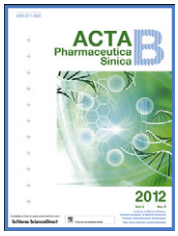
Acta Pharmaceutica Sinica B 2012;2(5):429–438

Institute of Materia Medica, Chinese Academy of Medical Sciences
Chinese Pharmaceutical Association

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

Quality control of cell-based high-throughput drug screening

Zhiyun Zhang[a], Ni Guan[a], Ting Li[a], Dale E. Mais[b], Mingwei Wang[a,*]

[a]*The National Center for Drug Screening and the State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*
[b]*MPI Research, Mattawan, MI 49071, USA*

Received 9 February 2012; revised 12 March 2012; accepted 19 March 2012

KEY WORDS

High-throughput screening;
Cell-based assay;
Quality control

Abstract The pharmaceutical industry is presently suffering difficult times due to low productivity of new molecular entities. As a major source of drug leads, high-throughput screening (HTS) has been often criticized for its 'dead end' lead compounds. However, the fruitful achievements resulting from HTS technology indicate that it remains a feasible way for drug innovation. Because of increasing considerations of earlier stage ADMET (absorption, distribution, metabolism, excretion and toxicity) in drug development, cell-based HTS is highly recommended in modern drug discovery for its ability to detect more biologically relevant characteristics of compounds in living systems. This review provides a systematic and practical description of vital points for conducting high quality cell-based HTS, from assay development to optimization, compound management, data analyses, hit validation as well as lead identification. Potential problems and solutions are also covered.

Abbreviations: R&D, research and development; ADMET, absorption, distribution, metabolism, excretion and toxicity; cAMP, cyclic adenosine monophosphate; GPCR, G protein-coupled receptor; NR, nuclear receptor.

*Corresponding author.
E-mail address: wangmw@mail.shcnc.ac.cn (Mingwei Wang).

Peer review under responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

ELSEVIER Production and hosting by Elsevier

2211-3835 © 2012 Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association. Production and hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.apsb.2012.03.006

1. Introduction

High-throughput screening (HTS), driven by the great progress in automation technology and combinatorial chemistry, has been widely implemented in drug discovery since the early 1990s and rapidly became one of the major sources of drug leads. Pharmaceutical companies, such as Pfizer and Glaxo SmithKline, were among the early leaders. In the past twenty years or so, many academic institutions joined the 'screening mania' and simultaneously, hundreds of screening centers appeared, as molecules available for screening continued to increase. However, in spite of constant increases in research and development (R&D) expenditures, the number of new chemical entities (NCEs) that reached to the market has actually decreased[1,2]. Analyses show that leads originating from HTS are responsible for over 60% of clinical trial failures[3,4]. Nonetheless, the data also indicate that among 58 drugs that were approved between 1991 and 2008, 19 were attributed to HTS[5]. Without question, HTS is still a feasible approach to drug innovation. The problem becomes one of how to improve the quality of leads arising from drug screening that may result in increased productivity of new molecular entity (NME) entering to the market place.

Since HTS has not substantially improved the drug discovery process and increased R&D spending has not led to a proportionate increase in new drug output, the pharmaceutical industry is looking back to the golden age of phenotype-oriented drug discovery[6] and considering ADMET (absorption, distribution, metabolism, excretion and toxicity) of leads at earlier stages in drug development. A potential way to do this is through the use of cell-based assays. Cell-based assays not only obtain potencies of compounds but also detect cytotoxicity, permeability and effects on growth at the same time, which can be viewed as predictors for late development. Cell-based assays accounted for 52.6% of all HTS efforts in 2006[7] and became more favorable in recent years. However, cell-based assays are generally more complicated than biochemical ones and their performance could be undesirable under certain circumstances. Thus, quality control is of paramount importance and will be discussed in detail below. Although several comprehensive reviews are accessible in the public domain, this article attempts to give key points relevant to carrying out high quality cell-based HTS in a systematic and practical manner with potential problems and solutions highlighted.

Basically, HTS program consists of five parts: target identification, reagent preparation, assay development, compound management and high-throughput screening[8]. Among them, target identification and reagent preparation are beyond the interest of this review, although both of them are vital for successful HTS. Instead, we will cover topics such as assay development and optimization, compound management and data analysis, as well as hit identification and lead validation. High content screening (HCS), as an important part of cell-based HTS, has attracted significant attention recently because of its multiplexing and functional cell based characteristics[9]. However, considering the complexity of its data analysis, HCS is not included in this discussion.

2. Assay development

Cell-based assays or screening models, as the fundamental ingredient of HTS, are approaches used for sensing functional changes of targets under the stimulatory or inhibitory effects of compounds. In biochemical assays, targets are generally specified, while for cell-based assays, the exact target is not required. It could be a specific molecule or a particular signaling pathway, even the whole cell. For example, in cell death assays[10], organisms such as bacteria, fungi, parasites and mammalian cells are directly used as screening models. These whole cell based screenings are highly physiologically relevant, thus providing opportunities to discover entirely novel drugs and drug targets. However, subsequent pharmacological characterization and target identification could be exhaustive. Most of the time, specific targets are decided as soon as screening assays are proposed.

Cell lines used for HTS can be roughly divided into two classes, primary and engineered cells. With technology advancements, such as HCS, ion channel patch-clamp and atomic force microscopy, screenings with primary cells become increasingly feasible and trendy[11]. Several selected primary cell types, originating from human or other species, are commercially available (e.g., Clonetics, Walkersville, MD, USA) and amenable to HTS. As far as mammalian cell based assays are concerned, large-scale primary cell culture still poses some difficulties. Therefore, engineered cells remain the major type of cell lines used in HTS. In the following discussion, we offer some general ideas and tips for generation of engineered cell lines and related detection methods.

2.1. Cell line generation

To provide sufficient signal output for detection, cell-based assays require high expression of targeted proteins, which in naive cells is often low and needs to be up-regulated through either transient or stable gene transfection[12]. The transient transformed cell lines briefly express high level of targeted proteins, but display relatively larger variances in expression quality due to transfection inefficiencies. Stable transformed cell lines consistently express targeted proteins over a long period of time, while their expression levels are usually not as high as transient ones. Both strategies can be employed in HTS[13], but stable expression is much more preferred because of reduced cost and less assay variation[14].

Generally, gene transfection requires primary knowledge about the sequence of a targeted protein for vector construction. It is noted that some genes of interest are protected by patents that prohibit commercial use. Alternative strategies must be sought under such a situation: one can either increase the expression level of a particular gene through activation of internal gene scripts[15,16] or introduce specific transcriptional factors[17]. Sometimes, simply increasing the expression of a targeted gene is not sufficient to yield a high signal output and genes involved in the same signaling pathway may also need to be enhanced. For instance, G protein enriched cell lines, such as $G_{\alpha}16$ or $G_{qi}5$ transfected CHO cells, are preferred to screening G protein coupled receptor (GPCR) modulators.

2.2. Detection methods

A variety of methods, such as reporter gene, fluorescence/bioluminescence resonance energy transfer (F/BRET), calcium mobilization and label-free detection, have been applied to

cell-based HTS. With different utilities, they provide valuable approaches to the screening of small molecular modulators.

2.2.1. *Reporter gene assay*

Many receptors, kinases and transcriptional factors regulate gene expression either directly or indirectly through signal transduction. By accessing these signal cascades, reporter gene technique offers great sensitivity for signal detection in cell-based assays. As we know, the sensitivity and efficiency of reporter gene systems are largely affected by both reporter gene and promoter upstream. In order to guarantee wide detecting windows and high resolutions, promoters with very low constitutive activities should be sought. This will ensure an optimal background and significant response in the presence of a stimulus. For reporter gene products, an appropriate level of stability is also vital. Long half-time may lead to an increased background resulting from excess accumulation of reporter gene encoded proteins. This could be caused by constitutive activation of promoters or lagging response to stimulations. Clearly, rational selection of both promoter and reporter gene according to the purpose and requirement of an assay is an indispensable step before HTS. Reporter genes that are commonly utilized today include green fluorescent protein (GFP), β-galactosidase, firefly luciferase, Renila luciferase and β-lactamase. The most widely used bioluminescent reporter genes are firefly luciferase from *Photinuspyralis* and Renilla luciferase derived from the sea pansy *Renillareniformis*. These bioluminescent reporter assays are of both high sensitivity and extraordinary accuracy. Because the readout is determined by efficiency of signal transduction and cellular physical/metabolic integrity, however, the false positive rate can be high[18]. Efforts in follow-up hit verification with secondary assays are generally required as a routine.

2.2.2. *cAMP measurement*

cAMP is one of the most important second messengers involved in a variety of physiological responses and is thus widely measured in HTS. Present methods for cAMP measurement include fluorescence polarization (FP)[19], HitHunter EFC technology based on β-galactosidase activity complementation, cyclic-nucleotide-gated ion channel (CNGC) coupled cAMP assay based on membrane potential[20], as well as cAMP-Glo assay developed by Promega. All of them reflect a decline of fluorescent signals in response to an increase in cAMP concentration. Other cAMP assays, based on reporter gene by splicing the reporter luciferase coding sequence under the control of Cre (causes recombination), measure elevations of cAMP levels exhibited as an increase in fluorescence/luminescence intensity. Obviously, this approach is labor intensive (requires more than 4 hours) and reflects cAMP concentration indirectly. Although cAMP measurements tend to offer more accurate and physiologically relevant results with low false positive hits, some of them require special apparatus/fittings (e.g., optical filters) and expensive regents (e.g., engineered antibodies fused with compensatory enzyme or fluorescent indicators).

2.2.3. *FRET and BRET techniques*

Protein–protein interaction is one of the key problems in mechanistic studies of critical cellular processes and diseases. Fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) are among a range of available methods for studies on such interaction. In brief, the resonance energy transfer refers to energy transfer between fluorescent/luminescent donor and acceptor molecules when they are in close proximity (1–10 nm) with properly oriented dipoles. Frequently used donors include blue fluorescent protein (BFP) and cyan fluorescent protein (CFP) for FRET, and luminescent products of luciferase for BRET. Commonly used acceptors include GFP, Venus, Citrine, YPet and mOrange. Efficiency of transfer and resolution of detection are greatly influenced by spectral properties of both donors and acceptors, in addition to expression ratio of donor and acceptor fusion proteins[21]. Donor–acceptor pairs with compatible spectral properties ensure efficient energy transfer and high signal to noise ratio. The emission spectrum of a donor must to some extent overlap the excitation spectrum of the acceptor to enable high efficient transfer, but a pronounced overlap would, on the other hand, sacrifice detection resolution resulting in poor signal to noise ratio. Moreover, efficiency is also markedly affected by comparative expression of donor *vs.* acceptor fusion proteins, especially when they form homogeneous dimmers[21]. Thus, it is critical to optimize protein expression pattern before prior to HTS campaign.

2.2.4. *Calcium mobilization*

Cellular Ca^{2+} is essential to a large number of physiological processes ranging from embryonic development to muscle contraction. In silent status, the cytosol Ca^{2+} is maintained at a very low level around 100 nM, however, it would burst to several mM with appropriate stimulation. The calcium influx is triggered by activation of Ca^{2+} channels on the plasma membrane or release of Ca^{2+} from endoplasmic reticulum. Today, numerous Ca^{2+} indicators have been developed for quantification of this important second messenger, involving protein indicators such as aequorin and chemical indicators such as Fluo-4. Protein indicators stably expressed in the cell could significantly reduce screening complexity, thereby increasing throughput compared with chemical indicators which may need additional steps for dye loading. One drawback relates to the extra efforts required for generation of cell lines that express desired indicators: some of them are now commercially available (e.g., aequorin). In contrast, chemical Ca^{2+} indicators are more conveniently used nowadays due to their wide range of spectral properties, large detection windows and robust performance.

2.2.5. *Label free methods*

Although reporter gene and fluorescent/luminescent probe based assays have been widely deployed in HTS with great successes, transfection of engineered cells and dye loading manipulation sometimes cause distortion in cell physiology, leading to loss of data fidelity and reliability[22]. The emergence of innovative label-free platforms aims at overcoming the limitations of label-based measurements by providing non-invasive means of detection. This approach enables real-time trace of kinetic information concerning a signal transduction process that is inevitably omitted by conventional traditional end-point assays. Current label-free systems used in cell-based investigations could be categorized into three major classes according to different detection principles: quartz microbalance, refraction index and cell impedance[22–25]. Quartz microbalance and refraction index are frequently employed

in evaluation of binding properties of drug candidates, and are applicable to HTS. Impedance could be used in screening for potential leads that modulate a variety of biological events such as cell attachment, apoptosis, migration and activities of specific receptors such as GPCRs and tyrosine kinases. For example, CellKey system and RT-CES are amenable for 96- and 384-well plate based HTS with a throughput around 20,000 samples per 6 hours. However, significant cost associated with special assay plates makes these methods less attractive in large-scale operations.

2.2.6. *Membrane potential*

Membrane potential can be determined by negatively charged indicators based on direct optical fluorescence or FRET[26]. The efficiency and accuracy of such measurement are largely dependent upon the redistribution of voltage-sensitive indicators: the slower the speed, the less the throughput. Since ion channels only respond transiently when activated, membrane potential changes must be promptly recorded. To increase the throughput one can engineer cells whose fast response channels are kept open during the redistribution process.

2.2.7. *Patch clamp*

This technique is the gold standard in monitoring physical changes of ion channel, while its traditional utility is confined by low throughput and high manipulation skill. Recently, remarkable improvements have been made in terms of its amenability to HTS. Contemporary patch clamp systems are mostly based on the planar chip recording method, such as IonWorks Series and PatchXpress 7000 developed by Molecular Devices[27,28], Qpatch Series from Sopohion and Patch linear from Nanion Technologies GmbH. Although these instruments are adapted to robust HTS, cost per compound is still considerable that hinders general uses at large. Nevertheless, high quality of biologically relevant data and large quantities of information generated by patch clamp make it an attractive tool to be further explored[29,30].

2.2.8. *Other approaches*

In addition to the methods described above, there remain a variety of approaches applicable to cell-based HTS, such as flow cytometry, HCS, atomic force microscopy and FP. HCS, an image analysis oriented technology, is becoming a core competence in many HTS laboratories because it captures multiplexing and functional cellular characteristics with remarkably rapid speed and high efficiency. HCS is beyond the scope of this review due to its complexity. While flow cytometry is presently capable of screening in small-scale modulators of cell cycle, apoptosis and proliferation, FP is widely utilized in protein–protein interaction related HTS.

Clearly, choice of HTS methods depends on the nature of a target in question. Most commonly used targets include receptors (GPCRs and nuclear receptors, NRs), ion channels and enzymes. GPCR assays often involve measurement of secondary messenger ($Ca^{2+}$and cAMP) and reporter gene. Other than these, GPCR modulators can also be screened by employing the label-free CellKey system developed by Medicine Devices. For NRs, specific ligands can be screened by fluorescence/luminescence, location-dependant reporter genes as well as FRET. To assess ion channel activities, membrane potential, secondary ion messenger (Ca^{2+}) and patch clamp techniques can be applied[27], whereas screening for modulators of enzymes always relies on their ability to catalyze transformation of pro-luminescent molecules to luciferase substrates. The latter can be further converted to bioluminescent molecules readily for luminescence detection (e.g., amino-luciferin for caspase-3/7 and methyl ester luciferin for monoamine oxidase). For screening without a specified target, such as induced pluripotent stem (iPS) cell stimulators or antibiotics, image-based methods (HCS and label-free assays) are preferable.

3. HTS optimization

Once an assay has been created, optimization is crucial before application to large scale screening. Pilot tests are usually carried out in small scale for a minor batch of known compounds to determine whether the assay is sufficiently feasible and reliable for HTS. Besides these, a successful HTS campaign also requires robust performance on a large scale, under automated and high speed conditions. Parameters (consistent cell culture, assay miniaturization, microtiter plate setting, assay automation, *etc.*) should all be carefully considered[31,32] during scale expansion.

3.1. *Cell culture*

Cell quality is critical in terms of generating consistent HTS results. However, sustaining a supply of high quality cells can always be a challenge, especially when employed cell lines show poor proliferation. Commonly, cells with the same passage age are cryo-preserved in large amounts before use, thawed and cultivated in flasks or plates to desirable confluence when HTS starts. They are then harvested and seeded into batches of 96/384 microtiter plates for screening. The redundant processes of cultivation and harvest before seeding may introduce variations in cell quality over time as well as decrease the speed and automation of screening. One of the alternative solutions may be the direct use of cryo-preserved cells. Without secondary thawing and cultivation, this approach can significantly accelerate the screening process and increase assay precision[33]. Obviously, this will require additional optimization (e.g., selection of cryo-preservation medium) to make sure that the recovered cells are fully functional and stable. Another solution is to utilize division arrested cells. It is known that cells can be irreversibly division arrested by mitomycin C at a dose that causes no apparent toxicity or obvious changes to signaling properties[34,35]. Division arrested technique allows HTS to be performed with cells in the same phase, thus increasing the accuracy of the output. Combination of these methods can provide a more precise way for running cell-based HTS. Fortunately, automated cell culture facilities specially developed for HTS to reduce labor intensity are now commercially available such as Cell Host system manufactured by Hamilton[36].

3.2. *Miniaturization*

In the early to mid-1990s, most HTS campaigns were carried out in 96-well microtiter plates. With advances in miniaturization and detection technologies, current HTS assays are usually performed in 384- or 1,536-well plates[37]. Miniaturization increases throughput, improves variation between wells

but reduces signal intensity. In the 384-well format, 10,000–40,000 cells per well are usually required to elicit consistent responses[38]. When the plate density increases to 1536 wells, the number of cells in each well decreases to thousands, which inevitably reduces the resistance to cell diversity resulting in large variances between wells. In addition, small volume in an extremely miniaturized assay is more sensitive to environmental changes, such as temperature and humidity, *etc*. If long incubation time is required, significant edge effects are likely to appear. Several successful examples of 1,536-well format HTS were reported in the literature[39], but concerns relative to the quality control of conducting extra large campaigns remain. It is understandable that the scale of miniaturization is closely associated with the nature of an assay and the detection capability of measuring devices. Assays based on endpoint detection, such as FRET/BRET, reporter gene and cAMP level, are generally amenable for miniaturization, whereas methods that measure kinetics, such as calcium mobilization, label-free impedance and patch clamp, are more likely preformed in relatively low throughput. Instruments are now available to allow kinetics based assays to be carried out in considerable throughput, for instance, the FLIPR system produced by Molecular Device conducts calcium influx measurement in 384-well format with high precision.

3.3. *Plate setting*

Currently, both 96- and 384-well plates are commonly used in cell-based HTS. Since the plates are made for different utilities, selecting the right one for a particular assay is of great importance. There are mainly three simple steps in this regard: (i) plate format and well design; (ii) material and color; and (iii) surface treatment. Special requirements are developed according to assay methods and detection instruments. For example, solid black polystyrene microplates are designed to reduce well-to-well cross-talk and background signal for fluorescent assays; solid white polystyrene plates are produced to decrease well-to-well cross-talk and background noise thereby enhancing specific readouts in luminescent assays. A useful plate selection guide is available in Corning's website, which may serve as a reference for assay development and optimization.

With the right plate, one should decide how to arrange an assay on it. Apart from testing samples, appropriate controls must be included in the same plate: positive and negative. Positive control that exhibits desired response is used to declare the validity of the assay and serves as a comparison to identified hits. Negative control usually produces no response and serves as the baseline or background, defining compounds with no activity. Normally, both are needed and set on side columns of the plate in an alternant and symmetrical way to avoid spatial variance[40,41], while testing samples are arranged inside.

3.4. *Automation*

It is automation and miniaturization that make HTS possible. Today, almost every single step of HTS can be automated but for ordinary screening centers, some of them are optional. The core of the automated system resides in one or more automatic multi-channel liquid handling instruments, which are essential to the production of high quality data. Before setting up a protocol, thorough understanding of the operation and precision of the automatic liquid handling system is indispensable. Certain strategies are required for the transfer of different kinds of liquids in high accuracy. Order of the steps, speed of dispensing, height of tips, mixing, changing tips and other related parameters should always be taken into account. Organization of each step in a proper working flow can significantly increase throughput. For example, dispensing of a glutinous substance should be slow and the tips should not be put too deep under the liquid surface to reduce hangover. Mixing can always improve accuracy if it does not disturb cell attachment[42]. Tips should also be changed if necessary to reduce cross contamination. In general, references can be found in technical support materials for particular handling system. A standard operating procedure (SOP) for liquid handling was recommended by Taylor et al.[43]. Smart apparatuses, such as the FLIPR system, have already added those special settings to their associated software.

3.5. *Other factors*

Optimization ought to include assay volume, cell number, reagent concentration, interval time, *etc*. Assay volume is partially decided by the plate density: 100–200 μL/well for a 96-well plate, 30–100 μL/well for a 384-well plate and 2.5–10 μL/well for a 1,536-well plate[37]. Cell number in each well varies among different cell types and may start at 10,000 per well in a 384-well format[38]. To obtain high signal with low cost, concentrations of reagents such as buffer, substrate and testing compound, *etc*. need to be adjusted as well. Usually, compounds are stored in DMSO which is toxic for living cells. Hence, compound concentration should not pose significant cytotoxicity in terms of DMSO tolerance[44], while reserving sufficient potency to achieve bioactivities. A smooth work flow and an appropriate time interval for each step are equally vital to an efficient HTS campaign.

Pilot assay will follow with positive and negative controls to determine Z' factor using the equation described previously[45]. A Z' factor in the range between 0.5 and 1 indicates that the assay is suitable to HTS. For Z' factors between 0 and 0.5, the assay is still acceptable, however, re-optimization is strongly recommended before large-scale application. If the Z' factor is less than 0, re-optimization must be carried out while the best choice may be to completely re-design the assay.

Although the Z' factor is a valuable indicator of assay quality, naive use can be misleading, especially if the value is around or less than 0.5. To exclude potential errors, the concept of power, which reflects the probability of finding a hit, was introduced[46]. Statistical analysis shows, for assays with Z' factor under 0.5, reduction of coefficient of variation (CV) is effective in elevating the power; in the case of constant CV (>0.15), a rise in signal to background ratio is important for increased power[46]. It appears that cell proliferation may significantly affect the CV while incubation time exerts a strong impact on signal to background ratio. Therefore, re-optimization should be adopted accordingly in each case with the concept of power in mind.

4. Compound management

Assays with reasonable Z' factors can be applied to large-scale HTS. However, before the start of a screening campaign, there

still exist several fundamental questions such as how many compounds need to be screened, which portion of the library should be covered and what kind of HTS strategies ought to be taken.

4.1. Compound selection

Generally speaking, the more compounds screened, the more hits that will be found. However, the budget for HTS is always limited and the cost for screening the whole collection, which may involve millions of compounds, is extremely high and the associated hit rate may be unacceptable. Academic laboratories are recommended to start with screening small collections of biologically active molecules. Several pragmatic strategies, which are guided by historical bioactivity data, have been adopted for library design with positive results[47]. Other than financial considerations, one of the other concerns for compound selection is the undesirable characteristics of sample structures. Some compounds in HTS libraries are too complex to reach protein targets in cells[48], while others that were frequently defined as hits in primary screening turned out to be artifacts in later validation due to undesirable structure features[49].

For rational compound selection, there are some tips that can be followed. Compounds should be soluble in the medium of choice, physically stable[50] and neither interfere with the assay nor chemically modify proteins. When interacting with cells, they should also be permeable, non-toxic, non-mutagenic, and bioavailable. Commonly, undesirable compounds can be excluded through computer-assisted physicochemical and functional group filters. The filters, first introduced by Lipinski, include molecular weight, calculated log *P*, number of hydrogen-bond donors and acceptors, rotatable bonds, polar surface area and charge state[51–53]. All of them are aiming to exclude 'non-drug like' compounds with potentially low oral bioavailability and solubility in bioassays. Application of these 'druggable' filters may yield libraries with relatively low hit rates. However, the quality of the hits will be markedly improved.

Compounds after this passage can be considered as candidates for screening. Inspection of structures prior to HTS is still advisable because some of them are sufficiently similar or of little novelty. Due to the fact that the ultimate goal for HTS is to discover new chemical entities with therapeutic relevance, compound libraries with great structural diversity are usually preferable[54]. Clustering is used to select representatives from screening libraries[55]. There are two competing demands that are difficult to reconcile. Before screening, we normally wish to cover as much of chemical space as possible, while noting that any given representative may not actually represent all the molecules in its class[56]. Clustering makes sense for expanding on hits once found, but its merit relative to the selection of a library member remains controversial as it is hard to confirm the validity.

Natural products and synthetic compounds inspired by them continue to attract considerable attention as a strategy to augment screening libraries[57,58]. Natural products often possess sound pharmacokinetic profiles, and nearly one half of the currently approved drugs are either mimicries or derived from them[59]. Moreover, a natural product that influences a discrete molecular event within a cell may not directly bind the target, and protein behavior (e.g., aggregation) is often subject to the influence of signaling pathways[60,61], which can create opportunities for therapeutic modulation and determination of pathological mechanisms. Many natural products are vastly different from synthetic compounds in structures[62]. For these reasons, we feel it prudent to deploy libraries that include natural products.

It is worth pointing out that screening an already established chemical space offers many advantages. These compounds are generally known to be biologically active, which increases the successful rate of drug discovery. Many small companies have now been started based on rescreening known compounds to develop drugs for novel indications. Normally, compounds in HTS libraries are dissolved in 96-well microplates. Cherry picking compounds from such plates to establish optimal library for particular screenings is extremely time consuming albeit possible. A simple solution is to focus on plate-based diversity rather than single compound-based diversity[63].

4.2. Pooling strategy

Pooling in HTS refers to the act of testing mixtures of compounds in primary screening. Because most compound libraries contain only a small portion of active compounds, pooling can significantly reduce the number of tests as oppose to the single compound per well approach. On the other hand, pooling is also controversial due to concerns surrounding potential molecular interactions in mixtures and chaos in signaling pathways induced by multiple molecules. Nonetheless, historical successes imply that pooling is still a practical and necessary part of HTS[64].

Pooling schemes can be roughly divided into two categories: adaptive and non-adaptive. For adaptive pooling[65,66], a fixed number of compound plates are combined into one plate for primary screening and each compound is screened for only one time. Compounds in positive wells are re-screened individually to identify the active hit(s). For example, a library with 100-compound plates could be pooled into 10 plates: each well contains 10 compounds. These 10 plates are screened and the 10 compounds in each positive well are re-screened individually. The adaptive pooling method is very efficient with a 10-fold increase in throughput. However, the false negative rate is high such that active compounds are easily missed as a result of system error. Moreover, results are delayed due to two-stage primary screening. In contrast, non-adaptive scheme involves the test of all pooled compounds in a single run. Typically each compound was assessed multiple times[67,68] depending on how many repeats are desired. For example, orthogonal pooling allows each compound to appear twice in the primary screening. During HTS implementation, the 100-compound plates are arranged as a 10×10 matrix, 10 different plates in the same row are combined into one plate, resulting in 10 row plates (called X-plate). In the same way, 10 different plates in the same column are pooled into another plate, leading to 10 column plates (called Y-plate). Thus, the entire library is reduced to 20 plates, each well contains 10 compounds, and all the compounds will appear twice in HTS, once in X-plate and the other in Y-plate. Some complex non-adaptive pooling strategies allow compounds to repeat more than twice, however, these are beyond our discussion here. In orthogonal pooling, each active compound is required to show activity in both wells of X- and Y-plate in order to be classified

as a hit. Although this method does not provide robustness, it does conserve both resources and time by conducting a reasonably efficient screening in a single run. Although false positives occur when an inactive compound is tested with different active compounds in both X- and Y-plate, such error tends to be rare, especially when the fraction of active compounds in library is small.

To prevent interactions among compounds during pooling, there are certainly some principles to remember. First, the fewer the compounds pooled the less the multiple interactions will appear. Second, pooling dissimilar compounds can reduce interactions. Finally, it would be better to systematically prevent specific chemical classes such as electrophiles and nucleophiles from being mixed. Computational methods have been developed to design adaptive pooling, which use structure information to optimize pools for maximal coverage of chemical spaces, while minimizing overlaps[69]. Because of repeats and one stage screening, non-adaptive pooling is more common in HTS. A simulation-based approach has been proposed to exploit the ability of non-adaptive pooling and to design better pooling strategies and decoding techniques[70].

5. Data analysis

Analyses of data generated by HTS are always challenging due to the amount of information and the complexity of statistics[40]. A 3-step statistical decision methodology is thus recommended by Shun et al.[71]. Step 1 is to determine the most appropriate data processing method and to establish criteria for quality control and hit identification. Step 2 is to perform a multilevel statistical and graphical review of the data to exclude those that failed to meet the standard. Step 3 is to apply the established criterion to the quality-assured data in order to identify active compounds. HTS data usually show some sort of distortion as a result of systematic errors. To correct and remove outliers, stringent quality control measures are required. Furthermore, hit identification is more than just finding highly active compounds, exclusion of false positive/negative samples should also be a priority during the selection.

5.1. *Systemic errors*

Many technical, procedural or environmental factors can cause systemic measurement errors. Without correction, such deviation can significantly curtail hit selection and hence, should be identified as quickly as possible[72,73]. Among which, spatial system error defined as differences between wells due to asymmetrical location is seen most frequently. It could result from a discriminating process operated by liquid handling or detection devices. For instance, in a poorly calibrated apparatus, data obtained from one side of the plate may differ drastically from that seen on the other side. Under most circumstances, such system errors are measurable, predictable[74] and correctable by applying statistical means like the median filter approach[75,76]. The other systemic error commonly encountered relates to the edge (or side) effect. Wells located on the edge of the plate always give incorrect results because of physical or environmental variances (e.g., evaporation). Normalization is thus crucial, particularly when positive and negative controls are arranged on the side of the plate. Statistical tools have been developed to estimate and correct this type of errors[75,77], and experimental studies suggest that incubating newly seeded plates at room temperature before placing them in a 37 °C CO_2 chamber could significantly reduce the edge effect[78]. An alternative method to avoid systemic errors is the use of whole sample area based data analysis instead of control based analysis to achieve both efficiency and accuracy[79].

5.2. *False positive and false negative*

In statistics, basically, there are two types (I and II) of errors that can be made when making interpretations. Type I error, also known as false positive, defines a wrong decision that is made when a test rejects a true null hypothesis, while type II error, or false negative, stands for a wrong decision made when a test fails to reject a false null hypothesis. In the case of HTS, false positive represents compounds which are identified as hits by statistical analysis but are not truly active; false negative is the mistake of failing to identify true active compounds as hits through statistical process. Repeats or follow-up studies, either in the same setting or through different assays, can always minimize such errors[40,80]. In the real situation, however, compounds are usually screened only once and repeats are not feasible in primary HTS. Pooling strategy allows each compound to be tested at least twice, thereby providing inherent repeats to reduce false positives.

5.3. *Hit identification*

Identification of active compounds as hits is the most exciting part of statistical process. Criteria or threshold for hit selection are established, and normally, hits are identified as samples that generate the highest measured activity, for example, compounds shown the highest 1% activity or whose responses exceed a fixed 'percent of control' threshold[40]. The latter actually determines the rate of false positive and false negative hits. A low threshold increases the number of hits identified and reduces the probability of false negatives, but inevitably raises type I errors. On the other hand, high threshold reduces false positives, but increases false negatives. Proper threshold is set according to the nature of HTS conducted. For tough targets which have proven to be difficult to find small molecule modulators, any hits with demonstrable activities should not be ignored. It follows that a lower threshold might be applied and false positives would be excluded at the validation stage. As far as well executed assays are concerned, a higher threshold is advisable in order to save labor in secondary screenings. There are times when HTS data are extremely noisy, low threshold does not necessarily improve the yield of hits but dramatically enhances the burden of validation. Under such circumstances, naive Bayes classifier is recommended for rational selection of potential hits[81]. Other than this, statistical tools such as compound set enrichment can also help us to identify active chemical series rather than just individual active compounds, which may prove to be useful for follow-up studies[82].

Although data analysis is essential, the focal point for a successful HTS resides in the screening process *per se*. Appropriate statistical methods only ascertain the conclusion drawn from the data set is empirically right and do not impact the

assay quality at all. They can also be employed during an HTS campaign to quickly react to or avoid quality problems[83].

6. Lead validation

Hits selected from statistical analysis need to be validated with repeats before making a conclusion. Assays used for validation could be the same as primary screenings, while methods that address the functionality of a compound are highly valuable in this stage. Different from biochemical assays, functional assessment is generally directed towards multiple targets, and an array of cell-based models may be required to exclude off-target effects. It is noted that confirmed hits are sometimes scarcely available and some of them display ambiguous activities in different assay systems, rendering the validation extremely frustrating. Therefore, the key is to capture as many potent hits as possible based on the primary data, and one way to achieve this goal is cherry picking additional compounds for validation in accordance with the Bayes classifier[81].

Validation can exclude most false positive compounds. Some false positives were so peculiar that they became candidates, but ended up with failure after wasting a great deal of time and resources. Studies have described the structural classes and known mechanisms of such non-lead like false positives[84,85], and computational approach to exclude such compounds is now available[85]. To increase the success rate of drug discovery, ideal leads should possess both structural novelty and drug-like properties, as proposed by Lipinski's 'rule of 5'[51].

7. Conclusions

Cell-based HTS as a more physiological relevant assay system has become increasingly popular in modern drug discovery. To execute a high quality HTS, comprehensive coordination is required throughout the screening process, including screening strategy, target selection, cell line generation, signal detection, statistical analysis, hit identification and lead validation.

Acknowledgments

This work was supported in part by grants from the Ministry of Science and Technology of China (2009ZX09302-001 and 2012ZX09304011), Shanghai Science and Technology Development Fund (11DZ2292200) and the CAS-Novo Nordisk Research Fund.

References

1. David E, Tramontin T, Zemmel R. Pharmaceutical R&D: the road to positive returns. *Nat Rev Drug Discov* 2009;**8**:609–10.
2. Prous J, Khurdayan VK. The story so far in R&D. *Drug News Perspect* 2007;**20**:7–15.
3. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;**3**:711–5.
4. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* 2011;**10**:188–95.
5. Perola E. An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs. *J Med Chem* 2010;**53**:2986–97.
6. Chaguturu R. Conference report: the screening Asia conference 2010: an update. *Future Med Chem* 2011;**3**:513–7.
7. Fox S, Farr-Jones S, Sopchak L, Boggs A, Nicely HW, Khoury R, et al. High-throughput screening: update on practices and success. *J Biomol Screen* 2006;**11**:864–9.
8. Landro JA, Taylor IC, Stirtan WG, Osterman DG, Kristie J, Hunnicutt EJ, et al. HTS in the new millennium: the role of pharmacology and flexibility. *J Pharmacol Toxicol Methods* 2000;**44**:273–89.
9. Nichols A. High content screening as a screening tool in drug discovery. *Methods Mol Biol* 2007;**356**:379–87.
10. Kepp O, Galluzzi L, Lipinski M, Yuan J, Kroemer G. Cell death assays for drug discovery. *Nat Rev Drug Discov* 2011;**10**:221–37.
11. Miller C, Slack M. Impact of new technologies for cellular screening along the drug value chain. *Drug Discov Today* 2010;**15**:384–90.
12. Jayawickreme CK, Kost TA. Gene expression systems in the development of high-throughput screens. *Curr Opin Biotech* 1997;**8**:629–634.
13. Chen J, Lake MR, Sabet RS, Niforatos W, Pratt SD, Cassar SC, et al. Utility of large-scale transiently transfected cells for cell-based high-throughput screens to identify transient receptor potential channel A1 (TRPA1) antagonists. *J Biomol Screen* 2007;**12**:61–9.
14. Johnston PA, Johnston PA. Cellular platforms for HTS: three case studies. *Drug Discov Today* 2002;**7**:353–63.
15. Song J, Doucette C, Hanniford D, Hunady K, Wang N, Sherf B, et al. Generation of cell lines for drug discovery through random activation of gene expression: application to the human histamine H3 receptor. *Assay Drug Dev Technol* 2005;**3**:309–18.
16. Harrington JJ, Sherf B, Rundlett S, Jackson PD, Perry R, Cain S, et al. Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat Biotechnol* 2001;**19**:440–5.
17. Liu PQ, Morton MF, Reik A, de la Rosa R, Mendal MC, Li XY, et al. Cell lines for drug discovery: elevating target-protein levels using engineered transcription factors. *J Biomol Screen* 2004;**9**:44–51.
18. Gabriel D, Vernier M, Pfeifer MJ, Dasen B, Tenaillon L, Bouhelal R. High throughput screening technologies for direct cyclic AMP measurement. *Assay Drug Dev Technol* 2003;**1**:291–303.
19. Pope AJ, Haupts UM, Moore KJ. Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. *Drug Discov Today* 1999;**4**:350–62.
20. Tang Y, Li X, He J, Lu J, Diwu Z. Real-time and high throughput monitoring of cAMP in live cells using a fluorescent membrane potential-sensitive dye. *Assay Drug Dev Technol* 2006;**4**:461–71.
21. Michelini E, Cevenini L, Mezzanotte L, Coppa A, Roda A. Cell-based assays: fuelling drug discovery. *Anal Bioanal Chem* 2010;**398**:227–38.
22. Zhou T, Marx KA, Warren M, Schulze H, Braunhut SJ. The quartz crystal microbalance as a continuous monitoring tool for the study of endothelial cell surface attachment and growth. *Biotechnol Prog* 2000;**16**:268–77.
23. Minor LK. Label-free cell-based functional assays. *Comb Chem High Throughput Screen* 2008;**11**:573–80.
24. Braunhut SJ, McIntosh D, Vorotnikova E, Zhou T, Marx KA. Detection of apoptosis and drug resistance of human breast cancer cells to taxane treatments using quartz crystal microbalance biosensor technology. *Assay Drug Dev Technol* 2005;**3**:77–88.
25. Abassi YA, Kirstein SL, Atienza JM, Xi B, Zhu J, Yu NC, et al. Live cell quality control and utility of real-time cell electronic sensing for assay development. *Assay Drug Dev Technol* 2006;**4**:545–53.
26. Gonzalez JE, Oades K, Leychkis Y, Harootunian A, Negulescu PA. Cell-based assays and instrumentation for screening ion-channel targets. *Drug Discov Today* 1999;**4**:431–9.

27. Lu Q, An WF. Impact of novel screening technologies on ion channel drug discovery. *Comb Chem High Throughput Screen* 2008;**11**:185–94.
28. Dale TJ, Townsend C, Hollands EC, Trezise DJ. Population patch clamp electrophysiology: a breakthrough technology for ion channel screening. *Mol Biosyst* 2007;**3**:714–22.
29. Castle N, Printzenhoff D, Zellmer S, Antonio B, Wickenden A, Silvia C. Sodium channel inhibitor drug discovery using automated high throughput electrophysiology platforms. *Comb Chem High Throughput Screen* 2009;**12**:107–22.
30. Bridgland-Taylor MH, Hargreaves AC, Easter A, Orme A, Henthorn DC, Ding M, et al. Optimization and validation of a medium-throughput electrophysiology-based hERG assay using IonWorks HT. *J Pharmacol Toxicol Methods* 2006;**54**:189–99.
31. Coma I, Clark L, Diez E, Harper G, Herranz J, Hofmann G, et al. Process validation and screen reproducibility in high-throughput screening. *J Biomol Screen* 2009;**14**:66–76.
32. Glick M, Bender A, Bojanic D, Davies JW, Crisman TJ, Mikhailov D, et al. Which aspects of HTS are empirically correlated with downstream success? *Curr Opin Drug Discov* 2008;**11**:327–37
33. TerWee JA, Chin CL, Watrin S, Tello RF, Rieder NJ, Lowell JD, et al. Increased consistency and efficiency in routine potency testing by bioassay with direct use of cryopreserved (ready-to-plate) cells. *J Immunol Methods* 2011;**370**:65–74.
34. Zhong Z, Fursov N, Cong M, Federici M, Platchek M, Haytko P, et al. Improving consistency of cell-based assays by using division-arrested cells. *Assay Drug Dev Technol* 2005;**3**:7–15.
35. Digan ME, Pou C, Niu H, Zhang JH. Evaluation of division-arrested cells for cell-based high-throughput screening and profiling. *J Biomol Screen* 2005;**10**:615–23.
36. Jain S, Sondervan D, Rizzu P, Bochdanovits Z, Caminada D, Heutink P. The complete automation of cell culture: improvements for high-throughput and high-content screening. *J Biomol Screen* 2011;**16**:932–9.
37. Mayr LM, Bojanic D. Novel trends in high-throughput screening. *Curr Opin Pharmacol* 2009;**9**:580–8.
38. Roda A, Michelini E, Cevenini L, Mezzanotte L, Coppa A. Cell-based assays: fuelling drug discovery. *Anal Bioanal Chem* 2010;**398**:227–38.
39. Maffia AM, Kariv I, Oldenburg KR. Miniaturization of a mammalian cell-based assay: luciferase reporter gene readout in a 3-microliter 1,536-well plate. *J Biomol Screen* 1999;**4**:137–42.
40. Nadon R, Malo N, Hanley JA, Cerquozzi S, Pelletier J. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;**24**:167–75.
41. Josiah S. Interpretation of uniform-well readouts. *Methods Mol Biol* 2009;**486**:177–92.
42. Orth AP, Hancock MK, Medina MN, Smith BM. Microplate orbital mixing improves high-throughput cell-based reporter assay readouts. *J Biomol Screen* 2007;**12**:140–4.
43. Taylor PB, Ashman S, Baddeley SM, Bartram SL, Battle CD, Bond BC, et al. A standard operating procedure for assessing liquid handler performance in high-throughput screening. *J Biomol Screen* 2002;**7**:554–69.
44. Waybright TJ, Britt JR, McCloud TG. Overcoming problems of compound storage in DMSO: solvent and process alternatives. *J Biomol Screen* 2009;**14**:708–15.
45. Zhang JH, Chung TDY, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 1999;**4**:67–73.
46. Wu ZJ, Sui YX. Alternative statistical parameter for high-throughput screening assay quality assessment. *J Biomol Screen* 2007;**12**:229–34.
47. Golebiowski A, Klopfenstein SR, Portlock DE. Lead compounds discovered from libraries: Part 2. *Curr Opin Chem Biol* 2003;**7**:308–325.
48. Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inform Comput Sci* 2001;**41**:856–64.
49. Irwin JJ. How good is your screening library? *Curr Opin Chem Biol* 2006;**10**:352–6.
50. McGovern SL, Caselli E, Grigorieff N, Shoichet BK. Aggregation: a common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Biochemistry* 2002;**45**:1712–22.
51. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001;**46**:3–26.
52. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 2002;**45**:2615–23.
53. Martin YC. A bioavailability score. *J Med Chem* 2005;**48**:3164–70.
54. Valler MJ, Green D. Diversity screening versus focused screening in drug discovery. *Drug Discov Today* 2000;**5**:286–93.
55. Bocker A, Derksen S, Schmidt E, Teckentrup A, Schneider G. A hierarchical clustering approach for large compound libraries. *J Chem Inform Model* 2005;**45**:807–15.
56. Harper G, Pickett SD, Green DVS. Design of a compound screening collection for use in high throughput screening. *Comb Chem High Throughput Screen* 2004;**7**:63–71.
57. Newman DJ, Cragg GM, Snader KM. Natural products as sources of new drugs over the period 1981-2002. *J Nat Prod* 2003;**66**:1022–37.
58. Clardy J, Walsh C. Lessons from natural molecules. *Nature* 2004;**432**:829–37.
59. Butler MS. Natural products to drugs: natural product-derived compounds in clinical trials. *Nat Prod Rep* 2008;**25**:475–516.
60. Diamond MI, Robinson MR, Yamamoto KR. Regulation of expanded polyglutamine protein aggregation and nuclear localization by the glucocorticoid receptor. *Proc Natl Acad Sci USA* 2000;**97**:657–61.
61. Pollitt SK, Pallos J, Shao JY, Desai UA, Ma AAK, Thompson LM, et al. A rapid cellular FRET assay of polyglutamine aggregation identifies a novel inhibitor. *Neuron* 2003;**40**:685–94.
62. Bajorath J. Chemoinformatics methods for systematic comparison of molecules from natural and synthetic sources and design of hybrid libraries. *J Comput Aid Mol Des* 2002;**16**:431–9.
63. Crisman TJ, Jenkins JL, Parker CN, Hill WAG, Bender A, Deng Z, et al. Plate cherry picking: a novel semi-sequential screening paradigm for cheaper, faster, information-rich compound selection. *J Biomol Screen* 2007;**12**:320–7.
64. Kainkaryam RM, Woolf PJ. Pooling in high-throughput drug screening. *Curr Opin Drug Discov* 2009;**12**:339–50.
65. Jones CM, Zhigljavsky AA. Comparison of costs for multistage group testing methods in the pharmaceutical industry. *Commun Stat-Theor M* 2001;**30**:2189–209.
66. De Bonis A, Gasieniec L, Vaccaro U. Optimal two-stage algorithms for group testing problems. *Siam J Comput* 2005;**34**:1253–70.
67. Motlekar N, Diamond SL, Napper AD. Evaluation of an orthogonal pooling strategy for rapid high-throughput screening of proteases. *Assay Drug Dev Technol* 2008;**6**:395–405.
68. Warrior U, Gopalakrishnan SM, Traphagen L, Freiberg G, Towne D, Humphrey P, et al. Maximizing the identification of leads from compound mixtures. *Lett Drug Des Discov* 2007;**4**:215–23.
69. Remlinger KS, Hughes-Oliver JM, Young SS, Lam RL. Statistical design of pools using optimal coverage and minimal collision. *Technometrics* 2006;**48**:133–43.
70. Thierry-Mieg N, Bailly G. Interpool: interpreting smart-pooling results. *Bioinformatics* 2008;**24**:696–703.
71. Shun TY, Lazo JS, Sharlow ER, Johnston PA. Identifying actives from HTS data sets: practical approaches for the selection of an appropriate HTS data-processing method and quality control review. *J Biomol Screen* 2011;**16**:1–14.
72. Dragiev P, Nadon R, Makarenkov V. Systematic error detection in experimental high-throughput screening. *BMC Bioinformatics* 2011;**12**:25.

73. Makarenkov V, Zentilli P, Kevorkov D, Gagarin A, Malo N, Nadon R. An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* 2007;**23**:1648–57.
74. Root DE, Kelley BP, Stockwell BR. Detecting spatial patterns in biological array experiments. *J Biomol Screen* 2003;**8**:393–8.
75. Bushway PJ, Azimi B, Heynen-Genel S, Price JH, Mercola M. Hybrid median filter background estimator for correcting distortions in microtiter plate data. *Assay Drug Dev Technol* 2010;**8**: 238–50.
76. Bushway PJ, Azimi B, Heynen-Genel S. Optimization and application of median filter corrections to relieve diverse spatial patterns in microtiter plate data. *J Biomol Screen* 2011;**16**: 1068–80.
77. Kevorkov D, Makarenkov V. Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen* 2005;**10**: 557–67.
78. Lundholt BK, Scudder KM, Pagliaro L. A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen* 2003; **8**:566–70.
79. Sewing A, Gribbon P, Lyons R, Laflin P, Bradley J, Chambers C, et al. Evaluating real-life high-throughput screening data. *J Biomol Screen* 2005;**10**:99–107.
80. Zhang JH, Wu X, Sills MA. Probing the primary screening efficiency by multiple replicate testing: a quantitative analysis of hit confirmation and false screening results of a biochemical assay. *J Biomol Screen* 2005;**10**:695–704.
81. Glick M, Klon AE, Acklin P, Davies JW. Enrichment of extremely noisy high-throughput screening data using a naive Bayes classifier. *J Biomol Screen* 2004;**9**:32–6.
82. Varin T, Gubler H, Parker CN, Zhang JH, Raman P, Ertl P, et al. Compound set enrichment: a novel approach to analysis of primary HTS data. *J Biomol Screen* 2010;**50**:2067–78.
83. Gunter B, Brideau C, Pikounis B, Liaw A. Statistical and graphical methods for quality control determination of high-throughput screening data. *J Biomol Screen* 2003;**8**:624–33.
84. Shoichet BK. Screening in a spirit haunted world. *Drug Discov Today* 2006;**11**:607–15.
85. Sink R, Gobec S, Pecar S, Zega A. False positives in the early stages of drug discovery. *Curr Med Chem* 2010;**17**:4231–55.