Title: GO-CRISPR: a highly controlled workflow to discover gene essentiality in loss-of-function screens

Authors:

Pirunthan Perampalam^{1,2}, James I. McDonald^{1,3}, Frederick A. Dick^{1,3,4*}

Affiliations:

1. London Regional Cancer Program, London Health Sciences Centre, London, ON, Canada, N6A 5W9
2. Department of Biochemistry, Western University, London, ON, Canada, N6A 5C1
3. Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada, N6A 5C1
4. Children's Health Research Institute, London, ON, Canada, N6A 4V2

*Correspondence: fdick@uwo.ca

Keywords:

CRISPR/Cas9, CRISPR screen, genome-wide CRISPR screen, dropout screen, loss-of-function, CRISPR screen analysis, data visualization, gene essentiality

SUMMARY

Genome-wide CRISPR screens are an effective discovery tool for genes that underlie diverse cellular mechanisms that can be scored through cell fitness. Loss-of-function screens are particularly challenging compared to gain-of-function because of the limited dynamic range of decreased sgRNA sequence detection. Here we describe **Guide-Only control CRISPR (GO-CRISPR)**, an improved loss-of-function screening workflow, and its companion software package, **Toolset for the Ranked Analysis of GO-CRISPR Screens (TRACS)**. We demonstrate a typical GO-CRISPR workflow in a non-proliferative 3D spheroid model of dormant high grade serous ovarian cancer and demonstrate superior performance to standard screening methods. The unique integration of the pooled sgRNA library quality and guide-only controls allows TRACS to identify novel molecular pathways that were previously unidentified in tumor dormancy. Together, GO-CRISPR and TRACS can robustly improve the discovery of essential genes in challenging biological scenarios.

INTRODUCTION

Gene editing using CRISPR/Cas9 technology has seen widespread adoption across most biomedical disciplines, including cancer research (Lytle et al., 2019; Shalem et al., 2015). In particular, the ability to multiplex CRISPR gene knockouts on a genome-wide scale has stimulated systematic interrogation of cell biology (Wang et al., 2014). Pooled single guide RNA (sgRNA) libraries are used to create single-gene knockouts in individual cells and selective pressure is applied through culture conditions or drug treatment. Genetic deficiencies that produce resistance or susceptibility are quantitated using sgRNA coding sequences as barcodes to compare gene knockout abundance between the start and end of the experiment (Shalem et al., 2014). CRISPR screens can therefore discover functional roles for genes and pathways not suggested by more traditional hypothesis-driven research.

Gain-of-function genome-wide CRISPR screens can lead to several orders of magnitude change in sgRNA sequence abundance because of resistant cell proliferation, unequivocally identifying resistance genes (Cai et al., 2020; Parnas et al., 2015; Sanson et al., 2018). Conversely, loss-of-function is more challenging to quantitate because complete disappearance of sgRNA sequences for a gene may represent technical failure of the screen design, or its execution (Koike-Yusa et al., 2014). In addition, knockout of an individual gene in the chosen culture condition may not cause lethality with complete penetrance (Thyme et al., 2016). Ultimately, identification of essential genes in loss-of-function screens has relied on prolonged periods of cell proliferation to separate the abundance of bystander sgRNA abundance from true deleterious changes (Shalem et al., 2014). For this reason, CRISPR screens have generally utilized rapidly proliferating 2D cell culture conditions. Scenarios such as the tumor microenvironment, metastasis and tumor dormancy, are better assessed in 3D culture models such as multicellular tumor spheroids or

organoids (Fujii et al., 2016; Jacob et al., 2020; Kenny et al., 2015; Vlachogiannis et al., 2018). However, the inability of organoids to quantitatively regenerate from individual cells upon subculture has prevented robust library representation (Ringel et al., 2020), and in some cases this has been compensated by screening more compact, partial genome libraries (Planas-Paz et al., 2019). Furthermore, most 3D spheroids exhibit slower growth kinetics due to hypoxia and necrosis which can further hamper detection of gene loss events (Zanoni et al., 2016). All of these factors likely contribute to stochastic loss of guides which can confound loss-of-function studies since current methods cannot distinguish these Cas9-independent events from bona fide loss-of-function due to gene editing. For these reasons, the classification of gene ‘essentiality’ is highly challenging in 3D culture conditions.

Therefore, there is a need for a screening method that can be adapted for a broad range of complex culture conditions that include low proliferation rates to identify essential genes. This motivated us to develop **Guide-Only control CRISPR (GO-CRISPR)**. GO-CRISPR is a scalable loss-of-function screening method that can be used to discover essential genes in standard monolayer (2D) or complex 3D culture conditions such as dormant tumor spheroids that exhibit arrested cell proliferation. To support broad usability, we also developed TRACS (**T**oolset for the **R**anked **A**nalysis of GO-CRISPR Screens) to automate the analysis of GO-CRISPR screens in an easy-to-use software package. Together, GO-CRISPR and TRACS allowed us to discover novel survival pathways in dormant ovarian cancer spheroids, whereas established CRISPR screening and analysis approaches were unable to find essential genes. We expect that this approach can be broadly applied to genome-wide loss-of-function CRISPR screens in low proliferation biological contexts.

RESULTS

The GO-CRISPR Workflow

The challenges presented by genome wide CRISPR screening in growth arrested populations of cells motivated us to develop a new workflow that could reveal critical insights into mechanisms of survival in cancer cell dormancy. We developed GO-CRISPR to overcome these challenges and its typical experimental workflow is illustrated in **Figure 1A**. CRISPR screens depend on high-level Cas9 expression to ensure maximum efficiency of gene disruption in Cas9-positive cells transduced with a pooled sgRNA library (L_0) (Shalem et al., 2014; Zhou et al., 2014). GO-CRISPR uniquely incorporates sequencing data from a parallel screen in which Cas9-negative cells are also transduced with the same pooled sgRNA library (L_0). Both the Cas9-positive and Cas9-negative cells are treated in an identical manner. Following antibiotic selection for sgRNA transduction and expansion into triplicate cultures, cells are harvested from the initial culture condition (T_0). Next, both Cas9-positive and Cas9-negative populations are exposed to the desired selective pressure or culture conditions (P_s) and cells are harvested from the final culture condition (T_f). Next-generation sequencing (NGS) is then used to quantitate the abundance of PCR-amplified sgRNA sequences from these 12 samples, as well the initial library preparation (L_0).

To evaluate GO-CRISPR screens, we developed the TRACS algorithm that integrates data from Cas9-positive and Cas9-negative populations to make gene essentiality predictions (**Figure S1**). It is based on assigning gene enrichment scores similar to the single gene score previously described by Wang *et. al.* (Wang et al., 2017). However, TRACS differs by calculating three different enrichment scores for each gene (**Figure 1A** in red). These include a Library Enrichment Score (Library ES) that compares sgRNA read counts for each gene between Cas9-negative cells and the library (L_0) to determine Cas9-independent non-gene-editing-related changes in

Figure 1

Figure 1: Typical experimental workflow for GO-CRISPR screening and analysis using TRACS.

(A) iOvCa147 High-grade serous ovarian cancer (HGSOC) cells were transduced with lentivirus expressing Cas9. High efficiency Cas9-positive cells (top row) and Cas9-negative cells (bottom row) were transduced with the GeCKO v2 pooled sgRNA library (L_0). After antibiotic selection, both Cas9 positive and negative cells were split into triplicates (x3) and maintained in initial culture conditions (T_0) before being transferred to suspension culture conditions in ULA plasticware (selective pressure, P_s) to induce spheroid formation and select for cell survival. Viable spheroid cells were then transferred to standard plasticware to facilitate reattachment in the final culture condition (T_f). The initial pooled sgRNA library (L_0) and Cas9-positive and Cas9-negative cells were collected at T_0 and T_f for sgRNA quantitation by NGS. TRACS was used to calculate Library, Initial and Final Enrichment Scores (ES) using read quantities from L_0 and Cas9-positive and Cas9-negative samples. **(B)** 3D plot output from TRACS illustrating the Library ES, Initial ES and Final ES for each gene. Genes highlighted in dark blue have low Library ES (determined by calculating the first quartile value across all Library ES; < 985 in this experiment). **(C)** Euler diagram showing the distribution of retained (in red) and discarded genes based on the Library ES (16,284 genes had Library ES > 985). **(D)** 2D scatter plot output from TRACS showing the distribution of Initial ES and Final ES for all genes. Genes highlighted in light blue (6,717 genes) met the low Library ES cutoff and had a negative Enrichment Ratio (ER) and $p_{adj} < 0.05$, indicating their sgRNA abundance decreases in T_f compared to T_0 .

abundance. This is an important consideration since pooled sgRNA library preparations do not uniformly represent all genes (Sanjana et al., 2014). An Initial Enrichment Score (Initial ES) is calculated by comparing sgRNA abundances for each gene in Cas9-positive cells relative to their abundances in Cas9-negative cells where they cannot direct gene editing. Lastly, a Final Enrichment Score (Final ES) determines sgRNA abundance between Cas9-positive and Cas9-negative cells following the exposure of both populations to the desired selective pressure or culture condition (P_s). For each gene, the Library ES, Initial ES and Final ES are weighted according to the number of sgRNAs that are detected for that gene. Thus, a relatively low Initial ES or Final ES indicate reduced sgRNA abundance in the Cas9-positive population and these scores incorporate a penalty for undetected sgRNAs to emphasize the most reliable sgRNA measurements. Finally, TRACS calculates an Enrichment Ratio (ER) that is the $\log_2$ -fold-change (LFC) value between the Final ES and Initial ES to reveal changes in relative abundance between T_0 and T_f culture conditions to detect sgRNAs that were depleted under the selective pressure (P_s), thereby identifying gene essentiality. The ER informs researchers if a gene shows essentiality for fitness (negative ER) or is non-essential (positive ER) in the experimental condition.

Discovering Ovarian Cancer Spheroid Vulnerabilities

To demonstrate the value of the GO-CRISPR and TRACS workflow, we performed a genome-wide screen in iOvCa147 high-grade serous ovarian cancer (HGSOC) cells. HGSOC is a highly metastatic disease in which cells detach from primary tumors and aggregate to form 3D spheroids in the abdomen (Matulonis et al., 2016). These spheroid cells are growth arrested and highly resistant to chemotherapy, emphasizing the need to discover their vulnerabilities to improve treatment (Bowtell et al., 2015). We designed a GO-CRISPR screen experiment (**Figure 1A**) to elucidate the genes and pathways that are critical to spheroid cell survival using ultra-low

attachment (ULA) plasticware to induce spheroid formation *in vitro* (MacDonald et al., 2017). Ovarian cancer cells undergo significant cell death in suspension culture while spheroids form, therefore after 48-hours we transferred cells back to standard plasticware to allow reattachment and purification of viable cells.

Following analysis with TRACS, we sought to discover genes that were most selectively required for survival in suspension conditions; these represent potential therapeutic targets for dormant ovarian cancer cell spheroids. **Figure 1B** displays each ES in a 3D plot that reveals the distribution of scores in each dimension and highlights genes with low Library ES in dark blue. A low Library ES means that a gene's sgRNA sequences were poorly represented at T_0 due to non-gene-editing events that occurred between viral transduction of the pooled sgRNA library and antibiotic selection. This is an important consideration because when a gene's Library ES is low, its initial sgRNA abundance is also low, and relatively small changes in sgRNA abundance can lead to extreme enrichment scores at T_0 (Initial ES) or T_f (Final ES) (**Figure S2A**). To avoid these false positives, we excluded the first quartile of Library ES from further analysis (Library ES < 985 in this experiment) (**Figure 1C and Figure S2B**). To discover genes essential for spheroid cell survival, we focused our attention on those that had a negative ER. In **Figure 1D**, genes highlighted in light blue met the Library ES cutoff (> 985) and had $ER < 0$ and $p_{adj} < 0.05$ at a false discovery rate (FDR) of 10%. We found 6,717 genes that met these criteria and the top 10 genes with the most negative ER are shown in **Table 1**. This data suggests these are the ten most essential genes required for spheroid cell viability in iOvCa147 cells.

Validation of TRACS Gene Essentiality Predictions

To determine the validity of gene essentiality predictions made by TRACS, we measured its ability to categorize the 1,000 non-targeting control (NTC) sgRNAs from the GeCKO v2 pooled

Table 1: Top 10 genes with the most negative Enrichment Ratio (ER) in TRACS

Gene	Library ES	Initial ES	Final ES	ER	p_{adj}
<i>AGPS</i>	2648.39	1384.00	16.61	-6.38	1.01×10^{-2}
<i>SLC2A11</i>	2421.17	2216.28	37.33	-5.89	1.06×10^{-2}
<i>ZC3H7A</i>	2397.67	2592.22	62.67	-5.37	5.19×10^{-3}
<i>PDCD2L</i>	1449.39	3018.56	82.67	-5.19	1.86×10^{-3}
<i>NPM1</i>	2966.72	2070.89	68.06	-4.93	9.78×10^{-3}
<i>KIAA1731</i>	3063.72	2050.61	83.78	-4.61	4.92×10^{-3}
<i>MAP3K6</i>	1741.61	2206.22	93.39	-4.56	2.76×10^{-4}
<i>SSH1</i>	1281.61	1347.67	57.50	-4.55	1.64×10^{-2}
<i>MFN2</i>	2779.67	2080.00	89.83	-4.53	1.16×10^{-2}
<i>CREBL2</i>	1449.78	2211.22	96.67	-4.52	6.69×10^{-3}

library. These NTC sgRNAs target non-coding intergenic sequences and should rank as non-essential (Sanjana et al., 2014). We computed a receiver operating characteristic curve (ROC) and determined the area under the curve (AUC) was 98.5%, demonstrating that TRACS correctly identified NTC sgRNAs as non-essential (**Figure 2A**). This is a critical control because amplified genome regions produce false essential calls among non-coding controls (Aguirre et al., 2016; Wang et al., 2015). HGSOC is characterized by extensive amplifications and deletions (Patch et al., 2015) and this data demonstrates TRACS eliminates this potentially confounding interpretation. For added validation, we used CRISPR/Cas9 to disrupt the top five genes with the most negative ER (**Table 1**) in iOvCa147 cells. Independent knockout of each gene showed significant loss of viability under suspension culture conditions (**Figure 2B**). Conversely, knockout of the top five genes with the most positive ER did not compromise viability, suggesting our GO-CRISPR screen approach reliably discovers loss-of-function events (**Figure 2C**).

GO-CRISPR and TRACS Identify Novel Pathways in HGSOC

To further explore the genes identified, we performed gene ontology and pathway enrichment analysis with genes that had a negative ER and $p_{adj} < 0.05$ and found 109 significantly enriched pathways (**Figure 2D**). Among these are cell cycle regulation (MacDonald et al., 2017), MAPK signaling (Sun et al., 2017) and TP53 signaling (Patch et al., 2015) which are known to be involved in HGSOC progression and metastasis. Remarkably, our analysis also found novel pathways that have not yet been implicated in HGSOC including Rho GTPase signaling and interleukin signaling. Together, these data demonstrate that GO-CRISPR and TRACS can robustly identify functionally connected genes to enable novel pathway discoveries.

Comparison of GO-CRISPR with conventional CRISPR screen workflows

Figure 2

Figure 2: Validation of TRACS gene essentiality predictions.

(A) The GeCKO v2 pooled library contains 1,000 non-targeting control (NTC) sgRNAs that should not elicit a change in cell fitness. We evaluated the ability of TRACS to classify these sgRNAs by computing a receiver operating characteristic curve (ROC). The area under the curve (AUC) was determined to be 98.5%, indicating TRACS accurately classifies these NTC sgRNAs as non-essential. **(B)** We evaluated the essentiality of the top five genes with the most negative ER: *AGPS*, *SLC2A11*, *ZC3H7A*, *PDCD2L*, *NPM1* (see **Table 1**). CRISPR/Cas9 was used to disrupt each gene in iOvCa147 cells and pure single-gene knockout populations were assayed for spheroid cell viability in suspension culture conditions. Disruption of these genes resulted in significantly reduced cell viability. Genes in bar graph are arranged from most negative ER to least negative ER. **(C)** We similarly knocked out the top five genes with the most positive ER (*EPS15*, *hsa-mir-761*, *RPAP1*, *SYAP1*, *TRAF3IP1*) and assayed for viability in suspension culture conditions. Disruption of these genes did not adversely affect viability. Genes in bar graph are arranged from smallest to largest ER. For B and C, each point represents a biological replicate (n = 6). Bars represent means and error bars represent standard deviation. Statistics was performed using one-way ANOVA; ** denotes $p < 0.01$; *** denotes $p < 0.001$; **** denotes $p < 0.0001$; ns denotes not significant ($p > 0.05$). **(D)** We performed gene ontology and pathway enrichment analysis with the 6,717 genes identified by TRACS to have negative ER and plotted the results. The minimum genes required for enrichment per category was set to 45 to ensure stringent selection of pathways. The dashed vertical line represents a p_{adj} value of 5×10^{-9} . Pathways to the right of this line have $p_{adj} < 5 \times 10^{-9}$ after controlling for FDR at 10%. Pathways labelled in blue are previously undescribed in HGSOC.

Formation of growth arrested HGSOc spheroids in suspension culture is a stressful process in which many cells die without being incorporated into a spheroid. Moreover, the communal nature of spheroids further suggests that individual gene loss events in single cells may be masked in loss-of-function CRISPR screens through non-cell autonomous mechanisms. Thus GO-CRISPR and TRACS were born out of the desire to screen a significantly challenging biological scenario. To fully illustrate the advantages of GO-CRISPR and TRACS, we have analyzed the triplicate replicates of T_0 and T_f solely in Cas9-expressing cells using MAGeCK-RRA, MAGeCK-MLE (Wang et al., 2019) and BAGEL (Hart et al., 2015) as this represents a commonly used CRISPR screen workflow that lacks guide only controls (**Figure S3**). A basic premise for genome-wide CRISPR screens using only Cas9-expressing cells is that poorly represented sgRNAs, or stochastic changes unrelated to the experiment in question, will be removed through statistical cutoffs. Analysis of this data using MAGeCK-RRA/MLE did not detect any essential genes using standard statistical cutoffs (**Figure S3A-D**), including the genes with the most negative ER that were found to be essential by TRACS (**Figure S3I**). BAGEL did not discover essential genes either (**Figure S3G-H**). We then removed statistical cutoffs in MAGeCK-RRA/MLE and found approximately 30% of top-ranked genes had low Library ES according to TRACS, reinforcing the previously described phenomenon of identifying false positives due to low initial sgRNA abundances (**Figure S3E-F and Figure S4**). Additionally, our computed ER discriminates essentiality of NTCs effectively (**Figure 2A**), whereas MAGeCK (without statistical cutoffs) frequently misclassifies NTCs as essential (**Figure S5A**). TRACS was also noticeably more reliable at identifying universally essential and non-essential gene sets (Hart et al., 2015) (**Figure S5B-C**). TRACS penalizes genes that have low sgRNA numbers and favors those with higher sgRNA values to further mitigate the effects of stochastic sgRNA loss and ensure that gene essentiality predictions

are made using the largest possible sample size (**Figure S6A**). Without statistical cutoffs, many MAGeCK-MLE top-ranked gene decisions are based on single gRNAs (**Figure S6B**). Overall, integrating data from the pooled sgRNA library and Cas9-negative populations allows TRACS to outperform other methods to accurately predict gene essentiality in a challenging low proliferation, suspension culture scenario.

DISCUSSION

The GO-CRISPR and TRACS workflow offers an important alternative to conventional genome-wide loss-of-function CRISPR screens. It rigorously identifies genes that contribute to survival and facilitates novel mechanistic discoveries in low proliferation culture conditions by controlling for the stochastic effects of Cas9-independent sgRNA loss. Most notably, we demonstrated the use of this screening workflow in a 3D ovarian cancer spheroid model to identify novel pathways that have not yet been described in HGSOC. This would not have been possible using conventional screening methods that lack guide-only controls.

To accommodate guide-only controls, we needed a new analysis pipeline. CRISPR screen analysis pipelines generally require an understanding of programming or advanced Unix/Linux knowledge to setup and manipulate raw NGS read files. In contrast, the TRACS software suite (<https://github.com/developerpiru/TRACS>) presents researchers with an easy-to-use graphical environment for analysis and data exploration. TRACS fully automates the analysis process – from raw NGS files to output – which will significantly reduce the barrier for many researchers to use GO-CRISPR. Furthermore, TRACS is fully scalable and can be deployed on a local workstation or a multi-CPU platform such as Amazon Web Services, Google Cloud Platform, or Microsoft Azure. We also provide example workflows and documentation to use TRACS on these platforms, including Docker containers for Linux, Mac OS and Windows that will automate setup.

We used the GeCKO v2 pooled sgRNA library (Sanjana et al., 2014) in our screen. However, the modularity of GO-CRISPR and TRACS will allow for the use of any pooled sgRNA library as long as Cas9 expression is separate from sgRNA viral delivery. In addition, the flexibility of TRACS in terms of unrestricted replicates and sgRNA library size will support the use of validated libraries, such as GeCKO v2, or custom libraries to answer novel questions across

biological systems of interest. Taken together, we anticipate GO-CRISPR and TRACS will open new opportunities for loss-of-function screens across diverse model systems and biological questions.

METHODS

Generation of Cas9-positive cells: High-grade serous ovarian cancer (HGSOC) iOvCa147 cells have previously been reported (MacDonald et al., 2017). They were transduced with viral particles encoding a Cas9 expression cassette (lentiCas9-Blast, Addgene #52962) to generate cells constitutively expressing Cas9 (Cas9-positive cells). Cells were selected with blasticidin (20 µg/mL). Single-cell clones were isolated by limiting dilution. Lysates were collected from clones and western blots were performed to determine Cas9 expression (Cell Signaling #14697). Cas9 editing efficiency was determined by viability studies using sgRNAs targeting selected fitness genes (*PSMD1*, *PSMD2*, *EIF3D*) and a non-targeting control (*LacZ*) as previously reported (Hart et al., 2015). A single clone showing the most effective Cas9 activity was selected for all further studies.

GeCKO v2 library preparation: HEK293T cells were transfected with the combined A and B components of the GeCKO v2 (Addgene #1000000048, #1000000049) whole genome library (123,411 sgRNAs in total) along with plasmids encoding lentiviral packaging proteins. Media was collected 2-3 days later and any cells or debris were pelleted by centrifugation at 500 x g. Supernatant containing viral particles was filtered through a 0.45 µm filter and stored at -80°C with 1.1 g/100 mL BSA.

GO-CRISPR screen in iOvCa147 cells: iOvCa147 Cas9-positive or Cas9-negative cells were transduced with virus collected as described above at a multiplicity of infection of 0.3 and with a predicted library coverage of >1000-fold. Cells were grown in media containing 2 µg/mL puromycin (Sigma #P8833) to eliminate non-transduced cells. Cells were maintained in complete media containing puromycin in all following steps. A total of 1.1×10^9 cells were collected and split into three groups consisting of approximately 3.0×10^8 cells each and were cultured for an

additional 2-3 days in complete media, then collected and counted. Triplicate samples of 6.2×10^7 cells were saved for sgRNA sequence quantitation at T_0 . The remaining cells (approximately 1.4×10^9 /set) were plated at a density of 2.0×10^6 cells/mL in each of twenty 10 cm ULA plates (total of 60 ULA plates). Following 2 days of culture, media containing spheroids was transferred to ten, 15 cm adherent tissue culture plates (total of 30 plates). The next day unattached spheroid cells were collected and re-plated onto additional 15 cm plates. This process was repeated for a total of 5 days at which point very few spheroids remained unattached. The attached cells were collected for DNA extraction and this population represents T_f . Complete media refers to DMEM/F12 media (Gibco #11320033) supplemented with 10% FBS (Wisent FBS Performance lot #185705), 1% penicillin-streptomycin glutamine (Wisent #450-202-EL) and 2 μ g/mL puromycin (Sigma #P8833).

High-throughput next generation sequencing (NGS): Cells were harvested and DNA was extracted using QIAmp Blood Maxi Kits (QIAGEN #51194). Genomic encoded sgRNA sequences were PCR amplified as previously described (Joung et al., 2017). Two rounds of PCR were performed. The initial round serves to increase the abundance of the initial sgRNA population, while the second round inserts barcodes necessary for identification of group and replicate number (sample barcode). PCR products were gel purified, quantitated by Qubit (Invitrogen), pooled and sequenced using an Illumina NextSeq 550 75-cycle high output kit (#20024906). FASTQ files were obtained containing raw reads and were demultiplexed to obtain individual FASTQ files for each sample. FASTQ files were processed accordingly for downstream analysis with TRACS, MAGeCK, or BAGEL.

Analysis with MAGeCK: FASTQ files were trimmed with Cutadapt (1.15) to remove adapter sequences and sample barcode identifiers. The library reference file (CSV) for the GeCKOv2

library was used in Bowtie2 (2.3.4.1) to align the initial library read FASTQ file and generate a BAM file (Samtools 1.7) in order to increase the read depth of the initial library. This library BAM file and trimmed FASTQ files for all samples were then inputted into the MAGeCK (0.5.6) count function to generate read counts. Differences in sgRNA abundance were computed using the MAGeCK-RRA (robust ranking aggregation) or MAGeCK-MLE (maximum likelihood estimation) methods. All plots and comparisons to TRACS were performed in R (3.6.2).

Analysis with BAGEL: BAGEL (0.91) was run using read counts generated by the MAGeCK (0.5.6) count function as described above. Standard non-essential and essential training gene sets were used as previously described (Hart et al., 2015). Bayes factors (BFs) obtained by BAGEL were plotted in R (3.6.2).

Analysis with TRACS: The library reference file containing a list of all sgRNAs and their sequences (CSV file), raw reads for the pooled sgRNA library (FASTQ files (L_0) and raw reads (FASTQ files) for all Initial (T_0) and Final (T_f) replicates for Cas9-positive and Cas9-negative cells (12 replicates) were loaded into TRACS (<https://github.com/developerpiru/TRACS>). TRACS then automatically trimmed the reads using Cutadapt (1.15). TRACS builds a Bowtie2 (2.3.4.1) index and aligns the trimmed initial sgRNA library read file to generate a BAM file using Samtools 1.7. MAGeCK (0.5.6) is then used to generate read counts from this library BAM file and all the trimmed sample FASTQ files. Instead of dropping all reads below a certain threshold (*e.g.* <30 counts), all reads were incremented by 1 to prevent zero counts and division by zero errors. The TRACS algorithm was then run using this read count file to determine the Library Enrichment Score (ES), Initial ES, Final ES and the Enrichment Ratio (ER) for each gene (see *The TRACS algorithm* section).

Data exploration using VisualizeTRACS: The VisualizeTRACS feature, part of the TRACS software suite, was then used to visualize and explore the data output from TRACS. Gene filtering (Library ES > 985, ER < 0, $p_{adj} < 0.05$ for our example ovarian cancer workflow) was performed, figures were generated and the final table of essential genes that met these criteria were downloaded for further analysis.

The TRACS algorithm: After read count preprocessing, TRACS first determines a Gene Score, GS , for every gene in the supplied library reference file by calculating the log₂-fold-change (LFC) from all sgRNAs for that gene for n replicates (minimum of 2 replicates required) of Cas9-positive and Cas9-negative samples:

$$GS_j = \sum_{i=1}^s \left(\log_2 \frac{[normalized\ sgRNA\ abundance]_{Cas9+}}{[avg\ normalized\ sgRNA\ abundance]_{Cas9-}} \right)$$

Where s is the number of unique sgRNAs for a gene j . This is done for each replicate such that for n replicates, there are n gene scores, GS , for a gene j . For each n replicates, the GS for all genes are then ranked in ascending order from 1 to x , where x is the rank of the gene with the highest GS in each respective replicate. TRACS then determines the Enrichment Score, ES_j , which is the average rank across all n replicates of a gene j , divided by the total number of sgRNAs, s , identified for that gene.

$$ES_j = \frac{\frac{1}{n} \sum_{i=1}^n GS\ rank}{s}$$

TRACS then determines the Enrichment Ratio, ER , for gene j by determining the LFC of the ES_{jTf} compared to ES_{jT0} .

$$ER_j = \log_2 \frac{ES_{jTf}}{ES_{jT0}}$$

TRACS calculates the p value for each gene using a paired t-test by pairing each of the n replicates together per gene per the initial (T_0) condition and final (T_f) condition. The Benjamini–Hochberg procedure is used to control the false discovery rate (FDR) at the user-defined level (10% in our example workflow).

After the ER is calculated, TRACS determines the distribution of Library ES values across all genes. The cutoff value for the Library ES was set to the first quartile for our example screen.

Pathway analysis: Using the final list of essential genes from TRACS, we performed gene ontology and pathway enrichment analysis using the ConsensusPathDB enrichment analysis test (Release 34 (15.01.2019)) for top-ranked genes of interest. p_{adj} values and ER values for each gene were used as inputs. The minimum required genes for enrichment was set to 45 and the FDR-corrected p_{adj} value cutoff was set to < 0.01 . The Reactome pathway dataset was used as the reference. For each identified pathway, ConsensusPathDB provides the number of enriched genes and a q value (p_{adj}) for the enrichment. Scatter plots were generated in R (3.6.2) using these values to depict the significant pathways identified.

Generation of single-gene knockouts: Gibson Assembly (NEB #E2611) was used to clone a pool of four sgRNAs per gene (*AGPS*, *SLC2A11*, *ZC3H7A*, *PDCD2L*, *NPM1*, *EPSI5*, *hsa-mir-761*, *RPAP1*, *SYAP1*, *TRAF3IP1*, and *EGFP*) into lentiCRISPR v2 (Addgene #52961). iOvCa147 cells were transduced with viral particles encoding a Cas9 and sgRNA expression cassettes. Cells were selected for 2-3 days in media containing 2 $\mu\text{g/mL}$ puromycin. Knockout cells were cultured for 72 hours in suspension conditions using ULA plasticware (2 x 10⁶ cells per well) to induce spheroid formation. Spheroid cells were then collected and transferred to standard plasticware to

facilitate reattachment for 24 hours. Reattached cells were fixed with fixing solution (25% methanol in 1x PBS) for 3 minutes. Fixed cells were incubated for 30 minutes with shaking in staining solution (0.5% crystal violet, 25% methanol in 1x PBS). Plates were carefully immersed in ddH₂O to remove residual crystal violet. Plates were incubated with detaining solution (10% acetic acid in 1x PBS) for 1 hour with shaking to extract crystal violet from cells. Absorbance of crystal violet at 590 nm was measured using a microplate reader (Perkin Elmer Wallac 1420) for each knockout and normalized to the EGFP control. Percent survival is inferred from relative absorbance.

Statistics: All error bars in the bar graphs represent standard deviation. Statistical significances were determined using two-way ANOVA. * denotes $P < 0.05$, *** denotes $P < 0.001$, **** denotes $P < 0.0001$ and ns denotes not significant ($P > 0.05$).

Abbreviations

AUC: Area under the curve; ER: Enrichment Ratio; ES: Enrichment Score; FDR: False discovery rate; GO-CRISPR: Guide-Only control CRISPR; GUI: Graphical user interface; HGSOC: High-grade serous ovarian cancer; L₀: pooled sgRNA library; LFC: Log₂-fold-change; NGS: Next generation sequencing; p_{adj} : adjusted p value; P_s: selective pressure; ROC: Receiver operating characteristic curve; SD: Standard deviation; sgRNA: single guide RNA; TRACS: Toolset for the Ranked Analysis of GO-CRISPR Screens; T₀: initial culture condition; T_f: final culture condition; ULA: Ultra-low attachment

Data and Code Availability

High throughput sequencing data from the GO-CRISPR screen is available from the GEO repository (accession number GSE150246). TRACS is available for download from the GitHub repository at <https://github.com/developerpiru/TRACS> or on Docker Hub at <https://hub.docker.com/r/pirunthan/tracs>. Complete documentation, reference sgRNA library file and TRACS output files are also available on the GitHub repository.

Authors' contributions

PP wrote the software, analyzed and validated the data and wrote the manuscript. PP and JIM performed the experiments. JIM and FAD wrote the manuscript.

Competing interests

The authors have no competing interests to declare

References

- Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., *et al.* (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* 6, 914-929.
- Boussouar, A., Tortereau, A., Manceau, A., Paradisi, A., Gadot, N., Vial, J., Neves, D., Larue, L., Battistella, M., Leboeuf, C., *et al.* (2019). Netrin-1 and its receptor DCC are causally implicated in melanoma progression. *Cancer Res.*
- Bowtell, D.D., Bohm, S., Ahmed, A.A., Aspuria, P.J., Bast, R.C., Jr., Beral, V., Berek, J.S., Birrer, M.J., Blagden, S., Bookman, M.A., *et al.* (2015). Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* 15, 668-679.
- Cai, M.Y., Dunn, C.E., Chen, W., Kochupurakkal, B.S., Nguyen, H., Moreau, L.A., Shapiro, G.I., Parmar, K., Kozono, D., and D'Andrea, A.D. (2020). Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double-Strand Break End Resection. *Cell Rep* 30, 2402-2415 e2405.
- Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., Ohta, Y., Toshimitsu, K., Nakazato, Y., Kawasaki, K., *et al.* (2016). A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. *Cell Stem Cell* 18, 827-838.
- Grandin, M., Meier, M., Delcros, J.G., Nikodemus, D., Reuten, R., Patel, T.R., Goldschneider, D., Orriss, G., Krahn, N., Boussouar, A., *et al.* (2016). Structural Decoding of the Netrin-1/UNC5 Interaction and its Therapeutical Implications in Cancers. *Cancer Cell* 29, 173-185.
- Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., *et al.* (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163, 1515-1526.
- Jacob, F., Salinas, R.D., Zhang, D.Y., Nguyen, P.T.T., Schnoll, J.G., Wong, S.Z.H., Thokala, R., Sheikh, S., Saxena, D., Prokop, S., *et al.* (2020). A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* 180, 188-204 e122.
- Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* 12, 828-863.
- Kenny, H.A., Lal-Nag, M., White, E.A., Shen, M., Chiang, C.Y., Mitra, A.K., Zhang, Y., Curtis, M., Schryver, E.M., Bettis, S., *et al.* (2015). Quantitative high throughput screening using a primary human three-dimensional organotypic culture predicts in vivo efficacy. *Nat Commun* 6, 6220.
- Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 32, 267-273.
- Lytle, N.K., Ferguson, L.P., Rajbhandari, N., Gilroy, K., Fox, R.G., Deshpande, A., Schurch, C.M., Hamilton, M., Robertson, N., Lin, W., *et al.* (2019). A Multiscale Map of the Stem Cell State in Pancreatic Adenocarcinoma. *Cell* 177, 572-586 e522.
- MacDonald, J., Ramos-Valdes, Y., Perampalam, P., Litovchick, L., DiMattia, G.E., and Dick, F.A. (2017). A Systematic Analysis of Negative Growth Control Implicates the DREAM Complex in Cancer Cell Dormancy. *Mol Cancer Res* 15, 371-381.
- Matulonis, U.A., Sood, A.K., Fallowfield, L., Howitt, B.E., Sehouli, J., and Karlan, B.Y. (2016). Ovarian cancer. *Nat Rev Dis Primers* 2, 16061.
- Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosch, I., Sanjana, N.E., *et al.* (2015). A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 162, 675-686.
- Patch, A.M., Christie, E.L., Etemadmoghadam, D., Garsed, D.W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P.J., *et al.* (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 521, 489-494.

- Planas-Paz, L., Sun, T., Pikiolek, M., Cochran, N.R., Bergling, S., Orsini, V., Yang, Z., Sigoillot, F., Jetzer, J., Syed, M., *et al.* (2019). YAP, but Not RSPO-LGR4/5, Signaling in Biliary Epithelial Cells Promotes a Ductular Reaction in Response to Liver Injury. *Cell Stem Cell* 25, 39-53 e10.
- Ringel, T., Frey, N., Ringnalda, F., Janjuha, S., Cherkaoui, S., Butz, S., Srivatsa, S., Pirkl, M., Russo, G., Villiger, L., *et al.* (2020). Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-beta Resistance. *Cell Stem Cell* 26, 431-440 e438.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11, 783-784.
- Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., *et al.* (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* 9, 5416.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., *et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84-87.
- Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 16, 299-311.
- Sun, C., Fang, Y., Yin, J., Chen, J., Ju, Z., Zhang, D., Chen, X., Vellano, C.P., Jeong, K.J., Ng, P.K., *et al.* (2017). Rational combination therapy with PARP and MEK inhibitors capitalizes on therapeutic liabilities in RAS mutant cancers. *Sci Transl Med* 9.
- Sung, P.J., Rama, N., Imbach, J., Fiore, S., Ducarouge, B., Neves, D., Chen, H.W., Bernard, D., Yang, P.C., Bernet, A., *et al.* (2019). Cancer-Associated Fibroblasts Produce Netrin-1 to Control Cancer Cell Plasticity. *Cancer Res* 79, 3651-3661.
- Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E., and Schier, A.F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* 7, 11750.
- Vlachogiannis, G., Hedayat, S., Vatsiou, A., Jamin, Y., Fernandez-Mateos, J., Khan, K., Lampis, A., Eason, K., Huntingford, I., Burke, R., *et al.* (2018). Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 359, 920-926.
- Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., *et al.* (2019). Integrative analysis of pooled CRISPR genetic screens using MAGECKFlute. *Nat Protoc* 14, 756-780.
- Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. *Science* 350, 1096-1101.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80-84.
- Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* 168, 890-903 e815.
- Zanoni, M., Piccinini, F., Arienti, C., Zamagni, A., Santi, S., Polico, R., Bevilacqua, A., and Tesei, A. (2016). 3D tumor spheroid models for in vitro therapeutic screening: a systematic approach to enhance the biological relevance of data obtained. *Sci Rep* 6, 19103.
- Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487-491.