Automated data management for reliable cell-based screening

Michael Kowalski, Staff Applications Scientist | Beckman Coulter Life Sciences, Indianapolis, IN
Timothy Sherrill, Staff Systems Engineer | Beckman Coulter Life Sciences, Indianapolis, IN
Tara Jones-Roe, Marketing Manager | Beckman Coulter Life Sciences, Indianapolis, IN

Abstract

Cell-based screens have numerous challenges that arise from assay complexity and throughput, however, the challenge of data management is often overlooked. Here we describe how an automated system that combines sample preparation and analysis maintains the data integrity that is essential for a robust and reliable screen.

Background

Screens have long been used as a means of biological interrogation. Cell-based screens can be used to investigate a specific biological pathway, such as by assaying effects of compound or nucleic acid libraries on the expression of a reporter gene or production of a protein of interest. This can be of particular use when there is too little information about a biological pathway to formulate a specific and testable hypothesis. In contrast, phenotypic screening is more interested in a general outcome rather than a specific pathway. An example of this is identifying compounds or gene overexpression/knockdown that induce cell death in a cancer cell line rather than focusing on inhibitors of a given gene or pathway that is known to be mutated in that cancer line.

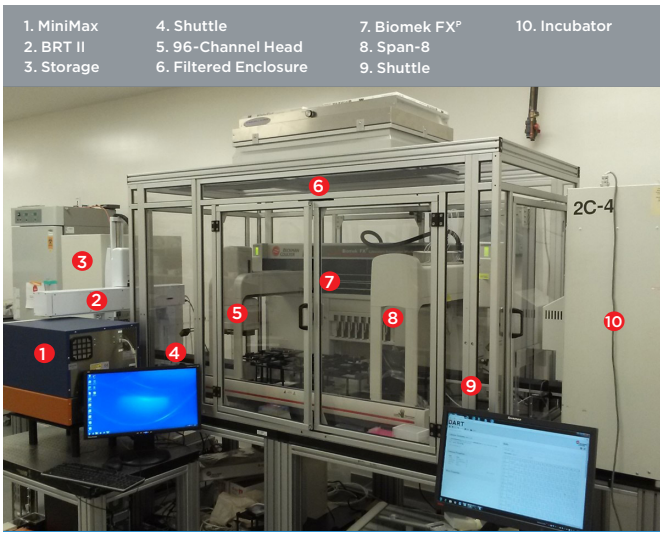


Figure 1. **Automated screening system.** The integrated system enabled the complete automation of a screen for miRNAs that affect cancer cell growth or chemosensitivity. Twelve 384-well plates were processed from cell transfection, to compound and reagent addition, to imaging analysis at multiple time points. All data were tracked using the Data Acquisition and Reporting Tool (DART 2.0).

Due to the high sample throughput typical of screens as well as the frequent complexity of cellular assays, both types of cell-based screening are capable of generating large quantities of data. Because of this, data management must be considered at all points of the screen – from design to execution to analysis. Here we will describe these considerations and demonstrate their application in an automated screen for miRNAs that affect cancer cell growth or susceptibility to chemotherapeutics.

Figure 1 shows an automated screening system built around the Biomek FXP Workstation. To enable higher throughput, the liquid handler was integrated to storage and incubation devices, as well as a SpectraMax i3x Multi-Mode Detection Platform with MiniMax 300 Imaging Cytometer (SpectraMax MiniMax) for imaging analysis. For our screen, HCT 116 colon carcinoma cells were transfected with a library of miRNA mimics and after 24 hours the cells were treated with 5-fluorouracil (5FU) and DRAQ7 to induce and identify cytotoxicity respectively. Cells were imaged in brightfield and 713 nm (DRAQ7+) shortly after compound addition (t=0) and following 24 and 48 hour incubations. Full details of the screen and results can be found in the application note entitled "A fully-automated, image-based screen for miRNAs that regulate cancer cell viability and chemosensitivity" (AAG-1686APP06.16).

Data in Screen Design and Preparation

When designing a cell-based screen, one must consider managing the data associated with each component of the screen. Examples include:

- Library data – what compounds or nucleic acids are present in each well?
- Cell data – what cell type and how many cells are added to each well?
- Treatment data – what transfection conditions or reagents are added to each well and when do these treatments happen?
- Assay data – what readouts are acquired for each well and at what time points?
- Data storage – where will this information be stored and how can it be accessed?

During the initial phase of screening preparation, the major data challenge is bringing data into the screening system. Initial information about the miRNA mimic library, including the miRNA name, well location, and sequence, was imported from a Microsoft® Excel® file that was provided by the manufacturer (Sigma Aldrich). We used the Biomek FX^P Workstation to resuspend and transfer 1920 miRNA mimics from the original twenty-four 96-well plates into six barcoded 384-well stock plates and the miRNA information accompanied the liquid transfer to the reformatted plates. This data transfer was accomplished within the standard Biomek Transfer Step without the need for any further programming. Additional miRNA controls and their corresponding data were also added to the 384-well plates. All of the miRNA data was linked to the 384-well plate barcodes and stored in the Data Acquisition and Report Tool (DART 2.0) Repository. This repository allowed the information to be accessed and added to during the screen as described below.

Data in Screen Execution

During execution of a screen, the main data challenge is to manage data across large numbers of samples and interactions. While million-plus compound library screens may be declining due to cost, compound and nucleic acid-based screening libraries typically consist of thousands to tens of thousands of samples. For the miRNA screen described above, transfection reagents were combined with miRNAs and stamped into replicate plates prior to cell addition, resulting in over 4,600 sample wells. When analyzed at two wavelengths at three time points, the screen results in nearly 28,000 assay data points.

One of the most essential aspects of data management during screen execution is ensuring that any assay data can be traced back to the correct library hit. Figure 2 illustrates the numerous instances of data generation and data transfers that occur due to the complexity of the cell-based screen and the time course for data acquisition. The vertical arrows in Figure 2 illustrate when screen data was accessed from or added to the DART Repository over multiple days. When the 384-well library plates described above were used for creating the transfection/assay plates, this library data was retrieved from the DART Repository via the plate barcode and transferred to the assay plates during the liquid transfer (horizontal arrow). Additional data for the transfection reagent, cell type, and compound were added to each plate during the respective liquid transfer steps. Finally, the imaging data for each time point was added to the assay plate and stored in the DART Repository. This enabled all the relevant data, from library information to results, to be directly associated with one another, ensuring no data loss throughout the process.

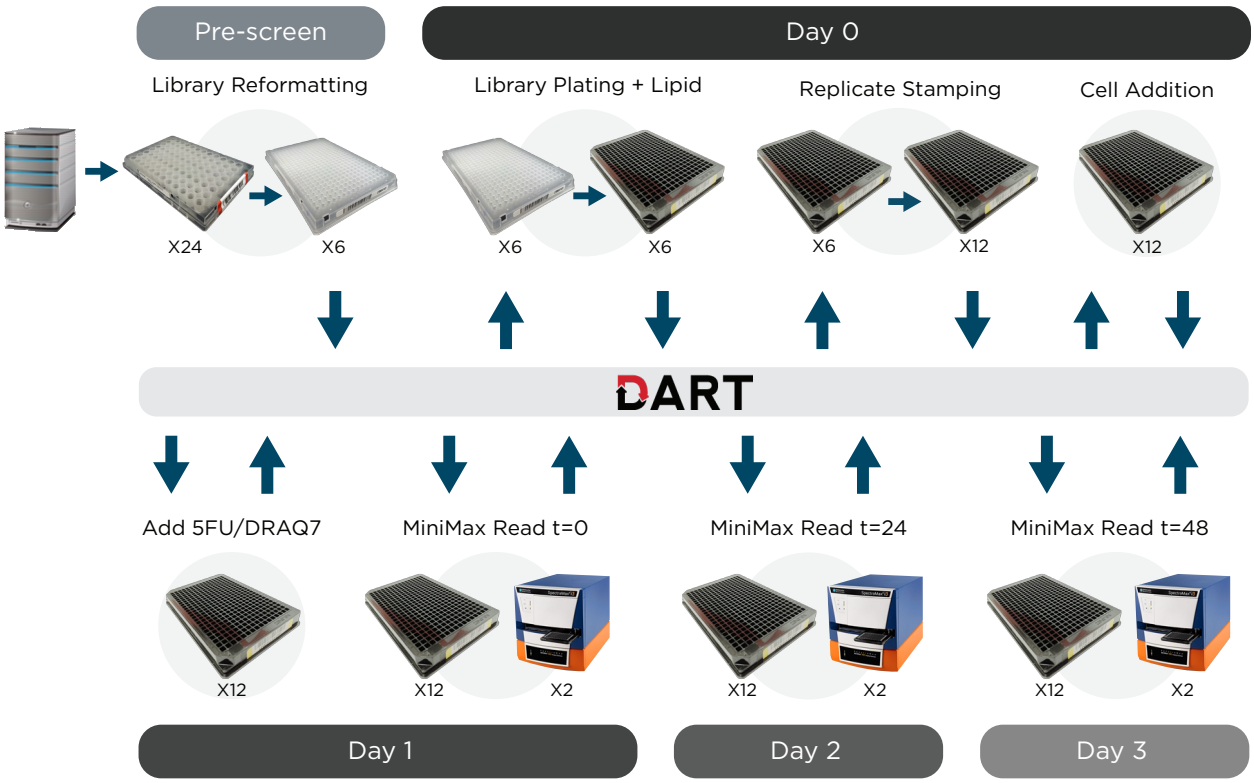


Figure 2. **Data progression through the miRNA screen.** Horizontal arrows indicate the flow of data between plates within an automated method. Vertical arrows indicate data flow between the Biomek FX^{P} Workstation or SpectraMax MiniMax and the DART Repository. This repeated data storage and retrieval allows the data to flow forward through the screen over multiple days, thereby ensuring data integrity.

By using plate barcodes to link the data throughout the process, DART 2.0 ensures data continuity, even if plates are moved between different analyzers or multiple Biomek Workstations. Through networking, these instruments can directly access the DART Repository or external data can be imported into the database. On our screening system, data continuity was aided by the physical integration of storage, liquid handling, and analyzers with barcode scanning of the plates en route from their storage location to ensure proper plate identification. Not only does this automated system reduce the manual time and effort required to set up runs compared to manually starting one of 24 imaging analyses every 15 minutes, it also reduces the opportunity for manual errors such as placing a plate in the wrong location or orientation.

In addition to the “well-level” data we have described thus far, “run-level” data, such as the start and finish times for each step of the screen and “plate-level” data such as time out of the incubator were also stored in the DART Repository. During screen execution, the DART Data Browser (Figure 3) allows you to quickly scan all three levels of data as they are generated. The data can be accessed remotely through a standard web browser and this real-time view of the system further removes the need to directly supervise the automation. One can track volume changes in plates as evidence of pipetting and the ability to monitor the real-time appearance of cytotoxicity data was particularly valuable to ensure the screen was progressing normally during the daily six hour imaging analysis.

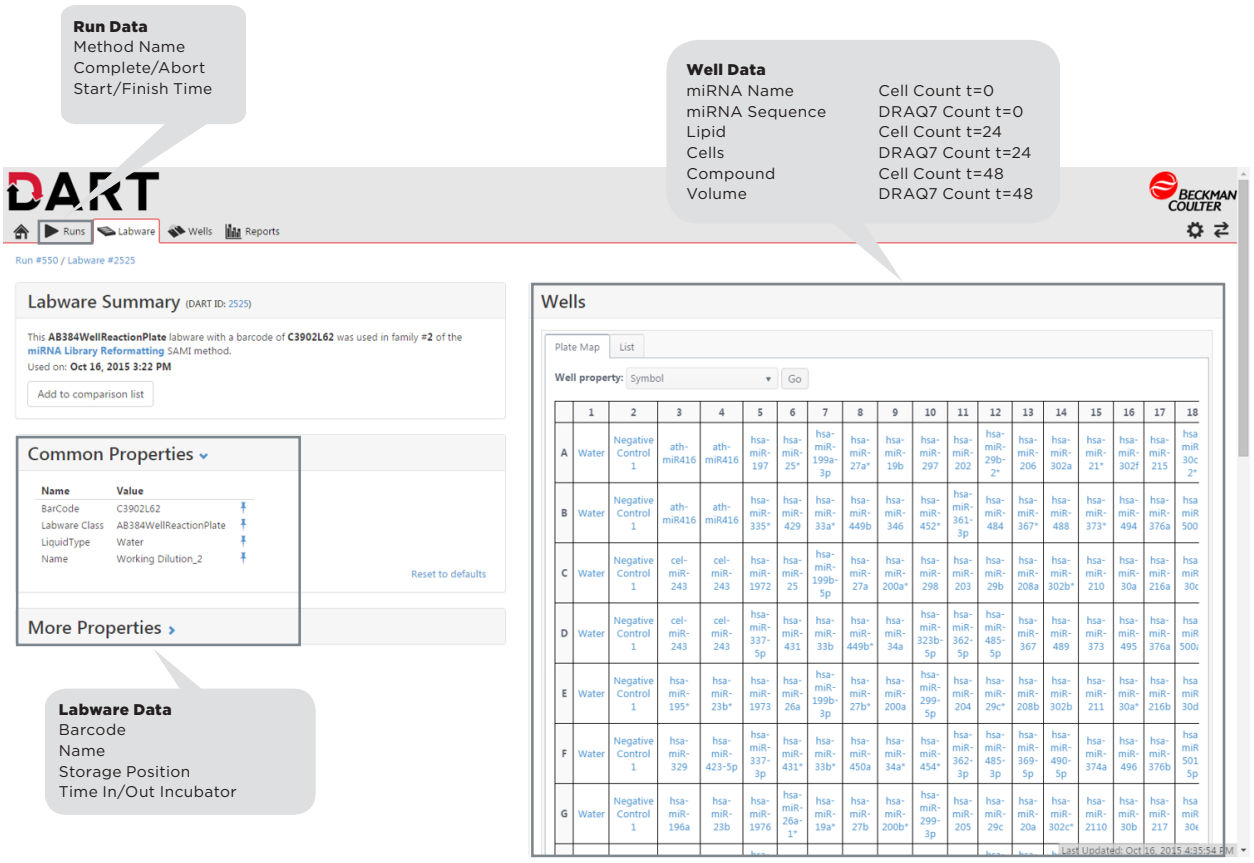
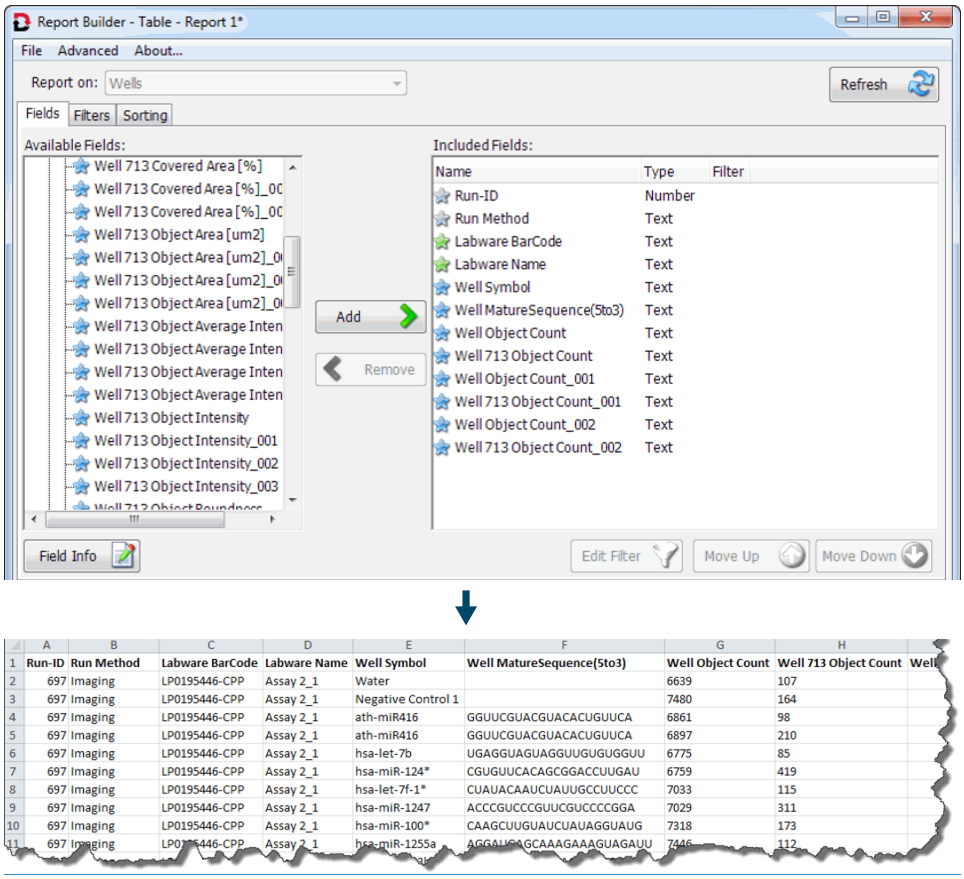


Figure 3. DART Data Browser. Boxed areas indicate the various levels of data stored in the DART Repository that are easily viewed within the DART Data Browser. The largest quantity of data is typically at the well-level and the many pieces of well-level data for a given plate are linked together by the plate barcode. Data for all plates processed in a given run, as well as run-level data (i.e. duration of the method) can be accessed from the "Runs" tab.

Data in Results Analysis

As illustrated above, there is an enormous amount of data generated during a screen, particularly if it is automated. By appending each new piece of data to a plate barcode, both data continuity and ease of use is greatly increased over compiling multiple data files at the end of the screen. However, much of this data may not be essential when it comes to a final analysis. For example, some run or plate data may only be necessary if there are abnormal results, otherwise only the assay results may be needed. In addition, one may only want the identification and sequence information for those miRNAs that generated the desired cell phenotype. DART 2.0 can generate reports that contain any of the available information, and the relevant information can easily be selected in the Report Builder software (Figure 4) or data can be directly exported from the DART Data Browser. These reports can then be opened in Microsoft® Excel® for data analysis. For the miRNA screen, dead cell counts (DRAQ7+) and total cell counts were used to calculate the cytotoxicity level and each well value was compared to the plate average to identify outlying hits.



	A	B	C	D	E	F	G	H
1	Run-ID	Run Method	Labware BarCode	Labware Name	Well Symbol	Well MatureSequence(5to3)	Well Object Count	Well 713 Object Count
2	697	Imaging	LP0195446-CPP	Assay 2_1	Water		6639	107
3	697	Imaging	LP0195446-CPP	Assay 2_1	Negative Control 1		7480	164
4	697	Imaging	LP0195446-CPP	Assay 2_1	ath-miR416	GGUUCGUACGUACACUGUUCA	6861	98
5	697	Imaging	LP0195446-CPP	Assay 2_1	ath-miR416	GGUUCGUACGUACACUGUUCA	6897	210
6	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-let-7b	UGAGGUAGUAGGUUGUGUGGUU	6775	85
7	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-miR-124*	CGUGUUCACAGCGGACCUUGAU	6759	419
8	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-let-7f-1*	CUAUACAAUCUAUUGCCUUCCC	7033	115
9	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-miR-1247	ACCCGUCCCGUUCGUCCCCGGA	7029	311
10	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-miR-100*	CAAGCUUGUAUCUAUAGGUAUG	7318	173
11	697	Imaging	LP0195446-CPP	Assay 2_1	hsa-miR-1255a	AGGAUGAGCAAAGAAAGUAGAUU	7446	112

Figure 4. DART Report Builder and sample report. DART Report Builder allows the user to select the relevant information from all the available run, labware, and well-level data and generate a report in a variety of formats (Microsoft® Excel® table shown).

For this particular screen, the resulting wells essentially contained dead cells, so at the end of the screen, plates were removed from the system and discarded. However, numerous screens would benefit from taking hits forward immediately following the screen. One example would be a screen to identify antibody-producing clones. The automated system described here can use real-time assay data to drive pipetting on the Biomek Workstation. In this scenario, wells that contained cells expressing antibody titers above a certain threshold, as measured by ELISA, could be expanded to larger well formats for continued growth while those below the threshold would be removed from the system. This is another example of how the integrity of a screen can be improved by linking data analysis directly to hit selection and maintenance without the need for error-prone interactions.

Discussion

A screen is only as reliable as the flow of data through it. An assay with a good data management:

- Identifies library samples, cell lines, and reagents as they are introduced to the system.
- Has well identity properties follow the sample through the process, whether to other plates (e.g. replicate or assay plates) or different wells (e.g. serial dilution).
- Captures and stores all information generated during the process and makes this information easily accessible. This allows the verification of an unexpected result or the identification of possible sources of error.
- Directly correlates analytical data with identity information so reports can be viewed with identity and analytical information side-by-side. Having to cross-reference the identity prevents the direct inspection of data and can lead to errors during analysis.

The automated system we have demonstrated here achieved these data goals during the screen for miRNAs that affect cancer cell growth or chemosensitivity. We identified 50 miRNAs that induced cytotoxicity in HCT 116 cells – either alone or in the presence of 5-fluorouracil. 23 of the 50 hits had published information suggesting a role in cancer, 21 of which were proposed to act as tumor suppressors. The two hits that were proposed oncogenes were only identified in the presence of 5-fluorouracil, suggesting that the miRNAs might be enhancing cytotoxicity by driving cell growth in the presence of the chemotherapeutic agent. The high correlation of predicted and actual results supports the idea that the data integrity of the system led to reliable screen results and gives us confidence that the 27 unknown hits may represent novel cancer targets.

Significant resources can be spent in follow-up investigations for screen hit validation. If errors have been introduced to the system, not only will many hits prove to be false positives and waste resources, but one may miss true hits by failing to accurately track them to the correct library sample. As screens are looking for rare events, it is essential that none of these valuable hits are lost due to poor data management.

For Beckman Coulter's worldwide office locations and phone numbers, please visit "Contact Us" at beckman.com

AAG-1735APP06.16