ARTICLE

doi:10.1038/nature14136

Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex

Silvana Konermann[1,2,3,4]*, Mark D. Brigham[1,2,3,4]*, Alexandro E. Trevino[1,2,3,4], Julia Joung[1,4], Omar O. Abudayyeh[1,2,3,4], Clea Barcena[1,2,3,4], Patrick D. Hsu[1,2,3,4], Naomi Habib[1], Jonathan S. Gootenberg[1,2,3,4,5], Hiroshi Nishimasu[6,7], Osamu Nureki[6] & Feng Zhang[1,2,3,4]

Systematic interrogation of gene function requires the ability to perturb gene expression in a robust and generalizable manner. Here we describe structure-guided engineering of a CRISPR-Cas9 complex to mediate efficient transcriptional activation at endogenous genomic loci. We used these engineered Cas9 activation complexes to investigate single-guide RNA (sgRNA) targeting rules for effective transcriptional activation, to demonstrate multiplexed activation of ten genes simultaneously, and to upregulate long intergenic non-coding RNA (lincRNA) transcripts. We also synthesized a library consisting of 70,290 guides targeting all human RefSeq coding isoforms to screen for genes that, upon activation, confer resistance to a BRAF inhibitor. The top hits included genes previously shown to be able to confer resistance, and novel candidates were validated using individual sgRNA and complementary DNA overexpression. A gene expression signature based on the top screening hits correlated with markers of BRAF inhibitor resistance in cell lines and patient-derived samples. These results collectively demonstrate the potential of Cas9-based activators as a powerful genetic perturbation technology.

Achieving systematic, genome-scale perturbations within intact biological systems is important for elucidating gene function and epigenetic regulation. Genetic perturbations can be broadly classified as either loss-of-function (LOF) or gain-of-function (GOF) on the basis of their mode of action. To date, various genome-scale LOF screening methods have been developed, including approaches employing RNA interference[1,2] and the RNA-guided endonuclease Cas9 from the microbial CRISPR (clustered regularly interspaced short palindromic repeat) adaptive immune system[3,4]. Genome-scale GOF screening approaches have largely remained limited to the use of cDNA library overexpression systems. However, it is difficult to capture the complexity of transcript isoform variance using these libraries, and large cDNA sequences are often difficult to clone into size-limited viral expression vectors. The cost and complexity of synthesizing and using pooled cDNA libraries have also limited their use. Novel technologies that overcome such limitations would enable systematic, genome-scale GOF perturbations at endogenous loci.

Programmable DNA-binding proteins have emerged as an exciting platform for engineering synthetic transcription factors for modulating endogenous gene expression[5–11]. Among the established custom DNA-binding domains, Cas9 is most easily scaled to facilitate genome-scale perturbations[3,4] owing to its simplicity of programming relative to zinc finger proteins and transcription activator-like effectors (TALEs). Cas9 nuclease can be converted into an RNA-guided transcription activator (dCas9–activator) via inactivation of its two catalytic domains[12,13] and fusion to transcription activation domains. These dCas9–activator fusions targeted to the promoter region of endogenous genes can then modulate gene expression[7–11]. Although the current generation of dCas9-based transcription activators is able to achieve upregulation of some endogenous loci, the magnitude of transcriptional upregulation achieved by individual single-guide RNAs (sgRNAs)[12] typically ranges from low to ineffective[8–11]. Tiling a given promoter region with several sgRNAs can produce more robust transcriptional activation[9–11], but this requirement presents enormous challenges for scalability, and in particular for establishing pooled, genome-wide GOF screens.

To improve and expand the applications of Cas9, we recently undertook crystallographic studies to elucidate the atomic structure of the Cas9–sgRNA–target DNA tertiary complex[14], thus enabling rational engineering of Cas9 and sgRNA. Here we report a series of structure-guided engineering efforts to create a potent transcription activation complex capable of mediating robust upregulation with a single sgRNA. Using this new activation system, we demonstrate activation of endogenous genes as well as non-coding RNAs, elucidate design rules for effective sgRNA target sites, and establish and apply genome-wide dCas9-based transcription activation screening to study drug resistance in a melanoma model. These results collectively demonstrate the broad applicability of CRISPR-based GOF screening for functional genomics research.

Structure-guided design of Cas9 complex

Transformation of the Cas9–sgRNA complex into an effective transcriptional activator requires finding optimal anchoring positions for the activation domains. Previous designs of dCas9-based transcription activators have relied on fusion of transactivation domains to either the amino or carboxy terminus of the dCas9 protein. To explore whether alternate anchoring positions would improve performance, we examined our previously determined crystal structure of the *Streptococcus pyogenes* dCas9(D10A/H840A) in complex with a single-guide RNA (sgRNA) and complementary target DNA[14]. We observed that the tetraloop and stem loop 2 of the sgRNA protrude outside of the Cas9–sgRNA ribonucleoprotein complex, with the distal 4 base pairs (bp) of each stem completely free of interactions with Cas9 amino acid side chains (Extended Data Fig. 1a). On the basis of these observations, along with

[1]Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, Massachusetts 02142, USA. [2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi Bunkyo, Tokyo 113-0032, Japan. [7]JST, PRESTO 2-11-16 Yayoi Bunkyo, Tokyo 113-0032, Japan.
*These authors contributed equally to this work.

©2015 Macmillan Publishers Limited. All rights reserved

functional data demonstrating that substitutions and deletions in the tetraloop and stem loop 2 regions of the sgRNA sequence do not affect Cas9 catalytic function[14] (Fig. 1a), we reasoned that the tetraloop and stem loop 2 could tolerate the addition of protein-interacting RNA aptamers to facilitate the recruitment of effector domains to the Cas9 complex (Fig. 1b).

We selected a minimal hairpin aptamer, which selectively binds dimerized MS2 bacteriophage coat proteins[15], and appended it to the sgRNA tetraloop and stem loop 2 (Extended Data Fig. 1b). We next tested whether MS2-mediated recruitment of VP64 to the tetraloop and stem loop 2 could mediate transcriptional upregulation more efficiently than a dCas9–VP64 fusion. As predicted, aptamer-mediated recruitment of MS2–VP64 to either tetraloop (sgRNA1.1) or stem loop 2 (sgRNA1.2) mediated three- and fivefold higher levels of *Neurog2* upregulation than a dCas9–VP64 fusion (sgRNA 1.0), respectively. Recruitment of VP64 to both positions (sgRNA 2.0) resulted in an additive effect, leading to a 12-fold increase over dCas9–VP64 (sgRNA 1.0). Combining sgRNA 2.0 with dCas9–VP64 instead of dCas9 provided an additional 1.3-fold increase in *Neurog2* upregulation (Fig. 1c). We further compared sgRNA 2.0 to a sgRNA previously described bearing two MS2-binding stem loops at the 3′ end (sgRNA + 2×MS2)[11] and found that sgRNA 2.0 drove 14- and 8.5-fold higher levels of transcription activation than sgRNA + 2×MS2 for *ASCL1* and *MYOD1*, respectively (Fig. 1d). This difference could be due to either improved positioning of MS2 stem loops or to dCas9 protection of internal MS2 stem loops from exonuclease degradation.

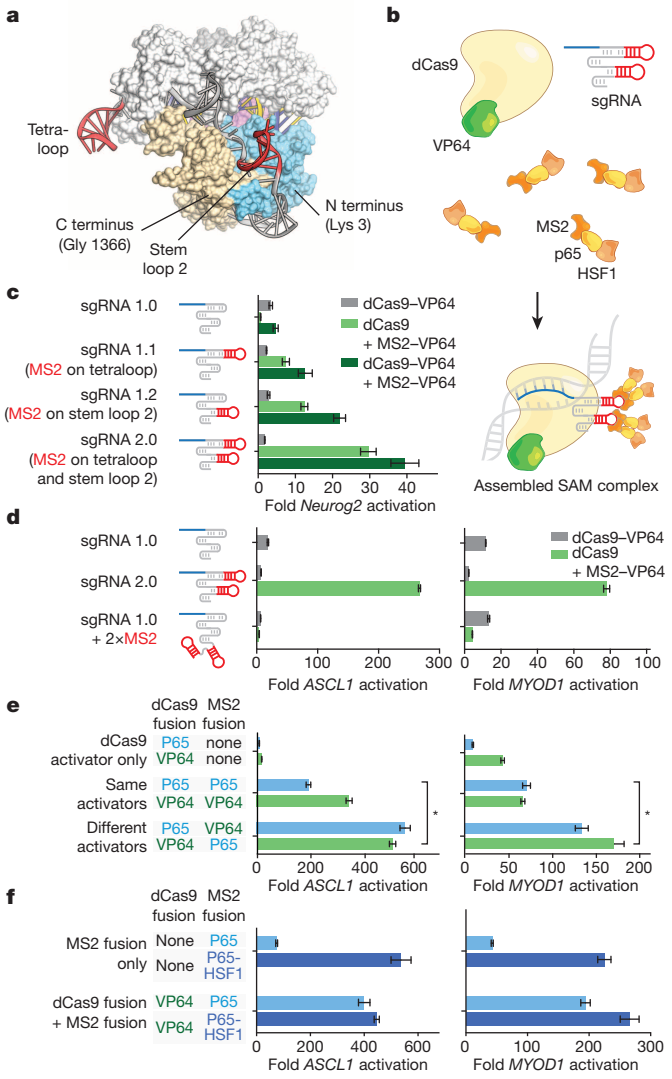


Figure 1 | Structure-guided design and optimization of an RNA-guided transcription activation complex. **a**, A crystal structure of the Cas9–sgRNA–target DNA ternary complex (PDB ID: 4OO8)[14] reveals that the sgRNA tetraloop and stem loop 2 are exposed. **b**, Schematic of the three-component SAM system. **c**, Design and optimization of sgRNA scaffolds for optimal recruitment of MS2–VP64 transactivators in Neuro-2a cells. **d**, MS2 stem loop placement within the sgRNA significantly affects transcription activation efficiency. **e**, Combinations of different activation domains act in synergy to enhance the level of transcription activation. **f**, Addition of the HSF1 transactivation domain to MS2–p65 further increases the efficiency of transcription activation. Experiments for **d**–**f** were performed in 293FT cells. All values are mean ± s.e.m. with $n = 3$ biological replicates. $^*P < 0.05$ based on Student's *t*-test.

To further improve the potency of Cas9-mediated gene activation, we considered how transcriptional activation is achieved in natural contexts, where endogenous transcription factors generally act in synergy with co-factors[16]. We thus hypothesized that combining VP64 with additional, distinct activation domains could improve activation efficiency. We chose the NF-κB trans-activating subunit p65 that, while sharing some common co-factors with VP64, recruits a distinct subset of transcription factors and chromatin remodelling complexes. For example, p65 has been shown to recruit AP-1, ATF/CREB and SP1 (ref. 17), whereas VP64 recruits PC4 (ref. 18), CBP/p300 (ref. 19), and the SWI/SNF complex[20].

We then varied the effector domain fused to dCas9 or MS2. Hetero-effector pairing of dCas9 and MS2 fusion proteins (for example, dCas9–VP64 paired with MS2–p65 or dCas9–p65 with MS2–VP64) provided over 2.5-fold higher transcription activation for both *ASCL1* and *MYOD1* than homo-effector pairing (for example, dCas9–VP64 paired with MS2–VP64 or dCas9–p65 with MS2–p65) (Fig. 1e). We further explored this concept of domain synergy by introducing the activation domain from human heat-shock factor 1 (HSF1)[21] as a third activation domain, and found that an MS2–p65–HSF1 fusion protein further improved transcriptional activation of *ASCL1* (12%) and *MYOD1* (37%) (Fig. 1f). Additional modifications to the sgRNA as well as Cas9 protein, including varying the nuclear localization signal (NLS), provided only minor improvements (Extended Data Fig. 1c–e). On the basis of these collective results, we concluded that the combination of sgRNA 2.0, NLS–dCas9–VP64 and MS2–p65–HSF1 comprises the most effective transcription activation system, and designated it synergistic activation mediator (SAM). For simplicity, we will refer to sgRNA 2.0 as sgRNA in subsequent discussions, unless noted otherwise.

Design rules for efficient sgRNAs

To evaluate thoroughly the effectiveness of SAM for activating endogenous gene transcription, we chose 12 genes that were previously found by several groups to be difficult to activate using dCas9–VP64 and individual sgRNA 1.0 guides[8,10,11]. For each gene, we selected 8 sgRNA target sites spread across the proximal promoter between −1,000 bp and the +1 transcription start site (TSS). For 9 out of 12 genes, the maximum level of activation achieved using dCas9–VP64 with any of the 8 sgRNA 1.0 guides was lower than twofold, while the remaining three genes (*ZFP42*, *KLF4* and *IL1B*) were maximally activated between two- and fivefold (Fig. 2a). In contrast, SAM stimulated transcription at least twofold for all genes and more than 15-fold for 8 out of 12 genes. SAM performed consistently better than sgRNA 1.0 + dCas9–VP64 for all 96 guides, with a median gain of 105-fold greater upregulation across all 12 genes (activation by SAM divided by activation by sgRNA 1.0 + dCas9–VP64).

Previous studies have demonstrated that the poor activation efficiency of single sgRNAs can be overcome by combining dCas9–VP64 with a pool of sgRNAs tiling the proximal promoter region of the target gene[9–11]. Therefore we compared the single sgRNA activation efficiency of SAM against dCas9–VP64 combined with a pool of 8 sgRNA 1.0 guides, all targeting the same gene. For 10 out of 12 genes, SAM with a single sgRNA performed more robustly than dCas9–VP64 with pools

©2015 Macmillan Publishers Limited. All rights reserved

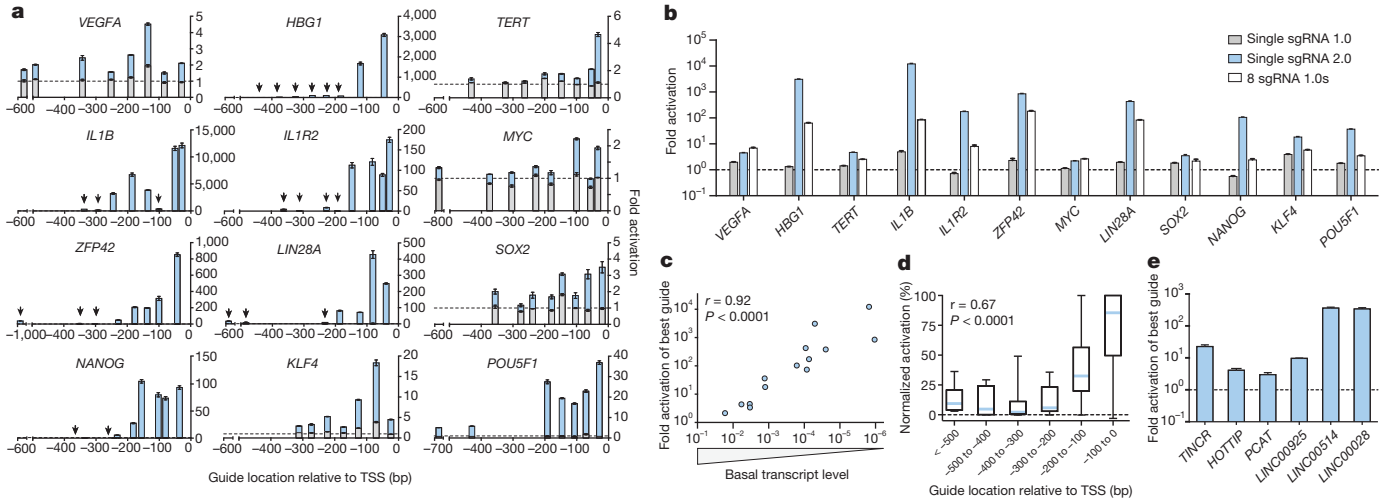


Figure 2 | Characterization of SAM-mediated gene and lincRNA activation and derivation of selection rules for efficient sgRNAs. **a**, Fold activation of 12 different genes plotted against the sgRNA location. sgRNA 1.0 with dCas9–VP64 (grey), sgRNA 2.0 with dCas9–VP64 and MS2–p65–HSF1 (blue). **b**, Comparison of activation efficiency of 12 target genes: dCas9–VP64 and a single sgRNA 1.0; dCas9–VP64 with a single sgRNA 2.0 and MS2–p65–HSF1, and dCas9–VP64 with a mixture of 8 sgRNA 1.0s. **c**, Efficiency of target gene activation as a function of baseline expression levels. **d**, Correlation of gene activation efficiency with sgRNA targeting position. Activation efficiency of each sgRNA for the same target gene is normalized against the highest-activating sgRNA. **e**, Fold activation of six lincRNA transcripts by SAM (best sgRNA out of 8 tested). All experiments were performed in 293FT cells. All values are mean ± s.e.m. with $n = 3$ biological replicates.

of 8 sgRNA 1.0 guides (Fig. 2b). Additionally, inclusion of a third activation domain, MS2–p65–HSF1 or MS2–p65–MyoD1, outperformed MS2–p65 alone (Extended Data Fig. 2a).

Next, we sought to determine factors that contribute to inter- and intragenic variability of activation efficiency by different sgRNAs. For inter-gene variability, differences in activation magnitudes could be due to epigenetic factors and/or variation in basal transcription levels. We were thus interested in correlating basal transcription with the level of transcription activation achieved using SAM. Using the relative transcriptional levels of target genes in control samples, we observed a highly significant correlation between the inverse of basal transcript level and the fold upregulation achieved using SAM (Fig. 2c; $r = 0.94$, $P < 0.0001$). This suggests that the basal expression level of each gene largely determines the level of activation.

To study the intragenic variability of SAM activity, we aggregated the activation data for all 96 guides and found the distance between the guide RNA target site and the TSS to be the strongest predictor of activation efficiency (Fig. 2d; $r = 0.67$, $P < 0.0001$). For all genes, the highest levels of activation were consistently achieved by targeting within the −200 bp to +1 bp window. This simple design guideline can inform the selection of efficient sgRNAs for gene activation.

We also sought to test whether SAM is able to activate non-coding elements in addition to protein-coding genes. We chose a diverse set of 6 lincRNAs and found that SAM mediated significant upregulation of each target (Fig. 2e), with MS2–p65–HSF1 or MS2–p65–MyoD1 leading to the highest levels of activation for each lincRNA ($P < 0.01$) (Extended Data Fig. 3). We also examined the effect of the most potent sgRNA for each lincRNA on the transcription of the nearest coding gene. Of all sgRNAs tested, only the sgRNA targeting *HOTTIP*—the only sgRNA located within 500 bp of the neighbouring gene's TSS—led to significant activation of its neighbour (Extended Data Fig. 2b).

Multiplex gene activation

The ability to simultaneously modulate gene expression at multiple loci would allow for a better understanding of complex genetic and regulatory networks. Using sets of two to ten sgRNAs, we observed successful activation of all target genes (>twofold) within all sgRNA combinations (Fig. 3a, b and Extended Data Fig. 4). As expected, most genes (excluding *IL1R2*) exhibited a decrease in the amount of upregulation achieved when concurrently targeted with 9 other genes. Interestingly, the relative activation levels of each gene changed between multiplex activation and single-gene activation experiments (Fig. 3a, b).

We asked if reduced activation of targets during multiplexing was due to the reduced amounts of sgRNA or SAM protein components. Surprisingly, diluting the sgRNA expression plasmid by tenfold in single-gene activation experiments did not reduce activation for all genes (Fig. 3c). We found that genes whose levels of activation are reduced upon sgRNA dilution also exhibited dampened levels of activation when multiplexed (Fig. 3d; $r = 0.94$, $P < 0.001$). In contrast, the activation efficiency of SAM was generally unperturbed by dilution of its protein components (dCas9–VP64 and MS2–p65–HSF1) (Extended Data Fig. 5). Activation efficiency remained stable particularly when all three components were diluted, retaining on average 90% activation efficiency across a 50-fold dilution range (Extended Data Fig. 5). This finding

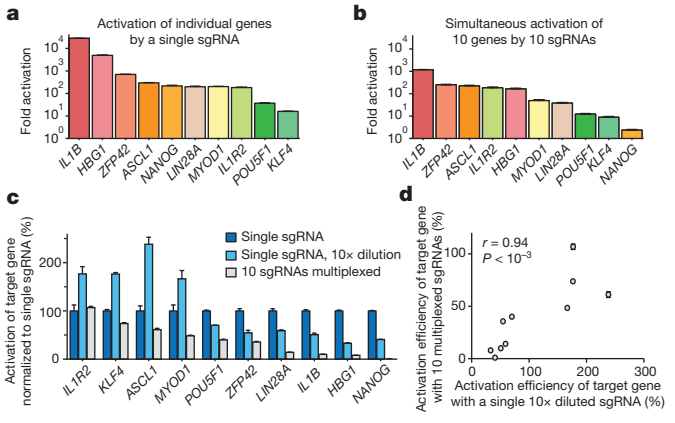


Figure 3 | Simultaneous activation of endogenous genes using multiplexed sgRNA expression. **a**, Activation of individual genes by single sgRNAs with dCas9–VP64 and MS2–p65–HSF1. **b**, Simultaneous activation of the same ten genes using a mixture of ten sgRNAs, each targeting a different gene. **c**, Effect of sgRNA dilution on gene activation efficiency. **d**, Correlation between the activation efficiency of a single tenfold diluted sgRNA and that of the same sgRNA delivered within a mixture of ten different-gene targeting sgRNAs. All values are mean ± s.e.m. with $n = 3$ biological replicates.

©2015 Macmillan Publishers Limited. All rights reserved

was particularly promising for genome-scale pooled screening applications, which rely on single-copy lentiviral integration.

Specificity of SAM-mediated activation

An important consideration for SAM use is its targeting specificity. Recent analysis of genome-wide dCas9-binding revealed significant concentration-dependent off-target binding[22], yet its effect on the specificity of transcription modulation remains unclear. To assess SAM specificity, we chose *HBG1/2* as our target gene, reasoning that globin genes would have few downstream targets that could confound our specificity analysis. We found that SAM specifically activated both *HBG1* and *HBG2* isoforms ($P < 0.05$, *t*-test after 0.01 false discovery rate (FDR) correction), which share the same TSS (Fig. 4 and Extended Data Fig. 6). We also tested two additional non-targeting sgRNAs with guide sequences that do not share perfect homology with the human genome. For all sgRNAs, we found only two additional genes, *S100A1* and *CYB5R2*, to be differentially expressed ($P < 0.05$, *t*-test after 0.01 FDR correction for multiple hypothesis testing) compared with green fluorescent protein (GFP)-expressing control (Extended Data Fig. 6) for both non-targeting guides. These results suggest that SAM-mediated gene activation is specific with minimal off-target activity.

Genome-scale gene activation screen

The ability to activate target genes using individual sgRNAs greatly facilitates the development of pooled, genome-scale transcriptional activation screening. To develop a SAM-based screening system, we generated lentiviral expression vectors that are able to drive robust transcription activation at low multiplicity of infection (MOI) (Extended Data Fig. 7a, b). Using this lentiviral system, we generated a genome-scale sgRNA library consisting of 70,290 guides, targeting every coding isoform from the RefSeq database (23,430 isoforms). For each gene, three sgRNAs were chosen to target sites within 200 bp upstream of the TSS, which was previously determined to provide more efficient activation (Fig. 2d and Fig. 5a).

Previously we applied genome-scale CRISPR knockout (GeCKO) screening[3] in A375 (BRAF(V600E)) melanoma cells to identify LOF mutations capable of mediating resistance against the BRAF inhibitor PLX-4720. Here we sought to use the new SAM sgRNA library to identify a complementary set of GOF changes that can confer BRAF inhibitor resistance (Fig. 5a).

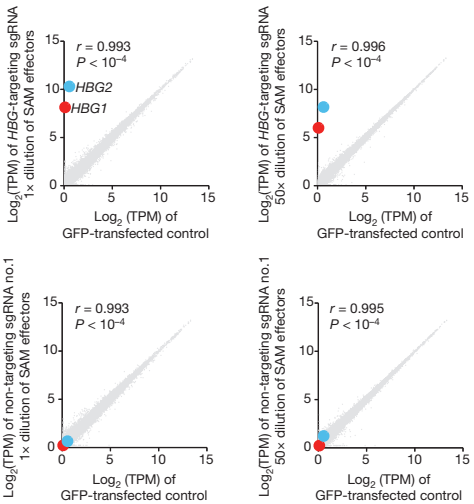


Figure 4 | Evaluation of SAM specificity. Expression levels in $\log_2$(transcripts per million (TPM)) values of all detected genes in RNA-seq libraries of GFP-transfected controls (*x* axis of all graphs) compared to (from left to right): SAM targeting *HBG1/2* genes in 1× dilution and 50× dilution, non-targeting control sgRNAs in 1× dilution and 50× dilution (*y* axis). Marked are the two most statistically significant differentially expressed genes (*t*-test *q* value < 0.05 with FDR correction): Red, *HBG1*; blue, *HGB2*. The average from $n = 3$ biological replicates is shown.

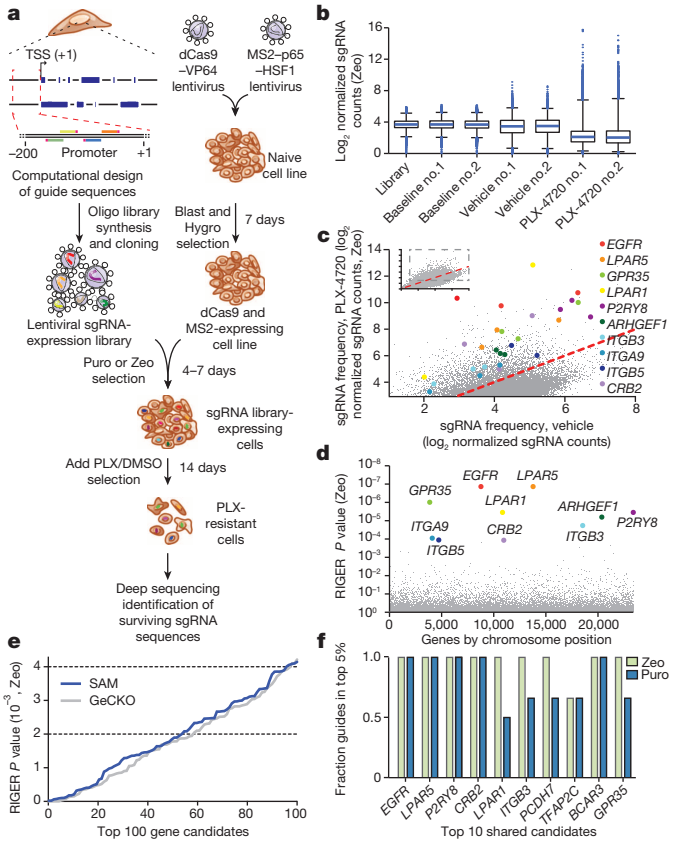


Figure 5 | Genome-scale gene activation screening identifies mediators of BRAF inhibitor resistance. **a**, Flow-chart of transcription activation screening using SAM. Blast, blasticidin; Hygro, hygromycin; Puro, puromycin; PLX, PLX-4720; Zeo, zeocin. **b**, Box-plot showing the distribution of sgRNA frequencies post lentiviral transduction for baseline (day 3), vehicle (day 21), and PLX-4720 (day 21) conditions from $n = 2$ infection replicates. **c**, Scatterplot showing enrichment of specific sgRNAs after PLX-4720 treatment. **d**, Identification of top candidate genes using the RIGER *P* value analysis based on the average of both infection replicates. **e**, Comparison of RIGER *P* values for the top 100 hits from SAM and GeCKO[3] PLX-4720 resistance screens. **f**, Consistency of sgRNAs for top screening hits. Fraction of unique sgRNAs targeting each gene that are in the top 5% of all sgRNAs is plotted.

We found that at 14 days post drug treatment, the sgRNA distribution was significantly different between cells treated with PLX-4720 and with vehicle, with the majority of sgRNAs exhibiting a reduced representation and a small set of guides showing high enrichment in PLX-4720-treated cells (Fig. 5b and Extended Data Fig. 7c). For a number of gene targets, multiple sgRNAs targeting the same gene were enriched in PLX-4720-treated cells (Fig. 5c) and the 10 most significant hits were distributed throughout the genome (Fig. 5d and Extended Data Fig. 7d). The significance of the *P* values of our top 100 hits determined by RNAi gene enrichment ranking (RIGER) (Supplementary Tables 1 and 2) was comparable to those observed for GeCKO screening[3] (Fig. 5e). In addition, for the top 10 shared hits between two independent screens (zeocin and puromycin selection for sgRNA expression), the fraction of effectively enriched guides per gene (present in the top 5% of all guides) was very high with 97% for zeocin and 81% for puromycin (89% ± 10.7% overall, compared to 78% ± 27% for the top 10 GECKO hits, Fig. 5f and Extended Data Fig. 7e).

Our screen results highlight a number of gene candidates that both confirm known PLX-4720 resistance pathways and suggest new mechanisms (Extended Data Fig. 7f). First, reactivation of the ERK pathway

©2015 Macmillan Publishers Limited. All rights reserved

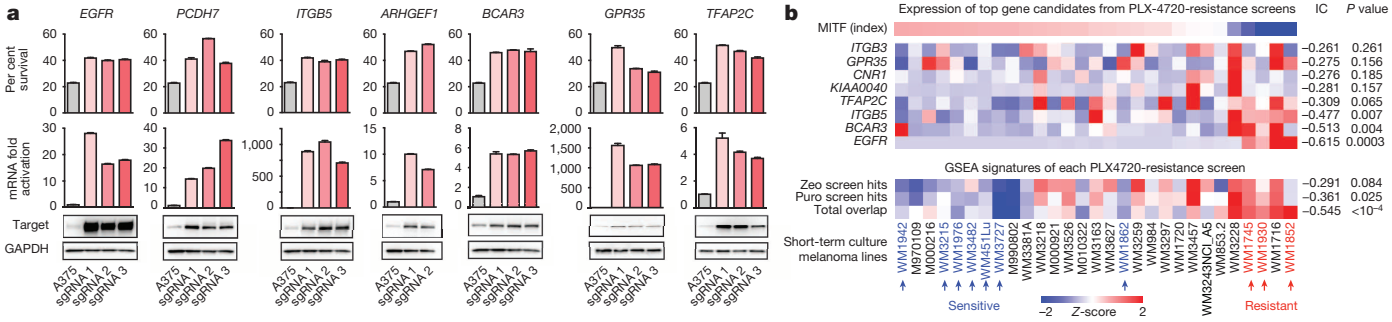


Figure 6 | Validation of top hits from genome-scale gene activation screen for PLX-4720 resistance mediators. **a**, Comparison of PLX-4720 resistance, transcription activation and protein upregulation in A375 cells for top screening hits. **b**, Expression levels of top hits and screen signatures are elevated in the resistant state of short-term BRAF(V600) melanoma cultures (see Methods for signature generation). The subset of samples which were previously tested for PLX-4720 sensitivity and resistance are indicated by blue and red arrows, respectively[39]. IC, information coefficient. All values are mean ± s.e.m. with $n = 4$ biological replicates.

is one of the main known resistance mechanisms[23,24], and two of our screening hits, *BCAR3* and *EGFR*, probably modulate downstream and upstream nodes of this pathway, respectively[25,26]. *EGFR* has been previously validated as a mediator of resistance to PLX-4720 through PI3K-AKT, in addition to ERK[26,27]. These two pathways are thought to be alternative routes of PLX-4720 resistance[24,28,29]. Furthermore, four out of the top 10 hits from our screen belong to the family of G-protein-coupled receptors (GPCRs: *GPR35*, *LPAR1*, *LPAR5* and *P2RY8*), which emerged as the top-ranked protein class conferring resistance to multiple MAP kinase inhibitors in melanoma cells in a recent screen using cDNA overexpression[30]. GPCRs signal through multiple downstream pathways including ERK and AKT as well as cAMP-PKA[31,32]. The final class of protein candidates from our screen belongs to the ITG receptor family, which is thought to interact with RTK and activate both ERK and AKT pathways[33,34].

To verify the results from the PLX-4720 resistance screen, we validated each of the top 13 genes. All sgRNAs from the screen that targeted these 13 genes conferred PLX-4720 resistance when individually expressed in A375 along with SAM (Fig. 6a and Extended Data Fig. 8a). We also verified that SAM was able to facilitate robust increase in target transcript (Fig. 6a and Extended Data Fig. 8b) and protein levels (Fig. 6a). Since 5 of our top candidates from the pooled SAM screen overlapped with hits from a previously conducted arrayed cDNA screen[30] (Extended Data Fig. 8c), we compared the relative efficacy of cDNA overexpression with SAM-mediated transcription activation. Interestingly, for these 5 targets, SAM led to at least similar levels of PLX-4720 resistance when compared with corresponding cDNA overexpression conditions (Extended Data Fig. 8a), despite cDNA leading to higher transcript levels (Extended Data Fig. 8d). Furthermore, we found that, for most genes, the levels of PLX-4720 resistance mediated by all three sgRNAs were comparable (Extended Data Fig. 8e).

In addition to validating our top screening hits through individual sgRNA or cDNA overexpression, we analysed the expression profile of our screening hits using four different data sets (CCLE[35,37], TCGA: https://tcga-data.nci.nih.gov/tcga/, short-term melanoma cultures[36], and pre/post treatment patient samples[38]). As shown previously[39], a distinct transcriptional state defines BRAF-inhibition sensitive and resistant states as described by activation of endogenous MITF/associated markers (for example, PMEL) and NF-κB-pathway activity/associated markers (for example, AXL), respectively (Fig. 6b and Extended Data Fig. 9b). Based on short-term melanoma data[36,39], we found that the expression of our top screening hits was significantly increased in the resistant state. Correspondingly, a gene expression signature based on the top screening hits (see Methods) correlated with a BRAF-inhibitor resistance state as defined previously[39] (Fig. 6b; total overlap, $P < 0.0001$). Further analysis performed using the CCLE, TCGA and pre/post treatment data set also revealed similar correlations (Extended Data Fig. 9).

Discussion

In summary, we have taken a structure-guided approach to design a dCas9-based transcription activation system for achieving robust, single sgRNA-mediated gene upregulation. By engineering the sgRNA to incorporate protein-interacting aptamers, we assembled a synthetic transcription activation complex consisting of multiple distinct effector domains modelled after natural transcription activation processes. Here we have shown that the SAM system is robust, specific, and can facilitate genome-scale gain-of-function screening when combined with a compact pooled sgRNA library. Our SAM-mediated screens exhibited a high degree of consistency and validation, with >80% effectively enriched guides per gene hit, and 100% validation of the top 10 hits.

Future engineering of the Cas9 complex based on structural information[14,40] will further expand the Cas9 toolbox[41]. Additional developments of the SAM system may be able to take advantage of the modularity and customizability of the sgRNA scaffold to establish a series of sgRNA scaffolds bearing different aptamers for recruiting distinct types of effectors in an orthogonal manner. For instance, replacement of the MS2 stem loops with PP7-interacting stem loops may be used to recruit repressive elements, potentially enabling multiplexed bidirectional transcriptional control.

Although we have taken initial steps towards defining selection rules for potent sgRNAs, future studies will reveal additional selection criteria that are critical for guide efficacy, such as sequence-intrinsic properties (Extended Data Fig. 10a–d). Applications of dCas9-based transcription modulators in positive and negative selection screens (Extended Data Fig. 10e, f)[42] will enable the dissection of many types of genetic elements, ranging from protein-coding genes to non-coding lincRNA elements. Furthermore, combining wild-type Cas9-mediated genome modifications with SAM-mediated recruitment of epigenetic modifiers will constitute powerful approaches for studying genome organization and regulation in diverse biological processes.

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 23 September; accepted 3 December 2014.
Published online 10 December 2014.

1. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428,** 431–437 (2004).
2. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303,** 832–835 (2004).
3. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).
4. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–84 (2014).
5. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. III. Toward controlling gene expression at will: specific regulation of the *erbB-2/HER-2* promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA* **95,** 14628–14633 (1998).

©2015 Macmillan Publishers Limited. All rights reserved

6. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnol.* **29,** 149–153 (2011).
7. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).
8. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500,** 472–476 (2013).
9. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10,** 977–979 (2013).
10. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10,** 973–976 (2013).
11. Mali, P, *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31,** 833–838 (2013).
12. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
13. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109,** E2579–E2586 (2012).
14. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).
15. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12,** 595–600 (1993).
16. Lemon, B. & Tjian, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14,** 2551–2569 (2000).
17. van Essen, D., Engist, B., Natoli, G. & Saccani, S. Two modes of transcriptional activation at native promoters by NF-κB p65. *PLoS Biol.* **7,** e73 (2009).
18. Kretzschmar, M., Kaiser, K., Lottspeich, F. & Meisterernst, M. A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. *Cell* **78,** 525–534 (1994).
19. Ikeda, K., Stuehler, T. & Meisterernst, M. The H1 and H2 regions of the activation domain of herpes simplex virion protein 16 stimulate transcription through distinct molecular mechanisms. Genes Cells **7,** 49–58 (2002).
20. Neely, K. E. *et al.* Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Mol. Cell* **4,** 649–655 (1999).
21. Marinho, H. S., Real, C., Cyrne, L., Soares, H. & Antunes, F. Hydrogen peroxide sensing, signaling and regulation of transcription factors. *Redox Biol.* **2,** 535–562 (2014).
22. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnol.* **32,** 670–676 (2014).
23. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468,** 968–972 (2010).
24. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468,** 973–977 (2010).
25. Musgrove, E. A. & Sutherland, R. L. Biological determinants of endocrine resistance in breast cancer. *Nature Rev. Cancer* **9,** 631–643 (2009).
26. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483,** 100–103 (2012).
27. Corcoran, R. B. *et al.* EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2,** 227–235 (2012).
28. Villanueva, J. *et al.* Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* **18,** 683–695 (2010).
29. Shi, H., Kong, X., Ribas, A. & Lo, R. S. Combinatorial treatments that overcome PDGFRβ-driven resistance of melanoma cells to ^{V600E}B-RAF inhibition. *Cancer Res.* **71,** 5067–5074 (2011).
30. Johannessen, C. M. *et al.* A melanocyte lineage program confers resistance to MAP kinase pathway inhibition. *Nature* **504,** 138–142 (2013).
31. Dorsam, R. T. & Gutkind, J. S. G-protein-coupled receptors and cancer. *Nature Rev. Cancer* **7,** 79–94 (2007).
32. Lappano, R. & Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Rev. Drug Discov.* **10,** 47–60 (2011).
33. Franke, T. F. PI3K/Akt: getting it right matters. *Oncogene* **27,** 6473–6488 (2008).
34. Desgrosellier, J. S. & Cheresh, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nature Rev. Cancer* **10,** 9–22 (2010).
35. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
36. Lin, W. M. *et al.* Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res.* **68,** 664–673 (2008).
37. Wilks, C. *et al.* The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database* **2014,** bau093 (2014).
38. Rizos, H. *et al.* BRAF inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact. *Clin. Cancer Res.* **20,** 1965–1977 (2014).
39. Konieczkowski, D. J. *et al.* A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer Discov.* **4,** 816–827 (2014).
40. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).
41. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).
42. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159,** 647–661 (2014).

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

Acknowledgements We would like to thank S. Shehata, K. Zheng, C. Johannessen, L. Garraway, O. Shalem and members of the Zhang laboratory for assistance and helpful discussions. O.O.A. is supported by a NSF Graduate Research Fellowship, J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship, H.N. is supported by PRESTO from JST and Grant-in-Aid for Scientific Research (B) from JSPS, O.N. is supported by the CREST program and JST, and F.Z. is supported by the NIMH (DP1-MH100706), the NINDS (R01-NS07312401), NSF, the Keck, Searle Scholars, Klingenstein, Vallee, and Simons Foundations, and Bob Metcalfe. CRISPR reagents are available to the academic community through Addgene, and associated protocols, support forum and computational tools are available via the Zhang laboratory website (http://www.genome-engineering.org).

Author Contributions S.K. and F.Z. conceived the project. S.K., M.D.B., A.E.T. and F.Z. designed the experiments. S.K., M.D.B., A.E.T., C.B., P.D.H. and J.J. performed experiments and analysed data. H.N. and O.N. helped with structural interpretation. N.H. performed the RNA-seq analysis. J.S.G. performed the depletion guide efficacy analysis. O.O.A. performed the analysis of clinical data sets. S.K., A.E.T., P.D.H. and F.Z. wrote the paper with help from all authors.

Author Information All reagents described in this manuscript have been deposited with Addgene (plasmid IDs 61422-61427 for SAM component plasmid and 61597 for the human SAM guide RNA library). RNA-seq data are available at BioProject under accession number PRJNA269048. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.Z. (zhang@broadinstitute.org).

©2015 Macmillan Publishers Limited. All rights reserved

METHODS

Sequences. DNA sequences for SAM components and sgRNA scaffolds are provided in Supplementary Sequences. sgRNA target sequences for characterization and optimization of SAM are listed in Supplementary Table 4.

Transient transfection experiments. Neuro-2a cells (Sigma-Aldrich) were grown in media containing 1:1 ratio of OptiMEM (Life Technologies) to high-glucose DMEM with GlutaMAX and sodium pyruvate (Life Technologies) supplemented with 5% HyClone heat-inactivated FBS (Thermo Scientific), 1% penicillin/streptomycin (Life Technologies), and passaged at 1:5 every 2 days.

HEK293FT cells (Life Technologies) were maintained in high-glucose DMEM with GlutaMAX and sodium pyruvate (Life Technologies) supplemented with 10% heat-inactivated characterized HyClone fetal bovine serum (Thermo Scientific) and 1% penicillin/streptomycin (Life Technologies). Cells were passaged daily at a ratio 1:2 or 1:2.5. For gene activation experiments, 20,000 HEK293FT cells per well were plated in 100 µl media in poly-D-lysine-coated 96-well plates (BD BioSciences). 24 h after plating, cells were transfected with a 1:1:1 mass ratio of sgRNA plasmid with gene-specific targeting sequence or pUC19 control plasmid:MS2-effector plasmid or pUC19:dCas9 plasmid, dCas9-effector plasmid or pUC19.

A total plasmid mass of 0.3 µg per well was transfected using 0.6 µl per well of Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Culture medium was changed 5 h after transfection. 48 h after transfection, cell lysis and reverse transcription were performed using a Cells-to-Ct kit (Life Technologies). Relative RNA expression levels were quantified by reverse transcription and quantitative PCR (qPCR) using TaqMan qPCR probes (Life Technologies, Supplementary Table 5) and Fast Advanced Master Mix (Life Technologies). qPCR was carried out in 5 µl multiplexed reactions and 384-well format using a LightCycler 480 Instrument II. Data was analysed by the $\Delta\Delta C_t$ method: target C_t values (FAM dye) were normalized to GAPDH C_t values (VIC dye), and fold changes in target gene expression were determined by comparing to GFP-transfected experimental controls.

Lentivirus production. HEK293T cells (Life Technologies) were cultured as described above for HEK293FT cells. 1 day before transfection, cells were seeded at ~40% confluency (12 × T225 flasks for library scale production, 1 × T25 flask for individual guide production). Cells were transfected the next day at ~80–90% confluency. For each flask, 10 µg of plasmid containing the vector of interest, 10 µg of pMD2.G and 15 µg of psPAX2 (Addgene) were transfected using 100 µl of Lipofectamine 2000 and 200 µl Plus Reagent (Life Technologies). 5 h after transfection the media was changed. Virus supernatant was harvested 48 h post-transfection, filtered with a 0.45-µm PVDF filter (Millipore), aliquoted, and stored at −80 °C.

Lentiviral transduction. A375 cells (Sigma) were cultured in RPMI 1640 (Life Technologies) supplemented with 10% FBS (Seradigm) and 1% penicillin/streptomycin (Life Technologies) and passaged every other day at a 1:4 ratio. Cells were transduced with lentivirus via spinfection in 12-well plates. 3×10^6 cells in 2 ml of media supplemented with 8 µg ml^{-1} polybrene (Sigma) were added to each well, supplemented with lentiviral supernatant and centrifuged for 2 h at 1,000*g*. 24 h after spinfection, cells were detached with TrypLE (Life Technologies) and counted. Cells were replated at low density (7.5×10^6 cells per T225 Flask) and a selection agent was added either immediately (zeocin, blasticidin and hygromycin, all Life Technologies) or 3 h after plating (puromycin). Concentrations for selection agents we determined using a kill curve: 0.5 µg ml^{-1} puromycin, 200 µg ml^{-1} zeocin, 10 µg ml^{-1} blasticidin, and 300 µg ml^{-1} hygromycin. Media was refreshed on day 2 and cells were passaged every other day starting on day 4 after replating. The duration of selection was 4 days for puromycin and 7 days for zeocin, hygromycin and blasticidin. Lentiviral titres were determined by spinfecting cells with 6 different volumes of lentivirus ranging from 0 to 600 µl and counting the number of surviving cells after a complete selection (3–6 days).

Design and cloning of SAM library. RefSeq coding gene isoforms with a unique TSS (total of 23,430 isoforms) were targeted with three guides each for a total library of 70,290 guides (Supplementary Table 6). Guides were designed to target the first 200 bp upstream of each TSS and subsequently filtered for GC content >25% and minimal overlap of the target sequence. After filtering, the remaining guides were scored according to predicted off-target matches as described previously[43], and three guides with the best off-target scores were selected. Cloning of the SAM sgRNA libraries was performed as previously described[3] with a minimum representation of 100 transformed colonies per guide.

Depletion and PLX-4720 screen. A375 cells stably integrated with SAM Cas9 and effector components were transduced with SAM sgRNA libraries as described above at an MOI of 0.2, with a minimal representation of 500 transduced cells per guide. Cells were maintained at >500 cells per guide during subsequent passaging. At 7 days post infection (DPI) (complete selection, see above), cells were split into vehicle (DMSO) and PLX-4720 conditions (2 µM PLX-4720 dissolved in DMSO, Selleckchem). Cells were passaged every 2 days for a total of 14 days of drug treatment. >500 cells per guide were harvested as a baseline at 3 DPI (4 days before treatment) and at 21 DPI (after 14 days of treatment) for gDNA extraction. Genomic DNA was extracted using the Zymo Quick-gDNA midi kit (Zymo Research). PCR of the virally integrated guides was performed on gDNA at the equivalent of >500 cells per guide in 96 parallel reactions using NEBnext High Fidelity 2X Master Mix (New England Biolabs) in a single-step reaction of 22 cycles. Primers are listed here: forward primer, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*NNNNNNNN*(1–10-bp stagger)GCTTTATATATCTTGTGGAAAGGACGAAACACC, 8 bp barcode indicated in italic; reverse primer, CAAGCAGAAGACGGCATACGAGAT*NNNNNNNN*GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCTGCCAAGTTGATAACGGACTAGCCTT, 8-bp index read barcode indicated in italic.

PCR products from all 96 reactions were pooled, purified using Zymo-Spin V with Reservoir (Zymo research) and gel extracted using the Zymoclean Gel DNA Recovery Kit (Zymo research). Resulting libraries were deep-sequenced on Illumina MiSeq and HiSeq platforms with a total coverage of >35 million reads passing filter per library.

NGS and screen hits analysis. NGS data were de-multiplexed using unique index reads. Guide counts (Supplementary Table 7) were determined based on perfectly matched sequencing reads only. For each condition, guide counts were normalized to the total number of counts per condition, and $\log_2$ counts were calculated based on these values. Ratios of counts between conditions were calculated as $\log_2((\text{count } 1 + 1)/(\text{count } 2 + 1))$ based on normalized counts.

RIGER analysis was performed using GENE-E based on the normalized day 14 $\log_2$ ratios (PLX-4720/DMSO) averaged over two independent infection replicates. All RIGER analysis used the Kolmogorov–Smirnov method as described previously[44], except for Fig. 6c, where the weighted average method was used in order to enable comparison to GeCKO values determined by that method.

Gene expression and pharmacological validation analysis. Gene expression data (CCLE, TCGA, short-term cultures, patient melanoma biopsies) and pharmacological data (CCLE, short-term cultures) were analysed to better understand the biological relevance of the top gene hits from the SAM screens. In the CCLE data set[35], gene expression data (RNA-sequencing, GCHub: https://cghub.ucsc.edu/datasets/ccle.html) and pharmacological data (activity area for MAPK pathway inhibitors) from BRAF(V600) mutant melanoma cell lines were used to compute the association between PLX-4720 resistance and the gene expression of each of the top hits. Additionally, gene expression signatures comprised of the top hits were generated using single-sample Gene Set Enrichment Analysis (ssGSEA)[45], and the associations between PLX-4720 resistance and these signatures were computed.

Gene expression data (Affymetrix GeneChip HT-HGU133) and PLX-4720 pharmacological data (GI_{50}, half-maximal growth inhibition concentration; only for a subset of the samples) from short term melanoma cultures (STC)[36] were also used for plotting the gene expression of top hits and their ssGSEA signature scores. Expression data for the STC samples were collapsed to maximum probe value per gene and pre-processed using robust spline normalization.

Gene expression (RNA-sequencing) and genotyping data were collected from 113 BRAF(V600)-mutant primary and metastatic patient tumours from The Cancer Genome Atlas (https://tcga-data.nci.nih.gov/tcga/) and these data were similarly used for determining the association between resistance and the expression of top hits/ssGSEA signature scores. Because pharmacological data was not available for the STCs (only a subset had PLX-4720 data) and the TCGA melanoma samples, a transcriptional state was plotted using marker genes and signatures[39] in order to identify samples resistant to BRAF-inhibition.

Gene expression data from 13 patients with BRAF(V600E) melanomas[38] was used for analysing the relationship between resistance and the expression of our top hits/ssGSEA signature scores. Because all the post-treatment tumours were resistant and not every sample had a paired on-treatment biopsy, we decided to order the samples by MITF expression in the pre-treatment samples to reflect the original PLX-4720 sensitivity state of the tumours. We then used the expression data in the post-treatment resistant tumours to plot the expression of top hits/ssGSEA signature scores. We also calculated the $\log_2$-fold change between each patient's post/pre paired samples and determined the number of patients that had at least a $\log_2$-fold change of 2 per top screen hit.

Single sample gene set enrichment analysis. While there was a significant association between the overexpression of some of our top individual SAM screen hits and resistance in three external cancer data sets, we sought a more robust scoring system independent of any single gene. Gene expression signatures were generated based on the set of top hits from each of the two SAM screens and for the overlap between them. Using single-sample Gene Set Enrichment analysis (ssGSEA), a score was generated for each sample that represents the enrichment of the SAM screen gene expression signature in that sample and the extent to which those genes are coordinately up- or downregulated. Additionally, signature gene sets from the Molecular Signature Database (MSigDB)[46] were used in order to fully map the

©2015 Macmillan Publishers Limited. All rights reserved

transcriptional BRAF-inhibitor resistant/sensitive states in the short-term culture and TCGA data sets as previously described[39].

Information coefficient for measuring associations in external data sets. 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 described[39]. 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 heatmap (pharmacological data, gene expression, and signature scores).

sgRNA sequence analysis. Depletion for each sgRNA was calculated as the ratio of counts (see "NGS and screen hits analysis") between day 3 and day 21. The sgRNAs corresponding to genes with significant depletion ($P < 0.05$ by RIGER analysis) in sgRNA-Puro and sgRNA-Zeo libraries were selected for analyses. These sgRNAs were analysed for nucleotide occurrence in the sgRNA sequence, distance from TSS, and guide strand relative to transcript orientation. For each variable, the correlation and significance with the sgRNA ratio was calculated by ordinary least squares linear regression.

PLX-4720 survival assay. A375 cells stably integrated with dCas9–VP64 and MS2–p65–HSF1 were transduced with individual guides from the top screening hits of the Zeocin and Puromycin screens (13 genes total, 3 sgRNAs per gene) as well as available cDNA at an MOI of <0.2 as described above. Cells were selected for guide expression with Zeocin (Life Technologies) for 5 days and replated at low density (3×10^3 cells per well in a 96-well plate). A375 cells and A375 cells expressing dCas9–VP64 and MS2–p65–HSF1 were plated as controls. Different concentrations of PLX-4720 (2 μM, 0.5 μM, 0.15 μM) or vehicle (DMSO) were added 3 h after plating. Cells were treated with PLX-4720 for 4 days before cell viability was measured using CellTiter-Glo Luminescent Cell Viability Assay (Promega). For qPCR quantification of target gene upregulation, cells were also plated at 5 DPI (3×10^4 cells per well in a 96-well plate) and harvested for mRNA 24 h after plating.

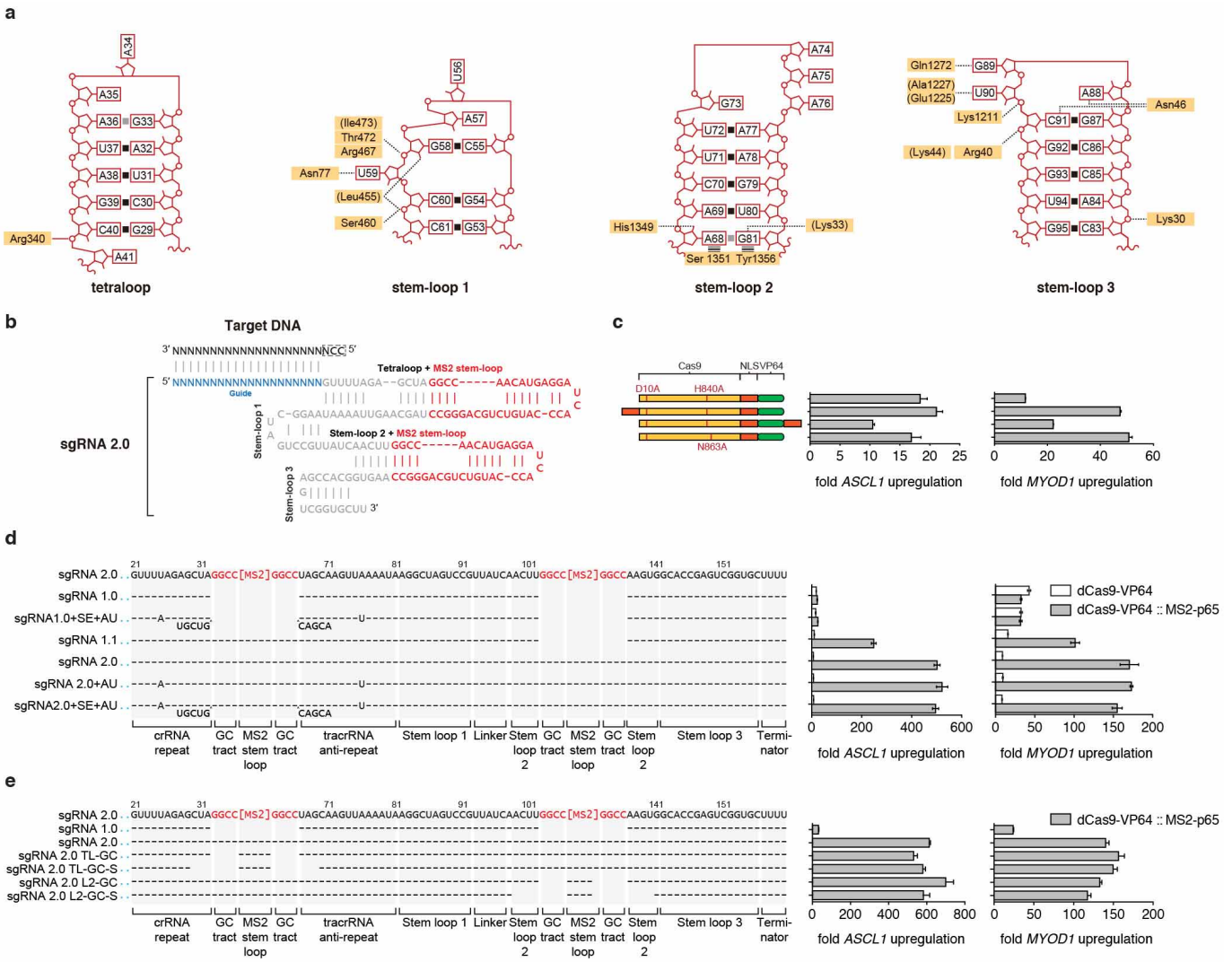
Western blot. Protein lysates were prepared with RIPA lysis buffer (Cell Signaling Technologies) containing a protease inhibitor cocktail (Roche). Samples standardized for protein with the Pierce BCA protein assay (Thermo Scientific) were boiled at 95 °C for 5 min under reducing conditions (except for GPR35 samples, which were incubated at 37 °C for 30 min). After denaturation, samples for probing proteins with lower or higher molecular weight were separated by 10–20% or 4–15% Criterion Tris-HCl gels (Bio-Rad) and electrotransferred onto a 0.2-μm or 0.45-μm polyvinylidene difluoride membrane (Millipore), respectively. Blots were blocked with 5% BLOT-QuickBlocker (VWR) and probed with different primary antibodies (anti-EGFR (rabbit polyclonal, SC-03, Santa Cruz Biotechnology, 1:1,000 dilution), anti-PCDH7 (rabbit polyclonal, HPA011866, Sigma-Aldrich, 1:1,000 dilution), anti-ITGB5 (rabbit polyclonal, SC-14010, Santa Cruz Biotechnology, 1:500 dilution), anti-ARHGEF1 (rabbit polyclonal, 11363-1-AP, Proteintech, 1:5,000 dilution), anti-BCAR3 (rabbit polyclonal, A301-671A, Bethyl Laboratories, 1:2,000 dilution), anti-GPR35 (rabbit polyclonal, 10007660, Cayman Chemical, 1:1,000 dilution), anti-TFAP2C (rabbit polyclonal, 2320, Cell Signaling Technology, 1:1,000 dilution, 2.5% bovine serum albumin, Sigma-Aldrich)) in 2.5% BLOT-QuickBlocker (VWR) unless noted otherwise overnight at 4 °C. Blots were then incubated with secondary antibody HRP-conjugated goat anti-rabbit IgG (7074, Cell Signaling Technology, 1:1,000 dilution) and HRP-conjugated GAPDH (rabbit monoclonal, 3683, Cell Signaling Technology, 1:2,000 dilution) in 2.5% BLOT-QuickBlocker (VWR) for 1 h at room temperature. Proteins with molecular weights similar to GAPDH (GPR35 and TFAP2C) were stripped with Restore Plus Western Blot Stripping Buffer (Thermo Scientific) before probing for GAPDH. SuperSignal West Pico and Femto Chemiluminescent Substrates (Thermo Scientific) were used for detection.

RNA sequencing and data analysis. Samples harvested for RNA sequencing were prepped with TruSeq Stranded mRNA Sample Prep Kit (Illumina) and deep-sequenced on the Illumina MiSeq platform (>9 million reads per condition). Bowtie2[47] index was created based on the human hg19 UCSC genome and known gene transcriptome, and paired-end reads were aligned directly to this index using Bowtie2 with command line options "-q–phred33-quals -n 2 -e 99999999 -l 25 -I 1 -X 1000 -a -m 200 –p 4–chunkmbs 512". Next, RSEM v1.27[48] was run with default parameters on the alignments created by Bowtie2 to estimate expression levels. RSEM's gene level expression estimates (tau) were multiplied by 1,000,000 to obtain transcript per million (TPM) estimates for each gene, and TPM estimates were transformed to log-space by taking $\log_2(\text{TPM}+1)$. The normalization between libraries was tested using an MA plot (mairplot function in Matlab V2013b). Genes were considered detected if their transformed expression level was equal to or above 1 (in $\log_2(\text{TPM}+1)$ scale). All genes detected in at least one library (out of three libraries per condition) were used to construct scatter plots comparing each of the six conditions to the control GFP condition, using the average across biological replicates with >80% alignment to the hg19 UCSC known gene transcriptome ($\log_2(\text{mean}(\text{TPM})+1)$ value per gene).

To find differentially expressed genes, we performed Student's *t*-test on each of the six conditions against the GFP condition. The *t*-test was run on all genes that had expression levels above $\log_2(\text{TPM}+1)>2.5$ in at least two libraries. This threshold was chosen as the minimal threshold for which the number of detected genes across all libraries was constant. Only genes that were significant (*P*-value pass 0.01 FDR correction) and had at least 1.5-fold change were reported and visualized using a heat map.

43. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31,** 827–832 (2013).
44. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105,** 20380–20385 (2008).
45. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).
46. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
48. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
49. Smalley, K. S. Understanding melanoma signaling networks as the basis for molecular targeted therapy. *J. Invest. Dermatol.* **130,** 28–37 (2010).
50. Wong, P. P. *et al.* Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor $p21^{cip}$ (*CDKN1A*). *Mol. Cell. Biol.* **32,** 1633–1644 (2012).
51. Hart, M. J. *et al.* Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by Gα13. *Science* **280,** 2112–2114 (1998).

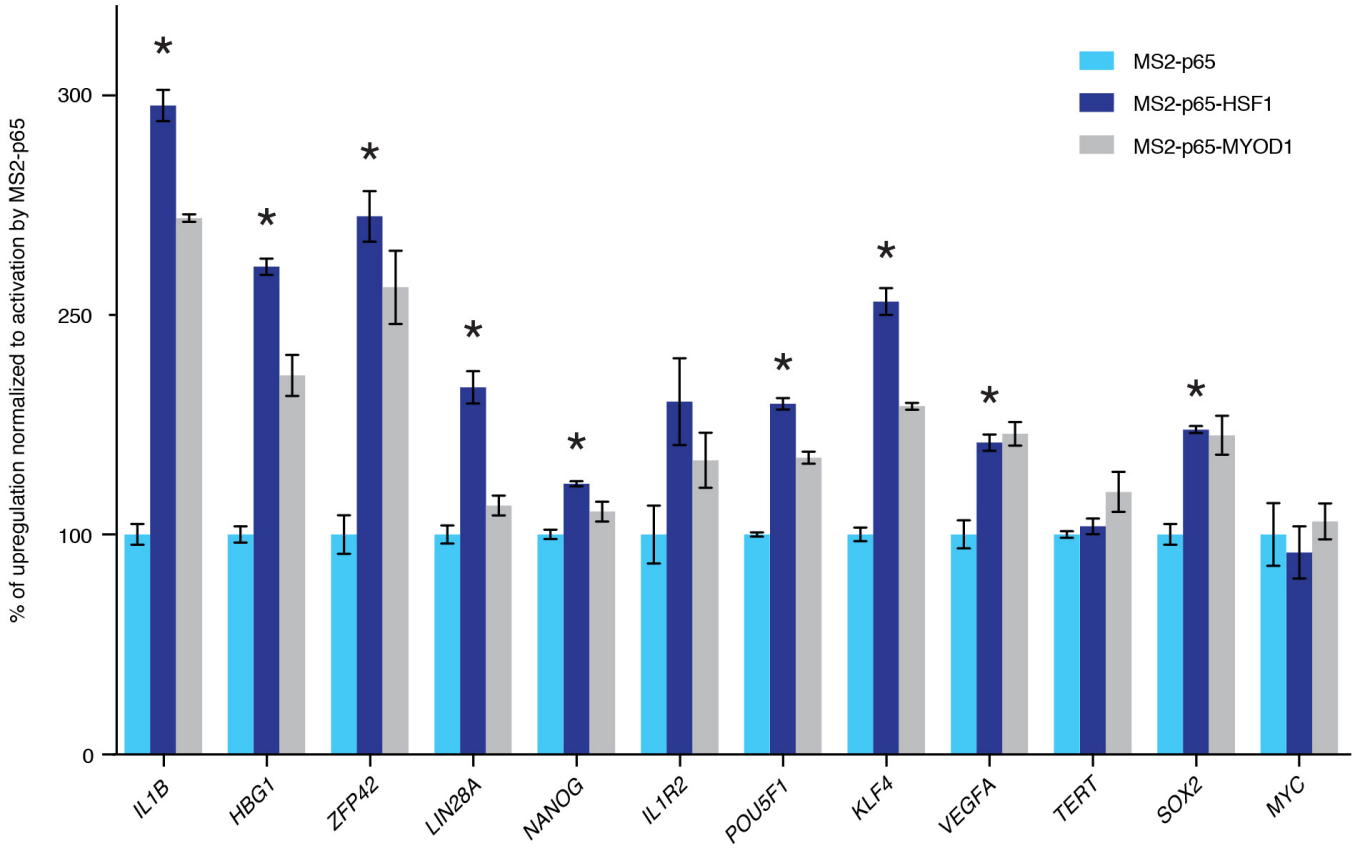
©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 1 | Structure-guided engineering of Cas9 sgRNA. **a**, Schematic of the sgRNA stem loops showing contacts between each stem loop and Cas9. Contacting amino acid residues are highlighted in yellow. Tetraloop and stem loop 2 do not make any contacts with Cas9, whereas stem loops 1 and 3 share extensive contacts with Cas9. **b**, sgRNA 2.0 with MS2 stem loops inserted into the tetraloop and stem loop 2. **c**, Addition of a second NLS or an alternative HNH domain inactivating point mutation in Cas9 improve efficiency of transcription activation for *MYOD1* moderately. **d**, dCas9–VP64 activators exhibit improved performance by recruitment of MS2–p65 to the tetraloop and stem loop 2. Addition of an AU flip or extension in the tetraloop does not increase the effectiveness of dCas9-mediated transcription activation. **e**, Tetraloop and stem loop 2 are amenable to replacement with MS2 stem loops. Base changes from the sgRNA 2.0 scaffold are shown at the respective positions, with dashes indicating unaltered bases and bases below dashes indicating insertions. Deletions are indicated by absence of dashes at respective positions. All figures are $n = 3$ and mean ± s.e.m.

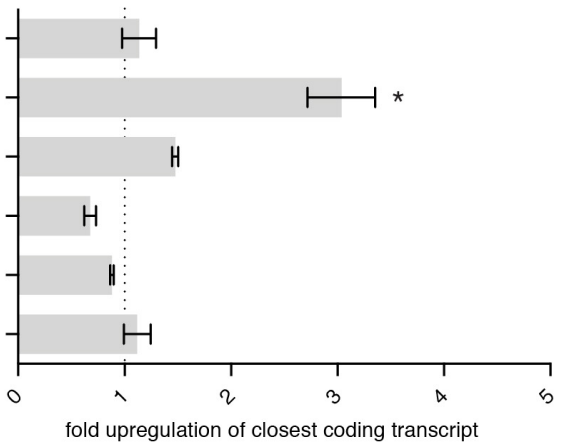
©2015 Macmillan Publishers Limited. All rights reserved

a



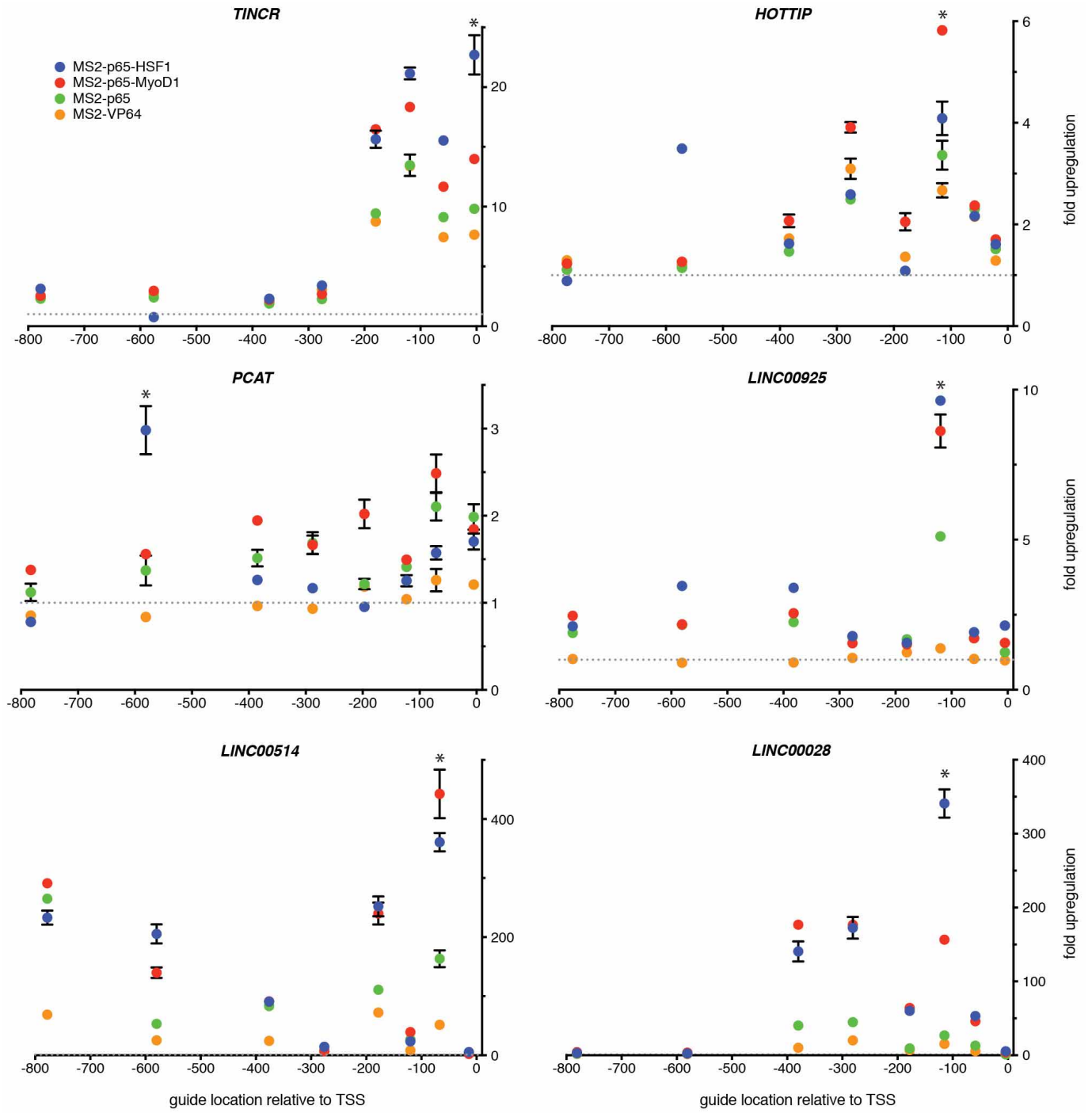
b

sgRNA target (fold activation)	closest coding transcript	transcript orientation	distance (kb)
TINCR (22.7)	***SAFB2***	same	19.0
HOTTIP (4.1)	***HOXA13***	opposed	0.32
PCAT (3.0)	***FAM84B***	opposed	454
LINC00925 (9.6)	***POLG***	opposed	43.2
LINC00514 (361)	***REM1***	same	0.87
LINC00028 (341)	***PKMYT***	opposed	8.5



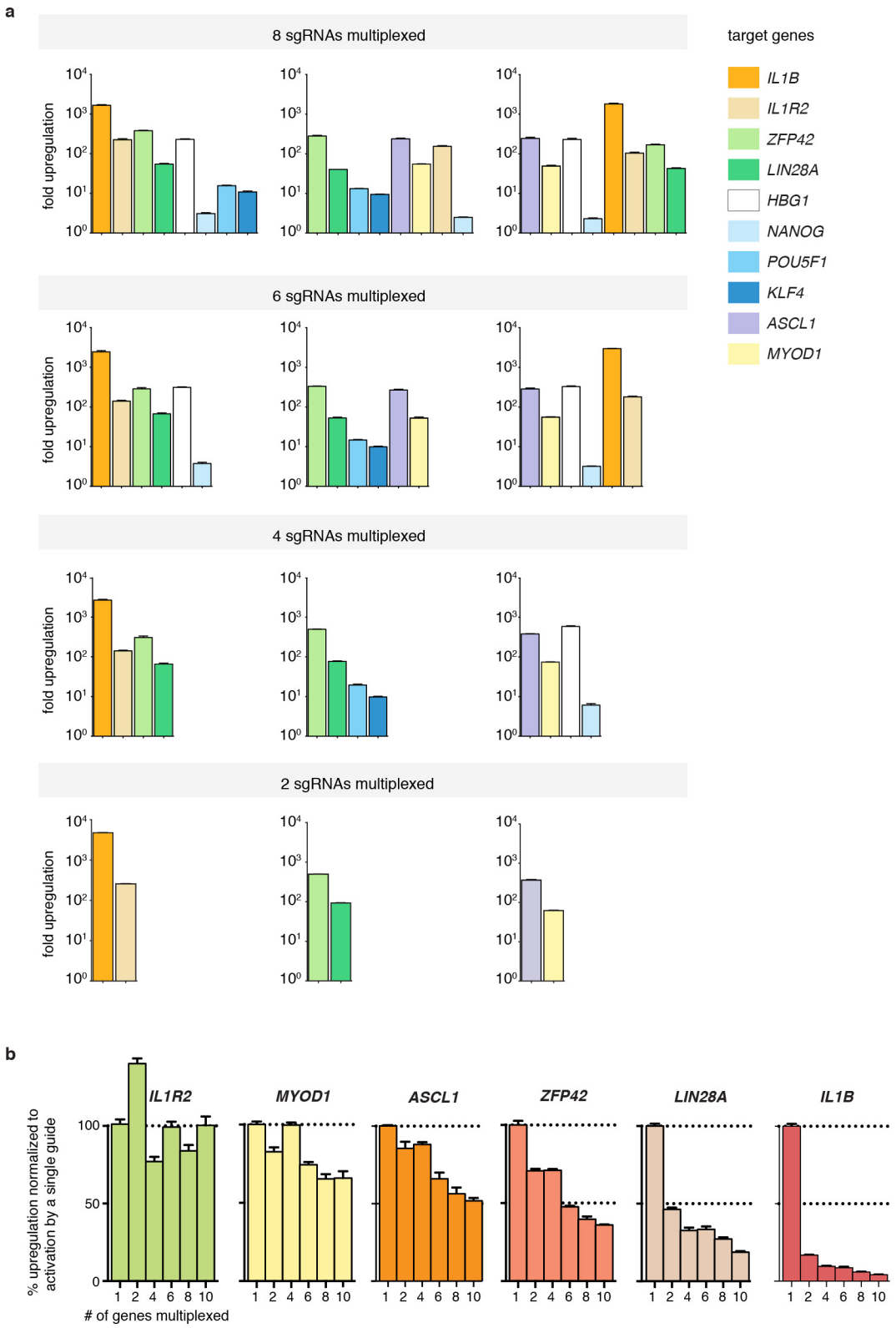
Extended Data Figure 2 | SAM mediates efficient activation of a panel of 12 coding genes and 6 lincRNAs. **a**, Comparison of the activation levels of 12 genes with dCas9–VP64 in combination with MS2–p65, MS2–p65–HSF1, or MS2–p65–MyoD1. MS2–p65–HSF1 mediated significantly higher levels of activation than MS2–p65 alone for 9 out of 12 genes. The best guide out of 8 tested for each gene (Fig. 2a) was used in this experiment. Activation levels for each type of MS2-fusion is presented as a percentage relative to the activation achieved using MS2–p65. **b**, Investigation of transcriptional changes in the closest coding transcripts for SAM-mediated activation of 6 lincRNAs. Direction of the coding transcript relative to the lincRNA and distance between transcription start sites are shown. Only targeting of HOTTIP resulted in a significant change in the levels of the closest coding transcript (HOXA13). The best guide out of 8 tested for each gene (Fig. 2e) in combination with dCas9–VP64 and MS2–p65–HSF1 was used in this experiment. All figures are $n = 3$ and mean ± s.e.m.

©2015 Macmillan Publishers Limited. All rights reserved



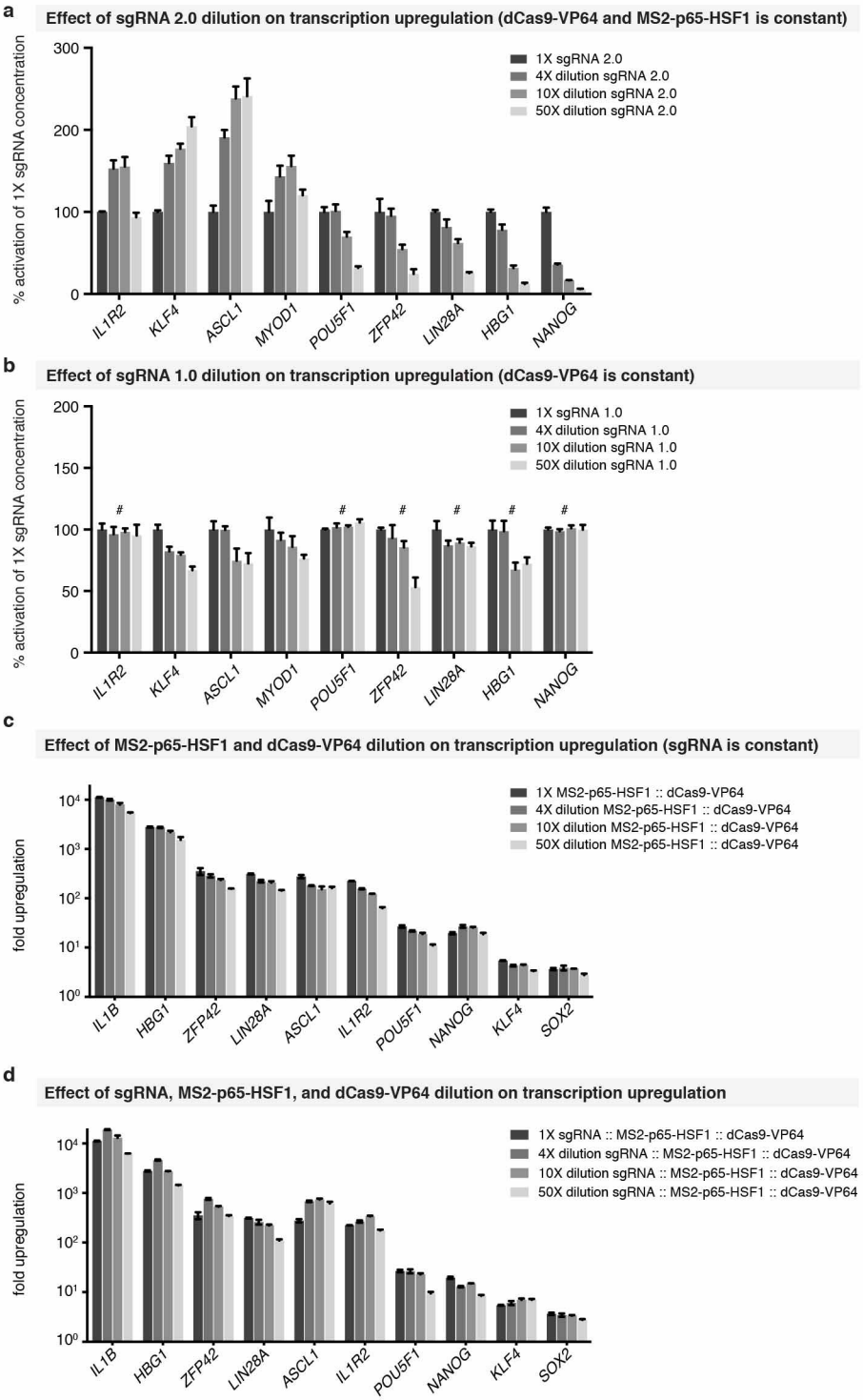
Extended Data Figure 3 | Activation of lincRNAs by SAM. Six lincRNAs, three characterized and three uncharacterized, were targeted using SAM. For each lincRNA, 8 sgRNAs were designed to target the proximal promoter region (+1 to −800 bp from the TSS) with 4 different MS2 activators (MS2–p65–HSF1, MS2–p65–MyoD1, MS2–p65, and MS2–VP64) in combination with dCas9–VP64. MS2 activators with a combination of 2 different domains (MS2–p65–HSF1 or MS2–p65–MyoD1) consistently provided the highest activation for each lincRNA, $*P < 0.01$ for MS2–p65–HSF1 or MS2–p65–MyoD1 versus MS2–p65. $n = 3$ and mean ± s.e.m. is shown.

©2015 Macmillan Publishers Limited. All rights reserved



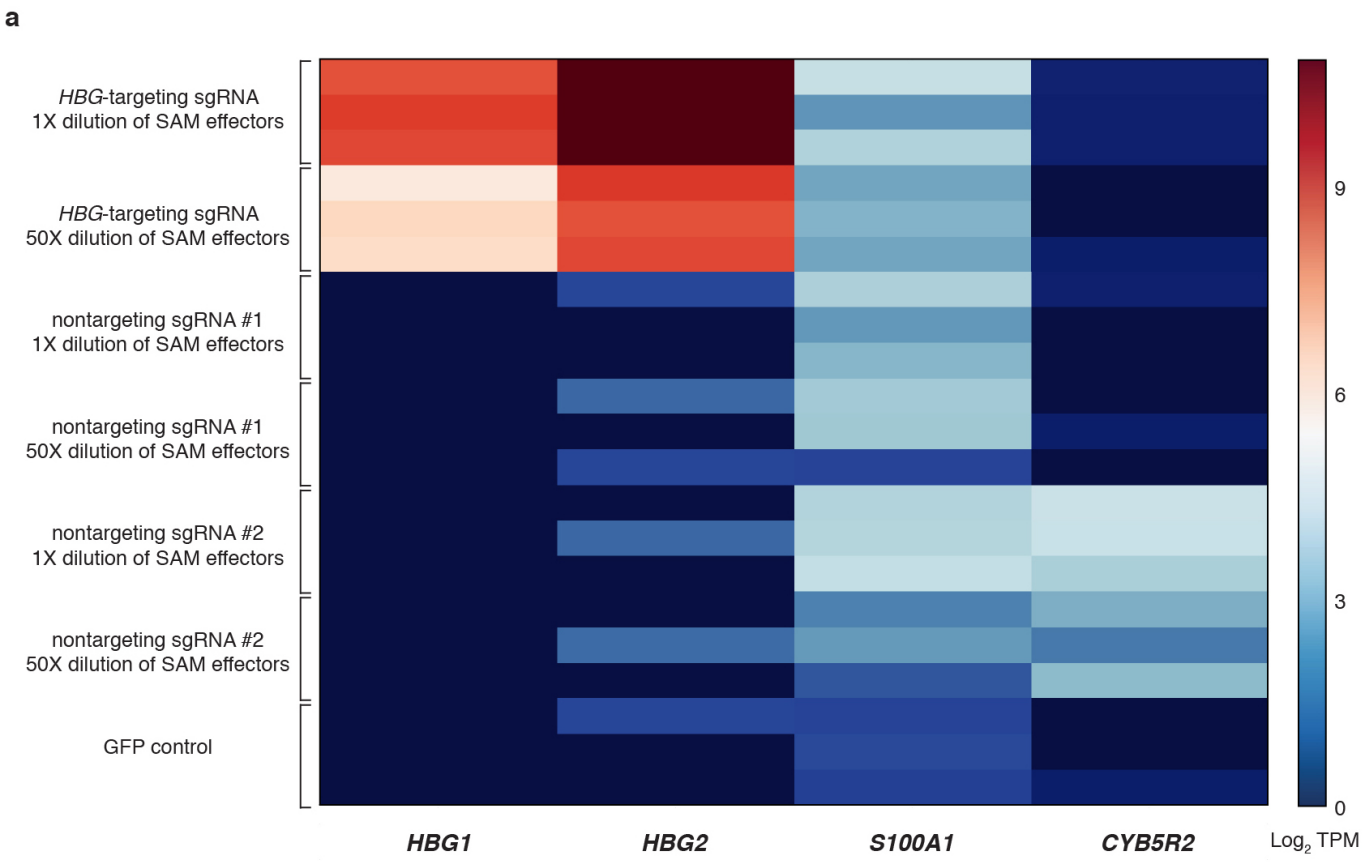
Extended Data Figure 4 | Multiplexed activation using SAM and activation of a panel of 10 genes as a function of SAM component dosage. **a**, Activation of a panel of 10 genes by combinations of 2, 4, 6 or 8 sgRNAs simultaneously. The mean fold upregulation is shown on a $\log_{10}$ scale. MS2–p65–HSF1 and dCas9–VP64 were used in this experiment. **b**, The relative activation efficiency of individual sgRNAs varies depending on the target gene and the degree of multiplexing. $n = 3$ and mean ± s.e.m. is shown.

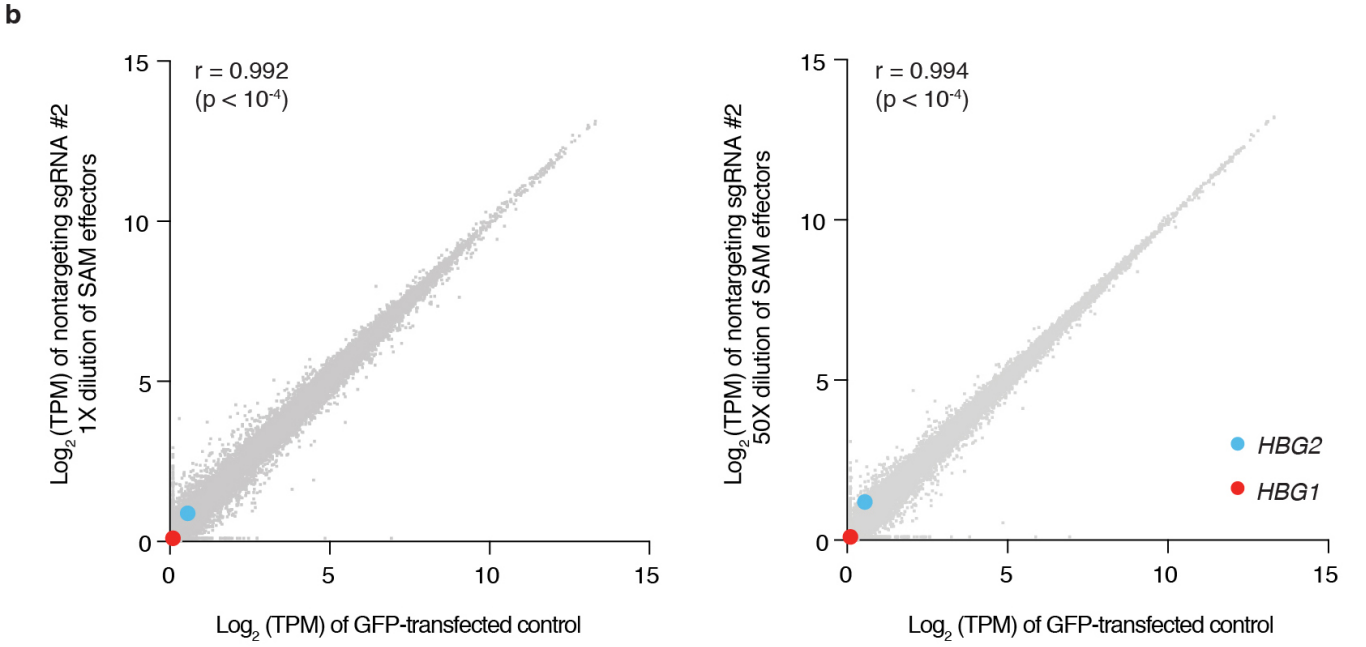
©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | The effect of guide and SAM-component dilution on target activation. **a**, The results for dilution of sgRNA 2.0 on target activation. **b**, The result for dilution of sgRNA 1.0 on target activation. # denotes an activation of <twofold at 1× guide dilution. **c**, Effect of MS2–p65–HSF1 and dCas9–VP64 dilution, at 1:1, 1:4, 1:10 and 1:50 of the original dosage for each component, on the effectiveness of transcription upregulation. The amount of sgRNA expression plasmid was kept constant. **d**, Effect of diluting all three SAM components (dCas9–VP64, MS2–p65–HSF1, and sgRNA) at 1:4, 1:10, and 1:50 of the original dosage for each component. Fold upregulation is calculated using GFP-transfected cells as the baseline. Error bars indicate s.e.m. and $n = 3$ for all figures.

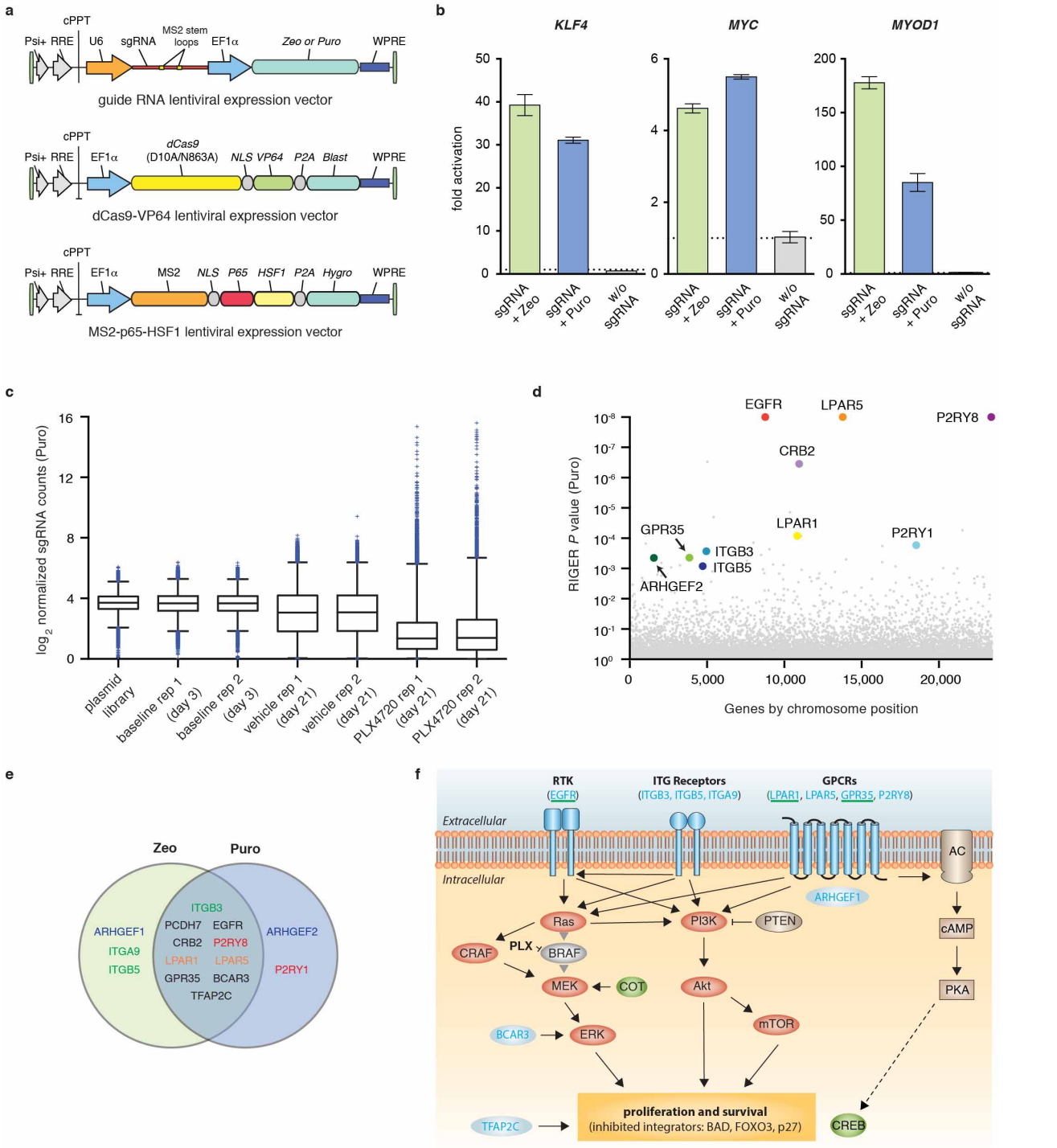
©2015 Macmillan Publishers Limited. All rights reserved





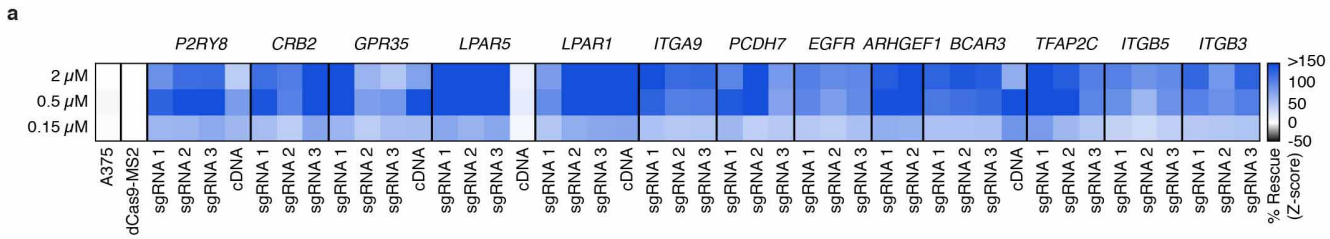
Extended Data Figure 6 | RNA-seq analysis of transcriptome changes mediated by SAM. **a**, A heat map of log(TPM) expression values of all statistically significant differentially expressed genes (*t*-test *q* value < 0.05 adjusted with FDR multiple hypothesis correction) found in any of the six experimental conditions compared to the GFP-transfected control. **b**, Expression levels in log(TPM) values of all detected genes in RNA-seq libraries of GFP-transfected controls (*x*-axis of all graphs) compared to (from left to right): non-targeting control sgRNA no. 2 in 1× dilution and 50× dilution (*y* axis). Marked are *HBG1* (red) and *HGB2* (blue).

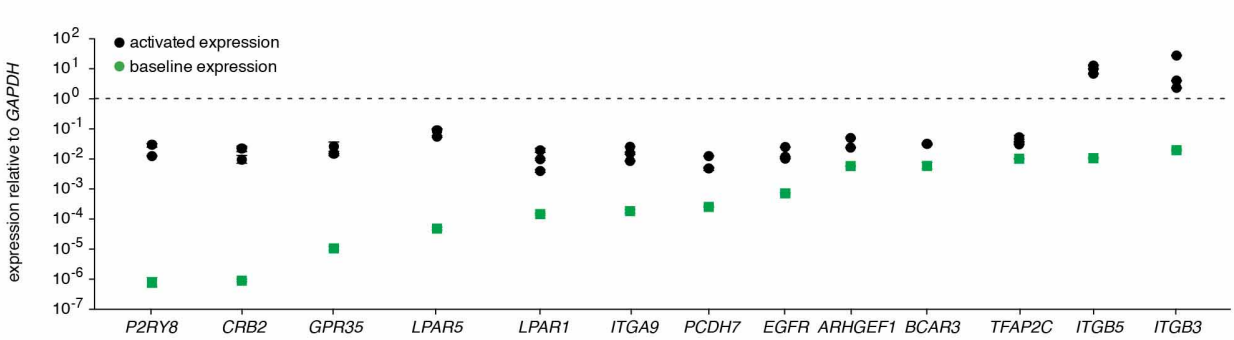
©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Genome-scale lentiviral screen using puromycin-resistant SAM sgRNA library. **a**, Design of three lentiviral vectors for expressing sgRNA, dCas9–VP64, and MS2–p65–HSF1. Each vector contains a distinct selection marker to enable co-selection of cells expressing all three vectors. **b**, Lentiviral delivery of SAM components was tested by first generating 293FT cell lines stably integrated with dCas9–VP64 and MS2–p65–HSF1, and subsequently transducing these cells with single-gene targeting lentiviral sgRNAs at MOI <0.2. Transcription activation efficiency is measured 4 days post sgRNA lentivirus transduction and selection with zeocin or puromycin. Activation is at least as effective as previously observed with transient transfection in all three cases. **c**, Box-plot showing the distribution of sgRNA frequencies at different time points post lentiviral transduction with the Puromycin library, after treatment with DMSO vehicle or PLX-4720. Two infection replicates are shown. **d**, Identification of top candidate genes using the RIGER *P* value analysis (KS method) based on the average of both infection replicates. Genes are organized by positions within chromosomes. **e**, Overlap between the top 20 hits from the zeocin and puromycin screens. Genes belonging to the same family are indicated by the same colour. There is a 50% overlap between the top hits of each screen as shown in the intersection of the Venn diagram. **f**, Relevant signalling pathways in BRAF inhibitor resistance. Reactivation of the Ras-ERK pathway as well as the parallel PI3K-Akt pathway have previously been implicated as two alternative resistance mechanisms to BRAF inhibitors[23,24,26–29]. Both pathways have been described as stimulating proliferation and survival[49]. BAD, FOXO and p27 are common inhibited downstream targets[49]. Recently, stimulation of the cAMP-CREB pathway by GPCRs has been described as a potential additional resistance mechanism[30]. Top candidates from our screen are indicated in blue and putative connections to all three pathways are shown[25,50,51]. Candidates previously validated to mediate PLX-4720 resistance are underlined in green[26,30]. COT and CREB are independently validated mediators of resistance[23,30].

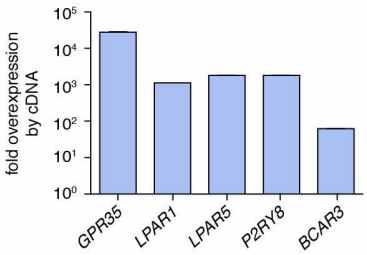
©2015 Macmillan Publishers Limited. All rights reserved

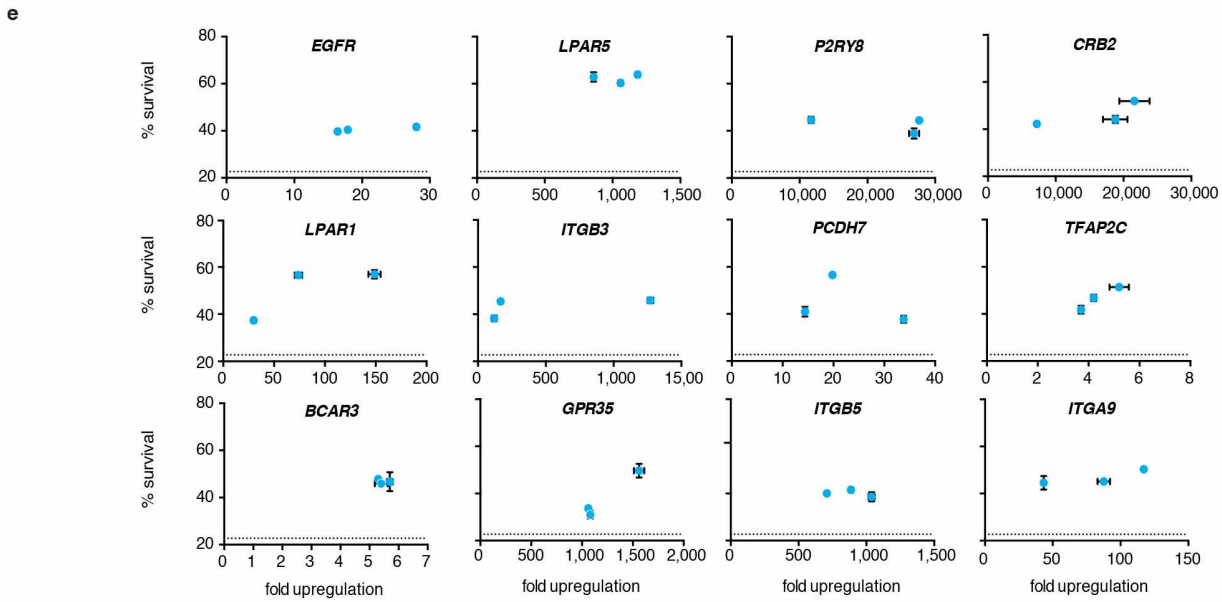




c

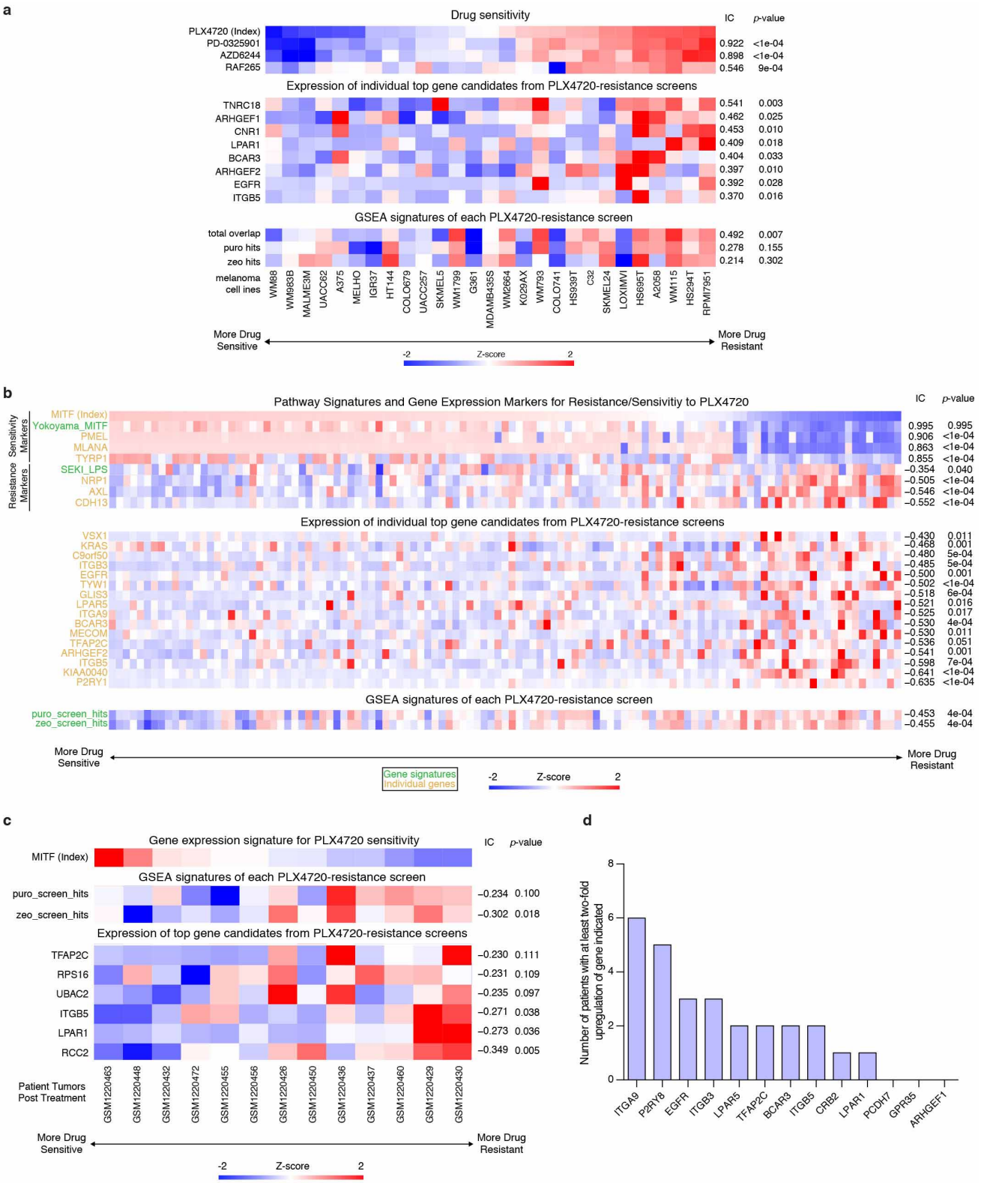
target	z-rank ORF	other family members in top 200 ORF hits
EGFR	229	---
LPAR5	>1,000	*LPAR1*, 4
P2RY8	74	---
CRB2	not included	---
LPAR1	30	---
ITGB3	not included	---
PCDH7	not included	---
TFAP2C	not included	---
BCAR3	309	---
GPR35	15	---
ITGB5	not included	---
ITGA9	not included	---
ARHGEF1	>1000	*ARHGEF2*, 3, 9, 19





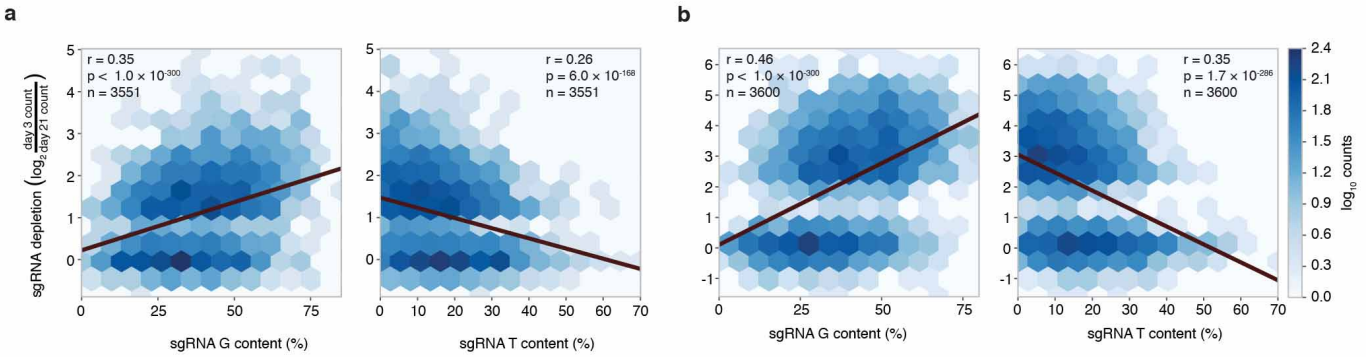
Extended Data Figure 8 | Individual validation of PLX-4720 resistance mediation by top screen hits. **a**, Validation of the top 10 Zeo screen hits and the top 10 shared hits (13 genes total). Every gene was independently activated by all three guides from the screen and tested for the ability to increase survival of A375 cells treated with three different concentrations of PLX-4720 (2μM, 0.5μM and 0.15μM). The z-score based on the % increase in survival relative to control (A375 cells transduced with dCas9–VP64 and MS2–p65–HSF1 alone) is shown for each guide and PLX-4720 concentration. Five cDNAs available from a previous large-scale gain-of-function PLX-4720 resistance screen were also included[30]. Every guide for each top hit mediates significant PLX-4720 resistance. **b**, The same panel of top hits exhibits a large range of basal expression levels and is effectively activated by all guides. The expression level relative to the housekeeping gene *GAPDH* is shown both at baseline as well as after activation by each individual guide. **c**, Ranks of the validated set of genes in the previous ORF screen. Six genes were not part of the cDNA library, five hits are shared (present in the top 3%) and only *LPAR5* and *ARHGEF1* were present but not highly ranked. Both of these genes had highly ranked members of the same family. **d**, Levels of overexpression from the five tested cDNA constructs. Transcript levels were higher for these five cDNAs than those mediated by SAM for the same genes. **e**, Correlation of survival at 2 μM PLX-4720 treatment and transcript upregulation achieved by individual guides. For most genes (9 out of 12 shown), the percent survival is very similar across transcript levels achieved by all three guides. Dotted lines indicate control survival.

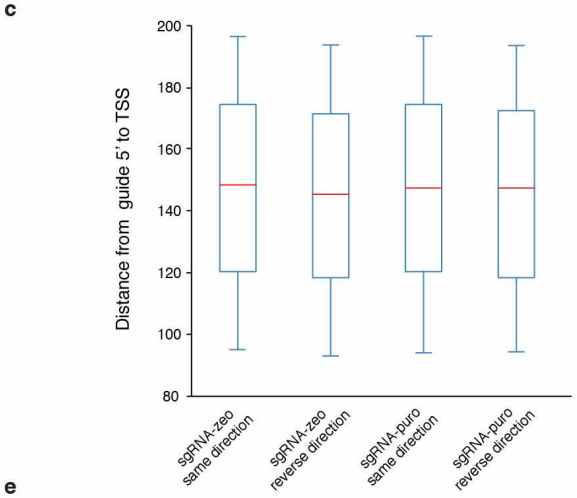
©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Expression of top hits and screen signatures are elevated in PLX-4720 resistant melanoma cell lines and patient samples. **a**, Heat map showing sensitivity to different drugs (top), expression of SAM top screen hits (middle), and SAM screen signature scores (bottom; see Methods for signature generation) in Cancer Cell Line Encyclopedia cell lines[35]. Drug sensitivities are measured as Activity Areas (AA). The melanoma cell lines are sorted by PLX-4720 drug sensitivity. RAF inhibitors: PLX-4720 and RAF265; MEK inhibitors: AZD6244 and PD-0325901. **b**, Heat map showing expression of gene/signature markers for BRAF-inhibitor sensitivity (top), expression of SAM top screen hits (middle) and screen signature scores (bottom) in different BRAF(V600) patient melanoma samples (primary or metastatic) from The Cancer Genome Atlas. **c**, Heat map showing MITF expression (top), screen signature scores (middle), and expression of SAM top screen hits (bottom) in different BRAF(V600E) patient melanoma biopsies post-treatment with BRAF inhibitors[38]. **d**, Bar chart showing the number of patients (out of 13 total) from **c** with at least a twofold change (post/pre-treatment) in gene expression of the top PLX-4720 screen hits in the post-treatment samples. 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. All heat maps show *Z* scores.

©2015 Macmillan Publishers Limited. All rights reserved

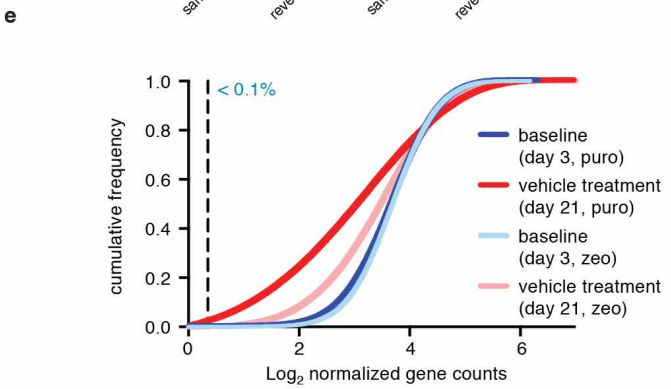




d

Condition	G content		T content	
	coefficient	p-value	coefficient	p-value
sg-Zeo	0.0968	5.9×10^{-97}	-0.0462	3.1×10^{-10}
sg-Puro	0.2201	3.2×10^{-173}	-0.1241	1.4×10^{-23}

Condition	Direction		distance from TSS(bp)	
	coefficient	p-value	coefficient	p-value
sg-Zeo	-0.0206	0.50	-1.2×10^{-4}	0.68
sg-Puro	0.038	0.44	-2.8×10^{-4}	0.58



f

depleted gene categories	
by sgRNA	by average of all 3 sgRNAs
PTEN signaling	molecular mechanisms of cancer
embryonic stem cell pluripotency	glioblastoma multiforme signaling
glioblastoma multiforme signaling	embryonic stem cell pluripotency
G protein-coupled receptor signaling	G protein-coupled receptor signaling
sphingosine-1-phosphate signaling	epithelial-mesenchymal transition
glioma signaling	G beta gamma signaling
STAT3 pathway	germ cell-Sertoli cell junction sig.
molecular mechanisms of cancer	ovarian cancer signaling

Extended Data Figure 10 | Guide depletion analysis to identify gene set enrichment and guide efficiency parameters. **a**, **b**, Heat maps of sgRNA nucleotide content versus depletion after 21 days. sgRNA targeting significantly depleted genes (from RIGER analysis) in sgRNA-zeo (**a**) or sgRNA-puro (**b**) screens were analysed for trends based on G or T content in the sgRNA sequence. sgRNA depletion is positively correlated with G content and negatively correlated with T content. Other bases analysed (A and C) had significant ($P < 0.0007$) but weak ($r < 0.2$) negative correlation. **c**, 90% of guides analysed fall within a 100-bp window < 200 bp from the TSS. Boxplots of distance from 5′ end of the guide to the TSS for sgRNA-zeo and sgRNA-puro in same and reverse direction (relative to target transcription). Whiskers span 5th to 95th quartile. **d**, Coefficients and *P* values for ordinary least squares predicting sgRNA depletion of significantly depleted genes from G content, T content, distance from 5′ end of the guide to the TSS and direction of guide. Only nucleotide content has a significant effect on depletion in this model, consistent with a high efficiency of guides within 200 bp of the TSS regardless of strand orientation (Fig. 2d). **e**, The cumulative frequency of sgRNAs 3 and 21 days after transduction in A375 cells is shown. Shift in the 21-day curve represents the depletion in a subset of sgRNAs. Less than 0.1% of all guides are not detected at day 3 (detected by less than 10 reads). **f**, Depleted guides (Supplementary Table 3) can be analysed for significant clustering of gene categories. Gene categories exhibiting significant depletion based on Ingenuity Pathway Analysis ($P < 0.01$ after Benjamini–Hochberg FDR correction) are shown. Categories based on the 1,000 most depleted guides individually (left) and the average of all 3 guides/gene (right). These categories include either positive or negative regulators of each pathway that reduce proliferation and survival.

©2015 Macmillan Publishers Limited. All rights reserved