CellPress

Neuron
Review

Cell-Based Screening: Extracting Meaning from Complex Data

Steven Finkbeiner,[1,2,*] Michael Frumkin,[3] and Paul D. Kassner[4]
[1]Director of the Taube/Koret Center for Neurodegenerative Disease and the Hellman Family Foundation Program in Alzheimer's Disease Research, Gladstone Institute of Neurological Disease, San Francisco, CA 94158, USA
[2]Departments of Neurology and Physiology, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Director of Engineering, Research, Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
[4]Director of Research, Amgen, Inc., 1120 Veterans Boulevard South, San Francisco, CA 94080, USA
*Correspondence: sfinkbeiner@gladstone.ucsf.edu
http://dx.doi.org/10.1016/j.neuron.2015.02.023

Unbiased discovery approaches have the potential to uncover neurobiological insights into CNS disease and lead to the development of therapies. Here, we review lessons learned from imaging-based screening approaches and recent advances in these areas, including powerful new computational tools to synthesize complex data into more useful knowledge that can reliably guide future research and development.

Introduction

As integrative biology (Blow, 2009) reshapes paradigms in cell biology, it is increasingly clear that many of the phenotypes we seek to measure in isolation are highly connected to each other (Collinet et al., 2010). In fact, cells and their phenotypes exhibit many features that qualify them as complex systems: phenotypes are often emergent properties of dynamic signaling pathways with nested feedback loops, non-linear signaling relationships, and the capacity to undergo adaptive changes.

Complex systems are challenging to understand using standard hypothesis-driven experimental approaches that aim to manipulate one variable at a time. Holding every other variable constant in complex biological systems may be impossible or require artificial measures that confound results. Certain biases are unavoidable, and the investigator's knowledge and conceptual framework limit the pace of discovery. Indeed, the instincts scientists have developed by studying well-defined simple biological systems may mislead them as much as guide them when applied to complex systems. Breakthroughs and paradigm shifts are infrequent and often result from serendipity rather than intention because the hypotheses that drive experimental design evolve slowly from past results.

Discovery or "hypothesis-free" approaches are an important alternative. Whereas hypothesis-driven research is linear, discovery approaches are massively parallel. Since the "system"—the cell in this case—tells the investigator which perturbation is relevant, discoveries can be unexpected and less biased than findings from hypothesis-driven approaches. One of the most common applications is imaging-based phenotypic screens in cells. Cell-based screens have provided novel biological insights into the genes that control cell morphology (Jones et al., 2009), chromosome segregation and structure (Neumann et al., 2006; Walter et al., 2010), cell division, migration and survival (Neumann et al., 2010), susceptibility to infection (Cronin et al., 2009), and regulators of the protein clearance pathway autophagy (Orvedahl et al., 2011). In neuroscience, cell-based screens (Al-Ali et al., 2013) have been used effectively to investigate regenerative approaches to multiple sclerosis (Deshmukh et al., 2013) and synaptogenesis (Sharma et al., 2013; Shi et al., 2011).

Commensurate analysis tools must be applied that treat cells as defined but complex systems (Freddolino and Tavazoie, 2012; Karr et al., 2012). Fortunately, large-scale computational facilities are changing the nature of data analysis. They have increased the ability to access and search data, improved visualization techniques and technologies, enabled the application of powerful statistical techniques to large complex data sets, and made it possible to apply previously computationally untenable machine learning (ML) techniques to build predictive models of complex biological systems.

In the sections that follow, we will review some of the lessons learned from past efforts with cell-based screens, some important considerations for those pursuing these approaches now and for the future, including the challenges and opportunities created by the massive amounts of data that these screens can generate. Our focus is imaging and cell-based screens applied to neurobiology, though the concepts and approaches described here are widely relevant. We will look at the methods for acquiring images, how images are analyzed, the value of cloud computing and ML, and the implications of all of this for the future of biology and medicine.

Model Systems

There are several key components of any screen, and the first is the cells to be examined (Figure 1). This choice is critical and should be driven by the biological question rather than expedience. The basic choice is between immortalized cells and primary cells, and more recently, cells derived from induced pluripotent stem cells (iPSCs).

Ultimately, a compromise between feasibility and biological relevance may be needed to conduct a screen. The trade-offs—what can and can't be learned from the in vitro system and the endpoints examined—need to be understood clearly before starting. For the purposes of this discussion, biological relevance is the extent to which lessons from simple systems that are feasible to use for screening hold true for the more complex systems that they are meant to model. In this regard, validating models for use as screening platforms can be complex. If the screen is focused on an aspect of biology observed in vivo,

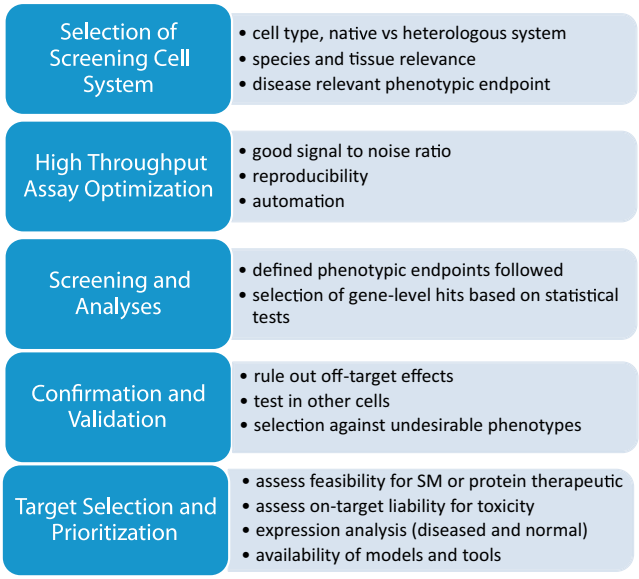

Figure 1. Generic Flow Scheme for Cell-Based High-Throughput Screening
The basic stages of target discovery and selection are outlined.

the emphasis for validation and ultimately model selection must be based on the ability of the in vitro model to replicate the critical in vivo biology. For screens focused on discovering treatments of a disease for which no effective therapies currently exist, the options for true validation are limited. Investigators must generally select a model based on a degree of face validity until an effective therapy is found, which can then be used to help validate and invalidate models. To be clear, no in vitro system will display the complexity of an intact organism, and not all biological insights will translate from in vitro to in vivo model systems or ultimately to human patients.

New Options with iPSCs

The Nobel Prize-winning discovery of cellular reprogramming by Takahashi and Yamanaka (2006), which led to the production of human iPSCs, offers new possibilities for disease models (Takahashi et al., 2007; Yu et al., 2007). Primary cells can be collected from people and reprogrammed into a stem or precursor cell that can be expanded and passaged (Churko et al., 2013; Hayes and Zavazava, 2013; Warren et al., 2010). In turn, iPSCs can be differentiated into cell types relevant to the disease, including subtypes of neurons and glia.

Protocols to make different brain cell types are being rapidly developed and improved. Many protocols involve the delivery of critical instructive factors to cells in culture at specific times and in a particular order to recapitulate key steps in development (Kim et al., 2014). For example, efficient protocols have been developed to make neural crest by dual-SMAD inhibition/WNT activation (Chambers et al., 2013). Protocols have been reported for making many brain cell types from stem cells, including dopaminergic neurons (Studer, 2012; Sun et al., 2013; Sundberg et al., 2013), motor neurons (Bilican et al., 2012; Boulting et al., 2011; Di Giorgio et al., 2007), forebrain-like neurons (HD iPSC Consortium, 2012), striatal neurons (Aubry et al., 2008), cortical interneurons (Maroof et al., 2013), retinal cells (Jin and Takahashi, 2012), oligodendrocytes (Czepiel et al., 2011; Ogawa et al., 2011; Wang et al., 2013; Yang et al., 2013), and astrocytes (Emdad et al., 2012; Serio et al., 2013). Neurons and neural progenitors can be produced directly from other types of somatic cells without having to first make those cells pluripotent (Ambasudhan et al., 2011; Kim et al., 2011; Vierbuchen et al., 2010).

The application of iPSCs to studying disease has generated the most excitement (Eglen and Reisine, 2011). For the first time, a skin or blood cell from a patient with a neurological or psychiatric disease can be reprogrammed to become a cell type of the nervous system, thereby creating a genetically faithful human model of disease (Churko et al., 2013; Hayes and Zavazava, 2013; Wray et al., 2012). Already, several models have been developed that exhibit disease-relevant phenotypes (Table 1) for Huntington's disease (HD) (HD iPSC Consortium, 2012; Zhang et al., 2010), amyotrophic lateral sclerosis (ALS) (Barmada et al., 2014; Bilican et al., 2012; Burkhardt et al., 2013; Donnelly et al., 2013; Egawa et al., 2012; Sareen et al., 2013; Serio et al., 2013), spinal muscular atrophy (Ebert et al., 2009), Parkinson's disease (Cooper et al., 2012; Skibinski et al., 2014), schizophrenia (Brennand et al., 2011), and Alzheimer's disease (AD) (Israel et al., 2012). In principle, genetic and SM screens can be conducted in what might be the most physiologically relevant cell-based model of neurological disease ever developed.

iPSCs might also help to solve one of the most vexing problems in drug development. Non-human models of neurological disease have a poor track record for predicting results of putative therapies in clinical trials (McGonigle, 2014; McGonigle and Ruggeri, 2014), including HD (Crook and Housman, 2011), ALS (Perrin, 2014), and AD (Mullane and Williams, 2013). Nearly all the compounds that were tested in human clinical trials and failed to show efficacy were supported by data showing that the drugs were effective in mice. There were two exceptions: tetrabenazine, a symptomatic therapy for HD, showed efficacy in mice, and riluzole had modest effects in a mouse model of ALS and extended the lives of ALS patients by a few months on average. Some discrepancy can be blamed on the design and execution of preclinical efficacy trials in mice (Perrin, 2014). But worrisome data suggest that more fundamental biological differences between mice and humans may be important. Humans and mice diverged in evolution over 65 million years ago, and many publications show that results from mice fail to reliably predict results from humans in drug absorption, distribution, metabolism, elimination, toxicity, bioavailability, carcinogenicity, teratogenicity, and efficacy, as well as disease pathophysiology. Known differences in pharmacodynamics (Richert et al., 2008; Xie et al., 2000) and toxicology (Carlson et al., 2009; Singh and Gupta, 1985) between humans and non-human models could affect drug safety. Differences exist in physiological responses and drug effects in human cells (e.g., neurons), compared to murine or other non-human counterparts (Berger et al., 2006; Castan et al., 1994; Curtis et al., 1997; Derian et al., 1995; Guo et al., 1989; Keshavaprasad et al., 2005; Kopin et al., 1997; Liang et al., 2010; Mattson et al., 1991; Okazaki et al., 1995; Penhoat et al., 1996; Rasakham