Methods 96 (2016) 12–26

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

A metric and workflow for quality control in the analysis of heterogeneity in phenotypic profiles and screens

Albert Gough [a,b,*], Tong Ying Shun [a], D. Lansing Taylor [a,b], Mark Schurdak [a,b]

[a] University of Pittsburgh Drug Discovery Institute, 3501 Fifth Avenue, Pittsburgh, PA, USA
[b] Dept. of Computational and Systems Biology, University of Pittsburgh, 3501 Fifth Avenue, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 9 August 2015
Received in revised form 12 October 2015
Accepted 13 October 2015
Available online 4 November 2015

Keywords:
Heterogeneity
Phenotypic profiling
High content screening
Systems biology
Drug discovery

ABSTRACT

Heterogeneity is well recognized as a common property of cellular systems that impacts biomedical research and the development of therapeutics and diagnostics. Several studies have shown that analysis of heterogeneity: gives insight into mechanisms of action of perturbagens; can be used to predict optimal combination therapies; and can be applied to tumors where heterogeneity is believed to be associated with adaptation and resistance. Cytometry methods including high content screening (HCS), high throughput microscopy, flow cytometry, mass spec imaging and digital pathology capture cell level data for populations of cells. However it is often assumed that the population response is normally distributed and therefore that the average adequately describes the results. A deeper understanding of the results of the measurements and more effective comparison of perturbagen effects requires analysis that takes into account the distribution of the measurements, i.e. the heterogeneity. However, the reproducibility of heterogeneous data collected on different days, and in different plates/slides has not previously been evaluated. Here we show that conventional assay quality metrics alone are not adequate for quality control of the heterogeneity in the data. To address this need, we demonstrate the use of the Kolmogorov–Smirnov statistic as a metric for monitoring the reproducibility of heterogeneity in an SAR screen, describe a workflow for quality control in heterogeneity analysis. One major challenge in high throughput biology is the evaluation and interpretation of heterogeneity in thousands of samples, such as compounds in a cell-based screen. In this study we also demonstrate that three heterogeneity indices previously reported, capture the shapes of the distributions and provide a means to filter and browse big data sets of cellular distributions in order to compare and identify distributions of interest. These metrics and methods are presented as a workflow for analysis of heterogeneity in large scale biology projects.

1. Introduction

Heterogeneity is a property of cellular systems that implies the presence of cell-to-cell variability with respect to one or more measurable traits (phenotypes) at the molecular-cellular level. Because heterogeneity is a property of the population of cells, it can be complex and therefore difficult to characterize as a single parameter. Large scale biology projects typically span many days or even months, and non-biological variation in heterogeneity over the course of a study could mask the true biological heterogeneity, suggesting the need for quality control. The details of the distributions of a trait can be visualized as a histogram of the cell population, but visualization of the large numbers of histograms generated in high throughput phenotypic screens is not practical without appropriate informatics tools, and qualitative comparisons are not sufficient for quality control. In this chapter we present a simple metric for quality control of heterogeneity, and an approach to evaluating heterogeneity in large-scale biology and drug discovery programs consisting of measurements on thousands to millions of samples, which can be applied to High Content Screening, High Throughput Microscopy, flow cytometry, digital pathology or any method that generates large datasets of cells [1–6].

1.1. Implications of cellular heterogeneity in biomedical research, drug discovery and diagnostics

It is now well established that cellular heterogeneity is an intrinsic feature observed in cell systems, even when composed

Abbreviations: QC-KS, a heterogeneity quality control metric that uses the KS statistic; HI, heterogeneity index; QE, quadratic entropy, an HI; nNRM, non-normality, an HI; %OL, percent outliers, an HI.

* Corresponding author at: University of Pittsburgh Drug Discovery Institute, 3501 Fifth Avenue, BST3 10052, Pittsburgh, PA 15260, USA.
E-mail address: Gough@pitt.edu (A. Gough).

http://dx.doi.org/10.1016/j.ymeth.2015.10.007
1046-2023/

of isogenic populations of cells. Cell-to-cell variability is not simply the result of intrinsic noise in molecular networks; rather it is often the result of deterministic regulatory molecular mechanisms that remain largely uncharacterized [7–9]. There is growing evidence that some heterogeneity is related to physiological and evolutionary adaptations to new challenges [10,11]. A recent study suggests that heterogeneity can be decomposed into different groups of biomarkers that are consistent with known signaling pathways, implying a mechanistic basis for the cell-to-cell variation [12]. It has also been shown that patterns of signaling heterogeneity can distinguish cellular populations with different drug sensitivities [13,14]. It has been proposed that combination therapies can be devised through computational analysis of heterogeneity [15]. Furthermore, the association between cellular heterogeneity and adaptation [10] suggests that ignoring the heterogeneity in the cellular response may lead to the selection of compounds to which cells will rapidly adapt, leading to a loss of efficacy in drug development programs. Thus, it is important to incorporate heterogeneity analysis in large scale screens in drug discovery, and phenotypic profiles in basic biomedical research, where the goal is to relate a cellular phenotype with some condition, such as a disease state, compound effect, protein knockdown, or other sample state or treatment.

Pluripotent stem cells are a platform with tremendous potential for development of patient specific disease models, for modeling biological development, and for regenerative medicine. However, stem cells exhibit heterogeneity on several levels: in the functional capacity to differentiate; in mRNA expression profiles; and in epigenetic and genetic state [16]. Studies of differentiating cells have found that heterogeneity reflects the presence of an evolving mixture of phenotypically distinct subpopulations, consistent with a hypothesis that differentiating cells transit through multiple robust and discrete phenotypic states [17–19]. Improved understanding and manipulation of the differentiation of stem cells will require tools to reliably characterize and monitor the evolution of the distributions of these subpopulations and their associated phenotypes.

In typical cell-based phenotypic assays, compound activity is characterized by the well average value of a feature or a combination of features, but the cell level data is usually ignored in order to achieve higher throughput. Metrics have been developed to address assay variability for large scale screens/profiles including signal-to-background and Z'-factor for inter-plate and inter-day variability [20–22]. However, these metrics do not address the biological heterogeneity present in individual cell populations. Therefore, new tools and criteria are required for analyzing and comparing biological heterogeneity in large scale screens/profiles, and for sorting and identifying perturbagens with potentially interesting or novel effects on the distribution of cell response among a large number of patterns of cellular activity. Measuring heterogeneity is particularly important when applying quantitative systems pharmacology (QSP) to drug discovery programs. A key tenet of our implementation of QSP is the need to have phenotypic assays that reflect the heterogeneity in patients as determined by a range of "omics" profiles.

1.2. Challenges in the analysis and interpretation of heterogeneity in high content phenotypic screens/profiles

There are several challenges to large scale analysis of heterogeneity and the comparison of heterogeneity between samples, assay runs, and laboratories. These challenges include: (1) large size and complexity of the cell level data in high content screens/profiles; (2) lack of a standard approach to quality control in measuring heterogeneity; (3) lack of standards for the characterization of the assay and detection system contributions to the measured heterogeneity; and (4) lack of adequate tools to quickly quantify, compare, review and interpret heterogeneous responses.

In large scale analysis of biological activity taking the average of a population is a quick way to reduce the data to a size that can be more easily managed, filtered and interpreted, but neglects the information contained in the distribution of cellular responses. The complexity of the cell distributions and the lack of a standard analytical approach for interpreting heterogeneity remains a barrier to incorporating such analyses into a project, despite evidence that investigation of heterogeneity leads to new insights into the underlying cellular functions [23,24]. Analyzing all the cellular data in a compound screen, RNA_i screen, genome-wide mutagenesis [25] or other large biology project is a "big data" problem that has not been adequately addressed. For a screen of 100,000 compounds and 1000 cells/well there will be 100 million data points for each feature (potentially 4–20 or many more) measured in the screen. Similarly, an SAR of ~240 compounds with 10 concentrations, triplicate wells, duplicate runs and $\geqslant$1000 cells/well, such as the data set evaluated in this chapter, can easily amount to well over 10 million cells. Data that is well organized in a database may be easily extracted by query, but efficiently analyzing 30–500 million or more data points acquired over several months for patterns in population distributions at least requires some thought and planning, and tools for manipulating and comparing distributions. In this chapter we present an approach to that process.

A second challenge in the heterogeneity analysis is quality control. The standardization of quality control in high throughput screening through the use of the Z'-factor [22,26], or the strictly standardized mean difference [27], has enabled comparison of assay performance over a wide range of assay types, routine monitoring of assay performance during a screen, and a statistical basis for understanding the impact of assay variability on the selection of hits from a screen. For dose–response assays, the V-factor, a generalization of the Z-factor, provides a robust QC measure that takes into account the complete dose–response, rather than just the maximum and minimum response [21,28]. To be able to compare cellular heterogeneity in order to interpret the biological meaning, standards and quality control are needed to ensure reproducibility. Present assay development guidelines for high content phenotypic screens address metrics for assay variability [20,21], but do not address the reproducibility of the distributions within the wells. The normalization of well average data to on plate controls adequately corrects for plate-to-plate variability at the well level but is not typically applied to the cell level data. It is important to emphasize that in typical use, good Z'-factors do not necessarily indicate that the cellular distributions (i.e. heterogeneity) within the wells are consistent from plate-to-plate, as we demonstrate here.

A third challenge in heterogeneity analysis is the lack of reference standards for calibration or characterization of most high content methods and in particular phenotypic methods. For example, in HCS spatial or temporal variation of illumination intensity or detection sensitivity can contribute "detection system variation" to the measured values. While approaches to detection system calibration and characterization have been published [14,29–31], so far standards such as those routinely used in flow cytometry [32] have not been widely adopted in imaging. Typically in screening, each plate will be normalized to the control wells [20,21]. This approach compensates for the average signal level on each plate; however, to compare distributions from plate-to-plate also requires normalization of the cell level data between plates. Optimally, attention to protocols during assay development, validation and screening/profiling can minimize the variation from plate-to-plate. Here we demonstrate that normalization can be an effective means to compensate for plate-to-plate variation in the range of intensities, but not necessarily the variation in the shape of the