UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

The Development and Application of a High-throughput, High-content System for the Investigation of Stem Cell Biology

Permalink

https://escholarship.org/uc/item/2nd7g57t

Author

McFarland`, Sean

Publication Date

2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Development and Application of a High-throughput, High-content System for the Investigation of Stem Cell Biology

By

Sean Kyle McFarland

Dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David V. Schaffer, Chair
Professor Douglas S. Clark
Professor Tejal A. Desai
Professor Evan W. Miller

Spring 2016

Abstract

The Development and Application of a High-throughput, High-content System for the Investigation of Stem Cell Biology

by

Sean Kyle McFarland

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor David V. Schaffer, Chair

The field of stem cell science is one with enormous potential for impact in both therapeutic applications and in understanding human development and homeostasis. It is increasingly appreciated that stem cells and stem cell behavior are governed by a complex, interwoven network of environmental signals with variable spatial and temporal presentation. While conventional molecular and cellular biology techniques have provided a fundamental foundation for stem cell investigation, advances in the adaptability and through-put of future laboratory work flows will be necessary to address questions in this ever-expanding parameter space. The goal of this dissertation, therefore, has been to instantiate just such a platform, and provide proof of concept evidence as to its utility in stem cell investigations. Design considerations and pipeline engineering are discussed, and data collected with the system in benchmarking, dose-response, and combinatorial experimental formats are provided to illustrate the experimental work the platform enables.

Table of Contents

List of Figures

List of Tables

List of Notes

List of Code

“Cassidy sought no euphoric interludes. They came, when they did, quite naturally and he was content to enjoy them privately. He ran not for crypto-religious reasons, but to win races, to cover ground fast. Not only to be better than his fellows, but better than himself. To be faster by a tenth of a second, by an inch, by two feet or two yards than he had been the week or year before. He sought to conquer the physical limitations placed upon him by a three-dimensional world (and if Time is the fourth dimension, that too was his province). If he could conquer the weakness, the cowardice in himself, he would not worry about the rest; it would come. Training was a rite of purification; from it came speed, strength. Racing was a rite of death; from it came knowledge. Such rites demand, if they are to be meaningful at all, a certain amount of time spent precisely on the Red Line, where you can lean over the manicured putting green at the edge of the precipice and see exactly nothing.”

- John L. Parker, Jr., Once a Runner

Dedication and Acknowledgments

There are innumerable people to thank for the culmination of experiences leading to this moment. The last seven years have been a roller coaster of ups and downs unlike anything I could have anticipated, and at times the only thing that kept it bearable was the love and support of so many colleagues, friends, and family. To all of you, I am forever indebted, and it is to all of you that this work is dedicated.

Particular thanks is owed to those that put up with me through the daily laboratory thick and thin. I ♥ my fellow Dark Siders (Riya Muckom, Brian Perea, Alyssa Rosenbloom) for all of their insightful collaboration and unwavering friendship, Dave Schaffer and Douglas Clark for being stellar (and inhumanly patient) advisers, Noem Ramey, everyone's mother away from home, Mary West, the valiant manager of the Shared Stem Cell Facility, Eric Granlund in the machine shop for his help creating custom parts for our equipment, and Jonathan Dordick, Greg Nierode, and Moo-Yeal Lee for their collegial advice on running the microarrayer,. I'm also sincerely thankful for my earlier experiences at Berkeley in the labs of Irina Conboy and Seung-Wuk Lee, where I learned many valuable lessons and got to work with many fantastic people.

On the administrative side, a quick shout out to Rebecca Pauling and Kristin Olson, the former and current BioE department masterminds that have kept the program running. Along similar lines, to Steven Conolly, for his advice, sincerity, and empathy.

To the BioE class of 2009, y'all are the best! Whether we were on campus, filming skits, hiking (Shire to Mount Doom, and everything in between), brunching, tinkering, camping, traveling to Grandma's … well, suffice to say it was never anything but non-stop, enjoyable adventure! Thank you.

Outside of academia, the Strawberry Canyon Track Club was a quintessential part of my experience, and a necessary outlet for my many lab-induced neuroses. I'm gracious for the menagerie of teammates I was honored to train, race, and live with over the years, and particularly for Carl Rose, an indelible coach and friend.

Above all else, though, I want to thank my family. For my parents, Jake and Gidge McFarland, I cannot even begin to fathom, much less put into words, how awe-inspiring you are. I'd be nothing without you. Aaron McFarland, my brother, you are the best possible friend (and partner in crime). And Nova. I love you more every day :)

Put simply, the lot of you mean the world to me, and I look forward to our many adventures to come. *Thank you*.

Introduction

Scientific inquiry is a discipline founded upon the bedrock of iterative observation and experimentation. Over the course of history, humanity's capacity to understand its environment and itself has advanced in lockstep with the capabilities that contemporary tools and methodologies have afforded. As technology has matured, the frontiers of investigation have likewise expanded, permitting scientists access to increasingly complex areas of study, with correspondingly exciting potential for application.

0.1 Stem Cell Potential and Complexity

The field of human stem cell research represents an archetypal example of this progress, born out of ongoing refinements in molecular and cell biology instrumentation and techniques. From conception, stem cells play a critical role in human development. In the early blastocyst, pluripotent cells within the inner cell mass proliferate and differentiate in a highly coordinated manner to give rise to all tissues of the body. These human embryonic[1] stem cells (hESCs), and their recently generated induced pluripotent[2,3] stem cell (iPSC) counterparts, have proven revolutionary in the way we think about and approach therapy, disease modeling, and development. Though outside the scope of this discussion, tissue-specific "adult" stem cells[4–8] are also critically important, persisting throughout life in tightly regulated niche microenvironments where they help to maintain normal tissue homeostasis. Together, all stem cells are identifiable by their capacity to self-renew, and by their ability to differentiate with varying degrees of potency.

By virtue of this capacity to proliferate and give rise to other cells, stem cells represent a resource of tremendous potential for the development of human therapies and modeling of disease states and basic biology. *In vitro*, expanding human stem cells can offer a scalable source for generating homologous or autologous cellular and tissue-level grafts for replacement therapies. Furthermore, *ex vivo* culture of donor cells creates a window of opportunity for genetic modification that can correct for phenotypic shortcomings in destination tissues, an application area that has received a substantial boost in recent attention with the advent of CRISPR/Cas9 genome editing techniques[9].

Disease-specific hiPSC lines can also be derived from patients suffering a variety of congenital illnesses, as has already been demonstrated for Parkinson's disease[10], spinal muscular atrophy[11], amyotrophic lateral sclerosis[12], type 1 diabetes[13], and others[14,15]. When re-differentiated into relevant affected cell types, retention of phenotypic deficits can allow for on-demand disease-specific models derived from a virtually inexhaustible supply of source cells[11]. This use of hiPSCs as disease models promises to enable the dissection of mechanisms implicated in pathology and the discovery of novel therapeutic compounds, all within the correct genetic and phenotypic contexts of cells affected by the disease.

All of this biological versatility is not without its costs, however, and so it is that the decisions governing stem cell behavior must necessarily integrate numerous signals from the ambient microenvironment, including soluble, insoluble, physical, and mechanical cues (Fig. 0.1). All of these

may interact with each other in antagonistic, orthogonal, contextual, or synergistic ways, substantially increasing the complexity of the input space. Soluble signals have been particularly well-studied, and have found broad applicability in the maintenance of pluripotency[16–19] and the specification of various cell fates, including epithelial[20], neural[8,21–23], cardiac[24–27], cartilaginous[28], hematopoietic[29,30], pancreatic[31,32], and hepatic[33,34] lineages, among others. Immobilized substrate cues of the extracellular matrix (ECM), like those that constitute the commercial substrate Matrigel, are also immensely important in establishing cell context. Various defined alternatives continue to emerge as our understanding of their roles improves[35–37] and have been shown to underpin the stable culture of stem cells of various types[38–41]. Mechanical stiffness of the substrate is increasingly appreciated for its influence on cell behavior in many models, as well, including in embryonic[42], mesenchymal[43], muscle[40], and neural stem cells (NSCs)[44]. Moreover, the growth of cells in 3D scaffolds instead of conventional 2D monolayers has attracted considerable attention, transcending limitations of conventional *in vitro* setting in ways that cal lead to more biomimetic functioning and phenotype of cultured cells[45–49]. Certain cell types can also be sensitive to transient cues such as shear[50,51] and electrical impulses[52,53], important in processes such as cardiogenesis. Taken together, it follows that tools providing the flexibility to test many conditions across multiple aspects of the culture environment may pave the way for deeper understandings of how different properties of the niche interact to dictate cell fate[54,55].

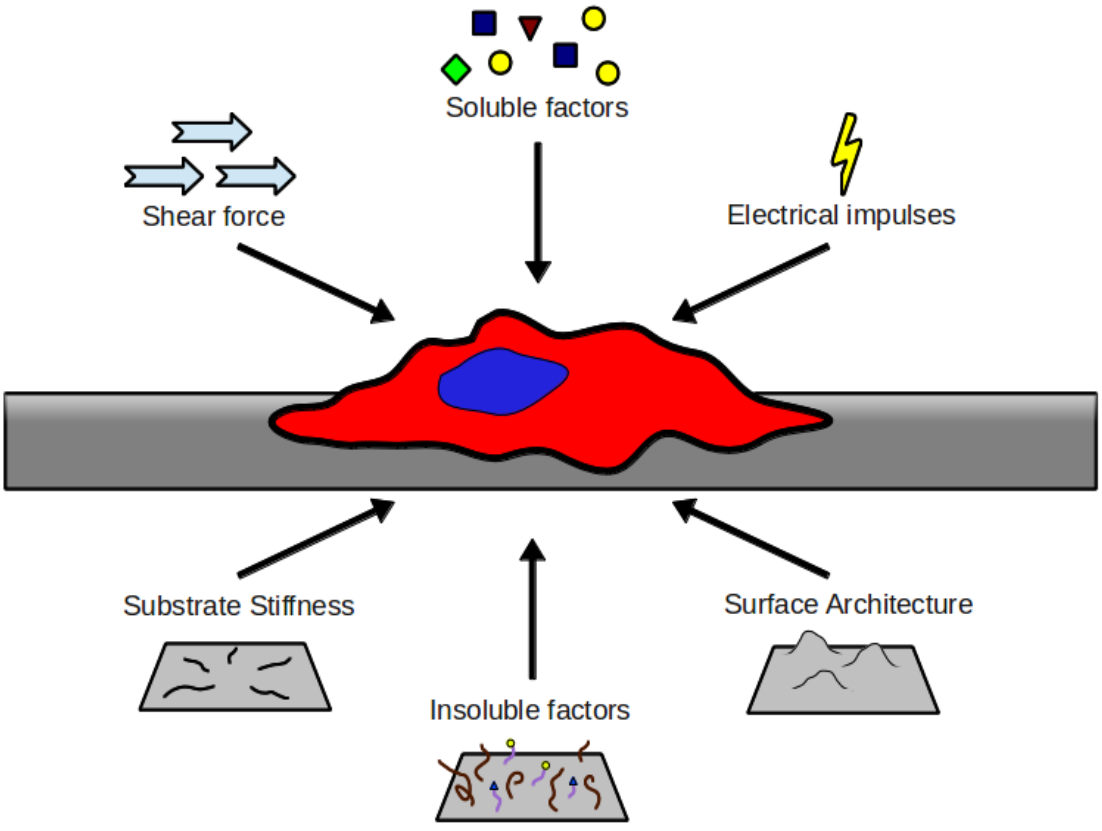

Figure 0.1. The stem cell microenvironment. Stem cell fate is determined by integration of a variety of signals, including shear forces, electrical impulses, substrate stiffness and architecture, and soluble and insoluble chemical cues. The spatiotemporal and combinatorial presentation of these stimuli are also critically important to cell decision-making.

0.2 Innovative Tools Enable Powerful Science

Along these lines, techniques leveraging robotics, automation, and small-scale fabrication have ushered in new paradigms of high-throughput experimentation, allowing researchers to screen libraries of cues orders of magnitude larger than what would be feasible using conventional methods. Through this lens, many traditional mechanisms important to cellular regulation, such as soluble factors and immobilized cues, have been explored in greater depth.

Leveraging gains in raw throughput, a number of studies have found success mining panels of small molecules for effects on cellular behavior. A portion of these endeavors have focused on hESCs, using markers such as Oct4 to screen thousands of small molecule candidates for applications in affecting self-renewal or differentiation[56,57], and in elucidating the biological mechanisms underlying these processes[57]. In a screen looking at tens of thousands of small molecules, Xu et al. discovered two that could enhance hESC survival when passaged as single cells[58]. Perhaps more importantly, though, subsequent investigations of the drugs' mechanisms led to evidence that E-cadherin disruption is causative in the massive die-off of singly passaged hESCs, a phenomenon that had been poorly understood through that time. Using a similarly sized library, a recent study led by Ben-David sought to address reservations surrounding clinical use of pluripotent cell-derived grafts over concerns that they may contain tumorigenic, undifferentiated contaminants[59]. The most potent agent discovered, an inhibitor of the enzyme SCD1, was found to kill pluripotent cells with an exclusivity sufficient to prevent undifferentiated cells from forming teratomas in grafted mice while having no adverse effects on the somatic cell population.

Substrate studies represent another area where high-throughput implementations have had a positive impact on the progression of the field. Pioneering work growing hESCs on various permutations of synthetic polymers revealed conditions capable of generating substantially pure cytokeratin-positive epithelial cells[60], with subsequent modifications to the system opening the doors to screening thousands of candidates in a variety of cell types[61]. Experiments of this scale were later put to the test pursuing a defined substrate composition for long-term self-renewal of various hESC and hiPSC lines, and identified a polymer that could accommodate growth in an undifferentiated state for more than five passages[62]. Substrate effects on cell performance are also dependent on deposited cues, and a variety of printing techniques have been employed to identify immobilized proteins capable of influencing neural stem cell[63] and hematopoietic stem cell[64] fate decisions. High-throughput methods investigating substrate elasticity[40] and surface topography[65] have also born fruit, identifying stiffnesses that permit previously impossible in vitro mouse muscle stem cell proliferation and architectures preferable for human mesenchymal stem cell differentiation, respectively.

Outside the realm of raw throughput, groups have also seen success tailoring screens that examine multiple factors in combination. Prudhomme et al., for instance, reported on a combinatorial regime drawn from four soluble and ECM cues in mESCs, elucidating roles played by 31 intracellular signaling components in decisions to proliferate or differentiate[66]. Another study employed a standard DNA spotter and off-the-shelf materials to investigate 32 different combinations of five ECM molecules, identifying sets that acted synergistically to affect hepatocyte function and mESC differentiation[67]. Similarly, studies in mNSCs utilized high-throughput methods to examine the interplay of Notch and Integrin ligands in survival and proliferation[68]. Most recently, 384 well plates with robotic liquid handling enabled researchers to optimize directed differentiation protocols from

human pluripotent cells to a variety of terminal neural fates, greatly decreasing the time necessary as well as achieving cranial motor neurons, which had not previously been derivable in culture[69]. In a different vein, cellular arrays have been used to great effect to study the toxicological profiles of drugs combined with various metabolizing enzymes, providing a controllable way to examine how the body's own chemistry can give rise to unforeseen toxic intermediates that might be difficult to isolate in animal models[70,71].

Sill other groups have sought to address growing concerns about the reliability[72] and relevance[73–76] of conventional 2D formats by pushing screens into the third dimension with hydrogel scaffolds. The parameters of PEG hydrogels, for example, have been assayed in array format to determine optimal conditions supporting hMSC viability[77]. In another report, mESCs were arrayed onto glass slides in alginate drops and tracked for maintenance of pluripotency and generation of neuroectoderm using *in situ* immunoassays and native reporters[78]. Proof of concepts using microfluidic methods have demonstrated that mESCs can be successfully encapsulated in agar microgels over a 35 fold variation in elastic modulus, permitting the examination of stiffness effects without compromising ease of cell manipulation[79]. As would be suggested by previous 2D observations[43], hESC lineage commitments have been shown to be sensitive to the elasticity of the encapsulating material[80].

0.3 Charting the Path Forward

Automated workflows have much to offer the scientific community, especially as investigators delve deeper into fields like stem cell biology where throughput, combinatorics, and dimensionality of the culture environment are fundamental to experimental success. Unfortunately, while platforms currently exist that may handle one of these aspects very well, they tend to be highly specialized, expensive, and difficult to adapt to subtle nuances from one experimental system to the next. In light of these facts, the goal of this dissertation was to develop a flexible system capable of high-throughput, combinatorial, 3D experimentation, integrate it with downstream high-content imaging and programmatic image analysis, and demonstrate its applicability as a complete pipeline for investigating fate decisions in human pluripotent stem cells. In **Chapter 1**, development of the experimental platform is described, starting from design goals and progressing through the problem solving necessary to implement desired features, establishing a pipeline that takes cells and culture conditions as inputs and yields high-content image data as output. **Chapter 2** discusses considerations and construction of the computational pipeline used to extract features from the image data sets and analyze them en masse to classify cellular outcomes. **Chapter 3** focuses on proof of concept experiments that served to simultaneously benchmark output and validate positive and negative controls. In Chapters 4 and 5, the flexibility of the platform is explored as two panels of molecules are examined for their effects on human pluripotent stem cell behavior. **Chapter 4** centers on their behavior in a dose-response format, while **Chapter 5** dips into combinatorial space and examines each of the agents' effects when paired with each of the other molecules in the screen.

Chapter 1: Platform Development

1.1 Design Goals

The following were identified as necessary design goals for the general purpose laboratory workflow:

1. High-throughput liquid handling to enumerate hundreds of unique conditions with replicates in a standard experiment.
2. High density culture substrate able to stably support proliferating or differentiating cells for a week or more.
3. Human pluripotent stem cell line to use as a model system.
4. Compliance with standard molecular biology assays (viability staining and immunostaining).
5. Temperature control from source plate to culture substrate to accommodate the handling of thermo-sensitive materials (especially hydrogels).
6. Humidity control during printing to mitigate drying during longer prints.
7. Automated imaging to acquire high-content data sets.
8. Automated feature extraction to quantify image-level and cell-level features.

As a starting point, the nascent platform would build upon infrastructure in place from previous cellular screening carried out between the labs of Douglas Clark and Jonathan Dordick[71,78,81,82]. At the time, this consisted of a stock 4 pin Digilab Omnigrid Micro DNA/Protein Arrayer with control computer[83], an entry level spotting device equivalent to what might be found in many universities' core facilities. This particular device utilizes synQUAD™ technology to aspirate and dispense liquid reagents in a non-contact fashion (Fig. 1.1a). It had been idle in storage for nearly three years, so the first order of business was to get it up and running so that its capabilities could be evaluated. Moo-Yeal Lee, a post-doc that had worked on the previous Clark-Dordick collaboration, kindly took a day to visit, help reconstruct it, and provide basic training on the device. Aside from needing a few new parts, it proved mostly functional. Out of the box, it included the capacity to print up to a few thousand arbitrarily addressable spots per hour and humidity control within the print chamber, checking off design criteria 1 and 6.

Since the conclusion of the previous toxicology work, new, specialized pillar/well chips had become available from Samsung that provided some interesting possibilities for design criterion 2 (Fig. 1.1b). While fewer cultures could be fit per standard slide footprint, capped at 532 by the pillar/well architecture, this configuration permitted each of those 532 cultures to be completely independent of each other, eliminating any media sharing or spot-spot interactions that might confound studies seeking to examine multiple soluble factors, alone or in combination. The simplicity of daily media changes, requiring only that a new well chip be printed with the desired conditions and re-stamped with the existing cell-carrying pillar chip (Fig. 1.1c), was appealing for its elimination of aspiration and wash steps and the potential it left open to incorporate temporal variation of soluble factors at daily intervals. The polystyrene make-up of the pillar/well chips would also be closer to the properties of standard tissue culture plastic, allowing results to be more readily cross-validated with conventional manual experiments. One slight drawback was that the pillar chips, perforated through their thickness for gas diffusion, would not be compatible with the vacuum-based mechanism used by the slide deck to keep substrates in place during printing, so a new slide deck would need to be fabricated.

Stocks of a human mesenchymal stem cell-derived induced pluripotent stem cell line (MSC-iPSC) were available in Schaffer lab at low passage number. It was documented as exhibiting stable pluripotency through typical maintenance and expansion and differentiation potential, at minimum, to neural lineage. Given its availability and track record, it seemed a good fit to satisfy design criterion 3. Prior to the generation of the data presented here, the stocks were discovered to have mycoplasma contamination. A new line of hiPSCs demonstrating similar growth and differentiation properties, TCTF[84], were acquired separately and used for all reported experiments.

While the previous toxicology work had made heavy use of plate readers to quantify aggregate staining intensities in cultured spots, the desire to move to high-content image acquisition for richer feature analysis, as per design criteria 7, would require an instrumentation upgrade. Fortunately, the core Shared Stem Cell Facility (SSCF), in close proximity to the lab space for easy access, housed a Molecular Devices Image Express Micro (IXM)[85] (Fig. 1.1d). While it did not natively support the Samsung pillar/well chip format, it did have the capacity to handle user-specified plate definitions, which the facility's manager Mary West was more than happy to help develop. A plate adapter would also be required to hold the chip in place during imaging, since it only supported conventional tissue culture plate footprints, so this was added to the list for the machine shop.

At this point, the pieces that would constitute the pipeline were starting to come together, though there were still a number of design criteria left to be addressed, alongside the new adapter requirements for the printer's slide deck and the IXM. Customization and optimization efforts began in earnest to tie up these loose ends.

1.2 Custom Platform Fabrications

In addition to the custom slide deck for the printer and chip holder for the IXM, custom solutions would be necessary to address design criterion 5 pertaining to temperature control. Work began with the slide deck.

As mentioned earlier, due to perforations in the pillar chips to facilitate gas exchange during culture, the existing slide deck with vacuum-based holders would not suffice to hold the substrates in place during printing. Working with the Eric Granlund and the college of Chemistry machine shop, a body was generated from measurements of the original slide deck and the vacuum ports were replaced with spring-loaded pogo pins to secure the chips (Fig. 1.1e). Since temperature control was also desired to warm or cool chips as needed, the deck was designed with an interior channel to pass liquid for heating and cooling the aluminum body (Figures A.1-6).

A similar approach was used for the printer's source plate platform (Figures A.7-9). Existing dimensions were lifted off of the stock part, and used to design a new part with an interior channel to accommodate cooling (Fig. 1.1f). Subsequent tests found this to not be adequate, however, because 96 well plates as used for the source plate are fabricated with an air gap between the bottom of the wells and the surface beneath. A 3D printer was used to prototype a well-bottom adapter to bridge the gap, which was subsequently rendered in aluminum in the machine shop to bring the source plate's temperature control capabilities to spec (Fig. 1.1g).

The printer head was less straightforward. Due to its small size and the need to not obstruct range of motion during printing, electrically driven Peltier coolers were favored. Despite several attempts in partnership with the electrical shop, however, the various thermoelectric plate configurations were not able to generate temperature gradients more than a few degrees from room temperature. In subsequent consultation with the machine shop, off the shelf brass fittings were found that could be mounted to either side of the printer head to act as a cooling jacket (Fig. 1.1h). Follow-up tests confirmed their ability to rapidly hit temperature targets down to 4ºC.

For the IXM adapter, a body was designed to the specifications of a standard 6 well cell culture plate, a format natively supported by the IXM, and fabricated in aluminum. Care was taken to match the edges so that the end product would remain compatible with the IXM's gasket fitting, which serves to help seal the imaging chamber and minimize evaporation. The area typically occupied by the wells was replaced with a window to permit observation by the objective below, and a pogo pin mechanism similar to that employed in the slide deck was used to lock the chip in position and ensure reproducible positioning from one experiment to the next (Fig. 1.1i).

Finally, to power all aspects of the temperature control, while allowing for the possibility of different temperatures to be assigned to different pieces, fluid from a Polar Series Accel 500 recirculating temperature control unit[86] and a ThermoCube 200[87] (Fig. 1.1j) were routed into the system with appropriately valved tubing.

While the pipeline would continue to undergo iterative refinement over time, with design criterion 5 satisfied, enough of the system was functional for pilot experiments to begin feeling out the parameter space for printing protocols (Fig. 1.2).

1.3 Printing Parameter Optimization and Experimental Design

The success of a printing protocol is inherently tied to the fidelity with which it can accurately array spots within a batch of chips. In the case of more complex prints, maximizing for speed can also become important to mitigate drying effects, but in general the goal is a straightforward one of getting the correct liquids to the correct locations on the target substrates. The settings to achieve this and all the ways the process can go wrong are not always so simple, though. Table 1.1 provides an overview of key considerations when developing a printing protocol.

Print protocols are designed within the AxSys software that serves to interface the controlling computer to the arrayer. This provides a graphical user interface, in which fundamental actions are used to build a printing protocol step by step. Rudimentary support is provided for implementing reusable functions, loops, iterated values, as well as for importing x,y positional information from external text files. While it falls far short of the utility one might expect in a modern programming language and there are some very non-intuitive peculiarities in how a few features are implemented, it provides enough adequate, if laborious, functionality. For the vast majority of prints on the Samsung chips, standards of 60 nl of material printed on the pillars and 800-850 nl of material printed in the wells were used.

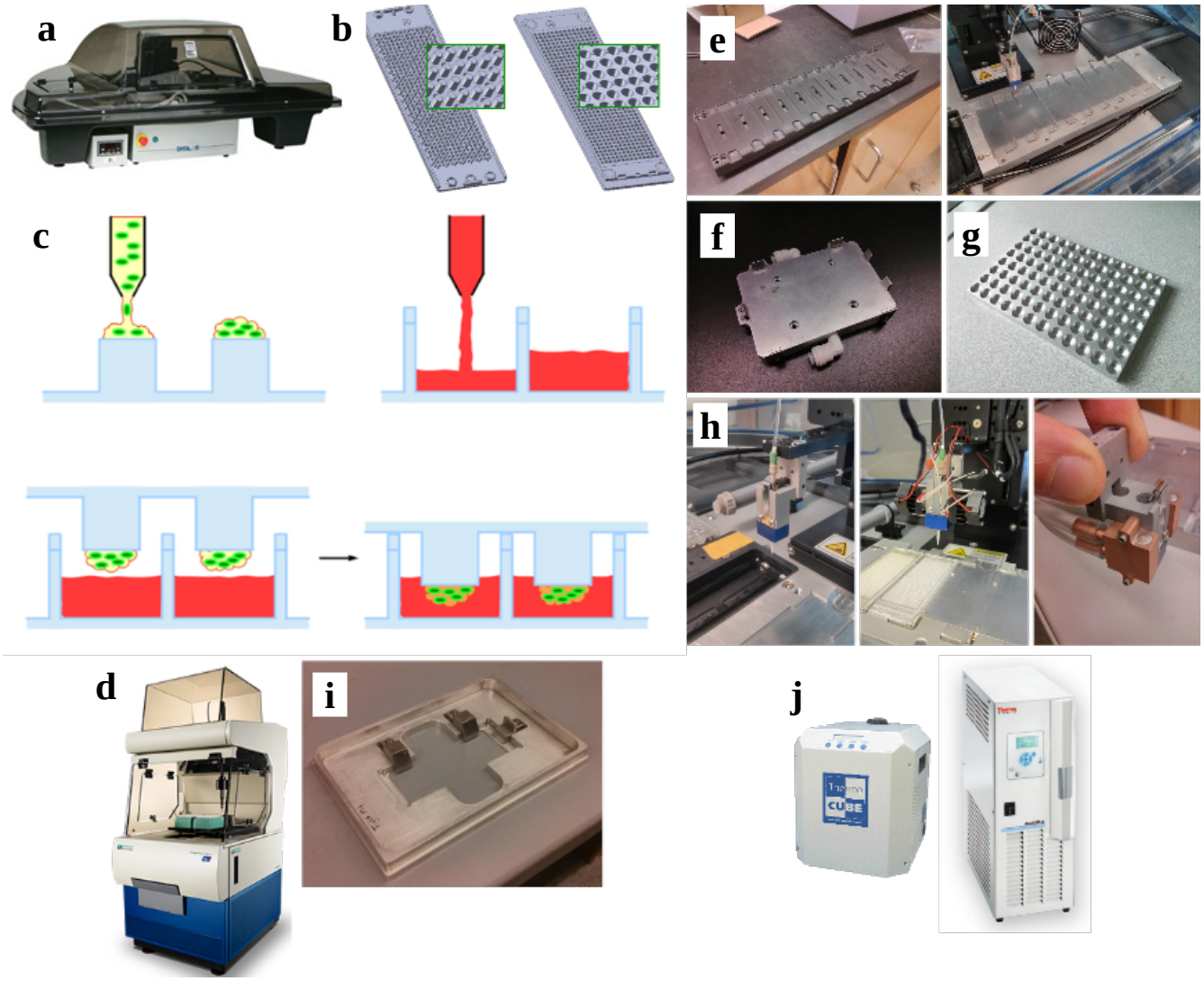

Figure 1.1. Components of the platform. (**a**) Digilab Omnigrid Micro non-contact printer and arrayer. (**b**) Samsung pillar/well chips. (**c**) Culture scheme. Cells are printed onto the pillar chip. Media is printed into the well chip. After cells have been given time to adhere, the pillar chip is inverted and submerged into the well chip containing media. (**d**) Molecular Devices Image Express Micro. (**e**) Stock (left) and custom fabricated slide deck with pogo-pin locking mechanisms and channels for liquid temperature control (right). (**f**) Custom fabricated source deck, with ports for liquid temperature control. (**g**) Custom fabricated adapter to bridge temperature control from the source deck to the source plate well bottoms. (**h**) Stock printer head (left), failed custom Peltier cooling jacket (middle), and final custom liquid temperature control jacket (right). (**i**) Custom fabricated IXM chip adapter for imaging in the IXM. (**j**) Thermocube 200 and Accela 500 recirculating temperature controllers used to manage component temperatures.

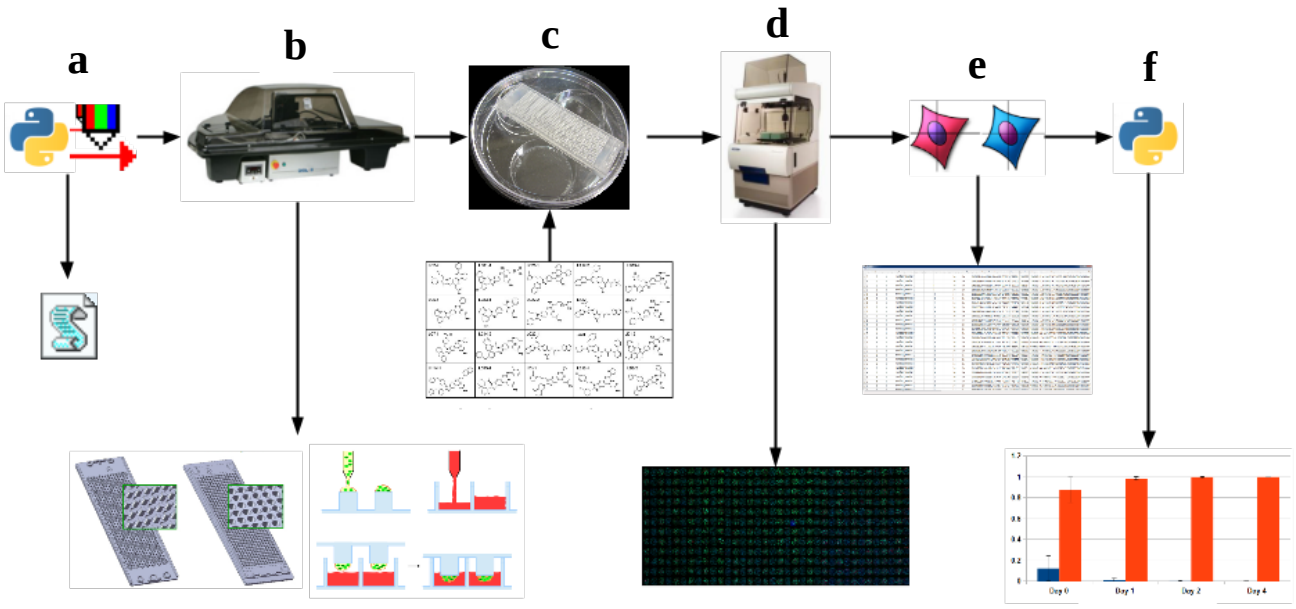

Figure 1.2. Overview of the platform workflow. (**a**) The process begins in AxSys, where the protocol file to run the arrayer is composed. Python may be used to supplement this protocol file with text files specifying positional information. (**b**) The source plates are prepared and the protocol file is used to array the liquids onto the pillar and well chips. (**c**) The pillar and well chips are stamped together, placed inside a humidity chamber, and grown in a standard incubator. Media changes occur daily by stamping into newly prepared well chips. Various agents may be added in these prints to treat the cells. (**d**) The chips are stained and taken to the IXM for automated imaging of each pillar/well spot. (**e**) CellProfiler is used to mine the images for features and store them in a database. (**f**) The feature database is analyzed using Python.

Extreme care had to be taken when using the arrayer to ensure that new print protocols would not damage the ceramic tip with any poorly planned movements. Replacement parts for the machine were not only exorbitantly over-priced, but their proprietary nature meant that work could be completely halted by the unresponsiveness or ineptitude of the supplier, Digilab. Unfortunately this was a difficulty that arose on multiple occasions, and more than once it took *months* to receive a necessary replacement part. Experiences with technical support were similarly disappointing. While this is being said in hindsight, the long-term success of this project would have likely been much better served starting from scratch with open source parts and software. Any head start the arrayer did afford were squandered ten times over dealing with issues related to its instability, proprietary parts, aged software, and poor documentation.

Table 1.1 Important Factors Determining the Success of a Printing Protocol

Raster Speeds	Higher values increase speed, but run the risk of reducing spot resolution, especially in line prints and for more viscous materials.
Syringe Speeds	Higher values increase speed, but can introduce bubbles during aspiration. Low values can lead to variability in printed volumes, especially during line prints.
Valve Open Time	Must be tuned to printed liquid viscosity. Too long will diminish the dispense pressure, decreasing resolution and even preventing drop ejection from the tip. Too short can lead to pressure imbalance, affecting droplet trajectory and clipping deposited volume.

Temperature Control	Cooling is critical for keeping thermo-switchable hydrogels as fluid as possible during the print. However, cooling runs the risk of generating condensation on components which can disrupt droplet deposition, especially if the chamber is being humidified.
Humidification	Critical in complex, longer duration prints to minimize spot drying. Increases risk of condensation and introduction of contamination.
Action Delays	Can become necessary, especially for handling of viscous materials, to allow time for the suction force to balance and all material to be aspirated. Must be tuned with downstream timings to ensure a secondary context runtime error isn't generated.
Preconditioning	Pre-prints that equalize the valve pressure. Important to maximize resolution and check for any problems with the deposition parameters, but also necessary to balance based on the amount of material that can be sacrificed and the aspiration volume needed.
Tip Sonication	Important, especially with "sticky" materials, to keep tip channel free of obstruction build-up that can inhibit print fidelity. Runs risk of damaging tip, and may require re-alignment every time tip is taken off and replaced on printer head.
Void Volumes	At minimum, a buffer of about 15 uL should be figured into any aspiration to account for mixing that will occur at the interface between the sample and the sheath fluid (water).
Print Volumes	The syringe pump holds a maximum of 250 uL. This is a critical consideration when determining how many aspirations from the source plate will be needed.
Overhead Minimization	To maximize efficiency, print protocols should be designed so as to minimize unnecessary movements, washes, etc.
Alignment	Alignment of the tip with the slide deck can drift over time or after maintenance, leading to botched printing resolution. Alignment checks every few weeks are strongly advisable.

As the number of chips and complexity increased for printed experiments, it became impractical to use the AxSys GUI alone to implement the printing protocols. There was also strong incentive to randomize the locations of all conditions on chip to help control for regional effects. Here, Python was used to generate text files containing the positional information for guiding the thousands of distinct movements required to print a batch of randomized chips. The identity of the condition printed at each row and column position on the chip was stored separately in a CSV file to sort all of the images after acquisition with the IXM.

1.4 Materials for 2D Cell Adhesion and 3D Encapsulation

Since the platform was intended to support both 2D and 3D experimentation, work had to be done to test various materials in both formats. Candidates were evaluated on a number of intrinsic and empirically determined metrics. For 2D coatings on which to attach cells, it was important that the cells adhere strongly and quickly to the treated surface and self-renew or differentiate robustly as appropriate for the media. At minimum, it would need to support human pluripotent stem cell culture. It was also desirable, though less a priority, that the coating be as defined as possible, xeno-free, and usable for 3D encapsulation such that results could be more easily compared between both formats. In 3D, it was important that the candidate be chemically or thermo-switchable, of low viscosity in its liquid state, readily adherent to the chip upon setting, and able to support both self-renewal and differentiation of encapsulated cells. As was the case for 2D, support for human pluripotent stem cells was a must, and well-defined, xeno-free options would be preferred, especially if the stiffness could be readily tuned.

With these considerations in mind, a variety of commercial and in-lab options were explored (summarized in Table 1.2). While there was strong desire to use a fully defined option or explore the tunable stiffness provided by the in-lab hyaluronic acid-DNA hybrid and the PEG-pNIPAM candidates, Matrigel's practicality and ease of use proved peerless. Notably, however, 3D drops containing cells from each of the materials all had issues adhering to the chip during immunostaining fixation and wash steps. As this reality emerged, parallel efforts were initiated to test various chip pre-treatment strategies and test tweaks to the conventional immunostaining protocol to obtain workable spot retention approaching 100%.

Table 1.2. Evaluation of Materials Suitable for 2D and 3D Applications.

Material	Source	Switchable	Printer Compatibility	Cell Compatibility	Notes
Matrigel	Corning	Thermo-	Great if kept cool. Low viscosity.	Robust standard for human PSCs expansion.	Undefined. Not bio-inert. Poor 3D adherence.
Mebiol	Cosmo Bio	Thermo-	Abysmal print quality. High viscosity even when chilled.	Robust growth and differentiation in conventional scale culture.	Fully defined. Bio-inert. Poor 3D adherence.
Alginate	Sigma	Chemical	Fully compatible. Low viscosity.	Poor compatibility with tested human PSCs.	Fully defined. Divalent cation polymerization is toxic. Poor 3D adherence.
Hyaluronic Acid – DNA Hybrid	Badri Ananthana-rayanan, Kumar Lab	Thermo- / Chemical	Poor print quality. High viscosity.	Demonstrated efficacy with neural stem cells. Human PSCs unknown.	Fully defined. *In situ d*ynamically tunable stiffness via DNA staples. Poor 3D adherence.
PEG-poly NIPAM	Barbara Eckert, Schaffer Lab	Thermo-	Mediocre print quality. Medium viscosity.	Similar to Mebiol.	Fully defined. Stiffness tunable at synthesis. Poor 3D adherence.

On the pre-treatment side, a number of options were reviewed for improvements to 3D Matrigel droplet retention. At the chip level, there were conventional polystyrene (PS) chips, as well as polystyrene-maleic anhydride (PS-MA) chips to be tested. Either of these could be pre-treated for varying amounts of time in a UV lamp box to increase hydrophilic character. Then, prior to printing the cells in Matrigel, a pre-print could be done with any of a number of potential cross-linking chemistries, including poly-L-lysine and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC). While some combinations of these showed marginal improvements, a number of toxicity concerns were raised, and spot retention was still far below 95%+.

By happenstance, conversation with a former lab colleague led to a trial run of pre-printed dopamine solution. While its structure is not entirely defined, the polymer that results has demonstrated broad spectrum adhesive properties[88]. In combination with some modifications to typical immunostaining procedure as will be discussed subsequently in section 1.6, spots attained near 100% retention through fixation, staining, and all washes, enabling the necessary capacity to probe cellular markers.

1.5 Optimizing Cell Culture at the nl Scale

Operating in a microarray format has the advantages of enabling many more conditions to be tested, and reducing costs associated with reagents used. However, when dealing with only a few tens or hundreds of nl, unique problems arise that are not typically a concern with conventional scale culture. Prominent among these are the risks of sample fluids evaporating, and of a variety of factors contributing to the formation of bubbles within the stamped pillar/well microenvironments.

The former was relatively straightforward to address, requiring only that the stamped pillar/well chips be stored in standard incubators inside a humidity chamber; a set of nested petri dishes that suspend the chip over a bath of water (Fig. 1.2c). This system was demonstrated to be very robust, preventing any noticeable loss in liquid levels within the wells for up to two days. Using this system, cells could be maintained for up to two weeks, with eventual crowding due to expansion being the only problem.

Bubbles within the pillar/well chips proved to be much harder to isolate and eliminate due to the variety of sources that can cause them. Drying, slight inaccuracies during stamping, excessive intra-pillar fluid build-up, and high protein content within the media were all found to be problematic in different circumstances. After eliminating drying due to unnecessary ambient exposure, sources of moisture that could lead to intra-pillar fluid build-up, and migrating conditions away from high protein content media like mTeSR, there was still not a surefire safeguard against bubble formation. The biggest breakthrough in this regard was the use of an anti-foaming agent. In particular, a 1:10,000 dilution of simethicone-based anti-foam C (Sigma) effectively reduced bubbles appearing in typical culture to zero, while showing no adverse effects on cell survival or phenotype.

1.6 On Chip Live/Dead Staining and Immunostaining

Among the assays that could be optimized for usage on chip, live/dead staining for cell viability and immunostaining for expression of markers were essential to the experimental plan. Cell viability is of central importance to judging the potential cytotoxic effects of conditions assayed. For immunostaining, which possesses the versatility to quantify a much wider array of cell markers, and hence, cellular state, a number of antibodies were selected for their pertinence to various aspects of the chosen human pluripotent stem cell system. Oct4, a nuclear-localized transcription factor that is a hallmark of the pluripotent state, was chosen to track the proportion of the cell population maintaining pluripotency. For identifying cells committing to an ectodermal fate, two markers were used. One, Pax6, is a nuclearly-localized transcription factor that indicates neural development. The other, Nestin, is an intermediate filament protein also characteristic of neurally committing cells. For mesoderm, a variety of antibodies were tested, and Brachyury, a transcription factor that plays a role in patterning and is strongly associated with mesodermal commitment, was chosen. Finally, for endoderm, the transcription factor GATA4 was selected for its association with endoderm commitment. In toto, these markers would allow for the pluripotency of stem cells to be tracked, along with their potential commitment to any of the three developmental lineages. Excluding Nestin, all are also transcription factors, and hence localize to the nucleus, facilitating subsequent operations counting and measuring co-staining in cells.

Fortunately, live/dead staining with the calcein AM, ethidium homodimer-based kit from Life Technologies was trivial to transition to the chip environment. Without any rigorous chemical fixations or wash steps, the reagents could simply be printed into a new well chip and stamped with the cell-carrying pillars to distinguish live and dead cells. Spot retention was never a problem in 2D or even the 3D Matrigel formats.

Immunostaining, however, engendered a number of significant challenges. As mentioned earlier, the fixation and washes resulted in unacceptable amounts of spot loss (often > 50%) for all of the 3D systems (Fig. 1.3a, c). In addition, the chip format was found to be exceptionally sensitive to bubble accumulation during the course of sample preparation. Efforts to optimize out these issues covered a wide range of parameters, and any potential solutions had to be carefully balanced against any compromises in image quality or increases background fluorescence.

As mentioned in earlier discussion, spot loss was approached from both angles of increasing 3D spot adhesion through chip pre-treatments and minimizing spot stress through tuning of the immunostaining fixation and watch steps. Multiple fixation methods were explored, including 2% and 4% paraformaldehyde (PFA), ethanol, methanol, and gluteraldehyde-formaldehyde mixtures quenched with sodium borohydride. TBS and PBS were both tested as buffers, with varying amounts of Triton X-100 and TWEEN 20 detergent applied in the various washing steps. The number of washes and incubation time for each wash were tested, as well as whether the washes were performed in bulk baths or wash-specific well chips. Due to the thermo-switchable nature of the hydrogels in play, temperature was an important factor, with the temperature of solutions and incubations attempted at 4ºC, 23ºC, and 37ºC.

Simultaneously, attention had to be paid to bubble problems unique to immunostaining. Not only do bubbles create severe artifacts while imaging, they can also occlude parts of the culture from primary or secondary antibody exposure. Detergents, which are paramount to permeabilizing cells, washing away excess antibody, and achieving optimal signal to noise ratios (SNR), become especially problematic at such small scales. The high protein content of the species-matched sera used to block non-specific binding further aggravates the situation. Moreover, presence of intra-pillar moisture from the wash baths created problems of its own, allowing fluids to be drawn out of the wells by capillary action, introducing bubbles and mixing with neighboring wells. Much like for the bubbles encountered in routine cell culture, anti-foam C was included in all immunostaining solutions. Detergent usage was minimized as much as possible without compromising SNR. Finally, a handful of drying methods were tested to address residual intra-pillar moisture, including a brief, motionless air dry, waving the chips in the air for fixed amounts of time, and using an aspirator from the outer side of the pillar chip to draw moisture out through the gas diffusion perforations without disturbing the fixed cells. Over-drying was quickly found to reduce image quality, so applying just the right amount proved key.

After many rounds of cross-testing these various parameters to jointly address spot loss and bubble formation, 3D immunostaining was reliably achieved by utilizing a polydopamine pillar pre-print (Fig. 1.3b, d), combined with the following protocol skeleton:

Note 1.1. Platform-optimized immunostaining protocol.
(All solutions and steps carried out at room temperature, except where noted otherwise.)

1. Place cell-carrying pillar chips into 2% PFA bath. Incubate 30 minutes.
2. Rinse chips by placing into a bath containing PBS for ~1 minute.
3. Block and permeabilize chips in a bath of PBS with 5% serum matching the species of secondary antibody and 0.1% Triton X-100. Incubate 45 – 60 minutes.
4. Rinse chips by placing into a bath containing PBS + 0.1% Triton X-100 for ~1 minute.
5. Rinse chips in three consecutive baths of PBS for ~1 minute each to remove any residual detergent.
6. Using an aspirator attached to a 2 ml pipet, pass the tip over the outer side of the pillar chip to remove intra-pillar moisture.
7. Stamp into freshly printed well chips containing PBS with primary antibodies and 5% secondary antibody species-matched serum, NO detergent. Incubate overnight at 37ºC in a humidity chamber.
8. Rinse chips by placing into a bath containing PBS + 0.1% Triton X-100 for ~1 minute.
9. Rinse chips in three consecutive baths of PBS for ~1 minute each to remove any residual detergent.
10. Using an aspirator attached to a 2 ml pipet, pass the tip over the outer side of the pillar chip to remove intra-pillar moisture.
11. Stamp into freshly printed well chips containing PBS with secondary antibodies, Hoechst 33342, and 5% secondary antibody species-matched serum, NO detergent. Incubate 2-4 hours at 37ºC in humidity chamber
12. Rinse chips by placing into a bath containing PBS + 0.1% Triton X-100 for ~1 minute.
13. Rinse chips in three consecutive baths of PBS for ~1 minute each to remove any residual detergent.
14. Using an aspirator attached to a 2 ml pipet, pass the tip over the outer side of the pillar chip to remove intra-pillar moisture.
15. Stamp into freshly printed well chips containing PBS and Hoechst 33342. Store in humidity chamber at 37ºC and image on the IXM as soon as possible. If protected from drying, the chips can be imaged up to a day later.

1.7 High-content Image Data Acquisition

Experiments on chip invariably culminate in the need to acquire thousands of images across three channels in a fashion that is reproducible from one experiment to the next. To do this effectively, an automated solution is required. The SSCF IXM was used for this purpose, though a number of special measures had to be taken to integrate the chip format.

As mentioned in 1.2, an adapter first had to be made to hold the chip in place during imaging, and the facility manager Mary West was consulted to create a custom plate definition file encoding every pillar/well position. Dependent upon specified exposure times, trial runs averaged ~40 minutes per chip. While the machine possesses native humidification control, the nl volumes inside the imaging chamber were still subject to drying and bubble formation by the end of a typical run. To address this and minimize bubble artifacts, a transparent adhesive film, similar to those used to seal PCR reaction plates, could be applied to prevent moisture loss through the gas exchange perforations in the pillar chip. While effective at diminishing evaporation everywhere but the chip edges, the adhesive contributes nontrivial background fluorescence to the green channel. By happenstance, it was later discovered that optimal resistance to drying could actually be achieved without the film by turning OFF the IXM's environmental control. While perhaps counterintuitive, this stemmed from the fact that (1) turning off the environmental control kept the chip at room temperature, as opposed to 37ºC, slowing evaporation and (2) though humidified air was being pushed into the chamber, it was discovered that the inlet for this flow was located directly above, and pointing down on, where the chip was positioned during imaging. By turning it off, the net effect of removing the air flowing directly onto the chip's surface was found to outweigh any benefits provided by having the more humid air.

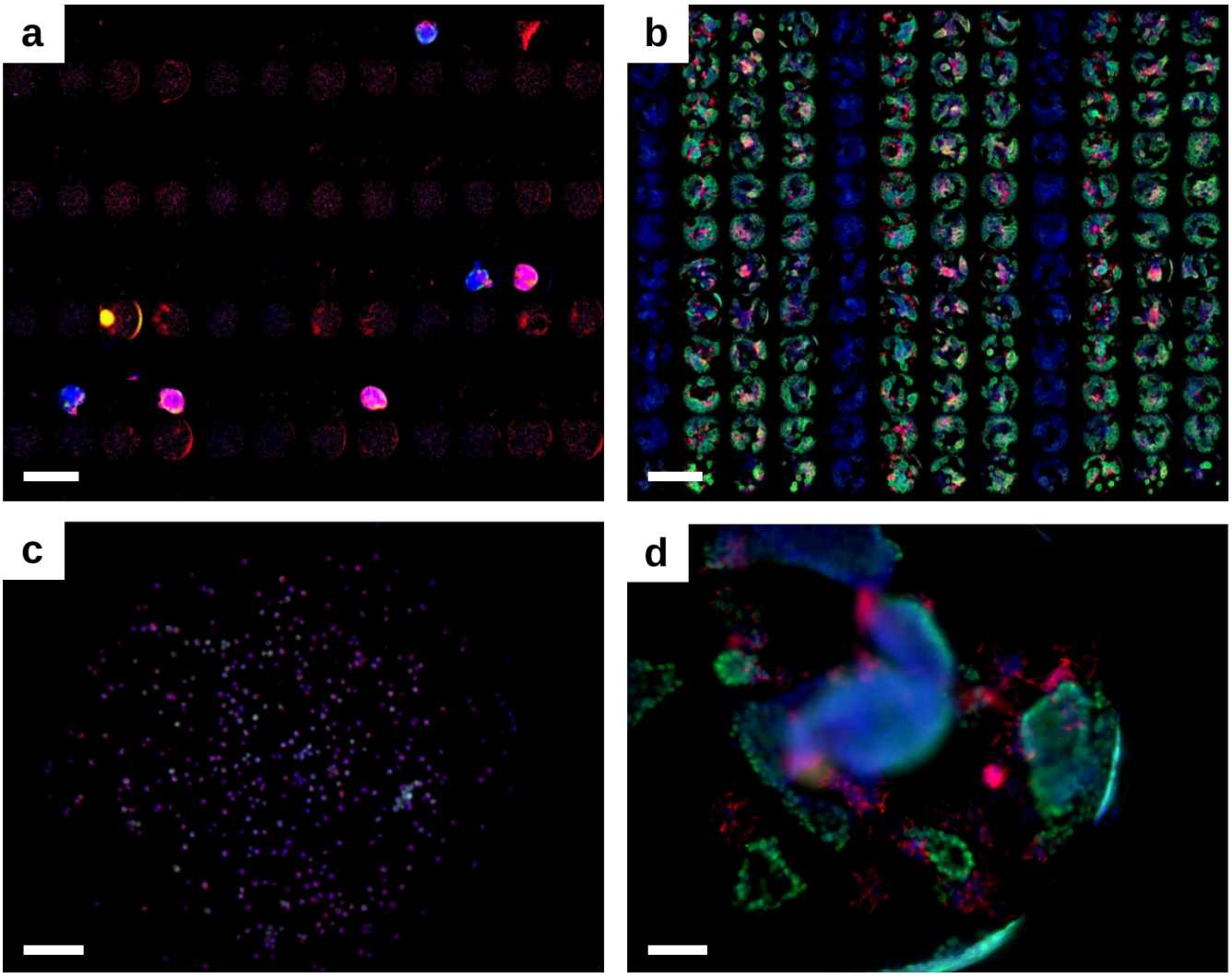

Figure 1.3. Comparison of spot retention post-immunostaining. Representative chip montages displaying spot retention for pillars pre-treated with (**a**) a medley of UV, poly-L-lysine, and EDC and (**b**) polydopamine and then printed with hiPSCs suspended in Matrigel. Not only does the polydopamine provide better spot retention, but it also demonstrates high biocompatibility. In (**c**), a close-up image of a single pillar from (**a**), cells are seen as small spheres and co-stain with ethidium homodimer, indicating the cells are nonviable. In (**d**), however, a zoom-in on the chip from (**b**), cells have expanded on chip, forming healthy colonies that stain for Oct4 (green) and Nestin (red). Scale bars in (**a**) and (**b**) = 1 mm, scale bars in (**c**) and (**d**) = 100 μm.

With these changes, chips could be reliably imaged in ~42 minutes and finish just before drying would start to encroach on the cell cultures. Subsequent upgrades to the IXM camera, hardware, and software further increased the speed of acquisition, notching the time down to an even more comfortable average of ~35 minutes per chip. Unfortunately, however, the system was not fast enough to acquire 3D image stacks across the entire chip. At most 2-3 slices could be acquired per position before drying would become an issue. While 2-3 images could be used with rudimentary success as input to image fusion algorithms to remove out of plane noise (Fig. A.10), this essentially meant that the bulk of the screening would necessarily need to focus on 2D systems until a more rapid-throughput (or confocal) high-content system could be employed.

Acquired images were output as raw 16 bit TIFF files averaging ~14 GB of hard drive space per chip, and it was not uncommon to image 4, up to as many as 8 chips a day with the system operating at full productivity. To keep memory clear, images were regularly backed up within 3 days of generation to keep space on the core facility's central server available for other users.

Chapter 2: Computational Pipeline for Data Analysis

As mentioned, image output from the IXM took the form of 16 bit TIFF gray scale images for each channel. For the typical chip, this meant 1638 images split between the DAPI, FITC / GFP, and Texas Red filters (it is not a perfect multiple of 532 because of certain constraints when defining the IXM plate definition which necessitated acquiring an extra row of images in each run). Manual feature extraction would not be feasible, so in order to translate these image data sets into meaningful quantification, a computational pipeline was developed to handle image pre-processing through feature extraction and ultimate analysis.

2.1 Image Pre-processing

Before processing could begin, a couple of clerical tasks needed to be automated to prepare the raw images for feature extraction:

1. Labeling each image with its corresponding condition metadata, as articulated in the CSV files generated during experimental design.
2. Down-scaling and conversion of images from the native, memory hungry 16 bit TIFF format to 8 bit JPEGs.

By default, the IXM exports images with a naming format that follows the convention:

```
YYYYMMDD-<Plate Name>-<Well>-<Site>-<Channel>.TIF
```

where `<Plate Name>` is the only user-specifiable value. Thus, for each chip image, the well and site values for each position needed to be mapped to corresponding row and column values, and then the "plate name" (which was used to specify the experiment name and identifying chip number within the experiment) would be used to look up the appropriate CSV file from the experimental design and rename the image according to the convention:

```
<Experiment>-<Dimension>-<TimePoint>-<Chip#>-<MediumAndDoseInfo>-
            <Row>-<Column>-<Antibodies>-<Channel>.jpg
```

To give a concrete example, an image might begin as:

```
20160212-DiffPairs2-Ecto-Chip0-2D-d5_A01_s1_w1.TIF
```

And end as:

```
DiffPairs2-2D-d5-i0-mSB431542[250nM]_IDE2[1000nM]-r00-c13-
                sHoechst_Nestin_Oct4-c1.jpg
```

Python's tool sets for file manipulation, CSV handling, and regular expressions were used to automate this in three steps, resulting in an ~60 fold reduction in memory footprint (Code B.1-3).

2.2 Feature Extraction

With the image data consolidated into a unified naming scheme and format, the stage was set for processing the images. Conventional approaches, such as using ImageJ to manually count cells, would not be feasible for the tens of thousands of images to process or the many more complicated features of interest. While hand-coding from scratch would certainly offer the most robust and customizable solution, time constraints dictated that a mature, purpose-made package be explored. The SSCF boasted access to Molecular Device's companion software, MetaMorph, with many appealing native integrations to the data output by the IXM. However, it came with a number of concerns, namely that it was proprietary and on shared time with other SSCF users.

Instead, the choice was made to go with CellProfiler[89], a Python-based image analysis suite developed at the Broad Institute. Its open source nature, rich documentation, wide feature set, respectable performance on parallel architectures, and rapid, live prototyping were all compelling reasons motivating the decision. Use of the software is very simple. In the graphical user interface, a batch of images is selected, a sequence of tunable processing modules is built up through which to pass each image (e.g. Identify Primary Objects, Measure Object Size/Shape, ...), and the collected information is output in the user's choice of a number of standard output formats.

After benchmarking various pipelines configurations, the following skeleton for standard processing was implemented to extract image- and object-level features from batches of images (Note B.1). The list of steps applied to the images and there associated outputs are listed below:

Note 2.1. Conceptual outline of the CellProfiler feature extraction pipeline.

1. Image metadata is extracted from the image file name.
2. Objects are identified in the blue channel. (Typically Hoechst-stained nuclei.)
    - Blue object count.
3. Objects are identified in the green channel. (Typically Oct4-stained nuclei.)
    - Green object count.
4. Objects are identified in the red channel. (Typically lineage-stained nuclei.)
    - Red object count.
5. Object relationships are recorded for co-staining blue, green, and red objects.
    - Identifying object number for any co-localized objects recorded.
6. Image level correlation between each pair of channels is calculated.
    - Pearson correlation coefficient, slope of least squares regression.
7. Image quality is measured.
    - Focus score, local focus score.
8. Image granularity is measured.
    - Fit of variously sized structure elements.
9. Blue, green, and red object intensity properties are measured.
    - Quantification of object intensity central tendency, variance, edge values, and location.
10. Image intensity properties are measured.
    - Statistical measures as for objects.
11. Blue, green, and red objects size and shape properties are measured.
    - Area, perimeter, form factor, solidity, extent, eccentricity, axis lengths, radius, etc.
12. Collected values are output to an HDF5 file.

2.3 Computational Hardware

Despite the built-in parallelization of the image processing pipeline, the hundreds of GB of image data accumulating on a weekly basis began to quickly outstrip the capacity afforded by standard personal computers. To scale with the growing computational load, a total of three data processing workstations were acquired over time to expedite feature extraction and downstream analysis.

There were a number of key considerations in sourcing these machines (Notes B.1-3). CPU parallelization is particularly important for maximizing image throughput, so each was fitted with a quad-core unit capable of virtualizing up to eight threads. RAM, which provides the memory buffer for holding the image data during processing, was set at 16 GB to provide ample accommodation for eight simultaneous threads. At the time of this writing, no image processing suites have yet integrated strong GPU support for maximum processing power, but each machine received a discrete GPU should such software become available.

One noteworthy problem that arose from the machines being assembled separately over a few years is that, though they possessed similar specifications, architectural differences in the components did lead to some discrepancies that required troubleshooting. Specifically, the pipeline file developed for CellProfiler was generating peculiar run-time errors on two of the three machines. The problem was traceable to a CellProfiler-independent problem with the zeromq library used to manage the multi-threading, as indicated by the emergence of the error in numerous user reports filed for other software. Courtesy of the helpful online forum community for CellProfiler, a workaround was found using the latest beta version of CellProfiler in headless mode, i.e. from the command line without a graphical user interface (Note B.4). While this restored the processing throughput to use all 3 computers, the beta version required each batch of images to be further broken up into sub-batches of eight, and was later found to output HDF5 files inconsistent with the structure of those generated by the stable release. As each of the feature databases were subsequently analyzed, additional code was needed to account for these discrepancies.

2.4 Feature Analysis

Since features were batched through CellProfiler approximately four chips at a time, results from any experiment with more than four chips (or any processed using the beta software) ended up spread across a number of HDF5 archives. Again, Python was used to consolidate these disparate files and correct for the reported feature inconsistencies between the HDF5 files generated in the stable and beta versions of CellProfiler (Code B.4).

Due to the fact that intensity can vary experiment to experiment, manual threshold cutoffs were not employed during feature extraction for fear of adversely affecting any experiments with lower than average intensity. This necessitated that additional object filtering be carried out after the fact to correct for image by image threshold differences. The strategy employed drew upon intra-experimental IgG intensity data to threshold out objects lower than 1-2 standard deviations above the experiment’s IgG mean.

Chapter 3: Platform Bootstrapping and Proof of Concept

As all the process elements of the pipeline fell into place, an initial batch of experiments, here referred to as “ImmMaster” or “Imm,” were designed to provide the pipeline with a closer to at-scale proof of concept that would simultaneously serve to vet the positive controls and antibody staining combinations to be used in future work.

3.1 ImmMaster Experimental Design

Each run of the ImmMaster experiment consisted of four chips, for a total of 2128 pillar/well samples. Four control media were tested (Table 3.1). Each of these control media were stained, in at least 56 replicates per run, with each of nine primary antibody combinations (Table 3.2).

Samples were printed on chip using standard line printing methods (Fig. 3.1). Cells were grown in 2D on Matrigel. Based upon reported marker appearance times in the tested media and the time constraints in play, a five day culture period was chosen, after which cells were fixed, stained, and imaged on the IXM. For one of the trials, an extra chip was printed with each of the four media and carried through nine days of culture to compare temporal differences in lineage marker expression.

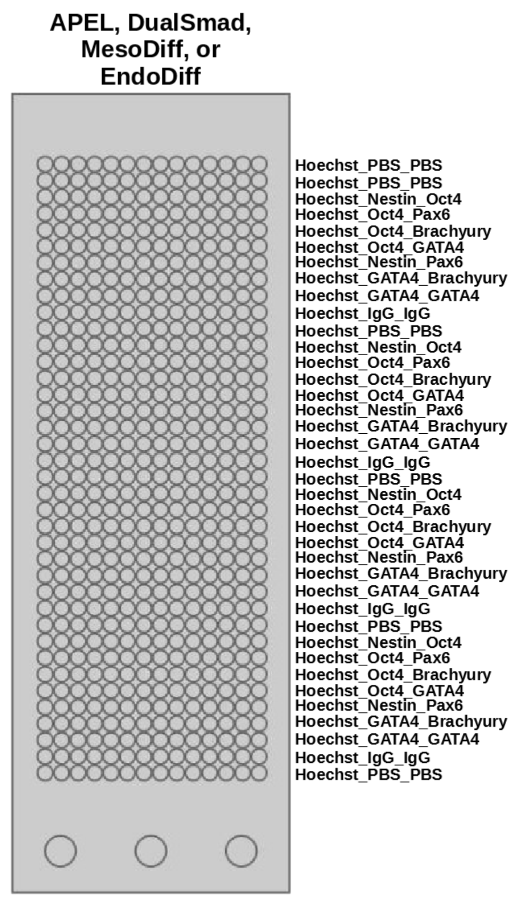

Figure 3.1. ImmMaster chip layout. 4 chips were used for each experimental run., one each of APEL, DualSmad, MesoDiff, or EndoDiff.

3.2 Cell Retention and Proliferation Characteristics

In contrast to earlier attempts at immunostaining on chip, spot retention was very high on many chips using the modified immunostaining protocol with hiPSCs grown in 2D on Matrigel (Fig. 3.2a). One will notice that there definitely seems to be a fluorescence difference going from left to right in each half of the chip. While the exact cause of this is unknown, it is likely tied to the printing method used for the media, which was done column-wise proceeding left to right (with the chip being done in two halves). This would be accounted for in later experimental plans through randomization of the printing protocol. Zooming in to a single pillar, staining is crisp with low background (Fig. 3.2b). There are some patches of cells that seem to be beginning to take on some mesodermal character, staining red for Brachyury, but it is not as pronounced as might otherwise be expected in MesoDiff. This seemed to generally be the case, with day 5 fixation just catching some cells beginning to commit. In general, cells

grew and proliferated well on chip, with MesoDiff treated cells in particular expanding very robustly (Fig. 3.2c). This was consistent with observations in conventional 6 well plates and in other experiments.

Table 3.1. Media used in ImmMaster experiments.

Alias	Base Media	Additives
APEL	APEL	-
DualSmad	APEL	100 nM LDN-193189 , 10 μM SB431542
MesoDiff	StemDiff Mesoderm Induction Media	-
EndoDiff	StemDiff Definitive Endoderm Kit	-

Table 3.2. Primary antibody combinations used in ImmMaster experiments.

Alias	Alexa 488	Alexa 594
Hoechst_PBS_PBS	PBS	PBS
Hoechst_Nestin_Oct4	Ms Anti-Nestin	Rb Anti-Oct4
Hoechst_Oct4_Pax6	Ms Anti-Oct4	Rb Anti-Pax6
Hoechst_Oct4_Brachyury	Ms Anti-Oct4	Rb Anti-Brachyury
Hoechst_Oct4_GATA4	Ms Anti-Oct4	Rb Anti-GATA4
Hoechst_Nestin_Pax6	Ms Anti-Nestin	Rb Anti-Pax6
Hoechst_GATA4_Brachyury	Ms Anti-GATA4	Rb Anti-Brachyury
Hoechst_GATA4_GATA4	Ms Anti-GATA4	Rb Anti-GATA4
Hoechst_IgG_IgG	Ms IgG	Rb IgG

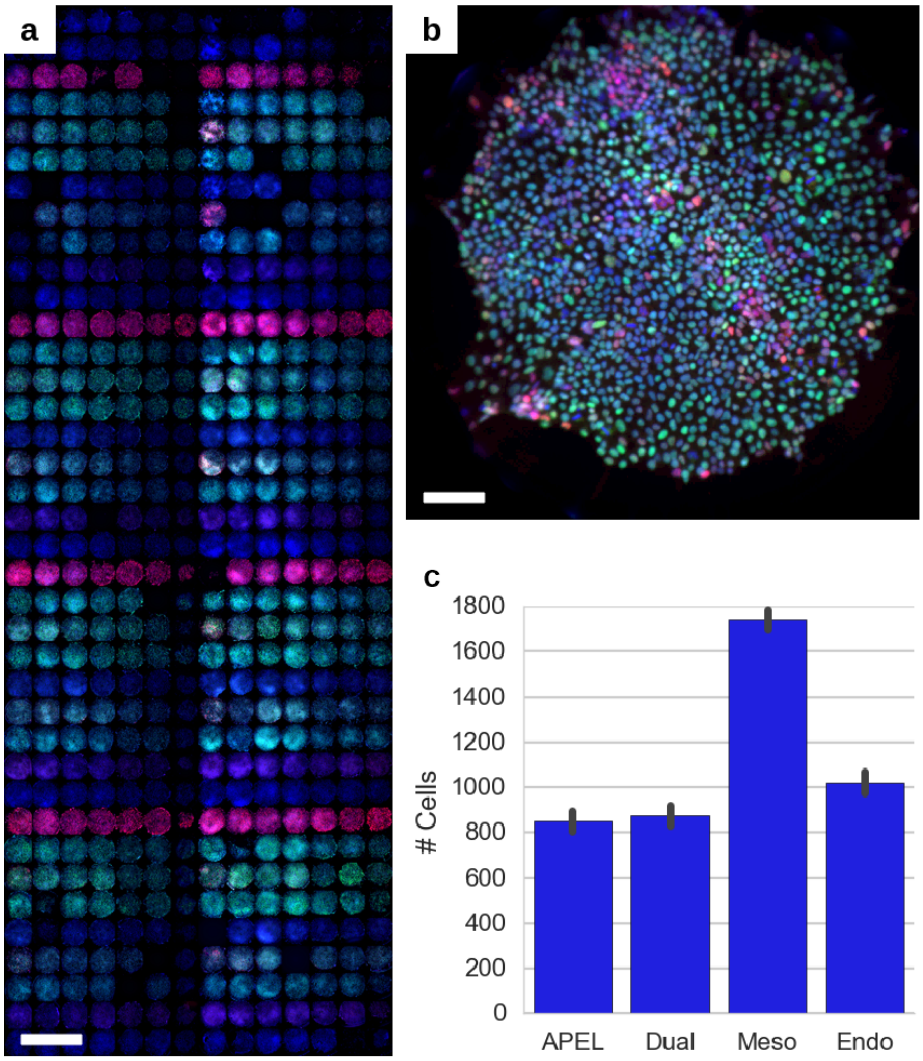

Figure 3.2. Summary appearance of cells on chip. Staining (as per Fig. 3.1) of (**a**) a montage of an entire chip of 532 cell cultures images grown in MesoDiff, scale bar = 1.75 mm, and (**b**) a culture growing on a single pillar staining for Hoechst (blue), Oct4 (green), and Brachyury (red), scale bar = 100 µm. (**c**) Quantification of average cells counted in pillars grown in each media.

3.3 Control Media Lineage Commitment

To establish precedent for immunostaining in downstream studies looking at hiPSC differentiation, panels of antibodies were chosen to study the early commitment of cells to each of the ectodermal, mesodermal, and endodermal fates. Sample images for each of the media and lineage antibody combinations at days five (Fig. 3.3) and nine (Fig. 3.4) of culture revealed a number of distinct staining patterns upon quantification (Fig. 3.5).

At day five (Fig. 3.3), the APEL basal media did not generate any particularly strong differentiation responses, with cells assuming a uniform distribution across the pillar

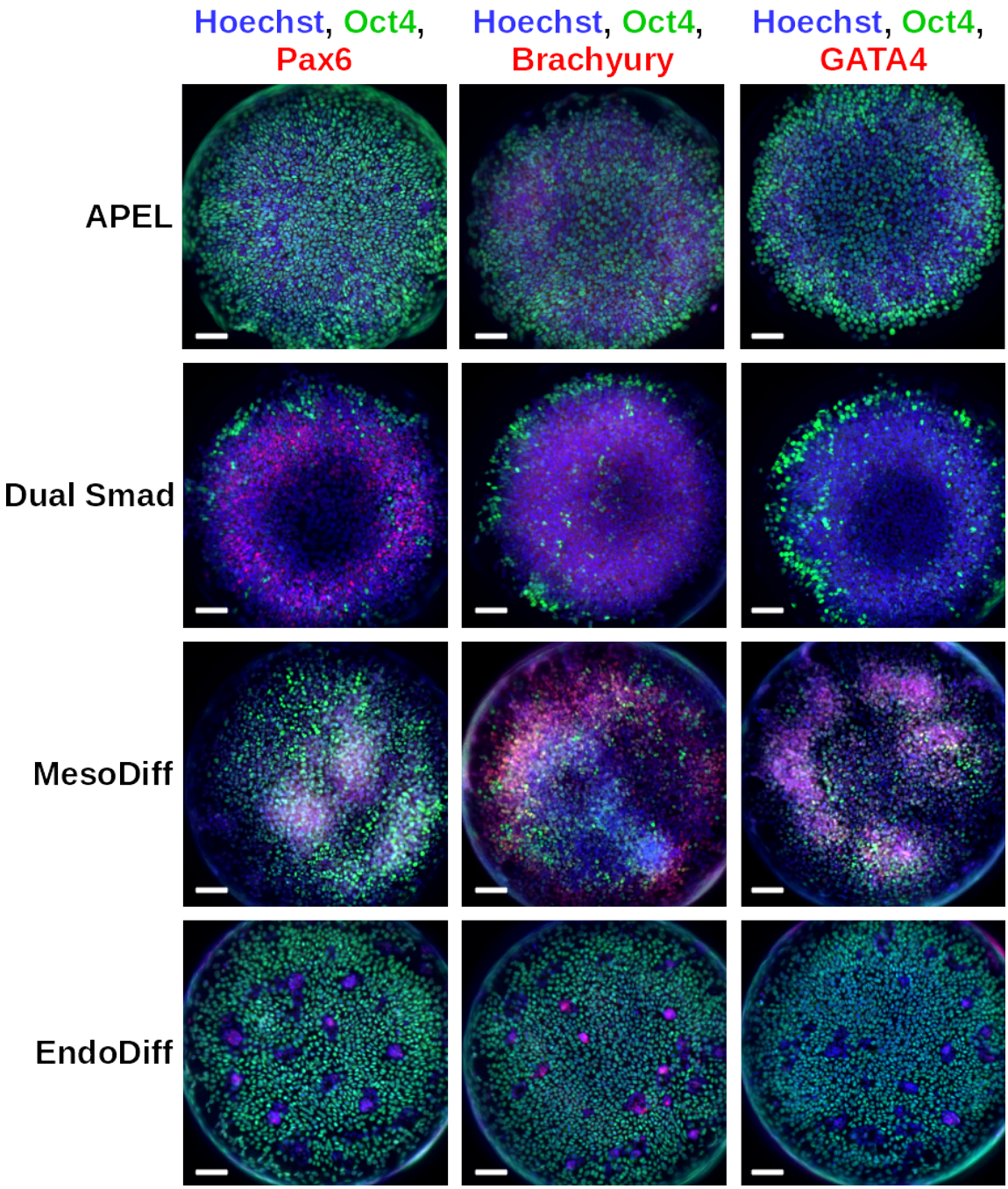

Figure 3.3. Sample images for ImmMaster media/marker pairings at day 5. Scale bars = 100 μm.

and strongly staining for the pluripotency marker Oct4. Dual Smad inhibition, on the other hand, very strongly suppressed expression of Oct4. Levels for of cells singly staining for Pax6, a neurectodermal marker, and Brachyury, a mesodermal marker, were elevated as compared to APEL. Interstingly, cell density tended to be higher along the pillar edges than in the center of the culture, and the few Oct4+ cells tended to be located along the culture fringe. The MesoDiff condition did not necessarily have as strong of a quenching effect on Oct4 staining, but did show elevated levels of singly and double-staining mesodermal and endodermal markers, Brachyury and GATA4. Lineage commitment also seemed to co-localize to large clusters irregularly distributed across the pillar. The EndoDiff condition

was not as strongly lineage marker-inducing as other conditions, but did exhibit the formation of very distinct small cellular islands across the pillar surface that tended to lack Oct4 and be focal points of differentiation markers.

At day nine (Fig. 3.4), cultures grown in APEL were similar to their day five counterparts. Cell density did not increase greatly, perhaps indicating a halt in expansion or cell attrition. In dual Smad inhibition cultures, Oct4 expression was nearly absent, as was also the case for all lineage markers. It is possible

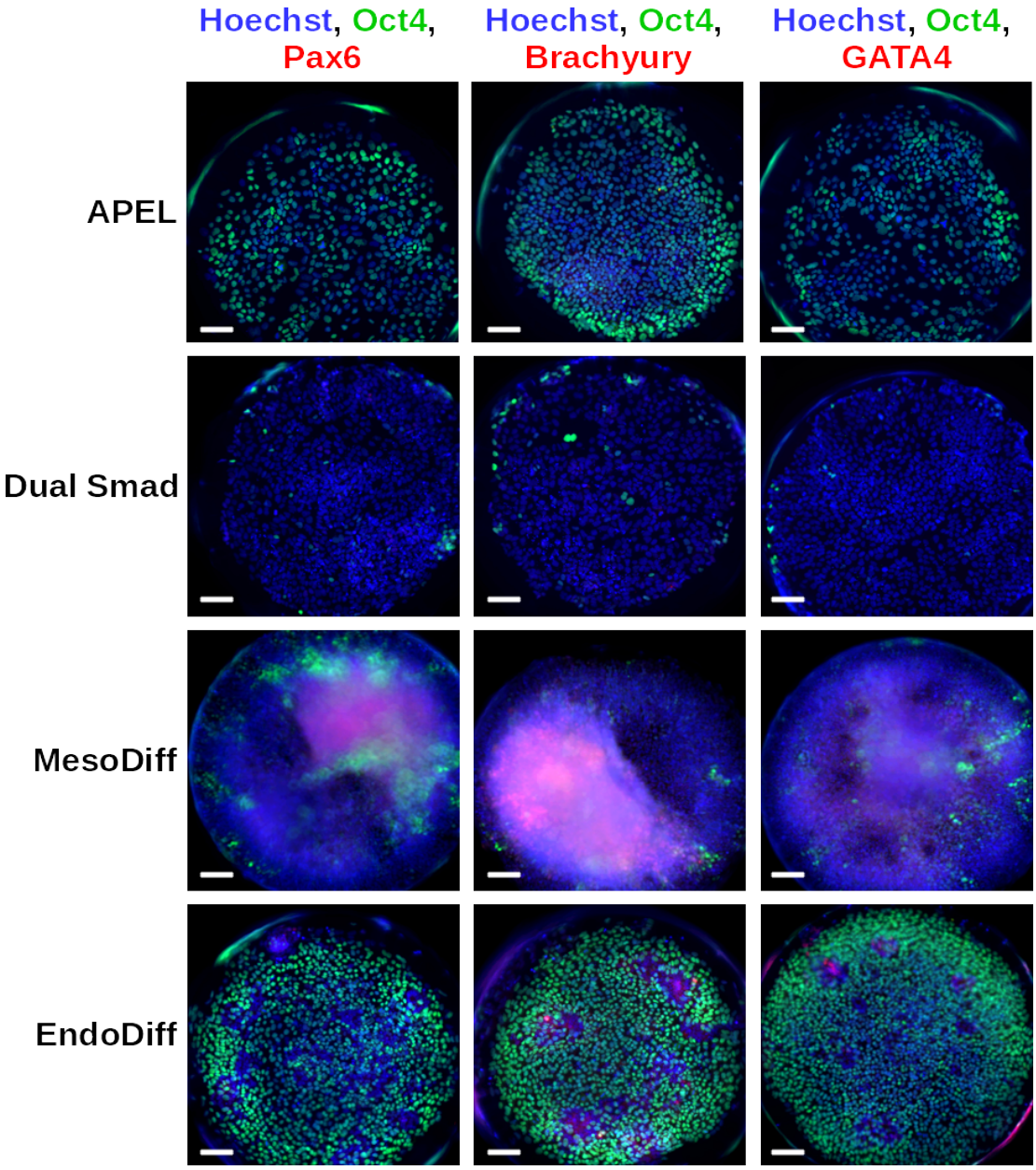

Figure 3.4. Sample images for ImmMaster media/marker pairings at day 9. Scale bars = 100 μm.

that day nine may have been too late to catch any emerging. The MesoDiff condition also saw a striking drop in Oct4 expression by day nine, with large increases in the amount of cells either single or double staining for a lineage marker. Cell density was exceedingly high, making cells difficult to

individually distinguish, a likely result of continued expansion of the cluster-like structures witnessed at day five. The EndoDiff condition continued to show general maintenance of Oct4 throughout the culture, with what appeared to be remnants of the small, dense cell clusters seen at day five.

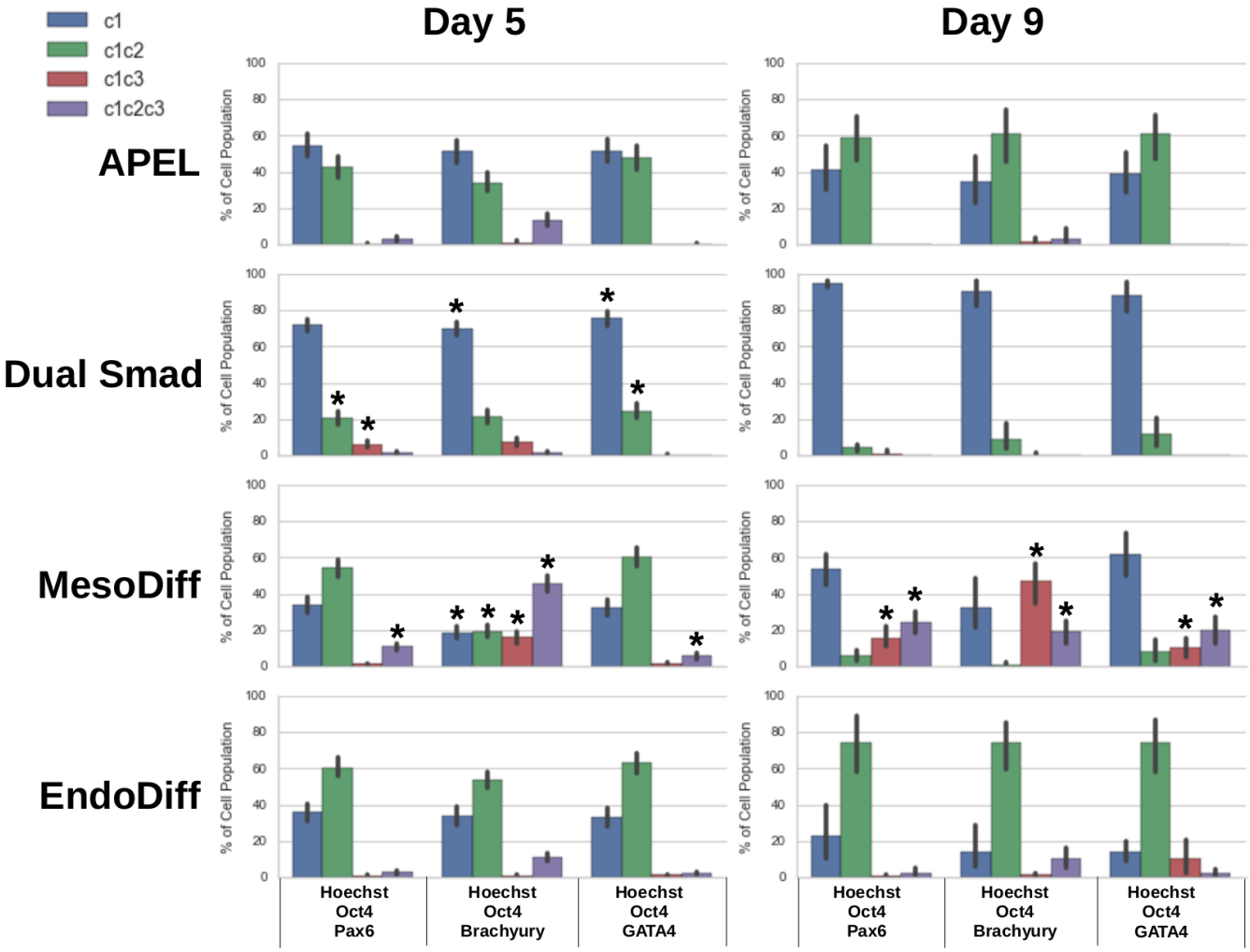

Figure 3.5. Control media marker staining. Expressed as the percent of the total cell population for reach condition. "*" indicates a significant difference compared to APEL on that day, $p < 0.05$, using the Mann-Whitney U test with Holmes Bonferroni correction.

3.4 Object-level Feature Correlations

Prompted by the distinct morphological features emerging in samples from each of the media conditions (Fig. 3.3-4), steps were taken to extract structural attributes from identified cell objects and cross-correlate them with staining information. From very high level perspectives, either globally (Fig. 3.6) across all images, or across all images within a given media or primary stain, there was not found to be any particularly significant cross-over between intensity- and structural-based features.

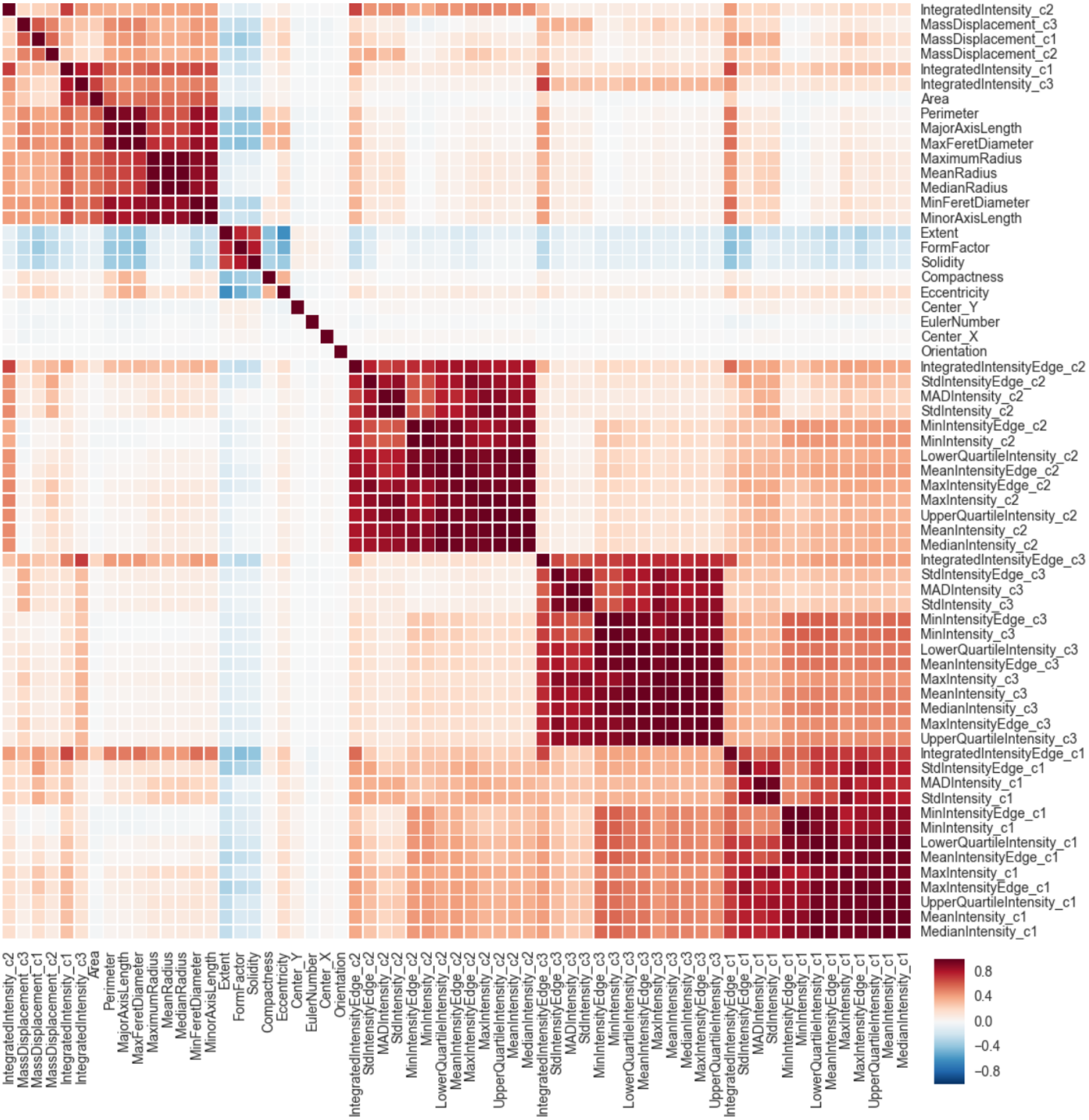

Figure 3.6. Correlation of object intensity and structural features. Matrix displaying the Pearson correlation coefficients across a subset of object metrics collected across all images during CellProfiler feature extraction.

This is not entirely surprising, however, since many of the distinguishing features discussed earlier are multi-cellular phenomenon resulting from heterogeneity in cell density that must necessarily take into account cells' positions and relationships to each other. While there are modules in CellProfiler that can get at cell neighbor information, they are very computationally intensive, and even with the purpose-built machines on hand, the time to pass all data through the pipeline would have been impractical.

Subsequent follow-up was carried out looking at object-level feature correlations with object classification, as indicated by marker staining. A random forest classifier implementation from Python's sklearn module was employed to attempt to machine learn relationships between individual cell objects' morphological parameters and staining identity. Unfortunately, predictive accuracy greater than 50% was not readily achievable across a variety of random forest parameterizations and training / test set splits. Two factors were identified as likely contributors to this impasse.

On one hand, the relative proportions of cell objects from each staining identity were grossly out of proportion within the training set. For particularly rare cell types, like GATA4+ nuclei, the algorithm essentially ignored them, because their relative impact on the final scoring was negligible. This might be addressable through implementation of a customized weighting scheme within the scoring algorithm, though the most robust solution would be to refine the experimental parameters to provide for more robust generation of higher numbers of the rarer cell types.

On the other hand, all stains used here were nuclear. While this is advantageous for counting and co-localization analysis, from a morphometric perspective nuclei tend to be much less expressive than what might be seen with cytoplasmic stains. Hence, future attempts at utilizing morphometric inputs for classification tasks would strongly benefit from the incorporation of a cytoplasmic or cytoskeletal stain (like actin) to better resolve the diversity of states the full cell body can assume in response to external stimuli.

3.5 Outlook and Perspectives

In a span of approximately three weeks, the ImmMaster experiments generated 25,536 images, totaling 120 GB of raw pixel data. Over a subsequent period of one week, CellProfiler extracted features from 25,203,929 identified cell objects in these images, yielding a final HDF5 feature database totaling just over 15 GB.

From a proof of concept standpoint, there were a number of valuable lessons learned for future experimental designs. Efforts to optimize immunostaining proved to be worthwhile, with staining robustness far surpassing initial attempts. To control for intrinsic structural errors, randomization would be advisable in future chip layouts and printing schemes. In addition, while a day nine endpoint for hiPSC differentiation might provide richer outcomes in terms of marker expression, growth of the cells for that long of a duration can lead to crowding issues on chip that complicate quantification. This is particularly evident in the MesoDiff conditions presented in Fig. 3.4, where the mitogenic potency of the media has forces the cells into largely indecipherable blobs. To account for this, special accommodations need to be made at print time to use adjusted seeding densities.

Even though this experiment examined a relatively small set of a few control media and antibody combinations, analysis has only just begun to scratch the surface of insights that can be gleaned from the rich image and feature sets. Traditional cell proliferation metrics and cell marker expressions were derived, shedding light on the dynamics of pluripotency maintenance and lineage commitment under each of the treatment conditions. Visual inspection of images led to identification of a number of morphological features characteristic of the different cultures, representing a ripe avenue for further investigation.

3.6 Materials and Methods

Cell Culture
Passage 30-40 TCTF Human induced pluripotent stem cells (hiPSCs) were maintained on tissue culture treated polystyrene 6 well plates (BD Falcon) coated as per manufacturer recommendation with hESC qualified Matrigel (Corning). mTeSR growth medium (StemCell Tech) supplemented with 1x mTeSR supplement (StemCell Tech) was refreshed daily. Cells were passaged 1:6 every 3 days using ReLeSR (StemCell Tech) and the manufacturer recommended protocol. All media was supplemented with 0.5% penicillin/streptomycin (Fisher).

Pillar/Well Chip Printing
Printing protocols for the pillar and well chips (Samsung) were prepared and validated in advance. The day before printing, pillar chips were stamped into well chips printed at 4ºC with hESC qualified Matrigel diluted as per manufacturer recommendations. These pillar/well chips were incubated overnight at 37ºC in humidity chambers. The following day, new well chips were printed with APEL (StemCell Tech) and stored until needed in humidity chambers at 37ºC. Cells were immediately passaged with ReLeSR, as above, re-suspended to ~3E6 cells/ml in APEL, and loaded onto the source plate. The prepared Matrigel-coated pillar chips were printed with cells and stored face-up in humidity chambers at 37ºC to allow cells to settle and adhere. After 20 minutes, pillar chips were stamped into the freshly prepared APEL well chips and returned to humidity chambers at 37ºC. All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution (Sigma).

Pillar/Well Chip Culture
On each subsequent day for five days, cells were stamped into freshly prepared well chips containing the appropriately printed APEL, DualSmad (APEL + 10 uM SB431542 (Tocris) + 100 nM LDN-193189 (Stemgent))[90], MesoDiff (StemCell Tech), or EndoDiff (StemCell Tech) media. All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution.

Immunostaining
Chips were stained as per Note 1.1 with various primary antibody combinations (Table 3.2). Materials used included 2% PFA (Santa Cruz), PBS (Gibco), Triton X-100 (Sigma), donkey serum (Sigma), 1:1000 Hoechst 33342 (Life Technologies), 1:200 Ms anti-Nestin (BD), 1:50 Ms anti-Oct4 (Santa Cruz), 1:50 Ms anti-GATA4 (Santa Cruz), 1:50 Rb anti-Oct4 (Santa Cruz), 1:200 Rb anti-Pax6 (BioLegend), 1:50 Rb anti-Brachyury (Santa Cruz), 1:50 Rb anti-GATA4 (Santa Cruz), 1:5000 Ms IgG (Abcam), 1:50 Rb IgG (Abcam), 1:500 Dk anti-Ms 488 (Jackson Immuno), 1:500 Dk anti-Rb 594 (Jackson Immuno). All solutions printed in well chips were supplemented with 1:10000 anti-foam C solution.

Data Analysis and Statistics
Images were re-formatted using Python, and a variety of image-level and object-level features were mined using CellProfiler[89]. HDF5 databases of said features were collated, analyzed, and plotted using Python. Statistical significance was calculated using the Mann-Whitney U test with Holmes Bonferroni correction in Python.

Chapter 4: Dose-Response Screening

To test the platform's performance in a realistic application, a new experiment was planned focusing on extracting dose-response data across a five log concentration range for two sets of compounds.

4.1 Dose-Response Experimental Design

The first panel, termed "Diff," consisted of 34 small molecules and growth factors with generally well-defined mechanisms of action on various cellular pathways (Table 4.1).

Table 4.1. "Diff" panel agents and information.

Agent	Target	ED50 / EC50	Chosen Middle Dose (mM)	Supplier
CHIR99021	GSK-3B	3 uM	0.0001	Stemgent
iCRT5	B-Catenin	20 nM	0.00002	Biovision
SB431542	Smad 2/3	100 nM	0.000025	Tocris
LDN193189	Smad 1/5/8	5 - 30 nM	0.00002	Stemgent
DAPT	Notch	20 nM	0.00002	SelleckChem
PD173074	FGF-R	5 - 21.5 nM	0.000001	StemCell Tech
OAC-1	Oct4	50 nM	0.00005	StemCell Tech
Prostaglandin E2	EP1,2,3,4	1 – 10 nM	0.00001	StemCell Tech
SP600125	JNK 1,2,3	110 nM	0.0001	StemCell Tech
SB202190	p38 MAPK	50 – 100 nM	0.00005	StemCell Tech
Go6983	PKC	5 – 75 nM	0.00001	StemCell Tech
Rosiglitazone	PPARy	30 – 100 nM	0.00005	StemCell Tech
ROCKi	Rho Kinase	10 uM	0.0001	Cell Guidance Systems
Purmorphamine	SHH	1.5 uM	0.0001	Stemgent
Cyclopamine	SHH	100 nM?	0.000025	CalBioChem
PD0325901	MEK/ERK	0.33 nM	0.0000005	StemCell Tech
Indolactam	PKC	10 nM	0.00001	StemCell Tech
PS-48	PI3k	10 uM	0.001	StemCell Tech
Ly294002	PI3k	1.4 uM	0.0001	StemCell Tech
Pifithrin-mu	p53	1 mM	0.0005	StemCell Tech
Pyrintegrin	Integrin	1 uM?	0.0001	StemCell Tech
Sinomenine	NFKB	50 uM	0.0005	StemCell Tech
ID-8	DYRK	500 nM	0.00005	StemCell Tech
AICAR	AMPK	1 mM	0.0005	StemCell Tech
IDE2	Activin / TGF-B	223 nM	0.0001	StemCell Tech
Trichostatin A	HDAC1	10 - 40 nM	0.00001	StemCell Tech
5-Azacytidine	DNMT	0.5 uM	0.0001	StemCell Tech
RepSox	TGF-B / Alk5	4-23 nM	0.00001	StemCell Tech
3-Deazaneplanocin A	EZH2 Lysine MT	1 uM	0.0001	StemCell Tech
Activin A	Smad 2/3	2 ng/ml	5 ng/ml	Peprotech
TGF-B1	Smad 2/3	0.2 ng/ml	5 ng/ml	R&D
BMP4	Smad 1/5/8	15 ng/ml	5 ng/ml	R&D
FGF2	FGFR	0.5 ng/ml	2 ng/ml	Peprotech
EGF	EGFR	0.1 ng/ml	1 ng/ml	Gold Biotech

The second panel, “Dev,” consisted of 34 agents of known and unknown teratogenicity (Table 4.2)[90].

Table 4.2. “Dev” panel agents and information.

Agent	Chosen Middle Dose (mM)	Supplier	Classification
Hydroxyurea	0.0001	Sigma	Strong
6-Aminonicotinamide	0.0001	Sigma	Strong
5-Bromo-2'-deoxyuridine	0.0001	Sigma	Strong
Methotrexate	0.0001	Sigma	Strong
Thalidomide	0.0001	Sigma	Strong
Retinoic Acid	0.00001	Sigma	Strong
EtOH	0.01%	VWR	Strong
Valproic Acid	0.001	StemCell Tech	Weak
Salicylic acid sodium salt	0.0001	Sigma	Weak
Dimethadione	0.0001	Sigma	Weak
Methoxyacetic acid	0.0001	Sigma	Weak
LiCl	0.0001	Sigma	Weak
Boric Acid	0.0001	Sigma	Weak
Dimethyl pthalate	0.0001	Sigma	Non
D-(+)-Camphor	0.0001	Sigma	Non
Diphenhydramine hydrochloride	0.0001	Sigma	Non
Penicillin G sodium salt	0.0001	Sigma	Non
Saccharin sodium hydrate	0.0001	Sigma	Non
Isoproterenol	0.0001	Sigma	?
DMSO	0.01%	Thermo Fisher	?
Cupric Sulfate	0.0001	Sigma	?
Ferric Chloride	0.0001	Sigma	?
Ascorbic Acid	0.0001	Sigma	?
Zinc Sulfate	0.0001	Sigma	?
NaCl	0.0001	Thermo Fisher	?
Histidine	0.0001	Thermo Fisher	?
Asparagine	0.0001	Thermo Fisher	?
Tyrosine	0.0001	Thermo Fisher	?
Glutamine	0.0001	Thermo Fisher	?
Aspartic Acid	0.0001	Thermo Fisher	?
Glycine	0.0001	Thermo Fisher	?
Leucine	0.0001	Thermo Fisher	?
Cadmium Sulfate	0.0001	Sigma	?
Riboflavin	0.0001	Thermo Fisher	?

In both cases, dose range values were chosen based upon values reported in the product or scientific literature, for stem cell targets wherever possible.

Each run of the Dose-Response experiment consisted of four chips for the Diff panel and four chips for the Dev panel, for 2128 pillar/well samples each. Three replicates of each agent at doses spanning five logs were printed per chip. Five control media were tested, with four to six replicates printed per chip (Table 4.3).

Table 4.3. Control media used in dose-response experiments.

Alias	Base Media	Additives
APEL	APEL	-
mTeSR	mTeSR	5x mTeSR Supplement, diluted to 1x
DualSmad	APEL	100 nM LDN-193189 , 10 µM SB431542
MesoDiff	StemDiff Mesoderm Induction Media	-
EndoDiff	StemDiff Definitive Endoderm Kit	-

Chips from both panels were stained with one of four primary antibody combinations (Table 4.4).

Table 4.4. Primary antibody combinations used in dose-response experiments.

Alias	Alexa 488	Alexa 594
Hoechst_Oct4_Pax6	Ms Anti-Oct4	Rb Anti-Pax6
Hoechst_Oct4_Brachyury	Ms Anti-Oct4	Rb Anti-Brachyury
Hoechst_Oct4_GATA4	Ms Anti-Oct4	Rb Anti-GATA4
Hoechst_IgG_IgG	Ms IgG	Rb IgG

Informed by the ImmMaster experimental results, chip layouts were moved to a randomized scheme to help account for potential printing errors. Cells would be cultured in 2D on Matrigel. Due to the crowding observed at day nine in the ImmMaster experiment and the necessities of time, a day five end point was chosen to fix and stain. Images would be examined from the perspectives of both lineage commitment, as in the ImmMaster experiment, and cell proliferation over the dose range.

4.2 Agent Effects on hiPSC Lineage Commitment

APEL was chosen to be the base media in which agents would be administered for its generally non-biased effects on cell fate. Absent any additives, cells grown in APEL in the Dev (Fig. 4.1a) and Diff (Fig. 4.1b) experiments demonstrated similar baseline marker expression, with a majority of cells not staining strongly for either robust pluripotency or any particular lineage commitment.

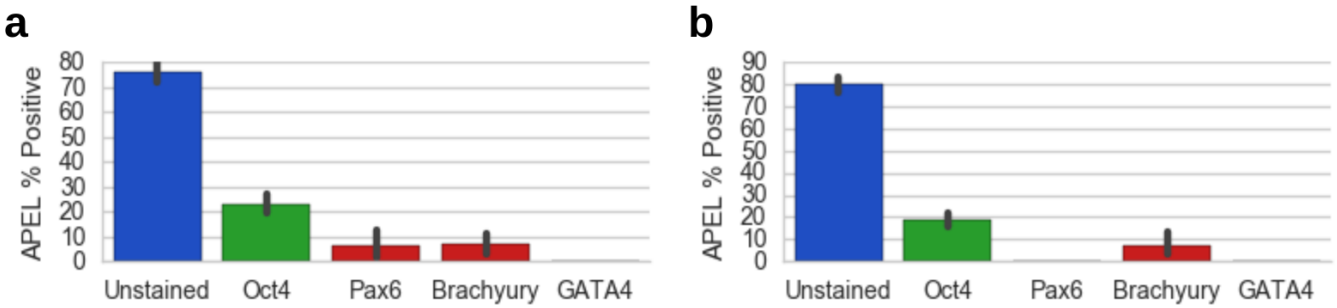

Figure 4.1. APEL antibody staining benchmarks. Percentage of cell objects in the APEL controls for (**a**) "Dev" and (**b**) "Diff" panels with unstained, Oct4+, Pax6+, Brachyury+, and GATA4+ identification.

Among the four control media run in both panels (Fig. 4.2), outcomes were largely in agreement with expectation. In mTeSR, a gold standard for maintenance of pluripotent cells, Oct4 was strongly upregulated, with most of the shifted population seeming to derive from the unstained nuclei in the base APEL condition. In the DualSmad condition Oct4 expression was downregulated, as had been observed in earlier experimental data, resulting in variable increases in unstained and lineage-staining cells.

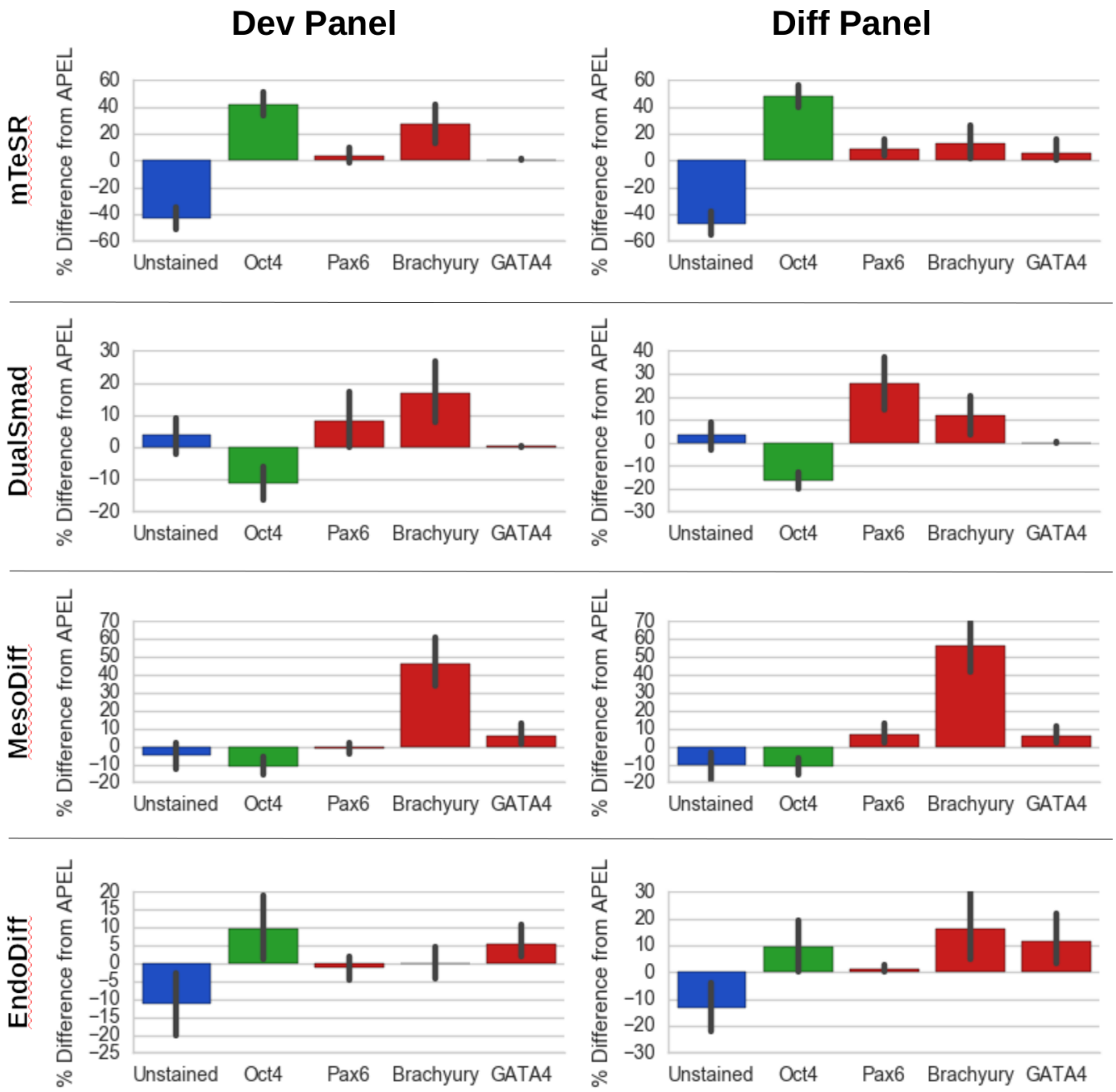

Figure 4.2. Control media antibody staining benchmarks relative to APEL. Values are presented as percentage point difference from APEL.

The MesoDiff condition, traditionally a strong inducer of differentiation, saw a large increase in the proportion of cells staining positive for the mesoderm marker Brachyury, as would be expected. Finally, in the EndoDiff condition, no particularly strong outcome was seen by day five in terms of marker expression.

Among the 68 agents tested between the Diff and Dev panels, a variety of outcomes were observed across the dose ranges. Perhaps the most common response seen was the loss of Oct4 staining over the increasing dosage range. Example data from the retinoic acid and SB431542 samples are shown (Fig. 4.3). Retinoic acid, a strong inducer of differentiation commonly used in applications like neural stem cell differentiation, very potently abolished Oct4 staining by day five. Similarly, SB431542, an inhibitor of TGF-B pathway signaling, had similar effects of Oct4 repression, which conforms with TGF-B's identified role in maintenance of pluripotency[91].

At the opposite end of the spectrum, a handful of agents were seen to maintain and enhance the pluripotency phenotype in treated cultures (Fig. 4.4). While just falling below the $p < 0.05$ cutoff, Activin A, a signaling molecule active within the TGF-B family signaling pathways, demonstrated an upward trend in Oct4 staining with increasing dose. Fibrobrast growth factor 2 (FGF2), commonly found as an additive in many stem cell maintenance media, had even stronger effects. Both have been previously published to have maintenance effects in pluripotent stem cells[92].

While rarer, there were also a few agents that enhanced departure from pluripotency towards any of a number of downstream differentiated fates (Fig. 4.5). Ly294002, a broad-spectrum inhibitor, had a correspondingly broad effect on cell states, creating measured, if variable, increases in all three germ-line markers. CHIR99021, a Wnt pathway activator, was found to strongly disrupt the Oct4+ cell population. This falls in line with previous reports detailing its activity prompting mesoderm differentiation[93].

There were also some unexpected results that came out of the screen. For instance, OAC-1, marketed as a small molecule that activates Oct4 expression and helps to maintain pluripotency, was found to have the complete opposite effect in this experimental system (Fig. 4.6). Over all dose ranges, it had a strong Oct4 destabilizing effect, forcing many cells into an unstained state. This defies intuitive explanation. In general, data exhibited more variability than would be ideal for running screens with three to four replicates to identify leads for various biological questions. Moreover, systematic error with on chip dilutions employed during printing confounded results for the second and fourth concentrations in each dose range. These issues will be discussed further in 4.4.

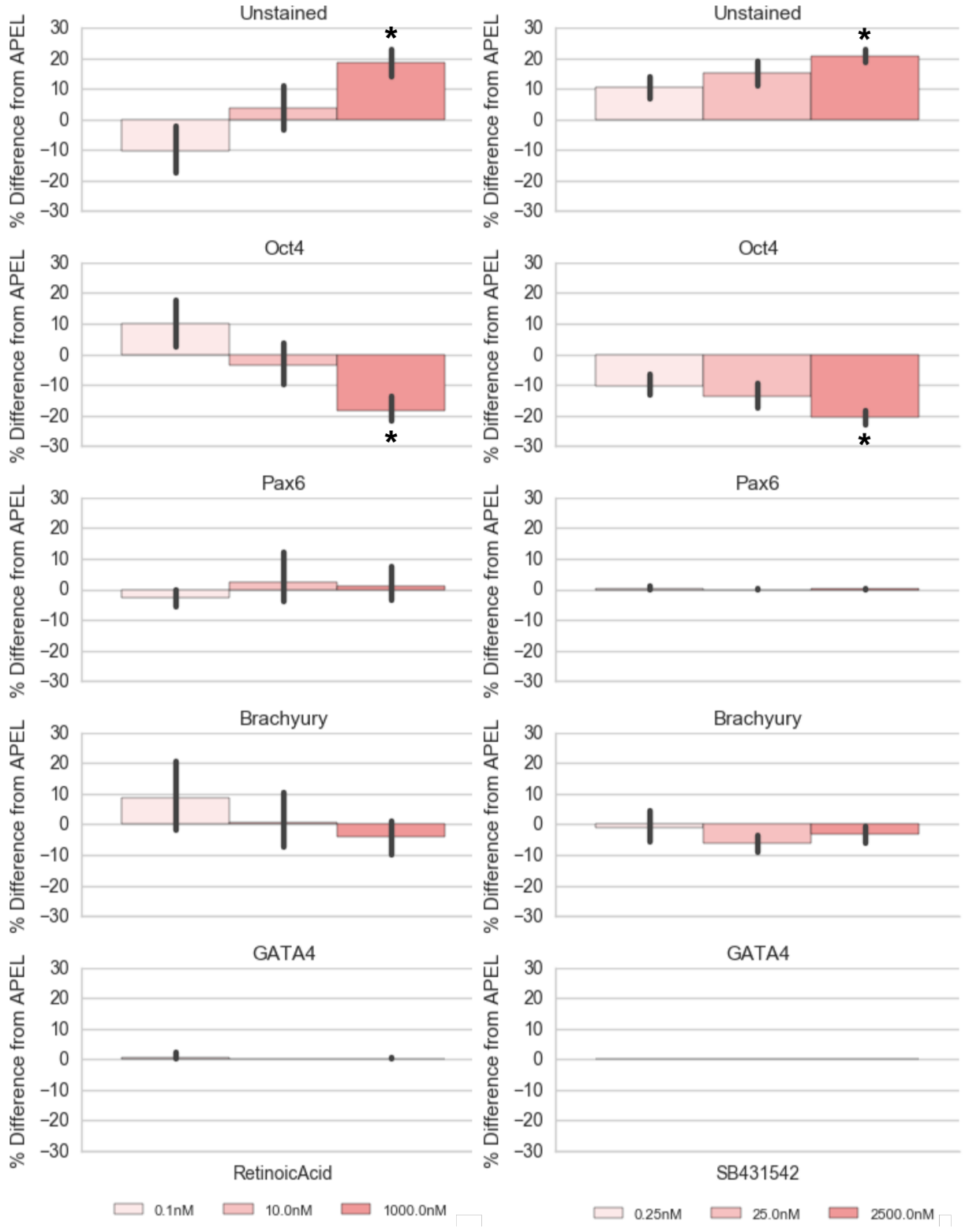

Figure 4.3. Sample of agents increasing the proportion of unstained cells. Values are presented as percentage point difference from APEL. "*" indicates difference from the APEL benchmark at $p < 0.05$ using Mann-Whitney U test with Holmes Bonferroni correction.

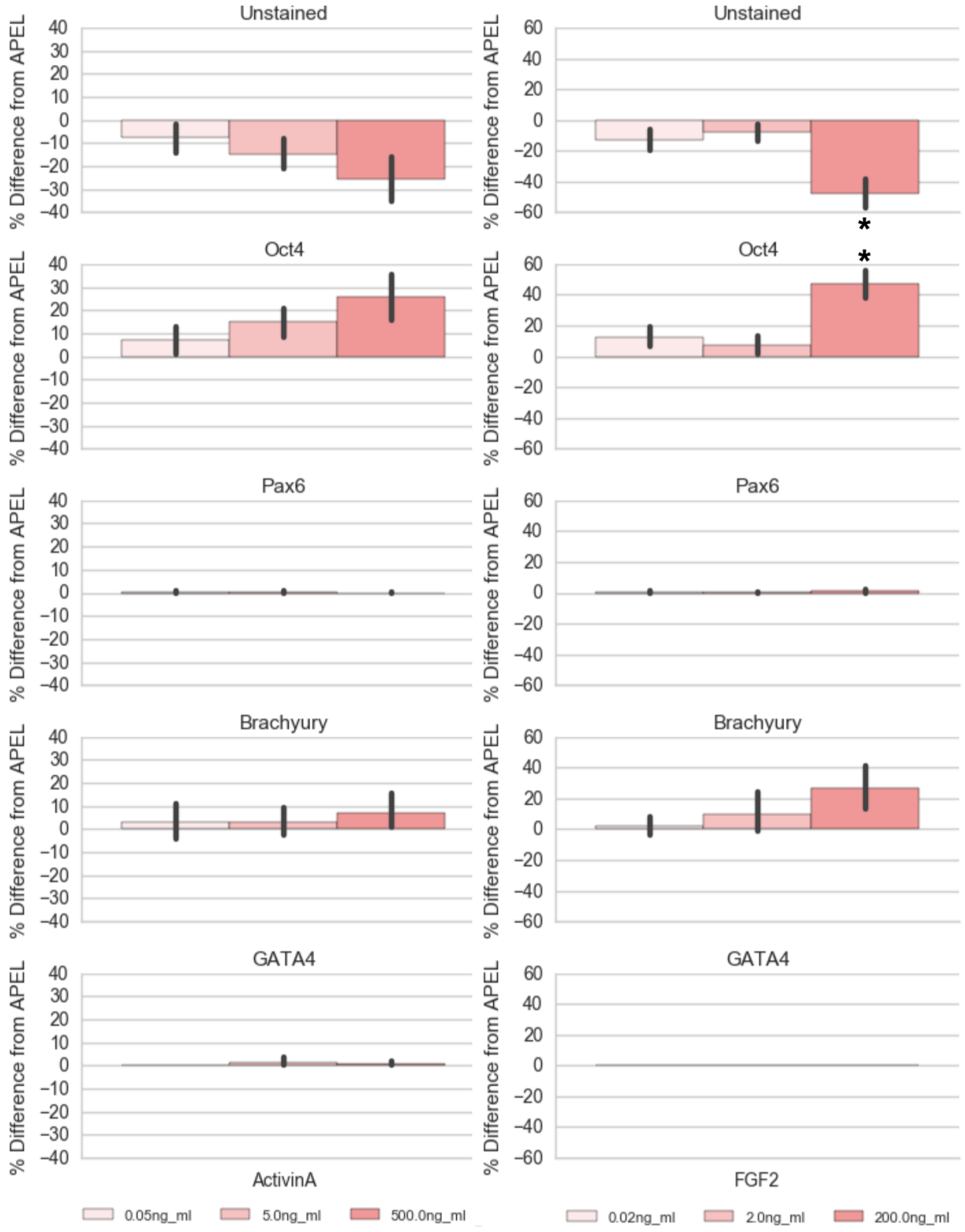

Figure 4.4. Sample of agents increasing the proportion of Oct4+ cells. Values are presented as percentage point difference from APEL. "*" indicates difference from the APEL benchmark at $p < 0.05$ using Mann-Whitney U test with Holmes Bonferroni correction.

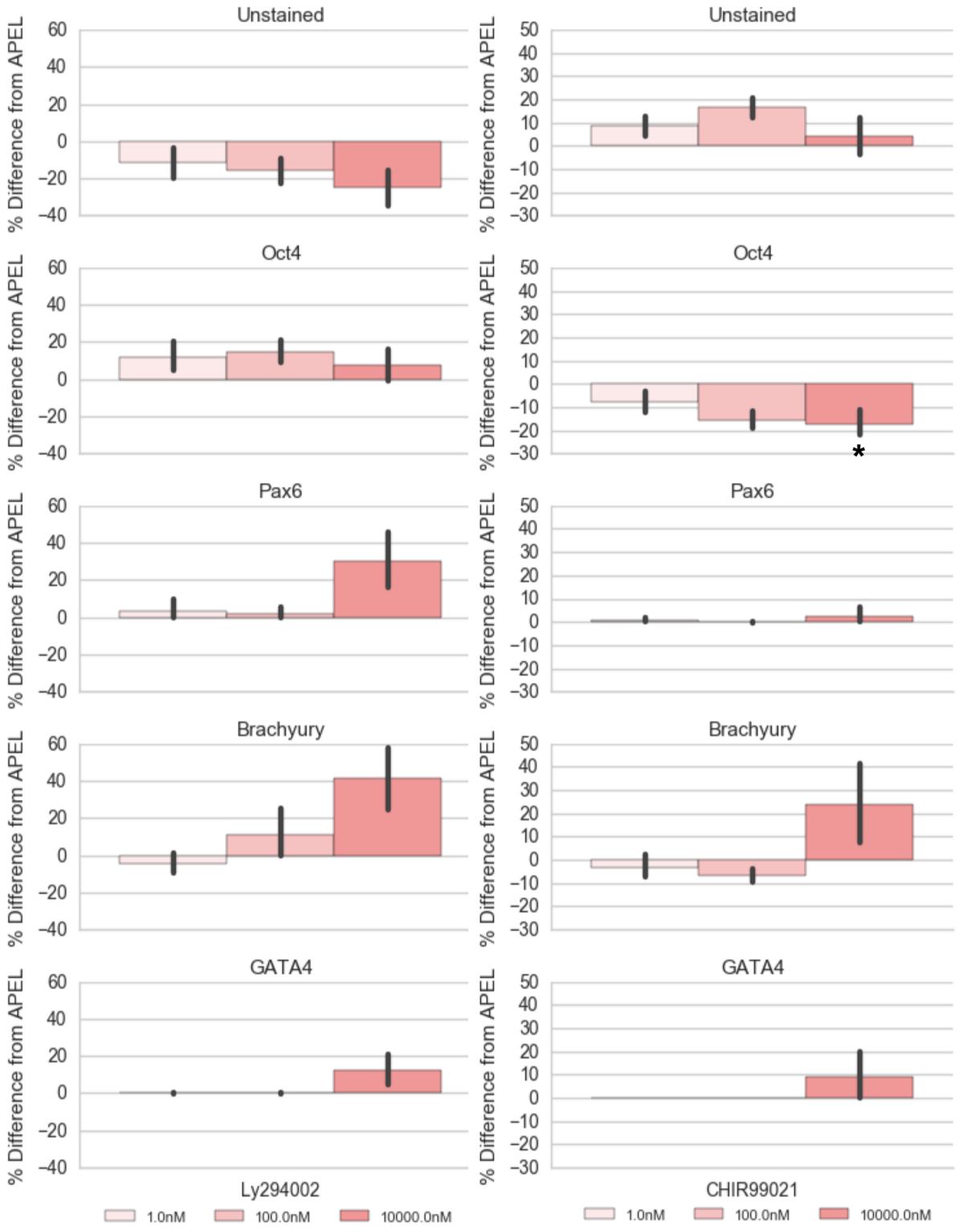

Figure 4.5. Sample of agents increasing the proportion of differentiating cells. Values are presented as percentage point difference from APEL. "*" indicates difference from the APEL benchmark at $p < 0.05$ using Mann-Whitney U test with Holmes Bonferroni correction.

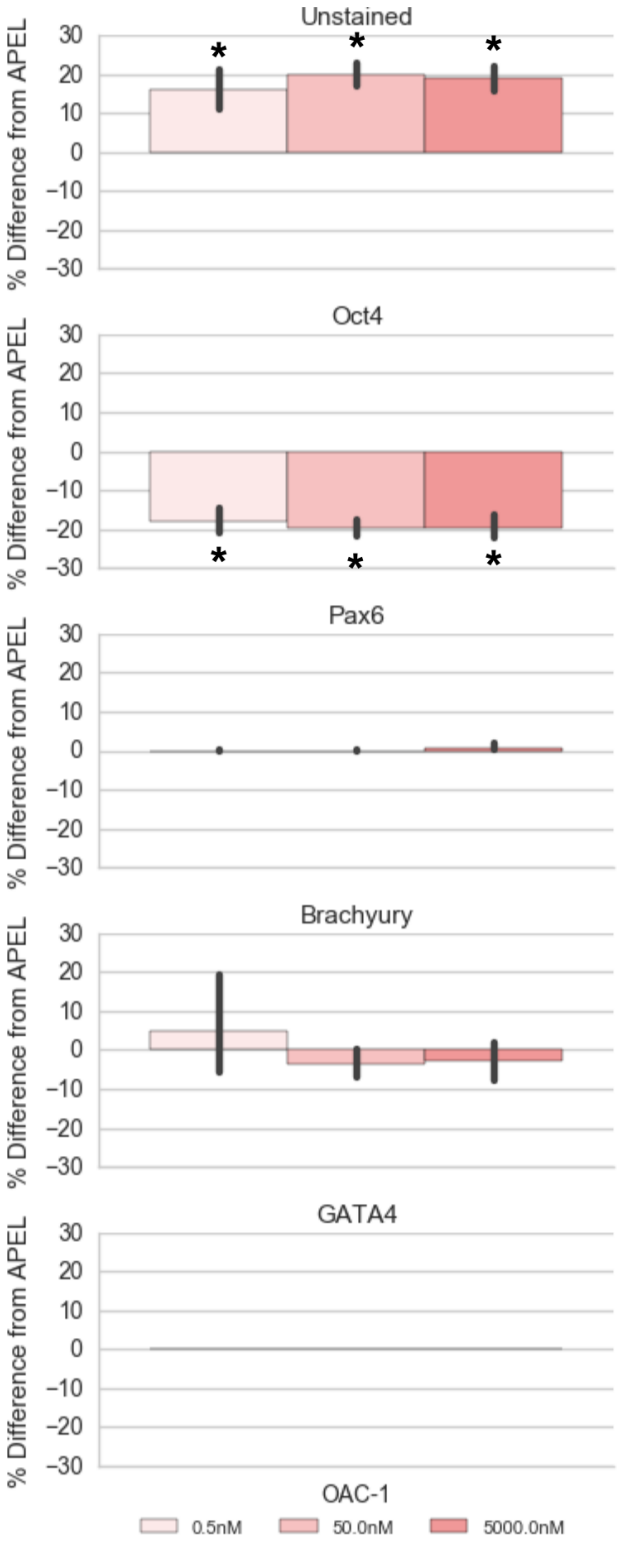

Figure 4.6. Depression of Oct4+ cells under OAC-1. Values are presented as percentage point difference from APEL. "*" indicates difference from the APEL benchmark at $p < 0.05$ using Mann-Whitney U test with Holmes Bonferroni correction.

4.3 Cell Proliferation Dose-Response Curves

Toxicology screening is an application area desperately in need of high-throughput augmentation. Many screens, both for traditional toxicity and developmental toxicity alike, rely heavily on “gold standard” animal models that are expensive, low-throughput, often discordant with each other, and, in worst case scenarios, can be misleading about outcomes in humans[90,94]. As a result, *in vitro* platforms that can address these concerns using human cell models in rapid, high-throughput pipelines stand to greatly impact the field.

Demonstration of an effective dose-response assay on this platform proved difficult, however. As can be seen in sample data for two tumor suppressing drugs (Fig. 4.7), the potential is there, and rough curves can be fit. However, the variability in the data acquired so far is much too high to make any strong conclusions, and will require further refinements to the platform to tighten the spread. For approximately 50% of agents tested, this was problematic enough that a sigmoid curve could not even be fit. Obviously this was exacerbated by general difficulties with the second and fourth doses, as mentioned in 4.2, reinforcing the need for further pipeline refinement to acquire crisp results.

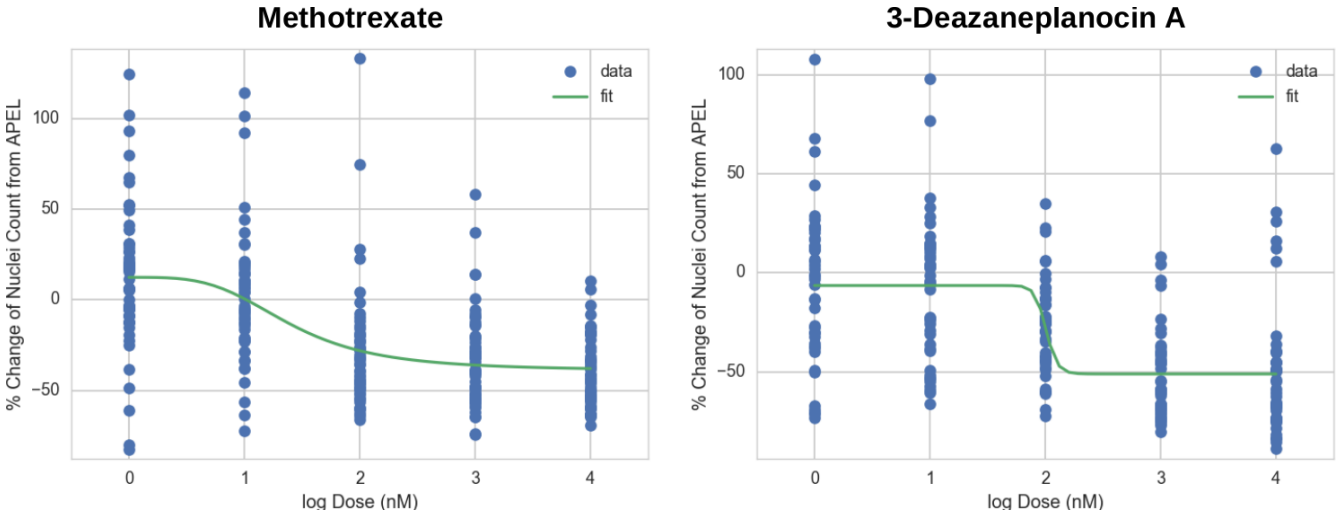

Figure 4.7. Sample dose-response curves. While a number of the compounds, like the tumor-suppressing drugs shown here, did display decreases in final cell numbers with increasing dose, in general the spread on cell number was too great to conclude anything substantial.

4.4 Outlook and Perspectives

In a span of approximately three weeks, the Dose-Response experiments generated 65,436 images, totaling to 360 GB of raw pixel data. Over a subsequent period of two weeks, CellProfiler extracted features from 56,238,060 identified cell objects in these images, yielding a final HDF5 feature database totaling just over 21.5 GB.

From the collected data, information could be gleaned about agent effects over a five log dose range on cellular pluripotency or lineage commitment, as indicated by the antibody stains, and agent effects on cell proliferation. There were a number of encouraging results matching what would be expected from reports in the product and scientific literature. Activin A and FGF2 both demonstrated expected stabilizing effects on the pluripotent state. Ly294002 and CHIR99021 both exhibited measurable, if not quite significant upregulation of various lineage markers characteristic of pluripotent cell differentiation. Results were adequately sensitive to observe maintenance or loss of the pluripotent state, the induction of specific differentiated lineage commitment, and perturbations in cell proliferation.

However, the platforms performance in these screens was not without caveat. Of note, peculiarities were observed in the second and fourth concentrations of the dose range. These are hypothesized to be attributable to systematic error inherent in how they were prepared on chip. While the first, third, and fifth concentrations were printed directly from the source plate, the second and fourth concentrations were prepared using the printer to carry out a 1/10 dilution with the corresponding one log higher stock on chip. This was critical to cutting down on printing overhead time to avoid chip drying. However, it appears that these dilutions printed with sub-optimal quality, potentially due to the removal of a preconditioning equilibration step and the order of magnitude difference in volume printed in these steps. As such, only the first, third, and fifth doses were at all reliable.

Moreover, additional issues emerged during scale-up. Due to the greatly increased complexity of the print, print times began to exceed 2 hours, which posed a serious drying problem for chips exposed for the duration. Steps to improve humidification, optimize all possible printing parameters, and remove overhead were taken to increase speed, resulting in an eventual reduction of run time to ~90 minutes. Even with these optimizations, the long duration also increased the incidence of chip contamination, particularly with a certain filamentous microorganism suspected to originate from the humidification apparatus.

In summary, while the platform certainly holds promise for screens in the mold of dose-response assays, a number of optimizations spanning print fidelity and speed will be necessary to achieve a power reasonable enough to fully leverage the throughput capacity of the platform. In combination with the cataloging and analysis of rich morphological feature data, this should enable much more reliable screening, as well as a firm foundation for future machine learning aimed at predicting developmental toxicity in high-throughput.

4.5 Materials and Methods

Cell Culture

Passage 35-45 TCTF Human induced pluripotent stem cells (hiPSCs) were maintained on tissue culture treated polystyrene 6 well plates (BD Falcon) coated as per manufacturer recommendation with hESC qualified Matrigel (Corning). mTeSR growth medium (StemCell Tech) supplemented with 1x mTeSR supplement (StemCell Tech) was refreshed daily. Cells were passaged 1:6 every 3 days using ReLeSR (StemCell Tech) and the manufacturer recommended protocol. All media was supplemented with 0.5% penicillin/streptomycin (Fisher).

Pillar/Well Chip Printing
Printing protocols for the pillar and well chips (Samsung) were prepared and validated in advance. The day before printing, pillar chips were stamped into well chips printed at 4ºC with hESC qualified Matrigel diluted as per manufacturer recommendations. These pillar/well chips were incubated overnight at 37ºC in humidity chambers. The following day, new well chips were printed with APEL (StemCell Tech) and stored until needed in humidity chambers at 37ºC. Cells were immediately passaged with ReLeSR, as above, re-suspended to ~3E6 cells/ml in APEL, and loaded onto the source plate. The prepared Matrigel-coated pillar chips were printed with cells and stored face-up in humidity chambers at 37ºC to allow cells to settle and adhere. After 20 minutes, pillar chips were stamped into the freshly prepared APEL well chips and returned to humidity chambers at 37ºC. All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution (Sigma).

Pillar/Well Chip Culture
On each subsequent day for five days, cells were stamped into freshly prepared well chips containing the appropriately printed media conditions (Table 4.1-3). All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution. Small molecule stocks were stored for no more than 5 days in the dark at 4ºC before being refreshed.

Immunostaining
Chips were stained as per Note 1.1 with various primary antibody combinations (Table 4.4). Materials used included 2% PFA (Santa Cruz), PBS (Gibco), Triton X-100 (Sigma), donkey serum (Sigma), 1:1000 Hoechst 33342 (Life Technologies), 1:50 Ms anti-Oct4 (Santa Cruz), 1:200 Rb anti-Pax6 (BioLegend), 1:50 Rb anti-Brachyury (Santa Cruz), 1:50 Rb anti-GATA4 (Santa Cruz), 1:5000 Ms IgG (Abcam), 1:50 Rb IgG (Abcam), 1:500 Dk anti-Ms 488 (Jackson Immuno), 1:500 Dk anti-Rb 594 (Jackson Immuno). All solutions printed in well chips were supplemented with 1:10000 anti-foam C solution.

Data Analysis and Statistics
Images were re-formatted using Python, and a variety of image-level and object-level features were mined using CellProfiler[89]. HDF5 databases of said features were collated, analyzed, and plotted using Python. Significance of results was tested using the Mann-Whitney U test with Holmes Bonferroni correction in Python.

Chapter 5: Pair-wise Interactivity Screening

One of the particular advantages of this pipeline is its ability to readily enumerate on chip conditions containing multiple agents. Given that agent interactivity, as manifested in additive, antagonistic, and synergistic effects, can be something difficult to study at any sort of scale with conventional techniques, and that such multi-factorial contexts can be critically important to cellular decision making, an additional set of experiments were designed and executed focusing on these aspects.

5.1 Pair-wise Experimental Design

The “Diff” panel, as per Chapter 4, consisted of 34 small molecules and growth factors with generally well-defined mechanisms of action on various cellular pathways (Table 5.1).

Table 5.1. “Diff” panel agents and information.

Agent	Target	ED50 / EC50	Dose (mM)	Supplier
CHIR99021	GSK-3B	3 uM	0.001	Stemgent
iCRT5	B-Catenin	20 nM	0.0002	Biovision
SB431542	Smad 2/3	100 nM	0.00025	Tocris
LDN193189	Smad 1/5/8	5 - 30 nM	0.0002	Stemgent
DAPT	Notch	20 nM	0.0002	SelleckChem
PD173074	FGF-R	5 - 21.5 nM	0.00001	StemCell Tech
OAC-1	Oct4	50 nM	0.0005	StemCell Tech
Prostaglandin E2	EP1,2,3,4	1 – 10 nM	0.0001	StemCell Tech
SP600125	JNK 1,2,3	110 nM	0.001	StemCell Tech
SB202190	p38 MAPK	50 – 100 nM	0.0005	StemCell Tech
Go6983	PKC	5 – 75 nM	0.0001	StemCell Tech
Rosiglitazone	PPARy	30 – 100 nM	0.0005	StemCell Tech
ROCKi	Rho Kinase	10 uM	0.001	Cell Guidance Systems
Purmorphamine	SHH	1.5 uM	0.001	Stemgent
Cyclopamine	SHH	100 nM?	0.00025	CalBioChem
PD0325901	MEK/ERK	0.33 nM	0.000005	StemCell Tech
Indolactam	PKC	10 nM	0.0001	StemCell Tech
PS-48	PI3k	10 uM	0.01	StemCell Tech
Ly294002	PI3k	1.4 uM	0.001	StemCell Tech
Pifithrin-mu	p53	1 mM	0.005	StemCell Tech
Pyrintegrin	Integrin	1 uM?	0.001	StemCell Tech
Sinomenine	NFKB	50 uM	0.005	StemCell Tech
ID-8	DYRK	500 nM	0.0005	StemCell Tech
AICAR	AMPK	1 mM	0.005	StemCell Tech
IDE2	Activin / TGF-B	223 nM	0.001	StemCell Tech
Trichostatin A	HDAC1	10 - 40 nM	0.0001	StemCell Tech
5-Azacytidine	DNMT	0.5 uM	0.001	StemCell Tech
RepSox	TGF-B / Alk5	4-23 nM	0.0001	StemCell Tech
3-Deazaneplanocin A	EZH2 Lysine MT	1 uM	0.001	StemCell Tech
Activin A	Smad 2/3	2 ng/ml	50 ng/ml	Peprotech
TGF-B1	Smad 2/3	0.2 ng/ml	50 ng/ml	R&D
BMP4	Smad 1/5/8	15 ng/ml	50 ng/ml	R&D
FGF2	FGFR	0.5 ng/ml	20 ng/ml	Peprotech
EGF	EGFR	0.1 ng/ml	10 ng/ml	Gold Biotech

The second panel, “Dev,” consisted of the same 34 agents as in Chapter 4 (Table 5.2).

Table 5.2. “Dev” panel agents and information.

Agent	Chosen Middle Dose (mM)	Supplier	Classification
Hydroxyurea	0.001	Sigma	Strong
6-Aminonicotinamide	0.001	Sigma	Strong
5-Bromo-2'-deoxyuridine	0.001	Sigma	Strong
Methotrexate	0.001	Sigma	Strong
Thalidomide	0.001	Sigma	Strong
Retinoic Acid	0.0001	Sigma	Strong
EtOH	0.10%	VWR	Strong
Valproic Acid	0.01	StemCell Tech	Weak
Salicylic acid sodium salt	0.001	Sigma	Weak
Dimethadione	0.001	Sigma	Weak
Methoxyacetic acid	0.001	Sigma	Weak
LiCl	0.001	Sigma	Weak
Boric Acid	0.001	Sigma	Weak
Dimethyl pthalate	0.001	Sigma	Non
D-(+)-Camphor	0.001	Sigma	Non
Diphenhydramine hydrochloride	0.001	Sigma	Non
Penicillin G sodium salt	0.001	Sigma	Non
Saccharin sodium hydrate	0.001	Sigma	Non
Isoproterenol	0.001	Sigma	?
DMSO	0.10%	Thermo Fisher	?
Cupric Sulfate	0.001	Sigma	?
Ferric Chloride	0.001	Sigma	?
Ascorbic Acid	0.001	Sigma	?
Zinc Sulfate	0.001	Sigma	?
NaCl	0.001	Thermo Fisher	?
Histidine	0.001	Thermo Fisher	?
Asparagine	0.001	Thermo Fisher	?
Tyrosine	0.001	Thermo Fisher	?
Glutamine	0.001	Thermo Fisher	?
Aspartic Acid	0.001	Thermo Fisher	?
Glycine	0.001	Thermo Fisher	?
Leucine	0.001	Thermo Fisher	?
Cadmium Sulfate	0.001	Sigma	?
Riboflavin	0.001	Thermo Fisher	?

In both cases, single doses were chosen near the higher end of the scale used in Chapter 4, reflecting the general elicitation of cellular response at the fourth dose and above.

Each run of the pair-wise experiment consisted of thirteen chips for the Diff panel and thirteen chips for the Dev panel, for 6916 pillar/well samples each. Three replicates of each agent were printed per experiment, in combination with each other agent and control media. The four control media were printed at four to six replicates per chip (Table 5.3). In a departure from the dose-response experiments, NeuroDiff medum was substituted for the DualSmad condition.

Table 5.3. Control media used in pair-wise experiments.

Alias	Base Media	Additives
APEL	APEL	-
NeuroDiff	StemDiff Neural Induction Medium	-
MesoDiff	StemDiff Mesoderm Induction Mediaum	-
EndoDiff	StemDiff Definitive Endoderm Kit	-

Chips from both panels were stained with one of four primary antibody combinations, with one chip in each run reserved for IgG staining (Table 5.4). Nestin was used in place of the Pax6 antibody used to identify ectodermal commitment in the dose-response experiments.

Table 5.4. Primary antibody combinations used in pair-wise experiments.

Alias	Alexa 488	Alexa 594
Hoechst_Oct4_Pax6	Ms Anti-Oct4	Rb Anti-Pax6
Hoechst_Nestin_Oct4	Ms Anti-Nestin	Rb Anti-Oct4
Hoechst_Oct4_GATA4	Ms Anti-Oct4	Rb Anti-GATA4
Hoechst_IgG_IgG	Ms IgG	Rb IgG

As in the dose-response experiments, chip layouts were randomized. Cells were in 2D on Matrigel. The removal of the printing overhead associated with doses facilitated much better printing fidelity and faster print times. To maintain consistency across experiments and comply with time constraints, day five was chosen as the endpoint for fixation and staining.

5.2. Control Benchmarking

As before, APEL was used as the base medium to benchmark cellular baselines and as the vessel for agent dilution and delivery from stock. High incidence of Nestin staining was observed in the basal state, which matches previous observations in this cell system (Fig. 5.1). Otherwise, benchmarks were largely in agreement across both the Dev and Diff batches of chips.

Among the three control media run in both panels (Fig. 5.2), outcomes largely followed with expectation. In the NeuroDiff condition, Nestin was variably upregulated, as is to be expected for cells being pushed towards the ectoderm lineage. The MesoDiff condition continued to demonstrate the strongest effects on differentiated marker expression, boosting Brachyury expression in cells committing to mesoderm. Finally, the EndoDiff condition seemed to exercise more activity than had previously been witnessed, with moderate increases in Brachyury and GATA4 staining. Of note, the Oct4 dynamics seemed to be a bit more robustly pluripotent in this set of experiments, with baseline levels higher in the APEL state and downregulation much less severe in the control differentiation media.

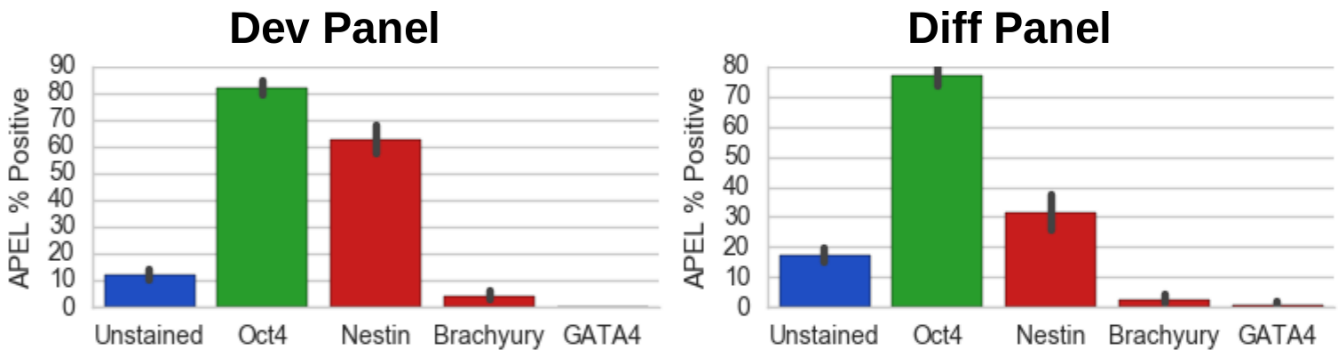

Figure 5.1. Baseline marker presentation in APEL. Values are presented as percent staining positive within the entire cell population. Aside from some discrepancies in the Nestin staining, values tended to agree well between experiments in the APEL control.

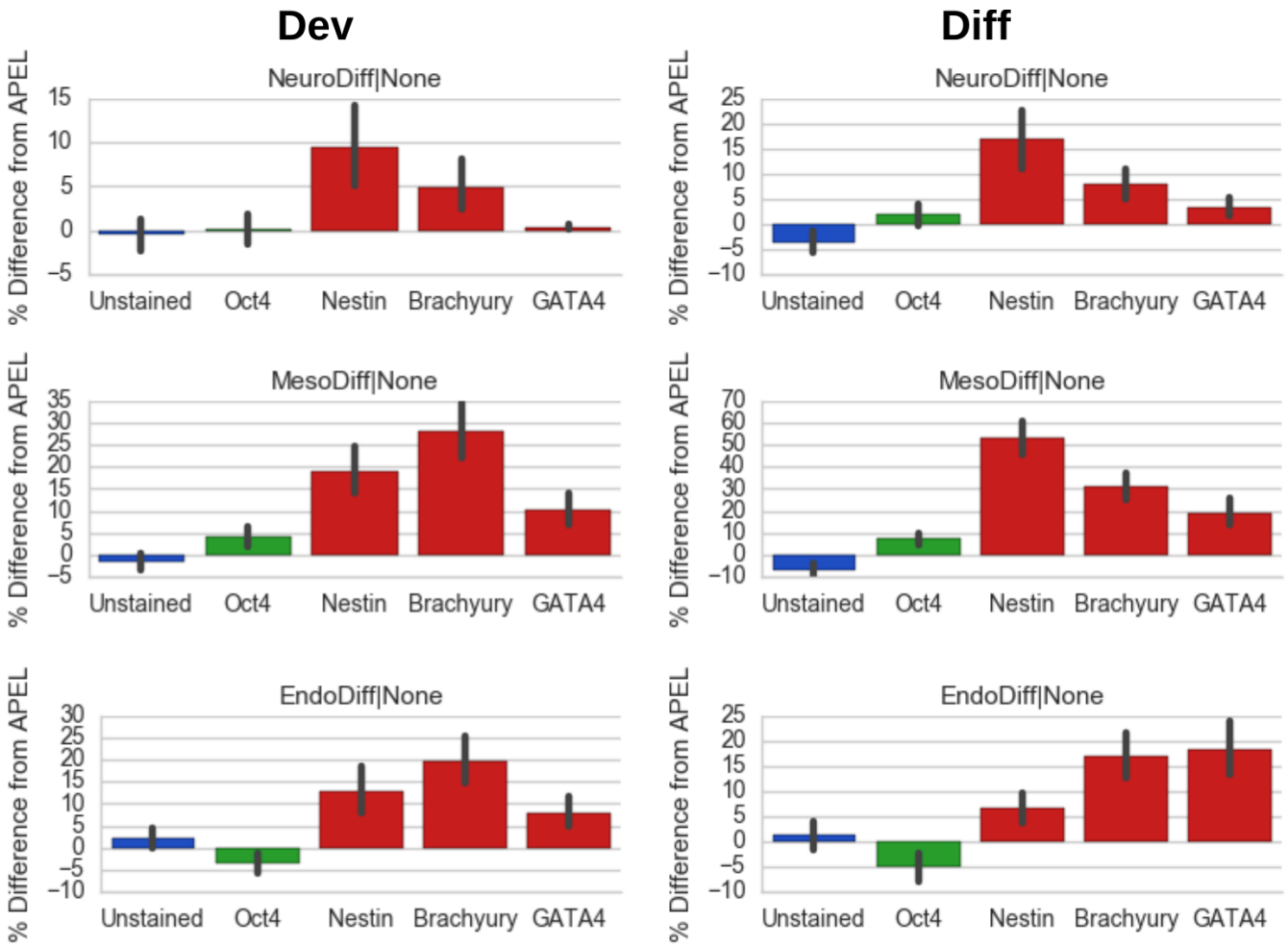

Figure 5.2. Baseline marker presentation in differentiation control media. Values are presented as the percentage point difference from population staining in the respective APEL controls.

5.3. Pair Effects on hiPSC Fate

In the understanding and control of stem cells, context is everything, so having a general-purpose system capable of pushing the experimental frontier beyond traditional single agent screening is increasingly important to progress in the field. While there remains work to be done to refine the hardware and computational pipeline employed here, the pair-wise screening of the Dev and Diff panels did yield a number of statistically significant interactions, of which two examples are provided in Figure 5.3.

PD-173072 is an inhibitor of FGF receptor-mediated signaling, and RepSox is an inhibitor of ALK5 receptor-mediated activity along the TGF-B / Activin signaling pathway. As discussed earlier, both of these pathways act to help keep stem cells pluripotent and self-renewing[92]. Combining the two induced nearly 60% of the cell population to shift from an Oct4+ state to unstained state after five days, a significant change from the APEL baseline and both of the agents when administered alone (Fig. 5.3a).

A similar effect was observed when two tumor-suppressing drugs, Trichostatin A and 5-Azacytidine, were jointly administered (Fig. 5.3b). Just as for PD173074 and RepSox, there was a large transition of the cell population from staining Oct4+ to not staining for any pluripotency or lineage markers. Whereas the PD173074 + RepSox combo did not severely effect cell proliferation, though (Fig. 5.3c), Trichostatin A + 5-Azacytidine clearly did, with a steep drop in cell number and density observed (Fig. 5.3d). This would implicate that signaling plays the primary role in the case of the former, while toxicity is the primary driving force in the latter, though further follow-up experiments would be necessary to tease apart the exact mechanisms in play.

5.4 Outlook and Perspectives

In a span of approximately two weeks, the Pair-wise experiments generated 121,296 images across more than 1500 unique conditions, totaling to 660 GB of raw pixel data. Over a subsequent period of three weeks, CellProfiler extracted features from 116,493,182 identified cell objects in these images, yielding a final HDF5 feature database totaling just over 61.3 GB.

From the collected data, insights could be gleaned about combinatorial agent effects on cellular pluripotency or lineage commitment, as observable by comparing the percent of the population staining for the various antibodies in each condition. Pairings were found that had significant, additive effects superseding the activity of either agent alone. As was the case in the example data looking at PD173074 + RepSox and Trichostatin A + 5-Azacytidine, outcomes reflecting both cell identity and proliferation could be discerned, in agreement with expectations from the existing knowledge base.

As with the dose-response experiments, though, the results tended to be confounded by variability among replicates. In hindsight this is to be expected, since the pair-wise experiments were initiated prior to the conclusion of the dose-response experiments, preventing lessons later learned from the dose-response results from being applied to iterate and refine experimental parameters and design. Since only one dose was used per agent, the pair-wise experiments did avoid the systematic error from on-chip dilution that compromised nearly 40% of the dose-response results. Despite the larger number

of unique conditions, the pair-wise print durations were actually shorter and less subject to problems with drying and contamination due to the reduction of the overhead inherent with the dose range.

In some places, the computational pipeline also began to show strain under the load of the proliferating unique conditions. As an example, calculation of the full Mann-Whitney U, Holmes Bonferroni-corrected significance matrix across the dose-response data for each marker-staining population required less than a minute of computational time. For this pair-wise data set containing 9,828,020 comparisons, these same calculations took upwards of nine hours. There are likely low-level optimizations to the code base that can be made to improve this, like replacing the append operations used to construct the matrix with a chunked approach, or developing a logic flow to slim down the number of comparisons calculated. Parallelizing the code is also important and can provide large gains. While parallelization was employed for some of the particularly time-intensive steps (like the manipulation of the feature HDF5 databases), broad implementation requires a fundamental restructuring of code architecture that can itself consume precious man hours to design and troubleshoot. Finally, the code base can also be migrated away from Python, a prototyping language that maximizes coder productivity at the expense of some computational slack, to a more production-oriented language like C++ or Julia.

To summarize, just like for the dose-response work, a number of encouraging results did emerge from the pair-wise experiments that point to the platform's ability to handle a diverse array of experimental formats at high-throughput. While further optimizations remain that could improve statistical power by increasing reproducibility across conditions and chips, time and additional independent replicates can help to compensate, making the platform capable of providing unique scientific insight into questions at scales not feasible for conventional throughput.

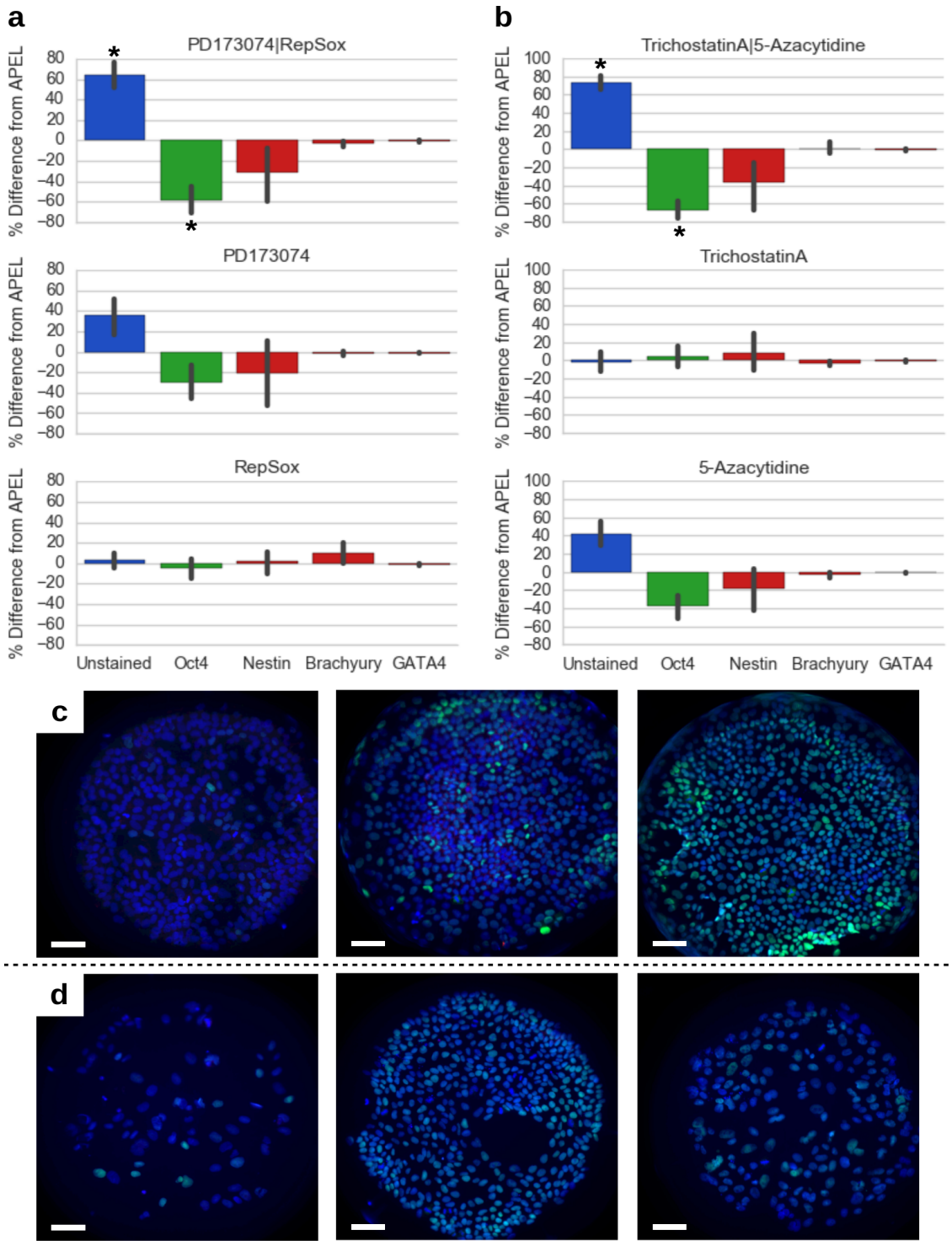

Figure 5.3. Examples of additive interaction within the screened agent pairs. Comparisons of the pairings (**a**) PD173074 | RepSox and (**b**) Trichostatin A | 5-Azacytidine with respective single agents in APEL. "*" indicates significant different to APEL baseline and single agents, $p < 0.05$, using the Mann-Whitney U test with Holmes Bonferroni correction. (**c**) From left to right, PD173074 + RepSox, PD17307, and RepSox conditions. (**d**) From left to right, Trichostatin A + 5-Azacytidine, Trichostatin A, and Azacytidine conditions. In both series, staining is for Hoechst (blue), Oct4 (green), and Brachyury (red). Scale bars = 100 μm.

5.5 Materials and Methods

Cell Culture
Passage 40-50 TCTF Human induced pluripotent stem cells (hiPSCs) were maintained on tissue culture treated polystyrene 6 well plates (BD Falcon) coated as per manufacturer recommendation with hESC qualified Matrigel (Corning). mTeSR growth medium (StemCell Tech) supplemented with 1x mTeSR supplement (StemCell Tech) was refreshed daily. Cells were passaged 1:6 every 3 days using ReLeSR (StemCell Tech) and the manufacturer recommended protocol. All media was supplemented with 0.5% penicillin/streptomycin (Fisher).

Pillar/Well Chip Printing
Printing protocols for the pillar and well chips (Samsung) were prepared and validated in advance. The day before printing, pillar chips were stamped into well chips printed at 4ºC with hESC qualified Matrigel diluted as per manufacturer recommendations. These pillar/well chips were incubated overnight at 37ºC in humidity chambers. The following day, new well chips were printed with APEL (StemCell Tech) and stored until needed in humidity chambers at 37ºC. Cells were immediately passaged with ReLeSR, as above, re-suspended to ~3E6 cells/ml in APEL, and loaded onto the source plate. The prepared Matrigel-coated pillar chips were printed with cells and stored face-up in humidity chambers at 37ºC to allow cells to settle and adhere. After 20 minutes, pillar chips were stamped into the freshly prepared APEL well chips and returned to humidity chambers at 37ºC. All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution (Sigma).

Pillar/Well Chip Culture
On each subsequent day for five days, cells were stamped into freshly prepared well chips containing the appropriately printed media conditions (Table 5.1-3). All media was supplemented with 0.5% penicillin/streptomycin and 1:10000 anti-foam C solution. Small molecule stocks were used for no more than 5 days stored at 4ºC in the dark before being refreshed.

Immunostaining
Chips were stained as per Note 1.1 with various primary antibody combinations (Table 5.4). Materials used included 2% PFA (Santa Cruz), PBS (Gibco), Triton X-100 (Sigma), donkey serum (Sigma), 1:1000 Hoechst 33342 (Life Technologies), 1:200 Ms anti-Nestin (BD), 1:50 Ms anti-Oct4 (Santa Cruz), 1:50 Rb anti-Oct4 (Santa Cruz), 1:50 Rb anti-Brachyury (Santa Cruz), 1:50 Rb anti-GATA4 (Santa Cruz), 1:5000 Ms IgG (Abcam), 1:50 Rb IgG (Abcam), 1:500 Dk anti-Ms 488 (Jackson Immuno), 1:500 Dk anti-Rb 594 (Jackson Immuno). All solutions printed in well chips were supplemented with 1:10000 anti-foam C solution.

Data Analysis and Statistics
Images were re-formatted using Python, and a variety of image-level and object-level features were mined using CellProfiler[89]. HDF5 databases of said features were collated, analyzed, and plotted using Python. Significance of results was tested using the Mann-Whitney U test with Holmes Bonferroni correction in Python.

Epilogue

Speaking objectively, this project has fallen far short of its original goals of using a high-throughput platform to model and treat neural and aging pathologies with stem cells. In fact, for most of the work's duration, biology has taken second-seat to the platform itself, with hardware failures, vendor incompetence, incompatibilities with standard biological assays, and innumerable other unforeseen difficulties necessitating a seemingly endless cycle of troubleshooting. Through all the frustration, though, something different and unexpected has emerged.

Over the course of years, the pipeline detailed in Chapters 1 and 2 has been built nearly from the ground-up to generalize to a diverse array of scientific needs in the high-throughput stem cell space. It is adaptive to cell type, culture dimensionality, a variety of coating and scaffolding materials, dose-wise and combinatorial experimental formats, and an array of read-outs that encompass cell viability, morphology, and marker expression. And it does all of this at a scale that dwarfs throughput of a conventional laboratory work flow, enabling terabytes of image data about thousands of unique conditions to be collected in months instead of years. Consider that all the data presented in Chapters 3, 4, and 5, just over 210,000 images looking at thousands of unique conditions, was generated in a time span of two months.

The system is still, of course, not perfect. As can be seen from the data presented in the foregoing, especially Chapters 4 and 5, further optimization will be required to achieve robust statistical power without relying on additional replicates. There are many ways these refinements might manifest, including iterative experimental design, benchmarking of inherent system and cell parameters, sharpened print fidelity, a generalized image processing and analysis pipeline, and gradual migration to open source infrastructure.

Whether through improvements like these, or simply through the addition of more replicates, the platform has now reached a level of maturity that can meaningfully contribute to a number of investigative directions. Dose-response experiments focusing on developmental toxicity can leverage the throughput of the platform, immunostaining for lineage markers, and ongoing work in morphological analysis to identify combinations of features capable indicative of developmental toxicity risk for screened chemicals. The platforms flexibility and throughput can also be turned to questions in understanding and optimally controlling cellular differentiation. Currently slow or inefficient processes for deriving certain cells of interest *ex vivo*, such as oligodendrocytes, could be significantly enhanced, both in terms of time to completion and overall differentiation efficiency, through comprehensive dosage, temporal, and combinatorial screening of select factors. Finally, the amount of data the platform produces is perfectly suited for efforts seeking to streamline and further automate biological data analysis. Application of machine learning and unsupervised clustering algorithms can reveal underlying structures in data that, due to scale, would be difficult or even impossible for manual human observation to identify. The deterministic nature of these approaches, if developed into a generalized pipeline, could also provide large benefits to the field as a whole. If developed into a standardized analytical pipeline, human error and differences in parameters used for quantification can be controlled, resulting in increased reproducibility and reliability across published literature.

Working on this project through the thick and thin has been enriching, if in a non-traditional sense. While the platform is not (yet) a prolific source of that universal metric of scientific success – publications – the diversity of obstacles that had to be overcome to get it to even function necessitated the integration of many disciplines, spanning CAD, 3D printing, mechanical engineering, materials science, programming, image analysis, and IT, all on top of traditional molecular and cell biology skill sets. It has also cultivated a keen appreciation for the well-oiled machine that is existing conventional laboratory techniques, and for the inertial barrier that keeps so many investigators from transitioning to cutting edge methods and instruments or striking out to develop their own.

In any case, rather than being a standard scientific exercise in taking things apart to understand them, this project has largely been the opposite, requiring that things (and expertise, and people) instead be brought together. Being a catalyst for that process, and the glimpse it has provided into a potential future where lab work is facilitated by machines and algorithms (perhaps in place of the sweat and tears of graduate students?) has been enlightening and refreshing.

References

1. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282,** 1145–1147 (1998).
2. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131,** 861–872 (2007).
3. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
4. Berardi, A. C., Wang, A., Levine, J. D., Lopez, P. & Scadden, D. T. Functional isolation and characterization of human hematopoietic stem cells. *Science* **267,** 104–108 (1995).
5. Bickenbach, J. R. Identification and Behavior of Label-retaining Cells in Oral Mucosa and Skin. *J. Dent. Res.* **60,** 1611–1620 (1981).
6. Doetsch, F., Caillé, I., Lim, D. A., García-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular Zone Astrocytes Are Neural Stem Cells in the Adult Mammalian Brain. *Cell* **97,** 703–716 (1999).
7. Beltrami, A. P. *et al.* Adult Cardiac Stem Cells Are Multipotent and Support Myocardial Regeneration. *Cell* **114,** 763–776 (2003).
8. Collins, C. A. *et al.* Stem Cell Function, Self-Renewal, and Behavioral Heterogeneity of Cells from the Adult Muscle Satellite Cell Niche. *Cell* **122,** 289–301 (2005).
9. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21,** 121–131 (2015).
10. Soldner, F. *et al.* Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136,** 964–977 (2009).
11. Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457,** 277–280 (2009).
12. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321,** 1218–1221 (2008).
13. Maehr, R. *et al.* Generation of pluripotent stem cells from patients with type 1 diabetes. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15768–15773 (2009).
14. Park, I.-H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134,** 877–886 (2008).
15. Lee, G. *et al.* Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461,** 402–406 (2009).
16. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* **19,** 971–974 (2001).
17. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nat. Methods* **3,** 637–646 (2006).
18. Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.* **24,** 185–187 (2006).
19. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8,** 424–429 (2011).
20. Metallo, C. M. *et al.* Human Embryonic Stem Cell-Derived Keratinocytes Exhibit an Epidermal Transcription Program and Undergo Epithelial Morphogenesis in Engineered Tissue Constructs. *Tissue Eng. Part A* **16,** 213–223 (2010).
21. Pankratz, M. T. *et al.* Directed Neural Differentiation of Human Embryonic Stem Cells via an Obligated Primitive Anterior Stage. *STEM CELLS* **25,** 1511–1520 (2007).

22. Suter, D. M., Preynat-Seauve, O., Tirefort, D., Feki, A. & Krause, K.-H. Phenazopyridine induces and synchronizes neuronal differentiation of embryonic stem cells. *J. Cell. Mol. Med.* **13,** 3517–3527 (2009).
23. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480,** 547–551 (2011).
24. Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat. Biotechnol.* **25,** 1015–1024 (2007).
25. Kattman, S. J. *et al.* Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8,** 228–240 (2011).
26. Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc. Natl. Acad. Sci.* **109,** E1848–E1857 (2012).
27. Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat. Protoc.* **8,** 162–175 (2013).
28. Nakagawa, T., Lee, S. Y. & Reddi, A. H. Induction of chondrogenesis from human embryonic stem cells without embryoid body formation by bone morphogenetic protein 7 and transforming growth factor β1. *Arthritis Rheum.* **60,** 3686–3692 (2009).
29. Niwa, A. *et al.* A Novel Serum-Free Monolayer Culture for Orderly Hematopoietic Differentiation of Human Pluripotent Cells via Mesodermal Progenitors. *PLoS ONE* **6,** e22261 (2011).
30. Wang, C. *et al.* TGFβ inhibition enhances the generation of hematopoietic progenitors from human ES cell-derived hemogenic endothelial cells using a stepwise strategy. *Cell Res.* **22,** 194–207 (2012).
31. Chen, S. *et al.* A small molecule that directs differentiation of human ESCs into the pancreatic lineage. *Nat. Chem. Biol.* **5,** 258–265 (2009).
32. Jiang, W. *et al.* In vitro derivation of functional insulin-producing cells from human embryonic stem cells. *Cell Res.* **17,** 333–344 (2007).
33. Hay, D. C. *et al.* Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proc. Natl. Acad. Sci.* **105,** 12301–12306 (2008).
34. Agarwal, S., Holton, K. L. & Lanza, R. Efficient Differentiation of Functional Hepatocytes from Human Embryonic Stem Cells. *STEM CELLS* **26,** 1117–1127 (2008).
35. Klim, J. R., Li, L., Wrighton, P. J., Piekarczyk, M. S. & Kiessling, L. L. A defined glycosaminoglycan-binding substratum for human pluripotent stem cells. *Nat. Methods* **7,** 989–994 (2010).
36. Rodin, S. *et al.* Long-term self-renewal of human pluripotent stem cells on human recombinant laminin-511. *Nat. Biotechnol.* **28,** 611–615 (2010).
37. Lei, Y. & Schaffer, D. V. A fully defined and scalable 3D culture system for human pluripotent stem cell expansion and differentiation. *Proc. Natl. Acad. Sci.* **110,** E5039–E5048 (2013).
38. Zhang, S.-C., Wernig, M., Duncan, I. D., Brüstle, O. & Thomson, J. A. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat. Biotechnol.* **19,** 1129–1133 (2001).
39. Lee, G., Chambers, S. M., Tomishima, M. J. & Studer, L. Derivation of neural crest cells from human pluripotent stem cells. *Nat. Protoc.* **5,** 688–701 (2010).

40. Gilbert, P. *et al.* Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. *Science* **329,** 1078–1081 (2010).
41. Ishii, T. *et al.* Effects of extracellular matrixes and growth factors on the hepatic differentiation of human embryonic stem cells. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **295,** G313–G321 (2008).
42. Li, Y. J., Chung, E. H., Rodriguez, R. T., Firpo, M. T. & Healy, K. E. Hydrogels as artificial matrices for human embryonic stem cell self-renewal. *J. Biomed. Mater. Res. A* **79,** 1–5 (2006).
43. Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell* **126,** 677–689 (2006).
44. Keung, A. J., de Juan-Pardo, E. M., Schaffer, D. V. & Kumar, S. Rho GTPases mediate the mechanosensitive lineage commitment of neural stem cells. *Stem Cells Dayt. Ohio* **29,** 1886–1897 (2011).
45. Itoh, M., Kiuru, M., Cairo, M. S. & Christiano, A. M. Generation of keratinocytes from normal and recessive dystrophic epidermolysis bullosa-induced pluripotent stem cells. *Proc. Natl. Acad. Sci.* **108,** 8797–8802 (2011).
46. Preynat-Seauve, O. *et al.* Development of human nervous tissue upon differentiation of embryonic stem cells in three-dimensional culture. *Stem Cells Dayt. Ohio* **27,** 509–520 (2009).
47. Schenke-Layland, K. *et al.* Recapitulation of the embryonic cardiovascular progenitor cell niche. *Biomaterials* **32,** 2748–2756 (2011).
48. Wang, X. & Ye, K. Three-Dimensional Differentiation of Embryonic Stem Cells into Islet-like Insulin-Producing Clusters. *Tissue Eng. Part A* **15,** 1941–1952 (2009).
49. Miki, T., Ring, A. & Gerlach, J. Hepatic Differentiation of Human Embryonic Stem Cells Is Promoted by Three-Dimensional Dynamic Perfusion Culture Conditions. *Tissue Eng. Part C Methods* **17,** 557–568 (2011).
50. Illi, B. *et al.* Epigenetic Histone Modification and Cardiovascular Lineage Programming in Mouse Embryonic Stem Cells Exposed to Laminar Shear Stress. *Circ. Res.* **96,** 501–508 (2005).
51. Jing, D., Parikh, A. & Tzanakakis, E. S. Cardiac Cell Generation From Encapsulated Embryonic Stem Cells in Static and Scalable Culture Systems. *Cell Transplant.* **19,** 1397–1412 (2010).
52. Sauer, H., Rahimi, G., Hescheler, J. & Wartenberg, M. Effects of electrical fields on cardiomyocyte differentiation of embryonic stem cells. *J. Cell. Biochem.* **75,** 710–723 (1999).
53. Chen, M. Q. *et al.* Current-Controlled Electrical Point-Source Stimulation of Embryonic Stem Cells. *Cell. Mol. Bioeng.* **2,** 625–635 (2009).
54. Ding, S. *et al.* Synthetic small molecules that control stem cell fate. *Proc. Natl. Acad. Sci.* **100,** 7632–7637 (2003).
55. Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D. A. & Benvenisty, N. Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11307–11312 (2000).
56. Desbordes, S. C. *et al.* High-Throughput Screening Assay for the Identification of Compounds Regulating Self-Renewal and Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell* **2,** 602–612 (2008).
57. Barbaric, I. *et al.* Novel regulators of stem cell fates identified by a multivariate phenotype screen of small compounds on human embryonic stem cell colonies. *Stem Cell Res.* **5,** 104–119 (2010).
58. Xu, Y. *et al.* Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc. Natl. Acad. Sci.* **107,** 8129–8134 (2010).

59. Ben-David, U. *et al.* Selective elimination of human pluripotent stem cells by an oleate synthesis inhibitor discovered in a high-throughput screen. *Cell Stem Cell* **12,** 167–179 (2013).
60. Anderson, D. G., Levenberg, S. & Langer, R. Nanoliter-scale synthesis of arrayed biomaterials and application to human embryonic stem cells. *Nat. Biotechnol.* **22,** 863–866 (2004).
61. Anderson, D. G., Putnam, D., Lavik, E. B., Mahmood, T. A. & Langer, R. Biomaterial microarrays: rapid, microscale screening of polymer–cell interaction. *Biomaterials* **26,** 4892–4897 (2005).
62. Brafman, D. A. *et al.* Long-term human pluripotent stem cell self-renewal on synthetic polymer surfaces. *Biomaterials* **31,** 9135–9144 (2010).
63. Ilkhanizadeh, S., Teixeira, A. I. & Hermanson, O. Inkjet printing of macromolecules on hydrogels to steer neural stem cell differentiation. *Biomaterials* **28,** 3936–3943 (2007).
64. Lutolf, M. P., Doyonnas, R., Havenstrite, K., Koleckar, K. & Blau, H. M. Perturbation of single hematopoietic stem cell fates in artificial niches. *Integr. Biol.* **1,** 59–69 (2009).
65. Unadkat, H. V. *et al.* An algorithm-based topographical biomaterials library to instruct cell fate. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16565–16570 (2011).
66. Prudhomme, W., Daley, G. Q., Zandstra, P. & Lauffenburger, D. A. Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signaling. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2900–2905 (2004).
67. Flaim, C. J., Chien, S. & Bhatia, S. N. An extracellular matrix microarray for probing cellular differentiation. *Nat. Methods* **2,** 119–125 (2005).
68. Roccio, M., Gobaa, S. & Lutolf, M. P. High-throughput clonal analysis of neural stem cells in microarrayed artificial niches. *Integr. Biol. Quant. Biosci. Nano Macro* **4,** 391–400 (2012).
69. Maury, Y. *et al.* Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat. Biotechnol.* **33,** 89–96 (2015).
70. Fernandes, T. G. *et al.* On-chip, cell-based microarray immunofluorescence assay for high-throughput analysis of target proteins. *Anal. Chem.* **80,** 6633–6639 (2008).
71. Fernandes, T. G., Diogo, M. M., Clark, D. S., Dordick, J. S. & Cabral, J. M. S. High-throughput cellular microarray platforms: applications in drug discovery, toxicology and stem cell research. *Trends Biotechnol.* **27,** 342–349 (2009).
72. Lussi, J. W., Falconnet, D., Hubbell, J. A., Textor, M. & Csucs, G. Pattern stability under cell culture conditions—A comparative study of patterning methods based on PLL-g-PEG background passivation. *Biomaterials* **27,** 2534–2541 (2006).
73. Yamada, K. M. & Cukierman, E. Modeling tissue morphogenesis and cancer in 3D. *Cell* **130,** 601–610 (2007).
74. Discher, D. E., Mooney, D. J. & Zandstra, P. W. Growth factors, matrices, and forces combine and control stem cells. *Science* **324,** 1673–1677 (2009).
75. Hazeltine, L. B., Selekman, J. A. & Palecek, S. P. Engineering the human pluripotent stem cell microenvironment to direct cell fate. *Biotechnol. Adv.* (2013). doi:10.1016/j.biotechadv.2013.03.002
76. Li, L., Zhou, Q., Voss, T. C., Quick, K. L. & LaBarbera, D. V. High-throughput imaging: Focusing in on drug discovery in 3D. *Methods San Diego Calif* **96,** 97–102 (2016).
77. Jongpaiboonkit, L., King, W. J. & Murphy, W. L. Screening for 3D Environments That Support Human Mesenchymal Stem Cell Viability Using Hydrogel Arrays. *Tissue Eng. Part A* **15,** 343–353 (2009).
78. Fernandes, T. G. *et al.* Three-dimensional cell culture microarray for high-throughput studies of stem cell fate. *Biotechnol. Bioeng.* **106,** 106–118 (2010).

79. Kumachev, A. *et al.* High-throughput generation of hydrogel microbeads with varying elasticity for cell encapsulation. *Biomaterials* **32,** 1477–1483 (2011).
80. Zoldan, J. *et al.* The influence of scaffold elasticity on germ layer specification of human embryonic stem cells. *Biomaterials* **32,** 9612–9621 (2011).
81. Fernandes, T. G. *et al.* On-Chip, Cell-Based Microarray Immunofluorescence Assay for High-Throughput Analysis of Target Proteins. *Anal. Chem.* **80,** 6633–6639 (2008).
82. Lee, M.-Y., Dordick, J. S. & Clark, D. S. Metabolic enzyme microarray coupled with miniaturized cell-culture array technology for high-throughput toxicity screening. *Methods Mol. Biol. Clifton NJ* **632,** 221–237 (2010).
83. Digilab; Precision towards Perfection. Available at: http://www.digilabglobal.com/--omnigrid-micro. (Accessed: 29th April 2016)
84. Lan, F. *et al.* Abnormal Calcium Handling Properties Underlie Familial Hypertrophic Cardiomyopathy Pathology in Patient-Specific Induced Pluripotent Stem Cells. *Cell Stem Cell* **12,** 101–113 (2013).
85. ImageXpress Micro XLS Widefield High-Content Analysis System. *Molecular Devices* Available at: https://www.moleculardevices.com/systems/high-content-imaging/imagexpress-micro-xls-widefield-high-content-analysis-system. (Accessed: 29th April 2016)
86. Polar Series Accel 500 LC Cooling/Heating Recirculating Chillers. Available at: http://www.thermoscientific.com/en/product/polar-series-accel-500-lc-cooling-heating-recirculating-chillers.html. (Accessed: 29th April 2016)
87. Solid State Cooling Systems: Thermocube 200, 300, 400 W Air Cooled, Liquid Recirculating Chiller for Lab, Laser, Semiconductor. Available at: http://sscooling.com/products/thermocube/thermocube-200-300-400/item/thermocube-air-cooled-200-300-400-w-liquid-recirculating-chiller-for-lab-laser-semiconductor. (Accessed: 29th April 2016)
88. Lee, H., Dellatore, S. M., Miller, W. M. & Messersmith, P. B. Mussel-Inspired Surface Chemistry for Multifunctional Coatings. *Science* **318,** 426–430 (2007).
89. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).
90. Genschow, E. *et al.* The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European Centre for the Validation of Alternative Methods. *Altern. Lab. Anim. ATLA* **30,** 151–176 (2002).
91. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Dev. Camb. Engl.* **132,** 1273–1282 (2005).
92. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118,** 4495–4509 (2005).
93. Lam, A. Q. *et al.* Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. *J. Am. Soc. Nephrol. JASN* **25,** 1211–1225 (2014).
94. Mellin, G. W. & Katzenstein, M. The saga of thalidomide. Neuropathy to embryopathy, with case reports of congenital anomalies. *N. Engl. J. Med.* **267,** 1184–1192 contd (1962).

Appendix A: Supplementary Material for Chapter 1

This appendix is the product of the collaboration with Eric Granlund in the UC Berkeley College of Chemistry machine shop to fabricate custom parts necessary to achieve pipeline functionality. Full resolution blueprints are available on file at the machine shop.

A.1 Custom Part Blueprints

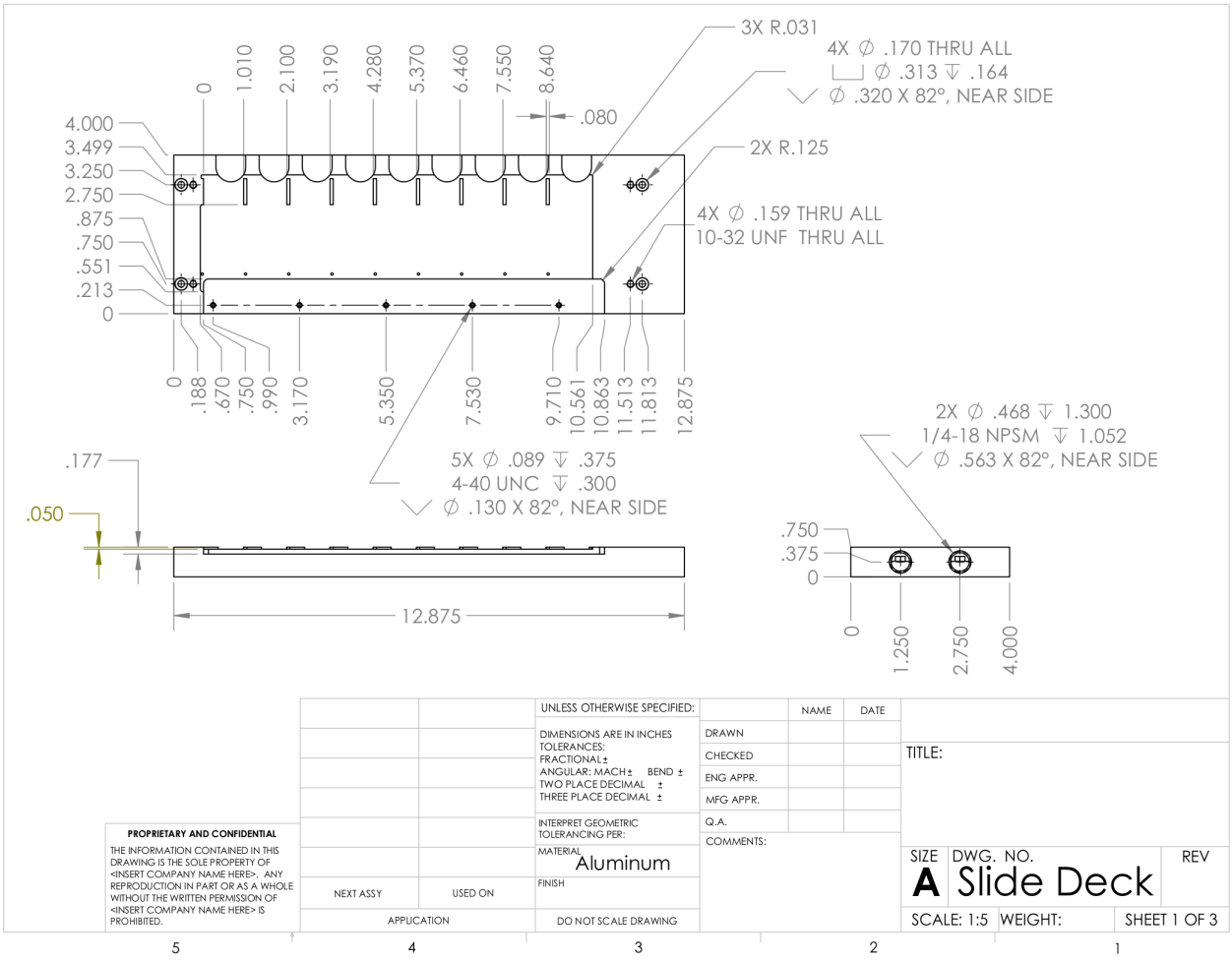

Figure A.1. Sheet 1 of 3, slide deck body.

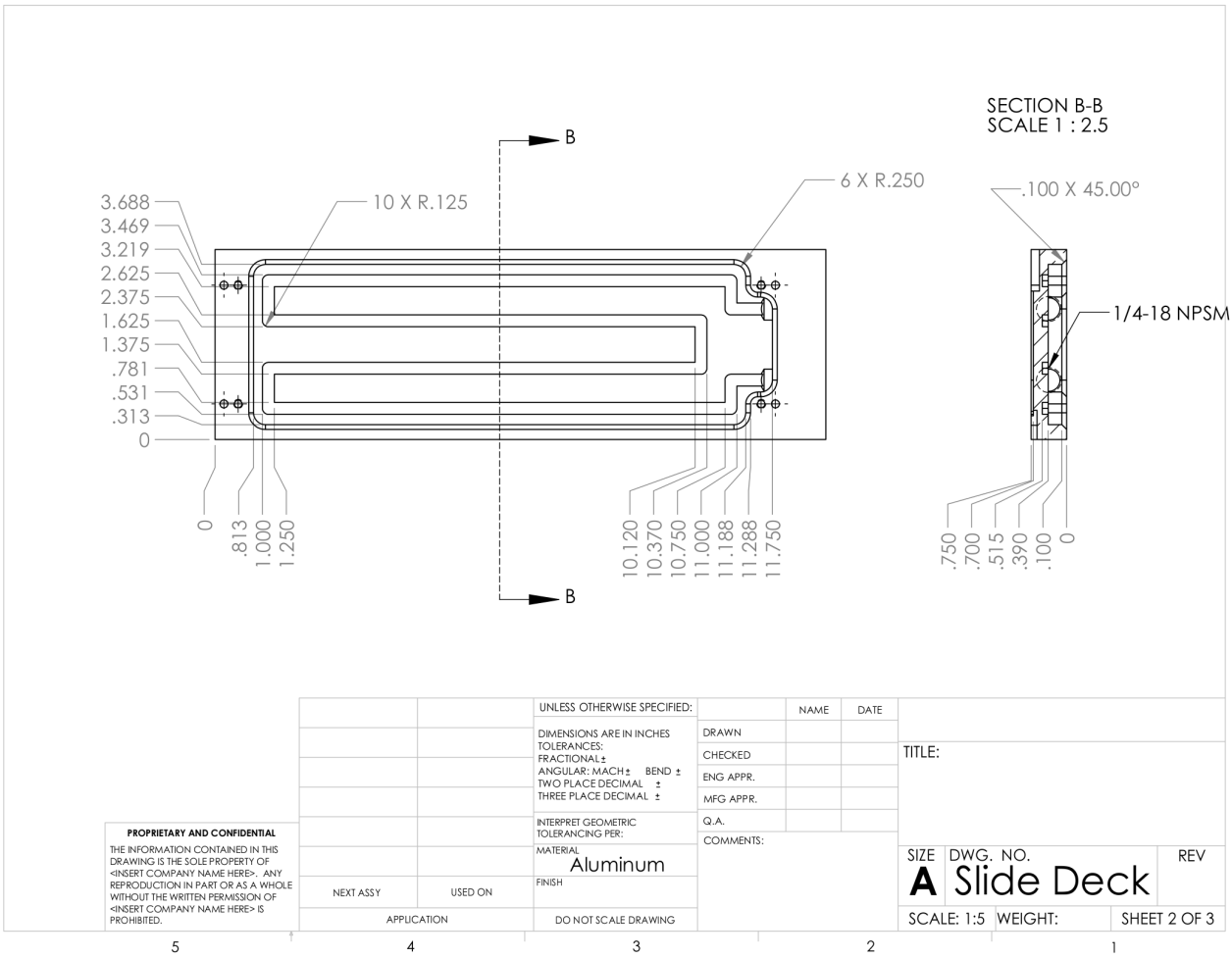

Figure A.2. Sheet 2 of 3, slide deck body.

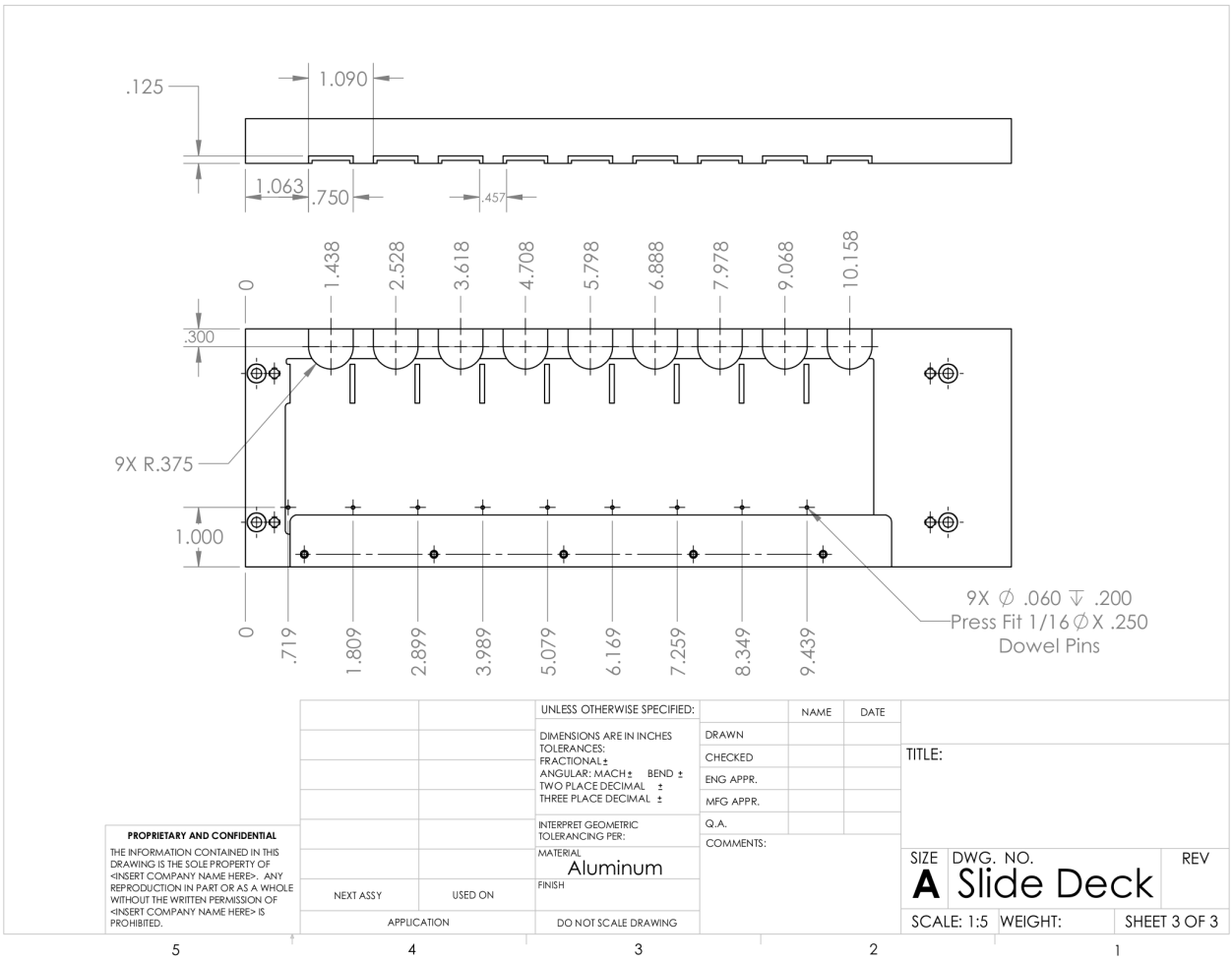

Figure A.3. Sheet 3 of 3, slide deck body.

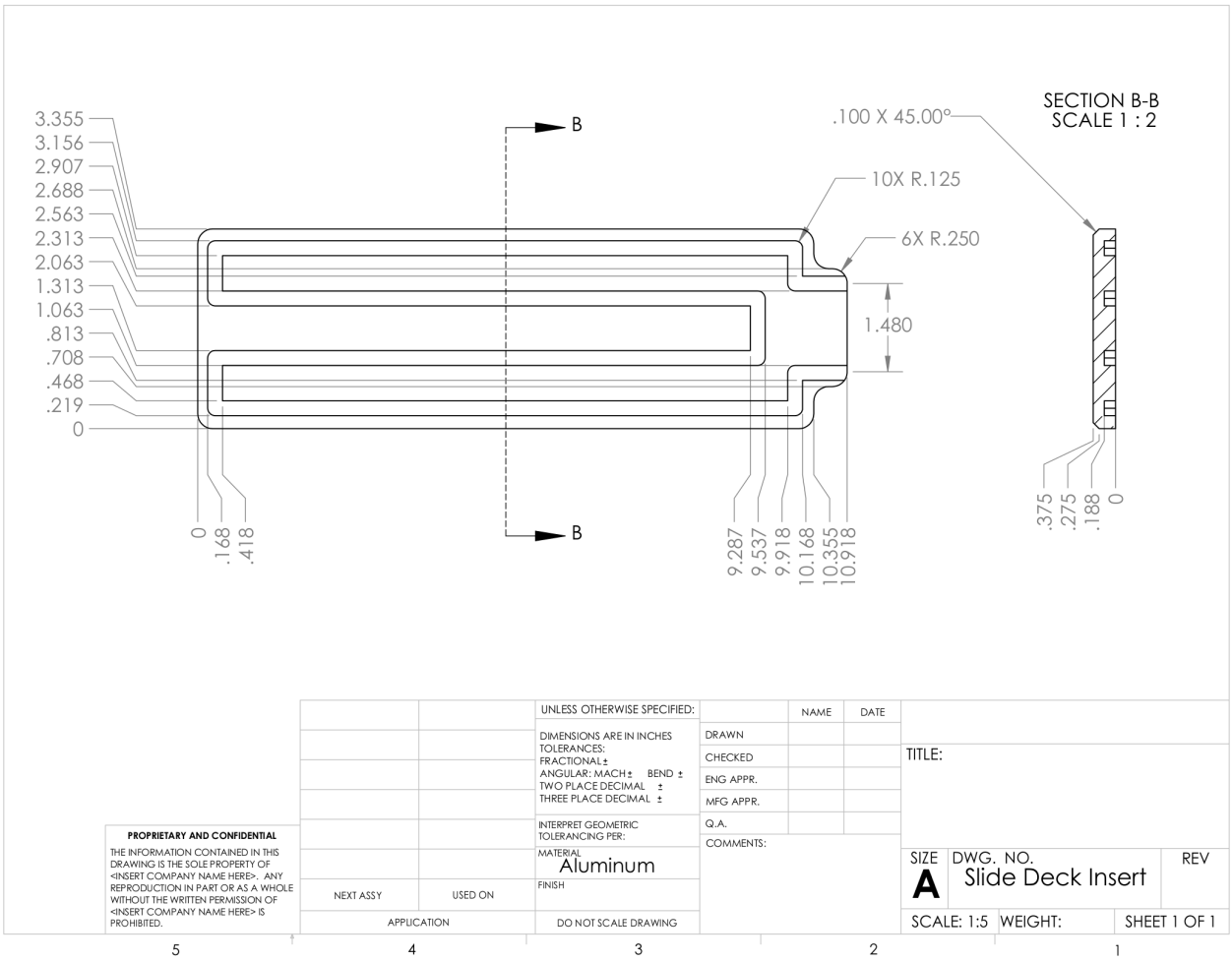

Figure A.4. Sheet 1 of 1, slide deck channel.

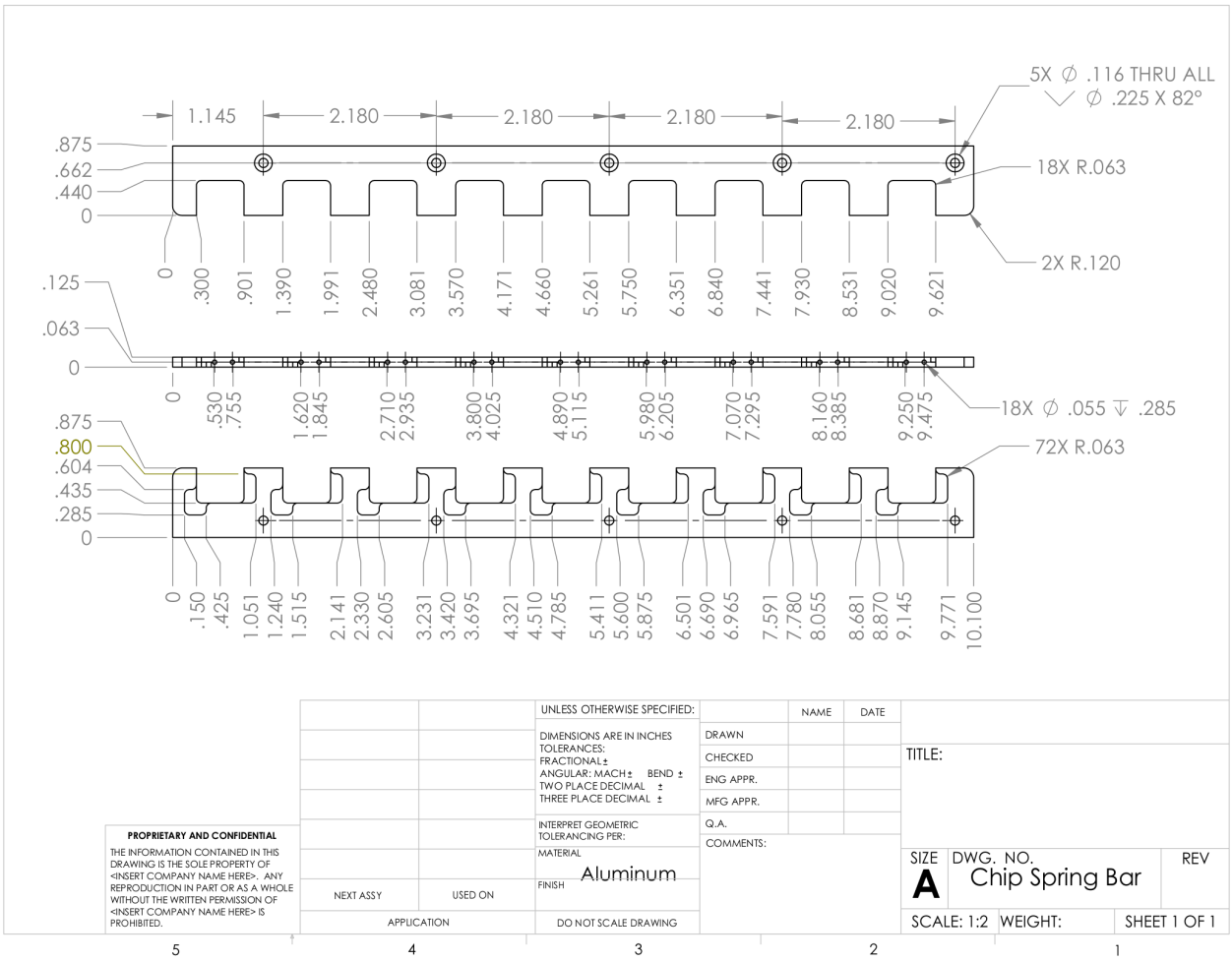

Figure A.5. Sheet 1 of 1, slide deck clip bars.

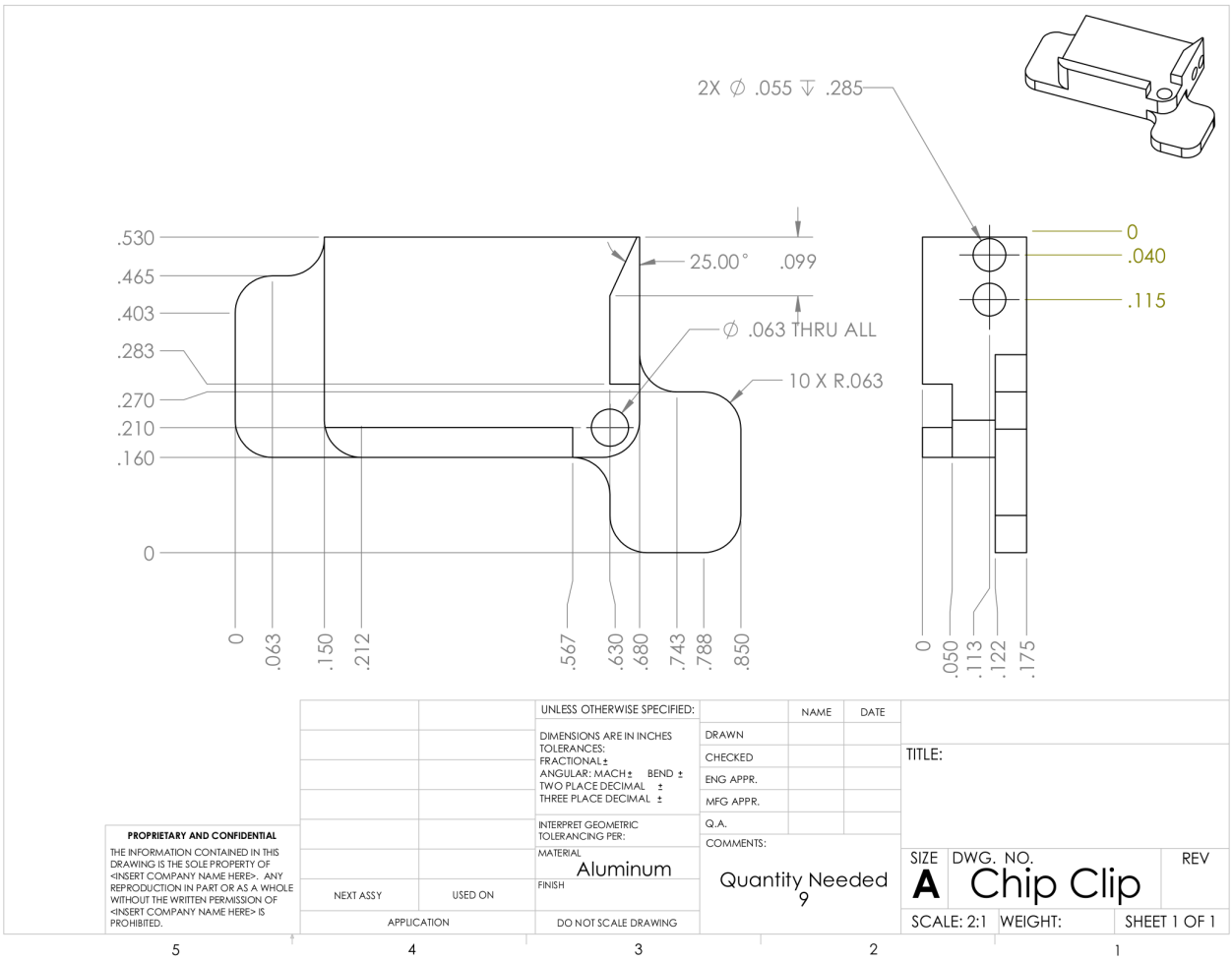

Figure A.6. Sheet 1 of 1, slide deck clip.

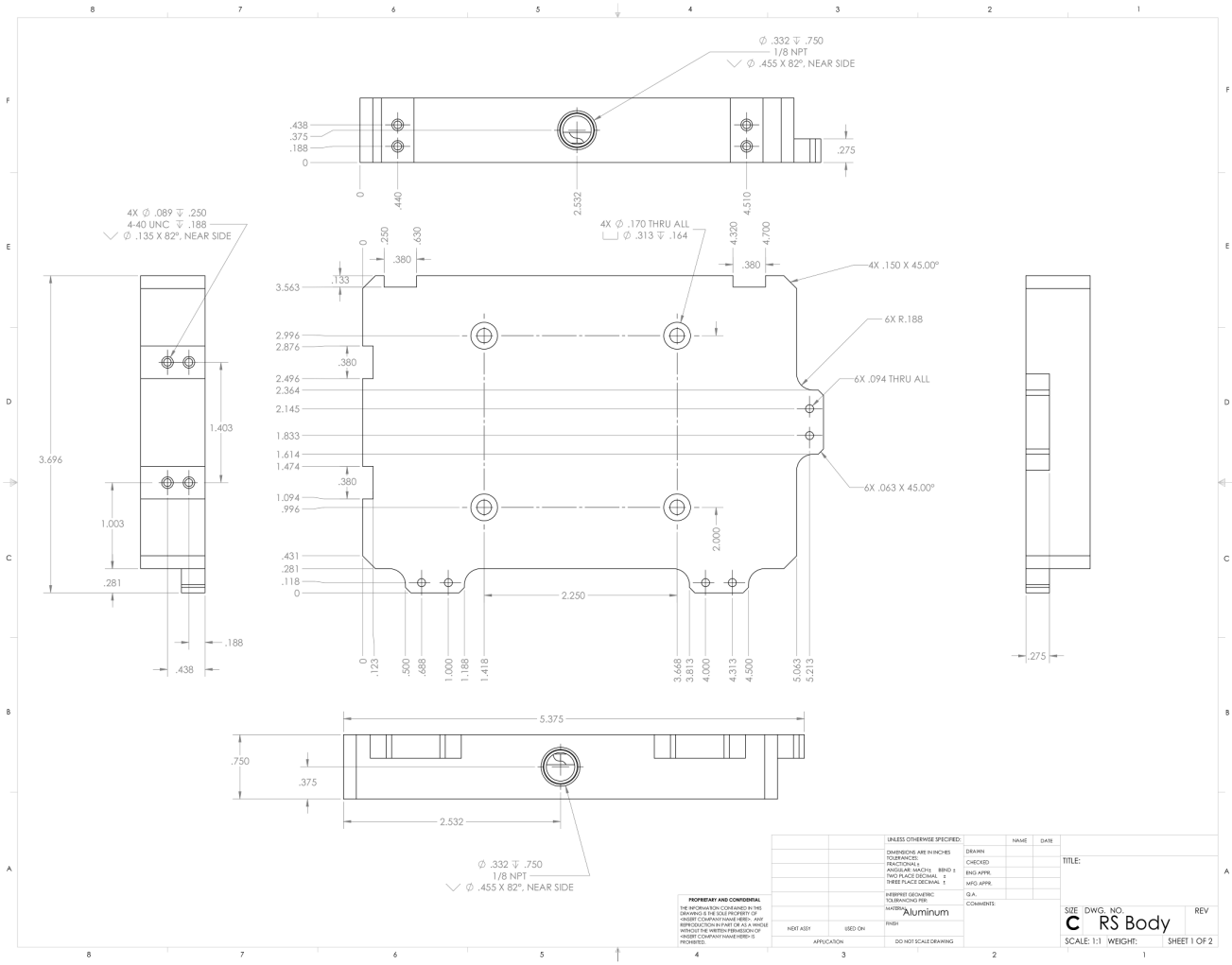

Figure A.7. Sheet 1 of 2, source deck body.

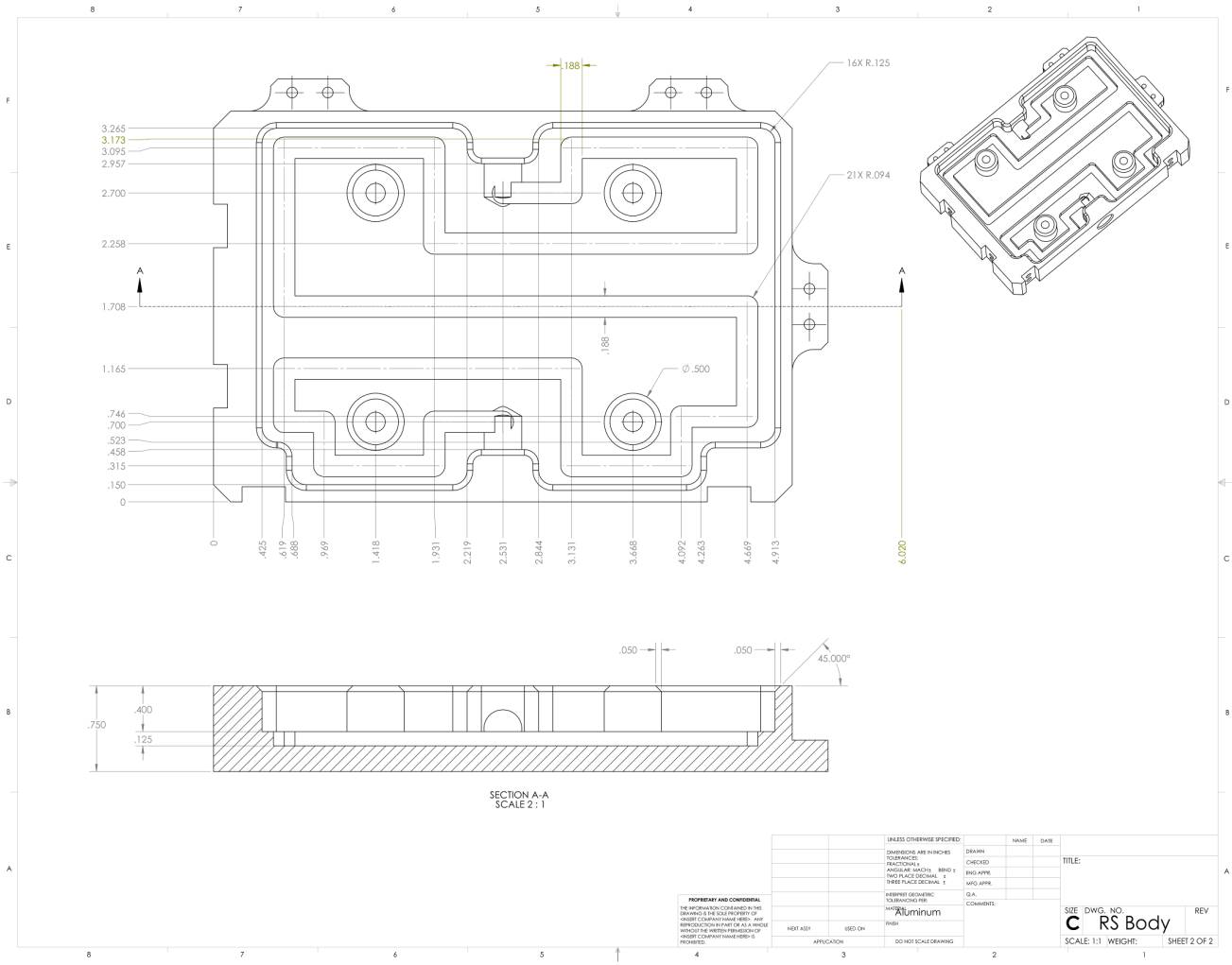

Figure A.8. Sheet 2 of 2, source deck channel.

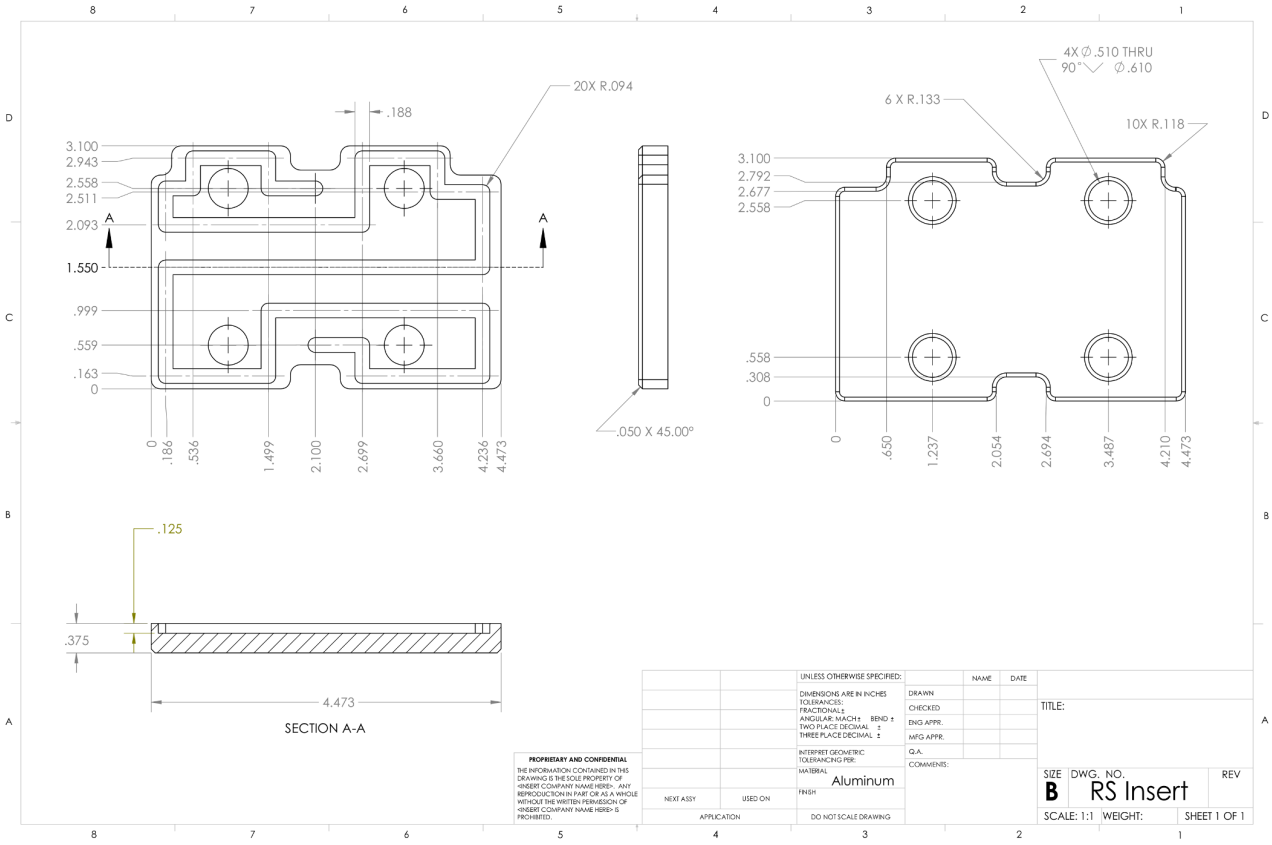

Figure A.9. Sheet 1 of 1, source deck channel.

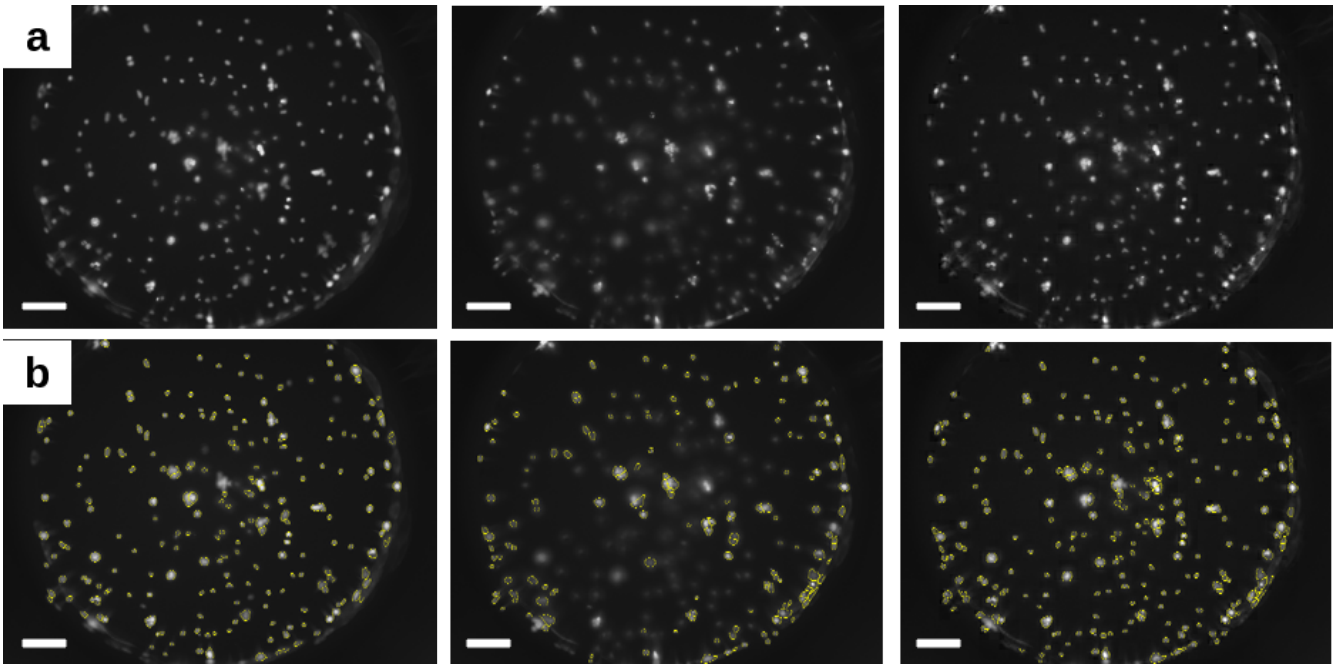

Figure A.10. Demonstration of wavelet-based image fusion algorithm. To attempt to address issues with out of plane noise in 3D chip cultures, stacks of 2-3 images were acquired for each sample and a wavelet-based algorithm was applied to fuse the images based on best focus. In (**a**), the left two images were slices of the same sample taken 40 μm apart. The right pane shows the fused outcome. In (b), yellow overlays are applied to cell objects identified by CellProfiler for each of the images in (**a**). From left to right, counts were 207, 157, and 242. While the image fusion algorithm was able to help resolve more cells and elminate out of plane artifacts, additional artifacts were introduced in the processing that had non-negligible effects on morphometric image assessment. Additional refinement would have been required, both to the algorithm and to the speed of stack acquisition, for high-throughput 3D experiments to be practical.

Appendix B: Supplementary Material for Chapter 2

This appendix contains content related to the computational pipeline developed in the experimental workflow, including Python code used to pre-process the images, a list of the features extracted using CellProfiler, component lists for the data processing computers, inputs for running CellProfiler headless at the command line, and Python code used to pre-process the HDF5 feature databases.

B.1 Python Scripts Used for Image Pre-processing

Code B.1. organize.py. Example pair-wise experiment script demonstrating image collection into a single directly with homogenization of the file name format.

```
# Run this script in parent directory for a given
# day to consolidate all image files into respective
# folders with naming in pattern:
# pP*rXXcYYwCsS.tif where p is plate name, r is row, c is col,
# w is wavelength, s is slice.

import glob
import os
import re

if 'thumbs' not in glob.glob('*'):
      os.mkdir('thumbs')

if 'images' not in glob.glob('*'):
      os.mkdir('images')

# Get the list of days in directory that
# haven't yet been sorted (based on whether
# or not there is a 'thumbs' directory for the
# thumbnail images.
plateList = [plate for plate in glob.glob('*_Plate_*') if not glob.glob(plate + \
      '\\organized')]

# Get nested list of thumb images for each day from filtered plateList.
# Note that this doesn't handle multiple days where there are no thumbnails.
thumbList = [file for file in [glob.glob(plate + '\\Time*\\*Thumb.TIF') for plate \
      in plateList]]

# Note that the index value (3rd group extracted) should always be e.g. 'Chip1'
etc.
immInfoRE = re.compile('\d{8}-(.*)-(.*)-Chip(\d)-(\d)D-
D(\d)_Plate.*\\\\.*\\\\.*_[A-Z](\d\d)_s(\d{1,3})_w(\d)')

# Move thumbs to thumbs folder to get them out of the way.
for thumbs in thumbList:
      for thumb in thumbs:
            experiment, stain, index, dim, day, well, site, channel = \
                  re.search(immInfoRE, thumb).groups()

            newName =  experiment + '-' + stain + '-' + dim + 'D-d' + day + '-i' \
```

```
                + index + '-w' + well + '-s' + site.zfill(3) + '-c' + \
                channel + '.tif'
            #print thumb + ' <===> ' + newName
            os.rename(thumb, 'thumbs\\' + newName)

# Now that the thumbnails are removed, get the actual images.
imageList = [file for file in [glob.glob(plate + '\\Time*\\*.TIF') for plate \
      in plateList]]

for count, images in enumerate(imageList):
      print count
      for image in images:
            experiment, stain, index, dim, day, well, site, channel = \
                re.search(immInfoRE, image).groups()

            if(well == '03' and int(site)%13 == 0):
                #print image
                pass
            else:
                newName =  experiment + '-' + stain + '-' + dim + 'D-d' + day \
                      + '-i' + index + '-w' + well + '-s' + site.zfill(3) + \
                      '-c' + channel + '.tif'
                #print image + ' <===> ' + newName
                os.rename(image, 'images\\' + newName)

# Create the an 'organized' folder so we know it is done for
# future reference.
for plate in plateList:
      os.mkdir(plate + '\\organized')
```

Code B.2. remapping.py. Example pair-wise experiment script demonstrating remapping of the sample names from the intermediate format generated in organize.py to a full enumeration of the metadata.

```
import csv
import glob
import re
import os
import numpy as np
import shutil

# Control key.
controlKey = {'C0':'APEL', 'C1':'NeuroDiff', 'C2':'MesoDiff', 'C3':'EndoDiff'}

###############
# Agent key.
AgentKey = {
      'Dose' : {'00':'CHIR99021[1000nM]', '01':'iCRT5[200nM]',
      '02':'SB431542[250nM]', '03':'LDN193189[200nM]', '04':'DAPT[200nM]',
      '05':'PD173074[100nM]', '06':'OAC-1[500nM]', '07':'ProstaglandinE2[100nM]',
      '08':'SP600125[1000nM]', '09':'SB202190[500nM]', '10':'Go6983[100nM]',
      '11':'Rosiglitazone[500nM]', '12':'ActivinA[50ng_ml]',
      '13':'TGF-B1[50ng_ml]', '14':'BMP4[50ng_ml]', '15':'FGF2[20ng_ml]',
      '16':'EGF[10ng_ml]', '17':'y-27632[1000nM]', '18':'Purmorpamine[1000nM]',
      '19':'Cyclopamine[250nM]', '20':'PD0325901[5nM]', '21':'Indolactam[100nM]',
      '22':'PS-48[10000nM]', '23':'Ly294002[1000nM]', '24':'Pifithrin-mu[5000nM]',
      '25':'Pyrintegrin[1000nM]', '26':'Sinomenine[5000nM]', '27':'ID-8[500nM]',
```

```
      '28':'AICAR[5000nM]', '29':'IDE2[1000nM]', '30':'TrichostatinA[100nM]',
      '31':'5-Azacytidine[1000nM]', '32':'RepSox[100nM]',
      '33':'3-DeazaneplanocinA[1000nM]'},
      'Dev' : {'00':'VPA[10000nM]', '01':'Hydroxyurea[1000nm]',
      '02':'6-Aminonicotinamide[1000nm]', '03':'SalicylicAcid[1000nM]',
      '04':'DimethylPthalate[1000nM]', '05':'BrdU[1000nM]',
      '06':'Methotrexate[1000nM]', '07':'Dimethadione[1000nM]',
      '08':'MethoxyaceticAcid[1000nM]', '09':'D-(+)-Camphor[1000nM]',
      '10':'Diphenhydramine[1000nM]', '11':'Penicillin[1000nM]',
      '12':'Saccharin[1000nM]', '13':'Thalidomide[1000nM]', '14':'LiCl[1000nM]',
      '15':'RetinoicAcid[100nM]', '16':'BoricAcid[1000nM]', '17':'Ethanol[0.1%]',
      '18':'DMSO[0.1%]', '19':'CuSO4[1000nM]', '20':'FeCl3[1000nM]',
      '21':'Isoproterenol[1000nM]', '22':'AscorbicAcid[1000nM]',
      '23':'ZnSO4[1000nM]', '24':'NaCl[1000nM]', '25':'Histidine[1000nM]',
      '26':'Asparagine[1000nM]', '27':'Tyrosine[1000nM]', '28':'Glutamine[1000mM]',
      '29':'AsparticAcid[1000nM]', '30':'Glycine[1000nM]', '31':'Leucine[1000nM]',
      '32':'CdSO4[1000nM]', '33':'Riboflavin[1000nM]'}}

# Stain key.
stainKey = {'Ecto':'Hoechst_Nestin_Oct4', 'Meso':'Hoechst_Oct4_Brachyury',
      'Endo':'Hoechst_Oct4_GATA4', 'IgG':'Hoechst_IgG_IgG'}

# Generate the row, column position from
# the well and site value for looking up
# condition in csv maps.
def getRowCol(wellVal, siteVal):
      row = (int(siteVal) - 1) % 13 + (int(wellVal) - 1) * 13
      col = 13 - (int(siteVal)-1) / 13

      return (row, col)

# Handle the chip maps. Load them into chipMaps dict.
chipMapRE = re.compile('\d{4}-\d{2}-\d{2}-(.*)')

exps = [x for x in glob.glob('*') if (('2D-D5' not in x) and ('2016' in x))]
for exp in exps:
      mapNames = [x for x in glob.glob(exp + '\\*\\*') if 'csv' in x]
      for mapName in mapNames:
            newName = re.search(chipMapRE, mapName).group(1).replace('\\', '-')
            shutil.copy(mapName, 'maps\\' + newName)

chipNames = glob.glob('maps\*.csv')

chipMaps = {}

for chipName in chipNames:
      condiList = ''
      csvFile = open(chipName, 'rt')
      for row in csvFile:
            condiList += row
      condiList = condiList[:-1]
      condiList = condiList.replace('\n', ',')
      condiList = condiList.split(',')
      chipMaps[chipName[5:-4]] = np.array(condiList).reshape((38,14))
      csvFile.close()

imageList = glob.glob('images\*.tif')
```

```
immInfoRE = re.compile('images\\\\(.*)-(.*)-(\d)D-d(\d)-i(\d)-w(\d\d)-s(\d\d\d)\
      -c(\d).tif')

### Now need to convert the site information into media and staining information.
### For each image in the images folder, use it's file name's index and site value
### to map them out of the file name, and control/media/agent + stain info in.
for image in imageList:
      if('IgG' in image):
            # Pull out the relevant info.
            experiment, stain, dim, day, index, well, site, channel = \
                  re.search(immInfoRE, image).groups()
            row, col = getRowCol(well, site)
            media = 'mTeSR'
            ab = stainKey[stain]

            # Compose the new name, and save.
            newName = experiment + '-' + dim + 'D-d' + day + '-i' + index + \
                  '-m' + media + '-r' + str(row).zfill(2) + '-c' + \
                  str(col).zfill(2) + '-s' + ab + '-c' + channel
            #print image + ' <===> ' + 'images\\' + newName + '.tif'
            os.rename(image, 'images\\' + newName + '.tif')

      else:
            # Pull out the relevant info.
            experiment, stain, dim, day, index, well, site, channel = \
                  re.search(immInfoRE, image).groups()
            row, col = getRowCol(well, site)
            condi = chipMaps[experiment + '-' + stain + '-chip' + index][row][col]

            # Determine the media.
            if '-' in condi:
                  condi = condi.split('-')
                  if ('Diff' in experiment):
                        media1 = agentKey['Dose'][condi[0]] if 'C' not in \
                              condi[0] else controlKey[condi[0]]
                        media2 = agentKey['Dose'][condi[1]] if 'C' not in \
                              condi[1] else controlKey[condi[1]]
                        media = '_'.join((media1, media2))
                  elif ('Dev' in experiment):
                        media1 = agentKey['Dev'][condi[0]] if 'C' not in \
                              condi[0] else controlKey[condi[0]]
                        media2 = agentKey['Dev'][condi[1]] if 'C' not in \
                              condi[1] else controlKey[condi[1]]
                        media = '_'.join((media1, media2))
            else:
                  media = controlKey[condi]

            # Determin the stain.
            ab = stainKey[stain]

            newName = experiment + '-' + dim + 'D-d' + day + '-i' + index + '-m' \
                  + media + '-r' + str(row).zfill(2) + '-c' + str(col).zfill(2) \
                  + '-s' + ab + '-c' + channel
            #print image + ' <===> ' + 'images\\' + newName + '.tif'
            os.rename(image, 'images\\' + newName + '.tif')
```

Code B.3. imConvert.py. Example pair-wise experiment script demonstrating conversion of the raw 16 bit TIFF images to 8 bit JPEG.

```
import numpy as np
import glob
import matplotlib.pyplot as plt
from PIL import Image, ImageFont, ImageDraw

# Function to convert image from 16 bit to 8 bit.
def to8(arg1):
      return (((arg1/65536.)*256).astype(int))

# Get the images.
imageList = glob.glob('images\\*.tif')

# Convert the images to 8 bit and save as jpg in new directory.
for image in imageList:
      im = plt.imread(image)
      im = Image.fromarray(to8(im).astype('uint8'))
      im.save(image[:-4].replace('images\\', 'jpgImages\\') + '.jpg')
```

B.2 List of Extracted CellProfiler Features

Note B.1. Catalog of image features extracted at the Experiment, Image, and Object level with CellProfiler.

- **Experiment Level**
  - CellProfiler_Version
  - ChannelType_c1
  - ChannelType_c2
  - ChannelType_c3
  - Default_InputFolder
  - Default_OutputFolder
  - Exit_Status
  - ImageQuality_ThresholdMeanOtsu_c1_2W
  - ImageQuality_ThresholdMeanOtsu_c2_2W
  - ImageQuality_ThresholdMeanOtsu_c3_2W
  - ImageQuality_ThresholdMedianOtsu_c1_2W
  - ImageQuality_ThresholdMedianOtsu_c2_2W
  - ImageQuality_ThresholdMedianOtsu_c3_2W
  - ImageQuality_ThresholdStdOtsu_c1_2W
  - ImageQuality_ThresholdStdOtsu_c2_2W
  - ImageQuality_ThresholdStdOtsu_c3_2W
  - ImageSet_Zip_Dictionary
  - Metadata_Tags
  - Modification_Timestamp
  - Pipeline_Pipeline
  - Run_Timestamp
- **Image Level**
  - Channel_c1
  - Channel_c2
  - Channel_c3
  - Correlation_Correlation_c1_c2
  - Correlation_Correlation_c1_c3
  - Correlation_Correlation_c2_c3
  - Correlation_Costes_c1_c2
  - Correlation_Costes_c1_c3
  - Correlation_Costes_c2_c1
  - Correlation_Costes_c2_c3
  - Correlation_Costes_c3_c1
  - Correlation_Costes_c3_c2
  - Correlation_K_c1_c2
  - Correlation_K_c1_c3
  - Correlation_K_c2_c1
  - Correlation_K_c2_c3
  - Correlation_K_c3_c1
  - Correlation_K_c3_c2
  - Correlation_Manders_c1_c2
  - Correlation_Manders_c1_c3
  - Correlation_Manders_c2_c1
  - Correlation_Manders_c2_c3
  - Correlation_Manders_c3_c1
  - Correlation_Manders_c3_c2
  - Correlation_Overlap_c1_c2
  - Correlation_Overlap_c1_c3
  - Correlation_Overlap_c2_c3
  - Correlation_RWC_c1_c2
  - Correlation_RWC_c1_c3
  - Correlation_RWC_c2_c1
  - Correlation_RWC_c2_c3
  - Correlation_RWC_c3_c1
  - Correlation_RWC_c3_c2
  - Correlation_Slope_c1_c2
  - Correlation_Slope_c1_c3
  - Correlation_Slope_c2_c3
  - Count_c1ob
  - Count_c2ob
  - Count_c3ob
  - ExecutionTime_01Images
  - ExecutionTime_02Metadata
  - ExecutionTime_03NamesAndTypes
  - ExecutionTime_04Groups
  - ExecutionTime_05IdentifyPrimaryObjects
  - ExecutionTime_06IdentifyPrimaryObjects
  - ExecutionTime_07IdentifyPrimaryObjects
  - ExecutionTime_08RelateObjects
  - ExecutionTime_09RelateObjects
  - ExecutionTime_10RelateObjects
  - ExecutionTime_11RelateObjects
  - ExecutionTime_12MeasureCorrelation
  - ExecutionTime_13MeasureImageQuality

- ExecutionTime_14MeasureGranularity
- ExecutionTime_15MeasureObjectIntensity
- ExecutionTime_16MeasureImageIntensity
- ExecutionTime_17MeasureObjectSizeShape
- FileName_c1
- FileName_c2
- FileName_c3
- Frame_c1
- Frame_c2
- Frame_c3
- Granularity_10_c1
- Granularity_10_c2
- Granularity_10_c3
- Granularity_1_c1
- Granularity_1_c2
- Granularity_1_c3
- Granularity_2_c1
- Granularity_2_c2
- Granularity_2_c3
- Granularity_3_c1
- Granularity_3_c2
- Granularity_3_c3
- Granularity_4_c1
- Granularity_4_c2
- Granularity_4_c3
- Granularity_5_c1
- Granularity_5_c2
- Granularity_5_c3
- Granularity_6_c1
- Granularity_6_c2
- Granularity_6_c3
- Granularity_7_c1
- Granularity_7_c2
- Granularity_7_c3
- Granularity_8_c1
- Granularity_8_c2
- Granularity_8_c3
- Granularity_9_c1
- Granularity_9_c2
- Granularity_9_c3
- Group_Index
- Group_Number
- Height_c1
- Height_c2
- Height_c3
- ImageNumber
- ImageQuality_Correlation_c1_20
- ImageQuality_Correlation_c2_20
- ImageQuality_Correlation_c3_20
- ImageQuality_FocusScore_c1
- ImageQuality_FocusScore_c2
- ImageQuality_FocusScore_c3
- ImageQuality_LocalFocusScore_c1_20
- ImageQuality_LocalFocusScore_c2_20
- ImageQuality_LocalFocusScore_c3_20
- ImageQuality_MADIntensity_c1
- ImageQuality_MADIntensity_c2
- ImageQuality_MADIntensity_c3
- ImageQuality_MaxIntensity_c1
- ImageQuality_MaxIntensity_c2
- ImageQuality_MaxIntensity_c3
- ImageQuality_MeanIntensity_c1
- ImageQuality_MeanIntensity_c2
- ImageQuality_MeanIntensity_c3
- ImageQuality_MedianIntensity_c1
- ImageQuality_MedianIntensity_c2
- ImageQuality_MedianIntensity_c3
- ImageQuality_MinIntensity_c1
- ImageQuality_MinIntensity_c2
- ImageQuality_MinIntensity_c3
- ImageQuality_PercentMaximal_c1
- ImageQuality_PercentMaximal_c2
- ImageQuality_PercentMaximal_c3
- ImageQuality_PercentMinimal_c1
- ImageQuality_PercentMinimal_c2
- ImageQuality_PercentMinimal_c3
- ImageQuality_PowerLogLogSlope_c1
- ImageQuality_PowerLogLogSlope_c2
- ImageQuality_PowerLogLogSlope_c3
- ImageQuality_Scaling_c1
- ImageQuality_Scaling_c2
- ImageQuality_Scaling_c3
- ImageQuality_StdIntensity_c1
- ImageQuality_StdIntensity_c2
- ImageQuality_StdIntensity_c3
- ImageQuality_ThresholdOtsu_c1_2W
- ImageQuality_ThresholdOtsu_c2_2W
- ImageQuality_ThresholdOtsu_c3_2W
- ImageQuality_TotalArea_c1
- ImageQuality_TotalArea_c2
- ImageQuality_TotalArea_c3
- ImageQuality_TotalIntensity_c1
- ImageQuality_TotalIntensity_c2
- ImageQuality_TotalIntensity_c3
- ImageSet_ImageSet
- Intensity_LowerQuartileIntensity_c1
- Intensity_LowerQuartileIntensity_c2
- Intensity_LowerQuartileIntensity_c3
- Intensity_MADIntensity_c1
- Intensity_MADIntensity_c2
- Intensity_MADIntensity_c3
- Intensity_MaxIntensity_c1
- Intensity_MaxIntensity_c2
- Intensity_MaxIntensity_c3
- Intensity_MeanIntensity_c1
- Intensity_MeanIntensity_c2
- Intensity_MeanIntensity_c3
- Intensity_MedianIntensity_c1
- Intensity_MedianIntensity_c2
- Intensity_MedianIntensity_c3
- Intensity_MinIntensity_c1
- Intensity_MinIntensity_c2
- Intensity_MinIntensity_c3
- Intensity_PercentMaximal_c1
- Intensity_PercentMaximal_c2
- Intensity_PercentMaximal_c3
- Intensity_StdIntensity_c1
- Intensity_StdIntensity_c2
- Intensity_StdIntensity_c3
- Intensity_TotalArea_c1
- Intensity_TotalArea_c2
- Intensity_TotalArea_c3
- Intensity_TotalIntensity_c1
- Intensity_TotalIntensity_c2
- Intensity_TotalIntensity_c3
- Intensity_UpperQuartileIntensity_c1
- Intensity_UpperQuartileIntensity_c2
- Intensity_UpperQuartileIntensity_c3
- MD5Digest_c1
- MD5Digest_c2
- MD5Digest_c3
- Metadata_FileLocation
- Metadata_Frame
- Metadata_Series
- Metadata_Well
- Metadata_channel
- Metadata_col
- Metadata_day
- Metadata_dim
- Metadata_exp

- Metadata_index
- Metadata_media
- Metadata_row
- Metadata_stain1
- Metadata_stain2
- Metadata_stain3
- ModuleError_01Images
- ModuleError_02Metadata
- ModuleError_03NamesAndTypes
- ModuleError_04Groups
- ModuleError_05IdentifyPrimaryObjects
- ModuleError_06IdentifyPrimaryObjects
- ModuleError_07IdentifyPrimaryObjects
- ModuleError_08RelateObjects
- ModuleError_09RelateObjects
- ModuleError_10RelateObjects
- ModuleError_11RelateObjects
- ModuleError_12MeasureCorrelation
- ModuleError_13MeasureImageQuality
- ModuleError_14MeasureGranularity
- ModuleError_15MeasureObjectIntensity
- ModuleError_16MeasureImageIntensity
- ModuleError_17MeasureObjectSizeShape
- PathName_c1
- PathName_c2
- PathName_c3
- Scaling_c1
- Scaling_c2
- Scaling_c3
- Series_c1
- Series_c2
- Series_c3
- Threshold_FinalThreshold_c1ob
- Threshold_FinalThreshold_c2ob
- Threshold_FinalThreshold_c3ob
- Threshold_OrigThreshold_c1ob
- Threshold_OrigThreshold_c2ob
- Threshold_OrigThreshold_c3ob
- Threshold_SumOfEntropies_c1ob
- Threshold_SumOfEntropies_c2ob
- Threshold_SumOfEntropies_c3ob
- Threshold_WeightedVariance_c1ob
- Threshold_WeightedVariance_c2ob
- Threshold_WeightedVariance_c3ob
- URL_c1
- URL_c2
- URL_c3
- Width_c1
- Width_c2
- Width_c3

- **Object Level**
  - AreaShape_Area
  - AreaShape_Center_X
  - AreaShape_Center_Y
  - AreaShape_Compactness
  - AreaShape_Eccentricity
  - AreaShape_EulerNumber
  - AreaShape_Extent
  - AreaShape_FormFactor
  - AreaShape_MajorAxisLength
  - AreaShape_MaxFeretDiameter
  - AreaShape_MaximumRadius
  - AreaShape_MeanRadius
  - AreaShape_MedianRadius
  - AreaShape_MinFeretDiameter
  - AreaShape_MinorAxisLength
  - AreaShape_Orientation
  - AreaShape_Perimeter
  - AreaShape_Solidity
  - Children_c1ob_Count
  - Children_c2ob_Count
  - Children_c3ob_Count
  - ImageNumber
  - Intensity_IntegratedIntensityEdge_c1
  - Intensity_IntegratedIntensityEdge_c2
  - Intensity_IntegratedIntensityEdge_c3
  - Intensity_IntegratedIntensity_c1
  - Intensity_IntegratedIntensity_c2
  - Intensity_IntegratedIntensity_c3
  - Intensity_LowerQuartileIntensity_c1
  - Intensity_LowerQuartileIntensity_c2
  - Intensity_LowerQuartileIntensity_c3
  - Intensity_MADIntensity_c1
  - Intensity_MADIntensity_c2
  - Intensity_MADIntensity_c3
  - Intensity_MassDisplacement_c1
  - Intensity_MassDisplacement_c2
  - Intensity_MassDisplacement_c3
  - Intensity_MaxIntensityEdge_c1
  - Intensity_MaxIntensityEdge_c2
  - Intensity_MaxIntensityEdge_c3
  - Intensity_MaxIntensity_c1
  - Intensity_MaxIntensity_c2
  - Intensity_MaxIntensity_c3
  - Intensity_MeanIntensityEdge_c1
  - Intensity_MeanIntensityEdge_c2
  - Intensity_MeanIntensityEdge_c3
  - Intensity_MeanIntensity_c1
  - Intensity_MeanIntensity_c2
  - Intensity_MeanIntensity_c3
  - Intensity_MedianIntensity_c1
  - Intensity_MedianIntensity_c2
  - Intensity_MedianIntensity_c3
  - Intensity_MinIntensityEdge_c1
  - Intensity_MinIntensityEdge_c2
  - Intensity_MinIntensityEdge_c3
  - Intensity_MinIntensity_c1
  - Intensity_MinIntensity_c2
  - Intensity_MinIntensity_c3
  - Intensity_StdIntensityEdge_c1
  - Intensity_StdIntensityEdge_c2
  - Intensity_StdIntensityEdge_c3
  - Intensity_StdIntensity_c1
  - Intensity_StdIntensity_c2
  - Intensity_StdIntensity_c3
  - Intensity_UpperQuartileIntensity_c1
  - Intensity_UpperQuartileIntensity_c2
  - Intensity_UpperQuartileIntensity_c3
  - Location_CenterMassIntensity_X_c1
  - Location_CenterMassIntensity_X_c2
  - Location_CenterMassIntensity_X_c3
  - Location_CenterMassIntensity_Y_c1
  - Location_CenterMassIntensity_Y_c2
  - Location_CenterMassIntensity_Y_c3
  - Location_Center_X
  - Location_Center_Y
  - Location_MaxIntensity_X_c1
  - Location_MaxIntensity_X_c2
  - Location_MaxIntensity_X_c3
  - Location_MaxIntensity_Y_c1
  - Location_MaxIntensity_Y_c2
  - Location_MaxIntensity_Y_c3
  - Number_Object_Number
  - ObjectNumber
  - Parent_c2ob
  - Parent_c3ob

B.3 Specifications of Data Processing Computers

Note B.2. Custom Built Data Processing Computer 1.
Supermicro X95AE-V Motherboard
Intel Xeon E3-1230 V2, 3.3GHz CPU
Kingston 2x8GB DDR3-1333 RAM
ASUS GTX660 Ti 2 GB GPU

Note B.3. Custom Built Data Processing Computer 2.
Asus Z97-E Motherboard
Intel Core i7-4790K 4.0GHz CPU
G.Skill 2x8GB DDR3-1866 RAM
MSI GTX980 Ti 6 GB GPU

Note B.4. Purchased Data Processing Computer 3.
ASUS ROG GL552VW-DH71 Laptop
Intel Core i7-6700HQ 2.6GHz CPU
Onboard 16 GB DDR4 RAM
Nvidia GTX 960M 2 GB GPU

B.4 Commands Used to Run CellProfiler Headless

Note B.4. CellProfiler Commands for Headless Operation.

```
CellProfiler.exe cpOutput\DevPairs1-Ecto1-h5 -c -r -p MegaQuantProduction.cpproj -f
1 -l 266
CellProfiler.exe cpOutput\DevPairs1-Ecto2-h5 -c -r -p MegaQuantProduction.cpproj -f
267 -l 532
CellProfiler.exe cpOutput\DevPairs1-Ecto3-h5 -c -r -p MegaQuantProduction.cpproj -f
533 -l 798
CellProfiler.exe cpOutput\DevPairs1-Ecto4-h5 -c -r -p MegaQuantProduction.cpproj -f
799 -l 1064
CellProfiler.exe cpOutput\DevPairs1-Ecto5-h5 -c -r -p MegaQuantProduction.cpproj -f
1065 -l 1330
CellProfiler.exe cpOutput\DevPairs1-Ecto6-h5 -c -r -p MegaQuantProduction.cpproj -f
1331 -l 1596
CellProfiler.exe cpOutput\DevPairs1-Ecto7-h5 -c -r -p MegaQuantProduction.cpproj -f
1597 -l 1862
CellProfiler.exe cpOutput\DevPairs1-Ecto8-h5 -c -r -p MegaQuantProduction.cpproj -f
1863 -l 2128
```

B.5 Python Scripts Used for HDF5 Pre-processing

Code B.4. mergeh5files.py. Example pair-wise experiment script demonstrating merging of the HDF5 experimental fragments generated while running CellProfiler headless.

```
from glob import glob
import h5py
import re
import numpy as np
from copy import deepcopy
import os
import pandas as pd
```

```
import skmUtils

def keyAppend(name):
      keyList.append(name)

# Get list of all h5 files.
h5list = glob('*.h5')

# Refine list to those that were broken into parts to process.
h5list = [x for x in h5list if len(filter(str.isdigit, x)) == 3]

# List of the experiment names that will be condensed.
expList = list(set([x[:-4] if x[-5] != '-' else x[:-5] for x in h5list]))

numRE = re.compile('.*/\d\d\d\d-\d\d-\d\d-\d\d-\d\d-\d\d/')

# For each of the experiments...
for expNum, exp in enumerate(expList):
      print 'Processing ' + exp + ', ' + str(expNum) + ' of ' + str(len(expList))
      # Get the respective constituent files.
      fileList = [x for x in h5list if exp in x]

      # Get the complete feature list.
      h5temp = h5py.File(fileList[0], 'r')
      print '\tProcessing ' + fileList[0]
      keyList = []
      h5temp.visit(keyAppend)
      keyList = [x for x in keyList if type(h5temp[x]) == h5py._hl.dataset.Dataset]
      imFeatures = [key for key in keyList if '/Image/' in key]
      obFeatures = [key for key in keyList if (('/c1ob/' in key or '/c2ob/' in \
            key or '/c3ob/' in key) and ('Relationship' not in key))]

      # Slim down to desired features paths.
      imFeatures = [x for x in imFeatures if 'Channel_' not in x]
      imFeatures = [x for x in imFeatures if 'ExecutionTime_' not in x]
      imFeatures = [x for x in imFeatures if 'Frame_' not in x]
      imFeatures = [x for x in imFeatures if 'Group_Index' not in x]
      imFeatures = [x for x in imFeatures if 'Group_Number' not in x]
      imFeatures = [x for x in imFeatures if 'Height_' not in x]
      imFeatures = [x for x in imFeatures if 'ImageSet_' not in x]
      imFeatures = [x for x in imFeatures if 'MD5Digest' not in x]
      imFeatures = [x for x in imFeatures if '_FileLocation' not in x]
      imFeatures = [x for x in imFeatures if '_Frame' not in x]
      imFeatures = [x for x in imFeatures if '_Series' not in x]
      imFeatures = [x for x in imFeatures if '_channel' not in x]
      imFeatures = [x for x in imFeatures if '_Well' not in x]
      imFeatures = [x for x in imFeatures if 'ModuleError_' not in x]
      imFeatures = [x for x in imFeatures if 'PathName_' not in x]
      imFeatures = [x for x in imFeatures if 'Scaling_' not in x]
      imFeatures = [x for x in imFeatures if 'Series_' not in x]
      imFeatures = [x for x in imFeatures if 'URL_' not in x]
      imFeatures = [x for x in imFeatures if 'Width_' not in x]
      imFeatures = [x for x in imFeatures if ('/index' not in x or \
            'ImageNumber/' in x)]

      obFeatures = [x for x in obFeatures if 'ObjectNumber' not in x]
      obFeatures = [x for x in obFeatures if ('/index' not in x)]
```

```
keyList = imFeatures + obFeatures

newKeyList = [re.sub(numRE, 'Measurements/0000-00-00-00-00-00/', x) for x \
      in keyList]

featureList = {x : h5temp[y][:] for x,y in zip(newKeyList, keyList)}

h5temp.close()

# Loop through the files and accumulate the features.
for fileName in fileList[1:]:
      h5temp = h5py.File(fileName, 'r')
      print '\tProcessing ' + fileName

      # Get the feature list.
      keyList = []
      h5temp.visit(keyAppend)
      keyList = [x for x in keyList if type(h5temp[x]) == \
            h5py._hl.dataset.Dataset]
      imFeatures = [key for key in keyList if '/Image/' in key]
      obFeatures = [key for key in keyList if (('/c1ob/' in key or \
            '/c2ob/' in key or '/c3ob/' in key) and ('Relationship' \
            not in key))]

      # Slim down to desired features paths.
      imFeatures = [x for x in imFeatures if 'Channel_' not in x]
      imFeatures = [x for x in imFeatures if 'ExecutionTime_' not in x]
      imFeatures = [x for x in imFeatures if 'Frame_' not in x]
      imFeatures = [x for x in imFeatures if 'Group_Index' not in x]
      imFeatures = [x for x in imFeatures if 'Group_Number' not in x]
      imFeatures = [x for x in imFeatures if 'Height_' not in x]
      imFeatures = [x for x in imFeatures if 'ImageSet_' not in x]
      imFeatures = [x for x in imFeatures if 'MD5Digest' not in x]
      imFeatures = [x for x in imFeatures if '_FileLocation' not in x]
      imFeatures = [x for x in imFeatures if '_Frame' not in x]
      imFeatures = [x for x in imFeatures if '_Series' not in x]
      imFeatures = [x for x in imFeatures if '_channel' not in x]
      imFeatures = [x for x in imFeatures if '_Well' not in x]
      imFeatures = [x for x in imFeatures if 'ModuleError_' not in x]
      imFeatures = [x for x in imFeatures if 'PathName_' not in x]
      imFeatures = [x for x in imFeatures if 'Scaling_' not in x]
      imFeatures = [x for x in imFeatures if 'Series_' not in x]
      imFeatures = [x for x in imFeatures if 'URL_' not in x]
      imFeatures = [x for x in imFeatures if 'Width_' not in x]
      imFeatures = [x for x in imFeatures if ('/index' not in x or \
            'ImageNumber/' in x)]

      obFeatures = [x for x in obFeatures if 'ObjectNumber' not in x]
      obFeatures = [x for x in obFeatures if ('/index' not in x)]

      keyList = imFeatures + obFeatures

      # Loop through the featuers...
      for key, newKey in zip(keyList, newKeyList):
            # If it's not a new set of features do nothing...
            if (np.array_equal(h5temp[key][:], featureList[newKey])):
```

```
                  pass
            # Else...
            else:
                  featureList[newKey] = np.concatenate( \
                        [featureList[newKey], h5temp[key][:]])

      h5temp.close()

# Make a handle for the new h5 file.
h5out = h5py.File(exp + '.h5')

# Initialize the data sets.
for newKey in newKeyList:
      h5out.create_dataset(newKey, data=featureList[newKey])

h5out.close()
```