Published in final edited form as:
Methods. 2012 June ; 57(2): 234–248. doi:10.1016/j.ymeth.2012.03.031.

High-throughput screening for genes that prevent excess DNA replication in human cells and for molecules that inhibit them

Chrissie Y. Lee[a], **Ronald L. Johnson**[b,1], **Jennifer Wichterman-Kouznetsova**[b], **Rajarshi Guha**[b], **Marc Ferrer**[b], **Pinar Tuzmen**[c], **Scott E. Martin**[c], **Wenge Zhu**[a,2,*], and **Melvin L. DePamphilis**[a,*]

[a]National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892-2753, United States

[b]NIH Center for Translational Therapeutics, Probe Development Branch, 9800 Medical Center Drive, Rockville, MD 20892-3370, United States

[c]NIH Center for Translational Therapeutics, Chemical Genomics Branch, 5 Research Court, Room 1A13, Rockville, MD 20850, United States

Abstract

High-throughput screening (HTS) provides a rapid and comprehensive approach to identifying compounds that target specific biological processes as well as genes that are essential to those processes. Here we describe a HTS assay for small molecules that induce either DNA re-replication or endoreduplication (i.e. excess DNA replication) selectively in cells derived from human cancers. Such molecules will be useful not only to investigate cell division and differentiation, but they may provide a novel approach to cancer chemotherapy. Since induction of DNA re-replication results in apoptosis, compounds that selectively induce DNA re-replication in cancer cells without doing so in normal cells could kill cancers *in vivo* without preventing normal cell proliferation. Furthermore, the same HTS assay can be adapted to screen siR-NA molecules to identify genes whose products restrict genome duplication to once per cell division. Some of these genes might regulate the formation of terminally differentiated polyploid cells during normal human development, whereas others will prevent DNA re-replication during each cell division. Based on previous studies, we anticipate that one or more of the latter genes will prove to be essential for proliferation of cancer cells but not for normal cells, since many cancer cells are deficient in mechanisms that maintain genome stability.

Keywords

High throughput screening; DNA re-replication; Endoreduplication; Endocycle; Fluorescence activated cell sorting; siRNA; RNAi; Cancer; Geminin; Emi1

*Corresponding authors: bcmwxz@gwumc.edu (W. Zhu), depamphm@mail.nih.gov (M.L. DePamphilis).
[1]Current address: Division of Cancer Biology, National Cancer Institute, National Institutes of Health, 6130 Executive Blvd, Rockville, MD 20852, United States.
[2]Current address: Department of Biochemistry and Molecular Biology, The George Washington University Medical School, 2300 I Street N.W., Washington, DC 20037, United States.

1. Introduction

1.1. DNA re-replication versus endoreduplication

DNA re-replication is an aberrant form of nuclear DNA replication that occurs in the absence of cell division and produces a mixture of partially and fully completed chromosomes containing stalled replication forks and damaged DNA (Fig. 1) [1]. The result is a population of cells with a heterogeneous DNA content between 4 N (four sets of chromosomes) and 8 N (eight sets of chromosomes), as detected by fluorescence activated cell sorting (FACS) using a flow cytometer (Fig 2B).

In contrast, multiple S phases in the absence of an intervening cytokinesis (termed endocycles) is a developmentally programmed form of excess genome duplication that can occur during terminal cell differentiation. Endocycles are characterized either as endoreduplication, (alternating S and G phases in the absence of M phases [2]), or as endomitosis (alternating S and aborted M phases [3]). In both cases, cells accumulate multiple copies of the nuclear genome under conditions that allow the cell to remain viable but nonproliferative. Endocycles produce a unique FACS profile in which cells accumulate with well-defined groups, each with a DNA content that is an integral multiple of its diploid content of 2 N (Fig. 2D).

1.2. Targeting cancer cells through induction of DNA re-replication

Initiation of DNA replication in human cells begins with assembly of prereplication complexes at replication origins (denoted as a ‘licensing reaction’; Fig. 3) as proliferating cells transit from meta-phase to G1 phase [4–6]. Prereplication complexes include at least 14 proteins: the six-subunit origin recognition complex (ORC), Cdc6, Cdt1, and the replicative DNA helicase Mcm(2–7). Once replication begins (S phase), the Mcm(2–7) helicase travels with the replication fork, leaving behind the helicase loader (ORC·Cdc6·Cdt1 complex) at the replication origin on one sister chromatid, and an empty replication origin on the other sister chromatid (Fig. 3). Eukaryotic cells contain multiple concerted mechanisms that prevent both the remaining helicase loader from recruiting additional Mcm(2–7) helicases and the empty replication origin from recruiting a new helicase loader (Fig. 4).

One of these regulatory pathways involves geminin, a small protein that prevents loading of Mcm(2–7) by binding to Cdt1. Previous studies have shown that siRNA suppression of geminin in human cells induces DNA re-replication (Fig. 1) only in cells derived from cancers [1]. Suppression of geminin in normal cells does not arrest cell proliferation, does not induce either detectable DNA damage or a DNA damage response, and does not induce apoptosis. Normal cells are resistant to suppression of geminin, because they contain at least three cyclin A (CcnA)-dependent pathways that inactivate Cdt1, ORC and Cdc6 (Fig. 4, [6]). If expression of both geminin and CcnA are suppressed, then DNA re-replication is induced in normal cells (Fig. 1), as well as in cancer cells [1]. Of the 23 cell types tested, only those derived from carcinomas, adenocarcinomas, glioblastomas, and osteosarcomas underwent DNA re-replication upon geminin deprivation; cells derived from normal colon, breast, lung, kidney, bone, and skin did not [1]. Moreover, cells derived from normal tissues and later immortalized with a transforming viral protein were also resistant to geminin

depletion, as were cells derived from melanoma and HeLa cells originally derived from a cervical cancer. Thus, drugs that selectively suppress geminin activity should have a high specificity for inducing DNA re-replication in a variety of different cancers with little or no effect on non-cancerous tissues. Since DNA re-replication arrests cell proliferation and induces apoptosis, such drugs would be excellent candidates for cancer chemotherapy.

This conclusion is further supported by the observation that siRNA against 'early mitotic inhibitor 1' (Emi1) also produced a phenotype in cancer cells comparable to that of siRNA against geminin and in normal cells comparable to that of siRNAs against both geminin and CcnA (Fig. 1). Emi1 binds to and inhibits the anaphase promoting complex (APC), an E3 ubiquitin ligase that targets cyclin A, cyclin B and geminin for degradation as cells exit metaphase (Fig. 5). Emi1, like geminin, is expressed from S phase through early M phase, where it prevents premature activation of the APC. Suppression of geminin in cancer cells lacking cyclin A-dependent regulation allows activation of Cdt1, which then, in the presence of functional ORC, Cdc6 and Mcm(2–7) proteins, induces DNA re-replication. Suppression of Emi1, however, allows for activation of the APC. APC activity results in degradation of both geminin and cyclin A, both of which are critical for preventing DNA re-replication in human cells (Fig. 4). Thus, siRNA suppression of Emi1 induces excess DNA replication in both cancer cells and normal cells [7,8], whereas siRNA suppression of geminin can induce DNA re-replication only in cancer cells [1].

1.3. High-throughput screening

High-throughput screening (HTS) allows a researcher to quickly conduct millions of chemical, genetic or pharmacological tests using robotics, data processing and control software, liquid handling devices, and sensitive detectors to execute a complex sequence of steps automatically. HTS assays frequently are designed to detect changes in metabolic activity, specific enzyme levels, cellular localization of a specific molecule, cell proliferation, or apoptosis. The targets for these assays can be purified proteins, cultured cells, or even whole animals such as *Caenorhabditis elegans* [9]. Purified protein targets offer simplicity in experimental design, good reproducibility, and well-defined target specificity. However, cell-based assays have the advantage of excluding molecules that either cannot enter the cell or that have undesired effects on cell proliferation or viability. Their disadvantage is that the identity of the protein(s) targeted generally remains to be determined. The HTS strategy described here minimizes this problem by using the same assay to identify both small molecules that induce excess DNA replication as well as genes that are essential to preventing excess DNA replication. Therefore, small molecules identified in the first HTS should target proteins identified in the second HTS.

The HTS assays described here are intended to elucidate pathways that restrict genome duplication to once per cell division and to identify novel inhibitors of these pathways that should be useful in studies of cell proliferation and differentiation. Our primary goal, however, is to identify therapeutically useful reagents that selectively induce DNA re-replication in cancer cells with little or no effect on the proliferation or viability of normal cells. While siRNA inhibition of geminin can achieve this effect, the clinical utility of RNA interference has not yet been realized [10]. Delivery of siRNA to specific tissues or cells can

be inefficient. In addition, siR-NA rarely eliminates enough of the targeted transcript to elicit a therapeutically useful response, and silencing of non-targeted gene transcripts can produce an unwanted outcome. In an effort to circumvent these problem s, a HTS assay to detect excess DNA replication (EDR) [11] was developed to test chemical libraries for small molecules that mimic the effects of siRNA suppression of geminin. Active compounds from this screen should prove useful in studies of genome duplication and cell division using purified protein, cell-based, and whole animal assays. The HTS assay was also adapted to screen siRNA libraries for other genes whose suppression produces a phenotype similar to that of geminin. Such genes might affect the level of geminin expression, localization to the nucleus, or interaction with other proteins. Such genes might provide additional, complimentary targets to improve the efficiency of geminin inhibitors (i.e. synthetic lethality). For example, siRNA screens have identified protein targets that improve the efficacy of proven chemotherapeutic agents such as paclitaxel [12–14] and poly-ADP ribose polymerase inhibitors [15]. Although the HTS assays described here require complex instrumentation and expertise, researchers funded with NIH grants are eligible for collaborations on small molecule HTS projects (see <http://ncats.nih.gov> for more information). Additionally, the Trans-NIH RNAi initiative is open to intramural NIH researchers (see <http://rnai.nih.gov> for more information). With increased funding, this service will become available to extramural researchers.

2. HTS assay for excess DNA replication

The ’excess DNA replication’ (EDR) assay measures excess DNA replication in the nuclei of cultured cells. This assay was developed first for HTS of small molecules that induce excess DNA replication in cancer cells but not in normal cells [11] and then adapted to screen siRNA libraries for genes that are essential in preventing excess DNA replication in cancer cells, but not in normal cells (Fig. 6).

2.1. Small molecule libraries

To screen libraries of small molecules for compounds that induce excess DNA replication, cells were plated at 250 cells/5μL/well into Aurora 1536-well clear bottom, black low-base plates (Nexus Biosystems) using a Multidrop Combi reagent dispenser (Thermo Scientific). The cells were cultured for 16 h at 37 °C in 5% CO_2 to allow them to adhere to the bottom of the plate before addition of a test compound to each well using an automated pin tool (Kalypsys) [16]. Treated cells were then incubated for 48 h, after which 1 μL Hoechst 33342, a bis-benzimide fluorescent dye that binds to the minor groove of DNA, was added to the cells (1:1000 in PBS) for 40 min to visualize nuclear DNA. Culture medium was removed using an automated plate washer (Kalypsys) and replaced with 6 μL/well of PBS. Nuclei in each well were then imaged and enumerated using an Acumen Explorer eX3 (TTP Lab-Tech) plate reader at 405 nm excitation and 420–500 nm emission. Each compound was tested at seven fivefold dilutions ranging from 3 to 46 μM to generate a dose response curve quantifying precise EC_{50} values (the concentration of an active compound that gives 50% of the maximum response). This strategy, termed ‘quantitative HTS’ (qHTS), and is refractory to variations in sample preparation [16].

2.2. siRNA libraries

HTS of small molecule libraries is effective when several different concentrations of each compound are assayed. However, this is not true for siRNA screens, since the frequency of off-target effects (OTEs) can increase with higher concentrations of siRNA [17]. OTEs are induced primarily by interactions between the siRNA and mRNA transcripts possessing limited sequence homology, particularly in the 3′-untranslated region. The result is an increase in the frequency of false positives, making identification of biologically significant genes difficult. OTEs can be minimized in three ways: by using low siRNA concentration, by chemically modifying siRNA, and by applying bioinformatics to siRNA design. There are two screening strategies for overcoming the problem of OTEs. One is to pool several different siRNAs targeted against the same gene in the hope of reducing OTEs arising from each member of the pool while maintaining on-target efficacy. The other approach is to analyze separately several different siRNAs targeted against the same gene and then select only those genes for which two or more siR-NAs gave a robust phenotype. Off-target genes often contain matches between the seed region of the siRNA (positions 2–7) and sequences in the 3′-UTR of the off-target gene [17,18]. The greater the number of seed matches in the 3′-UTR, the more likely that gene will be targeted. Most vendors take this into consideration during siRNA design, allowing users to minimize off-target effects.

The Ambion ‘*Silencer* Select Human Genome siRNA Library V4’ was used to identify genes that are essential to prevent excess DNA replication in cancer cells. This library contains three unique, non-overlapping siRNAs for each of 21,584 human targets. The siR-NAs were supplied in 384 well master plates from which 0.8 pmol of each siRNA in 2 μL was transferred to the wells of 384 well plates (Corning 3712) using a VPrep liquid handler (Velocity11, Agilent Technologies) integrated into a robotic platform (Biocel System, Agilent Technologies), leaving two columns of 16 wells in each plate for control siRNAs. To establish the baseline activity in this assay, each plate contained one column of wells with Ambion ‘*Silencer* Select negative control siRNA #2’. The sequence of these negative control siRNAs is designed not to target any gene, and they are reported to have minimal effects on gene expression with no detectable effects on cell proliferation, viability, or morphology. To establish the efficacy of this assay, each plate contained one column of wells with a siRNA that induces DNA re-replication. Anti-geminin siRNA was used as a positive control for cancer cells, and anti-Emi1 siRNA was used as a positive control for normal cells.

Transfection reagent (discussed below) was then added to each well in 20 μL of serum free media (McCoy’s 5A) using a WellMate dispenser (Thermo Scientific). Transfection reagent and siRNA were preincubated for 45 min at ambient temperature before adding 750 cells in 20 μL of media containing 20% serum (WellMate, Thermo Scientific). This yielded final transfection mixtures comprising 20 nM siRNA in media containing 10% serum (standard for growth of HCT-116 cells). This concentration was low enough to minimize OTEs. Note that this protocol, termed ‘reverse transfection’, adds the cells into the transfection-siRNA mix rather than adding the transfection-siRNA mix to cultured cells.

The transfected cells were cultured for 72 h, and then fixed by adding 40 μL of 4% paraformaldehyde in PBS and incubating for 45 min at ambient temperature. Cells were then washed twice with 40 μL of PBS and incubated with Hoechst 33342 for 1 h to stain DNA. Cells were washed a final time and left in 40 μL of fresh PBS before imaging. All of the above washing and dispensing steps were conducted with an automated plate washer (Biotek EL406). Nuclei in each well were then imaged using an ImageXpress Micro system (Molecular Devices) plate reader at 377 nm excitation and 447 nm emission.

3. Quantifying the amount of nuclear DNA in each cell

3.1. Small molecule libraries

The EDR assay described above images individual cells to quantify the amount of nuclear DNA in terms of its fluorescence intensity. Any compound that induces EDR is detected, regardless of its mechanism of action. To mitigate the time and processing demands inherent in image-based screens, a laser scanning plate imager was used that offers fast read times and adjustable resolution of data collection [19]. This detector scans the bottom of each well in a microtiter plate in order to enumerate individual fluorescent objects that adhere to the plate surface. Using total fluorescence intensity as a measure of nuclear DNA content, both quiescent and proliferating cells with a DNA content of less than 4 N (<4 N) were distinguished from cells that had excessively replicated their nuclear DNA content to greater than 4 N (>4 N). Individual cells containing a single spherical nucleus were distinguished from cell aggregates and irregularly shaped objects such as mitotic or unresolved closely adjacent nuclei that could be mistaken as >4 N DNA content by the shape of their fluorescence emission [11].

Fluorescent particles were grouped into four populations (Fig. 7): single nuclei with <4 N DNA content, single nuclei with >4 N DNA content, mitotic or closely grouped nuclei that could not be resolved as single nuclei, and objects well outside the size of nuclei, such as dust and fibers. The populations of unresolvable nuclei and physical debris were excluded from the analysis and represented a minority of the total population in general. The percentage of unresolvable nuclei varied depending on the cell type, cell density, and plating conditions but typically it was very low in conditions where cells stopped proliferating but higher in conditions of continued proliferation. By using minimum and maximum intensity thresholds, unresolvable nuclei and debris were excluded from the analysis thereby reducing the scoring of these artifacts as positives.

3.2. siRNA libraries

The ImageXpress that is integrated within the siRNA HTS screening platform is an automated microscope that produces images with greater resolution than a laser scanning microscope. Only cells adhering to the bottom of the well are scored. For HCT116 cells, several different siRNAs against geminin consistently induced ~10-fold higher amounts of EDR than that observed with negative control siRNA, and reduced the number of cells by ~5-fold, consistent with induction of DNA re-replication followed by apoptosis (Fig. 8). Similar results were obtained with siRNA against Emi1 in both HCT116 and MCF10A cells.

4. HTS cell proliferation assay

The HTS EDR assay identifies compounds that induce excess DNA replication, regardless of any effects on cell proliferation or viability. Thus, compounds that induce EDR in cancer cells may still be toxic to normal cells, even though the compound does not induce EDR in normal cells. Therefore, a cell proliferation index was devised in order to determine whether or not a compound inhibited cell proliferation or induced cell death. The cell proliferation index compares the number of nuclei present at the maximum effective concentration of each test compound relative to the number of nuclei in the DMSO treated controls. The equation is "% proliferation = (# $cells^{TEST}$/# $cells^{DMSO}$) – 1] × 100". Thus, compounds that do not affect cell division will have a cell proliferation index of zero. Compounds that prevent cell proliferation without killing the cells will have a proliferation index of about −50, because the number of control cells generally doubles within the two-day period of the assay. Compounds that cause cells either to lyse or to detach from the well bottom will have an index close to −100%. Compounds that induce cells to proliferate faster than control cells will have a positive cell proliferation index.

5. Optimizing HTS EDR assay conditions

5.1. Cell lines

The choice of cell lines for the qHTS EDR assay is based on several criteria. Cells must remain attached to the plate surface while undergoing automated aspiration and washing steps. Cells must have nuclei of uniform size and shape in order to measure their DNA content accurately. Cells must not aggregate, as the nuclei of cell clusters might be imaged as a single object wherein several diploid cells appear as one cell with >4 N DNA content. Ideal are cells that spread out evenly when plated and have a relatively high cytoplasm to nucleus ratio to maximize the distance between nuclei. Finally, the cells must be sensitive to induction of DNA re-replication by podophyllotoxin, in the case of the small molecule screen, and to siRNA against Emi1, in the case of the siRNA screen. In our experience, SW480 human colorectal adenocarcinoma cells, MCF10A non-tumorigenic human breast epithelial cells, and HeLa human cervical cancer cells worked well in the small molecule screen, and HCT116 human colorectal cancer cells and MCF10A worked well in the siRNA screen. Once compounds or genes are identified by HTS, then their phenotype must be confirmed on several other cell lines in order to establish a general conclusion [1].

5.2. Cell proliferation

The rate of cell division influences assay sensitivity, since faster dividing cells were clearly more sensitive to EDR within a limited time period. For example, cholera toxin stimulates the proliferation of MCF10A cells [20], and pilot EDR qHTS indicated that about twice as many active compounds were identified with MCF10A in the presence of cholera toxin than in its absence. The additional positives appeared to be weak inducers of EDR, which otherwise would have appeared as non-active compounds.

EDR cannot occur if the cells cannot enter S-phase. Therefore, plating cells at high density should be avoided in order to prevent cells reaching confluence during the course of the experiment and exiting their cell division cycle. The optimum initial cell density depends on

the volume of the well, the rate of cell proliferation, the length of time the cells are cultured during the screen, and the size of the cells. In general, passaging cells before they are 90% confluent and seeding cells at 30% confluency helps to prevent formation of focal aggregates that are difficult to disperse upon suspension and could result in false positives as a result of cell clumping. Cell aggregates can be removed from cell suspensions prior to plating by gravity filtration through a 40 μm basket filter. Conversely, if the cell density is too low, well-to-well variability increases as a consequence of small sample size. In our experience, a cell density of ~700 cells/well for 384 well plates, and 200–300 cells/well for 1586 well plates provided good results.

5.3. Small molecule screens

The effective concentration of active compounds can vary significantly. Therefore, screening compounds is most effective when done at several different concentrations in order to generate a dose response curve [16]. For our EDR screen of the LOPAC library (Table 1), we assayed compounds at seven different concentrations beginning at 3 nM and ending at 46 μM with incremental five-fold increases [11]. However, as the size of the small molecule library increases, it may become too expensive to assay so many concentrations. In our EDR screen of the MLSMR library (Table 1), six different concentrations were used beginning at 3 nM and ending at 9 μM. This eliminated compounds that were active only at very high concentrations. If a single concentration of the compound is to be assayed in the primary screen, then from 1 to 5 μM is a reasonable choice for cell-based assays, where active compounds are likely to show some efficacy but have less toxicity compared to higher concentrations. A positive control compound, such as podophyllotoxin (Fig. 2), should be selected from the active compounds to serve as a positive control in larger screens.qHTS can require large numbers of cells, but culturing cells continuously can lead to significant variation from one batch to the next. This issue was resolved by harvesting large batches of cells at 60–80% confluence, resuspending them to 2×10^7 cells/mL with 7.5% DMSO in complete culture medium. Cells should be healthy (>90% viability) and proliferating exponentially. Vials of cells are placed upright in a cyropreservation freezing container filled with isopropyl alcohol to insure a slow rate of freezing. Place them at −80 °C for 24 h and then store them at −140 °C. Cells should be thawed rapidly prior to use in a 37 °C water bath. To determine whether the frozen cells were acceptable for HTS, a small amount of frozen cells from each batch was evaluated based on the signal-to-background ratio and Z'-factor (see 'Section 5.6').

The signal-to-background ratio in the EDR assay also depends on the concentration of the test compound and the length of time that cells are exposed to the test compound. The optimal exposure time depends on the rate of cell proliferation and the fraction of cells in S phase during the exposure period. The exposure time must be long enough to allow EDR signals from test compounds to be high enough above the baseline to be reproducible, but short enough that a sufficient number of cells survive and remain attached to the plate. Cell adherence is essential since the laser scanning cytometer scores only those cells attached to the plate surface. For SW480 and MCF10A cells, a two-day incubation resulted in good assay performance [11].

5.4. siRNA screens

As with HTS of small molecules, small-scale pilot siRNA screens should be performed prior to large-scale screening in order to optimize assay conditions. To this end, siRNAs targeting the human kinome can be used, as kinases play key roles in a variety of cellular processes and kinase libraries are commercially available [21]. Pilot screens determine assay reproducibility, estimate the coefficient of variation (see ‘Section 5.6’) for the sample population as a whole, and identify genes involved in the biology under investigation. For example, in two independent pilot screens for genes essential to preventing EDR, seven of eight genes selected in the first pilot screen were selected during the second pilot screen, including many with known roles in DNA replication, thereby confirming assay reproducibility and biological relevance.

One significant difference between the EDR assay of small molecules and the EDR assay of siRNAs was the length of time required for an optimum response. The qHTS EDR assay of small molecules was carried out for two days [11], whereas transfection of cells with siRNA against geminin required three days for an optimal response [1]. Most siRNAs achieve maximal down-regulation of their target transcripts within 24 h, and the down-regulation of transcript levels is typically maintained for about 5 days. Loss of the cognate protein, of course, depends on its half-life, suggesting that a longer period of time would ensure maximal protein depletion. However, like higher concentrations of siRNA, longer times of siRNA exposure may increase nonspecific side effects and stochastic downstream responses. In the case of HTS of siRNAs using the EDR assay on HCT116 cells, 72 h yielded promising results.

Transfection conditions were optimized by testing several transfection reagents with siRNAs targeting essential genes. For HTS formats, reverse transfection (Section 2.2) is favored, because it allows printing and storage of siRNA libraries for long periods of time prior to experiments. Reverse transfection consists of combining siRNAs and transfection reagents in assay plate wells prior to the addition of cells. Since this procedure exposes cells to the siRNA-lipid mixture before they have attached to the bottom of the well, cell viability under these conditions must be confirmed first using standard viability detection reagents. The choice of a transfection reagent can depend on the cell line. Results may vary depending on variables such as the length of time over which transfection occurs and the concentration of the transfection reagent.

Biologically relevant controls should be included in each plate. Ideally, the sensitivity of cells to siRNAs targeting essential genes (a positive control for transfection) is balanced against the resilience of cells to potential toxicity of transfection. Experimental conditions can be validated further and fine-tuned using siRNAs that have been proven to target specific genes together with real-time PCR. One important caveat of siRNA screens is that some siRNAs may not transfect cells as efficiently as the siRNA used as a positive control, thereby resulting in false negatives. Since it is not feasible in HTS to confirm that expression of each gene is effectively reduced by a cognate siRNA, negative results cannot eliminate a gene’s involvement in the biological process under study. Only positive results can be confirmed and pursued.

5.5. Assay limitations

One limitation of the HTS EDR assay is that it does not measure the extent of excess DNA replication per cell, but only the fraction of cells with a DNA content >4 N. Therefore, a small molecule or siRNA that arrested cells in G2 or M phase could appear as a false positive by virtue of the fact that the larger than normal accumulation of cells with 4 N DNA content would include a larger than normal proportion of cells that appear as >4 N DNA content due to excess binding of Hoechst stain. Simply stated, a large G2/M peak in a FACS profile would spill over into the >4 N DNA area defined by the negative control. Thus, compounds or siRNAs that appear to induce excess DNA replication must be confirmed by FACS analysis.

Since the amount of Hoechst dye bound by cells is affected by the ratio of dye to cells, it was also possible that small molecules or siRNAs that cause cell death through mechanisms other than DNA re-replication could appear as false positives in the HTS EDR assay. Fewer nuclei might take up more dye per nuclei under our assay conditions, thereby increasing the fraction of nuclei that appear to contain >4 N DNA content. To determine the significance of this artifact, HCT116 cells were seeded at four different concentrations into 384 well plates. The cells were fixed and stained 24 h later, and the fraction of nuclei with >4 N DNA content was recorded as in a standard HTS. In confluent monolayers, the fraction of nuclei with >4 N DNA was 2%, whereas in wells with 8-fold fewer cells the fraction was 8%. Therefore, even when the cell proliferation index close to −100%, false positives are not likely.

5.6. Metrics

HTS generates massive amounts of data that must be validated and analyzed in order to identify those molecules with the greatest significance. Therefore, quality control metrics are used routinely to evaluate HTS assays [22]. They are employed during small pilot experiments to ensure that the assay is performing as expected and to ensure quality control during the primary screen. However, they are only guides that provide either confidence that the assay is working well, or alert the user that the assay is working poorly. The thresholds associated with any one metric are subjective.

The signal to background ratio (S/B) is the separation between the mean signals for positive and negative controls. For cell-based assays, S/B ratios of three or greater generally perform well. The S/ B for HTS EDR assays of small molecules and siRNAs was 4.5 for SW480 cells, 6 for MCF10A cells, and 6 for HCT116 cells (Fig. 9).

The Z′-factor (not to be confused with ‘z-score’) is a measure of the quality of a screen based on the positive and negative controls only. It incorporates both the S/B and the well-to-well variation among control wells [22]. The Z′-factor measures the number of standard deviations each plate departs from the mean values for both positive and negative controls. Thus, the Z′-factor indicates the probability that a positive result is genuine, and not simply a random event above background. Z′-factors between 0.5 and 1.0 are excellent; 0.5 is equivalent to a separation of 12 standard deviations between the mean values for positive and negative controls. HTS assays with Z′-factors less than 0.5 indicate a fair degree of

'noise' and thus contain less reliable data. The Z′-factors for EDR screens of small molecules and siRNAs was 0.6–0.7 on SW480 cells, MCF10A cells and HCT116 cells (Fig. 9; [11]).

The strictly standardized mean difference (SSMD) is another measure of the quality of the data. It is intended to reduce the number of false positives and false negatives by taking the ratio of the difference of the mean and standard deviations of positive and negative controls [23]. This method was used as an additional quality metric for the siRNA screen (Fig. 9), but not in selection of positive compounds. An SSMD >3 indicates strong controls and >5 very strong controls.

The coefficient of variation (CV) is the ratio of the standard deviation to the mean. The CV is useful for comparing the degree of variation from one data set to another, even if the means differ drastically. It is calculated to determine assay stability and the precision of liquid handling and detection instruments. The CV among different plates was derived from analysis of the positive and negative control siRNAs present on each plate (Fig. 9). CVs can vary from assay to assay, but plates with unusually low or high CVs indicate a problem in the assay, and the assays in these plates must be repeated. For example, if the standard deviation is 20% of the mean levels for multiple runs of positive and negative controls, then a majority of positives are expected to be false.

The coefficient of determination (r^2) is a fraction between 0 and 1 that is derived from the sum of the squares of the distance of the points from the best fit curve determined by a regression method. It is a measure of the quality of the regression model used to fit the points. A perfect fit with no scatter would yield an r^2 of 1.0. More realistically, r^2 values ≥0.9 indicate very good fits to the data. This was the case for the dose response curves of active compounds in the qHTS EDR assay. Furthermore, the potencies of active and inconclusive compounds were reproducible between qHTS runs for both SW480 and MCF10A cells. Comparison of log(EC_{50}) values from two independent screens fitted by linear regression revealed excellent reproducibility in both SW480 ($r^2 = 0.90$) and MCF10A ($r^2 = 0.81$) cells [11].

The median absolute deviation (MAD) is the median of the absolute deviations from the data's median: MAD = 1.4826 × median($|X_{ij} – \text{median}(X)|$), where X indicates all the values in the sample wells of a plate and X_{ij} indicates the sample well at row i and column j [22]. The constant 1.4826 is used to make MAD comparable to standard deviation when the data are distributed normally, thereby providing a robust standard deviation to use for selection of positive compounds. Thus, starting with the deviations from the data's median, the MAD is the median of their absolute values. The advantage of using MAD is that it is resistant to the presence of outliers (data points that lie far from the mean and therefore significantly affect the standard deviation). The positive results selected by the MAD-based method includes all those that would be selected by standard deviation based methods, as well as a significant number of additional positives. The MAD-based selection method also rescues physiologically relevant false negatives that would have been missed using the standard deviation-based method. Analysis of many HTS results reveals that signals greater than

3MAD above the baseline include all of the active molecules [23]. Signals greater than 5MAD contain the most reproducibly active molecules.

The z-score is used to normalize data by setting the mean of the original dataset to 0 and the standard deviation to 1. In the 'robust z-score' version, the mean is replaced by the median, and MAD replaces the standard deviation. The z-score is defined as $[X_i - \text{median}(X)]/\text{MAD}(X)$ where i is the i'th observation; X_i is the normalized signal for the i'th observation; X represents all observations, and MAD is the median absolute deviation. By transforming the data so that the median of the data is 0, the z-score becomes the number of MAD's separating a single result in a data set from the median of the data set (Fig. 10). As a result, siRNA's that lie at least XMAD from the median, can be identified simply by selecting siRNA's whose z-score was $\geq X$.

Heat maps are graphical representations of data where the values in a two-dimensional table are represented as colors. Larger values are represented by small dark gray or black squares (pixels) and smaller values by lighter squares. By creating heat maps of plate data, non-random positional biases can be quickly identified, such as those caused by poor dispenser performance.

6. Identifying active compounds in small molecule screens

After completion of an EDR screen, one or more plate correction methods [22] was applied to adjust for systematic artifacts in screening data, followed by normalization of the data to the positive and negative controls. For the EDR assay, control wells treated with DMSO at the concentration present in the test compounds was defined as 0% activity, and cells treated with podophyllotoxin at its maximum effective concentration (EC_{100}) was defined as 100% activity. The titration-response data were then fit using a custom algorithm [24] to obtain dose response curves for each of the test compounds.

Dose response data on the fraction of cells with >4 N DNA content were sorted into 'curve classes' [16]. The curve class is a heuristic assessment of the quality of a curve such as the presence of upper and lower asymptotes that allows for the identification of compounds scored as active from those that are scored as either inconclusive or non-active (Fig. 11). Based on the curve class and the curve fit parameters, a set of constraints were devised in order to identify a smaller subset of compounds that were most likely active and that selectively induced DNA re-replication in cancer cells. Active compounds were those whose dose response curve exhibited upper and lower asymptotes (classes 1.1 or 2.1) that were well fit ($r^2 \geq 0.9$) and had an efficacy greater than 80%. Compounds that produced well-fit curves (classes 1.2 or 2.2) together with an efficacy >50% were also considered active. Inconclusive compounds produced curves that were either poorly fitted ($r^2 < 0.9$) or active only at the highest concentration tested (class 3). Non-active compounds either did not yield a significant curve fit (class 4) or the efficacy of the curve fit was below threshold.

To confirm the active compounds identified in the initial screen, independent samples of selected compounds were tested in the HTS EDR assay. These confirmation assays typically tested each compound at 15 or more concentrations in order to obtain a better curve fit and a more accurate determination of the effective concentration.

7. Compounds that induce excess DNA replication in human cells

qHTS of the 1280 compounds in the LOPAC identified 15 that were active on SW480 cells and 53 that were inconclusive (Table 1; [11]). The active compounds had EC_{50} values from 0.013 (vincristine) to 16 μM (WIN 62,577) [11]. Of these 68 compounds, 13 had been reported previously to induce EDR in mammalian cells, thereby validating the qHTS results. Podophyllotoxin, as well as eight other compounds that interfere with microtubule dynamics, induced DNA re-replication in both cancer cells and normal cells (Table 2). Similarly, etoposide and two other compounds that inhibit Topoisomerase II also induced DNA re-replication in both cancer cells and normal cells.

SU6656, an inhibitor of Aurora Kinase B and certain Src kinases [25,26], was one of only three LOPAC compounds that selectively induced EDR in cancer cells. Remarkably, SU6656 induced endocycles in cancer cells, not DNA re-replication [11]. This distinction could only be determined from follow-up FACS analysis and apoptosis assays, because the number of cells in each well of the qHTS EDR assay is too small to distinguish reliably between the two manifestations of EDR. Unfortunately, SU6656 was toxic to both cancer cells and normal cells, thereby limiting its therapeutic applications.

In contrast, the aldehyde dehydrogenase inhibitor tetraethylthiuram disulfide (disulfiram, DSF) induced DNA re-replication and apoptosis in cancer cells at a 150-fold lower concentration than it did in normal cells, suggesting that DSF may be useful in cancer chemotherapy [11]. In fact, DSF has been used successfully to treat metastatic melanoma [27] and is currently in clinical trials for various other cancers (ClinicalTrials.gov, ‘Disulfiram’).

The only LOPAC compound that approximated the effects of siRNA against geminin was 3-phenylpropargylamine [11], a mono-amine oxidase inhibitor [28] that arrested cancer cell proliferation without affecting normal cell proliferation. Its effect on DNA replication may be related to the ability of monoamine oxidase inhibitors to inhibit Lysine-specific demethylase 1 (LSD1), the enzyme responsible for the oxidative demethylation of histone H3-Lys9 methylation that causes gene silencing [29,30].

Having validated the qHTS EDR assay in the pilot screen of LOPAC, a qHTS was performed on 343,078 compounds in the molecular libraries small molecule repository (MLSMR). This large scale screen performed reliably, as 9 of 10 LOPAC active compounds present in the MLSMR were recovered as active compounds (Table 1).

Of the 26,156 compounds that were scored as active or inconclusive on SW480 cells, 1191 had EC_{50} values that were ≥10-fold more active on SW480 cells than on MCF10A cells. Of these compounds, 890 either arrested MCF10A cell proliferation or induced apoptosis at high concentrations. Therefore, these active compounds induced DNA re-replication in SW480 cancer cells at concentrations that had adverse affects on MCF10A normal cells. However, 301 of the 1191 compounds had little effect on MCF10A cells, suggesting some specificity for cancer cells.

Of the 1191 selective compounds in the MLSMR that were selectively active on SW480 cells compared with their activity on MCF10A cells, very few exhibited promiscuous activity. Promiscuity was evaluated by examining 1200 different qHTSs performed at the NIH Chemical Genomics Center and counting the number of times a compound was tested in one or more of these assays versus how many times it was scored as active. A score of 0 indicated that the compound was inactive in all of the assays in which it was tested, whereas a score of 10 designated that it was active in all of the assays in which it was tested (i.e. highly promiscuous). A score of 5.5 indicated that it was active in half of the assays in which it was tested. These results demonstrate that, with few exceptions, the 1191 active compounds that were selective for cancer cells do not have a large number of targets (Fig. 12).

As the number of compounds screened increases, so too does the difficulty of identifying the most promising compounds. One approach to this problem is to search for relationships between the chemical or three-dimensional structure of a molecule and its biological activity [31], a process termed structure-activity analysis (SAR). SAR analysis uses a commercial or custom algorithm to cluster active compounds by structural features ranging from single atoms to complex substructures that may be responsible for evoking a particular biological effect. Thus, each cluster contains all of the structural analogs from the screened library that are annotated by their activity, potency, and other parameters used in the HTS. The goal is to identify a strong trend that suggests how subsequent modifications of the compounds of interest might improve its potency and reduce its toxicity. Typically, SAR analysis is done by clustering the active compounds into groups based on their structural similarity by comparing specific individual fragments of the compounds. The results are grouped into a structural series beginning with compounds that appear structurally unique. Then the common structural core of each series is searched against the library to find analogs that were scored as inconclusive or inactive. Such analysis is generally carried out only after the compounds of interest in the primary screen have been retested and confirmed in secondary screens and follow-up assays.

8. Identification of genes that prevent excess DNA replication

Identification of genes that prevent EDR was performed on the complete quality-controlled data set from the siRNA screen. Each sample well in the nearly 200 plates was scored for the fraction of nuclei with >4 N DNA content. As with small molecule HTS, quality control metrics were applied to the raw data from each of the plates (Fig. 9). The results revealed that the plates were clear of artifacts and did not require any plate corrections. The resulting siRNA signals were then analyzed in three ways. Analysis of the raw data from each plate avoids systematic errors that might be introduced by normalizing to the controls included in each plate. Alternatively, the raw data can be normalized to the median value of the negative controls present in each plate in order to correct for plate-to-plate variation. However, the fraction of nuclei with >4 N DNA content in the negative control siRNA treated cells is typically small, which means that tiny variations in the negative control will excessively magnify or diminish the values reported for the test wells. To circumvent this problem, the fraction of >4 N nuclei in each test well of a given plate was normalized to the median values of both the negative and positive siRNA control wells in the same plate in order to

generate a scale of 0–100% > 4 N nuclei, respectively. The results from each plate were normalized to the median plate-to-plate values of the negative control wells (~5% > 4 N DNA) and the positive control wells (~50% > 4 N DNA). In our experience, analysis of the raw data is preferred as long as the Z′-factor is excellent (see 'Section 5.6'). In that case, the raw data and data normalized to both positive and negative controls in its own plate were in excellent agreement.

In addition, cell-based assays used for siRNA screening often produce small dynamic ranges and significant variability due to cellular heterogeneity, transfection efficiency fluctuations, OTEs, and differences among genes in their sensitivity to siRNA. These properties make reliable selection of positive results in RNAi screening difficult. Thus, it is not uncommon to observe a coefficient of variation (see Section 5.6) for control siRNAs of 20–30 [32]. Under these circumstances, it is critical to reduce the number of false–negatives rather than false–positives from the initial screen, because false positives can be eliminated during the validation process whereas false negatives represent missed opportunities. The solution is to calculate the median of the absolute deviations from the data's median, termed the 'Median Absolute Deviation' (MAD).

The positive results obtained from raw data, from data normalized to the median of the negative controls, and from data normalized to both the positive and negative controls were each converted into a robust Z score [33]. For each normalization scheme, siRNAs were identified whose signal was ≥5MAD above the median sample signal across the entire screen (i.e. a robust Z score ≥5, Fig. 10). Genes were selected from the primary screen if at least two of the three siRNAs tested produced sufficient nuclei with >4 N DNA content to be >5MAD above the median.

For example, in the Ambion primary screen normalized to both positive and negative controls in each plate, only 69 genes (0.32% of the genome) had two or more siRNAs whose ability to induce excess DNA replication was ≥5MAD above the median. Both Emi1 and Gmnn were >12MAD above the median. Topoisomerase IIA (Top2a), another gene known to be essential for preventing DNA re-replication in cancer cells, was >10MAD above the median. Therefore, it also was included as a marker gene. The 69 genes ≥5MAD plus 95 genes ≥3MAD from the primary screen were subjected to confirmation analysis using four different siRNAs per gene from Qiagen that did not overlap with the three Ambion siRNAs. The additional genes were selected by 'enrichment analysis' (described below) of the 69 genes ≥5MAD. The results confirmed that at least 80% of the genes selected from the primary screen were essential for preventing EDR in HCT116 cells. Again the average signal for Emi1 and Gmnn siRNAs was >12MAD, and that for Top2a was >10MAD. Combining results from the Ambion primary screen and the Qiagen confirmation screen revealed 89 genes with 2–7 siRNAs ≥5MAD. Emi1 and Gmnn were 2 of only 11 genes whose signal was ≥12MAD, and Top2a was 1 of only 6 genes whose signal was ≥10MAD (Fig. 13B). Moreover, Emi1, Gmnn and Top2 were 3 of only 9 genes in which at least 6 of the 7 siRNAs tested were ≥5MAD, and all of these genes were ≥10MAD (Fig. 13A). These results strongly suggest that the phenotype of excess DNA replication was specific for suppression of gene these genes, and not the result of complex off target effects. Of the genes whose suppression registered ≥10MAD, at least 13 are known to regulate the mitotic

cell cycle, further validating the efficacy of the HTS EDR assay in identifying genes essential for preventing excess DNA replication in cancer cells. The fact that geminin, Emi1 (Fig. 4 and 5) and Top2a were among the top 20 genes essential for preventing EDR in HCT116 cells, and that Emi1 and Top2a also were essential in MCF10A cells (derived from normal human breast tissue) whereas geminin was essential only in the HCT116 cells (derived from a human colorectal cancer) validated the HTS EDR assay as a means for identifying human genes essential for preventing EDR in cancer cells, but not in normal cells.

9. Enrichment analysis

The final challenge is to determine the biological relationship among those genes selected by siRNA HTS. One approach to this problem, termed ‘enrichment analysis’, attempts to identify biochemical networks that are enriched preferentially for the genes selected by siRNA HTS. While this strategy does not exclude false positives, it does increase confidence that the genes selected by siRNA screening are biologically significant, and it identifies other genes that may participate in the same pathway. Enrichment analysis can be done by applying approaches such as the hypergeometric test, which measures the probability that two or more genes identified in the HTS would be related simply by chance. Applying one such method (GeneGo [34]), the 69 genes whose siRNA signals were greater than 5MAD above the median sample signal across the entire screen revealed that they were highly enriched for genes involved directly with regulating cell division, rather than genes involved in cell structure and metabolism that affect cell cycle regulation indirectly (Table 3). This type of analysis also helps to prioritize follow-up genes selected with lower MAD thresholds based on relationships with significantly enriched pathways.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) Roadmap for Medical Research, the NIH Molecular Libraries Program, the intramural research programs of the National Human Genome Research Institute and the National Institute of Child Health and Human Development, a Director’s Challenge grant awarded to M.L.D, and NIH grant 4R00CA136555 awarded to W.Z.

Abbreviations

HTS	high throughput screening
EDR	excess DNA replication
FACS	fluorescence activated cell sorting
siRNA	small interfering RNA
RNAi	RNA interference
OTEs	off-target effects

References

1. Zhu W, Depamphilis ML. Cancer Res. 2009; 69:4870–4877. [PubMed: 19487297]

2. Ullah Z, Lee CY, Lilly MA, DePamphilis ML. Cell Cycle. 2009; 8:1501–1509. [PubMed: 19372757]
3. Ullah Z, Lee CY, Depamphilis ML. Cell Div. 2009; 4:10. [PubMed: 19490616]
4. Depamphilis, ML.; Bell, SD. Genome Duplication. 1. Garland Science; London: 2010.
5. Sivaprasad, U.; Dutta, A.; Bell, SP. Assembly of pre-replication complexes. In: DePamphilis, ML., editor. DNA Replication and Human Disease. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2006. p. 63-88.
6. DePamphilis ML, Blow JJ, Ghosh S, Saha T, Noguchi K, Vassilev A. Curr Opin Cell Biol. 2006
7. Di Fiore B, Pines J. J Cell Biol. 2007; 177:425–437. [PubMed: 17485488]
8. Machida YJ, Dutta A. Genes Dev. 2007; 21:184–194. [PubMed: 17234884]
9. Giacomotto J, Segalat L. Br J Pharmacol. 2010; 160:204–216. [PubMed: 20423335]
10. Davidson BL, McCray PB Jr. Nat Rev Genet. 2011; 12:329–340. [PubMed: 21499294]
11. Zhu W, Lee CY, Johnson RL, Wichterman J, Huang R, DePamphilis ML. Mol Cancer Res. 2011; 9:294–310. [PubMed: 21257818]
12. Bartz SR, Zhang Z, Burchard J, Imakura M, Martin M, Palmieri A, Needham R, Guo J, Gordon M, Chung N, Warrener P, Jackson AL, Carleton M, Oatley M, Locco L, Santini F, Smith T, Kunapuli P, Ferrer M, Strulovici B, Friend SH, Linsley PS. Mol Cell Biol. 2006; 26:9377–9386. [PubMed: 17000754]
13. Swanton C, Marani M, Pardo O, Warne PH, Kelly G, Sahai E, Elustondo F, Chang J, Temple J, Ahmed AA, Brenton JD, Downward J, Nicke B. Cancer Cell. 2007; 11:498–512. [PubMed: 17560332]
14. Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, Minna JD, Michnoff C, Hao W, Roth MG, Xie XJ, White MA. Nature. 2007; 446:815–819. [PubMed: 17429401]
15. Turner NC, Lord CJ, Iorns E, Brough R, Swift S, Elliott R, Rayter S, Tutt AN, Ashworth A. EMBO J. 2008; 27:1368–1377. [PubMed: 18388863]
16. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP. Proc Natl Acad Sci USA. 2006; 103:11473–11478. [PubMed: 16864780]
17. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y, Baskerville S, Maksimova E, Robinson K, Karpilow J, Marshall WS, Khvorova A. Nat Methods. 2006; 3:199–204. [PubMed: 16489337]
18. Lin X, Ruan X, Anderson MG, McDowell JA, Kroeger PE, Fesik SW, Shen Y. Nucleic Acids Res. 2005; 33:4527–4535. [PubMed: 16091630]
19. Bowen WP, Wylie PG. Assay Drug Dev Technol. 2006; 4:209–221. [PubMed: 16712425]
20. Soule HD, Maloney TM, Wolman SR, Peterson WD Jr, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF, Brooks SC. Cancer Res. 1990; 50:6075–6086. [PubMed: 1975513]
21. Zhang JD, Koerner C, Bechtel S, Bender C, Keklikoglou I, Schmidt C, Irsigler A, Ernst U, Sahin O, Wiemann S, Tschulena U. PLoS ONE. 2011; 6:e22176. [PubMed: 21765947]
22. Kevorkov D, Makarenkov V. J Biomol Screen. 2005; 10:557–567. [PubMed: 16103415]
23. Chung N, Zhang XD, Kreamer A, Locco L, Kuan PF, Bartz S, Linsley PS, Ferrer M, Strulovici B. J Biomol Screen. 2008; 13:149–158. [PubMed: 18216396]
24. Wang Y, Jadhav A, Southal N, Huang R, Nguyen DT. Curr Chem Genomics. 2010; 4:57–66. [PubMed: 21331310]
25. Blake RA, Broome MA, Liu X, Wu J, Gishizky M, Sun L, Courtneidge SA. Mol Cell Biol. 2000; 20:9018–9027. [PubMed: 11074000]
26. Bain J, Plater L, Elliott M, Shpiro N, Hastie CJ, McLauchlan H, Klevernic I, Arthur JS, Alessi DR, Cohen P. Biochem J. 2007; 408:297–315. [PubMed: 17850214]
27. Brar SS, Grigg C, Wilson KS, Holder WD Jr, Dreau D, Austin C, Foster M, Ghio AJ, Whorton AR, Stowell GW, Whittall LB, Whittle RR, White DP, Kennedy TP. Mol Cancer Ther. 2004; 3:1049–1060. [PubMed: 15367699]
28. Alemany R, Olmos G, Garcia-Sevilla JA. Br J Pharmacol. 1995; 114:837–845. [PubMed: 7773544]

29. Liang Y, Vogel JL, Narayanan A, Peng H, Kristie TM. Nat Med. 2009; 15:1312–1317. [PubMed: 19855399]

30. Binda C, Valente S, Romanenghi M, Pilotto S, Cirilli R, Karytinos A, Ciossani G, Botrugno OA, Forneris F, Tardugno M, Edmondson DE, Minucci S, Mattevi A, Mai A. J Am Chem Soc. 2010; 132:6827–6833. [PubMed: 20415477]

31. Martin YC, Kofron JL, Traphagen LM. J Med Chem. 2002; 45:4350–4358. [PubMed: 12213076]

32. Stone DJ, Marine S, Majercak J, Ray WJ, Espeseth A, Simon A, Ferrer M. Cell Cycle. 2007; 6:898–901. [PubMed: 17438372]

33. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Nat Biotechnol. 2006; 24:167–175. [PubMed: 16465162]

34. Consortium GO. Nucleic Acids Res. 2008; 36:D440–D444. [PubMed: 17984083]

35. Seethala, R.; Zhang, L. Handbook of Drug Screening. 2. Informa Healthcare; New York: 2009.

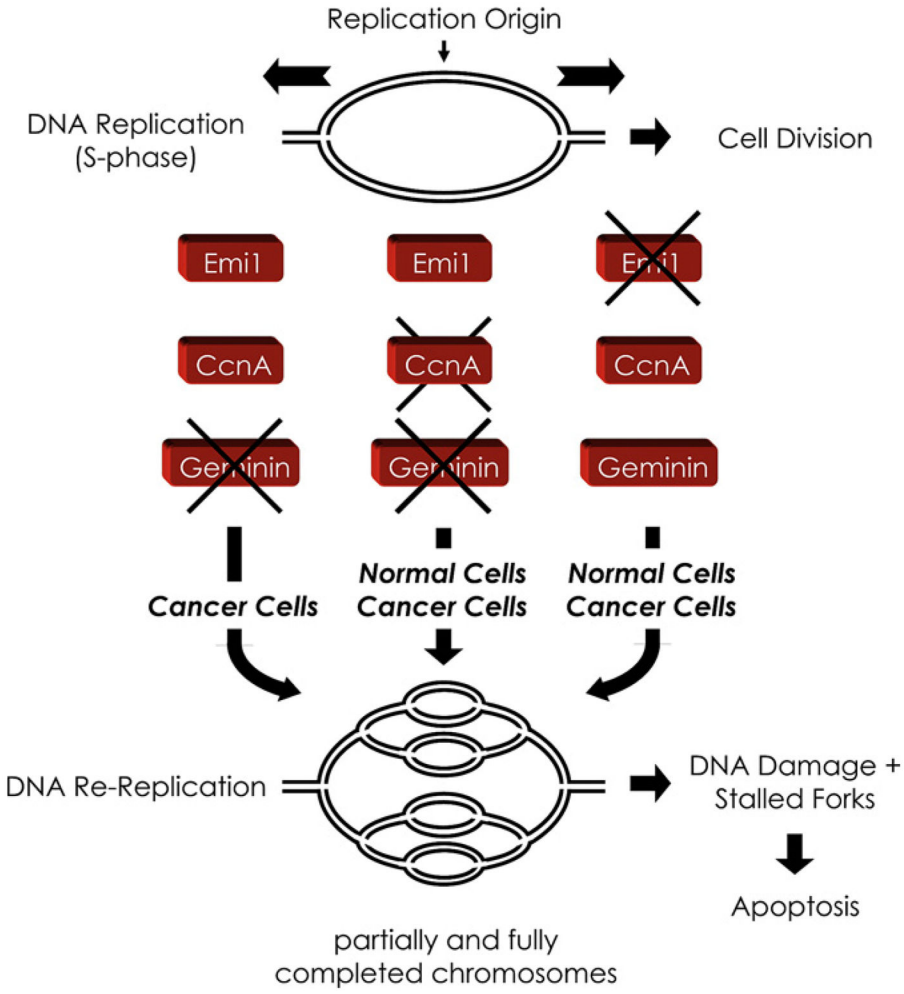

Fig. 1.

Induction of DNA re-replication. In cancer cells, siRNA suppression of geminin induces a second round of nuclear DNA replication prior to cell division that produces a mixture of partially and fully replicated chromosomes. This phenomenon is termed ‘DNA re-replication’ and results in apoptosis instead of cell division. In normal cells, suppression of geminin alone is insufficient to induce DNA re-replication, but suppression of both geminin and cyclin A2 (CcnA) does induce DNA re-replication. Thus, normal cells employ both geminin and at least one Cdk2·CcnA phosphorylation event (Fig. 3) to prevent initiation of DNA replication during S, G2 and early M phases, whereas cancer cells rely solely on geminin. Suppression of Emi1 alone can induce DNA re-replication in both normal cells and cancer cells, because Emi1 prevents premature activation of the ubiquitin ligase (anaphase promoting complex) that targets both geminin and cyclin A for destruction during M phase (Fig. 5).

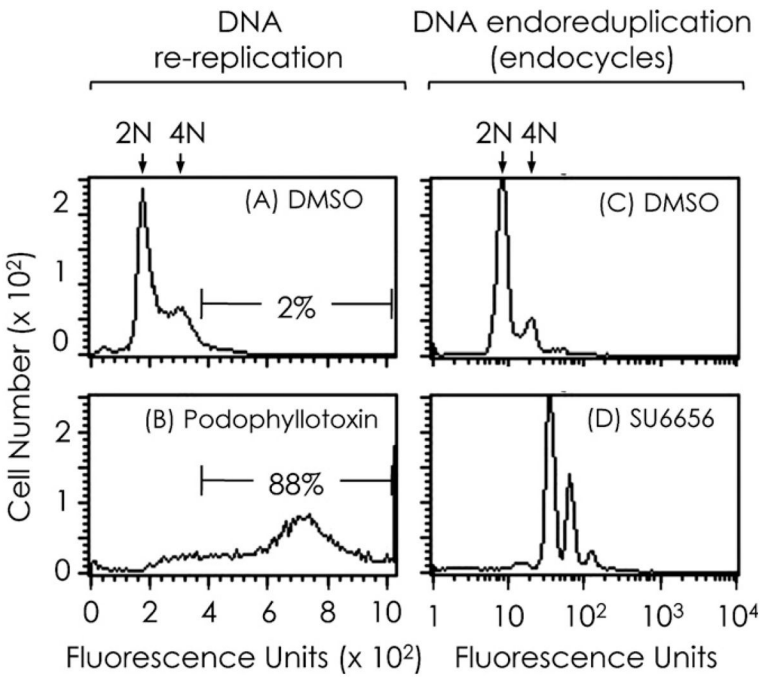

Fig. 2.

Induction of DNA re-replication and endocycles in cancer cells. Proliferating SW480 cells were treated for 2 days with 46 μM DMSO (panels A, C), or with 3 μM podophyllotoxin dissolved in DMSO (panel B), or with 6 μM SU6656 dissolved in DMSO (panel D). Their DNA content was then determined by FACS analysis and displayed on either a linear (A, B) or exponential (C, D) scale. Podophylotoxin induced DNA re-replication characterized by the accumulation of cells between 4 and 8 N DNA content. SU6656 induced endocycles characterized by the accumulation of cells with 8, 16 and 32 N DNA content. Reproduced from [11].

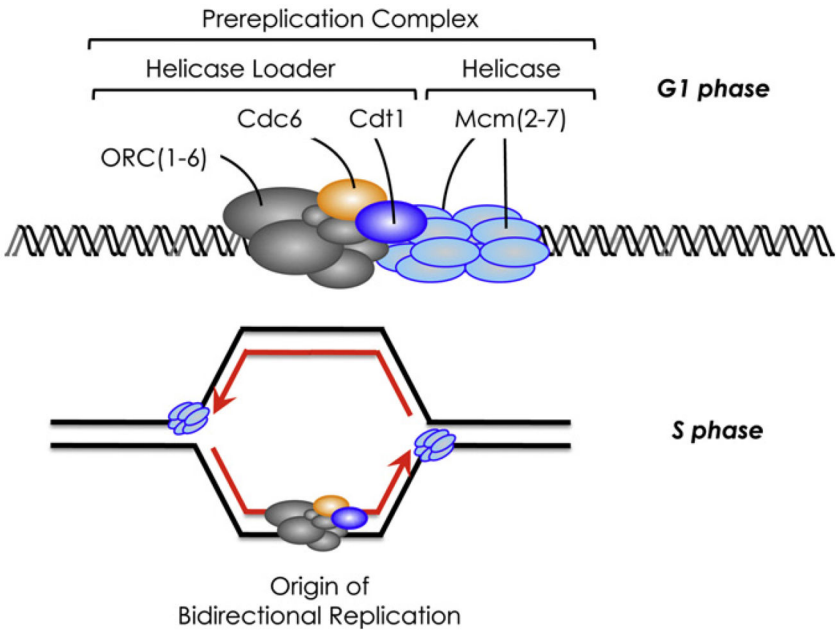

Fig. 3.

Prereplication complexes. During the transition from anaphase to G1 phase of the mammalian cell division cycle, the DNA helicase loader [ORC(1–6),·Cdc6, and Cdt1] loads two copies of the replicative DNA helicase [Mcm(2–7)] onto DNA replication origins to form a prereplication complex. Once S phase begins, the two helicases unwind DNA in opposite direction, each in concert with a DNA replication machine. The helicase loader remains at the replication origin of one of the two sister chromatids.

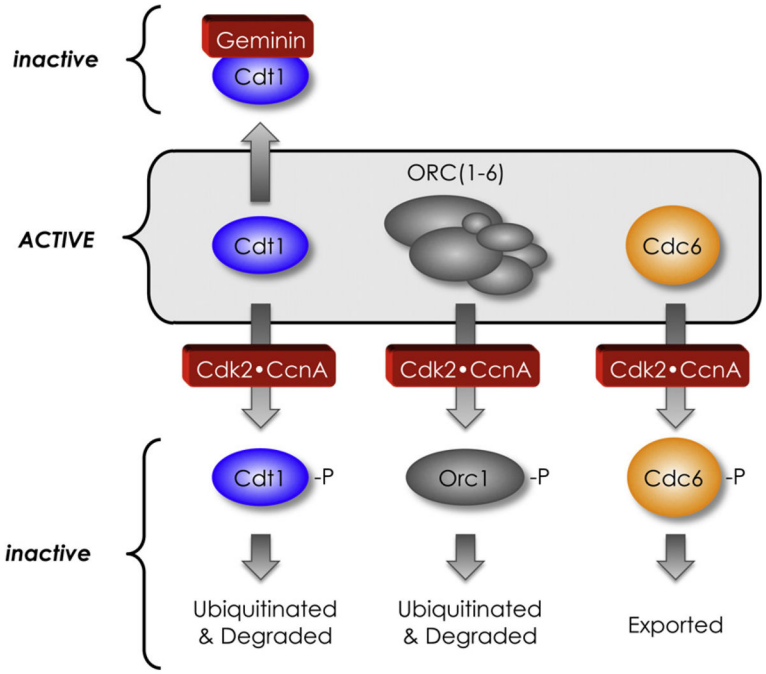

Fig. 4.

Inactivation of prereplication complexes. Once DNA replication begins, proteins essential for assembly and activity of prereplication complexes are inactivated to prevent a second round of DNA replication from occurring within a single cell division cycle. Binding to the protein geminin inactivates Cdt1. Cdk2·CcnA inhibits Cdt1, Orc1, and Cdc6 activities by phosphorylating these proteins. Cdt1-P and Orc1-P are ubiquitinated by CRL1·Skp2 and thereby targeted for degradation by the 26S proteasome. Cdc6-P is exported from the nucleus. Cdt1-P and Orc1-P are ubiquitinated by CRL1·Skp2. Non-phosphorylated Cdt1·PCNA·DNA complexes that form during S-phase are ubiquitinated by CRL4·Cdt2 (not shown).

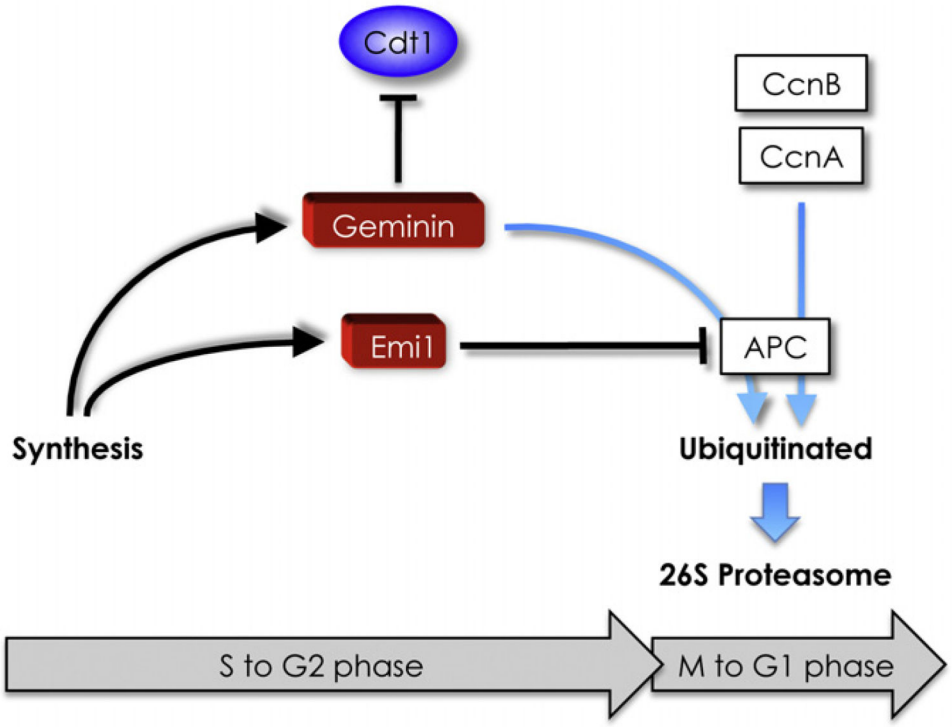

Fig. 5.

Geminin and Emi1 validate the siRNA HTS EDR assay. Geminin and Emi1 are present during S, G_2 and early M phases. Geminin inhibits Cdt1, a protein essential for assembly of prereplication complexes (Fig. 1). Emi1 inhibits the anaphase promoting complex (APC), a ubiquitin ligase that targets geminin, cyclin A (CcnA), and cyclin B (CcnB) for degradation. Thus, Emi1 prevents ubiquitin-dependent degradation of geminin, CcnA and CcnB. When Emi1 is degraded during mitosis, both geminin and CcnA are ubiquitinated by the APC and then degraded. Thus, siRNA suppression of geminin induces DNA re-replication in cells that rely solely on geminin to prevent prereplication complex assembly (e.g. cancer cells), whereas siRNA suppression of Emi1 induces DNA re-replication in cells that rely on either geminin or Cdk2·CcnA to prevent assembly of prereplication complexes (normal cells as well as cancer cells).

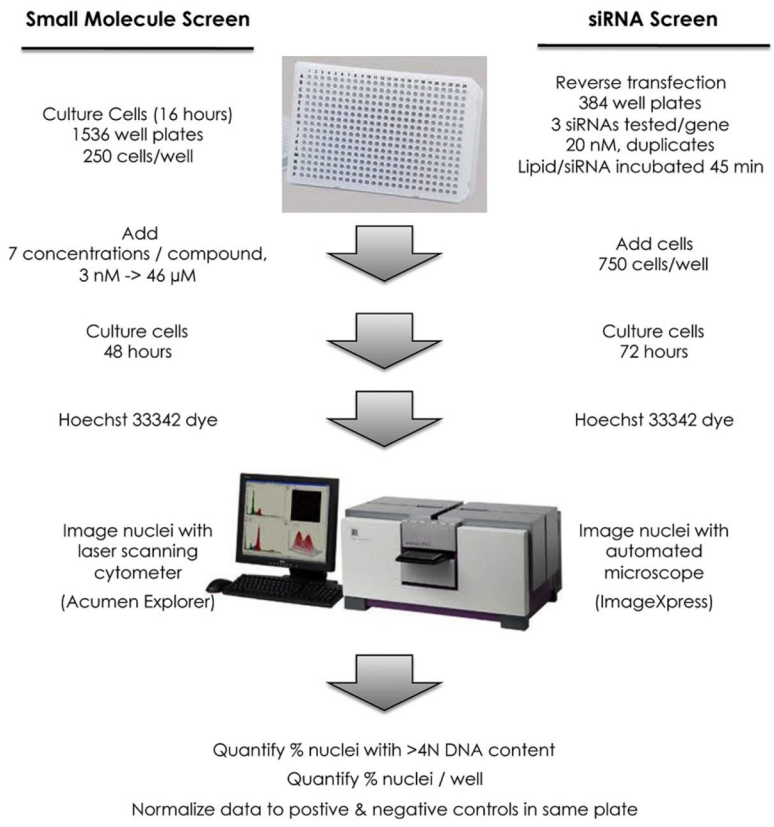

Fig. 6.

HTS of small molecules and siRNAs. The EDR assay can be used in HTS of either small molecules or siRNAs. The small molecule screen was performed with 1536-well plates and seeded with 250 cells/well. Each compound was added in a spectrum of concentrations from 3 to 46 µM and cultured for 48 h. For the siRNA screen, siRNA was incubated with lipid in 384 well plates for 45 min, and then 750 cells were added to each well and transfection was allowed to proceed for 72 h. In both assays, nuclear DNA was stained with Hoechst and the DNA content of each cell in each well was quantified using either a laser scanning cytometer (shown) or an automated microscope. The fraction of cells (i.e. nuclei) with >4 N DNA content was then quantified and analyzed as described [11].

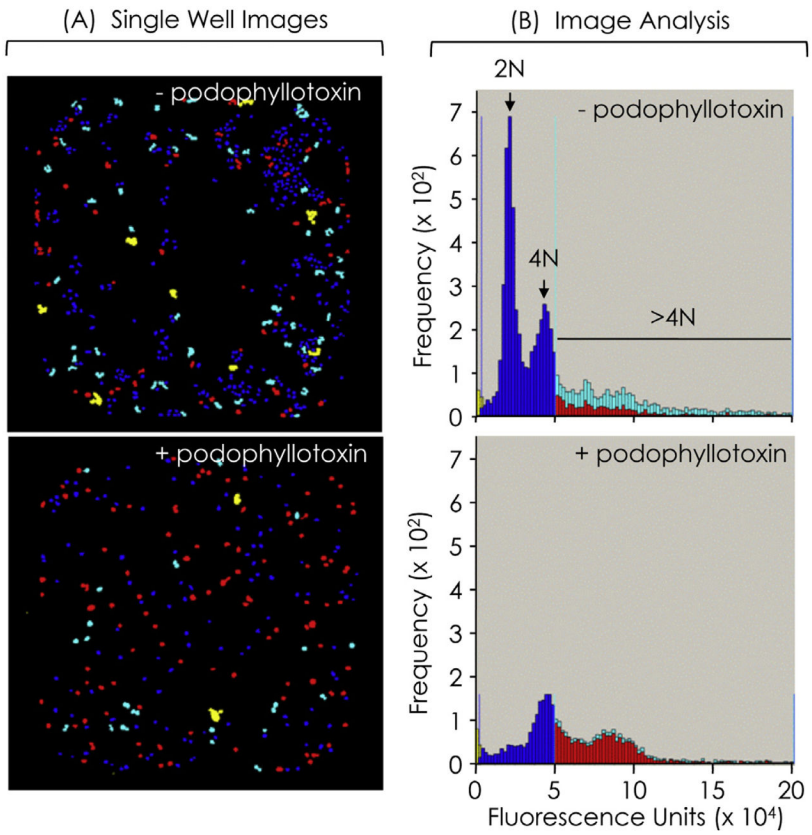

Fig. 7.

Detecting excess DNA replication in human cells. SW480 cells were incubated for 48 h at 37 °C and 5% CO_2 in medium containing 0.4% (65 mM) DMSO alone or with 3 μM podophyllotoxin. After addition of Hoescht and incubation for 1 h at room temperature, the cells were imaged by a laser scanning cytometer. Scans of a single well from a 1536-well plate are shown for DMSO minus or plus podophyllotoxin-treated control wells. Fluorescent objects were classified and colored as follows: nuclei with ≤4 N DNA content [dark blue, 3000–50,000 fluorescence units (FLU), 40–100% Gaussian shape)], nuclei with >4 N DNA content (red, 50,000–200,000 FLU, 40–100% Gaussian shape), unrecognizable as single nuclei (cyan, 3000–200,000 FLU,<40% Gaussian shape), and excluded fluorescent objects (yellow,<3000 or >200,000 FLU). Gaussian shape refers to the fit of the intensity profile to an ideal sphere (100%). Histograms using these color classifications were constructed from 16 wells each treated with DMSO alone (−podophyllotoxin) or with 3 μM podophyllotoxin (+podophyllotoxin) to reveal the frequency of each type of object. The results are analogous to a FACS profile in which the relative numbers of cells in G_1, S, and G_2/M phases of the cell cycle are determined. Reproduced from [11].

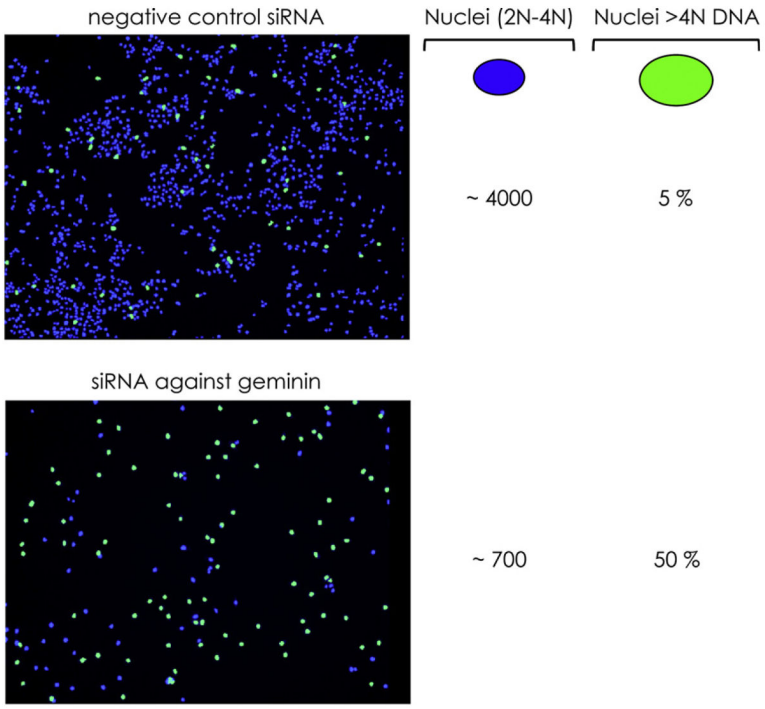

Fig. 8.

siRNA against geminin in the HTS EDR assay. HCT116 cells were reverse transfected with either 20 nM negative control siRNA or 20 nM siRNA against geminin for 72 h. DNA was stained with Hoechst (blue) and imaged using ImageXpress. Cells with >4 N DNA content are false-colored in green.

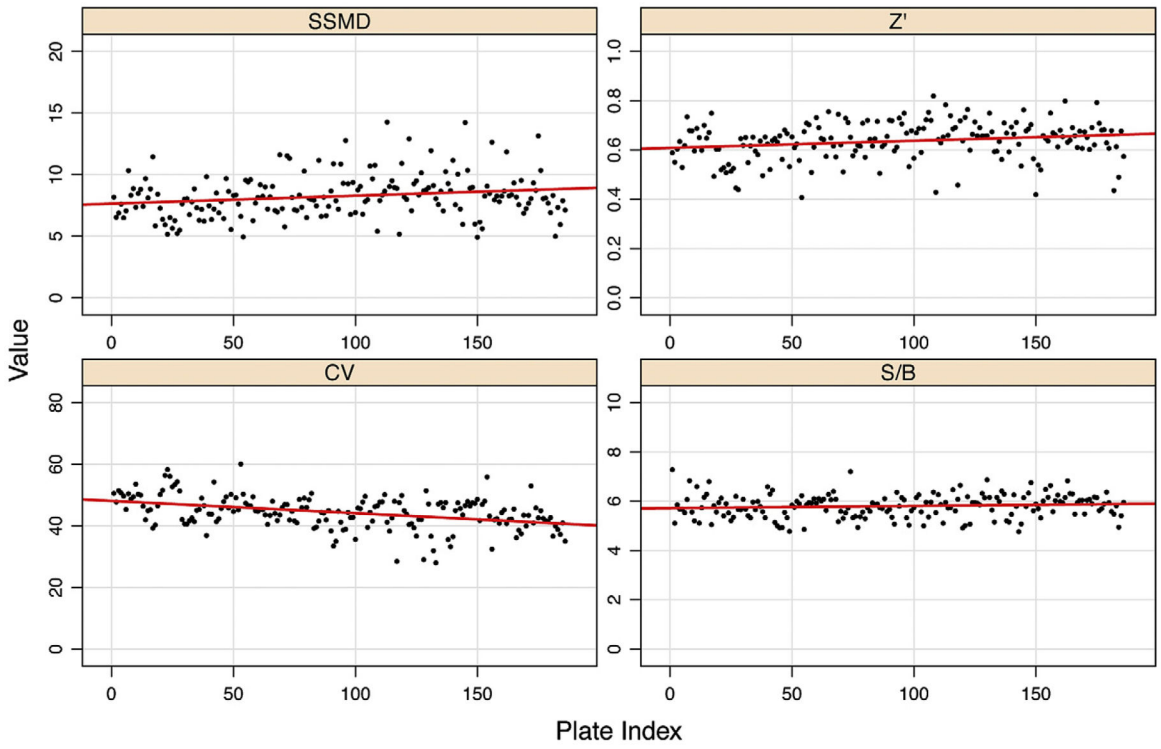

Fig. 9.

Examples of quality control metrics. SSMD, Z′-factor, CV, and S/B metrics were applied to the raw data from each plate in the siRNA HTS EDR assay of the Ambion human genome library. The red lines are linear regression plots to indicate the trend. The variation of each metric from plate to plate is indicated by the range of scatter in the data for the CV (28–60), S/B (5–7), SSMD (5–14), and Z′-factor (0.4–0.8).

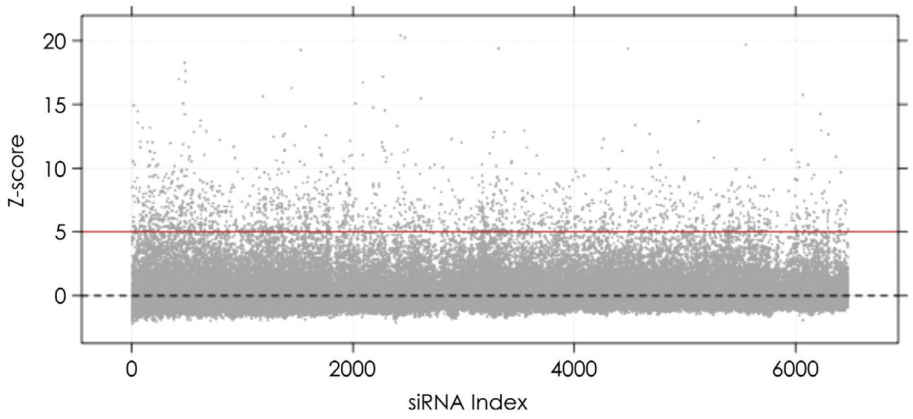

Fig. 10.

Example of robust Z scores. Data for the Ambion human genome library were normalized to their negative siRNA control. The dashed line represents the median signal for the whole screen. The red line indicates 5MAD above the median. Signals <3MAD above the median were considered negative results. Signals >3MAD but <5MAD were considered inconclusive results. Signals >5MAD were considered positive results.

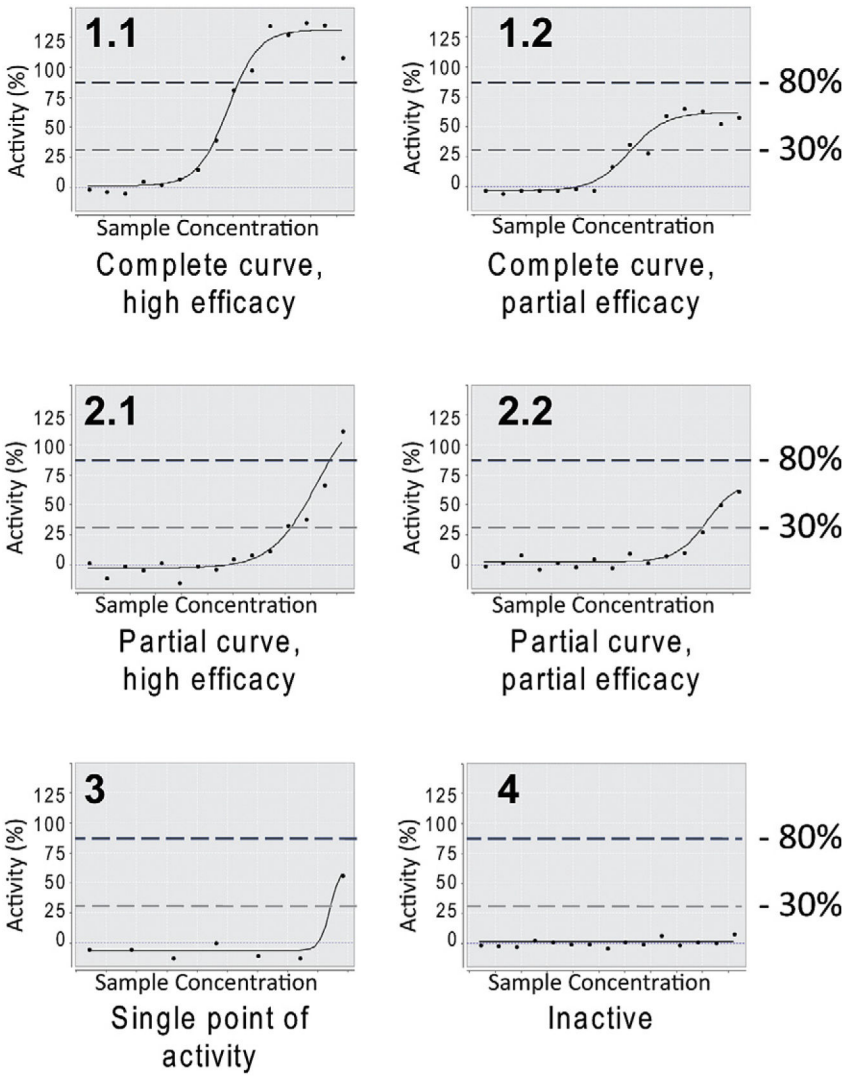

Fig. 11.

Curve Classes. Quantitative HTS relies on measuring the activity of each compound at a broad range of concentrations. Dose response curves can be subdivided into six classes based on curve fit and response efficacy. Compounds with curve classes 1.1, 1.2, 2.1, and 2.2 are considered active. Compounds with curve class three are considered inconclusive, and those with curve class four are nonactive. Compounds that do not fall into these six classes should be retested. Taken from [35].

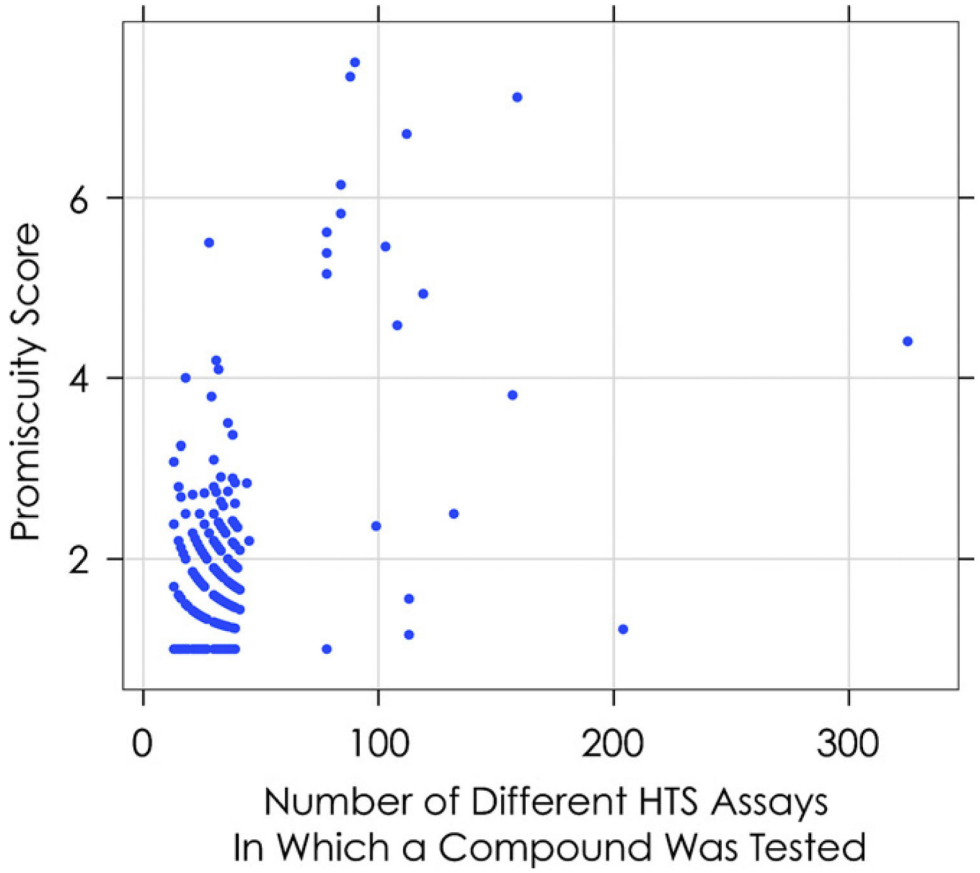

Fig. 12.

Promiscuity analysis. The activity of each of the 1191 compounds that selectively induced EDR in SW480 cells in the primary qHTS was recorded for each of the different HTS assays carried out by the NIH Chemical Genomics Center. A score of 0 indicates no activity in any of the assays, whereas a score of 10 indicates activity in all of the assays in which this compound was tested.

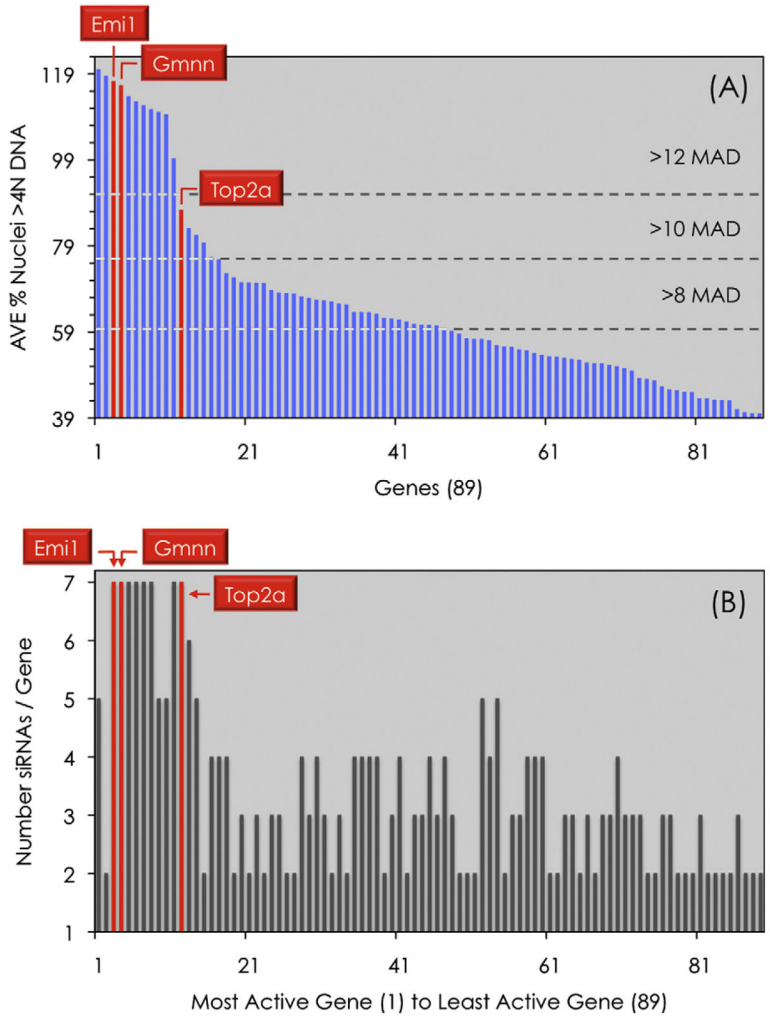

Fig. 13.

Identifying genes that prevent excess DNA replication in cancer cells. HCT116 cells were reverse transfected with 20 nM each of siRNAs representing 21,584 human genes. The data were normalized to both negative control siRNA and to the positive control siRNA against geminin. Of these, 69 genes had at least two siRNAs with a signal ≥5MAD above the median. These genes, together with an additional 84 genes whose signal was ≥3MAD above the median, but for which GeneGo identified relationships to one or more of the 69 genes selected from the primary screen, were again subjected to the same HTS assay, but using four independent siRNAs from Qiagen. Of these 153 genes, 89 had at least two siRNAs with a signal ≥5MAD above the median (panel A). The results were decidedly biphasic with Emi1, Gmnn and Top2a among the genes with the highest HTS signal and the maximum number of siRNAs per gene. The average fraction of nuclei with >4 N DNA content for each gene in panel A was determined from the 2–7 siRNAs that were ≥5MAD for each gene in panel B.

Table 1

Test compounds that induce excess DNA replication in SW480 cells.

Compounds	LOPAC[a]		MLSMR[b]	
	Number	%	Primary	%
Screened	1280	100	343,078	100
Active	15	1.2	5649	1.6
Inconclusive	53	4.1	18,938	5.5
Inconclusive, but high efficacy in SW480 cells			1569	0.3
Previously identified	13	1.0	9[c]	
[d] SW480 cell selective	2	0.16	890	0.26
[e] SW480 cell specific	1	0.08	301	0.09

[a] Library of pharmacologically active compounds.

[b] Molecular libraries small molecule repository.

[c] Ten of the 15 active compounds in LOPAC were also in the MLSMR, nine of which were active in the primary screen.

[d] EC_{50} values on SW480 were ≥10-fold less than on MCF10A cells (e.g. tetraethylthiuram disulfide, Table 2).

[e] EC_{50} values on SW480 were ≥10-fold less than on MCF10A cells, and their cell proliferation indices indicated no effect on proliferation or apoptosis of MCF10A cells (e.g. 3-phenylpropargylamine, Table 2).

Table 2

Selected compounds that induce excess DNA replication in human cells.

Compound	Activity		EC_{50} (μM)		Proliferation index	
	SW480	MCF10A	SW480	MCF10A	SW480	MCF10A
Podophyllotoxin	A	A	0.028	0.032	−48	−70
Etoposide	A	A	0.32	1.1	−45	−69
Tetraethylthiuram disulfide	A	I	0.20	22	−42	−73
3-Phenylpropargylamine	I	N	6.3	–	−50	+3
SU6656	I	I	4.5	3.2	−44	−61

A = active, I = inconclusive, N = nonactive, EC_{50} = concentration that gives 50% of maximum response.

Table 3

Top 10 categories for genes selected by siRNA HTS EDR assay.

Biological processes	**p-Value**	**Cell cycle pathways**	**p-Value**
Organelle fission	9.0E-14	Metaphase checkpoint	2.5E-10
M phase	1.5E-10	Spindle assembly chromosome separation	8.3E-09
Translational elongation	2.1E-10	Prometaphase	3.8E-08
Mitotic cell cycle	9.0E-10	Regulation by SCF complex	2.1E-07
Chromosome segregation	5.0E-09	Regulation by APC	3.6E-07
Mitosis	1.5E-07	Termination of DNA replication	3.1E-04
Cell division	1.8E-06	Nek kinase family	4.6E-04
Cytokinesis	2.5E-06	Ubiquitin-dependent Proteolysis	5.3E-03
Spindle attachment to kinetochore	1.1E-05	Initiation of mitosis	6.3E-03
Microtubule anchoring	1.7E-05	Cytoskeleton remodeling &integrins in cell motility	9.5E-03
		Metabolic Pathways	
		Lyso-phosphatidylserine	9.4E-03
		Phosphatidylinositol-4,5-diphosphate	1.4E-02
		1-acyl-glycerol_3-phosphoethanolamine	1.6E-02
		N-acyl-sphingosine phosphate pathway	1.8E-02
		Ceramide	1.4E-01
		(S)-citrulline	1.5E-01
		(L)-lysine pathways and transport	1.5E-01
		Lipid metabolism glycosphingolipid metabolism	1.7E-01
		1-linoleoyl-glycerol_3-phosphate	1.6E-01
		O-hexanoyl-(L)-carnitine	1.6E-01
		Lyso-phosphatidylserine	9.4E-03
		Phosphatidylinositol-4,5-diphosphate	1.4E-02
		1-acyl-glycerol_3-phosphoethanolamine	1.7E-02
		N-acyl-sphingosine phosphate	1.8E-02

p-Value is the probability of obtaining this result by random sampling from identical populations.