Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood

Julia Joung1,2,3,4, Jesse M. Engreitz2, Silvana Konermann2,3,4,†, Omar O. Abudayyeh2,3,4,5, Vanessa K. Verdine2,3,4,†, Francois Aguet2,†, Jonathan S. Gootenberg2,3,4,5, Neville E. Sanjana2,3,4, Jason B. Wright2,3,4, Charles P. Fulco5,6, Yuen-Yi Tseng7, Charles H. Yoon7, Jesse S. Boehm2, Eric S. Lander2,6,8, & Feng Zhang1,2,3,4

Mammalian genomes contain thousands of loci that transcribe long noncoding RNAs (lncRNAs)1–3, some of which are known to carry out critical roles in diverse cellular processes through a variety of mechanisms4–8. Although some lncRNA loci encode RNAs that act non-locally (in trans)3, there is emerging evidence that many lncRNA loci act locally (in cis) to regulate the expression of nearby genes—for example, through functions of the lncRNA promoter, transcription, or transcript itself4–8. Despite their potentially important roles, it remains challenging to identify functional lncRNA loci and distinguish among these and other mechanisms.

Here, to address these challenges, we developed a genome-scale CRISPR-Cas9 activation screen that targets more than 10,000 lncRNA transcriptional start sites to identify noncoding loci that influence a phenotype of interest. We found 11 lncRNA loci that, upon recruitment of an activator, mediate resistance to BRAF inhibitors in human melanoma cells. Most candidate loci appear to regulate nearby genes. Detailed analysis of one candidate, termed EMICERI, revealed that its transcriptional activation resulted in dosage-dependent activation of four neighbouring protein-coding genes, one of which confers the resistance phenotype. Our screening and characterization approach provides a CRISPR toolkit with which to systematically discover the functions of noncoding loci and elucidate their diverse roles in gene regulation and cellular function.

We have previously used the Cas9 synergistic activation mediator (SAM) to screen for protein-coding genes that confer resistance to the BRAF inhibitor vemurafenib in melanoma cells2, making this an ideal prototype for screening lncRNA loci (Supplementary Note 1). We designed a genome-scale single guide RNA (sgRNA) library that targeted 10,504 intergenic lncRNA transcriptional start sites (TSSs)2,10 (see Methods, Supplementary Table 1). We transduced A375 (BRAF(V600E)) human melanoma cells with the sgRNA library, cultured them in 2 μM vemurafenib or vehicle control for 14 days, and sequenced the distribution of sgRNAs (Fig. 1a, b and Extended Data Fig. 1a). RNAi gene enrichment ranking (RIGER) analysis3 identified 16 candidate loci that were significantly enriched (FDR < 0.05) in cells cultured with vemurafenib (Extended Data Fig. 1b–e and Supplementary Table 2), none of which had previously been functionally characterized.

Figure 1 | Genome-scale activation screen identifies lncRNA loci involved in vemurafenib resistance. a, A375 cells expressing SAM effectors were transduced with the pooled sgRNA library targeting more than 10,000 lncRNA TSSs and treated with vemurafenib or control for 14 days before deep sequencing. b, Box plot showing sgRNA frequencies after vemurafenib or control treatment from n = 4 infection replicates. c, Vemurafenib resistance and transcriptional activation in A375 cells upon SAM activation of candidate lncRNA loci. NT, non-targeting. All values are mean ± s.e.m. with n = 4. ****P < 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05.

1Department of Biological Engineering, MIT, Cambridge, Massachusetts 02139, USA. 2Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. 3McGovern Institute for Brain Research at MIT, Cambridge, Massachusetts 02139, USA. 4Department of Brain and Cognitive Science, MIT, Cambridge, Massachusetts 02139, USA. 5Division of Health Sciences and Technology, MIT, Cambridge, Massachusetts 02139, USA. 6Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. 7Department of Surgery, Brigham and Women’s Hospital, Boston, Massachusetts 02115, USA. 8Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. †Present addresses: Salk Institute for Biological Studies, La Jolla, California, USA (S.K.); New York Genome Center, New York, New York, USA (N.E.S.); Department of Biology, New York University, New York, New York, USA (N.E.S.).
To validate the screening results, we individually expressed the three most enriched sgRNAs targeting each of the top 16 candidate IncRNA loci in A375 cells. In all 16 cases, the sgRNAs conferred significant vemurafenib resistance ($P < 0.01$) (Extended Data Fig. 2), verifying the robustness of our screening approach. We performed RNA sequencing upon activation of each of the 11 loci that showed the strongest effects (Extended Data Fig. 2, Supplementary Table 3) and found global changes in gene expression consistent with vemurafenib resistance, supporting the functional relevance of these loci to the screening phenotype (Extended Data Fig. 3a).

Next, we turned to classifying the mechanisms by which activation of each of these loci might lead to resistance. These mechanisms could include (i) a non-local (trans) function of the IncRNA transcript\(^1\), (ii) a local (cis) function of the IncRNA transcript or its transcription in affecting the expression of a nearby gene\(^3,6–8\), (iii) a local function of the IncRNA promoter acting as a DNA element (enhancer)\(^2,9\); and, in theory, (iv) a local function of SAM (Supplementary Note 2). To study the first two of these mechanisms, which require the IncRNA or its transcription, we focused on the 6 of 11 loci at which SAM targeting led to robust upregulation of the IncRNA transcript (Fig. 1c, Supplementary Table 3). The remaining five loci may function through a mechanism other than activation of the IncRNA transcript (for example, the third and fourth mechanisms above; Supplementary Note 3 and Supplementary Table 4).

We investigated whether activating each of these six IncRNA loci might affect vemurafenib resistance through non-local or local functions. To test whether candidate IncRNAs contributed to vemurafenib resistance via non-local functions, we overexpressed cDNAs encoding each IncRNA from randomly integrated lentivirus. We did not find any that affected drug resistance (Extended Data Fig. 3b), suggesting that these loci are not likely to act through non-local functions (Supplementary Note 4 and Supplementary Table 3). To determine whether the phenotype might result instead from local functions of the IncRNA loci in regulating a nearby gene\(^3,6,7\), we examined the expression of all genes within 1 Mb of the targeted sites. At five of the six loci, SAM targeting led to differential expression of between one and eight nearby protein-coding genes (Supplementary Table 4); for the remaining locus, see Supplementary Note 5). For example, activation of NR_109890 upregulated its neighbouring gene EBF1 (Extended Data Fig. 3c), and activation of TCONS_00015940 led to dosage-dependent resistance.

Figure 2 | Activation of the EMICERI promoter results in dose-dependent upregulation of neighbouring genes. a, Genomic locus of EMICERI contains four neighbouring genes (EQTN, MOB3B, IFNK and C9orf72) and a topological domain. b, Expression of EMICERI and its neighbouring genes after targeting SAM to the EMICERI promoter. c, Expression of EMICERI and MOB3B upon tiling SAM across the EMICERI locus normalized to a non-targeting sgRNA. All values are mean ± s.e.m. with $n = 4$. ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$. ND, not detected.

Figure 3 | MOB3B mediates vemurafenib resistance through the Hippo signalling pathway. a, Vemurafenib resistance of A375 cells overexpressing each neighbouring gene or IncRNA. All values are mean ± s.e.m. with $n = 4$. ****$P < 0.0001$. NS, not significant. b, Western blots of A375 cells overexpressing MOB3B. For gel source images, see Supplementary Fig. 1. c, Proposed mechanism of MOB3B. d, Expression of gene or signature markers for BRAF inhibitor sensitivity (top), genes in the EMICERI locus (middle) and MOB3B overexpression RNA sequencing (RNA-seq) signature (bottom) in BRAF(V600) patient melanomas from The Cancer Genome Atlas (see Methods for signature generation). IC, information coefficient.
upregulation of four neighbouring genes (Fig. 2a, b). Thus, most candidate lncRNA loci appear to confer vemurafenib resistance by regulating the expression of one or more nearby genes, including some that have been previously implicated in vemurafenib resistance (for example, EBFI; see Supplementary Table 3).

To further elucidate the mechanism for one of these candidate local regulators, we focused on TCONS_00015940, which, when targeted, led to a dosage-dependent activation of the four closest nearby genes (EQTN, MOB3B, IFNK, and C9orf72) (Fig. 2a, b). The targeted site is a putative enhancer marked by assay for transposase-accessible chromatin (ATAC) and histone H3 lysine-27 acetylation (H3K27ac) that is proximal to the boundary of a topological domain (Fig. 2a and Extended Data Fig. 4). Upon examining this locus, we found that TCONS_00015940 comprises two separate transcripts (Extended Data Fig. 5a and Supplementary Note 6). We named these transcripts EMICERI I and EMICERII. The EMICERI promoter, which we targeted in our screen, produces two divergent transcripts that initiate about 66 bp apart: EMICERI and MOB3B, a protein-coding gene (Fig. 2a). Tiling SAM across this region indicated that targeting an approximately 200-bp region activated both of these genes (Fig. 2a, c). By contrast, targeting SAM to the promoters of the other three nearby genes did not produce coordinated transcriptional activation in the region (Extended Data Fig. 5b and Supplementary Note 7). These results demonstrate that the EMICERI–MOB3B promoter influences gene expression in a roughly 300-kb gene neighbourhood.

To determine how coordinated upregulation of the EMICERI gene neighbourhood led to vemurafenib resistance, we overexpressed CDNAs encoding each of the four protein-coding genes and EMICERI or EMICERII lncRNAs from randomly integrated lentivirus. Overexpression of one of the six genes, MOB3B, was sufficient to induce the resistance phenotype (Fig. 3a, Extended Data Fig. 6 and Supplementary Note 8). Notably, MOB3B is a kinase activator that is a parologue of MOB1A and MOB1B, known components of the Hippo signalling pathway, whose activation has been shown to confer vemurafenib resistance. MOB3B overexpression downregulated large tumor suppressor kinase 1 (LATS1) to activate the Hippo signalling pathway (Fig. 3b, c, Extended Data Fig. 7a–c and Supplementary Note 9). Activation of EMICERI and MOB3B conferred vemurafenib resistance in two additional vemurafenib-sensitive melanoma cell lines besides the one used in our initial screen (Extended Data Fig. 7d–f), and the effects of activation on gene expression correlated with a signature of vemurafenib resistance in patient melanomas (Fig. 3d, Extended Data Figs 3, 8 and Supplementary Note 9). Together, these results indicate that activation of the EMICERI locus confers vemurafenib resistance through upregulation of MOB3B and subsequent activation of the Hippo signalling pathway.

We next investigated whether upregulation of the EMICERI transcript is required for full MOB3B activation, because targeting SAM to the shared EMICERI–MOB3B promoter (Fig. 4a) could directly activate MOB3B to confer resistance. Accordingly, we used three perturbation methods to interfere with EMICERI transcription and observed effects on MOB3B.
First, to block transcription of EMICERI, we targeted dCas9 downstream of the MOB3B TSS. This intervention reduced the expression of EMICERI, MOB3B and the other neighbouring genes (Fig. 4a, b). We then used a bimodal perturbation system that uses an sgRNA without the SAM-recruitment sequences to target dCas9 to block EMICERI transcription and an sgRNA with the SAM-recruitment sequences to activate the promoter region (Fig. 4a, c). Different combinations of repression and activation sgRNAs demonstrated that the transcriptional levels of EMICERI and MOB3B were tightly coupled across several orders of magnitude ($r = 0.98, P < 0.0001$; Fig. 4d).

Second, we generated clonal A375 cell lines carrying insertions of three tandem polyadenylation signals (pAS) downstream of the EMICERI TSS, which eliminated most of the EMICERI RNA without disrupting the promoter sequence (Fig. 4e, Extended Data Fig. 9a–c and Supplementary Note 10). Upon SAM activation, the pAS-insertion clones showed reduced expression of EMICERI, MOB3B and the three other nearby genes compared to wild-type clones (Fig. 4f, g and Extended Data Fig. 9d–f), and, as expected, reduced vemurafenib resistance (Fig. 4h and Extended Data Fig. 9g, h). This provides genetic evidence that transcription of EMICERI is involved in MOB3B activation.

Third, in the context of SAM activation, we knocked down the EMICERI transcript by transient transfection with antisense oligonucleotides (ASOs; Fig. 4a), which can lead to cleavage of nascent transcripts by RNase H and transcriptional termination (Supplementary Note 11). ASOs targeting EMICERI reduced the expression of both EMICERI and MOB3B in a dosage-dependent manner (Fig. 4i and Extended Data Fig. 10a), consistent with the dCas9 and pAS insertion results.

These EMICERI perturbation experiments demonstrate that transcription of EMICERI is required for full activation of MOB3B, confirming that EMICERI is a functional noncoding locus that activates four neighbouring protein-coding genes and contributes to the screening phenotype. The precise mechanism may involve either the EMICERI transcript or its transcription.$^7,14$

To determine whether MOB3B transcription could reciprocally regulate EMICERI expression, we perturbed MOB3B transcription and observed the effects on EMICERI. We found that targeting dCas9 downstream of the MOB3B TSS successfully blocked MOB3B transcription and reduced expression of EMICERI and other neighbouring genes (Extended Data Fig. 10b); similarly, in the context of SAM activation, ASOs targeting MOB3B introns reduced the activation of both MOB3B and EMICERI (Fig. 4j and Extended Data Fig. 10c). Together, the EMICERI and MOB3B perturbation experiments suggested that transcription of both the lncRNA and the mRNA regulate one another in a positive feedback mechanism that then activates a broader gene neighbourhood, potentially through general processes associated with transcription.$^7,14$

An important challenge in understanding the regulatory logic of the genome has been to identify functional IncRNA loci and characterize their mechanisms. Here we demonstrate that genome-scale activation screens enable systematic identification of IncRNA loci that influence a specific cellular process. We provide a framework for distinguishing categories of IncRNA mechanisms, including non-local functions as well as a diverse array of possible local regulatory mechanisms. Notably, the candidate IncRNA loci we identified appear to act locally to regulate nearby gene expression, including a remarkable case involving coordinated activation of four nearby genes. Further application of our noncoding gain-of-function screening approach, together with loss-of-function screening methods$^{15–18}$ and our characterization strategy, will help to elucidate the complex roles of the noncoding genome in development and disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 30 June 2017.


Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Guttmann, C. M. Johannessen and M. Ghandi for helpful discussions and insights; A. Sayeed, R. Deasy, A. Rotem and B. Izar for generating the primary patient melanoma cell lines; and R. Belliveau, E. Chung, N. Riddle, K. Baugher, R. Macrea and the Zhang laboratory for support and advice. J.M.E. is supported by the Fannie and John Hertz Foundation. O.A.A. is supported by a Paul and Barbara Soros Fellowship and National Defense Science and Engineering Fellowship. J.S.G. is supported by a DOE Computational Science Graduate Fellowship. E.S.L. is supported by UM1HG008895 and funds from the Broad Institute. F.Z. is a New York Stem Cell Foundation-Robertson Investigator. F.Z. is supported by the NIH through NIGMS (RO1-GM07171); J.B.W. is supported by the NIH through NIDDK (F32-DK096822); C.P.F. is supported by the National Defense Science and Engineering Graduate Fellowship. F.Z. is supported by the NIH through NIMHD (3DP1-MH100706 and 1R01-MH110049); NSF, Howard Hughes Medical Institute, the New York Stem Cell Simons, Paul G. Allen Family, and Vaiilee Foundation; and James and Patricia Poitras, Robert Metcalfe, and David Cheng.

METHODS
Design and cloning of SAM IncRNA library. ReSeq noncoding RNAs (release 69) were filtered for IncRNA transcripts that were longer than 200 bp and did not overlap with ReSeq coding gene isoforms18. The ReSeq IncRNA catalogue was combined with the Cabili IncRNA catalogue and filtered for unique IncRNA TSSs, defined as TSSs that were >50 bp apart19. This resulted in 10,504 unique IncRNA TSSs that were targeted with ~10 sgRNAs each for a total library of 95,958 sgRNAs. sgRNAs were designed to target the first 800 bp upstream of each TSS and subsequently filtered for GC content, flanking the target sequence, and homopolymer stretch <4 bp. After filtering, the remaining sgRNAs were scored according to predicted off-target matches as described previously19, and six sgRNAs with the best off-target scores were selected in the first 200-bp region upstream of the TSS: one in the 200–300 bp region, one in the 300–400 bp region, one in the 400–600 bp region, and one in the 600–800 bp region. In regions with an insufficient number of possible sgRNAs, sgRNAs were selected from the neighbouring region closer to the TSS. The ideal location for sgRNA targeting to achieve maximal activation, either upstream or downstream of the TSS, may be unique for each IncRNA locus and dependent on the local regulatory context (for example, locations of TF binding sites). An additional 500 non-targeting sgRNAs were included as controls. Cloning of the SAM sgRNA libraries was performed as previously described with a minimum representation of 100 transformed colonies per sgRNA followed by next-generation sequencing (NGS) validation19.

Lentivirus production and transduction. For transduction, plasmids were packed into lentivirus through transfection of library plasmid with appropriate packaging plasmids (psPAX2: Addgene 12260; pMD2.G: Addgene 12259) using Lipofectamine 2000 (Thermo Fisher 11668019) and Plus reagent (Thermo Fisher 11514015) in HEK293FT (Thermo Fisher R70007) as described previously20. The HEK293FT cell line is on the list of commonly misidentified cell lines maintained by the International Cell Line Authentication Committee, and has not been authenticated or tested for mycoplasma contamination. HEK293FT cells are most commonly used for lentivirus production because they are easy to transfect and produce lentivirus very efficiently. Human melanoma A375 cells (Sigma-Aldrich 88113005) were cultured in R10 medium: RPMI 1640 (Thermo Fisher 61870) supplemented with 10% FBS (VWR 79068-085) and 1% penicillin/streptomycin (Thermo Fisher 15140122). A375 cells have not been tested for mycoplasma contamination. Cells were passaged every other day at a 1:3 ratio. Concentrations for selection agents were determined using a kill curve: 300 μg/ml zeocin (Thermo Fisher R25001), 10 μg/ml blasticidin (Thermo Fisher A1113903), and 300 μg/ml hygromycin (Thermo Fisher 10687010). Cells were transduced by spinfection and selected with the appropriate antibiotic as described previously20. During selection, medium was refreshed when cells were passaged every 3 days. The duration of selection was 7 days for zeocin and 5 days for hygromycin and blasticidin. Lentiviral titres were calculated by spincfecting cells with five different volumes of lentivirus and determining viability after a complete selection of 3 days19.

Vemurafenib resistance screen. The vemurafenib resistance screen was conducted similarly to a previously described genome-scale SAM coding gene screen21. A375 cells stably integrated with dCas9-VP64 (Addgene 61426) and MS2-P65-HSF1 (Addgene 61427) were transduced with the pooled sgRNA library (Addgene 61425) as described above at an MOI of 0.3 for a total of four infection replicates, with a minimal representation of 500 transduced cells per sgRNA in each replicate. Cells were maintained at ~500 cells per well for subsequent passaging. A375 cells or additional melanoma cell lines, A375 cells or additional melanoma cell lines (A2058, ATCC CRL-11471; COLO679, Sigma-Aldrich 87061210; UACC62, AddexBio C0002003) cultured in R10 medium were transduced with SAM effectors and EMICERI-targeting sgRNA 2 (Supplementary Table 7) or with control non-targeting sgRNA. Five days after transduction, cells were replated at low density (~105 cells per well in a 96-well plate; four biological replicates per condition), Vemurafenib (2μM) or control DMSO was added 3 h after plating and refreshed every 2 days for 3–4 days before cell viability was measured using CellTiter-Glo Luminescent Cell Viability Assay (Promega G7571). Significance testing was performed using two-sided Student’s t-test. For primary patient tumour-derived melanoma cell lines, cells were plated at low density (~105 cells per well in a 96-well plate; four biological replicates per condition) and vemurafenib was added 24 h after plating. Cells were treated for 3 days before cell viability was measured. For vemurafenib dose–response curves, the indicated concentrations of vemurafenib were added and the normalized per cent survival values were fitted with a nonlinear curve (log[Inhibitor]) vs normalized response; Prism 6). Significant differences in logIC50 values were determined using the extra sum-of-squares F test. qPCR quantification of transcript expression. A375 cells or additional melanoma cell lines stably integrated with SAM components were transduced with individual sgRNAs targeting the top candidate IncRNA loci (Supplementary Table 5), perturbing the EMICERI locus (Supplementary Table 7), or non-targeting control at an MOI of <0.5 and selected with zeocin for 5 days as described above. For cDNA overexpression, A375 cells were transduced with cDNA (Supplementary Table 6) or control GFP at an MOI of <0.5 and selected with hygromycin for 4 days. Cells were plated 5 days after transfection at 70% confluency (~3 × 104 cells per well in a 96-well plate; four biological replicates per condition, each with four technical replicates), and harvested for RNA 24 h after plating as described previously20. For transcripts that this method could not detect, cells transduced with the respective sgRNAs were plated 5 days after transfection (1.8 × 105 cells per well in a 24-well plate; three biological replicates per condition). RNA was harvested using the RNeasy Plus Mini Kit (Qiagen 74134) and 1 μg RNA was used for reverse transcription with the qScript Flex cDNA Kit (VWR 95049) and IncRNA-specific primers (Supplementary Table 8). After reverse transcription, TaqMan qPCR was performed with custom or readymade probes as described previously19 (Supplementary Tables 6, 8). Significance testing was performed using a two-sided Student’s t-test.

RNA sequencing and data analysis. A375 cells transduced with individual sgRNAs targeting candidate IncRNA loci or with control non-targeting sgRNAs (Supplementary Table 5) were plated 5 days after transfection at ~9 × 104 cells per well or ~1.8 × 105 cells per well, respectively, in a 24-well plate. Three biological replicates per condition were plated. For cDNA overexpression, A375 cells were transduced with cDNA (Supplementary Table 6) or control GFP at an MOI of <0.5 and selected with hygromycin for 4 days. Cells were treated with 2μM vemurafenib for 3 days before RNA was harvested as described above. The six candidate loci with the highest fold change in mRNA expression were prepared with TrueSeq Stranded Total RNA Sample Prep Kit with Ribo-Zero Gold (Illumina RS-122-2302) and all other samples were prepared with NEBNext Ultra RNA Library Prep Kit for Illumina (NEB E7350S) and NEBNext PAS mRNA Magnetic Isolation Module (NEB E490S). Libraries were deep-sequenced on the Illumina NextSeq platform (~9 million reads per condition). A Bowtie3 index was created, based on the human hg19 UCSC genome, and known gene and IncRNA transcriptomes were constructed as described above. Paired-end reads were aligned directly to this index using Bowtie2 with command line options ‘–qphred33–quals -n 2 -e 0.00000001 -I 25 -I 1 -X 1000 –churnk 512 -p 1 -m 200 –S’. Next, RSEM 1.2.224 was used to get read counts on the alignments created by Bowtie to estimate expression levels.

RSEM’s TPM estimates for each transcript were transformed to log-space by taking log10(TPM + 1). Transcripts were considered detected if their transformed expression level was equal to or above 1 (in log10(TPM + 1) scale). All genes detected in at least one library (out of three libraries per condition) were used to find
differentially expressed genes. For activation of IncRNA loci, the two-sided Student’s t-test was performed on each of the three replicates for each targeting sgrNA against both non-targeting sgrNAs. For MOB3B cDNA overexpression, the t-test was performed on the cDNA overexpression against GFP control. Only genes that were significant (P value > 0.05 FDR correction) were reported. For IncRNA loci activation, the genes overlapping all three targeting sgrNAs were reported as differentially expressed as a result of IncRNA loci activation. Power analysis for two-sided t-tests were performed on each targeting sgrNA against both non-targeting sgrNAs to determine the probability of correctly identifying a gene as differentially expressed.

For annotating EMICERI, TopHat2 was used to align RNA-seq reads from A375 transduced with sgrNA 2 or sgrNA 3 (Supplementary Table 7) with command line options ‘-so=sca-quals–num-threads =8–library-type fr-firststrand–transcriptome- max-hits =1–prefilter-multihits–keep–fasta-order’. To further investigate the mechanism for MOB3B overexpression, Ingenuity Pathway Analysis was applied to all genes that were differentially expressed with at least 1.2-fold change or less than 0.7-fold change and the most likely upstream regulator was reported.

Hi-C and chromatin immunoprecipitation with sequencing (ChIP–seq) in GM12878. In situ Hi-C data for GM12878 were obtained and visualized using a 2.5-kb-resolution KL-normalized observed matrix23. Hi-C data from seven cell lines suggested similar topological domain annotations as GM12878, suggesting that the TAD present in GM12878 is consistent across cell types. CTCF ChIP–seq for GM12878 and hg19 generated by the ENCODE Project Consortium24 was downloaded from the UCSC Genome Browser. CTCF motifs were identified using FIMO to search for the ‘V_CTCF_01’ and ‘V_CTCF_02’ position weight matrices from TRANSFAC as described previously25.

Assay for transposable and accessible chromatin sequencing (ATAC–seq). ATAC–seq samples were prepared as described previously17. The library was sequenced using Illumina NextSeq platform (1:3.6 million paired-end reads). Samples were aligned to the human hg19 UCSC genome using Bowtie2 with command line options ‘–n=25 mb=24 -s m=1 -X 2000’. For quality control, the duplicate read rate was measured using Picard-Tools Mark Duplicates (10–30%) and the mitochondrial read rate was measured by Bowtie alignment to chrM (<5%).

PhastCons sequence conservation. PhastCons data for primates (n = 10 animals), placental mammals (n = 33) and vertebrates (n = 46) for hg19 were downloaded from the UCSC Genome Browser and aligned to the EMICERI locus18. ChIP–seq for histone modifications. ChIP-seq samples were prepared as described previously27.

In brief, A375 cells were plated in T-225 flasks and grown to 70–90% confluence. Formaldehyde was added directly to the growth medium for a final concentration of 1% for 10 min at 37 °C to initiate chromatin fixation. The entire two-chip Hi-C procedure was performed using the EZ-Magna ChIP HiSense Chromatin Immunoprecipitation Kit (Millipore 1710460) according to the manufacturer’s protocol. Samples were pulse-sonicated with two rounds of 10 min (30 s on-off cycles, high frequency) in a rotating water bath sonicator (Diagenode Bioruptor) with 5 min on ice between each round. To detect histone modifications, antibodies (H3K3me2: Millipore 17-677, H3K4me3: Millipore 04-745, H3K27ac: Millipore 17-683) were optimized individually for each antibody to be 0.5 μl for one million cells. One microlitre of IgG (Millipore 12-370) was used for negative control.

After verifying that the IgG ChIP had minimal background, ChIP samples were prepped with NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645S) and deep-sequenced on the Illumina NextSeq platform (>60 million reads per condition). Bowtie2 was used to align paired-end reads to the human hg19 UCSC genome with command line options ‘-q X 500–sam-chunks 512’. Next, Model-based analysis of ChIP–seq (MACS) was run with command line options ‘–q g-h B -s–call-subpeaks’ to identify histone modifications.

Western blots were prepared for each treatment group. Whole cell lysates were isolated by scraping into the western lysis buffer (0.5% IG, 0.1% Triton X-100, 1× Protease Inhibitor Cocktail). Lysates were then subjected to SDS–PAGE and transferred onto nitrocellulose membranes for each blot. The membranes were blocked in Odyssey Blocking Buffer (TBS; LiCOR 925–50000) and probed with different primary antibodies (anti-PERK (Cell Signaling Technologies 4370, 1:2,000 dilution), anti-ERK (Cell Signaling Technologies 4695, 1:1,000 dilution), anti-pAkt (Ser473, Cell Signaling Technologies 4060, 1:1,000 dilution), and anti-LATS1 (Cell Signaling Technologies 3477, 1:1,000 dilution), anti-YAP/TAZ (Cell Signaling Technologies 8418, 1:1,000 dilution), anti-MST1 (Cell Signaling Technologies 3682, 1:1,000 dilution) and anti-ACTB (Sigma A5441, 1:5,000 dilution) overnight at 4 °C. Blots were then incubated with secondary antibodies: IRDye 680RD Donkey anti-Mouse IgG (LiCOR 925–68072) and IRDye 800CW Donkey anti-Rabbit IgG (LiCOR 925–32213) at 1:2,000 dilution in Odyssey Blocking Buffer for 1 h at room temperature. p-ERK and p-AKT blots were stripped with Restore PLUS Western Blot Stripping Buffer (Thermo Fisher 46430) before probing for ERK and AKT, respectively. Blots were then probed with monoclonal antibodies against GAPDH (Roche 10494303) with their antibody dilutions at 1:5000 for 1 h at room temperature. Blots were developed using the Odyssey Lumi imaging system (LiCOR).

Primary patient melanoma-derived cell lines. CLF_SFKCM_001_T and CLF_SFKCM_004_T melanoma tumour tissues were obtained from the Dana-Farber Cancer Institute hospital with informed consent and the cancer cell model line generation was approved by the ethical committee. Tumour tissues were dissected into tiny pieces using scalpels around 100 times. Dissected tissues were dissociated in collagenase–hyaluronidase (STEMCELL technologies 07912) medium for 1 h. Red blood cells were further depleted by adding ammonium chloride solution (STEMCELL Technologies 07800). The dissociated cells were plated with smooth muscle growing medium–2 (Lonza CC-3181) in a six-well plate and split when the well confluence reached 80%. Cells were passaged five times with 1:4 splitting ratio for sequencing verification. The confirmed BRAF V600E melanoma cell models were propagated for another 7–15 passages and preserved in a cryovial. We used passage 12 cells for this study. All cells were fed every 3–4 days.

Gene expression and pharmacological validation analysis. Gene expression data (CCLE, TCGA) and pharmacological data (CCLE) were analysed to better understand the biological relevance of EMICERI and MOB3B. Transcript expression in TCGA and CCLE samples was quantified as follows: 1) FASTQ files were generated from available BAM files using SamToFastq in Picard Tools (https://broadinstitute.github.io/picard/); 2) reads were aligned with STAR v2.5.2b using parameters from the GTEx consortium pipeline (https://github.com/broadinstitute/gtex-pipeline) and genome indexes were generated for read lengths of 48 bp (TCGA) and 101 bp (CCLE) (–sjdbOverhang option); 3) expression was quantified using RSEM v1.2.2221. For the alignment and quantification steps, annotations for TCONS_00011252, NR_034078, TCONS_00026344, TCONS_00015940_1, TCONS_00015940_2, and NR_109890 were appended to GENCODE 19 GTF (https://www.gencodegenes.org/releases/19.html).

Gene expression (RNA-seq) data were collected from 113 BRAF V600E–mutant primary and metastatic patient tumours from The Cancer Genome Atlas (TCGA: https://portal.gdc.cancer.gov/). Because pharmacological data were not available for the TCGA melanoma samples, signature gene sets were used to fully map the transcriptional BRAF-inhibitor-resistant or -sensitive states in TCGA as previously described26. The TCGA data set was used to determine the association between the resistance and the expression of candidate IncRNA loci or genes in the EMICERI locus. In addition, we sought a more robust scoring system independent of any single gene. Gene expression signatures were generated based on the genes that were differentially expressed (top 1,000 most differentially expressed) as a result of candidate IncRNA loci or MOB3B overexpression identified from RNA-seq. Using single-sample Gene Set Enrichment Analysis (ssGSEA)27, a score was generated for each sample that represents the enrichment of the gene expression signature in that sample and the extent to which those genes are coordinately up- or downregulated. Patient tumours were also sorted by EMICERI expression to determine the correlation between expression of EMICERI and its neighbouring genes.

In the CCLE data set28, gene expression data (RNA-sequencing, CGHub: https://portal.gdc.cancer.gov/) and pharmacological data (activity area for MAPK pathway inhibitors) from BRAF V600E mutant melanoma cell lines were used to compute the association between PLX-4720 resistance and the expression of genes in the EMICERI locus. Similar to the TCGA analysis, the MOB3B overexpression gene signature was determined using ssGSEA projected onto the CCLE RNA-sequencing data set. Gene expression scores were then used to determine the correlation between expression of EMICERI and its neighbouring genes.

To measure correlations between different features (signature scores, gene expression, or drug-resistance data) in the external cancer data sets, an information-theoretic approach (information coefficient; IC) was used and significance was measured using a permutation test (n = 10,000) as previously described29. The IC was calculated between the feature used to sort the samples (columns) in each data set and each of the features plotted in the heat map (pharmacological data, gene expression, and signature scores).

pAS Insertion. To truncate EMICERI, the following pAS sequences were inserted downstream of the EMICERI TSS: Synthetic pAS: AATTTTTTTTCTTACATC ATGATCTTGTTGT GGT TTTT TTGTTGTTGTTG

SV40 pAS: GTTTATGTCAGGCTAATATAGTTTACAATACAAAGCACTAGCA TCAACAAATTCTAACAAAGCTTTTTTTTCTCAGGATCTGTGATTTGTTGTTGCAACACTCAATGTTACATGTTG

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
PGK pAS: AAATTGATGATCTTTTAAACATGAGATGTCCTACTAAATGGAAGTTTTTTCTCTGATCATTGTTCAAGAGGTGAGAAGAGTACCTCATTCTTTGTGGATTTGTTTAATGATATGTTGATGATGATAATGTTTCATAGTTGGATATCATAATTTAAACCAAGAAAACAAAATTAAAGGGGCAGCTATATCTTTGCTCAGATCTATA

pAS clones were generated using Cas9-mediated homology-directed repair (HDR). Three sgRNAs targeting 103, 156 and 198 bp downstream of the EMICERI TSS (HDR sgRNA 1–3, Supplementary Table 9) and corresponding pAS HDR plasmids were used to insert pAS into each of the three copies of EMICERI in A375 cells. To construct the pAS HDR plasmids, for each sgRNA the HDR templates, which consisted of the 850–900-bp genomic regions flanking the sgRNA cleave site, were PCR amplified from A375 cell genomic DNA using KAPA HiFi HotStart ReadyMix (KAPA Biosystems KK2602). Then three pAS sequences in tandem (in the order listed above) flanked by the HDR templates were cloned into pUC19 (Addgene 50005). To insert pAS downstream of the EMICERI TSS, three rounds of HDR were performed with a different sgRNA and respective pAS HDR plasmid at each round, such that selected clones contained pAS sequences in one copy of EMICERI in the first round, two copies in the second round, and three copies in the third round. At each round of HDR, A375 cells were nucleasefied with 4 μg sgRNA and Cas9 plasmid (Addgene 52961) and 2.5 μg pAS HDR plasmid using SF Cell Line 4D-Nuclease X Kit L (Lonza V4XC-2024) according to the manufacturer’s instructions. Cells were then seeded sparsely (5 × 104 cells per well in a 96-well plate with four biological replicates per condition. Twenty-four hours after nucleasefication, cells were selected for the sgRNA plasmid with 1 μg/ml puromycin (Thermo Fisher A1113803) for 2 days and changes in transcript expression were determined by qPCR as described above.

ASO knockdown. ASOs targeting EMICERI, EMICERII and MOB3B were custom-designed using Exiqon’s Antisense LNA GapmeR designer (Supplementary Table 10) and a non-targeting ASO (Exiqon 300610) was included for control. ASOs were resuspended in water to a final concentration of 100 μM. A375 cells stably expressing SAM components dCas9-VP64 and MS2-p65-HSF1 were nucleasefied with 500 ng sgRNA (Supplementary Table 7; Addgene 73795) and 100 pmol ASO using the SF Cell Line 4D-Nuclease X Kit S (Lonza V4XC-2032) according to the manufacturer’s instructions. Cells were then seeded at 3 × 104 cells per well in a 96-well plate with four biological replicates per condition. Twenty-four hours after nucleasefication, cells were selected for the sgRNA plasmid with 1 μg/ml puromycin (Thermo Fisher A1113803) for 2 days and changes in transcript expression were determined by qPCR as described above.

Code availability. Code for IncRNA library design and the analyses described in this paper is available on Github (https://github.com/fenglab/LncRNA_Screen_Manuscript). Please note that the code is published as is for reference.

Data availability. Sequencing data for this study are available through the Gene Expression Omnibus (GSE99836) and Bioproject (PRJNA324504). All additional data are available from the authors upon reasonable request. The IncRNA library and associated SAM plasmids are available from Addgene under UBMTA.

Extended Data Figure 1 | Genome-scale activation screen for lncRNA loci involved in vemurafenib resistance. a, Scatterplots showing lncRNA-targeting and non-targeting sgRNA frequencies after vemurafenib (vemu) or control treatment from \(n = 4\) infection replicates. b, Scatterplot showing enrichment of sgRNAs targeting six candidate lncRNA loci. c, RIGER \(P\) values of the candidate lncRNA loci. d, RIGER \(P\) values for the top 100 hits from the previous SAM protein-coding gene screen compared to the SAM lncRNA loci screen. e, For each candidate lncRNA locus, ten sgRNAs were designed to target the proximal promoter region (800 bp upstream of the TSS). The relationship between the highest sgRNA enrichment in vemurafenib-treated compared to control condition across screening bioreps (\(n = 4\)) and respective spacer position suggests that sgRNAs targeting closer to the annotated TSS are not necessarily more effective, consistent with previous results.
Validation of candidate lncRNA loci for vemurafenib resistance. Vemurafenib resistance for A375 cells transduced with SAM and individual sgRNAs targeting the top 16 candidate lncRNA loci normalized to a non-targeting (NT) sgRNA. All values are mean ± s.e.m. with n = 4. **P < 0.001; ***P < 0.001; ****P < 0.01.
Extended Data Figure 3 | Activation of candidate lncRNA loci appears to mediate vemurafenib resistance by regulating expression of nearby genes. **a**, Heat map showing expression of genes and signature markers for BRAF inhibitor sensitivity (top), expression of candidate lncRNA loci (middle), and RNA-seq signature of gene expression changes upon activation of candidate lncRNA loci (bottom) in 113 different BRAF(V600) patient melanoma samples (primary or metastatic) from The Cancer Genome Atlas. All associations are measured using the information coefficient (IC) between the index and each of the features and P values are determined using a permutation test. Panels show Z-scores. **b**, Vemurafenib resistance of A375 cells overexpressing cDNAs encoding each candidate lncRNA or protein-coding gene normalized to GFP. GPR35 and LPAR1 are positive controls identified previously. The same set of sgRNAs targeted TCONS_00012395 and TCONS_00011252; NR_034078 and NR_034079; TCONS_00015940_1 and TCONS_00015940_2. **c**, Expression of NR_109890 and its neighbouring gene EBF1 after SAM activation of NR_109890. All values are mean ± s.e.m. with n = 4. ****P < 0.0001; ***P < 0.001; *P < 0.05. ns, not significant.
Extended Data Figure 4 | Topological domain in the *EMICERI* locus is consistent across cell types. Hi-C data and topological domain annotations (dotted lines) in the *EMICERI* locus from seven cell lines\(^2\). Heat map shows KR-normalized contact matrix at 5-kb resolution. Domain annotations for chromosome 9 were not available in K562, but the same topological domain structure is evident.
Extended Data Figure 5 | Dosage-dependent upregulation of the EMICERI locus is specific to activation of EMICERI at its conserved regulatory element. a, TopHat alignment of RNA-seq paired-end reads suggests that EMICERI is located at chr9:27,529,917–27,531,782 and EMICERII at chr9:27,535,71–27,540,711 (UCSC hg19) (Supplementary Note 6). A375 ATAC-seq and phastCons conservation scores for primates, placental mammals, and vertebrates at the EMICERI locus. Scale bar, 1 kb.
b, Expression of EMICERI and its neighbouring genes MOB3B, IFNK, EQTN and C9orf72 after transduction with sgRNAs targeting SAM to the promoters of neighbouring genes. All values are mean ± s.e.m. with n = 4.
* * * * P < 0.0001; *** P < 0.001; ** P < 0.01; * P < 0.05. ns, not significant.
ND, not detected.
Extended Data Figure 6 | Activation of EMICERI mediates vemurafenib resistance through MOB3B. a, Expression of the neighbouring genes or EMICERI or EMICERII after cDNA overexpression compared to GFP control. b, cDNA overexpression of top hits from the SAM protein-coding gene screen for vemurafenib resistance (GPR35 and LPAR1) or MOB3B compared to GFP control. c, Vemurafenib dose response curves for A375 cells overexpressing cDNA or GFP control. d, Vemurafenib half maximal inhibitory concentration (IC$_{50}$) for the same conditions as in c. e, ATARI gene-level scores from the Achilles Project that reflect genetic vulnerabilities of A375 cells. Lower ATARI gene-level scores indicate stronger dependency on the gene. Rank of MOB3B, 1,084; IFNK, 3,078; EQTN, 15,939. All values are mean ± s.e.m. with n = 4. **P < 0.01; *P < 0.05. ND, not detected.
Extended Data Figure 7 | Activation of EMICERI mediates vemurafenib resistance in melanoma cell lines. a–c, Western blots of A375 cells stably overexpressing MOB3B cDNA or GFP control after vemurafenib or control (DMSO) treatment. For gel source data, see Supplementary Fig. 1. d, Expression of EMICERI and MOB3B after SAM activation in different melanoma cell lines. e, Vemurafenib dose response curves for EMICERI activation in different melanoma cell lines. f, Vemurafenib half maximal inhibitory concentration (IC_{50}) for the same conditions as in e. All values are mean ± s.e.m. with n = 4. **P < 0.001; ***P < 0.001; ****P < 0.01; **P < 0.05. ns, not significant.
Extended Data Figure 8 | **EMICERI** expression is strongly correlated with **MOB3B** expression and vemurafenib sensitivity in melanoma cell lines and patient samples. **a**, Heat map showing expression of genes in the **EMICERI** locus in 113 different BRAF (V600) patient melanoma samples (primary or metastatic) from The Cancer Genome Atlas. Samples are sorted by **EMICERI** expression. **b**, Heat map showing expression of genes in the **EMICERI** locus in melanoma cell lines from the Cancer Cell Line Encyclopedia (CCLE) sorted by **EMICERI** expression. **c**, Heat map showing sensitivity to different drugs (top), expression of genes in the **EMICERI** locus (middle), and **MOB3B** cDNA overexpression RNA-seq signature (bottom; see Methods for signature generation) in melanoma cell lines from CCLE. Drug sensitivities are measured as Activity Areas. The melanoma cell lines are sorted by PLX-4720 (vemurafenib) drug sensitivity. RAF inhibitors: PLX-4720 and RAF265; MEK inhibitors: AZD6244 and PD-0325901. **d**, Expression of **EMICERI** and **MOB3B** in two primary patient-derived BRAF(V600E) melanoma cell lines. **e**, Vemurafenib dose response curves for the same cell lines. **f**, Vemurafenib half maximal inhibitory concentration (IC$_{50}$) for the same conditions as in **e**. All associations are measured using the information coefficient (IC) between the index and each of the features and $P$ values are determined using a permutation test. Heat maps show Z-scores. All values are mean ± s.e.m. with $n = 4$. ****$P < 0.0001$; *$P < 0.05$. 

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 9 | Transcriptional activation of EMICERI modulates expression of neighbouring genes to confer vemurafenib resistance. a, Gel confirming pAS insertion into all three copies of EMICERI for each pAS clone. For gel source data, see Supplementary Fig. 1. b, c, Basal expression of EMICERI and MOB3B for the wild-type and pAS clones. d–f, Expression of C9orf72, IFNK and EQTN after targeting SAM to EMICERI for the wild-type and pAS clones. g, Vemurafenib dose response curves for wild-type and pAS clones transduced with SAM and EMICERI-targeting or non-targeting (NT) sgRNAs. h, Vemurafenib half-maximal inhibitory concentration (IC50) for the same conditions as in g. All values are mean ± s.e.m. with n = 4. 

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 10 | Transcriptions of *EMICERI* and *MOB3B* act reciprocally to regulate each other. 

**a**. Expression of *EMICERI* and *MOB3B* after ASO knockdown of *EMICERI* in the context of SAM activation. 

**b**. Expression of *MOB3B* and neighbouring genes in A375 cells transduced with non-targeting (NT) or *MOB3B*-targeting sgRNAs and dCas9. 

**c**. Expression of *MOB3B* and *EMICERI* after ASO knockdown of *MOB3B* in the context of SAM activation. All values are mean ± s.e.m. with n = 4. ****P < 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05. ns, not significant.