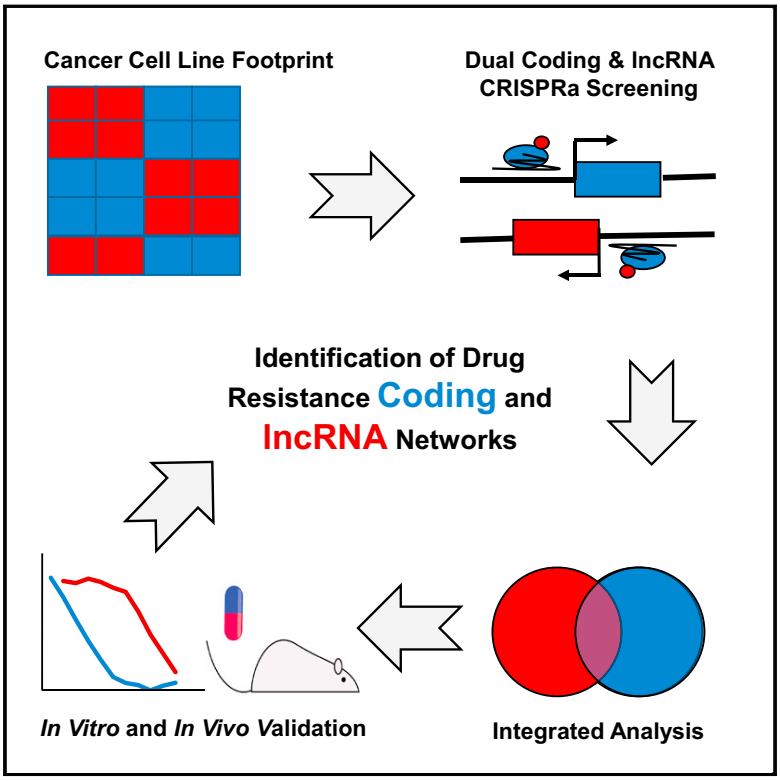
Article

Cell

An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance

Graphical Abstract



Authors

Assaf C. Bester, Jonathan D. Lee, Alejandro Chavez, ..., George M. Church, John G. Clohessy, Pier Paolo Pandolfi

Correspondence

ppandolf@bidmc.harvard.edu

In Brief

A CRISPR activation screen identifies both coding and noncoding pathways involved in resistance to chemotherapy.

Highlights

- Analysis of CCLE and CTD^2 identifies coding and lncRNA genes for Ara-C resistance
- Genome-wide CRISPRa screens identify functional coding/lncRNA resistance gene pairs
- Integrating computational, functional, and patient data reveals functional lncRNAs
- GAS6-AS2 lncRNA acts in a *cis*- and *trans*-manner to regulate GAS6/AXL signaling

Bester et al., 2018, Cell *173*, 649–664
April 19, 2018
https://doi.org/10.1016/j.cell.2018.03.052

CellPress

Article

Cell

An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance

Assaf C. Bester,[1,2] Jonathan D. Lee,[1,2,12] Alejandro Chavez,[3,4,5,12] Yu-Ru Lee,[1,2] Daphna Nachmani,[1,2] Suhani Vora,[3] Joshua Victor,[1,2] Martin Sauvageau,[8,10,11] Emanuele Monteleone,[6] John L. Rinn,[8] Paolo Provero,[6,7] George M. Church,[3] John G. Clohessy,[1,2,9] and Pier Paolo Pandolfi[1,2,13,*]

[1]Cancer Research Institute, Beth Israel Deaconess Cancer Center, Department of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

[2]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA

[3]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA

[4]Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

[5]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

[6]Department of Molecular Biotechnology and Health Sciences, and GenoBiToUS, Genomics and Bioinformatics Service, University of Turin, Turin, Italy

[7]Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute IRCCS, Milan, Italy

[8]Department of Stem Cell and Regenerative Biology, Harvard University, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

[9]Preclinical Murine Pharmacogenetics Facility and Mouse Hospital, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

[10]Functional Genomics and Noncoding RNAs Research Unit, Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC, Canada

[11]Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, QC, Canada

[12]These authors contributed equally

[13]Lead Contact

*Correspondence: ppandolf@bidmc.harvard.edu

https://doi.org/10.1016/j.cell.2018.03.052

SUMMARY

Resistance to chemotherapy plays a significant role in cancer mortality. To identify genetic units affecting sensitivity to cytarabine, the mainstay of treatment for acute myeloid leukemia (AML), we developed a comprehensive and integrated genome-wide platform based on a dual protein-coding and non-coding integrated CRISPRa screening (DICaS). Putative resistance genes were initially identified using pharmacogenetic data from 760 human pan-cancer cell lines. Subsequently, genome scale functional characterization of both coding and long non-coding RNA (lncRNA) genes by CRISPR activation was performed. For lncRNA functional assessment, we developed a CRISPR activation of lncRNA (CaLR) strategy, targeting 14,701 lncRNA genes. Computational and functional analysis identified novel cell-cycle, survival/apoptosis, and cancer signaling genes. Furthermore, transcriptional activation of the GAS6-AS2 lncRNA, identified in our analysis, leads to hyperactivation of the GAS6/TAM pathway, a resistance mechanism in multiple cancers including AML. Thus, DICaS represents a novel and powerful approach to identify integrated coding and non-coding pathways of therapeutic relevance.

INTRODUCTION

Although precision medicine and targeted therapies offer new hope for treating cancer, chemotherapy still remains the first, and last, line of defense for most patients. Cytarabine (1-β-d-arabinofuranosylcytosine, Ara-C) is a deoxycytidine analog that is used as part of a standard chemotherapeutic regimen for the treatment of acute myeloid leukemia (AML) (Ramos et al., 2015). However, ~30% to 50% of patients relapse with chemotherapy-resistant disease. Thus, there is an ever-present need to better understand the genetic and molecular mechanisms that contribute to chemotherapy resistance.

To date, studies on mechanisms leading to therapy resistance have focused on protein-coding genes, yet cancer development and progression cannot be fully explained by the coding genome (Huarte, 2015; Imielinski et al., 2012). The recent explosion in research and understanding related to the non-coding RNA (ncRNA) transcriptome has highlighted the importance of ncRNAs in biology (Hon et al., 2017; Iyer et al., 2015). Functional validation of various ncRNA species highlights the fact that these RNAs may play important roles in the pathogenesis of diseases including cancer (Schmitt and Chang, 2016).

One large group of ncRNAs is represented by long non-coding RNAs (lncRNA). LncRNAs can be either nuclear or cytoplasmic in localization and play roles in a diverse array of biological processes. As many nuclear lncRNAs behave in a *cis*-acting manner (Quinn and Chang, 2016), their study requires their expression from endogenous loci, and CRISPR technologies now facilitate the modulation of gene expression directly from the endogenous

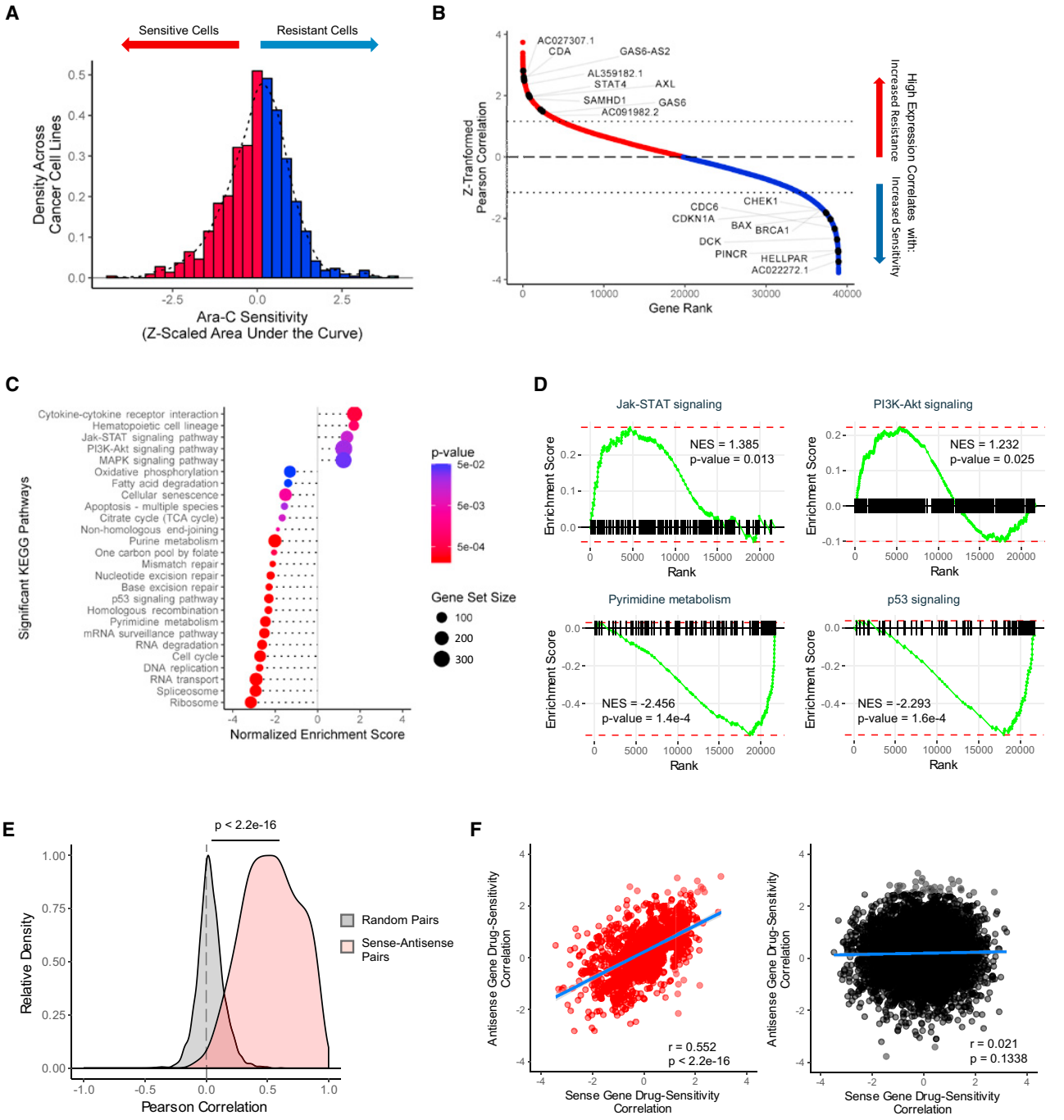


Figure 1. Identification of Protein-Coding and Non-coding Gene Biomarkers Correlated with Differential Ara-C Response

(A) Distribution of Ara-C drug sensitivities across 760 pan-cancer cell lines profiled by both CCLE and CTD^2 studies, quantified by their *Z* scaled area under the dose-response curve values after regressing out lineage-specific effects. See also Table S1.

(B) Distribution of *Z* scaled drug resistance-gene expression Pearson correlation values of all analyzed genes. Representative protein-coding and non-coding gene symbols enriched beyond a *Z* score threshold of ±1.16 are demarcated. See also Table S1.

(C) Summary of gene set enrichment analysis (GSEA) of protein-coding genes ranked by drug resistance-gene expression correlation values using annotated KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. See also Table S3.

(D) Representative KEGG pathways from GSEA of protein-coding genes ranked by drug sensitivity-gene expression correlation values as shown in (B) and (C). See also Table S3.

(legend continued on next page)

promoter (Joung et al., 2017a; Konermann et al., 2015). This approach has already been compellingly demonstrated using CRISPR interference (CRISPRi) to silence the expression of lncRNAs genome-wide (Liu et al., 2017).

Although we now have a wealth of high-throughput data delineating expression of coding and non-coding genes across hundreds of cancer cell lines (Barretina et al., 2012; Garnett et al., 2012), there remains a critical lack of integrated high-throughput functional characterization and validation of these data in a disease context. We therefore sought to develop an integrative and comprehensive CRISPR activation (CRISPRa) framework that would complement these publicly available ''Big Data'' databases to enable the discovery of functional human protein coding and lncRNA genes contributing to chemotherapy resistance. In doing so, we developed a dual coding and non-coding integrated CRISPRa screening (DICaS) platform and applied this integrative approach to identify genetic units and pathways that promote resistance to Ara-C treatment.

RESULTS

Pan-Cancer Cell Line Analysis of lncRNAs Associated with Drug Response

In order to comprehensively define resistance mechanisms to chemotherapy, we chose to examine cellular responses to Ara-C. We developed a computational strategy to identify genes that correlate with sensitivity or resistance to Ara-C by correlating pharmacological profiles from the Cancer Target Discovery and Development (CTD^2) database (Basu et al., 2013; Rees et al., 2016) with the transcriptomes of 760 corresponding cell lines from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) (Figure S1A). To identify high confidence gene targets, it is imperative to integrate analysis of as many cell lines as possible (Rees et al., 2016); however, we found that the cell line drug sensitivities formed a skewed distribution (Figure S1B), likely conferred by tissue of origin and histological subtype. Indeed, cancer cell type annotations explained a substantial amount of the variation in drug sensitivities (adjusted $R^2 = 0.5123$, ANOVA $p < 2.2e{-}16$) (Figure S1A), which were subsequently corrected (Figure S1C). Thus, using a linear regression model to remove these effects we established a normalized distribution of Ara-C sensitivity for the 760 cell lines analyzed (Figure 1A).

We subsequently performed a correlation analysis between drug sensitivities and gene expression levels across the 760 cell lines (Figure 1B; Table S1) and determined appropriate *Z* score thresholds (Figures S1D and S1E). Interestingly, genes involved in the metabolism of Ara-C were highly enriched, illustrating the applicability of such an approach to identifying chemotherapy resistance mechanisms. Low expression of deoxycytidine kinase (*DCK*) and equilibrative nucleoside transporter 1 (*ENT1/SLC29A1*) correlated with increased resistance to Ara-C ($Z = -2.51$ and -1.61, respectively), whereas high expression of cytidine deaminase (CDA) and SAM domain and HD domain 1 (SAMHD1) correlated with increased resistance (Herold et al., 2017; Schneider et al., 2017) ($Z = 2.54$ and 2.03, respectively) (Figure 1B). Interestingly, we also observed a number of cell-cycle and DNA damage regulators previously implicated in modulation of AraC sensitivity (Abraham et al., 2015; Lamba, 2009) (Figure 1B).

To define biological pathways predictive of Ara-C resistance we performed gene set enrichment analysis (GSEA) on the drug sensitivity-gene expression correlations using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations (Figure 1C; Table S3) (Kanehisa et al., 2014; Subramanian et al., 2005). We identified positive enrichment of cell survival signaling pathways, including the Jak-STAT (NES = 1.385, p = 0.013), PI3K-Akt (NES = 1.232, p = 0.025), and MAPK (NES = 1.222, p = 0.042) pathways and negative enrichment of the pyrimidine metabolic pathway (NES = −2.456, p = 0.00016), mechanisms related to DNA damage (e.g., p53 signaling pathway: NES = −2.293, p = 0.00016), and RNA regulatory mechanisms (e.g., RNA degradation: NES = −2.613, p = 1.6e−4) (Figures 1C, 1D, and S1F). To confirm their relevance in human AML, we correlated pre-treatment AML transcriptome profiles with corresponding disease-free survival data from 121 patients treated with Ara-C from The Cancer Genome Atlas (TCGA) (Ley et al., 2013) and identified a large number of enriched pathways shared with our cell line predictions, including oxidative phosphorylation (Farge et al., 2017) (NES = −1.994, p = 1.1e−4) and RNA regulatory mechanisms (e.g., RNA degradation: NES = −1.702, p = 0.0011) (Figures S1G and S1H).

As many non-coding genes act in a proximal and localized manner (Schmitt and Chang, 2016), we evaluated coding and non-coding cognate gene pairs for correlation with either resistance or sensitivity to Ara-C and compiled a genome wide set of 997 coding/non-coding sense/antisense gene pairs. Indeed, we observed a significant positive correlation between sense-antisense gene expression levels across the cell line panel (Pearson correlation, median R = 0.5312; Wilcoxon rank-sum test, $p < 2.2e{-}16$) (Figure 1E). Furthermore, cognate gene pairs demonstrated significant positive correlation in drug sensitivity (Pearson correlation, R = 0.5636, $p < 2.2e{-}16$) (Figure 1F). Importantly, analysis of these same cognate gene pairs among the TCGA AML patient cohort identified a similarly significant positive correlation (Figures S2A–S2C). Interestingly, cognate sense genes were found to be positively enriched in PI3K-Akt (NES = 1.426, p = 0.0764) and MAPK signaling pathways (NES = 1.787, p = 0.0040) (Figure S2D), implicating these sense-antisense gene pairs in a number of the previously identified enriched pathways.

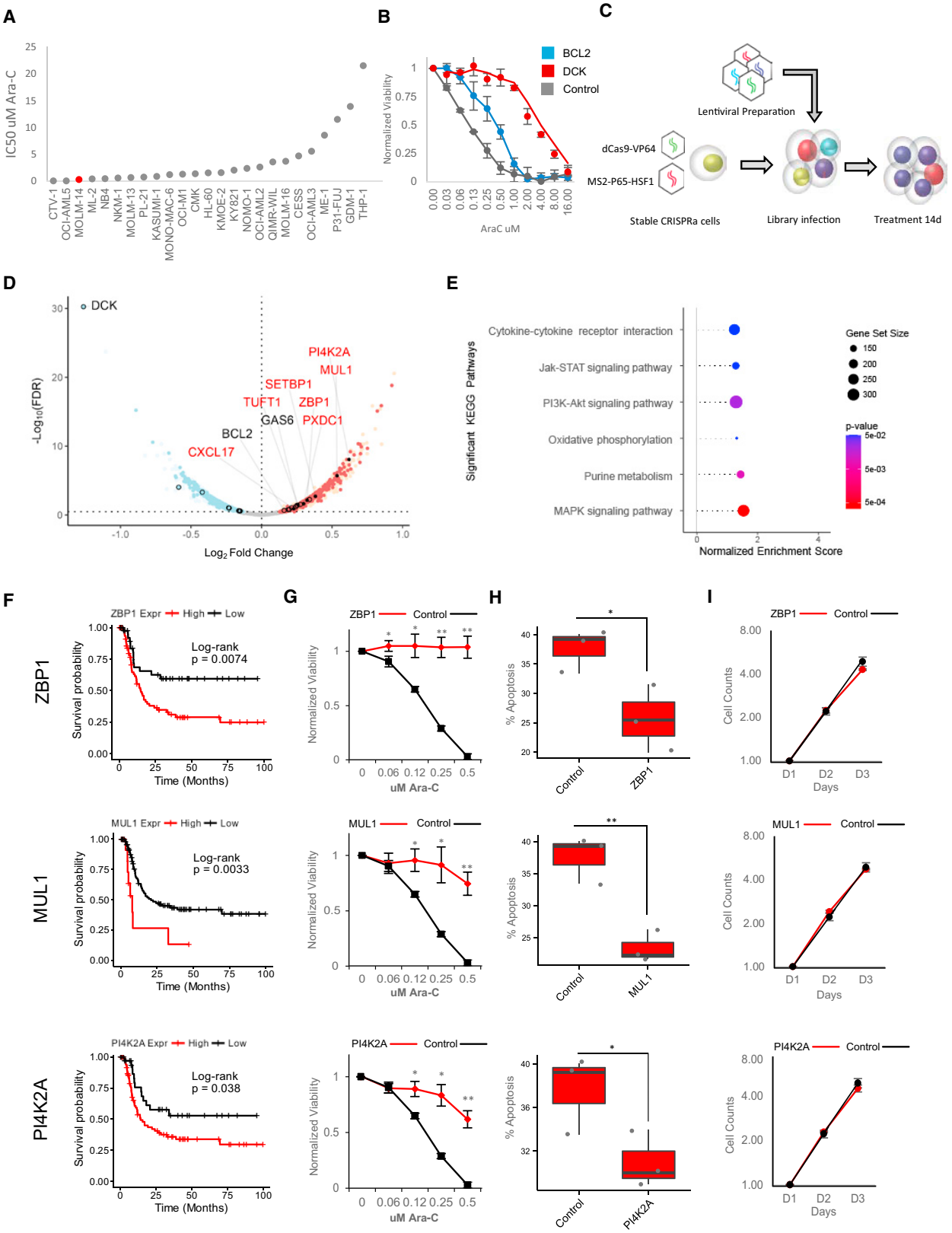
A CRISPRa Approach to Study AML Resistance to Ara-C

To functionally validate our predictive analysis in a high-throughput manner, we established a CRISPRa-based system

(E) Pearson correlation distributions of gene pair expression levels in the cancer cell line panel across 997 sense-antisense cognate gene pairs and 5,000 random protein coding-lncRNA gene pairs. Wilcoxon rank-sum test: $p < 2.2e{-}16$.

(F) Relationship of drug sensitivity-gene expression correlation values between protein coding-lncRNA gene pairs across 997 sense-antisense cognate gene pairs (left: Pearson's R = 0.552, $p < 2.2e{-}16$) and 5,000 random gene pairs (right: Pearson's R = 0.021, p = 0.1338).

See also Figures S1 and S2 and Table S2.



(legend on next page)

in AML cell lines to provide a comprehensive and integrative genome-wide study of both the coding and non-coding genes contributing to Ara-C resistance.

We identified the MOLM14 AML cell line to be the most appropriate model for our screening, as its IC_{50} (~0.13 μM) ranks it among the most sensitive AML cell lines (Yang et al., 2013) (Figure 2A). Overexpression of the anti-apoptotic B cell lymphoma 2 (*BCL2*) gene increased the IC_{50} of MOLM14 for Ara-C, while small hairpin RNA (shRNA)-mediated knockdown of DCK provided an even more significant protection, increasing its IC_{50} almost 300-fold (Figure 2B), confirming that sensitivity to Ara-C can be readily manipulated.

We also tested synergistic activation mediator (SAM)-mediated CRISPRa (Konermann et al., 2015) in MOLM14 cells as compared with two additional leukemia cell lines, K562 and HL60, and the previously validated HEK293T. Using a panel of validated single guide RNAs (sgRNAs) targeting the promoters of both coding (*TTN*, *RHOXF2*, *ASCL1*, *HBG1*) and non-coding (*MIAT*, *TUNA*) genes (Chavez et al., 2015, 2016), we established that the majority of sgRNAs gave the highest activation in MOLM14 among the leukemia cell lines (Figure S3A).

Genome-wide CRISPRa Screening of Protein-Coding Genes in AML

We next applied our CRISPRa platform to screen for protein-coding genes using a genome wide sgRNA library (Konermann et al., 2015) (Figure 2C). For library screening, cells were treated for 14 days with 0.25 μM Ara-C, and cell viability monitored over the treatment period (Figure S3B). Following treatment, sgRNA abundances were quantified and analyzed for quality control (Figures S3C–S3E). Transcript-level representation between T0 and T14 identified a host of genes enriched and depleted in Ara-C-treated cells (Figures 2D and S3F). Interestingly, both the correlation analysis and our forward genetic screen revealed *DCK* to be the most significantly depleted gene, thereby indicating that strong transcriptional activation of DCK by CRISPRa confers high sensitivity to Ara-C (Figures 1B and 2D). Indeed, this was confirmed by overexpressing the top-scoring *DCK* targeting sgRNA (Figure S3G). Furthermore, multiple genes suspected to modulate sensitivity to Ara-C were also identified (Table S4).

GSEA identified a number of pathways congruent with our cell line analysis (Figure 2E; Table S3). Importantly, we identified a large overlap of 2,411 genes significantly enriched/depleted in both our cell line and protein-coding CRISPRa screening (Figure S3H). We subsequently validated a subset of these genes, including *ZBP1*, *MUL1*, and *PI4K2A*, whose expression was associated with poor prognosis and decreased disease-free survival (Figure 2F). Cells expressing the relevant sgRNAs demonstrated increased survival upon treatment with Ara-C (Figures 2G and S3I), and a decrease in apoptosis (Figures 2H and S3J), thereby validating our findings. Importantly, the proliferative capacity of cells was not affected by the overexpression of these sgRNAs (Figure 2I).

Functional Genome-wide Screening of lncRNAs in AML

To study the functional roles of lncRNA genes in Ara-C resistance, we designed an sgRNA library using a comprehensive set of 14,701 lncRNA genes, covering all major classifications of lncRNAs (Figure 3A; Table S5). We designed at least 4 sgRNAs per lncRNA, accounting for 22,253 transcriptional start sites (TSSs), covering multiple TSSs per individual lncRNA. This resulted in a library with 88,444 targeting guides (Figure 3A). We termed this CRISPRa SAM-mediated approach "CRISPR activation of lncRNA" (CaLR).

To test our library, we picked sgRNAs targeting the *TUNA* lncRNA gene (n = 4 sgRNAs) (Figure S4A) and two alternative TSSs for the *MIAT* lncRNA gene (*MIAT-01*, n = 5 sgRNAs, *MIAT-06*, n = 4 sgRNAs), and we confirmed activation of each TSS using at least two sgRNAs (see Figures S3A, S4A, and S4B). An additional set of randomized sgRNAs was also tested on HEK293T and MOLM14, revealing that the majority of sgRNAs demonstrated transcriptional activation in at least one

Figure 2. CRISPRa Functional Screening of Coding Genes Modulating Ara-C Response

(A) Distribution of Ara-C IC_{50} values across a panel of AML cell lines.

(B) Effect of *BCL2* overexpression (blue) or *DCK* knockdown on sensitivity to Ara-C in MOLM14 cells. Data are represented as mean ± SD, n = 3.

(C) Schematic of CRISPRa pooled screening for the identification of genes whose activation modulates sensitivity to Ara-C in MOLM14 cells.

(D) Volcano plot summarizing the global changes in sgRNA representation of protein-coding genes before and after 14 days of treatment with Ara-C. A subset of genes validated herein (red text) or previously annotated (black text) to modulate Ara-C sensitivity are labeled. A FDR threshold of 0.339 was determined by receiver operating characteristic analysis (Figure S3F). Red, enrichment in the CRISPRa screening; blue, depletion in the CRISPRa screening; open black circles, genes previously associated with differential Ara-C sensitivity and above the significance threshold; filled black points, genes validated herein. See also Figures S3C–S3F and S3H, and Table S4.

(E) Summary of GSEA of protein-coding genes ranked by CRISPRa screening using annotated KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. See Table S3.

(F) Disease-free survival association with expression levels of ZBP1, MUL1, and PI4K2A, genes enriched in both protein-coding CRISPRa screening and drug sensitivity-gene expression correlation analyses among patients treated with Ara-C therapy within the TCGA-LAML patient cohort. ZBP1: VST expression level cutoff = 6.13 (low, n = 42; high, n = 79), log-rank test: p value = 0.0074. MUL1: VST expression level cutoff = 9.64 (low, n = 108; high, n = 13), log-rank test: p value = 0.0033. PI4K2A: VST expression level cutoff = 7.23 (low, 36; high, n = 85), log-rank test: p value = 0.038.

(G) Ara-C efficacy measurements in MOLM14 cells expressing sgRNAs targeting ZBP1, MUL1, or PI4K2A based on normalized MTS reads following 48 hr of treatment. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

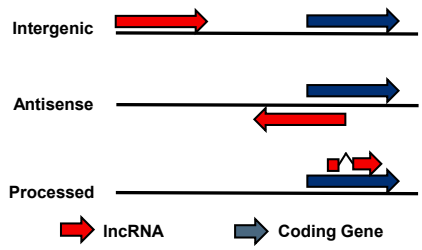
(H) Modulation of apoptotic response upon stable expression of sgRNAs targeting ZBP1, MUL1, or PI4K2A in MOLM14 cells. The percentage of apoptosis is determined by annexin V and propidium iodide (PI) staining of cells treated with 0.25 μM Ara-C for 72 hr. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

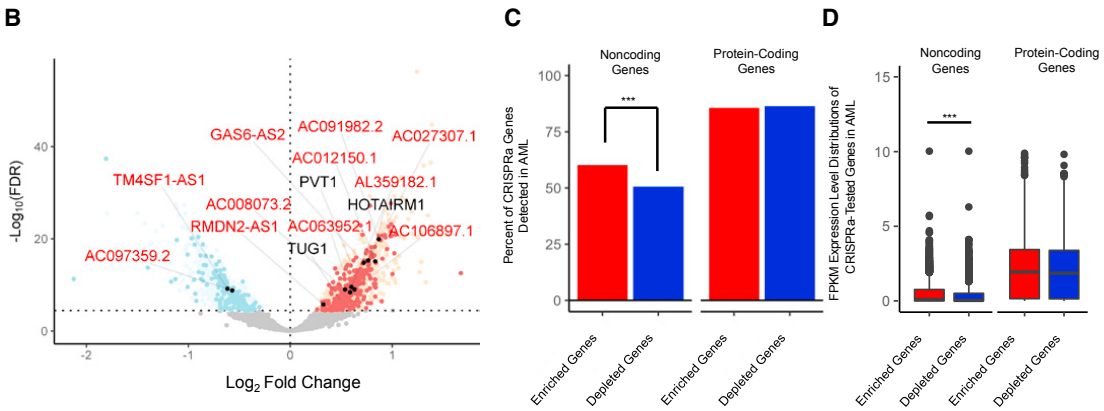
(I) Proliferation of unchallenged MOLM14 cells expressing sgRNAs targeting ZBP1, MUL1, or PI4K2A. Proliferation is quantified over 4 days (D1–D4). Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

See also Table S7.

A

Sub-library	Genes	TSS	sgRNAs
Human Intergenic lincRNA	7,563	10,781	42,836
Human Antisense	5,532	8,545	33,965
Human Processed Transcript	1,606	2,927	11,643
Non-Targeting Control			99
Total	**14,701**	**22,253**	**88,543**





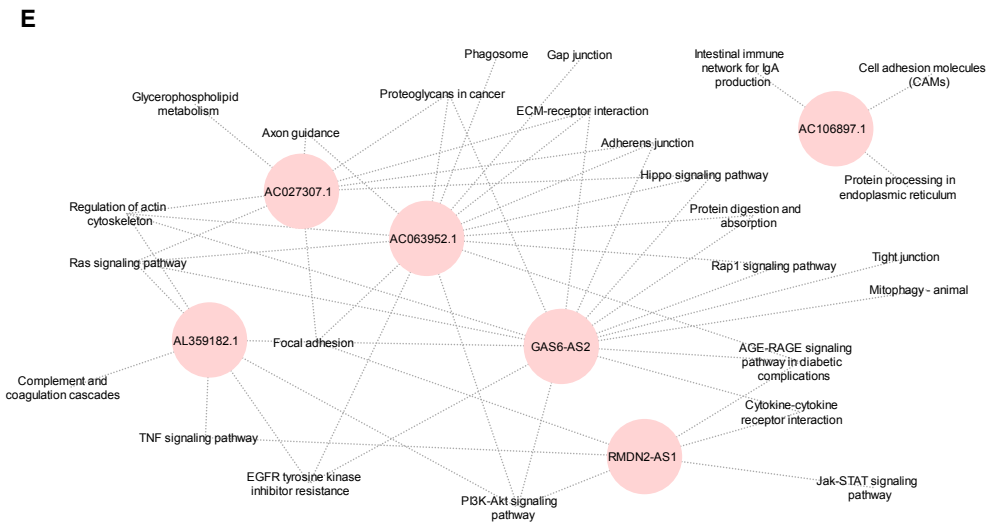


Figure 3. CRISPRa Functional Screening of Non-coding Genes Modulating Ara-C Response

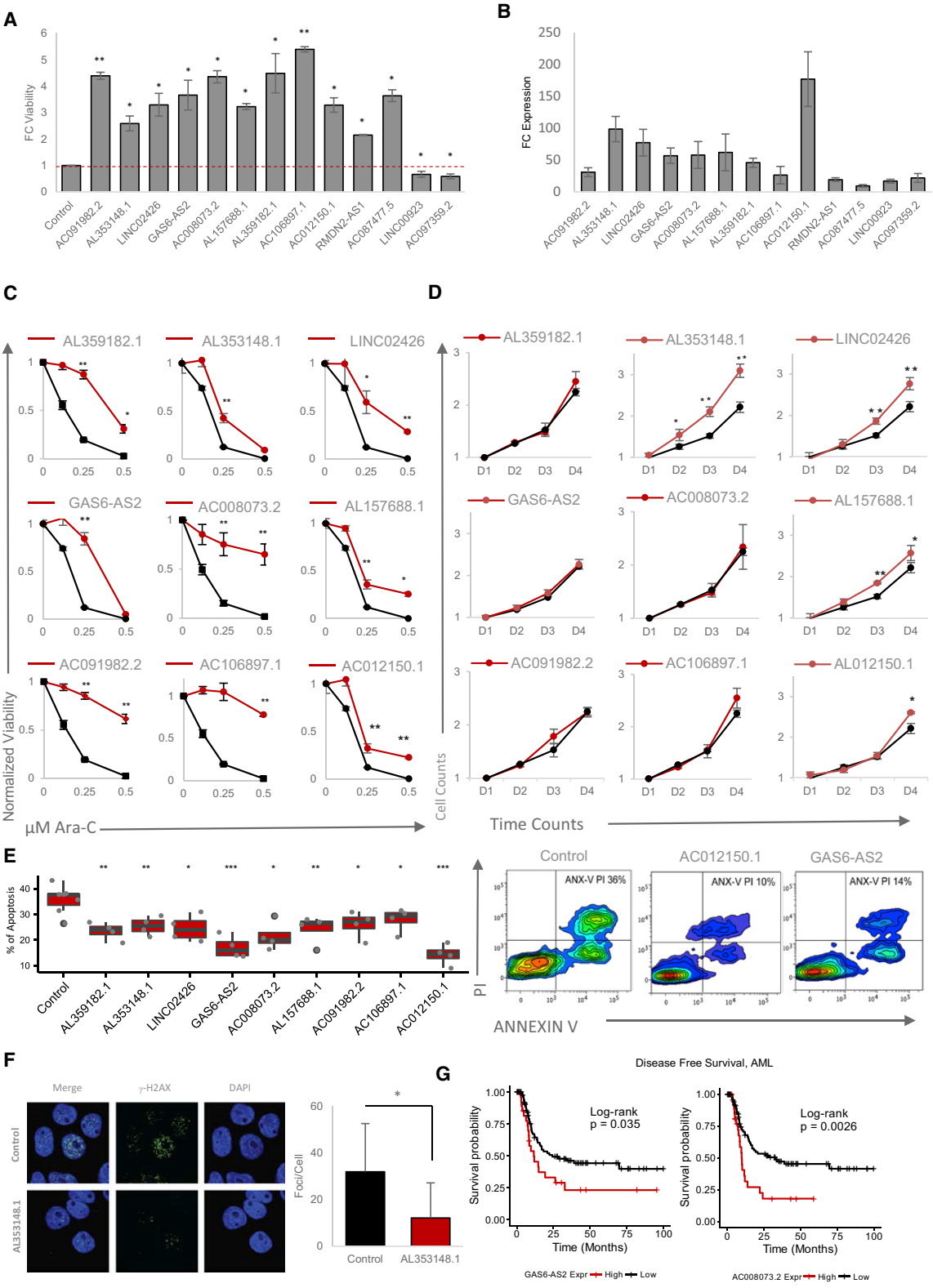
(A) Left: summary of the CaLR library design specifications, including lncRNA gene numbers, transcriptional start sites (TSS), and total sgRNA numbers. Right: relationships between coding genes and lncRNA genes for corresponding lncRNA classifications. See also Table S5.

(B) Volcano plot summarizing the global changes in sgRNA representation of non-coding genes before and after 14 days of treatment with Ara-C. A subset of genes either validated herein to modulate Ara-C sensitivity (red text) or previously annotated in various cancer-related pathways (black text) are labeled. A FDR threshold of 3.51e−5 was determined by analysis of nontargeting sgRNA negative controls at the transcript level (Figure S4H). Red points, enrichment in the CRISPRa screening; blue points, depletion in the CRISPRa screening; filled black points, genes validated herein. See also Figures S4E–S4I and Table S6.

(C) Percentages of significantly enriched or depleted protein-coding or non-coding genes from CRISPRa screens detected in the TCGA-LAML patient samples. Chi-square test: ***p = 6.92e−3,

(D) Gene expression level distributions of significantly enriched or depleted protein-coding or non-coding genes from CRISPRa screens detected in the TCGA-LAML patient samples. Wilcoxon rank-sum test: ***p = 5.4e−7.

(E) Guilt-by-association pathway annotation of enriched genes identified in the CaLR screen. KEGG pathway gene sets were used for this analysis. See also Table S7.



(legend on next page)

of these cell lines (Figure S4D). Next, we carried out screening using our CaLR library similar to that for the protein-coding library above.

After quantifying sgRNA abundance, library preparations were analyzed as above for potential technical bias (Figures S4E–S4G). In order to estimate the false-positive rate within our non-coding RNA screening, we included 99 non-targeting sgRNAs (Figure 3A). These 99 non-targeting sgRNAs behaved as expected (Figure S4G) and were utilized to determine an appropriate false discovery rate (FDR) cut-off to control for the false-positive rate (Figure S4H). Interestingly, several cancer-associated lncRNA genes were identified among enriched sgRNAs, including taurine up-regulated 1 (*TUG1*), HOXA transcript antisense RNA, myeloid-specific 1 (*HOTAIRM1*), and plasmacytoma variant translocation 1 (*PVT1*) (Figures 3B and S4I; Table S6).

Interestingly, expression analysis of lncRNAs and coding genes from the TCGA AML patient cohort revealed that enriched lncRNAs from our screen tended to be both detected at a higher rate ($p = 6.92e{-3}$) and expressed more highly than depleted lncRNAs ($p = 5.4e{-7}$), whereas a similar pattern was not observed among the enriched/depleted protein-coding mRNAs (Figures 3C and 3D).

Furthermore, guilt by association analysis of the enriched lncRNAs identified two distinct gene set networks: (1) oxidative phosphorylation and fatty acid metabolism, and (2) leukemia development and progression (Figure S4J). Enrichment of these pathways in the first network is reflective of the role of the mitochondria in regulating nucleotide metabolism, while specific pathways enriched in the latter network include leukemia-associated pro-survival pathways (e.g., interferon response, IL6/JAK/STAT3 signaling, tumor necrosis factor alpha [TNF-α]/nuclear factor κB [NF-κB] signaling) (Steelman et al., 2004; Stavropoulou et al., 2016).

We compiled a short list of novel annotated lncRNAs to characterize further, which were significantly enriched in both our functional screening and our cell line analysis (Figure 3B). Co-expression analysis to associate individual lncRNA transcript levels with their most highly correlated protein coding genes from our CCLE cell line panel identified many of the pathways uncovered in our global analysis, suggesting that these lncRNAs play roles in survival pathways know to affect leukemia and drug resistance (Figure 3E; Table S3).

Validation of Top lncRNA Candidates

To validate the findings from our screening experimentally, we chose 11 genes significantly enriched and two genes significantly depleted in our screening for further characterization (Figures 3B and 3E). Of these 13 genes selected from our screening, 10 were also found to be candidate genes predicted to influence Ara-C response in our cell line analysis. The enriched sgRNAs resulted in a significant protection over control cells, while the two depleted genes resulted in decreased viability in response to Ara-C (Figure 4A). Indeed, we confirmed increased lncRNA expression across the different sgRNAs examined (Figure 4B). To further characterize the ability of induced lncRNA expression to resist Ara-C cytotoxicity, we treated cells expressing the relevant sgRNAs with Ara-C (Figures 4C, S5A, and S5B). Expression of each enriched sgRNA resulted in decreased Ara-C sensitivity (Figures 4C and S5A), correlating with the protective effect observed in Figure 4A, while the depleted lncRNA genes also behaved as expected (Figure S5B).

To address how these lncRNAs may be promoting cell viability, we examined lncRNA ability to promote either increased proliferation or increased survival. Out of our candidate lncRNAs, only three appeared to promote proliferation in the absence of Ara-C (AL353148.1, LINC02426; AL157688.1) (Figures 4D, S5C, and S5D), suggesting that their enrichment might be facilitated by increased proliferation. On the other hand, while all sgRNAs were able to promote increased survival to some extent (Figures 4E and S5E), both AC012150.1 and GAS6-AS2 (also named GAS6 divergent transcript [GAS6-DT]) demonstrated a significant ability to attenuate apoptosis (Figure 4E, right panel). These results were further confirmed in an independent HL-60 hematopoietic cell line (Figures S5F–S5H).

Given that Ara-C promotes extensive genotoxic stress, we tested if our lncRNAs may affect the DNA damage response

Figure 4. Validation of CaLR Screening Results

(A) Fold change (FC) of MOLM14 cell viability treated with 0.25 μM Ara-C for 48 hr. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.

(B) Fold change (FC) of expression levels of targeted lncRNAs upon overexpression of enriched sgRNAs versus endogenous levels. Data are represented as mean ± SD, n = 3.

(C) Ara-C efficacy measurements in MOLM14 cells expressing sgRNAs targeting the indicating genes based on normalized MTS reads following 48 hr of treatment with the indicated concentrations of Ara-C. Data are represented as mean ± SD, n = 3, Welch two-sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$

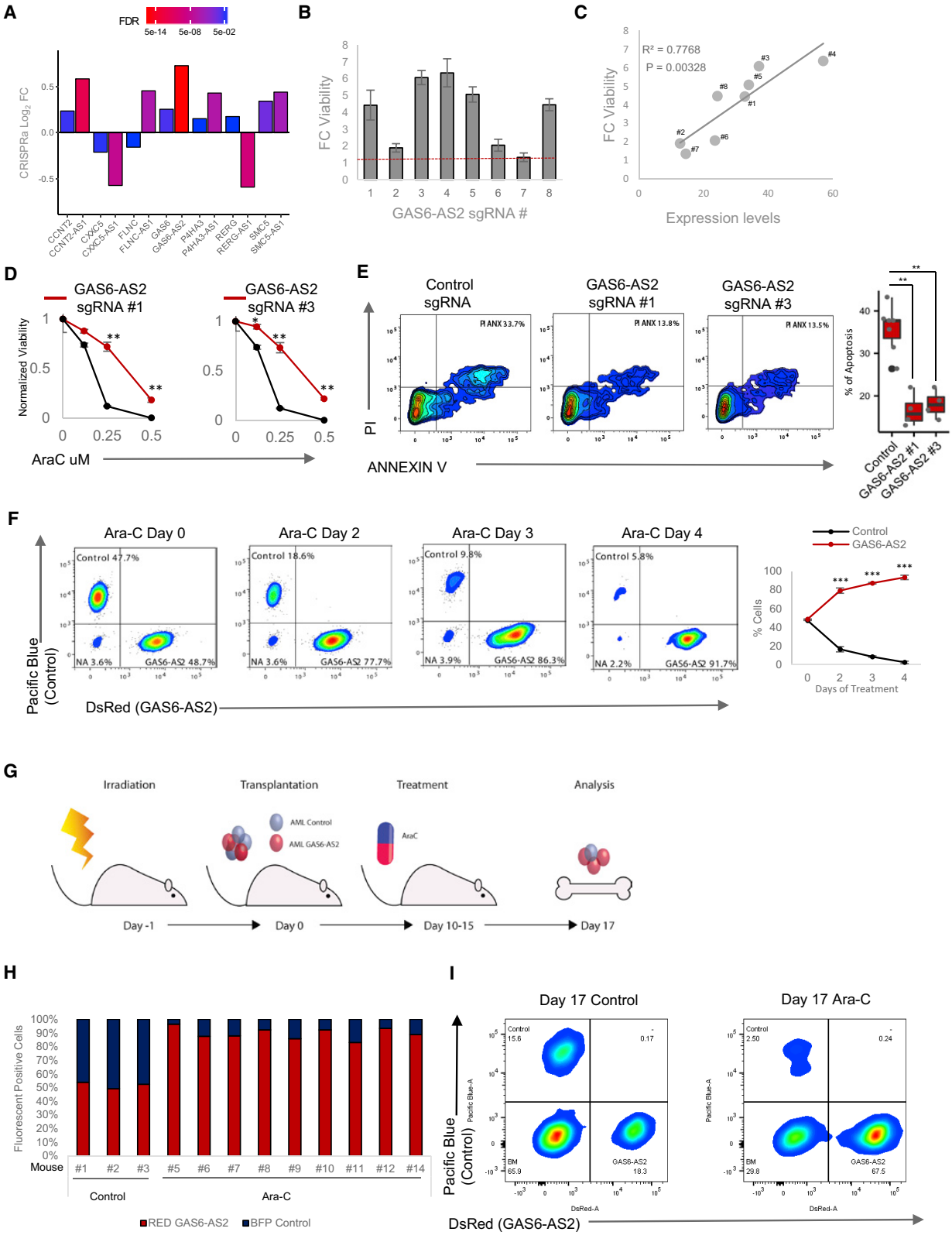
(D) Proliferation of unchallenged MOLM14 cells expressing sgRNAs targeting indicating genes. Proliferation is quantified over 4 days (D1–D4). Data are represented as mean ± SD, n = 3. Welch two-sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.

(E) Left: modulation of apoptotic response upon stable expression of sgRNAs targeting a panel of significantly enriched sgRNAs as determined through CaLR screening in MOLM14 cells. The percentage of apoptosis is determined by annexin V and propidium iodide (PI) staining of MOLM14 cells stably infected with individual sgRNAs and treated with 0.25 μM Ara-C for 72 hr. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$. Right: representative flow cytometry plots of annexin V/PI staining intensities corresponding to two sgRNAs promoting survival versus nontargeting control.

(F) Immunofluorescence images (left) for DAPI and phospho-γH2A.X staining in MOLM14 cells stably infected with sgRNAs targeting the lncRNA genes shown and treated with 25 μM Ara-C for 24 hr. Staining is quantified in the right panel. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.

(G) Disease-free survival association with expression levels of GAS6-AS2 and AC008073.2, genes enriched in both non-coding CRISPRa screening and drug resistance-gene expression correlation analyses among patients treated with Ara-C therapy within the TCGA-LAML patient cohort. GAS6-AS2: VST expression level cutoff = 3.38 (low, n = 92; high, n = 29), log-rank test: p value = 0.035. AC008073.2: VST expression level cutoff = 4.39 (low, n = 93; high, n = 28), log-rank test: p value = 0.0026.

See also Figure S5 and Table S7.



(legend on next page)

(DDR). Indeed, we found that the lncRNA AL353148.1 affected DDR response following Ara-C treatment (Figure 4F).

For two of our candidate lncRNAs (*GAS6-AS2* and *AC008073.2*), not only were they identified as candidates in both our cell line analysis and functional CaLR screening, but higher expression levels of these lncRNA genes were also associated with poor prognosis and decreased disease-free survival in AML patients treated with Ara-C (Figure 4G).

Taken together, these data reify our screening process as a platform to identify clinically relevant lncRNAs that may modulate Ara-C cytotoxicity through targeting a number of cellular processes.

lncRNA GAS6-AS2 Regulates the GAS/AXL Signaling Axis

We next integrated our computational analysis with both the coding and non-coding functional screens. Statistical analysis demonstrated significant enrichment of 7 sense-antisense gene pairs (Chi-square test, $p < 2.2e{-}16$) (Figures 5A and S2B). Of the 7 cognate pairs identified, *GAS6/GAS6-AS2* appeared to be one of the best candidate pairs for further analysis as both were highly enriched, *GAS6* is already known to play an important role in drug resistance in cancer, including AML (Figure S6A), while the role and function of *GAS6-AS2* remains unknown.

To confirm the on-target effect of our CRISPRa sgRNA, we overexpressed 8 different *GAS6-AS2*-targeting sgRNAs (Figure S6B). As expected, the majority of these sgRNAs led to a significant increase in cell survival (Figure 5B). Importantly, we found a strong correlation between the levels of *GAS6-AS2* activation and the resistance to Ara-C, indicating a dose-dependent specific effect of *GAS6-AS2* (Figures 5B and 5C). Similarly, expression of the strong inducer sgRNAs #1 and #3 promoted decreased sensitivity to Ara-C (Figures 5C and 5D) and a potent ability to reduce apoptosis (Figure 5E). Thus, the *GAS6-AS2* lncRNA appears to be a *bona fide* promoter of Ara-C resistance.

AML is known to develop as a multi-clonal disease, and resistant clones are frequently observed in early stages of the disease. The selective pressure of treatment leads to rapid clonal evolution and the emergence of resistant clones. Indeed, following Ara-C treatment of a mixed population of two MOLM14 cells, one expressing high GAS6-AS2 (labeled with red florescent protein) and the other expressing a non-targeting sgRNA (labeled with blue florescent protein), the GAS6-AS2 expressing clone emerged as dominant and was significantly enriched post-treatment (Figure 5F). These results were also confirmed *in vivo* as outlined in Figure 5G, with NSG mice engrafted with equal numbers of GAS6-AS2 overexpressing and control MOLM14 cells. Analysis of the bone marrow found significant enrichment ($p = 0.002$) of the GAS6-AS2-red cells (red/blue cells = 10.9 ± 6.4) (Figures 5H, 5I, S6D, and S6E). Importantly, within a non-treated cohort, both populations of cells were present in an equal ratio, demonstrating that GAS6-AS2 did not exert a proliferative advantage (Figures 5H and 5I). Furthermore, mice transplanted with GAS6-AS2 overexpressing cells alone had a greater tumor burden post-Ara-C treatment as compared with control cells (Figure S6F).

Several lncRNAs have been shown to exert their functional role by *cis*-regulation of neighboring genes (Rinn and Chang, 2012), and this is further supported by our genome wide analysis of sense-antisense cognate gene pairs (Figures 1E, 1F, and 5A). As *GAS6-AS2* lies in an antisense head-to-head manner with *GAS6* (Figure S6G), we hypothesized that the *GAS6/GAS6-AS2* cognate gene pair may function in this manner. Importantly, *GAS6-AS2* displayed nuclear (and cytoplasmic) localization (Figure S6H), suggesting that it may have the potential to regulate the *GAS6* locus. GAS6 is an important ligand for the TYRO3-AXL-MERTK (TAM) receptor tyrosine kinase signaling axis, controlling known pro-survival signals in AML (Wu et al., 2017). Indeed, up-regulation of GAS6/TAM signaling strongly correlates with resistance to chemotherapy and is a predictor of poor survival (Hong et al., 2008). In line with our hypothesis, GAS6 expression levels were found to be strongly correlated with *GAS6-AS2* expression upon CRISPRa modulation (Figure 6A). In addition, we also observed a striking correlation between these cognate gene pairs across the diverse 760 CCLE cell line panel (Pearson's $R = 0.8762$, $p < 2.2e{-}16$) (Figure 6B), as well as for a diverse set of primary human cancer types including AML (Figures 6C and S7A).

Activation of the GAS6/TAM pathway has been reported to promote MAPK, JAK/STAT, and NF-κB signaling (Schoumacher

Figure 5. GAS6-AS2 Promotes Drug Resistance *In Vitro* and *In Vivo*

(A) Integration of drug resistance-gene expression correlative analysis and forward genetic screenings identifies seven sense-antisense gene pairs that pass all significance thresholds, a higher number than expected by chance alone (Chi-square test: $p = 9.85e{-}7$).

(B) Fold change (FC) of MOLM14 cell viability treated with 0.25 μM Ara-C for 48 hr. Cells expressing individual sgRNAs targeting *GAS6-AS2*. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(C) Pearson correlation between cell viability versus *GAS6-AS2* expression level for each of the 8 sgRNAs targeting *GAS6-AS2*.

(D) Ara-C efficacy measurements in MOLM14 cells expressing sgRNAs #1 and #3 targeting GAS6-AS2 based on normalized MTS reads following 48 hr of treatment. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(E) Left: representative flow cytometry data of MOLM14 cells expressing either control or *GAS6-AS2*-targeting sgRNAs, treated with 25 μM Ara-C for 24 hr and labeled with viability (propidium iodide [PI]) and apoptotic (annexin V) markers. Right: percentage of apoptosis determined from quantification of staining results. Data are represented as mean ± SD, n > 3, Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(F) Competition assay between populations of MOLM14 control-Blue and MOLM14 GAS6-AS2-Red following 25 μM Ara-C treatment. Left: representative flow cytometry plots. Right: ratios between red and blue cells over time. Data are represented as mean ± SD, n > 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(G) Schematic of an orthotopic xenograft competition assay between control (blue) and GAS6-AS2 (red) MOLM14 cells with Ara-C treatment.

(H) Ratios of control (blue) versus GAS6-AS2 (red) MOLM14 cells from bone marrow of mice treated and analyzed at day 17 as outlined in (G).

(I) Representative flow cytometry results of cells harvested from mouse bone marrow 17 days following transplantation and treatment with vehicle or Ara-C for 5 days.

See also Figure S6.

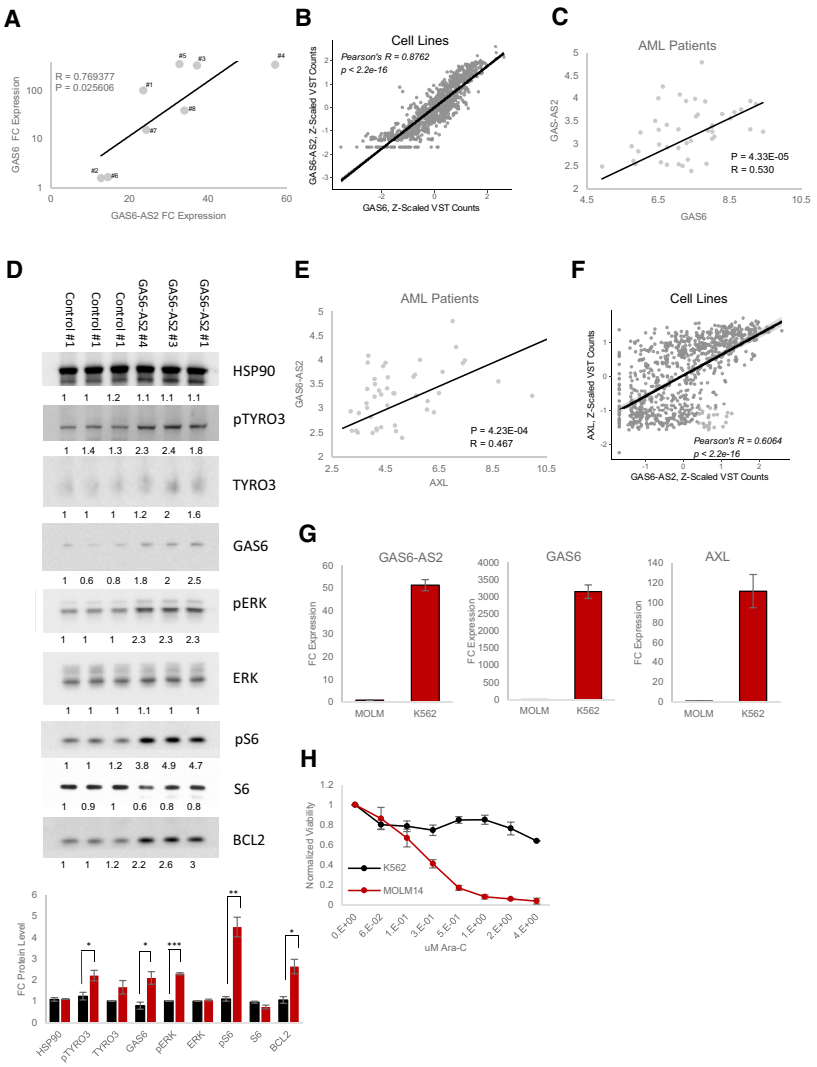


Figure 6. GAS6-AS2 Activates GAS6/TAM Signaling

(A) Pearson correlation between GAS6-AS2 and *GAS6* expression levels following GAS6-AS2 activation. Data are represented as mean of triplicate measurements.

(B) Pearson correlation between GAS6-AS2 and GAS6 expression levels across the 760 cancer cell lines analyzed (Figures 1A and 1B).

(C) Pearson correlation between GAS6-AS2 and GAS6 expression levels in AML patient samples.

(D) Western blot analysis of differential GAS6/TAM signaling activation in response to individual control or GAS6-AS2 sgRNA overexpression.

(E) Pearson correlation between GAS6-AS2 and AXL expression levels in AML patient samples.

(F) Pearson correlation between GAS6-AS2 and AXL expression levels across the 760 cancer cell lines analyzed (Figures 1A and 1B).

(G) Expression levels of GAS6-AS2, GAS6, and AXL in MOLM14 and K562 cell lines.

(H) Ara-C efficacy measurements in MOLM14 and K562 cell lines, based on normalized MTS reads following 48 hr treatment with the indicated concentrations of Ara-C. Data are represented as mean ± SD, n = 3.

See also Figure S7.

and Burbridge, 2017), with both MEK-ERK and S6K-RPS6 signaling axes being known downstream targets of TAM signaling (Xu et al., 2017). Western blot analysis of lysates from cells expressing three distinct *GAS6-AS2* targeting sgRNAs confirmed activation of the GAS6/TAM pathway (Figure 6D). Importantly, both pERK and pRPS6 were strongly phosphorylated in response to GAS6-AS2 activation (Figure 6D).

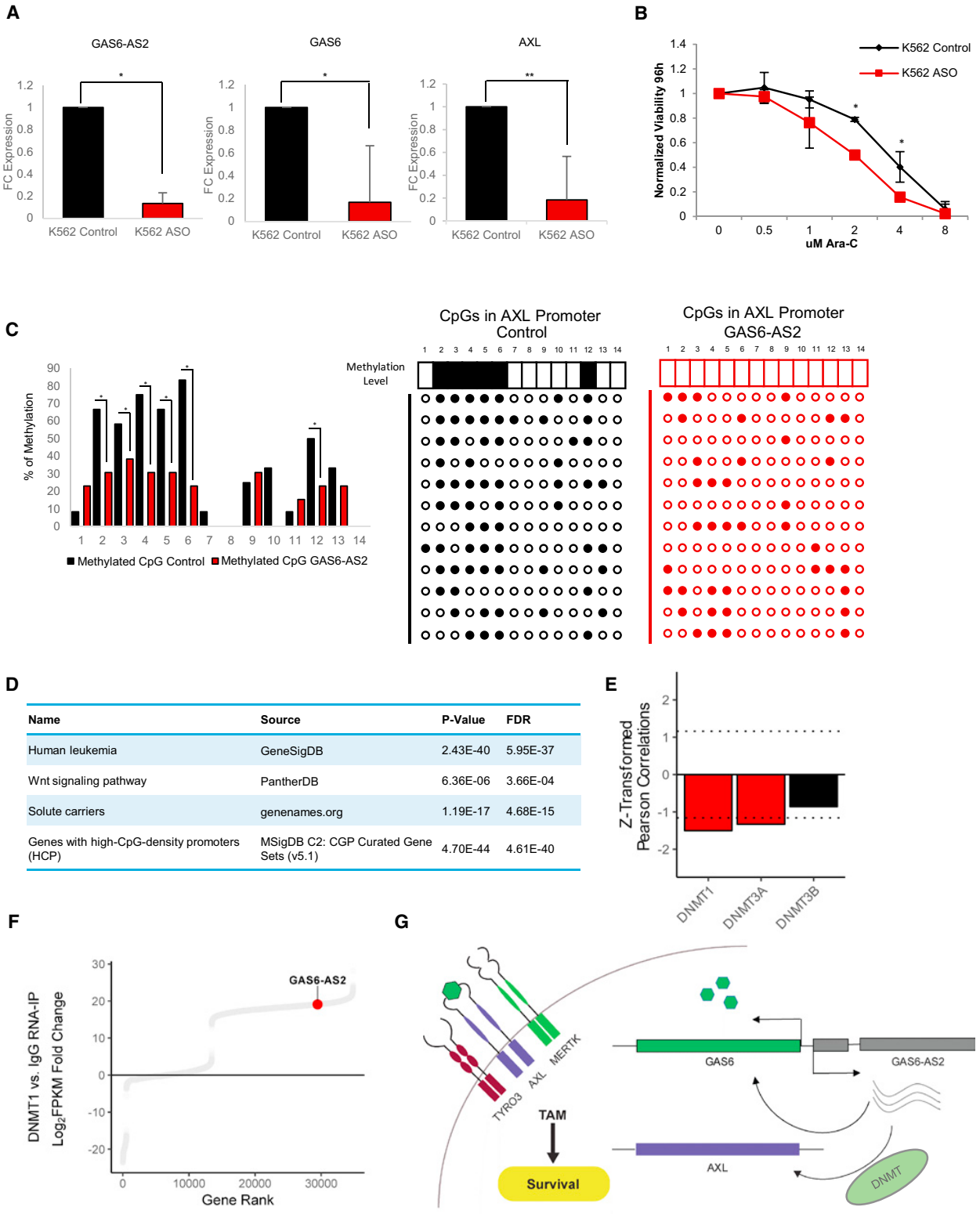
Surprisingly in a variety of cancer subtypes, including AML, GAS6-AS2 expression levels share strong correlations not only with GAS6 but also to its target receptor AXL (Figures 6E and S6B). AXL correlation was also mirrored in our 760 CCLE cell line panel (Pearson's R = 0.6064, p < 2.2e−16) (Figure 6F) and by overexpression of GAS6-AS2 in both MOLM14 and HEK293 cells *in vitro* (Figure S7C). These data suggested that GAS6-AS2 may be able to regulate the TAM receptor signaling axis at a number of levels.

To further investigate the role of GAS6-AS2 in regulating GAS6 and AXL, we took advantage of the K562 leukemia cell line, which we found to express high levels of GAS6-AS2, GAS6, and AXL relative to MOLM14 (Figure 6G) and that we demonstrate to be highly resistant to Ara-C treatment (Figure 6G). Interestingly, knockdown of GAS6-AS2 using two specific locked nucleic acid (LNA)-enhanced antisense oligonucleotides (ASOs) led to a significant decrease in both GAS6 and AXL mRNA levels (Figures 7A, S7I, and S7J) as well as an increased sensitivity of K562 cells to the activity of Ara-C (Figure 6B).

Previous studies found that AXL transcription is regulated by methylation of CpGs upstream of its TSS (Mudduluru and Allgayer, 2008). Direct methylation analysis using a bisulfite assay identified 6 highly methylated sites in the AXL promoter. Correspondingly, GAS6-AS2 overexpressing cells show significant decreases in methylation of these CpG sites (Figure 7C), suggesting that GAS6-AS has the potential to act in both a *cis*- and *trans*-acting manner.

To characterize the global function of GAS6-AS2 in cancer, we performed an unbiased *k*-means clustering based on coding and non-coding gene expression across 53 AML patients (Garzon et al., 2014) (Figure S7D). A large number of genes known to be regulated by promoter methylation were clustered together with GAS6-AS2 (Figure 7D), supporting our hypothesis that GAS-AS2 mediates CpG modification.

Based on these results, we hypothesized that GAS6-AS2 *trans*-activity may function through a DNA methyltransferase. An unbiased screening of our CCLE panel for candidate DNA methyltransferases that would correlate with Ara-C sensitivity identified decreased expression of DNMT1 and DNMT3A (Figure 7E). Importantly, we observed GAS6-AS2 to be significantly enriched in RNAs bound to DNMT1 using a DNMT1



Name	Source	P-Value	FDR
Human leukemia	GeneSigDB	2.43E-40	5.95E-37
Wnt signaling pathway	PantherDB	6.36E-06	3.66E-04
Solute carriers	genenames.org	1.19E-17	4.68E-15
Genes with high-CpG-density promoters (HCP)	MSigDB C2: CGP Curated Gene Sets (v5.1)	4.70E-44	4.61E-40

Figure 7. GAS6-AS2 Demonstrates *trans*-Regulation of AXL

(A) Fold change (FC) of GAS6-AS2, GAS6, and AXL in response to GAS6-AS2 knockdown via ASO in K562 cells. Data are represented as mean ± SD, n = 3. Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(legend continued on next page)

RNA-IP sequencing (RIP-seq) dataset (Figure 7F) (Di Ruscio et al., 2013). This suggests that GAS6-AS2 mediates *trans*-regulation of *AXL* by coordinating activity of DNMT proteins at the *AXL* promoter.

Thus, our data support a model whereby increased transcription and expression of *GAS6-AS2* promotes upregulation of both the GAS6 ligand and its TAM receptors to promote cellular survival and resistance to Ara-C treatment in AML (Figure 7G).

DISCUSSION

Although thousands of lncRNAs have now been detected and annotated in the human genome, the need to characterize their functions remains a critical challenge. Here, we developed a global approach to integrate computational analysis of cell line pharmacogenomic datasets with functional CRISPRa screens targeting coding and non-coding genes. This approach aimed to uncover integrated mechanisms regulating normal cellular homeostasis and disease and was applied to identifying functional lncRNAs modulating the cytotoxic effect of Ara-C, a front-line chemotherapy agent frequently used in the treatment of AML patients.

Because lncRNAs are poorly annotated, we developed a bioinformatic framework to facilitate the prioritization of candidate genes by their functional and physiological relevance. Using pharmacogenomic and transcriptomic data, we obtained a list of coding and non-coding genes whose expression levels are associated with cellular response to Ara-C. Our list identified many of the coding genes and pathways previously shown to regulate the response to Ara-C treatment. In addition, not only did this analysis reveal a large number of lncRNAs to influence response to Ara-C, but it also implicated a pattern of *cis*-regulation by lncRNAs on their adjacent cognate coding genes. Thus, this analysis provides us with a unique resource that can both deliver a wealth of novel predictive biomarkers for response to therapy and prioritize functionally relevant genes identified through functional screening.

For our purposes, functional screening was carried out with CRISPRa-based technologies using both an established protein-coding sgRNA library (Konermann et al., 2015) and a new genome wide non-coding sgRNA (CaLR) library. In adapting the CRISPRa technology, we found that appropriate cell models and optimization are critical. For our new CaLR library, we chose to generate sgRNAs targeting lncRNA genes that are well annotated. This, in turn, enabled extensive promoter coverage upstream of TSSs and targeting of multiple TSSs for individual lncRNA genes. In addition, this screening approach offers the advantage of driving lncRNA overexpression from the endogenous genomic locus, enabling us to capture *cis*-acting and nuclear lncRNA functions, which cannot be readily studied by traditional overexpression approaches (Shechner et al., 2015). It should be noted, however, that as some genes harbor other small non-coding RNAs including microRNAs and small nucleolar RNAs (snoRNAs) within intronic regions, driving expression from the endogenous promoter may also result in their expression.

Indeed, our novel CaLR approach identified lncRNAs that facilitate resistance to Ara-C treatment. These data demonstrate that many lncRNA genes are functionally relevant for cancer and modulate distinct cellular programs. Integrating coding and non-coding screening approaches also allows us to categorize lncRNA genes by function, and although we have applied these libraries to the identification of novel genes involved in chemotherapeutic resistance, this platform alone can be applied to the functionalization of lncRNAs across a wide range of biological questions.

Thus, complete integration of computational cell line analysis, coding/non-coding CRISPRa screens, and patient outcome data resulted in the discovery of a distinct set of 7 cognate sense-antisense gene pairs. Of these seven pairs, we pursued the *GAS6/GAS6-AS2* cognate gene pair as drivers of resistance to Ara-C in AML. We found that GAS6-AS2 functions through *cis*-regulation of its adjacent cognate gene, coding for the *GAS6* ligand, as well as the *trans*-regulation of its receptor *AXL* to drive aberrant downstream signaling of this pathway.

While each of these approaches as individual modules (computational and screening) has been shown to be useful to identify genes regulating specific cellular processes, each harbors inherent limitations and bias requiring extensive validation of hits. However, our integrated approach described here serves as a more powerful framework for the screening and discovery of protein-coding and non-coding networks regulating biological processes, thereby providing a resource to facilitate improvements toward the annotation and functionalization of non-coding RNAs at large. Our analysis suggests that there are a substantial number of coding and non-coding genes that, at a minimum, serve as predictive biomarkers that correlate with differential Ara-C responses and may serve as therapeutic targets for the tuning of Ara-C response through modulation of their expression levels. Indeed, this approach may facilitate the identification of novel high confidence and clinically relevant therapeutic opportunities across a broad spectrum of human diseases.

(B) Modulation of Ara-C response upon GAS6-AS2 knockdown via ASO in K562 cells. Data are represented as mean ± SD, n = 3, Welch two-sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(C) Methylation of CpG islands in the HEK293T AXL promoter following modulation of GAS6-AS2 expression. n = 12, Chi-square test: *p < 0.05. **p < 0.01, ***p < 0.001.

(D) Gene ontology analysis of coding genes clustered with GAS6-AS2 as determined by *k*-means clustering (cluster #6).

(E) Drug sensitivity-gene expression Pearson correlation values of DNA methyltransferases. Genes enriched beyond a *Z* score threshold of ± 1.16 are colored in red. See also Figure 1B.

(F) Distribution of FPKM-normalized transcript abundances associated with DNMT1 versus IgG.

(G) Model summarizing the mechanism by which GAS6-AS2 regulates GAS6/TAM signaling.

See also Figure S7.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Guide RNA cloning
  - RNA isolation and RT-qPCR
  - Determination of IC_{50}
  - Cellular proliferation and titer
  - Cell cycle analysis
  - Apoptosis
  - Immunofluorescence for detection of γH2A.X foci
  - Nuclear/cytoplasmic fractionation
  - Xenotransplantation of human leukemic cells
  - Western blotting
  - Sodium bisulfite conversion of DNA
  - Human cancer cell line analysis
  - Disease-free survival analysis
  - Human lncRNA CRISPRa library design
  - Human lncRNA CRISPRa library chip design and synthesis
  - Lentivirus production
  - Pooled lncRNA CRISPRa screening for Ara-C response
  - Pooled protein-coding CRISPRa screening for Ara-C response
  - CRISPRa screening deconvolution and analysis
  - Transcript abundances in AML
  - Global pathway annotation of enriched lncRNAs
  - K-means clustering of RNaseq from AML patient samples
  - Pan-cancer co-expression analysis
  - RNA-IP analysis
- QUANTITATIVE AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.03.052.

ACKNOWLEDGMENTS

The authors would like to thank Rory Kirchner of the Harvard Chan Bioinformatics Core, Harvard T.H. Chan School of Public Health, Boston, MA for assistance with computational analysis. The authors would like to thank Davide Corà and Claudio Isella of the Department of Oncology, and GenoBiToUS, Genomics and Bioinformatics Service, University of Turin, Italy for assistance with computational analysis. We thank the P.P.P. laboratory members for critical discussions. This work was supported by EMBO Long-Term Fellowship (ALTF 318-2013) and Fulbright awards to A.C.B. A.C. was funded by NCI (5T32CA009216-34) and a Burroughs Wellcome Fund Career Award for Medical Scientists. G.M.C was supported by NIH (RM1 HG008525 and P50 HG005550). P.P.P. is supported by an NIH NCI R35 grant (CA197529) and through support from the Ludwig Center at Harvard. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the NIH. The data used for the analyses described in this manuscript were obtained from the GTEx Portal. The results published here are fully or partially based on data generated by the Cancer Target Discovery and Development (CTD^2) Network established by the NCI Office of Cancer Genomics.

AUTHOR CONTRIBUTIONS

A.C.B. conceived the project, designed and performed experiments, interpreted results, and co-wrote the manuscript. A.C. designed and created the CaLR library, designed and performed experiments, and interpreted results. Y.-R.L., D.N., and J.V. performed experiments. S.V. and M.S. designed the CaLR library. A.C.B., J.D.L., E.M., and P.P. carried out computational analysis. J.L.R. and G.M.C. supervised experimental designs. J.G.C. designed and supervised experiments, interpreted results, and co-wrote the manuscript. P.P.P. conceived the project, supervised experimental designs, interpreted results, and co-wrote the manuscript. A.C.B., J.G.C., and P.P.P. provided overall project leadership.

DECLARATION OF INTERESTS

G.M.C. is a founder and advisor for Editas Medicine. G.M.C. has equity in Editas and Caribou Biosciences (for full disclosure list, please see http://arep.med.harvard.edu/gmc/tech.html). All other authors declare no competing interests.

Received: August 13, 2017
Revised: December 10, 2017
Accepted: March 21, 2018
Published: April 19, 2018

REFERENCES

Abraham, A., Varatharajan, S., Karathedath, S., Philip, C., Lakshmi, K.M., Jayavelu, A.K., Mohanan, E., Janet, N.B., Srivastava, V.M., Shaji, R.V., et al. (2015). RNA expression of genes involved in cytarabine metabolism and transport predicts cytarabine response in acute myeloid leukemia. Pharmacogenomics *16*, 877–890.

Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., et al. (2016). The Ensembl gene annotation system. Database (Oxford) *2016*, baw093.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell *154*, 1151–1161.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, J., Xu, H., Aronow, B.J., and Jegga, A.G. (2007). Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics *8*, 392.

Cheng, Z., Gong, Y., Ma, Y., Lu, K., Lu, X., Pierce, L.A., Thompson, R.C., Muller, S., Knapp, S., and Wang, J. (2013). Inhibition of BET bromodomain targets genetically diverse glioblastoma. Clin. Cancer Res. *19*, 1748–1759.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371–376.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

Farge, T., Saland, E., de Toni, F., Aroua, N., Hosseini, M., Perry, R., Bosc, C., Sugita, M., Stuani, L., Fraisse, M., et al. (2017). Chemotherapy-resistant human acute myeloid leukemia cells are not enriched for leukemic stem cells but require oxidative metabolism. Cancer Discov. *7*, 716–735.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

Garzon, R., Volinia, S., Papaioannou, D., Nicolet, D., Kohlschmidt, J., Yan, P.S., Mrózek, K., Bucci, D., Carroll, A.J., Baer, M.R., et al. (2014). Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc. Natl. Acad. Sci. USA *111*, 18679–18684.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. Cell *159*, 647–661.

Herold, N., Rudd, S.G., Ljungblad, L., Sanjiv, K., Myrberg, I.H., Paulin, C.B.J., Heshmati, Y., Hagenkort, A., Kutzner, J., Page, B.D.G., et al. (2017). Targeting SAMHD1 with the Vpx protein to improve cytarabine therapy for hematological malignancies. Nat. Med. *23*, 256–263.

Hon, C., Ramilowski, J., Harshbarger, J., Bertin, N., Rackham, O., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. Nature *543*, 199–204.

Hong, C.C., Lay, J.D., Huang, J.S., Cheng, A.L., Tang, J.L., Lin, M.T., Lai, G.M., and Chuang, S.E. (2008). Receptor tyrosine kinase AXL is induced by chemotherapy drugs and overexpression of AXL confers drug resistance in acute myeloid leukemia. Cancer Lett. *268*, 314–324.

Hu, X., Feng, Y., Zhang, D., Zhao, S.D., Hu, Z., Greshock, J., Zhang, Y., Yang, L., Zhong, X., Wang, L.-P., et al. (2014). A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. Cancer Cell *26*, 344–357.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. Nat. Med. *21*, 1253–1261.

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–1120.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., et al. (2017a). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017b). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

Kaimal, V., Bardes, E.E., Tabar, S.C., Jegga, A.G., and Aronow, B.J. (2010). ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. Nucleic Acids Res. *38*, W96-102.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–D205.

Kassambara, A., and Kosinski, M. (2017). Survminer: drawing survival curves using 'Ggplot2'. https://CRAN.R-project.org/package=survminer.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Lamba, J.K. (2009). Genetic factors influencing cytarabine therapy. Pharmacogenomics *10*, 1657–1674.

Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A., Hoadley, K., Triche, T.J., Jr., Laird, P.W., Baty, J.D., et al.; Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. *368*, 2059–2074.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., Weinstein, J.N., and Liang, H. (2015). TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. Cancer Res. *75*, 3728–3737.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*, eaah7111.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Mudduluru, G., and Allgayer, H. (2008). The human receptor tyrosine kinase Axl gene–promoter characterization and regulation of constitutive expression by Sp1, Sp3 and CpG methylation. Biosci. Rep. *28*, 161–176.

Perron, U., Provero, P., and Molineris, I. (2017). In silico prediction of lncRNA function using tissue specific and evolutionary conserved expression. BMC Bioinformatics *18* (*Suppl 5*), 144.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. *17*, 47–62.

Ramos, N.R., Mo, C.C., Karp, J.E., and Hourigan, C.S. (2015). Current approaches in the treatment of relapsed and refractory acute myeloid leukemia. J. Clin. Med. *4*, 665–695.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44* (W1), W83-9.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Saland, E., Boutzen, H., Castellano, R., Pouyet, L., Griessinger, E., Larrue, C., de Toni, F., Scotland, S., David, M., Danet-Desnoyers, G., et al. (2015). A robust and rapid xenograft model to assess efficacy of chemotherapeutic agents for human acute myeloid leukemia. Blood Cancer J. *5*, e297.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schmitt, A.M., and Chang, H.Y. (2016). Long noncoding RNAs in cancer pathways. Cancer Cell *29*, 452–463.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schneider, C., Oellerich, T., Baldauf, H.-M., Schwarz, S.-M., Thomas, D., Flick, R., Bohnenberger, H., Kaderali, L., Stegmann, L., Cremer, A., et al. (2017). SAMHD1 is a biomarker for cytarabine response and a therapeutic target in acute myeloid leukemia. Nat. Med. *23*, 250–255.

Schoumacher, M., and Burbridge, M. (2017). Key roles of AXL and MER receptor tyrosine kinases in resistance to multiple anticancer therapies. Curr. Oncol. Rep. *19*, 19.

Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. https://doi.org/10.1101/060012.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat. Methods *12*, 664–670.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431–432.

Stavropoulou, V., Kaspar, S., Brault, L., Sanders, M.A., Juge, S., Morettini, S., Tzankov, A., Iacovino, M., Lau, I.-J., Milne, T.A., et al. (2016). MLL-AF9 expression in hematopoietic stem cells drives a highly invasive AML expressing EMT-related genes linked to poor outcome. Cancer Cell *30*, 43–58.

Steelman, L.S., Pohnert, S.C., Shelton, J.G., Franklin, R.A., Bertrand, F.E., and McCubrey, J.A. (2004). JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis. Leukemia *18*, 189–218.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Therneau, T. (2015). A package for survival analysis in S. version 2.38. https://CRAN.R-project.org/package=survival.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wu, G., Ma, Z., Hu, W., Wang, D., Gong, B., Fan, C., Jiang, S., Li, T., Gao, J., and Yang, Y. (2017). Molecular insights of Gas6/TAM in cancer development and therapy. Cell Death Dis. *8*, e2700.

Xu, L., Chen, Y., Dutra-Clarke, M., Mayakonda, A., Hazawa, M., Savinoff, S.E., Doan, N., Said, J.W., Yong, W.H., Watkins, A., et al. (2017). BCL6 promotes glioma and serves as a therapeutic target. Proc. Natl. Acad. Sci. USA *114*, 3981–3986.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti P-H2A.X	Cell Signaling Technology	9718 (RRID:AB_2118009)
Rabbit anti Tyro3	Cell Signaling Technology	5585 (RRID:AB_10706782)
Rabbit anti P-MER/SKY	Invitrogen	PA5-37808 (RRID:AB_2554416)
Goat anti Gas6	R&D	AF885 (RRID:AB_2108079)
Rabbit pAB anti Gas6	abcam	Ab136249
Rabbit anti Bcl-2	Cell Signaling Technology	2872 (RRID:AB_10693462)
Mouse anti BrdU	Cell Signaling Technology	5292 (RRID:AB_10548898)
Rabbit anti p44/42 MAPK	Cell Signaling Technology	9102 (RRID:AB_330744)
Rabbit anti P-p44/42 MAPK	Cell Signaling Technology	4376 (RRID:AB_331772)
Rabbit anti S6 Ribosomal Protein	Cell Signaling Technology	2217
Rabbit anti P-S6 Ribosomal Protein	Cell Signaling Technology	2211 (RRID:AB_2721245)
Alexa Fluor Goat anti-mouse 488	Invitrogen	A-11001 (RRID:AB_2534069)
Bacterial and Virus Strains		
One Shot Stbl3	Invitrogen	C7373-03
lenti dCAS-VP64_Blast	Konermann et al., 2015	Addgene #61425
lenti MS2-P65-HSF1_Hygro	Konermann et al., 2015	Addgene #61426
Human CRISPR Activation Library	Konermann et al., 2015	Addgene #1000000057
pSB700 Cerulean-Zhang2.0	Chavez et al., 2016	Addgene #79378
pCDH-puro-Bcl2	Cheng et al., 2013	Addgene #46971
pMD2.G	Didier Trono	Addgene #12259
psPAX2	Didier Trono	Addgene #12260
shDCK	Sigma-Aldrich	TRCN0000009934
shDCK	Sigma-Aldrich	TRCN0000196649
shDCK	Sigma-Aldrich	TRCN0000196362
Chemicals, Peptides, and Recombinant Proteins		
5-Bromo-2-deoxyuridine (BrdU)	Sigma-Aldrich	B5002
Cytarabine	Sigma-Aldrich	PHR1787
FxCycle PI/RNase Staining Solution	Invitrogen	F10797
PureLinc DNase	Invitrogen	12185-010
Hygromycin B Gold	InvivoGen	Ant-hg-1
Zeocin	InvivoGen	Ant-zn-1
Blasticidin	InvivoGen	Ant-bl-1
Puromycin	Gibco	A11138-03
FastDigest Esp3I	Thermo Scientific	#FD0454
APC-Annexin V	BioLegend	640920
NEBNext® High-Fidelity 2X PCR Master Mix	New England Biolabs	#M0541
ATP	New England Biolabs	#B0756A
10X T4 Ligation Buffer	New England Biolabs	#B0202
T7 DNA ligase	New England Biolabs	#M0318
T4 PNK	New England Biolabs	#M0201
Critical Commercial Assays		
qPCR Lentivirus Titration Kit	abm	LC900
CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay	Promega	G5421

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TOPO TA Cloning Kit for Sequencing	Invitrogen	45-0030
EZ DNA Methylation-Direct Kit	ZYMO RESEARCH	D5020
QIAprep Spin Miniprep Kit	QIAGEN	27106
PureLink RNA Mini Kit	Invitrogen	12183025
PureLink HiPure Plasmid Filter Maxiprep	Invitrogen	K2100-17
RETROscript Kit	Invitrogen	AM1710
Experimental Models: Cell Lines		
Human: MOLM14 (Male)	N/A	N/A
Human: HL60 (Female)	N/A	N/A
Human: K562 (Female)	N/A	N/A
Human: HEK293T (Female)	N/A	N/A
Experimental Models: Organisms/Strains		
NSG (6-8 weeks old) mice (Female)	Jackson Laboratory	005557
Oligonucleotides		
GapmeR GAS6-AS2 #1 (ID 174250)	Exiqon	300600
GapmeR GAS6-AS2 #2 (ID 174298)	Exiqon	300600
GapmeR MALAT1	Exiqon	300600
GapmeR Negative Control A	Exiqon	300610
sgRNA TUNA	Invitrogen	GGCGGCGTCGGGGTCCCTAC
sgRNA TTN	Invitrogen	CCTTGGTGAAGTCTCCTTTG
sgRNA MIAT	Invitrogen	ATGCGGGAGGCTGAGCGCAC
sgRNA RHOXF2	Invitrogen	ACGCGTGCTCTCCCTCATC
sgRNA ASCL1	Invitrogen	CGGGAGAAAGGAACGGGAGG
sgRNA HBG	Invitrogen	CTTGACCAATAGCCTTGACA
sgRNA Control 1	Invitrogen	GCAGCTCGACCTCAAGCCGT
sgRNA Control 2	Invitrogen	GTACTCCAATCCGCGATGAC
sgRNA Control 3	Invitrogen	GTCGGTGATCCAACGTATCG
sgRNA Control 4	Invitrogen	GCGCCTTAAGAGTACTCATC
sgRNA Control 5	Invitrogen	GTATGGTGAGTAGTCGCTTG
CaLR sgRNA (See Table S5 for sequences)	Invitrogen	N/A
PCR sequencing primers (See Table S7 for oligos used)	Invitrogen/IDT	N/A
qPCR primers (See Table S7 for oligos used)	Invitrogen	N/A
Deposited Data		
CTRP Ara-C Drug Sensitivity Data	Basu et al., 2013; Rees et al., 2016	https://portals.broadinstitute.org/ctrp/
CCLE Gene Expression Data	Barretina et al., 2012	https://portal.gdc.cancer.gov/
DNMT1 IP- and IgG IP Sequencing Data	Di Ruscio et al., 2013	SRR358674, SRR358675
AML Patient Sample Gene Expression Data	Garzon et al., 2014	SRP050272
CRISPRa sgRNA Counts	This Paper	Tables S4 and S6
Software and Algorithms		
FlowJo 10	FlowJo	N/A
FlowLogic	Miltenyi Biotec	N/A
KALUZA	Beckman Coulter	N/A
Fiji imageJ	LOCI	Schindelin et al., 2012; Schneider et al., 2012
Image Lab	BIO RAD	N/A
R Statistical Software	https://www.r-project.org/	N/A
Picard	http://broadinstitute.github.io/picard/	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STAR	https://github.com/alexdobin/STAR	Dobin et al., 2013
SAMtools	http://samtools.sourceforge.net/	Li et al., 2009
Subread	http://subread.sourceforge.net/	Aken et al., 2016; Liao et al., 2014
biomaRt	http://bioconductor.org/packages/release/bioc/html/biomaRt.html	Durinck et al., 2005
DESeq2	http://bioconductor.org/packages/release/bioc/html/DESeq2.html	Love et al., 2014
fgsea	Bioconductor http://bioconductor.org/packages/release/bioc/html/fgsea.html	Sergushichev, 2016
CRAN: gProfiler	https://cran.r-project.org/web/packages/gProfileR/index.html	Reimand et al., 2016
Cytoscape	http://www.cytoscape.org/	Smoot et al., 2011
TCGAbiolinks	http://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html	Colaprico et al., 2016
CRAN: survival package in R	https://cran.r-project.org/web/packages/survival/index.html	Therneau, 2015
CRAN: survminer	https://cran.r-project.org/web/packages/survminer/index.html	Kassambara and Kosinski, 2017
CutAdapt	https://cutadapt.readthedocs.io/en/stable/	Martin, 2011
MAGeCK	https://sourceforge.net/p/mageck/wiki/Home/	Li et al., 2014
Funcpred	http://www.funcpred.com/	Perron et al., 2017
ToppCluster web server	https://toppcluster.cchmc.org/	Kaimal et al., 2010
Human lncRNA CRISPRa gRNA Library Design	https://github.com/vorasaurus/Genome-Wide-CRISPR-Guide-Construction	N/A
Other		
BD LSR II flow cytometer	BD	N/A
GALLIOS FLOW CYTOMETER	Beckman Coulter	N/A
ChemiDoc Imaging System	BIO RAD	N/A
StepOnePlus	Applied Biosystems	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Dr. Pier Paolo Pandolfi (ppandolf@bidmc.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

MOLM14 (male) and K562 (female) cell lines were maintained in RPMI supplemented with 10% FBS, 2 mmol/L glutamine, and 100 U/mL penicillin/streptomycin (Invitrogen). HEK293T (female) cells were maintained in DMEM supplemented with 10% FBS, 2 mmol/L glutamine, and 100 U/mL penicillin/streptomycin (Invitrogen). HL60 (female) cells were maintained in IMDM supplemented with 10% FBS, 2 mmol/L glutamine, and 100 U/mL penicillin/streptomycin (Invitrogen).

Adult female NSG mice (6–8 weeks old) mice were obtained from Jackson Laboratory (Bar Harbor, ME). Animals were used in accordance with a protocol reviewed and approved by the Institutional Animal Care and Use Committee at BIDMC (Protocol #082-2014).

Ara-C (Sigma-Aldrich) was diluted in sterile water to a working concentration of 1 mM.

METHOD DETAILS

Guide RNA cloning

sgRNAs were cloned as previously published (Chavez et al., 2016). Briefly, the sgRNA expression vector pSB700 Cerulean-Zhang2.0 (Addgene #79378) was modified by Gibson assembly to contain the *pac* puromycin resistance gene. The modified vector was digested with BsmBI and was gel-purified. A pair of 25 nt oligos containing the appropriate overhang was then ligated into the vector by mixing 1 μL of the cut vector (normalized to 100-200 ng/μL), 0.5 μL of each primer at a 100 μM stock concentration, 2 μL of 10x T4 DNA ligase buffer, and 1 μL of T4 DNA ligase (NEB #M0202T) into a total ligation volume of 20 μL. Ligations were left at room temperature overnight, and 1 μL of the ligation product was subsequently transformed into 10 μL of NEB Stable cells (NEB #C3040I). Resulting colonies were verified by Sanger sequencing.

RNA isolation and RT-qPCR

Total RNA was purified from cell lines and tissues using the PureLink RNA Mini Kit (Invitrogen). For qPCR analysis, 2 μg of total RNA was reverse transcribed into cDNA using the RETROscript Kit (Invitrogen). SYBR-Green qPCR analysis was then performed using Applied Biosystems StepOnePlus in accordance with the manufacturer's protocol.

Determination of IC_{50}

To determine the half maximal inhibitory concentration (IC_{50}), we grew cells for 48 h at a range of concentrations of Ara-C (32 to 0.03 μM). Cell numbers were measured by CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (Promega). Signal was measured using a Glomax Multi+ (Promega) plate reader. The normalized measurements were used to obtain survival curves and IC_{50} values. For DCK knockdown we used MISSION shRNA (Sigma-Aldrich TRCN0000009934, TRCN0000196649, TRCN0000196362), which were transduced according to standard protocols. For BCL2 overexpression we used pCDH-puro-Bcl2 (Addgene #46971), which was transduced according to standard protocols. Two different LNA-based GapmeR (Exiqon) antisense oligonucleotides (ASOs) were designed to target GAS6-AS2 specifically. The ASOs were transfected at a concentration of 50 μM using Lipofectamine 2000 according to standard protocols.

Cellular proliferation and titer

To determine cellular proliferation rates, 1e5 cells were plated in 24-well plates. Each day, 100 μL of cell suspension was transferred to a well of a 96-well plate, and cell titer was measured by the CellTiter 96 Aqueous Non-Radioactive Cell Proliferation Assay (Promega). Signal was measured using a Glomax Multi+ (Promega) plate reader.

Cell cycle analysis

Cells were treated with 0.03 mg/mL BrdU (Sigma-Aldrich B5002-100MG), incubated at 37°C for 30 minutes, and fixed in 70% ethanol. Then, the cells were treated with 1.5 M HCl for 30 minutes at room temperature and then incubated in 100 μL of 0.5 g bovine serum albumin (BSA) in 100 mL 1x PBS for 10 minutes. Mouse anti-BrdU antibody (Cell Signaling Technology #5292) was added at a 1:200 dilution and then incubated for 1 h at room temperature. Cells were subsequently incubated with goat ant-mouse IgG (H+L)-Alexa Fluor 488 (Thermo Fisher A-11001) for 30 minutes at room temperature. Propidium iodide (PI) staining was performed on cells resuspended in FxCycle PI/RNase staining solution (Invitrogen F10797). Fluorescence was measured using a CytoFLEX flow cytometer (Beckman Coulter) and analyzed by FlowJo v10.

Apoptosis

Cells were centrifuged, media was removed, and cells were resuspended in 100 μL of Annexin V Binding Buffer (BD PharMingen) containing 1 μg/ml propidium iodide (Sigma-Aldrich) and APC-Annexin V (BioLegend) for 15 minutes. 400 μL of Annexin V Binding Buffer (BD PharMingen) was then added to the samples. Fluorescence was measured using a BD LSR II flow cytometer or Gallios flow cytometer (Beckman Coulter) and then analyzed using the FlowLogic software (Miltenyi Biotec) or FlowJo V10.

Immunofluorescence for detection of γH2A.X foci

Cells were fixed in 3.7% formaldehyde in PBS for 10 min, permeabilized with 0.5% Triton in PBS, and blocked with 5% BSA in PBS. Mouse anti-γH2A.X primary antibody was used (Cell Signaling Technology #9718). Appropriate Alexa Fluor 488-conjugated secondary antibodies were then added (Invitrogen). Images were taken with a Zeiss LSM 880 confocal microscope. At least 50 nuclei per condition were analyzed.

Nuclear/cytoplasmic fractionation

Fractionation of the cells to separate nucleus and cytoplasm was performed using PARIS-kit (Ambion) following the manufacturer's protocol. The efficiency of the separation of the two fractions was assessed by qPCR amplification with primers targeting the nucleus-specific lncRNA MALAT1, followed by 2% agarose gel electrophoresis in 1x Tris-Acetate-EDTA buffer.

Xenotransplantation of human leukemic cells

Animals were used in accordance with a protocol reviewed and approved by the Institutional Animal Care and Use Committee at BIDMC (Protocol #082-2014). NSG mice were obtained from Jackson Laboratory (Bar Harbor, ME). Mice were housed in sterile conditions using high-efficiency particulate arrestance-filtered micro-isolators and fed with irradiated food and acidified water. Adult mice (6–8 weeks old) were sublethally irradiated with 250 cGy of total body irradiation 24 h before injection of leukemic cells. Cultured AML cell lines were washed twice in phosphate-buffered saline (PBS), cleared of aggregates and debris using a 0.2-mm cell filter, and suspended in PBS at a final concentration of 0.2 million cells per 200 μl of PBS per mouse for intravenous injection. Xenograft tumors were generated by injecting AML cells (in 200 μl of PBS) in the tail vein of NSG mice. Daily monitoring of mice for symptoms of disease (ruffled coat, hunched back, weakness, and reduced motility) determined the time of killing for injected animals with signs of distress (Saland et al., 2015). 10 days following transplantation, NSG mice were treated with Ara-C (Sigma-Aldrich) at a dose of 30 mg/kg/day for 5 days. At day 17, bone marrow was harvested from the mice, and cells were analyzed using a BD LSR II flow cytometer (BD Bioscience). The results were analyzed using FlowJo v10.3 and Microsoft Excel 2016.

Western blotting

Cells were lysed in RIPA buffer (Boston BioProducts) supplemented with protease (Roche) and phosphatase (Roche) inhibitors. Laemmli buffer with a 5% final concentration of β-mercaptoethanol was added to the samples, which were then boiled, separated on NuPAGE 4%–12% Bis-Tris gradient gels (Invitrogen), and transferred to polyvinylidine difluoride membranes (Immobilon P, Millipore). Membranes were then probed with the indicated antibodies. Antibodies for western blotting: Anti-TYRO3 (5585), anti-ERK (9102), anti-phospho-ERK (9101), anti-phospho-S6 (2211), anti-S6 (2217), anti-STAT3 (4904), and anti-phospho-STAT3 (9136) antibodies were purchased from Cell Signaling Technology. Anti-Actin (A3853) was purchased from Sigma Aldrich. Anti-phospho-TYRO3/MERTK (PA5-37808) was purchased from Invitrogen.

Sodium bisulfite conversion of DNA

DNA methylation status was examined by methylation-specific PCR of genomic DNA treated with sodium bisulfite using the EZ DNA Methylation-Direct kit (Zymo Research). Primers were used as described (Mudduluru and Allgayer, 2008).

Human cancer cell line analysis

Ara-C sensitivity data from human cancer cell lines were obtained from the Cancer Target Discovery And Development (CTD^2) database (Basu et al., 2013; Rees et al., 2016), and cell lines with corresponding RNaseq profiles available from the Cancer Cell Line Encyclopedia (CCLE) Genomic Data Commons (GDC) legacy archive were compiled, resulting in a total of 760 cell lines with matched Ara-C area under the dose-response curve (AUC) values and RNAseq profiles (Barretina et al., 2012). AUC values were regressed on annotations describing tissue of origin and cancer subtype (''ccle_primary_site,'' ''ccle_primary_hist,'' ''ccle_hist_subtype_1'') using linear regression and then Z-scaled (Table S1).

RNAseq datasets were downloaded as pre-aligned .bam files. Bam files were converted to .fastq files using Picard's SamToFastq (v1.130). Resultant .fastq sequencing files were then aligned to the hg38 reference genome with default STAR single-pass alignment (v2.5.2b) (Dobin et al., 2013). Realigned .bam files were sorted by name using SAMtools v1.3 (Li et al., 2009). Coordinates were mapped to Ensembl gene IDs with featureCounts (Subread v1.5.2) as denoted by the Ensembl v89 annotation using features ''-p -B -s 0'' (Aken et al., 2016; Liao et al., 2014).

Ensembl IDs were converted to gene names using the biomaRt package in R (Durinck et al., 2005). Genes were demarcated as either ''protein-coding'' or ''noncoding'' if the annotated gene_biotype was either ''protein_coding'' or one of ''lincRNA,'' ''antisense,'' ''processed_transcript,'' ''transcribed_processed_pseudogene,'' ''transcribed_unitary_pseudogene,'' ''transcribed_unprocessed_pseudogene,'' ''macro_lncRNA,'' ''scRNA,'' ''TEC,'' ''3prime_overlapping_ncRNA,'' ''sense_intronic,'' ''bidirectional_promoter_lncRNA,'' ''sense_overlapping,'' or ''unprocessed_pseudogene,'' respectively, to maintain consistency with noncoding annotations as obtained from analysis of the CaLR library annotations for downstream analysis. Transcript counts per gene were normalized by DESeq2, and genes with arithmetic means lower than 0.5 were filtered from further analysis (Love et al., 2014). Normalized counts were subjected to the varianceStabilizingTransformation (VST) function from DESeq2 prior to downstream analysis.

Pearson correlation coefficients between Ara-C lineage-regressed AUC values and VST-transformed gene expression levels were calculated and then Z-scaled, similarly as previously described (Rees et al., 2016) (Table S1). A significance threshold was determined by receiver operatic characteristic (ROC) analysis using a curated set of genes associated with differential Ara-C response as true positives (Table S2), which identified an optimal threshold of Z = 1.16. Gene set enrichment analysis was performed using fgsea (Sergushichev, 2016) using pathway annotations from the Kyoto Encyclopedia of Genes and genomes (KEGG) (Kanehisa and Goto, 2000). Pearson correlation coefficients to quantify co-expression between genes were assessed using VST-transformed gene expression levels. The threshold for significant Pearson correlation between gene pairs was determined as the Pearson correlation coefficient associated with the 99^{th} percentile among 5,000 random protein-coding/noncoding gene pairs. Sense-antisense cognate gene pairs were selected based on the presence of ''-AS'' in the noncoding gene symbol and the detection of the corresponding coding gene.

Guilt-by-association pathway annotation of individual lncRNAs was performed as follows. First, Pearson correlation coefficients between VST-transformed gene expression levels of each gene of interest and all other genes across the cancer cell line panel

were ranked, and the genes with Pearson correlation coefficients above the 99th percentile were used to determine enriched KEGG pathways using the gProfiler package with the parameters "max_p_value = 0.05, min_isect_size = 5" (Reimand et al., 2016). Network visualizations were produced in Cytoscape v3.4.0.

Disease-free survival analysis

151 RNAseq datasets preprocessed to read counts per gene via HTSeq from the TCGA LAML patient cohort were obtained through TCGAbiolinks in R (Colaprico et al., 2016). Of these available datasets, 123 patients were subjected to an Ara-C-containing treatment regimen, which were used for disease-free survival analysis. The resultant RNAseq count matrices were filtered of genes whose median read count was 0 and then subjected to the varianceStabilizingTransformation (VST) function from DESeq2 prior to downstream analysis. To determine relative associations of gene expression levels with disease-free survival, a Cox proportional hazards model was developed using sex, age over 60, white blood cell count over 16, and cytogenetic risk as covariates using the "coxph" function in the survival package in R (Therneau, 2015). The resultant coefficients per gene from the model were ranked and subjected to gene set enrichment analysis with the fgsea package in R using the KEGG pathway annotations. In building Kaplan-Meier curves, thresholds for the binary classification of gene expression were determined using the "surv_cutpoint" function, survival analysis was performed with the "survfit" function, and visualizations were obtained using the "ggsurvplot" function, all in the survminer package in R (Kassambara and Kosinski, 2017).

Human lncRNA CRISPRa library design

Transcripts from the Human GENCODE V22 (GRCh38, https://www.gencodegenes.org/releases/22.html) long non-coding RNA gene annotation set with the biotypes lincRNA, antisense, 3_prime_overlapping_ncRNA, processed_transcript, sense_intronic and sense_overlapping were retrieved and merged with the Broad human lincRNA catalog (Cabili et al., 2011). Annotated miRNA primary transcripts were then excluded from the list. Transcripts whose exons overlap extensively with RefSeq protein coding genes in the sense direction but are nevertheless categorized in one of the previously mentioned biotypes due to lack of predicted coding potential were also excluded. From the resulting list of 14,701 lncRNA loci, a set of 22,253 Transcription Start Sites (TSS) was obtained, which included transcripts belonging to the same gene. The nucleotides located 50-1000 bp upstream of each TSS were extracted from the UCSC Genome Browser, which has been determined to be the optimal region to design guides for gene activation (Gilbert et al., 2014). For each target location, two guides proximal to an NGG PAM sequence and separated by a minimum of 50 bp were selected on both the positive and negative strand, generating four possible guides as close as possible to each target TSS. Guide sequences containing degenerate nucleotides or inappropriate restriction sites were filtered from the final library, resulting in a library of 88,444 novel lncRNA targeting guides and 99 non-targeting guides validated in a previous study (Wang et al., 2014).

Human lncRNA CRISPRa library chip design and synthesis

For sgRNA library chip synthesis, each library member was designed to span 103 nucleotides in total, incorporating three separate 20 nt priming regions flanking the relevant sgRNA sequence. These regions included a universal 20 nt forward primer region, a 20 nt forward sub-pool primer region, a 43 nt region containing the variable sgRNA sequence, and a 20 nt universal reverse primer sequence. The library was partitioned into four sub-pools, which could be amplified individually through use of the sub-pool forward primers or altogether through use of the universal forward primer. The four sub-pools included were an antisense sub-library (33,965 elements), a lincRNA sub-library (42,836 elements), an alternate class sub-library (11,643 elements), and a non-targeting sub-library (99 elements). These human lncRNA CRISPRa sgRNA oligonucleotide libraries were synthesized on a standard Agilent array.

Lentivirus production

HEK293T cells were seeded 1 day before transfection. Cells were transfected the next day at 80%–90% confluency. For each dish, 15 μg of plasmid containing the vector of interest, 3.5 μg of pMD2.G, and 13 mg of psPAX2 (Addgene) were transfected using 90 μL Lipofectamine 2000 and 20 μL Plus Reagent (Life Technologies). Supernatant was collected 48 h and 72 h after transfection and filtered with a 0.45-mm PVDF filter (Millipore).

Pooled lncRNA CRISPRa screening for Ara-C response

MOLM14 cells were transduced with dCAS9-VP64 (Addgene #61425) and MS2-P65-HSF1 (Addgene #89308) simultaneously and cultured with high concentrations of Blasticidin (1 mg/mL) (Invivogen) and Hygromycin B (8 mg/mL) (Invivogen) for 14 days. The lentiviral CaLR library was generated in HEK293T as previously described. Supernatant containing viruses was collected 24-72 h after transfection, and viral titer was measured by qPCR Lentivirus Titration (Titer) Kit (abm). 72 h after transfection, stable MOLM14 SAM cells were infected with the CaLR library at a ratio of at least 500 cells/sgRNA. MOLM14 SAM CaLR cells were cultured in Puromycin (GIBCO) at a concentration of 1 ng/ml for 5 days. After 5 days, 500 cells/sgRNA were collected, and DNA was extracted using a standard phenol-chloroform protocol. The remaining cell pool, at a density of 500 cells/sgRNA, was subsequently cultured with Ara-C at a concentration of 0.25 μM (Sigma-Aldrich) for 14 days and then harvested for DNA extraction. Amplification of the specific sgRNAs was performed using NEBnext High Fidelity 2X Master Mix (New England Biolabs) in a single-step reaction as previously described (Joung et al., 2017b). In brief, a small amount of DNA from each sample was PCR amplified (20-36 cycles) using primers designed to

amplify the unique sgRNA from the genome (F: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TTA AGT AGA GGC TTT ATA TAT CTT GTG GAA AGG ACG AAA CAC C, R: CAA GCA GAA GAC GGC ATA CGA GAT TCG CCT TGG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT GCC AAG TTG ATA ACG GAC TAG CCTT). PCR products were run on an agarose gel to determine optimal PCR conditions. Next, the entire DNA of the sample was split to 48 PCR reactions, each amplified with a barcoded reverse primer and forward primers 1-10 (Table S7). PCR products were pooled and cleaned using QIAquick spin columns (QIAGEN). Products were tested for concentration and specificity using High sensitivity D1000 ScreenTape (Agilent) and qPCR. Libraries were pooled and sequenced using the Illumina NextSeq 500 platform (Harvard Medical School Biopolymers Facility, Boston).

Pooled protein-coding CRISPRa screening for Ara-C response

Protein-coding gene screening was performed using the human CRISPR activation library from Konermann et al. (2015) (Addgene #1000000057). Screening was performed as described for our CaLR screen.

CRISPRa screening deconvolution and analysis

The provided annotation library files were filtered sequentially using the following criteria: (1) duplicate entries of the same guide sequence, gene ID, and transcript ID, (2) guide RNA sequences annotated for more than one transcript, and (3) transcripts or genes represented by a single guide RNA. Raw .fastq sequencing files were trimmed with CutAdapt (v1.11) for the 5′ primer GCTTTATA TATCTTGTGGAAAGGACGAAACACC for the protein-coding gene library or GCTTTATATATCTTGTGGAAAGGACGAAACACCG for the noncoding gene library using the additional options "–m 20 –M 60" (Martin, 2011). Trimmed files were then matched to the guide sequences from the filtered library files using the MAGeCK count function (Tables S4, S5, and S6) (Li et al., 2014). Guide RNAs were further filtered such that they must be detected in at least half of the analyzed samples. Read counts were then subjected to MAGeCK mle analysis, modeling each experimental iteration as a separate batch, as suggested by principal component analysis, to obtain log fold changes and Wald test false discovery rates. Corresponding gene IDs were mapped to corresponding gene symbols via biomaRt (Durinck et al., 2005).

To determine an appropriate significance threshold for the CaLR screening results, the 99 nontargeting guides were combined to form a set of 21 simulated genes. The threshold for which 20/21 of these simulated genes would be considered not significant was determined to correspond to a false discovery rate of 3.51e-5, suggestive of an empirical false positive rate of 4.76% using this threshold.

An appropriate significance threshold for the protein-coding CRISPRa library was determined using a receiver operating characteristic (ROC) analysis as described above using a curated set of genes associated with differential Ara-C response (Table S2). Due to the absence of nontargeting guides in this library, an analysis to control the empirical false positive rate could not be performed in this analysis.

Transcript abundances in AML

151 RNaseq datasets preprocessed to normalized FPKM counts via HTSeq from the TCGA LAML patient cohort were obtained through TCGAbiolinks in R (Colaprico et al., 2016). FPKM values were $\log_2$-transformed after the addition of a unit pseudocount. Genes were considered detectable if their median $\log_2$ (FPKM + 1) values were greater than 0 and their 90^{th} percentile $\log_2$ (FPKM + 1) values were greater than 0.1, similar to previously used criteria (Hu et al., 2014).

Global pathway annotation of enriched lncRNAs

Predicted functional annotation of significantly enriched lncRNAs was performed using the funcpred tool (https://www.funcpred.com) (Perron et al., 2017). Briefly, this tool uses functional enrichment of the neighborhood of a lncRNA derived from co-expression networks based on GTEx expression profiling data to infer its putative function. We used a co-expression network based on all GTEx tissues and examined the functional enrichment of significant lncRNAs as defined by the Hallmark Gene Sets contained in the Molecular Signatures Database from the Broad Institute (http://software.broadinstitute.org/gsea/msigdb/). Interactions between functional annotations deemed significant using a 5% false discovery rate threshold were collated by annotations within each gene, and highly enriched interactions were demarcated as such when associations appeared at least 15 times. Networks of these highly enriched interactions were generated in Cytoscape using the "edge-weighted spring-embedded layout" function with minor adjustments for improved visualization (Smoot et al., 2011).

K-means clustering of RNaseq from AML patient samples

Raw .fastq files from Garzon et al. (2014) were downloaded from the SRA archive (SRP050272) and then subjected to preprocessing as described for the human cancer cell line analysis above. Transcriptome profiles were filtered out based on low mappability to either the hg38 genome (< 80%) or the Ensembl-annotated transcriptome (< 20%). Pearson correlation coefficients between gene pairs of interest were determined using VST-transformed gene counts. Prior to k-means clustering analysis, VST-transformed gene counts were normalized by gene using a Z-score scaling. The appropriate number of clusters was determined using the gap

statistic, and the k-means algorithm was performed with 17 clusters using Lloyd's algorithm. Functional annotation enrichment of the gene cluster containing GAS6-AS2 was performed using the ToppCluster web server (Kaimal et al., 2010). The server uses the ToppGene tool to perform enrichment analyses of gene ontology (GO) and pathways, as previously described (Chen et al., 2007).

Pan-cancer co-expression analysis

Data used for co-expression analyses between GAS6-AS2, GAS6, and AXL in different cancer types were obtained from the TANRIC database (Li et al., 2015) and the Firehose Broad GDAC portal.

RNA-IP analysis

Raw .fastq files corresponding to DNMT1 IP- and IgG IP-RNaseq datasets were obtained from the SRA archive (SRR358674, SRR358675) (Di Ruscio et al., 2013). Resultant .fastq sequencing files were then aligned to the hg38 reference genome with default STAR single-pass alignment (v2.5.2b) (Dobin et al., 2013). Realigned .bam files were sorted by name using SAMtools v1.3 (Li et al., 2009). Coordinates were mapped to Ensembl gene IDs with featureCounts (Subread v1.5.2) as denoted by Ensembl v89 annotation using features "-p -B -s 0" (Aken et al., 2016; Liao et al., 2014). Genes were then normalized by fragments per kilobase of transcript per million mapped reads (FPKM), with gene length determined by featureCounts, and then FPKM values were $\log_2$-transformed after the addition of a 1e-5 pseudocount. Log-transformed counts from DNMT1-IP were then subtracted from the log-transformed counts from IgG-IP, which segregated into three clusters representing transcripts associated predominantly with DNMT1, transcripts associated predominantly with IgG, and transcripts associated with both DNMT1 and IgG.

QUANTITATIVE AND STATISTICAL ANALYSIS

Statistical analysis was performed using R Statistical Software (https://www.r-project.org/) and Excel (Microsoft). All reported *P* values are two-tailed, and for all analyses, $p \leq 0.05$ is considered statistically significant, unless otherwise specified. Specific details concerning statistical tests for individual experiments are noted in the relevant Figure legends.

Replicates for individual experiment types are outlined here:

CRISPRa screening targeting protein coding genes, n = 2.
CRISPRa screening targeting lncRNA genes, n = 2.
All qPCR of cells expressing sgRNA, n > 3, groups were compared using Welch two sample t test.
All cell viability by MTS, n > 3, groups were compared using Welch two sample t test.
All apoptosis assay by ANNEXIN V, n = 3, groups were compared using Welch two sample t test.
DNA damage by γH2AX, n = 50, from 2 staining, groups were compared using Welch two sample t test.
DNA methylation by sodium bisulfite assay, n > 10, groups were compared using Chi-square test.
Cell cycle analysis by BrdU/PI, n = 3, groups were compared using Welch two sample t test.
Western blot, n = 3, groups were compared using Welch two sample t test.
Competition experiment in mouse model, n = 20, groups were compared using Welch two sample t test.
For detailed statistical analysis of the computational analysis see Method Details.

DATA AND SOFTWARE AVAILABILITY

Custom scripts were used for data preprocessing, statistical analysis, and data visualization as described previously. Code used in this study is available upon request by the Lead Contact Dr. Pier Paolo Pandolfi (ppandolf@bidmc.harvard.edu).

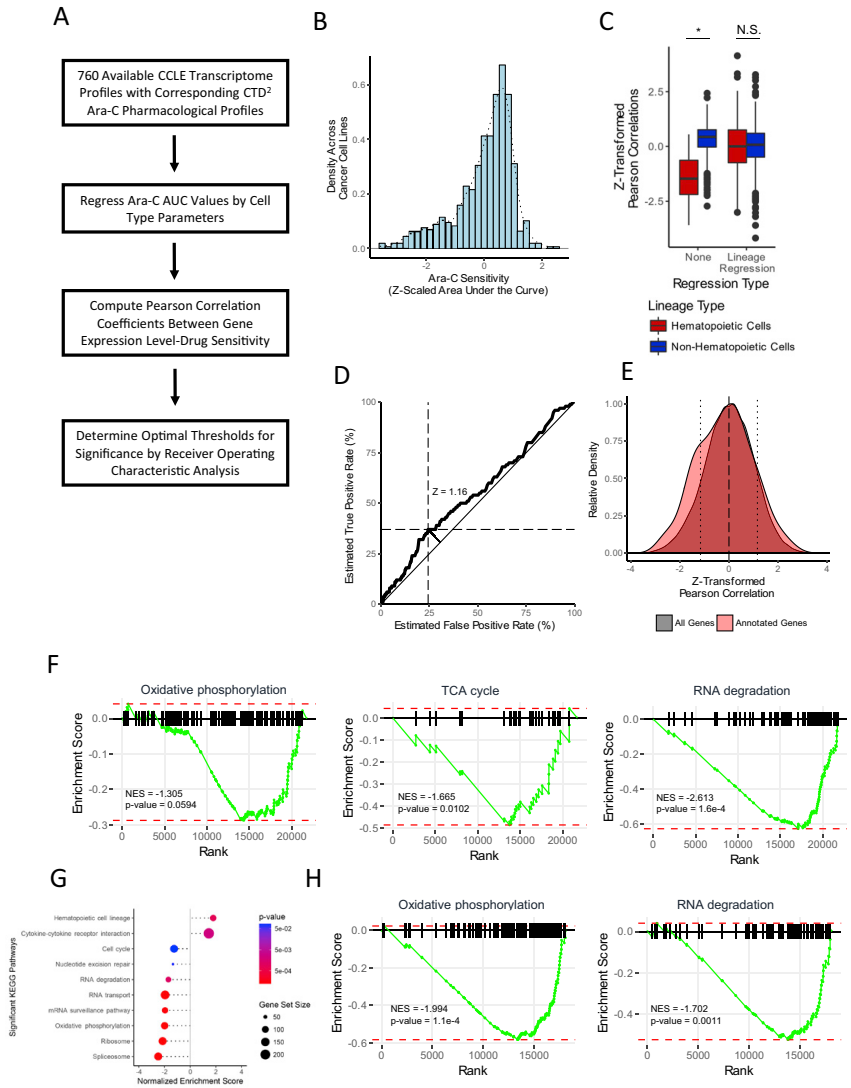


Figure S1. Identification of Protein-Coding and Non-coding Gene Biomarkers Correlated with Differential Ara-C Response, Related to Figure 1

(A) Pipeline for identification of protein-coding and noncoding gene biomarkers of differential Ara-C response across 760 cancer cell lines. See also Figures 1A and 1B.

(B) Distribution of Ara-C drug sensitivities across 760 pan-cancer cell lines profiled by both CCLE and CTD^2 studies, quantified by their Z-scaled area under the dose response curve (AUC) values without removing lineage-specific effects.

(C) Representative effect of regressing lineage-specific annotations from Ara-C AUC values between hematopoietic and non-hematopoietic cells. Wilcoxon rank-sum test: *p < 2.2e-16. N.S, p = 0.7208.

(D) Receiver operating characteristic (ROC) analysis of drug sensitivity-gene expression correlations to determine the optimal Z-score threshold (Z = 1.16). Genes curated from the literature as responsive to Ara-C (Table S2) passing this Z-score threshold were considered true positives, whereas other genes not within this curated list passing this Z-score threshold were considered false positives.

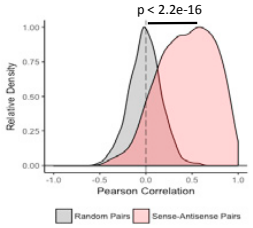
(E) Distributions of drug sensitivity-gene expression correlations for annotated genes (Table S2) versus all genes in the analysis.

(F) Representative KEGG pathways from GSEA of protein-coding genes ranked by drug sensitivity-gene expression correlation values as shown in Figures 1B and 1C. See Table S3.

(G) Summary of GSEA of protein-coding genes ranked by disease-free survival association strength using annotated KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Clinical and transcriptomic data from the TCGA-LAML patient cohort was used for this analysis. Disease-free survival association was quantified by the magnitude of the coefficient from a Cox proportional hazards model for each gene, with patient sex, age over 60, cytogenetic risk, and white blood cell count above 16 as covariates.

(H) Representative KEGG pathways from GSEA of protein-coding genes ranked by disease-free survival association strength as shown in Figures 1B and 1C. Clinical and transcriptomic data from the TCGA-LAML patient cohort was used for this analysis.

A



B

Gene Pair Type	# Pairs	# Gene Pairs, Both Genes Correlated with Ara-C Sensitivity	# Genes Pairs, Sense Gene Correlated with Cognate Antisense Gene	Overlap
Sense-Antisense Pairs	997	113	786	101
Random Pairs	5,000	291	50	10

C

Gene Pairs Correlated with Ara-C Resistance

ACOXL	ACOXL-AS1		INHBA	INHBA-AS1		P4HA3	P4HA3-AS1
B4GALT1	B4GALT1-AS1		IQCJ-SCHIP1	IQCJ-SCHIP1-AS1		PIK3IP1	PIK3IP1-AS1
BDNF	BDNF-AS		ISPD	ISPD-AS1		PINK1	PINK1-AS
CCDC183	CCDC183-AS1		ITGB5	ITGB5-AS1		PRICKLE2	PRICKLE2-AS2
CDKN2A	CDKN2A-AS1		KIRREL3	KIRREL3-AS2		PRICKLE2	PRICKLE2-AS3
COL5A1	COL5A1-AS1		KRT7	KRT7-AS		PRKG1	PRKG1-AS1
CPEB2	CPEB2-AS1		LACTB2	LACTB2-AS1		RBPMS	RBPMS-AS1
CXXC5	CXXC5-AS1		LIFR	LIFR-AS1		RERG	RERG-AS1
DOCK9	DOCK9-AS2		LOXL1	LOXL1-AS1		RMDN2	RMDN2-AS1
EGFR	EGFR-AS1		LRP1	LRP1-AS		ROR1	ROR1-AS1
EXOC3	EXOC3-AS1		MAMDC2	MAMDC2-AS1		SEMA5A	SEMA5A-AS1
FAM83A	FAM83A-AS1		MKX	MKX-AS1		SLC2A1	SLC2A1-AS1
FLNC	FLNC-AS1		MRVI1	MRVI1-AS1		SPON1	SPON1-AS1
FRMD6	FRMD6-AS1		NALCN	NALCN-AS1		TGFB2	TGFB2-AS1
FRMD6	FRMD6-AS2		NDST1	NDST1-AS1		THSD4	THSD4-AS1
FTCD	FTCD-AS1		NEURL1	NEURL1-AS1		TM4SF1	TM4SF1-AS1
FUT8	FUT8-AS1		NEXN	NEXN-AS1		TM4SF19	TM4SF19-AS1
GAS6	GAS6-AS2		NTRK3	NTRK3-AS1		VLDLR	VLDLR-AS1
GRM5	GRM5-AS1		OSBPL10	OSBPL10-AS1		ZNF667	ZNF667-AS1
GSN	GSN-AS1		OSMR	OSMR-AS1		ZNF793	ZNF793-AS1
IGFBP7	IGFBP7-AS1		P3H2	P3H2-AS1			

Gene Pairs Correlated with Ara-C Sensitivity

ADARB2	ADARB2-AS1		MYB	MYB-AS1
AP4B1	AP4B1-AS1		MYCBP2	MYCBP2-AS1
CCND2	CCND2-AS2		NOP53	NOP53-AS1
CCNT2	CCNT2-AS1		PABPC5	PABPC5-AS1
DCUN1D2	DCUN1D2-AS		PAXBP1	PAXBP1-AS1
DLGAP1	DLGAP1-AS4		PCCA	PCCA-AS1
DLGAP1	DLGAP1-AS5		PLCH1	PLCH1-AS1
ELFN1	ELFN1-AS1		PRC1	PRC1-AS1
ENTPD1	ENTPD1-AS1		RNF219	RNF219-AS1
FAM53B	FAM53B-AS1		SLC25A30	SLC25A30-AS1
FBXW7	FBXW7-AS1		SLCO4A1	SLCO4A1-AS1
FOXD2	FOXD2-AS1		SMAD1	SMAD1-AS2
GABPB1	GABPB1-AS1		SMC5	SMC5-AS1
GTF3C2	GTF3C2-AS1		SSBP3	SSBP3-AS1
HHATL	HHATL-AS1		THAP9	THAP9-AS1
INTS6	INTS6-AS1		TMPO	TMPO-AS1
INTS9	INTS9-AS1		USP3	USP3-AS1
KCNIP2	KCNIP2-AS1		ZBED3	ZBED3-AS1
LANCL1	LANCL1-AS1			
MAPKAPK5	MAPKAPK5-AS1			
MRGPRG	MRGPRG-AS1			

D

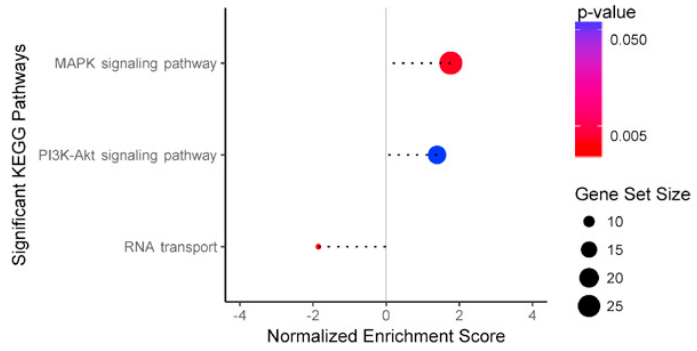
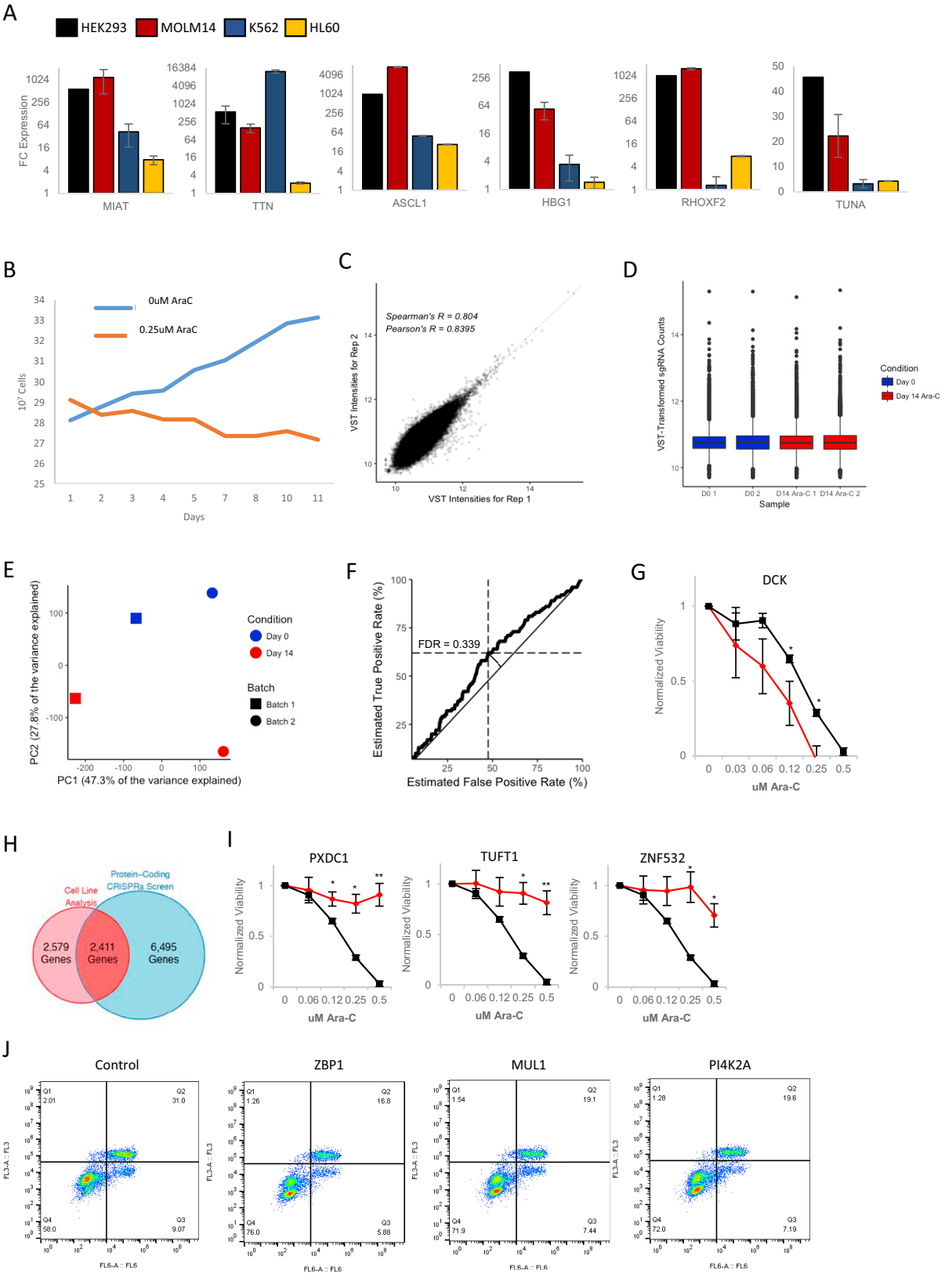


Figure S2. Identification of Sense-Antisense Gene Pairs Highly Correlative with Differential Ara-C Response, Related to Figure 1

(A) Pearson correlation distributions of gene pair expression levels in the TCGA-LAML patient samples across 997 sense-antisense cognate gene pairs and 5,000 random protein coding-lncRNA gene pairs. Wilcoxon rank-sum test: $p < 2.2e-16$.

(legend continued on next page)

(B) Summary of gene pair analysis to select pairs with correlated gene expression levels and high drug sensitivity-gene expression correlations. The enrichment of sense-antisense gene pairs is much larger than the enrichment of random coding-non-coding gene pairs. Chi-square test: $p < 2.2e-16$.
(C) List of protein-coding/cognate antisense gene pairs identified to be highly correlated in expression levels with each other and have significant drug sensitivity-gene expression correlations.
(D) Representative KEGG pathways from GSEA of cognate genes among sense-antisense gene pairs ranked by drug sensitivity-gene expression correlation values.



(legend on next page)

Figure S3. CRISPRa Functional Screening of Coding Genes Modulating Ara-C Response, Related to Figure 2

(A) Fold change (FC) of expression levels modulated by CRISPRa for a cohort of single sgRNAs representing various protein-coding and lncRNA genes modulated in HEK293T (black), MOLM14 (red), K562 (blue), or HL60 (yellow) cells. Data are represented as mean ± SD, n = 3.

(B) Cell titers of Ara-C treated or untreated MOLM14 over the time course of the CRISPRa screening.

(C) Correlation of the sgRNA levels represented in the sequencing libraries across experimental batches. Mapped sequencing read abundances were transformed using the variance-stabilizing transformation (VST) in DESeq2.

(D) Distributions of sgRNA levels across the four protein-coding CRISPRa sequencing libraries. Mapped sequencing read abundances were transformed using the variance-stabilizing transformation (VST) in DESeq2.

(E) Principal component analysis of sgRNA levels across the four generated sequencing libraries. PC, principal component.

(F) Receiver operating characteristic (ROC) analysis of protein-coding CRISPRa screening results to determine optimal FDR threshold (FDR = 0.339). Genes curated from the literature as responsive to Ara-C (Table S2) passing this Z-score threshold were considered true positives, whereas other genes not within this curated list passing this Z-score threshold were considered false positives.

(G) Modulation of Ara-C response upon stable expression of an sgRNA targeting DCK in MOLM14 cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.

(H) Overlap of protein-coding genes enriched in both the drug sensitivity-gene expression correlative analysis and the CRISPRa screen.

(I) Modulation of Ara-C response upon stable expression of sgRNAs targeting PXDC1, TUFT1, or ZNF532 in MOLM14 cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$

(J) Representative distributions of annexin V and propidium iodide co-staining intensities of MOLM14 cells expressing the indicated sgRNAs as determined by flow cytometry. See also Figure 2H.

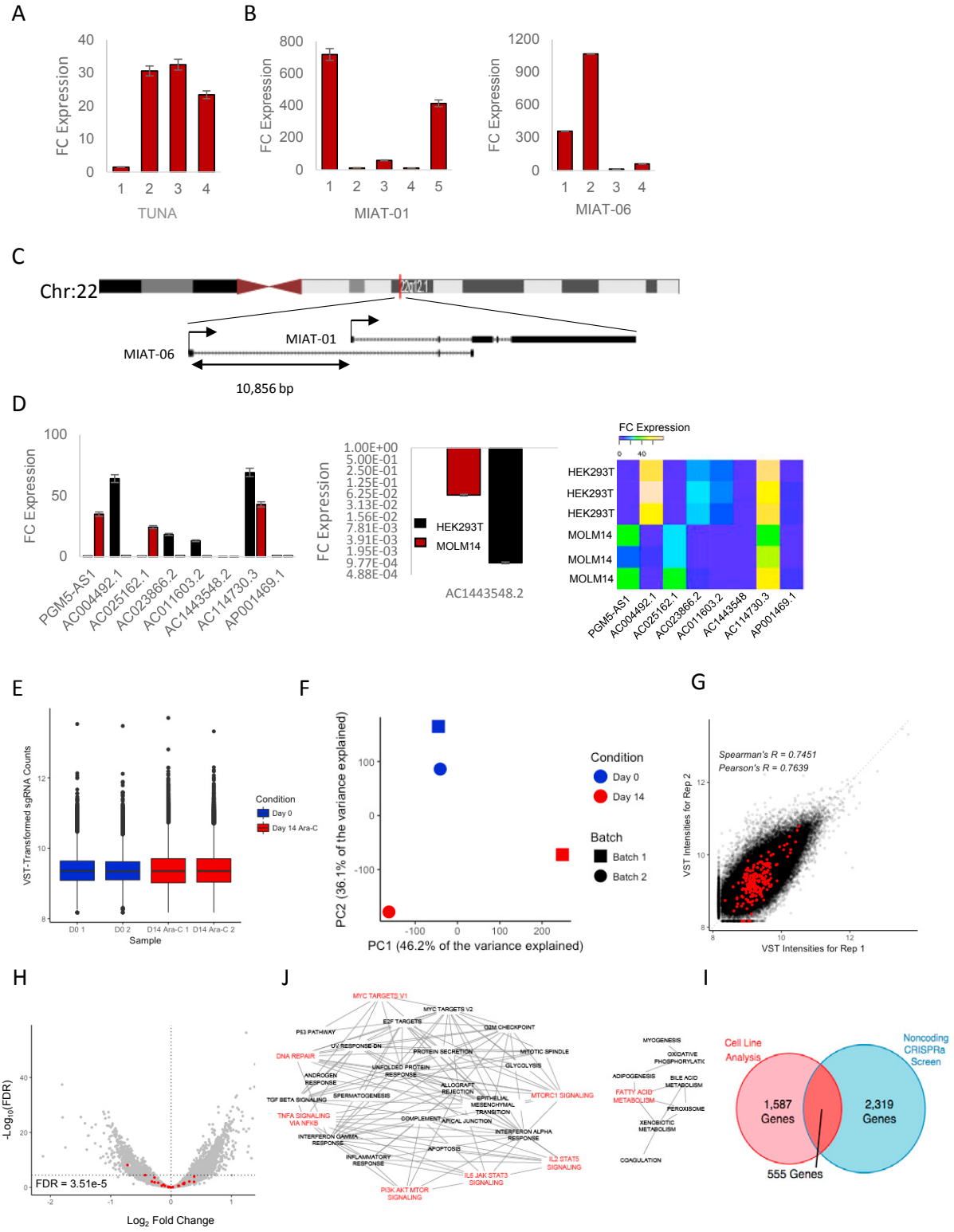


Figure S4. CRISPRa Functional Screening of Non-coding Genes Modulating Ara-C Response, Related to Figure 3

(A) Fold change (FC) of expression levels modulated by CRISPRa for sgRNAs predicted to target the *TUNA* lncRNA in MOLM14 cells. Data are represented as mean ± SD, n = 3.

(legend continued on next page)

(B) Fold change (FC) of expression levels by CRISPRa for sgRNAs predicted to target a previously annotated TSS (*MIAT-01*) of and an alternative, predicted TSS (*MIAT-06*) of the *MIAT* lncRNA gene in MOLM14 cells. Data are represented as mean ± SD, n = 3. See (C).
(C) Chromosomal localizations and predicted transcriptional start sites of MIAT-01 and MIAT-06 transcript isoforms.
(D) Left panels: fold change (FC) of expression levels modulated by CRISPRa with predicted sgRNAs for a cohort of lncRNA genes in either HEK293T (black bars) or MOLM14 (red bars) cells. Data are represented as mean ± SD, n = 3. Right panel: a heatmap representation of these data.
(E) Distributions of sgRNA levels across the four CaLR sequencing libraries. Mapped sequencing read abundances were transformed using the variance-stabilizing transformation (VST) in DESeq2.
(F) Principal component analysis of sgRNA levels across the four generated CaLR sequencing libraries. PC, principal component.
(G) Correlation of the sgRNA levels represented in the CaLR sequencing libraries across experimental batches. Mapped sequencing read abundances were transformed using the variance-stabilizing transformation (VST) in DESeq2.
(H) Analysis of nontargeting guides in the CaLR library to determine the optimal FDR threshold (FDR = 3.51e-5), corresponding to an empirical false positive rate of 4.76%. Genes above this FDR threshold were considered significant.
(I) Specific pathways identified in a guilt-by-association co-expression analysis for lncRNAs associated with enriched sgRNAs. MSigDB pathway gene sets were used for this analysis.
(J) Overlap of lncRNA genes enriched in both the drug sensitivity-gene expression correlative analysis and the CaLR screen.

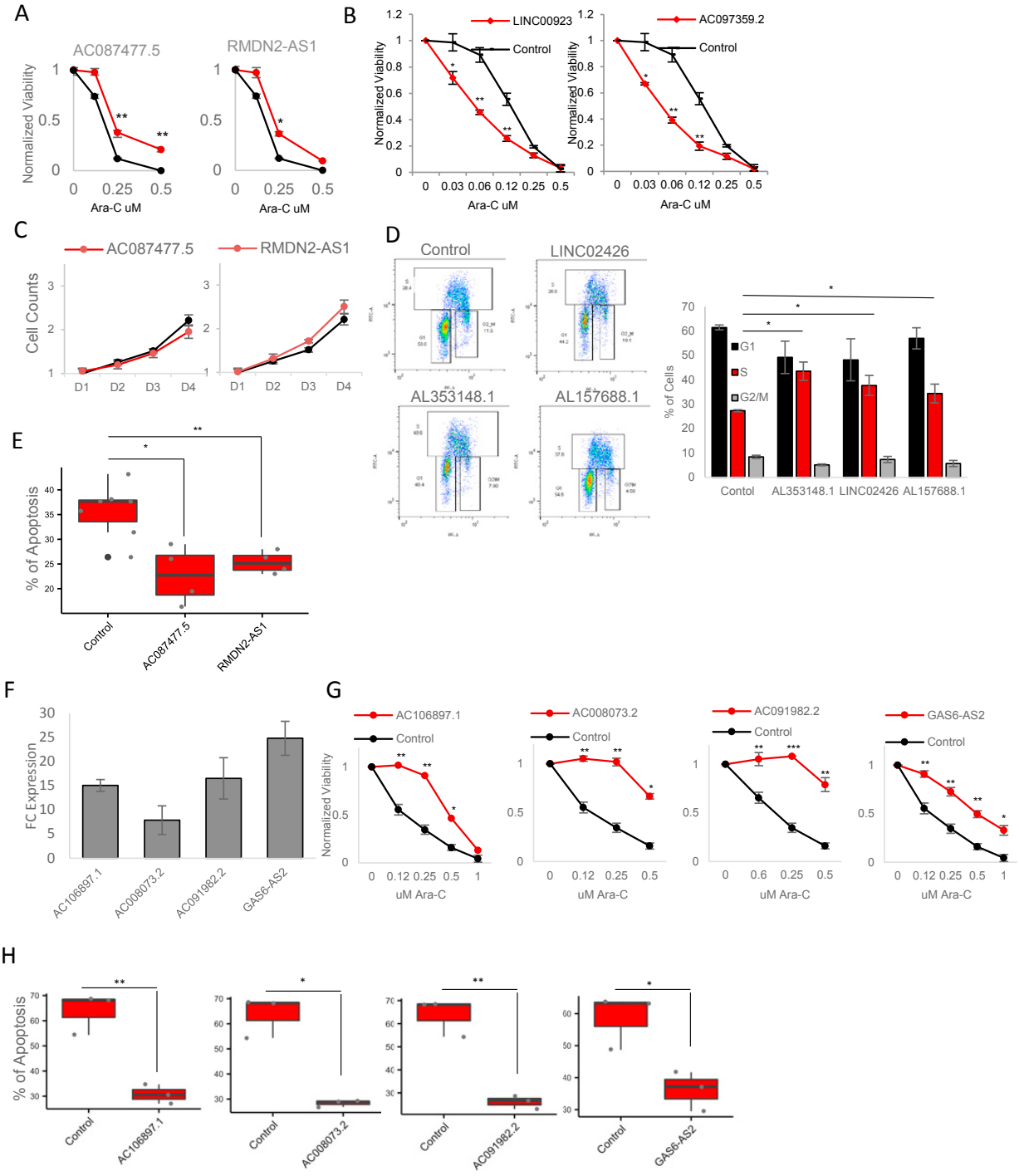


Figure S5. Validation of Enriched sgRNAs Targeting lncRNAs in the CaLR Screening, Related to Figure 4

(A) Modulation of Ara-C response upon stable expression of sgRNAs targeting CaLR-enriched lncRNAs in MOLM14 cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(B) Modulation of Ara-C response upon stable expression of sgRNAs targeting CaLR-depleted lncRNAs in MOLM14 cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(C) Modulation of cellular proliferation in the absence of Ara-C treatment upon stable expression of sgRNAs targeting AC087477.5 or RMDN2-AS1 in MOLM14 cells. Proliferation is quantified over four days (D1-D4). Data are represented as mean ± SD, n = 3.

(legend continued on next page)

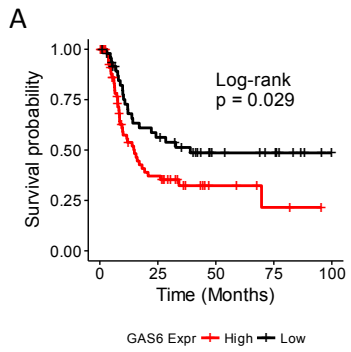
(D) Cell cycle analysis of MOLM14 cells with stable expression of sgRNAs targeting LINC02426, AL353148.1, AL157688.1, or control as determined by bromodeoxyuridine (BrdU) incorporation. Left panel: representative flow cytometry scatterplots. Right panel: quantification of cell cycle phases for each cell line. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(E) Modulation of apoptotic response upon stable expression of sgRNAs targeting AC087477.5, RMDN2-AS1, or control in MOLM14 cells. The percentage of apoptosis is determined by annexin V and propidium iodide (PI) staining of MOLM14 cells stably infected with individual sgRNAs and treated with 0.25 μM Ara-C for 24 hours. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

(F) Fold change (FC) of expression levels of targeted lncRNAs upon overexpression of enriched sgRNAs versus endogenous levels in HL60 cells. Data are represented as mean ± SD, n = 3.

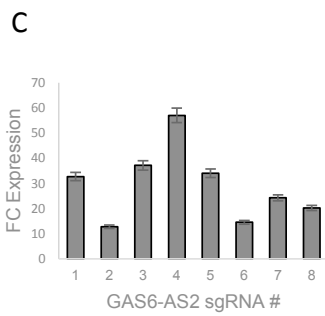
(G) Modulation of Ara-C response upon stable expression of sgRNAs targeting the indicated CaLR-enriched lncRNAs in HL60 cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

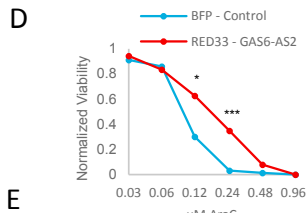
(H) Modulation of apoptotic response upon stable expression of sgRNAs targeting AC106897.1, A008073.2, AC091982.2, GAS6-AS2, or control in HL60 cells. The percentage of apoptosis is determined by annexin V and propidium iodide (PI) staining of HL60 cells stably infected with individual sgRNAs and treated with 0.25 μM Ara-C for 24 hours. Data are represented as mean ± SD, n > 3. Welch two sample t test: *p < 0.05. **p < 0.01, ***p < 0.001.

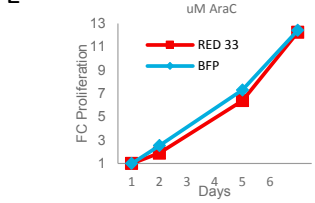


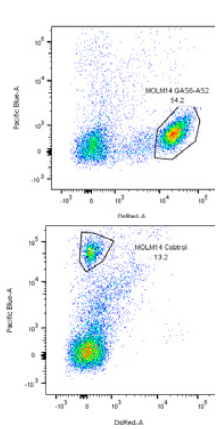
B

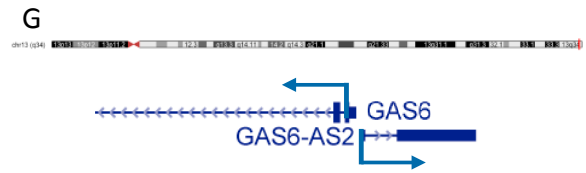
GAS6-AS2 sgRNAs	Target Sequence
#1	GCCGCCGCGATGTGACCTTC
#2	CGCCGCCGCCGCCGCCGCCG
#3	GGCGGCGGCGGCGGCGGCGG
#4	GCACCTCAAGCGCTCGGTCT
#5	ACGCCCAGACCGAGCGCTTG
#6	GCGCGGGGAGGCGGCGGGGG
#7	GCCCCGTCCGGGAGGGAGGT
#8	TGACCCCCCCACCTCCCTCC

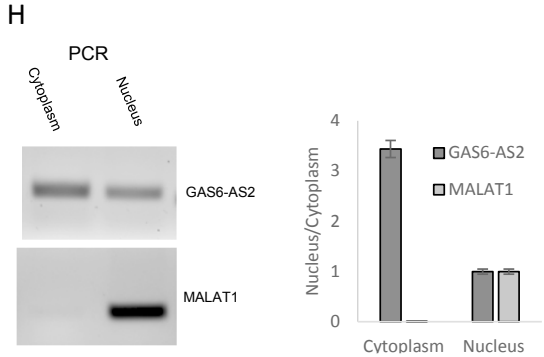


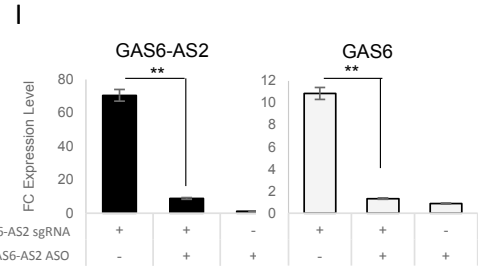












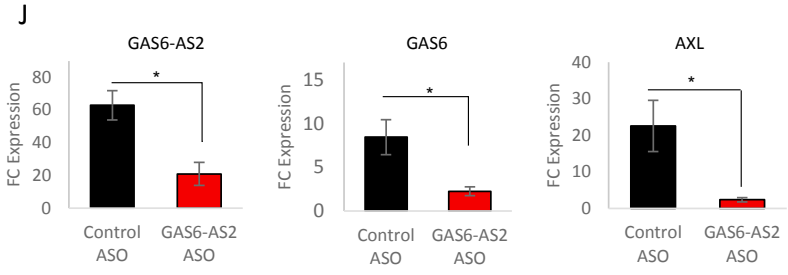


Figure S6. GAS6-AS2 Activates GAS6/TAM Signaling, Related to Figure 5

(A) Disease-free survival association with expression levels of GAS6, enriched in both protein-coding CRISPRa screening and drug sensitivity-gene expression correlation analyses, among patients treated with Ara-C therapy within the TCGA-LAML patient cohort. VST expression level cutoff = 9.20 (low, n = 49; high, n = 72), log-rank test: p value = 0.029.

(legend continued on next page)

(B) List of sgRNA sequences used to target the GAS6-AS2 TSS for CRISPRa. See Figures 5B and 5C.
(C) Fold change (FC) of GAS6-AS2 expression levels in MOLM14 cells for individual sgRNAs targeting *GAS6-AS2*. Data are represented as mean ± SD, n = 3.
(D) Ara-C dose-response profiles of MOLM14 cells with GAS6-AS2 overexpression (RED) or control (BLUE).
(E) Cellular proliferation curves of MOLM14 cells with GAS6-AS2 overexpression (RED) or control (BLUE).
(F) Representative flow cytometry scatterplots demonstrating tumor burden in xenograft murine models of AML following Ara-C treatment (upper panel – GAS6-AS2-RED, lower panel Control-BFP). See Figures 5H and 5I.
(G) Chromosomal localization and organization of the GAS6 and GAS6-AS2 genomic locus.
(H) Left panel: cellular localizations of GAS-AS2 and MALAT1 lncRNAs, as determined by nuclear fractionation followed by PCR. Right panel: nuclear/cytoplasmic ratio of GAS6-AS2 and MALAT1, based on qPCR. Data are represented as mean ± SD, n > 3.
(I) Fold change (FC) of GAS6-AS2 (left panel) and GAS6 (right panel) expression levels modulated by GAS6-AS2 overexpression via CRISPRa (sgRNA #4) and subsequent GAS6-AS2 knockdown via ASO in HEK293T cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.
(J) Fold change (FC) of GAS6-AS2, GAS6, and AXL expression levels modulated by GAS6-AS2 overexpression via CRISPRa (sgRNA #4) followed by either GAS6-AS2 or control ASO in HEK293T cells. Data are represented as mean ± SD, n > 3. Welch two sample t test: *$p < 0.05$. **$p < 0.01$, ***$p < 0.001$.

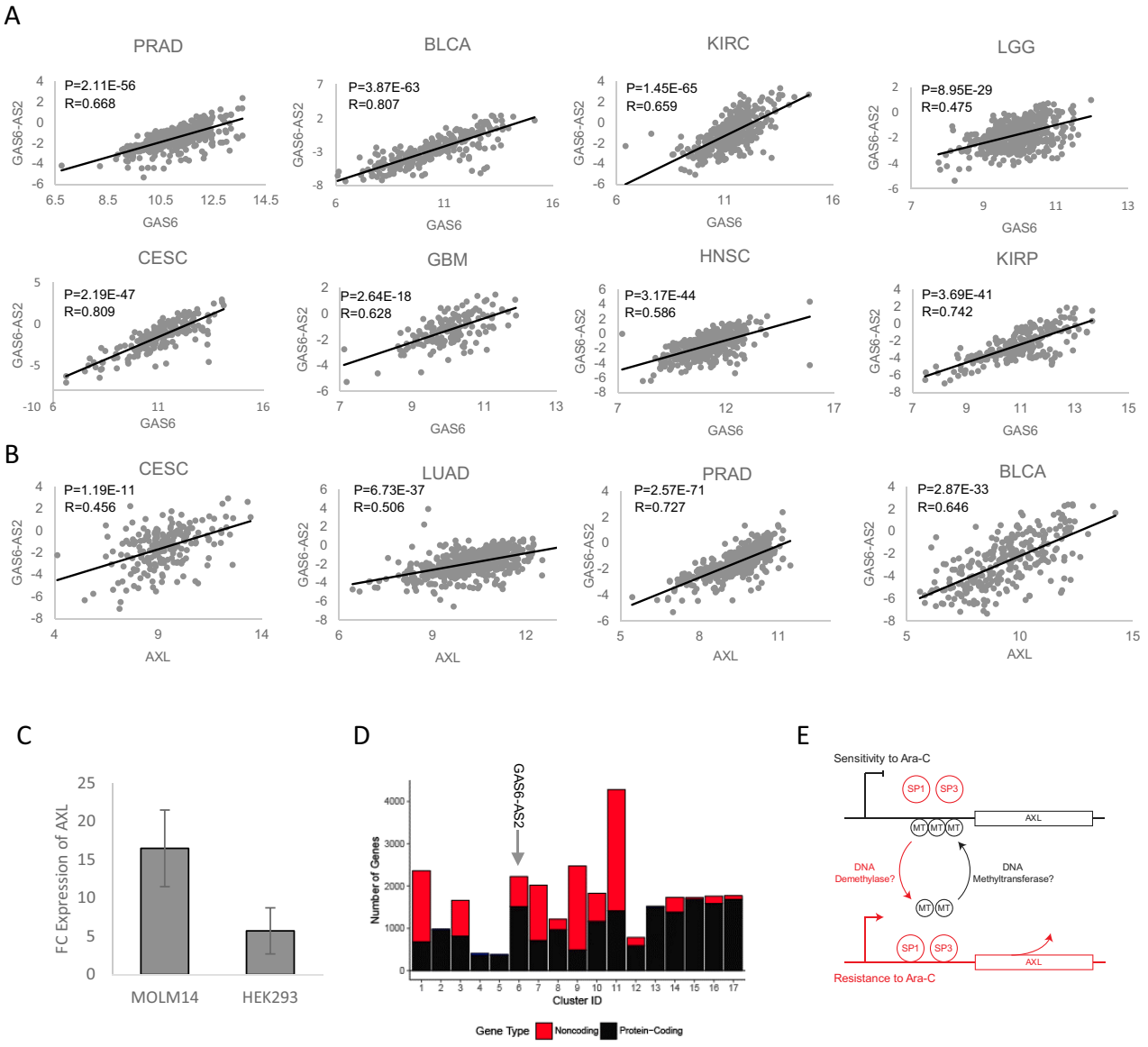


Figure S7. GAS6-AS2 Regulates AXL in *trans*, Related to Figures 6 and 7

(A) Relationship of GAS6-AS2 and GAS6 expression levels across a panel of patient tumor samples representative of different cancer subtypes from TCGA. PRAD – Prostate Adenocarcinoma; BLCA – Urothelial Bladder Carcinoma; KIRC – Kidney Renal Clear Cell Carcinoma; LGG – Lower Grade Glioma; CESC – Cervical Squamous Cell Carcinoma; GBM – Glioblastoma Multiforme; HNSC – Head-Neck Squamous Cell Carcinoma; KIRP – Cervical Kidney Renal Papillary Cell Carcinoma.

(B) Relationship of GAS6-AS2 and AXL expression levels across a panel of patient tumor samples representative of different cancer subtypes from TCGA. CESC – Cervical Squamous Cell Carcinoma; LUAD – Lung Adenocarcinoma; PRAD – Prostate Adenocarcinoma; BLCA – Urothelial Bladder Carcinoma.

(C) Fold change (FC) of AXL gene expression levels upon GAS6-AS2 CRISPRa targeting in MOLM14 and HEK293 cell lines. Data are represented as mean ± SD, n > 3. See also Figure 6E.

(D) K-means clustering of coding genes (black) and lncRNAs (red) based on gene expression levels in AML patients (Garzon et al., 2014). GAS6-AS2 is contained in cluster #6. See also Figure 7D.

(E) Schematic of putative modulation of CpG methylation and concomitant regulation of AXL gene expression.