References

Darzynkiewicz, Z., Juan, G., Li, X., Gorczyca, W., Murakami, T., and Traganos, F. (1997). Cytometry in cell necrobiology: Analysis of apoptosis and accidental cell death (necrosis). *Cytometry.* **27,** 1–20.

Lee, S., and Howell, B. J. (2006). High-content screening: Emerging hardware and software technologies. *Methods Enzymol.* **414**(this volume).

Mitchison, T. J. (2005). Small-molecule screening and profiling by using automated microscopy. *Chembiochem.* **6,** 33–39.

Prigent, C., and Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? *J. Cell Sci.* **116,** 3677–3685.

Vermes, I., Haanen, C., Steffens-Nakken, H., and Reutellingsperger, C. (1995). A novel assay for apoptosis flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled annexin V. *J. Immunol. Methods.* **184,** 39–51.

Villa, P., Kaufmann, S. H., and Earnshaw, W. C. (1997). Caspases and caspase inhibitors. *Trends Biochem. Sci.* **22,** 388–393.

[17] High-Throughput Screening for Modulators of Stem Cell Differentiation

By PAUL J. BUSHWAY and MARK MERCOLA

Abstract

Realizing the potential of stem cell biology requires the modulation of self-renewal and differentiation, both of which are incompletely understood. This chapter describes methods for the design, development, and implementation of cell-based screens of small molecules, genes and expressed proteins for modulation of stem and progenitor cell fate. These include the engineering of embryonic and other stem cells with gene promoter–reporter protein constructs and their application in automated screening. We discuss considerations of promoter reporter selection, assay development and implementation, and image acquisition, analysis, and data handling. Such black-box screens are useful for the identification of probes of developmental processes and should provide tools that will identify druggable targets for biochemical assays.

Introduction

Stem cells combine the potential for self-renewal with the ability to give rise to one or more differentiated cell types, and the ability to manipulate these cells offers unprecedented opportunities to increase our understanding of developmental biology and to provide a renewable source of cells to

METHODS IN ENZYMOLOGY, VOL. 414 0076-6879/06 $35.00

DOI: 10.1016/S0076-6879(06)14017-3

address crippling degenerative diseases by replacing lost or damaged tissue. Recent advances have identified stem cells in nearly all tissue compartments of the body, and, increasingly, it is becoming possible to isolate the stem cells, maintain them in culture, and, in certain instances, expand their numbers. Despite advances in basic stem cell biology, it remains quite challenging to stimulate them to differentiate efficiently into the cell types needed for clinical application, either in culture or after transplantation into damaged tissue. The reasons for inefficient differentiation include often extreme difficulty in obtaining large numbers of appropriate types of stem cells, poor survival in culture, and, upon transplantation, incomplete understanding of the natural molecules that promote differentiation. To promote efficient differentiation requires specific knowledge of the molecules and signaling pathways that are relevant to the given type of stem cell. Moreover, in those instances in which multiple cell types develop from a single class of stem cell, the molecules and signaling pathways must also be matched to the lineage of the particular differentiated cell type that is desired. Thus, before the biotechnological and medical potential of stem cells can be realized, it will be essential to have in hand molecules that can influence their survival, renewal, and differentiation.

The incomplete knowledge of natural modulators has prompted the development of cell-based assays to screen chemical and genome libraries for small molecules, genes, and proteins capable of controlling stem and progenitor cell biology. This chapter focuses on the development of cell-based assays and considers the advantages of this approach, as well as the limitations that their current design imposes on the assay development and implementation process.

Assay Design

General Considerations of Cell-Based Assay Technology Applied to Stem and Progenitor Cell Biology

Cell-based assays offer the ability to perform unbiased screens of libraries for molecules that modulate a wide range of biological phenomena (see other chapters in this volume). In the stem cell arena, screens have been implemented to identify small molecules that stimulate differentiation along multiple lineages (Ding *et al*., 2003; Wei *et al*., 2004; Wu *et al*., 2004). The readouts of these assays can be based on detailed analysis of microscopic images collected, capable of detecting subcellular events, such as localization of a protein (high-content screening, HCS) (e.g., Granas *et al*., 2005; Morelock *et al*., 2005; Starkuviene *et al*., 2004; Thomsen *et al*., 2005; Werner *et al*., 2006), which provides advantages over detection of a

particular cellular constituent using a plate reader-based assay (conventional high-throughput screening, HTS) because of the potential to reveal information about subcellular processes and detect relatively rare differentiation events against a background of nonresponding cells. This section focuses on the development of assays to measure the activity of individual marker gene promoters as a readout of differentiation because of their ubiquitous application in stem and developmental biology and because the fundamentals are the same as for detection of structural or biochemical features, which present more content.

Analysis of the wells is often best achieved through image acquisition by automated microscopy because this method affords the most precise means of reporting data on the basis of the number of cells that are affected (e.g., differentiate into a target cell type). This is particularly important when relatively few cells in a well might be expected to be competent to respond to the inducer, such as when cells in the well are heterogeneous [as in differentiating embryonic stem cells (ESCs) or primary cell material] or when the response is expected to vary stochastically by activity and the primary hit is modestly effective at the test dose or if toxicity or other confounding activities coexist. Detailed discussion of the HCS image analysis methods suitable for stem cell biology can be found in other chapters in this volume.

Promoter–Reporter Constructs

Some of the first stem cell-based screens were implemented using antibody binding to visualize differentiation (Ding *et al.*, 2003), and this continues to be a common approach for cell-based assays. Immunological detection of differentiation markers, however, has several obvious drawbacks for scale-up, including cost and availability of immunological and detection reagents and incorporation of wash steps that increase the risk of cell detachment.

The use of gene promoters to direct expression of reporter proteins is routine in developmental biology, and fluorescent proteins, such as green fluorescent protein (GFP) and dsRED variants, offer a straightforward detection method (Hadjantonakis *et al.*, 2002; Zhang *et al.*, 2002). If expressed as a fusion protein to a localization domain, a transcription factor or membrane-association domain for instance, the fluorescent protein can be used in real time to detect regulated translocation to spatially constrained compartments within cells. Fluoresent proteins are also engineered into real-time biosensors of numerous cellular-signaling processes, including detection of metabolites or protein–protein interactions indicative of flux through a signaling pathway (e.g., see Miyawaki *et al.*, 1997; Nalbant *et al.*, 2004;

Tallini *et al.*, 2006; Zaccolo and Pozzan, 2002; reviewed in Tavare *et al.*, 2001; Zaccolo and Pozzan, 2000; Zhang *et al.*, 2002). Alternatives to fluorescent proteins include enzymes that catalyze production of a fluorophore [e.g., β-lactamase directed cleavage of LiveBLAzer fluorescence resonance energy transfer (FRET) substrate (Invitrogen)], although these systems traditionally have lacked the ability to visualize subcellular localization of reporter proteins. Alternative fluorescent technologies that permit subcellular localization use covalent modification of the reporter protein with a small fluorescent probe, such as HaloTag (Promega).

A potential pitfall of cell-based, differentiation assays might be encountered when multiple steps in the differentiation program separate the readout (e.g., expression of a reporter protein at a particular stage of differentiation) from compound addition. Even a potently active compound might not be detected if culture conditions are not conducive to sustaining differentiation through successive steps until the readout is manifest. However, if a compound is active at a late stage of differentiation, the culture conditions might give rise to too few cells that are competent to respond. For these reasons, the readout of differentiation should follow compound addition as closely as possible, preferably within 24 to 48 h. This will also have the advantage of minimizing media changes with attendant reagent savings.

Lentiviral Technology to Create Stable Reporter Lines

Placing reporter proteins under the control of well-characterized differentiation-specific gene promoters requires that the transcriptional elements be located within a contiguous stretch of DNA suitable for introduction into cells. A number of relatively short, <10-kb promoter fragments have been described as capable of directing transgene expression to stem, progenitor, or terminally differentiated cell types, including the genes Oct4 for embryonic stem cells (Yang *et al.*, 2005), αMHC (Subramaniam *et al.*, 1991; Takahashi *et al.*, 2003) or MLC2V for cardiomyocytes (Yan *et al.*, 2003), PDX-1 for pancreatic tissue (Melloul *et al.*, 2002), and insulin for pancreatic β cells (Odagiri *et al.*, 1996). VSV-G pseudotyped HIV lentiviral vectors are well suited for transducing promoter–reporter cassettes incorporating such fragments. VSV-G pseudocoated vectors are generated by three plasmid (vector, packaging, and VSV-G plasmids) calcium phosphate cotransfection of 293T cells, yielding serum-free, viral supernatants with titers of 10^8 infectious viral particles/ml after concentration (as in Reiser, 2000) that can be stored at -80°. For infections, concentrated viral supernatant containing polybrene is added to stem cells, including murine or human ESCs, in suspension at a multiplicity of infection (MOI) of 10 to 100. The mixture is then plated as for routine culture.

Infection efficiency of ESCs under these conditions often exceeds 50%. The infection efficiency can be considerably higher for many other cell lines, but can be considerably lower for nondividing primary cells.

Targeted Introduction of Reporters into Marker Gene Loci by Homologous Recombination

The spatiotemporal patterns of gene expression that characterize many developmentally interesting genes require transcriptional elements that are distributed over large regions of DNA and thus these gene promoters exceed the carrying capacity of plasmid, HIV lentiviral, or common DNA viral transduction systems.

Homologous recombination in transgenic mice is routinely used to integrate reporter cDNAs under the control of marker DNA loci. Because targeted recombination is performed in mouse ESCs, ESCs harboring fluorescent proteins in reporter loci represent a potential source of cells for HCS and HTS.

Homologous recombination in other cell types is more difficult to achieve for a variety of reasons, such as poor transfection efficiency. An alternative is to target integration into reporter loci contained on bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs) because of their ability to contain large regions of genomic DNA and their stability. Commercially available libraries of BACs harboring upward of 150 kb of cargo DNA can be maintained and targeted by homologous recombination within *Escherichia coli*. Recombination-mediated genetic engineering of BACs, known as recombineering, is a powerful method for fast and efficient manipulation of large genomic regions for subsequent cell culture experiments. Protocols are available on the Recombineering Web site maintained by the NCI (http://www.recombineering.ncifcrf.gov). Recombineered BAC or PAC DNA is often prepared using a commercial kit (such as Qiagen) or using CsCl density sedimentation. Transfection into ESCs is often performed using a protocol involving lipid: DNA complex reagents (Bauchwitz and Costantini, 1998), but nonetheless is often inefficient and requires optimization (Montigny *et al*., 2003). An efficient method to overcome inefficient transfection by retrofitting a drug selection cassette has been developed by Wang *et al*. (2001).

Validation of Cell Lines

To confirm promoter–reporter fidelity, engineered cell lines must be tested for correct expression of the reporter transgenes by colocalization of transgene expression with expression of the endogenous gene by *in situ*

hybridization or immunohistochemistry. In the case of ESCs, differentiation of certain derivatives can be achieved in differentiating aggregates of cells maintained generally in suspension, termed embryoid bodies (EBs), (Robertson, 1987). Within 3 to 4 days, 65% of the cells express brachyury, a mesendodermal marker, by FACS analysis (unpublished results), by which time EBs can be either maintained floating or plated onto gelatin-coated cell culture plastic dishes or coverslips. In either case, beating cardiomyocytes arise by day 8. Because EBs are highly heterogeneous, immunohistochemistry or *in situ* hybridization is essential to verify coincident reporter protein expression with the spatiotemporal pattern of endogenous gene expression. For lineages that form poorly in EBs, such as later stages of the endocrine pancreas, mouse ESC reporter lines can be evaluated rigorously in chimeric mouse embryos obtained by injection of murine ESCs into preimplantation blastocysts that are then introduced into the uterus of pseudopregnant females (Robertson, 1987). At appropriate stages, founder chimeric embryos should be examined for ESC contribution and marker expression in relevant tissues to verify that fluorescent protein expression coincides with the spatiotemporal pattern of endogenous gene expression. The evaluation of reporter expression in human ESCs, for which the production of interspecies chimeras might be limited by poor contribution, as well as by law and material transfer agreements, verifying the fidelity of reporter expression, presents considerable challenges and is an argument in favor of targeting reporter proteins to human gene promoters in BACs to maximize the likelihood of faithful expression.

Assay Development

Confidence in the ability of an assay to resolve hits from background is based on its signal-to-noise ratio and the overall dynamic range. In the stem cell arena, it might be desirable to use cells that are not clonal and might even be primary cell isolates. The signal-to-noise ratios and dynamic range values in assays with these cells are likely to be more problematic than with established, clonal cell lines, resulting in an increased likelihood of false outcomes.

In general, high-throughput assays are considered robust if the coefficient of variation is less than 10% across trials and the signal-to-noise ratio and dynamic range results in a Z' value of >0.5. Z' is calculated according to the following formula:

$$Z' = 1 - (3 \times \mathrm{sd}[S_0] + 3 \times \mathrm{sd}[S_{\max}])/\mathrm{ABS}(\mathrm{mean}[S_{\max}] - \mathrm{mean}[S_0]),$$

where sd is standard deviation and S_0 and S_{max} are values obtained for negative [S_0] and positive (S_{max}) controls. An image-based HCS assay might very well be amenable to screening at lower Z' values [e.g., 0.3 (Granas *et al.*, 2005)] for several reasons, not least of which is the ability to visually confirm meaningful hits from image data. Evaluating an assay during development requires appropriate positive and negative controls. Physiologically relevant controls might not exist for a number of reasons, however, such as novelty of the assay for which no hit has yet been described. In such a case, it is imperative that the investigator optimize the assay using compounds or reagents that realistically approximate what would be expected from a scientifically interesting hit. Additionally, a pilot screening of about 10,000 compounds can be performed in the hope of blindly identifying an assay ''hit'' that can serve as a positive control. This is an undisciplined approach and is a risky and costly endeavor for a number of reasons.

- There is no expectation of the assay's physiological dynamic range.
- Scale-up and automation of the assay suffers from improper validation.
- No clear metric exists to validate success or failure of the screen on completion.
- Data extraction, quantification, and statistical inferences are nebulous.
- Elevated false positive and negative outcomes drive up confirmation time and cost.

Stem cell assays that use promoter–reporter readouts are often evaluated using image analysis algorithms because the number of responding cells might be a minor fraction of the total number of cells in the well. This would be true if the cell population were heterogeneous, such as differentiating ESCs, for which a small number of responders might be a meaningful positive outcome because only a small fraction of the cells would be expected to be competent to differentiate along one particular lineage. In this context, applying an algorithm that measures the signal generated by a few cells against background nonresponders might lead to a poor Z' value. Approaches that segment the images into single cells and evaluate the number of differentiating cells might circumvent this problem. Alternatively, a metric in which a given well is deemed ''on'' or ''off'' is determined relative to a threshold for activity could be imposed and validated statistically across many replicates. Such a strategy is worth using in a rapid and simple repeat of the primary screen when discrimination of hits by the primary screen is suboptimal.

Methods

Handling and Growth of Murine ESCs prior to Screening

The typical bench-top mESC differentiation assay calls for the growth on mouse embryonic fibroblasts (MEFs) on gelatin and subsequent differentiation within EBs. Each component is problematic in a high-content screening approach. The use of poly-D-lysine-coated tissue culture plastic was used to rapidly filter out MEFs during assay ramp-up. MEFs adhere tightly to poly-D-lysine-coated tissue culture plastic, whereas mESCs loosely tether to the surface. After 2 days in culture, the ESCs are then lifted from the culture dish with cold media, leaving MEFs behind. Two to three passages in this manner will typically clear the culture of all contaminating MEFs. A high level of residual MEFs reduces well-to-well reproducibility for several reasons, including that they might secrete factors that affect mESC differentiation.

Choice of matrix material influences numerous assay parameters, including maintaining the cells in a spread, two-dimensional environment more suitable for image analysis than EBs and the important consideration that certain matrix constituents will positively or negatively influence differentiation. Not surprisingly, evaluation of matrix components has been shown to optimize liver cell differentiation (Flaim *et al.*, 2005). For mESC cardiogenesis assays, we found that a combination of poly-D-lysine and gelatin (see later) was best at promoting mESC survival and adherence, while maintaining differentiation potency, during the assay.

In HC/HT formats it is not entirely unexpected to see deviations from the biology observed at the bench. One example of this was the longer assay development time observed when scaling the assay to a 384-well format. Differentiation in EBs and in larger well plate formats typically results in cardiomyocyte differentiation by 8 days. This time line reproducibly shifted to 10 to 11 days when converting the assay to a 384-well plate format. One possible explanation might be the change in cell number to media volume and alterations in media change procedure, which would alter the concentration of secreted factors that accumulate in conditioned medium.

Imaging and Processing a Stem Cell Differentiation Assay

Other chapters in this volume provide extensive discussion of the imaging processing of particular high-content screening assays. The expenditure of effort up front to increase the ability of the algorithm to recognize features of the cells that reflect the desired biological response will return

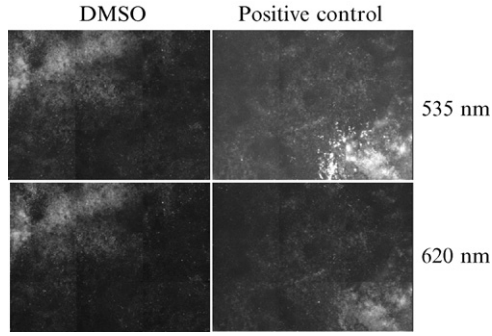

FIG. 1. Comparison of a ''strongly'' positive focus as compared to a negative outcome in an ESC-based cardiogenesis assay. Each panel consists of nine-image, single-well mosaics of images. Images were acquired at 460 nm (nuclear, not shown), 535 nm (specific channel, upper panels), and 620 nm (nonspecific channel, lower channels); emissions were gathered on the INCell 1000 with a 10× objective and 4 × 4 binning. The focus of eGFP+ cardiomyocytes in this case coincides with increased autofluorescence, probably reflecting cell death in cells that had grown into a multilayered structure. From an imaging perspective, this dictates image capture of as much of the well as possible so as to maximize the likelihood of capturing a positive event. The appearance of a high-intensity 535-nm signal is localized to the lower right quadrant of the positive control well.

the investment by improving the hit identification. This sections discusses a few points relevant to acquisition and processing of images for cell differentiation assays. The number of images acquired per well depends on the incidence of responding cells. Figure 1 demonstrates the appearance of a ''strong'' positive focus of differentiation as compared to a negative outcome in promoter–reporter ESC cardiogenesis assay in which relatively few cells are expected to respond to the inducing stimulus. Nine image single-well mosaics of both 535-nm (specific channel, eGFP) and 620-nm (nonspecific channel, autofluorescence) emissions were gathered on the GE INCell 1000 with a 10× objective and 4 × 4 binning. From an imaging perspective, this dictates image capture of as much of the well as possible so to maximize the likelihood of capturing a positive event. To achieve this end, one can capture many images with a higher power objective or fewer with a low-power objective. The requirement for spatial resolution will ultimately drive the decision of objective and binning. Promoter–reporter assays that measure gene activity produce a simple yes/no response; thus, lower resolution is preferable due to speed and memory efficiency.

In imaging a HC/HT screening campaign, processing speed, computational power, and memory constraints should be considered for overall process efficiency. A typical image captured with a 10× objective, 12-bit

depth, with 4 × 4 binning will approximate 170 kB in standard TIFF image format. Table I summarizes the approximate data load one might expect as image quality increases. Process time and storage requirement will increase proportionately.

It will be essential to include operations on an image used to correct systemic problems due to the image capture process. Examples of image processing include correction of the illumination field (commonly referred to as flat-field correction), removal of corrupt pixels, noise suppression, and so on. The need for processing is exemplified in Fig. 2, which shows the (x,y) pixel coordinate map with intensity on the z axis of a blank well from a 384-well plate. Most of the variance is due to imperfections in the illuminating field with occasional punctuate artifacts.

Quantification of the images involves the calculation of segmented pixels in an image (usually an area encompassing the region of interest) dictated by user preferences. The summed time of these two processes can rapidly saturate the computational power of even the most powerful workstations when dealing with large volumes of high-resolution images. As shown earlier, from the perspective of cost effectiveness, it is preferable to use only the minimum resolution required to achieve the computational objective confidently. However, it might be desirable to capture high-resolution images if archival for subsequent data mining for other purposes is anticipated. Using the example of the cardiomyocyte differentiation assay, three colors and nine fields were captured per well with a 10× objective on the GE INCell1000 with 12-bit images binned 4 × 4. The data load of this process on a single 384-well plate is approximately 3.7 GB. This is the lowest resolution that can be achieved on the INCell1000 as of February 2006. This resolution exceeds that which is necessary to resolve "positive" events defined by promoter–reporter activity. Furthermore, simplicity in processing and quantification is also desirable to enhance

TABLE I
APPROXIMATE DATA LOAD

Objective	12-bit image binning	Number of images	Image file size in MB	Three-color imaging	Image file size/384-well plate in MB
10×	1 × 1	1	2.832	8.496	3262
10×	2 × 2	1	0.708	2.124	815.6
10×	4 × 4	1	0.177	0.531	203.9
20×	1 × 1	1	11.32	33.98	13049
20×	2 × 2	1	2.832	8.496	3262
20×	4 × 4	1	0.708	2.124	815.6

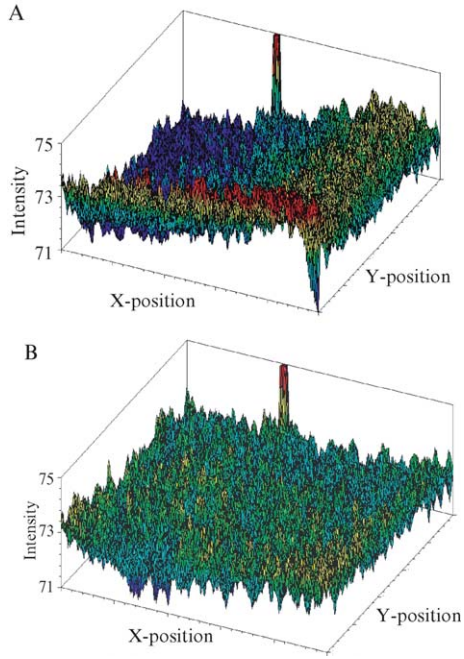

FIG. 2. The need for image correction and processing is shown in the (x,y) pixel coordinate maps with intensity on the z axis of a blank well from a 384-well plate. Most of the variance in an uncorrected image (A) is due to imperfections in the illuminating field with occasional punctuate artifacts. The correction applied here (B) is referred to commonly as flat-field correction. Variance in the illumination field is a normal component of CCD-based image capture and is corrected easily using calibration images gathered after each imaging run.

efficiency. In the ESC-based cardiogenesis assay, the emphasis was on calculation and summarizing events; in the assay described here, 1 well = 1 sample = 1 test. To process images rapidly in as simple a manner possible, an image mask was generated by calculating the global average of each image field applying and accepting all pixels above this threshold plus an empirically defined differential (percentage of global average). After application of a size exclusion filter, the result is the image processing scheme depicted in Fig. 3.

Notes on Data Handling

The success of a HC/HT screening assay depends on the handling of quantified data generated from the images. A general assumption can be made of a population of random samples; given enough samples, the population will

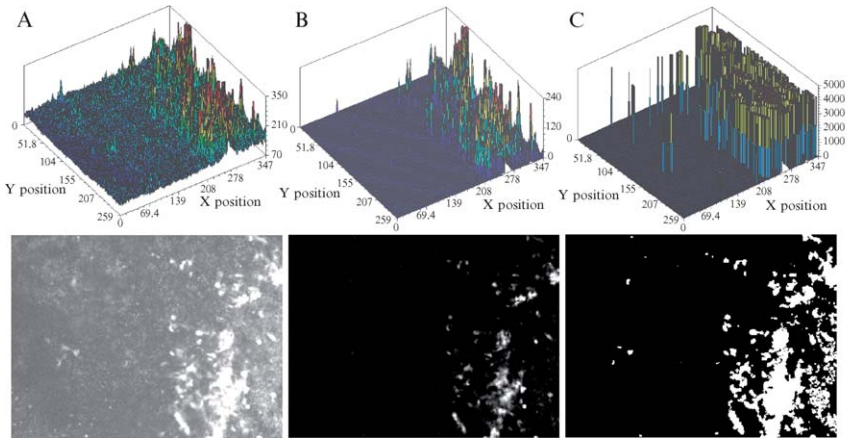

FIG. 3. An image-processing scheme of a ESC-based cardiomyocyte assay. (A) The image and three-dimensional area plot represents the image as it is captured from the microscope. (B) The same image after application of the process for removing nonspecific signal from the image calculation using general statistics from the entire image. (C) The image and plot represent the generation of a two bit-mask, which can be applied to segment images collected in all wavelengths (e.g., 460, 535, 620, 680 nm) to identify features present within the mask. One can apply many methods to achieve a similar result; however, all image-processing algorithms must be validated on appropriate test cases to ensure fidelity to the desired outcome.

have a normal distribution. When zeros and extreme outliers are removed and the population is normalized across all well plates, one should expect a near-normal distribution. In the mESC cardiomyocyte example, the desired result was to obtain a small number of hits that biased mESC lineage toward a cardiomyocyte outcome. In the primary screen, the majority of compounds would be expected to have little or no effect, with a small population of wells harboring artifactual fluorescence due to fluorescent compounds or toxicity. This can be summarized numerically in two categories: zeros (background level) and numbers of extraordinary scale relative to the population at large. Figure 4 depicts this graphically, with the intensity value of the wells on the *x* axis in quantiles and the frequency of observations on the *y* axis. Figure 4A shows that zeros are present in the population of data points at the greatest abundance, skewing the normality of the data population. Figure 4B shows curve normality after filtering out all zeros and artifacts. The box plot (Fig. 4D) depicts near and far outliers distributed greater than 1.5× the value of the interquartile region (IRQ, which equals the distance between

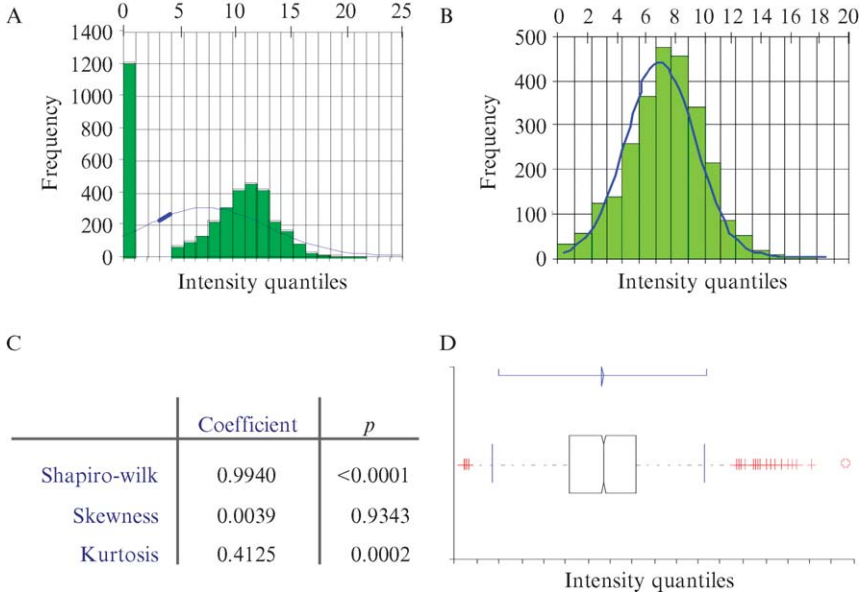

FIG. 4. Graphic summary of raw and filtered data obtained during the screening process. (A) Using the ESC-based cardiomyocyte differentiation assay as an example, samples can be parsed numerically into two categories: zeros and numbers of extraordinary scale relative to the population at large. The frequency distribution curve shows the intensity value of the observation after normalization, in quantiles, is represented on the *x* axis and the number of observations falling in that quantile is represented on the *y* axis. Zeros are present in the population of data points at the greatest abundance (reporter OFF), skewing the normality of the data population. (B) The frequency distribution curve shows normality after filtering out all zeros and artifacts. (C) General statistics for the distribution indicating a slightly peaked curve with kurtosis at 0.4 and only slight right tail skew at 0.004. (D) The box plot depicts the intensity distribution of compounds with the boundaries of the box reflecting the 25 and 75th percentile and the internal bar set at the 50th percentile. Blue tick marks represent 2.5 and 97.5th percentiles. Red + and O symbols reflect near and far outliers, respectively, lying greater than 1.5 IQRs above the upper quartile. IQR is the interquartile region, the region between the 25 and the 75th percentiles. Summarizing data in such a way makes it possible to view the assay results in a top-down view and highlights any important concerns with the data population that need to be addressed.

25 and 75th percentiles) above the upper quartile with red + and O symbols, respectively. Using general statistics to summarize data makes it possible to view the assay results in a top-down view and highlights any concerns with the data population that should be addressed.

Conclusion

Illustrations were provided to describe the approaches taken to develop assays for stem cell differentiation. The major advantages of black-box, cell-based screens are that the compounds identified are selected based on biological function and can serve as powerful probes of basic mechanisms of cell phenomena, such as stem cell differentiation. A consideration with such screens is achieving sufficient dynamic range and a signal-to-noise ratio with minimal plate-to-plate and day-to-day variation to discern hits. This can be achieved with judicious selection of promoter–reporter constructs and image analysis techniques. As data handling and storage capabilities and complementary algorithms for data mining improve, it will be practical to archive image data for subsequent analysis long after the initial screening campaign. We envision that this will enhance probe discovery in the future.

Appendix

As an illustration, a protocol is provided for an ESC cardiogenesis screen, which would need to be adapted to suit differentiation along alternate lineage or for other cell types, so is provided as an illustration.

Growth of ESCs

Cells are a murine embryonic stem cell line carrying the αMHC promoter (REF) fused to a fluorescent eGFP reporter (REF). Cells are recovered from a fresh thaw by coculture with MEFs and 10% fetal calf serum (FCS)-enriched DMEM fortified with penicillin and strepfamycin, non-essential amino acids (NEAA), Na-pyruvate, β-mercaptoethanol, and leukemia inhibitory factor (LIF). On third passage cells are converted to poly-D-lysine-coated plates (0.01%; 300 kDa) and weaned to screening growth media with identical formulation of recovery media with the exception of 7% FCS. Cells are passaged twice prior to seeding 384- or 1536-well plates.

Cell Plating

The 384- and 1536-well plates are prepared with a matrix of poly-D-lysine and gelatin at 0.005 and 0.05%, respectively. The matrix is left on the plates overnight at room temperature and removed and washed once with water. Residual water is aspirated from the plates, which are dried at 42° overnight. Seeding densities determined for 384- and 1536-well plate formats are 750 to 1000 and 175 to 225 cells per well, respectively. Twofold shifts from these approximations can result in failure of the assay due to lack of

sufficient cells or excessive spontaneous differentiation. Cells are resuspended in a half-well volume of growth medium without LIF and applied to the plates and left to grow for 2 full days.

Assay Execution

On days 2 and 4, compound fortified medium is applied to the wells. In brief, the compound is prepared in CRG8 growth media at 2× final concentration and seeding volume equivalent added. Residual compound fortified media is then diluted twofold prior to second addition on day 4. Compounds are removed on day 6, and fresh growth medium is reapplied to cultures every other day. Ideally, each plate should contain positive and negative controls.

Assay Termination

On approximately day 10 (± 1 day) the assay is terminated by evacuation of media from wells and the addition of 4% paraformaldehyde for 45 min. Plates are then washed twice in 1× phosphate-buffered saline(PBS) and then stained in 1 μg/ml of 4′,6′-diamidine-2′-phenylindole dihydrochloride (DAP1) in 1× PBS for 30 min. DAPI is removed from the wells with one wash of 1× PBS and preserved in a solution of 50% glycerol and water.

Quantification

Images gathered from the GE INCell 1000 are flat-field corrected and quantified using an image subtraction technique utilizing the image global average and a global average based differential to remove the nonspecific 535-nm fluorescent signal. This process is executed using the GE software package Developer Toolbox v1.5 (formerly known as IRI). An image bit mask produced from the 535-nm channel is passed through a size-exclusion filter of approximately 3px and applied to both 535- and 620-nm channels. Average and integrated densities in addition to area measurement are provided as output to data tables.

Data Handling

The data set is normalized across plates. We generally normalize using medians and application of simple linear scalers within each plate. A common alternative is to rely on control samples in each plate. Data are then power transformed according to the data set for visual statistical viewing of data. Data are binned and filtered according to user criteria (rank, stdev, morphology, correlation across gathered excitations, frequency, etc.).

Acknowledgments

The authors acknowledge the help of Suzette Farber-Katz, Maria Barcova, Pam Itkin-Ansari, and Fred Levine for developing the MIN6 insulin promoter screen and Joaquim Teixeira, Tomo Takahashi, and Richard T. Lee (Brigham and Women's Hospital, Boston) for developing the αMHC screen. This work would not have been possible without the many helpful discussions of Fred Levine, Jeff Price, Joaquim Teixeira, Pam Itkin-Ansari, Maria Barcova, Rosa Guzzo, and Steve Vasile at the BIMR. The research was supported by grants from the Juvenile Diabetes Foundation and the NIH (HL059502, HL071913 and DK068715) to MM.

References

Bauchwitz, R., and Costantini, F. (1998). YAC transgenesis: A study of conditions to protect YAC DNA from breakage and a protocol for transfection. *Biochim. Biophys. Acta* **1401,** 21–37.

Ding, S., Wu, T. Y., Brinker, A., Peters, E. C., Hur, W., Gray, N. S., and Schultz, P. G. (2003). Synthetic small molecules that control stem cell fate. *Proc. Natl. Acad. Sci. USA* **100,** 7632–7637.

Flaim, C. J., Chien, S., and Bhatia, S. N. (2005). An extracellular matrix microarray for probing cellular differentiation. *Nature Methods* **2,** 119–125.

Granas, C., Lundholt, B. K., Heydorn, A., Linde, V., Pedersen, H. C., Krog-Jensen, C., Rosenkilde, M. M., and Pagliaro, L. (2005). High content screening for G protein-coupled receptors using cell-based protein translocation assays. *Comb. Chem. High Throughput Screen* **8,** 301–309.

Hadjantonakis, A. K., Macmaster, S., and Nagy, A. (2002). Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2,** 11.

Melloul, D., Marshak, S., and Cerasi, E. (2002). Regulation of pdx-1 gene expression. *Diabetes* **51**(Suppl. 3), S320–S325.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997). Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388,** 882–887.

Montigny, W. J., Phelps, S. F., Illenye, S., and Heintz, N. H. (2003). Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells. *Biotechniques* **35,** 796–807.

Morelock, M. M., Hunter, E. A., Moran, T. J., Heynen, S., Laris, C., Thieleking, M., Akong, M., Mikic, I., Callaway, S., DeLeon, R. P., Goodacre, A., Zacharias, D., and Price, J. H. (2005). Statistics of assay validation in high throughput cell imaging of nuclear factor kappaB nuclear translocation. *Assay Drug Dev. Technol.* **3,** 483–499.

Nalbant, P., Hodgson, L., Kraynov, V., Toutchkine, A., and Hahn, K. M. (2004). Activation of endogenous Cdc42 visualized in living cells. *Science* **305,** 1615–1619.

Odagiri, H., Wang, J., and German, M. S. (1996). Function of the human insulin promoter in primary cultured islet cells. *J. Biol. Chem.* **271,** 1909–1915.

Reiser, J. (2000). Production and concentration of pseudotyped HIV-1-based gene transfer vectors. *Gene Ther.* **7,** 910–913.

Robertson, E. J. (1987). ''Teratomacarcinomas and Embyronic Stem Cells: A Practical Approach.'' IRL Press, Oxford.

Starkuviene, V., Liebel, U., Simpson, J. C., Erfle, H., Poustka, A., Wiemann, S., and Pepperkok, R. (2004). High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. *Genome Res.* **14,** 1948–1956.

Subramaniam, A., Jones, W. K., Gulick, J., Wert, S., Neumann, J., and Robbins, J. (1991). Tissue-specific regulation of the alpha-myosin heavy chain gene promoter in transgenic mice. *J. Biol. Chem.* **266,** 24613–24620.

Takahashi, T., Lord, B., Schulze, P. C., Fryer, R. M., Sarang, S. S., Gullans, S. R., and Lee, R. T. (2003). Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes. *Circulation* **107,** 1912–1916.

Tallini, Y. N., Ohkura, M., Choi, B. R., Ji, G., Imoto, K., Doran, R., Lee, J., Plan, P., Wilson, J., Xin, H. B., Sanbe, A., Gulick, J., Mathai, J., Robbins, J., Salama, G., Nakai, J., and Kotlikoff, M. I. (2006). Imaging cellular signals in the heart *in vivo*: Cardiac expression of the high-signal Ca^{2+} indicator GCaMP2. *Proc. Natl. Acad. Sci. USA* **103,** 4753–4758.

Tavare, J. M., Fletcher, L. M., and Welsh, G. I. (2001). Using green fluorescent protein to study intracellular signalling. *J. Endocrinol.* **170,** 297–306.

Thomsen, W., Frazer, J., and Unett, D. (2005). Functional assays for screening GPCR targets. *Curr. Opin. Biotechnol.* **16,** 655–665.

Wang, Z., Engler, P., Longacre, A., and Storb, U. (2001). An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res.* **11,** 137–142.

Wei, Z. L., Petukhov, P. A., Bizik, F., Teixeira, J. C., Mercola, M., Volpe, E. A., Glazer, R. I., Willson, T. M., and Kozikowski, A. P. (2004). Isoxazolyl-serine-based agonists of peroxisome proliferator-activated receptor: Design, synthesis, and effects on cardiomyocyte differentiation. *J. Am. Chem. Soc.* **126,** 16714–16715.

Werner, T., Liebisch, G., Grandl, M., and Schmitz, G. (2006). Evaluation of a high-content screening fluorescence-based assay analyzing the pharmacological modulation of lipid homeostasis in human macrophages. *Cytometry A* **69,** 200–202.

Wu, X., Ding, S., Ding, Q., Gray, N. S., and Schultz, P. G. (2004). Small molecules that induce cardiomyogenesis in embryonic stem cells. *J. Am. Chem. Soc.* **126,** 1590–1591.

Yan, X., Price, R. L., Nakayama, M., Ito, K., Schuldt, A. J., Manning, W. J., Sanbe, A., Borg, T. K., Robbins, J., and Lorell, B. H. (2003). Ventricular-specific expression of angiotensin II type 2 receptors causes dilated cardiomyopathy and heart failure in transgenic mice. *Am. J. Physiol. Heart Circ. Physiol.* **285,** H2179–H2187.

Yang, H. M., Do, H. J., Oh, J. H., Kim, J. H., Choi, S. Y., Cha, K. Y., and Chung, H. M. (2005). Characterization of putative cis-regulatory elements that control the transcriptional activity of the human Oct4 promoter. *J. Cell Biochem.* **96,** 821–830.

Zaccolo, M., and Pozzan, T. (2000). Imaging signal transduction in living cells with GFP-based probes. *IUBMB Life* **49,** 375–379.

Zaccolo, M., and Pozzan, T. (2002). Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* **295,** 1711–1715.

Zhang, J., Campbell, R. E., Ting, A. Y., and Tsien, R. Y. (2002). Creating new fluorescent probes for cell biology. *Nature Rev. Mol. Cell. Biol.* **3,** 906–918.