Image-based pooled genetic screens for complex cellular phenotypes

by

Luke Benjamin Funk

B.S., Engineering
LeTourneau University, 2016

Submitted to the Harvard-MIT Program in Health Sciences and Technology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Medical Engineering and Medical Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Signature of Author .....
Harvard-MIT Program in Health Sciences and Technology
April 29, 2022

Certified by .....
Paul C. Blainey, PhD
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by .....
Emery N. Brown, MD, PhD
Professor of Computational Neuroscience and Health Sciences and Technology
Director, Harvard-MIT Program in Health Sciences and Technology

Image-based pooled genetic screens for complex cellular phenotypes

by

Luke Benjamin Funk

Submitted to the Harvard-MIT Program in Health Sciences and Technology
on April 29, 2022 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Medical Engineering and Medical Physics

ABSTRACT

Biological processes are organized in hierarchical interactions of molecules, cells, tissues, and organisms. Cells perform complex functions individually, which when misregulated can result in disease states affecting the entire organism. However, knowledge of the genetic and molecular basis for many cellular phenomena is incomplete, limiting the ability to reverse disease states and engineer biological function. Although recent technologies have enabled scalable functional genomics approaches such as pooled CRISPR screening, the cellular phenotypes that can be linked to gene function in a pooled screen have been restricted to measurement by sequencing, and are often a step removed from the biological process of interest. In contrast, microscopy provides a high-throughput and flexible means to measure a wide range of biologically-relevant phenotypes. Here, we apply an image-based pooled screening approach based on *in situ* sequencing to understand the contributions of protein-coding genes to a wide range of cellular processes. Specifically, we combine pooled CRISPR/Cas9 genomic perturbations of 5,072 fitness-conferring genes with microscopy-based visualization of DNA, DNA damage response, actin, and microtubules across more than 31 million human cells. By leveraging the complex phenotypes resulting from each perturbation, we identify co-functional genes across diverse cellular activities, revealing novel gene functions and associations. Additionally, we demonstrate pooled CRISPR screening combined with live-cell imaging of more than 400,000 cell division events to further identify unexpected contributions to chromosome segregation. Altogether, this work demonstrates image-based pooled genetic screening as a scalable approach to measure and understand genetic contributions to complex phenotypes and cellular functions.

Thesis Supervisor: Paul C. Blainey
Title: Associate Professor of Biological Engineering

Thesis Committee Members

Mark Bathe, PhD (Chair)
Professor of Biological Engineering
Massachusetts Institute of Technology

Paul C. Blainey, PhD (Thesis Supervisor)
Associate Professor of Biological Engineering
Massachusetts Institute of Technology

Iain M. Cheeseman, PhD
Professor of Biology
Massachusetts Institute of Technology

Table of contents

Acknowledgements

My undeniably positive experience in graduate school is the result of both my initial conditions (prior relationships and education) and the various environmental factors experienced along the way (new relationships, challenges, and opportunities). It would be impossible to list all of these contributions here, but below I attempt to capture my gratitude for a small subset.

The work presented in this thesis is fundamentally the result of the support and mentorship of Paul Blainey, whose never-failing optimism kept my motivation going numerous times. I am similarly thankful for the continued mentorship and collaboration of Iain Cheeseman, who has gone above and beyond in helping me navigate projects and graduate school in general. Additionally, Mark Bathe has provided invaluable encouragement and has always been willing to discuss topics big or small.

My time in graduate school would not have been complete without the time I spent in Woods Hole. On two occasions, my experiences at the MBL refreshed my love of science and kick-started significant progress in the lab. I am particularly grateful for the ability to work with both Rob Phillips and Jan Funke in separate collaborations initiated at the MBL.

I was very privileged in graduate school to be able to work daily with the amazing scientists in the Blainey lab. From my early wanderings around ill-defined projects up to the present, I was encouraged, challenged, and valued by many lab members. I have strived to do the same for others, and am especially thankful for “Team Lasagna,” for their technical help and friendships over the years: David Feldman, Avtar Singh, Becca Carlson, Anna Le, Russell Walton, and Jake Qiu among others. Additionally, the lab would not be functioning without the logistical and relational support of Emily Botelho. The work in this thesis would also not have been possible without the important contributions of Kuan-Chung Su, Jimmy Ly, Marek Nagiec, and Jeff Cottrell. There are numerous others I had the privilege of getting to know during graduate school around the Broad, MIT, the MBL, Janelia, and during the many Zoom conversations held in my search for a future job. Each of these interactions contributed meaningfully to this thesis, although in ways that would be difficult to measure.

Although I believe I took full advantage of the rich learning environment available during graduate school at MIT, there are two powerful minds in my life that learned significantly more than I did in shorter time frames: my daughters Joy and Jean. It has been amazing to watch them grow, and they have both provided much-needed distraction and motivation since their entry into the world earlier in graduate school, even if the distraction wasn’t always appreciated in the moment. The beautiful individuals that Joy and Jean are growing into are primarily reflections of my wife Colleen. Colleen has willingly sacrificed more than I could fathom in the last six years, and deserves congratulations for this moment at least as much as myself. Colleen, you are strong and I am eternally grateful for your support. Finally, I am thankful for my parents and siblings, who have provided support, advice, and relationships from my childhood through the present that formed who I am.

Chapter 1. Introduction

Portions of this chapter and the corresponding figures are adapted from the publication “Pooled genetic perturbation screens with image-based phenotypes” in *Nature Protocols* (*1*), with the following authors: David Feldman*, Luke Funk*, Anna Le, Rebecca J. Carlson, Michael D. Leiken, FuNien Tsai, Brian Soong, Avtar Singh, and Paul C. Blainey. (*equal contributions)

1.1 Perturbation-based functional genomics

Biological phenomena arise from hierarchical interactions of molecules, cells, groups of cells, and organisms. At subcellular length scales, interactions between molecules are governed by their structure, stoichiometry, and spatial organization. Molecular structures encoded by the genome, including proteins and functional RNAs, are inheritable and drive the evolution of cellular functions and emergent properties. Understanding the functions and interactions of such molecules can lead to exogenous control over biological processes and higher-order phenotypes, including therapeutic intervention in disease progression. Recently, the sequencing of full genomes has led to thorough enumeration of protein-coding and non-coding genetic components (*2*, *3*), although their biological functions are unclear in many cases. Now, a major goal of biology is to characterize the role of each genetically-encoded functional molecule across cellular processes.

One common experimental approach to functional characterization is to remove or disrupt a genetic element of interest, such as knocking out a protein-coding gene, and then evaluate the resulting phenotype at the molecular, cellular, or organismal level as compared to a matched, unperturbed control sample. A consistent genetic background between comparison samples enables clean evaluation of the exact contributions of the perturbed genetic element, which can be performed in a screen across many genetic perturbations. Historically, genetic screens were performed in model organisms, such as yeast, *C. elegans,* and *D. melanogaster*, using random mutagenesis or by leveraging convenient genetic properties of the organisms (*4–6*). Often the phenotypes of interest were identified visually on the organism level, and successfully linked molecular functions of genes to high-level phenotypes (*7–10*). Although human gene function is of particular interest due to the potential for insights into disease processes, such experimental perturbation studies are ethically and practically infeasible at the organism level. Additionally, mutagenesis experiments with mammalian-derived cell cultures were historically difficult due to the lack of a meiotic phase in culture, limiting the use of classical genetics approaches (*11*).

However, new tools for precise and scalable genetic perturbations have since been developed, enabling efficient genetic screening in human cells.

An initial breakthrough in programmatic genetic perturbations came with synthetic utilization of RNA interference pathways to knockdown the expression of genes (*12*). This built on top of newly-available genome sequences to enable scalable and designable gene perturbations, and was widely employed in large-scale genetic screens in human cells (*13*–*16*). However, RNAi reagents were plagued with off-target effects which ultimately limited their usability (*17*). More recently, CRISPR-based RNA-guided nuclease systems were identified in prokaryotes and further developed for application in other organisms (*18*, *19*), exhibiting higher specificity and efficiency than RNAi. In the most common CRISPR application in human cells, the Cas9 endonuclease is expressed in combination with a single guide RNA (sgRNA) containing a 20 nucleotide sequence complementary to a target genetic locus. Cas9 efficiently and precisely makes a double-strand break at the targeted loci, which is then natively repaired via error-prone DNA repair pathways. This double-strand break repair often results in random genomic insertions or deletions ("indels") that ideally disrupt the correct expression or function of the gene. The programmability and precision of such genetic perturbations, in combination with advances in sequence measurement technologies, has made large-scale genetic perturbation screens in human cells feasible and reliable as discussed in the following section.

1.2 Scalable genetic screens via pooled perturbations

In addition to having an efficient means of genetic perturbation, large-scale screens require an equally-scalable method for measuring the resulting phenotypes. This is commonly achieved by pooled screening, where many genetic perturbations are delivered in a single combined sample and phenotypes are measured across a pooled library of cells, each containing a separate perturbation. However, pooled screens fundamentally require a method to link individual genotypes (either the perturbation identity or the actual resulting genetic alteration) to corresponding phenotypes. The simplest approach to this problem is enrichment-based screening, where perturbations are stably preserved in the genotype of each cell and a selection step is applied, such as growth under conditions preferentially enriching a desired phenotype. The genetic disruptions resulting in the target phenotype are then identified by measuring the relative change in abundance of genotypes before and after enrichment as a proxy for cellular fitness in the applied conditions, typically using next-generation sequencing (NGS; Fig. 1.1). This approach

has been successfully applied both with RNAi and CRISPR reagents (*13–15*, *20–22*). A practical advantage of pooled screening is that the full screen is handled as a single cell population, without the need for expensive laboratory automation often required in large-scale arrayed screens to maintain many individually-perturbed cell populations. Additionally, pooled screens are exceptionally well-controlled, as all control perturbations (e.g., CRISPR sgRNAs with no target site in the human genome) are always present within the same batch (e.g., cell culture flask) as screening perturbations and see identical environmental conditions. Control and screening perturbations also have their phenotypes measured simultaneously within the same sample, which further limits batch-correlated measurement noise.

However, the requirement to link phenotypes of interest to a growth advantage has fundamentally limited the range of phenotypes accessible via pooled screening. Slightly more diverse phenotypes are possible via direct selection of cell populations based on fluorescent reporter systems using fluorescence-activated cell sorting (FACS). However, both cell fitness- and FACS-based screens fundamentally project complex phenotypes into a single-dimensional measurement, and can only do so one phenotype at a time. Additionally, perturbation genotypes are not uniquely traceable to individual cells and the resulting enrichments reflect either population- or lineage-averaged phenotypes, depending on the experimental approach (*20–24*). More recently, pooled CRISPR screens have been developed that read out single-cell resolved perturbation phenotypes using molecular profiling methods, such as single-cell RNA sequencing or mass cytometry (*25–30*). These approaches simultaneously capture a genotype and phenotype for each cell, identifying the former via either a proxy barcode or the perturbation sequence itself. The acquired phenotypes also represent much more complex information than enrichment screens, with thousands of phenotype features measured per cell that can be leveraged to identify genes involved in multiple different biological pathways in a single screen. Rather than unidimensional, averaged per-perturbation enrichments, these screens yield a rich matrix of cells by phenotypic features in which each cell is labeled by perturbation genotype (Fig. 1.1). However, the information accessible by such molecular profiling methods in pooled screening formats is limited to phenotypes observable via genomic sequencing or simple proteomic measurements. These measurements focus primarily on a single mode of biological function, the regulation of molecular abundance. These abundance measurements are often a step removed from the biological phenomena of interest, and at best provide poor proxy measurements for phenotypes such as intercellular interactions, cell morphology, spatial organization of proteins and organelles, and dynamic signaling events. Furthermore, the

application of molecular profiling methods in pooled screens has been limited in scale, primarily due to the corresponding cost. While the focus on sequencing-based phenotype measurements in pooled screens originates from technical considerations, it also reflects a current preoccupation with sequencing technologies in modern biology, which may not always be the best hammer for a given nail. As demonstrated by historical examples of laborious genetic screens in model organisms for high-level phenotypes beyond the scale of molecular abundance (*7–10*), there is potential for much to be learned from expanding the current range of phenotype measurements compatible with modern pooled screening in human cells.

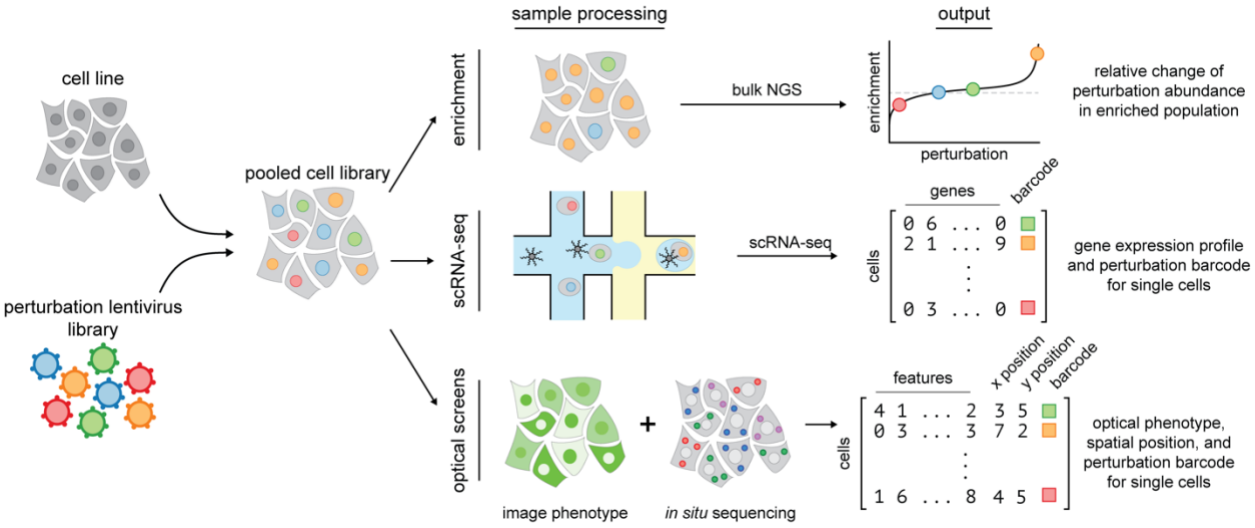

Figure 1.1. Pooled screening approaches. In pooled screening, a population of cells is subjected to a library of genetic perturbations, such as guide RNAs for CRISPR screens. Enrichment, single-cell profiling and optical-based assays are three approaches for phenotypic readout. Enrichment-based screens determine population-level changes in perturbation abundance by bulk NGS following an applied selection step. Single-cell profiling and optical screens do not require an enrichment step and instead rely on information-rich phenotypic measurements. Single-cell assays pair perturbation barcodes to a cell phenotype, such as transcriptome measurements, at single-cell resolution. Through *in situ* sequencing, optical pooled screens pair image-based phenotypes with perturbation barcodes, also at single-cell resolution.

1.3 Using microscopy to measure phenotypes

For several centuries, light microscopy has been an invaluable tool for observation in biology, and was regularly employed for phenotype measurement in classical genetic screens (*7–10*, *31*). Microscopy offers a plethora of options for monitoring the phenotypic state of cells, many of which

provide information complementary to molecular measurements of RNA or protein abundances. Fixation of cells followed by fluorescent labeling with antibodies, RNA or DNA hybridization probes, or small-molecule affinity reagents allows measurement of spatial distributions in approximately five distinct channels, with many additional channels potentially available using sequential detection and/or hyperspectral imaging approaches (*32*–*36*). In living cells, the abundance and localization of protein and RNA molecules can be visualized by genetic fusion or binding to fluorescent reporters. Additionally, fluorescent reporters can relate a wide range of biochemical states in living cells, ranging from ion concentrations and membrane potentials to kinase activity (*37*–*40*). Time-lapse imaging can track cells longitudinally, enabling high-resolution measurements of dynamic phenotypes, such as progression through the cell cycle or the time to activate or relax a signaling response (*41*–*45*). Image-based assays can also employ mixtures of cell types that are optically distinguishable (e.g., by reporter or marker expression) to more accurately model a physiological environment or identify interactions between cells (*46*–*48*). While image phenotype measurements can often be quantified via straightforward metrics, such as mean fluorescence intensity or cross-correlation between channels, machine learning techniques have shown substantial enhancements in classifying cell behaviors by extracting meaningful features from pixel-level raw data or higher-level descriptors (*49*–*53*), as also discussed in section 3.4.

Due to their versatile nature, image-based cell assays have been used successfully for a wide variety of genetic screens with arrayed perturbations, including genome-wide RNA interference and targeted CRISPR-based screens characterizing genes involved in mitosis and cell cycle progression (*16*, *54*), membrane trafficking (*55*), autophagy (*56*), viral and bacterial infection (*57*–*60*), and cellular morphology (*61*). However, the complexity and cost of performing large-scale arrayed genetic screens have limited their feasibility for many applications and potential users as they require expensive or customized automation to deliver precise amounts of each perturbation, culture individual cell populations, and image phenotypic assays at scale. Maintaining arrayed cell populations is particularly challenging with assays that require longer periods between perturbation and phenotypic measurement (e.g., assays requiring cell differentiation after perturbation) as differential perturbation efficiency and fitness effects accumulated over time can increase well-to-well variability and biological noise, in addition to the burden of maintaining a large number of wells in culture over time. CRISPR-based perturbations especially suffer from this limitation as they typically require more time to modulate target gene activities than RNA interference—often several days. Additionally, new arrayed perturbation libraries are expensive

to synthesize and will therefore lag technical developments in fast-moving fields such as CRISPR perturbations. For these reasons, comprehensive arrayed CRISPR-based imaging screens have been largely impractical. A method to link phenotypes measured using microscopy to perturbation genotypes in the context of pooled screening would alleviate many of these limitations by leveraging the technical advantages of pooling genetic perturbations discussed previously.

1.4 Combining microscopy phenotypes with pooled genetic perturbation screens

Several recent technologies enable image-based pooled screens in bacterial and mammalian systems, linking perturbation identity to phenotype measurement by physically retrieving relevant cell subpopulations or by using *in situ* optical barcoding of genetic perturbations. Approaches that physically retrieve subpopulations of cells are in essence an extension of enrichment-based pooled screening. In one approach, advances in flow cytometry have enabled fast acquisition of cell images while simultaneously sorting cells based on measured image attributes (*62*). This enables FACS-like genetic screens of phenotypes more complex than fluorescence intensity of a reporter system, but the image resolution is limited and does not enable screening of dynamic, time-resolved phenotypes. Other approaches first image the full pooled cell library using standard microscopy, then physically retrieve cells with phenotypes of interest by either photoactivation of individual cells followed by FACS selection (*63*–*66*) or by using magnetic manipulation to select cells grown on microraft arrays (*67*). Similar to selection-based screens, subpopulation retrieval approaches measure bulk enrichment of perturbations in a few predefined phenotypic bins, limiting the characterization of cell-level heterogeneity. The need to *a priori* specify the phenotypes of interest also inherently limits the ability to identify unexpected cellular phenotypes. However, physical separation enables subsequent phenotypic characterization of cells from the screen, such as deep molecular profiling of relevant cell states and additional functional assays not compatible with the conditions of the screen.

Optical barcoding methods for image-based pooled screens were initially developed using iterative fluorescence *in situ* hybridization (FISH; *68*–*71*). This approach employs a separate DNA barcode within the perturbation delivery vector, which is optically read-out *in situ* within each individual fixed cell on the imaging plate using FISH-based methods. Optical barcoding thus enables matching of perturbation identity within a pooled experiment to phenotypic data acquired by methods such as immunofluorescence or live-cell imaging on a cell-by-cell basis, providing a

rich matrix of individual genotyped cells by phenotypic features analogous to pooled screens using molecular profiling approaches. However, despite exceptional detection sensitivity not limited by a reverse transcription step, FISH methods employ only limited barcode signal amplification and thus require relatively high optical magnification for detection, which limits screening throughput. Additionally, long barcodes with multiple hybridization sites are required to identify perturbations, necessitating bespoke library cloning methods and random pairing of perturbations and barcodes.

In this thesis, genotype is linked to phenotype within pooled CRISPR screens at single-cell resolution by directly sequencing cellular perturbation identity using *in situ* sequencing-by-synthesis (SBS; Fig. 1.2A). Perturbation identities are deduced from mRNA containing the CRISPR sgRNA sequence itself, enabled by the now-standard CROPseq lentiviral sgRNA vector (*72*), thus simplifying library design and cloning over FISH-based approaches. sgRNA sequences are read out in fixed cells via padlock-based *in situ* sequencing (*73*, *74*), a process involving padlock probe hybridization and gap filling, rolling circle amplification (RCA), and *in situ* SBS (Fig. 1.2B). Earlier optimizations to this protocol improved both the number and brightness of sequencing reads, enabling high-throughput optical pooled screening with perturbations successfully identified for a large fraction of cells when sequenced with low (10X) magnification (*75*). In previous work, >80% of identified sequencing reads exactly matched the designed set of library sequences over 12 cycles of SBS, allowing optical pooled screening with genetic libraries containing thousands of perturbations (*75*).

1.5 Large-scale image-based screening for complex biological phenotypes

Due to the flexibility and throughput of phenotype measurement using microscopy, optical pooled screens are an advantageous method for scalable screening of complex phenotypes. In many ways, this also represents a return to approaches analogous to those of classical genetic screens in model organisms, where the molecular basis of high-level visual phenotypes were successfully identified. This is in contrast to high-dimensional pooled screening approaches based on single-cell sequencing measurements, which are fundamentally limited to measurements of molecular abundance that are a step removed from many relevant cellular functions. However, despite the significant technological progress in developing optical pooled screens, prior published work has only demonstrated image-based pooled screens in human cells at limited scale ($3x10^3$

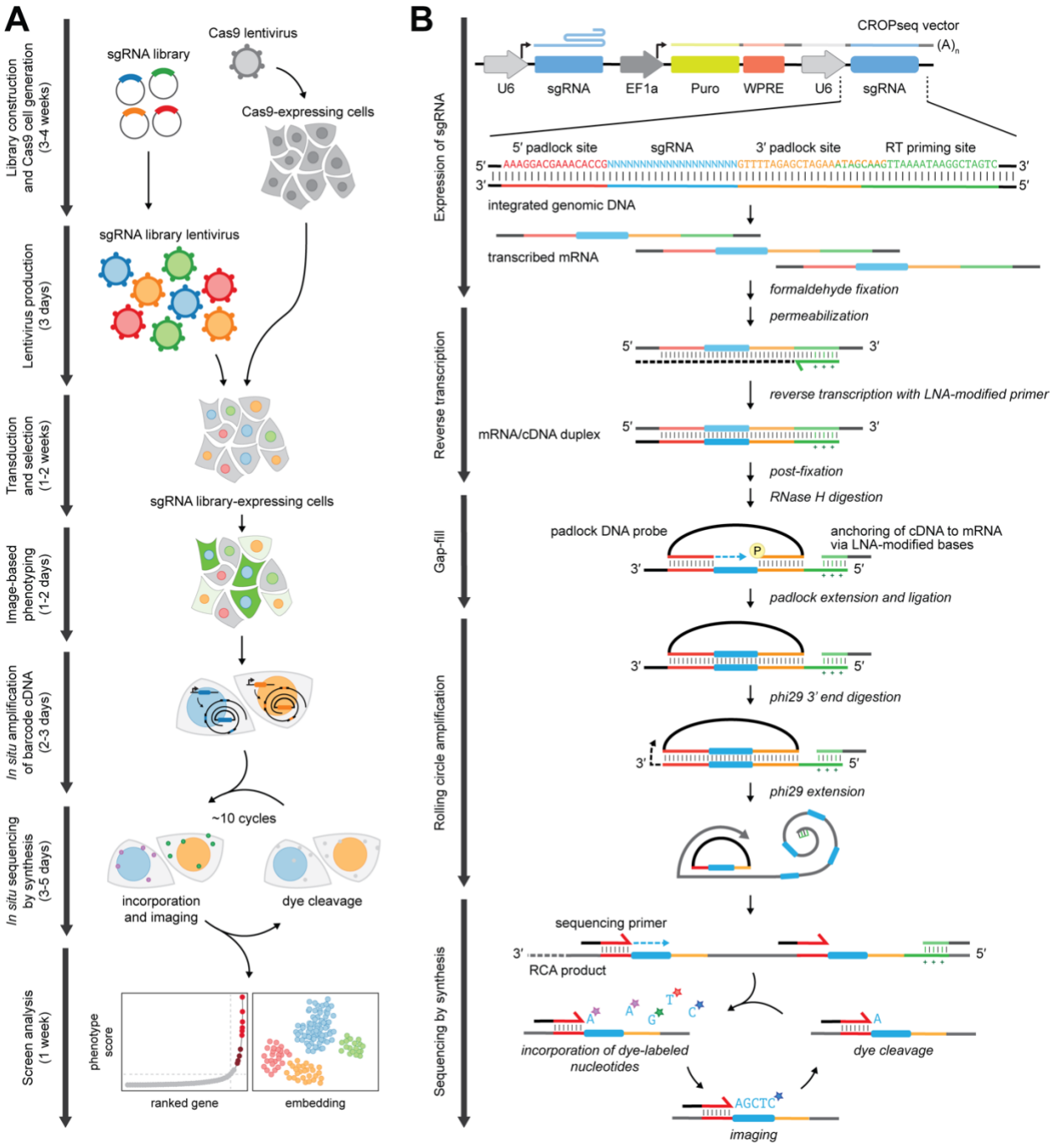

Figure 1.2. Overview of optical pooled screening. (**A**) Experimental workflow. First, a pooled sgRNA library is designed, packaged into lentivirus and delivered into Cas9-expressing cells. A live-cell or fixed-cell imaging assay is performed to generate an optical phenotypic profile of individual cells. The sgRNA spacer sequences in each cell are then amplified *in situ* and read out by SBS, consisting of cycles of dye incorporation, imaging and cleavage. Finally, sgRNA-encoded perturbations are mapped to phenotypic scores at the single-cell level, with candidate genes identified through various statistical approaches. (**B**) Schematic of the *in situ* SBS process. The

sgRNA is expressed as a polyadenylated mRNA transcript from an integrated copy of the CROPseq vector. After fixation and permeabilization, an LNA-modified primer is used to reverse transcribe a cDNA copy of the sgRNA sequence. After glutaraldehyde and formaldehyde post-fixation, the mRNA is digested and a padlock probe is hybridized to cDNA regions flanking the sgRNA sequence. The padlock probe is then extended and ligated to copy the sgRNA sequence into a single-stranded circularized DNA. This circularized DNA serves as a template for RCA with Phi29 polymerase, which produces tandem repeats of the sgRNA spacer sequence. These sequences are read out by successive cycles of SBS.

perturbations at most) and still relatively simple phenotypes (e.g., nuclear translocation of a reporter protein; *75*). Thus, the full potential of image-based pooled screening has yet to be realized.

The work presented in **Chapter 2** demonstrates advances in each of these areas, with efficient image-based pooled screening of more than 10^7 human cells across 20,445 individual perturbations targeting 5,072 genes. For each cell, phenotype images are collected using four stains visualizing distinct biological structures, yielding high-dimensional measurements of cellular morphology, molecular abundance, and spatial organization (~1,000 extracted image features in total) that we combine to gain a global understanding of perturbed cellular phenotypes present in the screen. We also demonstrate the feasibility of image-based pooled screening for dynamic and complex live-cell phenotypes, collecting time-lapse movies of over 400,000 mitotic cell division events across 239 gene targets.

In addition to the technical advances presented in this work, we also leverage the vast amount of matched perturbation genotype and phenotype data to provide new insights for the diverse functions of essential human genes. Identifying the roles of essential genes in specific biological processes is a long-standing challenge across organisms (*4*–*6*), primarily due to the necessary inviability that follows gene perturbation. Here, we precisely control the timing between gene knockout and sample fixation to acquire meaningful phenotype measurements prior to cell death, an alternative to the hypomorph approach used in other organisms (*76*). While many of the gene associations identified in our screen are not readily identifiable from visual inspection of the knockout cell images, quantitative validation of the dataset against prior functional databases

demonstrates that the relationships between the complex phenotypes represent meaningful biological functions. Furthermore, we find new genes involved in ribosome biogenesis, proteasome, and Integrator complex function that were concurrently identified by independent studies. We additionally identify and validate new roles for multiple mRNA processing factors and membrane transporters in the regulation of mitotic function. Together, the work presented in Chapter 2 demonstrates systematic characterization of the contributions of essential genes to complex phenotypes and cellular functions using large-scale optical pooled screening.

Finally, in **Chapter 3** we discuss the future of image-based pooled screening. Applying the presented approach to diverse biological models (such as in **Appendix A**) with various modes of genetic perturbation has a strong potential to build basic understanding of disease processes and identify novel therapeutic targets. In addition to opportunities for technical improvements in optical pooled screening, we discuss how the resulting datasets present new opportunities and challenges for biological discovery.

Chapter 2. The phenotypic landscape of essential human genes

This chapter is adapted from *The phenotypic landscape of essential human genes*, a manuscript under review with the following authors: Luke Funk*, Kuan-Chung Su*, Jimmy Ly, David Feldman, Avtar Singh, Paul C. Blainey, and Iain M. Cheeseman (*equal contributions). An earlier version of this manuscript is available as a preprint on bioRxiv (*77*).

2.1 Abstract

Understanding the basis for cellular growth, proliferation, and function requires determining the contributions of essential genes to diverse cellular processes. Here, we combined pooled CRISPR/Cas9-based functional screening of 5,072 fitness-conferring genes in human cells with microscopy-based visualization of DNA, DNA damage response, actin, and microtubules. Analysis of >31 million individual cells revealed measurable phenotypes for >90% of genes. Using multi-dimensional clustering based on hundreds of quantitative phenotypic parameters, we identified co-functional genes across diverse cellular activities, revealing novel gene functions and associations. Pooled live-cell screening of ~450,000 cell division events for 239 genes further

identified functional contributions to chromosome segregation. Our work establishes a resource detailing the effects of disrupting core cellular processes that defines the functional landscape of essential human genes.

2.2 Introduction

For a human cell to grow, proliferate, and function, it must carry out a variety of essential cellular processes, including transcription, mRNA splicing, translation, vesicle trafficking, proteolysis, DNA replication, and cell division. CRISPR/Cas9-based pooled genetic screens have revolutionized the ability to test the functional requirements for cell growth and proliferation by enabling the potent disruption of thousands of individual genetic elements in single experiments (*21*). However, most current screening approaches, including those based on fluorescence-activated cell sorting (FACS) of cell populations (*78*, *79*), produce a single scalar measurement of barcode enrichment or depletion that summarizes the contributions of each perturbation to cellular phenotypes at the population level. In cellular fitness screens using these approaches, it is thus rarely possible to distinguish essential genes that function in distinct cellular processes. To improve the differentiation of complex phenotypes, recent studies have combined pooled functional genetic screens with multi-dimensional measurements at limited scales, including single-cell profiling of transcriptional cell states (*25*). Defining the specific contributions of essential genes to core cellular processes requires quantitative analysis of complex cellular phenotypes, many of which can be directly visualized using microscopy. Leveraging the power of microscopy, recent work utilized targeted photoactivation of fluorescence in cells exhibiting specific optical phenotypes to enable visual CRISPR screens (*63*, *64*, *66*). However, these approaches similarly produce a single enrichment score for each gene, one predefined phenotype at a time. The ability to interrogate and systematically compare a large and diverse array of cell biological phenotypes simultaneously across thousands of genomic perturbations represents an important unmet goal for functional studies. Here we use optical pooled screening (*1*, *75*) to combine large-scale Cas9-based targeting of essential genes with microscopy and image-based profiling of single-cell resolved cell biological phenotypes at a large scale (Fig. 2.1A).

2.3 A large-scale, image-based pooled CRISPR screen of essential genes

To determine the functional contributions of essential genes in cultured human cells, we first identified a set of fitness-conferring genes based on combined evidence from multiple Cas9- and transposon-based genetic screens (*80*–*88*; Fig. 2.S1A-B; Section 2.13). This approach defined a collection of 5,072 genes that contribute to optimal cellular fitness, although we note that not every gene will be required for fitness in a given cell line. To create a library of CRISPR sgRNAs targeting this gene collection, we selected four sgRNA sequences per target gene from existing sgRNA libraries (*88*–*90*), prioritizing guides with evidence of high on-target efficiency and low off-target activity (Section 2.13.1). In addition, we selected 250 "non-targeting" sgRNAs that lack targets in the human genome as negative controls. Together, this constituted a library of 20,445 total sgRNAs.

We delivered the sgRNA library to HeLa cells containing an integrated, doxycycline-inducible Cas9 construct (*91*) using the CROPseq-puro-v2 lentiviral vector that contains an optimized sgRNA scaffold (*75*, *91*, *92*; Section 2.13.1, 3, 5). Based on trial image-based screens (Fig. 2.S1C) and an analysis of sgRNA depletion from the cell library at 3 and 5 days post-Cas9 induction (Fig. 2.S1D), we defined a time point at 78 hours post-Cas9 induction to maximize phenotype observability. This approach balances the time required for Cas9 activity and protein depletion with negative fitness effects that deplete knockout cells from the population. After fixing the cell population at 78 hours post-Cas9 induction, we amplified the sgRNA sequences in situ as described previously (Fig. 2.1A; *1*, *75*). Following amplification, we stained and imaged cells for DNA (DAPI), DNA damage response (γH2AX; anti-phospho-Ser139 H2AX antibody), microtubules (anti-α-tubulin antibody), and filamentous actin (phalloidin; Fig. 2.1C). These stains were chosen to visualize diverse cell biological behaviors, including nuclear morphology, cell size, DNA damage response, cytoskeletal structures, cell cycle stage, and mitotic chromosome alignment.

Following the completion of phenotype imaging, we performed in situ sequencing-by-synthesis to identify the sgRNA present in each individual cell (Fig. 2.1A, C; Fig. 2.S1E; *1*, *75*), allowing us to directly assess the phenotypic consequences of disrupting each target gene. We extracted 1,084 phenotypic parameters from each individual cell image, including measurements of the intensity and subcellular distribution of each stain, colocalization of stains, and cellular and nuclear size

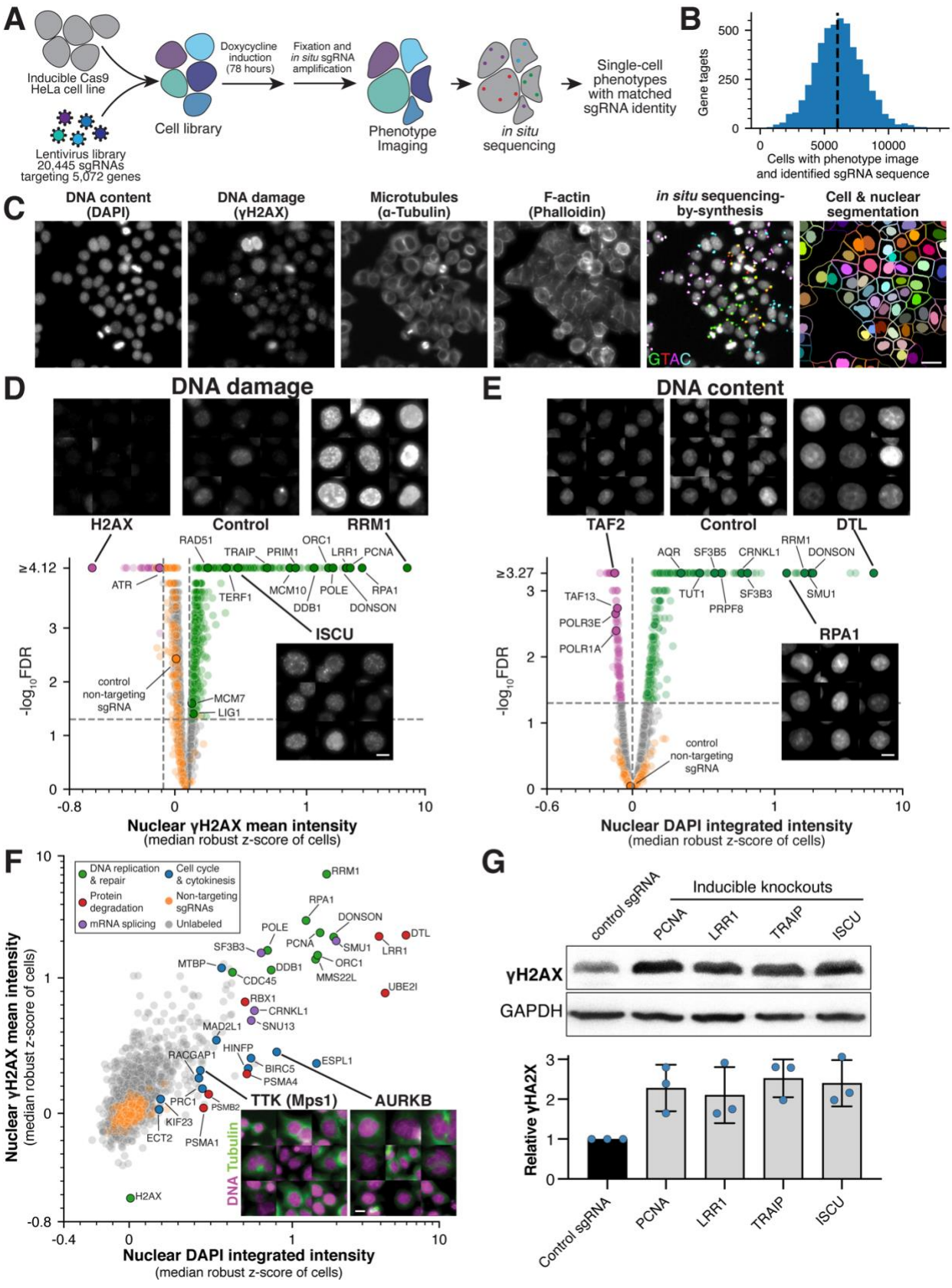
A
Inducible Cas9 HeLa cell line
Lentivirus library 20,445 sgRNAs targeting 5,072 genes
Cell library
Doxycycline induction (78 hours)
Fixation and in situ sgRNA amplification
Phenotype Imaging
in situ sequencing
Single-cell phenotypes with matched sgRNA identity
B
Gene tagets
Cells with phenotype image and identified sgRNA sequence
C
DNA content (DAPI)
DNA damage (γH2AX)
Microtubules (α-Tubulin)
F-actin (Phalloidin)
in situ sequencing-by-synthesis
Cell & nuclear segmentation
GTAC
D
DNA damage
H2AX
Control
RRM1
Nuclear γH2AX mean intensity (median robust z-score of cells)
-log10 FDR
control non-targeting sgRNA
ISCU
E
DNA content
TAF2
Control
DTL
RPA1
Nuclear DAPI integrated intensity (median robust z-score of cells)
F
DNA replication & repair
Cell cycle & cytokinesis
Protein degradation
Non-targeting sgRNAs
mRNA splicing
Unlabeled
Nuclear γH2AX mean intensity (median robust z-score of cells)
Nuclear DAPI integrated intensity (median robust z-score of cells)
TTK (Mps1)
AURKB
DNA Tubulin
G
control sgRNA
Inducible knockouts
PCNA
LRR1
TRAIP
ISCU
γH2AX
GAPDH
Relative γHA2X
Control sgRNA

Figure 2.1. Large-scale image-based pooled CRISPR screen identifies essential genes with roles in genome integrity. (**A**) Experimental workflow for the fixed-cell, image-based pooled CRISPR screen (also see section 2.13.5). (**B**) Histogram showing the number of cells analyzed for each gene target with an acquired phenotype image and single sgRNA sequence mapped *in situ*. (**C**) Example image from the pooled screen showing the phenotype stain channels (DNA, γH2AX, tubulin, and F-actin) together with a matched field-of-view of fluorescent *in situ* sequencing (Laplacian-of-Gaussian filtered) and cell segmentation. Scale bar, 25 μm. (**D**) Volcano plot for mean nuclear γH2AX intensity across gene targets in the screen. Selected images from the screen show γH2AX staining for example cells to highlight specific targets whose knockout results in increased (green) or decreased (magenta) DNA damage relative to random samples of cells expressing targeting sgRNAs (orange; FDR<0.05). The median robust z-score was measured across cells with the same sgRNA, and aggregated to the gene level by taking the median of sgRNAs targeting the same gene (see section 2.13.8). The median robust z-score is plotted on a symmetric log scale (linear between -1 and 1), with effect size thresholds set as the 2.5 and 97.5 percentiles of non-targeting sgRNA scores. Raw P-values were computed by comparing gene targets to bootstrapped null distributions of cells expressing targeting sgRNAs (see section 2.13.8), with false discovery rate (FDR) estimated using the Benjamini-Hochberg procedure. Scale bar, 10 μm. (**E**) Volcano plot as in (D) for integrated nuclear DAPI intensity, along with example images of DAPI staining for gene knockouts that result in increased or decreased DNA content. Raw P-values were computed by comparing gene targets to bootstrapped null distributions of cells expressing non-targeting sgRNAs (see section 2.13.8), followed by FDR estimation as in (D). Scale bar, 10 μm. (**F**) Scatter plot comparing the relationship between DNA damage and DNA content. A subset of gene knockouts result in particularly increased DNA content due to cell division failure; labeled genes are colored by functional category. Example images show tubulin (green) and DNA (magenta) to highlight polyploid and multinucleate cells. Scale bar, 10 μm. (**G**) Western blot (top) and quantification (bottom) confirming the presence of increased DNA damage based on cellular γH2AX levels for genes identified in the image-based pooled screen and PCNA as a positive control. Each sample represents a distinct cell line with inducible Cas9 expression and a stably-expressed sgRNA targeting either the indicated gene or a negative control single copy locus. Blue data points indicate independent replicates. For each sample, γH2AX intensity was referenced to its GAPDH loading control and then normalized by negative control γH2AX relative intensity.

and shape (Section 2.13.7-8). Interphase and mitotic cells showed distinct baseline phenotypes in our dataset, thus we classified cells as mitotic or interphase with a support vector classifier using a subset of extracted phenotype parameters and conducted downstream analyses separately (Section 2.13.8; Fig. 2.S1F). Together, this approach yielded microscopy images, extracted phenotypic measurements, and matched sgRNA identities for 31,884,270 individual cells with a median of 6,119 cells per gene target across each set of four sgRNAs (Fig. 2.1B). Image montages and phenotypic parameters of interphase and mitotic cells are available for exploration through the companion interactive web portal (https://vesuvius.wi.mit.edu/).

2.4 Interphase nuclear phenotypes reveal established and novel regulators of genomic integrity

Maintaining genomic integrity is critical to ensuring proper cellular function, as DNA mutations and chromosome imbalances result in genome instability, misregulated gene expression, cell inviability, and oncogenic cell states. Cells utilize a range of DNA damage-sensors, DNA repair mechanisms, and cell cycle checkpoints to recognize and correct genomic aberrations and protect the genome against spontaneous DNA damage, DNA replication-induced errors, and chromosome segregation defects (*93*). To identify genes that are required for genome integrity, we analyzed nuclear phenotypic parameters in interphase cells from our screen that monitor DNA damage (mean γH2AX nuclear intensity) and total DNA content (integrated DAPI nuclear intensity; Fig. 2.1D-E). We defined summary phenotype scores for each gene target as the median robust z-score of cells relative to the local intermixed population expressing non-targeting control sgRNAs (Section 2.13.8). Gene targets that displayed decreased γH2AX intensity in interphase cells relative to random samples of cells expressing targeting sgRNAs included H2AX itself and ATR, which is involved in directing the γH2AX phosphorylation event (Fig. 2.1D, *93*). Reciprocally, of the 5,072 genes targeted in the screen, we observed 1,258 genes whose disruption resulted in significantly increased γH2AX intensity (Section 2.13.8). The top scoring hits included many factors with known roles in DNA replication (e.g., RRM1, PCNA, DNA Polymerase subunits, Primase subunit 1, DNA Ligase 1, RPA1/2/3, ORC and MCM2-7 subunits), DNA repair (e.g., RAD51, REV3L, ATAD5, DONSON, DTL, DDB1;), and telomere protection (TERF1/2, RTEL1; Fig. 2.1D; *93*, *94*). For most genes, increases in mean γH2AX intensity corresponded to increases in the number of γH2AX foci, although a minority of genes (RRM1/2, RPA1, CHEK1, and POLA1) demonstrated diffuse staining consistent with widespread DNA damage (Fig. 2.S2A-C). In addition to the established players in DNA replication and repair, we

noted that a substantial portion of the genes that demonstrated increased γH2AX intensity are targeted by sgRNAs with multiple genomic target loci and therefore likely exhibit strong Cas9-associated DNA damage. This was particularly pronounced for sgRNAs whose target sites are spread across multiple chromosomes (Fig. 2.S2D).

Many gene targets that caused increased DNA damage also resulted in increased DNA content (Fig. 2.1E-F), including the knockouts of DNA replication and repair factors listed above with the exception of DNA Ligase 1, RAD51, REV3L, and TERF1/2. Gene knockouts with increased γH2AX and DNA content were also enriched for spliceosome components (Fig. 2.S3A-B), consistent with reports that disrupting mRNA splicing results in a DNA damage response (*95*). We observed an overall correlation between DNA damage and total DNA content ($r = 0.62$), although some knockouts displayed strong increases in DNA content but less severe DNA damage (Fig. 2.1F). This includes proteasome 20S core particle subunits and gene targets whose disruption prevents cytokinesis (AURKB, BIRC5, CDCA8, PRC1, KIF23, ECT2; *96*, *97*) or that allow cells to progress through cell division without segregating their chromosomes (ESPL1, TTK/Mps1, MAD2L1; *98*). Targeting each of these cytokinesis and chromosome segregation genes results in more cells with increased DNA content and nuclear area due to tetraploidy or multinucleation (Fig. 2.1F; Fig. 2.S3C-E).

To validate our approach for identifying essential genes involved in genomic integrity, we further investigated selected gene targets that displayed increased DNA damage, including the E3 ubiquitin ligase subunits LRR1 and TRAIP, the mitochondrial iron-sulfur cluster biogenesis gene ISCU. In each case, we generated individual cell lines with inducible Cas9 expression and a single sgRNA targeting the corresponding genes (*91*). Based on Western blotting, we noted a substantial increase in γH2AX levels compared to a control sgRNA with a single genomic target site following ISCU, LRR1, and TRAIP depletion, up to similar γH2AX levels as a PCNA positive control knockout (Fig. 2.1G). The effect of ISCU knockout is consistent with the requirement for iron-sulfur clusters in the enzymatic activity of proteins involved in DNA metabolism (*99*), and LRR1 and TRAIP have recently been reported to play roles in replisome disassembly (*100*). As an additional validation of our approach, we compared our results to previous screens for gene knockouts that result in sensitization or resistance to treatment with genotoxic agents (*101*). Of the 583 previously-identified DNA damage-related genes also present in our screen, we find that 190 (32.6%) also demonstrate either increased or decreased γH2AX staining despite the absence of directly induced DNA damage in our approach. We also found 870 genes with changes in

γH2AX staining that were not identified in the previous genotoxicity response screens. These genes are enriched for spliceosome, proteasome, RNA polymerase, and protein export factors that would be depleted in the untreated negative control population of the genotoxicity enrichment screens (Fig. 2.S3F), highlighting the complementarity of these approaches and the sensitivity of our screen to identify specific phenotypes even for genes that confer strong baseline fitness effects.

Together, this analysis validates our image-based screening strategy to identify diverse players in genome integrity and highlights the importance of multiple genes in DNA replication and repair.

2.5 Identification of essential genes controlling cytoskeletal function

To direct cellular proliferation, structure, organization, and mechanical force production, cells rely on complex and dynamic cytoskeletal networks involving actin and microtubule polymers (*102*, *103*). Our screen measured interphase cytoplasmic features for both filamentous-actin (F-actin) and microtubules, enabling the broad identification of essential genes involved in cytoskeletal processes and cellular organization. An analysis of interphase mean F-actin intensity revealed 460 gene knockouts with decreased intensity relative to non-targeting controls and 899 genes with increased intensity (Fig. 2.2A; Fig. 2.S4A). Among these genes, we identified established factors required for regulating actin assembly and dynamics. For example, knockout of the actin depolymerization and severing factor cofilin (CFL1) or the F-actin capping protein CAPZB, which acts to block actin elongation, resulted in substantially increased actin polymer levels. In contrast, depletion of RHOA or ARHGEF7, which regulate the formation of actin fibers, resulted in strongly decreased actin levels. Although the Arp2/3 complex plays an established role in nucleating actin assembly (*102*), we found that its loss resulted in increased cellular actin intensity. However, this was coupled with a substantial decrease in interphase cell area (Fig. 2.S4B). This suggests that disrupting selected components of the actin cytoskeleton also perturbs cellular adhesion, resulting in reduced cell-substrate contacts and an increase in mean cytoplasmic actin intensity due to altered cell shape. Indeed, we observed a similar phenotype for 251 genes that include the adhesion components RAC1, TLN1, CRKL, ILK, and Integrin subunits (ITGAV, ITGB1, and ITG5; Fig. 2.S4B-C). The remaining 648 genes with increased actin intensity do not show decreased cellular area and are likely to be independent of changes to cell adhesion. The gene target whose loss resulted in the largest increase of mean F-actin intensity in both interphase and mitotic cells

was the E3 ubiquitin ligase KCTD10, along with its partners CUL3 and RBX1 (Fig. 2.2A; Fig. 2.S4D). Recent work implicated KCTD10 in restricting actin assembly during cell migration or developmentally-programmed cell fusion (*104*, *105*), but our analysis suggests a general role for this E3 ubiquitin ligase in regulating actin assembly.

We also identified multiple factors regulating interphase tubulin levels. Mean tubulin intensity was significantly decreased for 492 gene targets when compared to non-targeting controls, including genes encoding tubulin proteins (TUBA1B/C, TUBB, TUBB4B), tubulin-specific chaperones (TBCC/D/E), and factors required for tubulin folding and complex assembly (CCT chaperonins/TRiC complex and prefoldin subunits; Fig. 2.2B). Reciprocally, we observed an increased mean tubulin intensity for 639 knockouts. However, as noted above, cytoplasmic proteins such as actin and tubulin can display increased mean stain intensity under conditions where cell area is reduced due to altered substrate adhesion (Fig. 2.S4B-C, F). Thus, we compared actin and tubulin intensity to identify gene targets that selectively affect one stain (Fig. 2.2C). We observed substantially increased tubulin fluorescence without strong increases in actin intensity for Casein kinase I delta (CSNK1D), which has been suggested to regulate microtubule-associated proteins (*106*), and subunits of the CCR4-NOT complex (CNOT1/4/10/11), which functions in post-transcriptional mRNA regulation (*107*). Overall, 191 genes exhibited increased tubulin intensity without increased actin intensity or decreased cell area. In summary, this analysis of interphase cytoplasmic actin and tubulin intensity identifies the contributions of diverse molecular players for roles in controlling cytoskeletal assembly and dynamics.

2.6 Analysis of morphological phenotypes reveals a tight correspondence between cellular and nuclear size

In addition to measuring stain intensity for each marker, we also measured multiple morphological parameters including nuclear and cellular area. We noted substantial differences in median interphase cell area across the different gene targets, ranging in segmented area from 319 μm^2 to 583 μm^2 (Fig. 2.2D; Fig. 2.S5A). Consistent with a required role for protein production in cell growth, targeting ribosome and ribosome biogenesis genes resulted in substantially reduced cell area (Fig. 2.S5B). In contrast, gene targets with roles in DNA replication and repair, mRNA splicing, and proteasome function displayed increased cell areas (Fig. 2.S5B), suggesting continued cell growth in the absence of further division. Using individual inducible knockout cell lines, we were able to confirm these substantial changes in cell size for DTL and DONSON gene

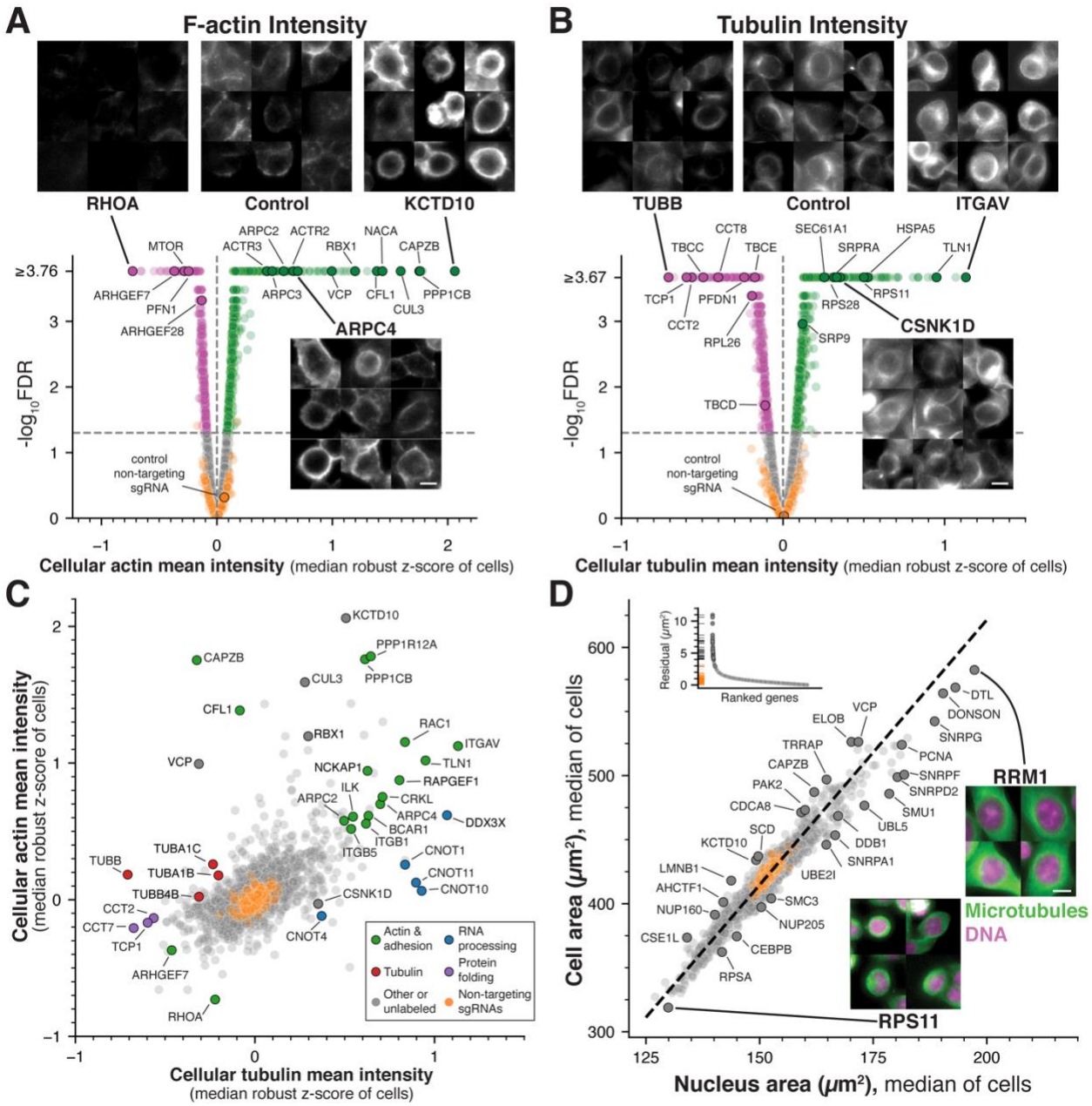

Figure 2.2. Identification of essential genes regulating cytoskeletal structures and cellular organization. (**A**) Volcano plot for mean cellular F-actin (phalloidin) intensity across gene targets in the screen. Selected images from the screen show phalloidin staining for example cells to highlight specific targets that result in increased (green) or decreased (magenta) cellular actin levels relative to non-targeting control sgRNAs (orange; FDR<0.05). The median robust z-score was measured across cells with the same sgRNA, and aggregated to the gene level by taking the median of sgRNAs targeting the same gene (see section 2.13.8). Raw P-values were computed by comparing gene targets to bootstrapped null distributions of cells expressing non-targeting sgRNAs (see section 2.13.8), with false discovery rate (FDR) estimated using the Benjamini-

Hochberg procedure. (**B**) Volcano plot as in (A) for mean cellular tubulin intensity, along with example images of tubulin staining for gene knockouts that result in increased or decreased tubulin levels. (**C**) Scatter plot comparing the relationship between actin and tubulin stain intensity highlights gene targets that selectively affect one cytoskeletal element. A subset of knockouts impact both actin and tubulin mean intensity measurements by disrupting cellular adhesion and decreasing the segmented cell area (see also Fig. S4B, C, F). Labeled genes are colored by functional category. (**D**) Scatter plot showing the comparison between cellular and nuclear area across gene targets. These morphological features are highly correlated across scales and conditions (r = 0.96). The median of area measurements was computed across cells with the same sgRNA and aggregated to the gene level by taking the median of sgRNAs targeting the same gene. Orthogonal regression was performed to identify gene targets that result in an altered nuclear:cytoplasmic area ratio (dotted line). Labeled genes are also highlighted in the distribution of regression residuals (inset). Example images display DNA (magenta) and tubulin (green) staining for gene targets that result in increased or decreased cell and nuclear size. Scale bars, 10 µm.

knockouts (Fig. 2.S5C-D). Nuclear size similarly varied widely across gene targets, including increased nuclear area for five genes previously identified in a prior photoactivation-based CRISPRi imaging screen (AURKB, CDCA8, FBXO5, TICRR, and RAD51; *64*).

Strikingly, we observed a strong correlation between cell area and nuclear area across all tested gene targets (r = 0.96; Fig. 2.2D). Prior work has suggested that cells actively regulate the size ratio between their nucleus and cytoplasm (*108*). Our analysis demonstrates that, across a wide range of cell sizes and functional perturbations, this relationship is closely maintained. Although we observed a clear relationship between nuclear size and cell size, we identified a limited number of gene targets whose depletion disrupted this coordinated scaling (Fig. 2.2D). A subset of gene targets displayed abnormally large nuclei for their given cell size, including RNA splicing factors (e.g., SMU1 and UBL5), the nuclear pore complex member NUP205, and DNA replication factors (e.g, PCNA, DDB1, RRM1). We also identified gene knockouts that displayed decreased relative nuclear size, including lamin B1 (LMNB1), the nucleocytoplasmic transport protein CSE1L, and the nuclear pore components NUP160 and AHCTF1, consistent with roles in controlling nuclear integrity and function. Together, this analysis demonstrates that cell biological parameters from a

large-scale screen can be used to provide insights into the control of cellular morphology and organization.

2.7 Phenotypic clustering of interphase cellular parameters defines co-functional genes

We next sought to take advantage of the full range of identifiable phenotypes in the rich image data from our screen to reveal additional gene activities required for cellular function. To define the phenotypic landscape of essential genes, we selected and aggregated 472 non-redundant parameters extracted from each interphase cell image to create a summary phenotypic profile for each gene target (Section 2.13.8). We then visualized these high-dimensional profiles using the PHATE algorithm (*109*) and performed clustering to identify genes with similar interphase phenotypes (Section 2.13.8; Fig. 2.3A; Fig. 2.S6A-C). Based on a comparison of knockout phenotypes to non-targeting guides, 4,665 of the 5,072 gene targets in our screen display a measurable interphase phenotype (Fig. 2.S6A-B). Of the remaining 407 gene knockouts, only 55 genes displayed strong fitness effects at 5 days post-Cas9 induction based on sgRNA depletion from our library (Fig. 2.S6B). Thus, the 352 genes without a measurable phenotype or fitness effect are likely not required for cellular fitness in HeLa cells at the tested time point following Cas9 induction.

To evaluate our PHATE-based clustering strategy, we compared our results with databases of protein-protein interactions, co-essential gene pairs, and curated protein complexes (*110–112*). We found that the relative similarity of gene phenotype profiles from our screen was increased among gene pairs identified as functionally-related in each external database (Fig. 2.3B; Section 2.13.8). Based on a minimal set of CORUM complexes, our clustering results accurately recall 19% of gene pairs from annotated protein complexes while maintaining a precision of 43% (Fig. 2.S7; Section 2.13.8). This increases to a recall of 58% and precision of 81% when restricting this analysis to genes within the top 10% of summarized phenotype strength. Overall, 64 of 222 clusters were enriched for at least one CORUM protein complex, and 55 clusters were enriched for at least one KEGG pathway. Of the 166 tested KEGG pathways, 58 were enriched in at least one identified cluster, while the remaining functional categories not enriched in any cluster included many metabolic, biosynthetic, cellular signaling, and cell death pathways (apoptosis, ferroptosis, necroptosis), which correspond to processes not explicitly evaluated in our approach.

Consistent with our unbiased evaluation of the PHATE clustering analysis, we noted clear functional relationships between the genes within a given cluster, allowing us to identify major clusters primarily composed of gene targets with roles in transcription, RNA processing, translation, protein degradation, DNA replication and damage response, cell cycle control, or other core cellular processes (Fig. 2.3A, C-F; Fig. 2.S8; https://vesuvius.wi.mit.edu/). Strikingly, the clustering behaviors also allowed us to distinguish high-resolution functional sub-categories within each cellular process. For example, despite a shared role in translation, we identified separable clusters containing established 40S ribosome subunits (cluster 66), 60S ribosome subunits (cluster 23), tRNA ligases and eIF2 translation initiation subunits (cluster 14), distinct clusters for factors involved in 40S ribosome biogenesis (cluster 136) and 60S ribosome biogenesis (cluster 15), and several others which included nucleolar proteins, RNA helicases, and additional factors involved in translation initiation or ribosome biogenesis (clusters 21, 112, 203, and 216; Fig. 2.3C, D). Knockouts for the genes within each of these clusters resulted in reduced nuclear and cellular areas, but displayed differences in other cellular phenotypic parameters, such as actin and tubulin staining intensities, enabling distinction among these functional sub-categories (Fig. 2.3D).

Similarly, we identified multiple clusters containing 26S proteasome subunits with phenotypic differences that allowed segregation of 20S core particle subunits (167) from the 19S regulatory particle ATPase (106) and non-ATPase components (213), as well as a cluster containing components of the COP9 Signalosome (200), which controls ubiquitin-dependent processes (Fig. 2.S8A). Finally, we observed multiple distinct clusters for the core transcriptional machinery, including TFIID subunits (192), RNA Polymerase II subunits (199), two clusters comprised of Mediator complex subunits, General Transcription Factors (GTFs), and mRNA export factors (8 and 60), and separate clusters containing RNA Polymerase I (155) and RNA Polymerase III (45) components (Fig. 2.3E-F). Interestingly, although our cellular imaging did not include membrane-targeted markers for cellular organelles, such as the Golgi or Endoplasmic Reticulum, we identified multiple distinct clusters comprised of vesicle trafficking components. For example, we identified a cluster (201) containing the coatamer subunits (COPA/B1/B2/G1/Z1, ARCN1), SNAP proteins (NAPA and GOSR2), and cholesterol biosynthesis proteins HMGCS1 and HMGCR, a second cluster (54) containing signal recognition particle (SRP19/54), ESCRT proteins (UBAP1, CHMP6, and VPS28), clathrin, and additional vesicle trafficking proteins, as well as a cluster (140) containing the exocyst complex (EXOC1/4/5/7) and glycosylation machinery (Fig. 2.S9A). This suggests that specific cell morphological changes resulting indirectly from perturbing vesicle

trafficking and organelle function can be detected by our extracted image parameters, beyond what is easily distinguishable by visual inspection.

In summary, the phenotypic clustering using quantitative parameters extracted from cell images provides a fine-grained picture of the distinct functional contributions of specific protein sub-complexes to core cellular processes.

2.8 Phenotypic clustering provides novel insights into gene functions and pathway relationships

The coherent phenotypic clustering of known co-functional gene targets provides potential predictions of cellular function for additional co-clustering genes. For example, our interphase phenotypic clustering revealed similarities between knockouts of the key signaling proteins KRAS and BRAF with multiple mitochondrial components, such as mitochondrial ribosome subunits and proteins involved in mitochondrial respiration (NADH dehydrogenase and Cytochrome, cluster 149; Fig. 2.S9C), consistent with roles for KRAS and BRAF signaling in maintaining normal metabolic homeostasis (*113–115*). In independent experiments, we additionally identified overall disruption of mitochondrial content in KRAS and BRAF knockout cells, similar to knockouts of co-clustering mitochondrial factors (Fig. 2.S9D). Similarly, although several clusters containing transcriptional regulators exhibited related phenotypes, we identified a cluster (121) with a distinct phenotypic profile that contains both of the master regulator Myc and Max transcription factors, along with multiple other transcriptional regulators (FOXN1, ILF3, SP2, ZBTB11, nuclear respiratory factor 2 genes GABPA and GABPB1), chromatin remodeling factors (ZMYND8, H3K36 methyltransferase SETD2), and E3 ubiquitin ligase components (KEAP1, DDA1; Fig. 2.S9E). This clustering suggests that these factors may either be specifically required for Myc expression (as is the case for ZMYND8, see ref. *116*) or work with Myc to promote downstream expression at its target promoters. By analyzing Myc mRNA expression and protein levels in individual gene knockouts from this cluster, we confirmed a role for SETD2 in regulating Myc expression (Fig. 2.S9F-G).

Our interphase clustering analysis further implicated poorly characterized gene targets in specific cellular activities. For example, we nominated C1orf131 as a putative nucleolar component involved in ribosome biogenesis based on its membership in cluster 21 (Fig. 2.3D), a prediction that was recently confirmed by others (*117*). Similarly, AKIRIN2 clustered with the 20S core

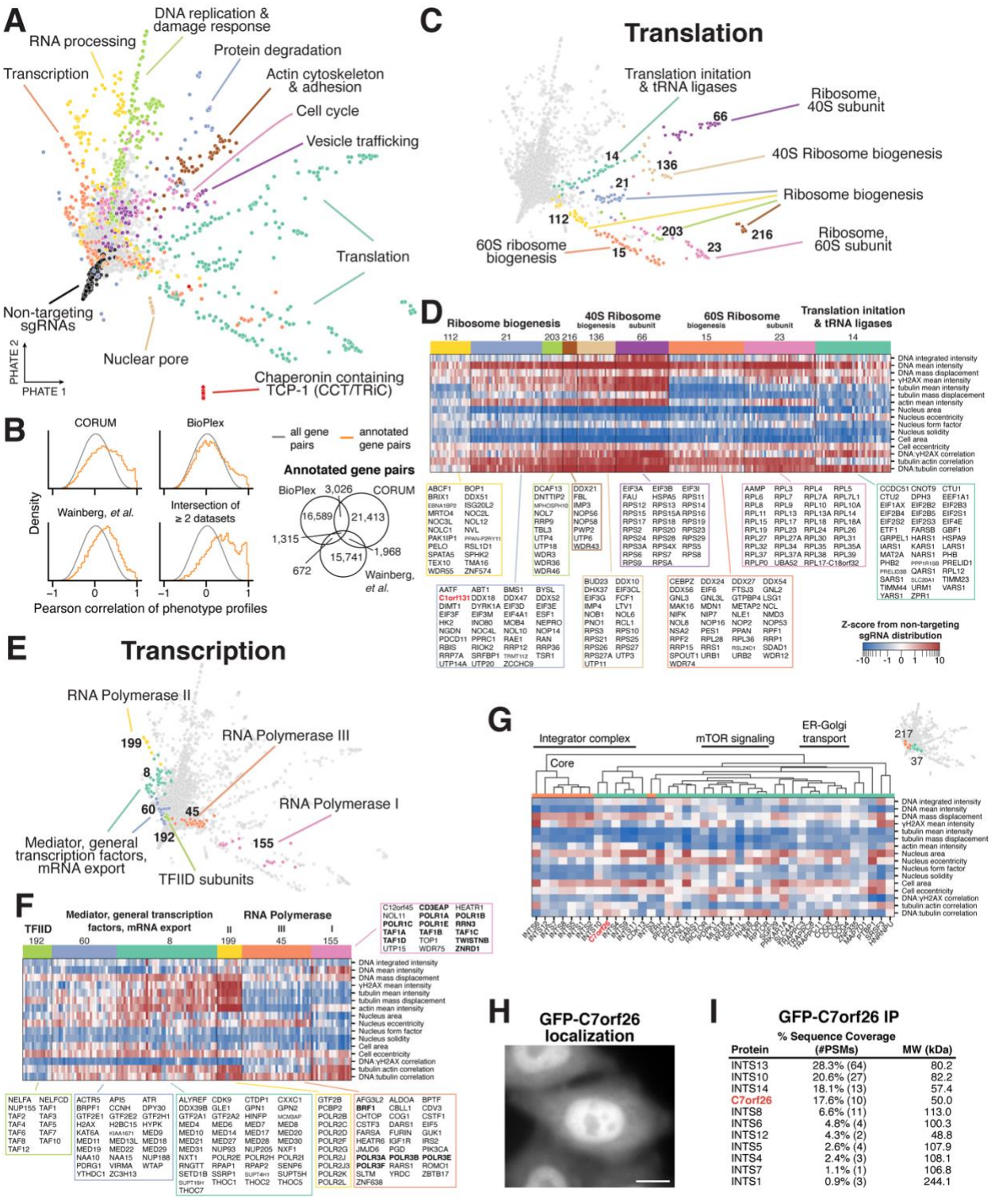

Figure 2.3. Clustering of multi-dimensional interphase phenotypes reveals co-functional essential genes. (**A**) Two-dimensional representation of the interphase phenotype landscape of gene targets in the primary screen visualized using PHATE (*39*) with hundreds of summary phenotype parameters, and then clustered to form groups of genes with similar phenotypes (see

section 2.13.8). Each dot represents a single gene target, colored corresponding to the indicated functional category of grouped clusters. (**B**) Distributions of the Pearson correlation between image-based interphase phenotype profiles from the primary screen for all gene pairs (gray) or gene pairs annotated as co-functional in the labeled external databases (orange), indicating increased phenotype similarity between known co-functional genes across the full dataset. (**C**) Individual clusters of genes related to translation from (A) identify fine-grained sub-categories of gene function in ribosome biogenesis, translation initiation, and individual ribosome subunits. Functional descriptions of labeled cluster numbers summarize the roles of the contained gene targets. (**D**) Heat map of interphase knockout phenotypes corresponding to the translation clusters in (C) for a manually-selected subset of phenotype parameters (see section 2.13.8). This highlights the phenotypic similarity of gene targets within functionally-coherent clusters, but clear distinctions between separate clusters despite broadly related roles in translation. All genes from each cluster are listed. Parameters are presented as z-scores from the distribution of non-targeting sgRNAs, visualized on a symmetric log scale (linear between -1 and 1). (**E**) Individual clusters of genes related to transcription from (A) indicate separate clusters for components of each type of RNA polymerase, TFIID, and related complexes. (**F**) Heat map as in (D) corresponding to the clusters in (E) highlighting the phenotypic similarities and differences that define each cluster of genes with transcriptional functions. (**G**) Heat map as in (D) of interphase clusters 37 and 217, demonstrating the phenotypic similarity between C7orf26 knockouts with those of the Integrator complex. Hierarchical clustering (top) within these clusters using the Pearson correlation of PCA-projected phenotype profiles (see section 2.13.8) indicates particularly strong similarities between C7orf26 (red) and INTS10, and predicts that the uncharacterized gene C7orf26 is co-functional with established Integrator subunits. (**H**) Fluorescent image of human cells expressing GFP-C7orf26 demonstrates nuclear localization. Scale bar, 10 μm. (**I**) Mass spectrometry from an immunoprecipitation of GFP-C7orf26 in human cells relative to controls indicates that C7orf26 associates with subunits of the Integrator complex.

particle proteasome subunits (cluster 167, Fig. 2.S8A), and was recently described as a required proteasome nuclear import factor (*118*). In addition, HNRNPD was present in cluster 197 together with METTL3 and METTL14 (Fig. 2.S9B), which form the core heterodimer that writes m6A mRNA modifications, consistent with the emerging role of m6A modifications in promoting HNRNPD associations with mRNA (*119*).

We also identified the uncharacterized gene C7orf26 in cluster 37 as displaying phenotypes closely related to those observed in knockouts of known subunits of the Integrator complex, an RNA endonuclease involved in RNA processing (*120*) present in clusters 37 and 217 (Fig. 2.3G). C7orf26 and Integrator complex knockouts resulted in reduced interphase tubulin intensity without corresponding changes in actin intensity. To evaluate this co-clustering relationship, we generated a cell line stably expressing GFP-C7orf26. This GFP-C7orf26 fusion localized to the nucleus (Fig. 2.3H), consistent with Integrator complex localization (*121*). Using affinity purifications coupled to mass spectrometry, GFP-C7orf26 pull-downs specifically isolated multiple Integrator complex subunits, with particularly robust levels of INTS13, INTS10, and INTS14 (Fig. 2.3I; also see ref. *122*). This is consistent with the strong phenotype similarity in our screen between C7orf26 and INTS10 knockouts (Fig. 2.3G). INTS10, INTS13, and INTS14 were recently shown to comprise a functional subunit of the Integrator complex that associates the cleavage module with target RNA (*123*). Our data thus suggests C7orf26 may interact with this sub-complex, consistent with concurrent studies (*124*, *125*). Thus, the phenotypic clustering of this dataset identifies established interacting partners and provides predictive insights to identify novel associations and co-functional players across key cell biological processes with the potential for many additional discoveries.

2.9 Analysis of mitotic phenotypes identifies requirements for proper cell division

We next analyzed the phenotypes observed in mitotic cells for each gene target. In total, 2.6% of the cells visualized in our microscopy-based screen were present in the mitotic phase of the cell cycle (median of 157 mitotic cells per gene target). In the presence of mitotic errors, cells activate the spindle assembly checkpoint and arrest in mitosis (*98*). Therefore, an increased fraction of obtained cell images that are identified as mitotic for a given gene (i.e., mitotic index) can reflect a mitotic disruption (Fig. 2.4A). We measured an increased mitotic index for gene knockouts targeting established components of the kinetochore and mitotic spindle. In contrast, we observed a reduced mitotic index for components of the spindle assembly checkpoint, including MAD2L1, BUB1B, and TTK (Mps1).

Similar to our analysis of interphase cells, we next selected 876 non-redundant measurements from the extracted image parameters of mitotic knockout cells, including the overlap of tubulin

and DNA staining as a measure of mitotic chromosome alignment, and clustered gene targets with similar phenotype profiles (Fig. 2.4B-C; Fig. 2.S6D-F; Section 2.13.8). In addition, we conducted a manual visual analysis for each gene, with two individuals blindly and independently scoring image montages for the presence of mitotic defects. Overall, we found a strong correspondence between the automated and manual scoring (Fig. 2.S10A-B). From our computational analysis, we identified multiple mitotic clusters with functionally-related genes (Fig. 2.4B-C), despite decreased cell counts and increased baseline morphological heterogeneity as compared to interphase cells. For example, we observed close clustering of CKAP5 and the entire Augmin complex (mitotic cluster M214), tubulin subunits with tubulin folding factors and the CCT chaperonin (M205), DNA replication factors (M34), factors required for chromosome alignment including kinetochore components (M11), and spindle and centrosome components (M109). In addition, we identified clusters for gene targets with established roles in mRNA splicing (M6 and M33), proteasome function (M88), and ribosome function (M0, M17, and M21) indicating that mitotic phenotype parameters are able to distinguish these functional categories. We found that this high-dimensional computational analysis provided a complementary but distinct measurement of mitotic phenotypes as compared to mitotic index (Fig. 2.4A). Visual analysis of cell image montages further allowed us to distinguish phenotypic clusters and individual gene targets for their specific roles during mitosis (Fig. 2.4C). For example, we detected reduced microtubule density following depletion of the tubulin chaperone TBCC, chromosome mis-alignment following knockout of kinetochore components and additional factors (e.g., CENPC, SKA1, and the spliceosome gene SMU1), monopolar spindles associated with the targeting of KIF11 or PLK4, and short mitotic spindles in knockouts of CKAP5 or Augmin subunits (e.g., HAUS6; Fig. 2.4C).

To evaluate our analysis of mitotic phenotypes, we additionally compared our findings to those from MitoCheck (*16*), a prior genome-wide siRNA-based arrayed microscopy screen for mitotic phenotypes in HeLa cells. Of the 293 identified genes from MitoCheck that were also present in our screen, we found that 79 demonstrated an aberrant mitotic index or measureable image-based mitotic phenotype, while 70 additional genes scored significantly in at least one interphase phenotype category consistent with downstream consequences of mitotic defects. Conversely, we identified 799 genes with mitotic phenotypes in our screen that were not identified in MitoCheck, which are enriched for proteasome, cell cycle, and DNA replication genes among other relevant pathways (Fig. 2.S10C). The mitotic phenotypes uniquely observed in our study also include many canonical mitotic regulators such as kinetochore components (e.g, CENPC,

NDC80, SPC24/25, SKA1/2/3), Augmin complex subunits (HAUS1/2/5-8), and factors involved in centrosome function (CEP85, PLK4, STIL, SASS6, NEDD1, TUBGCP2/3/4/6). Our approach thus represents a significant improvement over prior systematic approaches to identifying mitotic regulators, enabled by Cas9-based gene perturbation and pooled phenotype acquisition.

In addition to established mitotic players, predicted roles in mitosis also emerged for poorly characterized genes based on co-clustering with well-defined mitotic functions. In particular, we found that ZNF335 clustered with spindle proteins (including TACC3 and TPX2), gamma-tubulin complex proteins (TUBGCP2/3/6), and proteasome non-ATPase (PSMD) subunits (cluster M109; Fig. 2.4C). Examination of mitotic cells from the screen targeting ZNF335 and co-clustered gene targets suggested the presence of spindle defects (Fig. 2.4C). We confirmed this phenotype by generating an inducible knockout cell line for ZNF335, which displayed a substantially increased proportion of cells with monopolar spindles (Fig. 2.4D) and a reduction in centriole numbers (Fig. 2.4E). GFP-ZNF335 localizes to the nucleus in interphase cells (Fig. 2.S10D), but did not display localization to the spindle or centrioles and did not associate with established mitotic factors in immunoprecipitation experiments (not shown). RNA-sequencing analysis of gene expression in ZNF335 knockouts also did not show changes in established spindle factors, but revealed a significant decrease in the expression of the non-ATPase 1 proteasome subunit (PSMD1) and a subunit of the gamma-tubulin ring complex (TUBGCP6; Fig. 2.4F), which both displayed similar monopolar spindle phenotypes in our screen (Fig. 2.4C). Together, this suggests a likely role for ZNF335 in centrosome and spindle function through regulating the expression of specific proteasome or centrosome factors, and provides a potential explanation for why ZNF335 mutations are observed in human microcephaly (*126*), as is the case for many centrosome components (*127*). In addition, we identified dozens of genes with measurable mitotic phenotypes, but less clear phenotypic clustering, that have not been implicated previously as having roles in cell division. We investigate these genes further in a targeted secondary screen below. In summary, this analysis demonstrates the utility of pooled large-scale image-based screening to identify complex interphase and mitotic phenotypes.

2.10 A pooled live-cell imaging-based screen for mitotic defects

Based on the large number of genes with unexpected mitotic phenotypes and the power of microscopy-based approaches to directly visualize these phenotypes in detail (*16*, *128*), we performed a secondary pooled live-cell screen to analyze these factors further. First, we defined

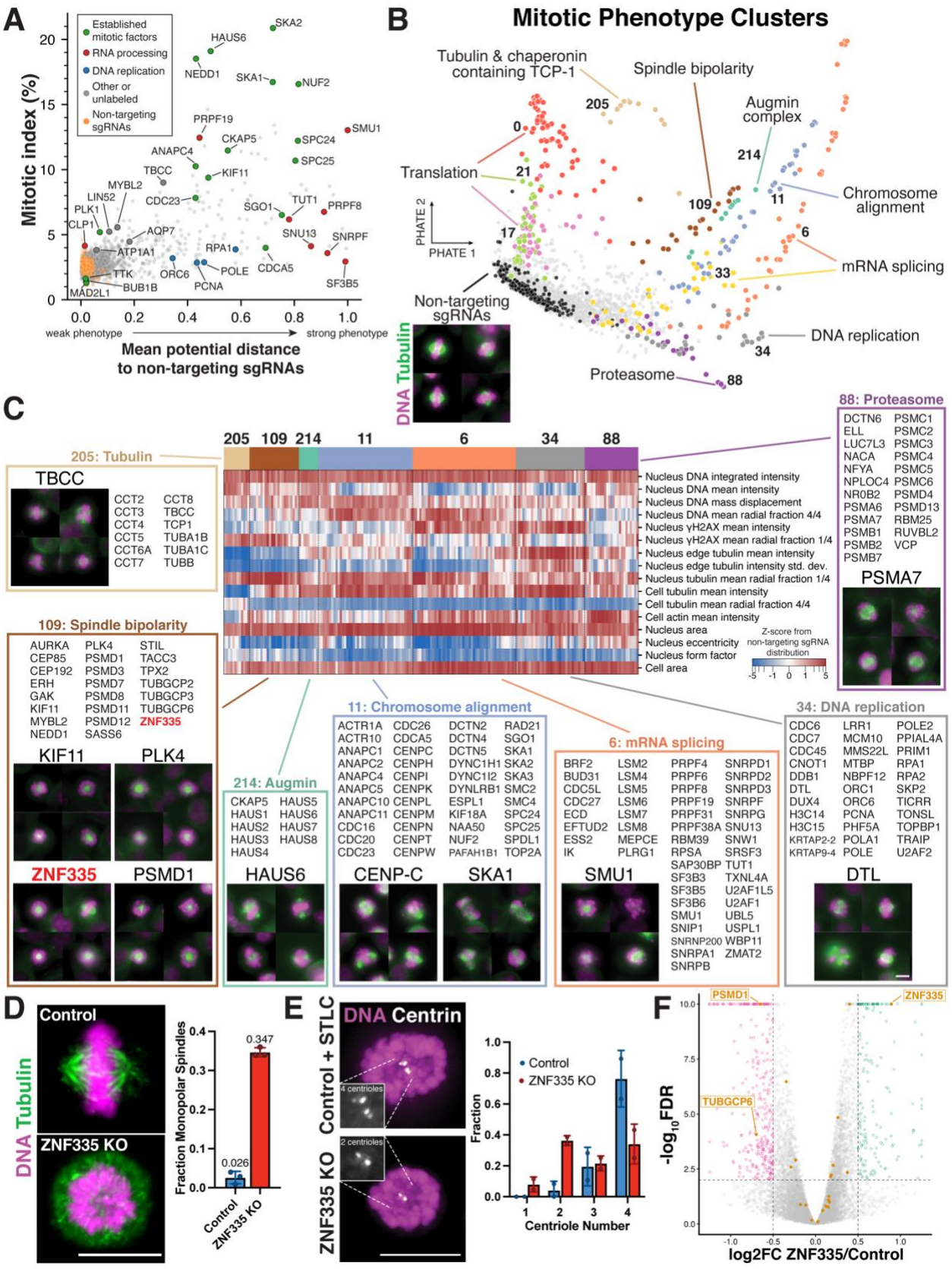
A
Established mitotic factors
RNA processing
DNA replication
Other or unlabeled
Non-targeting sgRNAs
Mitotic index (%)
Mean potential distance to non-targeting sgRNAs
weak phenotype
strong phenotype
B
Mitotic Phenotype Clusters
Tubulin & chaperonin containing TCP-1
Spindle bipolarity
Augmin complex
Translation
Chromosome alignment
mRNA splicing
DNA replication
Proteasome
Non-targeting sgRNAs
DNA Tubulin
C
205: Tubulin
109: Spindle bipolarity
214: Augmin
11: Chromosome alignment
6: mRNA splicing
34: DNA replication
88: Proteasome
Z-score from non-targeting sgRNA distribution
D
Control
ZNF335 KO
DNA Tubulin
Fraction Monopolar Spindles
E
DNA Centrin
Control + STLC
ZNF335 KO
4 centrioles
2 centrioles
Fraction
Centriole Number
F
PSMD1
ZNF335
TUBGCP6
-log10 FDR
log2FC ZNF335/Control

Figure 2.4. Mitotic phenotypes uncover essential genes required for cell division. (**A**) Scatter plot showing mitotic index (proportion of mitotic cells) of the imaged cell population for each gene target compared to a summary score of image-based mitotic phenotype strength computed by PHATE (*39*) high-dimensional analysis. The summary phenotype score is the mean PHATE potential distance to non-targeting control sgRNAs for each gene, normalized between 0 and 1. This highlights the complementary but distinct information provided by computational analysis of image-based mitotic phenotypes. Labeled gene targets are colored by functional category and highlight factors with increased cell division defects. (**B**) Two-dimensional representation of the mitotic phenotype landscape of gene targets in the primary screen visualized using PHATE with hundreds of summary phenotype parameters, and then clustered to form groups of genes with similar phenotypes (see section 2.13.8). Each dot represents a single gene target, colored corresponding to the indicated cluster. Functional descriptions of clusters summarize the roles of the contained gene targets. (**C**) Heat map of mitotic knockout phenotypes corresponding to the clusters in (B) for a manually-selected subset of phenotype parameters (see section 2.13.8). Also shown are example images of mitotic cells from gene targets in the identified clusters, visualizing DNA (magenta) and microtubules (green). Scale bar, 10 μm. This highlights the phenotypic similarity of gene targets within clusters and the ability to separate distinct mitotic functions. Furthermore, this analysis implicated the role of ZNF335 in mitotic spindle bipolarity. All gene targets from selected clusters are listed. Parameters are presented as z-scores from the distribution of non-targeting sgRNAs, visualized on a symmetric log scale (linear between -1 and 1). (**D**) Left, immunofluorescence images showing individual cell lines stably expressing a single control sgRNA or a sgRNA targeting ZNF335. Right, bar plot of the corresponding fraction of mitotic cells with monopolar spindles; each data point represents one experiment with >100 cells. This demonstrates the reproducible strong effect of knocking out ZNF335 on spindle assembly. Images are deconvolved maximum intensity projections of fixed cells stained for microtubules (anti-alpha-tubulin) and DNA (Hoechst). (**E**) Example images (left) of DNA (Hoechst, magenta) and Centrin (grayscale) stains from monopolar ZNF335 knockout cells along with quantification of reduced centriole numbers (right) compared to monopolar control cells generated by STLC treatment (n>88 cells per condition). Insets show magnified regions. Scale bars, 10 μm. (**F**) Volcano plot of differential gene expression following ZNF335 knockout. Yellow data points represent genes co-clustering with ZNF335 in the image-based screen in mitotic cluster 109 (C). Of the genes in this cluster, PSMD1 and TUBGCP6 are significantly decreased in ZNF335 KO cells. 296 genes are downregulated (magenta) and 177 genes are upregulated (green) in ZNF335 KO cells (FDR < 0.01, $\log_2$ effect size > 0.5).

a list of 228 genes with unexpected mitotic phenotypes, and additionally selected 11 positive control genes with established roles in diverse mitotic processes. We generated a lentiviral library of Cas9 guides targeting these genes, with 2 sgRNAs per gene and 50 non-targeting sgRNAs (526 total sgRNAs; Section 2.13.1). The sgRNA library was transduced into a HeLa cell line containing doxycycline-inducible Cas9 and a constitutively-expressed H2B-mCherry fusion to visualize chromatin (Fig. 2.5A). We conducted time-lapse imaging of the pooled cell population for 24 hours with time points at 10 minute intervals, starting after either 48 or 72 hours of Cas9 expression in separate time courses. Following the acquisition of the time-lapse images, we immediately fixed the cell population and amplified the sgRNA sequences in situ to identify the gene targeted in each cell, as described for the fixed-cell screen (Fig. 2.5A; Section 2.13.6). After tracking cell lineages through each time course and using a support vector classifier to identify mitotic cells, we obtained time-lapse movies for 451,434 total cell division events, with a median of 1,381 division events per gene target (Fig. 2.S11A-C; Section 2.13.10). This enabled us to generate time-lapse montages of tracked cells for each gene target, with each cell temporally aligned to mitotic entry (Fig. 2.5B-C; Fig. 2.S11D; https://vesuvius.wi.mit.edu/).

Using this automated live-cell analysis of cell division, we calculated the mean duration of division events for each gene target, as well as the fraction of cells that enter mitosis during the time course (Fig. 2.5B). As expected based on the presence of mitotic defects in the primary screen, 197 of 239 tested gene knockouts, including 10 of 11 positive controls, displayed altered mitotic duration or entry relative to non-targeting controls in at least one time course (Fig. 2.5B; Fig. 2.S12). The spindle assembly checkpoint component BUB1B demonstrated decreased mitotic duration, highlighting the ability to identify either mitotic delay or acceleration. ANLN, the positive control not exhibiting a strong effect, is required for cytokinesis and thus its loss may not significantly affect the measured phenotypes. We were also able to distinguish gene targets with established or predicted roles in DNA replication or repair (e.g., DTL, LRR1, TICRR, MMS22L; *93*, *100*, *129–131*), based on their increased mitotic duration but reduced fraction of cells entering mitosis (Fig. 2.5B), indicative of defective mitotic entry. We next visually inspected each time-lapse montage for the presence of mitotic phenotypes, including lagging chromosomes or delayed chromosome alignment (Fig. 2.5C; Fig. 2.S11D). From these observed phenotypes, we selected 28 genes of interest with chromosome alignment defects to conduct targeted downstream analyses. In each case, we generated individual cell lines with a single sgRNA targeting the corresponding gene and conducted both fixed- and live-cell microscopy to identify phenotypes for each individual gene knockout at higher spatial and temporal resolution (Fig. 2.5D; Fig. 2.S13A;

Fig. 2.S14E). All of the 28 selected gene targets displayed clear defects in chromosome alignment or segregation. Of these genes, a subset displayed defects in bipolar spindle assembly, including short spindles (TCP1). We were also able to distinguish gene targets with roles in DNA replication or DNA damage (e.g., LRR1, TRAIP, ISCU, TICRR, MMS22L), which resulted in multipolar spindles or misaligned chromosomes (Fig. 2.S13A). Unexpectedly, amongst the gene targets whose knockouts resulted in misaligned chromosomes, we identified two membrane-bound transporters - the plasma membrane-localized aquaporin AQP7 and the sodium/potassium-transporting ATPase ATP1A1 (Fig. 2.5D). Notably, AQP7 is the only aquaporin that is broadly essential for cellular viability in the DepMap database (*82*). Individual AQP7 and ATP1A1 knockout cell lines both displayed a reproducible delay in chromosome alignment and an extended mitotic duration (Fig. 2.5E-F), but we did not observe defects in bipolar spindle assembly (Fig. 2.5D) or kinetochore assembly (Fig. 2.S13B-C). As both membrane transporters are involved in maintaining a proper intracellular osmotic environment, we tested the effect of treating cells with 300 Da polyethylene glycol (PEG 300), which is known to create hyperosmotic stress (*132*). PEG300 treatment of control cells resulted in a qualitatively similar mitotic phenotype, including chromosome misalignment and a mitotic delay (Fig. 2.5E-F). By titrating PEG300 concentration, we also observed an additive mitotic defect from combining hyperosmotic stress with ATP1A1 and AQP7 knockouts (Fig. 2.5E-F). Taking into consideration the combined evidence, we propose that AQP7 and ATP1A1 are required to create an internal osmotic cellular environment that promotes proper chromosome segregation, revealing an unanticipated role for osmolarity and its regulation in mitotic fidelity. Thus, our pooled live-cell screen confirms the observed mitotic defects from the primary screen and reveals roles for diverse gene targets in mitotic progression and fidelity.

2.11 LIN52, CLP1, and RNPC3 are required for the correct expression of kinetochore assembly factors

To define the basis for the observed mitotic phenotypes, we next tested the function of the kinetochore, the key player in mediating interactions between centromere DNA and microtubule polymers during cell division (*133*). In particular, we evaluated the localization of the inner kinetochore centromere-specific histone CENP-A and the outer kinetochore microtubule-binding protein Ndc80 for the 28 individual inducible knockout cell lines described above. Of these genes, 25 displayed only modest or no changes in the kinetochore recruitment of Ndc80 (Fig. 2.S13B), suggesting that kinetochore assembly is largely intact. In contrast, for the CLP1, RNPC3, and

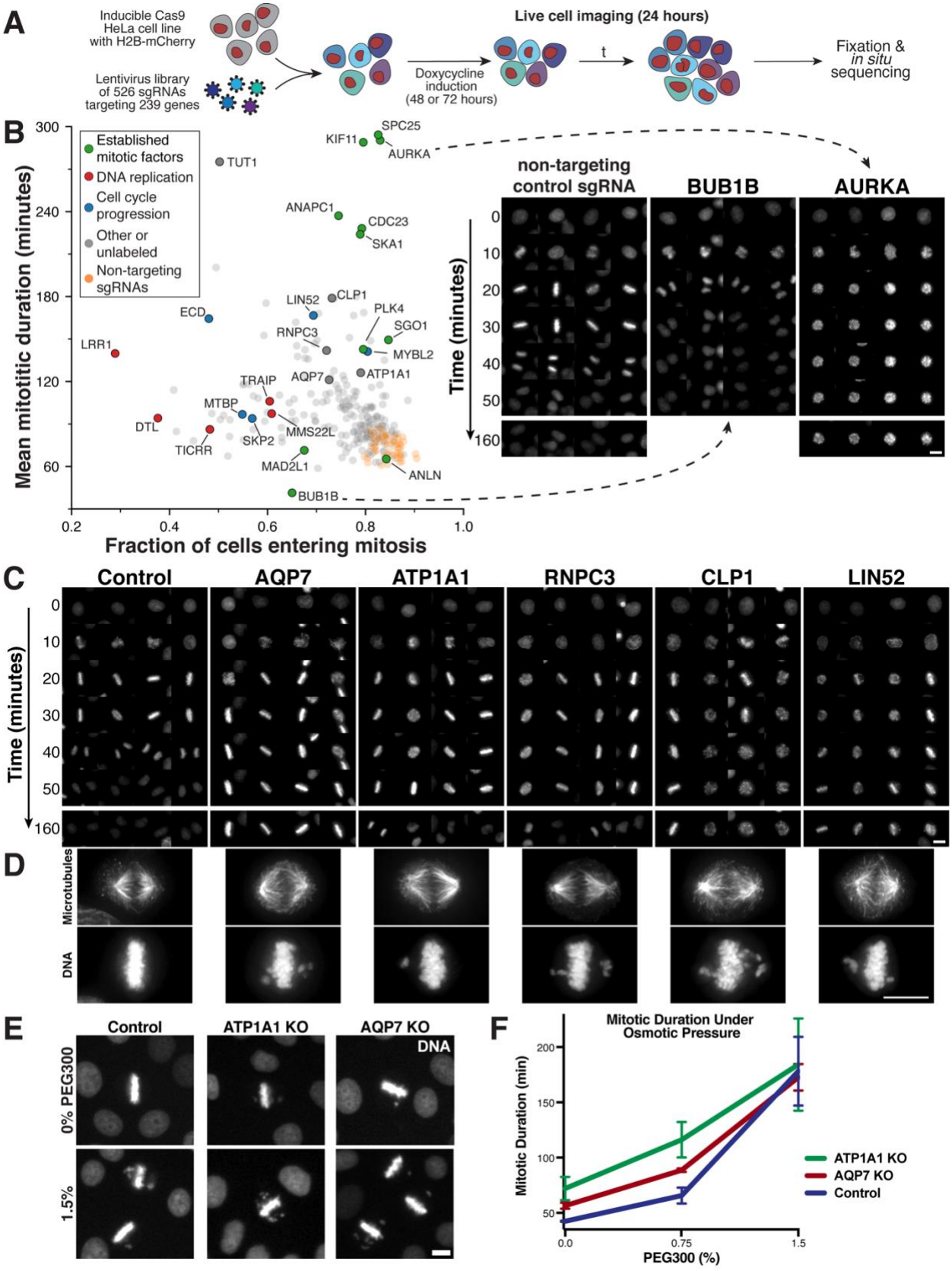
A
Inducible Cas9 HeLa cell line with H2B-mCherry
Lentivirus library of 526 sgRNAs targeting 239 genes
Doxycycline induction (48 or 72 hours)
Live cell imaging (24 hours)
t
Fixation & in situ sequencing
B
Established mitotic factors
DNA replication
Cell cycle progression
Other or unlabeled
Non-targeting sgRNAs
Mean mitotitic duration (minutes)
Fraction of cells entering mitosis
KIF11
SPC25
AURKA
TUT1
ANAPC1
CDC23
SKA1
LIN52
CLP1
PLK4
SGO1
ECD
RNPC3
MYBL2
LRR1
TRAIP
AQP7
ATP1A1
MTBP
DTL
MMS22L
SKP2
TICRR
MAD2L1
ANLN
BUB1B
non-targeting control sgRNA
BUB1B
AURKA
Time (minutes)
C
Control
AQP7
ATP1A1
RNPC3
CLP1
LIN52
Time (minutes)
D
Microtubules
DNA
E
Control
ATP1A1 KO
AQP7 KO
DNA
0% PEG300
1.5%
F
Mitotic Duration Under Osmotic Pressure
Mitotic Duration (min)
PEG300 (%)
ATP1A1 KO
AQP7 KO
Control

Figure 2.5. A pooled live-cell screen identifies gene targets required for mitotic progression. (**A**) Schematic of the experimental workflow for the live-cell, image-based pooled CRISPR screen using a cell line expressing an H2B-mCherry fusion (also see section 2.13.6). (**B**) Left, scatter plot comparing the fraction of cells that enter mitosis within the 24 hour time course and the mitotic duration of observed cell division events. This live-cell analysis identifies genes with defects in mitotic entry or progression. Plotted values represent the mean of sgRNAs targeting the same gene. Labeled genes are colored by functional category, including positive control mitotic factors in green. Right, example images of H2B-mCherry fluorescence from the live-cell screen at the indicated time points after mitotic entry for knockouts of established cell division components. Targeting the spindle assembly checkpoint gene BUB1B results in an acceleration of mitosis, whereas knockouts of AURKA, a key mitotic kinase, result in a mitotic arrest. (**C**) Example time course montages from the live-cell screen as in (B) demonstrating mitotic delay and mitotic defects for selected target genes. (**D**) Immunofluorescence images showing individual cell lines stably expressing a single sgRNA targeting each gene of interest to enable visualization of phenotypes at higher resolution across a single population (see also Fig. S13A, S14E). Images are deconvolved maximum intensity projections of fixed cells stained for microtubules (anti-alpha-tubulin) and DNA (Hoechst). Scale bars, 10 μm. (**E**) Example images and (**F**) mitotic duration quantification from time-lapse imaging of control, AQP7, and ATP1A1 inducible knockout cells incubated with varying PEG300 concentrations to induce hyperosmotic stress. Hyperosmotic stress alone induces chromosome alignment delays in control cells similar to AQP7 and ATP1A1 knockouts, while an additive effect on mitotic delay is observed when incubating knockout cells with PEG300. n>50 cells per datapoint.

LIN52 inducible knockouts, we observed a substantial reduction in Ndc80 localization and total Ndc80 protein levels (Fig. 2.6A; Fig. 2.S13D). Similarly, we found that 27 of the tested knockouts did not strongly alter the kinetochore localization of CENP-A, with the exception of the LIN52 inducible knockout which also demonstrated decreased total CENP-A protein levels (Fig. 2.6B; Fig. 2.S13C-D). Based on these changes in kinetochore assembly, we chose to focus on LIN52, CLP1, and RNCP3 to evaluate their contributions to mitosis.

LIN52 is a component of the DREAM complex, comprised of E2F family transcription factors, LIN9/37/52/54, MYBL1/2, RBL1/2, RBBP4, and TFDP1/2, which acts together with FOXM1 as a

transcriptional regulator for cell cycle genes (*134*, *135*). GFP-LIN52 localizes to the nucleus (Fig. 2.S14A), and in immunoprecipitation-mass spectrometry experiments we found that LIN52 associates with LIN9/37/54, RBBP4, and RBBP7, but not other established DREAM complex proteins (Fig. 2.S14B). Correspondingly, in our fixed-cell screen LIN52 displayed similar interphase phenotypes to LIN9/37/54, RBBP4 and RBBP7 (cluster 46; Fig. 2.6C), but not with the other DREAM-related genes present in the screen (MYBL2, TFDP1, E2F1, E2F3, E2F6, FOXM1), highlighting the predictive power of the clustering analysis. Consistent with the phenotypic co-clustering and physical interactions, we observed chromosome misalignment, a mitotic delay, and substantial changes to kinetochore assembly in knockouts of LIN52, LIN9, and LIN54 (Fig. 2.S14C-D). Based on RNA-seq analysis of LIN52 knockout cells arrested in mitosis, we found a pervasive decrease in the expression of diverse cell division genes (Fig. 2.6D). This includes substantial downregulation of multiple subunits of the Ndc80 complex as well as CENP-A and its associated deposition machinery, providing an explanation for the broad defects in kinetochore assembly. Thus, although LIN52 knockout cells are able to progress through the cell cycle and enter mitosis (Fig. 2.5B), the downregulation of these mitotic players provides an explanation for the observed aberrant chromosome alignment and mis-segregation. In contrast, we did not detect altered kinetochore protein levels or chromosome misalignment for FOXM1 knockouts and only observed a modest change in CENP-A and Ndc80 localization in MYBL2 knockouts (Fig. 2.6A-B; Fig. 2.S13D; Fig. 2.S14E), suggesting a potent role for the LIN52 sub-complex in the expression of cell division components.

CLP1 is a component of the pre-mRNA cleavage complex II (*136*). In our interphase phenotypic clustering analysis, CLP1 is closely associated with the direct interacting partner PCF11 (*136*) and another gene involved in transcription termination, ZC3H4 (Fig. 2.6E; *137*, *138*). This suggests the observed CLP1 phenotype is related to its role in mRNA 3' end formation. We performed RNA-sequencing analysis of CLP1 knockout cells, which revealed wide-spread defects in transcription termination (Fig. 2.S14F) as well as a global decrease in mRNA expression (Fig. 2.6F). Amongst the genes that evaded downregulation, PCF11 displayed significantly increased relative gene expression, consistent with autoregulation of the pre-mRNA cleavage complex (*139*). Reciprocally, we observed particularly strong downregulation of selected cell division components, including the Ndc80 complex subunits Spc24 and Spc25. The downregulation of these proteins may explain the selective loss of the Ndc80 complex (Fig. 2.6A-B), and the chromosome mis-alignment defects in CLP1 knockouts (Fig. 2.5C-D).

Finally, we analyzed RNCP3. In our fixed-cell screen, RNPC3 displayed an interphase phenotype closely related with multiple components of the minor spliceosome machinery (Fig. 2.6G), consistent with prior work (*140*). Based on RNA-seq analysis of RNPC3 knockouts arrested in mitosis, we found pervasive issues in the splicing of minor introns across diverse genes, coupled with substantial down-regulation of these minor intron-containing mRNAs (Fig. 2.6I). Amongst these genes, we identified the Ndc80 complex subunit SPC24 as being specifically mis-spliced and downregulated (Fig. 2.6H-I; Fig. 2.S14G; see also *141*). Strikingly, we were able to rescue the chromosome alignment defects in RNPC3 knockout cells through the exogenous expression of a spliced SPC24 cDNA (Fig. 2.6J). Thus, the role of RNPC3 in the minor spliceosome and the selective requirement for minor intron splicing in the production of the SPC24 mRNA explains the observed outer kinetochore assembly defects and chromosome mis-segregation phenotype in RNPC3 knockouts (Fig. 2.5C-D; Fig. 2.6A).

Together, this work provides a molecular explanation for the cell division phenotypes of LIN52, CLP1, and RNPC3 knockouts observed in our large-scale optical pooled screens, with defects in the production of critical players in chromosome segregation and kinetochore assembly. These analyses also highlight the ability of our image-based phenotypic clustering to identify functional relationships and define roles for diverse factors in complex phenotypes.

2.12 Pooled image-based screens define the phenotypic landscape of cellular functions

Our pooled microscopy-based analysis of tens of millions of individual knockout cells for thousands of essential and fitness-conferring human genes defines their functional contributions to diverse biological processes. By obtaining quantitative information for diverse image-based parameters that are directly comparable across a large cell population, this approach identifies co-functional gene targets with fine-grained resolution to distinguish roles in specific cellular processes and protein complexes. Studies analyzing proteome-wide protein interactions, single-cell transcriptional responses to thousands of genetic perturbations, and coordinated gene expression across biological contexts have also defined large-scale molecular networks (*111*, *124*, *142*). The precision and breadth of the clustering behaviors reported here highlight the ability of quantitative image-based phenotypic profiling to provide a similar scale of functional information, with complementary but distinct insights. Although we focused primarily on aggregate behaviors of predefined image features across the population of imaged cells for each gene

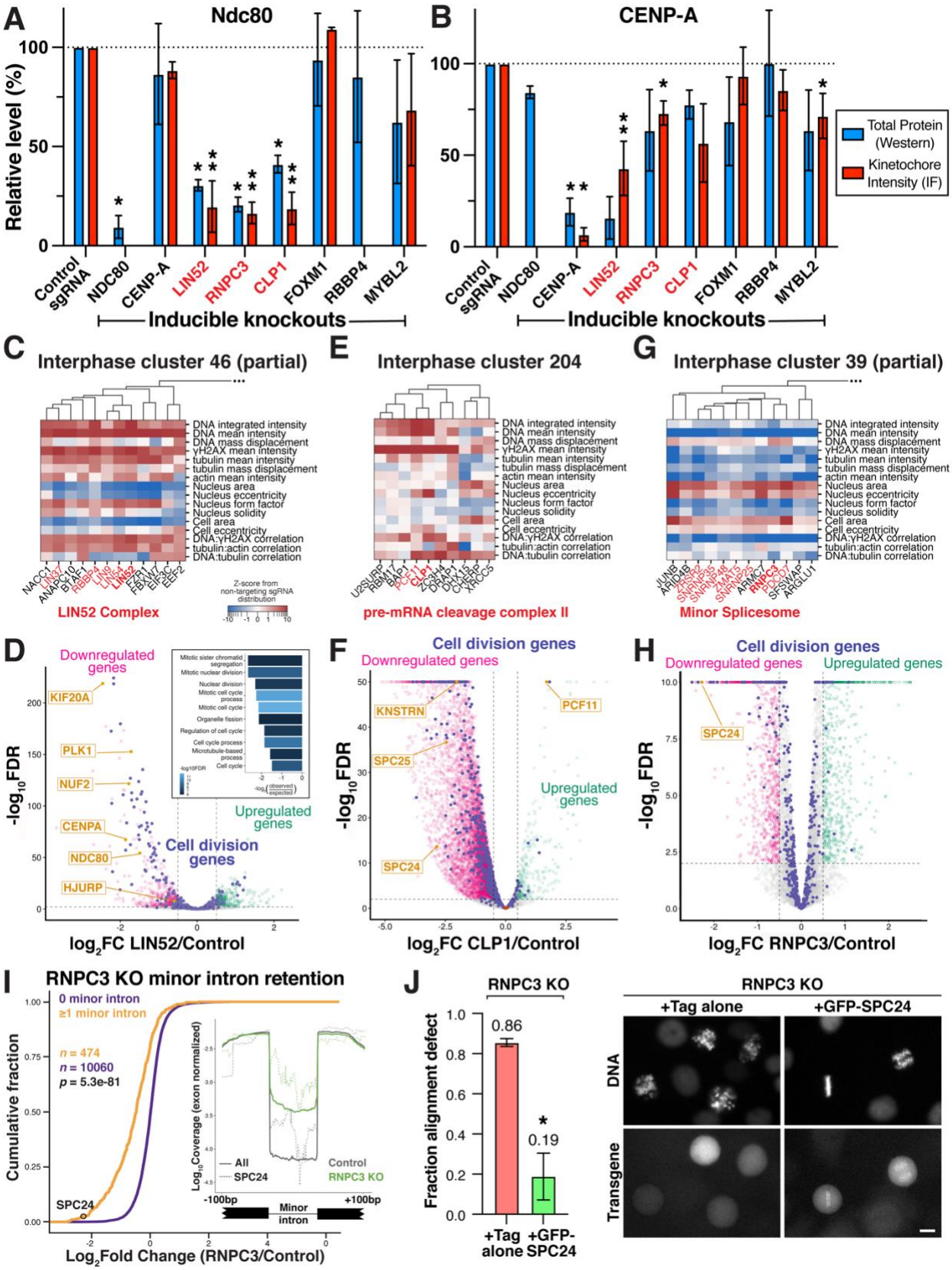
A
Ndc80
Relative level (%)
Control sgRNA
NDC80
CENP-A
LIN52
RNPC3
CLP1
FOXM1
RBBP4
MYBL2
Inducible knockouts
B
CENP-A
Total Protein (Western)
Kinetochore Intensity (IF)
C
Interphase cluster 46 (partial)
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean intensity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation
LIN52 Complex
Z-score from non-targeting sgRNA distribution
E
Interphase cluster 204
pre-mRNA cleavage complex II
G
Interphase cluster 39 (partial)
Minor Splicesome
D
Downregulated genes
Upregulated genes
Cell division genes
KIF20A
PLK1
NUF2
CENPA
NDC80
HJURP
-log10FDR
log2FC LIN52/Control
F
Cell division genes
KNSTRN
SPC25
SPC24
PCF11
log2FC CLP1/Control
H
Cell division genes
log2FC RNPC3/Control
I
RNPC3 KO minor intron retention
0 minor intron
≥1 minor intron
n = 474
n = 10060
p = 5.3e-81
Cumulative fraction
Log2Coverage (exon normalized)
All
SPC24
Control
RNPC3 KO
-100bp
Minor intron
+100bp
Log2Fold Change (RNPC3/Control)
J
RNPC3 KO
Fraction alignment defect
0.86
0.19
+Tag alone
+GFP-SPC24
RNPC3 KO
+Tag alone
+GFP-SPC24
DNA
Transgene

Figure 2.6. Lin52, Clp1 and RNPC3 functions to promote proper kinetochore assembly and chromosome segregation. (**A**) Bar plot showing total protein level (blue) and kinetochore-localized intensity (red) of the outer kinetochore microtubule-binding protein NDC80 relative to a control sgRNA in the indicated inducible knockout cell lines. NDC80 total protein levels and kinetochore-localized intensity are substantially decreased for CLP1, RNPC3, and LIN52 knockouts. N=2 biological replicates for total protein levels, which were normalized to GAPDH. N=2-10 biological replicates for kinetochore measurements, each replicate represents the median kinetochore signal from >10 cells. Both measurements were further normalized relative to controls from the same experiment. *P<0.05, **P<0.01 by two-tailed independent T-test relative to corresponding control samples**.** ND, no data. (**B**) Bar plot as in (A) showing total protein level (blue) and kinetochore-localized intensity for the inner kinetochore centromere-specific histone CENP-A relative to a control sgRNA in the indicated inducible knockout cell lines. CENP-A total protein levels and kinetochore-localized intensity are strongly decreased for LIN52 knockouts. Experiment design as in (A). *P<0.05, **P<0.01 by two-tailed independent T-test relative to corresponding control samples. (**C**) Phenotype heat map and hierarchical clustering for a subset of primary screen interphase cluster 46 genes. This highlights the similarities between LIN52 and its previously defined interacting partners (red). Hierarchical clustering (top) was performed using the Pearson correlation of PCA-projected phenotype profiles (see section 2.13.8). Parameters are presented as z-scores from the distribution of non-targeting sgRNAs, visualized on a symmetric log scale (linear between -1 and 1). (**D**) Volcano plot of LIN52 knockout differential expression demonstrates the requirement of LIN52 for the expression of both inner and outer kinetochore components. Genes involved in cell division processes are indicated in purple. Significance threshold FDR < 0.01, $\log_2$ effect size > 0.5 for up- (green) and down-regulated genes (magenta). Inset, GO term analysis of LIN52 downregulated genes shows a significant enrichment of mitotic genes (inset). (**E**) Heat map of primary screen interphase cluster 204 phenotypes as in (C), demonstrating an association of knockout phenotypes for the pre-mRNA cleavage complex II factors CLP1 and PCF11. ZC3H4, an additional gene involved in transcriptional termination, also shows a similar knockout phenotype. (**F**) Volcano plot of differential expression as in (D) following CLP1 knockout, identifying a global decrease in mRNA abundance in these cells, identified by normalizing to library spike-in control RNA (brown). Expression of outer kinetochore genes SPC24 and SPC25 is particularly sensitive to loss of CLP1. The interacting 3' end processing gene PCF11 is upregulated in CLP1 knockout cells, consistent with autoregulation of this complex (*70*). (**G**) Heat map of a subset of primary screen interphase cluster 39 phenotypes as in (C), demonstrating tight clustering of minor spliceosome components including RNPC3. (**H**)

Volcano plot of differential gene expression as in (D) after RNPC3 knockout. SPC24 is the only outer kinetochore component significantly downregulated in RNPC3 knockout cells. (**I**) Cumulative distributions of mRNA fold change from RNPC3 knockout cells for transcripts containing at least 1 minor intron (orange) are significantly downregulated compared to transcripts with no minor introns (purple). Statistical significance between cumulative distributions was assessed using the Mann-Whitney U test. Inset, minor introns are retained in RNPC3 knockout cells (green), including the minor introns in SPC24 (dotted lines), possibly explaining the decreased SPC24 transcript levels and kinetochore assembly defects in these cells. (**J**) Left, bar plot showing fraction of mitotic RNPC3 knockout cells displaying chromosome alignment defects when expressing a fluorescent Tag only (NeonGreen) or GFP-SPC24. Data points indicate independent replicates of >100 cells. * = P<0.01 by T-test relative to the tag alone control**.** Right, representative images of DNA (H2B-mCherry) and transgene localization for live RNPC3 knockout cells expressing Tag only or GFP-SPC24. Scale bar, 10 µm.

target, future studies leveraging the distribution of single-cell phenotypes or applying convolutional neural networks to directly learn phenotype representations will enable additional resolution and insights for understanding gene functions. In addition to providing an expansive and powerful resource for the analysis of phenotypes resulting from the disruption of essential genes in our companion interactive web portal (https://vesuvius.wi.mit.edu/), this work provides multiple predictions for the contributions of incompletely characterized genes to fundamental cellular functions. We anticipate that this type of scalable and information-rich cell biological genomic screening will enable future studies that will yield additional key insights across numerous cellular phenotypes and conditions.

2.13 Methods

2.13.1 Library design and cloning

The primary screen library of fitness-conferring genes was defined based on evidence from multiple published sources. First, we used data from the Broad Institute DepMap project (*81*, *82*, *85*) to identify genes that are broadly fitness-conferring in a variety of cell lines. Specifically, we selected genes with a genetic dependency probability of >0.35 in at least 10% of the >600 tested cell lines (Fig. 2.S1A, B), resulting in 3,991 selected genes. We subsequently chose 1,081

additional genes that had evidence of essentiality in at least 2 other published screens (*80*, *83*, *84*, *86–88*). CRISPR sgRNA sequences were selected from published libraries (*88–90*), with simultaneous optimization of sgRNA performance (e.g., on- and off-target efficiency) and minimization of 5' sequence length required to demultiplex all sgRNAs during in situ sequencing. In total, we selected 20,445 sgRNA sequences, including 4 sgRNAs each for all but one gene target (3 sgRNAs targeting RGPD5) and 250 non-targeting control sgRNAs, with a minimum Levenshtein distance of 2 between the leading 11-nucleotide 5' sequence for all possible pairs of sgRNAs. We note that, for some groups of genes with high sequence homology, it is not possible to design distinct targeting sgRNAs for each gene. For groups of genes where the full lists of possible sgRNAs collected from previously published libraries were identical, a single set of 4 sgRNAs was chosen to target these genes collectively. Two sgRNAs per gene were selected for the 239 genes in the live cell screen based on performance in the fixed-cell screen, in addition to 50 non-targeting guides selected using the 5' sequence optimization described above. Targeting and non-targeting sgRNA libraries were designed as separate subpools of synthesized oligo arrays (Agilent) and independently cloned into CROPseq-puro-v2 (Addgene #127458) as described previously (*1*).

For expression of fusion proteins, H2B (pKC96) was amplified from a template retroviral construct (*143*) and SPC24 (pKG422) from pJAG261 (gift from Jagesh Shah), while C7orf26 (NP_076972.2; pKC509) and LIN52 (Q52LA3.1; pKC518) were human codon-optimized and synthesized (Twist Biosciences). Gene fragments were ligated into an mCherry, GFP, or EGFP pBABE-based vector (Addgene #44432). ZNF335 (pKC530) was amplified from Synthetic construct Homo sapiens clone IMAGE:100066405 andligated into a EGFP lentiviral vector. sgRNA constructs for individual inducible knockout cell lines were generated by primer annealing and ligation into sgOPTI (*144*). A control sgRNA with a single target site within the non-essential LBR gene was used for comparison of all follow-up experiments (HS1, *145*).

2.13.2 Tissue culture

HeLa and HEK293 cells were cultured in DMEM with sodium pyruvate and GlutaMAX (Life Technologies 10569044) or 2 mM L-glutamine supplemented with 10% heat-inactivated fetal bovine serum (Sigma F4135) and 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific 15140122).

2.13.3 Virus production, transduction and selection

Prior to lentiviral production of screening sgRNA libraries, the corresponding targeting and non-targeting plasmid pools were mixed (final non-targeting sgRNA pool fraction of 5% for the primary fixed-cell screen, 9.5% for the secondary live-cell screen). Lentiviral production and transduction were performed as described previously for libraries (*1*, *75*) or single targets (*87*).

Retrovirus was generated by transfecting VSVG packaging plasmid and pBABE-based vectors containing H2B-mCherry, EGFP-C7orf26, EGFP-Lin52, GFP-SPC24 fusions or mNeonGreen into HEK293-GP cells with Effectene (Qiagen) for transduction as described previously (*146*). Transduced cells were enriched by FACS (GFP-SPC24) or selected with 375 µg/ml hygromycin (Invitrogen).

2.13.4 Fluorescence microscopy

All screening datasets were acquired using a Nikon Ti-2 inverted epifluorescence microscope with automated stage control, hardware autofocus, and an Iris 9 sCMOS camera (Teledyne Photometrics). All hardware was controlled using NIS-Elements AR, and a CELESTA light engine (Lumencor) was used for fluorescence illumination. In situ sequencing cycles were acquired using a 10X 0.45 NA CFI Plan Apo Lambda objective (Nikon MRD00105) and 2x2 pixel binning with the following laser lines, filters, and exposure times for each channel: DAPI (408 nm laser excitation with 0.8% power, custom Chroma dual-band 408/473 dichroic and emission filter set, 50 ms exposure), Miseq G (545 nm laser with 30% power and Semrock FF01-543/3 excitation filter, Chroma T555LPXR dichroic filter, Chroma ET575/30 emission filter, 200 ms exposure), Miseq T (545 nm laser excitation with 30% power, Chroma T565LPXR dichroic filter, Semrock FF01-615/24 emission filter, 200 ms exposure), Miseq A (635 nm laser excitation with 30% power, Chroma ZET635RDC dichroic filter, Semrock FF01-680/42 emission filter, 200 ms), Miseq C (635 nm laser excitation with 30% power, Chroma ZET635RDC dichroic filter, Semrock FF01-732/68 emission filter, 200 ms exposure). Fixed-cell primary screen phenotype images were acquired using a 20X 0.75 NA CFI Plan Apo Lambda objective (Nikon MRD00205) using DAPI (as before), FITC (473 nm laser excitation, custom Chroma 408/473 filter set), Alexa Fluor 594 (same settings as MiSeq T), and Alexa Fluor 750 (750 nm laser excitation, Semrock FF765-Di01 dichroic filter, custom ET820/110 Chroma emission filter) fluorescence channels. For the live-cell secondary screen, timelapse phenotype images were acquired using the 20X objective lens, an mCherry

fluorescence channel (same settings as MiSeq T), and a microscope enclosure with temperature and CO_2 control along with passive humidification (Okolab H201).

Immunofluorescence images of single knockout cell lines were taken on the Deltavision Ultra (Cytiva) system using a 60x/1.42NA objective and deconvolution. For kinetochore component quantification, z-sections at 0.2 µm intervals were taken using a 100X/1.45NA objective. For time lapse imaging of individual inducible knockouts and EGFP fusion cell lines, we used a Nikon Eclipse microscope equipped with an ORCA-Fusion BT sCMOS camera (Hamamatsu) using a Plan Fluor 20X/0.5 NA (live cells) or 40x/1.3NA (EGFP) objective lens.

2.13.5 Fixed-cell optical pooled CRISPR screen

For the fixed cell screen, HeLa-TetR-Cas9 (A7) cells were transduced with the 20,445 sgRNA library in CROPseq-puro-v2 and selected with 2 µg/mL puromycin (Thermo Fisher Scientific A1113803) for 4 days. Cas9 expression was induced with 2 µg/mL doxycycline for 78 hours, and the cell library was seeded into eight 6-well glass-bottom plates (Cellvis P06-1.5H-N) at a density of 300,000 cells per well (~30,000 cells/cm^2) 48 hours prior to fixation. Cells were fixed with 4% paraformaldehyde in PBS for 30 minutes, followed by in situ amplification as described previously (*1*, *75*). After rolling circle amplification, cells were stained with rabbit anti-gamma H2A.X (phospho S139) antibody (Abcam ab81299, 1:2000 dilution in PBS with 3% BSA) for 1 hour at room temperature. Cells were washed twice with PBS-T (PBS with 0.05% Tween-20), then stained with mouse anti-alpha-tubulin-FITC antibody (Sigma F2168, 1:500 dilution), goat anti-rabbit antibody disulfide-linked to Alexa Fluor 594 (Invitrogen 31212, Thermo Fisher Scientific A10270, custom conjugation; 1:500 dilution), and Alexa Fluor Plus 750 Phalloidin (Thermo Fisher Scientific A30105, 1:1000 dilution) in PBS with 3% BSA for 45 minutes at room temperature. After washing with PBS-T three times, well plates were replaced with 200 ng/mL DAPI in 2X SSC and imaged for cellular phenotypes using the microscope configuration described above with 4 z-slices at 1.5 µm intervals. Following phenotype imaging, Alexa Fluor 594 was cleaved from disulfide-linked antibodies by incubating cells in 50 mM TCEP in 2X SSC for 1 hour at room temperature, followed by three washes with PBS-T. Finally, 11 cycles of in situ sequencing-by-synthesis were performed as described previously (*1*, *75*). In parallel with the optical pooled screen, cells expressing the same sgRNA library were induced with 1 µg/mL doxycycline, and then doxycycline media was refreshed every day for 2 more days. Cells were harvested on days 0 (pre-induction), 3, and 5 post-Cas9 induction and genomic DNA was extracted using PureLink (Invitrogen). sgRNA sequences were then PCR amplified using Q5 hotstart (NEB) with primers

oDF344 (ACACGACGCTCTTCCGATCTcttgtggaaaggacgaaac) and oDF112 (CTGGAGTTCAGACGTGTGCTCTTCCGATCaagcaccgactcggtgccac) before addition of index barcodes and sequencing on an Illumina HiSeq using sequencing primer oKC651 (ACACTCTTTCCCTACACGACGCTCTTCCGATCTcttgtggaaaggacgaaacaccg).

2.13.6 Live-cell optical pooled CRISPR screen

HeLa-TetR-Cas9 cells expressing an H2B-mCherry fusion protein (cKC556) were transduced with the live-cell screening library of 526 sgRNA sequences. Cells were selected with 2 μg/mL puromycin (Thermo Fisher Scientific A1113803) for 3 days. Cas9 expression was induced with 2 μg/mL doxycycline for 48 (day 3 time course) or 72 hours (day 4 time course) prior to the beginning of live-cell imaging, and cells were seeded into 6-well glass-bottom plates (Cellvis P06-1.5H-N) at a density of 300,000 or 350,000 cells per well (~35,000 cells/cm^2) 24 hours prior to imaging. Each time course was performed in three batches on separate days. Immediately before imaging, cells were washed once with PBS, and then replaced with imaging media consisting of phenol red-free DMEM with L-glutamine and HEPES (Thermo Fisher Scientific 21063029) supplemented with 10% heat-inactivated fetal bovine serum (Sigma F4135) and 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific 15140122). Live-cell imaging was performed using the microscope configuration described above and 2 z-slices spaced at either 4 or 5 μm intervals. Cells were imaged for 24 hours at 10 minute time intervals, immediately fixed with 4% paraformaldehyde in PBS for 30 minutes, then processed through in situ amplification and sequencing-by-synthesis following the same protocol as the fixed-cell screen.

2.13.7 Screening image analysis

In situ sequencing spots were identified and barcode sequences extracted using our previously described workflow (*1*, *75*, *147*). In addition, phenotype images were acquired at a higher magnification than in situ sequencing images, and thus the datasets were computationally aligned to match cell identities. This alignment was completed by computing the Delaunay triangulation of nuclei centroids for each phenotype and sequencing image tile, and then comparing triangulations between images from the two datasets to find matching tiles and cell identities. Overall, approximately 60% of all segmented cells were included in the final dataset (Fig. 2.S1E), with remaining cells unused due to no in situ sequencing spots matching a designed sgRNA sequence, sequencing reads mapping to multiple gene targets, or an inability to match cell identities between the sequencing and phenotype images.

Phenotype images from both screens were first maximum intensity projected to compress z-slices into a single plane, and then a retrospective flat-field correction was applied to reduce effects from uneven illumination (*148*). Nuclei were semantically segmented by applying a local intensity threshold to the DAPI channel and then performing morphological operations to remove aberrant holes and particles. Individual nucleus instances were then segmented using the watershed algorithm. In the fixed-cell phenotype data, semantic segmentation of cytoplasmic foreground was achieved by thresholding a Gaussian-filtered copy of the phalloidin (actin) channel (sigma of 3 pixels), followed by morphological operations. Cell instances were identified by applying the watershed algorithm with nuclear segmentations as seeds. Phenotype parameters were extracted from nuclear and cellular segmentations for each channel by implementing image features from CellProfiler (*149*), scikit-image (*150*), and mahotas (*151*) as Python functions operating on scikit-image RegionProperties objects. Image segmentation, phenotype feature extraction, and in situ sequencing analysis were performed in parallel on a per-image tile basis using the Snakemake workflow manager (*152*).

2.13.8 Fixed-cell screen phenotype analysis

After aligning the phenotype and sequencing datasets, a subset of features were transformed to approximate normal distributions. All features for each cell were then normalized using the median and median absolute deviation of the population of cells carrying non-targeting sgRNAs within the same well (robust z-score). This internal control procedure was used to reduce batch effects between wells and plates that may be caused by intensity differences or cell density effects. Mitotic and interphase cells were identified using a support vector classifier (scikit-learn (*153*) svm.SVC implementation, default parameters) trained with 2,514 annotated cells on a subset of 182 features that demonstrated the highest average difference between annotated mitotic and interphase cells (Fig. 2.S1F). Cell-level measurements were then re-normalized from the raw data as before, but within interphase and mitotic cells separately.

Summary phenotype measurements were computed for each gene target by taking the median of z-scored parameters for all cells targeted by a single sgRNA sequence, then aggregating to the gene level by taking the median across sgRNAs targeting the same gene. Raw p-values for a subset of summary parameters were computed by comparing gene scores to null distributions of corresponding bootstrapped summary scores from cells expressing non-targeting sgRNAs. Separate null distributions were defined for each gene target by first performing 100,000 cell

sampling repetitions to produce a distribution of bootstrapped non-targeting sgRNA scores for each cell sample size of the targeting sgRNAs. These guide-level null distributions were then correspondingly sampled 100,000 times for each group of sgRNAs targeting the same gene and aggregated to produce gene-level null distributions with matched cell and guide sample sizes. The Benjamini-Hochberg procedure was applied to obtain the reported FDR q-values. An FDR threshold of 0.05 was used for defining significance for all parameters. This process was modified for the mean nuclear γH2AX intensity measurements, as the non-targeting cells do not provide an adequate null population for this phenotype given the lack of Cas9-induced DNA breaks. In this case, bootstrapped null distributions were generated by sampling from all cells expressing any targeting sgRNA. This approach, combined with an effect size threshold at the 2.5 and 97.5 percentiles of the non-targeting sgRNA scores, resulted in a conservative identification of mean nuclear γH2AX intensity phenotypes beyond baseline DNA damage resulting from Cas9 nuclease activity.

For the high-dimensional analysis, pairs of features with a Pearson correlation greater than 0.9 were iteratively excluded, and additional features with low variance or only a few unique discrete values across the dataset were removed. This resulted in a set of 472 features for the interphase dataset and 884 features for the mitotic dataset, selected independently from the full list of 1,084 extracted features. Further feature redundancies were reduced by applying principal component analysis (PCA) and retaining the components that explain 95% of the variance in the datasets (103 components for the interphase data, 530 components for the mitotic data). The PHATE manifold learning and visualization algorithm (*109*) was then used to produce two-dimensional representations of the phenotypic landscape of gene targets (default parameters except n_pca=None). To cluster knockout phenotypes, the PHATE diffusion operator affinity graph was supplied as input to the Leiden algorithm, which optimizes cluster modularity (*154*). The Leiden resolution parameter was chosen by analyzing the robustness of clustering solutions to the subsampling of gene-level data with varying resolution (resolution = 10 for interphase dataset, resolution = 9 for mitotic dataset; Fig. 2.S6C, F). For visualization of differences between phenotype profiles and corresponding clusters in the presented heatmaps (Fig. 2.3D, F, G, Fig. 2.4C, etc.), we first selected clusters to highlight based on the presence of known functional groups of genes, then iteratively selected a minimal set of phenotype parameters that together discriminated the various clusters. Features with clear explanation were prioritized to enable interpretability, resulting in 16 interphase and 16 mitotic phenotype features. Where hierarchical clustering results are presented, this was performed using average linkage (UPGMA) of the

Pearson correlation between PCA-projected phenotype profiles. In parallel with the computational phenotype analysis, two individuals independently scored mitotic phenotypes from the primary screen by visually inspecting montages of mitotic cells from each gene target and assigning a phenotype severity score from 1 to 9 (Fig. 2.S10A, B). During this process, the scoring individuals were blinded to the gene identities associated with each montage of cells.

2.13.9 Comparisons to external data

To compare our interphase phenotype profile similarities to existing datasets of co-functional genes (Fig. 2.3B), we used the CORUM 3.0 core set of protein complexes (*110*), BioPlex 3.0 HEK293T interactions (*111*), and the co-essential gene pairs defined in Wainberg, et al. (*112*). For CORUM, all possible pairs of genes within the same complex were considered as co-functional. In each dataset, all annotated co-functional gene pairs were included for the correlation analysis in Fig. 2.3B if both genes were targeted in our fixed-cell screening library. For the precision-recall analysis (Fig. 2.S7), a subset of minimal CORUM complexes were selected that contained limited overlap with other complexes (defined as containing at least 3 genes from the screening library with at least 2/3 of the full complex represented, and removing the largest complexes that share more than 10% of gene-pairs with smaller CORUM complexes), resulting in 9,781 gene pairs included across 292 annotated complexes. Precision and recall were defined as indicated in Fig S7A. When restricting precision-recall to varying overall phenotype strengths, the indicated quantiles were used as thresholds on the interphase PHATE mean potential distance to non-targeting sgRNAs (as presented in Fig. 2.S6A-B), only evaluating pairs of genes where both members met the given threshold. For the tested similarity metrics in Fig. 2.S7B, Pearson correlation coefficients were calculated between the PCA-projected phenotype profiles, the PHATE diffusion operator affinity graph was used as described above, and the UMAP fuzzy simplicial set affinity graph was computed using the PCA-projected phenotype profiles. To generate precision and recall curves, the following parameters were swept for the corresponding clustering method to vary the resolution of clustering results: the Leiden algorithm resolution parameter, the DBSCAN epsilon parameter (with min_samples=1 using sklearn.cluster.dbscan), and the maximum number of clusters for hierarchical clustering (criterion="maxclust" with scipy.cluster.hierarchy.fcluster). In all analyses of KEGG pathway enrichment, pathways categorized as "Organismal Systems," "Human Diseases," or "Drug Development" were excluded. For cluster enrichment of CORUM complexes, complexes were included that contained at least 3 genes from the screening library with at least 2/3 of the full complex represented.

To compare to prior DNA damage response screens (*101*), all genes identified as either sensitizing and resistance-conferring to genotoxic agents were compared to all genes in the fixed-cell screen exhibiting significant increases or decreases in γH2AX staining (Fig. 2.S3F). From MitoCheck, all genes identified in at least one of the four main phenotype categories ("mitotic arrest/delay," "binuclear," "polylobed," and "grape") were considered as exhibiting mitotic phenotypes. These genes were compared to those in our fixed-cell screen demonstrating strong image-based mitotic phenotype profiles (using the mean PHATE potential distance to non-targeting sgRNAs, selecting gene targets above the 95th percentile of non-targeting sgRNAs; Fig. 2.S6D-E) or significantly altered mitotic index ($P<0.05$ by permutation test with 10,000 permutations of sgRNA-gene assignments; Fig. 2.S10C).

2.13.10 Live-cell screen phenotype analysis

Following nuclear segmentation of the time lapse data, cells were tracked across frames using the TrackMate implementation of the linear assignment problem approach to particle tracking (*155*, *156*). The cost of linking nuclei in consecutive frames was set as the squared distance between centroids, with maximum linking distance set to 60 pixels (~18 μm). Track gaps up to 2 frames were allowed, in addition to track merges and splits. Tracked cell lineages that did not last for the full 24 hour time-course were excluded from analysis.

The sgRNA assigned to each tracked cell lineage in the phenotype data was determined by matching cell identities between the in situ sequencing images and the final time point of the time course. Similar to the fixed-cell screen, individual cell feature measurements were normalized using the median and median absolute deviation of the non-targeting control cell population from the same well and time point to reduce batch effects and correct for temporal intensity variations. Interphase, mitotic, and apoptotic cells were classified using a support vector machine (scikit-learn svm.SVC, linear kernel) with 2,514 annotated cells using 81 features selected from the full set of 116 extracted features by iteratively removing pairs with Pearson correlation >0.9 (Fig. 2.S11A). However, due to the difficulty of separating mitotic and apoptotic cells based on H2B-mCherry fluorescence alone, these categories were later combined into a single, broad "mitotic" bin. Cell division events were defined as a contiguous sequence of at least 2 frames of mitotic-classified cells immediately followed by a split in the track into 2 daughter cells (Fig. 2.S11B). Also included as cell division events were continuous sequences of mitotic cells that start in the first frame or reach the end of the acquired time course, if the observed mitotic duration was at least

as long as the average mitotic duration of non-targeting control cells in the same well. Mitotic duration was measured as the time difference between the first and last frame of the cell division event. The fraction of cells entering mitosis was calculated as the fraction of tracked lineages containing at least one cell division event as defined above. Both measurements were aggregated to the gene level by taking the average of sgRNAs targeting the same gene. Since many genes exhibited a stronger phenotype at either the Day 3 or Day 4 time point, likely due to differences in protein depletion timing, the strongest phenotype was selected for plotting in Fig. 2.5B by selecting the time course with the highest absolute difference in mitotic duration compared to the mean of non-targeting sgRNAs.

2.13.11 GFP immunoprecipitation and Mass-spectrometry

IP-MS experiments were performed as described previously (*157*). EGFP-C7orf26 and EGFP-LIN52 cells were mitotically enriched with 10μM STLC overnight, harvested and washed in PBS and resuspended 1:1 in 1X Lysis Buffer (50 mM HEPES, 1 mM EGTA, 1 mM $MgCl_2$, 100 mM KCl, 10% glycerol, pH 7.4) then frozen in liquid nitrogen. Cells were thawed after addition of an equal volume of 1.5X lysis buffer supplemented with 0.075% Nonidet P-40, 1X Complete EDTA-free protease inhibitor cocktail (Roche), 1 mM phenylmethylsulfonyl fluoride, 20 mM beta-glycerophosphate, 1 mM sodium fluoride, and 0.4 mM sodium orthovanadate. Cells were then lysed by sonication and cleared by centrifugation. The supernatant was mixed with Protein A beads (Biorad) coupled to rabbit anti-GFP antibodies (Cheeseman lab) and rotated at 4°C for 1 hour. Beads were washed five times in wash buffer (50 mM HEPES, 1 mM EGTA, 1 mM $MgCl_2$, 300 mM KCl, 10% glycerol, 0.05% NP-40, 1 mM dithiothreitol, 10 μg/mL leupeptin/pepstatin/chymostatin, pH 7.4). After a final rinse in wash buffer without detergent, bound protein was eluted with 100 mM glycine pH 2.6. Eluted proteins were precipitated by addition of $1/5^{th}$ volume trichloroacetic acid at 4°C overnight. Precipitated proteins were reduced with TCEP, alkylated with iodoacetamide, and digested with mass-spectrometry grade trypsin (Promega) using S-Trap (Protifi) according to the manufacturer's instructions. Peptides were separated by liquid chromatography and analyzed on an Orbitrap Elite mass spectrometer (Exploris 480, Thermo Fisher) with FAIMS Pro Interface (Thermo Fisher). Data were analyzed using Proteome Discoverer Software (Thermo Fisher).

2.13.12 Western Blotting

Cells expressing individual sgRNAs were induced in 1 µg/mL doxycycline for 2 to 5 days before lysis in Laemmli buffer and incubation at 95°C for 5 min. For mitotic samples, cells were harvested by mitotic shake off and, when necessary, after an overnight 10 µM STLC incubation. Samples were separated by SDS-PAGE and semi-dry transferred to nitrocellulose. Membranes were blocked for 30 min in blocking buffer (5% BSA for H2A.X; for all others, milk in TBS with 0.1% Tween-20) before incubation with primary antibodies: anti-phospho-H2A.X (Ser139, Millipore clone JBW301; 1:1000), anti c-Myc (Abcam, ab32072; 1:1000), anti-CENP-A (Clone 3-19, Invitrogen; 1:2000), or anti-"Bonsai"/NDC80 (*158*; 0.5 µg/mL). This was followed by HRP-conjugated secondary antibody (Kindle Biosciences) incubation at 1:1000 dilution. To detect GAPDH as a loading control, HRP-conjugated antibody (Abcam, ab185059) was applied at 1:20,000 dilution. Membranes were imaged with a KwikQuant Imager (Kindle Biosciences) and quantified using Image Studio software (LI-COR).

2.13.13 Arrayed imaging experiments with inducible knockout cell lines

Inducible knockout cell lines for immunofluorescence were seeded on poly-L-lysine (Sigma-Aldrich) coated coverslips and fixed in PHEM with 4% formaldehyde for 10 min at 37°C (microtubule staining) or ice cold methanol. Coverslips were washed with PBS, permeabilized with 0.2% Triton X-100 in PBS, and blocked in Abdil buffer (20 mM Tris-HCl, 150 mM NaCl, 0.1% Triton X-100, 3% bovine serum albumin, 0.1% NaN_3, pH 7.5). Anti-alpha-tubulin (DM1A, Sigma; 1:3000 dilution), anti-Centrin (*159*,1µg/mL), anti-CENP-A (Clone 3-19, Invitrogen; 1:1000 dilution) and anti- "Bonsai"/NDC80 (*158*; 1 µg/mL) antibodies in Abdil buffer were used for primary staining. Cy2- and Cy5-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories) were diluted 1:500 with 1 µg/mL Hoechst-33342 (Sigma-Aldrich) in Abdil for subsequent staining. Slides were mounted with ProLong Gold Antifade (Invitrogen) prior to imaging using the microscope configuration described above.

For quantifications of Ndc80 and CENP-A kinetochore stain intensity, sections of cells were maximum intensity projected and cropped in Fiji (*160*). Integrated fluorescence intensity of mitotic kinetochores was measured with a custom pipeline in CellProfiler (*149*). The median intensity of a 5-pixel wide region surrounding each kinetochore was used to background subtract each measurement.

For live analysis of individual knockout and RNPC3 rescue cell lines, cells were induced with 1 µg/mL doxycycline for 3 days, refreshing doxycycline media each day. On day 3 or 4 post-Cas9 induction, cells were moved to CO_2-independent media (Gibco) supplemented with 10% FBS, 100 U/mL penicillin and streptomycin, and 2 mM L-glutamine before imaging using the microscope configuration described above in 12-well polymer-bottomed plates (Cellvis). For the hyperosmotic stress experiments, polyethylene Glycol (PEG) 300 (TCI) was applied to the media at the indicated concentrations (w/v%) and incubated for 6 h prior to imaging. For mitochondrial imaging in Fig. 2.S9D, MitoTracker Orange CMTMRos (Molecular Probes) was applied at 25 nM for 30 min before imaging. For MitoTracker image analysis, nuclei were segmented using the CellPose segmentation algorithm (*161*) with a Hoechst stain, then a cytoplasmic ring was defined by morphologically dilating the nuclei by 10 pixels.

2.13.14 RNA-sequencing and analysis of inducible knockout cell lines

Inducible knockout cells were seeded in 1 µg/mL doxycycline, and doxycycline media was refreshed each day for 3 days before harvest of a mitotically-enriched cell population by shake-off on day 5. Control and ZNF335 knockout cells were additionally treated with 10µM STLC for 12 h prior to harvest. Cells were washed in PBS before snap-freezing pellets of 500,000 cells in liquid nitrogen. RNA was purified using TRIzol reagent (Life Technologies) according to manufacturer's instructions. 2ug of purified total RNA was mixed with 0.7 pg and 1.3 pg polyadenylated Nano luciferase and Firefly luciferase spike-in mRNA, respectively. KAPA mRNA HyperPrep kit with poly(A) selection was used to prepare libraries. Libraries were sequenced with the Illumina NovaSeq 6000 platform, 100x100 bp paired-end reads.

Reads were trimmed to remove any poly(A) sequences using Cutadapt (v3.7) (*162*) with the parameters "--minimumlength 1 -a A{25}". Reads were mapped to the human genome (Gencode v25) using STAR v2.7.1a (*163*) with the parameters "--runMode alignReads --outFilterMultimapNmax 1 --outFilterType BySJout --outSAMattributes All --outSAMtype BAM SortedByCoordinate". Aligned reads were quantified using htseq-count (0.11.0) (*164*). A read cutoff of at least 20 reads for each gene sample was applied before further analysis. Differential expression analysis was performed using DESeq2 (*165*). Differentially-expressed genes are defined as $\log_2$ effect size > 0.5 and FDR < 0.01. For the Clp1 RNA-seq data, the relative abundance of spike-in mRNAs was used as the sizeFactor for DESeq2 instead of median normalization due to 3' end processing defects resulting in global mRNA downregulation. The Minor Intron Database v1.2 (*166*) was used to reference minor intron containing genes.

ShinyGO v0.75 (*167*) was used to identify enriched GO terms from the GO Biological Process database. Enrichment analysis was performed within the downregulated genes from LIN52 knockout cells; genes that surpassed the 20 read cutoff and did not show differential expression were used as the background set.

Clp1 and RNPC3 meta plots were generated using the deepTools v3.5 package (*168*). The aligned reads were converted to RPKM-normalized coverage using bamCoverage with the parameters "--outFileFormat bigwig --normalizeUsing RPKM --binSize 1". For analysis of transcription termination, 500bp upstream of the transcription termination site from longest isoform per gene was used for the annotation file for meta plots and the following parameters were used for computeMatrix "--binSize 1 --regionBodyLength 300 --downstream 1000" to generate a matrix of coverages. For analysis of minor introns, the following parameters were used for computeMatrix "--binSize 1 --regionBodyLength 100 --upstream 100 --downstream 100" with the minor intron annotation file from Minor Intron Database v1.2 (*166*). Normalized coverage per bin was obtained using plotProfile with the parameter "--averageType mean" and the average coverage of two biological replicates was plotted. For minor intron analysis, the RPKM-normalized coverage in all bins was further normalized to the summed flanking exon coverage to correct for the decreased mRNA abundance of minor intron containing genes in RNPC3 knockouts.

2.13.15 Reverse transcription and qPCR

Total RNA was purified as described for RNA-sequencing. 1 µg of total RNA was used in a cDNA synthesis reaction with the Maxima First Strand cDNA synthesis kit (Thermo Scientific) according to the manufacturer's protocol. The cDNA was subjected to qPCR using the PowerUP SYBR Green Master Mix (Thermo Fisher) according to the manufacturer's protocol or end point PCR using 2x Q5 polymerase mix (NEB). For qPCR, a standard curve was used for quantitative assessment of mRNA levels and normalized to GAPDH mRNA. Myc primers: 5'CCTTCTCTCCGTCCTCGGAT3' and 5'CTTCTTGTTCCTCCTCAGAGTCG3'; SPC24 primers: 5'GGCTCAACTTTACCACCAAGTTAG3' and 5'CACCAGACTCCAGAGGTAGTCG3; GAPDH primers: 5'TCGGAGTCAACGGATTTGGT3' and 5'TTCCCGTTCTCAGCCTTGAC3'.

2.14 Supplemental figures

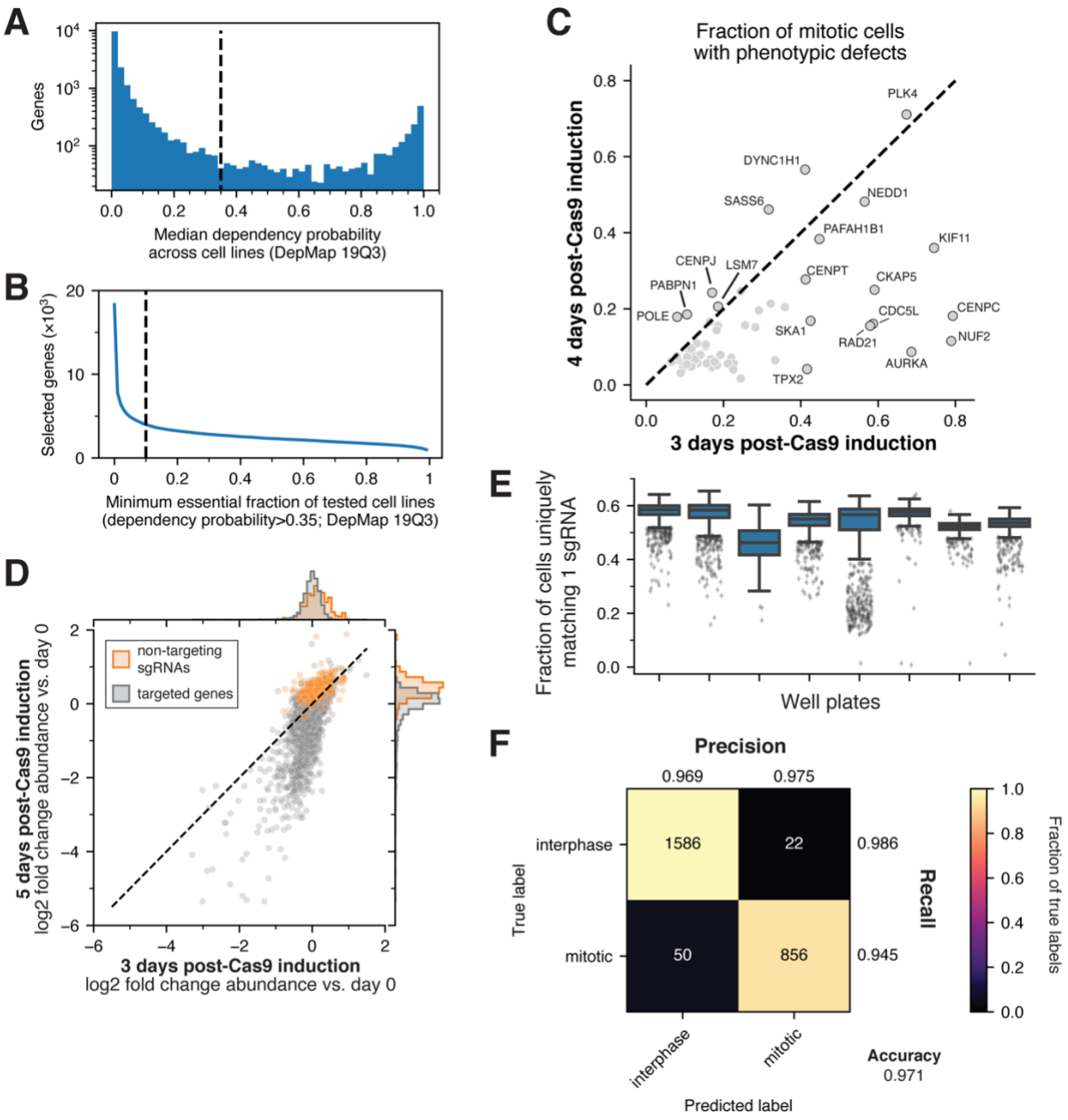

Figure 2.S1. Optimization of image-based pooled screening for essential gene function. (**A**) Histogram of median dependency probability across cell lines in the DepMap dataset, indicating the threshold chosen (0.35) for defining individual essential cell lines for each gene in (B). (**B**) Number of genes identified as essential for at least the specified fraction of tested cell lines in DepMap. Genes that were essential in at least 10% of tested cell lines were selected for the screen, in addition to those selected from additional sources (Section 2.13.1). (**C**) Scatter plot showing the results from trial screens of 400 gene targets. This compares the fraction of mitotic

cells with visually-identified phenotypic defects for established cell division factors at 3 and 4 days post-Cas9 induction. Overall, mitotic phenotypes were more commonly observed at the earlier time point. (**D**) Scatter plot showing mean change in abundance within the 20,445 sgRNA primary screen library at 3 and 5 days post-Cas9 induction, both time points relative to pre-induction (day 0). N=2 screen replicates were performed, averaged across sgRNAs targeting the same gene. Orange indicates non-targeting control sgRNAs. Many gene targets begin to drop out of the population at day 5 due to fitness defects. Based on this data and from (C), 78 hours post-Cas9 induction was chosen as the fixation time point for our image-based screen to maximize observable phenotypes. (**E**) Boxplot demonstrating *in situ* sequencing quality in our fixed-cell image-based pooled screen. Sequencing quality was consistent across the eight imaging plates, with the majority of imaging tiles exceeding 50% of cells with sequencing reads that uniquely match a single sgRNA sequence from the library. N = 1,665 or 1,998 imaging tiles in each plate column. (**F**) Confusion matrix demonstrating performance of the support vector classifier in distinguishing interphase and mitotic cells, 5-fold cross-validation with N=2,514 manually annotated cell images.

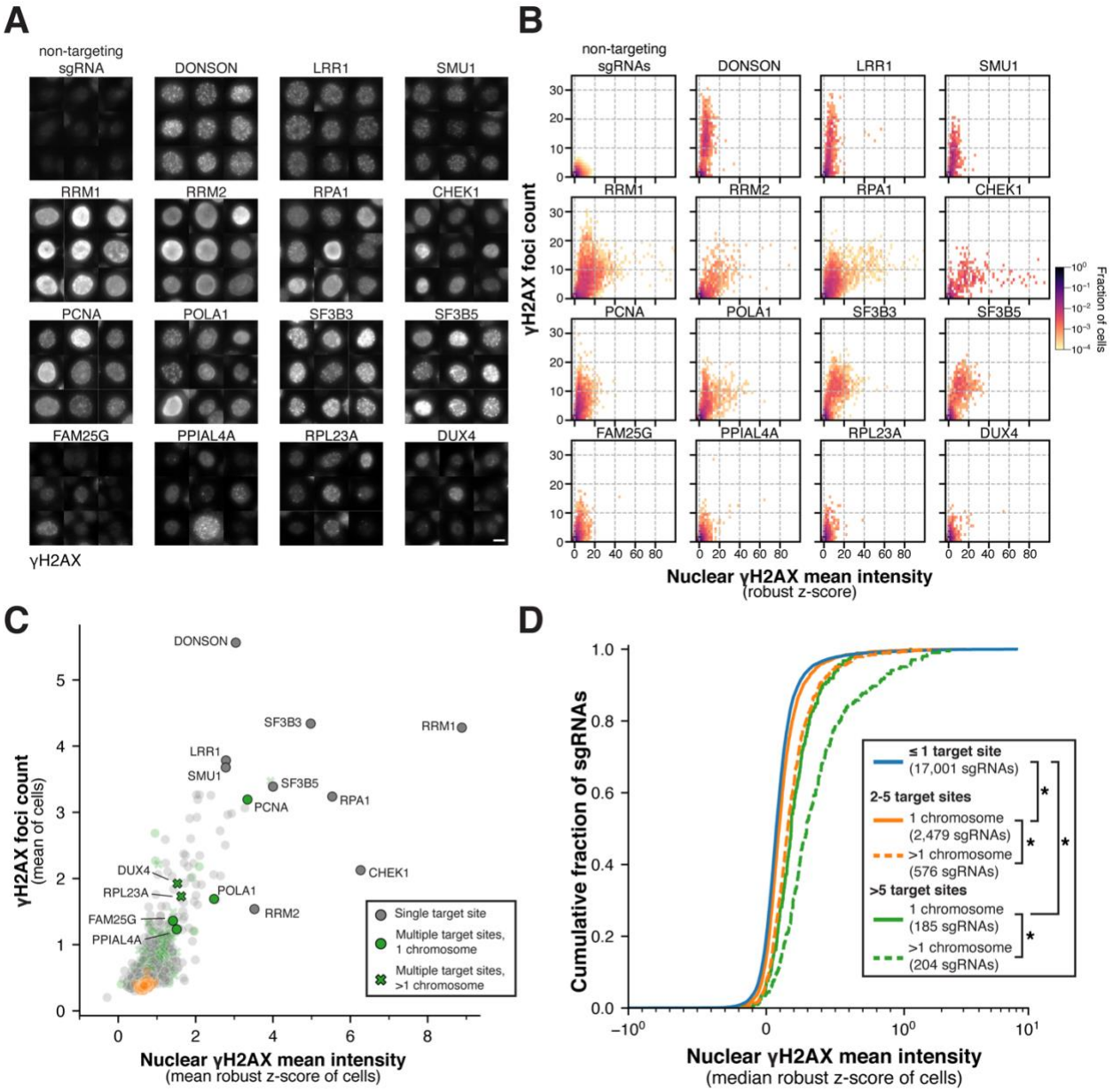

Figure 2.S2. γH2AX localization and Cas9-induced DNA damage. (**A**) Example cell images demonstrating punctate (DONSON, LRR1, SMU1, etc.) and diffuse (e.g., RRM1/2, RPA1, CHEK1) γH2AX staining patterns. Scale bar, 10 μm. (**B**) Bivariate histograms of γH2AX foci counts and mean nuclear γH2AX intensity, displaying single-cell distributions for cells expressing non-targeting sgRNAs (top left) and selected gene targets (corresponding to A) showing increased γH2AX foci, nuclear γH2AX mean intensity, or both. Histogram bins containing less than 1 in 10^4 total cells for a given gene target are not displayed. (**C**) Scatter plot comparing gene-level summaries of γH2AX foci counts to nuclear γH2AX mean stain intensity. A minority of genes demonstrate separate increases in either diffuse γH2AX staining or γH2AX foci counts. Genes

with at least 2 sgRNAs targeting multiple sites within one chromosome (green circles) or across multiple chromosomes (green crosses) demonstrate similar overall trends as genes targeted by single target site sgRNAs (gray circles). (**D**) Cumulative distribution plots of mean nuclear γH2AX intensity (DNA damage phenotype) sgRNA scores, with sgRNAs grouped by number and location of target sites. Non-targeting control sgRNAs and sgRNAs targeting a single genomic locus (blue) include the vast majority of sgRNAs in the library and displayed minimal DNA damage on average in the screen. In contrast, sgRNAs with increasing numbers of target sites (orange, green) tend to display stronger DNA damage phenotypes, in particular when the target sites are spread across multiple chromosomes (dotted lines). Genomic target sites are defined as the total number of cutting frequency determination (CFD) bin 1 matches (see *19*). *$P<10^{-9}$ by 1-sided Mann-Whitney U test.

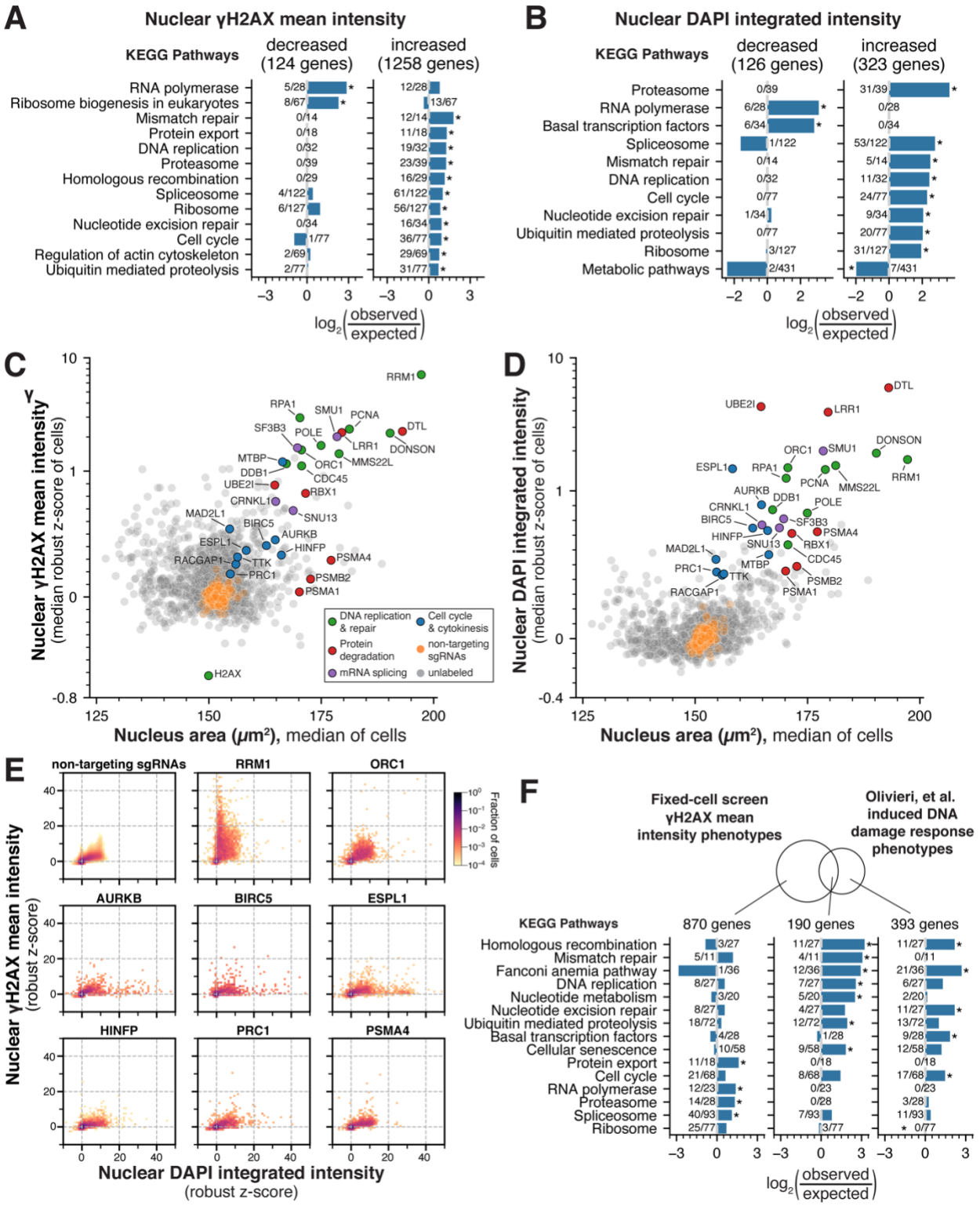

Figure 2.S3. Analysis of interphase nuclear phenotypes. (**A**) Bar graph indicating over-representation of KEGG pathways among gene targets exhibiting decreased or increased nuclear γH2AX mean intensity. *FDR<0.05 (**B**) Bar graph of over-representation analysis results as in (A)

among gene targets with decreased or increased nuclear DNA (DAPI) integrated intensity. *FDR<0.05 (**C**) Scatter plot showing summary gene scores (see Section 2.13.8) for mean nuclear γH2AX intensity compared to nuclear area, showing a subset of gene knockouts with increases in both γH2AX and nuclear area. Summary γH2AX scores are plotted on a symmetric log scale (linear between -1 and 1) and labeled genes are colored by functional category. (**D**) Scatter plot showing summary gene scores for integrated nuclear DNA (DAPI) intensity compared to nuclear area as in (C). DNA content is relatively constant across gene targets exhibiting a range of nuclear areas, although a subset demonstrates increased nuclear area and DNA. Summary DAPI scores are plotted on a symmetric log scale (linear between -1 and 1) and labeled genes are colored by functional category. (**E**) Bivariate histograms of integrated nuclear DNA intensity and mean nuclear γH2AX intensity, displaying single-cell distributions for all cells expressing non-targeting sgRNAs (top left) and selected gene targets. Knockouts of genes that regulate chromosome segregation or cytokinesis result in more cells with increased DNA content, but only modest increases in γH2AX intensity. Histogram bins containing less than 1 in 10^4 of the total cells for a given gene target are not displayed. (**F**) Comparison of γH2AX phenotype genes identified in our screen with the DNA damage-associated genes identified in Olivieri, *et al*. (*101*). Bottom, bar graph of KEGG pathway over-representation in the indicated gene sets against the background of all genes present in both datasets.

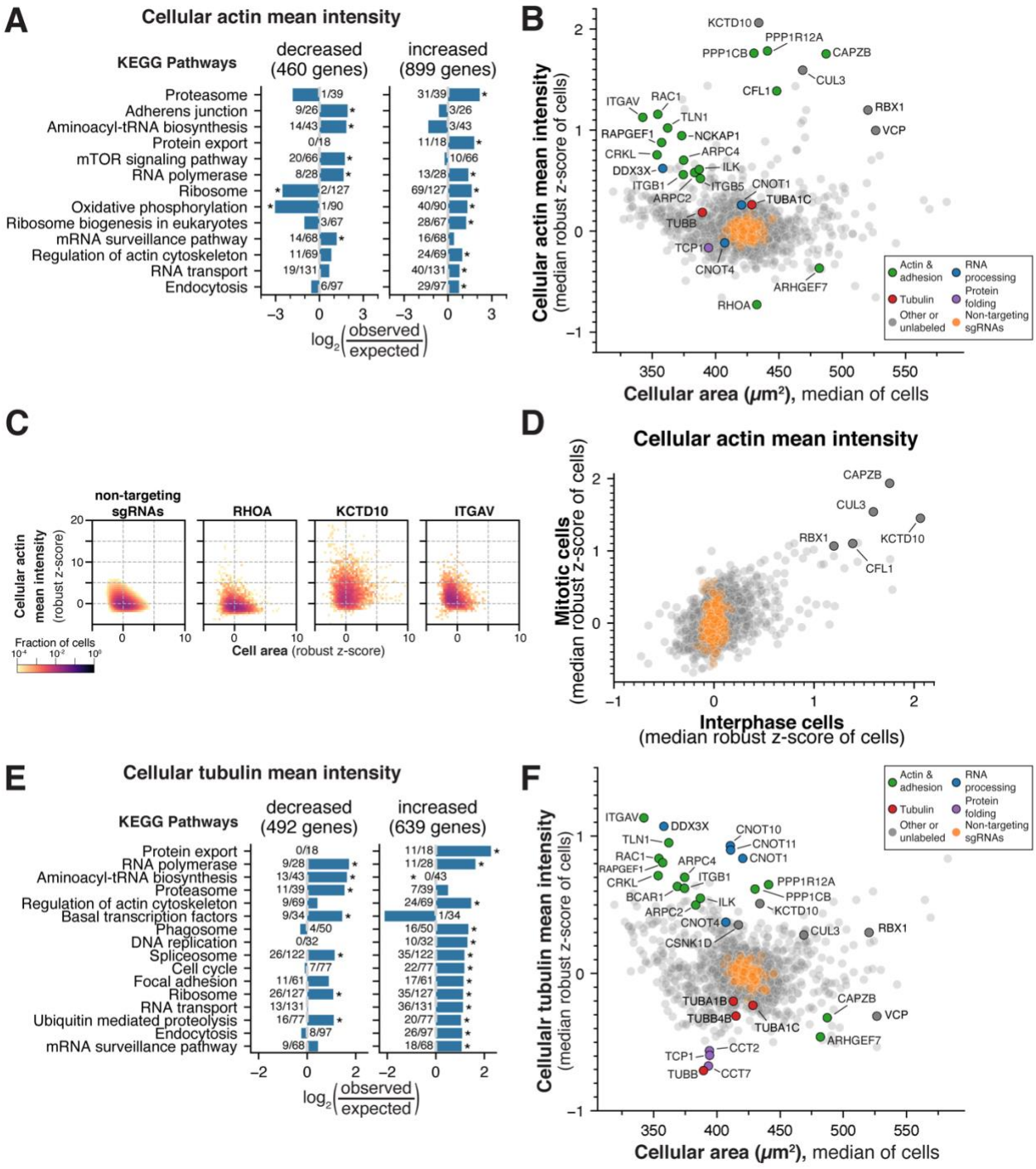

Figure 2.S4. Analysis of interphase cytoskeletal phenotypes. (**A**) Bar graph indicating over-representation of KEGG pathways among gene targets with decreased or increased mean cellular actin (phalloidin) intensity in interphase cells. *FDR<0.05 (**B**) Scatter plot indicating summary gene scores (see Section 2.13.8) for mean cellular actin intensity compared to cell area. A subset of gene knockouts display increased actin staining together with decreased cell area due to disrupted cellular adhesion. Labeled genes are colored by functional category. (**C**) Bivariate

histograms of mean cellular actin intensity and cellular area, displaying single-cell distributions for all cells expressing non-targeting sgRNAs (left) and selected gene targets. Knockouts of genes that regulate cellular adhesion (e.g., ITGAV) show a distribution of cells shifted toward lower cellular area and correspondingly increased mean actin intensity. Histogram bins containing less than 1 in 10^4 total cells for a given gene target are not displayed. (**D**) Scatter plot comparing mean cellular actin intensity summary scores between interphase and mitotic cell populations, indicating factors that robustly affect actin structures throughout the cell cycle. (**E**) Bar graph of KEGG pathway over-representation analysis results as in (A) for gene targets demonstrating decreased or increased mean cellular tubulin intensity. (**F**) Scatter plot showing summary gene scores for mean cellular tubulin intensity compared to cell area as in (B). Similar to actin, a subset of gene knockouts display increased tubulin staining in combination with decreased cell area.

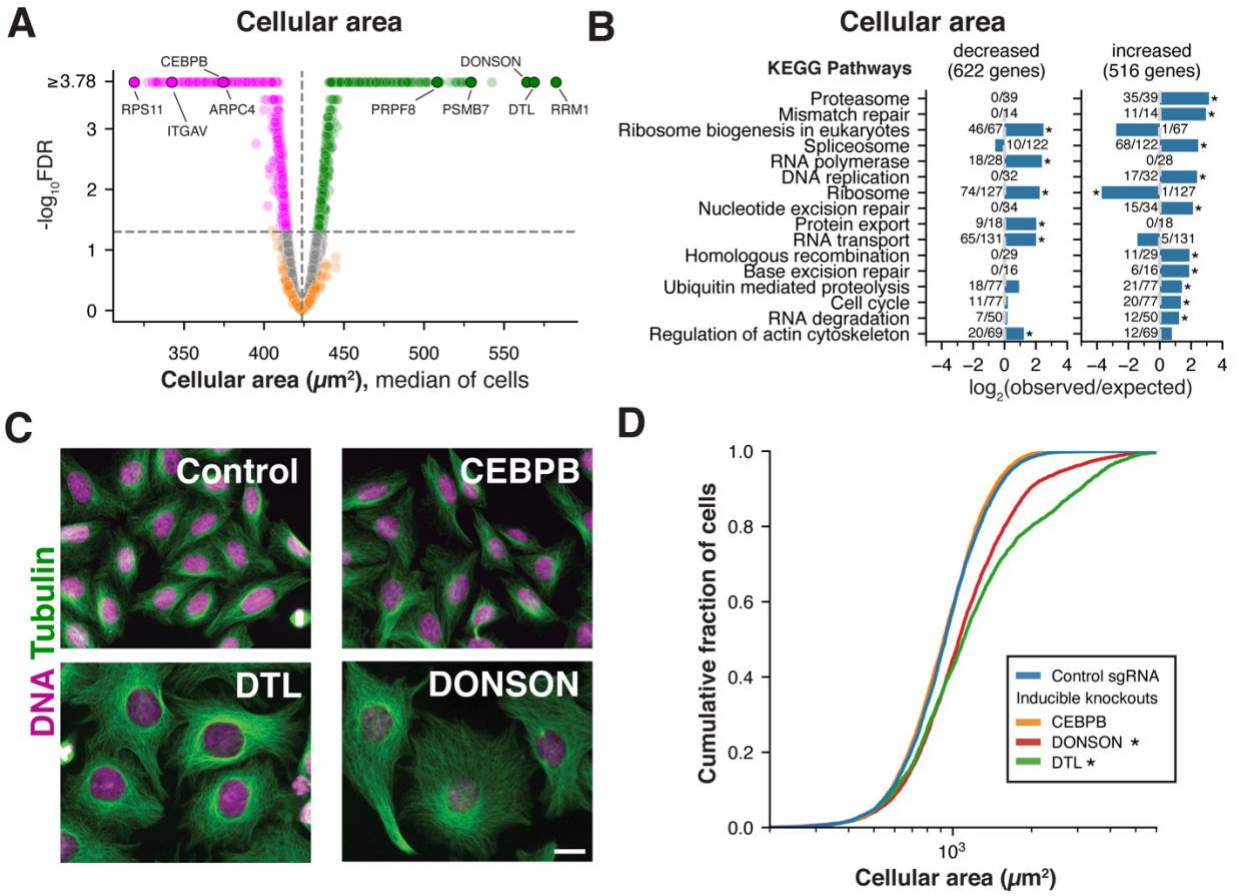

Figure 2.S5. Analysis and validation of cell area changes. (**A**) Volcano plot for interphase cell area across gene targets in the screen, showing a wide range of decreased (magenta) and increased (green) cell areas (FDR<0.05). Raw P-values were computed by comparing gene targets to a bootstrapped null distribution of cells expressing non-targeting sgRNAs (Section 2.13.8), with false discovery rate (FDR) estimated using the Benjamini-Hochberg procedure. (**B**) Bar graph of over-representation analysis for gene targets that result in decreased or increased cell area. *FDR<0.05. (**C**) Example images and (**D**) cumulative distributions of inducible knockout cells validating increased cell area for DTL and DONSON knockouts. *P<10^{-10} by Mann-Whitney U test relative to control sgRNA, N>3,700 cells per gene target. Scale bar, 10 μm.

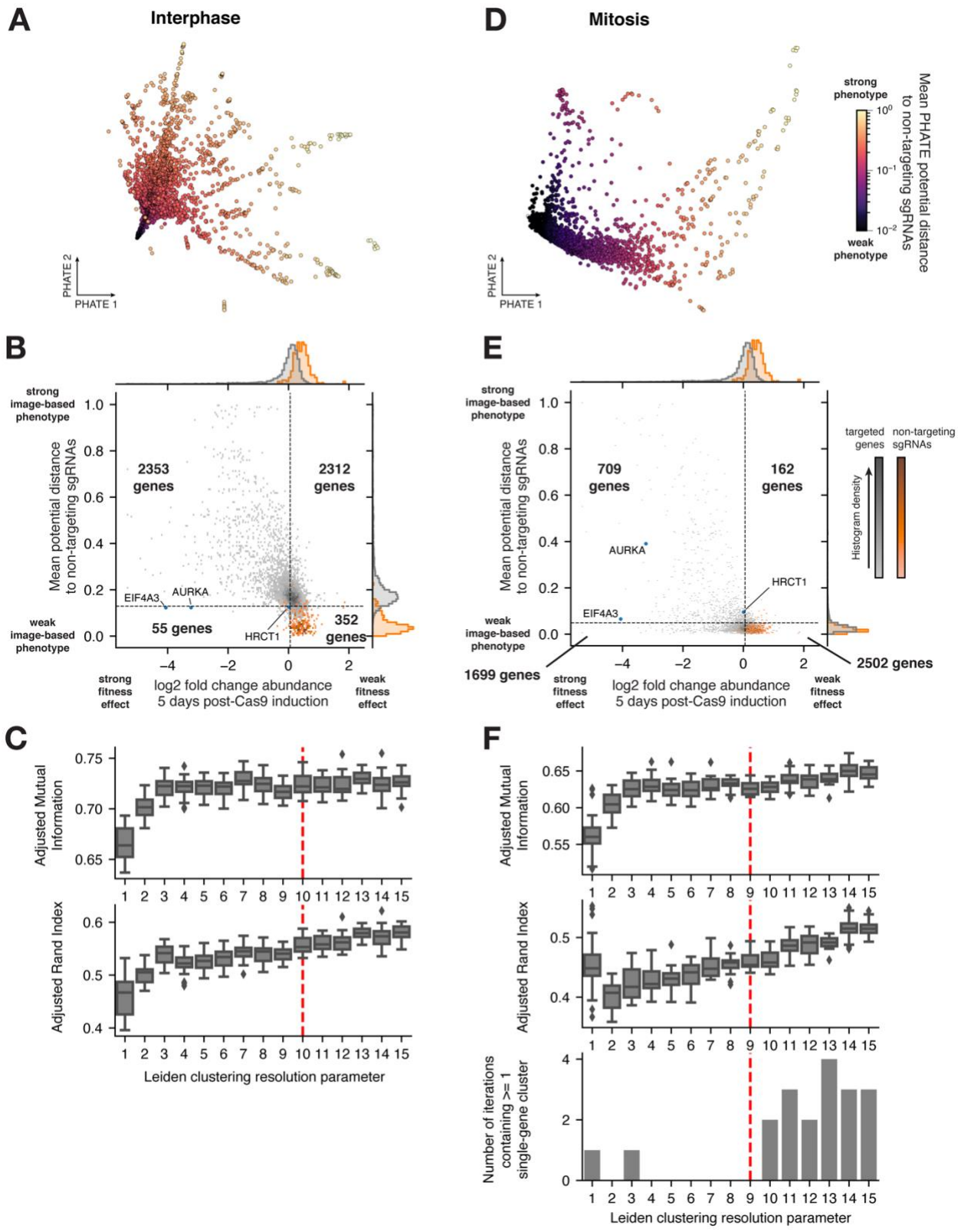

Figure 2.S6. PHATE analysis of multi-dimensional phenotypes. (**A**) Two-dimensional representation of the interphase phenotype landscape of gene targets in the primary screen using

PHATE (35; Section 2.13.8). Data points are colored by the mean potential distance to non-targeting control sgRNAs computed by PHATE for the interphase phenotype profiles, normalized between 0 and 1. Gene targets with increasing phenotype strength intuitively radiate outward from a dense region containing non-targeting sgRNAs. (**B**) Bivariate histogram showing the joint distribution of image-based interphase phenotype strength from (A) together with the strength of knockout fitness effect in the screening cell line (mean change in sgRNA abundance within the library after 5 days of Cas9 induction, N=2 screen replicates averaged across sgRNAs targeting the same gene, data from Fig. 2.S1D). >90% of gene targets exhibit a measurable interphase phenotype in the image-based screen (potential distance greater than the 95th percentile of non-targeting sgRNAs). Of the remaining 407 genes, only 55 demonstrate a meaningful fitness effect in the tested cell line (log2 fold change abundance less than the 5th percentile of non-targeting sgRNAs). Labeled genes are those that display a fitness effect and no interphase phenotype, but do show a measurable mitotic phenotype in (E). (**C**) Boxplot illustrating selection of the Leiden clustering (64) resolution parameter for interphase phenotypes. For 20 repetitions at each resolution, 90% of gene targets were sampled without replacement and clustered using PHATE and Leiden algorithms in series. Each sampled cluster solution was then compared to the full dataset clusters using adjusted mutual information (top) and adjusted Rand index (bottom). The dotted line indicates the chosen resolution, selected based on the plateau in robustness of clustering solutions. (**D**) Two dimensional representation of the mitotic phenotype landscape of gene targets as in (A), colored by the mean potential distance to non-targeting control sgRNAs computed by PHATE for the mitotic phenotype profiles. (**E**) Bivariate histogram showing the joint distribution of image-based mitotic phenotype strength from (D) together with the strength of knockout fitness effect in the screening cell line as in (B). The threshold for measurable mitotic phenotypes is the 95th percentile of potential distance among non-targeting control sgRNAs. Labeled genes indicate those that display a fitness effect and mitotic phenotype, but do not exhibit an interphase phenotype in (B). (**F**) Boxplot illustrating selection of the Leiden clustering resolution parameter for mitotic phenotypes using the same procedure as (C). A resolution parameter of 9 was chosen in part due to the increased presence of single-gene clusters with resolution ≥10 (bottom).

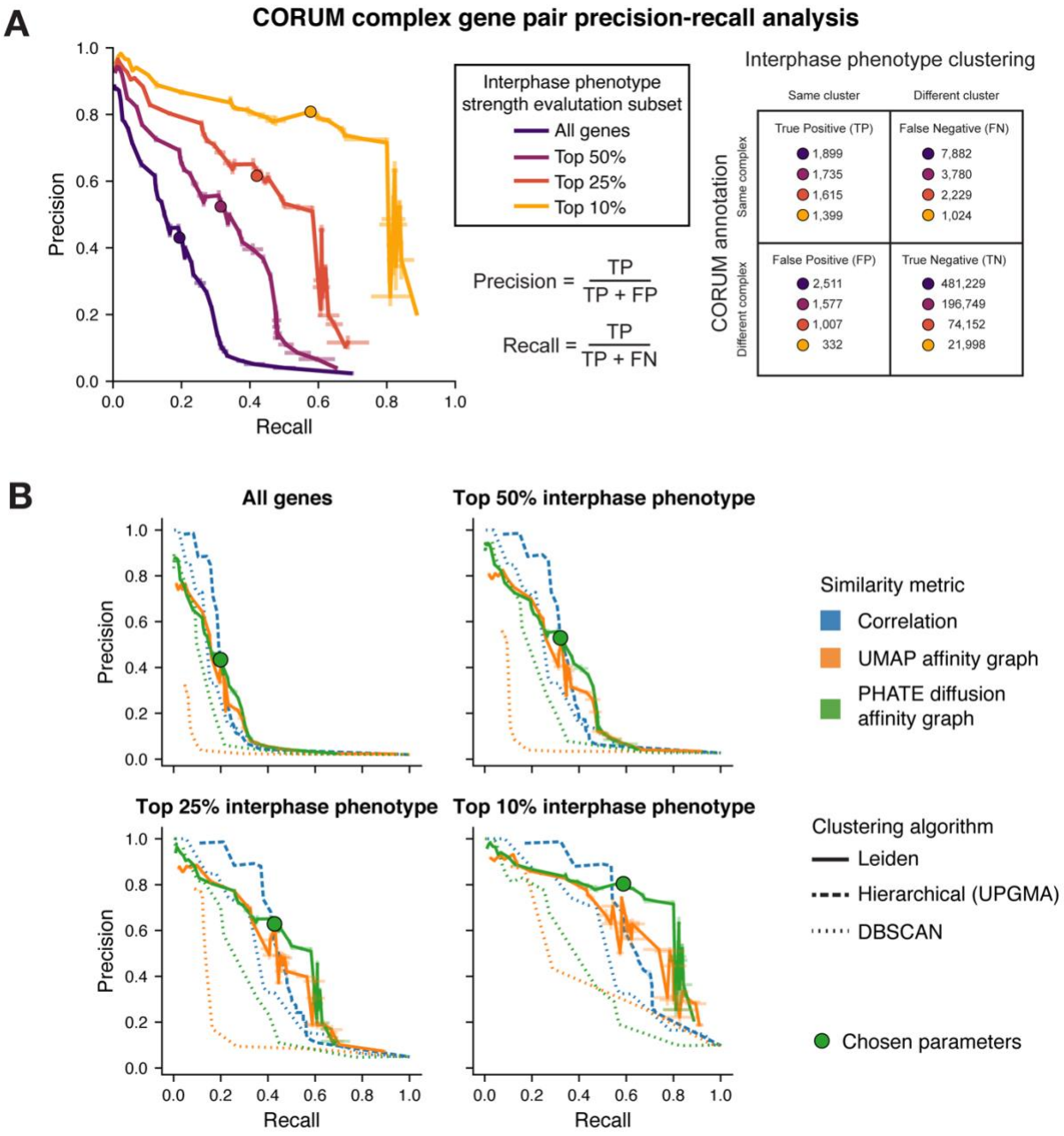

Figure 2.S7. Precision-recall analysis of interphase phenotype clustering with CORUM complex gene pairs. (**A**) Precision-recall curves for clustering gene pairs from a subset of 292 minimal CORUM complexes using the Leiden clustering algorithm with the PHATE diffusion affinity graph (Section 2.13.8), restricted to varying overall interphase phenotype strength. Error bars indicate the standard deviation of 10 runs with different random seeds. When evaluating gene pairs within the top 10% of phenotype strength, this approach achieves gene pair recall of 59% and precision of 81%. (**B**) Precision-recall curves as in (A) for alternative clustering approaches (Section 2.13.8). At higher recall (>0.2 when evaluating across the full dataset),

Leiden clustering with the PHATE diffusion affinity graph achieves the highest precision. At lower recall, hierarchical clustering of phenotype profile correlations demonstrates the highest precision. Error bars for PHATE and UMAP similarity metrics indicate standard deviation of 10 runs with different random seeds.

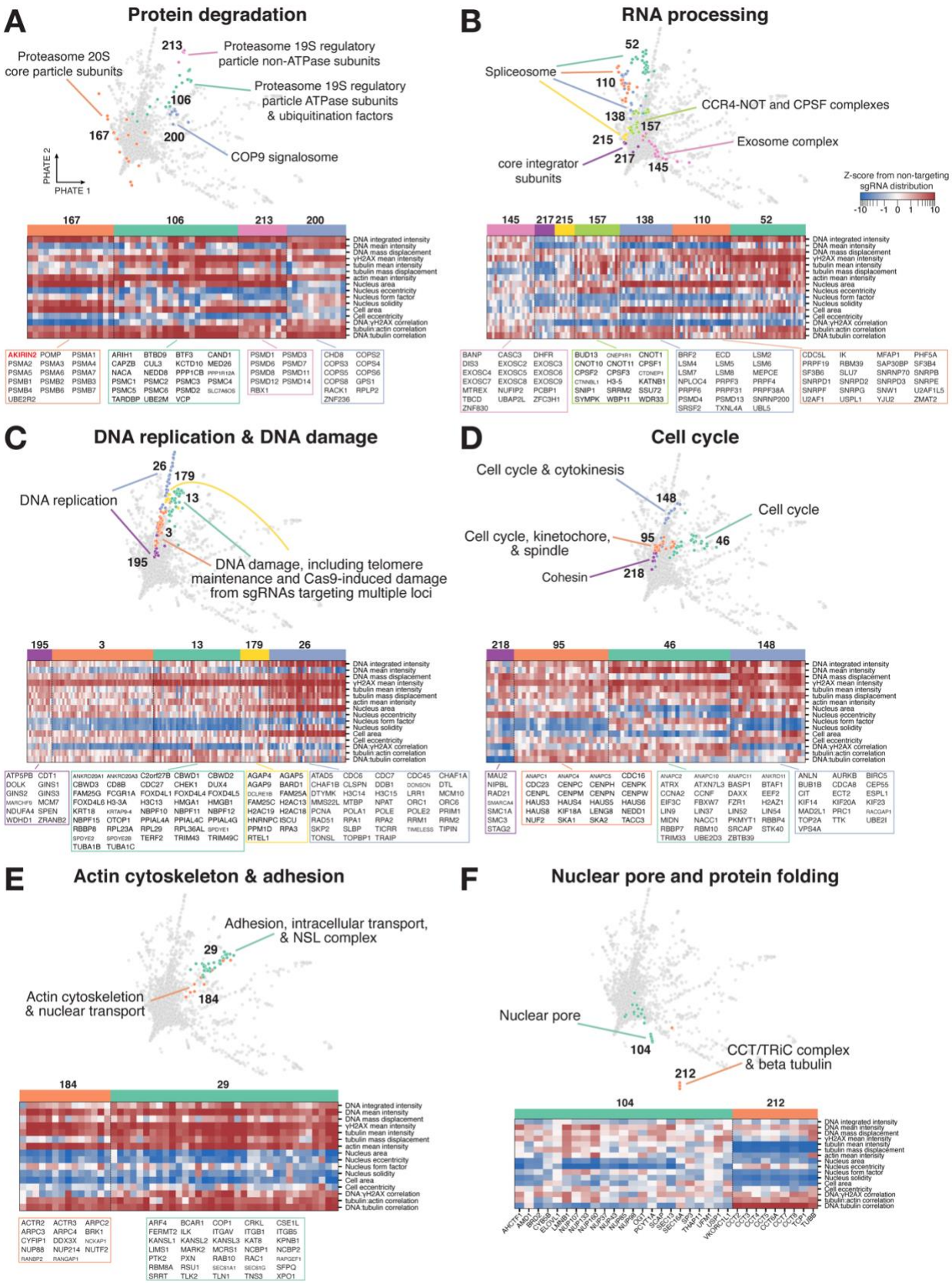
A
Protein degradation
Proteasome 20S core particle subunits
213
Proteasome 19S regulatory particle non-ATPase subunits
Proteasome 19S regulatory particle ATPase subunits & ubiquitination factors
106
167
200
COP9 signalosome
PHATE 2
PHATE 1
167 106 213 200
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean intensity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation
AKIRIN2 POMP PSMA1
PSMA2 PSMA3 PSMA4
PSMA5 PSMA6 PSMA7
PSMB1 PSMB2 PSMB3
PSMB4 PSMB6 PSMB7
UBE2R2
ARIH1 BTBD9 BTF3 CAND1
CAP2B CUL3 KCTD10 MED26
NACA NEDD8 PPP1CB PPP1R10A
PSMC1 PSMC2 PSMC3 PSMC4
PSMC5 PSMC6 PSMD2 SLC7A6OS
TARDBP UBE2M VCP
PSMD1 PSMD3
PSMD6 PSMD7
PSMD8 PSMD11
PSMD12 PSMD14
RBX1
CHD8 COPS2
COPS3 COPS4
COPS5 COPS6
COPS8 GPS1
RACK1 RPLP2
ZNF236
B
RNA processing
Spliceosome
52
110
138
157
CCR4-NOT and CPSF complexes
215
217
145
Exosome complex
core integrator subunits
Z-score from non-targeting sgRNA distribution
-10 -1 0 1 10
145 217 215 157 138 110 52
BANP CASC3 DHFR
DIS3 EXOSC2 EXOSC3
EXOSC4 EXOSC5 EXOSC8
EXOSC7 EXOSC8 EXOSC9
MTREX NUFIP2 PCBP1
TBCD UBAP2L ZFC3H1
ZNF830
BUD13 CNEPHR1 CNOT1
CNOT10 CNOT11 CPSF1
CPSF2 CPSF3 CTDNEP1
CTNNBL1 H3-5 KATNB1
SNIP1 SRRM2 SSU72
SYMPK WBP11 WDR33
BRF2 ECD LSM2
LSM4 LSM5 LSM6
LSM7 LSM8 MEPCE
NPLOC4 PRPF3 PRPF4
PRPF6 PRPF31 PRPF38A
PSMD4 PSMD13 SNRNP200
SRSF2 TXNL4A UBL5
CDC5L IK MFAP1 PHF5A
PRPF19 RBM39 SAP30BP SF3B4
SF3B6 SLU7 SNRNP70 SNRPB
SNRPD1 SNRPD2 SNRPD3 SNRPE
SNRPF SNRPG SNW1 U2AF1L5
U2AF1 USPL1 YJU2 ZMAT2
C
DNA replication & DNA damage
26
179
DNA replication
13
3
195
DNA damage, including telomere maintenance and Cas9-induced damage from sgRNAs targeting multiple loci
195 3 13 179 26
ATP5PB CDT1
DOLK GINS1
GINS2 GINS3
MARCHF9 MCM7
NDUFA4 SPEN
WDHD1 ZRANB2
ANKRD20A1 ANKRD20A3 C2orf27B CBWD1 CBWD2
CBWD3 CD8B CDC27 CHEK1 DUX4
FAM25G FCGR1A FOXD4L1 FOXD4L4 FOXD4L5
FOXD4L6 H3-3A H3C13 HMGA1 HMGB1
KRT18 KRTAP9-4 NBPF10 NBPF11 NBPF12
NBPF15 OTOP1 PPIAL4A PPIAL4C PPIAL4G
RBBP8 RPL23A RPL29 RPL36AL SPDYE1
SPDYE2 SPDYE2B TERF2 TRIM43 TRIM49C
TUBA1B TUBA1C
AGAP4 AGAP5
AGAP9 BARD1
DCLRE1B FAM25A
FAM25C H2AC13
H2AC19 H2AC18
HNRNPC ISCU
PPM1D RPA3
RTEL1
ATAD5 CDC6 CDC7 CDC45 CHAF1A
CHAF1B CLSPN DDB1 DONSON DTL
DTYMK H3C14 H3C15 LRR1 MCM10
MMS22L MTBP NPAT ORC1 ORC6
PCNA POLA1 POLE POLE2 PRIM1
RAD51 RPA1 RPA2 RRM1 RRM2
SKP2 SLBP TICRR TIMELESS TIPIN
TONSL TOPBP1 TRAIP
D
Cell cycle
Cell cycle & cytokinesis
148
Cell cycle
Cell cycle, kinetochore, & spindle
95
46
Cohesin
218
218 95 46 148
MAU2
NIPBL
RAD21
SMARCA4
SMC1A
SMC3
STAG2
ANAPC1 ANAPC4 ANAPC5 CDC16
CDC23 CENPC CENPH CENPK
CENPL CENPM CENPN CENPW
HAUS3 HAUS4 HAUS5 HAUS6
HAUS8 KIF18A LENG8 NEDD1
NUF2 SKA1 SKA2 TACC3
ANAPC2 ANAPC10 ANAPC11 ANKRD11
ATRX ATXN7L3 BASP1 BTAF1
CCNA2 CCNF DAXX EEF2
EIF3C FBXW7 FZR1 H2AZ1
LIN9 LIN37 LIN52 LIN54
MIDN NACC1 PKMYT1 RBBP4
RBBP7 RBM10 SRCAP STK40
TRIM33 UBE2D3 ZBTB39
ANLN AURKB BIRC5
BUB1B CDCA8 CEP55
CIT ECT2 ESPL1
KIF14 KIF20A KIF23
MAD2L1 PRC1 RACGAP1
TOP2A TTK UBE2I
VPS4A
E
Actin cytoskeleton & adhesion
Adhesion, intracellular transport, & NSL complex
29
Actin cytoskeletion & nuclear transport
184
184 29
ACTR2 ACTR3 ARPC2
ARPC3 ARPC4 BRK1
CYFIP1 DDX3X NCKAP1
NUP88 NUP214 NUTF2
RANBP2 RANGAP1
ARF4 BCAR1 COPI CRKL CSE1L
FERMT2 ILK ITGAV ITGB1 ITGB5
KANSL1 KANSL2 KANSL3 KAT8 KPNB1
LIMS1 MARK2 MCRS1 NCBP1 NCBP2
PTK2 PXN RAB10 RAC1 RAPGEF1
RBM6A RSU1 SEC24A1 SEC24C SFPQ
SRRT TLK2 TLN1 TNS3 XPO1
F
Nuclear pore and protein folding
Nuclear pore
104
CCT/TRiC complex & beta tubulin
212
104 212
AHCTF1 AMD1 BRD8 CYB5B DVL1 ELMB1 LMNB1 NUP107 NUP133 NUP153 NUP35 NUP43 NUP50 OGT PCF11A SUMO1 SEC13 SPC1 THAP1 UFM1 USP11 WDR61 CCT2 CCT3 CCT4 CCT5 CCT6A CCT7 CCT8 TCP1 TUBB
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean intensity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation

Figure 2.S8. Interphase phenotypes enable detailed clustering of specific functional categories. Two-dimensional PHATE representations of interphase phenotype clusters and corresponding heat maps of a manually-selected subset of specific parameters for a broad range of functional categories, as in Fig. 2.3. Distinct and coherent phenotypes are observable for genes involved in processes such as (**A**) protein degradation (including the recently-characterized gene AKIRIN2); (**B**) RNA processing; (**C**) DNA replication and DNA damage; (**D**) cell cycle function; (**E**) actin cytoskeleton and cellular adhesion; and (**F**) nuclear pore function and protein folding. Numbers indicate individual interphase cluster identities. All genes from selected clusters are listed below each heatmap. Parameters are presented as z-scores from the distribution of non-targeting sgRNAs, visualized on a symmetric log scale (linear between -1 and 1).

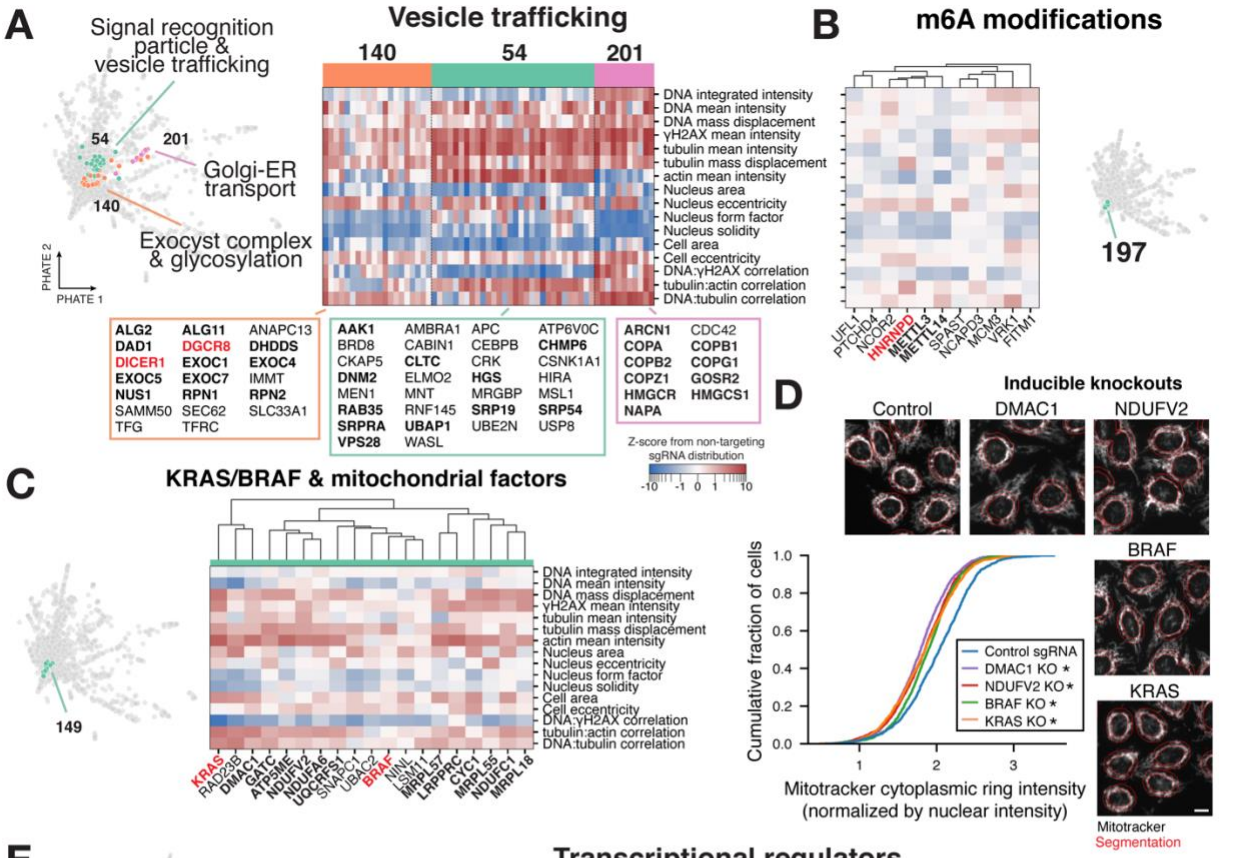
A
Signal recognition particle & vesicle trafficking
Golgi-ER transport
Exocyst complex & glycosylation
54
201
140
PHATE 2
PHATE 1
Vesicle trafficking
140
54
201
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean intensity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation
ALG2 ALG11 ANAPC13
DAD1 DGCR8 DHDDS
DICER1 EXOC1 EXOC4
EXOC5 EXOC7 IMMT
NUS1 RPN1 RPN2
SAMM50 SEC62 SLC33A1
TFG TFRC
AAK1 AMBRA1 APC ATP6V0C
BRD8 CABIN1 CEBPB CHMP6
CKAP5 CLTC CRK CSNK1A1
DNM2 ELMO2 HGS HIRA
MEN1 MNT MRGBP MSL1
RAB35 RNF145 SRP19 SRP54
SRPRA UBAP1 UBE2N USP8
VPS28 WASL
ARCN1 CDC42
COPA COPB1
COPB2 COPG1
COPZ1 GOSR2
HMGCR HMGCS1
NAPA
Z-score from non-targeting sgRNA distribution
-10 -1 0 1 10
B
m6A modifications
197
UFL1 PTCHD4 NCOR2 HNRNPD METTL3 METTL14 SPATS2 NCAPD3 MCM3 VRK1 FITM1
C
KRAS/BRAF & mitochondrial factors
149
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean interisity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation
KRAS RAD23B DMAC1 GATC ATP5ME NDUFI2 NDUFA6 UQCRFS1 SNAPC1 UBAC2 BRAF NINL MRPL37 LRPPRC CYC1 MRPL55 NDUFC1 MRPL18
D
Inducible knockouts
Control
DMAC1
NDUFV2
BRAF
KRAS
Cumulative fraction of cells
Control sgRNA
DMAC1 KO *
NDUFV2 KO *
BRAF KO *
KRAS KO *
Mitotracker cytoplasmic ring intensity (normalized by nuclear intensity)
Mitotracker
Segmentation
E
Transcriptional regulators
121
19
146
39
MYC-related
Other transcriptional regulators
DNA integrated intensity
DNA mean intensity
DNA mass displacement
γH2AX mean intensity
tubulin mean intensity
tubulin mass displacement
actin mean intensity
Nucleus area
Nucleus eccentricity
Nucleus form factor
Nucleus solidity
Cell area
Cell eccentricity
DNA:γH2AX correlation
tubulin:actin correlation
DNA:tubulin correlation
CCDC174 DDA1 FAM32A
FOXN1 GABPA GABPB1
HNRNPM ILF3 KEAP1
MAX MEMO1 MYC
N6AMT1 POU5F1B RAB11FIP4
SETD2 SLC38A2 SON
SP2 SPAG5 ZBTB11
ZMYND8
AKAP8L AP1M1 ARID3A ASF1B
BORCS8 C19orf67 CCDC130 CRLF1
CRTC1 DCAF15 DHX9 DHX35
EP300 FKBP8 GPATCH1 HOMER3
ISYNA1 JUND KXD1 MAST3
MEF2B NCAN NR2C2AP PDE4C
PGPEP1 PIK3R2 RFXANK SEMA6B
SLC25A42 SRRM1 SRSF11 SSBP4
TNRC6A UNC13A WDR83 ZC3H18
ZNF333 ZNF431 BORCS8-MEF2B
ATP13A1 DDI2
DNAJB11 JOSD1
MAEA NFE2L1
NFYA NFYB
NFYC RB1CC1
RUNX1 TM2D1
TM2D2 TM2D3
TOX4 UBE2H
UBE4B WDR26
YPEL5
AP2M1 AP2S1 ARGLU1
ARID4B ARMC7 CCNK
CDK7 ELL GRB2
HNRNPH1 JUNB KLF5
KLF16 MED12 MED15
MED23 PDCD7 PPP6C
PPP6R3 PRPF40A RBM25
RNPC3 SFSWAP SNRNP25
SNRNP35 SNRNP48 SRSF7
TBP TCF3 WDR25
ZMAT5 ZNF207 ZRSR2
F
Inducible knockouts
Control sgRNA
MAX
MYC
KEAP1
SETD2
SP2
ZBTB11
ZMYND8
MYC
GAPDH
G
MYC Expression
mRNA (RT-PCR)
Protein (Western)
Relative Level (%)
150
100
50
0
Control sgRNA
MAX
MYC
KEAP1
SETD2
SP2
ZBTB11
ZMYND8
Inducible knockouts

Figure 2.S9. Interphase cluster analysis reveals novel functional associations for established factors. Two-dimensional PHATE representations of interphase phenotype clusters and corresponding heat maps of a manually-selected subset of specific parameters, as in Fig. 2.3. (**A**) Gene targets involved in vesicle trafficking and related processes exhibit distinct image-based phenotypes, despite the absence of membrane-targeted phenotype stains in the screen. (**B**) Phenotype clustering suggests a co-functional role of HNRNPD with m6A modifications, as well as (**C**) a relationship between mitochondrial function and KRAS/BRAF signaling. (**D**) Example images and cumulative distributions of MitoTracker staining for control and knockout cells for KRAS, BRAF, and two co-clustering mitochondrial factors. Each knockout demonstrates disrupted mitochondrial content as compared to control cells, supporting the phenotypic association of KRAS and BRAF with mitochondrial factors in the primary screen. *$P<10^{10}$ by Mann-Whitney U test relative to control sgRNA, N>1,400 cells per gene target. Scale bar, 10 μm. (**E**) Transcriptional regulators show interrelated phenotypes, with an apparent distinct phenotype for cluster 121 containing MYC and MAX, suggesting the presence of additional MYC-associated factors. *P<0.05 by two-tailed independent T-test relative to corresponding controls. (**F**) Western blot and quantification (**G**) of MYC mRNA and protein expression following knockout of several genes from cluster 121 in (**E**). SETD2 demonstrates substantially decreased mRNA and protein expression of MYC, confirming the functional association identified via interphase phenotype clustering.

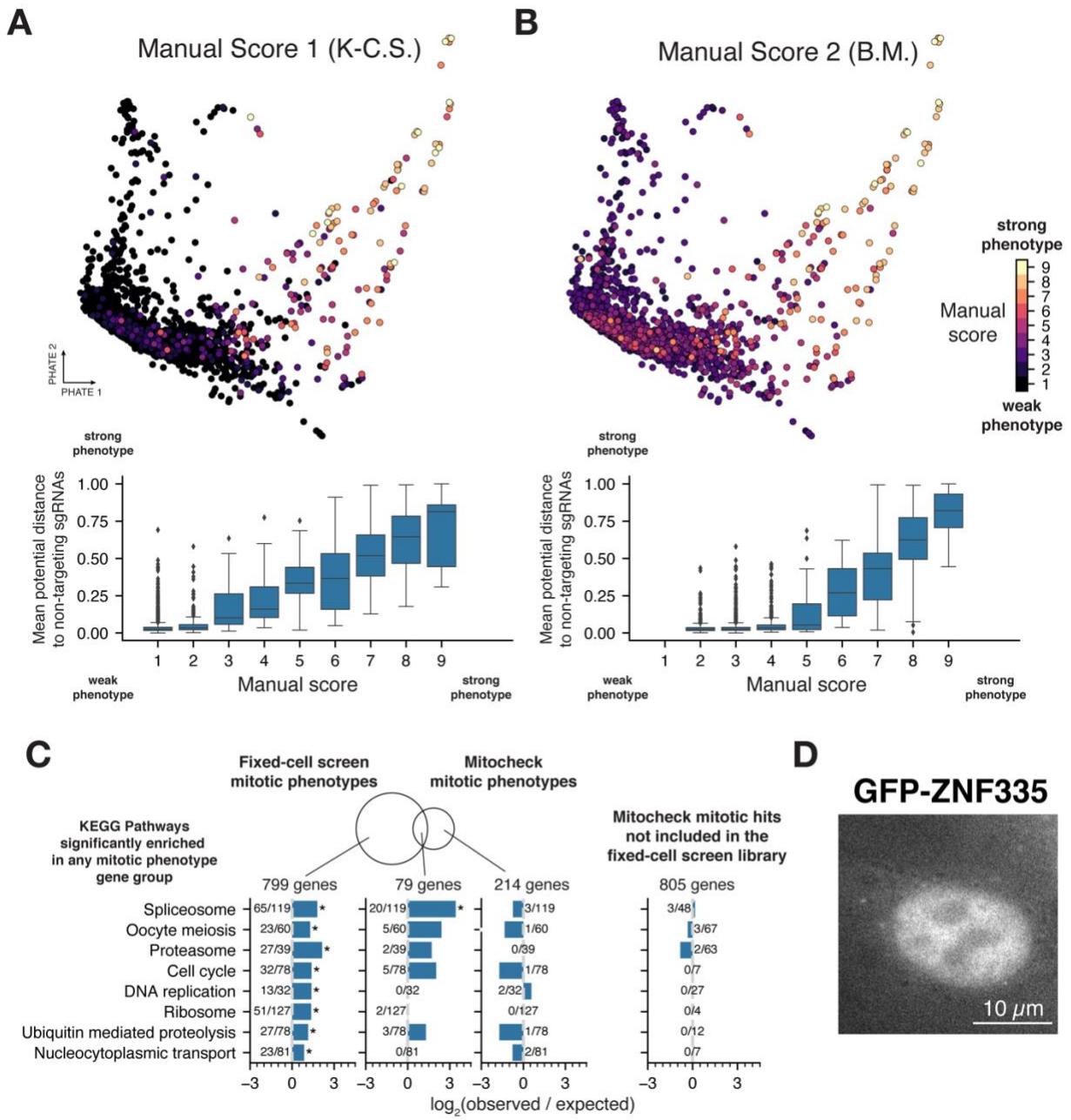

Figure 2.S10. Fixed-cell mitotic phenotype analysis. (**A** and **B**, top) Two-dimensional PHATE representations of mitotic phenotypes, colored by manual phenotype severity scores independently assigned by two individuals using anonymized gene labels. (A and B, bottom) Box plots demonstrating strong agreement between manual phenotype scores and computational phenotype strength (mean potential distance to non-targeting sgRNAs from mitotic PHATE analysis, normalized between 0 and 1). (**C**) Comparison of mitotic phenotype genes identified in our fixed-cell screen to mitotic phenotypes found by MitoCheck (*16*). Mitotic phenotypes in the fixed-cell screen were defined as overall phenotype strength greater than the 95th percentile of

non-targeting sgRNAs (as in Fig. 2.S6E) or significantly altered mitotic index (sgRNA permutation test, P<0.05). Bottom, bar graphs indicating KEGG pathway enrichment for all pathways enriched in any of the indicated gene sets. For the gene sets corresponding to the Venn diagram, the background for enrichment analysis was all genes present in both screens. For the set of mitotic hit genes identified in MitoCheck that were not present in our screen (far right), the background for enrichment was all genes screened in MitoCheck not present in our screen. *FDR<0.05. (**D**) Example image indicating nuclear localization of GFP-tagged ZNF335. Scale bar, 10 μm.

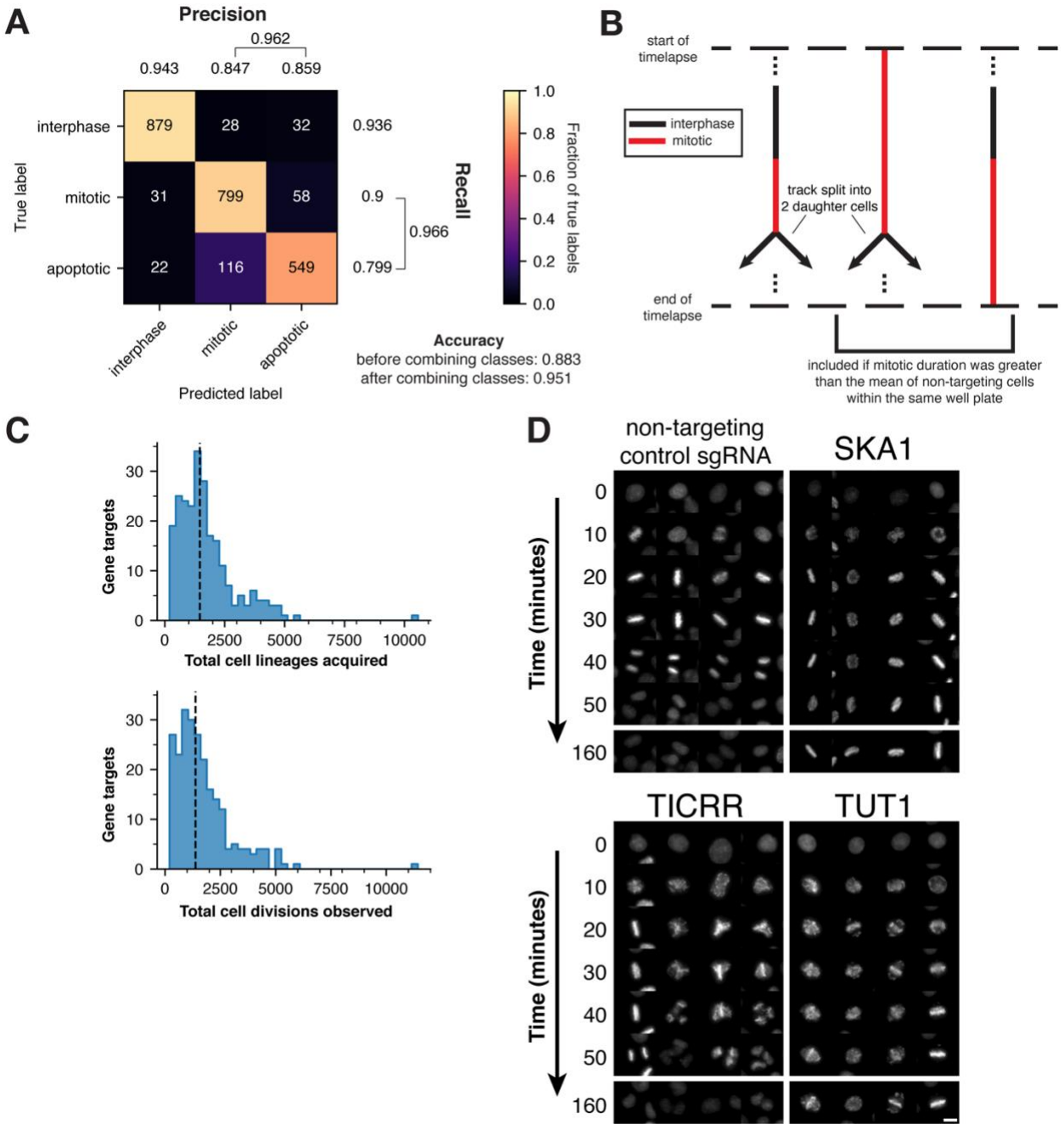

Figure 2.S11. Optical pooled screening for live-cell mitotic phenotypes. (**A**) Confusion matrix demonstrating performance of the support vector classifier in distinguishing interphase, mitotic, and apoptotic cells from the live-cell screen. 5-fold cross-validation with N=2,514 manually annotated cell images. Due to the relative difficulty of differentiating mitotic and apoptotic cells from H2B-mCherry fluorescence alone, the mitotic and apoptotic classes were combined after inference (cross-validation precision and recall after combining classes indicated by brackets). (**B**) Criteria for identifying a cell division event in the live-cell screen analysis. Cell lineages that

were not tracked across the entire time course were excluded. (**C**) Histograms of the total cell lineages acquired (top) and cell divisions observed (bottom) across both day 3 and day 4 time courses for each gene target. (**D**) Example images of H2B-mCherry fluorescence from the live-cell screen at the indicated time points after mitotic entry for selected gene targets. Each displayed knockout demonstrates increased mitotic duration and chromosome alignment defects relative to the non-targeting control sgRNA. Control images are reproduced from Fig. 2.5B. Scale bar, 10 μm.

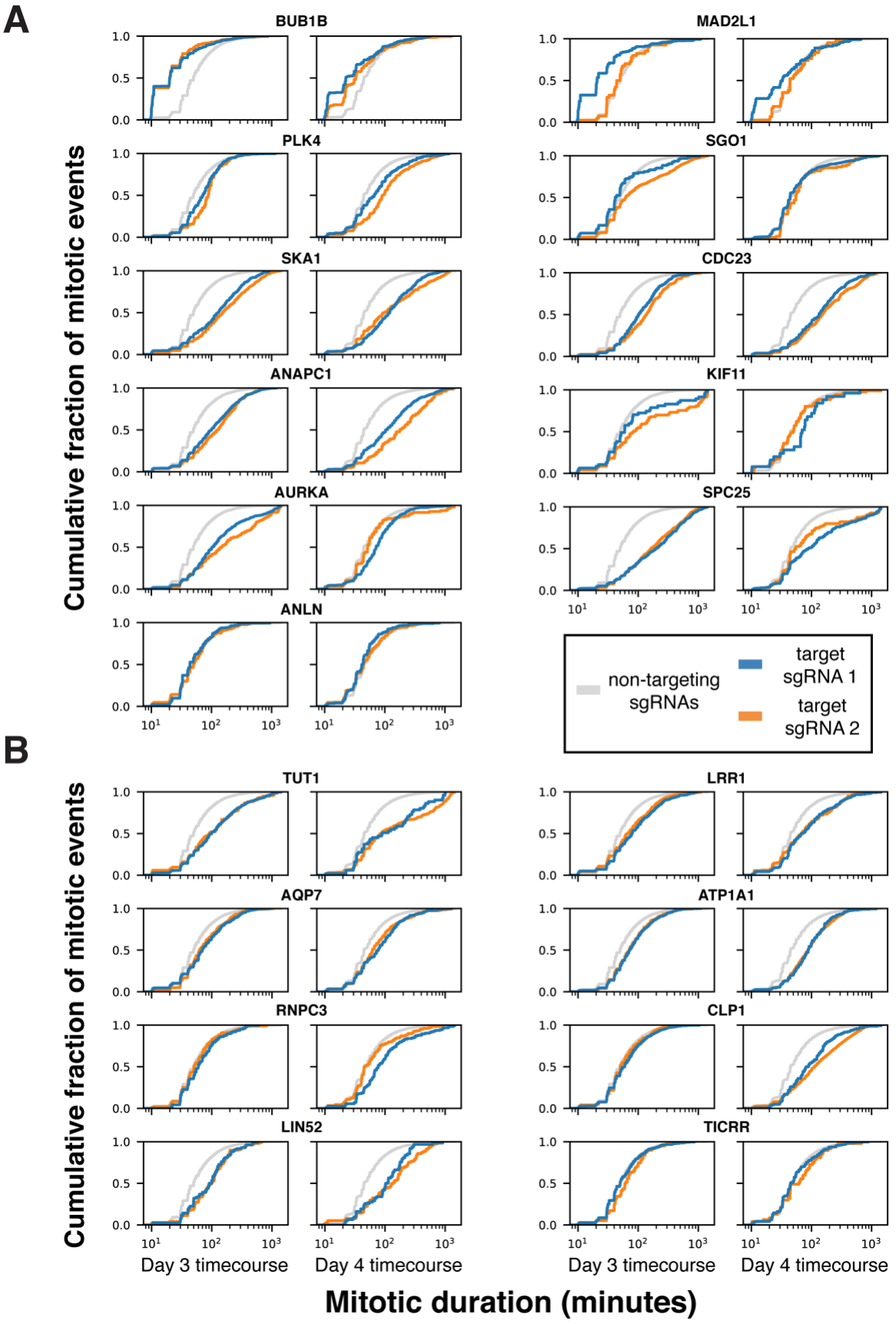
A
B
BUB1B
MAD2L1
PLK4
SGO1
SKA1
CDC23
ANAPC1
KIF11
AURKA
SPC25
ANLN
TUT1
LRR1
AQP7
ATP1A1
RNPC3
CLP1
LIN52
TICRR
Cumulative fraction of mitotic events
non-targeting sgRNAs
target sgRNA 1
target sgRNA 2
Day 3 timecourse
Day 4 timecourse
Day 3 timecourse
Day 4 timecourse
Mitotic duration (minutes)

Figure 2.S12. Mitotic duration distributions from the live-cell screen. Cumulative distributions of mitotic duration for individual sgRNAs (blue, orange) compared to all mitotic events of non-targeting cells (gray) across both day 3 (left) and day 4 (right) post-Cas9 induction time course experiments. (**A**) Positive control genes included in the screen and (**B**) selected genes identified from analyzing the screen data.

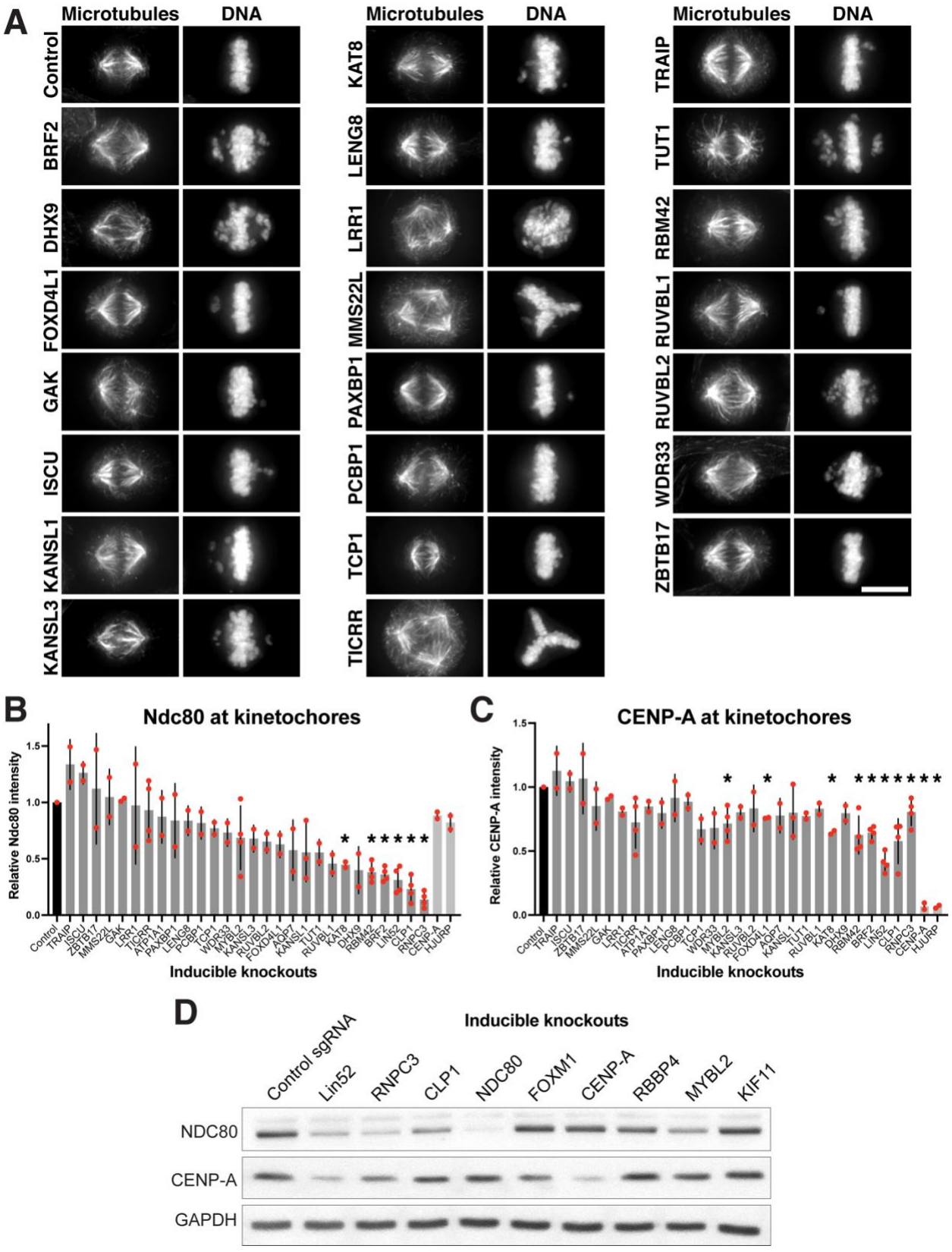
A
Microtubules
DNA
Control
BRF2
DHX9
FOXD4L1
GAK
ISCU
KANSL1
KANSL3
KAT8
LENG8
LRR1
MMS22L
PAXBP1
PCBP1
TCP1
TICRR
TRAIP
TUT1
RBM42
RUVBL1
RUVBL2
WDR33
ZBTB17
B
Ndc80 at kinetochores
Relative Ndc80 intensity
Inducible knockouts
C
CENP-A at kinetochores
Relative CENP-A intensity
Inducible knockouts
D
Inducible knockouts
Control sgRNA
Lin52
RNPC3
CLP1
NDC80
FOXM1
CENP-A
RBBP4
MYBL2
KIF11
NDC80
CENP-A
GAPDH

Figure 2.S13. Targeted analysis of mitotic phenotypes. (**A**) Immunofluorescence images of individual cell lines stably expressing a single sgRNA targeting each gene of interest to confirm live-cell pooled screen phenotypes and enable visualization at higher resolution across a single population. Images are deconvolved maximum intensity projections of fixed cells stained for microtubules (anti-alpha-tubulin) and DNA (Hoechst). Scale bar, 10 μm. (**B**) Bar plot showing kinetochore-localized intensity of the outer kinetochore microtubule-binding protein NDC80 in inducible knockout cell lines for all 29 genes pursued from the live-cell screen, along with CENP-A and HJURP controls. Each data point represents the median kinetochore signal of one experiment for >10 cells per gene target. Values are normalized relative to negative control cells from the same experiment. *P<0.01 by two-tailed independent T-test relative to negative control cells. (**C**) Bar plot of kinetochore-localized intensity for the inner kinetochore centromere-specific histone CENP-A in inducible knockout cell lines; experiment design as in (B). *P<0.01 by two-tailed independent T-test relative to control cells. (**D**) Western blot of CENP-A and NDC80 total protein levels for a subset of inducible gene knockouts.

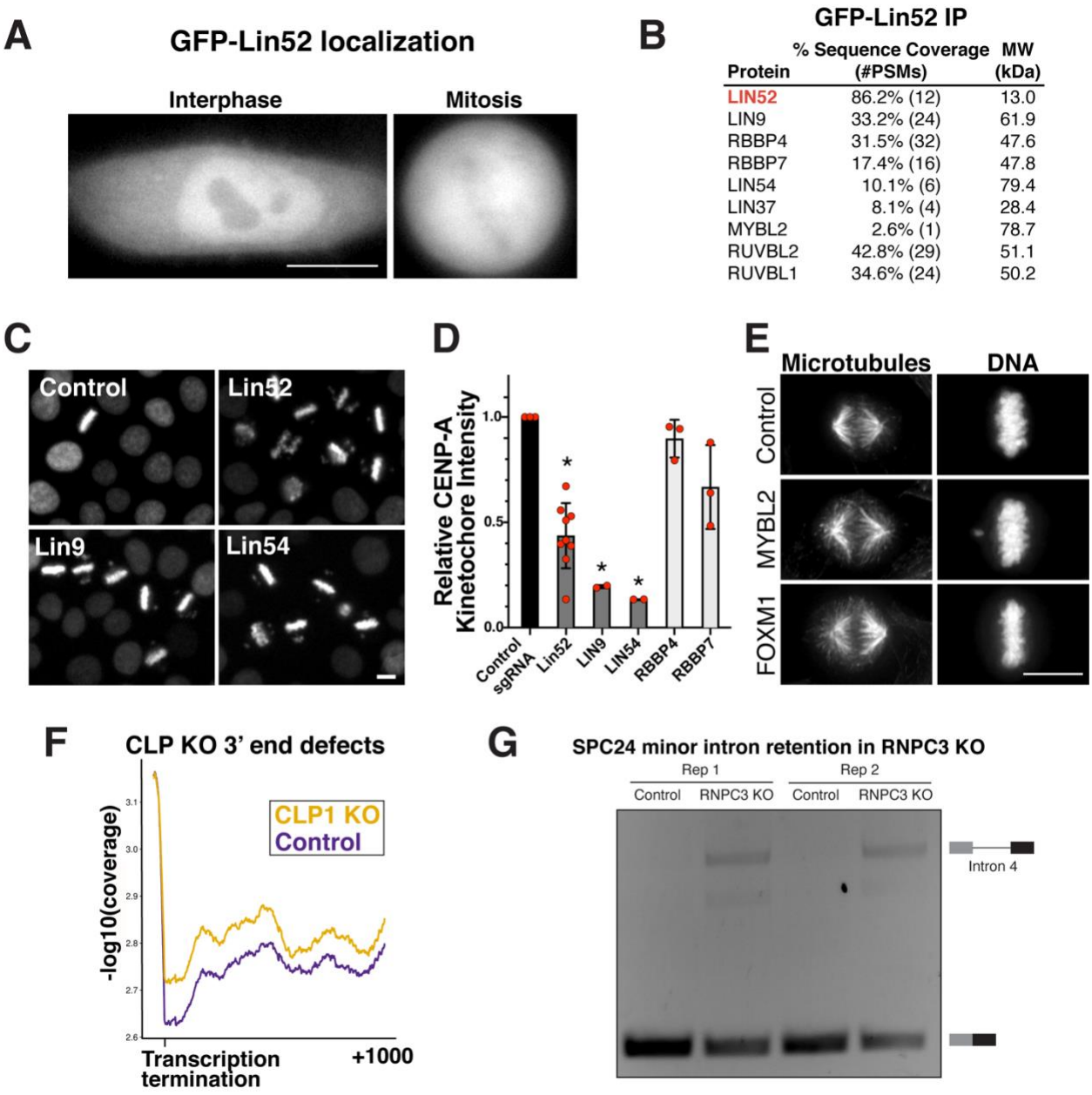

Figure 2.S14. Additional evidence for the roles of LIN52 and RNPC3 in kinetochore assembly. (**A**) Fluorescence images of human cells expressing GFP-LIN52, indicating LIN52 nuclear localization in interphase cells and non-specific localization in mitotic cells. (**B**) Mass spectrometry analysis of an GFP-LIN52 immunoprecipitation from mitotically-enriched cells relative to controls, indicating that LIN52 associates with a subset of expected factors, but not the entire DREAM complex. (**C**) Images from time-lapse fluorescence imaging of individual knockout cell lines expressing H2B-mCherry, demonstrating similar mitotic phenotypes for LIN52, LIN9, and LIN54 knockouts. (**D**) Bar plot showing kinetochore-localized intensity for the inner

kinetochore centromere-specific histone CENP-A in inducible knockout cell lines of LIN52-associated genes. LIN52, LIN9, and LIN54 each demonstrate a significant decrease in CENP-A kinetochore localization. *P<0.01 by two-tailed independent T-test relative to control sgRNA. (**E**) Immunofluorescence images of microtubules (anti-alpha-tubulin) and DNA (Hoechst) in inducible knockout cell lines identify chromosome alignment defects for the DREAM complex component MYBL2, but not FOXM1. (**F**) Metagene analysis of transcription termination in CLP1 knockouts. The increased density of reads downstream of annotated transcription termination sites suggests a defect in 3' end mRNA processing. (**G**) RT-PCR validation of SPC24 minor intron retention, showing an intron 4-containing SPC24 amplicon after RNPC3 KO that is not present in control samples. Scale bars, 10 µm.

Chapter 3. Future directions for image-based pooled screens

While the work presented in this thesis demonstrates the power of the current approach to image-based pooled screens, there are many opportunities for further technical and conceptual advancements. Here, we discuss current throughput limitations, potential to expand applications to diverse biological models and screening modalities, and challenges and opportunities associated with the resulting datasets.

3.1 Addressing current limitations of optical pooled screening

The primary factor limiting the scale of an optical pooled screen is the time and labor required to process and image samples during the iterative cycles of *in situ* SBS. Although the protocol is relatively straightforward, it is repetitive and laborious. Additionally, at each cycle of sequencing the imaging plate must be carefully aligned to the same microscope stage position as previous cycles. To reduce sample handling time and limit the risks of microscope misalignment, automation of the sequencing steps would be highly advantageous for improving the throughput of optical pooled screens. This may be achieved by integrating automated fluidics control into a standard well plate configuration, or developing a purpose-built flow cell system. Alternatively, pre-existing integrated sequencing systems could be adapted for the purpose of *in situ* SBS. The *in situ* SBS process used in optical pooled screens is identical to the SBS process performed within standard NGS instruments and use much of the same hardware and reagents, thus

presenting an opportunity to repurpose retired NGS instruments. In particular, others have developed well-documented open-source software for controlling each of the components within Illumina HiSeq 2500 instruments (*169*), which we have had preliminary success implementing on a retired instrument. Beyond automating the *in situ* SBS protocol, further throughput improvements may be achieved by increasing the speed of imaging iterative sequencing cycles. Simple optimizations such as implementing fast external filter wheels and externally-triggered hardware synchronization could likely improve imaging speed over our current systems by an order of magnitude. An alternative approach could involve optimizing two-color SBS chemistries for *in situ* sequencing in combination with simultaneous acquisition using a dual camera microscope configuration, although accurate base calling of *in situ* sequencing reads will likely prove more difficult with two-color sequencing schemes.

In addition to sequencing throughput limitations, barcode mRNA detection sensitivity and background staining of sequencing dyes are important challenges for optical pooled screening that need to be addressed. Although prior work estimates that only 5% of expressed barcode mRNA are captured by the presented protocol in HeLa cells (*170*), this has not been quantitatively measured and likely will vary significantly across cell models. Further optimizations to the *in situ* sequencing protocol should involve direct measurements of the fraction of expressed target mRNAs that are successfully captured and amplified with the padlock probes, likely enabled by FISH-based measurements. The detection sensitivity of *in situ* sequencing is likely limited by the reverse transcription and padlock probe gap-filling steps, which may be improved by optimizing each reaction and the fixation conditions before and after reverse transcription. These enzymatic reaction bottlenecks could potentially be eliminated from the protocol by implementing direct (no RT) barcode mRNA detection and/or gapless padlock probe libraries to match a given perturbation library (*73*, *171*, *172*). Optimizing *in situ* barcode detection efficiency is especially important as the protocol is applied to new biological models, which will likely exhibit variable barcode expression levels.

The reagents used for *in situ* SBS in optical pooled screens are derived from those used for sequencing in standard NGS instruments. While these reagents are undoubtedly highly optimized for their intended use of sequencing purified DNA libraries, the components may not be optimal for targeted SBS in the complex environment of fixed cells. The sequencing dyes currently used for optical pooled screens exhibit strong non-specific affinity for the fixed cell matrix, and accumulation of this binding over many cycles of SBS can result in decreased *in situ* read calling

accuracy, low screening efficiency, and unusable data. Thus, as new commercial solutions for NGS and spatial genomics come onto the market, it would be advantageous to carefully test each set of sequencing reagents in the context of *in situ* sequencing. Additionally, the buffers used for washing steps between cycles of sequencing dye incorporation are optimized for removing non-specifically bound dyes from a relatively simple flow cell surface, which is much different than the intracellular environment present in our application. Wash buffers or pre-incorporation blocking steps could be further developed for the specific purpose of reducing non-specific staining in fixed cell cultures.

In summary, multiple areas of throughput and efficiency improvements remain open for further development of optical pooled screening into a robust and scalable platform. This will become increasingly important as the sizes of perturbation libraries grow, and will also enable repeated library screening under multiple experimental conditions. Increasing the robustness of the protocol will also further enable expanding applications into a wide variety of new biological models and screening modalities as discussed in the following sections.

3.2 Expanding biological models

3.2.1 *In vitro* models

Although all of the published studies using optical pooled screening have thus far been performed in cancer cell lines, there is immense potential for expanding image-based pooled screening applications in more diverse biological models. Additional *in vitro* cell models that are likely candidates for image-based screening are hTERT-immortalized cell lines, specialized cell types derived from differentiating induced pluripotent stem cells (iPSCs) or direct reprogramming, primary cells isolated from model organism or human donors, or co-cultures of multiple cell types. Image-based pooled screens in these model systems have the potential to identify genes involved in cell type-specific functional phenotypes or disease processes. As many cellular pathologies can likely be directly measured using microscopy, there is a strong potential for discovering novel therapeutic targets with optical pooled screening. Although the advantages of using diverse cell models are clear, this will likely come with decreases in screening efficiency which may necessitate additional optimization. In particular, expression levels of barcode mRNA are likely to vary significantly across cell models. This may be alleviated by testing multiple promoters for barcode expression and optionally including other perturbation vector elements (e.g., to limit

methylation-induced gene silencing in models that require differentiation; see ref. *173*). Differentiation protocols and primary cell cultures may also result in debris within the imaging plates that could cause increases in non-specific background staining of sequencing reagents. Cell culture protocols should be carefully optimized, and modifications to the *in situ* SBS protocol can be considered as discussed in section 3.1. Complex cell morphologies, such as neurons and other polarized cell types, may cause difficulty in accurate cell segmentation, an important step for assigning *in situ* sequencing reads to individual cells. Thus, cell segmentation algorithms developed for specific cell model applications may need to be developed, or a generalist algorithm such as CellPose may be adequate (*161*).

For post-mitotic cells, antibiotic selection of transduced cells is inefficient and the transduced cells do not expand to increase library coverage. Screening with post-mitotic cells will consequently be inefficient and require larger transduction scales to achieve a given final cell library size. However, replacing antibiotic resistance markers in the perturbation delivery vector with a fluorescent protein may help identify transduced cells in the final population. For differentiation-based assays, it is likely advantageous to transduce and expand cells prior to differentiation. However, the scale of screens possible with *in vitro* differentiated cell types will depend on the differentiation efficiency. The most fruitful screening applications in iPSC-derived cell lines will likely involve differentiation via overexpression of lineage-committing transcription factors (*174–176*), where differentiation is exceptionally efficient and fast, which likely also limits issues of methylation-induced gene silencing. Despite the challenges associated with increasing complexity of *in vitro* cell model systems, we have already demonstrated the feasibility of optical pooled screens in embryonic hippocampal neurons derived from rat models, as discussed in **Appendix A**.

3.2.2 *In vivo* models

An inherent property of microscopy datasets is the acquisition of spatial information concerning intracellular components and the cells themselves. The spatial relationships between cells is of particular interest for potential applications of image-based pooled screening to *in vivo* biological models. While *in vivo* pooled CRISPR screens have been performed with enrichment and scRNA-seq based phenotype measurements (*177–180*), these approaches do not measure tissue organization or intercellular interactions, which likely contribute to most *in vivo* phenotype measurements. One recent study used protein-based barcodes to perform pooled CRISPR screens with spatially-resolved transcriptomic and histopathologic phenotype measurements (*181*), although restricted to the small scale of 35 gene targets. As *in situ* sequencing protocols

have been previously used in tissue samples for other purposes (*73*, *182*), there is potential for its use in *in vivo* pooled CRISPR screens. This will bring new technological challenges, including likely increased non-specific accumulation of sequencing dyes. The approaches suggested in section 3.1 may help to partially reduce this issue, and incorporating tissue clearing protocols may also provide improvements. Optimizations to the *in situ* amplification reactions may also be necessary, or non-optical identification of perturbation barcodes could be employed. Several sequencing-based approaches to measure spatial expression of mRNAs have been developed (*183*, *184*), which could be adapted for identification of perturbation barcode mRNA and matched with microscopy-based phenotype measurements from alternating tissue slices. Besides technological challenges, the scalability of *in vivo* pooled CRISPR screens is limited by the total number of cells that can be targeted *in vivo* by perturbation vectors or perturbed *ex vivo* and then delivered to a model organism. Thus, applications of *in vivo* pooled screens should be carefully considered and may be inherently limited to small library sizes.

3.3 Screening modalities

While CRISPR knock-outs are a simple and efficient way to introduce genetic perturbations, this approach may not be suitable for all biological models or questions. Optical pooled screening can be trivially extended to other Cas9-based genomic perturbations, including direct disruption of non-coding genetic elements and down- or up-regulation of gene expression using CRISPRi or CRISPRa. This is achieved by using a cell model expressing the corresponding enzyme and delivering a sgRNA library targeting the appropriate genomic loci, and may be particularly important in disease-relevant cell models with active p53 responses (*185–187*). Screens evaluating phenotypic effects of overexpressing open reading frames (ORFs) of wild-type or variant alleles are also possible by associating a unique barcode to each individual ORF (*188*). Optimization of differentiation protocols is a particular area of potential application for optical pooled screens, as overexpression of single or combined transcription factors could be tested in a pooled format, with the resulting cell states evaluated morphologically or with functional cell type assays (*189*). More complex perturbations that enable specific genomic alterations via base or PRIME editing (*190–192*) are also likely amenable for incorporation into image-based pooled screening workflows. When using either base and PRIME editing, the actual editing outcome of a given perturbation is highly important for its effect on the measured phenotype, as opposed to the reliance on random results of DNA damage repair processes with Cas9-based knockout screening. Directly measuring these editing outcomes in individual cells may be possible with *in*

situ sequencing, especially in pooled screens tiling perturbations across one or a few genomic loci, by including additional padlock probes to amplify the targeted sequence.

As image-based pooled screens identify perturbations within each individual cell, measuring effects of combinatorial perturbations introduces almost no additional protocol complexity. Combinatorial perturbation screening is attractive for identifying interacting or redundant factors controlling a given phenotype. This is particularly important in disease-associated processes that likely act through the function of multiple genes, in contrast to the primary focus of current target-finding functional genomics on single perturbations. This may be achieved by sequential infections of perturbation libraries to be screened against each other (*75*) or defined pairs of sgRNAs synthesized in a single construct for use with Cas9 or other CRISPR enzymes such as Cas12a (*193–197*). Alternatively, high MOI infections can be employed to carry out "compressed" interaction screens, where the relative infrequency of genetic interactions is leveraged to identify consistent effects of combined perturbations across random independent collisions (*198*). This could reduce the total number of cells needed for screening a large number of gene pairs, although there are corresponding constraints on the sparsity and order of functional gene interactions that may not be met with the focused libraries that are often required with combinatorial screens. For combinatorial perturbations in optical pooled screens, the limited sensitivity of perturbation identification mentioned in section 3.1 becomes especially important when correct perturbation assignment to each cell depends on successful identification of multiple barcodes.

Beyond screening perturbations of genetic elements, the technology developed for optical pooled screening has potential for enabling high throughput image-based experiments for other applications. For example, several CRISPR technologies have been developed for efficient endogenous tagging of proteins with fluorescent proteins or other fusions (*199–201*). Already, others have combined this approach with *in situ* sequencing to visualize changes in localization for a pool of hundreds of endogenously tagged proteins in response to drug treatments (*202*). Such experiments have the potential to efficiently screen therapeutic drug candidates and identify mechanisms of action. Image-based pooled screens could also be employed for functional screening of designed protein structures. This could be useful for high-throughput optimization of fluorescent reporter systems like kinase activity reporters (*37*), or for developing *de novo* protein functions for therapeutic or other purposes (*203*, *204*).

3.4 Leveraging complex phenotype measurements

Acquiring an image-based pooled screen dataset with even a few phenotype stains can yield hundreds of measurements for each of the individual cells, as the localization and intensity of fluorophores can be numerically summarized in many non-redundant image features. There are many open questions in the field of how to handle such large datasets and extract meaningful insights, and further expanding phenotype imaging channels with the approaches mentioned in section 1.3 will exacerbate these issues. In Chapter 2, we demonstrated that a vast amount of meaningful information is accessible with a relatively simple approach using pre-defined phenotypic features and aggregating information to the level of gene targets. However, this is not necessarily the optimum approach to understanding the full information available in the acquired complex phenotypes.

3.4.1 Using single-cell level information

A primary question that must be addressed when first approaching an image-based screening dataset is the value of the single-cell level information. While the importance of single-cell information has been much-discussed for genomics measurements of the complex cell communities found *in vivo*, this information has been arguably underutilized in the context of pooled perturbation screens with single-cell RNA-sequencing phenotypes (*25*, *72*, *124*). Part of this likely is due to the observed cell-level variation arising from multiple sources that are difficult to distinguish, in particular perturbation efficiency in addition to meaningful biological heterogeneity. Thus, aggregation of single-cell phenotypes, especially using statistical approaches robust to outliers, may be necessary to reduce sources of noise inherent to single-cell measurements. For image-based pooled screens, the large scale of the acquired phenotype measurement matrices (e.g., ~10^3 phenotype features by $3x10^7$ cells in Chapter 2) also limits the feasibility of analyzing the full datasets at the single-cell level.

However, there are several intermediate approaches to analyzing single-cell resolved datasets that may balance the corresponding advantages and challenges. First, individual cells can be classified into phenotypic categories, and perturbations compared by the fraction of cells present in each phenotype class. This has the advantage of directly identifying changes in cell state occupancy, although information will be lost by reducing continuous phenotypes into discrete bins. Practically, this could be achieved by training a supervised machine learning model to place cells into pre-defined phenotypic categories, or by using an unsupervised clustering approach to

automatically identify groups of phenotypically similar cells directly from the acquired data. A supervised classifier has the particular advantage of enabling model training on a smaller subset of cell measurements at reduced computational cost, with the resulting algorithm then applied to the full dataset in parallelized chunks. Meanwhile, most clustering algorithms are not implemented for distributed computation and require the full dataset to be held in memory simultaneously.

Alternatively, aggregation of single-cell measurements could be performed in a way that accounts for distributional changes rather than alterations in simple summary statistics of central tendencies. For example, in the original optical pooled screen of a transcription factor translocation phenotype, one-dimensional translocation distributions for each gene target were compared to the distribution of negative control cells by integrating the difference in the cumulative distributions of cells (*75*). This is known as the Wasserstein or Earth Mover's distance, which is well-developed for low-dimensional distributions in the field of optimal transport theory. However, applying this concept for high-dimensional phenotypes at-scale is more difficult, and could require computationally expensive distance measurements between all possible pairs of perturbations to identify similarities or differences between gene targets. Graph-based methods for applying optimal transport concepts to high-dimensional data are currently under development (*205*, *206*), but applying them to the dataset in Chapter 2 has proven infeasible thus far due to the computational cost at the given scale of data.

Finally, the single-cell level data may be used to identify and correct for confounding variables prior to aggregation of perturbation-level phenotype profiles for downstream analysis and interpretation. This approach was well demonstrated by the MIMOSCA approach developed for scRNA-sequencing based perturbation screens, where a linear model of perturbation effects on gene expression phenotypes was trained while simultaneously accounting for several known technical and biological covariates (*25*). Removing such confounding variables helps isolate phenotypic changes unique to the perturbation, but must be applied carefully to maintain identification of perturbation effects that result purely in cell state proportion changes.

3.4.2 Learning phenotype representations directly from images

While profiling of predefined image features results in easily interpretable measurements and has proven to be powerful for identifying biological phenotypes in many cases (e.g., as presented in Chapter 2), this approach is limited by human expectations of phenotype effects and provides no guarantee of optimality. Meanwhile, advances in machine learning theory along with the

associated computational hardware provide new opportunities for how to understand large datasets. In particular, models based on convolutional neural networks (CNNs) have emerged as advantageous for image classification and representation learning, as CNNs explicitly use the spatial structure of images and in many ways operate analogously to real-world imaging systems. Furthermore, machine learning models, such as those constructed with CNNs, perform best when they are trained on large and diverse datasets such as those acquired using image-based pooled screening, leading to a clear opportunity for successful application. Specifically, CNN-based models could learn phenotype representations directly from the acquired images combined with knowledge of the perturbation identities, instead of the current approach of hand-designed phenotype image features defined *a priori*.

When developing a machine learning model, the primary design decision is the choice of objective that the model parameters should be optimized for. In representation learning, the optimized objective can be separated from the actual features that are used for downstream analysis, such as an intermediate result from within the model. A common supervised learning approach for learning representations of images is to combine several convolutional layers operating on the raw data, followed by "fully-connected" layers that transform the convolutional, spatially-aware features into a vector of probabilities that the given sample belongs to each user-defined category. The model parameters are then optimized such that the model performs well at predicting the image label, which in theory also causes image properties relevant to separating the individual classes to be summarized into the intermediate convolutional features. These intermediate features can then be used for downstream tasks, including analysis to understand intraclass variation and relationships between the classes. This approach has been applied to understanding phenotypes in image-based screens (*52*), where the class label predicted for each cell during training is the chemical or genetic perturbation identity. However, this is prone to model overfitting, as it assumes all perturbation phenotypes are separable from each other, a condition which often does not hold.

An important alternative method for direct feature learning from images are models that employ autoencoders. At their base level, autoencoders optimize the ability to encode an image into a low-dimensional "latent" representation, and then decode this representation back into the original image. The model parameters are thus trained to minimize the difference between the original image and the image decoded from the latent representation. The latent representations can then be used for further analysis, which are more faithful representations of the images themselves

and less prone to overfitting than those learned using simple perturbation labels in supervised learning contexts. However, autoencoder-derived features will contain much information that is irrelevant to the differences between perturbation phenotypes, including baseline cell structure and cell orientation in the image. In a previous application where representations of subcellular protein localizations were learned from microscopy images, the condition-specific information within such representations was accentuated by including an additional task in the model training regimen that tries to predict the cell's condition (in this case, which protein is visualized) from the latent features (*53*, *199*). Alternatively, others have demonstrated with natural images the advantages of learning relevant feature embeddings using multiple versions of the same image, each independently subjected to condition-irrelevant image transformations or distortions prior to model input. The parameters are then optimized such that features are consistent when extracted from the paired examples (*207*).

A related CNN-based approach we have pursued is to explicitly disentangle the perturbation-irrelevant information in images from the image features that are unique to a given gene target. This in theory can be achieved using two parallel encoder paths, one of which is trained adversarially such that the encoded features explicitly cannot be used to predict the perturbation identity and thus mostly contain information unrelated to the specific perturbation. The latent features from the parallel paths are then combined through a common decoder network that is optimized to reproduce the original image. In theory, the latent representation produced by the second encoder network only contains information relevant to a given perturbation and would provide an information-rich input for downstream analysis. Overall, this approach could overcome the challenges of optimizing a model to predict perturbation identity by instead doing the inverse: removing from the learned representations information that cannot be used to predict the perturbation identity. However, the practical implementation of this approach is still under development.

As outlined, there are many opportunities to use CNNs or other machine learning approaches to learn phenotype representations directly from raw images, used in place of or in combination to human-designed phenotype measurements. While the extracted features from CNNs will likely be processed downstream identically to predefined features, the hope is that they will exhibit higher information density and be able to capture more complex image features, as they are optimized for specific, although often abstract, objectives without the limitations of human intuition. These approaches will be especially useful for screens collecting phenotypes across many

imaging channels, where there are often only weak biological priors of how to measure and compare the phenotypes. In screen designs where a specific biological phenotype is known, phenotype extraction from images should focus on measuring the desired phenomena, which may or may not require machine learning or CNN optimization. Furthermore, computational costs for training and applying such models may be prohibitive. This can be partially ameliorated by starting with a network pre-trained on other image datasets (e.g., natural images from ImageNet) and then training a shallow network on top of these features to adapt them to a specific application. Additionally, the conventional approach is to train such models on the majority of a dataset (70-80%), monitor model optimization using a validation subset (10-20%), and reserve for downstream processing only the remaining samples (often 10%). This may be problematic when screening throughput is limited, or when other factors (e.g., perturbation efficiency) result in only a small fraction of cells exhibiting the phenotype of interest. Finally, it is yet to be seen how transferable model parameters and designs are between different image-based screening applications, potentially resulting in time consuming end-to-end model design and optimization for each new application. Despite these caveats, image-based pooled screens are ideal datasets for representation learning approaches due to their scale and relative consistency across perturbation conditions.

3.4.3 Understanding biological functions from perturbation phenotypes

Genetic screening in model systems has long been an essential method for understanding biological functions (*4*–*6*, *11*). However, the ability to identify and characterize the functions of genes from perturbation phenotypes is complicated by several organizing principles in biology, particularly hierarchy, pleiotropy, and redundancy. The hierarchical nature of biology is present even within individual cells, with proteins and other molecules associating into complexes, complexes and macromolecules organizing into organelles, and biological pathways functioning across multiple scales. In a genetic perturbation experiment, a measured phenotype may arise from functional disruption at any point in a pathway. It is often challenging or impossible to identify the exact level at which a gene is acting directly from the phenotypic data collected in a screen, particularly for one-dimensional measurements such as overall cell fitness. However, with complex phenotypic measurements across multiple dimensions, such as enabled with optical pooled screening, subtle differences between similar phenotypes can be identified that may provide more specific functional information. This view influences how the collected data should be processed, as discrete grouping of phenotypes without consideration of their relationships will obscure this information. We demonstrate an initial such approach in Chapter 2, employing an

algorithm for low-dimensional visualization that maintains global relationships, and clustering at different scales to understand the hierarchy of gene functions. This could be further developed in the future by implementing algorithms that explicitly learn hierarchical relationships from perturbation phenotype measurements. Additionally, the flexibility of optical pooled screens could be leveraged to directly measure related phenotypes at multiple scales, such as expression levels and spatial organization of both transcripts and proteins, combined with higher-level cellular phenotypes of interest. For alternative high-dimensional pooled screens, such as those using single-cell RNA-sequencing, phenotypes are necessarily measured at the molecular scale of transcript abundance, without the intrinsic ability to connect these measurements to other scales such as high-level cellular behaviors.

Pleiotropy, where a single gene is involved in multiple distinct biological functions, introduces similar complications to the simple picture of genotype-phenotype associations. In a genetic screen, this manifests as an observed phenotype measurement from a single perturbation actually resulting from the additive effects of multiple disrupted functions. Similar to the case of hierarchy, these effects typically cannot be de-convoluted with one-dimensional measurements, but may be possible with more complex phenotypes where a perturbation may exhibit a phenotype that is quantitatively represented by a combination of others. A recent study inferring gene pleiotropy from many independent pooled fitness screens demonstrates this approach (*125*), which likely could be adapted to learn pleiotropic function within a single image-based pooled screen with high-dimensional phenotype measurements.

Finally, the pervasive redundancy of biological systems presents inherent limitations to genetic screens, as the disruption of a single gene may be buffered by other cellular mechanisms. Unlike the inference limitations associated with hierarchical and pleiotropic biological organization, redundancy cannot be overcome by merely acquiring and analyzing more data of the same type. Instead, identifying redundant gene functions in high throughput requires new experimental approaches. Performing identical screens across many different biological systems has shown some progress in identifying redundant elements (*112*), but is limited by existing biological variation and thus is likely unable to uncover redundant regulation of the most essential biological processes. A more direct approach to high-throughput screening for redundant factors is to perform combinatorial perturbation screens. As discussed in section 3.3, this is highly amenable to optical pooled screening, and should be implemented in future work.

Pooled perturbation screening has developed into a primary approach for discovering the functions of biological molecules, and the continuing development of experimental and computational frameworks for these studies points to their ongoing use well into the future. However, orthogonal methods, including large-scale protein-protein interaction and gene co-expression studies, will remain as effective approaches and will likely enable complementary discoveries of biological functions. Overall, this work demonstrates one perturbation screening approach, image-based pooled screening, as particularly advantageous for identifying the molecular basis for complex phenotypes and cellular functions at scale.

Appendix A. Optical pooled screening in primary neurons

This work was completed in collaboration with Marek Nagiec and Jeff Cottrell, formerly of the Stanley Center for Psychiatric Research at the Broad Institute.

A.1 Motivation

While human cancer cell lines provide convenient *in vitro* models for fundamental biological phenomena such as those explored in Chapter 2, they are limited in their ability to recapitulate cell type-specific phenotypes and disease processes. For example, screening for factors involved in specialized neurological functions requires culture models optimized for this specific purpose. One such model involves dissecting embryonic hippocampal neurons from rats, which can then be cultured *in vitro* for extended periods of time. This approach is advantageous as it uses neurons resulting from normal development pathways, rather than iPSC-derived models that are generated through artificial differentiation processes that may not fully reach biologically-relevant endpoint cell states. Thus, *in vitro* primary neuron models can be used to recapitulate basic neurological functions and disease states much closer to their native context. However, primary cell models present significant challenges for genetic perturbation screens. Primary neurons are post-mitotic, which presents technical challenges in acquiring a sufficient number of virally-transduced cells and also severely limits the ability to screen for enrichment-based phenotypes. While the image-based pooled screening method discussed throughout this thesis enables an effective phenotyping approach beyond mitosis-dependent enrichment, the challenges of screening in the more complex cell model and without antibiotic selection still remain. Here, we

perform technical optimizations and a small-scale trial screen to enable optical pooled screening in primary rat embryonic hippocampal neurons.

A.2 Technical optimizations

In order to make image-based pooled screening efficient in primary neurons, we optimized our approach to maximize the fraction of transduced cells with successful Cas9-mediated gene disruption and also the fraction of transduced cells with identifiable *in situ* sequencing spots. To optimize these two metrics, we tested several CROPseq-derived lentiviral sgRNA delivery vectors with alternative RNA polymerase II (Pol II) promoters and sgRNA scaffold sequences. Ideally, using a Pol II promoter with strong expression in a specific cell type will enable higher expression of barcode mRNA for efficient *in situ* SBS. We compared our standard EF1a CROPseq promoter to the human Synapsin promoter, which provides strong expression in many neuronal cell types (*208*). Meanwhile, an optimized sgRNA scaffold sequence developed previously (*92*), as also used in Chapter 2, will likely increase the fraction of transduced cells with successful knockouts. We also replaced the CROPseq antibiotic selection marker with an ORF encoding the mNeonGreen fluorescent protein to enable identification of transduced cells. The multiplicity of infection (MOI) of the lentivirus infection provides an additional parameter to adjust the total number of transduced cells. However, this comes with limitations as very high MOI will result in multiple sgRNA integrations within individual cells, which are potentially detrimental to identifying phenotypic effects from each individual gene perturbation.

We first evaluated knockout efficiency of the different CROPseq variants using 5 control sgRNAs targeting 3 genes in an arrayed infection experiment (Fig. A.1A). The fraction of cells with successful knockout was identified by immunofluorescence staining for the targeted protein with validated antibodies. Despite sgRNA and gene target level variability, we found that using the optimized sgRNA scaffold (CROPseq-v2) overall resulted in a larger fraction of cells with disrupted protein expression. The highest overall knockout efficiency was observed when the optimized sgRNA scaffold was combined with the EF1a promoter. To evaluate *in situ* sequencing efficiency, we performed a small-scale pooled lentiviral infection with 38 sgRNAs in each of the three tested CROPseq vectors (Fig. A.1B). Here we find that the optimized scaffold combined with either promoter resulted in a lower fraction of cells with at least 1 identified *in situ* sequencing spot, especially when combined with the Synapsin promoter. Overall, the results presented here suggest a trade-off between gene perturbation efficiency and barcode mRNA expression when

using the CROPseq vector design. Furthermore, the optimized sgRNA scaffold used in CROPseq-v2 was partially designed to increase RNA polymerase III (Pol III) expression of the sgRNA, while the Synapsin promoter likely increases the Pol II expression of the barcode mRNA in this cell type. Together these effects may result in interference between the two RNA polymerases due to the proximity of the promoters and overlap between the barcode mRNA transcript and one copy of the integrated sgRNA expression cassettes. Although the decrease in *in situ* sequencing efficiency will be problematic for large-scale screens, we chose to proceed with the CROPseq-v2-EF1a-mNeon vector as it maintains high knockout efficiency, which is absolutely required for measuring phenotypic effects of gene perturbations.

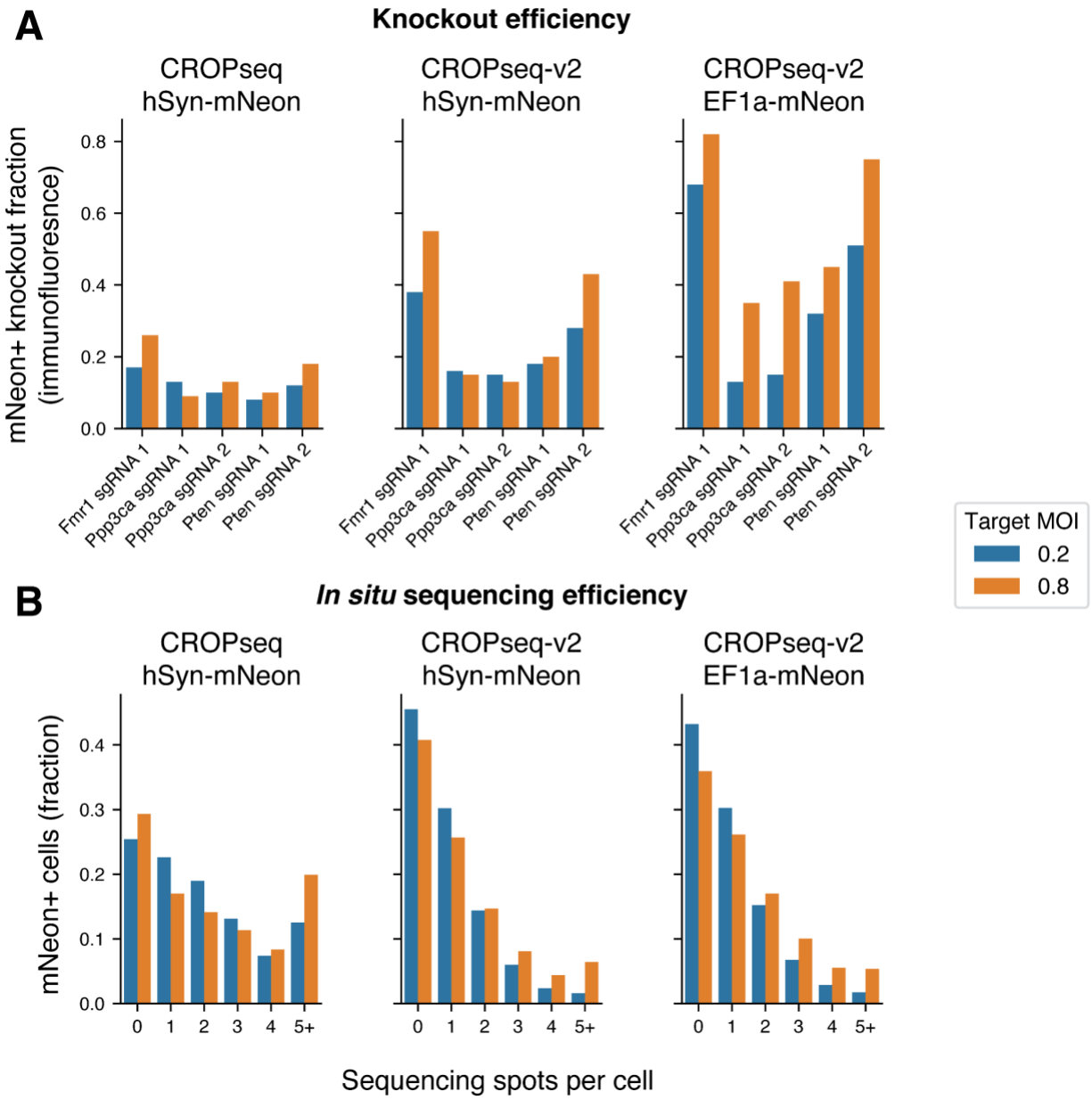

Figure A.1. Optimizing optical pooled screens in primary neurons. (**A**) Cas9-mediated knockout efficiency with three CROPseq-derived vectors, tested across 5 sgRNAs in arrayed infection experiments. Knockout efficiency was measured in transduced (mNeon positive) cells by immunofluorescence with antibodies staining for the corresponding proteins. (**B**) Detection efficiency of *in situ* sequencing spots with the same vectors as in (A). Sequencing spots within mNeon positive cells were identified from one cycle of *in situ* sequencing.

A.3 Trial screen

Following the optimization of our sgRNA delivery vector in section A.2, we performed a small-scale image-based pooled CRISPR screen using CROPseq-v2-EF1a-mNeon to validate the full screening approach in primary neurons. For the phenotype measurements, we chose to use a straightforward immunofluorescence antibody stain for a Calcineurin phosphorylation event (Ppp3ca phospho-S469), which is being developed by our collaborators as a general marker of neuronal activity associated with synaptic plasticity. This phosphorylation event can be conveniently induced *in vitro* by stimulating N-methyl-d-aspartate (NMDA) receptors. We chose a set of 7 genes which are expected to be involved in regulating this signaling event (Ppp3ca, Grin1, Fmr1, Kcnq2, Nrgn, Pten, and Scn2a), and designed 4 sgRNAs targeting each of these genes in addition to 10 non-targeting negative control sgRNAs for a total of 38 sgRNAs. After lentiviral infection and 14 days *in vitro* (DIV), separate wells were treated with tetrodotoxin (TTX; suppresses activity), TTX followed by NMDA (activating), or Bicuculline (activating). The cultures were then fixed and processed for *in situ* sequencing as in Chapter 2, and 10 cycles of *in situ* SBS were acquired (Fig. A.2A). Transduced cells were identified using the mNeon marker, with high-expressing mNeon cells demonstrating increasing *in situ* sequencing read counts (Fig. A.2B). Overall, more than 80% of transduced cells had an *in situ* sequencing read successfully mapped to the designed library (Fig. A.2C). However, the MOI of the infection appeared to be higher than anticipated, and only 50-60% of mNeon positive cells had *in situ* sequencing reads uniquely mapping to a single sgRNA sequence (Fig. A.2D-F).

Following *in situ* identification of the sgRNAs expressed in each transduced cell, this information was combined with phenotype measurements of neuronal activation via Calcineurin pS469 staining. Overall, phosphorylation increased strongly between unstimulated or TTX-suppressed conditions and Bicuculline or NMDA activating stimulations within cells expressing non-targeting sgRNAs (Fig. A.3A). However, for the cells with targeting sgRNAs, only very minor differences in phosphorylation levels were observed, and in only two genes (Ppp3ca, Grin1), when comparing to the non-targeting sgRNAs (Fig. A.3B). These two genes were expected to have

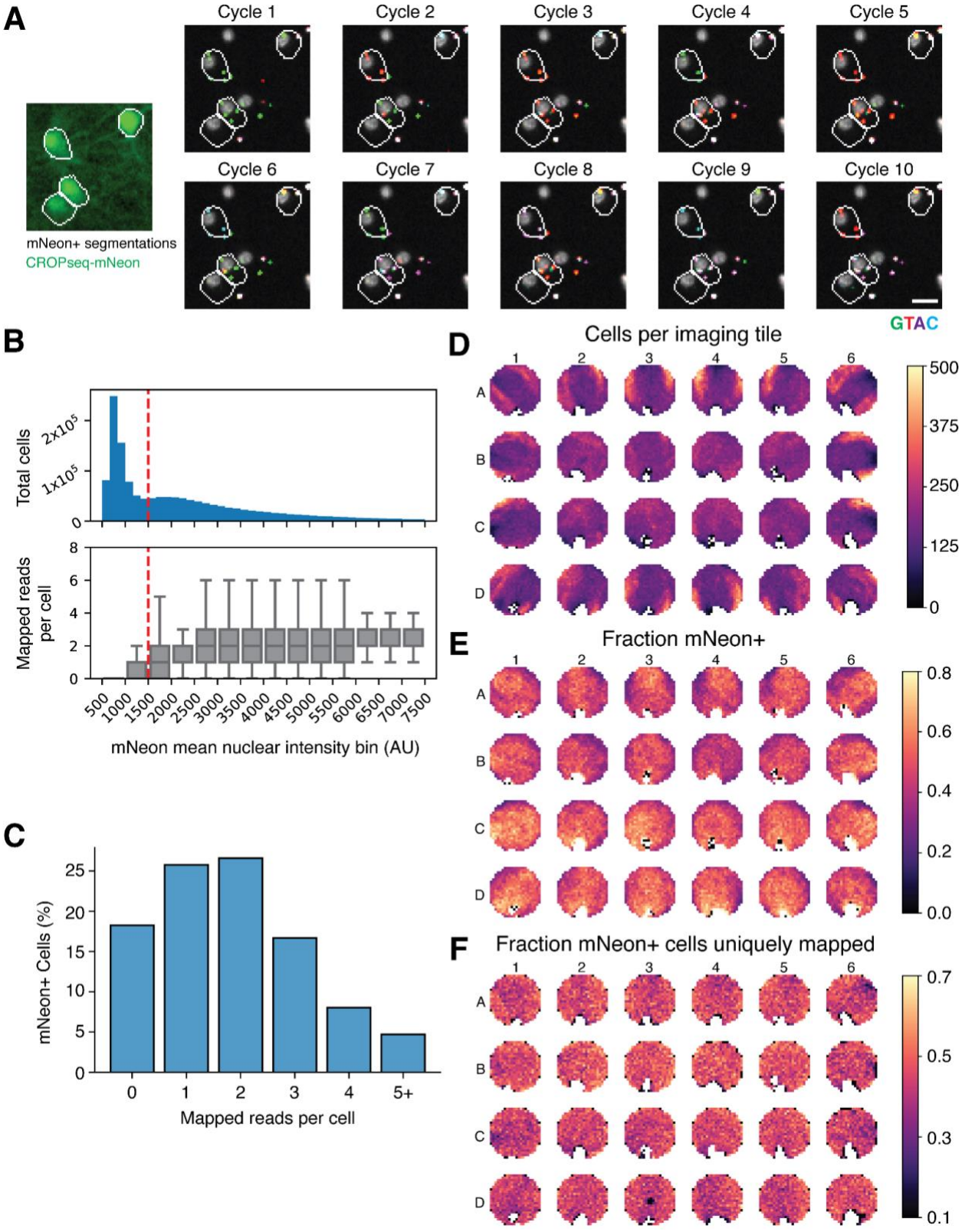

Fig. A.2. Quality control of a trial screen in primary neurons. (**A**) Example images from 10 cycles of *in situ* SBS in primary neuron cultures (sequencing channels are Laplacian-of-Gaussian

filtered). Scale bar, 20 µm. (**B**) Increased mNeon expression corresponded to increases in *in situ* sequencing reads and was used to identify transduced cells. (**C**) Across the datasets, more than 80% of transduced cells contained at least one *in situ* sequencing reads that matched a designed sgRNA sequence. (**D-F**) Plate heatmaps demonstrating throughput of the screen in terms of total plated cells (D), fraction of transduced (mNeon+) cells (E), and the fraction of transduced cells with sequencing reads uniquely mapping to designed sequences (F).

the strongest effect sizes, as Ppp3ca encodes for the Calcineurin protein targeted by the antibody, and Grin1 is a subunit of the NMDA receptors that transduce the signaling event into the neurons. The remaining gene targets are slightly less directly involved in the Ppp3ca pS469 signaling event, but the complete absence of effect is surprising. This, combined with the weak effects observed in Ppp3ca and Grin1 gene targets even with high cell representation (>3,000 cells represented in each distribution shown in Fig. A.3B), points to likely inefficiencies in Cas9 gene targeting in this trial screen. At this current stage, the weak effects for positive control genes makes larger screens infeasible despite the relative success of applying *in situ* SBS in primary neuron cultures. In the future, perturbation efficiency in this system should be optimized by further iterations of the sgRNA or Cas9 delivery vector. In these experiments, Cas9 expression was delivered using an adeno-associated virus (AAV) vector. Additional approaches should be considered including procuring neurons from engineered Cas9-expressing rat models or directly delivering Cas9 ribonucleoprotein (RNP) via electroporation (*209*). Additionally, the overall throughput of primary cell screens is limited in the low MOI regime as transduced cells cannot be easily enriched, resulting in many of the imaged cells being uninformative. A FACS-based selection may be feasible in some cell types (although likely not neurons), while high MOI experiments may be possible by employing compressed experiment approaches similar to those discussed in Section 3.3 for combinatorial interaction screening.

A.4 Methods

Hippocampal neurons were dissected from rat (*Rattus norvegicus)* E18 embryos, and plated on glass-bottom imaging plates (Cellvis P24-1.5H-N) coated with poly-D-lysine. After 2 or 5 DIV, the cultures were infected with AAV8-Cas9 (*210*) and sgRNA libraries in the CROPseq-derived vectors described in section A.2. At 16-21 DIV, the cultures were optionally treated with 0.5 µM TTX for 30 minutes, 0.5 µM TTX for 30 minutes followed by 20 µM NMDA stimulation for 10

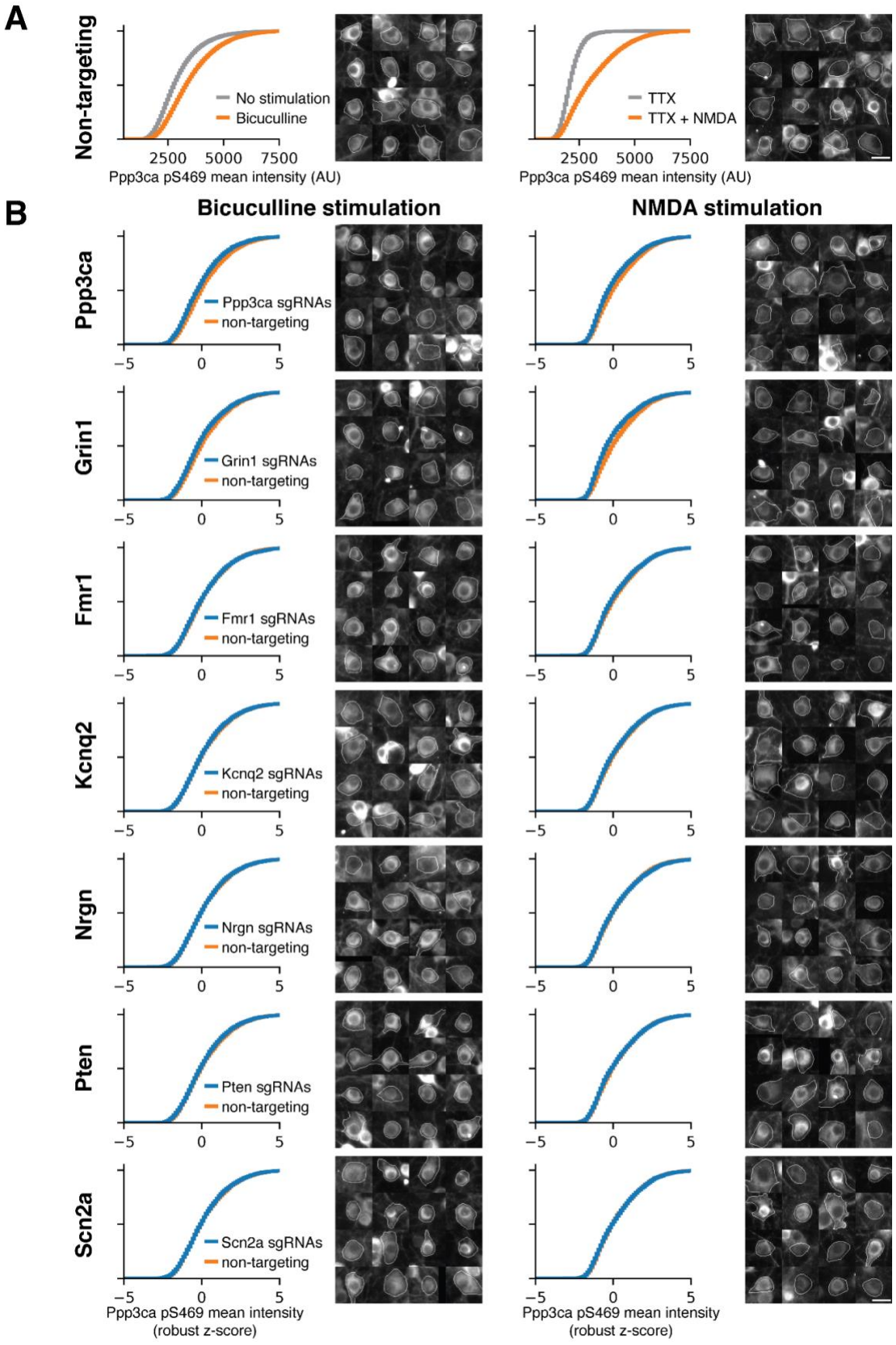
A
Non-targeting
No stimulation
Bicuculline
2500 5000 7500
Ppp3ca pS469 mean intensity (AU)
TTX
TTX + NMDA
2500 5000 7500
Ppp3ca pS469 mean intensity (AU)
B
Bicuculline stimulation
NMDA stimulation
Ppp3ca
Ppp3ca sgRNAs
non-targeting
Grin1
Grin1 sgRNAs
non-targeting
Fmr1
Fmr1 sgRNAs
non-targeting
Kcnq2
Kcnq2 sgRNAs
non-targeting
Nrgn
Nrgn sgRNAs
non-targeting
Pten
Pten sgRNAs
non-targeting
Scn2a
Scn2a sgRNAs
non-targeting
−5 0 5
Ppp3ca pS469 mean intensity
(robust z-score)
Ppp3ca pS469 mean intensity
(robust z-score)

Figure A.3. Phenotype measurements from the primary neuron trial screen. (**A**) Cumulative distribution plots of all cells expressing non-targeting sgRNAs and example images showing increases in Calcineurin (Ppp3ca) pS469 phosphorylation after stimulation by Bicuculline or NMDA. (**B**) Cumulative distributions for all cells expressing sgRNAs targeting the indicated gene and example images following Bicuculline or NMDA stimulation. Mean intensity was normalized using the median and median absolute deviation of the non-targeting cell population from within the same well to generate robust z-scores. Scale bars, 20 µm.

minutes, or 20 µM Bicuculline stimulation for 5 minutes. Cells were then immediately fixed with 4% paraformaldehyde for 15 minutes. *In situ* sequencing and corresponding image analysis was performed as described in Chapter 2. The sgRNAs used for the primary neurons experiments were designed against the *Rattus norvegicus* target genes using either CRISPick (https://portals.broadinstitute.org/gppx/crispick/public) or E-CRISP (*211*).

References

1. D. Feldman, L. Funk, A. Le, R. J. Carlson, M. D. Leiken, F. Tsai, B. Soong, A. Singh, P. C. Blainey, Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).
2. S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A. V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V. Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V. Maduro, T. Marschall, A. M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sović, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, The complete sequence of a human genome. *Science*. **376**, 44–53 (2022).
3. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R.

Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S.-P. Yang, R.-F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, M. J. Morgan, International Human Genome Sequencing Consortium, C. for G. R. Whitehead Institute for Biomedical Research, The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, I. of M. B. Department of Genome Analysis, GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, T. I. for S. B. Multimegabase Sequencing Center, Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, L. A. H. G. C. Cold Spring Harbor Laboratory, GBF—German Research Centre for Biotechnology:, also includes individuals listed under other headings): *Genome Analysis Group (listed in alphabetical order, U. N. I. of H. Scientific management: National Human Genome Research Institute, Stanford Human Genome Center:, University of Washington Genome Center:, K. U. S. of M. Department of Molecular Biology, University of Texas Southwestern Medical Center at Dallas:, U. D. of E. Office of Science, The Wellcome Trust:, Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).

4. S. L. Forsburg, The art and design of genetic screens: yeast. *Nat. Rev. Genet.* **2**, 659–668 (2001).
5. E. M. Jorgensen, S. E. Mango, The art and design of genetic screens: Caenorhabditis elegans. *Nat. Rev. Genet.* **3**, 356–369 (2002).
6. D. St Johnston, The art and design of genetic screens: Drosophila melanogaster. *Nat. Rev. Genet.* **3**, 176–188 (2002).

7. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in Drosophila. *Nature*. **287**, 795–801 (1980).
8. P. Nurse, Genetic control of cell size at cell division in yeast. *Nature*. **256**, 547–551 (1975).
9. K. Kume, H. Cantwell, F. R. Neumann, A. W. Jones, A. P. Snijders, P. Nurse, A systematic genomic screen implicates nucleocytoplasmic transport and membrane growth in nuclear size control. *PLOS Genet.* **13**, e1006767 (2017).
10. S. Brenner, THE GENETICS OF CAENORHABDITIS ELEGANS. *Genetics*. **77**, 71–94 (1974).
11. S. Grimm, The art and design of genetic screens: mammalian culture cells. *Nat. Rev. Genet.* **5**, 179–189 (2004).
12. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. **391**, 806–811 (1998).
13. P. J. Paddison, J. M. Silva, D. S. Conklin, M. Schlabach, M. Li, S. Aruleba, V. Balija, A. O'Shaughnessy, L. Gnoj, K. Scobie, K. Chang, T. Westbrook, M. Cleary, R. Sachidanandam, W. Richard McCombie, S. J. Elledge, G. J. Hannon, A resource for large-scale RNA-interference-based screens in mammals. *Nature*. **428**, 427–431 (2004).
14. K. Berns, E. M. Hijmans, J. Mullenders, T. R. Brummelkamp, A. Velds, M. Heimerikx, R. M. Kerkhoven, M. Madiredjo, W. Nijkamp, B. Weigelt, R. Agami, W. Ge, G. Cavet, P. S. Linsley, R. L. Beijersbergen, R. Bernards, A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*. **428**, 431–437 (2004).
15. J. M. Silva, K. Marran, J. S. Parker, J. Silva, M. Golding, M. R. Schlabach, S. J. Elledge, G. J. Hannon, K. Chang, Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science*. **319**, 617–620 (2008).
16. B. Neumann, T. Walter, J.-K. Hériché, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Cetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wünsche, V. Satagopam, M. H. A. Schmitz, C. Chapuis, D. W. Gerlich, R. Schneider, R. Eils, W. Huber, J.-M. Peters, A. A. Hyman, R. Durbin, R. Pepperkok, J. Ellenberg, Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*. **464**, 721–727 (2010).
17. A. L. Jackson, S. R. Bartz, J. Schelter, S. V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, P. S. Linsley, Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
18. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. **339**, 819–823 (2013).
19. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. **337**, 816–821 (2012).
20. O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*. **343**, 84–87 (2014).
21. T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*. **343**, 80–84 (2014).
22. J. Joung, S. Konermann, J. S. Gootenberg, O. O. Abudayyeh, R. J. Platt, M. D. Brigham, N. E. Sanjana, F. Zhang, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
23. G. Michlits, M. Hubmann, S.-H. Wu, G. Vainorius, E. Budusan, S. Zhuk, T. R. Burkard, M. Novatchkova, M. Aichinger, Y. Lu, J. Reece-Hoyes, R. Nitsch, D. Schramek, D. Hoepfner, U. Elling, CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat. Methods*. **14**, 1191–1197 (2017).

24. B. Schmierer, S. K. Botla, J. Zhang, M. Turunen, T. Kivioja, J. Taipale, CRISPR/Cas9 screening using unique molecular identifiers. *Mol. Syst. Biol.* **13**, 945 (2017).
25. A. Dixit, O. Parnas, B. Li, J. Chen, C. P. Fulco, L. Jerby-Arnon, N. D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T. M. Norman, E. S. Lander, J. S. Weissman, N. Friedman, A. Regev, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. **167**, 1853-1866.e17 (2016).
26. B. Adamson, T. M. Norman, M. Jost, M. Y. Cho, J. K. Nuñez, Y. Chen, J. E. Villalta, L. A. Gilbert, M. A. Horlbeck, M. Y. Hein, R. A. Pak, A. N. Gray, C. A. Gross, A. Dixit, O. Parnas, A. Regev, J. S. Weissman, A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. **167**, 1867-1882.e21 (2016).
27. D. A. Jaitin, A. Weiner, I. Yofe, D. Lara-Astiaso, H. Keren-Shaul, E. David, T. M. Salame, A. Tanay, A. van Oudenaarden, I. Amit, Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. **167**, 1883-1896.e15 (2016).
28. A. Wroblewska, M. Dhainaut, B. Ben-Zvi, S. A. Rose, E. S. Park, E.-A. D. Amir, A. Bektesevic, A. Baccarini, M. Merad, A. H. Rahman, B. D. Brown, Protein Barcodes Enable High-Dimensional Single-Cell CRISPR Screens. *Cell*. **175**, 1141-1155.e16 (2018).
29. A. J. Rubin, K. R. Parker, A. T. Satpathy, Y. Qi, B. Wu, A. J. Ong, M. R. Mumbach, A. L. Ji, D. S. Kim, S. W. Cho, B. J. Zarnegar, W. J. Greenleaf, H. Y. Chang, P. A. Khavari, Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell*. **176**, 361-376.e17 (2019).
30. E. P. Mimitou, A. Cheng, A. Montalbano, S. Hao, M. Stoeckius, M. Legut, T. Roush, A. Herrera, E. Papalexi, Z. Ouyang, R. Satija, N. E. Sanjana, S. B. Koralov, P. Smibert, Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods*. **16**, 409–412 (2019).
31. A. J. M. Wollman, R. Nudd, E. G. Hedlund, M. C. Leake, From Animaculum to single molecules: 300 years of the light microscope. *Open Biol.* **5**, 150019 (2015).
32. S. K. Saka, Y. Wang, J. Y. Kishi, A. Zhu, Y. Zeng, W. Xie, K. Kirli, C. Yapp, M. Cicconet, B. J. Beliveau, S. W. Lapan, S. Yin, M. Lin, E. S. Boyden, P. S. Kaeser, G. Pihan, G. M. Church, P. Yin, Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
33. S. Codeluppi, L. E. Borm, A. Zeisel, G. La Manno, J. A. van Lunteren, C. I. Svensson, S. Linnarsson, Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods*. **15**, 932–935 (2018).
34. J.-R. Lin, M. Fallahi-Sichani, P. K. Sorger, Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* **6**, 8390 (2015).
35. G. Gut, M. D. Herrmann, L. Pelkmans, Multiplexed protein maps link subcellular organization to cellular states. *Science*. **361** (2018), doi:10.1126/science.aar7042.
36. T. Zimmermann, J. Rietdorf, R. Pepperkok, Spectral imaging and its applications in live cell microscopy. *FEBS Lett.* **546**, 87–92 (2003).
37. S. Regot, J. J. Hughey, B. T. Bajar, S. Carrasco, M. W. Covert, High-Sensitivity Measurements of Multiple Kinase Activities in Live Single Cells. *Cell*. **157**, 1724–1734 (2014).
38. D. R. Hochbaum, Y. Zhao, S. L. Farhi, N. Klapoetke, C. A. Werley, V. Kapoor, P. Zou, J. M. Kralj, D. Maclaurin, N. Smedemark-Margulies, J. L. Saulnier, G. L. Boulting, C. Straub, Y. K. Cho, M. Melkonian, G. K.-S. Wong, D. J. Harrison, V. N. Murthy, B. L. Sabatini, E. S. Boyden, R. E. Campbell, A. E. Cohen, All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods*. **11**, 825–833 (2014).
39. Y. Zhao, S. Araki, J. Wu, T. Teramoto, Y.-F. Chang, M. Nakano, A. S. Abdelfattah, M. Fujiwara, T. Ishihara, T. Nagai, R. E. Campbell, An Expanded Palette of Genetically

Encoded Ca2+ Indicators. *Science*. **333**, 1888–1891 (2011).
40. M. Z. Lin, M. J. Schnitzer, Genetically encoded indicators of neuronal activity. *Nat. Neurosci.* **19**, 1142–1153 (2016).
41. D. E. Nelson, A. E. C. Ihekwaba, M. Elliott, J. R. Johnson, C. A. Gibney, B. E. Foreman, G. Nelson, V. See, C. A. Horton, D. G. Spiller, S. W. Edwards, H. P. McDowell, J. F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D. B. Kell, M. R. H. White, Oscillations in NF-κB Signaling Control the Dynamics of Gene Expression. *Science*. **306**, 704–708 (2004).
42. J. G. Albeck, G. B. Mills, J. S. Brugge, Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. *Mol. Cell*. **49**, 249–261 (2013).
43. C. Cohen-Saidon, A. A. Cohen, A. Sigal, Y. Liron, U. Alon, Dynamics and Variability of ERK2 Response to EGF in Individual Living Cells. *Mol. Cell*. **36**, 885–893 (2009).
44. G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A. J. Levine, M. B. Elowitz, U. Alon, Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat. Genet.* **36**, 147–150 (2004).
45. A. Sakaue-Sawano, H. Kurokawa, T. Morimura, A. Hanyu, H. Hama, H. Osawa, S. Kashiwagi, K. Fukami, T. Miyata, H. Miyoshi, T. Imamura, M. Ogawa, H. Masai, A. Miyawaki, Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*. **132**, 487–498 (2008).
46. T. Biederer, P. Scheiffele, Mixed-culture assays for analyzing neuronal synapse formation. *Nat. Protoc.* **2**, 670–676 (2007).
47. P. Scheiffele, J. Fan, J. Choih, R. Fetter, T. Serafini, Neuroligin Expressed in Nonneuronal Cells Triggers Presynaptic Development in Contacting Axons. *Cell*. **101**, 657–669 (2000).
48. N. Varadarajan, B. Julg, Y. J. Yamanaka, H. Chen, A. O. Ogunniyi, E. McAndrew, L. C. Porter, A. Piechocka-Trocha, B. J. Hill, D. C. Douek, F. Pereyra, B. D. Walker, J. C. Love, A high-throughput single-cell analysis of human $CD8^+$ T cell functions reveals discordance for cytokine secretion and cytolysis. *J. Clin. Invest.* **121**, 4322–4331 (2011).
49. E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, Deep learning for cellular image analysis. *Nat. Methods*. **16**, 1233–1246 (2019).
50. B. T. Grys, D. S. Lo, N. Sahin, O. Z. Kraus, Q. Morris, C. Boone, B. J. Andrews, Machine learning and computer vision approaches for phenotypic profiling. *J. Cell Biol.* **216**, 65–71 (2016).
51. J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington, A. E. Carpenter, Data-analysis strategies for image-based cell profiling. *Nat. Methods*. **14**, 849–863 (2017).
52. J. C. Caicedo, C. McQuin, A. Goodman, S. Singh, A. E. Carpenter, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 9309–9318.
53. H. Kobayashi, K. C. Cheveralls, M. D. Leonetti, L. A. Royer, Self-Supervised Deep Learning Encodes High-Resolution Features of Protein Subcellular Localization. *bioRxiv* (2022), p. 2021.03.29.437595, , doi:10.1101/2021.03.29.437595.
54. Ž. Strezoska, M. R. Perkett, E. T. Chou, E. Maksimova, E. M. Anderson, S. McClelland, M. L. Kelley, A. Vermeulen, A. van B. Smith, High-content analysis screening for cell cycle regulators using arrayed synthetic crRNA libraries. *J. Biotechnol.* **251**, 189–200 (2017).
55. C. Collinet, M. Stöter, C. R. Bradshaw, N. Samusik, J. C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M. S. Mueller, W. E. Nagel, E. Fava, Y. Kalaidzidis, M. Zerial, Systems survey of endocytosis by multiparametric image analysis. *Nature*. **464**, 243–249 (2010).
56. A. Orvedahl, R. Sumpter, G. Xiao, A. Ng, Z. Zou, Y. Tang, M. Narimatsu, C. Gilpin, Q. Sun, M. Roth, C. V. Forst, J. L. Wrana, Y. E. Zhang, K. Luby-Phelps, R. J. Xavier, Y. Xie, B.

Levine, Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature*. **480**, 113–117 (2011).
57. A. Karlas, N. Machuy, Y. Shin, K.-P. Pleissner, A. Artarini, D. Heuer, D. Becker, H. Khalil, L. A. Ogilvie, S. Hess, A. P. Mäurer, E. Müller, T. Wolff, T. Rudel, T. F. Meyer, Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*. **463**, 818–822 (2010).
58. J. Mercer, B. Snijder, R. Sacher, C. Burkard, C. K. E. Bleck, H. Stahlberg, L. Pelkmans, A. Helenius, RNAi Screening Reveals Proteasome- and Cullin3-Dependent Stages in Vaccinia Virus Infection. *Cell Rep.* **2**, 1036–1047 (2012).
59. H. Agaisse, L. S. Burrack, J. A. Philips, E. J. Rubin, N. Perrimon, D. E. Higgins, Genome-Wide RNAi Screen for Host Factors Required for Intracellular Bacterial Infection. *Science*. **309**, 1248–1251 (2005).
60. H. S. Kim, K. Lee, S.-J. Kim, S. Cho, H. J. Shin, C. Kim, J.-S. Kim, Arrayed CRISPR screen with image-based assay reliably uncovers host genes required for coxsackievirus infection. *Genome Res.* **28**, 859–868 (2018).
61. R. de Groot, J. Lüthi, H. Lindsay, R. Holtackers, L. Pelkmans, Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol. Syst. Biol.* **14**, e8064 (2018).
62. D. Schraivogel, T. M. Kuhn, B. Rauscher, M. Rodríguez-Martínez, M. Paulsen, K. Owsley, A. Middlebrook, C. Tischer, B. Ramasz, D. Ordoñez-Rueda, M. Dees, S. Cuylen-Haering, E. Diebold, L. M. Steinmetz, High-speed fluorescence image–enabled cell sorting. *Science*. **375**, 315–320 (2022).
63. N. Hasle, A. Cooke, S. Srivatsan, H. Huang, J. J. Stephany, Z. Krieger, D. Jackson, W. Tang, S. Pendyala, R. J. Monnat Jr., C. Trapnell, E. M. Hatch, D. M. Fowler, High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* **16**, e9442 (2020).
64. X. Yan, N. Stuurman, S. A. Ribeiro, M. E. Tanenbaum, M. A. Horlbeck, C. R. Liem, M. Jost, J. S. Weissman, R. D. Vale, High-content imaging-based pooled CRISPR screens in mammalian cells. *J. Cell Biol.* **220** (2021), doi:10.1083/jcb.202008158.
65. J. Lee, Z. Liu, P. H. Suzuki, J. F. Ahrens, S. Lai, X. Lu, S. Guan, F. St-Pierre, Versatile phenotype-activated cell sorting. *Sci. Adv.* **6**, eabb7438 (2020).
66. G. Kanfer, S. A. Sarraf, Y. Maman, H. Baldwin, E. Dominguez-Martin, K. R. Johnson, M. E. Ward, M. Kampmann, J. Lippincott-Schwartz, R. J. Youle, Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J. Cell Biol.* **220** (2021), doi:10.1083/jcb.202006180.
67. E. C. Wheeler, A. Q. Vu, J. M. Einstein, M. DiSalvo, N. Ahmed, E. L. Van Nostrand, A. A. Shishkin, W. Jin, N. L. Allbritton, G. W. Yeo, Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. *Nat. Methods*. **17**, 636–642 (2020).
68. M. J. Lawson, D. Camsund, J. Larsson, Ö. Baltekin, D. Fange, J. Elf, In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
69. D. Camsund, M. J. Lawson, J. Larsson, D. Jones, S. Zikrin, D. Fange, J. Elf, Time-resolved imaging-based CRISPRi screening. *Nat. Methods*. **17**, 86–92 (2020).
70. G. Emanuel, J. R. Moffitt, X. Zhuang, High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods*. **14**, 1159–1162 (2017).
71. C. Wang, T. Lu, G. Emanuel, H. P. Babcock, X. Zhuang, Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl. Acad. Sci.* **116**, 10842–10851 (2019).
72. P. Datlinger, A. F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L. C. Schuster, A. Kuchler, D. Alpar, C. Bock, Pooled CRISPR screening with single-cell

transcriptome readout. *Nat. Methods*. **14**, 297–301 (2017).
73. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*. **10**, 857–860 (2013).
74. C. Larsson, I. Grundberg, O. Söderberg, M. Nilsson, In situ detection and genotyping of individual mRNA molecules. *Nat. Methods*. **7**, 395–397 (2010).
75. D. Feldman, A. Singh, J. L. Schmid-Burgk, R. J. Carlson, A. Mezger, A. J. Garrity, F. Zhang, P. C. Blainey, Optical Pooled Screens in Human Cells. *Cell*. **179**, 787-799.e17 (2019).
76. D. K. Breslow, D. M. Cameron, S. R. Collins, M. Schuldiner, J. Stewart-Ornstein, H. W. Newman, S. Braun, H. D. Madhani, N. J. Krogan, J. S. Weissman, A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods*. **5**, 711–718 (2008).
77. L. Funk, K.-C. Su, D. Feldman, A. Singh, B. Moodie, P. C. Blainey, I. M. Cheeseman, The phenotypic landscape of essential human genes. *bioRxiv* (2021), p. 2021.11.28.470116, , doi:10.1101/2021.11.28.470116.
78. K. J. Condon, J. M. Orozco, C. H. Adelmann, J. B. Spinelli, P. W. van der Helm, J. M. Roberts, T. Kunchok, D. M. Sabatini, Genome-wide CRISPR screens reveal multitiered mechanisms through which mTORC1 senses mitochondrial dysfunction. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2022120118 (2021).
79. J. Nieuwenhuis, A. Adamopoulos, O. B. Bleijerveld, A. Mazouzi, E. Stickel, P. Celie, M. Altelaar, P. Knipscheer, A. Perrakis, V. A. Blomen, T. R. Brummelkamp, Vasohibins encode tubulin detyrosinating activity. *Science*. **358**, 1453–1456 (2017).
80. V. A. Blomen, P. Májek, L. T. Jae, J. W. Bigenzahn, J. Nieuwenhuis, J. Staring, R. Sacco, F. R. van Diemen, N. Olk, A. Stukalov, C. Marceau, H. Janssen, J. E. Carette, K. L. Bennett, J. Colinge, G. Superti-Furga, T. R. Brummelkamp, Gene essentiality and synthetic lethality in haploid human cells. *Science*. **350**, 1092–1096 (2015).
81. J. M. Dempster, J. Rossen, M. Kazachkova, J. Pan, G. Kugener, D. E. Root, A. Tsherniak, Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *bioRxiv*, 720243 (2019).
82. DepMap, Broad., DepMap 19Q3 Public. figshare. Dataset doi:10.6084/m9.figshare.9201770.v2.
83. T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K. R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F. P. Roth, O. S. Rissland, D. Durocher, S. Angers, J. Moffat, High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. **163**, 1515–1526 (2015).
84. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife*. **5**, e19760 (2016).
85. R. M. Meyers, J. G. Bryan, J. M. McFarland, B. A. Weir, A. E. Sizemore, H. Xu, N. V. Dharia, P. G. Montgomery, G. S. Cowley, S. Pantel, A. Goodale, Y. Lee, L. D. Ali, G. Jiang, R. Lubonja, W. F. Harrington, M. Strickland, T. Wu, D. C. Hawes, V. A. Zhivich, M. R. Wyatt, Z. Kalani, J. J. Chang, M. Okamoto, K. Stegmaier, T. R. Golub, J. S. Boehm, F. Vazquez, D. E. Root, W. C. Hahn, A. Tsherniak, Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
86. K. Tzelepis, H. Koike-Yusa, E. De Braekeleer, Y. Li, E. Metzakopian, O. M. Dovey, A. Mupo, V. Grinkevich, M. Li, M. Mazan, M. Gozdecka, S. Ohnishi, J. Cooper, M. Patel, T. McKerrell, B. Chen, A. F. Domingues, P. Gallipoli, S. Teichmann, H. Ponstingl, U. McDermott, J. Saez-Rodriguez, B. J. P. Huntly, F. Iorio, C. Pina, G. S. Vassiliou, K. Yusa,

A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* **17**, 1193–1205 (2016).
87. T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, D. M. Sabatini, Identification and characterization of essential genes in the human genome. *Science*. **350**, 1096–1101 (2015).
88. T. Wang, H. Yu, N. W. Hughes, B. Liu, A. Kendirli, K. Klein, W. W. Chen, E. S. Lander, D. M. Sabatini, Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*. **168**, 890-903.e15 (2017).
89. J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, D. E. Root, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
90. T. Hart, A. H. Y. Tong, K. Chan, J. Van Leeuwen, A. Seetharaman, M. Aregger, M. Chandrashekhar, N. Hustedt, S. Seth, A. Noonan, A. Habsid, O. Sizova, L. Nedyalkova, R. Climie, L. Tworzyanski, K. Lawson, M. A. Sartori, S. Alibeh, D. Tieu, S. Masud, P. Mero, A. Weiss, K. R. Brown, M. Usaj, M. Billmann, M. Rahman, M. Constanzo, C. L. Myers, B. J. Andrews, C. Boone, D. Durocher, J. Moffat, Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Bethesda Md*. **7**, 2719–2727 (2017).
91. K. L. McKinley, I. M. Cheeseman, Large-Scale Analysis of CRISPR/Cas9 Cell-Cycle Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects. *Dev. Cell*. **40**, 405-420.e2 (2017).
92. Y. Dang, G. Jia, J. Choi, H. Ma, E. Anaya, C. Ye, P. Shankar, H. Wu, Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* **16**, 280 (2015).
93. A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kaçmaz, S. Linn, Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
94. S. P. Bell, A. Dutta, DNA Replication in Eukaryotic Cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
95. C. Pederiva, S. Böhm, A. Julner, M. Farnebo, Splicing controls the ubiquitin response during DNA double-strand break repair. *Cell Death Differ.* **23**, 1648–1657 (2016).
96. M. Carmena, M. Wheelock, H. Funabiki, W. C. Earnshaw, The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat. Rev. Mol. Cell Biol.* **13**, 789–803 (2012).
97. T. D. Pollard, B. O'Shaughnessy, Molecular Mechanism of Cytokinesis. *Annu. Rev. Biochem.* **88**, 661–689 (2019).
98. P. Lara-Gonzalez, J. Pines, A. Desai, Spindle assembly checkpoint activation and silencing at kinetochores. *Semin. Cell Dev. Biol.* **117**, 86–98 (2021).
99. R. Shi, W. Hou, Z.-Q. Wang, X. Xu, Biogenesis of Iron–Sulfur Clusters and Their Role in DNA Metabolism. *Front. Cell Dev. Biol.* **9**, 2676 (2021).
100. F. Villa, R. Fujisawa, J. Ainsworth, K. Nishimura, M. Lie-A-Ling, G. Lacaud, K. P. Labib, CUL2LRR1, TRAIP and p97 control CMG helicase disassembly in the mammalian cell cycle. *EMBO Rep.* **22**, e52164 (2021).
101. M. Olivieri, T. Cho, A. Álvarez-Quilón, K. Li, M. J. Schellenberg, M. Zimmermann, N. Hustedt, S. E. Rossi, S. Adam, H. Melo, A. M. Heijink, G. Sastre-Moreno, N. Moatti, R. K. Szilard, A. McEwan, A. K. Ling, A. Serrano-Benitez, T. Ubhi, S. Feng, J. Pawling, I. Delgado-Sainz, M. W. Ferguson, J. W. Dennis, G. W. Brown, F. Cortés-Ledesma, R. S. Williams, A. Martin, D. Xu, D. Durocher, A Genetic Map of the Response to DNA Damage in Human Cells. *Cell*. **182**, 481-496.e21 (2020).
102. T. D. Pollard, Actin and Actin-Binding Proteins. *Cold Spring Harb. Perspect. Biol.* **8**, a018226 (2016).
103. H. V. Goodson, E. M. Jonasson, Microtubules and Microtubule-Associated Proteins. *Cold*

Spring Harb. Perspect. Biol. **10**, a022608 (2018).
104. L. Li, W. Zhang, Y. Liu, X. Liu, L. Cai, J. Kang, Y. Zhang, W. Chen, C. Dong, Y. Zhang, M. Wang, W. Wei, L. Jia, The CRL3BTBD9 E3 ubiquitin ligase complex targets TNFAIP1 for degradation to suppress cancer cell migration. *Signal Transduct. Target. Ther.* **5**, 1–9 (2020).
105. F. Rodríguez-Pérez, A. G. Manford, A. Pogson, A. J. Ingersoll, B. Martínez-González, M. Rape, Ubiquitin-dependent remodeling of the actin cytoskeleton drives cell fusion. *Dev. Cell*. **56**, 588-601.e9 (2021).
106. D. Zyss, H. Ebrahimi, F. Gergely, Casein kinase I delta controls centrosome positioning during T cell activation. *J. Cell Biol.* **195**, 781–797 (2011).
107. M. A. Collart, The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *WIREs RNA*. **7**, 438–454 (2016).
108. H. Cantwell, P. Nurse, Unravelling nuclear size control. *Curr. Genet.* **65**, 1281–1285 (2019).
109. K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, S. Krishnaswamy, Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
110. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
111. E. L. Huttlin, R. J. Bruckner, J. Navarrete-Perea, J. R. Cannon, K. Baltier, F. Gebreab, M. P. Gygi, A. Thornock, G. Zarraga, S. Tam, J. Szpyt, B. M. Gassaway, A. Panov, H. Parzen, S. Fu, A. Golbazi, E. Maenpaa, K. Stricker, S. Guha Thakurta, T. Zhang, R. Rad, J. Pan, D. P. Nusinow, J. A. Paulo, D. K. Schweppe, L. P. Vaites, J. W. Harper, S. P. Gygi, Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*. **184**, 3022-3040.e28 (2021).
112. M. Wainberg, R. A. Kamber, A. Balsubramani, R. M. Meyers, N. Sinnott-Armstrong, D. Hornburg, L. Jiang, J. Chan, R. Jian, M. Gu, A. Shcherbina, M. M. Dubreuil, K. Spees, W. Meuleman, M. P. Snyder, M. C. Bassik, A. Kundaje, A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat. Genet.* **53**, 638–649 (2021).
113. R. Haq, J. Shoag, P. Andreu-Perez, S. Yokoyama, H. Edelman, G. C. Rowe, D. T. Frederick, A. D. Hurley, A. Nellore, A. L. Kung, J. A. Wargo, J. S. Song, D. E. Fisher, Z. Arany, H. R. Widlund, Oncogenic BRAF Regulates Oxidative Metabolism via PGC1α and MITF. *Cancer Cell*. **23**, 302–315 (2013).
114. G.-Y. Liou, H. Döppler, K. E. DelGiorno, L. Zhang, M. Leitges, H. C. Crawford, M. P. Murphy, P. Storz, Mutant KRas-Induced Mitochondrial Oxidative Stress in Acinar Cells Upregulates EGFR Signaling to Drive Formation of Pancreatic Precancerous Lesions. *Cell Rep.* **14**, 2325–2336 (2016).
115. F. Weinberg, R. Hamanaka, W. W. Wheaton, S. Weinberg, J. Joseph, M. Lopez, B. Kalyanaraman, G. M. Mutlu, G. R. S. Budinger, N. S. Chandel, Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc. Natl. Acad. Sci.* **107**, 8788–8793 (2010).
116. Z. Cao, K. A. Budinich, H. Huang, D. Ren, B. Lu, Z. Zhang, Q. Chen, Y. Zhou, Y.-H. Huang, F. Alikarami, M. C. Kingsley, A. K. Lenard, A. Wakabayashi, E. Khandros, W. Bailis, J. Qi, M. P. Carroll, G. A. Blobel, R. B. Faryabi, K. M. Bernt, S. L. Berger, J. Shi, ZMYND8-regulated IRF8 transcription axis is an acute myeloid leukemia dependency. *Mol. Cell*. **81**, 3604-3622.e10 (2021).
117. S. Singh, A. Vanden Broeck, L. Miller, M. Chaker-Margot, S. Klinge, Nucleolar maturation of the human small subunit processome. *Science*. **373**, eabj5338 (2021).
118. M. de Almeida, M. Hinterndorfer, H. Brunner, I. Grishkovskaya, K. Singh, A. Schleiffer, J.

Jude, S. Deswal, R. Kalis, M. Vunjak, T. Lendl, R. Imre, E. Roitinger, T. Neumann, S. Kandolf, M. Schutzbier, K. Mechtler, G. A. Versteeg, D. Haselbach, J. Zuber, AKIRIN2 controls the nuclear import of proteasomes in vertebrates. *Nature*, 1–6 (2021).
119. H. Song, X. Feng, H. Zhang, Y. Luo, J. Huang, M. Lin, J. Jin, X. Ding, S. Wu, H. Huang, T. Yu, M. Zhang, H. Hong, S. Yao, Y. Zhao, Z. Zhang, METTL3 and ALKBH5 oppositely regulate m6A modification of TFEB mRNA, which dictates the fate of hypoxia/reoxygenation-treated cardiomyocytes. *Autophagy*. **15**, 1419–1437 (2019).
120. D. Baillat, M.-A. Hakimi, A. M. Näär, A. Shilatifard, N. Cooch, R. Shiekhattar, Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell*. **123**, 265–276 (2005).
121. J. N. Jodoin, P. Sitaram, T. R. Albrecht, S. B. May, M. Shboul, E. Lee, B. Reversade, E. J. Wagner, L. A. Lee, Nuclear-localized Asunder regulates cytoplasmic dynein localization via its role in the integrator complex. *Mol. Biol. Cell*. **24**, 2954–2965 (2013).
122. A. Malovannaya, Y. Li, Y. Bulynko, S. Y. Jung, Y. Wang, R. B. Lanz, B. W. O'Malley, J. Qin, Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl. Acad. Sci.* **107**, 2431–2436 (2010).
123. K. Sabath, M. L. Stäubli, S. Marti, A. Leitner, M. Moes, S. Jonas, INTS10–INTS13–INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat. Commun.* **11**, 3422 (2020).
124. J. M. Replogle, R. A. Saunders, A. N. Pogson, J. A. Hussmann, A. Lenail, A. Guna, L. Mascibroda, E. J. Wagner, K. Adelman, J. L. Bonnar, M. Jost, T. M. Norman, J. S. Weissman, Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *bioRxiv* (2021), , doi:10.1101/2021.12.16.473013.
125. J. Pan, J. J. Kwon, J. A. Talamas, A. A. Borah, F. Vazquez, J. S. Boehm, A. Tsherniak, M. Zitnik, J. M. McFarland, W. C. Hahn, Sparse dictionary learning recovers pleiotropy from human cell fitness screens. *Cell Syst.* (2022), doi:10.1016/j.cels.2021.12.005.
126. Y. J. Yang, A. E. Baltus, R. S. Mathew, E. A. Murphy, G. D. Evrony, D. M. Gonzalez, E. P. Wang, C. A. Marshall-Walker, B. J. Barry, J. Murn, A. Tatarakis, M. A. Mahajan, H. H. Samuels, Y. Shi, J. A. Golden, M. Mahajnah, R. Shenhav, C. A. Walsh, Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation and differentiation. *Cell*. **151**, 1097–1112 (2012).
127. D. Jayaraman, B.-I. Bae, C. A. Walsh, The Genetics of Primary Microcephaly. *Annu. Rev. Genomics Hum. Genet.* **19**, 177–200 (2018).
128. G. Goshima, R. Wollman, S. S. Goodwin, N. Zhang, J. M. Scholey, R. D. Vale, N. Stuurman, Genes Required for Mitotic Spindle Assembly in Drosophila S2 Cells. *Science*. **316**, 417–421 (2007).
129. W. Piwko, L. J. Mlejnkova, K. Mutreja, L. Ranjha, D. Stafa, A. Smirnov, M. M. Brodersen, R. Zellweger, A. Sturzenegger, P. Janscak, M. Lopes, M. Peter, P. Cejka, The MMS22L–TONSL heterodimer directly promotes RAD51-dependent recombination upon replication stress. *EMBO J.* **35**, 2584–2601 (2016).
130. U. Jo, W. Cai, J. Wang, Y. Kwon, A. D. D'Andrea, H. Kim, PCNA-Dependent Cleavage and Degradation of SDE2 Regulates Response to Replication Stress. *PLoS Genet.* **12**, e1006465 (2016).
131. A. Kumagai, A. Shevchenko, A. Shevchenko, W. G. Dunphy, Treslin Collaborates with TopBP1 in Triggering the Initiation of DNA Replication. *Cell*. **140**, 349–359 (2010).
132. H. M. Taïeb, D. S. Garske, J. Contzen, M. Gossen, L. Bertinetti, T. Robinson, A. Cipitria, Osmotic pressure modulates single cell cycle dynamics inducing reversible growth arrest and reactivation of human metastatic cells. *Sci. Rep.* **11**, 13455 (2021).
133. I. M. Cheeseman, The kinetochore. *Cold Spring Harb. Perspect. Biol.* **6**, a015826 (2014).
134. M. Fischer, G. A. Müller, Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Crit. Rev. Biochem. Mol. Biol.* **52**, 638–662 (2017).

135. J. Esterlechner, N. Reichert, F. Iltzsche, M. Krause, F. Finkernagel, S. Gaubatz, LIN9, a Subunit of the DREAM Complex, Regulates Mitotic Gene Expression and Proliferation of Embryonic Stem Cells. *PLOS ONE*. **8**, e62882 (2013).
136. M. A. Ghazy, J. M. B. Gordon, S. D. Lee, B. N. Singh, A. Bohm, M. Hampsey, C. Moore, The interaction of Pcf11 and Clp1 is needed for mRNA 3′-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic Acids Res.* **40**, 1214–1225 (2012).
137. C. Estell, L. Davidson, P. C. Steketee, A. Monier, S. West, ZC3H4 restricts non-coding transcription in human cells. *eLife*. **10**, e67305 (2021).
138. L. M. I. Austenaa, V. Piccolo, M. Russo, E. Prosperini, S. Polletti, D. Polizzese, S. Ghisletti, I. Barozzi, G. R. Diaferia, G. Natoli, A first exon termination checkpoint preferentially suppresses extragenic transcription. *Nat. Struct. Mol. Biol.* **28**, 337–346 (2021).
139. K. Kamieniarz-Gdula, M. R. Gdula, K. Panser, T. Nojima, J. Monks, J. R. Wiśniewski, J. Riepsaame, N. Brockdorff, A. Pauli, N. J. Proudfoot, Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. *Mol. Cell*. **74**, 158-172.e9 (2019).
140. B. Verma, M. V. Akinyi, A. J. Norppa, M. J. Frilander, Minor spliceosome and disease. *Semin. Cell Dev. Biol.* **79**, 103–112 (2018).
141. B. de Wolf, A. Oghabian, M. V. Akinyi, S. Hanks, E. C. Tromer, J. J. E. van Hooff, L. van Voorthuijsen, L. E. van Rooijen, J. Verbeeren, E. C. H. Uijttewaal, M. P. A. Baltissen, S. Yost, P. Piloquet, M. Vermeulen, B. Snel, B. Isidor, N. Rahman, M. J. Frilander, G. J. P. L. Kops, Chromosomal instability by mutations in the novel minor spliceosome component CENATAC. *EMBO J.* **40**, e106536 (2021).
142. THE GTEX CONSORTIUM, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. **369**, 1318–1330 (2020).
143. K. E. Gascoigne, K. Takeuchi, A. Suzuki, T. Hori, T. Fukagawa, I. M. Cheeseman, Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell*. **145**, 410–422 (2011).
144. K. L. McKinley, N. Sekulic, L. Y. Guo, T. Tsinman, B. E. Black, I. M. Cheeseman, The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol. Cell*. **60**, 886–898 (2015).
145. J. van den Berg, A. G. Manjón, K. Kielbassa, F. M. Feringa, R. Freire, R. H. Medema, A limited number of double-strand DNA breaks is sufficient to delay cell cycle progression. *Nucleic Acids Res.* **46**, 10132–10144 (2018).
146. J. P. Morgenstern, H. Land, Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
147. D. Feldman, L. Funk, Pooled genetic perturbation screens with image-based phenotypes, OpticalPooledScreens. (2021), , doi:https://doi.org/10.5281/zenodo.5002684.
148. S. Singh, M.-A. Bray, T. Jones, A. Carpenter, Pipeline for illumination correction of images for high-throughput microscopy. *J. Microsc.* **256**, 231–236 (2014).
149. C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo, A. E. Carpenter, CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* **16**, e2005970 (2018).
150. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python. *PeerJ*. **2**, e453 (2014).
151. L. P. Coelho, Mahotas: Open source software for scriptable computer vision. *J. Open Res. Softw.* **1**, e3 (2013).
152. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. **28**, 2520–2522 (2012).
153. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P.

Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
154. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
155. K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, G. Danuser, Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods*. **5**, 695–702 (2008).
156. J.-Y. Tinevez, N. Perry, J. Schindelin, G. M. Hoopes, G. D. Reynolds, E. Laplantine, S. Y. Bednarek, S. L. Shorte, K. W. Eliceiri, TrackMate: An open and extensible platform for single-particle tracking. *Methods*. **115**, 80–90 (2017).
157. I. M. Cheeseman, A. Desai, A Combined Approach for the Localization and Tandem Affinity Purification of Protein Complexes from Metazoans. *Sci. STKE*. **2005**, pl1–pl1 (2005).
158. J. C. Schmidt, H. Arthanari, A. Boeszoermenyi, N. M. Dashkevich, E. M. Wilson-Kubalek, N. Monnier, M. Markus, M. Oberer, R. A. Milligan, M. Bathe, G. Wagner, E. L. Grishchuk, I. M. Cheeseman, The Kinetochore-Bound Ska1 Complex Tracks Depolymerizing Microtubules and Binds to Curved Protofilaments. *Dev. Cell*. **23**, 968–980 (2012).
159. C. B. Backer, J. H. Gutzman, C. G. Pearson, I. M. Cheeseman, CSAP localizes to polyglutamylated microtubules and promotes proper cilia function and zebrafish development. *Mol. Biol. Cell*. **23**, 2122–2130 (2012).
160. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods*. **9**, 676–682 (2012).
161. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods*. **18**, 100–106 (2021).
162. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10–12 (2011).
163. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
164. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
165. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
166. A. M. Olthof, K. C. Hyatt, R. N. Kanadia, Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics*. **20**, 686 (2019).
167. S. X. Ge, D. Jung, R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. **36**, 2628–2629 (2020).
168. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
169. K. Pandit, J. Petrescu, M. Cuevas, W. Stephenson, P. Smibert, H. Phatnani, S. Maniatis, An open source toolkit for repurposing Illumina sequencing systems as versatile fluidics and imaging platforms. *Sci. Rep.* **12**, 5081 (2022).
170. D. Feldman, thesis, Massachusetts Institute of Technology (2019).
171. S. Liu, S. Punthambaker, E. P. R. Iyer, T. Ferrante, D. Goodwin, D. Fürth, A. C. Pawlowski, K. Jindal, J. M. Tam, L. Mifflin, S. Alon, A. Sinha, A. T. Wassie, F. Chen, A. Cheng, V. Willocq, K. Meyer, K.-H. Ling, C. K. Camplisson, R. E. Kohman, J. Aach, J. H. Lee, B. A. Yankner, E. S. Boyden, G. M. Church, Barcoded oligonucleotides ligated on RNA amplified

for multiplexed and parallel in situ analyses. *Nucleic Acids Res.* **49**, e58 (2021).
172. X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G. P. Nolan, F.-A. Bava, K. Deisseroth, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. **361**, eaat5691 (2018).
173. N. Pfaff, N. Lachmann, M. Ackermann, S. Kohlscheen, C. Brendel, T. Maetzig, H. Niemann, M. N. Antoniou, M. Grez, A. Schambach, T. Cantz, T. Moritz, A ubiquitous chromatin opening element prevents transgene silencing in pluripotent stem cells and their differentiated progeny. *Stem Cells*. **31**, 488–499 (2013).
174. R. Tian, M. A. Gachechiladze, C. H. Ludwig, M. T. Laurie, J. Y. Hong, D. Nathaniel, A. V. Prabhu, M. S. Fernandopulle, R. Patel, M. Abshari, M. E. Ward, M. Kampmann, CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*. **104**, 239-255.e12 (2019).
175. R. Tian, A. Abarientos, J. Hong, S. H. Hashemi, R. Yan, N. Dräger, K. Leng, M. A. Nalls, A. B. Singleton, K. Xu, F. Faghri, M. Kampmann, Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. *Nat. Neurosci.* **24**, 1020–1034 (2021).
176. N. M. Dräger, S. M. Sattler, C. T.-L. Huang, O. M. Teter, K. Leng, S. H. Hashemi, J. Hong, G. Aviles, C. D. Clelland, L. Zhan, J. C. Udeochu, L. Kodama, A. B. Singleton, M. A. Nalls, J. Ichida, M. E. Ward, F. Faghri, L. Gan, M. Kampmann, A CRISPRi/a platform in iPSC-derived microglia uncovers regulators of disease states. *bioRxiv* (2022), p. 2021.06.16.448639, , doi:10.1101/2021.06.16.448639.
177. R. D. Chow, C. D. Guzman, G. Wang, F. Schmidt, M. W. Youngblood, L. Ye, Y. Errami, M. B. Dong, M. A. Martinez, S. Zhang, P. Renauer, K. Bilguvar, M. Gunel, P. A. Sharp, F. Zhang, R. J. Platt, S. Chen, AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. *Nat. Neurosci.* **20**, 1329–1341 (2017).
178. R. T. Manguso, H. W. Pope, M. D. Zimmer, F. D. Brown, K. B. Yates, B. C. Miller, N. B. Collins, K. Bi, M. W. LaFleur, V. R. Juneja, S. A. Weiss, J. Lo, D. E. Fisher, D. Miao, E. Van Allen, D. E. Root, A. H. Sharpe, J. G. Doench, W. N. Haining, In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature*. **547**, 413–418 (2017).
179. X. Jin, S. K. Simmons, A. Guo, A. S. Shetty, M. Ko, L. Nguyen, V. Jokhi, E. Robinson, P. Oyler, N. Curry, G. Deangeli, S. Lodato, J. Z. Levin, A. Regev, F. Zhang, P. Arlotta, In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science*. **370**, eaaz6063 (2020).
180. M. Kuhn, A. J. Santinha, R. J. Platt, Moving from in vitro to in vivo CRISPR screens. *Gene Genome Ed.* **2**, 100008 (2021).
181. M. Dhainaut, S. A. Rose, G. Akturk, A. Wroblewska, S. R. Nielsen, E. S. Park, M. Buckup, V. Roudko, L. Pia, R. Sweeney, J. Le Berichel, C. M. Wilk, A. Bektesevic, B. H. Lee, N. Bhardwaj, A. H. Rahman, A. Baccarini, S. Gnjatic, D. Pe'er, M. Merad, B. D. Brown, Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell*. **185**, 1223-1239.e20 (2022).
182. X. Chen, Y.-C. Sun, H. Zhan, J. M. Kebschull, S. Fischer, K. Matho, Z. J. Huang, J. Gillis, A. M. Zador, High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell*. **179**, 772-786.e19 (2019).
183. R. R. Stickels, E. Murray, P. Kumar, J. Li, J. L. Marshall, D. J. Di Bella, P. Arlotta, E. Z. Macosko, F. Chen, Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
184. S. Vickovic, G. Eraslan, F. Salmén, J. Klughammer, L. Stenbeck, D. Schapiro, T. Äijö, R. Bonneau, L. Bergenstråhle, J. F. Navarro, J. Gould, G. K. Griffin, Å. Borg, M. Ronaghi, J. Frisén, J. Lundeberg, A. Regev, P. L. Ståhl, High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods*. **16**, 987–990 (2019).

185. E. Haapaniemi, S. Botla, J. Persson, B. Schmierer, J. Taipale, CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
186. R. J. Ihry, K. A. Worringer, M. R. Salick, E. Frias, D. Ho, K. Theriault, S. Kommineni, J. Chen, M. Sondey, C. Ye, R. Randhawa, T. Kulkarni, Z. Yang, G. McAllister, C. Russ, J. Reece-Hoyes, W. Forrester, G. R. Hoffman, R. Dolmetsch, A. Kaykas, p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
187. G. Schiroli, A. Conti, S. Ferrari, L. Della Volpe, A. Jacob, L. Albano, S. Beretta, A. Calabria, V. Vavassori, P. Gasparini, E. Salataj, D. Ndiaye-Lobry, C. Brombin, J. Chaumeil, E. Montini, I. Merelli, P. Genovese, L. Naldini, R. Di Micco, Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell*. **24**, 551-565.e8 (2019).
188. O. Ursu, J. T. Neal, E. Shea, P. I. Thakore, L. Jerby-Arnon, L. Nguyen, D. Dionne, C. Diaz, J. Bauman, M. M. Mosaad, C. Fagre, A. Lo, M. McSharry, A. O. Giacomelli, S. H. Ly, O. Rozenblatt-Rosen, W. C. Hahn, A. J. Aguirre, A. H. Berger, A. Regev, J. S. Boehm, Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat. Biotechnol.*, 1–10 (2022).
189. A. H. M. Ng, P. Khoshakhlagh, J. E. Rojo Arias, G. Pasquini, K. Wang, A. Swiersy, S. L. Shipman, E. Appleton, K. Kiaee, R. E. Kohman, A. Vernet, M. Dysart, K. Leeper, W. Saylor, J. Y. Huang, A. Graveline, J. Taipale, D. E. Hill, M. Vidal, J. M. Melero-Martin, V. Busskamp, G. M. Church, A comprehensive library of human transcription factors for cell fate engineering. *Nat. Biotechnol.* **39**, 510–519 (2021).
190. R. Cuella-Martin, S. B. Hayward, X. Fan, X. Chen, J.-W. Huang, A. Taglialatela, G. Leuzzi, J. Zhao, R. Rabadan, C. Lu, Y. Shen, A. Ciccia, Functional interrogation of DNA damage response variants with base editing screens. *Cell*. **184**, 1081-1097.e19 (2021).
191. R. E. Hanna, M. Hegde, C. R. Fagre, P. C. DeWeirdt, A. K. Sangree, Z. Szegletes, A. Griffith, M. N. Feeley, K. R. Sanson, Y. Baidi, L. W. Koblan, D. R. Liu, J. T. Neal, J. G. Doench, Massively parallel assessment of human variants with base editor screens. *Cell*. **184**, 1064-1080.e20 (2021).
192. S. Erwood, T. M. I. Bily, J. Lequyer, J. Yan, N. Gulati, R. A. Brewer, L. Zhou, L. Pelletier, E. A. Ivakine, R. D. Cohn, Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.*, 1–11 (2022).
193. M. Dede, M. McLaughlin, E. Kim, T. Hart, Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol.* **21**, 262 (2020).
194. P. C. DeWeirdt, K. R. Sanson, A. K. Sangree, M. Hegde, R. E. Hanna, M. N. Feeley, A. L. Griffith, T. Teng, S. M. Borys, C. Strand, J. K. Joung, B. P. Kleinstiver, X. Pan, A. Huang, J. G. Doench, Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat. Biotechnol.* **39**, 94–104 (2021).
195. R. A. Gier, K. A. Budinich, N. H. Evitt, Z. Cao, E. S. Freilich, Q. Chen, J. Qi, Y. Lan, R. M. Kohli, J. Shi, High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nat. Commun.* **11**, 3455 (2020).
196. J. P. Shen, D. Zhao, R. Sasik, J. Luebeck, A. Birmingham, A. Bojorquez-Gomez, K. Licon, K. Klepper, D. Pekin, A. N. Beckett, K. S. Sanchez, A. Thomas, C.-C. Kuo, D. Du, A. Roguev, N. E. Lewis, A. N. Chang, J. F. Kreisberg, N. Krogan, L. Qi, T. Ideker, P. Mali, Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods*. **14**, 573–576 (2017).
197. J. A. Vidigal, A. Ventura, Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nat. Commun.* **6**, 8083 (2015).
198. B. Cleary, A. Regev, The necessity and power of random, under-sampled experiments in biology. *ArXiv201212961 Q-Bio Stat* (2020) (available at http://arxiv.org/abs/2012.12961).
199. N. H. Cho, K. C. Cheveralls, A.-D. Brunner, K. Kim, A. C. Michaelis, P. Raghavan, H.

Kobayashi, L. Savy, J. Y. Li, H. Canaj, J. Y. S. Kim, E. M. Stewart, C. Gnann, F. McCarthy, J. P. Cabrera, R. M. Brunetti, B. B. Chhun, G. Dingle, M. Y. Hein, B. Huang, S. B. Mehta, J. S. Weissman, R. Gómez-Sjöberg, D. N. Itzhak, L. A. Royer, M. Mann, M. D. Leonetti, OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science*. **375**, eabi6983 (2022).
200. J. L. Schmid-Burgk, K. Höning, T. S. Ebert, V. Hornung, CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **7**, 12338 (2016).
201. Y. V. Serebrenik, S. E. Sansbury, S. S. Kumar, J. Henao-Mejia, O. Shalem, Efficient and flexible tagging of endogenous genes by homology-independent intron targeting. *Genome Res.* **29**, 1322–1328 (2019).
202. A. Reicher, A. Koren, S. Kubicek, Pooled protein tagging, cellular imaging, and in situ sequencing for monitoring drug action in real time. *Genome Res.* **30**, 1846–1855 (2020).
203. I. Anishchenko, S. J. Pellock, T. M. Chidyausiku, T. A. Ramelot, S. Ovchinnikov, J. Hao, K. Bafna, C. Norn, A. Kang, A. K. Bera, F. DiMaio, L. Carter, C. M. Chow, G. T. Montelione, D. Baker, De novo protein design by deep network hallucination. *Nature*. **600**, 547–552 (2021).
204. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature*. **537**, 320–327 (2016).
205. W. S. Chen, N. Zivanovic, D. van Dijk, G. Wolf, B. Bodenmiller, S. Krishnaswamy, Uncovering axes of variation among single-cell cancer specimens. *Nat. Methods*. **17**, 302–310 (2020).
206. A. Tong, G. Huguet, A. Natik, K. MacDonald, M. Kuchroo, R. Coifman, G. Wolf, S. Krishnaswamy, Diffusion Earth Mover's Distance and Distribution Embeddings. *ArXiv210212833 Cs* (2021) (available at http://arxiv.org/abs/2102.12833).
207. J. Zbontar, L. Jing, I. Misra, Y. LeCun, S. Deny, Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *ArXiv210303230 Cs Q-Bio* (2021) (available at http://arxiv.org/abs/2103.03230).
208. S. Kügler, E. Kilic, M. Bähr, Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Ther.* **10**, 337–347 (2003).
209. P. Y. Ting, A. E. Parker, J. S. Lee, C. Trussell, O. Sharif, F. Luna, G. Federe, S. W. Barnes, J. R. Walker, J. Vance, M.-Y. Gao, H. E. Klock, S. Clarkson, C. Russ, L. J. Miraglia, M. P. Cooke, A. E. Boitano, P. McNamara, J. Lamb, C. Schmedt, J. L. Snead, Guide Swap enables genome-scale pooled CRISPR–Cas9 screening in human primary cells. *Nat. Methods*. **15**, 941–946 (2018).
210. L. Swiech, M. Heidenreich, A. Banerjee, N. Habib, Y. Li, J. Trombetta, M. Sur, F. Zhang, In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106 (2015).
211. F. Heigwer, G. Kerr, M. Boutros, E-CRISP: fast CRISPR target site identification. *Nat. Methods*. **11**, 122–123 (2014).