nature
genetics

Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma

Berkley E. Gryder [1]*, Silvia Pomella [1,2], Carly Sayers[3], Xiaoli S. Wu[4,5], Young Song[1],
Anna M. Chiarella [6], Sukriti Bagchi[3], Hsien-Chao Chou [1], Ranu S. Sinniah[1], Ashley Walton[1],
Xinyu Wen[1,7], Rossella Rota[2], Nathaniel A. Hathaway[6], Keji Zhao [8], Jiji Chen[9], Christopher R. Vakoc[4],
Jack F. Shern[3], Benjamin Z. Stanton [1,10] and Javed Khan [1]*

Core regulatory transcription factors (CR TFs) orchestrate the placement of super-enhancers (SEs) to activate transcription of cell-identity specifying gene networks, and are critical in promoting cancer. Here, we define the core regulatory circuitry of rhabdomyosarcoma and identify critical CR TF dependencies. These CR TFs build SEs that have the highest levels of histone acetylation, yet paradoxically the same SEs also harbor the greatest amounts of histone deacetylases. We find that hyperacetylation selectively halts CR TF transcription. To investigate the architectural determinants of this phenotype, we used absolute quantification of architecture (AQuA) HiChIP, which revealed erosion of native SE contacts, and aberrant spreading of contacts that involved histone acetylation. Hyperacetylation removes RNA polymerase II (RNA Pol II) from core regulatory genetic elements, and eliminates RNA Pol II but not BRD4 phase condensates. This study identifies an SE-specific requirement for balancing histone modification states to maintain SE architecture and CR TF transcription.

More than 1,500 transcription factors (TFs) are encoded by the human genome[1]. Some TFs are used across all human cell types (such as the general transcription factors[2]), whereas many TFs are restricted to a particular time and place in development[3,4]. In a given cell type, a few CR TFs, expressed at the highest levels, tend to dominate and determine the placement of large histone acetylation deposits, termed SEs[5], which form around a mosaic array of CR TF binding sites and drive cell-type-specific gene expression[6]. CR TFs are themselves driven by a subset of the SEs they form, and can be co-opted as essential dependencies in cancer[7,8].

CR TFs function by recruiting acetylation writers (such as the CBP/p300 proteins), readers (such as BRD4) and erasers (such as histone deacetylases, HDACs), among many other co-activators, to create SEs[9]. The entire axis of histone acetylation is essential for CR TF transcription[10]. Although the need to chemically add or recognize acetylation for enhancer-driven RNA Pol II transcription is well documented[11–14], it is not understood why CR TFs recruit HDAC-containing complexes to SEs.

Here, we determine and dissect the essential regulatory networks underlying childhood rhabdomyosarcoma (RMS) in primary tumors and cell lines, and use this disease context to mechanistically interrogate the consequences of hyperacetylation at the chromatin template. By using a combination of RNA sequencing (RNA-seq), single-cell RNA-seq and nascent ChRO-seq (chromatin run-on and sequencing), we find that CR TFs have a rapid and high sensitivity to HDAC inhibition. Spike-in normalized ChIP-Rx (chromatin immunoprecipitation with reference exogenous genome) and AQuA-HiChIP shows that hyperacetylated histones spread and

disrupt the three-dimensional (3D) organization of SEs, which destabilizes CR TF and RNA Pol II binding at SEs and dissolves RNA Pol II but not BRD4 condensate assembly. Thus, although histone acetylation is considered an 'active' chromatin modification, its deposition must be tempered and controlled to facilitate SE-driven core regulatory transcription.

Results

RMS core regulatory nodes include SOX8 and are selectively required for growth. To understand the epigenetic networks driving RMS, we sought to identify its regulatory circuitry. We performed analysis of SE-associated TFs across 21 RMS samples, both primary tumors and cell lines. Because RMS also relies on myogenic TFs, we cross-analyzed seven samples from the muscle lineage. SEs were defined with H3K27ac ChIP-seq experiments, for which we incorporated sample-matched RNA-seq data. For a given SE-associated TF_a (expressed at a level of at least 4 transcripts per million (TPM) in RNA-seq), the circuitry input (in degree) was calculated as the number of all TFs ($motif_b + motif_c$ and so on) with a motif present in nucleosome-depleted valleys of TF_a's SE. The output for TF_a (Fig. 1a) was calculated as the total number of SEs (with putative TF target genes; SE1, SE3, SE4 and so on) that had the motif of TF_a (out degree). Total connectivity (in + out degree, normalized to 1 representing maximum connectivity in the sample) predicted the TFs with high connectivity, the 'core' of the regulatory circuitry (Fig. 1a). In RMS samples, CR TFs formed four modules: first, a pan-RMS module that includes MYOD1 and MYOG; second, a fusion-positive (FP)-RMS only module that includes MYCN and FOXO1 (the SE regulating *PAX3-FOXO1* and *PAX7-FOXO1* gene
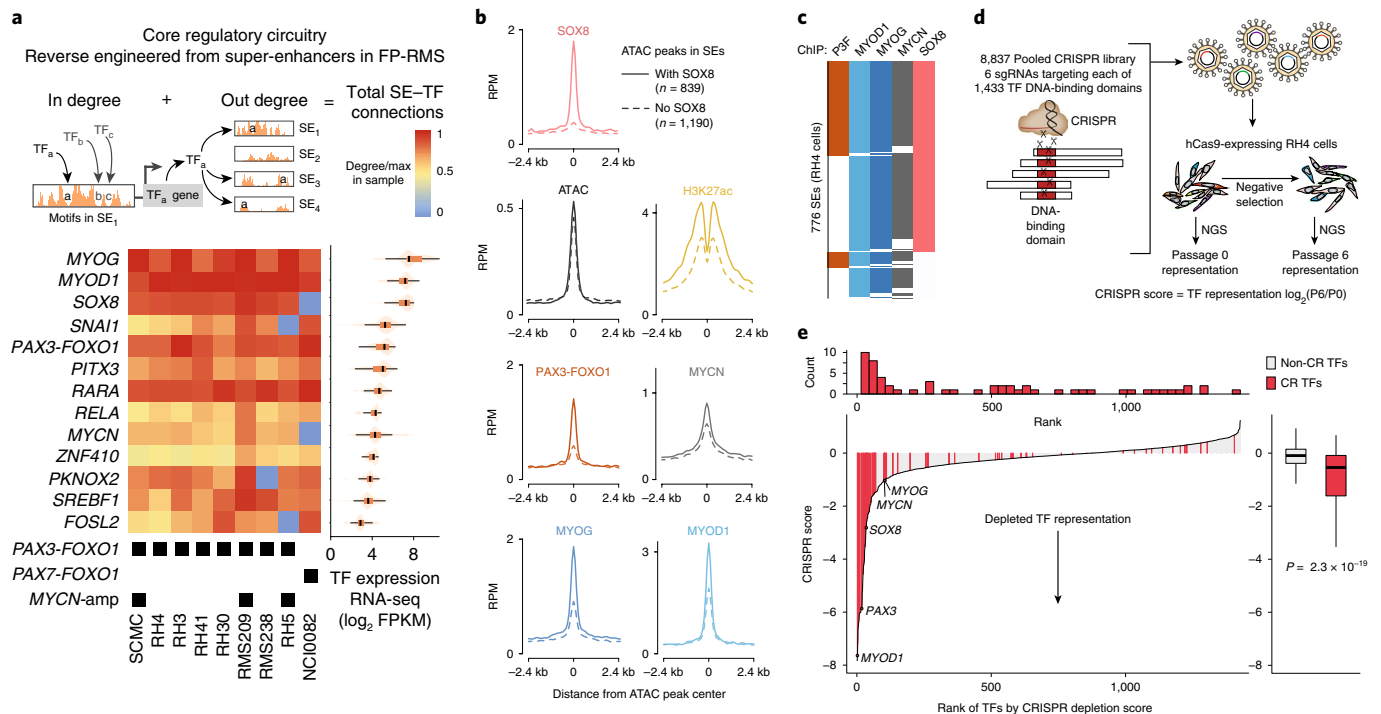
Fig. 1 | Core regulatory circuitry includes SOX8 and is critical for FP-RMS. a, Core regulatory circuitry identified by analysis of motif networks in SE associated TFs. Black arrows depict TF_a acting on its own SE, gray arrows indicate TF_b and TF_c that arise from elsewhere in the genome and act on the SE of TF_a. Wide gray arrow represents transcription of the TFa gene. Heatmap shows predicted CR TFs found in FP-RMS cell lines and tumors ($n=9$ independent samples), clustered and colored by degree of connectivity (scaled so that 1 represents the maximum connectivity in the sample). Expression of CR TFs is plotted on the right (box plots show quartiles, and whiskers show the $1.5\times$ interquartile range, with data distribution as violin plots). **b**, Core regulatory TF validation ChIP-seq in RH4 ChIP-seq (an FP-RMS cell line). Metagene plots at ATAC-seq peaks in SEs were divided into SOX8-bound (solid lines, $n=839$ peaks) and SOX-unbound (dotted lines, $n=1,190$ peaks). **c**, Prevalence and co-occurrence of CR TFs in SEs in RH4 cells. **d**, Functional genetic screening to evaluate TF dependencies in FP-RMS cells, using pooled CRISPR–Cas9 methodology. Each sgRNA ($n=8,837$) is designed to recruit Cas9 to the DNA sequence coding for the DNA-binding domain(s) of each human TF ($n=1433$). **e**, Depletion of sgRNAs targeting CR TFs, compared to all non-CR TFs, over six passages of negative selection (P0, passage 0; P6, passage 6). Top panel shows a histogram of the number of CR TFs (red bar). Lower panel shows a CRISPR score for each TF as an average $\log_2$[fold change] of sgRNAs ($n=6$ distinct sequences per TF) targeting DNA-binding domains. Right panel shows summary statistics in box (quartiles) and whisker ($1.5\times$ inter-quartile range) plot, with P value comparing the extent of depletion between CR TFs and all other TFs calculated with an unpaired, two-sided student's t-test with Welch's correction.

fusions')[15,16]; third, a fusion-negative (FN)-RMS module unified by the consistent presence of PAX7 and AP1-family TFs; and fourth, a normal muscle-specific module of CR TFs that includes Nur77 (encoded by *NR4A1*) and MEF2D (Supplementary Fig. 1a).

CR TF prediction identified SOX8 as consistently high-scoring across all PAX3-FOXO1 samples (Fig. 1a). SOX8 was validated by ChIP-seq, which revealed that it co-localizes with the other CR TFs in FP-RMS (Fig. 1b). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks in SEs that contain SOX8 ($n=839$) are more strongly bound by all other CR TFs and have the largest H3K27ac signal (Fig. 1b). SOX8 binds to 623 of 776 SEs in RH4 cells (Fig. 1c). Among SOX family members, *SOX8* was most highly expressed (Supplementary Fig. 1b) and overexpressed, compared to normal tissue (Supplementary Fig. 1c). Furthermore, histone acetylation network modeling placed SOX8 as a central hub (Supplementary Fig. 1d). Western blot analysis also showed that SOX8 was present at the protein level in two primary FP-RMS tumors (Supplementary Fig. 1e). These data support the inclusion of SOX8 as a previously unrecognized component of the core regulatory circuitry in RMS.

Analysis of Project Achilles CRISPR data demonstrated that *SOX8* and other CR TFs are essential to the growth of FP-RMS (Supplementary Fig. 1f), and uniquely so compared to 389 other cell lines across both pediatric and adult malignancies[17]. To validate this, we targeted six single guide RNAs (sgRNAs) per DNA-binding domain of all TFs in a pooled fashion (Fig. 1d) in RH4 cells. This cell line harbors 50 CR TFs, with an average of 1.8 SEs near to each CR TF gene locus (Supplementary Fig. 2a); these 50 TFs are the most connected among 113 TFs that are SE associated and not 'pan-essential' (that is, those TFs depleted across all cell lines with a median CRISPR depletion score of less than −0.4) (Supplementary Fig. 2b). We found that the CR TFs in FP-RMS, including SOX8, were critical to the maintenance of cancer cell proliferation (Fig. 1e). Those CR TFs with a depletion score of less than −1 (on a $\log_2$ scale) were defined as the 'top' CR TFs ($n=13$) in RH4 cells (Supplementary Fig. 2c,d). The dependence on these TFs agrees with recent evidence suggesting that myogenic factors are oncogenic in RMS[18].

Functional dissection of core regulatory circuitry reveals antimyogenic attribute of SOX8 in RMS. The hallmark t(2;13) (q35;q14) translocation of FP-RMS creates the fusion gene *PAX3-FOXO1*. As the primary oncogenic driver, we considered the possibility that other CR TFs (SOX8, MYOD1, MYOG) may support the growth of FP-RMS by contributing to the transcription of *PAX3-FOXO1*. ChIP-seq showed binding of these factors to the enhancer constituents of one another (Fig. 2a, with the exception of PAX3-FOXO1 showing no peaks near *MYOG*). To assay the 3D connectivity among these elements and their putative target genes,
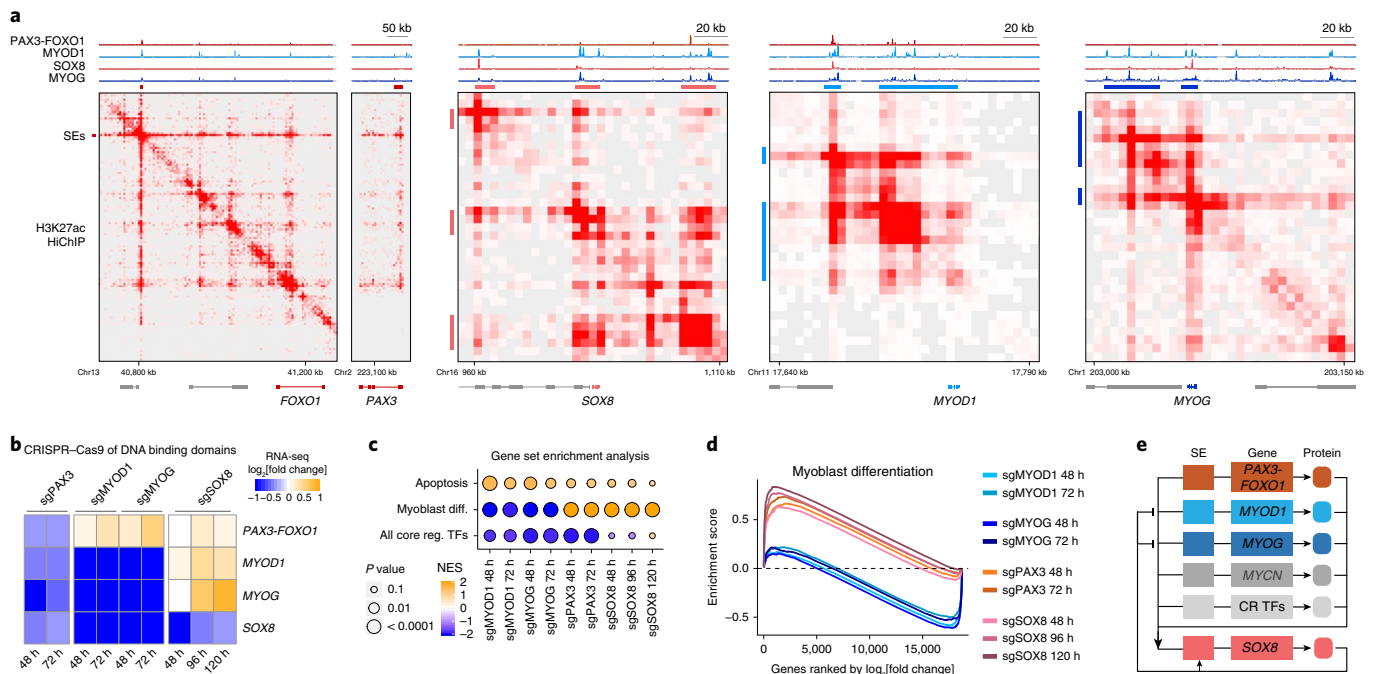
Fig. 2 | Genetic dissection of core regulatory network reveals a SOX8-mediated myogenic blockade in RMS. a, Connectivity of CR TFs shown by ChIP-seq of PAX3-FOXO1, MYOD1, MYOG and SOX8 at SEs that interact with the promoters of *PAX3*, *MYOD1*, *MYOG* and *SOX8*. Heatmaps depict interaction frequency between genomic elements, assayed by H3K27ac HiChIP (one representative of two experiments in RH4 cells). **b**, Co-regulation of CR TFs evaluated in RH4 cells with CRISPR–Cas9 targeted to the DNA binding domains of PAX3, MYOD1, MYOG or SOX8. Experiments were repeated orthogonally with shRNA with similar results (Supplementary Fig. 2e). **c**, Gene set enrichment analysis shows divergent transcriptional impact of individual CR TF disruption in RH4 cells (gene sets available in Supplemental Table 2). Similar results were obtained using shRNA (Supplementary Fig. 2f). *P* value was generated using the GSEA algorithm of the enrichment score relative to the null distribution calculated with 1,000 permutations. Each bubble represents one gene set analyzed against an individual RNA-seq experiment (sgRNA target versus non-targeting sgRNA) at the indicated time points. **d**, Activation and enrichment of myoblast differentiation genes upon sgRNA-mediated disruption of PAX3-FOXO1 or SOX8 were also seen upon shRNA knockdown (Supplementary Fig. 2g), in contrast to depletion of the myogenic differentiation program upon *MYOD1* or *MYOG* knockout. **e**, Model of auto-regulatory feed-forward (PAX3-FOXO1, MYOD1, MYOG, MYCN and other CR TFs) and negative feedback (from SOX8) circuitry in FP-RMS.

we performed HiChIP of H3K27ac, and identified direct interactions between SEs, nearby typical enhancers and each TF gene (Fig. 2a). The SE on chromosome 13 (near *FOXO1*) is connected to many additional enhancers and to the *PAX3* promoter on chromosome 2. Yet, individual dissection of CR TF function (using the two most potent sgRNAs per CR TF from our pooled screen) showed that expression of PAX3-FOXO1 was not reduced by disruption of MYOD1, MYOG or SOX8, whereas all factors were negatively regulated by disruption of PAX3-FOXO1 itself (Fig. 2b). Disruption of either MYOD1 or MYOG resulted in strong transcriptional depletion of *MYOD1*, *MYOG*, *SOX8* and other less prominent CR TFs (Fig. 2b,c).

SOX8 disruption caused upregulation of *MYOD1* and *MYOG* that persisted, suggesting a unique negative-regulatory role in FP-RMS. Consistent with this idea, it has been reported that SOX8 is a marker for muscle satellite cells that inhibits myogenesis[19]. We performed gene set enrichment analysis (GSEA) for genes upregulated during differentiation from myoblasts to myotubes. This program is strongly downregulated by CRISPR–Cas9 disruption of MYOD1 or MYOG, while perturbation of PAX3-FOXO1 or SOX8 activates these myogenic genes (Fig. 2c,d). Additionally, disruption of PAX3, MYOD1 or MYOG (but not SOX8) caused downregulation of CR TFs in RH4 cells (Fig. 2d). The distinct characteristics of these CR TFs were orthogonally confirmed by short hairpin RNA (shRNA) knockdown (Supplementary Fig. 2e–g). As PAX3-FOXO1 directly upregulates *MYOD1*, yet FP-RMS is unable to complete the MYOD1-driven myogenic differentiation program, the discovery that PAX3-FOXO1 also drives *SOX8* supports a core regulatory

model in which both positive and negative regulation (Fig. 2e) are able to lock RMS in its de-differentiated state.

HDACs are used by CR TFs. Gene activation is strongly positively correlated with histone acetylation. The most abundant sites of H3K27ac in the RMS epigenome, especially SEs[15], are bound by PAX3-FOXO1. Yet, PAX3-FOXO1 co-immunoprecipitates with the enzymes that erase this mark, nuclear HDACs[20]. SEs are associated with high levels of HDAC1 and HDAC2 binding in mouse embryonic stem cells[5,9].

To determine whether PAX3-FOXO1 co-binds with HDACs on the epigenome, we performed ChIP-seq on all three nuclear HDACs (HDAC1, HDAC2 and HDAC3). We found that each of these HDACs co-associated with PAX3-FOXO1 and other CR TFs at SEs (Fig. 3a). Increased HDAC genomic binding predicted greater expression (independent of CR TF binding), and CR TF targets were much more often found among genes with the highest levels of HDACs at their promoters (Fig. 3b). We next sub-classified genes by transcriptional output, finding a higher level of HDAC binding to the epigenome as expression increased, with the strongest association at CR TF targets (Supplementary Fig. 3a). We found that HDAC binding was strongest at the approximately 9,000 sites co-bound by all three HDACs, and that 91% of these triply-bound sites were co-occupied with CR TFs (Supplementary Fig. 3b). Our data highlighted that HDAC1, HDAC2 and HDAC3 binding is positively correlated with CT TF networks at the chromatin level. This raised the question of whether the removal of excess acetylation is essential for transcription at core regulatory domains.
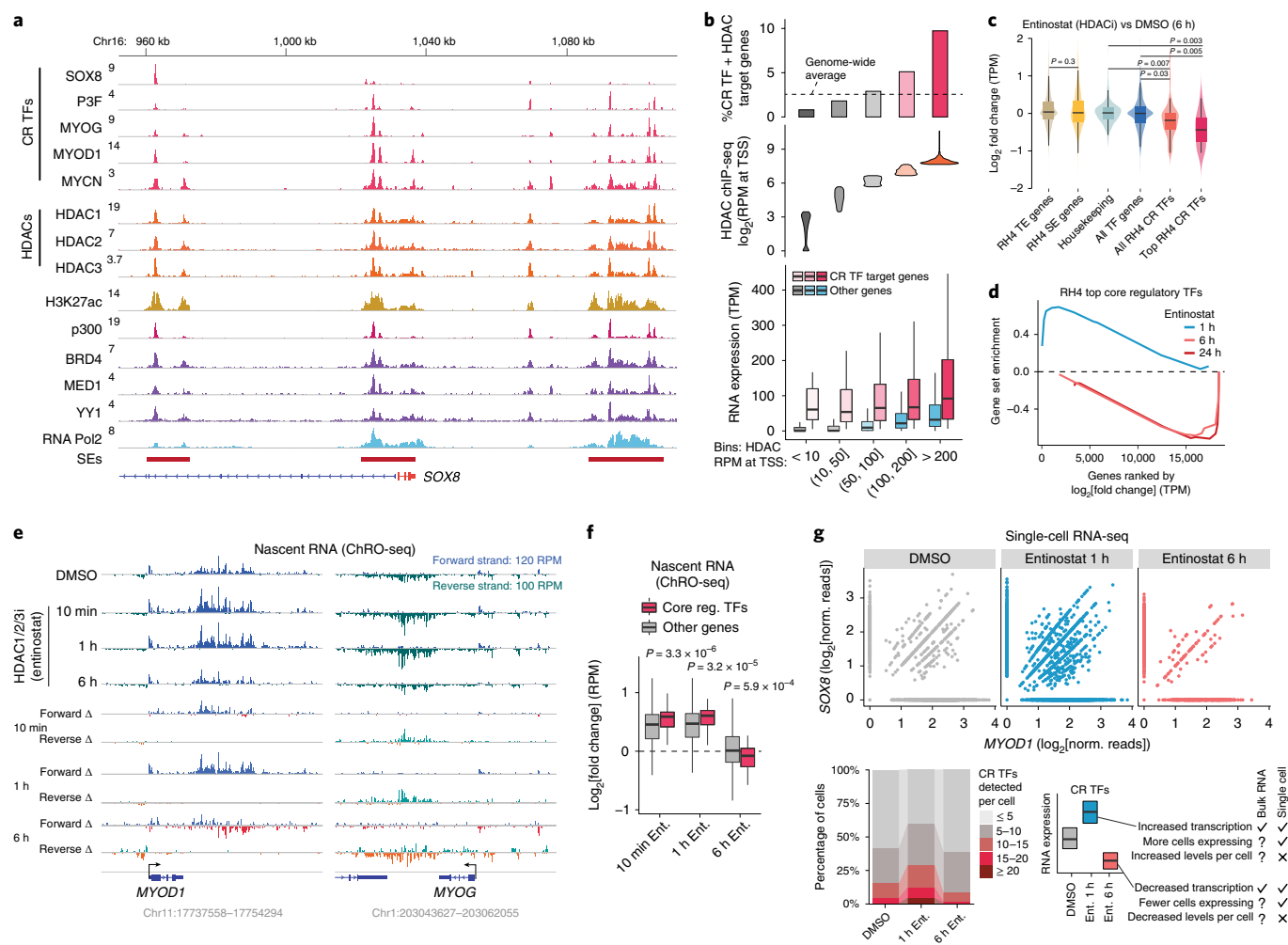
Fig. 3 | Core regulatory TFs require HDAC1, HDAC2 and HDAC3 for their transcription. a, Genome browser view at the *SOX8* locus, including three SEs (red bars) shown by ChIP-seq to be heavily decorated by CR TFs, HDACs (nuclear isoforms 1, 2 and 3), histone acetylation, other co-activators and RNA Pol II. Each track is one representative of two independent ChIP-seq experiments, except for HDAC3, MYCN, YY1 and MED1, which were performed once, and PAX3-FOXO1 and H3K27ac ChIP-seq data, which were performed more than four times across different RH4 cell passages. **b**, Enrichment of CR TFs in SEs associated with genes that have high amounts of HDAC bound to their promoters (top, dotted line indicates genome-wide average), genes divided into five categories of HDAC ChIP-seq density (middle, violin plots), which correlated positively with expression (bottom, box plots with median and quartiles, and whiskers show the 1.5× inter-quartile range). Bins were defined using a single representative HDAC2 ChIP-seq, and similar binding was seen with HDAC1 and HDAC3 (see Supplementary Fig. 3a,b). **c**, Changes in gene expression upon HDAC1, HDAC2 and HDAC3 inhibition with Entinostat for 6 h in RH4 cells. Violin plots are overlaid with box plots of median and quartiles, and whiskers show the 1.5× inter-quartile range. *P* values were calculated with a two-tailed unpaired *t*-test with Welch's correction. **d**, Time-course RNA-seq followed by gene set enrichment reveals that Entinostat causes an initial increase in core regulatory transcription (1h), followed by a selective downregulation at 6 h and 24 h in RH4 cells. False discovery rate (FDR) was calculated by GSEA with 1,000 permutations for each timepoint, and was $q = 0.006$ for 1h, $q = 0$ for 6 h, and $q = 0$ for 24 h. **e**, ChRO-seq to measure nascent RNA shows transcriptional increase at CR TFs *MYOD1* and *MYOG* in 10 min, which was sustained for the first hour, and decreased at 6 h of Entinostat treatment (1 μM in RH4 cells). **f**, Nascent RNA changes over time course treatment with Entinostat, as measured by ChRO-seq and summarized by box plots (median with quartiles, whiskers show the 1.5× interquartile range in the data). A two-sided, unpaired *t*-test with Welch's correction was used to calculate *P* values. **g**, Single-cell RNA-seq analysis of cells treated with DMSO ($n = 2,925$ cells) or Entinostat for 1h ($n = 3,805$ cells) or 6 h ($n = 3,240$ cells). Top panel shows an increased number of cells expressing the CR TFs *SOX8* and *MYOD1* at 1h of Entinostat and a decreased number at 6 h, compared to cells treated with DMSO. Bottom-left panel shows single-cell RNA-seq data of CR TFs abundance per cell in RH4 cells. These experiments were performed once. Bottom-right panel shows heuristic box plots and describes some relevant advantages of single-cell RNA-seq over bulk RNA-seq.

Dynamics of CR TF transcription upon HDAC inhibition.

The high levels of HDACs at loci encoding CR TF genes could translate into a selective transcriptional vulnerability. RNA-seq after treatment with an inhibitor selective for nuclear HDACs, Entinostat, revealed a marked decrease in expression of *SOX8*, *MYOD1*, *MYOG* and *MYCN* (Supplementary Fig. 3c). All core regulatory genes in RH4 cells, especially the highest-expressing CR TFs, were selectively downregulated by inhibition of HDAC1, HDAC2 and

HDAC3, whereas housekeeping genes, other TFs, or even SE genes generally were not downregulated (Fig. 3c). Interestingly, the first hour of HDAC inhibition resulted in increased expression of CR TFs, before a sharp decrease in expression at 6 h and continued suppression at 24 h (Fig. 3d and Supplementary Fig. 3d). Genetic depletion of individual HDACs revealed incomplete transcriptional hinderance of CR TFs (Supplementary Fig. 3e). To examine whether selective changes in CR TFs was owing to differences in

nascent transcription at these loci (as opposed to differences only in RNA transcript stability), we performed ChRO-seq[21] in RH4 cells treated with Entinostat for 10 min, 1 h and 6 h. We found a rapid initial increase and eventual decrease in nascent transcription at CR TF loci, exemplified at *MYOD1* and *MYOG* (Fig. 3e) and seen across all CR TFs (Fig. 3f). Furthermore, we calculated relative messenger RNA stability using combined analysis of ChRO-seq and total RNA-seq[22], and found indeed that CR TF transcripts are much less stable (Supplementary Fig. 3f–h), allowing relatively tight kinetics between transcriptional changes and net mRNA reduction, especially for the least stable transcripts, such as *MYCN* (as reported for *MYC* in previous studies[23]).

As enhancers control transcriptional bursts[24], we reasoned that the initial increase at 10 min and 1 h did not represent an increased number of transcripts in each cell, but an increase in the proportion of cells actively engaged. By use of single-cell RNA-seq (scRNA-seq) we tested and confirmed that the proportion of single cells expressing any given CR TF increased (Fig. 3g) whereas the expression levels of any given TF did not increase. It appears that co-expression of any two specific factors (considering *SOX8* and *MYOD1* as an example, Fig. 3g) is uncommon, but we interpret this as an artifact of sparsity and limited detection in scRNA-seq. The increase in expression of apoptotic genes, unlike the increase in expression of CR TFs, was the result of an increased quantity of reads per cell, not only an increased proportion (Supplementary Fig. 3i); also, while CR TF expression decreased at 6 h, expression of apoptotic genes continued to increase. Therefore, the scRNA experiments indicate that acute hyperacetylation does not necessarily increase the maximum output of SE controlled CR TFs, but activates transcriptionally resting cells, before the eventual halting of CR TF transcription at later time points.

Enhancer spreading and loss of architectural integrity at CR TFs. We reasoned that increased histone acetylation at CR TFs may explain the initial burst and subsequent crash of CR TF transcription. In previous studies of H3K27ac changes with HDAC inhibitors, ChIP-seq showed drug-induced spreading of acetylation, and yet a decrease in acetylation at enhancer peaks[25,26]. We reasoned that, because these previous studies normalized by sequencing depth, a global increase in H3K27ac could not have been detected even if it was occurring, which prevented correct interpretation of the data. To evaluate this, we incorporated reference exogenous *Drosophila* chromatin (ChIP-Rx)[27]. Indeed, analysis and normalization without spike-in gave an apparent decrease in H3K27ac at SEs, whereas reference normalization revealed a strong increase (Supplementary Fig. 4a). The boundaries of SEs were stable within the first hour of HDAC inhibition with focal increases in acetylation. In contrast, by 6 h, we observed spreading (or outward diffusion) of acetylated chromatin beyond SE boundaries (Fig. 4a).

CR TFs require looping machinery, such as cohesin and YY1, to bring TF-bound SEs into close spatial proximity to their gene body, and to form insulated neighborhoods with CTCF[28,29]. Drug-induced hyperacetylation at SEs, which constitutes a new chemical composition of the chromatin template, may alter looping. To investgate this, we began with ChIP-Rx of YY1 and BRD4, in RH4 cells treated with DMSO or Entinostat for 6 h. YY1 and BRD4 bound more strongly at both CTCF sites and the surrounding chromatin upon treatment with Entinostat for 6 h (Supplementary Fig. 4b,c). YY1 primarily occupied SE constituents and target promoters, whereas RAD21 (a member of the cohesin complex) bound preferentially to CTCF-occupied sites (insulators, Supplementary Fig. 4d). Upon HDAC inhibition, RAD21 binding increased subtly, whereas YY1 loading increased markedly (Supplementary Fig. 4d). In a similar manner, BRD4 showed increased binding, whereas acetylation writer p300 and erasers HDAC2 and HDAC3 showed a mild binding reduction (Supplementary Fig. 4d). Histone deacetylase inhibitors (HDACi) caused accessibility (ATAC-seq) increases at nucleosome depleted

valleys in SEs (Supplementary Fig. 4d) flanked by regions of H3K27ac spread, which highlights the connection between genomic deposition of H3 acetylation and accessibility.

Molecules that inhibit HDAC isoforms non-selectively have been used in previous studies that found histone acetylation spreading[25,26]. Entinostat also achieves spreading, but inhibits only the class I nuclear HDAC isoforms HDAC1, HDAC2 and HDAC3. These HDACs co-occupy SEs with near identical patterning (Supplementary Fig. 5a). In an attempt to further narrow down the isoforms of HDAC required for CR TF transcription, we used inhibitors selective for HDAC1 and HDAC2 (Merck60) or HDAC3 (LW3). Strong downregulation of *MYOD1* (Supplementary Fig. 5a) required pharmacological inhibition of all three class I HDACS, and either HDAC1 and HDAC2i or HDAC3i alone were unable to increase histone acetylation to the same extent as HDAC1, HDAC2 and HDAC3i (Fig. 4b and Supplementary Fig. 5b). We confirmed that 3 days of CRISPR-mediated individual disruption of HDAC1, HDAC2 or HDAC3 caused only a slight increase in H3K27ac, and no spreading (Supplementary Fig. 5c). Triple inhibition of HDAC1, HDAC2 and HDAC3 with Entinostat increases histone acetylation not only at H3K27 but also at H2BK5 and H4K16, and to a lesser extent at H3K36 (Fig. 4b).

To test whether two-dimensional SE spreading was associated with aberrant engagement of SEs to loci encoding CR network promoters or other genomic targets, we devised a modified HiChIP[30,31] to capture global changes in contact frequencies that may be masked without external normalization. Termed AQuA-HiChIP, the method incorporated the addition of an identical amount of fixed mouse cells to human cells immediately prior to in situ Hi-C contact generation. After enrichment (in this case of H3K27ac) with an antibody possessing cross-species reactivity, contacts were captured by biotin and library preparation was performed on streptavidin beads. The ratio of mouse and human contact pairs was quantified, and the AQuA contact frequency (reference normalized contacts per million) allowed us to define absolute changes in 3D SE dynamics upon HDACi. Indeed, the hyperacetylation structures in normal HiChIP were masked, but were uncovered by AQuA-HiChIP (Supplementary Fig. 6a).

AQuA-HiChIP in RH4 cells revealed that upon HDACi, new interactions invade unmodified chromatin, such as the 40-kb (kilobase) gap between *MYOD1* SEs (Fig. 4c). These new interactions were not observed in classical HiChIP but were uncovered by the AQuA-HiChIP methodology. Although many aberrant new interactions grow, the prominent SE-to-SE interaction at *MYOD1* is diminished (Fig. 4d). The apparent 'spreading' of AQuA-HiChIP is greater than that seen in ChIP-Rx (Supplementary Fig. 5c), but each 3D contact pair can be composed of both a direct and DNA-ligation-mediated association with an acetylated histone. The spreading phenomenon was observed at other CR TFs that are regulated by one or more SEs, such as *SOX8* (Supplementary Fig. 6b), *PAX3-FOXO1*, *FOXM1*, *JUN*, *MYCN*, *MYOG*, *RARA* and *SIX2* (Supplementary Fig. 6c). New hyperacetylation-induced interactions are seen at SE pairs genome-wide but are confined by CTCF boundaries (Fig. 4e).

The loss of SE contacts, and the gain of excessive aberrant contacts, may help to explain why transcription of CR TFs is most sensitive to HDAC inhibition. But if the role of HDACs at SEs is to prevent spreading, it is not clear why they are bound to the epicenters (with p300) rather than the boundaries. We reasoned that the shape of binding indicates nucleosome movement and turnover[32]: p300 and HDACs do not directly overlap H3K27ac, but are immediately adjacent, and the signal of H3K27ac at SEs tapers off over distances (>2,000 bp) much greater than that of a single nucleosome (147 bp). Thus, interference with the catalytic balance of acetylation writers by inhibition of HDACs at the epicenter may then affect spreading at distal boundaries in a tapered fashion
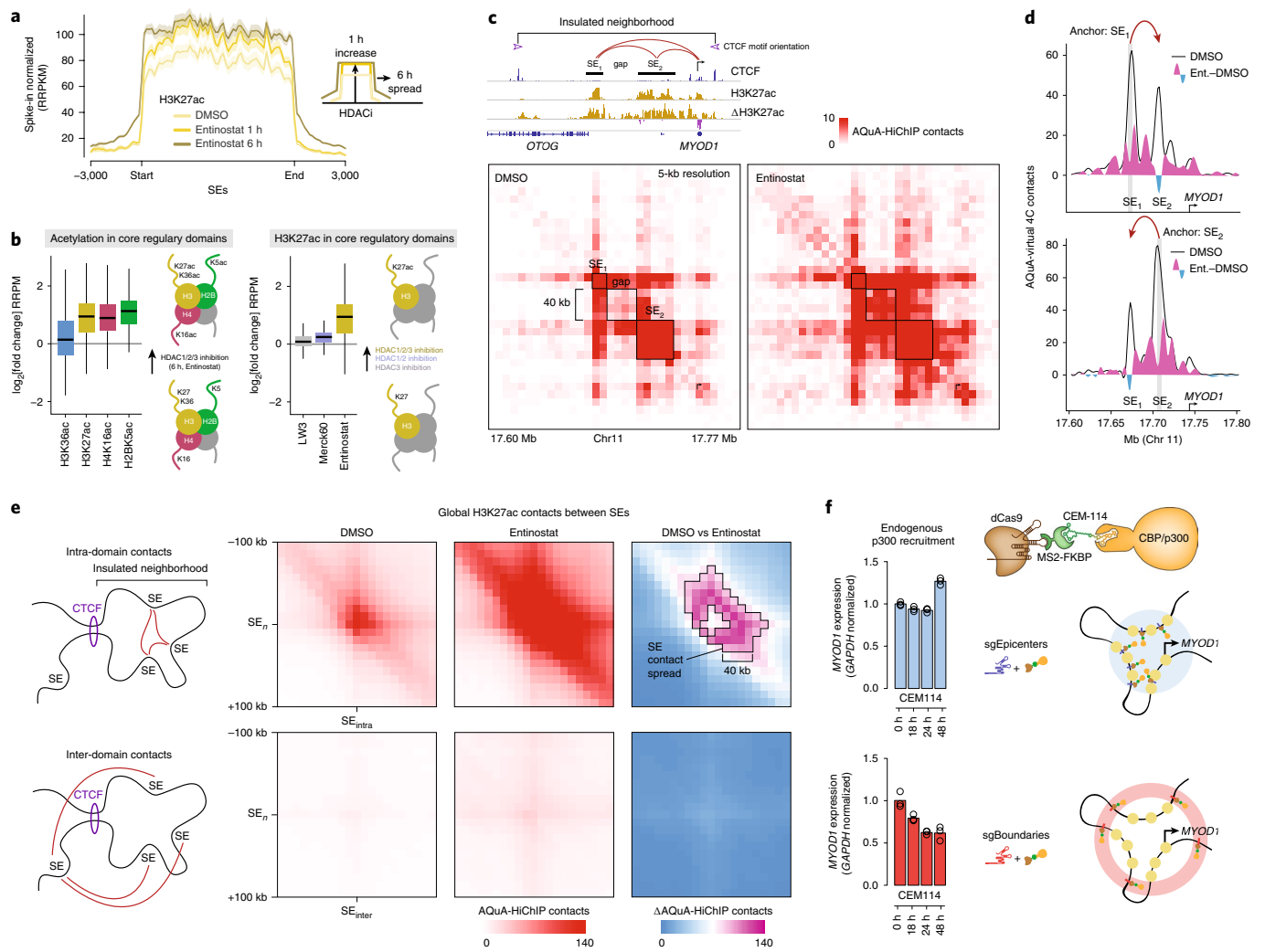
Fig. 4 | AQuA-HiChIP shows disruption of SE architecture by hyperacetylation. a, SE dynamics upon HDAC inhibition revealed an acute increase in H3K27ac after 1h, followed by a spread beyond the endogenous SE boundary at 6h. ChIP-Rx with exogenous spike-in is reported as reference-normalized reads per kilobase per million mapped reads (RRPKM). Shading shows the s.e.m. **b**, Acetylation changes quantified on diverse lysine residues on histone subunits H3 (K36, K27), H2B (K5) and H4 (K16) by ChIP-Rx. Reference-normalized reads per million mapped reads (RRPM) differences for diverse acetylation sites on histones are shown for core regulatory domains after treatment with either DMSO or Entinostat for 6h (left), or for H3K27ac upon treatment with HDAC1 and HDAC2 inhibitor Merck60, HDAC3 inhibitor LW3 or HDAC1, HDAC2 and HDAC3 inhibitor Entinostat (right). All experiments were carried out in RH4 cells treated for 6h with 1μM of the indicated inhibitors. Box plots show the median and quartiles, and whiskers show the 1.5× interquartile range. **c**, AQuA-HiChIP identifies that SE-mediated contacts spread within the insulated neighborhood of CR TF *MYOD1*. Contact map is shown at 5 kb resolution (5-kb by 5-kb contact squares), and is scaled to AQuA normalized contacts per million. **d**, Gained aberrant contacts and lost SE-to-SE contacts are visualized by AQuA-Virtual 4C at *MYOD1* from viewpoint anchor SE$_1$ (top) and SE$_2$ (bottom). Virtual 4C is representative of two replicate biotin captures and library preparations, both with similar results, and agrees with non-virtual 4C experiments at *MYOD1* SEs under the same treatment conditions. **e**, SE contact spreading as seen by AQuA-HiChIP aggregate peak analysis (APA) plots of all SE-to-SE contacts within insulated neighborhoods (SE$_n$ to SE$_{intra}$, top) or SEs near to but outside of insulated neighborhood CCCTC-binding factor (CTCF) boundaries (SE$_n$ to SE$_{inter}$, bottom). Resolution is shown at 10-kb by 10-kb squares. **f**, Endogenous p300 recruitment to *MYOD1* SE elements. CEM-114, bi-functional FKBP-binder and p300 bromodomain binder) enables dCas9-guided recruitment of p300 to MYOD1 SE epicenters (sgEpicenters) or SE boundaries (sgBoundaries) in RH4 cells. Triplicate values are one representative of two independent cell treatments, each with two quantitative PCR with reverse transcription (RT-qPCR) replicates that gave similar results.

by increasing the local concentration of acetylated histones. This agrees with the shape and breadth of spreading induced by Entinostat. To directly recapitulate transcriptional downregulation by increase in histone acetylation past the boundary regions at a single SE, we chose to navigate p300 to specific locations using dCas9 with chemical-induced proximity binding[33]. Thus, we engineered RH4 cells to stably express dCas9 and MS2-FKPB to enable sgRNA-directed recruitment of endogenous p300 by means of a bifunctional small molecule, chemical epigenetic modifier-114

(CEM-114) (FK506 linked to a binder of the bromodomain of p300)[34]. Targeting of the SE epicenters had little impact on transcription (possibly because p300 abundance was already high at these locations), but recruitment of p300 past the SE boundaries caused downregulation of *MYOD1* (and only upon co-administration of 50 nM CEM-114, Fig. 4f).

HDAC inhibition decommissions binding of CR TFs and RNA Pol II at SEs. To further investigate the directness of the effects
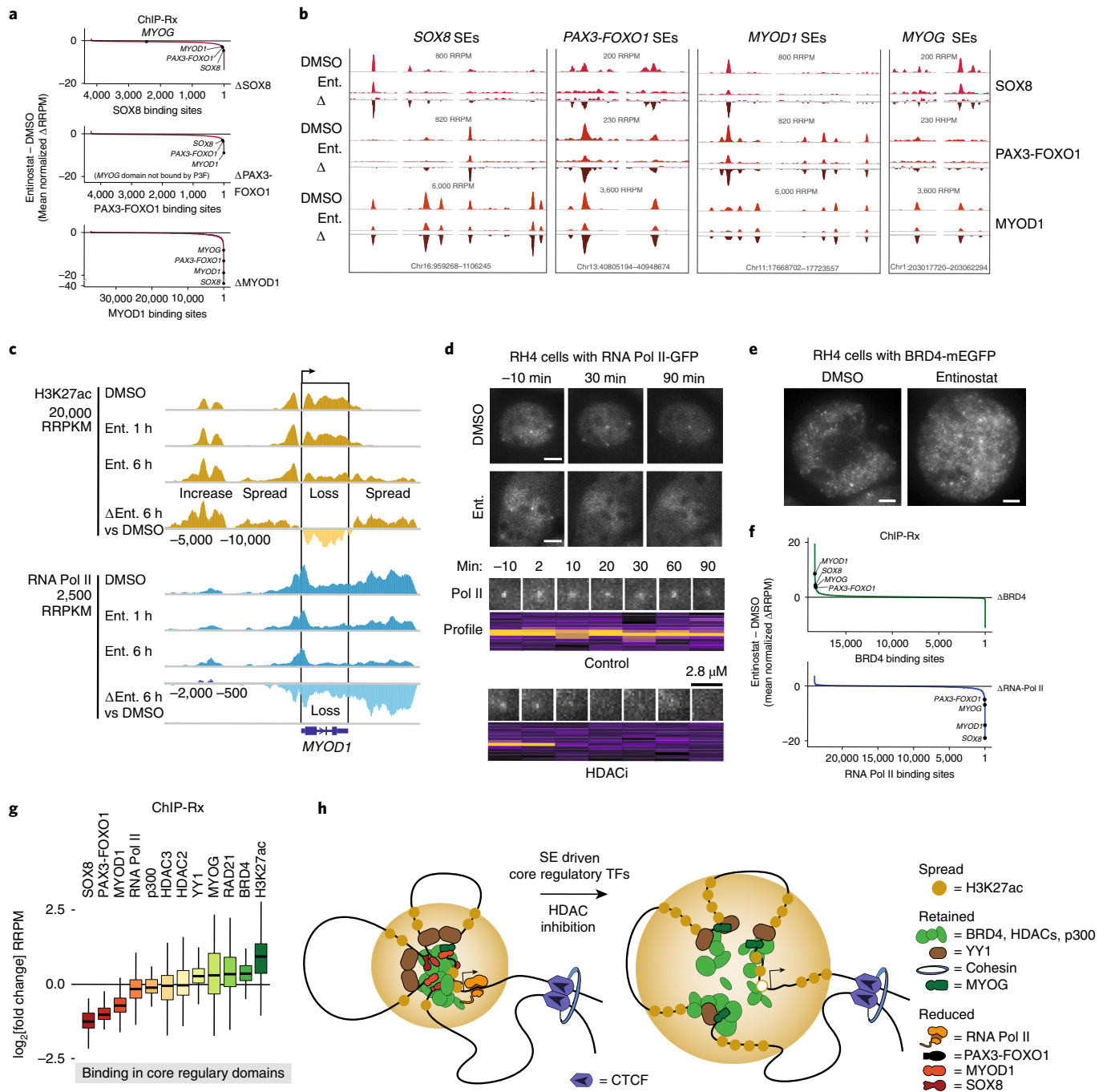
Fig. 5 | SE clusters and phase condensates are disrupted by hyperacetylation. a, ChIP-Rx binding sites for CR TFs, ranked by change in binding upon 6 h of Entinostat treatment in RH4 cells. **b**, Entinostat-induced changes in binding of CR TFs shown at *cis*-regulatory elements for *SOX8, PAX3-FOXO1, MYOD1* and *MYOG* in RH4 cells. **c**, RNA Pol II unloading along all genic positions of *MYOD1* at 1 h and 6 h of Entinostat treatment, and associated changes in H3K27ac, as measured by ChIP with spike-in normalization (ChIP-Rx). **d**, Clusters of RNA Pol II tagged with GFP (in live FP-RMS cells, RH4), imaged in a time course with or without addition of HDAC inhibitor Entinostat. Min, time since beginning of time course. Scale bar, 5 μM. Representative single puncta of RNA Pol II clusters over time are shown (2.8 μM squares) from control or HDAC inhibitor treated RH4 cells. Cross-section profiles of pixel brightness for corresponding time points are shown below image frames. Images are from wide-field (see Online Methods) and are representative of 50 images across 2 independent experiments with similar results. Alternative imaging modality iSIM gave similar results to those obtained with wide-field, and is shown in Supplementary Fig. 7f. **e**, Live cell imaging of BRD4 (endogenously tagged with mEGFP in RH4 cells) treated for 6 h with DMSO (left) or Entinostat (right). Scale bar, 5 μM. Images are representative of 20 images across 2 independent experiments with similar results. **f**, Quantification of RNA Pol II (top) and BRD4 (bottom) binding changes upon HDAC1, HDAC2 and HDAC3 inhibition with Entinostat (1 μM, RH4 cells) by ChIP-Rx signal. **g**, HDAC inhibition-induced changes in binding of key factors within core regulatory domains, measured by ChIP-Rx in RH4 cells, summarized by box plots that represent the median and quartiles, with whiskers showing the 1.5 × interquartile range. **h**, A model whereby histone hyperacetylation disrupts normal looping interactions, reduces binding of certain CR TFs, reduces binding of RNA Pol II, and causes loss of transcription at SE-dependent CR TF genes.

of HDAC1, HDAC2 and HDAC3 inhibition, we next performed ChIP-Rx of PAX3-FOXO1, MYOD1, SOX8, MYOG after 6h treatment of RH4 cells with Entinostat. We then ranked each CR TF binding site by the total amount of change, and found PAX3-FOXO1, SOX8 and MYOD1 to be reduced at the SEs regulating these important CR TFs (Fig. 5a,b), whereas MYOG showed very subtle changes in binding (data not shown). In addition, hyperacetylation caused RNA Pol II to be largely removed, exemplified at *MYOD1* (Fig. 5c), which loses H3K27ac in the gene body and undergoes H3K27ac spreading into the transcriptional end site (TES) region. The loss of RNA Pol II at CR TFs coincides with increased H3K27ac in the TES region; this effect was not seen at typical TFs (Supplementary Fig. 7a). The ratio of genic H3K27ac spread was greatest for shorter TFs (including *SOX8*, *MYOG*, *MYOD1* and *MYCN*), compared to all Pol2 bound genes, regular TFs or even other SE genes (Supplementary Fig. 7b,c). Reduction of Pol2 in the gene body and at the TES may occur by inhibition of pause release, but HDACi-induced pausing was ruled out at *MYOD1* as the decrease did not coincide with an increase of Pol2 at the transcription start site (TSS)-proximal pause site. Instead, unloading of Pol II was seen at the promoter, the TSS region and the gene body. CR TFs exhibit more elongation in the ground state than do other genes (Supplementary Fig. 7d, left), but whereas HDAC inhibition induces pausing genome-wide (Supplementary Fig. 7d) in agreement with previous work[25], this does not occur at CR TFs (Supplementary Fig. 7d, right).

Disruption of RNA polymerase clusters. Many genomic sites lose RNA Pol II upon HDACi, but Pol2 clusters found at SEs are the most disrupted (Supplementary Fig. 7e; *t*-test, $P < 2.2 \times 10^{-16}$). As a result of our ChIP-Rx observations, we reasoned that such clusters would be visibly dissipated by HDACi. Recent high-resolution imaging of RNA Pol II in live cells revealed clustering of up to 80 molecules of RNA-Pol II (ref. [35]), and these clusters reside in liquid–liquid phase separated droplets at SEs[36]. In RH4 cells stably transduced with the RNA Pol II subunit RPB3 tagged with GFP, we identified RNA Pol II clusters in single cells that were stable over time (Fig. 5d). These RNA Pol II-GFP clusters rapidly dissipated upon HDAC inhibition (Fig. 5d). The size and brightness of RNA Pol II puncta are asymmetrically distributed, similar to those of SEs, and are reduced by HDACi (Supplementary Fig. 7f).

PAX3-FOXO1 and the other CR TFs both recruit and rely on BRD4 to mediate their output at SEs. To evaluate whether BRD4 forms condensates in RMS, we knocked-in monomeric enhanced green fluorescent protein (mEGFP) to the endogenous human *BRD4* gene in RH4 cells, and found that it forms the predicted structures. In contrast to the rapid disappearance of RNA Pol II puncta, BRD4 puncta are resilient to hyperacetylation. These visual changes were reflective of ChIP-Rx binding changes; whereas RNA Pol II is lost at CR TF gene loci, BRD4 is gained (Fig. 5f). In studying the binding of many relevant chromatin factors by using ChIP-Rx, we found that Entinostat reduces binding of three of the four essential CR TFs that were measured, and reduces RNA Pol II, whereas the competing acetylation enzymes (p300, HDACs) are stable, and YY1, RAD21 and BRD4 are increased (Fig. 5g). These divergent responses to hyperacetylation may be underpinned by rules of inclusion or exclusion from phase condensates at SEs that may be elucidated by future work.

The loss of SE architectural integrity by hyperacetylation of histones and increased accessibility, with concomitant dissipation of RNA Pol II clusters, provides a plausible mechanism by which a drug with a non-gene-specific target (HDAC) could have a focused effect on only certain genes that are particularly dependent on co-localization with large enhancer elements (Fig. 5h). Thus, 3D genome architecture can be targeted as a vulnerability in RMS by modulation of HDAC activity.

Discussion

Core regulatory TFs not only govern epigenetic status in FP-RMS, but also represent disease-critical targets[5,7]. We discovered and confirmed that the developmental gene *SOX8* is critical among CR TFs. SOX8 regulates early neural crest development[37]. CR TFs MYOD1 and MYOG lead to pro-myogenic differentiation, in RMS as in satellite muscle cells, whereas SOX8 counteracts the ability of these factors to complete the muscle lineage. As SOX8 has SEs bound by PAX3-FOXO1, it provides a mechanistic link to the anti-differentiation activity of the fusion. SOX8 is reminiscent of SOX2 among CR TFs in embryonic cell circuitry[4], as both factors promote the dedifferentiated state, yet these are distinct in that SOX2 is a positive regulator of embryonic CR TFs, and SOX8 is a negative regulator of RMS CR TFs. This opens an important avenue of future research in RMS biology.

Phase condensates form compartments of concentrated components needed for biochemical reactions[38], and SEs perform this function at the chromatin level[39]. Furthermore, oncogenic TFs can position phase condensates[40]. The choice of components (that is, which proteins, nucleic acid polymers and small molecules exist in a phase condensate) is driven by biophysical properties, and post-translational acetylation is a means of decreasing a fundamental property, charge. Our data show that SEs do not fit the classical paradigm that increased acetylation and decreased HDAC activity are associated with higher levels of transcription. We instead suggest a paradigm whereby, at critical CR TF loci, transcription cannot be maintained without the counterbalance to acetyltransferases, HDACs. This model is supported by HDACi-induced erosion of SE boundaries, the invasion of acetylated histones into surrounding unmarked chromatin, the dissipation of 3D enhancer loops, the unloading of RNA Pol II at CR TFs and dissolving of RNA Pol II clusters upon hyperacetylation. When considered together, our findings begin to shed light on why HDACs are essential to CR network circuitry. Furthermore, our results offer a new mechanism of action for reinterpreting historical studies of HDACi in cancer therapeutics, perhaps especially for TF-driven cancers. Future experiments will help to determine how overall charge balance and distribution of modifications on the disordered histone tail influences the formation of larger condensates and modulates transcription. These properties will shape an important new framework for interpreting chemical epigenomics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0534-4.

References

1. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
2. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
3. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
4. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
5. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
6. Saint-André, V. et al. Models of human core transcriptional regulatory circuitries. *Genome Res.* **26**, 385–396 (2016).
7. Durbin, A. D. et al. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.* **50**, 1240–1246 (2018).

8. Lin, C. Y. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).

9. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

10. Gryder, B. E. et al. Chemical genomics reveals histone deacetylases are required for core regulatory transcription. *Nat. Commun.* **10**, 3004 (2019).

11. Allfrey, V., Faulkner, R. & Mirsky, A. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl Acad. Sci. USA* **51**, 786–794 (1964).

12. Pogo, B., Allfrey, V. & Mirsky, A. RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proc. Natl Acad. Sci. USA* **55**, 805–812 (1966).

13. Kanno, T. et al. BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat. Struct. Mol. Biol.* **21**, 1047–1057 (2014).

14. Moriniere, J. et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**, 664–668 (2009).

15. Gryder, B. E. et al. PAX3-FOXO1 establishes myogenic super enhancers and confers BET bromodomain vulnerability. *Cancer Discov.* **7**, 884–899 (2017).

16. Stewart, E. et al. Identification of therapeutic targets in rhabdomyosarcoma through integrated genomic, epigenomic, and proteomic analyses. *Cancer Cell* **34**, 411–426.e19 (2018).

17. Meyers, R. M. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).

18. Tenente, I. M. et al. Myogenic regulatory transcription factors regulate growth in rhabdomyosarcoma. *eLife* **6**, e19214 (2017).

19. Schmidt, K., Glaser, G., Wernig, A., Wegner, M. & Rosorius, O. Sox8 is a specific marker for muscle satellite cells and inhibits myogenesis. *J. Biol. Chem.* **278**, 29769–29775 (2003).

20. Böhm, M. et al. Helicase CHD4 is an epigenetic coregulator of PAX3-FOXO1 in alveolar rhabdomyosarcoma. *J. Clin. Invest.* **126**, 4237–4249 (2016).

21. Chu, T. et al. Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat. Genet.* **50**, 1553–1564 (2018).

22. Blumberg, A. et al. Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data. Preprint at *bioRxiv* https://doi.org/10.1101/690644 (2019).

23. Sharova, L. V. et al. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* **16**, 45–58 (2009).

24. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).

25. Greer, C. B. et al. Histone deacetylases positively regulate transcription through the elongation machinery. *Cell Rep.* **13**, 1444–1455 (2015).

26. Sanchez, G. J. et al. Genome-wide dose-dependent inhibition of histone deacetylases studies reveal their roles in enhancer remodeling and suppression of oncogenic super-enhancers. *Nucleic Acids Res.* **46**, 1756–1776 (2018).

27. Orlando, D. A. et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).

28. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

29. Dowen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).

30. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).

31. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).

32. Chory, E. J. et al. Nucleosome turnover regulates histone methylation patterns over the genome. *Mol. Cell* **73**, 61–72.e3 (2019).

33. Stanton, B. Z., Chory, E. J. & Crabtree, G. R. Chemically induced proximity in biology and medicine. *Science* **359**, eaao5902 (2018).

34. Chiarella, A. M. et al. Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-019-0296-7 (2019).

35. Cho, W.-K. et al. RNA Polymerase II cluster dynamics predict mRNA output in living cells. *eLife* **5**, e13617 (2016).

36. Cho, W.-K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

37. Weider, M. & Wegner, M. SoxE factors: Transcriptional regulators of neural differentiation and nervous system development. *Semin. Cell Dev. Biol.* **63**, 35–42 (2017).

38. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

39. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).

40. Boulay, G. et al. Cancer-specific retargeting of BAF complexes by a prion-like domain. *Cell* **171**, 163–178.e19 (2017).

Methods

Cell lines and primary tumors. Cell lines were tested for mycoplasma within one or two passages of each experiment and were found to be negative, and cell line identities were ensured by RNA-seq and genotyping. RH4, RH3, RH5 and RH41 were kind gifts from P. Houghton, SCMC from J. Shipley, RD, CTR and Birch from L. Helman. NIH3T3 and HEK293T cells were purchased from ATCC. Validation was performed by DNA fingerprinting AmpFlSTR Identifiler PCR Amplification Kit (Life Technologies, 4322288). Cell lines were grown at 5% CO_2 and 37 °C in DMEM medium supplemented with 10% fetal bovine serum (FBS) and penicillin and streptomycin. Primary tumors were acquired using the National Cancer Institute (NCI) coordinated ClinOmics protocol as described previously[41].

ChIP-seq and ChIP-Rx. ChIP-seq was performed as previously described[42,43]. In brief, cells or tumor tissue samples were formaldehyde fixed (1% final concentration) for 14 min, Dounce homogenized, pelleted and resuspended in ChIP buffer with protease inhibitors (Active Motif, 53040). Then, samples were sheared for 27 cycles (1 cycle represented by 30 s of sonication and 30 s of rest) with the Active Motif EpiShear Probe Sonicator. Sheared chromatin samples were immunoprecipitated overnight at 4 °C with antibodies (listed in Supplementary Table 6) and were purified by agarose beads (Active Motif ChIP-IT Sigh Sensitivity Kit, 53040). For ChIP-seq with reference exogenous spike-in chromatin, we added *Drosophila* chromatin (Active Motif, 53083) with a *Drosophila*-specific antibody (Active Motif, 61686) to each condition, careful to keep both the amounts of human chromatin and *Drosophila* chromatin constant across drug- or vehicle-treated samples. In each experiment, validation of enrichment was assessed using qPCR with the ChIP-IT qPCR Analysis Kit (Active Motif, 53029) and the primers listed in Supplementary Table 1. ChIP-seq or Chem-seq DNA libraries were made with the Illumina TruSeq ChIP Library Prep Kit, and DNA was size selected with the SPRIselect reagent kit (250–300 bp insert fragment size). Libraries were multiplexed and sequenced using the NextSeq500 High Output Kit v2 (FC-404-2005, 75 cycles) on an Illumina NextSeq500 machine. For each ChIP-seq and Chem-seq sample, 25,000,000–35,000,000 reads were generated.

ATAC-seq library preparation. Assays of transposase-accessible chromatin (ATAC) were performed as previously described[44,45]. In brief, 50,000 cells were isolated, and nuclei were generated by incubating on ice with 500 µl lysis buffer (RSB with 0.1% Tween-20) for 10 min. The resulting nuclei were centrifuged at 500*g* for 10 min and resuspended in 1× Tagment DNA buffer (Illumina) with 2.5 µl Tagment DNA Enzyme (Illumina), and were then incubated at 37 °C for 30 min. For each transposition reaction, the volume was 50 µl. The transposition mixtures were quenched with 500 µl PB buffer (Qiagen) and were purified by standard protocol with a MinElute PCR purification kit. Each ATAC library was amplified with Nextera primers for 16 PCR cycles and was purified with Agencourt AMPure XP (Beckman Coulter) to remove excess primers. The resulting ATAC libraries were sequenced with NextSeq500 with paired-end reads.

Analysis of ChIP-seq and ATAC-seq data. ChIP-seq and ATAC-seq reads were aligned to the human genome version hg19 using BWA version 0.7.17, and were visualized in IGV after being extended to the average fragment size and binned to 25 bp using the count function of igvtools. Samples with spike-in data were additionally normalized to reads per million mapped dm3 reads[27]. Peaks were called using MACS2 (version 2.1.1.20160309, https://github.com/taoliu/MACS) using the 'narrow' mode for all targets reported in this paper, as they form sharp genomic peaks (rather than broad swaths, as is seen for H3K27me3). Parameters for MACS2 usage: [--format BAM --control input.bam --keep-dup all --pvalue 0.0000001]. Regions that were called as peaks but that are known to be spurious mapping artifacts were removed before any further analysis (reference locations for sites black-listed by the ENCODE consortium, https://sites.google.com/site/anshulkundaje/projects/blacklists). Motif analysis was performed on peaks called from MACS2, using findMotifsGenome.pl from HOMER version 4.9.1 (http://homer.ucsd.edu/homer/index.html). SEs were identified with the ROSE2 package (https://github.com/linlabbcm/rose2) by using a stitching parameter of 12,500 bp. In analysis in which we calculated the ChIP-seq read signal in 'core regulatory domains', we defined the genomic coordinates of the domain of each CR TF by stitching together its gene body with all associated SEs within 200 kb.

CR TF network analysis. The degree of inward and outward binding (that is, TF network prediction for CR TFs) was analyzed from SE calls in RMS cell lines and tumors for which H3K27ac ChIP-seq data and paired RNA-seq data were available. The scripts used are available at https://pypi.org/project/coltron and require ROSE2 SE calls for input. In brief, the following computational steps for each sample are performed: a list of active TFs is generated by RNA-seq (with a cutoff of 4 TPM), each SE from ROSE2 is then associated with the nearest expressed TF (with a distance cutoff of 0.5 Mb), then a list of coordinates for each candidate TF is generated. Valleys within SEs are then identified using bamliquidator (https://github.com/BradnerLab/pipeline/wiki/bamliquidator), FASTA sequences for each valley are generated, then TF motifs are identified and network connections are mapped, from which the 'in degree' and 'out degree' values are tabulated for each TF[8].

RNA-seq sample preparation and data analysis. In brief, RNA was extracted from RMS cell lines or tumors, as well as from cell lines treated with drugs, using the RNeasy mini kit (Qiagen). PolyA-selected RNA libraries were prepared and sequenced on an Illumina HiSeq2000. RNA-seq reads were aligned to hg19 using STAR version 2.5.3a. Gene expression values were calculated as TPM or fragments per kilobase of exon per million mapped fragments (FPKM) using RSEM version 1.3.0 and the UCSC hg19 reference at the gene level. For GSEA[46] the TPM values for each protein coding gene were compared to DMSO by a log_2[fold change] comparison, followed by rank-ordering. Bubble plots of enrichment output from GSEA analysis of custom and public gene sets were created in R using custom scripts (https://github.com/GryderArt/VisualizeRNAseq).

ChRO-seq. ChRO-seq was performed on formaldehyde-fixed (same as ChIP-seq) RH4 cells treated with DMSO (6 h), or Entinostat (10 min, 1 h or 6 h). All buffers were made in RNase-free water, and care was taken to maintain a completely RNase-free environment, including frequent use of RNaseZap (Ambion, AM9780) on all surfaces, containers and gloves. Drug treated, fixed and frozen cell pellets were thawed and resuspended in 1 ml of freshly prepared, ice-cold NUN buffer (0.3 M NaCl, 1 M Urea, 1% NP-40, 0.2 mM EDTA, 7.5 mM $MgCl_2$, 20 mM HEPES, 1 mM DTT, 1× PIC (protease inhibitor cocktail), 20 units ml⁻¹ SUPERase•In RNase Inhibitor), vortexed vigorously for 30 s, then 400 µl of additional NUN buffer was added and the sample was vortexed for 30 s. Samples were incubated for 30 min at 4 °C and 1,500 r.p.m. in a thermomixer, then centrifuged for 30 min at 4 °C and 12,500*g* to pellet the chromatin. NUN supernatant was removed, and the pellet was washed 3 times with 50 mM Tris-HCl, centrifuged at 10,000*g* (5 min, 4 °C). Samples were then resuspended in 50 µl cold Storage Buffer (50 mM Tris-HCl, 25% glycerol, 5 mM $MgAc_2$, 0.1 mM EDTA, 5 mM DTT, 40 units ml⁻¹ SUPERase•In) and were sonicated (30 s cycles, 30% amplitude, 5 cycles, kept at 4 °C). Run-on was performed at 37 °C for 5 min after adding 50 µl of run-on master mix (containing 60 µM biotin-11-CTP, 60 µM biotin-11-UTP, 400 µM ATP, 400 µM GTP, 10 mM Tris-HCl, 5 mM Mg, 300 mM KCl, 0.8 units µl⁻¹ SUPERase•In, 1% Sarkosyl and 1 mM DTT), and the reaction was stopped with 500 µl Trizol LS and placed on ice. Adapters with a random hexamer were used as unique molecule indices (UMI). Base hydrolysis, streptavidin bead binding, RNA adapter ligation and reverse transcription steps were performed as previously described[21]. Sequencing was performed in paired-end mode to a depth of approximately 24 million reads per condition. Read clean-up and processing was performed using the Proseq 2.0 pipeline (https://github.com/Danko-Lab/proseq2.0) and visualized in IGV. ChRO-seq sample statistics are available in Supplementary Table 5.

Estimation of mRNA stability. To estimate mRNA stability and half-life, we compared RNA (TPM values calculated as above from total RNA-seq, exons and UTRs only) and nascent RNA-seq from ChRO-seq experiments (nascent TPM, with reads counted for the whole transcript, including exons, introns, UTRs and transcriptional end site regions). We adopted the same assumptions as in a recent report from the Danko and Siepel laboratories[22] (except that we used ChRO-seq data in place of PRO-seq data), namely that the relative decay rates of transcripts are proportional to the ratio of RNA production (nascent RNA, in TPM) divided by the steady-state transcript levels (mRNA). We then applied a first-order half-life equation $T_{1/2}^{CR} = \ln(2)/(\text{mRNA decay rate})$, where CR denotes estimation from ChRO-seq/RNA-seq data. We used matched datasets from RH4 cells treated with DMSO for 6 h to define relative mRNA half-life values for protein coding genes in FP-RMS.

Pooled CRISPR screening. The design of human TF sgRNA pooled libraries was described in a previous study[47]. In brief, annotations of TF DNA binding domain were retrieved from the NCBI Conserved Domains Database. On average, 6 different sgRNAs were designed to target each individual domain of 1,427 TFs in the human genome. All of the sgRNAs with a high predicted off-target effect were then excluded. Specific domain targeting and positive or negative control oligonucleotides were synthesized in duplicate using an array platform (Twist Bioscience) and then were PCR cloned into BsmbI-digested LRG2.1 vector using the Gibson Assembly kit (New England Biolabs (NEB)). Deep-sequencing analysis was then performed on a MiSeq instrument (Illumina) to verify that 100% of the designed sgRNAs were cloned in the LRG2.1 backbone and the quantity of >95% of the sgRNA was within a fivefold measurement of the mean amount.

CRISPR-based negative selection genetic screenings were performed in non-clonal RMS cell lines with stable Cas9 expression (LentiV-Cas9-Puro vector). Lentivirus of pooled sgRNA library for TF library was produced in HEK293T. In brief, HEK293T cells were transfected with pooled TF library sgRNAs and helper packaging plasmid (pVSVG and psPAX2) with polyethylenimine (PEI 25000) transfection reagent to produce lentivirus. HEK293T cells were plated 1d before transfection at 70%–80% confluency in 10 cm tissue culture dishes. For one 10 cm dish of HEK293T cells, 10 µg of plasmid DNA, 5 µg of VSVG, 7.5 µg of psPAX2 and 32 µl of 1 mg ml⁻¹ PEI were mixed, incubated and added to the cells. Medium was changed at 8 h post transfection, and lentivirus-containing supernatant was collected at 24 h, 48 h, 56 h and 72 h post transfection and samples were pooled together.

For lentiviral infection of the TF library, target RMS cells were mixed with the virus and 4 μg ml⁻¹ polybrene, and were then centrifuged at 1,700 r.p.m. for 20 min at 25 °C. Medium was changed at 24 h post infection. Virus titer was measured by infection of the cells with serially diluted virus. To ensure a single copy sgRNA transduction per cell, multiplicity of infection (m.o.i.) was set to approximately 0.3–0.4. To maintain the representation of sgRNAs during the screen, the number of sgRNA-positive cells was kept to at least 1,000 times the number of sgRNAs in the library. Cells were harvested at initial (day 3 post infection) and final (around 12 doubling times after the initial passage) time points. Genomic DNA was extracted using the QIAamp DNA mini kit (QIAGEN) according to the manufacturer's instructions.

Sequencing libraries were constructed as described previously[48]. In brief, sgRNA cassettes (~200 bp) were amplified by PCR from genomic DNA, followed by end-repair with T4 DNA polymerase (NEB, B02025), DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210L) and T4 polynucleotide kinase (NEB M0201L), and the addition of a 3′ A-overhang with Klenow Fragment (3′-5′ exo⁻) (NEB). The DNA fragments then were ligated to diversity-increased custom barcodes with a Quick ligation kit (NEB, M2200L), and were attached to Illumina paired-end sequencing adaptors with Phusion master mix (Thermo Fisher Scientific, F548L). The final libraries were pooled together in equal molar ratios and were sequenced by MiSeq (Illumina) with the MiSeq Reagent Kit v3 (Illumina).

The sequencing data were de-multiplexed and were trimmed to contain only the sgRNA sequence cassettes. The read counts of each individual sgRNA were calculated, with no mismatches to the reference sgRNA sequence, as described previously[48]. The total read counts were normalized to each sample. Individual sgRNAs with a read count lower than 50 at the initial time point were discarded. The average $\log_2$[fold change] of abundance of all sgRNAs against a specific domain or gene was calculated. Data for TF dependencies, paired with RNA-seq expression of TFs in RMS cells, along with Achilles data used to categorize pan-essential TFs, are available in Supplementary Table 4. Follow-up CRISPR experiments used the top two sgRNA sequences (the two that had the greatest effect on cell proliferation among all six sgRNA sequences for a given gene) from the pooled screen, per target, and these lentiviral constructs are available upon request.

shRNA lentivirus production and cell infection. HEK293T cells were transfected with either pLKO.1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 (Sigma Aldrich). Transfection medium was replaced 24 h later and 48 h after transfection the lentiviral containing medium was collected, centrifuged to remove cell debris, and the supernatant was filtered. For lentiviral infection, RH4 cells were transduced at an m.o.i. of 5–10 for 24 h with polybrene (5 μg ml⁻¹; Sigma) and 10% FCS. After a further 24 h cells were selected with 0.75 μg ml⁻¹ puromycin for 2 d.

Small-molecule compounds. All molecules were dissolved in DMSO to a final concentration of 10 mM, and were diluted to a final DMSO concentration of <0.03% by volume in DMEM for cell culture experiments. Entinostat was generously supplied by the Developmental Therapeutics Program (NCI, NIH).

Single-cell RNA-sequencing. RH4 cells were treated with either DMSO for 1 h or with 1 μM Entinostat for 1 h or 6 h. Drop-seq was subsequently performed as previously communicated[49]. In brief, the treated cells were trypsinized, washed, filtered through a 40 μm filter and resuspended in PBS with 0.01% BSA at a concentration of 100 cells μl⁻¹. Barcoded beads (ChemGenes, Macosko-2011-10) were resuspended in complete lysis buffer at a concentration of 120 cells per μl. Cells, beads and oil (Bio-Rad, 186-4006) were loaded into syringes and then pumped through a microfluidic chip (FlowJEM) at a rate of 4 ml h⁻¹ (beads and cells) or 12 ml h⁻¹ (oil) to create nanoliter-sized droplets containing a single bead and/or cell. The outflow of droplets was collected. Droplets were disrupted with perfluorooctanol (Sigma, 370533-25 G) to release the beads, which were then washed several times. Reverse transcription of the mRNAs bound to the capture beads was performed using Maxima H Minus RTase (Thermo Scientific, EP0753) and template switch oligonucleotides (Supplementary Table 3). Single-stranded nucleotides were removed by exonuclease I treatment (Thermo Scientific, EN0581). Beads were counted, and 20 PCR reactions, each containing 8,000 beads, were prepared and performed using 2× Kapa HiFi Hotstart Readymix (Kapa Biosystems, KK2605) and a custom SMART PCR primer. The resulting complementary DNA was purified and concentrated using AMPure XP beads (Beckman Coulter, A63880). The quality of the purified cDNA was assessed using an Agilent TapeStation, and then a DNA library was prepared on 600 pg of the cDNA using a Nextera XT kit (Illumina, FC-131-1024) with the custom New-P5-SMART PCR hybrid oligonucleotide. The DNA library was purified using AMPure XP beads and sequenced using a NextSeq 500/550 High Output v2 kit (75 cycles) (Illumina, FC-404-2005) using a custom Read 1 primer. For each experimental condition, mRNA captured from around 8,000 cells was sequenced on an Illumina NextSeq500.

Single-cell bioinformatic analysis. Sequencing data were processed following the Drop-seq Computational Protocol and using the Drop-seq software tools v.1.2 from the McCarroll lab (http://mccarrolllab.com/wp-content/uploads/2016/03/

Drop-seqAlignmentCookbookv1.2Jan2016.pdf) with the default parameters. Barcodes were extracted and reads were aligned to the hg19 genome. Cells containing more than 500 unique genes were kept for downstream processing. The Seurat version 2.3.0 R package was used for downstream data analysis. The per cent mitochondrial reads were calculated and cells with more than 15% mitochondrial reads were discarded from further analysis. The gene expression measurement per cell was normalized by the total expression and was multiplied by the scaling factor of 10,000 and log transformed. A canonical correlation analysis was run to combine results from all three single-cell runs. The CR TF list used in this analysis was defined as the complete candidate list of SE regulated TFs in RH4 cells without any thresholds for inward or outward binding.

AQuA-HiChIP. For a detailed AQuA HiChIP protocol, please see the associated Protocol Exchange article (https://protocolexchange.researchsquare.com/article/nprot-7121/v1). In brief, NIH3T3 (mouse) cells were grown and fixed in parallel with human cancer cells (RH4) treated with DMSO or Entinostat for 6 h. Two million fixed mouse cells were mixed with seven million fixed RH4 cells, for each condition. Gentle lysis was used to release the nuclei that were then permeabilized in 0.5% SDS (10 min at 62 °C) and then quenched with 10% Triton X-100, followed by digestion with MboI restriction enzyme (2 h at 37 °C) and heat inactivation (20 min at 62 °C). Biotin was incorporated (biotin-14-dATP, Thermo Scientific, 19524-016) with DNA Polymerase I, Large (Klenow) Fragment (NEB, M0210) for 1 h at 37 °C, followed by in situ ligation with T4 DNA ligase (4 h at 25 °C). Nuclei were then pelleted, sonicated (11 min shearing 'on' time with 30 s on, 30 s off, using an Active Motif Epi-shear probe sonicator at 30% power), and immunoprecipitated with cross-species reactive antibody against H3K27ac (Active Motif, 39133). After overnight incubation (rotating at 4 °C), the ChIP reactions were buffer-exchanged into Dynabeads Protein A (ThermoFischer Scientific, 10002D) and incubated (2 h, 4 °C). The ChIP reaction was then washed, eluted and treated with Proteinase K (45 min at 55 °C), and crosslinks were reversed by increasing the temperature to 67 °C for 2 h, followed by purification. Biotin capture and washing was performed with Dynabeads M-280 Streptavidin (ThermoFischer Scientific, 11205D), followed by on-bead library preparation (end-repair, A-tailing, adapter ligation and library amplification)[50]. Amplified library DNA was purified with AMPure beads, quantified and multiplexed, and paired-end sequenced with Illumina NextSeq High Output v2 kit (150 cycles, FC-404-2002) to generate 100–200 million read pairs per condition.

AQuA-HiChIP bioinformatic analysis. A step-by-step bioinformatic protocol for AQuA-HiChIP, with associated code, is available at https://github.com/GryderArt/AQuA-HiChIP. In brief, paired end reads were mapped once to the human genome (hg19), and in parallel were mapped to the mouse genome (mm10), using Bowtie2 within the HiC-pro pipeline[51]. Mapped read pairs were then filtered first for junction validity (containing two fragments from restriction digestion), then duplicates were removed. Contact maps were converted to Juicebox compatible *.hic file format using the hicpro2juicebox.sh tool (https://github.com/nservant/HiC-Pro) provided with HiC-pro, and matrices were extracted using the Juicer[52] dump tool (https://github.com/theaidenlab/juicer/wiki/Data-Extraction). AQuA factor incorporation was done by calculating the contacts per million (CPM; Raw contact frequency matrix × 1,000,000/total valid contacts), then converting to reference normalized contacts per million (RCPM; CPM × hg19 valid contacts/mm10 valid contacts). Virtual 4C analysis was performed by isolating a single row by coordinate of interest (the virtual viewpoint anchor) from a dense RCPM normalized matrix within a region of interest. For aggregate peak analysis (APA) plots, Juicer was used (juicer_tools apa -r 10000,5000 -k NONE -n 0) to generate raw APA data matrices (the sum of contact frequencies for all pairs of SE to SE peaks, both inter-domain and intra-domain, in both DMSO- and HDAC- treated cells) were normalized to RCPM, then plotted in R using pheatmap. For a detailed bioinformatic protocol, see ref. [53].

CEM activation. RH4 cells were transduced with dCas9-HA (with blasticidin resistance) and MS2-FKBPx2 (with hygromycin resistance, selected with 20 μg ml⁻¹ blasticidin and 200 μg ml⁻¹ hygromycin, and expanded for 3–4 passages. Once established, we used RH4-dCas9-MS2-FKBPx2 cells for CEM activation using a guide RNAs with MS2 loops to bind MS2-FKPBx2, and a heterobifunctional small molecule CEM-114 (FK506 linked to a p300 bromodomain binder) to recruit endogenous p300. Guide RNAs were transduced overnight prior to activation with 50 nM CEM-114 for various durations. Constructs for guide RNAs included tandem insertion of each set (four sgRNAs for sgBoundary around the MYOD1 super enhancer boundaries and six sgRNAs for sgEpiCenters targeting MYOD1 SE centers) into a lentiviral backbone (plasmids available upon request).

Super-resolution imaging of RNA Pol II-GFP and BRD4-mEGFP. RH4 cells were stably transduced by lentivirus with RBP3 (RNA Pol II subunit) tagged with GFP (construct of a gift of D. Larson, NIH) and purified by FACS sorting. Endogenous CRISPR–Cas9-mediated knock-in of mEGFP to BRD4 in RH4 cells was carried out by homology directed repair (constructs a gift from Dr. Richard Young, MIT), followed by sorting for positive cells, as described previously[39].

Super-resolution structured-illumination microscopy was performed with a home-built instant structured illumination microscope (iSIM)[54]. The iSIM was equipped with an ×60 oil-immersion objective lens (Olympus, N.A. = 1.42), a stage top incubator (Okolab) and an XYZ automated stage (ASI). A 488 nm laser (Genesis MX-Series, Coherent) was used as the excitation source and the emission was collected by an sCMOS camera (pco.edge 4.2, PCO) after passing through a 488 notch filter (NFD01-488). The microscope stage, lasers, acousto-optic tunable filter and the camera were controlled through custom written Python scripts. Imaging volumes were acquired every 10 min, using a 0.25 μm z spacing. Imaging was performed at 37 °C.

Widefield imaging was conducted on an Olympus IX 81 microscope equipped with an ×60 silicone oil-immersion objective lens (Olympus, N.A. 1.3), a live cell environmental control chamber (INUBG2A-PI Tokai Hit), xy automated stage (ASI) and piezo z-stage (Mad City Labs). A light-emitting diode (LED) light source (Excelitas Technologies, Xcite 110LED) with single band filter (Semrock, FF02-472/30) provided excitation and the emission was collected by the same objective passing through a bandpass filter (Semrock, 520/35) in front of the electron-multiplying charge-coupled device (EMCCD) (Andor Technologies, iXon DU-888). The microscope stage, LED and the camera were controlled through μ-manager (2.0 Beta)[55]. Imaging volumes with 0.5 μm z spacing were collected every 10 min for time-lapse imaging, at multiple xy stage positions. Imaging was performed at 37 °C.

Western blot. Whole-cell lysates were prepared with RIPA buffer (supplemented with protease and phosphatase inhibitors). Cells were lysed by sonication, incubated for 30 min at 4 °C, and then lysates were centrifuged for 15 min. The protein concentration of the resulting supernatant was estimated by BCA assay. Then 40 μg of sample was separated on 4–12% Bis-Tris gels and transferred to PVDF membrane, blocked for 1 h in 5% non-fat milk in Tris buffered saline and Tween-20 (TBS-T). Membranes were incubated at 4 °C overnight with 1:1,000 diluted SOX8 (Abcam, ab104245) primary antibody, washed three times in TBS-T, then incubated in 1:10,000 diluted HRP-labeled secondary antibody (Santa Cruz Biotechnology, sc-2004) at room temperature for 1 h, washed an additional three times with TBS-T and then developed with chemiluminescent reagents (Thermo Fischer Scientific, SuperSignal West Femto).

Statistics and reproducibility. Statistical tests were performed either in GraphPad Prism (GraphPad Software, version 7), R (the Foundation for Statistical Computing, version 3.5.1 from), or GSEA (Broad Institute, version 4.0). For distribution violin box plots and statistical calculations of significance in Fig. 3c, the size (n) of the gene sets were: RH4 TE (typical enhancer) genes (n = 7,446), RH4 SE (super-enhancer) genes (n = 1,591), Housekeeping genes (n = 389), All TF genes (n = 1,427), All RH4 CR (core regulatory) TFs (n = 50) and Top RH4 CR TFs (n = 13). All gene sets are available in Supplementary Table 2. Gene set enrichment P values, NES values and FDR values reported throughout are calculated with 1,000 permutations in the GSEA software, run in pre-ranked mode. CRISPR for human TFs in RMS cells was performed and sequenced as a single experiment per cell line, and was performed across four RMS cell lines (data available in Supplementary Table 4). ChRO-seq experiments were performed at four different time points to increase confidence in the data trends, and were performed once each (Fig. 3e). ChIP-seq experiments were often performed across multiple FP-RMS cell lines or primary tumors rather than performing 'biological' replicates in the same cell line; in some instances we increased confidence in the accuracy of results by performing an orthogonal experimental condition or related ChIP target (for example, we performed ChIP-seq for HDAC1, HDAC2 and HDAC3, rather than HDAC1 in triplicate). Related to Fig. 3, for HDAC1 and HDAC3 we performed ChIP-seq once each, and although these were from different RH4 cell passages these single experiments showed high concordance with the HDAC2 ChIP-seq that was repeated twice in RH4 and once in each of three other FP-RMS samples (RH5, SCMC and a PDX tumor, data not shown) all of which provided very similar ChIP-seq profiles. ChIP-seq of CR TFs MYOG, MYOD1 and SOX8 were repeated twice independently with very similar results; MYCN, YY1 and MED1 were performed once and were bound to expected regions; BRD4 results are representative of six independent ChIP-seq experiment in RH4 cells; and PAX3-FOXO1 ChIP-seq data were representative of four experiments across different RH4 cell passages with similar results. For histone acetylation after DMSO or Entinostat treatment (related to Fig. 4), we saw similar profiles and changes (spreading) among diverse histone lysine modifications (H3K27ac, H3K16ac, H2BK5ac), and these experiments were performed once each from different RH4 cell preparations; orthogonally the H3K27ac HiChIP experiments were qualitatively similar (when flattened) to H3K27ac ChIP-seq results (see Supplementary Fig. 5c). AQuA-HiChIP data are representative of two separate biotin-captures, library preparations and sequencing runs that gave similar results. H3K27ac ChIP-seq profiles in FP-RMS were similar in 38 different independent experiments across 14 different cell lines, tumors and PDX models (15 of these ChIP-seq experiments of H3K27ac were performed from RH4 cells, as independently repeated control experiments for various chemical or genetic perturbations, and all provided very similar profiles).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data reported herein are made publicly available through the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo). The GEO accession number for all ChIP-seq, ChIP-Rx, AQuA-HiChIP and RNA-seq data is GSE116344. Code is available at https://github.com/GryderArt. A list of related resources and reagents used in our study can be found in Supplementary Table 6, which includes a list of all software and data used herein and their sources.

References
41. Chang, W. et al. Multi-dimensional ClinOmics for precision therapy of children and adolescent young adults with relapsed and refractory cancer: a report from the center for cancer research. *Clin. Cancer Res.* **22**, 3810–3820 (2016).
42. Barski, A. et al. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
43. Yohe, M. E. et al. MEK inhibition induces MYOG and remodels super-enhancers in RAS-driven rhabdomyosarcoma. *Sci. Transl. Med.* **10**, eaan4470 (2018).
44. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
45. Hodges, H. C. et al. Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72 (2018).
46. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
47. Tarumoto, Y. et al. LKB1, salt-inducible kinases, and MEF2C are linked dependencies in acute myeloid leukemia. *Mol. Cell* **69**, 1017–1027.e6 (2018).
48. Shi, J. et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
49. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
50. Kidder, B. L. & Zhao, K. Efficient library preparation for next-generation sequencing analysis of genome-wide epigenetic and transcriptional landscapes in embryonic stem cells. *Methods Mol. Biol.* **1150**, 3–20 (2014).
51. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
52. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
53. Stanton, B., Khan, J., Gryder, B. Absolute quantification of architecture (AQuA-HiChIP) enables measurement of differential chromatin interactions. *Protoc. Exch.* https://doi.org/10.1038/protex.2018.130 (2018).
54. York, A. G. et al. Instant super-resolution imaging in live cells and embryos via analog image processing. *Nat. Methods* **10**, 1122–1126 (2013).
55. Edelstein, A. D. et al. Advanced methods of microscope control using μManager software. *J. Biol Methods* **1**, e10 (2014).

Author contributions
B.E.G. and J.K. conceived the project. B.E.G. wrote the manuscript. All authors contributed to the interpretation of data and editing of the manuscript. B.E.G., S.P., C.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0534-4.

Correspondence and requests for materials should be addressed to B.E.G. or J.K.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Berkley E Gryder and Javed Khan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on <u>statistics for biologists</u> may be useful.

Software and code

Policy information about <u>availability of computer code</u>

Data collection	Sequencing data collection was performed with Illumina's NextSeq 550 system. Imaging data collection was done with the MSIM custom Python scripts available here (https://code.google.com/archive/p/msim/source) and detailed in Nature Methods (doi: 10.1038/nmeth.2687).
Data analysis	Software and Algorithms IGV tools (v2.6.2) https://software.broadinstitute.org/software/igv/igvtools Proseq (v2.0) https://github.com/Danko-Lab/proseq2.0 HiC-pro (v2.11.1) https://github.com/nservant/HiC-Pro Juicebox (v1.11.08) https://github.com/aidenlab/Juicebox Juicer (v1.5.6) https://github.com/aidenlab/juicer STAR (v2.5.3a) https://github.com/alexdobin/STAR RSEM (v1.3.2) https://github.com/deweylab/RSEM MSIM package (v1p0) https://code.google.com/archive/p/msim/source R statistical package drc Christian Ritz https://cran.r-project.org/web/packages/drc/drc.pdf Seurat (v2.3.0) Satija Lab https://github.com/satijalab/seurat/ Drop-seq software tools (v.1.2) McCarroll lab http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf) NGSplot (v. 2.63) Shen Lab https://github.com/shenlab-sinai/ngsplot MACS (v 2.1.1.20160309) Zhang et al 2008 https://github.com/taoliu/MACS

BWA (v 0.7.17) Li and Durbin, 2009 http://bio-bwa.sourceforge.net/
HOMER (Hypergeometric Optimization of Motif EnRichment) version 4.9.1 Heinz et al, 2010 http://homer.ucsd.edu/homer/index.html
ImageJ NIH https://imagej.nih.gov/ij/
Graphpad Prism (v7.01) Graphpad Software https://www.graphpad.com/scientific-software/prism/
ROSE2 Charles Lin Lab https://github.com/linlabbcm/rose2
Coltron Lin et al, 2016 https://pypi.org/project/coltron/
Bamliquidator version 1.3.4 John DiMatteo https://github.com/BradnerLab/pipeline/wiki/bamliquidator
GSEA software version 2.2.0 Subramanian et al, 2005 https://software.broadinstitute.org/gsea/
Bedtools version 2.27.1 Quinlan and Hall, 2010 http://bedtools.readthedocs.io/en/latest/
RNA-seq Plotting R Scripts This paper https://github.com/GryderArt/VisualizeRNAseq

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data reported herein is made publicly available through the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The GEO accession number for all ChIP-seq, ChIP-Rx, AQuA-HiChIP and RNA-seq data is GSE116344.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each ChIP-seq and RNA-seq experiment, approximately 30 million uniquely mapped reads were generated. For AQuA-HiChIP, 100 million reads were generated per sample. For single cell RNA-seq experiments, for each condition we sampled 8000 cancer cells. Additional details about group size (n = number of tumor samples used to generate the violin plots for RNA-seq of a gene of interest, for example) are included in the Figure Legends or main text.
Data exclusions	Regions of the genome on the ENCODE blacklist (comprised of, for instance, highly repetitive regions) were excluded from called peaks prior to all downstream analysis (ie, prior to super enhancer calling in the case of H3K27ac ChIP-seq data)
Replication	Replication of super enhancer calling, core regulatory transcription factor identification and network analysis in RMS cells (RH4) was performed by testing 5 additional cell lines (FP-RMS) and for SMS-CTR we validated its circuitry by testing an additional 6 FN-RMS cell lines. To validate the cell lines, we performed ChIP-seq on a total of 8 primary tumor samples from infant or adolescent patients. Replication of top SEs was successful, and showed remarkable consistency between cell lines and primary tumors and fidelity to subtype. Core regulatory factors were further validated by their expression by RNA-seq of FN-RMS cell lines (n = 17), FN-RMS primary tumors (n = 67), FP-RMS cell lines (n = 23), FP-RMS primary tumors (n = 39), exemplified in figure 1 and supplemental figure 1. For ChIP-seq of individual factors, we have only presented one experimental dataset per factor, although in many cases we have performed these experiments in biological duplicate or triplicate, or across multiple cell lines with multiple antibodies (data not presented). For instance, peaks found for SOX8, MYOG, MYOD, PAX3-FOXO1, BRD4, HDAC2, RAD21, CTCF were reproducible across different passages of the same cell line, PAX3-FOXO1 and MYOD were reproducible across multiple cell lines, while some targets were only tested once (HDAC3, YY1, H2BK5ac, H4K16ac). H3K27ac was reproducible across all cell lines and tumors of the disease subtype, and multiple passages of the same cell line (for instance, the most repeated was RH4 H3K27ac ChIP-seq, performed 15 times always with a consistent H3K27ac profile).
Randomization	This work involved no animal studies or randomized clinical trial data. All RMS patients whose samples were obtained were not the subject of randomization at the time of sampling.
Blinding	Blinding was not relevant to our study because the our study only needed diagnosis to perform analysis within and across diagnosis groups, and did not need patient identification (which we are blinded to). Tumor data generated in this study (ChIP-seq) was unlinked to any patient identification. Epigenetic and transcriptomic analysis of primary tumors and cell lines were done with diagnosis knowledge in hand.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All cell lines and reagents described in the methods section are either commercially available or available upon request to the authors. Materials not available to the public include the rare and limited primary tumor samples used to generate data herein.

Antibodies

Antibodies used	Rabbit polyclonal anti-HDAC1, Abcam, Cat# 7028, RRID:AB_305705, Lot# GR225006-9, dilution: 0.05 µg/µL Rabbit polyclonal anti-HDAC2, Abcam, Cat# 7029, RRID:AB_305706, Lot# GR88809-14, dilution: 0.1 µg/µL Rabbit polyclonal anti-HDAC3, Abcam, Cat# ab7030, Lot# GR3202599-2, dilution: 0.05 µg/µL Rabbit polyclonal anti-SOX8, Abcam, Cat# ab104245, RRID:AB_10974591, Lot# GR73485-1, dilution: 0.025 µg/µL Rabbit polyclonal anti-H3K27ac, Active Motif, Cat# 39133, RRID:AB_2561016, Lot# 28518012, dilution: 0.025 µg/µL Rabbit polyclonal anti-Pol2 (N-20), Santa Cruz, Cat# sc-899, RRID:AB_632359, dilution: 0.075 µg/µL Rabbit polyclonal anti-H3K36ac, Active Motif, Cat# 39379, RRID:AB_2614977, Lot# 29108001, dilution: 0.05 µg/µL Anti-H2BK5ac, Active Motif, Cat# 39123, Lot# 01008001, dilution: 0.05 µg/µL Anti-H4K16ac, Millipore, Cat# 07-327, Lot# 3068450, dilution: 0.05 µg/µL Anti-MYOD, Santacruz, Cat# sc-760, Lot# E057, dilution: 0.0125 µg/µL Mouse monoclonal anti-PAX3-FOXO1, In house, PFM2, Lot# M4.8, dilution: 0.1 µg/µL Anti-MYOG, Thermo, Cat# MA5-11486, Lot# RJ2277382, dilution: 0.05 µg/µL Rabbit polyclonal anti-p300 (C-20) X, Santa Cruz, Cat# sc-585 X, Lot# D0615, dilution: 0.04 µg/µL Rabbit polyclonal anti-BRD4, Bethyl, Cat# A301-985A50, Lot# A301-985A100-5, dilution: 0.05 µg/µL Rabbit polyclonal anti-RAD21, Abcam, Cat# ab992, RRID:AB_2176601, Lot# GR3194001-1, dilution: 0.05 µg/µL Rabbit polyclonal anti-YY1, Abcam, Cat# ab109237, RRID:AB_10890662, Lot# GR3179296-8, dilution: 0.04 µg/µL Goat polyclonal anti-rabbit IgG-HRP, Santa Cruz, Cat# sc-2004, RRID:AB_631746, dilution: 1:10,000 Spike-in Antibody (Drosophila specific H2Av), Active Motif, Cat# 61686, Lot# 00419007, dilution: 0.01 µg/µL
Validation	Antibody validations are available on these websites: www.abcam.com, www.scbt.com/scbt and www.activemotif.com In the case of the antibodies for SOX8, which were previously validated by western blot but not validated for ChIP-seq, we evaluated enrichments at predicted SOX binding site in RMS cells by ChIP-qPCR and reviewed the products here: https://www.abcam.com/SOX8-antibody-ab104245/reviews/57941

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human cancer cell lines: RH4, RH41, RH3, RH5 Peter Houghton Lab http://gsbs.uthscsa.edu/faculty/peter-houghton-ph.d SCMC line is from Dr. Janet Shipley, RD, SMS-CTR and Birch from Dr. Lee Helman.
Authentication	Validation was performed by DNA fingerprinting AmpFlSTR® Identifiler® PCR Amplification Kit (Catalog Number 4322288) by Life Technologies. Additional validation was done by RNA-seq comparison to historic cell line RNA-seq data.
Mycoplasma contamination	Cell lines were tested for mycoplasma frequently within one or two passages of each experiment, and were found to be mycoplasma negative.
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used in this study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	All tumors were from de-identified and unlinked tissues banks. Patients were known to be pediatric and the diagnosis of rhabdomyosarcoma were known. Specific age and sex and other information were unknown.

| Recruitment | All tumors were from de-identified and unlinked tissues banks. |

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| Data access links *May remain private before publication.* | To review ChIP-seq and RNA-seq data, use GEO accession GSE116344:
Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116344
Enter token incneucolzutdgj into the box

To review AQuA-HiChIP data, use GEO accession GSE120770:
Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120770
Enter token cpapeueuthsvxgf into the box |
| Files in database submission | Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7_allValidPairs.hic
Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7_allValidPairs.hic
Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7_R1.fastq.gz
Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7_R1.fastq.gz
Sample_RH4_D6_H3K27ac_HiChIP_HKJ22BGX7_R2.fastq.gz
Sample_RH4_Ent6_H3K27ac_HiChIP_HKJ22BGX7_R2.fastq.gz
Sample_RH4_HDAC2_005_C_H7WKVBGXX_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_SOX8_016_C_HJ73HBGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH3_H3K27ac_016_C_HJ73HBGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_bioJQ1_016_C_HJ73HBGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_H3K27ac_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent1_H3K27ac_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_H3K27ac_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_Pol2_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent1_Pol2_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_Pol2_018_C_HWC77BGXY_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH30_H3K27ac_016_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_H3K36ac_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_H3K36ac_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_RAD21_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_RAD21_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_HDAC2_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_HDAC2_021_C_H5JKVBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_HDAC1_024_C_HLFMLBGX3_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_RAD21_025_C_HFH72BGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_RAD21_025_C_HFH72BGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_YY1_025_C_HFH72BGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_YY1_025_C_HFH72BGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_DMSO_6hr_ATAC_H7H5CBGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent_1hr_ATAC_H7H5CBGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent_6hr_ATAC_H7H5CBGX5_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_HDAC2_005_C_H7WKVBGXX_R1.fastq.gz
Sample_RH4_SOX8_016_C_HJ73HBGXY_R1.fastq.gz
Sample_RH3_H3K27ac_016_C_HJ73HBGXY_R1.fastq.gz
Sample_RH3_input_016_C_HJ73HBGXY_R1.fastq.gz
Sample_RH4_D6_bioJQ1_016_C_HJ73HBGXY_R1.fastq.gz
Sample_RH4_D6_H3K27ac_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH4_Ent1_H3K27ac_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH4_Ent6_H3K27ac_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH4_D6_Pol2_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH4_Ent1_Pol2_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH4_Ent6_Pol2_018_C_HWC77BGXY_R1.fastq.gz
Sample_RH30_H3K27ac_016_C_H5JKVBGX3_R1.fastq.gz
Sample_RH30_input_016_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_D6_H3K36ac_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_Ent6_H3K36ac_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_D6_RAD21_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_Ent6_RAD21_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_D6_HDAC2_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_Ent6_HDAC2_021_C_H5JKVBGX3_R1.fastq.gz
Sample_RH4_D6_HDAC1_024_C_HLFMLBGX3_R1.fastq.gz
Sample_RH4_D6_RAD21_025_C_HFH72BGX5_R1.fastq.gz
Sample_RH4_Ent6_RAD21_025_C_HFH72BGX5_R1.fastq.gz
Sample_RH4_D6_YY1_025_C_HFH72BGX5_R1.fastq.gz
Sample_RH4_Ent6_YY1_025_C_HFH72BGX5_R1.fastq.gz
Sample_RH4_DMSO_6hr_ATAC_H7H5CBGX5_R1.fastq.gz |

```
Sample_RH4_Ent_1hr_ATAC_H7H5CBGX5_R1.fastq.gz
Sample_RH4_Ent_6hr_ATAC_H7H5CBGX5_R1.fastq.gz
Sample_RH4_DMSO_6hr_ATAC_H7H5CBGX5_R2.fastq.gz
Sample_RH4_Ent_1hr_ATAC_H7H5CBGX5_R2.fastq.gz
Sample_RH4_Ent_6hr_ATAC_H7H5CBGX5_R2.fastq.gz
RH4_D1_T.gene.TPM.txt
RH4_D24_T.gene.TPM.txt
RH4_D6_T.gene.TPM.txt
RH4_ENT1_T.gene.TPM.txt
RH4_ENT24_T.gene.TPM.txt
RH4_ENT6_T.gene.TPM.txt
Sample_RH4_D1_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_D1_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_D24_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_D24_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_D6_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_D6_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_ENT1_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_ENT1_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_ENT24_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_ENT24_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_ENT6_T_HVNVFBGX2_R1.fastq.gz
Sample_RH4_ENT6_T_HVNVFBGX2_R2.fastq.gz
Sample_RH4_D6_HDAC3_035_C_HMG7VBGX7_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_HDAC3_035_C_HMG7VBGX7_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_PAX3FOXO1_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_PAX3FOXO1_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_MYOD_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_MYOD_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_SOX8_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_SOX8_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_MYOG_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_MYOG_039_C_HL3LVBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_r2_H3K27ac_042_C_HLWGCBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_LW3_H3K27ac_042_C_HLWGCBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Merck60_H3K27ac_042_C_HLWGCBGX9_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_H2BK5ac_043_C_HFKMHBGXB_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_H2BK5ac_043_C_HFKMHBGXB_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_H4K16ac_043_C_HFKMHBGXB_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_Ent6_H4K16ac_043_C_HFKMHBGXB_peaks.narrowPeak.nobl.GREAT.bed
Sample_RH4_D6_HDAC3_035_C_HMG7VBGX7_R1.fastq.gz
Sample_RH4_Ent6_HDAC3_035_C_HMG7VBGX7_R1.fastq.gz
Sample_RH4_D6_PAX3FOXO1_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_Ent6_PAX3FOXO1_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_D6_MYOD_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_Ent6_MYOD_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_D6_SOX8_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_Ent6_SOX8_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_D6_MYOG_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_Ent6_MYOG_039_C_HL3LVBGX9_R1.fastq.gz
Sample_RH4_D6_r2_H3K27ac_042_C_HLWGCBGX9_R1.fastq.gz
Sample_RH4_LW3_H3K27ac_042_C_HLWGCBGX9_R1.fastq.gz
Sample_RH4_Merck60_H3K27ac_042_C_HLWGCBGX9_R1.fastq.gz
Sample_RH4_D6_H2BK5ac_043_C_HFKMHBGXB_R1.fastq.gz
Sample_RH4_Ent6_H2BK5ac_043_C_HFKMHBGXB_R1.fastq.gz
Sample_RH4_D6_H4K16ac_043_C_HFKMHBGXB_R1.fastq.gz
Sample_RH4_Ent6_H4K16ac_043_C_HFKMHBGXB_R1.fastq.gz
```

Genome browser session (e.g. UCSC)

We viewed all datasets in IGV, not a public browser website. TDF files, a compressed version of BEDGRAPHS for IGV viewing, can be made available to reviewers if needed.

Methodology

Replicates

ChIP-seq replicates were performed across multiple cell lines and in most cases multiple ChIP-seq experiments per target in the same cell line. For AQuA-HiChIP experiments, the immunoprecipitation was performed in technical triplicate and pooled before library preparation, and biotin capture and library preparation were performed twice as independent duplicates and achieved near identical ratios of human to mouse contacts across all samples.

Sequencing depth

All ChIP-seq and ChIP-Rx experiments were performed in single-end read mode, 75 base pairs. Spike in reads were measured in parrallel mapping to dm3 and hg19, with results as follows:
ChIP-Rx Sample name Human_Reads
Sample_RH4_D6_BRD4_024_C_HLFMLBGX3 33045619
Sample_RH4_D6_HDAC2_021_C_H5JKVBGX3 23460819

Sample_RH4_D6_p300_024_C_HLFMLBGX3 26822752
Sample_RH4_D6_RAD21_021_C_H5JKVBGX3 16714974
Sample_RH4_D6_RAD21_025_C_HFH72BGX5 51895856
Sample_RH4_D6_YY1_025_C_HFH72BGX5 23373134
Sample_RH4_Ent6_BRD4_024_C_HLFMLBGX3 72847125
Sample_RH4_Ent6_HDAC2_021_C_H5JKVBGX3 24487174
Sample_RH4_Ent6_p300_024_C_HLFMLBGX3 22568713
Sample_RH4_Ent6_RAD21_021_C_H5JKVBGX3 22368744
Sample_RH4_Ent6_RAD21_025_C_HFH72BGX5 37802129
Sample_RH4_Ent6_YY1_025_C_HFH72BGX5 40957167
Sample_RH4_D6_H3K27ac_018_C_HWC77BGXY 29336224
Sample_RH4_D6_Pol2_018_C_HWC77BGXY 45354877
Sample_RH4_Ent1_H3K27ac_018_C_HWC77BGXY 31486542
Sample_RH4_Ent1_Pol2_018_C_HWC77BGXY 33191147
Sample_RH4_Ent6_H3K27ac_018_C_HWC77BGXY 38211649
Sample_RH4_Ent6_Pol2_018_C_HWC77BGXY 35643654
Sample_RH4_D6_H3K36ac_021_C_H5JKVBGX3 29947004
Sample_RH4_Ent6_H3K36ac_021_C_H5JKVBGX3 22926282

ChIP-Rx Sample name SpikeIn_Drosophila_Reads
Sample_RH4_D6_BRD4_024_C_HLFMLBGX3 1072779
Sample_RH4_D6_HDAC2_021_C_H5JKVBGX3 423584
Sample_RH4_D6_p300_024_C_HLFMLBGX3 3786163
Sample_RH4_D6_RAD21_021_C_H5JKVBGX3 949948
Sample_RH4_D6_RAD21_025_C_HFH72BGX5 813980
Sample_RH4_D6_YY1_025_C_HFH72BGX5 659906
Sample_RH4_Ent6_BRD4_024_C_HLFMLBGX3 2023651
Sample_RH4_Ent6_HDAC2_021_C_H5JKVBGX3 606145
Sample_RH4_Ent6_p300_024_C_HLFMLBGX3 4025804
Sample_RH4_Ent6_RAD21_021_C_H5JKVBGX3 1263186
Sample_RH4_Ent6_RAD21_025_C_HFH72BGX5 414133
Sample_RH4_Ent6_YY1_025_C_HFH72BGX5 781349
Sample_RH4_D6_H3K27ac_018_C_HWC77BGXY 650454
Sample_RH4_D6_Pol2_018_C_HWC77BGXY 3661245
Sample_RH4_Ent1_H3K27ac_018_C_HWC77BGXY 663539
Sample_RH4_Ent1_Pol2_018_C_HWC77BGXY 2647315
Sample_RH4_Ent6_H3K27ac_018_C_HWC77BGXY 586269
Sample_RH4_Ent6_Pol2_018_C_HWC77BGXY 3687619
Sample_RH4_D6_H3K36ac_021_C_H5JKVBGX3 694793
Sample_RH4_Ent6_H3K36ac_021_C_H5JKVBGX3 373765

Antibodies

same antibodies listed above.

Peak calling parameters

Peaks were called using MACS2 (version 2.1.1.20160309, https://github.com/taoliu/MACS) using "narrow" mode for all targets reported in this paper, as they form sharp genomic peaks. Parameters for MACS2 usage: [--format BAM --control input.bam --keep-dup all --pvalue 0.0000001]. Regions called as peaks which are known to be spurious mapping artifacts were removed before any further analysis (reference locations for sites black-listed by the ENCODE consortium, https://sites.google.com/site/anshulkundaje/projects/blacklists).

Data quality

Samples are listed with the number of peaks at a threshold of $p < 1\times10^{-7}$ in the 2nd column, and, in the 3rd column, the number of peaks with at least 5-fold or more enrichment are reported.
SampleFile PEAKS_PVALUE7 PEAKS_FOLD_CHANGE5
Sample_RH4_H3K27ac_001_C_H5TLGBGXX 69895 66719
Sample_RH4_HDAC2_005_C_H7WKVBGXX 31422 29444
Sample_SCMC_H3K27ac_008_C_HHC7JBGXX 39084 38125
Sample_RH5_H3K27ac_008_C_HHC7JBGXX 37036 38696
Sample_RD_H3K27ac_008_C_HHC7JBGXX 41078 38039
Sample_RH18_H3K27ac_009_C_HHC7KBGXX 58143 54262
Sample_CTR_D48_H3K27ac_010_C_HMKGLBGXX 38141 36786
Sample_CTR_T48_H3K27ac_010_C_HMKGLBGXX 33382 33273
Sample_RH4_SOX8_016_C_HJ73HBGXY 5190 5069
Sample_RH3_H3K27ac_016_C_HJ73HBGXY 40904 35683
Sample_RH3_P3F_016_C_HJ73HBGXY 4609 4365
Sample_RH4_D6_bioJQ1_016_C_HJ73HBGXY 21905 20612
Sample_RMS559_H3K27ac_016_C_HJ73HBGXY 54360 41203
Sample_RMS238_H3K27ac_007_C_HHF57BGXX 4988 3107
Sample_RMS209_H3K27ac_007_C_HHF57BGXX 31464 33674
Sample_NCI0082_H3K27ac_007_C_HHF57BGXX 29736 28428
Sample_RMS206_H3K27ac_007_C_HHF57BGXX 28789 30773
Sample_RMS216_H3K27ac_007_C_HHF57BGXX 33656 34696
Sample_RMS008_H3K27ac_007_C_HHF57BGXX 32943 31325
Sample_NCI0075_H3K27ac_007_C_HHF57BGXX 23894 25122

```
Sample_RH4_D6_H3K27ac_018_C_HWC77BGXY 62043 58677
Sample_RH4_Ent1_H3K27ac_018_C_HWC77BGXY 63599 60301
Sample_RH4_Ent6_H3K27ac_018_C_HWC77BGXY 71043 64212
Sample_RH4_D6_Pol2_018_C_HWC77BGXY 27014 24297
Sample_RH4_Ent1_Pol2_018_C_HWC77BGXY 21338 19232
Sample_RH4_Ent6_Pol2_018_C_HWC77BGXY 29228 26854
Sample_Birch_H3K27ac_016_C_H5JKVBGX3 27818 27711
Sample_RH30_H3K27ac_016_C_H5JKVBGX3 37404 36761
Sample_RH4_D6_H3K36ac_021_C_H5JKVBGX3 21309 20147
Sample_RH4_Ent6_H3K36ac_021_C_H5JKVBGX3 13744 13319
Sample_RH4_D6_RAD21_021_C_H5JKVBGX3 13350 13351
Sample_RH4_Ent6_RAD21_021_C_H5JKVBGX3 21278 21256
Sample_RH4_D6_HDAC2_021_C_H5JKVBGX3 14590 14266
Sample_RH4_Ent6_HDAC2_021_C_H5JKVBGX3 28588 27817
Sample_RH4_D6_HDAC1_024_C_HLFMLBGX3 75609 72600
Sample_RH4_D6_RAD21_025_C_HFH72BGX5 37633 36978
Sample_RH4_Ent6_RAD21_025_C_HFH72BGX5 39160 38543
Sample_RH4_D6_YY1_025_C_HFH72BGX5 24398 23081
Sample_RH4_Ent6_YY1_025_C_HFH72BGX5 18963 17921
Sample_RH4_DMSO_6hr_ATAC_H7H5CBGX5 15440 13650
Sample_RH4_Ent_1hr_ATAC_H7H5CBGX5 29204 27063
Sample_RH4_Ent_6hr_ATAC_H7H5CBGX5 49207 45234
```

Software

Motif analysis was performed on peaks called from MACS2, using findMotifsGenome.pl from HOMER version 4.9.1 (http://homer.ucsd.edu/homer/index.html). Super enhancers were identified using the ROSE2 package (https://github.com/linlabbcm/rose2) employing stitching parameter of 12500 bp. The scripts used for core regulatory analysis are available here (https://pypi.org/project/coltron/). For peak comparisons, we used Bedtools version 2.27.1 (http://bedtools.readthedocs.io/en/latest/). Plots of ChIP-seq data were made with NGSplot v. 2.63 (https://github.com/shenlab-sinai/ngsplot).