SUPPLEMENTARY METHODS

Data used

Gene expression data in 5 DLBCL cells including LY1, LY4, HBL1, TOLEDO and DHL6 under DMSO (24h) and JQ1 (24h) treatment conditions were obtained from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) accession code GSE45630 (Chapuy et al. 2013); malignant peripheral nerve sheath tumor cells 90-8TL: GSE62500 (De Raedt et al. 2014); and DMSO and JQ1 (4h) treatment data in HepG2: GSE51143 (Picaud et al. 2013). AR, ESR1, PPARG, NOTCH, and POU5F1 (also known as OCT4) ChIP-seq data were retrieved from GSM980662, GSM470419, GSM534493, GSM1252933, and GSM803438, respectively. NR3C1 (GR) ChIP-seq data from A549 cells were obtained from the NCBI Short Read Archive with accession number of SRA008630. Gene expression data in LNCaP under untreated and DHT (6h) conditions were obtained from GSE7868 (Wang et al. 2007), in MCF7 cells under untreated and E2 (12h) conditions from GSE11324 (Carroll et al. 2006). In adipose stromal cell, we used the gene expression at different time points across the differentiation process, pre-adipocytes at day -2 and day 14 (GSE20697) (Mikkelsen et al. 2010). We analyzed NOTCH regulated gene expression using CUTLL cells under NOTCH off (GSId3) and NOTCH on (GSI washout 4h) conditions (GSE29544) (Wang et al. 2011). For GR regulated gene expression analysis, non-treated and 6h dexamethasone treatment gene expression were used (GSE17307) (Muzikar et al. 2009). In the analysis of OCT4 regulated gene expression, we compared triplicates of 3 days knockdown of OCT4 with shRNA targeting luciferase (GSE21200) (Kunarso et al. 2010). For the diverse gene expression responses analyses, we downloaded human chemical and genetic perturbations gene sets from MSigDB (Liberzon et al. 2011) (www.broadinstitute.com/gseq/msigdb), filtering out the gene sets with less than 100 genes or all genes located in single chromosome, 671 gene sets remained. H3K27ac ChIP-seq datasets used in the MARGE analysis are summarized in Supplemental Tables 1 and 2.

ChIP-seq data analyses pipeline

ChIP-seq data sets were aligned to the human genome (GRCh38/hg38) or mouse genome (GRCm38/mm10) using BWA (Li and Durbin 2009) (Version 0.7.10) with the following command: bwa aln -q 5 -l 32 -k 2 -t 8 {*index*} {*fastq*}, bwa samse {*index*}{*in.sai*}{*fastq*}. We used the peak calling algorithm MACS2 (Zhang et al. 2008) (github.com/taoliu/MACS) to identify the ChIP-enriched regions with the following command: macs2 callpeak –t {*treat*} –c {*control*} --SPMR -B –q 0.01 --keep-dup 1 --nomodel –g hs(mm), --extsize=146 for H3K27ac and TF ChIP-seq, and extsize=100 for DNase-seq. The parameter --SPMR was used to normal each dataset by the sequencing depth and to generate bedGraph signal tracks.

Microarray Data analyses

Affymetrix microarray gene expression data were normalized using the standard multichip average (RMA) package in R (Irizarry et al. 2003), differential expression analysis were performed with linear model for microarray (LIMMA) (Smyth 2004). In the analysis of BET-inhibitor repressed genes the down-regulated gene sets for DLBCL cell lines were defined using the cutoffs: fold-change ≤ 0.5 and FDR ≤ 0.01. For the HepG2 and 90-8TL cases the differential gene expression signal was weaker so it was necessary to use less stringent cut-offs to carry out any analysis. In the case of HepG2 the fold-change cutoff was ≤ 0.66 and no FDR cutoff was used. In the case of 90-8TL the fold-change cutoff was ≤ 0.66 and the FDR cutoff was ≤ 0.05.

H3K27ac ChIP-seq signal tracks

We converted MACS2 generated bedGraph files to more compressed bigWig files using the UCSC bedGraphToBigWig script (http://hgdownload.cse.ucsc.edu/admin/exe/). We extracted the sequencing depth normalized H3K27ac ChIP-seq signal from bigwig files using the UCSC BigWigSummary script.

Super-enhancer and super-enhancer associated gene detection using ROSE
We downloaded the ROSE (Whyte et al. 2013; Lovén et al. 2013) software (*https://bitbucket.org/young_computation/rose.git*), and detected super-enhancers and super-enhancer related genes using ROSE with default parameters. Super-enhancer associated genes were ranked by the strength of super-enhancers. Genes regulated by the same super-enhancer were assigned an equal rank. The highest ranked super-enhancer was assigned to any gene associated with multiple super-enhancers.

Receiver operating characteristic curve (ROC) and AUC values
We used the R package ROCR to draw the ROC curves and calculate the AUC values. In Figure 1 and Supplemental Figure 3, JQ1 down-target genes (FDR ≤ 0.01, fold- change ≤ 0.5) were set as the positives, and the others as negatives. For our regulatory potential and relative regulatory potential method, we ranked genes by p or p^* from high to low. Since ROSE determines only several hundred super-enhancer related genes, when drawing the ROC and calculating the AUC value for ROSE, typical enhancer related genes were included as well. We ranked genes associated with typical enhancers in the same way as super-enhancer genes, so that all information from ROSE output was considered. In the promoter based analysis of H3K27ac we calculated the H3K27ac normalized read count in regions 1kb up and downstream of the TSS.

High and low CpG genes
We downloaded the file with CpG island information from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/cpgIslandExt.txt.gz), and defined the high CpG island regions with observed-to-expected CpG ratio (last column in the file) larger than 0.85 as the high CpG regions, and the rest as low CpG regions. Genes with high CpG islands in their promoters (1kb up/down-stream of gene TSS) are classified as the high CpG genes, and genes with low CpG islands or without CpG islands in their promoters are classified as the low CpG genes. Finally, We got 13247 high CpG genes and 15361 low CpG genes.

Keratinocyte differentiation lncRNA analysis.
The late differentiation gene set was obtained from Supplementary Table 1 of Lopez-pajares et al. 2015. This gene set, based on DNA microarray analysis, contains protein-coding genes that gradually increased over a time course of keratinocyte differentiation, peaking at days 5–7. Using MARGE-express we trained a model to predict gene expression in this system. Another study by the same group carried out a RNA-seq analysis of a similar system of progenitor and differentiating keratinocytes. This study reported, in their Supplementary Table 1, a list of 258 of annotated non-coding RNAs that changed over the differentiation time course (Kretz et al. 2013). This table reports FPKM values for days 0, 3 and 6 of the differentiation time course. We used the following criterion for an increasing signature: FPKM day 6 / day 0 > 2 and the expression level increases monotonically, day 6 > day 3 > day 0. Similarly we used the following criterion for a decreasing signature: FPKM day 0 / day 6 > 2 and the expression level decreases monotonically, day 6 < day 3 < day 0. P-values comparing scores between groups where calculated using the Wilcoxon rank sum test.

Union of DNase I Hypersensitive Sites (UDHS)
We collected 458 DNase-seq human and 116 mouse datasets from the GEO database and processed them with the ChIP-seq data analysis pipeline. All DNase-seq reads were extended to 100 bp fragments from their 5' ends and peaks were identified using MACS2 for each DNase-seq dataset under an FDR cutoff of 0.01 and fold enrichment cutoff of 4. Then DHS peaks from all DNase-seq datesets were put together, with each peak trimmed to 50 bp centered at the peak summit. Overlapping peaks were merged as one peak. In this way we obtained 2,723,010 union DNase-seq peaks in human (hg38) and 1,529,448 in mouse (mm10).

Clustering and heatmaps

For each H3K27ac ChIP-seq sample, we calculated each RefSeq gene's regulatory potential *p,* and associated each RefSeq identifier with a gene symbol. For genes that have multiple-TSSs, we used the one with the highest median of regulatory potentials across all 365 samples. We made the union gene set (15136 genes) containing the top 1000 genes with highest regulatory potentials for each sample. We computed the square root of *p*, and selected the 2000 genes with the highest coefficients of variation to use in the heatmap. The square root *p* matrix was normalized within each sample by median scaling. We clustered genes using *k*-means on the normalized square root of *p* across 365 human H3K27ac samples, with k=10. The samples are clustered using hierarchical clustering with euclidean distance and average linkage.

MSigDB Analysis

Gene sets of at least 100 genes that include genes on more than one chromosome were selected from the chemical and genetic perturbation sets of the molecular signatures database http://software.broadinstitute.org/gsea/msigdb. Regression was carried out using MARGE-express. Gene sets were designated breast if their brief descriptions contain either of the words breast or mammary. Similarly blood from blood, haematopoietic, leukemia, lymphoma, lymphocyte or myeloma; liver from liver, hepatocellular, or hepatoblastoma; neuron from neuroblastoma, cortex, glioma or glioblastoma; prostate from prostate; colon from colon or colorectal; lung from lung.

Gene Ontology Analyses

In the gene ontology (GO) analysis, we used a subset of 26 data sets that were analyzed in the Hnisz super-enhancer study (Hnisz et al. 2013). For each dataset (tissue), ROSE super-enhancer associated genes were analyzed using DAVID (david.ncifcrf.gov) (Huang et al. 2007). To make a fair comparison of GO analyses of ROSE and MARGE predictions, we selected the same number of genes based on the relative regulatory potential as were associated with super-enhancers by ROSE. We then compared the GO categories defined by Hnisz et al in terms of the enrichment significance of ROSE with the relative regulatory potential. In addition, we carried out a GO analysis based on the 500 genes in each of the 26 data sets with the greatest relative regulatory potentials and compared the enrichment of the most significant 3 categories with a GO analysis of the top 500 super-enhancer genes in the same data sets.

IGV browser tracks

IGV Browser tracks (Supp. Fig. 12) indicate the MACS2 normalized H3K27ac ChIP-seq signal. The bottom track in Supp. Fig. 12 depicts the MARGE-cistrome reweighted H3K27ac signals based on the 10 selected H3K27ac samples. The combined signal of location *i* was calculated via the formula $\sum_j \lambda_j^* c_{kj}$ where c_{kj} indicates the signal of region k in *j* sample, and λ^* is the optimal centroid (details about the centroids will be described in MARGE pipeline).

Cell culture

LNCaP-abl cells were maintained in phenol red-free RPMI 1640 medium with 10% charcoal/dextran-treated FBS.

RNA interference and RNA-seq

The siRNA smart pool oligos targeted *AR*, *E2F1*, *FOXA1*, *FOXM1*, *MALAT1*, *EZH2*, *KDM1A*, *UTX* and *RAD21*, were purchased from Dharmacon, Cells were transfected with 20nM siRNA oligos by RNAiMax reagent (Invitrogen) in 12 well plates according to manufacturer's instructions. The total RNAs were isolated by TRIzol (Invitrogen), followed by library construction using the TruSeq RNA Library Prep Kit (Illumina) for Illumina HiSeq.

RNA-seq data analyses

RNA-seq data were processed using CuffLinks (Version 2.0.2) (Trapnell et al. 2012) with default parameters. RPKM levels on all annotated genes with official gene symbols were obtained from

the Cufflinks output for each sample. The expression data matrix was converted to $\log_2$ RPKM scale and normalized by quantile normalization. For each gene in each siRNA sample, a z-score was calculated using the mean and standard deviation of this gene in the 3 siControl samples and adjusted using a hierarchical model (Ji and Liu 2010). Down-regulated genes with adjusted z-score ≤-2 were used as input gene gets for MARGE analyses.

ChIP-seq experiments

ChIP experiments for FOXA1 and H3K27ac in LNCaP-abl cells were performed as previously described (He et al. 2010), and the antibody for FOXA1 and H3K27ac were ab5089 (Abcam) and ab4729 (Abcam) respectively. Library construction was performed using the ChIP-seq DNA sample Prep Kit (Illumina) according to the manufacture's instruction; followed by high-througput sequencing with Illumina HiSeq.

References

Carroll JS, Meyer C a, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al. 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**: 1289–97.

Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MGM, Qi J, Rahl PB, Sun HH, Yeda KT, Doench JG, et al. 2013. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**: 777–90.

De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, Helin K, Hornick JL, Mautner V, Kehrer-Sawatzki H, et al. 2014. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. *Nature* **514**: 247–251.

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova A a, Hoke H a, Young R a. 2013. Super-enhancers in the control of cell identity and disease. *Cell* **155**: 934–47.

Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, et al. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**: W169–75.

Irizarry RA, Hobbs B, Beazer-barclay YD, Antonellis KJ, Scherf UWE, Speed TP. 2003. Exploration , normalization , and summaries of high density oligonucleotide array probe level data. 249–264.

Ji H, Liu XS. 2010. Analyzing 'omics data using hierarchical models. *Nat Biotechnol* **28**: 337–40.

Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, et al. 2013. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**: 231–5.

Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42**: 631–4.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–60.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**: 1739–40.

Lopez-pajares V, Qu K, Kretz M, Khavari PA, Zarnegar BJ, Boxer LD, Rios EJ, Tao S, Kretz M, Khavari PA. 2015. Article A LncRNA-MAF : MAFB Transcription Factor Network Article A LncRNA-MAF : MAFB Transcription Factor Network Regulates Epidermal Differentiation. *Dev Cell* **32**: 693–706.

Lovén J, Hoke H a, Lin CY, Lau A, Orlando D a, Vakoc CR, Bradner JE, Lee TI, Young R a. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**: 320–34.

Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, Rosen ED. 2010. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**: 156–69.

Muzikar KA, Nickols NG, Dervan PB. 2009. Repression of DNA-binding dependent glucocorticoid receptor-mediated gene expression. **2009**.

Picaud S, Wells C, Felletar I, Brotherton D, Martin S, Savitsky P. 2013. RVX-208 , an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 1–13.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–78.

Wang H, Zou J, Zhao B, Johannsen E, Ashworth T, Wong H, Pear WS, Schug J, Blacklow SC, Arnett KL, et al. 2011. Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells. *Proc Natl Acad Sci U S A* **108**: 14908–13.

Wang Q, Li W, Liu XS, Carroll JS, Jänne O a, Keeton EK, Chinnaiyan AM, Pienta KJ, Brown M. 2007. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* **27**: 380–92.

Whyte W a, Orlando D a, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young R a. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307–19.

Zhang Y, Liu T, Meyer C a, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.