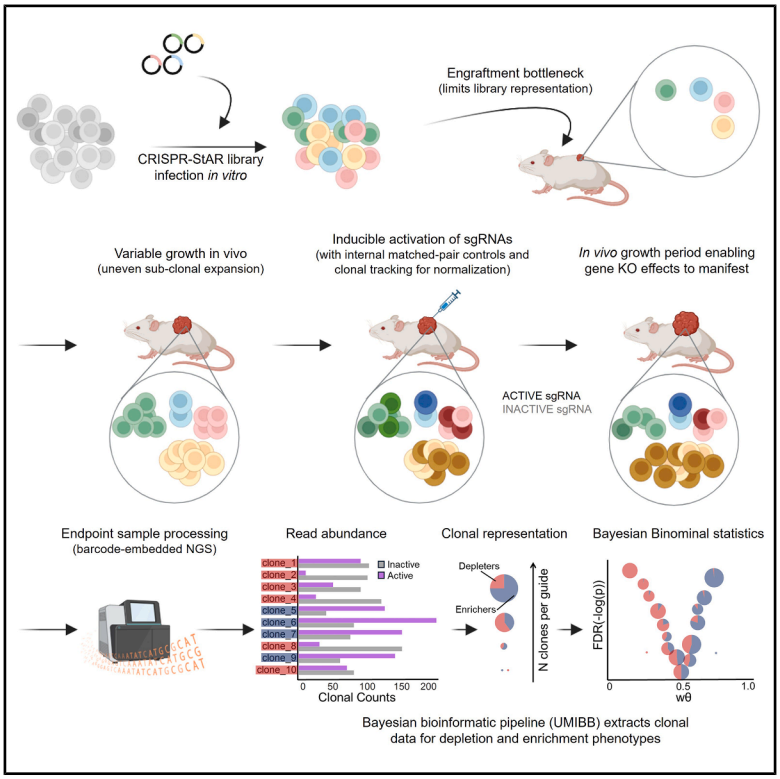
Cell Reports
Methods

Article

Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens

Graphical abstract



Highlights

- CRISPR-StAR enables inducible, matched-pair intra-tumor controls *in vivo*
- Bayesian UMIBB pipeline enables clonal-resolution analysis for gene phenotypes
- ~30,000-sgRNA libraries can be functionalized using only 30 tumors
- Up to 7-fold reduction in animal usage compared with conventional *in vivo* screens

Authors

Silvia Fenoglio, Yi Yu, James Tepper, ..., Xuewen Pan, Jannik N. Andersen, Teng Teng

Correspondence

tteng@tangotx.com

In brief

Fenoglio et al. demonstrate the application of CRISPR-StAR, an inducible *in vivo* screening platform with internal intra-tumor normalization, and developed a Bayesian bioinformatics pipeline to extract clonal depletion and enrichment signals. This integrated approach enables scalable, high-fidelity *in vivo* screens with up to 7-fold fewer animals.

Fenoglio et al., 2026, Cell Reports Methods 6, 101470
July 20, 2026
https://doi.org/10.1016/j.crmeth.2026.101470

CellPress

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens

Silvia Fenoglio,[1] Yi Yu,[1] James Tepper,[1] Lauren Grove,[1] Alborz Bejnood,[1] Samuel R. Meier,[1] Ashley H. Choi,[1] Hsin-Jung Wu,[1] Annabel Devault,[1] Shangtao Liu,[1] Binzhang Shen,[1] Tenzing Khendu,[1] Hannah Stowe,[1] Esther C.H. Uijttewaal,[2] Minjie Zhang,[1] Brian B. Haines,[1] Erik Wilker,[1] Alan Huang,[1] Daniel Schramek,[3,4] Ulrich Elling,[2] Xuewen Pan,[1] Jannik N. Andersen,[1] and Teng Teng[1,5,*]

[1]Tango Therapeutics Inc, Boston, MA, USA
[2]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr. Bohr-Gasse 3, Vienna BioCenter (VBC), Vienna, Austria
[3]Centre for Molecular and Systems Biology, Lunenfeld–Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada
[4]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[5]Lead contact
*Correspondence: tteng@tangotx.com
https://doi.org/10.1016/j.crmeth.2026.101470

MOTIVATION Conventional *in vivo* CRISPR dropout screens suffer from poor library representation due to engraftment bottlenecks and uneven clonal expansion, requiring large animal cohorts and masking microenvironment-dependent vulnerabilities. We demonstrate the implementation of inducible CRISPR-StAR with matched active/inactive controls and Bayesian analysis to normalize guides locally, improve reproducibility, and scale *in vivo* screening with fewer animals.

SUMMARY

CRISPR-StAR (stochastic activation by recombination) is an inducible pooled screening system that activates gene knockout after tumor engraftment and provides matched internal controls for guide-level normalization. In this study, we explore the scalability and reproducibility of this approach for *in vivo* cancer screens. Through barcode-embedded sequencing and the development of a Bayesian analysis pipeline, we screened a 30,000-sgRNA library in A549 xenografts, achieving reproducible dropout and enrichment phenotypes using just ~30 tumors. Across additional xenograft models, single tumors yielded reliable functional annotation for ~1,000 genes. Comparing *in vivo* and *in vitro* screens uncovered tumor suppressor effects detectable only *in vivo*; for example, KMT2C and KMT2D knockouts produced contrasting growth and transcriptional programs. Together with our R analysis package, we show that CRISPR-StAR enables scalable *in vivo* dependency mapping that complements *in vitro* resources and reduces animal use by up to 7-fold versus conventional dropout screens, improving methodological rigor at genome-scale clonal resolution.

INTRODUCTION

Pooled *in vitro* CRISPR-Cas9 screening approaches, exemplified by large-scale efforts such as DepMap,[1] have transformed our understanding of genetic dependencies in human cancer. These methodologies have been instrumental in expanding key concepts such as synthetic lethality, lineage dependency, and oncogene addiction, providing a rich catalog of novel targets for therapeutic intervention.[1,2] For example, early functional genomics studies revealed that MTAP-deleted cancer cells are uniquely sensitive to PRMT5 inhibition,[3–5] motivating the development of methylthioadenosine (MTA)-cooperative PRMT5 inhibitors such as AMG193,[6] MTRX1719,[7] and vopimetostat (TNG462).[8] Similar loss-of-function screens identified synthetic lethality between *STAG2* and *STAG1*,[9,10] as well as between *PTEN* loss and inhibition of mTORC2 components,[11] underscoring the power of functional genomics to uncover context-specific cancer vulnerabilities.

Building on this foundational work in RNAi- and CRISPR-based genetic screens,[1,12] we aimed to address the limitations of *in vitro* screens by enabling robust knockout (KO) screens in the noisier and more complex context of *in vivo* tumor biology.[13,14] While *in vitro* systems have been invaluable, they often fail to recapitulate critical interactions with the tumor microenvironment, immune system, and nutrient gradients, factors that profoundly influence tumor progression and therapeutic response.[13,15–17] By studying these dynamics *in vivo*, we hope to uncover novel context-dependent vulnerabilities that are undetectable in conventional tissue culture systems.

Historically, *in vivo* genetic screens have faced several major challenges, including (1) random library representation due to

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

low cellular engraftment efficiency, (2) variability in subclonal expansion resulting from intra-tumoral heterogeneity, and (3) overrepresentation of outlier clones, all of which can confound data interpretation.[18] To address these challenges, we implemented the CRISPR-StAR (Stochastic Activation by Recombination) platform.[14] This technology builds on the now-common use of unique molecular identifiers (UMIs) as clonal barcodes[19–23] to identify outlier clones and provide robust statistics for each sgRNA. However, this approach alone fails to overcome bottlenecks from inconsistent tumor take rates or variable clonal growth. The CRISPR-StAR platform resolves these issues by integrating an inducible Cre/Cas9 dual system.[14] In this method, the sgRNA library vector contains a stop cassette flanked by two pairs of lox sites, preventing the expression of functional sgRNAs until Cre recombinase is activated. Upon 4-hydroxytamoxifen (4-OHT) treatment, CreERT2 activity generates two constructs with equal frequency: an active, functional sgRNA that allows for target gene KO and an inactive sgRNA that serves as a clonal control within the same microenvironment. This provides a dynamic internal control for each tumor-initiating clone, enabling clonal "lineage" tracking[14] that mitigates library representation bias and accounts for tumor heterogeneity.

Several computational tools, such as MAGeCK,[24] ScreenBEAM,[25] BAGEL,[26] and JACKS,[27] have been developed to analyze the pooled CRISPR screen data. However, these tools commonly use methods that model bulk sgRNA abundance to determine gene-level impact and are not well suited to the discrete clone-resolved data generated by CRISPR-StAR. To effectively analyze the complex next-generation sequencing (NGS) data generated by CRISPR-StAR, we developed a dedicated computational pipeline capable of extracting meaningful sgRNA dropout signals from individual tumors. This UMIBB pipeline (available on GitHub) utilizes embedded barcodes to achieve high-resolution clonal analysis. By conducting *in vivo* dropout screens in the A549 lung cancer xenograft model, followed by statistical analysis and computational downsampling, we demonstrate that 30,000 sgRNAs can be functionalized *in vivo* using tumors from just 30 mice. This represents a log-order reduction in animal usage relative to conventional approaches,[13] made possible by the unique ability to track clonal dynamics and incorporate internal normalization, improving efficiency and reproducibility in *in vivo* genetic screening.[14] Of note, our new UMIBB pipeline enables versatile, barcode-based clonal analysis for a wide range of applications, from CRISPR-StAR screens to other studies utilizing UMIs such as barcoding studies[28] or organoid studies.[29–31]

By incorporating a parallel *in vitro* comparator arm, we further identified genes whose loss promotes tumor growth exclusively *in vivo*. This set was enriched for epigenetic regulators, including members of the COMPASS family and SWI/SNF chromatin remodeling complexes. Single-gene KO studies demonstrated that loss of *KMT2C* markedly accelerated tumor growth *in vivo*, whereas *KMT2D* KO had the opposite effect. Neither gene exhibited a phenotype in 2D culture, highlighting the unique insights that can emerge from large-scale *in vivo* genetic screening.

Collectively, the inducible CRISPR-StAR platform, coupled with a novel bioinformatic framework, substantially enhances the fidelity, scalability, and interpretability of *in vivo* functional genomics screens. By temporally controlling sgRNA library activation in established tumors, the CRISPR-StAR platform overcomes key bottlenecks of traditional *in vivo* KO screening methodologies[13,14] with the CSTAR-UMIBB analysis package being broadly accessible to the research community.

RESULTS

CRISPR-StAR: A controlled dual-lox-based inducible system for expression of matched-pair sgRNAs with unique molecular barcodes for clonal tracking

In CRISPR-StAR,[14] the sgRNA expression cassette contains a stop element within the tracrRNA sequence that is flanked by two incompatible lox site pairs (*LoxP* and Lox5171), preventing guide activity prior to recombination (Figure 1A). Upon 4-OHT induction of CreERT2, recombination resolves the cassette into two outcomes occurring at similar frequency: (1) an active sgRNA that restores guide function and enables target gene KO and (2) an inactive construct that is transcribed but nonfunctional. At endpoint, the ratio of active vs. inactive (A/I) reads provides a within-clone measure of KO-associated fitness, where depletion (A/I < 1) or enrichment (A/I > 1) reflects the impact of target loss *in vivo* (Figure 1B).

Each library element also carries an embedded molecular barcode, so individual tumor clones are uniquely identified by the combination of its barcode and sgRNA sequence in the endpoint dataset.[20] This design provides matched-pair internal controls and enables clonal-level tracking of *in vivo* KO phenotypes while generating discrete and UMI-rich data, motivating an analysis framework that aggregates phenotypes across many clonal events rather than relying solely on guide-level abundance.

UMIBB: A Bayesian-based pipeline for clonal-resolved analysis of CRISPR-StAR datasets

To analyze CRISPR-StAR datasets, we developed a computational pipeline, UMIBB (Bayesian beta-binomial), designed to leverage the internal matched-pair active/inactive controls and the large number of independent clonal events captured per tumor (Figures 1C and S1). Because the ratio of active to inactive sgRNA counts (A/I ratio) provides a direct measure of sgRNA-induced phenotypes at endpoint, we implemented a Bayesian inference model to quantify dropout and enrichment across clones. In this framework, each clone is treated as a Bernoulli trial, for which the probability of observing sgRNA depletion (A/I < 1) or enrichment (A/I > 1) can be modeled by a beta-binomial distribution (Figures 1B and 1C).

To summarize phenotypic trends across clones targeting the same gene, we defined a weighted-theta score (wθ) ranging from 0 to 1. This score reflects the weighted average frequency of enriched active constructs across sgRNAs targeting the same gene, with weights proportional to the number of informative clones per sgRNA. A wθ score <0.5 indicates depletion, while scores >0.5 indicate enrichment. Importantly, the wθ score provides an interpretable measure of phenotype robustness, while the statistical significance of each phenotype is captured by the Bayesian-derived, false discovery rate (FDR)-adjusted *p* value (Figures 1C and S1).

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

CellPress
OPEN ACCESS

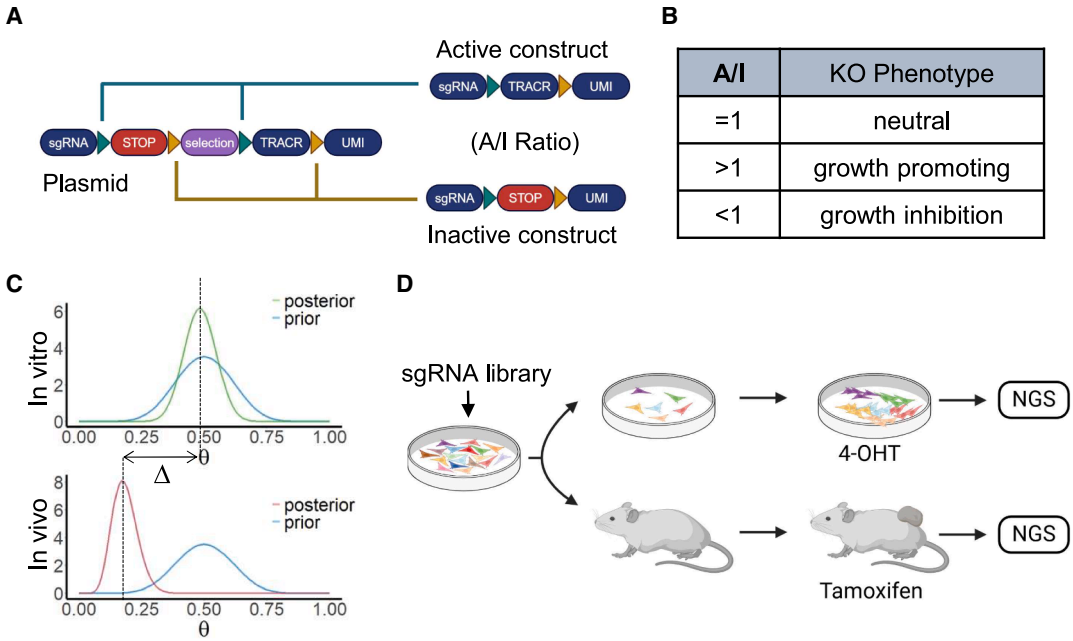


Figure 1. Overview of the CRISPR-StAR system and the Bayesian analysis (UMIBB) of screening results

(A) CRISPR-StAR vector design enables Cre-mediated stochastic recombination via two pairs of incompatible lox sites (*LoxP* and Lox5171), generating (with equal efficiency) either active (restored sgRNA-tracrRNA, resulting in gene knockout) or inactive (tracrRNA-deleted) constructs.

(B) Knockout phenotypes are quantified as the ratio of active to inactive reads (A/I ratio).

(C) Bayesian modeling framework for estimating wθ and Δ values in the UMIBB pipeline.

(D) CRISPR-StAR screen setup. *In vitro*: 4-OHT (300 nM) was added to the culture media for 4 days to induce recombination. *In vivo*: tamoxifen was administered to tumor-bearing mice to induce recombination.

To compare conditions (e.g., *in vivo* vs. *in vitro*), we applied a Bayesian statistical test of differences in binomial proportion estimates (wθ) between the two conditions, using non-targeting control (NTC) clones to set priors (Figure 1C). To quantify the magnitude of differential effect, we introduced the weighted delta (wΔ) score, defined as the absolute difference in wθ values between the two conditions. This metric captures the extent to which a gene's phenotype diverges *in vivo* versus *in vitro*. The statistical significance of wΔ scores was evaluated using FDR-adjusted *p* values, providing confidence in this differential score (Figures 1C and S1).

Validation of CRISPR-StAR for large-scale *in vivo* screening in human tumor xenografts

To optimize the CRISPR-StAR platform for human xenograft models, we infected the non-small cell lung cancer cell line A549 with both the CreERT2 and Cas9 constructs and selected for single-cell clones that can (1) tolerate the combined nuclease activities of the two enzymes and (2) exhibit a high Cre recombination rate and Cas9 editing efficiency. This minimized confounding effects from CreERT2 activation and ensured high screening quality. Indeed, pilot xenograft studies showed equivalent tumor growth with and without tamoxifen treatment in the resulting A549 clones (Figure S2A).

We next conducted a large-scale pooled CRISPR screen using the A549 model *in vitro* and *in vivo* (n = 118 tumors) (Figure 1D) with a 30,000-sgRNA library (four guides per gene). Clones with less than 10 reads (active and inactive combined counts) were excluded for noise reduction. The number of clones and the sum of active and inactive counts in each experimental group before and after QC are provided in Tables S1–S3.

A key feature of CRISPR-StAR is the near-balanced recombination outcomes, with stochastic activation yielding an A/I ratio near 1 for effective noise reduction.[14] As such, it is critical that Cre recombination is induced when each clone has expanded sufficiently to ensure reproducibility of this stochastic event. *In vivo*, this was achieved by delaying tamoxifen treatment until tumors were established (Figure S2B). We confirmed sustained Cas9 expression at the time of recombination by immunohistochemistry (Figure S2C) and observed efficient Cre-mediated recombination in tumors (Figure S2D), both prerequisites for robust gene editing.

In the *in vitro* screening arm, we included an intentional bottleneck post-library transduction to mimic *in vivo* engraftment constraints, followed by a 50-fold expansion prior to 4-OHT induction (Figures 1D and S2E). We first compared the performance of the UMIBB analysis, which took advantage of the CRISPR-StAR clonal data structure, versus the performance of the MAGeCK analysis, which represented the conventional CRISPR screening analysis based on guide-level counts only (Figure S2F). The UMIBB analysis demonstrated better statistical sensitivity and robustness (precision = 0.97, recall = 0.92, F1_Score = 0.95) than conventional MAGeCK maximum-likelihood estimation (MLE) analysis (precision = 1, recall = 0.68,

CellPress
OPEN ACCESS

Cell Reports Methods
Article

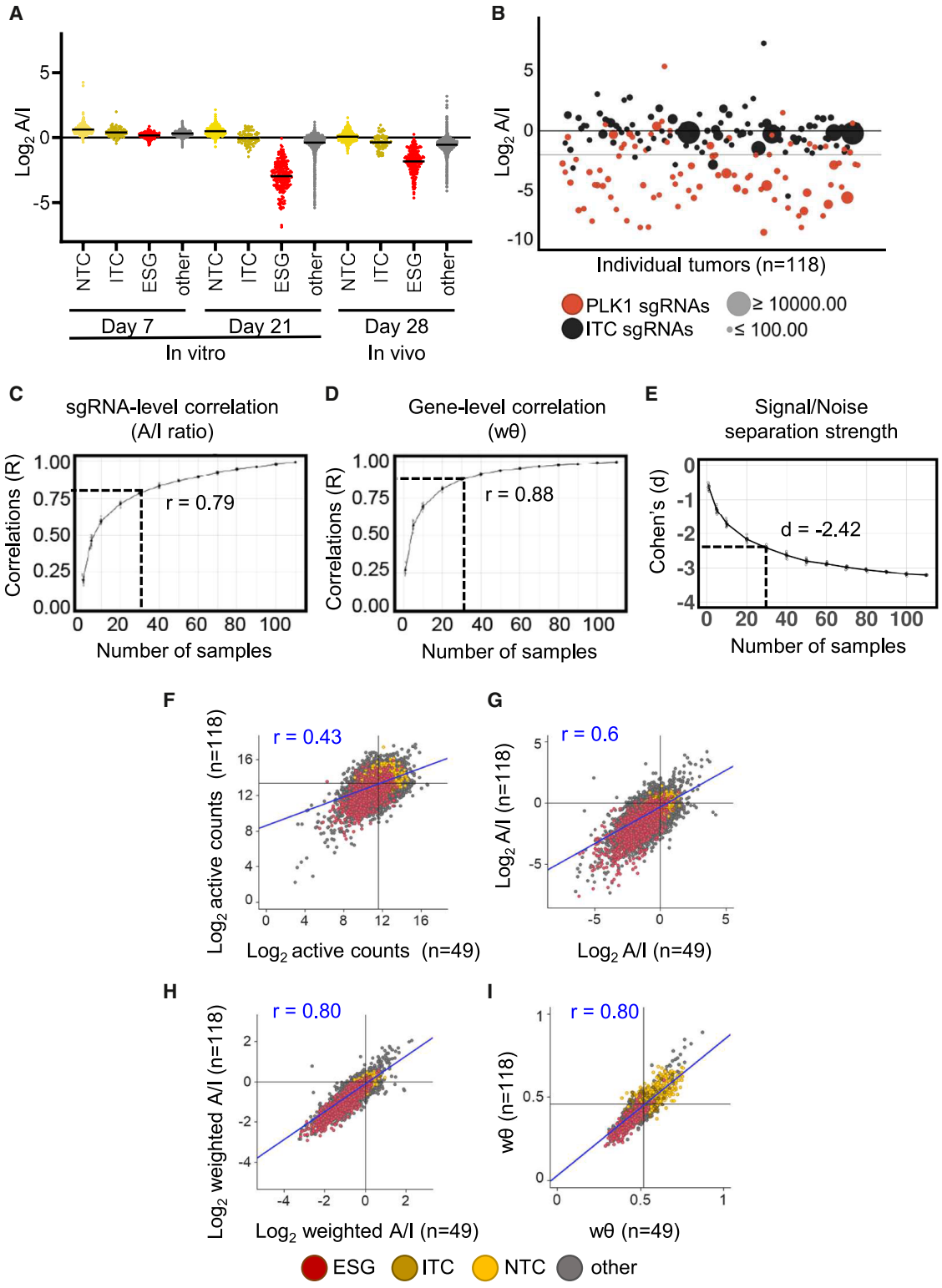


Figure 2. Analysis of a 30K sgRNA CRISPR-StAR screen and downsampling performance

(A) Clonal log_2A/I ratios, averaged across sgRNAs and grouped by gene type, are plotted over time post 4-OHT (*in vitro*) or tamoxifen (*in vivo*) treatment. The library includes 198 essential genes (ESGs, 792 sgRNAs), 72 sgRNAs targeting intronic regions (ITC) and 499 non-targeting sgRNAs (NTC), and 6,925 other genes (27,700 sgRNAs).

(legend continued on next page)

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

F1_Score = 0.81) in terms of differentiating known essential genes (defined as a Chronos gene effect score < −1 in >90% of cell lines in DepMap/Achilles[1]) from non-essential genes (defined as a Chronos gene effect score > −0.3 in all of cell lines in DepMap/Achilles[1]).

In the *in vivo* arm, aggregating data from 118 tumors provided comprehensive library representation (Figure S3A), with each sgRNA supported by ≥40 distinct clones, comparable to the *in vitro* coverage (Figure S3B). The number of library-integrated, clonally expanded cells that contributed meaningfully to each tumor plateaued at ∼30,000 clones independent of final tumor size (Figure S3C), indicating a limit on the number of clones capable of successful engraftment and outgrowth. Injecting 5-fold more or fewer cells did not impact either tumor growth or clonal complexity (Figure S3D), supporting an engraftment bottleneck in this A549 xenograft model. Consistent with CRISPR-StAR assumptions, NTC guides exhibited a near-unity A/I ratio (mean = 1.032, SD = 0.017; p = 0.055, t test) *in vivo*, confirming the balanced recombination of CRISPR-StAR (Figure S3E).

The CRISPR-StAR screen reliably distinguished core essential genes (ESGs) from NTCs and intron-targeting controls (ITCs) in both *in vitro* and *in vivo* arms (Figure 2A). As expected, across multiple experiments, sgRNA dropout was not observed at day 7 post-induction, highlighting the lag time required for recombination and gene editing; thus, we extended the *in vitro* screening endpoints to 21 days and the *in vivo* endpoint to 28 days (Figure 2A). Despite partial library representation within individual tumors, informative dropout signals were extracted using clonal $\log_2$A/I ratios. sgRNAs targeting *PLK1* showed consistent depletion across tumors, whereas NTCs were stable and ITCs exhibited only modest cutting-associated effects that remained distinct from essential gene dropout by day 28 (Figure 2B).

Although variability in clonal $\log_2$A/I ratios was observed—likely reflecting both editing efficiency and NGS limitations—our analysis pipeline effectively distinguished true biological signal from technical noise. By plotting the clonal size (as measured by total sequencing reads per clone) against the $\log_2$A/I ratio in each clone, we generated density maps to visualize the overall behaviors of large numbers of clones (Figure S3F). As expected, larger clones exhibited reduced noise, likely due to mitigation of PCR- and NGS-related distortions at higher read counts (Figure S3F).

Simulation and benchmarking of the UMIBB computational pipeline

Having established UMIBB performance on experimental CRISPR-StAR datasets, including robust separation of essential genes from controls and improved reproducibility via matched active/inactive normalization, we performed simulation-based benchmarking to quantify method robustness across sequencing depth and recombination regimes that commonly vary between *in vivo* tumors. Specifically, we developed a simulation framework that generates paired active and inactive clonal counts under tunable A/I recombination ratios, while explicitly modeling NGS sampling noise, guide-to-guide variability in editing/fitness effects, and stochastic clonal growth. Simulated datasets were processed using the same filtering as experimental data (excluding clones with <10 total reads).

We first simulated a single-arm CRISPR-StAR experiment comprising a 250-gene library (30 NTCs, 10 essential genes as depletion controls that inhibit tumor growth, 10 tumor suppressor genes [TSGs] as enrichment controls, and 200 neutral genes), with four guides per gene and 200 clones per guide. Under baseline conditions of balanced recombination (A/I recombination ratio = 1) and high sequencing depth (4 M reads), the median wθ for NTCs centered at 0.5, and the NTC wθ distribution defined an empirical noise boundary. Using an FDR-adjusted $p < 0.05$ and an NTC-derived wθ range [0.45, 0.55] to call significant phenotypes, UMIBB recovered all engineered true positives (i.e., essential genes showing a depletion phenotype [wθ < 0.45] and TSGs showing an enrichment phenotype [wθ > 0.55]) while excluding neutral genes (Figure S4A). Across sequencing depths (100K to 4 M reads; 10 replicates per depth), increased depth yielded more clones passing QC (combined counts >10), improving statistical power (higher true positive rate [TPR], lower false positive rate [FPR]; Figures S4B and S4C), whereas low depth expanded the NTC wθ range, reflecting increased uncertainty and reduced separability of signal from noise when limited depletion/enrichment events are available as evidence (Figure S4D).

To assess robustness to recombination imbalance, we varied the A/I ratio in single-arm simulations. UMIBB maintained strong sensitivity (TPR > 0.75) when recombination remained within a moderately balanced range (0.5–2.0) (Figure S4E) and normalization preserved NTC calibration, keeping the NTC median wθ close to 0.5 even under skewed recombination ratios (Figure S4F). Finally, we simulated a two-arm dataset (20 NTCs, 10 essential genes, 10 TSGs, 200 neutral genes, and 10 genes engineered to exhibit differential growth between arms) under otherwise matched settings. In this two-condition framework, the differential metric wΔ for NTCs centered at 0, and an empirical NTC boundary of −0.05 to 0.05 captured most noise hits (Figure S4G). Across sequencing depths, increased coverage improved power to detect engineered differential genes (higher TPR, lower FPR; Figure S4H), consistent with the requirement for sufficient informative clonal events to shrink

(B) Example dropout of the essential gene PLK1 across 118 individual tumors at endpoint compared to intron-cutting control (ITC). Each point represents the weighted average $\log_2$(A/I) ratio per gene per tumor; circle size reflects total read count per gene.
(C and D) Downsampling analysis of *in vivo* data showing (C) sgRNA-level ($\log_2$A/I) and (D) gene-level (wθ) correlation values between subsets of tumors and the full 118-tumor dataset. x axis: number of tumors in the downsampled subset; y axis: average correlations of the randomly selected tumor subset versus the full 118-tumor dataset (n = 10). Dotted line indicates correlations when 30 random tumors are sampled (r = 0.79 for A/I ratio; r = 0.88 for wθ).
(E) Sensitivity analysis showing signal/noise separation strength between downsampled subsets of tumors and the full 118-tumor dataset. x axis: number of tumors in the downsampled subset; y axis: average Cohen's d value of the randomly selected tumor subset versus the full 118-tumor dataset (n = 10).
(F–I) Reproducibility of true biological replicates (118 vs. 49 mice) for the 30K sgRNA library screen: (F) correlation of sgRNA-level $\log_2$ active counts (r = 0.43; proxy for traditional screens without internal controls), (G) correlation of sgRNA-level $\log_2$A/I ratios leveraging internal normalization (r = 0.6), (H) correlation of sgRNA-level $\log_2$A/I ratios weighted by clonal size (r = 0.8), and (I) correlation of gene-level wθ score from the CSTAR-UMIBB analysis (r = 0.8).

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

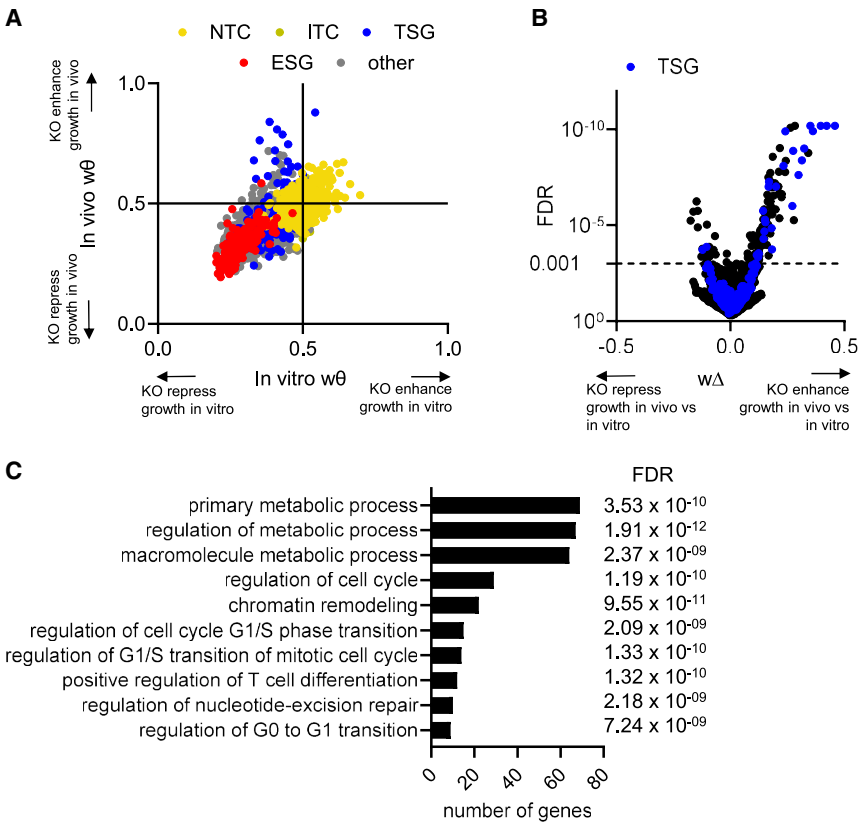


Figure 3. CRISPR-StAR screen identifies *in vivo*-specific gene phenotypes

(A) Comparison of gene-level wθ scores from *in vivo* and *in vitro* arms of the 30K sgRNA library screen. Colors indicate sgRNA categories: non-targeting controls (NTCs, yellow), intronic controls (ITCs, gold), tumor suppressor genes[32] (TSGs, blue), essential genes (ESGs, red), and all others (gray).

(B) CSTAR-UMIBB analysis comparing *in vivo* vs. *in vitro* phenotypes. *x* axis: Δ value attributed by the CSTAR-UMIBB analysis for *in vivo*-specific effect. *y* axis: false discovery rate (FDR). Dotted line marks FDR = 0.001 threshold used to identify 173 hits.

(C) Gene Ontology (GO) pathway enrichment analysis of top *in vivo*-selective hits from (B), showing selected pathways and associated FDR values. Analysis performed using https://geneontology.org/.[33,34]

the NTC-defined noise band (Figure S4I). Sensitivity for differential phenotypes was highest under near-balanced recombination, with strong performance maintained over a narrower recombination range (0.75–1.33) (Figure S4J), while wΔ noise levels remained comparably constrained by NTCs (Figure S4K). Collectively, these simulations show that UMIBB remains well calibrated on negative controls and robust to reduced sequencing depth and recombination imbalance, maintaining FDR <0.05 across tested scenarios.

In silico downsampling reveals minimal tumor requirement for high-power screening (~30,000 sgRNAs require 30 tumors)

Given the depth and complexity of our dataset, we hypothesized that our CRISPR-StAR screen may be statistically overpowered and that fewer tumors could yield reliable dropout signals. To test this, we performed *in silico* downsampling of the A549 xenograft screen (n = 118 tumors), analyzing sgRNA-level $\log_2$(A/I) ratios and gene-level wθ scores (Figures 2C and 2D). Dropout signals remained highly correlated with the full dataset even with substantial reductions in sample size. Notably, a cohort of just 30 tumors was sufficient to maintain strong concordance with the full 118-tumor screen at both the sgRNA (r = 0.79) and gene level (r = 0.88). The effect size, measured by Cohen's d (d = −2.42) of wθ values between essential genes and NTCs, showed that strong separation of signal from noise could be achieved in 30 tumor samples (Figure 2E). Together, these results indicate that CRISPR-StAR can retain high statistical power with markedly reduced animal usage relative to conventional xenograft dropout screens, which would have used hundreds of mice (Figures 2C–2E).

Experimental validation of screening depth and reproducibility of CRISPR-StAR

To experimentally validate the downsampling predictions and assess reproducibility across biological replicates, we next repeated the A549 CRISPR-StAR screen using just 49 xenograft tumors. We first evaluated the reproducibility of sgRNA dropout based solely on active read counts, mimicking conventional CRISPR-UMI[20] screening without internal controls. This approach yielded low correlation across replicates (r = 0.43; Figure 2F), highlighting the inherent noise in *in vivo* pooled screens.[19–23] Incorporating inactive guides for sgRNA-level normalization substantially improved replicate concordance (r = 0.60; Figure 2G). Integrating clone size and matched-pair normalization via UMIBB further increased reproducibility (r = 0.80, Figure 2H), with gene-level wθ scores also high correlated between replicates (r = 0.80, Figure 2I). Together, these results demonstrate robust and reproducible sgRNA phenotyping *in vivo* with CRISPR-StAR despite a ~60% reduction in sample size. Based on clone engraftment rates, downsampling simulations, and biological replication, we estimate that approximately 1,000 sgRNAs can be confidently resolved per tumor in this model, substantially enhancing screening efficiency while minimizing animal use.

Identification of *in vivo*-specific phenotypes using CRISPR-StAR

To pinpoint *in vivo*-specific effector genes, we compared gene-level wθ scores derived from *in vitro* and *in vivo* screens (Figure 3A). This analysis identified a subset of genes that selectively promoted tumor growth *in vivo* with minimal effects on

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

CellPress
OPEN ACCESS

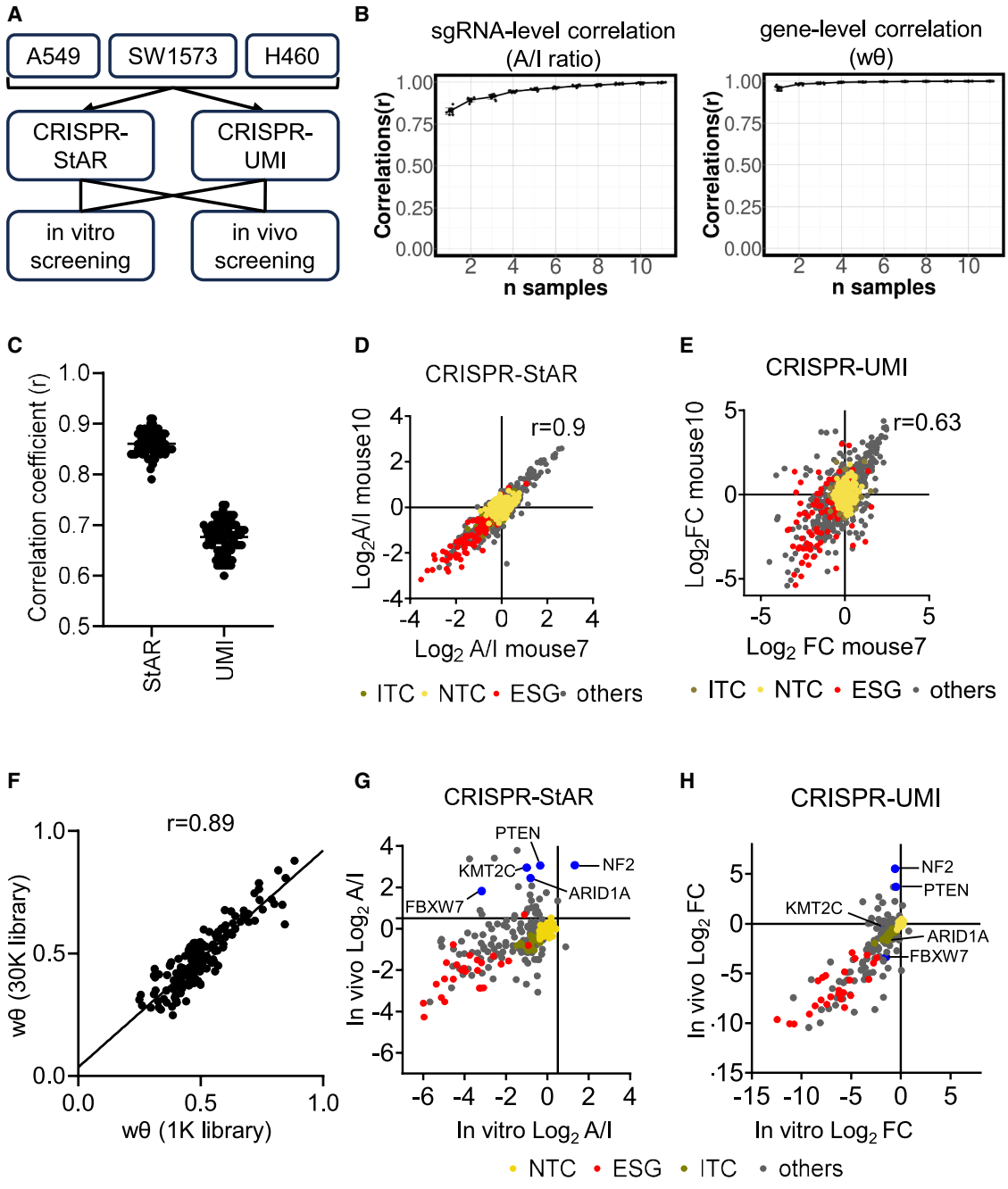


Figure 4. Validation of primary screening hits and comparison of CRISPR-StAR vs. CRISPR-UMI

(A) Schematic of mini-pool validation screen using a 1K sgRNA library targeting top enrichers and depletors from the 30K screen. Screens were performed in CRISPR-StAR and constitutive CRISPR-UMI systems across three non-small cell lung cancer (NSCLC) cell lines.

(B) Downsampling analysis of CRISPR-StAR *in vivo* samples. sgRNA-level $\log_2$(A/I) and gene-level wθ correlations were computed between subsets and the full 12-tumor dataset. *x* axis: number of tumors sampled; *y* axis: average correlation.

(C) Average pairwise Pearson correlation (r) of sgRNA profiles across all tumor pairs within each arm (CRISPR-StAR and CRISPR-UMI). Full pairwise plots shown in Figure S5B.

(D) Representative correlation between two CRISPR-StAR tumors, showing clonal average $\log_2$(A/I) ratios per sgRNA.

(E) Representative correlation between two CRISPR-UMI tumors, showing $\log_2$ fold change of sgRNA counts between day 28 (study endpoint) and day 0 (injection).

(F) Concordance (scatterplot) of gene-level wθ values between the original 30K screen and the 1K validation mini-pool. Pearson correlation shown.

(G) Gene-level comparison of *in vivo* vs. *in vitro* wθ values from the CRISPR-StAR 1K mini-pool validation screen.

(H) Gene-level comparison of *in vivo* vs. *in vitro* fold change from the CRISPR-UMI 1K mini-pool validation screen.

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

proliferation *in vitro*. Applying a stringent significance threshold (ΔFDR < 0.001) on the weighted Δ score, we identified 173 genes with differential scores *in vivo* vs. *in vitro*. Among these, KO of 148 genes selectively enhanced tumor growth *in vivo* (Figure 3B).

Notably, this *in vivo*-selective gene set was enriched for TSGs,[32] with ~20% of the *in vivo* enrichers categorized as TSGs, while only 2.7% of non-significant genes were TSGs (Figure 3B), consistent with prior studies.[35] Furthermore, there was a marked overrepresentation of epigenetic regulators among the *in vivo*-enriched genes (Table S4), including members of the COMPASS family and SWI/SNF chromatin remodeling complexes. Gene Ontology enrichment analysis[33,34] confirmed that chromatin organization and remodeling pathways were significantly overrepresented among the *in vivo*-specific hits (Figure 3C), suggesting a role for epigenetic regulation in tumor suppression *in vivo*. Pathway analysis further highlighted that metabolic processes were substantially different between the *in vitro* and the *in vivo* conditions, consistent with distinct selection pressures in tumors relative to standard culture conditions.[36]

Mini-pool gene validation confirms *in vivo*-specific hits across multiple cancer cell lines

To validate *in vivo*-specific hits across models and to confirm that a single tumor can functionalize ~1,000 sgRNAs, we built a 910-sgRNA mini-pool library targeting top-ranked genes. This library was cloned into both a traditional constitutive CRISPR-UMI backbone[20] and the inducible CRISPR-StAR system[14] (Figure 4A), allowing head-to-head comparison. While CRISPR-UMI enables clonal tracking, constitutive sgRNA expression leads to pre-engraftment gene editing and early *in vitro* phenotype. It also lacks the matched-pair internal controls of CRISPR-StAR,[14,20] increasing susceptibility to outlier-driven bias. The validation screens were performed across three human cancer cell lines (A549, NCI-H460, and SW1573) using comparable animal numbers per arm to ensure adequate UMI library representation for subsequent bioinformatics analysis. Indeed, across all three xenograft models, both platforms reliably separated core essential genes from NTCs (Figure S5A), confirming technical robustness.

Importantly, CRISPR-StAR achieved consistent results at the predicted resolution of ~1,000 sgRNAs per tumor, supported by sufficient clonal coverage (Figure 4B). Computational downsampling analysis of the dataset showed that a single tumor could recapitulate the signal obtained from a 12-tumor cohort with a correlation coefficient of r = 0.96 for wθ and r = 0.83 for sgRNA-level A/I ratio (Figure 4B). Pairwise analysis demonstrated higher reproducibility across individual animals in the CRISPR-StAR arm compared with CRISPR UMI (Figure S5B). This is also evident when the coefficient scores of all possible intra-tumor comparisons were summarized in the dot plot (Figure 4C). As a representative example, the correlation between mouse 10 and mouse 7 in the CRISPR-StAR arm shows a correlation coefficient of r = 0.9 (Figure 4D), while the correlation of 2 representative animals in the UMI cohort has a lower correlation coefficient of r = 0.63 (Figure 4E). These data demonstrate the ability of single-tumor CRISPR-StAR screens to resolve meaningful phenotypes at this scale.

In contrast, although UMI-based data also showed high correlation when endpoint counts were normalized to plasmid counts (data not shown), closer examination revealed significant pre-transplant drift in sgRNA representation—biases introduced during *in vitro* expansion of the cell pool prior to xenografting due to the *in vitro* gene KO effects. This drift reduced the reproducibility in single-tumor UMI screens unless the data were normalized to the pre-injection pool (Figure S5C). CRISPR-StAR, by contrast, allows direct *in vivo* assessment without dependence on pre-transplant sgRNA distributions, thereby reducing batch effects and improving signal fidelity.

Unlike conventional CRISPR screening methods, where increased library size can diminish resolution due to competition among sgRNAs, CRISPR-StAR's internal control structure ensures that sgRNA dropout measurements are independent of library composition. As a result, gene-level dropout signals in the mini-pool validation screen closely matched those observed in the original 30K-library screen (Figure 4F), despite differences in sgRNA sequences and library context. No additional enrichment or dropout was observed in the mini-pool screen, indicating screen saturation. Collectively these findings demonstrate that CRISPR-StAR enables efficient discovery and validation of *in vivo*-specific hits across models and supports high-fidelity screening of up to 1,000 sgRNAs per tumor.

Comparison of *in vivo*-specific hits between UMI and CRISPR-StAR libraries

We next analyzed *in vivo*-specific hits identified by both the CRISPR-StAR and UMI-based libraries in A549 xenograft models (Figures 4G and 4H). Several gene dependencies that were robustly detected using CRISPR-StAR were not captured in the constitutive UMI KO screen (i.e., *FBXW7*, *KMT2C*, and *ARID1A*). This discrepancy likely reflects fundamental design differences: CRISPR-UMI induces gene KO before engraftment, potentially eliminating cells with early *in vitro* fitness defects that impair tumor formation. In contrast, CRISPR-StAR activates gene KO only after tumors are established, allowing interrogation of functional selection exclusively within the *in vivo* tumor context.

Extending this validation to three non-small cell lung cancer cell lines revealed distinct sets of *in vivo*-selective dependencies in each model (Figures S6A and S6B), underscoring the cell-line-specific nature of tumor vulnerabilities and the necessity of context-aware screening strategies. Overall, this comparison highlights the enhanced sensitivity of CRISPR-StAR for identifying *in vivo* gene dependencies that may go undetected with conventional UMI-based *in vivo* or *in vitro* screens.

Divergent roles of KMT2C and KMT2D in tumor growth revealed by *in vivo* CRISPR-StAR and validated by single-gene KO studies

To complement our pooled CRISPR-StAR screens, we conducted single-gene validation of the chromatin regulators KMT2C and KMT2D. Both genes encode lysine-specific methyltransferases that function within COMPASS complexes, regulating gene expression programs critical for differentiation and development.[37] Although they are often considered tumor

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

CellPress
OPEN ACCESS

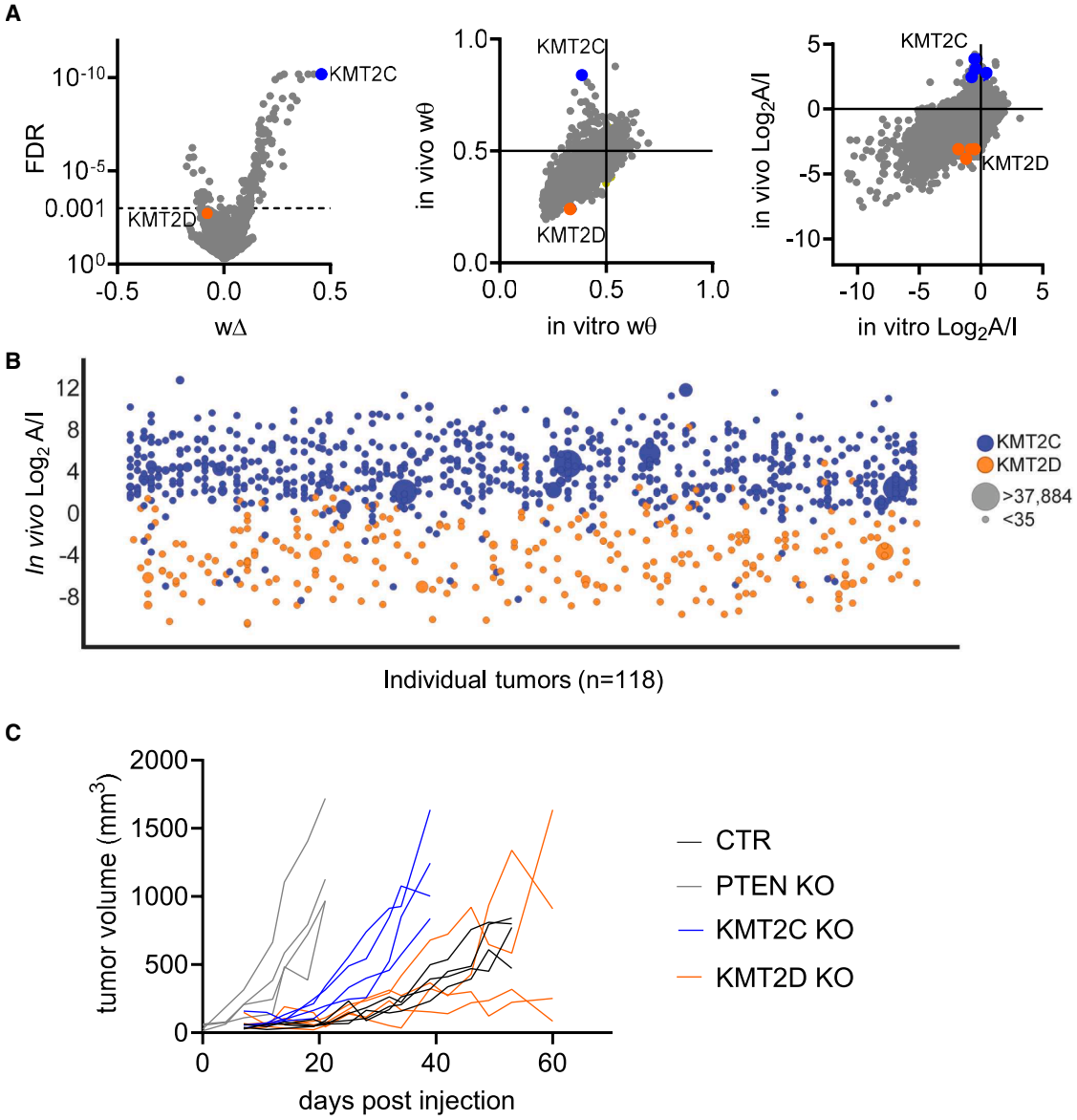


Figure 5. KMT2C and KMT2D have opposite *in vivo* phenotypes in A549

(A) CSTAR-UMIBB analysis highlighting divergent effects of *KMT2C* and *KMT2D*. Left: Δ values (*in vivo* vs. *in vitro*) vs. FDR, with *KMT2C* and *KMT2D* marked. Middle: *in vivo* vs. *in vitro* wθ gene scores. Right: sgRNA-level average $\log_2$(A/I) ratios *in vivo* vs. *in vitro*, with *KMT2C* and *KMT2D* sgRNAs highlighted.

(B) Distribution of $\log_2$(A/I) ratios for individual clones pertaining to either *KMT2C* or *KMT2D* sgRNAs across tumors. Each point represents a clone; each *x* axis position is a separate tumor. Circle size reflects clone size (total sgRNA read counts).

(C) *In vivo* tumor growth curves of A549 xenografts with knockout of *KMT2C*, *KMT2D*, *PTEN* (positive control), or intron-targeting control (CTR). Tumor volumes plotted individually over time in NCG mice.

suppressors and functionally redundant,[38–43] our *in vivo* screen in A549 xenografts revealed divergent phenotypes.

Specifically, KMT2C KO consistently enhanced tumor growth, indicating a tumor-suppressive function, whereas KMT2D KO impaired tumor progression, suggesting it is required for *in vivo* tumor maintenance (Figure 5A, left). These opposing effects were supported by gene-level wθ values (Figure 5A, middle) and reproducible $\log_2$A/I ratios across all four sgRNAs per gene (Figure 5A, right). Clonal-resolution analysis of $\log_2$A/I ratios further confirmed the trend: KMT2C KO clones were consistently enriched, while KMT2D KO clones were depleted across tumors (Figure 5B).

To validate these findings, we generated isogenic A549 cell lines harboring CRISPR KOs of KMT2C or KMT2D, along with intron-targeting and PTEN control sgRNAs (Figures S7A–S7E). When these lines were implanted subcutaneously into immunodeficient mice, KMT2C KO accelerated tumor growth, whereas KMT2D KO did not exhibit the same phenotype (Figure 5C).

CellPress
OPEN ACCESS

Cell Reports Methods
Article

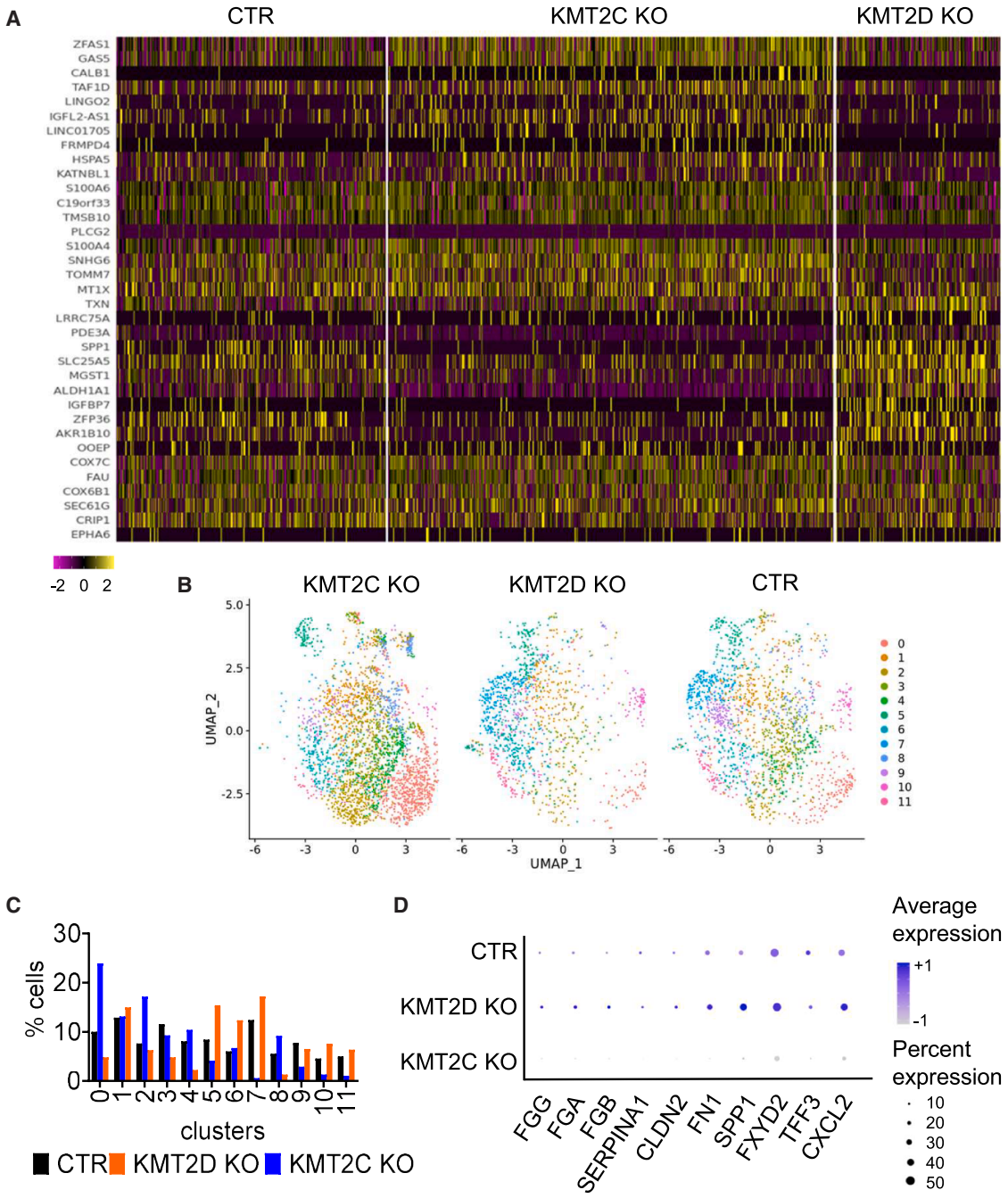


Figure 6. Single-cell RNA-seq analysis of KMT2C and KMT2D KO tumors

(A) Heatmap of top differentially expressed genes in KMT2C KO tumor vs. control sample based on single-cell RNA-seq data (each block represents one single cell; expression of the same genes is shown for the KMT2D KO tumors).

(B) UMAP clustering of single-cell transcriptomes from KMT2C KO, KMT2D KO, and control tumors.

(C) Quantification of each cluster shown in (B) expressed as the percentage of total cells per group.

(D) Key genes comprising cluster 7 in (B) and (C). The average expression of each gene in either control, KMT2D KO, or KMT2C KO samples was plotted. Dot size represents the percentage of cells expressing each gene; color indicates average expression level across all cells within each group.

Finally, we performed single-cell RNA sequencing on time-matched endpoint tumors to validate that underlying transcriptional mechanisms were driving these divergent phenotypes. As expected from COMPASS-mediated enhancer regulation,[37] KMT2C loss led to broad transcriptional changes compared with intron-targeting negative controls (CTR), with the most differentially expressed genes between the two genotypes shown in the heatmap (Figure 6A) and top differentially regulated

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

pathways as measured by gene set enrichment analysis (GSEA) shown in Figure S7F. Interestingly, many genes were regulated in the opposite direction in KMT2D KO tumors (Figure 6A), indicating reciprocal regulatory roles for these methyltransferases in tumor growth and transcriptional control.

Unbiased UMAP (uniform manifold approximation and projection) clustering of single-cell transcriptomes also revealed distinct tumor cell populations associated with each genotype (Figure 6B). Quantification of the percentage of cells in each cluster compared with the total pool revealed multiple clusters with distinct representation between KMT2C and KMT2D KO tumors (Figure 6C). Notably, a cluster of cells (cluster 7), characterized by high expression of cell-cell interaction genes in control and KMT2D KO tumors, was completely absent in KMT2C KO tumors (Figures 6B–6D). This is also consistent with the GSEA analysis (Figure S7F), where the coagulation pathway is one of the most differentially regulated gene sets between KMT2C KO and intron-cutting controls. Together, these data highlight the power of CRISPR-StAR to uncover divergent *in vivo* functions of structurally related genes.

DISCUSSION

CRISPR-based functional genomics has transformed cancer research by enabling systematic identification of genetic dependencies that underlie tumor proliferation, survival, and immune evasion. Landmark efforts such as DepMap and Achilles[1] projects have mapped essential genes across hundreds of cell lines, revealing tumor-context vulnerabilities and synthetic lethal interactions that continue to guide therapeutic development. However, *in vitro* systems are limited in capturing key features of the tumor microenvironment, including cellular heterogeneity, stromal interactions, and nutrient or oxygen gradients that shape tumor behavior and therapeutic response.[13,15]

Our study addresses this gap through implementation of CRISPR-StAR in mouse models of human cancer, enabling activation of gene editing in established xenograft tumors.[14] By temporally separating engraftment from gene perturbation, CRISPR-StAR minimizes pre-engraftment selection and captures gene dependencies that emerge within the *in vivo* tumor niche. Coupled with a computational pipeline, implementing a Bayesian statistical framework to exploit internally matched sgRNA controls and clonal barcoding, this approach enables scripted high-resolution quantification of gene dropout and enrichment from noisy *in vivo* pooled screens.

Conventional *in vivo* CRISPR dropout screens are constrained by both engraftment bottlenecks (i.e., less than 2%–5% of injected cells contribute to the xenograft tumor formation) and uneven clonal expansion due to heterogeneous growth within the tumor microenvironment.[18] This leads to low sgRNA library representation, often addressed by increasing animal numbers, an approach that is resource intensive and only partially compensates for the underlying variability. Using clonal barcode tracing, such as CRISPR-UMI,[20] can help reduce the required coverage and exclude outlier clones; however, constitutive editing introduces gene KO before engraftment and can confound *in vitro* fitness effects with *in vivo* selection. Inducible Cas9 models[44] can delay editing but may be more variable in editing efficiency *in vivo* due to differences in induction, doxycycline exposure, and animal-to-animal response, and they do not directly correct for heterogeneous clonal growth within tumors. By contrast, CRISPR-StAR is a screening platform tailored for high-resolution *in vivo* screening[14] that achieves improved resolution by activating sgRNAs after tumor establishment, generating internal, matched-pair active/inactive sgRNA controls for each clone and directly addressing random variation. Our data build on the initial application and explore the resolution power of CRISPR-StAR by generating comprehensive single-tumor data.

Although tools like MAGeCK can be adapted to analyze CRISPR-StAR data by treating active and inactive clone counts as sgRNA abundance,[14] this approach is practical only when clonal representation and sample sizes are relatively small. In our experience, large CRISPR-StAR screens can be challenging to process using existing pipelines optimized for guide-level abundance, where the discrete clonal read counts may violate negative binomial model assumptions and are subject to false-positive findings due to outlier clones.[14] While prefiltering has been recommended to mitigate these effects, outlier exclusion is inherently arbitrary, motivating analytical frameworks that directly model clonal events.

To address these limitations, we developed UMIBB, a Bayesian beta-binomial framework that models paired active/inactive counts at the clone level as discrete depletion/enrichment events and aggregates evidence across many clonal observations per gene. This event-based representation limits the influence of extreme outlier counts because each clone contributes at most one depletion/enrichment observation, rather than dominating a gene statistic through raw read abundance. Therefore, the UMIBB approach does not require prefiltering of outlier clones as the adapted MAGeCK procedure.[24] On the other hand, the clonal size is often log-normally distributed, with many very small clones where stochastic sampling can yield missing active or inactive counts; these are filtered to reduce noise, consistent with prior UMI-based screening practices.[20] UMIBB models observed depletion/enrichment events using a beta-binomial framework with prior parameters estimated from the combined events in NTC guides. This further strengthens the robustness of the UMIBB method, as the posterior statistic combines the model prior with the observed evidence and requires a large enough number of events to move the posterior estimate away from the NTC noise range to reach significance. Tied events are always included in the null hypothesis events (i.e., not depleted or not enriched) in tests, making the UMIBB analysis even more stringent.

To further quantify robustness and address concerns about sensitivity to technical variation, we developed a simulation framework tailored to the matched-pair, barcode-resolved structure of CRISPR-StAR datasets. Simulations demonstrated that UMIBB remains robust under reduced sequencing depth and recombination imbalance, maintaining false-positives rates below 0.05. Because UMIBB depends on clonal depletion/enrichment events, the statistical power decreases only when clone numbers per guide are very low (<3), whereas robust detection is achieved with $\geq$7 clones per guide on average. By effectively utilizing UMI abundance, UMI representation, NTCs, and matched-pair controls in a combined statistical framework,

UMIBB leverages the clonal information within the CRISPR-StAR system, improving the robustness of the data by acknowledging the outlier effect.

Our study protocol provides a proof of concept for the scalability and resolution of CRISPR-StAR. Using statistical downsampling and validation screens, we estimate that a single tumor can resolve ~1,000 sgRNAs in the A549 model. Compared to conventional *in vivo* dropout screens requiring >100 mice to achieve sufficient library coverage for a 30K sgRNA library, CRISPR-StAR reduces animal use by up to 7-fold without sacrificing statistical power. As such, it makes it a more sustainable and scalable approach for large-scale *in vivo* screens. In our study, we were able to cover approximately one-third of the human genome (30K sgRNA library) with a representation of 40 independent clones (i.e., "experiments") per sgRNA. It should be noted that the exact number of evaluable clones per mouse is likely to differ from cell line to cell line in relation to the specific engraftment potential of the cell line model and the mouse strain. For example, NCIH292 and MIAPACA2 cell lines yield double the number of evaluable clones per mouse compared with our A549 models. However, regardless of the intrinsic engraftment potential of each model, CRISPR-StAR coupled with the UMIBB analysis pipeline can enable significant reductions in animal usage.

One of the most advantageous aspects of CRISPR-StAR is its use of inactive sgRNAs as an internal reference, which allows for reliable data extraction even from under-represented screens by focusing on sgRNAs retained in the inactive pool at the endpoint. Traditional CRISPR screening methods lack this reference population and rely on increasing animal numbers to achieve library representation. Additionally, CRISPR-StAR's internal controls also enhance cross-experimental comparisons. Based on our clone representation and dropout correlation analyses, the reproducibility of true biological replicates yielded highly correlated data across experiments.

Beyond platform performance, a central biological observation is that many TSGs exhibit minimal phenotypes in 2D culture yet exert strong effects *in vivo*. These genes, frequently mutated or deleted in human cancers, include PTEN, FBXW7, KMT2C, and ARID1A in our datasets (Figures 4G and 5C) highlighting the value of interrogating gene function under physiological constraints that are absent from conventional culture conditions. Indeed, prior studies have exemplified that certain TSGs only manifest a phenotype in soft agar cultures or *in vivo* conditions.[35] For instance, STK11 is altered in approximately 14% of lung adenocarcinoma[45–48]; however, its genetic KO or re-expression does not significantly impact cell proliferation in *in vitro* 2D culturing conditions, despite its well-established role in promoting tumorigenesis *in vivo*.[49] It was also expected that metabolic process would be among the most enriched GSEA categories for *in vivo*-specific effects (Figure 3C), aligning with known differences between nutrient/oxygen-limited tumor environments and engineered media conditions.

Notably, we identified 173 genes with divergent *in vivo* vs. *in vitro* phenotypes, including 148 genes whose loss selectively enhanced tumor growth *in vivo*. These genes were enriched for epigenetic regulators, particularly COMPASS and SWI/SNF complex members. Despite being paralogous and often considered functionally redundant, KMT2C and KMT2D displayed opposing *in vivo* roles in the A549 screen: KMT2C loss promoted tumor growth, while KMT2D KO impaired it. Pooled single-gene KO validation consistently confirmed the growth-enhancing effect of KMT2C loss, while KMT2D KO showed variable effects, potentially reflecting low frameshift efficiency in the pooled KMT2D KO population (Figure S7E) or competitive dropout of KMT2D-deficient clones in the screening context. Single-cell transcriptomic profiling revealed reciprocal transcriptional programs including loss of cell-cell interaction pathways (i.e., fibrinogen family genes) upon KMT2C KO (Figures 6B–6D), consistent with prior reports linking fibrinogen alpha chain (FGA) loss to increased cell migration and invasion in A549 and other models.[50,51]

Looking ahead, we envision CRISPR-StAR as a foundation for building a complementary *in vivo* dependency map that augments *in vitro* resources such as DepMap and Achilles by capturing vulnerabilities that emerge only within tumor microenvironments. By integrating temporal control of perturbation, internal matched-pair normalization, and clonal resolution, CRISPR-StAR overcomes key limitations of traditional *in vivo* screening approaches and enables reproducible, large-scale functional genomics screens. When combined with UMIBB, this Bayesian framework not only facilitates meta-analysis across independent experiments, enhancing statistical power, but also provides a practical platform that can be incorporated into future cancer target discovery pipelines.

Limitations of the study

Although essential gene depletion was robust in our *in vivo* screen, we observed more *in vivo*-specific enrichers than depletors in the A549 model. This likely reflects a variety of factors. First, *in vivo* selection may preferentially amplify growth-promoting clones, while depleted ("dying") clones can persist as residual DNA and continue contributing to sequencing reads at endpoint; in contrast, routine passaging *in vitro* accelerates physical loss of unfit cells and reduces their contribution to the measured pool. Second, our library represents ~1/3 of coding genes and is biased toward the druggable genome, potentially under-sampling *in vivo*-essential dependencies that fall outside this target space. Third, while cell line-derived xenograft (CDX) models capture important microenvironmental features absent from standard culture, they do not fully recapitulate human tumor complexity; incorporating additional microenvironmental components (e.g., fibroblast co-engraftment) or adapting CRISPR-StAR to immunocompetent settings may reveal additional classes of *in vivo* dependencies and depletion phenotypes.

In addition, implementing CRISPR-StAR requires careful cell engineering and quality control. We found that combined nuclease activities from CreERT2 activation and Cas9 editing can reduce cell viability *in vitro*, motivating single-cell cloning to select derivatives that tolerate transient CreERT2 activation while maintaining high Cas9 activity. Although we optimized tamoxifen dosing regimens to increase recombination without compromising tolerability, recombination efficiencies remained below ~60% across models. Notably, unrecombined vectors are excluded by our NGS PCR amplification strategy and therefore do not contribute to endpoint sequencing reads.

Cell Reports Methods
Article

CellPress
OPEN ACCESS

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the corresponding author, Dr. Teng Teng (tteng@tangotx.com).

Materials availability

Reagent generated in this study will be made available on request, but we may require a license/payment and/or a completed materials transfer agreement if there is potential for commercial application.

Data and code availability

- The primary and processed sequencing data of the CRISPR-StAR screen are deposited to the Gene Expression Omnibus database: GSE328830 and the Zenodo repository: https://doi.org/10.5281/zenodo.19323460). All other data reported in this paper will be shared by the lead contact upon request.
- The CSTAR-UMIBB source code is packaged into an R-library and available at Github: https://github.com/tangotx/CRISPRstar and has been archived at Zenodo: https://doi.org/10.5281/zenodo.19323460
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This study is funded by Tango Therapeutics. The graphical abstract was created in BioRender. Teng, T. (2026) https://BioRender.com/sfhia0y.

AUTHOR CONTRIBUTIONS

Conceptualization, S.F., Y.Y., A.H., D.S., U.E., X.P., J.N.A., and T.T.; data curation, S.F., Y.Y., J.T., L.G., A.B., S.R.M., A.H.C., H.-J.W., A.D., S.L., B.S., T.K., H.S., and E.C.H.U.; formal analysis, S.F., Y.Y., L.G., S.R.M., A.H.C., H.-J.W., T.K., and E.C.H.U.; investigation, S.F., Y.Y., J.T., L.G., A.B., S.R.M., A.H.C., H.-J.W., A.D., S.L., B.S., T.K., H.S., and E.C.H.U.; methodology, S.F. and Y.Y.; software, Y.Y., A.B., S.R.M., A.H.C., and T.K.; supervision, S.F., Y.Y., M.Z., B.B.H., E.W., A.H., X.P., J.N.A., and T.T.; validation, S.F., Y.Y., J.T., L.G., A.B., S.R.M., A.H.C., H.-J.W., A.D., S.L., B.S., T.K., H.S., and E.C.H.U.; visualization, S.F., Y.Y., L.G., A.B., S.R.M., J.N.A., and T.T.; writing – original draft, S.F., Y.Y., J.N.A., and T.T.; writing – review and editing, S.F., Y.Y., D.S., U.E., X.P., J.N.A., and T.T.

DECLARATION OF INTERESTS

All authors were employees, shareholders, and/or consultants of Tango Therapeutics at the time of their contributions to this body of work. D.S., U.E., and X.P. are also co-founders and shareholders of ViVerita Therapeutics.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used Microsoft CoPilot to check grammar and spelling and enhance sentence structure. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2026.101470.

Received: September 15, 2025
Revised: March 2, 2026
Accepted: May 1, 2026

REFERENCES

1. Arafeh, R., Shibue, T., Dempster, J.M., Hahn, W.C., and Vazquez, F. (2025). The present and future of the Cancer Dependency Map. Nat. Rev. Cancer *25*, 59–73. https://doi.org/10.1038/s41568-024-00763-x.
2. Huang, A., Garraway, L.A., Ashworth, A., and Weber, B. (2020). Synthetic lethality as an engine for cancer drug target discovery. Nat. Rev. Drug Discov. *19*, 23–38. https://doi.org/10.1038/s41573-019-0046-z.
3. Mavrakis, K.J., McDonald, E.R., 3rd, Schlabach, M.R., Billy, E., Hoffman, G.R., deWeck, A., Ruddy, D.A., Venkatesan, K., Yu, J., McAllister, G., et al. (2016). Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. Science *351*, 1208–1213. https://doi.org/10.1126/science.aad5944.
4. Marjon, K., Cameron, M.J., Quang, P., Clasquin, M.F., Mandley, E., Kunii, K., McVay, M., Choe, S., Kernytsky, A., Gross, S., et al. (2016). MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. Cell Rep. *15*, 574–587. https://doi.org/10.1016/j.celrep.2016.03.043.
5. Kryukov, G.V., Wilson, F.H., Ruth, J.R., Paulk, J., Tsherniak, A., Marlow, S.E., Vazquez, F., Weir, B.A., Fitzgerald, M.E., Tanaka, M., et al. (2016). MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. Science *351*, 1214–1218. https://doi.org/10.1126/science.aad5214.
6. Radzimierski, A., Bobowska, A., Stachowicz, A., Kuś, K., Kozłowska-Tomczyk, K., Ludwig-Słomczyńska, A., Podkalicka-Gołda, P., Gołas, A., Żukowska, M., Lebed, P., et al. (2024). Abstract 4598: Discovery of novel MTA-cooperative PRMT5 inhibitors as targeted therapeutics for MTAP-deleted cancers. Cancer Res. *84*, 4598. https://doi.org/10.1158/1538-7445.am2024-4598.
7. Engstrom, L.D., Aranda, R., Waters, L., Moya, K., Bowcut, V., Vegar, L., Trinh, D., Hebbert, A., Smith, C.R., Kulyk, S., et al. (2023). MRTX1719 is an MTA-cooperative PRMT5 inhibitor that exhibits synthetic lethality in preclinical models and patients with MTAP deleted cancer. Cancer Discov. *13*, 2412–2431. https://doi.org/10.1158/2159-8290.cd-23-0669.
8. Cottrell, K.M., Briggs, K.J., Tsai, A., Tonini, M.R., Whittington, D.A., Gong, S., Liang, C., McCarren, P., Zhang, M., Zhang, W., et al. (2025). Discovery of TNG462: A Highly Potent and Selective MTA-Cooperative PRMT5 Inhibitor to Target Cancers with MTAP Deletion. J. Med. Chem. *68*, 5097–5119. https://doi.org/10.1021/acs.jmedchem.4c03067.

9. van der Lelij, P., Lieb, S., Jude, J., Wutz, G., Santos, C.P., Falkenberg, K., Schlattl, A., Ban, J., Schwentner, R., Hoffmann, T., et al. (2017). Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. eLife *6*, e26980. https://doi.org/10.7554/elife.26980.

10. Lelij, P. van der, Newman, J.A., Lieb, S., Jude, J., Katis, V., Hoffmann, T., Hinterndorfer, M., Bader, G., Kraut, N., Pearson, M.A., et al. (2020). STAG1 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. Life Science Alliance *3*, e202000725. https://doi.org/10.26508/lsa.202000725.

11. Chen, C.-J., Sgritta, M., Mays, J., Zhou, H., Lucero, R., Park, J., Wang, I.-C., Park, J.H., Kaipparettu, B.A., Stoica, L., et al. (2019). Therapeutic inhibition of mTORC2 rescues the behavioral and neurophysiological abnormalities associated with Pten-deficiency. Nat. Med. *25*, 1684–1690. https://doi.org/10.1038/s41591-019-0608-y.

12. McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell *170*, 577–592.e10. https://doi.org/10.1016/j.cell.2017.07.005.

13. Kuhn, M., Santinha, A.J., and Platt, R.J. (2021). Moving from in vitro to in vivo CRISPR screens. Gene Genome *2*, 100008. https://doi.org/10.1016/j.ggedit.2021.100008.

14. Uijttewaal, E.C.H., Lee, J., Sell, A.C., Botay, N., Vainorius, G., Novatchkova, M., Baar, J., Yang, J., Potzler, T., van der Leij, S., et al. (2025). CRISPR-StAR enables high-resolution genetic screening in complex in vivo models. Nat. Biotechnol. *43*, 1848–1860. https://doi.org/10.1038/s41587-024-02512-9.

15. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

16. Huch, M., Knoblich, J.A., Lutolf, M.P., and Martinez-Arias, A. (2017). The hope and the hype of organoid research. Development *144*, 938–941. https://doi.org/10.1242/dev.150201.

17. Miller, T.E., Liau, B.B., Wallace, L.C., Morton, A.R., Xie, Q., Dixit, D., Factor, D.C., Kim, L.J.Y., Morrow, J.J., Wu, Q., et al. (2017). Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. Nature *547*, 355–359. https://doi.org/10.1038/nature23000.

18. Nolan-Stevaux, O., Tedesco, D., Ragan, S., Makhanov, M., Chenchik, A., Ruefli-Brasse, A., Quon, K., and Kassner, P.D. (2013). Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment. PLoS One *8*, e67316. https://doi.org/10.1371/journal.pone.0067316.

19. Esk, C., Lindenhofer, D., Haendeler, S., Wester, R.A., Pflug, F., Schroeder, B., Bagley, J.A., Elling, U., Zuber, J., von Haeseler, A., et al. (2020). A human tissue screen identifies a regulator of ER secretion as a brain-size determinant. Science *370*, 935–941. https://doi.org/10.1126/science.abb5390.

20. Michlits, G., Hubmann, M., Wu, S.-H., Vainorius, G., Budusan, E., Zhuk, S., Burkard, T.R., Novatchkova, M., Aichinger, M., Lu, Y., et al. (2017). CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. Nat. Methods *14*, 1191–1197. https://doi.org/10.1038/nmeth.4466.

21. Schmierer, B., Botla, S.K., Zhang, J., Turunen, M., Kivioja, T., and Taipale, J. (2017). CRISPR/Cas9 screening using unique molecular identifiers. Mol. Syst. Biol. *13*, 945. https://doi.org/10.15252/msb.20177834.

22. Rogers, Z.N., McFarland, C.D., Winters, I.P., Naranjo, S., Chuang, C.-H., Petrov, D., and Winslow, M.M. (2017). A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nat. Methods *14*, 737–742. https://doi.org/10.1038/nmeth.4297.

23. Eirew, P., O'Flanagan, C., Ting, J., Salehi, S., Brimhall, J., Wang, B., Biele, J., Algara, T., Lee, S.R., Hoang, C., et al. (2022). Accurate determination of CRISPR-mediated gene fitness in transplantable tumours. Nat. Commun. *13*, 4534. https://doi.org/10.1038/s41467-022-31830-2.

24. 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. https://doi.org/10.1186/s13059-014-0554-4.

25. Yu, J., Silva, J., and Califano, A. (2015). ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. Bioinformatics *32*, 260–267. https://doi.org/10.1093/bioinformatics/btv556.

26. Hart, T., and Moffat, J. (2016). BAGEL: a computational framework for identifying essential genes from pooled library screens. BMC Bioinform. *17*, 164. https://doi.org/10.1186/s12859-016-1015-8.

27. Allen, F., Behan, F., Khodak, A., Iorio, F., Yusa, K., Garnett, M., and Parts, L. (2019). JACKS: joint analysis of CRISPR/Cas9 knockout screens. Genome Res. *29*, 464–471. https://doi.org/10.1101/gr.238923.118.

28. Kebschull, J.M., and Zador, A.M. (2018). Cellular barcoding: lineage tracing, screening and beyond. Nat. Methods *15*, 871–879. https://doi.org/10.1038/s41592-018-0185-x.

29. Abid, T., Goodale, A.B., Kalani, Z., Wyatt, M., Gonzalez, E.M., Zhou, K.N., Qian, K., Novikov, D., Condurat, A.-L., Bandopadhayay, P., et al. (2023). Genome-wide pooled CRISPR screening in neurospheres. Nat. Protoc. *18*, 2014–2031. https://doi.org/10.1038/s41596-023-00835-6.

30. Hansen, S.L., Larsen, H.L., Pikkupeura, L.M., Maciag, G., Guiu, J., Müller, I., Clement, D.L., Mueller, C., Johansen, J.V., Helin, K., et al. (2023). An organoid-based CRISPR-Cas9 screen for regulators of intestinal epithelial maturation and cell fate. Sci. Adv. *9*, eadg4055. https://doi.org/10.1126/sciadv.adg4055.

31. Mukhare, R., Gandhi, K.A., Kadam, A., Raja, A., Singh, A., Madhav, M., Chaubal, R., Pandey, S., and Gupta, S. (2025). Integration of Organoids With CRISPR Screens: A Narrative Review. Biol. Cell *117*, e70006. https://doi.org/10.1111/boc.70006.

32. Lenoir, W.F., Morgado, M., DeWeirdt, P.C., McLaughlin, M., Griffith, A.L., Sangree, A.K., Feeley, M.N., Esmaeili Anvar, N., Kim, E., Bertolet, L.L., et al. (2021). Discovery of putative tumor suppressors from CRISPR screens reveals rewired lipid metabolism in acute myeloid leukemia cells. Nat. Commun. *12*, 6506. https://doi.org/10.1038/s41467-021-26867-8.

33. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

34. Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., Hill, D.P., et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.

35. Martin, T.D., Patel, R.S., Cook, D.R., Choi, M.Y., Patil, A., Liang, A.C., Li, M.Z., Haigis, K.M., and Elledge, S.J. (2021). The adaptive immune system is a major driver of selection for tumor suppressor gene inactivation. Science *373*, 1327–1335. https://doi.org/10.1126/science.abg5784.

36. Zhu, X.G., Chudnovskiy, A., Baudrier, L., Prizer, B., Liu, Y., Ostendorf, B.N., Yamaguchi, N., Arab, A., Tavora, B., Timson, R., et al. (2021). Functional Genomics In Vivo Reveal Metabolic Dependencies of Pancreatic Cancer Cells. Cell Metab. *33*, 211–221.e6. https://doi.org/10.1016/j.cmet.2020.10.017.

37. Cenik, B.K., and Shilatifard, A. (2021). COMPASS and SWI/SNF complexes in development and disease. Nat. Rev. Genet. *22*, 38–58. https://doi.org/10.1038/s41576-020-0278-0.

38. Langille, E., Al-Zahrani, K.N., Ma, Z., Liang, M., Uuskula-Reimand, L., Espin, R., Teng, K., Malik, A., Bergholtz, H., Ghamrasni, S.E., et al. (2022). Loss of Epigenetic Regulation Disrupts Lineage Integrity, Induces Aberrant Alveogenesis, and Promotes Breast Cancer. Cancer Discov. *12*, 2930–2953. https://doi.org/10.1158/2159-8290.cd-21-0865.

39. Alam, H., Tang, M., Maitituoheti, M., Dhar, S.S., Kumar, M., Han, C.Y., Ambati, C.R., Amin, S.B., Gu, B., Chen, T.-Y., et al. (2020). KMT2D Deficiency

Impairs Super-Enhancers to Confer a Glycolytic Vulnerability in Lung Cancer. Cancer Cell *37*, 599–617.e7. https://doi.org/10.1016/j.ccell.2020.03.005.

40. Wang, G., Chow, R.D., Zhu, L., Bai, Z., Ye, L., Zhang, F., Renauer, P.A., Dong, M.B., Dai, X., Zhang, X., et al. (2020). CRISPR-GEMM Pooled Mutagenic Screening Identifies KMT2D as a Major Modulator of Immune Checkpoint Blockade. Cancer Discov. *10*, 1912–1933. https://doi.org/10.1158/2159-8290.cd-19-1448.
41. Chow, R.D., Guzman, C.D., Wang, G., Schmidt, F., Youngblood, M.W., Ye, L., Errami, Y., Dong, M.B., Martinez, M.A., Zhang, S., et al. (2017). AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. Nat. Neurosci. *20*, 1329–1341. https://doi.org/10.1038/nn.4620.
42. Zhang, Z., Christin, J.R., Wang, C., Ge, K., Oktay, M.H., and Guo, W. (2016). Mammary-Stem-Cell-Based Somatic Mouse Models Reveal Breast Cancer Drivers Causing Cell Fate Dysregulation. Cell Rep. *16*, 3146–3156. https://doi.org/10.1016/j.celrep.2016.08.048.
43. Chen, C., Liu, Y., Rappaport, A.R., Kitzing, T., Schultz, N., Zhao, Z., Shroff, A.S., Dickins, R.A., Vakoc, C.R., Bradner, J.E., et al. (2014). MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. Cancer Cell *25*, 652–665. https://doi.org/10.1016/j.ccr.2014.03.016.
44. Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. *33*, 390–394. https://doi.org/10.1038/nbt.3155.
45. Mazor, T., de Bruijn, I., AlHamad, R., Chennault, C., Dubin, C., Easton-Marks, J., Fu, Z., Gross, B., Haynes, C., Higgins, D.M., et al. (2024). Abstract 1249: cBioPortal for Cancer Genomics. Cancer Res. *84*, 1249. https://doi.org/10.1158/1538-7445.am2024-1249.
46. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. *2*, 401–404. https://doi.org/10.1158/2159-8290.cd-12-0095.
47. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.
48. de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., et al. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res. *83*, 3861–3867. https://doi.org/10.1158/0008-5472.can-23-0816.
49. Hollstein, P.E., Eichner, L.J., Brun, S.N., Kamireddy, A., Svensson, R.U., Vera, L.I., Ross, D.S., Rymoff, T.J., Hutchins, A., Galvez, H.M., et al. (2019). The AMPK-Related Kinases SIK1 and SIK3 Mediate Key Tumor-Suppressive Effects of LKB1 in NSCLC. Cancer Discov. *9*, 1606–1627. https://doi.org/10.1158/2159-8290.cd-18-1261.
50. Wang, M., Zhang, G., Zhang, Y., Cui, X., Wang, S., Gao, S., Wang, Y., Liu, Y., Bae, J.H., Yang, W.-H., et al. (2020). Fibrinogen Alpha Chain Knockout Promotes Tumor Growth and Metastasis through Integrin–AKT Signaling Pathway in Lung Cancer. Mol. Cancer Res. *18*, 943–954. https://doi.org/10.1158/1541-7786.mcr-19-1033.
51. Han, X., Liu, Z., Cui, M., Lin, J., Li, Y., Qin, H., Sheng, J., and Zhang, X. (2024). FGA influences invasion and metastasis of hepatocellular carcinoma through the PI3K/AKT pathway. Aging (Albany NY) *16*, 12806–12819. https://doi.org/10.18632/aging.206011.
52. Chylinski, K., Hubmann, M., Hanna, R.E., Yanchus, C., Michlits, G., Uijttewaal, E.C.H., Doench, J., Schramek, D., and Elling, U. (2019). CRISPR-Switch regulates sgRNA activity by Cre recombination for sequential editing of two loci. Nat. Commun. *10*, 5454. https://doi.org/10.1038/s41467-019-13403-y.
53. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
54. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278. https://doi.org/10.1186/s13059-015-0844-5.
55. Li, H., Verma, Y., Hockemeyer, D., and Soldner, F. (2022). CRISPResso analysis v1. https://doi.org/10.17504/protocols.io.b4nxqvfn.
56. Simoneau, A., Engel, J.L., Bandi, M., Lazarides, K., Liu, S., Meier, S.R., Choi, A.H., Zhang, H., Shen, B., Martires, L., et al. (2023). Ubiquitinated PCNA Drives USP1 Synthetic Lethality in Cancer. Mol. Cancer Ther. *22*, 215–226. https://doi.org/10.1158/1535-7163.mct-22-0409.
57. Princeton University Press (1952). Vol. I: The American Soldier : Adjustment during army life by S.A. Stouffer, E.A. Suchman, L.C. de Vinney, S.A. Star, R.M. Williamsjr. , 1949, XII p. 600 p., $ 7.50. - Vol. II: The American Soldier : Combat and its aftermath by S.A. Stouffer, A.A. Lumsdaine, M.H. Lurnsdaine, R.M. Williamsjr , M.B. Smith, I.L. Janis, S.A. Star, L.S. Cottrelljr. , 1949, 675 p., $ 7.50. - Vol. III: Experiments on Mass Communication by C.I. Hovland, A.A. Lumsdaine, F.D. Sheffield, 1949, Xp. 345 p., $ 5.00. - Vol. IV: Measurement and Prediction by S.A. Stouffer, I. Guttman, E.A. Suchman, P.F. Lazarfeld, S.A. Star, J.A. Clausen. 1950, X p. 756 p., $ 10.00. Princeton N. J. Bull. l'Inst. Rech. économiques Soc. *18*, 744–745. https://doi.org/10.1017/s0770451800047771.

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PTEN Rabbit Monoclonal antibody	Cell Signaling Technology	Cat# 9559; RRID:AB_390810
Vinculin Rabbit Monoclonal antibody	Cell Signaling Technology	Cat# 13901; RRID:AB_2728768
Actin Rabbit Monoclonal antibody	Cell Signaling Technology	Cat# 4970; RRID:AB_2223172
Guide-it™ Cas9 Monoclonal Antibody (Clone TG8C1)	Takara Bio	Cat# 632628
Bacterial and virus strains		
Endura ElectroCompetent cells	Bioresearch Technologies	Cat#60242-2
Chemicals, peptides, and recombinant proteins		
T4 DNA ligase	New England Biolabs	Cat# M0202L
DMEM	Thermo Fisher Scientific	Cat#11965118
Lipofectamine 3000	Thermo Fisher Scientific	Cat# L30000015
Opti-MEM	Gibco	Cat# 31985-062
Polybrene	Sigma-Aldrich	Cat# TR-1003-G
Matrigel	Corning	Cat# 354263
Tamoxifen	Sigma Aldrich	Cat# T5648
Esp3I	NEB	Cat# R0734L
Fetal Bovine Serum	Gemini Bio Products	Cat# 100-500
Blasticidin	Thermo Fisher Scientific	Cat# A1113903
Puromycin	Thermo Fisher Scientific	Cat# A1113803
Neomycin	Thermo Fisher Scientific	Cat# 10131027
Hygromycin B	Thermo Fisher Scientific	Cat# 10687010
Corn oil	Sigma Aldrich	Cat# C8267
Critical commercial assays		
Lonza MycoAlert Detection Kit	Lonza	Cat# LT07-318
Chromium Next GEM Single-cell 3' Reagent Kits v3.1, 4rxns	10x Genomics	Cat# 1000269
Chromium Next GEM Chip G Single Cell Kit, 48 rxns	10x Genomics	Cat# 1000120
Chromium Nuclei Isolation Kit WITH RNase inhibitor, 16 rxn (3′/5′, Multiome & Fixed RNA Assay Compatible)	10x Genomics	Cat# 1000494
Chromium NextGem Single Cell Multiome ATAC + Gene Expression Reagent Bundle, 4 rxn	10x Genomics	Cat# 1000285
Dual Index Kit TT, Set A (96 rxn)	10x Genomics	Cat# 1000215
NextSeq™ 2000 P2 XLEAP-SBS™ Reagent Kit (100 Cycles)	Illumina	Cat# 20100987
NextSeq™ 2000 P2 XLEAP-SBS™ Reagent Kit (200 Cycles)	Illumina	Cat# 20100986
Deposited data		
CRISPR-StAR screen primary NGS data	GEO	GSE328830
CRISPR-StAR screen processed NGS data	Zenodo	https://doi.org/10.5281/zenodo.19323460
Experimental models: Cell lines		
A549	ATCC	Cat# CCL-185; RRID:CVCL_0023
NCI H460	ATCC	Cat# HTB-177; RRID:CVCL_0459
SW1573	ATCC	Cat# CRL-2170; RRID:CVCL_1720

(*Continued on next page*)

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lenti-X 293T	Takara Bio	Cat# 632180
Experimental models: Organisms/strains		
Mouse: NOD-Prkdcem26Cd52Il2rg/NjuCrl	Charles River Laboratorie	RRID:IMSR_CRL:572
Oligonucleotides		
ITC gRNA	AGATCTTACCTCCTTAGACA	N/A
PTEN gRNA	GGTTTGATAAGTTCTAGCTG	N/A
KMT2C gRNA	AGACCGTGCCAGATACACAT	N/A
KMT2D gRNA	TGAGTGGCAGCTCCATAAGG	N/A
Recombinant DNA		
CRISPR-StAR vector	Uijttewaal, E.C.H. et al.[14]	N/A
CreERT2 construct	Uijttewaal, E.C.H. et al.[14]	N/A
Lenti Virus Packaging Mix	Cellecta	Cat# CPCP-K2A
CRISPR-SWITCH	Chylinski, K et al.[52]	N/A
Software and algorithms		
MAGeCK v 0.5.9.2	Li, W et al.[24]	RRID:SCR_025016
Cellranger v6.1.2	10x Genomics	RRID:SCR_023221
Seurat v4.3	Hao, Y et al.[53]	RRID:SCR_016341
MAST	Finak, G et al.[54]	RRID:SCR_016340
CRISPResso	Li, H et al.[55]	RRID:SCR_021538
GraphPad PRISM	GraphPad	RRID:SCR_002798
Spotfire	Tibco	RRID:SCR_008858
UMIBB	This paper	https://github.com/tangotx/CRISPRstar, https://doi.org/10.5281/zenodo.19323460
Other		
245 mm2 bioassay agar dishes	Teknova	Cat# L6010
2-layer CellSTACK	Corning	Cat# 3313
NextSeq 2000	Illumina	N/A
Chromium X	10X Genomics	N/A
DNA Clean & Concentrator-5 spin column	Zymo Researcg	Cat# DNA Clean & Concentrator-5

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

A549 (Cat# CCL-185; RRID:CVCL_0023), NCI H460 (Cat# HTB-177; RRID:CVCL_0459) and SW1573 (Cat# CRL-2170; RRID:CVCL_1720) cell lines were purchased from ATCC (Manassas, VA). Lenti-X cells (Cat# 632180) was purchased from Takara Bio.

Mouse models

6-8 weeks old female NCG mice (strain NOD-Prkdcem26Cd52Il2rg/NjuCrl; RRID:IMSR_CRL:572) were purchased from Charles River Laboratory (strain 572). All animal protocols were approved by the Charles River Institutional Animal Care & Use Committee (IACUC).

METHOD DETAILS

Cloning of sgRNA and cDNA constructs

All sgRNA and cDNA constructs are built with the 3rd generation lentiviral system. CRISPR-StAR vectors were modified based on previously described constructs[14] to aid cell line engineering and NGS workflows, with the following changes: (1) a puromycin selection cassette was used instead of the original neomycin cassette, (2) an i7 sequencing adapter was added downstream of the UMI barcode for PCR and Illumina sequencing, and (3) PacI restriction sites were added flanking the CRISPR-StAR component to enable enzymatic digestion and enrichment of the target sequences during NGS library preparation. CreERT2 constructs were used as previously described.[14] Individual sgRNAs oligonucleotides were synthesized by Integrated DNA Technologies (IDT,

Newark, NJ) and cloned into destination vectors using T4 DNA ligase (New England Biolabs, M0202L). All plasmid constructs were verified by Sanger sequencing at Genewiz (Boston, MA).

Cell culture and cell line engineering

All cells were cultured in DMEM (Thermo Fisher Scientific, #11965118) supplemented with 10% FBS and maintained at 37°C under 5% CO_2. Cell line authentication was performed at LabCorp (Burlington, NC) and routinely tested for mycoplasma contamination using Lonza MycoAlert Detection Kit (Lonza, LT07-318).

To produce lentivirus, transfection mix containing the respective vector, a lentiviral packaging mix (Cellecta, CPCP-K2A), Lipofectamine 3000 (Thermo Fisher Scientific, L30000015) were prepared in Opti-MEM (Gibco, 31985-062) to transfect the Lenti-X cells (Takara Bio, Cat# 632180). Fresh DMEM +30% FBS were used to replace the transfection mix after 16 h. Viral supernatant was collected 48 h post-transfection, filtered through a 0.45 μm membrane, aliquoted, and stored at −80°C.

For infection, lentivirus was added to target cells with 8 μg/mL polybrene (Sigma-Aldrich, TR-1003-G). Media was refreshed overnight. Antibiotic selection was initiated 48 h post-infection according to manufacturer's recommended doses.

To generate CRISPR-StAR clones, Cas9 expressing cells were engineered with CreERT2 construct[14] and seeded by limited dilution for single-cell cloning. Established single-cell clones were screened with CRISPR-SWITCH construct as described previously.[52] The clone exhibiting the highest activity upon 4-OHT treatment with minimal leakage was selected for CRISPR-StAR screen.

Animal studies

For tumor inoculation, engineered cell lines were resuspended in Matrigel (Corning, 354263) and PBS (Sigma-Aldrich) 1:1 solution at a final concentration of 50,000 cells/μL. Each animal was injected with 200 μL cell suspension subcutaneously in the upper flank. Tumor growth and animal welfare were monitored twice a week. Tamoxifen (Sigma Aldrich, T5648) was delivered in corn oil (Sigma Aldrich, C8267. Concentration 100 mg/mL) at a dose of 75 mg/kg by intraperitoneal injection (IP), once a day for 2 consecutive days.

CRISPR screen

The 30K and 1K library oligonucleotide pools were custom synthesized from Twist Biosciences (San Francisco, CA). The oligo pools were PCR-amplified, purified and cloned into destination vectors via Golden Gate Assembly with Esp3I (NEB, R0734L). The ligated library plasmids were purified using spin columns (Zymo Research, D4014) and electroporated into Endura ElectroCompetent cells (Bioresearch Technologies, 60242-2). Following recovery, transformed cells were plated onto 245 mm^2 bioassay agar dishes (Teknova, L6010) and incubated overnight incubation at 32°C. Colonies were scraped into fresh LB medium and pelleted by centrifugation. The final plasmid library was prepared using Qiagen Maxiprep Kits. The plasmid pool was then used for transfection utilizing 2-layer CellSTACK vessels (Corning, 3313). Viral supernatants were aliquoted and stored in −80°C freezer until further use.

To perform the 30K CRISPR-StAR screen, the A549 CRISPR-StAR clone was infected at low MOI at library representation of 1000x. In the *in vitro* arm, an initial bottleneck at 50X was applied post-library transduction to simulate clonal selection pressure observed *in vivo*. This was followed by a 50-fold expansion before recombination was induced with 4-hydroxytamoxifen (4-OHT). Cells were cultured for 21 days before to endpoint sample collection. For the *in vivo* arm, the library infected A549 CRISPR-StAR cell pool was expanded without bottlenecking and injected subcutaneously into 118 mice at 10 million cells per mouse. Tamoxifen treatment was initiated once tumors reached 150mm^3. Tumors were allowed to grow for 28 days before mice were euthanized and tumors harvested.

Genomic DNA was extracted from both *in vitro* and *in vivo* samples and subjected to PCR amplification targeting regions surrounding the CRISPR-StAR components. Amplified PCR products were purified with KAPA beads and sequenced on an Illumina NextSeq 2000 platform.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantitative analyses were performed using the sample sizes and biological replicates described in the manuscript, figure legends and specific STAR Methods sections below.

CRISPR-StAR screen NGS sequence alignment and read count normalization

FASTQ files are mapped to the reference file consisting of both guide sequences and barcode sequences using custom scripts. After mismatches and low-quality reads are removed, only aligned reads are used to measure the abundance of UMIs with an active or inactive form of construct respectively. Very small clones with less than 10 reads (active and inactive combined counts) are removed from further analysis. A pseudo count of 0.5 is added when the active or inactive count is 0.[56,57]

CRISPR-StAR screen normalized A/I ratio defines clonal depletion or enrichment events per guide

In each sample, the active count and inactive count of each clone is normalized by the ratio between the sum of counts of the active non-targeting control (NTC) guides and the sum of counts of the inactive NTC guides. A/I Ratio (R_{si}) of the normalized active count (n_{Asi}) vs. inactive count (n_{Isi}) of the same clone UMI(i) of sgRNA (s) is calculated to determine if the active form guide was enriched ($R_{si}>1$) or depleted($R_{si}<1$) comparing to the inactive form of the same clone. Guide level log2(A/I ratio) was derived to show the effect

sizes by averaging of $\log_2(R_{si})$ weighted by combined counts of each clone, then normalized by the median $\log_2(R_s)$ of guides targeting essential genes.

Unique molecular identifier Bayesian Beta-binomial (UMIBB) analysis for UMI-CRISPR pooled screening data

We applied the UMIBB analysis previously reported by Simoneau A.,et.,al.[56] in the pooled screening dataset of A549 cell line. Briefly, the counts in case group vs. control group of the same UMI clone are compared to determine the number of depletion or enrichment events per sgRNA, Then the probability of depletion or enrichment for each sgRNA is modeled by a beta-binomial distribution and testing in a Bayesian framework.

UMIBB analysis for one group CRISPR-StAR samples

Let M_s be the total distinct clones of sgRNA (s), E_s is the number of enriched clones (R_{si}>1). It can be modeled as Bernoulli trials process follows beta-binomial distribution with a conjugate prior,

$$E_s \sim Bin(\theta, M_s)$$

$$\theta \sim Beta(\alpha, \beta) \quad \text{(Equation 1)}$$

where θ is the prior probability of active count greater than inactive count in a clone. The prior probability θ and parameters α, β are estimated using the observed numbers of enriched clones (E_{NTC}) across all the clones with NTC guides (M_{NTC}). The prior θ is equal or very close to 0.5, because the NTC clones are expected to have same chance of enrichment or depletion if there is no systematic bias. Markov chain Monte Carlo (MCMC) approximation is used to fit the posterior probability distribution of θ, which combined the prior and the likelihood functions using Bayes' theorem. The posterior estimate (θ_{Es}) for clones showing an enrichment of active form measures the effect size for the sgRNA. P-value ($p_{\theta Es}$) is estimated by MCMC to test the null hypothesis that θ_{Es} is less than or equal to the mean prior (θ_{NTC}). Similarly, we build beta-binomial models using the number (D_s) of depleted clones (R_{si}<1) and M_s, then derived the posterior estimate (θ_{Ds}) for clones showing a depletion of active form and associated *p*-value ($p_{\theta Ds}$). Then, the gene level depletion test statistics ($w\theta_{Dg}$,$p_{\theta Dg}$) and enrichment test statistics ($w\theta_{Eg}$,$p_{\theta Dg}$)are calculated using the weighted Stouffer's *Z* score method[57] weighted by the number of distinct clones in each sgRNA(M_s) targeting gene (g). Because most clones are either depleted or enriched for the active form, only very few have equal number of active and inactive form, the sum of $w\theta_{Eg}$ and $w\theta_{Dg}$ is equal or very close to one. So we further consolidate the statistics by using the lower *p*-value from the depletion and enrichment tests as the gene-level *p*-value. The relative frequency of clonal enrichment frequency ($w\theta_g$) is defined by Equation 2, which is plotted as the effect size of genes in the volcano-plot. Finally, the gene-level *p*-values are adjusted for the multiple testing by the False Discovery Rate (FDR) procedure.

$$p_{\theta g} = \min(p_{\theta Dg}, p_{\theta Eg})$$

$$w\theta_g = \begin{cases} 1 - w\theta_{Dg}, p_{\theta Dg} < p_{\theta Eg} \\ w\theta_{Eg}, p_{\theta Dg} \geq p_{\theta Eg} \end{cases} \quad \text{(Equation 2)}$$

UMIBB analysis comparing CRISPR-StAR samples between two experiment groups

To compare the difference of active form guide enrichment probability between two experiment groups, we use two independent beta-binomial models to fit the observed number of clones (E_{s1}, M_{s1}) and (E_{s2}, M_{s2}) in the two experimental groups in the same way as defined previously in the single group analysis.

$$E_{s1} \sim Bin(\theta_1, M_{s1})$$

$$\theta_1 \sim Beta(\alpha_1, \beta_1)$$

$$E_{s2} \sim Bin(\theta_2, M_{s2})$$

$$\theta_2 \sim Beta(\alpha_2, \beta_2) \quad \text{(Equation 3)}$$

The priors θ_1 and θ_2 are estimated based on the number of active enriched NTC clones (E_{NTC1},M_{NTC1}) and (E_{NTC2},M_{NTC2}) among NTC clones in each group separately(3). The posterior estimates (θ_{Es1},θ_{Es2}) are derived by Bayesian inference as describe before.

$$\Delta_{Es} = \theta_{Es1} - \theta_{Es2} \quad \text{(Equation 4)}$$

The Δ_{Es} is the difference between the posterior proportion estimates of the two experimental groups(4). P-value ($p_{\Delta Es}$) is estimated by MCMC to test the null hypothesis that Δ_{Es} is equal to zero.

Similarly, we fit the observed number of clones (D_{s1}, M_{s1}) and (D_{s2}, M_{s2}) for their depletion probability in the two experimental groups, and tested the null hypothesis that the difference (Δ_{Ds}) of their posterior estimates (θ_{Ds1},θ_{Ds2}) is equal to zero (5).

$$\Delta_{Ds} = \theta_{Ds2} - \theta_{Ds1} \quad \text{(Equation 5)}$$

Then, the gene level depletion test statistics of two groups ($w\Delta_{Dg}$,$p_{\Delta Dg}$) and enrichment test statistics ($w\Delta_{Eg}$,$p_{\Delta Eg}$)are calculated using the weighted Stouffer's *Z* score method weighted by the number of distinct clones in each sgRNA targeting gene (g) combined across the two groups. The gene level statistics were consolidated to use the greater significant value of ($p_{\Delta Dg}$,$p_{\Delta Eg}$) and its associated $w\Delta$ value for the gene (6). $w\Delta_g$ is the difference of posterior estimates of clonal enrichment probabilities between the two experimental groups, the *p*-values ($p_{\Delta g}$) of gene-level tests are adjusted by the FDR procedure.

$$p_{\Delta g} = \min\left(p_{\Delta Dg}, p_{\Delta Eg}\right)$$

$$w\Delta_g = \begin{cases} w\Delta_{Dg}, p_{\Delta Dg} < p_{\Delta Eg} \\ w\Delta_{Eg}, p_{\Delta Dg} \geq p_{\Delta Eg} \end{cases} \quad \text{(Equation 6)}$$

The analysis pipeline described above is incorporated into a R package CSTAR-UMIBB, in which MCMC approximation is conducted through JAGS (Just Another Gibbs Sampler) with default parameters (iterations = 15,000, chains = 3, warm-up = 1,000).

Simulation framework of barcoded CRISPR screens

To evaluate the performance of UMIBB analysis algorithms, we developed a hierarchical stochastic simulation framework to model the read counts of barcoded clones with a well-defined ground truth mimicking realistic experimental conditions. The model can simulate clonal read counts of active and inactive guides targeting genes in different categories, such as NTC, neutral genes, essential (depleted) genes, tumor suppressor (enriched) genes, and genes with different growth effects in two experimental conditions. The simulation models biological and technical variation at gene level, guide level and clonal levels. At gene level, each gene *i* is assigned a base growth effect γ_i, representing the "true" fitness change upon complete loss-of-function. In our benchmarking configuration, γ is assigned via explicit vectors to represent a spectrum of depletion ($\gamma \in [0.2,0.65]$) and enrichment ($\gamma \in [1.5,5.0]$) for active counts; for inactive counts, NTC or neutral genes γ is assigned 1. To model variability in sgRNA synthesis, on-target efficiency, and PCR amplification bias, each guide *j* is assigned a quality score (Q_j) drawn from a Gamma distribution:

$$Q_j \sim Gamma(\alpha = 10, \beta = 1)$$

To represent stochasticity due to clonal expansion and the "heavy-tailed" nature of clonal counts distributions, individual barcodes *k* within a guide are assigned weights W_k drawn from a Log-Normal distribution with $\sigma_{tail} = 2$ to mimic clonal counts distributions in real experiments.

$$W_k \sim Log - Normal(0, \sigma_{tail})$$

At the clonal level, we also modeled the stochastic DNA repair outcomes following Cas9-induced double-strand breaks. For each clone, the probability of getting a complete (loss-of-function, i.e., Frameshifts), partial (i.e., In-frame Indels) or wildtype (i.e., synonymous) outcomes are set at 0.7, 0.2 and 0.1 respectively. An editing outcome is randomly assigned by the editing outcome probability with a corresponding functional impact I_{ijk}, where $I = 0$ for complete outcomes, $I = 0.5$ for partial functional outcomes, and $I = 1$ for wildtype outcomes. The functional fitness of a specific clone ω_{ijk} is calculated as:

$$\omega_{ijk} = 1 - \left[(1 - \gamma_i) \times \left(1 - I_{ijk}\right)\right]$$

This ensures that "escapee" clones maintain higher fitness even when targeting essential genes, accurately reflecting phenotypic mosaicism. The underlying frequencies for the inactive (FI_{ijk}) and active (FA_{ijk}) cells are calculated as:

$$FI_{ijk} \propto Q_j \times W_k$$

$$FA_{ijk} \propto Q_j \times W_k \times \omega_{ijk}$$

A dropout probability (D = 0.1) is introduced to simulate the random technical noise during NGS sampling. And the imbalance of Cre induced recombination ratio (R) between active vs. inactive outcomes is modeled as:

$$R_A = \{\text{R if R} < 1;\ 1 \text{ if R} \geq 1\}$$

$$R_I = \{\text{R if R} > 1;\ 1 \text{ if R} \leq 1\}$$

Together, a mask is applied on the underlying frequencies via a Bernoulli process:

$$FI_M \sim FI \odot Bernoulli[(1\text{-}D) \times R_I]$$

$$FA_M \sim FA \odot Bernoulli[(1\text{-}D) \times R_A],$$

where active or inactive clones are randomly masked to 0 counts. Final raw counts $\boldsymbol{C_A}$,$\boldsymbol{C_I}$

Please cite this article in press as: Fenoglio et al., Temporal control of sgRNA library activation unlocks large-scale *in vivo* CRISPR screens, Cell Reports Methods (2026), https://doi.org/10.1016/j.crmeth.2026.101470

CellPress
OPEN ACCESS

are generated using Multinomial Sampling to account for the competitive nature of NGS libraries, where the total number of reads is constrained by a fixed library size N:

$$C_I \sim Multinomial(N, FI_M)$$

$$C_A \sim Multinomial(N, FA_M)$$

The variance of aggregated guide level counts in this model follows an overdispersed relationship relative to the mean μ, approximating the Negative Binomial distribution common in NGS data:

$$Var(C) \approx \mu + \phi\mu^2$$

In this study, we applied the above framework to simulate clonal counts for various types of genes defined by underlying growth effect (γ): including NTCs or neutral genes (γ = 1), essential genes ($\gamma \in$[0.2,0.65]), tumor suppressor genes($\gamma \in$[1.5,5.0]), and differential growth genes with different growth effect in experimental arms A and B ($r_A = 1 and r_B \in$[0.2,0.65];or $r_A = 1 and r_B \in$[1.5,5.0]). We benchmarked the performance of UMIBB in various scenarios set by the sequencing depth (library size N from 100K to 4M reads) or underlying recombination ratio (R ∈[0.25,4.0]).

Guide level MAGeCK analysis

Clonal active counts and inactive counts are aggregated to generate the guide level active counts and inactive counts, which are used as input for MAGeCK (v 0.5.9.2) MLE analysis (RRID:SCR_025016).

Animal sequencing counts data down-sampling analysis

Down-sampling of the original full experimental dataset was performed by the sample() function without replacement in R. We randomly took 10 subsets of down-sampling samples for each sample size (n animals) group and calculated the statistics (log2(A/I ratio), $w\theta_g$) for each subsets individually. Pearson correlation coefficients (R) are calculated against the statistics of the original full experimental dataset.

Single-cell RNAseq

Tumors were collected from either control A549 or KMT2C or KMT2D KO xenograft models and subjected to 10x genomic protocol for tumor dissociation for single-cell RNAs sequencing (protocol#CG000147). Dead cells were removed from single-cell suspensions following 10x genomic instruction (protocol#CG000093). Final prepared single-cell suspension was then processed with Chromium Next GEM Single-cell 3' Reagent Kits v3.1 following the manufacturer's protocol (#CG00031) and subjected to Illumina sequencing on NextSeq 2000.

Single-cell RNAseq analysis

The fastq files from 10x Chromium were demultiplexed and mapped with Cellranger v6.1.2 (RRID:SCR_023221) using a GRCh38 reference provided by 10x Genomics. The feature barcode matrix was then loaded into R for quantification and normalization by the Seurat v4.3 (RRID:SCR_016341) package.[53] Cells classified as negatives or doublets were removed. Cells with more than 2500 genes or fewer than 200 genes were removed, along with cells with a mitochondrial expression fraction above 0.05. Cell cycle effects were calculated using *CellCycleScoring* and regressed out with *ScaleData*. Differentially expressed genes between KMT2C and intron cutting control samples were identified using the MAST package,[54] with the top ones visualized using *DoHeatmap*. Gene expression values from the KMT2D samples were added to the heatmap. We performed graph based Louvain clustering with *FindNeighbors* and *FindClusters* on dimensionality reduced data from *RunPCA* and *RunUMAP*, visualizing 12 Louvain-identified clusters. Percent cell expressions were determined using the *DotPlot* function.