss of H3K9me3 at multiple classes of centromeric satellites, non-centromeric repeats, and telomeric repeats that also exhibit an increase in RNA expression (Fig. 2f). Increased SVA, Alu, and ERV, specifically ERVK, expression was observed in the *ERH* knockdown when analyzed by RepEnrich (Extended Data Fig. 4d,e) and SalmonTE (Fig. 1b). Repeat expression as measured by RNA-seq correlated with decreased H3K9me3; correlation was especially high for Alu and ERVK elements while tRNAs exhibited much H3K9me3 loss but did not display elevated expression (Extended Data Fig. 4d). Most other classes of repeats, including rDNA, which are under control of H3K9me2 and not H3K9me3[38], were unchanged (Extended Data Fig. 4e). We observe an increase in H3K27me3 levels at several classes of satellite repeats observed to lose H3K9me3 (Extended Data Fig. 4f) consistent with previous observations that H3K27me3

may compensate for H3K9me3 loss[39]. In summary, human ERH has gained new functions in repressing genes for diverse lineages as well as satellite repeats.

srHC proteins impede gene activation during reprogramming

To test the hypothesis that activating transcription factors may be needed for the expression of heterochromatic genes of alternative cell fates, we repeated the siRNA depletion of srHC proteins with the addition of the hepatic transcription factors FOXA3, HNF1α, and HNF4α, which elicit reprogramming of fibroblasts to induced hepatocytes[23] (hiHeps) (Fig. 1a). Strikingly, we observed far more extensive activation of srHC genes in the presence of hiHep reprogramming factors, especially for srHC embedded hepatic genes (Fig. 3a). Of the 97 proteins targeted for depletion, 71 allowed an upregulation of at least 93 srHC genes (over 1% of all srHC genes). As suspected, we observed an enrichment of DNA binding motifs for hiHep reprogramming factors in the promoters of hepatic srHC genes that were activated by srHC protein knockdown (Fig. 3b). Analysis of ERV, LINE and SINE sequences downloaded from RepeatMasker for hiHep factor motifs revealed FOXA3 motifs to frequently occur in LINE elements (Extended Data Fig. 4g), potentially accounting for the LINE activation in our siRNA knockdowns during hiHep reprogramming (Fig. 3a). Of four srHC protein knockdowns (SUV39H1, RBMX, ERH, XRN2) tested for increasing hiHep reprogramming efficiency at 14 days, as measured by the percent cells expressing the hepatic cell surface marker ASGPR1 which is encoded by a gene not located in srHC, only depleting RBMX enhanced reprogramming, as seen previously[2] (Extended Data Figure 4h). Visualizing patterns of srHC genes across the genome that were activated by knockdowns revealed that there are chromosomal regions where multiple srHC proteins co-repress blocks of srHC genes, while in other chromosomal regions, srHC genes remain recalcitrant to nearly all perturbations (Extended Data Fig. 5; see expanded example regions), perhaps reflecting the H3K9me3-srHC subset bias from our original proteomic study of srHC proteins.

srHC protein groups target distinct heterochromatin types

To identify which groups of srHC proteins may exhibit specificity for common targets, we performed a nearest neighbor clustering of the 73 most impactful proteins, based upon the overlapping sets of srHC genes they repressed. Four clusters emerged, with 9 hub nodes at the center of multiple peripheral nodes (Fig. 3c). Although the clustering was based upon genetic evidence, it independently validated previously reported protein interactions, including XIST-mediated gene silencing components RBM15, YTHDC1[40] and ZC3H13[41] around the central hub of cluster 2, HUSH complex interactors PPHLN1 and SETDB1[42] in a sub-cluster of cluster 2, and paraspeckle components NONO, FUS and SFPQ[43] in cluster 3 (Fig. 3c).

We assessed heterochromatin and functional parameters of the srHC genes that were selectively regulated by each cluster. Cluster 1 proteins, which includes the hub proteins ERH and HMGA1, as well as SUV39H1, repress srHC genes with high levels of H3K9me3, low levels of H3K27me3, and high srHC across their gene bodies and flanking regions (Fig. 3d, meta-analysis; Extended Fig. 6, individual gene examples). Cluster 1 represses genes

relating to cell adhesion and tissue-specific processes (Extended Data Fig. 6, 7a) and were especially enriched for B-compartment[44] localization (Extended Data Fig. 7b).

Cluster 2 proteins repress srHC genes with moderate levels of H3K9me3, low H3K27me3, and high srHC levels across the gene body, but low in flanking regions (Fig. 3d; Extended Data Fig. 6; Supplementary Fig. 1), and elevated DNA methylation and H4K20me1 (Extended Data Fig. 7b). srHC genes repressed by cluster 2 span a wide range of biological functions, including metabolism and cell cycle (Extended Data Fig. 7a). Cluster 2 hub proteins MYBBP1A and ZNF622, as well as peripheral cluster 2 proteins, including CBX1, RPF2, MPHOSPH10, and CCDC137, have been previously observed to exhibit complete or partial nucleolar localization[45], and knockdown or loss of MYBBP1A or ZNF622 has been shown to cause significant nucleolar abnormalities[46]. The KMT SETDB1 occupies a different cluster 2 subcluster, along with PPHLN1 (Fig. 3c), a component of the HUSH complex that recruits SETDB1[42].

Cluster 3, with hub proteins CEBPZ and FUS, preferentially represses srHC genes not highly enriched for H3K9me3 or H3K27me3, corresponding to our previously annotated srHC unmarked genes[2], yet with high srHC levels across the gene body and with euchromatic flanking regions (Fig. 3d; Extended Data Fig. 6; Supplementary Fig. 1) and elevated DNA methylation and H4K20me1 (Extended Data Fig. 7b). Cluster 3 selectively represses srHC genes enriched for the biological function of immune response including elements of interferon signaling, a known repressive target of SFPQ[47] (Extended Data Fig. 7a). Cluster 3 also selectively represses srHC genes encoding miRNA target genes by the TargetScan microRNA database[48] (Extended Data Fig. 7c), consistent with the roles of cluster 3 proteins FUS, SFPQ, and HNRNPA2B1 in the miRNA pathway[49,50] and the presence of cluster 3 proteins in paraspeckles[51].

Cluster 4 encompasses hub proteins GATAD2A and ZNF438, as well as the PRC2 complex member SUZ12, and represses srHC genes with high H3K9me3 and high H3K27me3, including many of our previously annotated srHC H3K9me3/K27me3 dual marked genes[2], and that possess high srHC across their gene bodies and flanking regions (Fig. 3d; Extended Data Fig. 6; Supplementary Fig. 1). Cluster 4 members GATAD2A and GATAD2B are components of the NuRD complex, which mediates deacetylation of histones and facilitates recruitment of PRC2 for H3K27me3 methylation[52]. Across all four clusters, we detect varying degrees of H3K9me3 and H3K27me3 heterogeneity across the srHC gene bodies and promoters uniquely regulated by each cluster (Extended Data Fig. 7d). This likely indicates complex maintenance of srHC genes with various histone marks.

In summary, the distinct heterochromatin profiles targeted by each srHC protein cluster validates the genetic method of determining the clusters by the common gene expression changes upon srHC protein knockdown, and demonstrates that srHC proteins selectively repress genes with particular heterochromatin subtypes and biological functions.

Binding of srHC hub proteins to repressed srHC genes

We performed ChIP-seq on the 8 srHC hub proteins and assessed their binding profiles over srHC genes upregulated by each of the srHC hub knockdowns during hiHep reprograming (Fig. 4a). ERH binding was enriched at ERH repressed genes as well as genes repressed by HMGA1 (Fig. 4a), which clustered with ERH based upon shared repressed genes (Fig. 3c). We observed that while several of the srHC hub proteins, particularly HMGA1, MYBBP1A, FUS, and GATAD2A, demonstrated enrichment over the srHC genes upregulated by their depletion, they also displayed enrichment over srHC genes repressed by hubs of other clusters (Fig. 4a). In cases of overlapping enrichment at srHC genes by two factors, there was enrichment over shared repressed srHC genes but not genes functionally repressed solely by the other factor (Extended Data Figure 8a).

ChIP-seq for endogenous ERH demonstrated that it preferentially associates with H3K9me3 domains in a broad pattern (Fig. 4b, c, d), similar to its observed binding pattern in *S. pombe*[15], including in human euchromatic H3K9me3 domains (Fig. 4d; Extended Data Fig. 8b), where it also represses euchromatic H3K9me3 genes (Extended Data Fig. 8c).

We observed that the chromatin immunoprecipitations for various srHC proteins, especially ZNF622 and CEBPz, were enriched for large chromatin fragments relative to the input size profile (Extended Data Fig. 8d); consistent with these factors binding to srHC regions that generate larger fragments during sonication[2]. We had performed paired-end sequencing without size selection to preserve the large fragments in the amplified libraries and observed that large fragments were enriched in srHC domains (Extended Data Fig. 8e). In summary, srHC hub proteins are bound to genes that they functionally repress.

srHC protein maintenance of repressive histone modifications

Quantitative confocal H3K9me3 imaging of fibroblast cells treated with the two independently targeting siRNAs against all 97 target proteins was used to assay global H3K9me3 by indirect immunofluorescence at the single cell level (Extended Data Fig. 9a). Depletion of hub node srHC proteins ERH, HMGA1, MYBBP1A, ZNF622, CEBPz, and FUS decreased total nuclear H3K9me3 levels more than peripheral node srHC proteins, with the exceptions of cluster 4 hub proteins GATAD2A and ZNF438, whose depletions had limited effects on H3K9me3 levels (Fig. 5a, exemplary primary data; b, summary of single-cell analysis of all knockdowns; see also Extended Data Fig. 9). Depletion of XRN2 caused the greatest global decrease in H3K9me3 levels of any peripheral cluster protein (Fig. 5b), consistent with XRN2 regulating H3K9me3 heterochromatin in *S. pombe*[53]. XRN2 in *C. elegans* was recently shown to be involved in H3K27me3 heterochromatin[54], a mark we also observed to decrease by siXRN2 (Extended Data Fig. 9b). YTHDC1 was the only srHC protein whose knockdown elicited significantly elevated nuclear H3K9me3 (Fig. 5b; Extended Data Fig. 9c), likely due to its role as a regulator of mRNA stability of MAT2A, which catalyzes production of the methyl donor S-adenosylmethionine[55] for the HMTs.

Cluster 1 knockdowns caused widespread decreases in H3K9me3 global levels (Fig. 5a, b), but limited decreases of H3K27me3 (Extended Data Fig. 9d, e). Surprisingly, knockdowns of cluster 2 hub proteins resulted in the greatest observed decreases in both H3K9me3 and H3K27me3 (Fig. 5a, b; Extended Data Fig. 9d, e), despite not activating srHC genes with high H3K9me3 and H3K27me3 profiles (Fig. 3d; Extended Data Fig. 7). Cluster 3 knockdowns caused small decreases in global H3K9me3 (Fig. 5a, b) and H3K27me3 (Extended Data Fig. 9d, e). Cluster 4, which contains the hub nodes GATAD2A and ZNF438, as well as SUZ12, exhibited no appreciable effect on H3K9me3 levels (Fig. 5b), consistent with enrichment for repressing H3K27me3 marked srHC genes (Fig. 3d). Despite causing the greatest activation of H3K27me3 marked srHC genes, depletion of cluster 4 node proteins GATAD2A and ZNF438 caused the smallest decrease in H3K27me3 levels globally by immunofluorescence, indicating that the de-repression observed in these knockdowns apparently does not function through global loss of K27 methylation (Extended Data Fig. 9d, e). The extent to which global H3K9me3 levels were depleted after srHC protein knockdown correlates (R=−0.31, p=0.006) with increasing numbers of hepatic srHC genes that could be activated during hiHep reprogramming (Extended Data Fig. 9f).

ERH globally maintains chromatin states at srHC genes

To focus on locus-specific changes caused by siRNA depletion of the cluster 1 hub protein ERH, which represses genes with the strongest H3K9me3 signals (Fig. 3d; Extended Data Fig. 6), we performed H3K9me3 and H3K27me3 ChIP-seq and srHC-seq[5] in siControl and siERH knockdowns (Extended Data Fig. 10a). We first plotted 9275 srHC genes based upon their mean siControl H3K9me3, H3K27me3, and srHC gene body chromatin states in two dimensions, using t-Distributed Stochastic Neighbor Embedding (t-SNE) (Fig. 6a). The graphical representation illustrates the existence of H3K9me3, H3K27me3, H3K9me3/H3K27me3 dual marked and unmarked srHC genes, as we have previously shown[2], as well as their relative srHC levels (Fig. 6a). Varying levels for each mark can be found across srHC genes and drive the t-SNE separation (Extended Data Fig. 10b). ERH depletion led to a twofold or greater decrease in H3K9me3 at 76% of srHC genes (Fig. 6b). Unexpectedly, ERH depletion also led to a twofold or greater loss of H3K27me3 at 67% of srHC genes, though mostly for residual H3K27me3 genes that were not previously annotated as H3K27me3 marked[2] (Fig. 6b). Most significantly, ERH depletion decreased srHC markedly at nearly all srHC genes, regardless of their initial histone modification state (9274 of 9275 total) (Fig. 6b; Extended Data Fig. 10c). Consequently, while the hiHep factors in the *ERH* knockdown activated 13.4% of hepatic srHC genes, they also activated, to a lesser extent, srHC genes in alternative lineages including neural and pluripotency/iPSC (Fig. 6c; Extended Data Fig. 10d). Motif analysis of promoters of non-hepatic srHC genes activated in siERH+hiHep conditions but not siERH alone identified enriched motifs corresponding to factors expressed in hiHep treated fibroblasts (Extended Data Fig. 10e).

With *ERH* knockdown, H3K9me3 loss was variable across several classes of key hepatic genes, with H3K9me3 decreasing markedly at genes encoding cytochrome P450s, but not at those encoding glucuronidases (Extended Data Fig. 10f), indicating network specificity. H3K9me3 was unchanged or gained at 12% and 11% of srHC genes respectively (Extended Data Fig. 10g), with the H3K9me3 gain highly specific for initially unmarked srHC genes

(Fig. 6a). Similarly, a H3K27me3 gain was observed at 18% of srHC genes, primarily at srHC genes initially unmarked by H3K27me3 (Fig. 6b; Extended Data Fig. 10h). We presume these changes to be secondary consequences of loss of ERH at its many target genes.

At srHC genes exhibiting marked H3K9me3 loss upon ERH knockdown during hepatic reprogramming, we observed higher rates of activation (Fig. 6d) among genes with strong hepatic transcription factor motifs (Extended Data Fig. 10i, j). Conversely, for a subset of srHC genes exhibiting H3K9me3 loss that also gain H3K27me3, we observed lower rates of activation (Fig. 6e; Extended Data Fig. 10j), indicating that H3K27me3 may compensate for H3K9me3 loss at some genes. We conclude that ERH has global roles in maintaining chromatin states at srHC genes and modulates the interplay between H3K9me3, H3K27me3, and srHC.

Discussion

We identified and deeply profiled the genetic functions of 97 heterochromatin-associated proteins and revealed unexpected complexity in the basis for heterochromatic gene and repeat element silencing. We also demonstrated how transiently overcoming such silencing can enable alternative-fate genes to be accessed, providing a framework for enhancing the ability to genetically reprogram cells. We find that srHC protein depletion does not usually result in the activation of heterochromatin-embedded genes, but rather imparts permissibility for relevant transcription factors to induce alternative-fate genes. The point illustrates how a transient, global diminution of heterochromatin does not obligatorily cause extensive nonspecific gene activation, but rather how the transcriptional response can be largely predicted by the type of transcription factors expressed in a cell and the exogenous factors that may be added.

We found that ERH is key to global H3K9me3 maintenance in human cells and, remarkably, is evolutionarily conserved with *S. pombe* with regard to its repression of meiotic and germ cell genes. ERH also represses many other lineage-specific genes and satellite repeats, apparently providing a primary mechanism for heterochromatin in metazoans, which lack the RNA-dependent RNA polymerase complex seen in *S. pombe* and plants. Significantly, we found that ERH is required to maintain the H3K9me3 methyltransferase SUV39H1 on heterochromatin.

Previously we had discerned four distinct types of sonication-resistant heterochromatin, including that which was enriched for H3K9me3, H3K27me3, both modifications, or neither[2]. Our collective knockdown analysis of srHC proteins showed that the proteins clustered by their action on genes that grouped into different srHC subtypes, but with a resolution and degrees of overlap not discerned in our original assessment (Fig. 3d). While depletion of many of the tested srHC proteins triggered widespread changes in heterochromatin histone modifications, locus specific analysis revealed that while H3K9me3 loss corresponded to srHC gene activation, H3K27me3 levels remained high or even increased at transcriptionally de-repressed srHC genes, illustrating that for those genes, the H3K9me3-based mechanisms was the dominant repressive effector.

Our finding that clusters of proteins regulate specific srHC subtypes and different gene classes provides a framework to explore how the srHC proteins govern heterochromatin regulation at selective regions of the genome. Further work will address the mechanisms governing the regulation of the srHC subtypes, enabled by the apparent collaboration between srHC proteins predicted by our clustering. The present study has expanded our understanding of the complexity of heterochromatin maintenance in vertebrates, uncovered ERH as a key player in mammalian H3K9me3 regulation, and provides a resource of srHC proteins and gene targets at heterochromatin subtypes that can be modulated to enhance cellular reprogramming.

Methods

Cell culture

Human BJ foreskin fibroblasts were obtained from Stemgent (08–0027) at passage 6 and cultured in Eagle's Minimum Essential Medium (EMEM) (Sigma-Aldrich M2279) supplemented with 10% fetal bovine serum (FBS, Hyclone SH30071) and 2mM L-glutamine (GIBCO 25030149) at 37°C and 5% CO_2. HepG2 cells were obtained from ATCC (HB-8065) and cultured in Dulbecco's Modified Eagle Medium (Hyclone SH30022.02) supplemented with 10% fetal bovine serum (FBS, Hyclone SH30071) and 2mM L-glutamine (GIBCO 25030149) at 37°C and 5% CO_2. Published reports with HepG2 cells have been previously noted as being of various origins[56]. For our assays only HepG2 cells directly from ATCC were used and the reported differences in liver specific function between real and misidentified HepG2 lines would not be expected to impact the H3K9me3 levels we measured in these cells.

Lentivirus Production

Lentiviral plasmids pWPI.1-FOXA3, pWPI.1-HNF1A, and pWPI.1-HNF4A were kindly provided by the laboratory of Lihian Hui[23]. 293T cells were seeded in 10-cm dishes at a density of 8*10^5 cells and cultured in DMEM High Glucose (Thermo Fisher Scientific 11995) supplemented with 10% FBS (Hyclone SH30071). After 24 hours, each dish of cells was co-transfected with a total of 5 μg of DNA, 2.5 μg of expression vector, 1.7 μg of packaging plasmid psPAX2 (Addgene 12260) and 0.8 μg of envelope plasmid PMD2.G (Addgene 12259), and 30 μL of transfection reagent Fugene6 (Promega E2691) along with 570 μL of OptiMEM-I Reduced Serum Media (Thermo Fisher Scientific 31985070). Fugene6 was first diluted in OptiMEM, vortexed for 1 second, and incubated for 5 minutes at room temperature. Transfection mixes were generated by adding the DNA to the diluted Fugene6, vortexing for one second, and incubating for 15 minutes at room temperature. The mix was then added to the cells in a dropwise manner. 16 hours post-transfection, the media was replaced with 10 mL of fresh culture media. 60 hours post media change, the media containing the viral particles was collected, and centrifuged for 10 minutes at 2000 rpm at 4°C to remove cell debris, and filtered on a 0.45 uM syringe filter (Millipore SLHV033RS). Next, the supernatant was concentrated using ultracentrifugation at 25,000 rpm at 4°C for 1.5 hours with an SW-32 swinging bucket centrifuge (Beckman Coulter). The viral pellet was resuspended in 1:100 volume of plain high glucose DMEM overnight at 4°C, and then stored at −80°C. Viral titer was determined by immunostaining of FOXA3, HNF1A, and

HNF4A in BJ fibroblasts 3 days post infection at various serial dilutions of concentrated virus. Dilutions producing 10–35% of cells expressing the transgene were used to calculate the multiplicity of infection (MOI), and the titer was calculated using the relationship MOI = $(-1) * \ln(1 - $ [proportion infected]).

hiHep Reprogramming

hiHep reprogramming was conducted as previously described[2,23].

siRNA transfection experiments

All knockdown experiments were performed using two cycles of siRNA transfection, three days apart as previously described[2]. All Silencer Select siRNAs from Thermo Fisher used in experiments can be found in Supplementary Table 1. For hiHep reprogramming experiments the first cycle of siRNA transfection was performed one day after the initial treatment with the hiHep lentivirus cocktail.

Western blotting

Whole-cell protein extracts were prepared by resuspending cells in RIPA extraction buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1% Na-deoxycholate, 1% IGEPAL CA-630, 0.1% SDS) supplemented with Protease Inhibitor Cocktail (Roche 11873580001). Suspensions were incubated on ice for 10 min and sonicated for 15 seconds on HI using a Diagenode Bioruptor UCD-200. Samples were centrifuged at 20,000 g for 10 min (4°C) to pellet debris, and the supernatant was transferred to new tubes. Protein content was quantified by BCA assay (Thermo Fisher Scientific 23227). Protein samples were mixed with 4X NuPAGE LDS Sample Buffer (Thermo Fisher Scientific NP0007) and 10X NuPAGE Sample Reducing Agent (Thermo Fisher Scientific NP0009), and were denatured at 70°C for 10 min. Samples were loaded in NuPAGE Novex 4%–12% Bis-Tris Protein Gels (NP0335), and run using NuPAGE Running Buffers (NP0001; NP0002). Wet transfer to PVDF membranes (100V for 1.5 hr) was performed using NuPAGE Transfer Buffer (NP0006) containing 20% methanol, and membranes were blocked overnight in 5% nonfat dairy milk in TBS-T (20 mM Tris, pH 7.5, 150 mM NaCl, 0.1% Tween-20). Primary antibodies were diluted in 1% milk/TBS-T at the following concentrations: anti-GAPDH (1:1000, Santa Cruz Biotechnology sc-365062), anti-CREM (1:1000, ThermoFisher PA5–81971), anti-MAZ (1:1000, ThermoFisher PA5–61710), anti-ETV5 (1:1000, ThermoFisher PA5–30023), anti-NFYB (1:1000, ThermoFisher PA5–31913), anti-H3K9me3 (1:1000, Abcam ab8898), anti-ERH (1:1000, Millipore Sigma HPA002567), anti-SUV39H1 (1:1000, Bethyl A302–127A), anti-SUV39H2 (1:1000, Cell Signaling 8729S) or anti-SETDB1 (1:1000, Cell Signaling 93212S). HRP-conjugated secondary antibodies (BioRad 1706515, 1706516) were diluted 1:15,000 in 1% milk/TBS-T. Blots were developed using Super- Signal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific 34080) and visualized with an Amersham Imager 600 (GE Healthcare Life Sciences).

Immunofluorescence

Cells were grown on 96 well glass bottom plates (MatTek PBK96G-1.5–5-F) coated with collagen I (Corning 354236) and were plated at 3,000 cells per well 7 days prior to

fixation. At least 2 wells of siControl treated fibroblasts were included on all 96 well plates processed for imaging to allow normalization and minimize variation in staining intensity. To prepare for immunofluorescence cells were washed twice briefly with PBS, and fixed in 4% paraformaldehyde (Electron Microscopy Sciences 15714) in PBS for 10 min at room temperature. Fixed cells were washed twice with PBS, permeabilized with ice-cold 0.1% Triton X-100 in PBS for 1 minute and washed twice with TBS-T (20mM Tris-HCL pH 7.4, 150 mM NaCl, 0.05% Tween-20). Samples were blocked with 4% donkey serum (Sigma-Aldrich D9663) in PBS for 1–2 hours at room temperature. Primary antibody staining was performed in 4% donkey serum in PBS for 1 hours at room temperature, using the following concentrations: anti-H3K9me3 (1:500, Abcam ab8898), anti-H3K9me2 (1:500, Active Motif 39683;1:500, Thermo Fisher 49–1007). Cells were washed 3 times with TBS-T and then incubated with AlexaFluor 488- or 594-conjugated secondary antibodies raised in donkey (1:500, Thermo Fisher Scientific A32790, A32754) in PBS for 45 min at room temperature, protected from light. Samples were washed 3 times with PBS, counterstained with 1 μg/mL DAPI (Thermo Fisher, D1306) in PBS for 10 min. Cell were washed once more with PBS and stored at 4°C protected from light until imaging.

Imaging and analysis

Images were taken on a Zeiss LSM800 confocal microscope using a Plan-Apochromat 63× oil immersion objective. Complete images were composed of stitching together nine 512×512 16-bit images taken using bidirectional scanning. A stack ranging from 10 to 20 z slices with a step size of 0.52um was used. Between 100–300 cells were imaged per sample condition for quantitative analysis. Image analysis was performed using MATLAB (R2018a). Images were first manually corrected to remove cells that overlapped. The DAPI channel was then binarized using Otsu's method to create a mask of each nuclei. The nuclei mask gave the total number of nuclei for each condition as well as location and size for each nucleus. Objects smaller than 1500 pixels and partially imaged nuclei on the border of the image were removed leaving only complete, non-overlapping nuclei. Using the mask, the signal and location of every pixel within the nuclei mask was extracted from the H3K9me3 channel. For each nucleus, the average H3K9me3 signal was taken. Intensity values for each cell was normalized to the average intensity from cells on the same plate treated with siControl non-targeting siRNAs. Two independently targeting siRNAs, each performed in duplicate were used for each target for quantification.

Confluency quantification

Images for quantification of confluency were taken using a Nikon TE2000 inverted microscope. For each siRNA target two independently targeting siRNAs were used and two images were collected per siRNA. Images were processed using the PHANTAST library for MATLAB to perform local contrast thresholding and cell density estimation[57]. Default parameters for PHANTAST were used and confluency values were normalized to the mean values measured for siControl.

Flow cytometry

Two biological replicates of day 14 hiHep cells for each treatment were washed with PBS and dissociated from the plate into a single cell suspension with Accutase (Stem Cell

Technologies 07922) at 37°C for 5 min. Cells were washed twice with PBS and fixed in 4% paraformaldehyde in PBS for 15 min at room temperature. Cells were blocked in blocking solution (3% donkey serum in PBS) for 1 hour. Staining was performed in blocking solution for 1 hour at room temperature with anti-ASGPR1 (1:200, BD Biosciences 563655, lot 7097548). Cells were washed three times with PBS, resuspended in water then analyzed on an Accuri C6 and data were collected for all cells. All gating and quantification of population was performed in FlowJo (V10.5). Initial FSC/SSC gating was performed and applied uniformly to all samples.

RNA Isolation and quantitative reverse transcription PCR Assessing siRNA knockdown efficiency and expression of pre-RNAs

Total RNA was isolated using the ZR-96 Quick-RNA kit (Zymo Research R1052), this includes an on-column DNAse I digestion. The samples were then eluted in 30 μL RNAseq Free ddH2O. cDNA was prepared using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher 4368814). Primers were designed for each transcript targeted by an siRNA (Supplementary Table 2). For assessing the expression of pre-RNAs primers were designed with one primer targeting an intron. qPCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher 4367659), and data were normalized using the GAPDH primer as an endogenous control (Supplementary Table 2). qPCR reactions were run in 384-well format on an 7900HT Real-Time PCR machine (Thermo Fisher 4329001), using the following thermal cycler protocol: 50°C for 2 min, 95°C for 10 min, followed by 45 cycles of 95°C for 15s then 60°C for 1 min. For SYBR-based qPCR reactions, a dissociation curve was generated to verify that a single PCR product was generated.

RNA-seq Library Preparation

75–100 ng of isolated RNA was used to prepare RNA-seq libraries for sequencing. All custom primers used in cDNA synthesis are listed below. An oligo-dT primer was used to perform poly-A enrichment and was incubated with the sample for 5 minutes. First-strand cDNA synthesis was performed with Superscript II Reverse Transcriptase (Invitrogen 18064022) and a template switch oligo by incubating for 1.5 hours. Second-strand cDNA synthesis was performed with the KAPA HiFi HotStart Ready Mix (Roche 7958927001) and an IS primer with 3 cycles of PCR amplification. 5ng of cDNA was fragmented using the Tn5 enzyme-adaptor compound. 15 cycles of PCR were then carried out with barcoded primers compatible with the BGISEQ-500. The 300–500bp DNA fragments were selected and purified. The fragments were then heat-denatured and one of the single strands was circularized with DNA ligase to obtain a single strand circular DNA library. The remaining single strand was digested with the exonuclease. The sequencing process was conducted according to the BGISEQ-500 protocol as described[58].

Primers used in cDNA prep:

Oligo-dT: 5′-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3′

TSO: 5′-AAGCAGTGGTATCAACGCAGAGTACrGrG+G-3′

(rG= riboguanosine, +G= locked-nucleic acid)

IS Primer: 5′-AAGCAGTGGTATCAACGCAGAGTAC-3′

RNA-seq data processing

Reads with low quality, adapter, high N rate or poly-A sequences were filtered out from the raw FASTQ data before alignment using SOAPnuke[59]. After filtering, we obtained ~33 million paired 50 base reads per sample on average. Clean reads were aligned to the hg38 reference and gene expression levels were quantified using RSEM[60].

Batch effect correction of RNA-seq data

Due to the large number of samples sequenced batch effect differences were detected and adjusted for using Combat batch effect correction algorithm[61] in R (v.3.5.2) Per gene reads were normalized to total sample reads and Combat batch effect correction algorithm was applied to normalized data to ensure that relative gene expression distributions were the same between sequencing batches. To minimize the chance that batch effect correction would lead to false positive expression gene models with 1 or fewer mapping reads were set to a value of 0 after batch effect correction.

RNA-seq data analysis

Batch effect corrected read counts were analyzed by the DESeq2 (v.1.22.2) package in R (v.3.5.2). Differentially expressed genes were determined in a pairwise manner. srHC genes were considered upregulated if they had a log fold change value>0 and an adjusted p-value<=0.05.

Clustering of proteins by shared repressed genes and analysis of clusters

73 proteins were selected for clustering analysis based upon the results from siRNA depletion during hiHep reprogramming showing that their depletion resulting in the upregulation of at least 93 (>1% of the total 9275) srHC genes. A pairwise distance score was calculated between all 73 proteins as the number of common repressed genes multiplied by the mean fraction of shared genes for each member of the pair; with higher scores indicating greater similarity. Clustering was performed by merging each protein into the cluster of its nearest neighbor. Proteins were called as hubs if they had at least 4 direct nearest neighbor connections and their depletion upregulated at least 300 srHC genes during hiHep reprogramming. GO analysis of gene sets regulated by each cluster was performed using the PANTHER Classification system tool[62] for statistical overrepresentation test and a list of all srHC genes as background. Enrichment analysis of miRNA target sequences was performed using the TargetScan tool[48].

Repeat analysis by SALMONTE

Expression of transposable elements genome wide were analyzed from the RNA-seq data using SalmonTE[29]. Salmon was run using Python (v.3.6.3) with the parameter –exprtype TPM. The quantification files produced were then analyzed for differential expression between knockdown samples and non-targeting controls samples in R (v.3.4.2) using EdgeR (v.3.20.9) and the GLM method.

Repeat Analysis by RepEnrich

RepEnrich[37] was used to perform repeat expression analysis on the RNA-seq samples, and to quantify reads from repetitive elements in the H3K9me3 ChIP-seq samples. Samples were first aligned to the hg19 genome using bowtie1 v0.12.9 (parameters -t -m 1 -S –max) which outputs unique mapping and multimapping reads into separate files. The output Sam files was then converted to a bam file with samtools v0.1.19. RepEnrich was run using python v2.7.3 and a repeatMasker annotation retrieved from UCSC with simple repeats removed. The fractional counts produced were then analyzed for differential expression between knockdown samples and non-targeting controls samples in R (v.3.4.2) using EdgeR (v3.20.9) and the GLM method.

ChIP-seq Chromatin Isolation, Immunoprecipitation, and Library Preparation

Chromatin was prepared from 10 cm plates of ~80% confluent BJ fibroblasts with and without siRNA knockdowns. Crosslinking, processing and sonication of cells was performed as described previously[2].

To perform the chromatin immunoprecipitation, Protein G magnetic Dynabeads (Thermo Fisher, 10004D) were used along with the antibody of interest, either anti-H3K9me3 (Abcam ab8898, Lot GR3291043 –1) anti-H3K27me3 (Active Motif 39155, Lot 31618020), anti-ERH (Millipore Sigma HPA002567), anti-HMGA1 (Cell Signaling 7777S), anti-MYBBP1A (Millipore Sigma HPA005466), anti-ZNF622 (Bethyl A304–075A), anti-CEBPz (Bethyl A303–153A), anti-FUS (ThermoFisher PA5–52610), anti-GATAD2A (Bethyl A302–356A) or anti-ZNF438 (Millipore Sigma HPA039843). Additionally, for H3K9me3 and H3K27me3 ChIP, Drosophila spike-in chromatin (Active Motif cat 53083) and Drosophila H2Av antibody (Active Motif 104597, Lot 00419007) were added in order to normalize for global changes in histone modification levels during computational analysis. Per ChIP, 3 μg of human chromatin were used, along with 2 μg of the antibody of interest, additionally 12.5 ng of Drosophila chromatin and 1 μg of H2Av antibody were used along with 25 μL of Dynabeads. 1/10th of the samples were saved to be used as input. Antibody conjugation, binding, washing, elution and purification was performed as described previously[2].

The NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7103) was used to prepare the libraries for sequencing following the manufacturer's protocol, with the following modifications. The size selection step was performed by selecting first with 15 μL of Ampure XP beads (Beckman Coulter A63881), and next by selecting with 60 μL of Ampure XP beads. For ChIP of the srHC proteins no size selection was performed. The libraries were eluted in 17 μL of 0.1X TE buffer. 10 rounds of PCR amplification were performed. The libraries were eluted in 33 μL of TE during the final elution.

srHC-seq Sample Preparation and Library Preparation

The srHC-seq sample preparation was performed as previously described[6] with the following modifications. Chromatin was prepared from 10 cm plates of ~80% confluent BJ fibroblasts with and without siRNA knockdowns. Additionally, a three-part fragment separation of the sonicated DNA was performed using AMPure XP beads (Beckman Coulter

A63881). For small sonication-sensitive fragment isolation, 0.7 volumes (14 μL) of beads were added to the isolated, sonicated DNA, incubated, and the small fragments were saved in the supernatant. The large sonication-resistant and medium sized fraction were isolated from the beads, and then resuspended in 50 μL of TE. 0.2 volumes (10 μL) of beads were added to the large sonication-sensitive fragments and medium size fragments. Medium DNA was isolated from the beads, and the large fragments were saved in the supernatant and subjected to further sonication. The medium fragments were not sequenced. Size selection efficacy was confirmed (Agilent 5067–4626). The NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7103) was used to prepare the large and small fragment libraries for sequencing following the manufacturer's protocol.

Alignment and visualization of ChIP-seq and srHC-seq data

Sequenced read were aligned to the human hg38 genome assembly and to the Drosophila melanogaster dm6 genome assembly (to identify reads from the spike in) using STAR (v.2.3.0e). STAR output files were converted to .bam files using samtools v1.1, and then to .bed files using bedtools (v.2.20.1; bamtobed). To generate input-normalized genome coverage tracks, BED files were converted to bedgraph using bedtools (genomecoverageBed) and normalized to the number of millions of reads sequenced (rpm). The fraction of reads from the spike-in drosophila chromatin was calculated from the aligned .bed files as the number of reads aligning to dm6 divided by the number of reads aligning to hg38 and the ratio of the input to the IP spike-in fractions were used to scale the global signal. For ChIP-seq, input subtraction was performed on a basepair by basepair basis by subtracting the corresponding input sample from each sample's normalized bedgraph. srHC-seq normalized bedgraphs were further processed by dividing the large fragment file by the corresponding small fragment file on a basepair by basepair basis then taking the log2 of the resulting ratio to assess the relative sonication resistance of the region. Normalization of ChIP-seq data across domains of variable length was performed by subdividing the domain into 500 bins then calculating the average signal for each bin. t-Distributed Stochastic Neighbor Embedding of ChIP-seq and srHC-seq data was performed in Matlab (R. 2019a) using the tsne function with mean gene body H3K9me3, H3K27me3 and srHC signals as input.

Analysis of published ChIP-seq data.

For this analysis (Fig. 3d; Fig. 6f, g; Extended Data Fig. 4a; Supplementary Fig. 1; Extended Data Fig. 6; Extended Data Fig. 10j) H3K9me3 ChIP-seq data from untreated BJ fibroblasts was downloaded from GEO accession GSE87039. For this analysis (Fig. 3d; Fig. 5f, g; Supplementary Fig. 1; Extended Data Fig. 6; Extended Data Fig. 9b) Gradient-seq data from untreated BJ fibroblasts was downloaded from GEO accession GSE87039. For this analysis (Fig. 3d; Supplementary Fig. 1; Extended Data Fig. 6) H3K27me3 ChIP-seq data from Human foreskin fibroblasts was downloaded from GEO accessions GSM817237, GSM817240 and GSM958154. For this analysis (Extended Data Figure 8b) the A/B compartment enrichment was downloaded from the 4D nucleome data portal (accession 4DNFID41C3X7), the H4K20me1 ChIP-seq was downloaded from the GEO accessions GSM521917 and GSM521915 and DNA methylation data were downloaded from GEO accession GSM1127120.

Motif analysis

Analysis of gene promoters for enriched motifs was performed in HOMER (v.4.10). For finding positions of known motifs, position weight matrices were downloaded from Jaspar[63] FOXA3(MA1683.1), HNF1a(MA0046.2), HNF4a(MA114.2), MAZ(MA1522.1) and CREM(MA0609.2). Motif identification was performed the findMotifs.pl command in HOMER (v.4.10; parameters -start −200 -end 200 -find) using the indicated position weight matrices. A threshold score of 4.5 was used for identifying motif occurrence except where otherwise noted.

Sucrose gradient separation

Cells fixed in 4% paraformaldehyde were incubated in hypotonic lysis buffer (10mM HEPES pH 7.4, 10mM KCl, 0.05% NP-40) for 30 minutes. Nuclei were collected by centrifugation at 300g for 10 minutes. Nuclei was then incubated in low salt buffer (10mM Tris-HCl pH 7.4, 0.2 $MgCl_2$, 1% Triton x-100) for 30 minutes. Chromatin was collected by centrifugation at 13,000×g for 10 minutes. Sucrose gradient separation was performed as previously described[1].

Data visualization

All heat maps, violin plots, box plots, and clustering diagrams were generated using Matlab (R. 2019a). All boxplots show the median as the center line with box limits corresponding to upper and lower quartiles and whiskers covering 1.5× the interquartile range.

Statistics and Reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blided to allocation during and experiments and outcome assessment. Repeated measurements between two samples were analyzed for significance by two-tailed Student's t test. All statistical tests, resulting P values and observation numbers are indicated in the figure panels or in the figure legends. Western blot experiments where one replicate is shown were repeated twice with similar results.

Data availability

RNA-seq, ChIP-seq and srHC-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE154233. Previously published sequencing data that were re-analysed here are available from GEO under accession codes GSE87039, GSM817237, GSM817240, GSM958154, GSM521917, GSM521915 and GSM1127120 or from the 4D nucleome under accession code 4DNFID41C3X7. All other data supporting the finding of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

Code availability

No novel programs or algorithms were utilized. All code for data analysis and visualization is available by request.
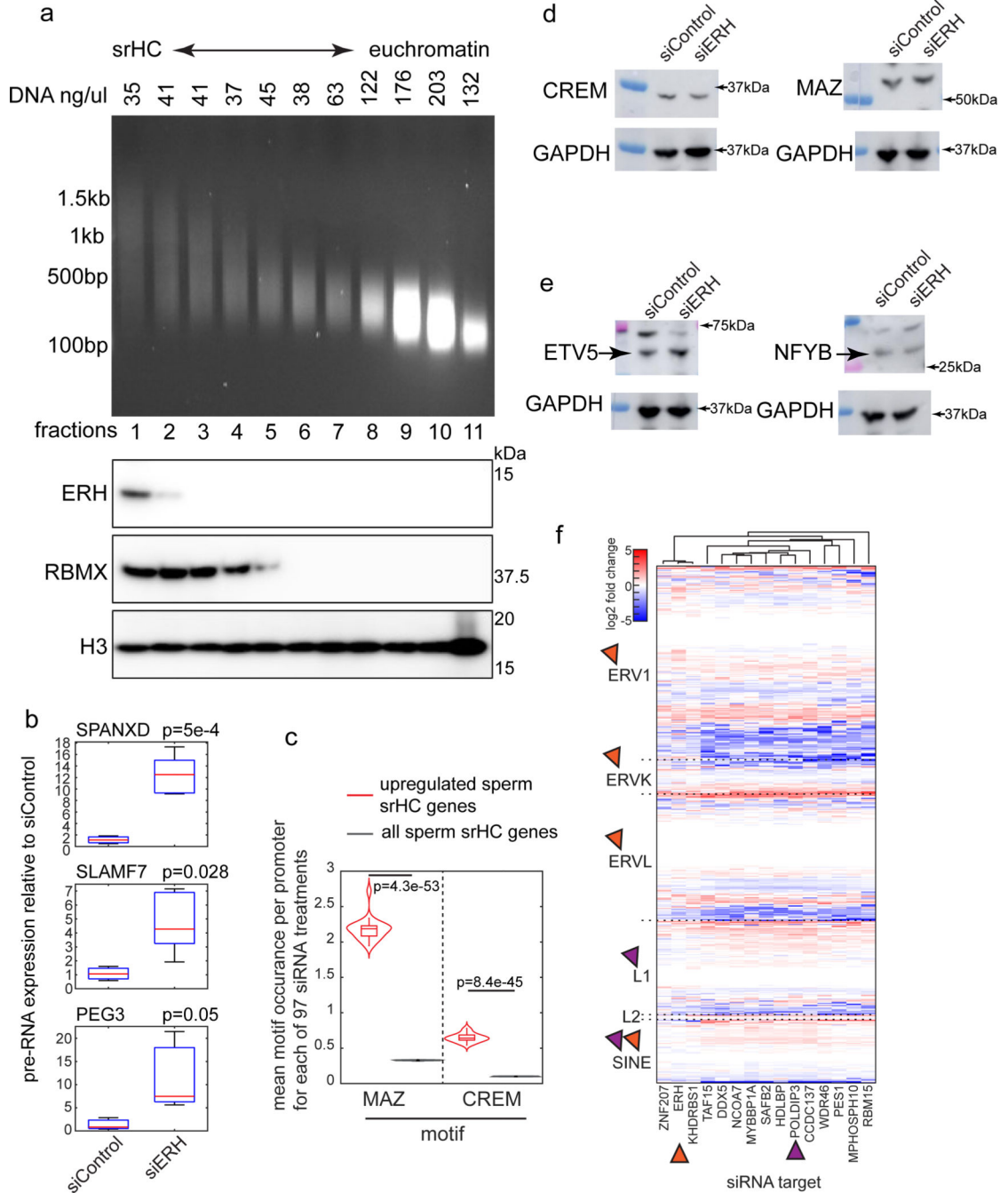
Extended Data
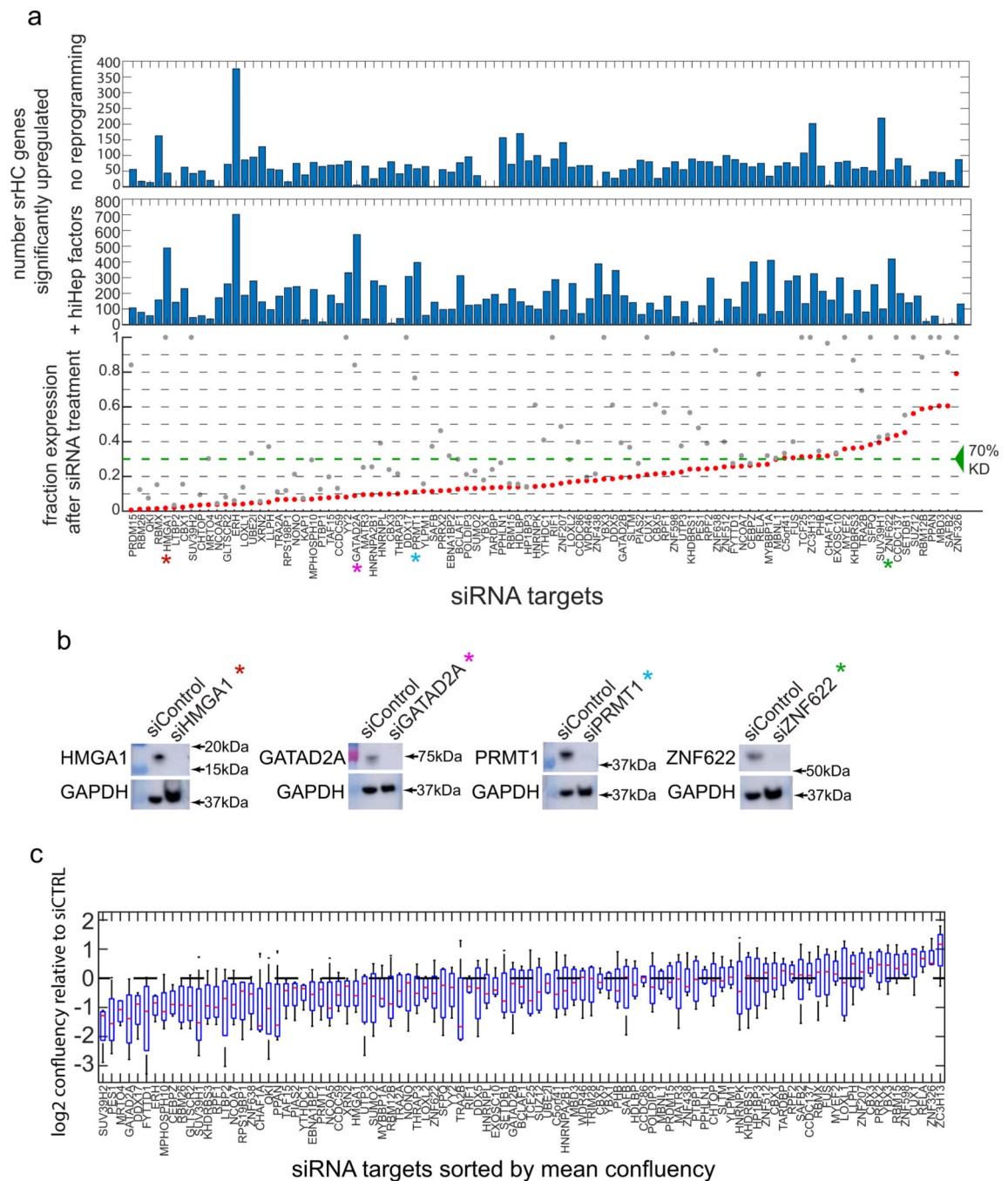
**Extended Data Fig. 1. Validation of siRNA efficiency and confirmation of RNA-seq results by qPCR.**

a, qPCR quantification of knockdown for all siRNAs used in this study. siRNAs used for treating cells for RNA-seq analysis indicated in red. Above bar graphs show the number of srHC genes significantly upregulated (DESeq2, Benjamini multiple test corrected Wald test p-value ⩽ 0.05 and log2(foldchange)>0) by each knockdown in reprogramming and non-reprogramming conditions. (n=2 biological replicates per siRNA, two siRNAs per target)

b, Protein depletion efficiency for select srHC proteins in study, asterisk color corresponds to knockdown in (**a**). Arrows indicate location of indicated molecular weight markers. Experiment repeated independently 2 times with similar results. Unprocessed blots are provided as source data. **c,** Quantification of cell confluency using PHANTAST[55] from phase contrast images (n=4 images for each condition; two independently targeting siRNAs per target, two replicates per siRNA). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values.

Extended Data Fig. 2. Extended analysis of top noTF srHC proteins.

a, Sucrose gradient fractionation of sonicated DNA, DNA concentration of each fraction indicated (20ul loaded per lane), followed by western blot probing for ERH with RBMX and H3 as controls. Experiment repeated independently 2 times with similar results. **b**, qRT-PCR in siControl (n=3) and siERH treated (n=3 for each of two different siRNAs) human fibroblasts of pre-RNAs of genes upregulated at mRNA level by ERH depletion (two tailed Student's t-test). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. **c,** Spermatogenic TF motif enrichment in srHC spermatogenesis genes upregulated and not upregulated by each of the n=97 siRNA targets (two tailed Student's t-test). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. **d,** Western blots for transcription factors involved in spermatogenesis to assess levels in fibroblast whole cell lysate. **e,** Western blots for transcription factors with motifs enriched in promoters of srHC genes activated by ERH knockdown. **f**, Expanded Salmon TE heatmaps for knockdowns showing highest repeat activation (arrows indicate subtypes with the highest percent of upregulation by indicated knockdown). P values from b,c are denoted in the panels. Statistical information and unprocessed blots are provided as source data.
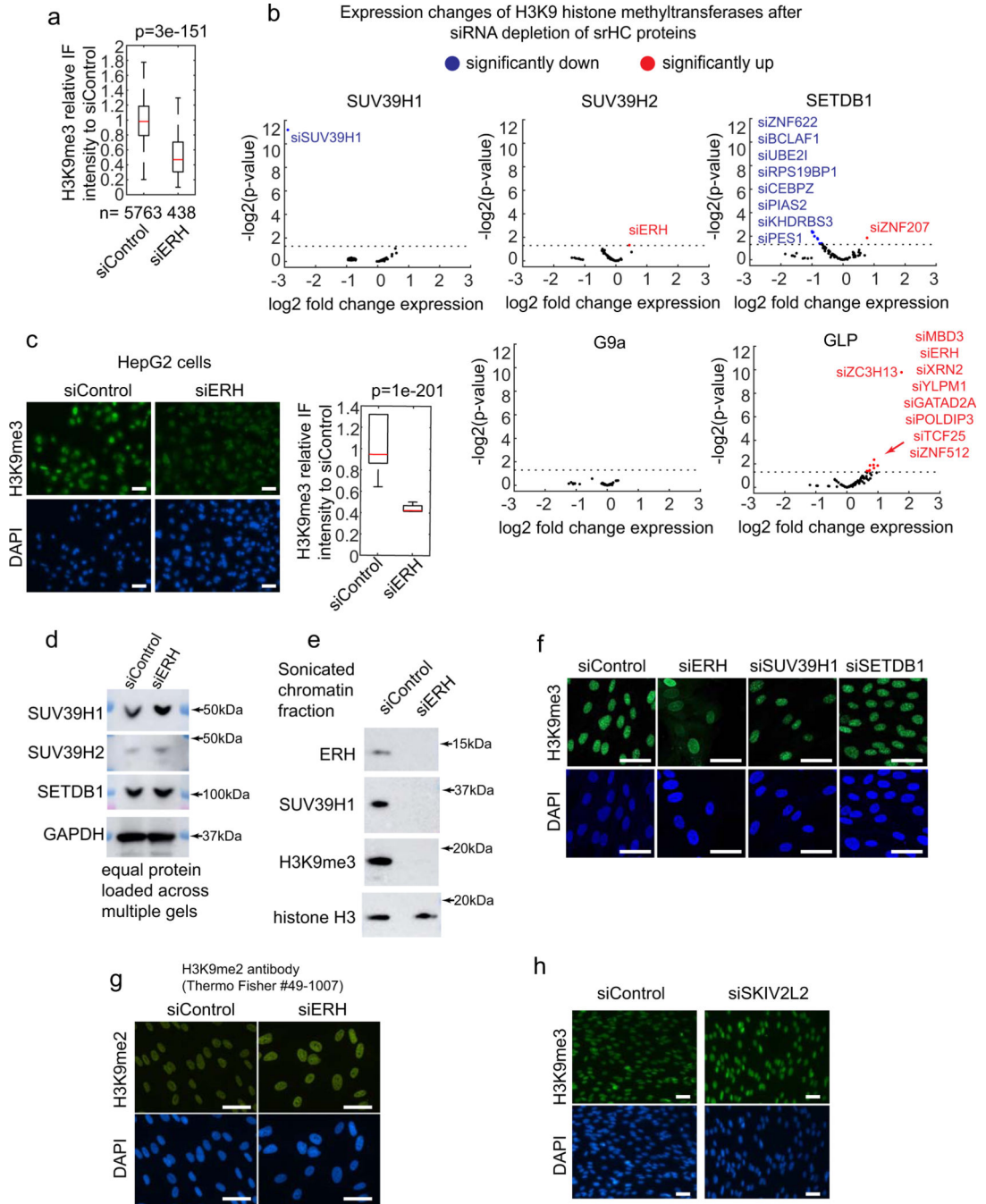
Extended Data Fig. 3. ERH depletion causes a global decrease in H3K9me3 but does not decrease H3K9 HMT expression or H3K9me2 levels.

a, Quantification of H3K9me3 immunofluorescence in siControl and siERH treated human fibroblasts (Student's two tailed t-test). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. N numbers are denoted in the panel and represent the number of cells imaged per treatment. **b,** Volcano plots showing expression change and significance of indicated H3K9 histone methyltransferase for 97 siRNA knockdowns by RNA-seq (n=2). siRNA knockdowns

causing significant (DEseq2, Benjamini multiple test corrected Wald test p-value 0.05 and log2(foldchange)>0) upregulation (red) or downregulation (blue) are listed within graph. **c**, H3K9me3 immunofluorescence (left) and quantification (right) in siControl (n=352 cells) and siERH (n=365 cells) treated HepG2 cells (two tailed Student's t-test). Experiment repeated independently 2 times with similar results. Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. P values denoted in panel. **d**, Western blots for H3K9me3 histone methyltransferases in siControl and siERH treated human fibroblasts. Experiment repeated independently 2 times with similar results. **e**, Western blots for ERH, SUV39H1 and H3K9me3 in the sonicated chromatin fraction from siControl and siERH treated human fibroblasts. Arrows indicate location of indicated molecular weight markers. Experiment repeated independently 2 times with similar results. **f**, H3K9me3 immunofluorescence comparison between siERH and HMTs siSUV39H1 and siSETDB1. Experiment repeated independently 2 times with similar results. **g**, H3K9me2 (green) and DAPI (blue) immunofluorescence in siControl and siERH treated human fibroblasts using an alternative antibody. Experiment repeated independently 2 times with similar results. **h**, H3K9me3 immunofluorescence in siControl and siSKIV2L2 treated human fibroblasts. Scale bars indicate 50μm for all images. Experiment repeated independently 2 times with similar results. Unprocessed blots are provided as source data.

Extended Data Fig. 4. ERH regulates gametogenic genes and a subset of repeat elements.
a, RNA-seq tracks showing SPANX cluster expression in siERH and several additional srHC protein knockdowns; same scale used for all mRNA-seq tracks. **b**, Protein sequence alignment of the ERH interacting domain of Mmi1 (95–122) and corresponding regions, determined by full length protein alignment, of the closest human orthologs. Purple arrow indicates tryptophan residue previously observed in a separate study[2] to be important for ERH interaction. **c**, R-Deep database showing control and RNase treated fractionation and mass spectrometry detection of ERH. Other proteins observed to fractionate with ERH in
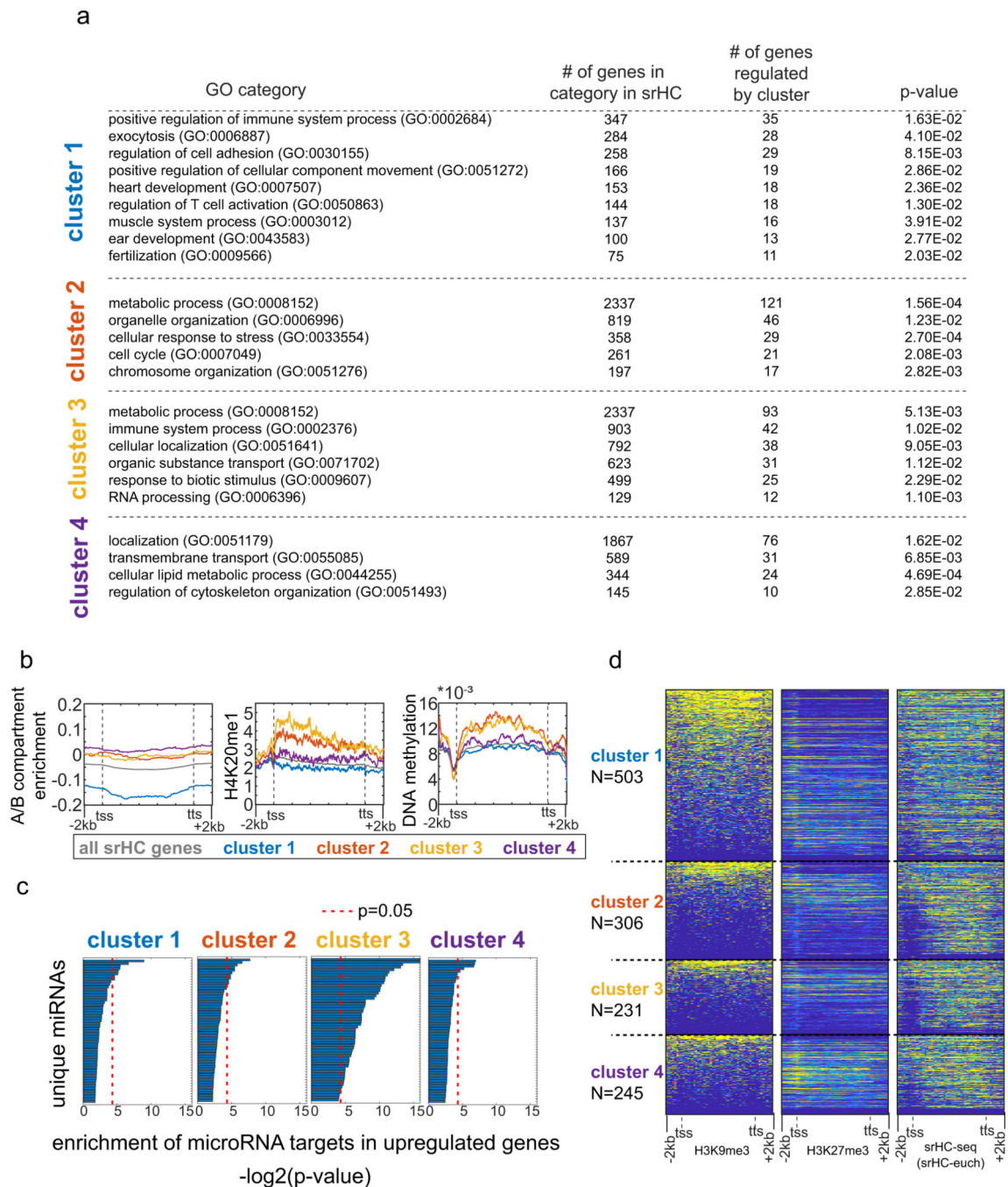
fraction 22 which also exhibit a RNase induced shift listed in red box. **d**, Correlation of RepEnrich analysis of H3K9me3 and expression changes of repeat element classes in siERH relative to siControl. Alu and ERVK elements showed the greatest negative correlation between H3K9me3 and expression and are plotted separately. R correlation coefficient and p-value calculated using corrcoef in MATLAB; p-value calculated as the corresponding two-sided p-value for the t-distribution with n-2 degrees of freedom. N numbers stated in panel and represent the number of distinct repeat elements of indicated class. **e**, Violin plots of H3K9me3 and expression fold change (log2(siERH/siControl)) as determined by RepEnrich for the repeat element classes exhibiting significant H3K9me3 and expression changes (two tailed Student's t-test). N numbers stated in panel and represent the number of distinct repeat elements of indicated class. **f**, H3K27me3 changes at satellite repeats in siERH relative to siControl. **g**, Motif occurrence for FOXA3, HNF1α and HNF4α in ERV (n=1641 element sequences), LINE (n=471 element sequences) and SINE (n=147 element sequences) elements (two tailed Student's t-test). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. **h**, Quantification of flow cytometry assay of siRNA+hiHep cells stained for the hepatic marker ASGPR1 (two tailed Student's t-test to siControl). N numbers stated in panel and represent the number of cells quantified to calculate percent positive per condition. P values in d,e,g,h are denoted in the panels.
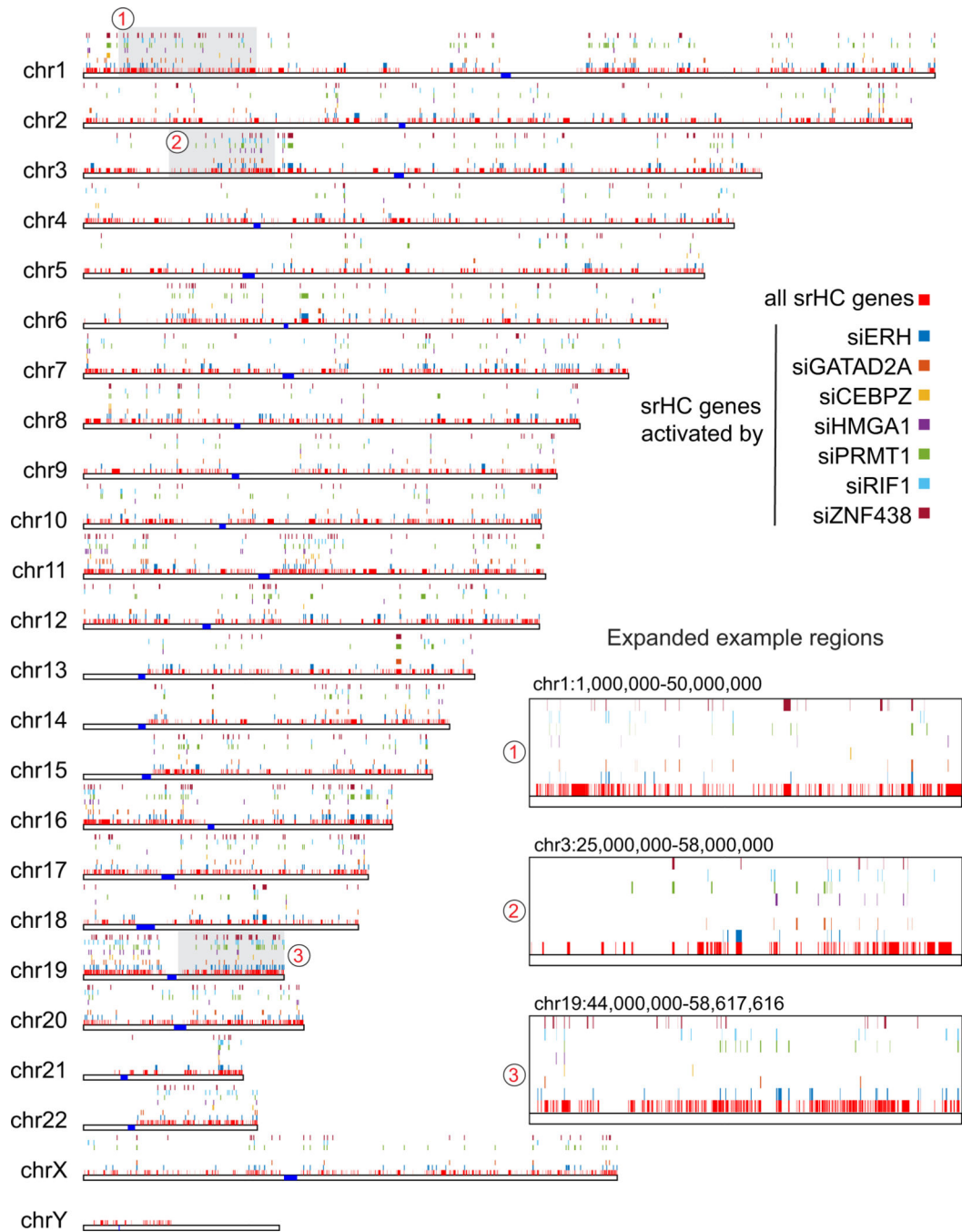
Extended Data Fig. 5. Chromosomal positions of srHC gene activation for noTF and hiHep.
Chromosomal positions for hiHep activation of srHC genes by indicated knockdowns.
Expanded example regions 1–3 are marked by grey background.
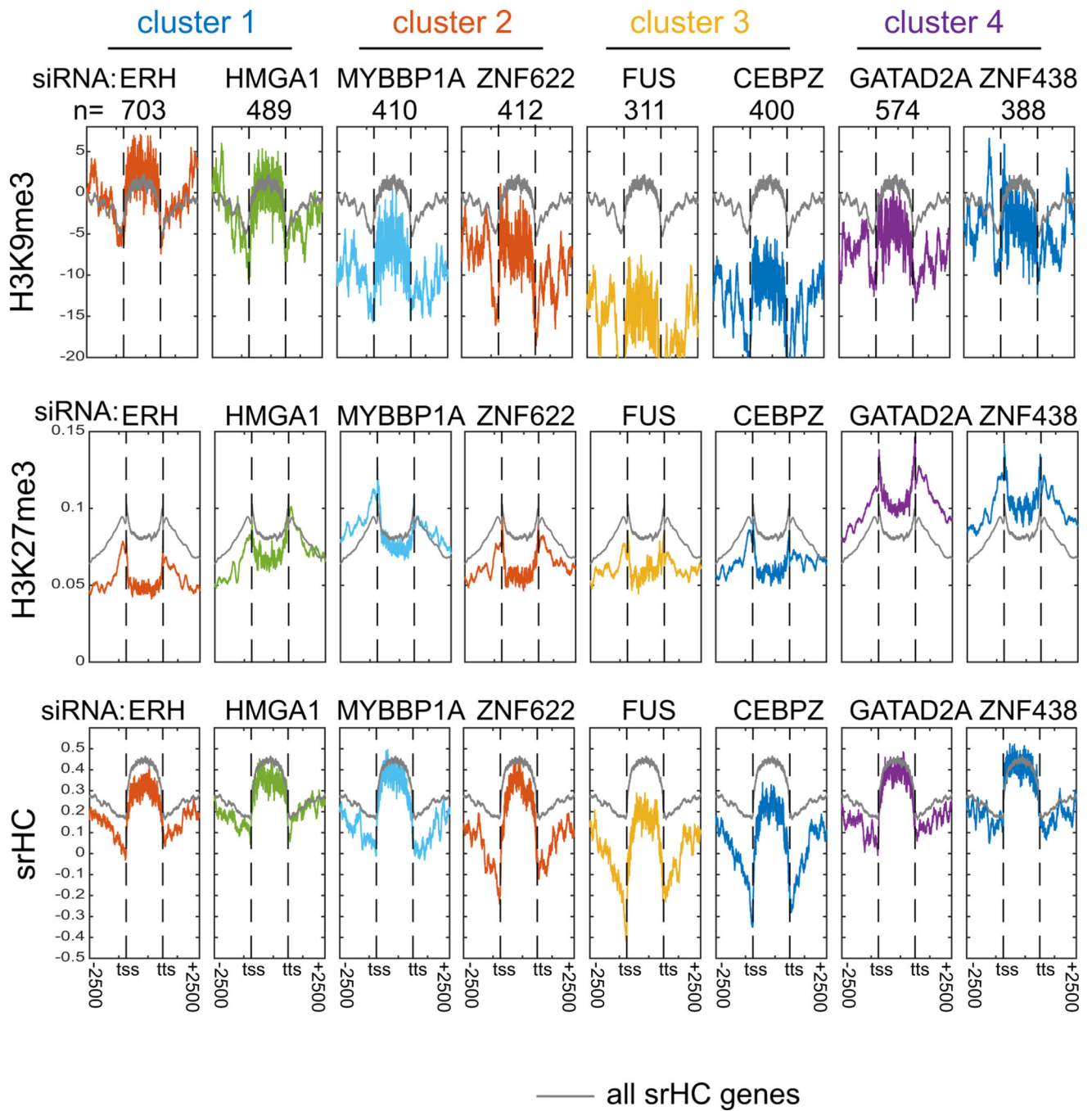
Extended Data Fig. 6. Histone profiles of genes targeted by hub proteins.
a,b,c, H3K9me3 (**a**), H3K27me3 (**b**) and srHC (**c**) meta profiles of genes targeted by hub proteins 2.5kb upstream of TSS to 2.5kb downstream of TTS.

a

	GO category	# of genes in category in srHC	# of genes regulated by cluster	p-value
cluster 1	positive regulation of immune system process (GO:0002684)	347	35	1.63E-02
	exocytosis (GO:0006887)	284	28	4.10E-02
	regulation of cell adhesion (GO:0030155)	258	29	8.15E-03
	positive regulation of cellular component movement (GO:0051272)	166	19	2.86E-02
	heart development (GO:0007507)	153	18	2.36E-02
	regulation of T cell activation (GO:0050863)	144	18	1.30E-02
	muscle system process (GO:0003012)	137	16	3.91E-02
	ear development (GO:0043583)	100	13	2.77E-02
	fertilization (GO:0009566)	75	11	2.03E-02
cluster 2	metabolic process (GO:0008152)	2337	121	1.56E-04
	organelle organization (GO:0006996)	819	46	1.23E-02
	cellular response to stress (GO:0033554)	358	29	2.70E-04
	cell cycle (GO:0007049)	261	21	2.08E-03
	chromosome organization (GO:0051276)	197	17	2.82E-03
cluster 3	metabolic process (GO:0008152)	2337	93	5.13E-03
	immune system process (GO:0002376)	903	42	1.02E-02
	cellular localization (GO:0051641)	792	38	9.05E-03
	organic substance transport (GO:0071702)	623	31	1.12E-02
	response to biotic stimulus (GO:0009607)	499	25	2.29E-02
	RNA processing (GO:0006396)	129	12	1.10E-03
cluster 4	localization (GO:0051179)	1867	76	1.62E-02
	transmembrane transport (GO:0055085)	589	31	6.85E-03
	cellular lipid metabolic process (GO:0044255)	344	24	4.69E-04
	regulation of cytoskeleton organization (GO:0051493)	145	10	2.85E-02

Extended Data Fig. 7. Extended cluster specific analysis.
a, Gene Ontology analysis for statistical overrepresentation of biological processes for srHC genes uniquely repressed by each cluster; non-redundant GO categories shown (p-value calculated by PANTHER Statistical overrepresentation test, denoted in panel). N numbers represent number of genes in GO category repressed by srHC protein member of indicated cluster and are stated in panel. **b**, Profiles of A/B compartment enrichment[43], H4K20me1 and DNA methylation by bisulfide sequencing across genes uniquely regulate by each cluster. **c**, TargetScan database analysis for enrichment of miRNA target sequences

in srHC genes uniquely repressed by each cluster; -log2(p-value) shown for the top 55 enriched miRNA target sequences per cluster (p-value calculated using the statistical model developed in Agarwal et al., 2017). **d**, Heatmap of H3K9me3, H3K27me3 and srHC-seq profiles of srHC genes uniquely regulated by each srHC protein cluster and sorted by mean H3K9me3 level within each cluster set. N numbers represent number of srHC gene profiles depicted and are stated in the panel.
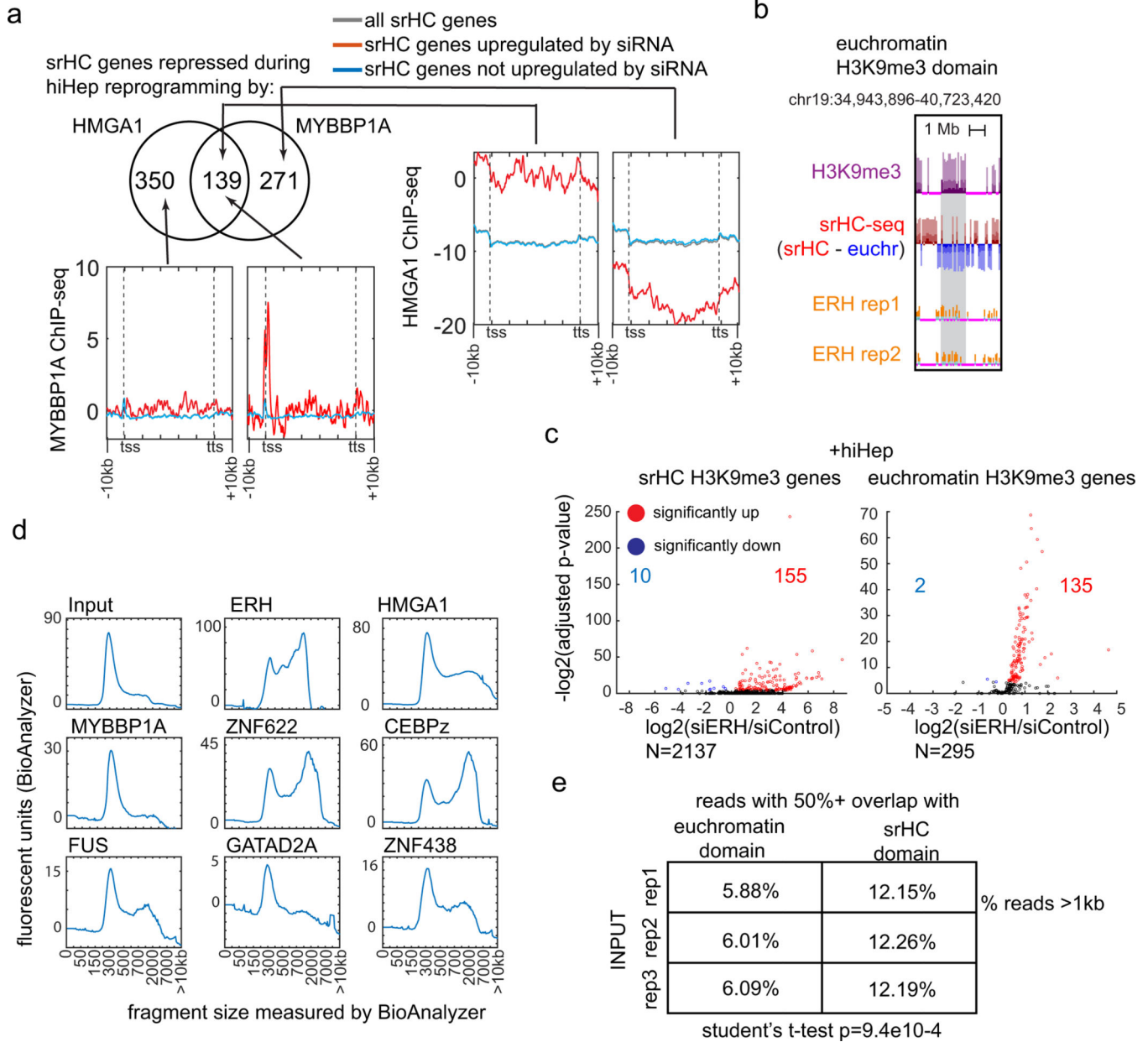
Extended Data Fig. 8. srHC protein ChIP-seq.
a, ChIP-seq profiles of HMGA1 and MYBBP1A at srHC genes repressed by both factors or repressed by the other factor. **b**, Browser track showing ERH at a H3K9me3 marked euchromatin domain. **c**, srHC and euchromatin H3K9m3 gene expression changes in

siERH+hiHep. N numbers represent the number of srHC or euchromatin srHC genes and are stated in the panel. The number of genes in each set significantly up (red) or down (blue) are indicated. **d**, DNA fragment size profiles by BioAnalyzer of INPUT and eluted DNA after ChIP of the indicated srHC protein. **e**, Table showing the percent of DNA fragments with length>=1kb as determined by paired-end sequencing for reads mapping with over 50% overlap with euchromatin or srHC domains (two tailed Student's t-test, n=3 biological replicates sequenced as INPUT).
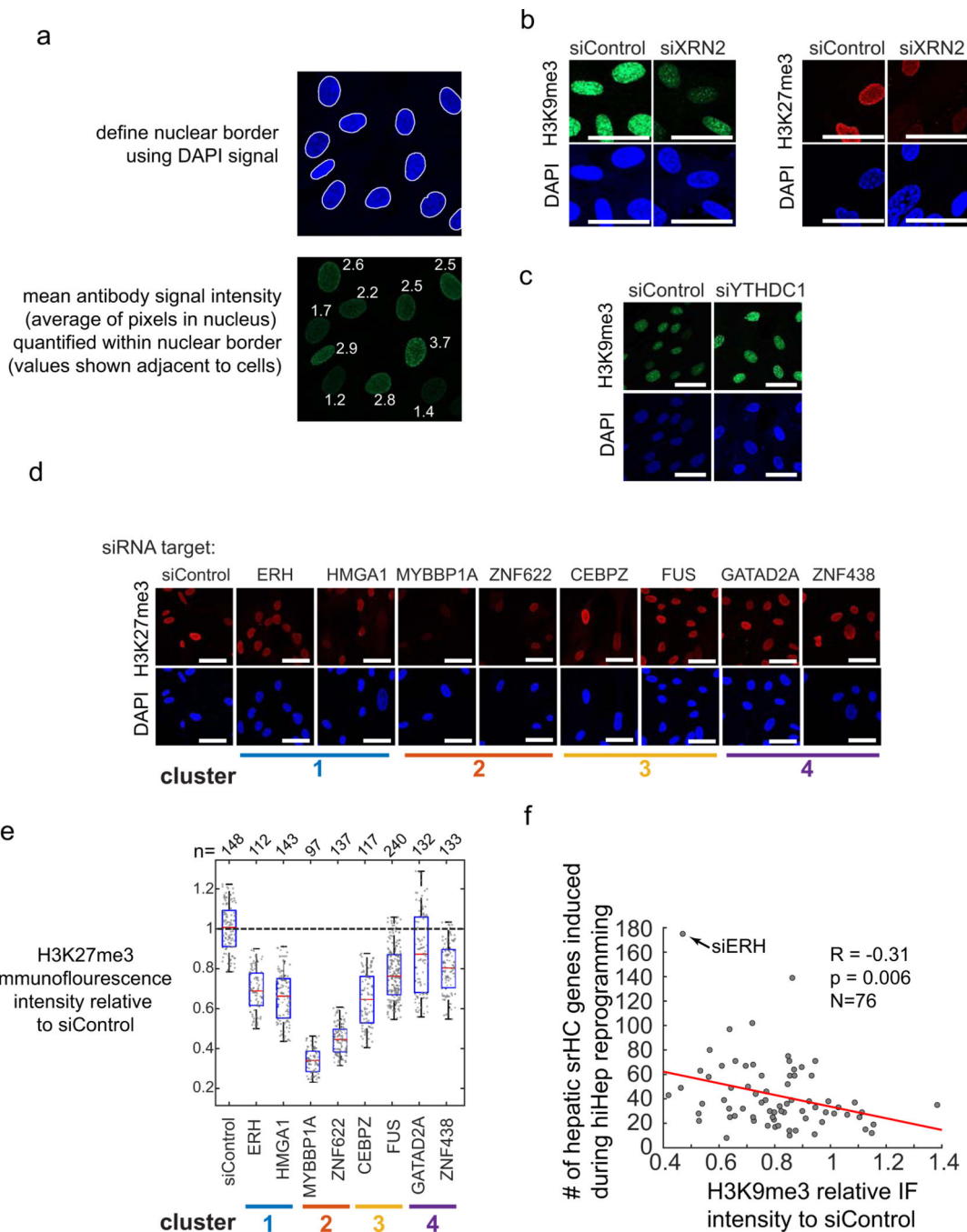
Extended Data Fig. 9. Quantification of confocal images and extended analysis.
a, Methodology for defining nuclear border and quantifying fluorescence intensity. **b**, H3K9me3 and H3K27me3 immunofluorescence in siControl and siXRN2 treated human fibroblasts. Images representative of 2 experiments. **c**, H3K9me3 immunofluorescence in siControl and siYTHDC1 treated human fibroblasts. Images representative of 4 experiments. **d,e** Representative images (**d**) and quantification (**e**) of H3K27me3 for cells depleted by siRNA for hub proteins. Images representative of 2 experiments. Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. N numbers for e represent the number of cells imaged and are stated in the panel. **f**, Plot showing correlation of H3K9me3 immunofluorescence intensity relative to siControl vs the number of hepatic srHC genes induced during hiHep reprogramming for the target srHC proteins in Fig. 5b. R correlation coefficient and p-value calculated using corrcoef in MATLAB; p-value calculated as the corresponding two-sided p-value for the t-distribution with n-2 degrees of freedom. N numbers stated in panel and represent the number of siRNA targets. All scale bars indicate 50 μm.
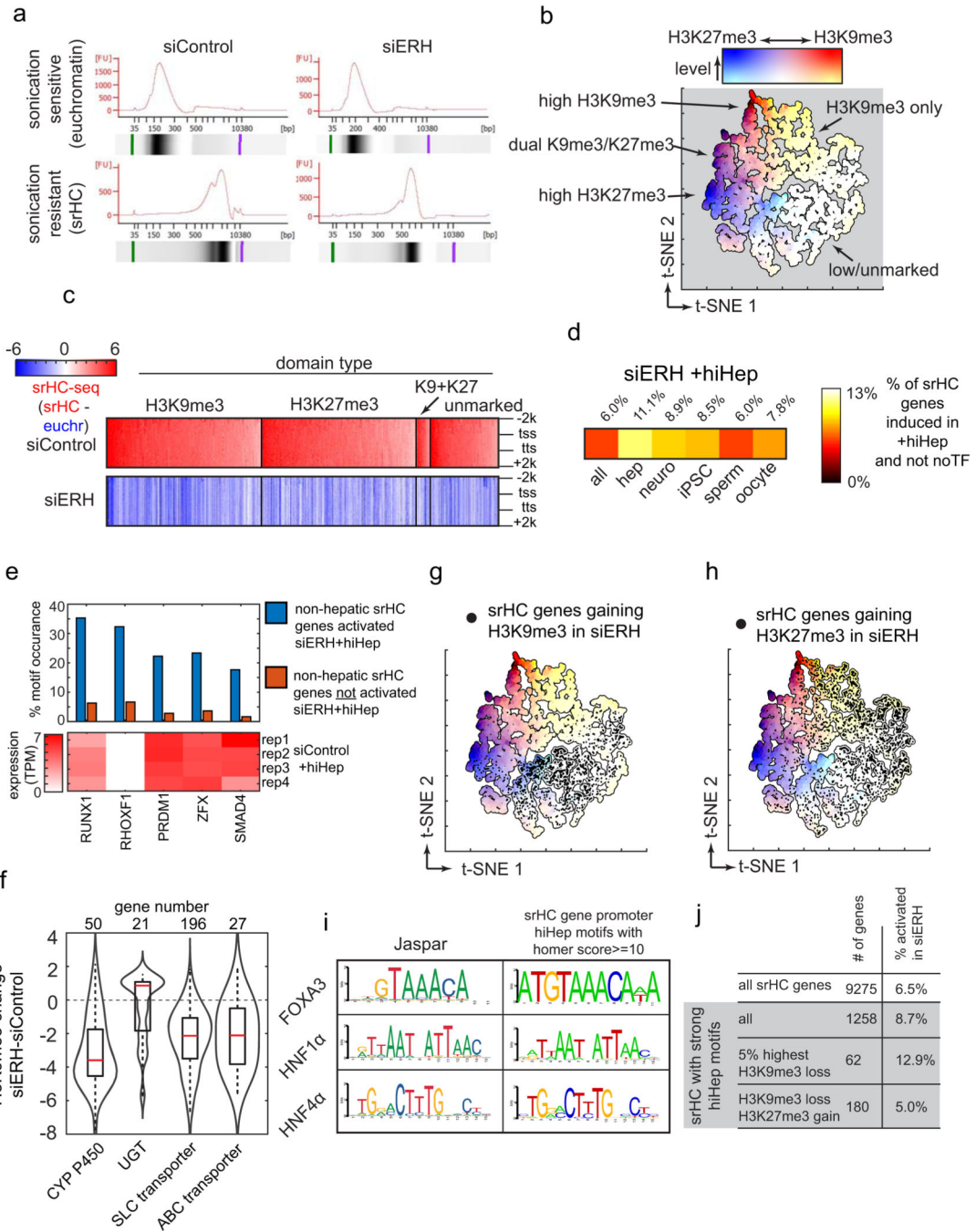
Extended Data Fig. 10. Expanded analysis of ChIP-seq and srHC-seq in siERH.
a, DNA size profiles of sonicated fractions from siControl and siERH. **b**, Combinatorial
H3K9me3 and H3K27me3 levels for srHC genes in siControl treated human fibroblasts.
c, Heatmap displaying enrichment of H3K9me3, H3K27me3, dual-marked, and unmarked
srHC gene subtypes in siControl and siERH from 2kp upstream of TSS to 2kp downstream
of TTS. **d**, Activation of srHC alternative lineage genes in siERH relative to siControl
during hiHep reprogramming not activated by siERH without hiHep factors. **e**, Percent
occurrence of top 5 motifs enriched in promoters of non-hepatic genes upregulated in

siERH+hiHep conditions and corresponding expression of the putative targeting factors in 4 siControl+hiHep replicates. **f**, H3K9me3 changes at classes of key hepatic genes, cytochrome p450 (n=50), UGT (n=21), SLC transporter (n=196) and ABC transporter (n=27), in siERH compared to siControl. Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. **g,h**, Location of srHC genes gaining H3K9me3 (**g**) and H3K27me3 (**h**) on t-SNE embedding. **i**, hiHep motifs from Jaspar database and identified motifs in promoter regions (tss +/− 200) of srHC genes with motif scores >=10. **j**, Table showing total gene numbers and activation rates in siERH for sets of srHC genes defined by presence of strong hiHep motifs and specific changes in H3K9me3 and H3K27me3 levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Becker JS, Nicetto D & Zaret KS H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. 32, 29–41 (2016). [PubMed: 26675384]

2. Becker JS et al. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. Mol. Cell 68, 1023–1037.e15 (2017). [PubMed: 29272703]

3. Margueron R. & Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature 469, 343–349 (2011). [PubMed: 21248841]

4. Bulut-Karslioglu A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol. Cell 55, 277–290 (2014). [PubMed: 24981170]

5. Nicetto D. et al. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science 363, 294–297 (2019). [PubMed: 30606806]

6. Wang C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat. Cell Biol 20, 620–631 (2018). [PubMed: 29686265]

7. Pace L. et al. The epigenetic control of stemness in CD8+ T cell fate commitment. Science 359, 177–186 (2018). [PubMed: 29326266]

8. Soufi A, Donahue G. & Zaret KS Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004 (2012). [PubMed: 23159369]

9. Schotta G. et al. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 18, 1251–1262 (2004). [PubMed: 15145825]

10. Rea S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406, 593–599 (2000). [PubMed: 10949293]

11. Schultz DC, Ayyanathan K, Negorev D, Maul GG & Rauscher FJ SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 16, 919–932 (2002). [PubMed: 11959841]

12. Johnson WL et al. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. Elife 6, (2017).

13. Zhang K, Mosch K, Fischle W. & Grewal SIS Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol 15, 381–388 (2008). [PubMed: 18345014]

14. Allshire RC & Madhani HD Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol 19, 229–244 (2018). [PubMed: 29235574]

15. Sugiyama T. et al. Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. Mol. Cell 61, 747–759 (2016). [PubMed: 26942678]

16. Xie G. et al. A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. Nat Commun 10, 251 (2019). [PubMed: 30651569]

17. Hazra D, Andri V, Palancade B, Rougemaille M. & Graille M. Formation of S. pombe Erh1 homodimer mediates gametogenic gene silencing and meiosis progression. Sci Rep 10, 1034 (2020). [PubMed: 31974447]

18. Djupedal I. et al. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. 19, 2301–2306 (2005). [PubMed: 16204182]

19. Kato H. et al. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science 309, 467–469 (2005). [PubMed: 15947136]

20. Onodera Y. et al. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120, 613–622 (2005). [PubMed: 15766525]

21. Pinzón N. et al. Functional lability of RNA-dependent RNA polymerases in animals. PLoS Genet. 15, e1007915 (2019).

22. Bannister AJ et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120–124 (2001). [PubMed: 11242054]

23. Huang P. et al. Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell 14, 370–384 (2014). [PubMed: 24582927]

24. Liu M-L et al. Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. Nat Commun 4, 2183 (2013). [PubMed: 23873306]

25. Cacchiarelli D. et al. Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. Cell 162, 412–424 (2015). [PubMed: 26186193]

26. Hermann BP et al. The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep 25, 1650–1667.e8 (2018). [PubMed: 30404016]

27. Zhang Y. et al. Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. Mol. Cell 72, 1021–1034.e4 (2018). [PubMed: 30472193]

28. Cheloufi S. et al. The histone chaperone CAF-1 safeguards somatic cell identity. Nature 528, 218–224 (2015). [PubMed: 26659182]

29. Jeong H-H, Yalamanchili HK, Guo C, Shulman JM & Liu Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. Pac Symp Biocomput 23, 168–179 (2018). [PubMed: 29218879]

30. Lewis ZA et al. DNA methylation and normal chromosome behavior in Neurospora depend on five components of a histone methyltransferase complex, DCDC. PLoS Genet. 6, e1001196 (2010).

31. Kouprina N. et al. Evolutionary diversification of SPANX-N sperm protein gene structure and expression. PLoS ONE 2, e359 (2007). [PubMed: 17406683]

32. Bao X. et al. Capturing the interactome of newly transcribed RNA. Nat. Methods 15, 213–220 (2018). [PubMed: 29431736]

33. Caudron-Herger M. et al. R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. Mol. Cell 75, 184–199.e10 (2019). [PubMed: 31076284]

34. Wang C. et al. A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. Nucleic Acids Res. 44, 969–982 (2016). [PubMed: 26673708]

35. Scarola M. et al. Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun 6, 7631 (2015). [PubMed: 26158551]

36. Kalyana-Sundaram S. et al. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell 149, 1622–1634 (2012). [PubMed: 22726445]

37. Criscione SW, Zhang Y, Thompson W, Sedivy JM & Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics 15, 583 (2014). [PubMed: 25012247]

38. Zhu Z. et al. PHF8 is a histone H3K9me2 demethylase regulating rRNA synthesis. Cell Res. 20, 794–801 (2010). [PubMed: 20531378]

39. Peters AHFM et al. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell 12, 1577–1589 (2003). [PubMed: 14690609]

40. Patil DP et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature 537, 369–373 (2016). [PubMed: 27602518]

41. Knuckles P. et al. Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m6A machinery component Wtap/Fl(2)d. Genes Dev. 32, 415–429 (2018). [PubMed: 29535189]

42. Tchasovnikarova IA et al. GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science 348, 1481–1485 (2015). [PubMed: 26022416]

43. Yamazaki T. et al. Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Mol. Cell 70, 1038–1053.e7 (2018). [PubMed: 29932899]

44. Krietenstein N. et al. Ultrastructural Details of Mammalian Chromosome Architecture. Mol Cell 78, 554–565.e7 (2020). [PubMed: 32213324]

45. Thul PJ et al. A subcellular map of the human proteome. Science 356, (2017).

46. Liu Y. et al. PWP1 Mediates Nutrient-Dependent Growth Control through Nucleolar Regulation of Ribosomal Gene Expression. Dev. Cell 43, 240–252.e5 (2017). [PubMed: 29065309]

47. Collier SP, Collins PL, Williams CL, Boothby MR & Aune TM Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. J. Immunol 189, 2084–2088 (2012). [PubMed: 22851706]

48. Agarwal V, Bell GW, Nam J-W & Bartel DP Predicting effective microRNA target sites in mammalian mRNAs. Elife 4, (2015).

49. Zhang T. et al. FUS Regulates Activity of MicroRNA-Mediated Gene Silencing. Mol. Cell 69, 787–801.e8 (2018). [PubMed: 29499134]

50. Villarroya-Beltri C. et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nat Commun 4, 2980 (2013). [PubMed: 24356509]

51. Yamazaki T. & Hirose T. The building process of the functional paraspeckle with long non-coding RNAs. Front Biosci (Elite Ed) 7, 1–41 (2015). [PubMed: 25553361]

52. Reynolds N. et al. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J. 31, 593–605 (2012). [PubMed: 22139358]

53. Chalamcharla VR, Folco HD, Dhakshnamoorthy J. & Grewal SIS Conserved factor Dhp1/Rat1/Xrn2 triggers premature transcription termination and nucleates heterochromatin to promote gene silencing. Proc. Natl. Acad. Sci. U.S.A 112, 15548–15555 (2015). [PubMed: 26631744]

54. Mattout A. et al. LSM2–8 and XRN-2 contribute to the silencing of H3K27me3-marked genes through targeted RNA decay. Nat. Cell Biol (2020) doi:10.1038/s41556-020-0504-1.

55. Shima H. et al. S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. Cell Rep 21, 3354–3363 (2017). [PubMed: 29262316]

References for Methods

56. van Pelt JF, Decorte R, Yap PSH & Fevery J. Identification of HepG2 variant cell lines by short tandem repeat (STR) analysis. Mol Cell Biochem 243, 49–54 (2003). [PubMed: 12619888]

57. Jaccard N. et al. Automated method for the rapid and precise estimation of adherent cell culture characteristics from phase contrast microscopy images. Biotechnol Bioeng 111, 504–517 (2014). [PubMed: 24037521]

58. Huang J. et al. A reference human genome dataset of the BGISEQ-500 sequencer. Gigascience 6, 1–9 (2017).

59. Chen Y. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7, 1–6 (2018).

60. Li B. & Dewey CN RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011). [PubMed: 21816040]

61. Johnson WE, Li C. & Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127 (2007). [PubMed: 16632515]

62. Mi H, Muruganujan A. & Thomas PD PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 41, D377–386 (2013). [PubMed: 23193289]

63. Fornes O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 48, D87–D92 (2020). [PubMed: 31701148]
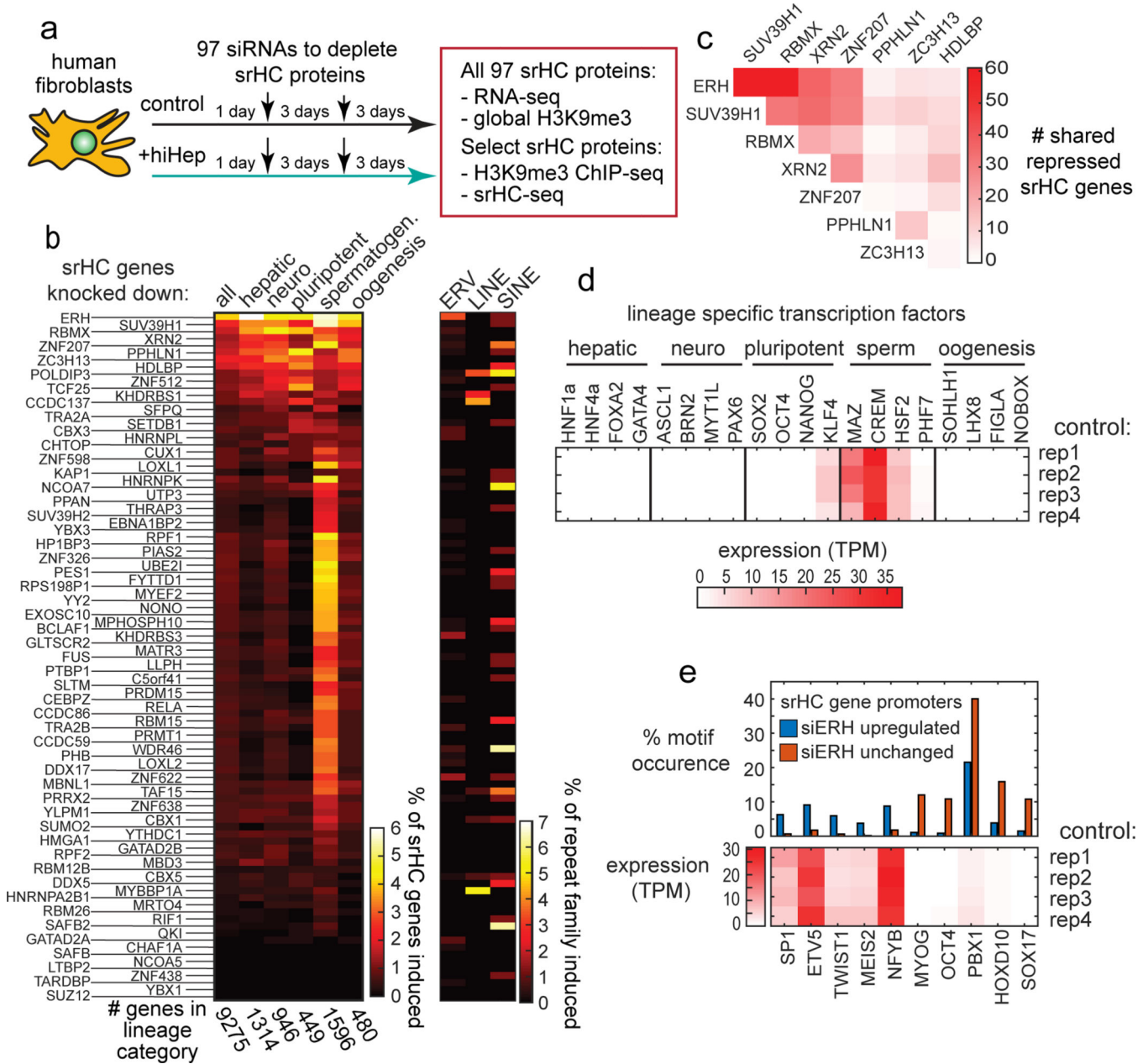
Figure 1 |.

Heterochromatin associated proteins maintain repression of genes and repetitive elements. **a**, Experimental design of siRNA treatments and sample collection. **b**, Fraction of srHC genes and repetitive elements upregulated vs siControl (Bejamini multiple test corrected Wald test p-value 0.05 and log2(foldchange)>0) by each knockdown for specified gene lineage category or repetitive element class in human fibroblasts; hierarchically clustered by target gene across all lineage categories. DESeq2 results are provided as source data. **c**, Number of all repressed srHC genes shared between the 8 indicated srHC proteins. **d**, expression in human fibroblasts across four siControl treated replicates of canonical transcription factors for hepatic, neuro, pluripotent, sperm and oocyte lineages. **e**, Transcription factor binding

site motifs enriched in promoters of srHC genes induced or not induced by *ERH* knockdown and their RNA expression in control cells.
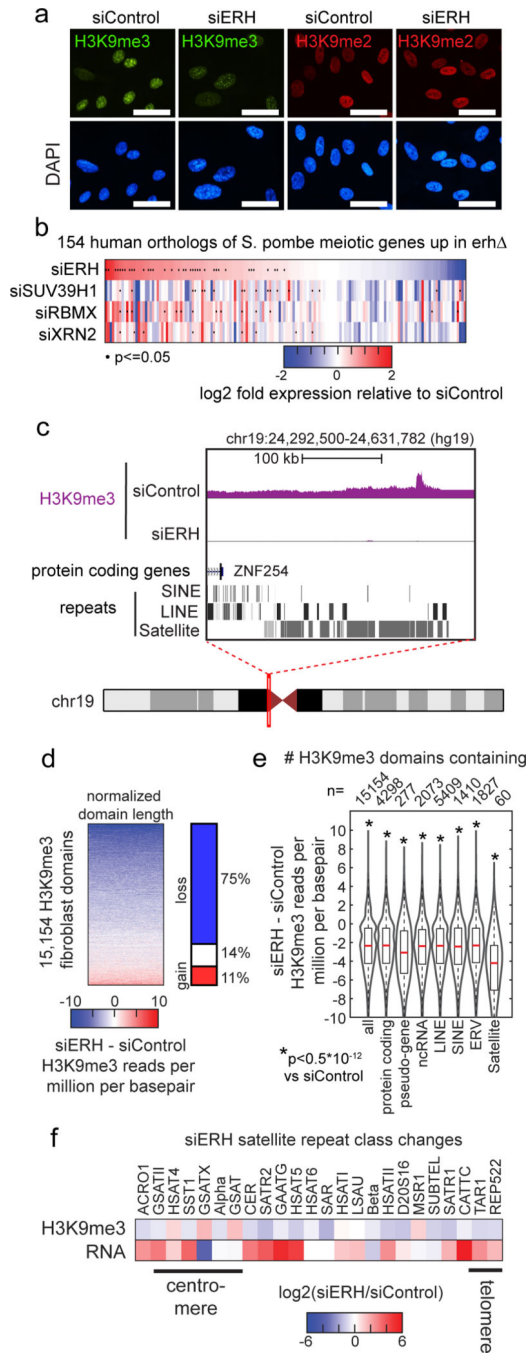
Figure 2 |.

ERH functions through conserved mechanisms to maintain H3K9 methylation and gene repression. **a**, Immunofluorescence of H3K9me3 (green), H3K9me2 (red) and DAPI (blue) after siControl or siERH siRNA treatment in human fibroblasts. Experiment repeated independently 8 times with similar results. Scale bars indicate 50 μm. **b**, Expression changes of 154 human gene orthologs of *S. pombe* meiotic genes upregulated in erhΔ compared to other knockdowns (DESeq2, Benjamini multiple test corrected Wald test p-value ≤ 0.05 and log2(foldchange)>0). Gene names, lfc and p-value provided as source data. **c**, Gene track of

example H3K9me3 domain showing H3K9me3 levels in siControl and siERH and locations of incident protein coding genes and repeats. **d**, Heatmap of H3K9me3 ChIP-seq in siERH minus siControl at 15154 length normalized H3K9me3 domains, and the corresponding percentages of domains with H3K9me3 loss, gain or no change. **e**, H3K9me3 changes for domains containing at least one of protein coding genes, pseudo-genes, non-coding RNA, LINE, SINE, ERV or Satellite repeats (one sample t-test *p<$0.5*10^{-12}$). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. N numbers are denoted in the panel and represent the number of annotated srHC domains containing each type of gene or repeat. **f**, Analysis of H3K9me3 levels and RNA expression by RepEnrich[37] for classes of satellite repeats after siERH knockdown. Statistical information is provided as source data.
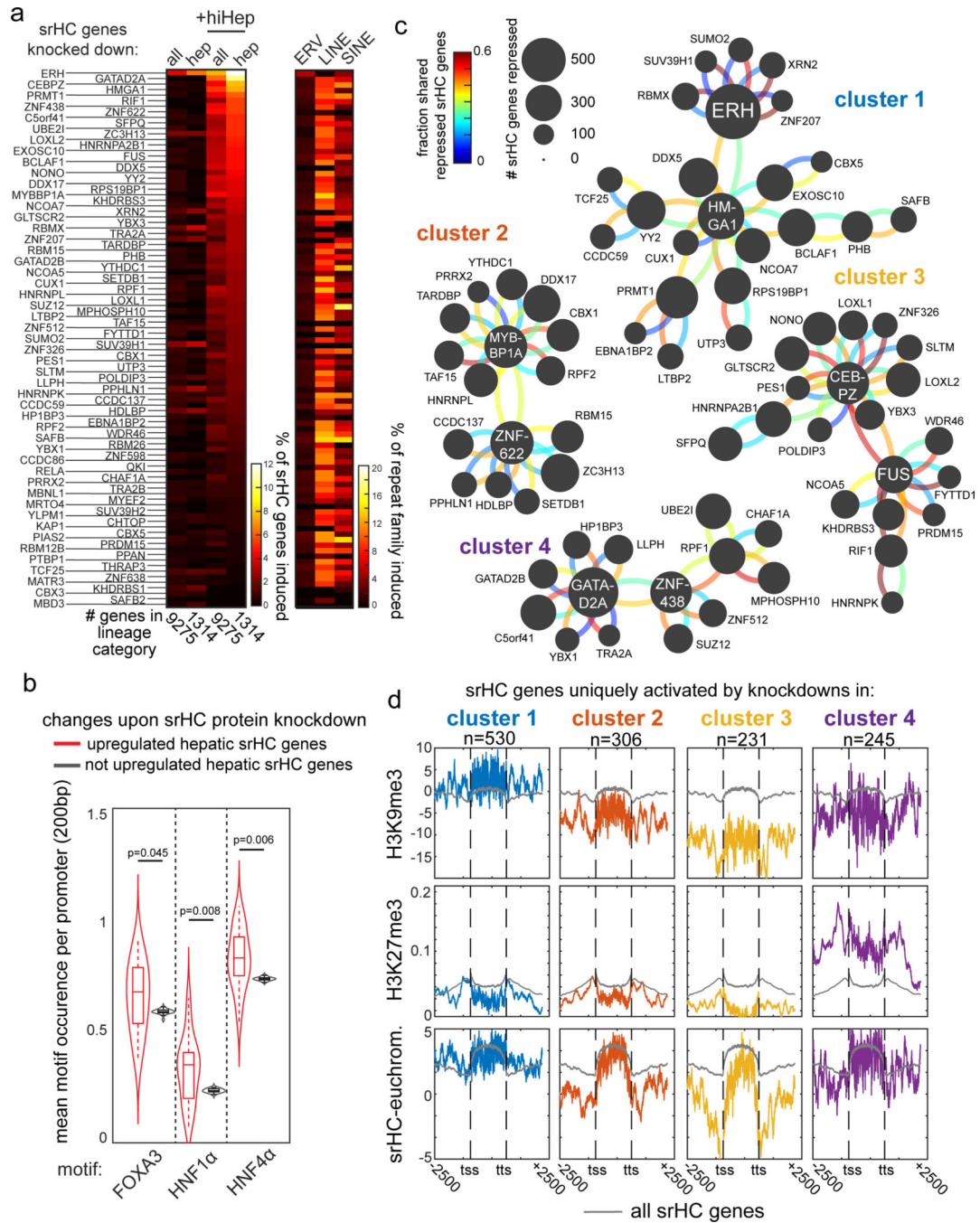
Figure 3 |.

Heterochromatin associated proteins function cooperatively and distinctly to regulate silencing during reprogramming. **a**, Heatmap representation of the fraction of srHC genes upregulated vs siControl (DESeq2, Benjamini multiple test corrected Wald test p-value 0.05 and log2(foldchange)>0) by each knockdown for specified gene lineage category in normal fibroblasts and hiHep reprogrammed cells. DESeq2 results are provided as source data. Knockdown targets ordered by number of hepatic genes upregulated in hiHep conditions. **b**, Violin plots showing hiHep transcription factor motif enrichment in all

hepatic gene promoters (black) and in activated genes (red) by all siRNA knockdowns upregulating at least 25 srHC hepatic genes during hiHep reprogramming (n=63 independent siRNA targets, one tailed Student's t-test). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values. P values denoted in panel. **c**, Graph of nearest neighbors clustering analysis of srHC genes upregulated by each knockdown (black nodes) during hiHep reprogramming, and display of fraction total genes shared with the connected knockdown (colored connections) represented in a counterclockwise manner. **d**, Histone profiles, input subtracted, for H3K9me3, H3K27me3 and srHC for srHC genes upregulated uniquely by each cluster. Statistical information is provided as source data.
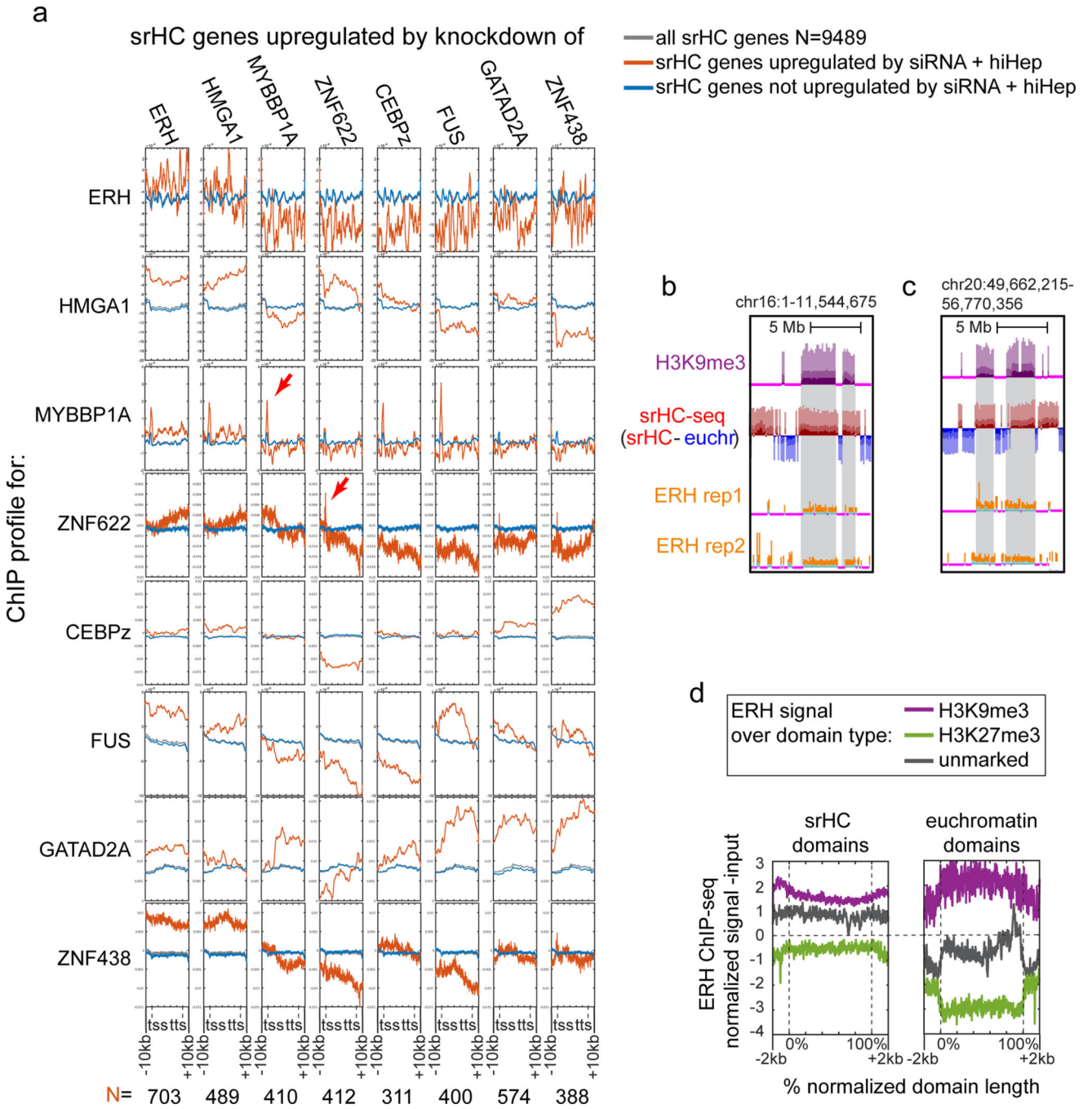
Figure 4 |.

srHC proteins bind repressed target genes. **a**, ChIP-seq profiles of 8 srHC proteins across the gene body +/−10kb of all srHC genes (grey), srHC genes upregulated by the indicated siRNA + hiHep condition (red) and srHC genes not upregulated by the indicated siRNA + hiHep (blue). **b, c,** Browser tracks showing H3K9me3, sonication resistance and two replicates of ERH ChIP-seq at regions of chromosome 16 (**b**) and chromosome 20 (**c**). **d**, ERH ChIP-seq signal across srHC and euchromatin H3K9me3, H3K27me3 and unmarked domains.
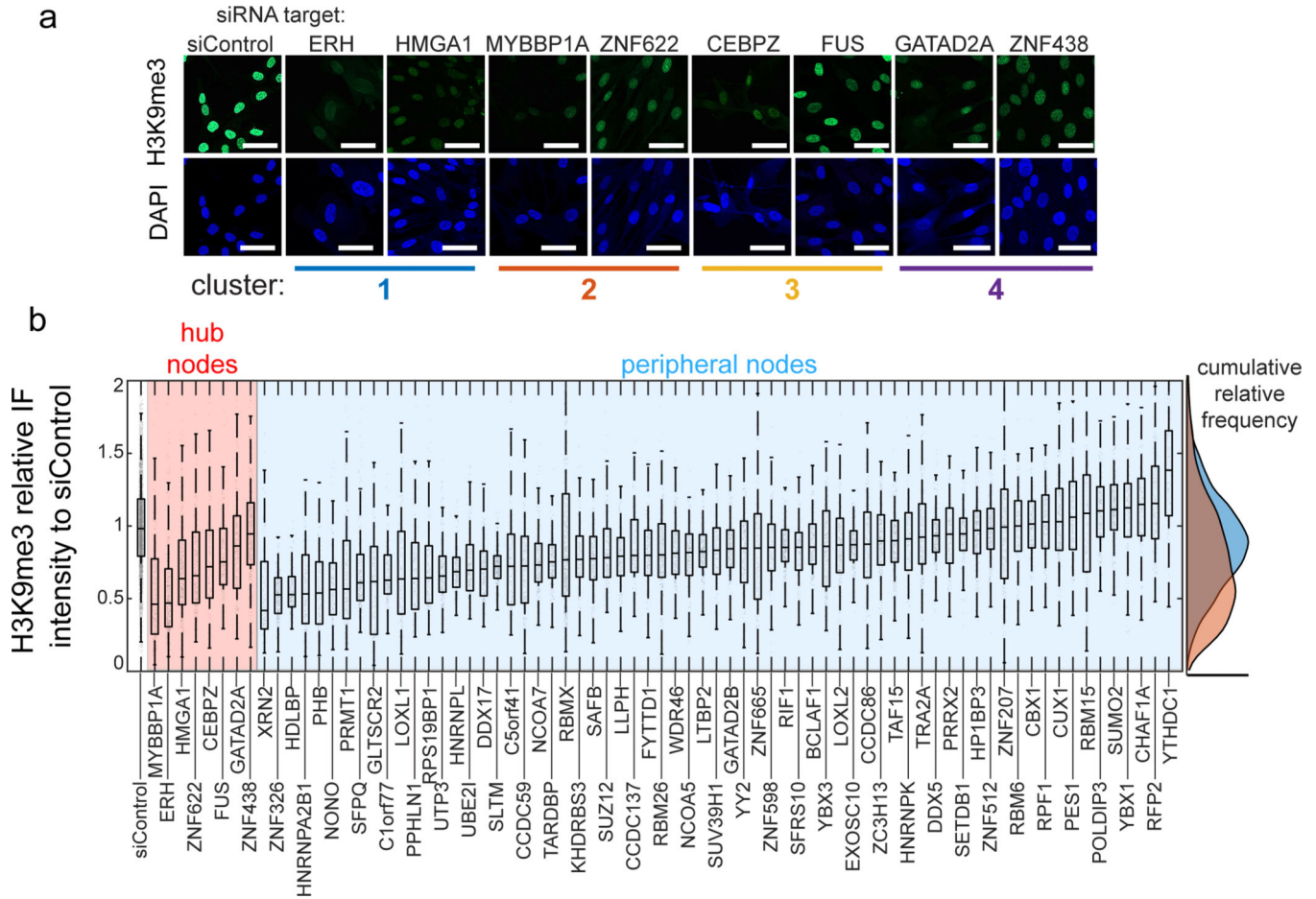
Figure 5 |.

srHC protein clusters regulate heterochromatin histone modifications. **a**, Example IF images showing H3K9me3 (green) changes and DAPI (blue) with hub node knockdowns in human BJ fibroblasts. Scale bars indicate 50μm. Image is representative of 4 experiments. **b**, Quantification of H3K9me3 immunofluorescence for knockdowns of hub and peripheral node srHC proteins relative to siControl (n>100 nuclei for each treatment). Boxplot center, bounds and whiskers represent the median, 25–75% range and minimum to maximum values respectively. Statistical information is provided as source data.
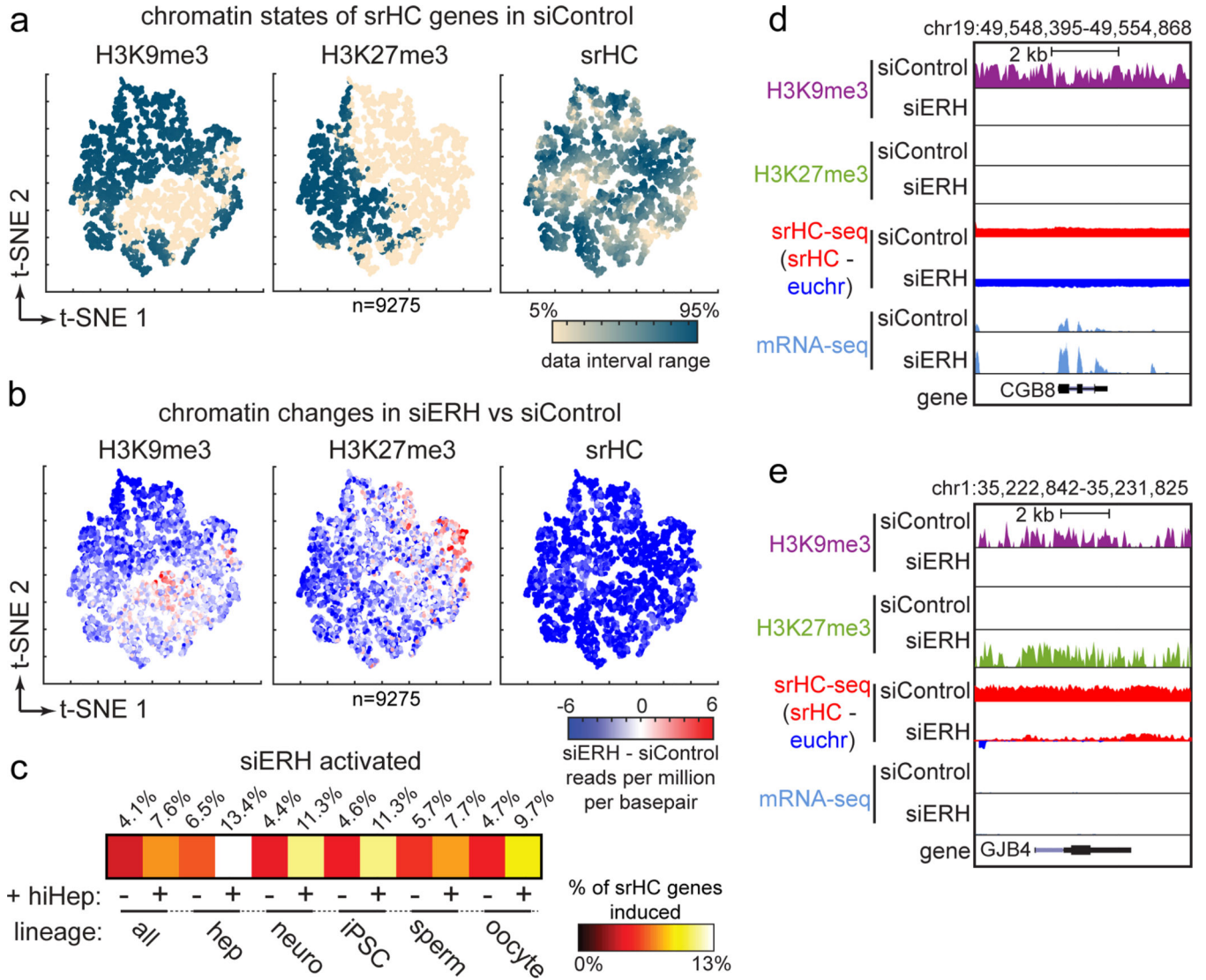
Figure 6 |.

Locus specific and global changes in heterochromatin drive de-repression of srHC genes. **a**, t-SNE plots of H3K9me3, H3K27me3 and srHC levels across 9275 srHC genes. Colorbar ranges from bottom 5% to upper 95% of data points. **b**, t-SNE plots of changes in H3K9me3, H3K27me3 and srHC levels in siERH relative to siControl. **c**, Heatmap displaying percentage of alternative lineage srHC gene activation by siERH during hiHep reprogramming. **d,e**, Browser track showing H3K9me3, H3K27me3, srHC and expression changes for example srHC gene losing H3K9me3 (**d**) and example srHC gene losing H3K9me3 and gaining H3K27me3 (**e**).