IMAGE-BASED SCREENING OF HIGH-PERFORMING CLONES USING PHOTOACTIVATED CELL SORTING VIA DUAL PHOTOPOYLMERIZED MICROWELL ARRAYS

T. Sun[1] and J. Voldman[1*]

[1]Research Laboratory of Electronics, Massachusetts Institute of Technology, USA

ABSTRACT

We demonstrate an image-based dual-photoactivated cell-sorting method to screen high-performing Chinese Ovary Hamster (CHO) cells. Cellular organelles including mitochondria, lysosome and nuclei were stained and imaged at high resolution. Each cell was characterized by a unique signature comprised of >100 quantitatively extracted phenotypic features from the images. We show that low- and high-performing cells can be successfully distinguished by applying a machine learning-based classifier to the high dimensional cell signature matrix. Follow the cell classification, we presented a single cell sorting technology to retrieve the desired cells by dual-photopolymerization.

KEYWORDS: High-performing clone screen, Machine learning, photopolymerization, Single cell sorting

INTRODUCTION

Biologics (e.g., recombinant antibodies) are an increasing proportion of therapeutics, and are used to treat diseases ranging from arthritis to cancer. The process of producing a biologic entails introducing the specific gene into a host cell system, typically CHO cells, followed by screening hundreds of cell clones to hopefully find a few (10~30) producing enough biologic. The selected clones are then expanded by several passages and finally evaluated in larger-scale bioreactors. In order to isolate the best clones with the highest possible productivity and growth rate, it is common to spend several weeks during the multi-well plate screening step, which is time-consuming and laborious. Methods that select better-performing clones earlier in the process reduce development time and cost and are of great interest to the biotech industry. Strategies such as GFP-tagging the biologic to make clones flow sortable[1-3] have been explored to improve clone selection but have not been widely adopted by industry. The critical challenge is primarily due to the complexity of the protein synthesis and secretion process, which occurs inside the cell and associates with different cellular organelles. In the industrial cell line development, transcription from DNA to mRNA is probably not the limiting step, since the expressing vectors inside the nuclear are amplified through multiple rounds to provide high mRNA levels[4]. This suggests that the downstream process of translation may be the limiting factor, in which cellular organelles such as mitochondria get involved. Therefore, it is possible that phenotypic markers based on cellular organelles may be predictive of eventual clone performance. Based on this hypothesis, we are exploring whether visual phenotypic markers based on cell organelles correlate with downstream performance. Here we present a quantitative image-based prediction method that allows, for the first time, classification of low- and high-performing clones simply by imaging them using light microscopy. We couple this imaging and classification approach with a novel microfluidic cell sorting technology that combines photopolymerization with microwell arrays to sort the desired clones at high throughput (Fig 1B).

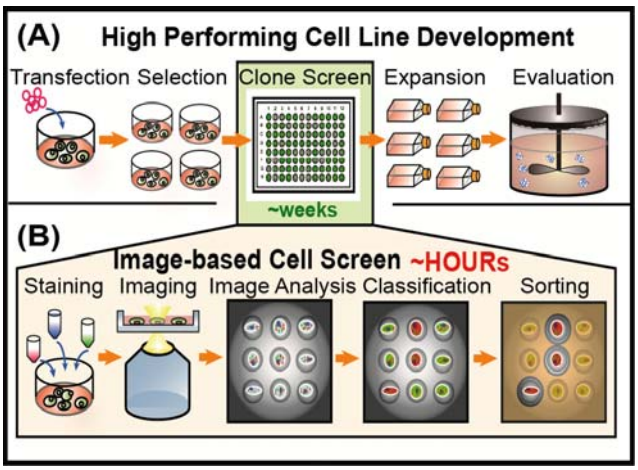

Figure 1: High-performing recombinant antibody clone selection. (A) Industrial cell line development process, requiring 4-6 months. The ability to pick up the ''best'' clones early in the process can decrease development time. (B)To speed up the clone selection, we developed an image-based cell screen method, involving live cell staining and imaging, quantitative image analysis, machine learning classification and microfluidic single cell sorting.

IMAGE-BASED CELL CLASSIFICATION

We received low-and high-performing clones from our collaborators at Pfizer, named as CHO_L (low-performing), which secretes 147 mg/L and CHO_H (high-performing), which secretes 422 mg/L. We viably stained different cellular organelles in the CHO cells, including mitochondria, lysosome and nuclei (Fig. 2A). We tested the cell viability after staining by continuously culturing the stained and unstained cells for 3 days and compared the growth rates between the two, showing

no difference (not shown). Phase and fluorescent images of cells were taken using a wide field fluorescent microscope with a 60X oil-immersion objective to image the details of the cellular structures. To eliminate imaging variations due to overall intensity from cell to cell, which can introduce artefacts, we developed an automated image processing and analysis pipeline that normalized and quantitatively extracted the morphological signature of every individual cell. This morphological signature was composed of >100 imaged cellular phenotypic features across phase and fluorescent channels, covering cell area, shape, intensity, texture, granularity, etc. (Fig. 2B).

To classify the cells according to their quantitative signatures, we used support vector machine (SVM), a machine learning algorithm to generate a classifier for CHO_L and CHO_H classification. We randomly picked up two known CHO_L and CHO_H populations as the training samples. The CHO_L population was given the score of '–1' while the CHO_H population was given the score of '+1'. The SVM algorithm computationally determines an optimal hyper-plane (classifier) in the multi-dimensional data space (cell signature matrix) to best separate the two populations in the training samples. The optimal classifier was determined by calculating 5-fold cross validation accuracy in the training samples and finding the SVM classifier parameters for achieving the maximum accuracy. After establishing the classifier, we applied this classifier to score independent datasets of CHO_L and CHO_H from the training samples. For each cell, if its computed score is >= +1, we classified it as a high-performing cell (CHO_H), while its computed score is <= –1, we classified it as a low-performing cell (CHO_L). For the cells receiving scores between –1 and +1, we treated those as unclassified cells. Figure 2C shows the scatter plot of SVM score vs. a feature value of the pure independent low- and high-performing cells, demonstrating the clear classification of two populations according to the SVM score distribution.

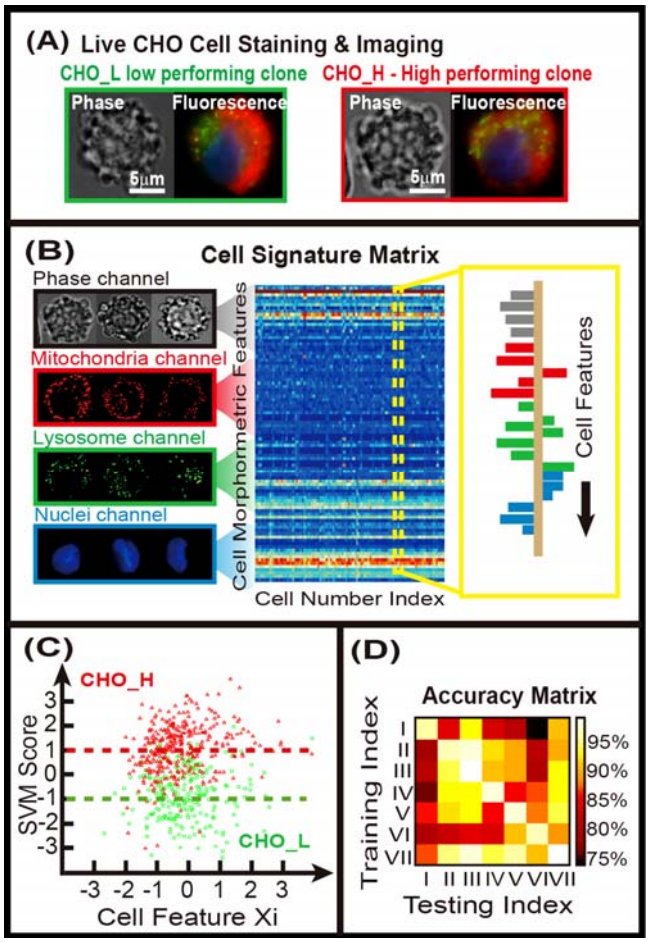

Figure2: Live-cell staining, imaging and classification. (A) Images of CHO cells (B)Image analysis to extract quantitative information from imaged cell organelles. (C) Scatter plot showing classification of independent pure CHO_L and CHO_H samples. (D) Accuracy matrix showing the classification accuracy across multiple independent experiments, demonstrating the robustness of the method.

It is also interesting to know which specific features significantly contribute to the classification. We employed different feature selection algorithms, e.g., Fisher Score, to rank all the features in the morphological signature. We found that the features from the phase channel, such as cell area, shape, have modest contribution to the classification, since CHO_L and CHO_H cells have similar size distribution and there is a wide heterogeneity in the shape of the cells in culture. Several phenotypic features categorized as texture in mitochondria and lysosome showed significant contributions (not shown). However, no single feature can dominate since the phenotypic features in the same fluorescent channel are not completely independent from each other. To test the robustness of the classifier, we performed the classification across 7 independent experiments. We took one experimental data set as the training samples for establishing the classifier and applied the classifier to the other experimental datasets. Classification accuracy is calculated as the ratio between the number of the true positive classified cells to the number of the total cells (Fig. 2D), showing that our image-based classification method is robust.

SINGLE CELL SORTING

To use the imaged information for cell line development, we next developed a method to retrieve classified high-producing clones. Existing approaches to cell sorting, such as flow sorting, are inadequate as they do not image. To enable image-based sorting, we developed a dual photoactivated sorting technology to retrieve the desired cells that uses two sequential photopolymerization steps using complementary photopolymers (Fig. 3A). The approach combines a photo-patterned microwell array with light-directed encapsulation of cells. We first photo-pattern a microwell array out of an optical photopolymer for massively parallel single-cell trapping (Fig. 3B). A microfabricated PDMS stamp was used to pattern microwells in optical adhesive polymer under UV activation. The stamp consisted of micro-pillars (50 μm diameter and

height), which was made from a photolithographically defined master. A droplet of photocurable pre-polymer was deposited onto a glass-bottom dish. The PDMS stamp was then gently pressed onto the droplet. After ultraviolet (UV) exposure, the PDMS stamp was peeled off, leaving cured microwell arrays on top of the glass. Using this microwell array we were able to image biologic-producing cells (Fig. 3B) under a high magnification oil-immersion objective, which has a short working distance of 130 μm. To sort the desired cells, an alignment mark image was generated, with black round-dot features (200 μm diameter) corresponding to locations of desired cells. After aligning the transparency mask to the back of the dish, opaque mask features resided beneath desired cells. We then mixed a second pre-polymer solution consisting of cell culture media, a photoinitiator, and poly(ethylene glycol) diacrylate (PEGDA) monomer. We added the pre-polymer to the dish and shined ultraviolet (UV) light from a standard fluorescence microscope light source through the transparency and into the dish. The pre-polymer then crosslinked into a hydrogel in all unmasked regions, encapsulating the undesired cells. The mask features shielded the desired cells away from the UV exposure, leaving them isolated in the unpolymerized region (Fig. 3C). To retrieve desired cells, we simply sorted by washing (Fig. 3D).

CONCLUSION

We demonstrate the first-ever classification of low- and high-performing clones from by imaging them using light microscopy. Our results show a promising way to predict the secretion performance of cells predicted by computationally learning the phenotypic features of cellular organelles. Coupled with our dual photoactivated single-cell sorting technology, which allows to isolate and retrieve cells according to their phenotypic markers in both of one-time shot and time lapse scenarios, our screening method is poised to significantly improve the biologic process efficiency and reduce the manufacturing cost.

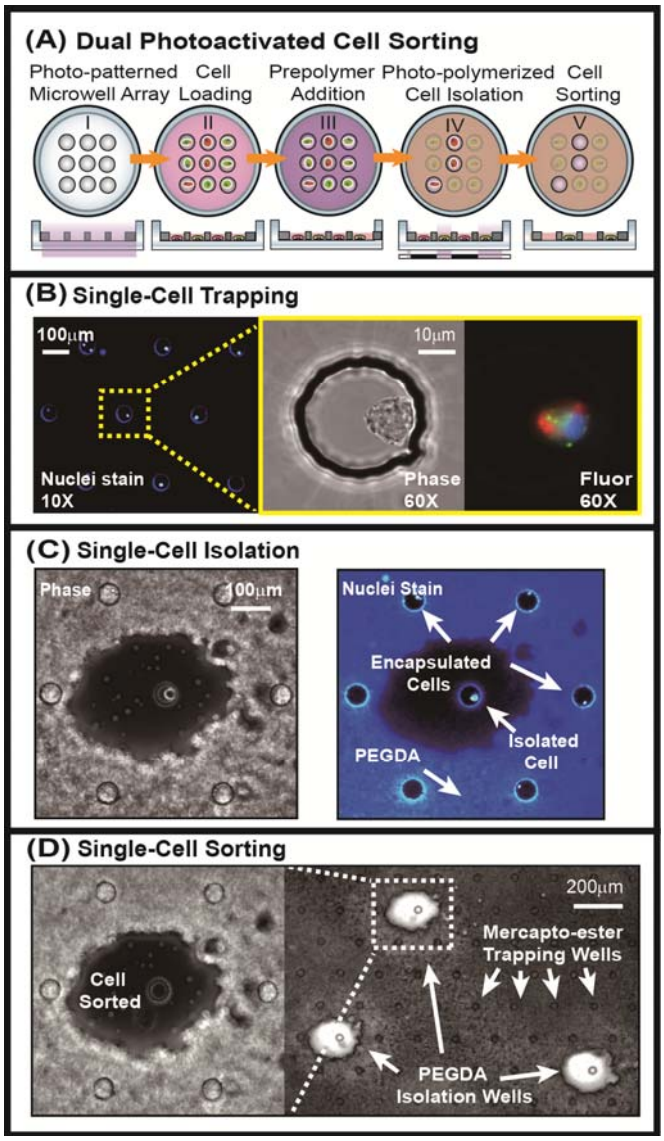

Figure 3: Microfluidic single-cell sorting. (A)Dual photoactivated cell sorting pipeline. (B)Image showing trapped single CHO cells in photo-patterned microwell arrays. (C) Images showing targeted cells isolated from surrounding cells after photopolymerization. (D) Closed-up and zoomed-out images of PEGDA isolation wells after cell sorting.

ACKNOWLEDGEMENTS

We acknowledge Dr. Lin Zhang from Pfizer Global Biologics R&D Laboratories at St. Louis for valuable discussions and project funding from Pfizer.

REFERENCES

[1] Meng, Y. G, et. al., "Green fluorescent protein as a second selectable marker for selection of high producing clones from transfected CHO cells" *Gene*, vol.242, pp. 201-207, 2000.

[2] Hanania, E. G., et. al., "Automated in situ measurement of cell-specific antibody secretion and laser-mediated purification for rapid cloning of highly-secreting producers", *Biotechnol. Bioeng*., vol.91, pp. 872-876, 2005.

[3] Demaria, C. T., et. al., "Accelerated clone selection for recombinant CHO cells using a FACS-based high-throughput screen", *Biotechnol. Prog*., vol.23, pp.465-472, 2007.

[4] Jayapal, K. P., et. al., "Recombinant protein therapeutics from CHO cells – 20 years and counting", *Chem. Eng. Prog.*, vol.103, pp.40-47, 2007.

CONTACT

*J. Voldman, tel: +1-617-253-2094; VOLDMAN@MIT.EDU