Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma

Berkley E. Gryder1,4, Silvia Pomella1,2, Carly Sayers3, Xiaoli S. Wu4,5, Young Song1, Anna M. Chiarella6, Sukriti Bagchi3, Hsien-Chao Chou1, Ranu S. Sinniah1, Ashley Walton1, Xinyu Wen1,7, Rossella Rota2, Nathaniel A. Hathaway6, Keji Zhao8, Jiji Chen9, Christopher R. Vakoc4, Jack F. Shern3, Benjamin Z. Stanton6,10 and Javed Khan1,11

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 genome1. Some TFs are used across all human cell types (such as the general transcription factors2), whereas many TFs are restricted to a particular time and place in development3–5. 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 SEs6, which form around a mosaic array of CR TF binding sites and drive cell-type-specific gene expression7. CR TFs are themselves driven by a subset of the SEs they form, and can be co-opted as essential dependencies in cancer7,8.

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 SEs9. The entire axis of histone acetylation is essential for CR TF transcription10. Although the need to chemically add or remove 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, 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 (motiff + motiff + motiff and so on) with a motif present in nucleosome-depleted valleys of TFf’s SE. The output for TFf (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 TFf (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
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, acting on its own SE, gray arrows indicate TF and TF, that arise from elsewhere in the genome and act on the SE of TF. Wide gray arrow represents transcription of the TFs 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 x inter-quartile 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 = 1,190 peaks). (c) Prevalence and co-occurrence of CR TFs in SEs in RH4 cells. Functional genetic screening peaks) and SOX-unbound (dotted lines, n = 1,433). (d) Metagene plots at ATAC-seq peaks in SEs near to each CR TF gene locus (Supplementary Fig. 2a); these SEs 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 (Supplementary Fig. 1a). 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. 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 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.

Functional dissection of core regulatory circuitry reveals anti-myogenic 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,
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. 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) knockout (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, are bound by PAX3-FOXO1. Yet, PAX3-FOXO1 co-immunoprecipitates with the enzymes that erase this mark, nuclear HDACs. SEs are associated with high levels of HDAC1 and HDAC2 binding in mouse embryonic stem cells.

To determine whether PAX3-FOXO1 co-binds with HDACs on the epigenome, we performed ChIP-seq on all three nuclear HDACs on 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. 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.

Fig. 3 | Genomic HDAC binding to core regulatory networks at SEs. a, Chromatin immunoprecipitation–sequencing (ChIP-seq) of all three HDACs was performed on SEs (chr2:17,167,162–17,171,675) at core regulatory promoters discovered with ChIP-seq. b, Gene set enrichment analysis shows enrichment at core regulatory TF (CR TF) target genes (log2-fold change). c, Heatmaps depict interaction frequency between genomic elements, assayed by H3K27ac HiChIP (one representative of two experiments in RH4 cells). d, 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). e, 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.
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 hindrance 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\(^\text{21}\) 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\(^\text{21}\), 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\(^\text{23}\)).

As enhancers control transcriptional bursts\(^\text{24}\), 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)\(^\text{22}\) 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\(^\text{17,18}\). 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)\(^\text{25}\). 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\(^\text{28,29}\). Drug-induced hyperacetylation at SEs, which constitutes a new chemical composition of the chromatin template, may alter looping. To investigate 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\(^\text{22,23}\), 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\(^\text{17,21}\) 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 (kilo-base) 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\(^\text{21}\); 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 1 h, followed by a spread beyond the endogenous SE boundary at 6 h. ChIP-Rx with exogenous spike-in is reported as reference-normalized reads per kilobase per million mapped reads (RPKM). 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 (RPPM) differences for diverse acetylation sites on histones are shown for core regulatory domains after treatment with either DMSO or Entinostat for 6 h (left), or for H3K27ac upon treatment with HDACi (right). All experiments were carried out in RH4 cells treated for 6 h with 1 µM of the indicated inhibitors. Box plots show the median and quartiles, and whiskers show the 1.5x 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, (top) and SEn, (bottom). Vertical 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 (SEn, to SEintra, top) or SEs near to but outside of insulated neighborhood CTCF-binding factor (CTCF) boundaries (SE, to SEinter, 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. Thus, we engineered RH4 cells to stably express dCas9 and MS2-FKBP 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). 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. Images are representative of 20 images across 2 independent experiments with similar results. e, Quantification of RNA Pol II (top) and BRD4 (bottom) binding changes upon HDAC1, HDAC2 and HDAC3 inhibition with Entinostat (1 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). Images are representative of 20 images across 2 independent experiments with similar results.
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 Pol2 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 work28, 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 x 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. 15), and these clusters reside in liquid–liquid phase separated droplets at SEs29. 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 HDACi 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 targets27. We discovered and confirmed that the developmental gene SOX8 is critical among CR TFs. SOX8 regulates early neural crest development37, CR TFs MYOD1 and MYOG lead to pro-myogenic differentiation, in RMS as in satelllite 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 circuitry3, as both factors promote the de-differentiated 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 reactions38, and SEs perform this function at the chromatin level27. Furthermore, oncogenic TFs can position phase condensates37. 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.

Received: 3 October 2018; Accepted: 23 October 2019; Published online: 29 November 2019

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 circuits. 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. Garzon, R. et al. Myc amplification and expression in human tumors. N. Engl. J. Med. 359, 161–173 (2008).
9. Nikolic, M. et al. MYCN amplification in neuroblastoma: clinical implications and potential therapeutic targets. Neuro-Oncol. 12, 121–136 (2010).
10. Pitteri, S. et al. Role of the MYC family in cancer. Adv. Cancer Res. 116, 277–313 (2013).
8. Lin, C. Y. et al. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57–62 (2016).
9. Hinna, 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. Rong, S. 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. Sox5 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. Bromodomain-condensed gene loci: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
39. Rabani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Bromodomain-condensed gene loci: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
40. Boulay, G. et al. Cancer-specific retargeting of BAF complexes by a prion-like domain. *Cell* **171**, 163–178.e19 (2017).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
ChIP-seq and ChIP-Rx. ChIP-seq was performed as previously described42,43.

ATAC-seq library preparation. Assays of transposase-accessible chromatin (ATAC) were performed as previously described29,30. In brief, 50,000 cells were isolated, and nuclei were generated by incubating on ice with 500 μl of lysis buffer (0.3 M NaCl, 1 M urea, 1% NP-40, 0.2 mM EDTA, 7.5 mM MgCl₂, 20 mM HEPES, 1 mM DTT) and nuclei were generated by incubating on ice with 500 μl of run-on master mix (containing 60 μM ATP, 400 μM GTP, 10 mM Tris-HCl, 5 mM MgCl₂, 300 mM KCl, 0.8 μl units⁻¹ SUPERase•In, 1% Sarkosyl and 1 mM DTT), and the reaction was stopped with 500 μl Trizol LS and placed on ice. With adapters a random hexamer were used as unique molecule indices (UMI).  

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 sequenced on an Illumina HiSeq 2500 RNA-seq was 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 GSEA46 the TPM values for each protein coding gene were compared to DMSO by a log([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).

CRISPR screening. The design of human TF sgRNA pooled libraries 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 each sgRNA was then performed on a Qubit fluorometer (ThermoFisher) to determine the amount of each sgRNA as quantified after the cloning step.

Pooled CRISPR screening. The design of human TF sgRNA pooled libraries 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 each sgRNA was then performed on a Qubit fluorometer (ThermoFisher) to determine the amount of each sgRNA as quantified after the cloning step.

CRISPR-based negative selection genetic screenings were performed in non-clonal RMS cell lines with stable Cas9 expression (LentV-Cas9 Puro vector). Lentivirus of pooled sgRNA library for TF library was produced with HEK293T. In brief, HEK293T cells were transfected with pooled TF library sgRNAs and helper packaging plasmid (pVSVG and pSPAX2) with polyethyleneimine (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 each 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 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 single round of infection per cell, multiplicity of infection (m.o.i.) was set to approximately 0.3–0.4. To maintain the representation of sRNAs during the screen, the number of sRNA-positive cells was kept to at least 1,000 times the number of sRNAs 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 previously49, 13. In brief, sRNA 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, F544L). 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 sRNA sequence cassettes. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. 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 sRNA sequence cassettes. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA sequences for a given gene) from the pooled screen, per target, and these lentiviral constructs are available upon request.

The final libraries were pooled together in equal molar ratios and were sequenced by MiSeq (Illumina) with the MiSeq Reagent Kit v3 (Illumina).

**shRNA lentivirus production and cell infection.** HEK293T cells were transfected with either pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA sequences for a given gene) from the pooled screen, per target, and these lentiviral constructs are available upon request.

The final libraries were pooled together in equal molar ratios and were sequenced by MiSeq (Illumina) with the MiSeq Reagent Kit v3 (Illumina).

**shRNA lentivirus production and cell infection.** HEK293T cells were transfected with either pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA 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 pKO1-puro shRNA Scramble or specific shRNA (MYOD1, MYOG, MYCN, SOX8) and helper packaging plasmid (pVSVG and psPAX2) with Fugene 6 transfection reagent. The read counts of each individual sRNA were calculated, with no mismatches to the reference sRNA sequence, as described previously49, 13. The total read counts were normalized with respect to cell proliferation among all six sRNA sequences for a given gene) from the pooled screen, per target, and these lentiviral constructs are available upon request.
Super-resolution structured-illumination microscopy was performed with a home-built instant structured illumination microscope (iSIM)54. The iSIM was equipped with an x60 oil-immersion objective lens (Olympus, N.A. = 1.42), a stage-top incubator (Okalab) and an XYY automated stage (ASI). A 408 nm laser (Genesio 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 nm notch filter (NFDF01-488). 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 x60 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 (Osilites Technologies, Xcite 110LED) and single band filters (Edgewood, FF04-472/30) 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 μManager software. 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 and 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 HRP-labeled antibodies (Abcam, housekeeping, HDACs). The membranes were 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-computed 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). ChRQ-seq experiments were performed at four different time points to increase confidence in the data trends, and were performed once each (Fig. 3c). ChRQ-seq experiments were often performed across multiple FP-RMS cell lines or primary tumors in order to prevent ‘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. 3f, 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, MYOD1 and MED1 were repeated once and were bound to expected regions; BRD4 results are representative of six independent ChIP-seq experiments 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 divergent histone lysine modifications (H3K27ac, H3K16ac, and H2BK5ac), and these experiments were performed once each from different RH4 cell preparations; orthogonally the H3K27ac H3ChIP experiments were qualitatively similar (when flattened) to H3K27ac ChIP-seq results. AQUA H3ChIP data are representative of two separate bioin-capture, library preparations and sequencing runs that gave similar results. Related to Fig. 5, the AQUA 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).

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.

References

41. Chang, W. et al. Multi-dimensional ClinOomics 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 Methylation in the Human Genome. Cell 129, 823–837 (2007).

43. Yohe, M. E. et al. MEK induction induces MYO9 and remodels super-enhancers in RAS-driven rhabdomyosarcoma. Sci. Transl. Med. 10, eaan4470 (2018).

44. Buenrostro, J. D., Greissi, 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. Huygens, H. C. et al. Domains-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, sial-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 nanoplot 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., Gnyder, B. Absolute quantification of architecture and processing. Cell Syst. 3, 15545–15550 (2005).

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 microscale control using μManager software. J. Biol Methods 1, e10 (2014).

Acknowledgements

We would like to thank M. Yuhe, D. Levins, J. Klein, W. Kuehl, B. Hawley, C. Greer and E. Chory for thoughtful discussions regarding experiments and the manuscript. We are grateful to C. Danko and S. P. Chou for advice, protocols and code involved with performing ChRQ-seq. We are indebted to J. Qi and L. Wu for providing selective HDAC inhibitors LW3 and Merck60. We also thank P. Brown of the Structural Genomics Consortium for providing their set of other epigenetic molecules. We thank the CCR Genomics Core at the National Cancer Institute, NIH, Bethesda, MD, USA, for sequencing the single-cell RNA. This work was supported by the NCI, NIH. S.P. is a recipient of a Fondazione Veronesi fellowship. R. Rota was supported by Associazione Italiana Ricerca sul Cancro (AIRC 15312) and Italian Ministry of Health (PE-2013-02355271). N. Hathaway is supported in part by grant R01GM118653 from the US National Institutes of Health. C. R. Yako and X. S. Wu were supported by the Pershing Square Sohn Cancer Research Alliance, Northwell Health Translational Research Award, the Christina Renna Foundation, The Clark Gilles Foundation, the Friends of T.J. Foundation and the Michelle Paternoster Foundation for Sarcoma Research. B.Z.S. is supported as a St. Baldrick’s Foundation Scholar. Finally, we are indebted to H. Vishwasrao and H. Shroff for contributing their expertise on high-resolution imaging. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services. Mention of trade names, commercial products, or organizations does not indicate endorsement by the US Government.

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.
and Y.S. performed ChIP-seq and RNA-seq experiments and generated data. B.E.G. and B.Z.S. conceived and performed AQuA-HiChIP experiments. B.E.G., A.W., X.W. and H.-C.C. wrote scripts and pipelines for bioinformatic analysis. B.E.G. and J.C. designed and performed imaging experiments. B.E.G., S.B., R.S.S. and A.M.C. performed dCas9-based recruitment experiments under supervision of N.A.H. J.F.S., B.Z.S., K.Z., C.R.V. and J.K. supervised the work and mentored the first author. S.P. performed western blot and shRNA experiments under supervision of R.R. X.S.W. designed and performed domain focused CRISPR screening under supervision of C.R.V. B.E.G. and J.K. made final edits to the manuscript.

**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 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 Authors & Referees and the Editorial Policy Checklist.

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).

- The exact sample size (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 central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (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)

Software and code

Policy information about availability of computer code

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

- IGV tools (v2.6.2) https://software.broadinstitute.org/software/igv/igvtools
- Proseq (v2.0) https://github.com/Danko-Lab/proseq2.0
- HiC-pro (v2.11.3) https://github.com/raeservant/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
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 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 unlabeled 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

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/VisualizerRNAseq

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.
**Materials & experimental systems**

| Involved in the study |
|------------------------|
| Unique biological materials |
| Antibodies |
| Eukaryotic cell lines |
| Palaeontology |
| Animals and other organisms |
| Human research participants |

**Methods**

| Involved in the study |
|------------------------|
| ChIP-seq |
| Flow cytometry |
| MRI-based neuroimaging |

**Unique biological 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# GR7348S-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# R2277382, dilution: 0.05 μg/μL
  - Rabbit polyclonal anti-p300 (C-20), 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_10809662, 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**

**Cell line source(s)**

Human cancer cell lines: RH4, RH41, RH3, RHS 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® Identifier® 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**

**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

tables containing file names and links
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 parallel 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
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 < 1x10^-7 in the 2nd column, and, in the 3rd column, the number of peaks with at least 5-fold or more enrichment are reported.
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).