UNIVERSITÉ DU QUÉBEC À MONTRÉAL

NOUVELLES MÉTHODES INFORMATIQUES ET STATISTIQUES POUR ANALYSER LES DONNÉES DU CRIBLAGE À HAUT DÉBIT

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN INFORMATIQUE

PAR

IURIE CARAUS

AVRIL 2018

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my two supervisors, Professor Vladimir Makarenkov from Université du Québec à Montréal (UQAM) and Professor Robert Nadon from McGill University for their patience, encouragement, and motivation. I appreciate the efforts, ideas and financial support they have provided to make my PhD work productive and stimulating. Their guidance helped me hugely during my PhD investigations and thesis writing. My colleagues and friends at the Department of Computer Science of UQAM and Génome Québec Innovation Centre at McGill University have also contributed immensely to the completion of my PhD dissertation. I would like to thank Nadia Tahiri, Bogdan Mazoure, Elika Garg, Dunarel Badescu, Alexandru Guja, Vitalie Samoila, Haig Djambazian, Ashot Harutyunyan and Abdulaziz A. Alsuwailem for their assistance with research materials, helpful discussions, and comments. This research was supported by Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT).

Special thanks are addressed to my family: I would like to thank my wife, my son and my parents for supporting me spiritually, as well as for understanding and encouraging me to follow my dreams throughout this multi-year adventure.

Iurie Caraus

Montreal, April 18th, 2018

DEDICATED

To my lovely parents: My mother Sima Rusnac and my father Nicolae Caraus

PREFACE

This thesis presents new statistical methods and programs for the analysis of high-throughput screening (HTS) and high-content screening (HCS) data. When Dr. Makarenkov agreed to become my supervisor, he introduced me to Dr. Nadon from the Department of Human Genetics of McGill University and Génome Québec Innovation Centre. Drs. Makarenkov and Nadon proposed me to work on the development of new statistical methods and software in the field of HTS, explaining to me that HTS assays were prone to various systematic errors and that this topic was very relevant. They introduced me to their scientific groups. When I began my literature review, I understood the urgent need to develop efficient and accurate statistical methods to detect and eliminate systematic errors from screening data. Although in the HTS field there exist already a number of systematic error elimination methods, I realized that this scientific direction had not been studied in detail. Prof. Makarenkov, Prof. Nadon and I formulated together some important statistical issues that could be solved using new methods and software. I spent a lot of time in the laboratory to study the HTS process, formulating questions to the technical staff of the laboratory, studying the recommended scientific literature, and thus getting a better understanding of the origins of systematic error problem in screening technologies.

In the first paper, published in *Briefings in Bioinformatics*, we reviewed systematic errors typical for HTS and HCS screens. This work led to the writing of my second article, published in *Bioinformatics*, where we proposed three new multiplicative error elimination methods. Based on the results of the second article, we wrote the third article, submitted to *SLAS Discovery,* which presents different additive and multiplicative bias models, including four new models. Dr. Makarenkov offered his expertise in HTS/HCS techniques. He helped me define the main objectives of my

PhD thesis. Dr. Nadon gave me advices and recommendations concerning the use of statistical methods. Both Prof. Makarenkov and Prof. Nadon helped me write and review the three articles and the thesis.

This PhD thesis includes multidisciplinary materials, so that it can be of interest to statisticians, bioinformaticians, and life scientists. The Introduction section of this thesis does not include a comprehensive literature review because it is present in the first article (Chapter I). My PhD thesis contains an introduction chapter, the three manuscript chapters, and a conclusion chapter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

HTS	High-Throughput Screening
HCS	High-Content Screening
MAD	Median Absolute Deviation
cDNA	Complementary DNA
RNAi	RNA interference
NLMBE	Non-Linear Multiplicative Bias Elimination
SPAWN	Spatial Polish Well Normalization
PMP	Partial Mean Polish
mPMP	Multiplicative PMP
CVM	Cramer-von-Mises test
AD	Anderson-Darling test
dsRNA	double-stranded RNA
siRNAs	small interfering RNA
shRNAs	short hairpin RNA
esiRNAs	endoribonuclease-prepared siRNAs

RÉSUMÉ

Le criblage à haut débit (CHD) et le criblage à haut contenu (CHC) sont des techniques expérimentales efficaces permettant aux chercheurs d'identifier un petit nombre de candidats potentiels parmi des millions de composés chimiques (ou par exemple, de petites molécules, d'ADN complémentaire, de petits ARN interférents) pouvant devenir de nouveaux médicaments. Durant les dernières décennies, les nombreux centres CHD/CHC ont été créés dans les campus universitaires (Peter et Roy 2011). Au cours des dernières années, les résultats des CHD/CHC ont trouvé des grandes applications dans la recherche biologique, par exemple pour l'étude des maladies orphelines comme le paludisme et la fibrose kystique (Brewer 2009; Okiyoneda et Lukacs 2012; Preuss et al. 2012).

Néanmoins, les données CHD/CHC peuvent contenir des erreurs systématiques (ou des biais spatiaux). Ces erreurs affectent généralement de manière significative toutes les mesures expérimentales, en augmentant ainsi le nombre de faux positifs et de faux négatifs retournés pas des méthodes de recherche de composés actifs. L'application des méthodes statistiques appropriées permet d'éliminer ou de réduire l'effet d'erreurs systématiques dans les données CHD/CHC. Plusieurs chercheurs (Brideau et al. 2003, Makarenkov et al. 2007, Dragiev et al. 2011 et 2012) ont montré que les méthodes d'élimination des erreurs systématiques peuvent être appliquées avec succès aux données CHD/CHC expérimentales. Dans cette thèse, nous proposons de nouvelles méthodes et protocoles statistiques servant à réduire l'impact d'erreurs systématiques dans les analyses CHD/CHC. La thèse est divisée en trois parties principales correspondant à nos trois articles.

Le premier article examine les technologies de criblage existantes et leurs biais associés. Ici nous décrivons les différents types d'erreurs systématiques caractéristiques aux données CHD/CHC. Nous parlons de l'avantage des mesures répliquées et randomisées pour obtenir une meilleure précision des résultats dans les campagnes CHD/CHC. Les principales méthodes statistiques qui sont utilisées pour éliminer les erreurs systématiques, essentiellement de type additif, sont également présentées. Dans ce premier article, nous évaluons la grandeur de l'erreur systématique présente dans les données CHD expérimentales publics. Nous proposons également un protocole de prétraitement des données général, adopté à l'analyse des données de criblage.

Le deuxième article présente trois nouvelles méthodes statistiques pour éliminer les erreurs systématiques multiplicatives. Pour détecter l'erreur systématique, nous utilisons le test non-paramétrique de Mann-Whitney *U*. Les biais spatiaux présents dans les essais sont corrigés via la résolution d'un système d'équations nonlinéaires ou par les procédures itératives d'élimination de biais. Nous montrons que les nouvelles méthodes de correction de données suppriment bien des erreurs systématiques multiplicatives présentes dans les lignes et les colonnes de chaque plaque de l'essai considéré.

Le troisième article propose de nouvelles méthodes statistiques pour éliminer les erreurs systématiques du type additif et multiplicatif, en considérant les différentes interactions possibles entre ces erreurs. Nous utilisons les tests de Cramer-von-Mises et d'Anderson-Darling pour estimer la qualité de l'ajustement des valeurs originales par des valeurs corrigées et pour déterminer ainsi le meilleur modèle pour les données d'intérêt.

MOTS-CLÉS : biais spatial, criblage à haut débit, criblage à haut contenu, erreur systématique, erreur additive, erreur multiplicative

SUMMARY

High-Throughput Screening (HTS) and High-Content Screening (HCS) are effective experimental techniques that allow researchers to identify a small number of potential drug candidates among millions of chemical compounds, cDNAs or RNAis. Over the last few decades, many HTS and HCS centers have been created in academic campuses (Peter and Roy 2011). Recently, HTS assays have also found large applications in biological research, e.g., in studying orphan diseases like malaria and cystic fibrosis (Brewer 2009; Okiyoneda and Lukacs 2012; Preuss et al. 2012).

However, as has been noted in many studies (Brideau et al. 2003, Makarenkov et al. 2007, Dragiev et al. 2011 and 2012), HTS and HCS data are usually severely affected by systematic errors (i.e., spatial biases). These errors lead to an important increase in the number of false positive and false negative hits (e.g., active compounds that have the potential to be developed into new medications). One possible solution to this problem is to apply statistical methods designed to minimize the impact of systematic errors on experimental HTS/HCS data. To this end, this thesis proposes new statistical methods, data pre-processing protocols, and software for reducing systematic errors from experimental high-throughput screening assays. The thesis is organized by publications.

The first paper reviews existing screening technologies and their related biases. It describes the different types of systematic errors present in HTS and HCS data. The existing statistical methods and models proposed to eliminate systematic errors are also reviewed. In the first article, we also assess the magnitude of systematic error in experimental HTS data and propose a general data pre-processing protocol which can be recommended for the analysis of the current or next generation screening data.
The second paper presents three new statistical methods for spatial bias correction meant to minimize the impact of multiplicative spatial biases. In our study, the presence of bias in rows and columns of a given plate is identified using the non-parametric Mann-Whitney *U* test. Our data correction methods modify the data only in the bias-affected rows and columns. The usefulness of the new methods is demonstrated by a simulation study as well as by the analysis of publicly available ChemBank data.

In the third paper, we consider six bias correction models: two existing models and four new models. These models account for different possible interactions between additive and multiplicative spatial biases. We use the Cramer-von-Mises and Anderson-Darling tests to estimate the goodness-of-fit of the raw data by the corrected data and to select the most appropriate (additive, multiplicative or mixed) spatial bias model for the data at hand.

KEYWORDS: high-throughput screening, high-content screening, systematic error, spatial bias, additive error, multiplicative error

INTRODUCTION

0.1. High-Throughput Screening and High-Content Screening

High-Throughput Screening (HTS) and High-Content Screening (HCS) are popular biotechnological methods allowing researchers to test a large number of chemical compounds in order find a small proportion (around 1% or less) of potential drug candidates (i.e., chemical compounds with very high activity levels, also called hits). Nowadays, HTS and HCS assays are regularly used by the modern pharmaceutical industry at the first step of the drug discovery process (Dove 2003; Kaiser 2008; Silber 2010; Lachmann et al. 2016). It has been also demonstrated that these technologies can be used successfully for studying various diseases, including the orphan diseases like malaria and cystic fibrosis (Brewer 2009; Okiyoneda and Lukacs 2012; Preuss et al. 2012). As indicated in Malo et al. (2006) and Sirois et al. (2005), the drug development process usually contains the four following main steps (Figure 0.1):

1. Primary Screen: preparation of screening assay and primary hit detection;
2. Counter screen and Secondary screen: hit identification for biological relevance and hit confirmation;
3. Structural-activity relationship (SAR) and medical chemistry: identification of leads;
4. Clinical trials: clinical compound selection, entry into human studies and regulatory approval for a new drug.

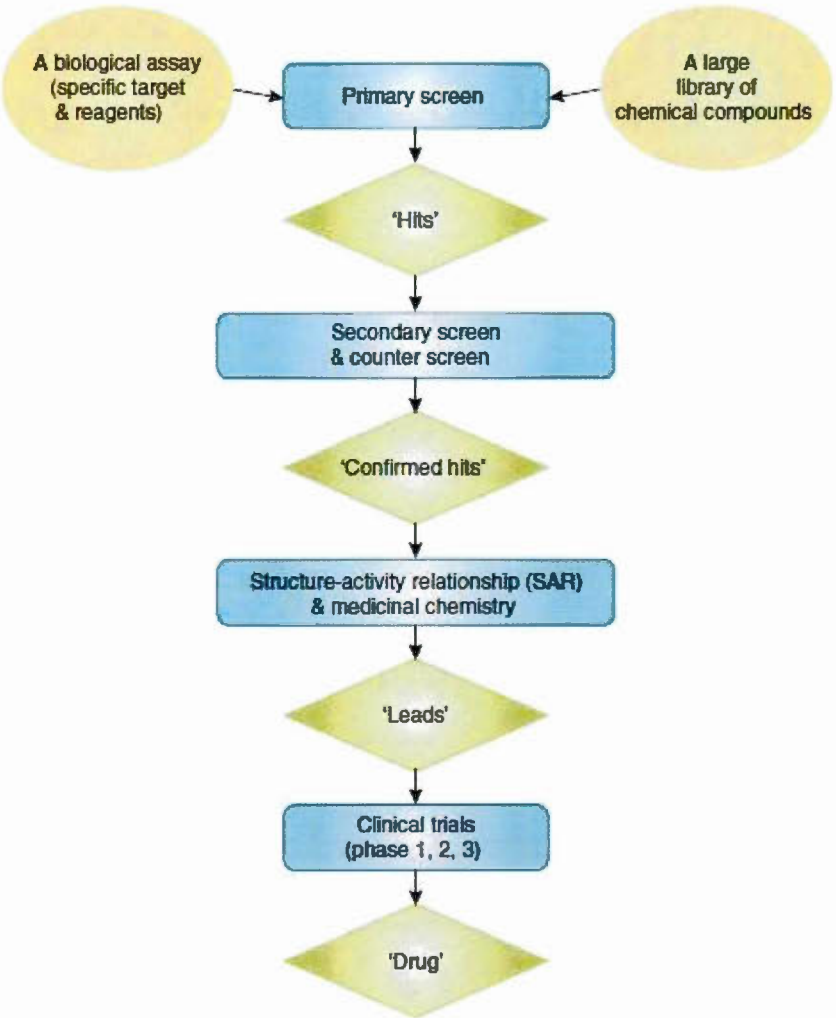


Figure 0.1 The drug development process according to Malo et al. (2006).

During a typical HTS/HCS campaign, lasting over a few days, millions of chemical compounds and thousands of microtiter plates are usually analyzed (Agresti et al. 2010).

In an HTS/HCS primary assay, the selected library of chemical compounds is screened against a specific biological target to measure the intensity of the related inhibition or activation signal.

Figure 0.2 presents typical HTS plates. In general, a standard HTS plate contains either 96, or 384, or 1536 wells.

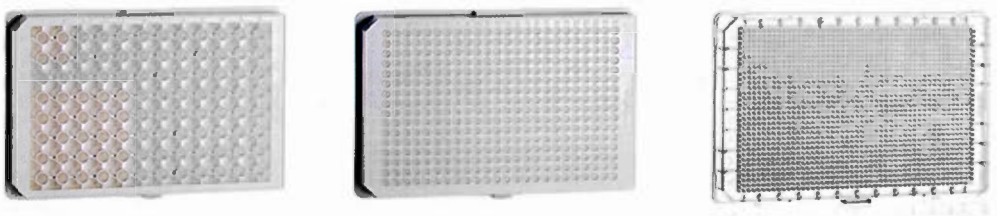


Figure 0.2 Typical HTS plates (96 wells, 384 wells and 1536 wells, respectively); source: https://shop.gbo.com/en/row/products/0110/0110_0040.

The wells are arranged in a rectangular matrix pattern. The robots introduce an appropriate biological target culture into every well of a given plate, such as a protein, a small molecule, a cell, or a siRNA. After the incubation period, the measurements are taken to get the evaluations of biological activity in all wells of the plate using specialized automated analysis machines. The measurement of a well is usually considered as a single numeric value. Thus, the measurements of a plate obtained during the screening process can be represented by a matrix of numerical values. At the end of the screening process, statistical methods should be used to find the hits (i.e., the compounds with the highest activity levels).

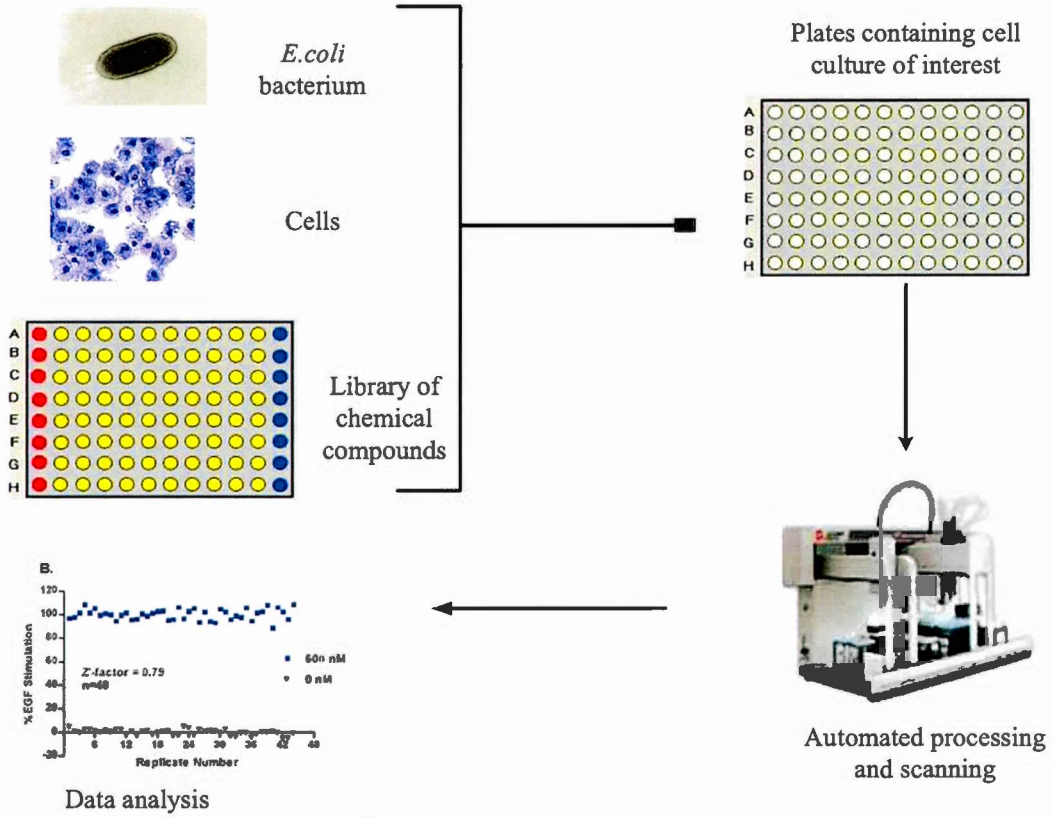


Figure 0.3 A typical HTS/HCS procedure. Blue colored wells contain negative controls, red colored wells contain positive controls and yellow colored wells contain the library compounds.

Figure 0.3 presents an example of a typical HTS/HCS process (Menke 2002; Janzen and Bernasconi 2009; Macarrón and Hertzberg 2011). This process uses the robotic liquid handling, automatic processing and scanning of plates, as well as some statistical tools to detect chemical compounds (i.e., hits) providing the best inhibition of the *E.coli* bacterium. The presented assay uses 96-well plates.

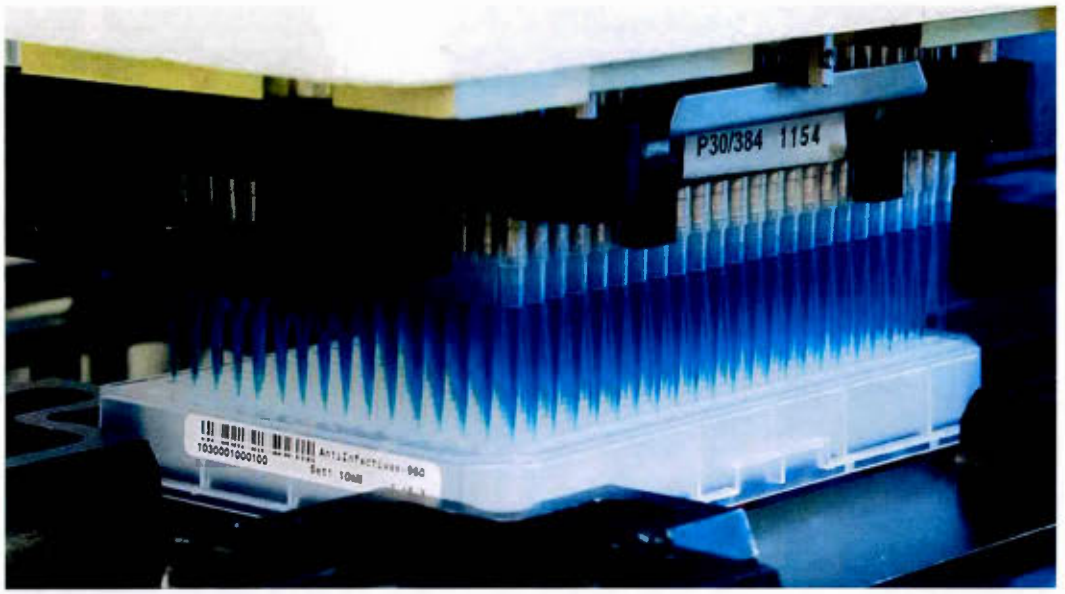


Figure 0.4 HTS robotic equipment using hundreds of pipettes; source: http://whyfiles.org/2014/stem-cell-advance.

Recent improvements in drug discovery and in high-throughput screening permitted obtaining high resolutions even at the cellular level (Noah 2010). The corresponding screening technology was named High-Content Screening (HCS). The detailed information about cell structure can be now obtained by extracting multicolor fluorescence signals. The main sub-categories of HTS and HCS technologies are the following: small molecules, complementary DNAs (cDNAs), and RNA interference (RNAis). These sub-categories vary with respect to the target of interest.

In the pharmaceutical industry the term small molecule is related to a specific biological target such as a specific nucleic acid or protein (Cram101, 2012). Small-molecules are used in drug discovery for target validation, assay development, secondary screening, pharmacological property assessment and lead optimization (Lazo et al. 2007). cDNA libraries are manufactured from messenger RNAs (mRNA molecules) (Brown and Song 2000). The transformation of unstable mRNA template,

via reverse transcription, produces stable cDNA. Nowadays, the cDNA libraries are a very effective tool used in drug discovery. The biological process, in which a short RNA molecule suppresses gene expression by targeting specific mRNA molecules for degradation is defined as RNA interference (RNAi). The researchers call this process silencing. This technique is frequent in drug discovery. Scientists use the RNAis to silence genes. Nowadays, four types of RNAi reagents are applied in RNAi screening technologies: dsRNAs (double-stranded RNA), siRNAs (small interfering RNA), shRNAs (short hairpin RNA), and endoribonuclease-prepared siRNAs (esiRNAs) (Mohr et al. 2010).

A more detailed description of small molecule, cDNA and RNAi screens is available in Chapter I. The main biases characteristic for these screens are presented in details in Section 0.3.

0.2 Systematic error in HTS/HCS screens

Unfortunately, raw compound activity measures can be disturbed by two additional factors: *random errors* and *systematic errors* (or *spatial biases*) (Makarenkov et al. 2007). Random error lowers screening precision and may affect false positive and false negatives rates (Box et al. 2005). Its negative impact can be minimized by using duplicate measurements (Malo et al. 2010) or by blocking (Murie et al. 2015). Ramadan et al. (2007) showed that random errors can be the cause of nonspecific phenotypes in specific wells and lead to higher rates of false positives and false negatives. In contrast, Dragiev et al. (2011) suggested that random error, on the opposite of systematic error, usually has a minimum impact on the hit selection process.

Definition *Systematic error, or systematic bias, can be defined as a variability of measurements taken at the same plate or assay locations, consisting of their systematic under or over-estimation* (Kevorkov and Makarenkov 2005).

Systematic error, or systematic bias, has been observed in many HTS/HCS assays (Brideau et al. 2003, Makarenkov et al. 2007, Lachmann et al. 2016). The origin of systematic error can be biological, human or mechanical. For instance, it can be due to different robotic and environmental factors, such as robotic failure, reader effect, pipette malfunctioning or other liquid handling anomalies, variation in the incubation time or temperature difference, and lighting or air flow abnormalities, present over the course of the entire screen (Heyse, 2002; Makarenkov et al., 2007).
Different types of systematic error, including edge and batch effects (Soneson et al. 2014), are described in detail in Chapter I of this thesis.

0.3 Small molecule, cDNA and RNAi screens and the related systematic errors

Small molecule screens: Small molecules can be either natural or artificial. These molecules are associated with a particular biopolymer. Kim et al. (2007) showed that a lot of small molecules are cytotoxic in hepatocyte replicon cell lines. These molecules are very susceptible to cytotoxic or cytostatic agents. Cytotoxic effects may be confused with antiviral activity when cytotoxic effects have lower luciferase signal reducing cells' vital activity. Chan et al. (2009) indicated that a substantial fraction of small molecules show aggregating behaviour. It usually happens because of nonspecific attachment to target proteins, resulting in false positive hits in experimental screening assays. Thus, the capability to effectively detect the substances with aggregating characteristics may rationalize the screening procedure by excluding unstable molecules from further examination. Harmon et al. (2010) indicated that many pathogenic Gram-negative bacteria represent a type three

secretion system (TTSS). TTSS translocates effectors proteins into the cytosol of their eukaryotic cell targets. The authors mentioned that the molecules that interfere with assembly of TTSS are not efficiently detected in experimental HTS screens. They postulated that systematic error in small molecule screens often appears due to the fact that TTSS is preassembled prior to exposure to compounds and cells. Wootten et al. (2013) mentioned that the individual G protein-coupled receptor (GPCRs) may be found in multiple receptor conformations and may be the cause of numerous functional reactions, both G protein- and non G protein-mediated. This concept, referred to as biased agonism, is also known as functional selectivity, or stimulus systematic bias, or ligand-directed signaling, or ligand systematic bias (Kenakin 2011). We should mention that some interesting cases of ligand systematic bias for the glucagon-like peptide-1 were discussed in Willard et al. (2012).

cDNA screens: *cDNA* libraries are essential tools for discovery and validation of novel drug targets in functional genomics applications, but they are not exempt of different biases. In one of earlier studies, Kopczynski et al. (1998) suggested that the representation of clones is underestimated or overestimated by sequence analysis encoding membrane-targeted proteins in the rough microsomes libraries. It happens due to systematic error related to cytosolic and nuclear proteins. Carninci et al. (2000) indicated that the process of generating full-length cDNA libraries may have some specific issues. The authors mentioned that the process of preparation a full-length cDNA is more effective for short mRNAs than for long transcripts. Moreover, cloning and propagation are trickier for long cDNAs than for short cDNAs, therefore creating further size bias. Carninci and Shibata (2000) observed that the plasmid libraries were related to a cDNA-size cloning bias. This bias manifests itself as an increased cloning effectiveness of short cDNAs. Furthermore, the increase of cDNA clones can be related to the plasmid length during library amplification before normalization and/or subtraction. It means that the long clones are usually underrepresented after bulk amplification of the library. To overcome the issues

related to the library amplification, Carninci and Shibata developed a specific cDNA data correction method. Fosså et al. (2004) demonstrated the aptitude of biopanning to enrich TAAs (tumor-associated antigens) from tumor cDNA libraries under determined experimental conditions. However, this enrichment may be related to the loss of positive clones as well as to systematic error. Systematic error arises here from non-immunological factors (for example: inefficient protein presentation or delayed growth of phage-infected bacteria). Wan et al. (2006) suggested that the oligonucleotide synthesis is not entirely effective. It means that the probability of the presence of synthesis bias augments with every base added. Wan et al. suggested using only high-quality oligonucleotides because synthesis bias may be either included into the amplified product or may generate some other off-target effects.

RNAi screens: Recently, RNA interference (RNAi) screening has made great progress, evolving from a biological phenomenon into an effective method of drug discovery (Sharma and Rao 2009). Birmingham et al. (2009) mentioned that the transfection process is the main source of variability in siRNA screens. Transfection process produces cell stress and influences cell viability. It may have variable and indirect phenotypic influence in cellular assays. The target gene product can be decreased or practically erased in the cell by RNAi reagents. Due to these mechanistic factors, RNAi reagents require 48–72 hours for maximal effect, whereas small molecules can immediately affect their protein targets within a few hours. Such longer time intervals between cell plating and assay endpoint lead to a greater impact of cell culture and environmental variation on phenotypes and cause some important assay variability in RNAi screens. Zhang et al. (2008) underlined that a good plate design in HTS RNAi is needed to identify systematic error as well as to determine which normalization and data correction techniques should be used to reduce the impact of systematic error on both quality control and hit selection processes. Auer and Doerge (2010) discussed so-called *lane effect* in RNAi screens. The authors indicated that lane effect includes any systematic error that appear from the item at

which the sample is introduced into the flow cell until the samples are removed from the sequencing machine (i.e., badly-organized sequencing cycles and errors in base calling). Several publications have appeared in recent years to highlight that the main issue in RNAi screening is *off-target effect* (Birmingham et al. 2006; Buehler et al. 2012; Chen et al. 2013; Das et al. 2013). This kind of error appears when a siRNA is processed by the RNA-Induced Silencing Complex (RISC) and down-regulates unintended targets. Echeverri et al. (2006) suggested that an initial origin of off-target effects is a comparatively huge tolerance for mismatches between the siRNA guide strand, the ultimate targeting molecule and the complementary target mRNA sequence, outside of the short 'core' targeting region. Sharma and Rao (2009) described three different types of off-target effect. Firstly, when a siRNA sequence is the same or nearly identical to a sequence present in an unrelated mRNA, the final degradation of the unrelated mRNA can establish a confounding phenotype. Thus, the final result is a false-positive hit. We can avoid this problem only if the related siRNA sequences are selected very carefully. Secondly, a siRNA may operate as microRNA and produce depletion of non-target proteins through mRNA degradation or translational block, where the 'seed region' of a siRNA pairs with a weakly complementary sequence in the 3' untranslated region of an unrelated mRNA. This problem is intrinsic to RNAi screens and is very difficult to solve. The last type of off-target effect is characteristic for mammalian cells. It is well known that even short siRNA can switch on the antiviral type I interferon response in a sequence-independent way, particularly when saturating doses of siRNA or shRNA are supplied.

0.4 Systematic errors specific to HCS

HCS-specific *systematic biases* can be divided into two types according to their spatial distribution (see Table 0.1):

1. *Intra-well (per well) bias* appears when cells are not distributed uniformly. It is arising while gathering data from a given well location. It usually happens when cells are clumped, which also changes cell adherence and morphology characteristics. Such an effect is called *cell count distribution systematic bias.* It affects post-processing steps and can skew results significantly if not enough images per well are taken. Another type of *intra-well bias* can appear because of cell cycle distribution heterogeneity when cells are in different stages of cell cycle, and thus treatments are affecting them differently. This usually has an effect when analyzing treatment response and cell structure (Snijder and Pelkmans 2011).

2. *Intra-image (per-image) bias* consisting of microscope-related errors is arising while capturing images. One of the issues here is a non-uniformity of background light intensity distribution, which is slowly varying as well as systematic change of the spatial distribution of light in images. Such an effect can add or subtract intensities at any pixel location, thus affecting data quantification and statistical analysis, which, in turn, can affect cell segmentation and florescence measurements (Lo et al. 2012).

Systematic bias can also occur if the focus is not maintained throughout all captured images. Out of focus images can impact on the cell segmentation (Bray et al. 2012). They can also lead to the issues in the identification of thresholds that consistently separate objects of interest from background, thus reducing the accuracy of object segmentation and causing bias in the measured cell metrics that depend on both the pixel intensities and segmentation step (Lo et al. 2012). It is worth noting that image segmentation is vital for the viability HCS data. Identification of cells or sub-cellular structures and the related morphological "features" (such as fluorescent intensity, object shape, size and texture) are fully dependent on image segmentation. We should mention that, unfortunately, some "HCS-unfriendly" cell lines, which are still very clinically pertinent, for example human tumor cell line SK-BR-3, grow only in

complex patterns. For such cell lines, one must be very careful when doing image segmentation, because errors in segmentation can arise more frequently, resulting in the inappropriate designation of partial or multiple cells as single cells.

Another common problem in HCS is that of fluorophore saturation and debris, which can affect the intensity measurement. It can occur if the settings are not optimal for every image, resulting in a higher than maximum signal for at least one pixel. Such an effect reduces the measured intensity value of the affected pixel, systematically biasing the intensity results (Brown 2007). In a typical HCS screen, in which hundreds or thousands of images are segmented, the review, classification and correction of the acquired images is a complex task requiring multiple statistical and medical skills. Systematic errors in these screens should be removed using appropriate correction and normalization methods (Hill et al. 2007).

Table 0.1 Systematic biases specific to HCS technologies

Per well (intra-well)				**Per image (intra-image)**		
Cell count distribution		*Cell cycle distribution*		*Out-of-focus bias*	*Non-uniformity systematic bias*	*Fluorophore saturation and debris*
Clumping of cells	Cells adherence error	Morphology-related error	Cell structure error	Segmentation error		Intensity measurement error

0.5 Data normalization methods

Normalization techniques used in HTS/HCS allow for transforming the measurements of a given plate (or well location) in order to make them comparable over different plates (or well locations) of the same assay. We should mention that data normalization methods are not specifically designed to eliminate systematic

errors. The following simple normalization methods are commonly used in HTS/HCS technologies:

Z-score: is a simple and well known method which allows one to transform the original data set into a normalized data set with the mean of 0 and the standard deviation of 1. The mean value of a given plate is first subtracted from all of its measurements. After that the values of all measurements are divided by the standard deviation of the plate. The equation that describes Z-score is as follows: $\hat{x}_{ij} = \frac{x_{ij} - \mu}{\sigma}$, where μ and σ are, respectively, the mean and the standard deviation the plate's measurements.

Robust Z-score: is similar to *Z-score*, but uses the *MAD* statistic instead of the standard deviation of the plate. Also, in the *Robust Z-score* the median is applied in place of the mean to get a more robust statistical estimate. *Robust Z-score* normalization is described by the following equation: $\hat{x}_{ij} = \frac{x_{ij} - med}{MAD}$, where *med* is the median of all measurements of a given plate and *MAD* is its Median Absolute Deviation. The *robust Z-score* normalization is more robust to outliers compared to *traditional Z-scores*. For more details about the existing data normalization techniques the reader is referred to Chapter I.

0.6 Statistical tests to detect the presence of systematic error in screening data

Once the raw experimental data are normalized, statistical tests can be applied to them to detect the presence of systematic error. These tests allow for a judicious use of bias elimination methods and software which should be applied with caution. Makarenkov *et al.* (2007) indicated, for example, that spatial bias correction methods may introduce different biases into the data at hand when applied to bias-free data. We will actively use bias detection techniques in Chapter II, in which we present new

bias correction methods assuming that the biased rows and columns of all plates are known. The main tests that have been used for bias detection in screening technologies are summarized below:

Welch's t-test (*Welch 1947*) is an adaptation of *Student's t-test.* This test does not assume that the two data samples are drawn from populations with equal variances and that these populations have the same size. The measurements of a particular plate (or of a hit distribution surface) are divided into two samples: the first sample includes the values of the tested row or column and the second sample includes the rest of the plate's measurements. The null hypothesis, H_0, here is that the considered row or column does not contain systematic error. The equation that defines *Welch's t-test* is as follows: $t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$. The measurements of a given plate are subdivided into two samples: S_1 - the first sample that includes the measurements of the tested row or column and S_2 - the second sample that includes the rest of the plate's measurements. Here, μ_1 is the mean of sample S_1 and μ_2 is the mean of sample S_2, sample S_1 contains N_1 elements and sample S_2 contains N_2 elements, and s_1^2 and s_2^2 are the respective variances of S_1 and S_2.

The Mann-Whitney U test is equivalent to the Wilcoxon rank sum test (Wilcoxon 1945; Gibbons and Chakraborti 2011). It is a nonparametric test. As in many other nonparametric procedures, the ranks of the scores are calculated. We combine two populations and we assign the rank to each of the scores. The first lowest measurement obtains rank 1; the second lowest measurement gets rank 2, and so on. This procedure verifies the null hypothesis, H_0, which postulates that the two compared population distributions are identical.

The Kolmogorov-Smirnov goodness-of-fit test (D'Agostino and Stephens 1986) preceded by the *Discrete Fourier Transform* (*DFT*) procedure (Cooley and Tukey

1965; also see Dragiev et al. 2011). To apply the DFT procedure to raw data we need first to unroll the plate's measurements matrix into a linear sequence of measurements. There are two natural ways to do so:

- to construct this sequence beginning with the first row of the plate, followed by the second one, and so on, and
- to construct this sequence beginning with the first column of the plate, followed by the second one, and so on.

DFT detects frequencies of signals that repeat each two, three, four, and so on, positions in the sequence and computes the amplitudes of each eventual frequency component. We apply the Kolmogorov-Smirnov goodness-of-fit test to calculate the density spectrum y_k^p which occurs under the null hypothesis of no effect. The test statistic D can be determined using the following equation: $D = \max_{1 \le k \le N \times (N_R + N_C)} \left(F(y_k^p) - \frac{k-1}{N}, \frac{k}{N} - F(y_k^p) \right)$, where $F(y_k^p)$ is the number of measurements in the density spectrum that are lower than or equal to y_k^p. If values of D are large, this suggest the rejection of the null hypothesis which postulates that the tested measurements have been drawn from random normally distributed data.

0.7 Systematic error correction methods used in screening technologies

The main purpose of systematic error correction methods is to eliminate the plate-specific and assays-specific spatial biases within the considered plate or assay. Plate-specific biases affect the data located in a particular row or column of a given plate. Assay-specific biases affect the data of the same well location over all plates of a given assay. Here, we present the most important methods, which have been used for bias correction in the HTS/HCS technologies. More methods are presented in Chapter I of this thesis.

R-score: The robust regression procedure is used to fit the regression model to the data in the R-score method. To assess the method's parameters, the *rlm* function from the MASS package of the R statistical language is applied. The equation that describes the *R-score* model (Wu et al. 2008) is as follows: $x_{ijp} = \mu_p + R_{ip} + C_{jp} + r_{ijp}$, where x_{ijp} is the compound measurement in row i and column j of plate p, μ_p is the mean of plate p, R_{ip} is the row bias affecting row i of plate p, C_{jp} is the column bias affecting column j of plate p, and r_{ijp} is the residual in well (i, j) of plate p.

B-score: On the p^{th} plate, the residual $r_{ijp} = x_{ijp} - \widetilde{x}_{ijp}$ of the measurement located on the intersection of row i and column j is calculated by applying a two-way median polish procedure (Tukey 1977; Brideau et al. 2003). The residual is obtained as the difference between the raw value x_{ijp} and the fitted value $\widetilde{x}_{ijp}$. The model is defined as follows: $x_{ijp} = \mu_p + R_{ip} + C_{jp} + r_{ijp}$, where μ_p is the mean measurement of plate p, R_{ip} is the bias affecting row i of plate p, and C_{jp} is the bias affecting column j of plate p. For every plate, the adjusted median absolute deviation (*MAD*) is calculated from r_{ijp}'s. The *B-score* estimates are obtained from the following equation:

$$B\text{-}score = \frac{r_{ijp}}{MAD_p}, \text{ where } MAD_p = med\{|r_{ijp} - med(r_{ijp})|\}.$$

Partial Mean Polish (PMP): PMP (Dragiev et al. 2012), or additive PMP, is another extension of the well-known median polish procedure (Tukey 1977). PMP is an iterative method, in which the measurement correction is applied only to the rows and columns of a given plate that are affected by spatial bias (the bias location is assumed to be known in this method).

Well correction: This is an assay-specific correction method that normalizes the data along the well locations of a given assay (Makarenkov et al. 2007). At first, Z-score normalization is carried out within each plate of the assay. Then, a linear least-square approximation is carried out for the measurements of each well location of the assay.

Finally, Z-score normalization is applied once again, but this time for the measurements of all well locations of the assay.

Diffusion Model: This method was designed to minimize the edge effect in the HTS RNAi screens (Carralot et al. 2012). The method relies on a diffusion model using the Laplace operator: $\frac{\partial \tilde{b}(i,j,t)}{\partial t} = c \times \Delta \times \tilde{b}(i,j,t)$, where $\tilde{b}(i,j,t)$ is the evaluated spatio-temporal diffusion field in well (i, j) at time t (i.e., evaluated spatial bias), c is the diffusion coefficient, and Δ is the Laplace operator.

Spatial and Well Normalization (*SPAWN*): The *SPAWN* method first applies a trimmed mean polish procedure on individual plates to minimize row and column (i.e., plate-specific) spatial biases (Murie et al. 2013). It was shown that the trimmed mean approach has good a robustness (Malo et al. 2010; Murie et al. 2015). The R-score model is used to fit the data. Second, a well normalization step is carried out to determine spatial bias template, using the median of the scores at well location (i, j) computed over all plates of the assay. The spatial bias template scores are subtracted from the scores obtained by the median polish procedure. Thus, *SPAWN* allows one to minimize both plate-specific and assay-specific spatial biases.

0.8 Hit identification

The goal of any HTS/HCS campaign is to identify the compounds with the highest activity levels (i.e., hits). Statistical and bioinformatics tools can allow practitioners to ameliorate the precision and exactitude of the detected hits using an appropriate experimental design and analytical methods. This data analysis step is critical in high-throughput screening.

Some practitioners select as screening positives a fixed number, or a fixed percentage, of top scoring compounds (for example 1%, Nelson 2004). Compounds

whose activity exceeds a fixed percent-of-control threshold may also be considered as hits (Malo et al. 2006). A key limitation of this strategy is that it is rather arbitrary and suffers from the absence of any probability model.

The second strategy for hits detection is to find the compounds whose activity exceeds a threshold that is a function of the mean and the standard deviation of the data at hand. As reported by Makarenkov and Zentilli (Makarenkov et al. 2007), the hit selection thresholds are usually established using the μ-$c\sigma$ formula for inhibition assays (here the hits are the values that are lower than this threshold) and μ+$c\sigma$ formula for activation assays (here the hits are the values that are higher than this threshold), where the mean value μ and the standard deviation σ are computed separately for each plate. The constant c is usually set to 3.0, 2.5 or 3.5.

A number of more robust statistical strategies for the hit identification have been described by Malo et al. (2006) and Birmingham et al. (2009). Among them we can mention the Random Variance Model (RVM, Wright and Simon 2003), the quartile-based hit identification procedure in RNAi screens, which establishes upper and lower hit selection thresholds based on number of interquartile ranges (Zhang et al. 2006), and an accurate Strictly Standardized Mean Difference (SSMD) method, calculating the ratio between the difference of the means and the standard deviation of the difference between positive and negative controls (Zhang 2007).

For more details about the existing hit identification techniques the reader is referred to Chapter I.

0.9 Data randomization and replicate measurements

Randomization is a very important part in a number of experimental technologies. In 1925, Fisher introduced the concept of randomization in which experimental units are assigned to groups or treatment in a manner that the probability of assignment to any particular group or treatment is equal and unbiased (Fisher 1925). The work of Fisher indicates that the placement of testing units has to depend on a random unbiased process. The main advantage of randomization in screening technologies is that randomized experimental units can distribute the error in a way that does not introduce discrepancy to the experiment (Box 2006; Hall 2007; Murie et al. 2015). Consequently, the compound placement, both within each plate and each well location of an HTS assays, should be randomized in order to reduce the impact of systematic compound placement on the outcome of screening experiments.

The hit identification accuracy can be also improved using replicates. Replicates offer the twin advantage of obtaining a greater precision of activity measurements and that of estimating the measurements variability (Malo et al. 2006). Due to the cost issue, primary screens each compound is usually evaluated only once. But in secondary screens the use of replicates is strongly recommended (Murie et al. 2013). We should mention that Malo et al. (2006) recommended the application of replicates even in primary screens. Nowadays it is standard practice to get at least three replicates per measurement, assuming that these replicates provide the benefits which exceeds the cost of short-dated considerations (Lee et al. 2000; Nadon and Shoemaker 2002). Obviously, replicated samples in HTS/HCS screens should be evaluated under the identical experimental conditions.

A detailed discussion of the advantages of the data randomization and replication procedures is provided in Chapter I.

0.10 Thesis content

My thesis is organized by articles. It includes an Introduction, three article chapters (Chapters I-III) and a Conclusion. The first paper reviews existing screening technologies and their related biases. It describes the different types of systematic errors present in HTS and HCS data. The existing statistical methods and models proposed to eliminate systematic errors are also reviewed. In the first article, we also assess the magnitude of systematic error in experimental HTS data and propose a general data pre-processing protocol which can be recommended for the analysis of the current or next generation screening data.

The second paper presents three new statistical methods for spatial bias correction meant to minimize the impact of multiplicative spatial biases. In our study, the presence of bias in rows and columns of a given plate is identified using the non-parametric Mann-Whitney *U* test. Our data correction methods modify the data only in the bias-affected rows and columns. The usefulness of the new methods is demonstrated by a simulation study as well as by the analysis of publicly available ChemBank data. In the third paper, we consider six bias correction models: two existing models and four new models. These models account for different possible interactions between additive and multiplicative spatial biases. We use the Cramer-von-Mises and Anderson-Darling tests to estimate the goodness-of-fit of the raw data by the corrected data and to select the most appropriate (additive, multiplicative or mixed) spatial bias model for the data at hand. We analyze the data generated by the four HTS technologies (homogeneous, microorganism, cell-based and gene expression), the three HCS technologies (area, intensity and cell-count) and the unique small-molecule technology represented in ChemBank (Seiler et al. 2008).The new methods presented in this thesis have been implemented in the C# and R programming languages and included in our *AssayCorrector* software. This program was implemented in R and is freely available at CRAN (the *AssayCorrector* program has been developed with my colleague Bogdan Mazoure, Master's student at McGill University).

CHAPTER I

DETECTING AND OVERCOMING SYSTEMATIC BIAS IN HIGH-THROUGHPUT SCREENING TECHNOLOGIES: A COMPREHENSIVE REVIEW OF PRACTICAL ISSUES AND METHODOLOGICAL SOLUTIONS

This chapter is a reproduction of the following article: Iurie Caraus, Abdulaziz A. Alsuwailem, Robert Nadon, and Vladimir Makarenkov. "Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions." *Briefings in Bioinformatics* (2015), 16(6) 974-986.

1.1 Abstract

Significant efforts have been made recently to improve data throughput and data quality in screening technologies related to drug design. The modern pharmaceutical industry relies heavily on high-throughput screening (HTS) and high-content screening (HCS) technologies, which include small molecule, complementary DNA (cDNA) and RNA interference (RNAi) types of screening. Data generated by these screening technologies are subject to several environmental and procedural systematic biases which introduce errors into the hit identification process. We first review systematic biases typical of HTS and HCS screens. We highlight that study design issues and the way in which data are generated are crucial for providing unbiased screening results. Considering various data sets, including the publicly available ChemBank data (Seiler et al. 2008), we assess the rates of systematic bias in experimental HTS by using plate-specific and assay-specific error detection tests. We describe main data normalization and correction techniques and introduce a general

data pre-processing protocol. This protocol can be recommended for academic and industrial researchers involved in the analysis of current or next generation high-throughput screening data.

1.2 Introduction

There has been a growing interest in the development of high-throughput screening technologies over the last few decades (Shelat and Guy 2007), largely because screening methods promoted by the pharmaceutical industry have played a key role in drug discovery. The increasing computing power and miniaturization of screening equipment now allow for carrying out high-throughput screening analyses even in small academic laboratories. The most popular screening technologies used in drug design are high-content screening (HCS) (Giuliano et al. 2003) and high-throughput screening (HTS) (Brideau et al. 2003). Their different subcategories include small molecule (Inglese et al. 2007), complementary DNA (cDNA) (Chiao et al. 2005) and RNA interference (RNAi) (Auer and Doerge 2010) types of screening. In a typical HCS or HTS campaign, hundreds of terabytes of experimental data concerning molecule activity, specificity, and physiological and toxicological properties can be generated. These data should be processed using appropriate data mining and statistical methods and protocols in order to identify promising drug candidates (i.e., hits). One of the key challenges that need to be answered during the analysis of HCS and HTS data is the identification and successful elimination of bias (i.e., systematic error) in the measurements. In this review, we discuss the existing types of bias common to all high-throughput screening technologies and discuss their negative impact on the hit selection process. We underline the necessity of randomization of screened samples and indicate the advantages of using replicate measurements. We present the methods intended to detect systematic error and those designed to correct the data affected by it. We argue that the latter methods should be applied only when

the presence of a specific type of systematic error in the data has been confirmed by a suitable statistical test (Dragiev et al. 2011). Furthermore, we provide suggestions concerning which data normalization and correction techniques should be applied in various practical situations. Finally, we present a broad-spectrum data pre-processing protocol that can be used for the correction and analysis of screening data prior to assay quality estimation and hit selection steps. This protocol can also be used for detecting and removing bias in future HTS technologies involving sequential screening of multiple plates. To illustrate the results of our analyses, we examine publically available HTS and HCS data generated at the McGill University HTS laboratory (Figure 1.1), Chemistry Department of Princeton University (Figure 1.2), McMaster University laboratory - Data screened for McMaster Data Mining and Docking Competition (Figure 1.3), as well as those provided by the largest public HTS/HCS database (Seiler et al. 2008), maintained by Harvard University's Broad Institute (Figure 1.4).

1.3 Screening technologies and related biases

1.3.1 HTS and HCS technologies and their subcategories

In this review, we focus on the two most widely used screening technologies: High-throughput screening (HTS) and high-content screening (HCS). In a typical HTS/HCS primary assay, the selected library of chemical compounds is screened against a specific biological target to measure the intensity of the related inhibition or activation signal (Malo et al. 2006). The size of the compound library can vary from hundreds to millions of items. Compounds are allocated across disposable microtiter plates of different sizes, typically including 96, 384 or 1536 wells. Well locations within a plate follow a rectangular matrix pattern. Each compound is usually placed in a single well. A suitable biological target culture (e.g., cells or a bacterial enzyme) is then added to each well of the plate. It is common to conduct unreplicated HTS

experiments, although, as we show next, it is much more appropriate to obtain at minimum duplicate measurements. Processing the assay plates by HTS robotic equipment consists of a number of experimental wet-lab steps, including incubation, rising, and reagent additions to the biological culture of interest. Once the incubation period is over, the plates are scanned to obtain measures of biological activity characterizing the selected compounds. It is worth noting that the obtained raw activity levels depend not only on putative biological activity, but also on systematic and random errors affecting the given screen. Data analysis steps, including statistical procedures for data normalization and data correction, should then be carried out to identify hits.

The increasing capacity of computer storage devices along with improvements in automation have allowed the use of HTS technologies to achieve resolution at the cellular level (Noah 2010). This related technology is called high-content screening (HCS). HCS is a screening method with multiple readouts that is based on microscopic imaging from a cell-based assay (Smellie et al. 2006). HCS obtains detailed information of cell structure by extracting multicolor fluorescence signals. HCS has three advantages relative to other screening techniques: (a) Cell-based analysis achieves high physiological correspondence, especially regarding drug screening; (b) Single cell analysis captures the heterogeneity of cell populations as well as the related individual response to treatments; (c) HCS generally has low false-positive and false-negative rates (Kozak et al. 2009). Thus, HCS technologies are commonly used in all areas of contemporary drug discovery, including primary compound screening, post-primary screening capable of supporting structure–activity relationships, early evaluation of ADME properties and complex multivariate drug profiling (Zanella et al. 2010). The Mytocheck (see http://mitocheck.org) and ChemBank databases (Seiler et al. 2008) are among the rare online resources containing publically available HCS data.

Different subcategories of HTS and HCS technologies exist, depending on the target of interest. They comprise altering protein function using small molecules, increasing gene function using cDNA libraries and manipulating gene function using RNAi.

(1) *Small molecules*: A "small molecule", which can be either natural or artificial, is defined in pharmacology as a molecule associated with a particular biopolymer – for example a nucleic acid or a protein (Cram101, 2012). There is currently a significant interest in extending efforts to discover small molecules targeting proteins encoded in the genomes of humans and pathogenic organisms (Lazo et al. 2007). Furthermore, small-molecule screening technologies have applications in other areas of drug discovery, such as target validation, assay development, secondary screening, pharmacological property assessment and lead optimization. The combination of principles of molecular pharmacology with modern high-throughput (Inglese et al. 2007) and high-content (Korn and Krausz 2007) technologies is critical for the success of these discoveries.

(2) *cDNA library*: High quality, full-length *cDNA* libraries are essential for discovery and validation of novel drug targets in functional genomics applications (Brown and Song 2000). The discovery of reverse transcriptase permitted the transformation of unstable mRNA molecules into stable complementary DNA (cDNA) molecules. A comprehensive review of cDNA HCS can be found in (Buchser et al. 2014), and that of cDNA HTS in (Chiao et al. 2005; Honma et al. 2001).

(3) *RNA interference* (*RNAi*): In the past decade, RNA interference (RNAi) has made great progress, evolving from a biological phenomenon into an effective method of drug discovery (Sharma and Rao 2009). The two main advantages of RNAi screens compared to classical genetic screens are: (a) sequences of all identified genes are instantaneously identified and (b) lethal mutations are simple to determine because mutant recovery is not required (Boutros and Ahringer 2008). The four types of RNAi reagents currently used in cell-based HTS are the following: dsRNAs, siRNAs, shRNAs and endoribonuclease-prepared siRNAs (esiRNAs) (Mohr et al. 2010). An important issue in genome-wide RNAi investigation is to combine both experimental

and computational approaches to obtain high-quality RNAi HTS assays and to overcome off-target effects (Amberkar et al. 2013; Buehler et al. 2012; Zhang et al. 2008). A recent review by Knapp and Kaderali focuses on the analysis of RNAi HCS data and presents an approach for statistical processing of high-content microscopic screens (Knapp and Kaderali 2012).

1.3.2 Systematic error in screening technologies

As with all biotechnologies, screening data are prone to both *random* and *systematic errors*. Random error, which varies among measured HTS compounds, lowers screening precision and likewise affects false positive and false negatives rates. Its adverse effects can be greatly minimized by obtaining at least duplicate measurements (Malo et al. 2010). Systematic error (i.e., systematic or spatial bias) can be defined as the systematic under or over-estimation of measurements taken at the same plate or assay location (Kevorkov and Makarenkov 2005). Systematic errors can be the cause of nonspecific phenotypes in specific well, row or column locations and thus lead to higher false positive and false negative rates (Dragiev et al. 2011; Ramadan et al. 2007). Its adverse effects can be minimized by the application of data correction methods and study design procedures such as randomization and blocking (Malo et al. 2006; Murie et al. 2015).

Systematic error can be due to various technological and environmental factors, such as robotic failure, reader effect, pipette malfunctioning or other liquid handling anomalies, unintended differences in compound concentration related to agent evaporation, variation in the incubation time or temperature difference, as well as lighting or air flow abnormalities present over the course of the screening campaign (Heyse 2002; Makarenkov et al. 2007). Thus, bias causing systematic under- or over-estimation of biological activity measurements can cause some inactive compounds

to be incorrectly identified as hits (i.e., *false positives*) and some active compounds to remain undetected (i.e., *false negatives*). Systematic error can be well, row or column dependent. It can affect compounds placed either to the same well, row or column location over all plates of the assay (i.e., *assay-specific error*) or those located in a particular row or column of a single plate (i.e., *plate-specific error*) (Dragiev et al. 2012).

Some specific positional effects appearing in HTS/HCS screens as a consequence of bias are summarized below. One often overlooked hurdle of HTS technologies is the *batch effect* (Leek et al. 2010). A batch effect, i.e., bias present in some continuous subsets of the data and absent in others, occurs when some continuous groups of plates are affected by laboratory conditions which vary during the experiment. Although batch effects are hard to detect in low-dimensional assays, HTS technologies provide enough data to detect and remove them (Leek et al. 2010). The *edge effect*, also called *border effect*, is another type of systematic error that consists in systematic under or over-estimation of the measurements located on the plate's edges. Carralot et al. (2012) indicated that although most repetitive errors in RNAi HTS can be generally controlled, some biases, such as edge effects, cannot be easily corrected due to well-to-well discrepancies inherent in the spatial structure of the plate. The cause of this effect is often unclear but medium evaporation or uneven treatment of the entire plate surface might be contributing factors (Armknecht et al. 2005). Similarly to the plate-specific edge effect, a more general assay-specific *row*, *column*, or *well location effects* can occur in both HTS and HCS screens when the data located in a particular row, column or well location are systematically over or under-estimated across all the plates of the assay. On the other hand, a systematic *intra-image bias*, consisting of the microscope-related errors, arises while capturing images in HCS. One of the issues here is a non-uniformity of background light intensity distribution, which is a slowly varying and systematic change of the spatial

distribution of light in images. Such an effect can add or subtract intensities at any pixel location, thus affecting cell segmentation and florescence measurements, which, in turn, affect data quantification and statistical analysis (Lo et al. 2012).

Cell population context can also create systematic bias in high-content cellular screens and thus significantly influence results of HCS campaigns (Snijder et al. 2009). A method allowing for normalizing and scoring statistically microscopy-based RNAi screens has been recently proposed (Knapp and Kaderali 2012). This method exploits individual cell information of hundreds of cells per knockdown. The application of the proposed method and software (Knapp et al. 2011) led to the identification of new host dependency factors of the hepatitis C and dengue viruses as well as to higher reproducibility of results of the screening experiments.

Figure 1.1a illustrates the presence of edge effects (e.g., the measurements in column 2 are systematically overestimated) in the Harvard 164-plate assay (Helm et al. 2003; Kevorkov and Makarenkov 2005). This assay consists of a screen of compounds inhibiting the glycosyltransferase MurG function of *E. coli*. Here, the binding effect of MurG to a fluorescent (fluorescein-labeled) analogue of UDP-GlcNAc was estimated. In this example, the threshold of μ–2σ was applied to identify hits. The HTS Corrector software (Makarenkov et al. 2006) was used to calculate raw (Figure 1.1a) and B-score corrected (Figure 1.1b) hit distribution surfaces (i.e., a hit distribution surface gives the number of hits per well location found over all plates of the assay). The edge effect observed in column 2, and partially in row V, in the raw data was successively eliminated by the B-score procedure (Brideau et al. 2003).

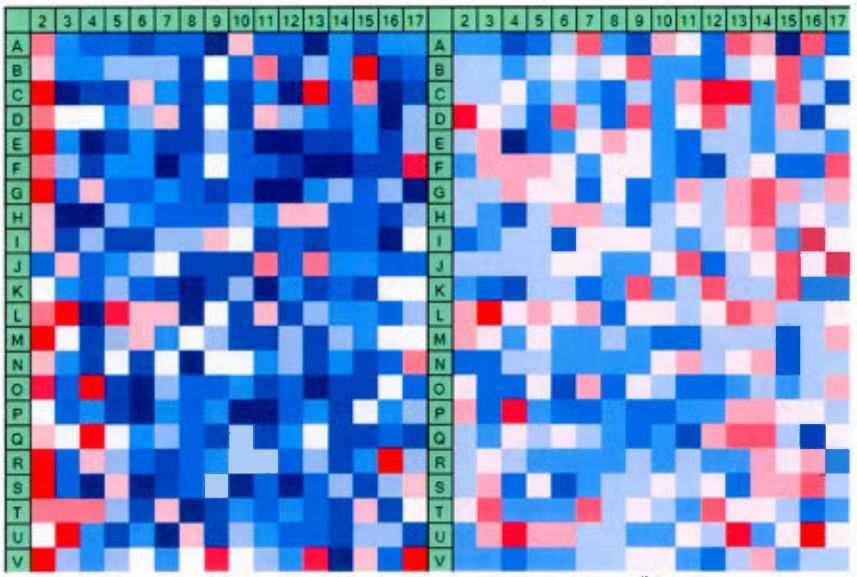


Figure 1.1 Systematic error in experimental HTS data (Harvard's 164-plate assay (Helm, et al., 2003)). Hit distribution surfaces for the μ-2σ hit selection threshold are shown for: (a) Raw data; (b) B-score corrected data. Well, row and column positional effects are illustrated. The data are available at http://www.info2.uqam.ca/~makarenkov_v/HTS/home.php/downloads/Harvard_164.zip.

Similarly, image non-uniformity bias in HCS can be approximated and corrected by combining multiple images to generate a single image with an expected random spatial distribution of intensity values (Lo et al. 2012). Such an approximation represents the overall effect of bias on the imaging field estimated using an image-averaging technique (Vassal et al. 2006). This positional bias can be detected by comparing the center of the image to its edges. In most cases, there is at least a two-fold increase in brightness between center and edges. Figure 1.2 illustrates non-uniformity bias present in a (96-well x 4-field) HCS plate of microtubule polymerization status screened in the HCS laboratory of McGill University.

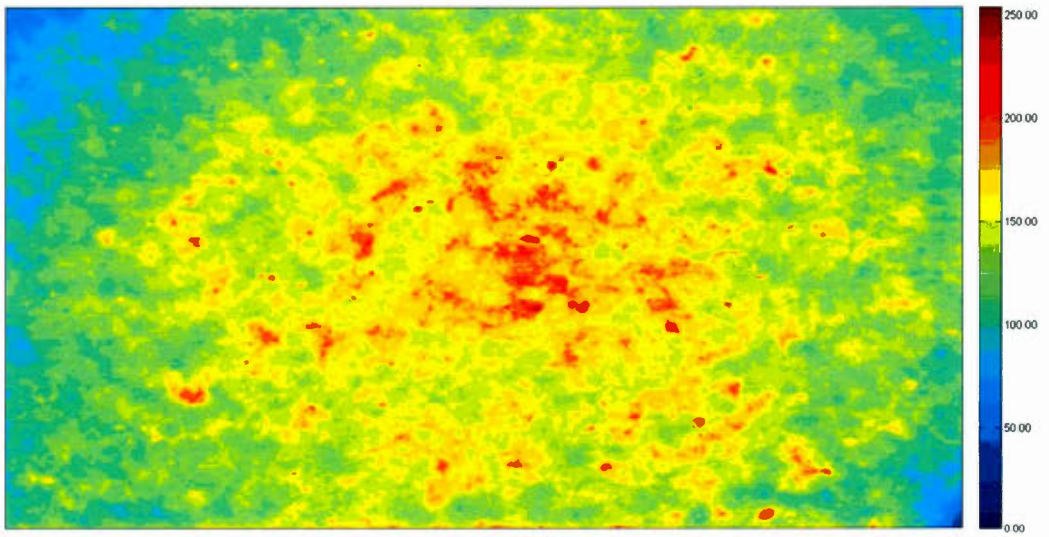


Figure 1.2 Non-smoothed foreground (non-uniformity bias) for images of a single (96-well x 4-field) HCS plate of microtubule polymerization status generated in the HCS laboratory of McGill University is shown. The data are available at: http://nadon-mugqic.mcgill.

1.4 Methods and results

1.4.1 Data randomization and use of controls

The primary aim of statistical practice consists in estimating experimental error, and in the case of systematic error, in reducing the negative effect of this error (Box et al. 2005). Experimental design and statistical analysis methods should be applied to accomplish these objectives, although often underused in screening practice (Murie et al. 2015). A fundamental approach for error reduction in experimental design must include control and randomization techniques (Hays 1994); R.A. Fisher introduced the concept of randomization in which experimental units are assigned to groups or treatment in a manner that the probability of assignment to any particular group or

treatment is equal and unbiased (Fisher 1925). The main advantage of randomization in screening technologies is that randomized experimental units can distribute the error in a way that does not introduce discrepancies to the experiment (Box 2006; Murie et al. 2015; Verdugo et al. 2009). Thus, order of plate processing and compound placements both within each plate and across replicate plates of HTS/HCS assays should be randomized in order to reduce the impact of systematic bias on the outcome of screening experiments.

Controls contain compounds with well-known biological activity. Positive controls provide maximum possible activity measurements and negative controls provide minimum possible activity measurements. Controls are used in control-based normalization methods to render the screening data comparable across different plates and to establish assay background levels. Ideally, controls should be located randomly within plates, but in practice, only the first and the last columns of the plate are typically available for controls. The related systematic edge effect can be reduced by alternating the positive and negative controls in the available wells, so that they appear equally on each of the plate's rows and columns (Malo et al. 2006). If the edge effect affects the control wells, it will also affect all of the plate's measurements because they are normalized relative to the control activities. Randomization of the position of compounds in the replicated experiments is also very important, but unfortunately, is often limited due to practical considerations when automatic spotting approaches or some of the available statistical pipelines (e.g., cellHTS in BioConductor (Boutros et al. 2006)) not supporting control randomization are used. RNAi controls generally exhibit more inter-well variability than small molecule controls because of variations in transfection efficiencies (Birmingham et al. 2009). Cell-based biological controls are especially problematic because cell clumping or evaporation within different plate areas can lead to different growth conditions and thus to position-related bias (Carralot et al. 2012; Dragiev et al. 2011).

1.4.2 Advantages of replicated measurements

Replicates offer the twin advantage of obtaining a greater precision of activity measurements and that of estimating the measurements variability (Malo et al. 2006). The use of replicates allows one to reduce the uncertainty associated to a single measurement (i.e., standard error of the mean), as indicated in Formula 1.1:

$$100\times\left(1-{}^{1}\!/\!_{\sqrt{n}}\right)\%, \tag{1.1}$$

where n is the number of replicates. Thus, carrying out two replicated screens reduces imprecision by 29%; carrying out three replicated screens reduces imprecision by further 13%; and, carrying out four replicated screens reduces the imprecision by additional 8% (i.e., eliminating in total 50% of imprecision associated with a single measurement). Therefore, the replicates make minimally and moderately active compounds simpler to detect. Two types of replicates exist: technical and biological ones. Technical replicates, which address the variability of the process, are repeated measurements of the same sample that represent independent measures of the random noise associated with equipment or protocols. Biological replicates, which mainly address the variability of the population but also reflect the variability of the process, are separate biological samples that were treated using the same protocol. When the sample population is unknown or has a higher variability, more biological replicates are needed. Increasing the number of technical replicates is important for a more variable technical protocol or when new screening equipment is used. Generally, biological variability is considerably greater than technical variability, so it is to our advantage to commit resources to sampling biologically relevant variables (Blainey et al. 2014). When planning for replication, researchers have to determine the proportion of variability induced by each experimental step to design statistically independent replicates and distribute the capacity for replication of the experiment across steps. Recognizing that obtaining even the minimal requirement of two replicates can be prohibitively expensive for some screens, Murie et al. (Murie et al.

2013) introduced the single assay-wide variance experimental (SAVE) design which can generate statistical tests of biological activity based on replication of only a small subset of plates.

Figure 1.3 illustrates the presence of the batch effect in the *McMaster Test dataset* including the original and replicated sets of plates (Elowe et al. 2005). The McMaster Test assay consisted of a sequence of 625 plates, each of which was screened twice (8x12-well plates were used; the first and the last columns of each plate contained controls – these columns are not displayed here; the remaining 80 wells contained different compounds meant to inhibit the *E. coli*'s dihydrofolate reductase). The well (8, 9) (i.e., well (H,10) – according to McMaster annotation; it is highlighted by a green box in Figure 1.3) displays a hit only in the replicated plates R1, R2, R3, R4, R5, R6, R7, R9 and R10, but not in the original plates 1, 2, 3, 4, 5, 6, 7, 9 and 10. This batch effect is absent in the replicated plate R8 and disappears starting from the replicated plate R11. Three of the hit compounds: MAC-0120363 (plates 1 and R1), MAC-0121481 (plates 3 and R3) and MAC-0121668 (plates 5 and R5) were initially recognized as *Average Hits* by the McMaster competition organizers (the list of average hits contained 96 compounds whose average measurements, computed over the original and replicated screens, were lower than or equal to 75% of the reference control average), but all of them were then rejected as false positives when the dose-response relationship analysis of the selected compounds was carried out (Elowe et al. 2005). It is worth noting that only 96 of 50 000 screened compounds in this assay were recognized as *Average Hits*.

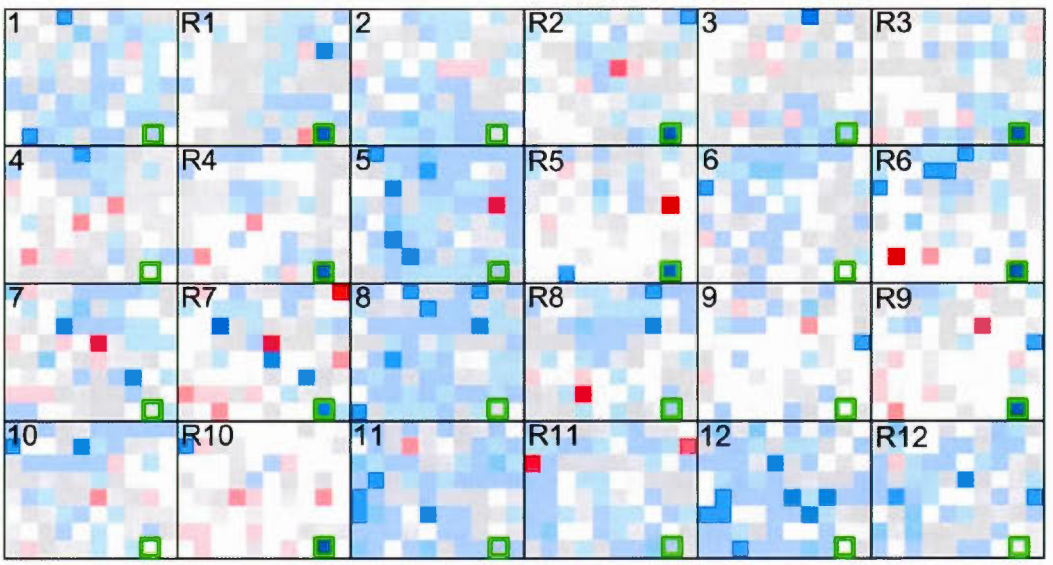


Figure 1.3 Batch positional effect appearing in the *McMaster Test* assay screened during the McMaster Data Mining and Docking Competition (Elowe et al. 2005). The first 24 plates of the assay are shown (12 original and 12 replicated plates; the plate number is indicated on each plate; the replicated plates are indicated by the letter R). Each original plate is followed by its replicate. Hits are shown in blue. Green boxes emphasize well (8, 9) on each plate (i.e., well H10, according to the McMaster annotation) where the batch effect occurs. The data are available at http://www.info2.uqam.ca/~makarenkov_v/HTS/home.php/downloads/McMaster_1250.zip

1.4.3 Identification of hits

The identification of hits is the primary goal of any HTS/HCS campaign. Some screeners select as screening positives a fixed number, or a fixed percentage, of top scoring compounds. Compounds whose activity exceeds a fixed percent-of-control threshold may also be considered as hits (Gagarin et al. 2006; Malo et al. 2006). A wide range of more sophisticated hit identification techniques is available nowadays.

Birmingham and colleagues (Birmingham et al. 2009) reviewed the existing hit selection methods, which can be classified as small-molecule derived methods and RNAi-specific techniques. Small-molecule derived methods include: selection of samples whose screening activity exceeds a fixed threshold, which usually equals mean - 3 standard deviations for inhibition assays and mean + 3 standard deviations for activation assays (Malo et al. 2006); a robust to outliers improvement of the previous approach, using median instead of mean and median absolute deviation instead of standard deviations (Müller et al. 2005); for assays using replicated measurements, the difference in means between replicates for each condition can be assessed with multiple t-tests (Birmingham et al. 2009); finally, the Random Variance Model (RVM), which uses a weighted average of the compound-specific variance and an estimate of the typical variance of all of the compounds, has shown to be appropriate for small molecule HTS data with performance superior to that of standard t-tests (Wright and Simon 2003; Malo et al. 2006). RNAi-specific techniques include: quartile-based hit identification procedure, which establishes upper and lower hit selection thresholds based on number of interquartile ranges (i.e., above or below the first and third quartiles of the data) (Douglas Zhang et al. 2006); an accurate Strictly Standardized Mean Difference (SSMD) method, which computes the ratio between the difference of the means and the standard deviation of the difference between positive and negative controls (Zhang 2007) and, the redundant siRNA Activity (RSA) analysis method, designed for screeners interested in information about multiple RNAi reagents tested for each gene, which assigns p-values to all reagents of a single gene (König et al. 2007).

1.4.4 Data normalization techniques which correct for overall plate bias only

Data normalization in HTS consists in data transformation allowing for data comparability across different plates of the same assay (Birmingham et al. 2009). The

following simple types of data normalization, which do not correct for spatial systematic biases, are commonly used in screening technologies.

Control Normalization is a control-based normalization method using the measurements of both positive and negative controls (Formula 1.2).

$$\hat{x}_{ij} = \frac{x_{ij} - \mu_{neg}}{\mu_{pos} - \mu_{neg}}, \tag{1.2}$$

where x_{ij} is the raw measurement of the compound located in well (i, j), $\hat{x}_{ij}$ is the normalized value of the raw measurement x_{ij}, μ_{pos} is the mean of positive controls of the plate and μ_{neg} is the mean of negative controls of the plate.

Median Percent Inhibition (MPI) normalization is carried out as follows (Formula 1.3):

$$\hat{x}_{ij} = 100 \times \left(1 - \frac{x_{ij}}{med}\right), \tag{1.3}$$

where *med* is the median of all measurements of the plate.

Z-score normalization is defined as follows (Formula 1.4):

$$\hat{x}_{ij} = \frac{x_{ij} - \mu}{\sigma}, \tag{1.4}$$

where μ and σ are, respectively, the mean and the standard deviation of all measurements of the plate.

Robust Z-score normalization can account for different scale and variability effects across HTS plates. It is less likely to produce biased scores because of outlying values of highly active compounds. Robust Z-score normalization is similar to Z-score except that the median is used instead of the mean and the median absolute deviation (*MAD*) is considered instead of the standard deviation to obtain the outlier resistant dispersion estimates (Formula 1.5):

$$\hat{x}_{ij} = \frac{x_{ij} - med}{MAD}, \tag{1.5}$$

where *MAD* is the median absolute deviation of measurements of the plate.

1.4.5 Systematic error detection tests

Several error correction methods and software have been recently developed to minimize the impact of systematic bias (Dragiev et al. 2011). These methods and software should, however, be used with caution. Makarenkov et al. (2007) demonstrated that systematic error correction methods can introduce systematic bias when applied on error-free HTS data. The introduced bias may be less important as in the case of the well correction procedure (Makarenkov et al. 2007) or very important as in the case of the B-score method (Brideau et al. 2003). Thus, the presence or absence of systematic bias in raw HTS data must be first confirmed by the appropriate statistical tests (Dragiev et al. 2011; Gibbons and Chakraborti 2011; Koziol 2010; Welch 1947). Systematic error detection tests that work well with screening data are summarized below.

Welch's t-test: This test is based on the classical two-sample Welch's t-test for the case of samples with various sizes and unequal variances(Welch 1947). Two variants of this test can be considered in the framework of HTS/HCS analysis. The first variant concerns its application to each row and each column of every plate of the assay. The second variant concerns its application to the assay's hit distribution surface. The measurements of the given plate (or of the hit distribution surface) are subdivided into two samples: the first sample contains the measurements of the tested row or column, while the second sample includes the remaining plate's measurements. The null hypothesis, H_0, here is that the considered row or column does not contain systematic error. For the two considered samples, S_1 with N_1 elements and S_2 with N_2 elements, the two sample variances, s_1^2 and s_2^2, are first calculated. Welch's t-test statistic can then be computed using Formula 1.6:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}, \tag{1.6}$$

where μ_1 is the mean of sample S_1 and μ_2 is the mean of sample S_2. The t-test value is then compared to the critical value corresponding to the chosen statistical significance level α in order to decide whether H_0 should be rejected or not. Welch's t-test is usually applied when the data are normally distributed but the sample variances may differ. However, for moderately large samples and a one-tailed test, this statistic is relatively robust to violations of the normality assumption.

χ^2 goodness-of-fit test: This test can be used to establish the presence or absence of systematic error in a hit distribution surface (Dragiev et al. 2011). The null hypothesis H_0 here is the same as in Welch's t-test. The rejection region of H_0 is $P(\chi^2 > C_\alpha) > \alpha$, where C_α is the χ^2 distribution critical value, corresponding to the chosen parameter α and the number of degrees of freedom. For a hit distribution surface with N_R rows and N_C columns, one can test the presence of systematic error in a given row *r* by calculating the χ_r^2 statistic (Formula 1.7):

$$\chi_c^2 = \sum_{i=1}^{N_C} \frac{(x_{rj} - E)^2}{E} \tag{1.7}$$

where x_{rj} is the j^{th} value in row *r*, E is the hits count of the whole hit distribution surface divided by the number of wells ($N_R \times N_C$). The number of degrees of freedom here is N_R -1.

In the same way, the columns of the hit distribution surface affected by systematic error can be tested by computing the test statistic χ_c^2 (Formula 1.8):

$$\chi_c^2 = \sum_{i=1}^{N_R} \frac{(x_{ic} - E)^2}{E} \tag{1.8}$$

The number of degrees of freedom here is N_C - 1.

Systematic error affecting a particular well location (*i, j*) and appearing along all plates of the assay can be also identified by computing the χ^2 statistic (Dragiev et al. 2011) (Formula 1.9):

$$\chi^2 = \sum_{i=1}^{N_R} \sum_{j=1}^{N_C} \frac{(x_{ij} - E)^2}{E}. \tag{1.9}$$

The number of degrees of freedom here is $N_R \times N_C$ -1. The following main assumptions should be met for this test: (a) the observations are independent of each other, and (b) the expected hits count in each well location of the hit distribution surface should be at least 5. The number of degrees of freedom here is $N_R \times N_C$ -1. The following main assumptions should be met for this test: (a) the observations are independent of each other, and (b) the expected hits count in each well location of the hit distribution surface should be at least 5.

Kolmogorov-Smirnov test preceded by Discrete Fourier Transform: This method consists of Discrete Fourier Transform (DFT) (Cooley and Tukey 1965) signal analysis method followed by the Kolmogorov-Smirnov (KS) test (D'Agostino and Stephens 1986). It is included in some commercial software intended to detect systematic error in screening data (e.g., in the *Array Validator* program described in (Kelley 2003)). The KS test is a non-parametric test having the advantage of making no assumption about the distribution of data. As recently has been shown, Welch's t-test usually outperforms the χ^2 goodness-of-fit test and the KS test preceded by DFT in the context of HTS analysis (Dragiev et al. 2011). A comprehensive simulation study involving artificially generated HTS data was carried out to compare the three above-mentioned tests in a variety of practical situations. The success rate of the t-test was usually above 90%, regardless the plate size, the type, and the magnitude of systematic error, whereas the values of Cohen's kappa coefficient for this test suggested its superior performance, in the case of large plates and high level of systematic bias (Dragiev et al. 2011).

Mann-Whitney-Wilcoxon (MWW) test: This test verifies whether two samples of measurements are identical. First, a suitable Type I error probability, α, is chosen for the test and the data in two samples of interest, X_1 and X_2, are ranked. The MWW test (Gibbons and Chakraborti 2011) is based on Formula 1.10:

$$z = \frac{W_1 - \frac{N_1 \times (N_1 + N_2 + 1)}{2} + C}{\sigma_W}, \tag{1.10}$$

where N_1 and N_2 are the sizes of samples X_1 and X_2, and $W_1 = \sum_{k=1}^{N_1} Rank(X_{1k})$ is the sum of the ranks of the first sample measurements. The correction factor, C, equals 0.5, if the rest of the numerator of z is negative, or equals -0.5, otherwise. The standard deviation, σ_W, is determined using (Formula 1.11):

$$\sigma_W = \sqrt{\frac{N_1 \times N_2 \times (N_1 + N_2 + 1)}{12}} \quad . \tag{1.11}$$

As this is a non-parametric test, it does not make assumptions about the underlying data distribution.

Rank products test: Consider the expression levels of n genes for k_1 independent replicates in sample X_1, and k_2 independent replicates in sample X_2. Let X_{ijm} be the expression level of the i^{th} gene in the j^{th} replicate of the m^{th} sample, where $1 \le i \le n$, $1 \le j \le k_m$, $1 \le m \le 2$. By ranking the expression levels X_{1jm}, X_{2jm}, ..., X_{njm} within each replicate j, we form the vectors $R_{ijm} = rank(X_{ijm})$, where $1 \le R_{ijm} \le n$ and $1 \le m \le 2$. The suitable two-sample version of Breitling's Rank products statistic, RP, for the i^{th} gene can then be calculated by using (Formula 1.12) (Breitling et al. 2004; Koziol 2010):

$$RP_i = \left(\prod_{j=1}^{k_1} R_{ij1} \right)^{1/k_1} / \left(\prod_{j=1}^{k_2} R_{ij2} \right)^{1/k_2} . \tag{1.12}$$

Genes associated with sufficiently large or small RP_i values are marked for further consideration. A few assumptions for this non-parametric test are the following (Breitling et al. 2004): (a) relevant expression changes affect only a minority of genes, (b) measurements are independent between replicated plates (or screens); (c) most changes are independent of each other, and (d) measurement variance is about equal for all genes. The MWW and Rank products tests have been successively applied in the RNAi screening (Rieber et al. 2009)

To estimate the magnitude of systematic bias in experimental HTS data, we carried out a series of tests using the data extracted from the largest public HTS/HCS database (Seiler et al. 2008) Figure (1.4a) reports the average row and column systematic error rates in raw HTS measurements obtained from 41 HTS assays (735 plates in total) aimed at the inhibition of the *E. coli* bacterium. In this analysis, we considered all HTS assays related to the *E. coli* inhibition, that were available in ChemBank (Seiler et al. 2008) as to April 2014. The presented results were obtained by using Welch's t-test (Equation 1.6) with different values of the parameter α = 0.01, 0.025, 0.05, 0.075 and 0.1. The null hypothesis here was that the considered row or column did not contain any systematic bias. Figure (1.4b) illustrates the average hit distribution surface error rates for raw data. The presence of systematic errors in an assay can be determined through the analysis of its hit distribution surface depicting the total hit counts per well location over all plates of the assay (Makarenkov et al. 2007). Thus, we estimated over all assay's plates the number of measurements with the values lower than the μ-$c\sigma$ threshold, where the mean value μ and the standard deviation σ were computed separately for each plate; the constant c was gradually set to 2.5, 3.0 and 3.5 to account for the most popular hit selection thresholds. Here also, Welch's t-test was used to determine the presence or absence of systematic error. Similarly, Figure (1.4c) presents the average row and column error rates for the background-subtracted measurements for the same 41 HTS assays (background-subtracted data were also extracted from ChemBank (Seiler et al. 2008)), and Figure (1.4d) shows the average hit distribution surface error rates for the background-subtracted data. The Matlab 8.2 package (Gilat A. 2014) was used in our computations. The presented graphics suggest that the row and column systematic bias is common to experimental HTS assays (i.e., plate-specific error) – at least 30% of rows and columns in the raw data and 20% of rows and columns in the background-subtracted data were affected by systematic bias (Figure 1.4a and c). Moreover, systematic error is even more visible when analyzing hit distribution surfaces (i.e., assay-specific error) – at least 50% of raw hit distribution surfaces and

65% of background-subtracted hit distribution surfaces were affected by systematic error (Figure 1.4b and d).

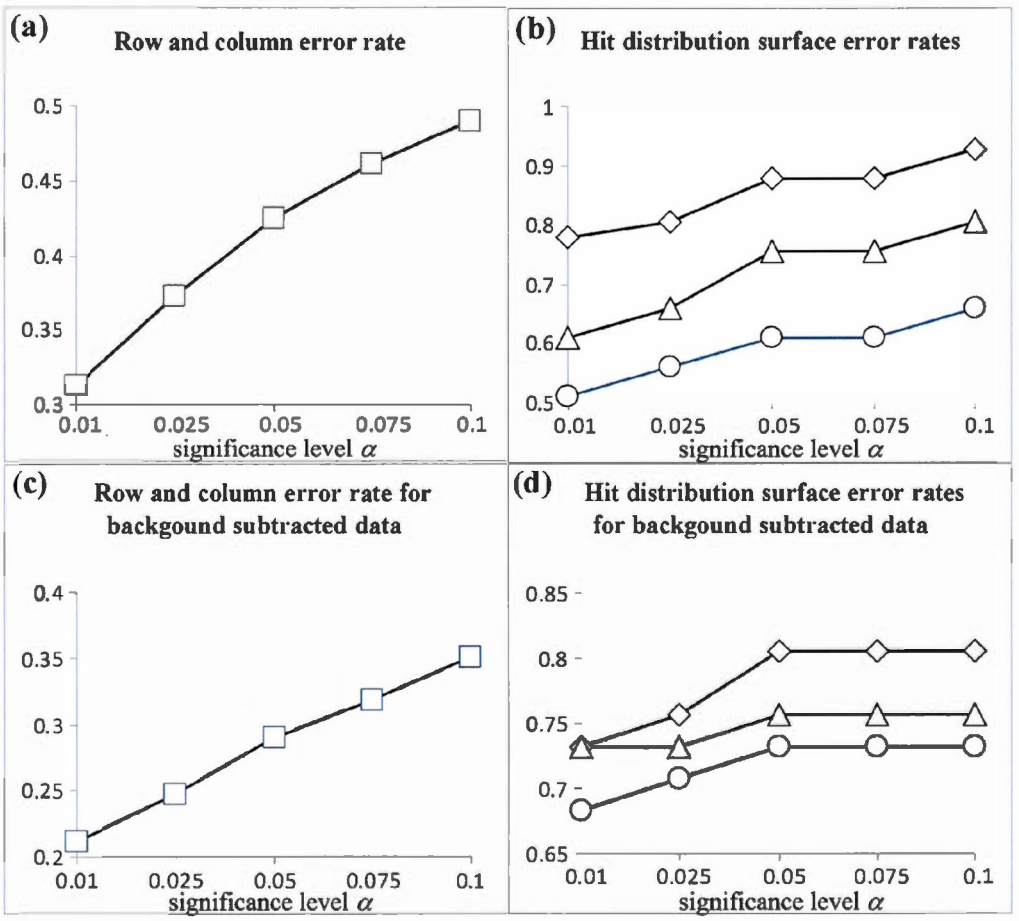


Figure 1.4 Proportion of rows and columns affected by systematic bias in 41 experimental HTS assays (735 plates in total; control wells were ignored) aiming at the inhibition of the *Escherichia coli*. Experimental data were extracted from the Harvard University HTS databank (i.e., ChemBank (Seiler et al. 2008)). Here we show: (a) Overall row and column error rate for raw data; (b) hit distribution surface error rates for raw data; (c) overall row and column error rate for background-subtracted data; (d) hit distribution error rate for background subtracted data. The following hit selection thresholds were used to identify hits and establish hit distribution surfaces of the assays: μ-2.5σ (◊), μ-3σ (Δ) and μ-3.5σ (○), where μ and σ are, respectively, the mean and standard deviation of the plate's measurements.

1.4.6 Data normalization techniques which correct for plate-specific and assay-specific spatial systematic biases

This section describes the statistical methods that are used for minimizing plate-specific and assay-specific (i.e., across-plate well-location bias) systematic biases in screening technologies. Most of these methods allow the correction of overall plate bias as well.

$$x_{ijp} = \mu_p + R_{ip} + C_{jp} + r_{ijp}, \tag{1.13}$$

R-scores: This plate-specific correction method relies on Formula 1.13:
where x_{ijp} is the compound measurement in row i and column j of plate p, μ_p is the mean of plate p, R_{ip} is the row bias affecting row i of plate p, C_{jp} is the column bias affecting column j of plate p and r_{ijp} is the residual in well (i, j) of plate p. These parameters can be estimated using, for example, the *rlm* function from the MASS package of the *R* language (Venables et al. 1994). The R-scores (Wu et al. 2008) are the model's residuals rescaled by dividing them by the standard deviation estimate of the regression function.

B-scores: This method corrects the raw plate measurements by iteratively eliminating possible row and column positional bias (Brideau et al. 2003). The statistical model for the raw measurement x_{ijp} is similar to (Formula 1.13). The B-scores method relies on a two-way median polish (MP) procedure (Tukey 1977) carried out separately for each plate of the assay to obtain the estimates of x_{ijp}, μ_p, R_{ip} and C_{jp}. The residual r_{ijp} of the measurement in well (i, j) is then calculated as the difference between the raw measurement x_{ijp} and its fitted value $\widetilde{x}_{ijp}$: $r_{ijp} = x_{ijp} - \widetilde{x}_{ijp}$. Finally, the obtained residuals are divided by the median absolute deviation of plate p (Formula 1.14):

$$B\text{-}score = \frac{r_{ijp}}{MAD_p}, \text{ where } MAD_p = med\{|r_{ijp} - med(r_{ijp})|\}. \tag{1.14}$$

A variant of the *B*-scores method used in HCS (Gosai et al. 2010) considers the mean true activity value, μ_{ijp}, in well (i, j) in Formula (1.13), instead of μ_p.

Well correction: This assay-specific correction method proceeds by data normalization along the well locations of the assay (Makarenkov et al. 2006; Makarenkov et al. 2007). At first, Z-score normalization (Formula 1.4) is performed within each plate of the assay. The following two steps are then carried out. First, a linear least-square approximation is performed for the measurements of each well location of the assay (this well-specific approximation is done across all plates of the assay). Second, Z-score normalization of the fitted measurements obtained from regression is carried out independently for each well location of the assay (still across all plates of the assay).

Robust well correction: This is another assay-specific data correction procedure. Each plate is normalized using robust Z-scores (Formula 1.5) and then the entire set of plates is ordered by date of processing and a robust regression line is fit to the data. This fitting is carried out independently for each well location across all plates of the assay as in the Well correction method. The obtained normalized residuals are considered as final corrected scores (Murie et al. 2015).

Diffusion Model: This model is designed to eliminate the edge effect in the HTS RNAi screens (Carralot et al. 2012). The process-specific diffusion process is described by the following parabolic differential equation (Formula 1.15):

$$\frac{\partial \tilde{b}(i,j,t)}{\partial t} = c \times \Delta \times \tilde{b}(i,j,t), \tag{1.15}$$

where $\tilde{b}(i,j,t)$ is the evaluated spatiotemporal diffusion field in well (i, j) at time t (i.e., evaluated systematic bias), c is the diffusion coefficient and Δ is the Laplacian operator. The following boundary conditions are considered (Formula 1.16):

$$\begin{cases} \tilde{b}(i,j,t) = U_1, \forall (i,j) \in Z^2 \setminus \Gamma, \\ \tilde{b}(i,j,t=0) = U_0, \forall (i,j) \in \Gamma, \end{cases} \tag{1.16}$$

where U_0 and U_1 are the model's positive parameters; the model also assumes that:

- at the initial time $t = 0$ of the dispensing, there is no edge effect on the given plate;
- the effect strength depends on a physical difference between the inside parameter Γ and outside parameter $Z^2 \setminus \Gamma$ of the given plate.

Loess correction method: The loess error correction method evaluates the plate's row and column effects by fitting a loess curve to each row and column of the given plate (Baryshnikova et al. 2010; Murie et al. 2015). The loess correction is defined by Formula 1.17:

$$\tilde{x}_{ij} = x_{ij} \times \left(\frac{\bar{r}_i}{r_{ij}} \right) \times \left(\frac{\bar{c}_j}{c_{ij}} \right), \tag{1.17}$$

where x_{ij} is the raw measurement in well (i, j), $\tilde{x}_{ij}$ is the adjusted measurement in this well, $\bar{r}_i$ is the mean of the fitted loess curve for row i, $\bar{c}_j$ is the mean of the fitted loess curve for column j, r_{ij} is the value of the fitted row loess curve for row i and column j, and c_{ij} is the value of the fitted column loess curve for row i and column j.

Median Filter: The median filter method (Bushway et al. 2010) adjusts the intensity value of the given well (i, j) using the median of the intensity values of the nearby wells. First, a row median filter, whose filter window includes the wells located on the same row i, within k wells of well (i, j), is carried out. Second, a standard median filter procedure, its filter window includes the wells located within l wells of well (i, j), is applied. The constants k and l usually equal 3 for the 1536-well plates, and 1 and 2 for the 96-well plates. The method relies on Formula 1.18 to compute the adjusted measurements:

$$\tilde{x}_{ij} = x_{ij} \times \left(\frac{med_p}{med_{wij}} \right), \tag{1.18}$$

where med_p is the median intensity of plate p and med_{wij} is the median intensity of wells included in the filter window of well (i, j).

SPatial And Well Normalization (SPAWN): This two-step procedure gradually applies a trimmed mean polish procedure on individual plates in order to minimize row and

column systematic effects (Murie et al. 2013). The considered statistical model relies on Formula 1.13. Then, a well normalization step is carried out to determine spatial bias template, SBT_{ij}, which is the median of the scores at well location (i, j) computed over all plates of the assay. The spatial bias template scores are subtracted from the scores obtained by the median polish procedure: $\hat{r}_{ijp} = r_{ijp} - SBT_{ij}$. Finally, the resulting scores are rescaled by dividing them by the median absolute deviation of the plate. Thus, SPAWN corrects for both plate-specific and assay-specific biases.

Matrix Error Amendment (MEA) and *Partial Mean Polish (PMP)*: These algebraic methods are designed to modify only those rows and columns of the given plate that are affected by systematic bias (Dragiev et al. 2012). MEA and PMP methods rely on prior information concerning the presence and absence of systematic error in the rows and columns of the given plate. Such information can be obtained using a specific version of Welch's t-test or the χ^2 goodness-of-fit test (see previous section). One of the main advantages of the PMP method over MP and B-scores (Brideau et al. 2003) is that PMP does not reduce the original data to residuals, keeping the corrected measurements on the same scale with the original ones.

Table 1.1 reports the discussed data normalization techniques recommended for the analysis of HTS and HCS data along with the underlying assumptions regarding their practical application.

Table 1.1 Data normalization methods recommended for the analysis of HTS and HCS data, and classified according to the context in which they should be applied. The available software implementations are also indicated.

Method	Overall plate bias correction	Spatial systematic bias correction	Removes plate-specific spatial bias	Removes assay-specific spatial bias	Data randomization assumption	Software available
Control normalization	Yes	No	No	No	Control placements are preferably randomized within plates	In many software; readily implemented in MS Excel and R
Median percent inhibition	Yes	No	No	No	No any	In many software; readily implemented in MS Excel and R
Z-score	Yes	No	No	No	No any	In many software; readily implemented in MS Excel and R
Robust Z-score	Yes	No	No	No	No any	SIGHTS(Murie, et al., 2015)
R-score	Yes	Yes	Yes	Yes (RW function in SIGHTS)	Samples randomized within plates	rlm (MASS package) function in R, SIGHTS
B-score	Yes	Yes	Yes	No	Samples randomized within plates	BioConductor (cellHTS, RNAither) (Boutros, et al., 2006), HTS-Corrector (Makarenkov, et al., 2006), HCS-Analyzer(Ogier and Dorval, 2012), HTS-Helper(Dragiev, et al., 2012)
Well correction	Yes	Yes	No	Yes	Samples randomized across different well locations of the assay	HTS-Corrector, HTS-Helper, SIGHTS
Robust well correction	Yes	Yes	No	Yes	Samples randomized across different well locations of the assay	SIGHTS
Diffusion model (HCS)	Yes	Yes	Yes	No	Samples randomized within plates	HCS-Analyzer
Loess correction	Yes	Yes	Yes	No	Samples randomized within plates	SIGHTS, BioConductor (RNAither)
Median filter	Yes	Yes	Yes	No	Samples randomized within plates	SIGHTS, HMF (Bushway, et al., 2010)
SPAWN	Yes	Yes	Yes	Yes (SPAWNW function in SIGHTS)	Samples randomized within plates and across different well locations	SIGHTS
Matrix error amendment	No	Yes	Yes	No	Samples randomized within plates	HTS-Helper
Partial mean polish	No	Yes	Yes	No	Samples randomized within plates	HTS-Helper
IQEM (HCS)	No	Yes	Yes	No	Samples randomized within plates	IQEM code, MATLAB (Lo et al. 2012)

Various plots that use robust statistical indices have been also suggested for detecting shifts and trends across time in large screening campaigns. Systematic bias within plates can be detected with visualization methods such as 2-dimensional heat maps and 3-dimensional wire plots, although typical plate-specific bias patterns are more easily detected with auto-correlation plots that show the degree of correspondence between wells at various "lags" (e.g., adjacent or separated by one well) (Murie et al. 2015). Finally and somewhat counterintuitively, screens with few active compounds should show low correlations between replicate plates; for these screens, scatterplots which show high correspondence between replicate plates indicate across-plate well-specific bias rather than good biological reproducibility (Murie et al. 2015).

1.5 Discussion and conclusion

We reviewed current knowledge on systematic bias affecting raw data in HTS and HCS technologies. First, we discussed the causes of systematic bias and its impact on the selection of correct hits in HTS and HCS experiments. The main steps of HTS and HCS screening protocols were presented along with the subcategories of screening technologies, including small molecule, cDNA and RNAi screens. Positional bias effects characteristic of screening technologies, comprising batch effects, edge effects and well location effects, were discussed in detail. We highlighted that randomization of experimental units and use of replicates can significantly reduce the magnitude of systematic error. Data normalization techniques which correct for overall plate bias were presented, followed by the description of systematic error detection tests specific to screening technologies. Finally, we discussed error correction methods, indicating under which assumptions and for which kind of spatial bias each of them should be used. In particular, we underlined the distinction between the plate-specific and assay-specific systematic biases and pointed out that data correction methods should be applied only if the presence of systematic bias was confirmed by the appropriate statistical tests. Otherwise, an unwanted bias can be introduced into error-free data.

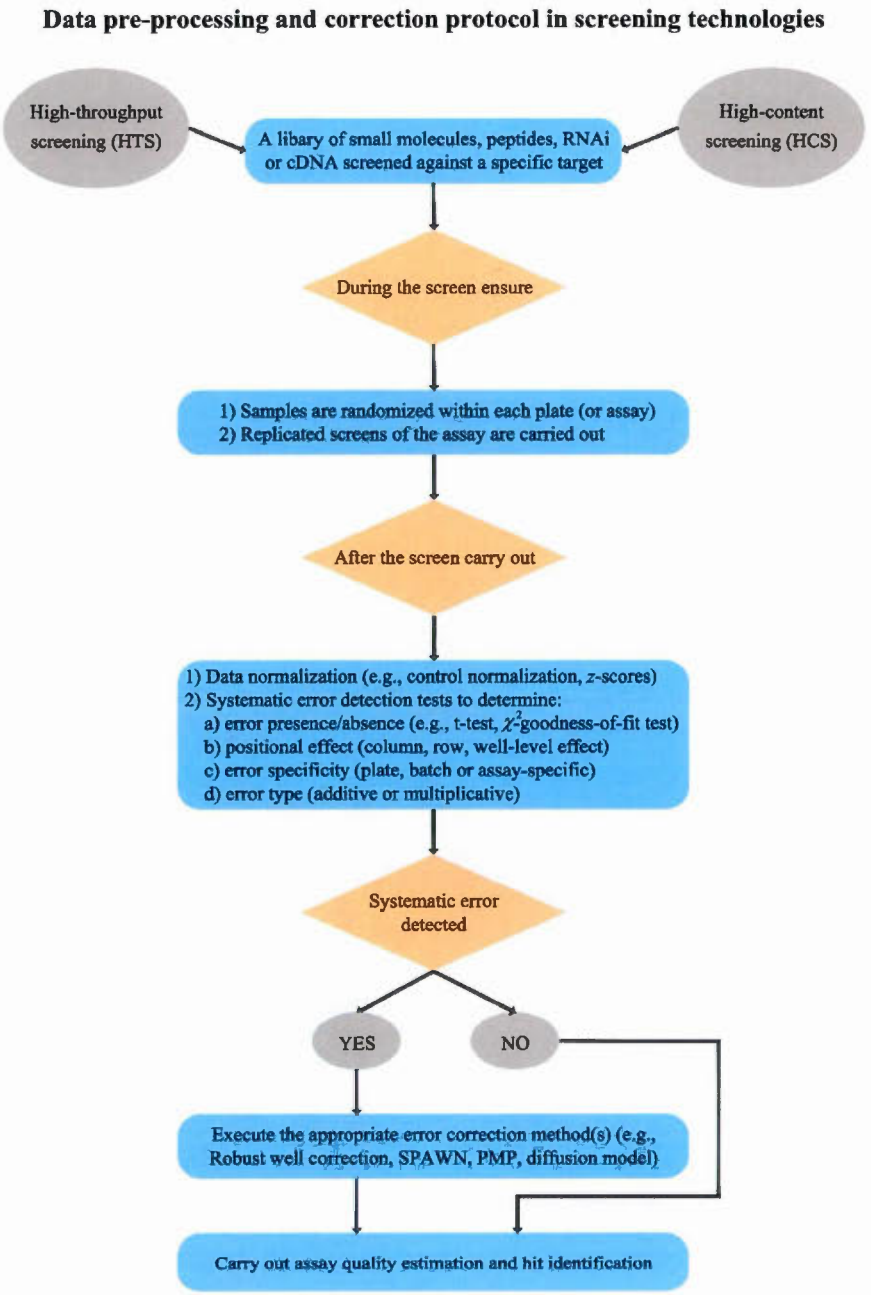


Figure 1.5 Recommended data pre-processing and correction protocol to be performed prior to the hit identification step in high-throughput and high-content screening.

In order to summarize our presentation, we describe here a general data pre-processing and correction protocol (Figure 1.5), which could be used as a guide by academic and industrial researchers involved in the analysis of current or next generation screening data. The first required step concerns general design of a screening campaign. The compound locations within each plate, as well as over all plates of the assays, should be randomized in order to reduce the impact of systematic bias on the outcome of screening experiments. Moreover, whenever the campaign funding allows, several replicates of the compound library should be screened. Replicated screens provide both a greater precision of activity measurements and the ability to assess measurement variability (Malo et al. 2006). Once the assay measurements have been established, the appropriate data normalization procedure should be carried out to ensure the data comparability over different plates and screening conditions. Afterwards, systematic error detection tests should be carried out to confirm the presence or absence of systematic error in raw data (e.g., Welch's t-test or χ^2 goodness-of-fit test). In particular, these tests can be applied to identify: (1) positional effects of systematic error, including row, column and well location biases; (2) error specificity, including plate, batch and assay-specific biases; (3) type of systematic error, including additive (e.g., Robust well correction, SPAWN or PMP methods can be applied to eliminate this type of bias) and multiplicative (e.g., diffusion model can be applied to eliminate this type of bias) biases. If systematic error was not detected in the data, then no any correction method needs to be applied to them to avoid the risk of introduction of additional biases (Dragiev et al. 2011). Otherwise, the appropriate error correction method, preferably including a success of control step, should be carried out. Once systematic bias is minimized, assay quality estimation and hit identification steps can be carried out. It is worth noting that the plate-specific correction methods (e.g., PMP) can sometimes be applied in combination with the assay-specific correction methods (e.g., Robust well correction). First, Welch's t-test can be carried out independently for each individual plate of the assay to detect the plate's rows and columns affected by systematic bias.

The measurements affected by bias can be subsequently corrected by using the PMP method, which keeps the corrected data on the same scale with the original ones. Second, Welch's t-test can be performed over the hit distribution surface of the entire assay. If the test identifies the presence of systematic bias on the surface, the Robust well correction procedure can be carried out to remove the assay-specific bias. An alternative solution to this problem could be provided by the methods which correct for both plate-specific and assay-specific biases (e.g., SPAWN).

Key Points

- We reviewed current knowledge on systematic bias affecting experimental HTS and HCS data.
- Study design issues and the way in which data are generated are crucial for providing unbiased screening results. Unfortunately, these key steps are often ignored by HTS practitioners.
- Data correction methods should be applied only if the presence of systematic error has been confirmed by the appropriate statistical tests.
- Discussed sources of systematic bias and presented statistical methods and software intended to correct experimental screening data provide a unifying framework when considering new screening technologies.
- We presented a general data pre-processing and correction protocol which can be used as a guide by academic and industrial researchers involved in the analysis of current or next generation screening data.

CHAPTER II

DETECTING AND REMOVING MULTIPLICATIVE SPATIAL BIAS IN HIGH-THROUGHPUT SCREENING TECHNOLOGIES

This chapter is a reproduction of the following article: Caraus Iurie, Bogdan Mazoure, Robert Nadon and Vladimir Makarenkov. "Detecting and removing multiplicative spatial bias in high-throughput screening technologies", *Bioinformatics* (2017), doi: 10.1093/bioinformatics/btx327.

2.1 Abstract

2.1.1 Motivation

Considerable attention has been paid recently to improve data quality in high-throughput screening (HTS) and high-content screening (HCS) technologies widely used in drug development and chemical toxicity research. However, several environmentally- and procedurally-induced spatial biases in these screens decrease measurement accuracy, leading to increased numbers of false positives and false negatives in hit selection. Although effective bias correction methods and software have been developed over the past decades, almost all of these tools have been designed to reduce the effect of additive bias only. Here, we address the case of multiplicative spatial bias.

2.1.2 Results

We introduce three new statistical methods meant to reduce multiplicative spatial bias in screening technologies. We assess the performance of the methods with synthetic and real data affected by multiplicative spatial bias, including comparisons with current bias correction methods. We also describe a wider data correction protocol that integrates methods for removing both assay and plate-specific spatial biases, which can be either additive or multiplicative.

2.1.3 Conclusions

The methods for removing multiplicative spatial bias and the data correction protocol are effective in detecting and cleaning experimental data generated by screening technologies. As our protocol is of a general nature, it can be used by researchers analyzing current or next-generation high-throughput screens.

2.1.4 Availability and Implementation

The *AssayCorrector* program (AssayCorrector), implemented in R, is available at CRAN.

2.2 Introduction

Growing interest of the pharmaceutical industry has stimulated the development of several effective screening techniques such as high-throughput screening (HTS) and high-content screening (HCS). Modern HTS and HCS screening campaigns allow for examining hundreds of thousands of chemical compounds and generating gigabytes of experimental data (Lachmann et al. 2016). These data need to be extensively

analyzed using appropriate statistical methods and protocols to detect potential drug candidates, called hits (Birmingham et al. 2009; Malo et al. 2006). A typical HTS assay is organized as a sequence of microtiter plates, featuring a grid of wells, which contain test samples. The most common plate formats consist of 96, 384, 1536 and 3456-well plates.

Spatial bias within plates (i.e., positional bias or systematic error) remains one of the major hurdles of experimental screening campaigns. It can be caused by a number of technical or environmental factors, including reader and pipette effects, liquid processing anomalies, unintended variations in compound concentration associated with agent evaporation, irregular changes in the incubation time, or temperature, lighting and air flow fluctuations (Heyse 2002; Makarenkov et al. 2007). Spatial bias is evident as systematic under- or over-estimation of specific screen measurements (Kevorkov and Makarenkov 2005). It is a constant source of false positives (i.e., inactive compounds incorrectly identified as hits) and false negatives (i.e., undetected active compounds) in screening technologies (Birmingham et al. 2009). Typically, spatial bias affect either compounds placed in the same well, row or column location over all plates of the assay (i.e., *assay-specific error*) or compounds from a specific row or column of a given plate (i.e., *plate-specific error*) (Dragiev et al. 2012). As we will show in this paper, spatial bias can also be of the additive or multiplicative nature.

Figure 2.1 illustrates an example of positional bias that affects the RNAi HIV inhibition assay screened at Pasteur Institute of Korea (Carralot et al. 2012). This cell-based assay uses HeLa P4 LTR-EGFP 2B4 cells, engineered to express the HIV cellular-entry receptors CD4 and CCR5. Both assay-specific (Figure 2.1a) and plate-specific (Figure 2.1b) spatial biases are present in this assay. Moreover, these biases exhibit opposite trends. On one hand, higher hit counts can be observed in the middle of the hit distribution surface (Figure 2.1a; hit counts per well location are depicted),

whereas lower hit counts can be observed at its edges (in rows A, O and P and in columns 23 and 24). On the other hand, the measurements in rows A, C and P and in columns 23 and 24 of Plate 7 (Figure 2.1b) are systematically overestimated (hits correspond to the lowest measurements in inhibition assays).

Several data normalization and quality control techniques have been proposed to allow for an efficient evaluation and validation of HTS and HCS assays (Birmingham et al. 2009; Brideau et al. 2003; Carralot et al. 2012; Dragiev et al. 2011; Dragiev et al. 2012; Kevorkov and Makarenkov 2005; Makarenkov et al. 2007; Malo et al. 2006; Shun et al. 2011; Zhang et al. 1999; Zhang 2008). The most popular data normalization methods used to compare experimental measurements across different plates of a given assay are Percent of Control (POC), Normalized Percent Inhibition (NPI), Z-score and robust Z-score (Birmingham et al. 2009; Malo et al. 2006). These methods, however, do not correct for spatial bias. A number of correction tools for the detection and removal of spatial bias from experimental screening data have been proposed (Caraus et al. 2015; Mpindi et al. 2015).

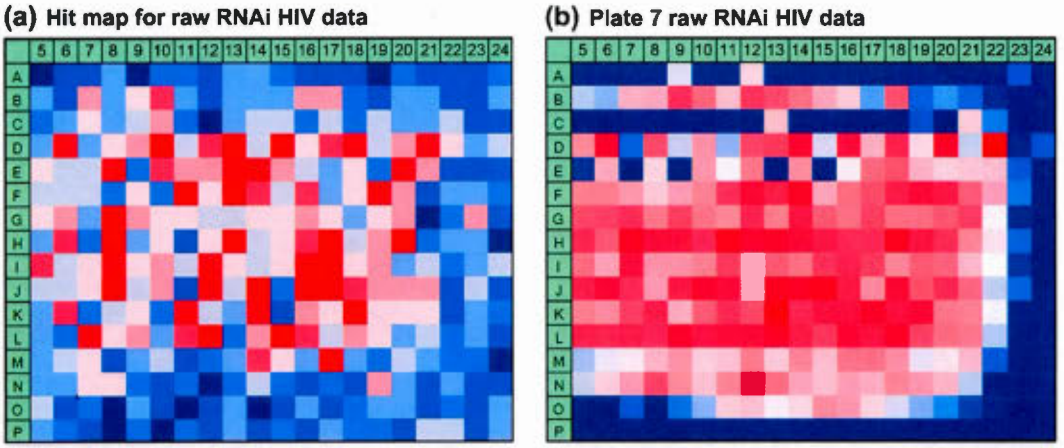


Figure 2.1 Edge effects in the RNAi HIV inhibition assay, screened at Pasteur Institute of Korea (Carralot et al. 2012), shown within: (a) the assay hit distribution surface (i.e., overall hit counts per well location are depicted) and (b) Plate 7 raw activity levels. The hit selection threshold of μ-1.219σ (or 1% of hits) was used, where μ and σ were the mean and the standard deviation of the plate measurements, respectively. The samples whose measurements were lower than the threshold value were selected as hits. In (a), higher hit counts are in red and lower hit counts are in blue. In (b), higher raw measurements are in blue and lower raw measurements are in red. The non-target wells located in columns 1 to 4 were ignored in our analysis.

The popular B-score method (Brideau et al. 2003) relies on the median polish procedure (Tukey 1977) in order to remove plate-specific spatial bias. Well Correction (Makarenkov et al. 2007) removes assay-specific bias using Z-score normalization and linear regression across well locations. R-score (Wu et al. 2008) is a plate-specific bias correction method which fits a robust linear model to experimental well measurements. Dragiev et al. (2012) presented the partial mean polish method intended for removing the additive plate-specific spatial bias from the data by correcting the biased measurements only (i.e., aPMP method). The SPAWN

method (Murie et al. 2013) uses an iterative approach based on trimmed mean polishing to minimize row and column systematic effects within a given plate. Then, a well normalization step can be carried out to create a spatial bias template, which is used to correct assay-specific bias. However, almost all data correction methods, including B-score, R-score, aPMP and SPAWN, employed in screening technologies are designed to remove additive type of spatial bias. A notable exception is the diffusion model (Carralot et al. 2012), proposed to eliminate edge effects in RNAi HTS due to multiplicative bias.

The additive spatial bias model can be described by Equation (2.1):

$$\hat{x}_{ijp} = x_{ijp} + R_{ip} + C_{jp} + \varepsilon_{ijp}, \tag{2.1}$$

whereas the multiplicative bias model by Equation (2.2):

$$\hat{x}_{ijp} = x_{ijp} \times R_{ip} \times C_{jp} + \varepsilon_{ijp}, \tag{2.2}$$

where $\hat{x}_{ijp}$ is the resulting (biased) measurement value in well (i,j) of plate p, x_{ijp} is the original error-free measurement, R_{ip} is the bias affecting row i of plate p, C_{jp} is the bias affecting column j of plate p and ε_{ijp} is the random error in well (i,j) of plate p.

The most straightforward approach for removing multiplicative spatial bias in screening technologies consists of the use of a logarithmic transformation of raw measurements, followed by the application of one of the above-mentioned additive bias correction methods. However, a typical data preprocessing step in HTS consists of normalizing raw measurements (e.g., plate or well-wise using Z-score) prior to their correction and/or hit selection. Thus, the logarithmic transformation cannot be applied in many cases because the normalized measurements contain negative values. Furthermore, the multiplicative PMP method presented in Section 2.2.3 allows for minimizing multiplicative spatial bias by modifying only the biased measurements and keeping the corrected data on the same scale with the unbiased raw data. This property makes the data corrected by the proposed multiplicative PMP method directly comparable to those corrected by the additive PMP method (Dragiev et al.

2012). This is further exploited in our general data processing algorithm presented in Section 2.2.5, where we show how to select the most appropriate bias correction model (additive or multiplicative) for the data at hand. Obviously, such a model selection algorithm could not be carried out if one of the two methods would work with the raw data, while the other with the log-transformed data.

Once spatial bias has been removed from the data, assay quality estimation and hit identification steps can be carried out (Birmingham et al. 2009). The most popular hit selection procedure identifies as hits the samples whose measurements are lower (for inhibition assays) or higher (for activation assays) than the selected μ-$c\sigma$ threshold, where μ is the mean, σ is the standard deviation of the targeted group of measurements, and c is the selected constant (often varying between 1 and 3).

In this article, we describe and compare three novel methods for removing the multiplicative spatial bias from experimental screening data. We then present a comprehensive bias correction protocol, which can be used to remove both additive and multiplicative spatial biases, as well as both assay and plate-specific biases. We apply the methods to real data to correct the RNAi measurements generated during a genome-wide siRNA screen aimed at studying HIV-host interactions (Carralot et al. 2012) and determine the dominant type of plate-specific spatial bias characterizing the four HTS screening categories (homogeneous, microorganism, cell-based and gene-expression) available in the ChemBank database (Seiler et al. 2008).

2.3 Methods

2.3.1 Three new methods for correcting multiplicative bias

Here, we describe and compare three new methods for correcting multiplicative spatial bias in screening technologies. These methods are designed to remove plate-specific bias. Ideally, compound measurements and hit counts should be uniformly

distributed over a given plate and the hit distribution surface of the assay. The first method, called NLMBE, solves a system of nonlinear algebraic equations in which the unknowns correspond to spatial biases, which affect specific rows and columns of a given plate. The second method, called multiplicative PMP (mPMP), is based on a multiplicative partial mean polish procedure in which the mean of each row and each column affected by spatial bias is adjusted iteratively with respect to the mean of the unbiased plate measurements. Rows and columns affected by spatial bias can be detected, for example, by the Mann–Whitney *U* test (Gibbons and Chakraborti 2011; Wilcoxon 1945). This information is required by both NLMBE and mPMP. The third method, called multiplicative B-score, is an adaptation of a 2-way median polish procedure (Tukey 1977) to the case of multiplicative bias.

2.3.2 Non-Linear Multiplicative Bias Elimination (NLMBE)

The first method proceeds by solving a system of nonlinear equations in which the unknowns correspond to systematic errors affecting the measurements of a given plate. Let x_{ijp} be experimental measurements of plate p, where (i =1, …,m) and (j =1, …, n), having m rows and n columns,. Assume that some rows and/or columns of p contain spatial bias (as indicated for example by the Mann-Whitney *U* test). Let μ be the mean (or median) of the measurements of plate p that do not contain spatial bias. The following system of nonlinear equations can be composed:

$$R_{ip} \times (\sum_{j=1}^{n} C_{jp} \times x_{ijp} / W_{ij}) = n \times \mu, \text{ for all } i=1,...,m, \tag{2.3}$$

$$C_{jp} \times (\sum_{i=1}^{m} R_{ip} \times x_{ijp} / W_{ij}) = m \times \mu, \text{ for all } j=1,...,n, \tag{2.4}$$

where R_{ip} is the multiplicative bias affecting row i of plate p, C_{jp} is the multiplicative bias affecting column j of p and W_{ij} is the systematic measurement offset in well (i,j) across all plates of the assay. W_{ij} can be estimated directly from the assay background

surface. This system of nonlinear equations is obtained by summing up all multiplicative bias terms corresponding to rows and columns of p. It includes $m + n$ equations and at most $m + n$ unknowns (i.e., biases R_{ip} and C_{jp} affecting the rows and columns of plate p). The values of R_{ip} and C_{jp} that correspond to row i and column j not affected by bias are equal to 1. Since in practice only a few rows and columns per plate are affected by spatial bias, this system will have more equations than unknowns. Such systems can be solved, for example, by using the Levenberg-Marquardt (Moré 1978) method based on nonlinear least-squares. Note that the Levenberg-Marquardt algorithm has a quadratic convergence rate. The corrected plate measurements $\hat{x}_{ijp}$ (Equation 2. 2) can then be calculated taking into account the system's solution, i.e., the obtained values of R_{ip} ($i = 1,\ldots,m$) and C_{jp} ($j = 1,\ldots,n$).

2.3.3 Multiplicative PMP method (mPMP)

Our second method is based on an iterative procedure in which the mean of each row and each column affected by systematic error is gradually adjusted with respect to the mean of the plate measurements not affected by spatial bias. Assume that rows r_1, r_2, $\ldots,r_k$ and columns c_1, c_2, …, c_l of plate p contain multiplicative spatial bias. First, for any row i affected by spatial bias ($i = r_1, r_2, \ldots, r_k$), we calculate:

$$x_{ijp} = \mu \times x_{ijp} / (\mu_i \times W_{ij}), \text{ for all } j=1,\ldots, n, \tag{2.5}$$

and for any column j affected by systematic bias ($j = c_1, c_2, \ldots, c_l$), we calculate:

$$x_{ijp} = \mu \times x_{ijp} / (\mu_j \times W_{ij}), \text{ for all } i=1,\ldots, m, \tag{2.6}$$

where μ_i is the mean of row i, μ_j is the mean of column j, μ is the mean of the plate's measurements that are not affected spatial bias and W_{ij} is the systematic measurement offset in well (i,j) across all plates of the assay. When a large number of hits or outliers are expected, the means of the plate's unbiased measurements, rows and columns should be replaced by the corresponding medians in order to obtain more

robust parameter estimates. This iterative procedure should be repeated until a desired convergence threshold is reached. The time complexity of mPMP is $O(mnI)$, where m and n are the plate dimensions and I is the number of iterations required for convergence. In practice, this method converges after a few iterations. Importantly, mPMP is usually much faster than NLMBE. The main advantages of the NLMBE and mPMP algorithms are that they modify only biased data and keep raw and corrected plate measurements on the same scale.

2.3.4 Multiplicative B-score method

We also present the multiplicative version of the well-known B-score algorithm (Brideau et al. 2003). The conventional (additive) B-score is a robust data correction procedure widely used in experimental screening technologies. Our multiplicative B-score transformation assumes the following bias model:

$$\hat{x}_{ijp} = \mu_p \times R_{ip} \times C_{jp} \times W_{ij}, \tag{2.7}$$

where $\hat{x}_{ijp}$ is the estimated (biased) activity measurement in well (i,j) of plate p, μ_p is the average of plate p, R_{ip} is the spatial bias affecting row i, C_{jp} is the spatial bias affecting column j and W_{ij} is the systematic measurement offset in well (i,j). Our method is based on a 2-way median polish procedure (Tukey 1977) in which subtractions are replaced by divisions in order to remove multiplicative spatial bias from all rows and all columns of p. The residual, r_{ijp}, of the measurement in well (i,j) is defined as the difference between the raw measurement x_{ijp} and its fitted value $\hat{x}_{ijp}$:

$$r_{ijp} = x_{ijp} - \hat{x}_{ijp}, \tag{2.8}$$

where x_{ijp} is the raw measurement in well (i,j) of plate p. Finally, the B-score is calculated as follows:

$$Bscore = \frac{r_{ijp}}{MAD_p}, \quad MAD_p = median\{|r_{ijp} - median(r_{ijp})|\}, \tag{2.9}$$

where MAD_p is the adjusted median absolute deviation obtained from the residuals of plate p. The time complexity of the multiplicative B-score method is also $O(mnl)$.

2.3.5 General data correction protocol

Here we present a complete bias correction protocol that can be used to remove both multiplicative and additive spatial biases, which can be assay or plate-specific.

First, *assay-specific bias* can be removed from a given assay by applying either the Well Correction (Makarenkov et al. 2007), based on Z-score normalization, or SPAWN (Murie et al. 2013), based on the robust Z-scores normalization, method. These methods normalize the measurements of specific well locations in which the presence of spatial bias has been detected (Dragiev et al. 2011). It is worth noting that the conventional Z-scores, when applied well-wise, allow for removing both additive and multiplicative spatial biases (Brideau et al. 2003). Following these normalizations, some data will become negative, thus making the use of the logarithmic transformation impossible.

Second, we propose the following algorithm to detect and remove *plate-specific spatial bias*. This algorithm should be applied in turn on all plates of a given assay.

Carry out the Mann–Whitney U test on each plate of the assay in order to detect biased rows and columns. This test will allow us to compare the sum of ranks of a given row or column to the sum of ranks of the rest of the plate's measurements.

If (spatial bias is detected in some row(s) or column(s) of the plate), **then**:

1. **Use** the additive PMP method (Dragiev et al. 2012) to correct the plate's measurements.

2. **Use** the multiplicative PMP method to correct the plate's measurements.

3. **Carry out** the Kolmogorov-Smirnov two-sample test in order to compare first the distributions of unbiased measurements with those corrected by the additive PMP, and, second, the distributions of unbiased measurements with those corrected by the multiplicative PMP. **Compute** the p-values associated with these two corrections.

4. **If** (either the additive or multiplicative p-value from the previous step is larger than the selected significance level α), **then apply** the correction algorithm that yields the highest p-value (i.e., additive or multiplicative PMP) to remove spatial bias from the plate's data;

otherwise, the bias model for this plate is *undetermined.*

Here, the Mann-Whitney U test is applied on a plate-by-plate basis. The measurements of a considered row or column of a given plate compose the first vector used in the Mann-Whitney U test and the rest of the plate's data compose the second vector used in this test. If enough evidence for the presence of spatial bias, expressed through the test's p-value, is obtained, the corresponding row or column is flagged as biased and removed from the computation. If no biased rows or columns have been found at the current iteration, the procedure is stopped. In our study, the maximum number of iterations allowed by the algorithm was limited to 50% of the total number of rows (when the presence of bias in rows was examined) and 50% of the total number of columns (when the presence of bias in columns was examined) of a given plate.

It is worth noting that when the plate's background estimation is close to zero, the Mann-Whitney U test should not detect any biased row or column within a given plate. We compared the results of the Mann-Whitney U test to those of the t-test used by Dragiev et al. (2012) in terms of spatial bias detection, and found that the Mann-Whitney U test is more robust in this context and thus better suited for spatial bias

detection purposes in screening technologies. The main advantages of the Mann-Whitney U test compared to the t-test are that it does not make any distributional assumptions and is more robust to outliers.

2.4. Results

To evaluate the performance of the three spatial bias correction methods, we carried out simulations with artificially generated screening data. Afterwards, we compared the most successful of them, multiplicative PMP, to the existing data correction techniques such as B-score (Brideau et al. 2003) and diffusion model (Carralot et al. 2012; Ogier and Dorval 2012) considering RNAi HIV inhibition assay, screened at Pasteur Institute of Korea (Carralot et al. 2012). We also used our general data correction protocol to assess the extent of plate-specific bias across the data corresponding to different HTS categories available in ChemBank (Seiler et al. 2008).

2.4.1. Simulation study

Our simulation study was conducted using randomly generated 1000-plate assays. The three considered plate sizes were as follows: 96-well plates (8 rows × 12 columns), 384-well plates (16 rows × 24 columns) and 1536-well plates (32 rows × 48 columns). The values of inactive activity measurements followed a normal distribution with parameters ($\mu = 7.344$ and $SD = 1$), where μ and SD were the mean and the standard deviation of the plate's measurements. Active compounds (hits) were randomly added to the plates to obtain assays with the following hit percentages: 0%, 0.5%, 1%, 2%, 3%, 4% and 5%. Hit locations on each plate were randomly chosen following a uniform distribution. Hit measurements were generated via sampling from a normal distribution with parameters $\sim N(\mu\text{-}1.67SD,\ SD)$. A

multiplicative spatial bias was randomly assigned to some rows and columns of all plates of the assay. The bias value was selected following a normal distribution with parameters $\sim N(0, C)$, with C equal to 0, $0.1SD$, $0.2SD$, $0.3SD$, $0.4SD$ and $0.5SD$.

In our simulations, the number of rows and columns of a given plate which could be affected by systematic bias was limited to a maximum of 4 rows and 4 columns for 96 and 384-well plates, and a maximum of 5 row and 5 columns for 1536-well plates. A small random noise was also added to both hit and non-hit measurements on all plates. The noise values followed a normal distribution with parameters $\sim N(0, 0.5SD)$. The biased measurements of a given plate were generated using Equation (2.2). The four following data correction methods were tested in our simulations: No Correction, NLMBE, mPMP, and multiplicative B-score. The Mann-Whitney U test was used in the NLMBE and mPMP methods to identify the rows and columns affected by spatial bias. The hits were chosen globally across all assays by using the hit selection threshold of μ_{hs}-1.67SD_{hs} (i.e., all measurements lower than μ_{hs}-1.67SD_{hs} were chosen as hits, where μ_{hs} and SD_{hs} were the mean and the standard deviation of a given assay after the addition of both hits and spatial bias). The performance of our data preprocessing techniques was assessed by measuring the total number of false positives (FP) and false negatives (FN) as well as by computing the hit detection rate (i.e., the true positive rate) for all methods.

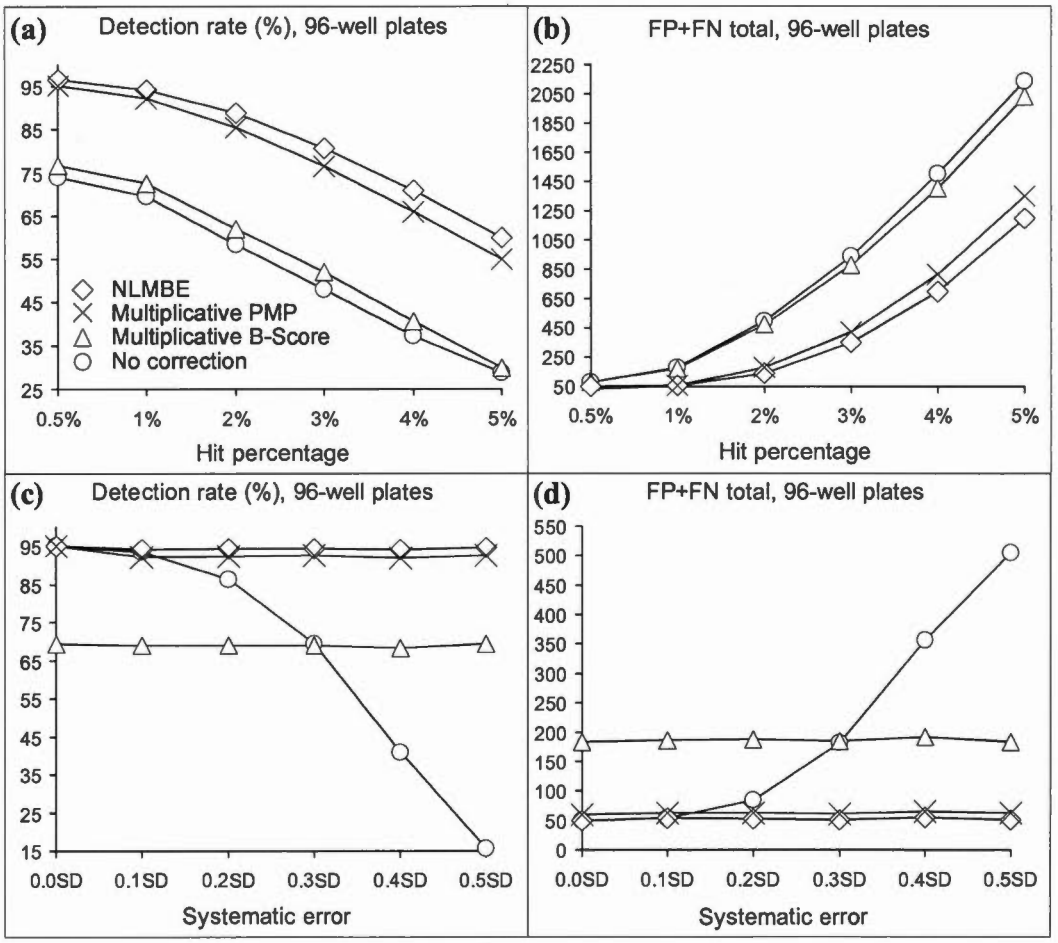


Figure 2.2 True positive rate, and combined false positive and false negative rate for 96-well plate assays obtained under the condition that a maximum of 4 rows and 4 columns of each plate were affected by spatial bias. Panels (a and b) present the results for datasets with the fixed standard deviation of spatial bias, equal to 0.3*SD*. Panels (c and d) present the results for datasets with the fixed hit percentage rate of 1%.

Two series of experiments were conducted, by varying either the hit percentage or the bias level. In the first set of experiments, 1000 different assays with a fixed standard deviation of bias, equal to 0.3*SD*, and the hit percentage rate varying from 0% to 5%

were generated for each plate size (there are no true positives for the case of 0% of hits; Figures 2.2-2.4a and b). In the second series of experiments, 1000 different assays with the fixed hit percentage of 1% and the standard deviation of bias varying from 0 to 0.5*SD* were generated for each plate (Figure 2.2-4c and d). Figures 2.2, 2.3 and 2.4 present the average results generated by the four compared methods for the 96-well, 384-well and 1536-well plates, respectively. The results of our simulations suggest that the NLMBE and mPMP methods clearly outperformed the No Correction procedure regardless of plate size, hit percentage and spatial bias variance (see Figures 2.2 to 2.4). Moreover, NLMBE and mPMP usually outperformed the multiplicative B-score method. This trend is most noticeable with 96-well plates. The results of multiplicative B-score improved with the increase in plate size. It is worth noting that the multiplicative B-score method was very prone to generating false positives, especially for 96 and 384-well plates (see also (Mpindi et al. 2015) for a discussion on drawbacks of this method). The NLMBE method generally achieved slightly better performances than mPMP in terms of detection rate (i.e., true positive rate). However, mPMP was slightly better than NLMBE in terms of the combined false positive and false negative rate. Considering that mPMP converges much fast than NLMBE, it can be recommended for reducing multiplicative spatial bias in HTS. Thus, the mPMP method was used in our experiments with real data presented in the next sections.

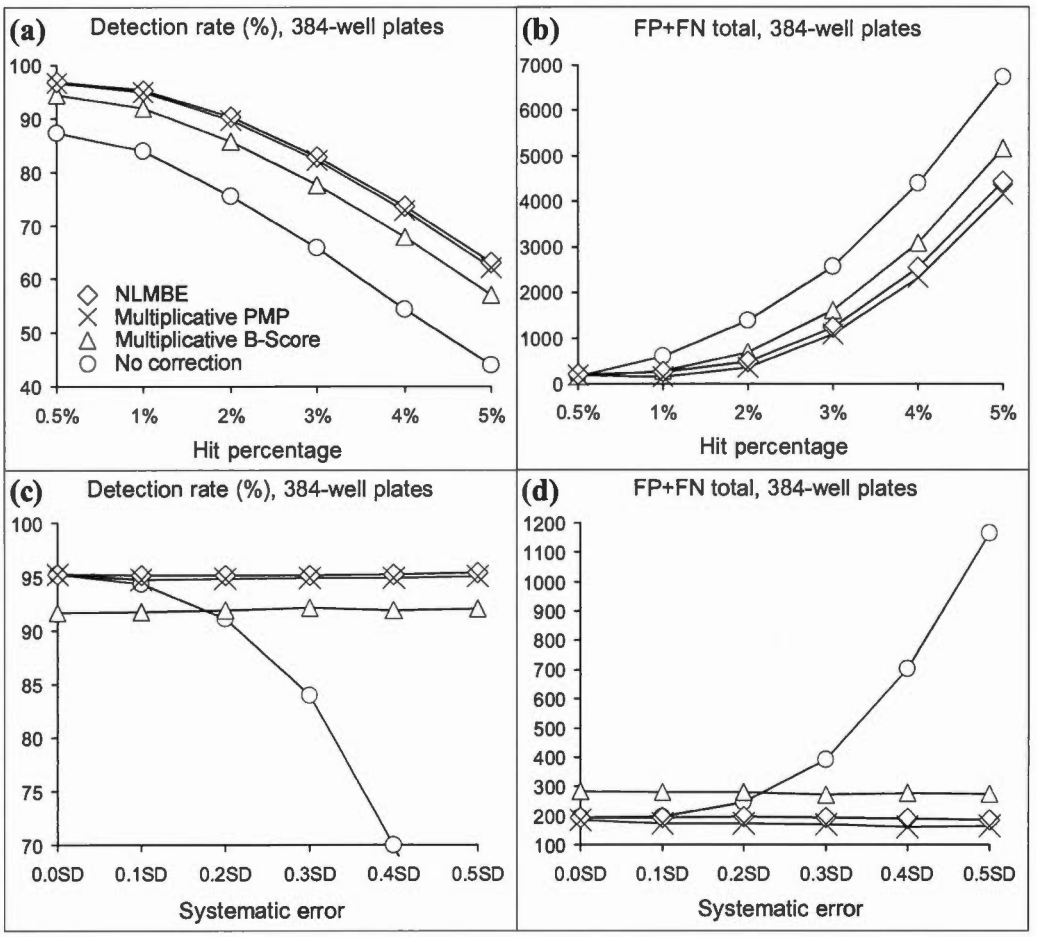


Figure 2.3 True positive rate, and combined false positive and false negative rate for 384-well plate assays. Figure 2.2 panel description applies here.

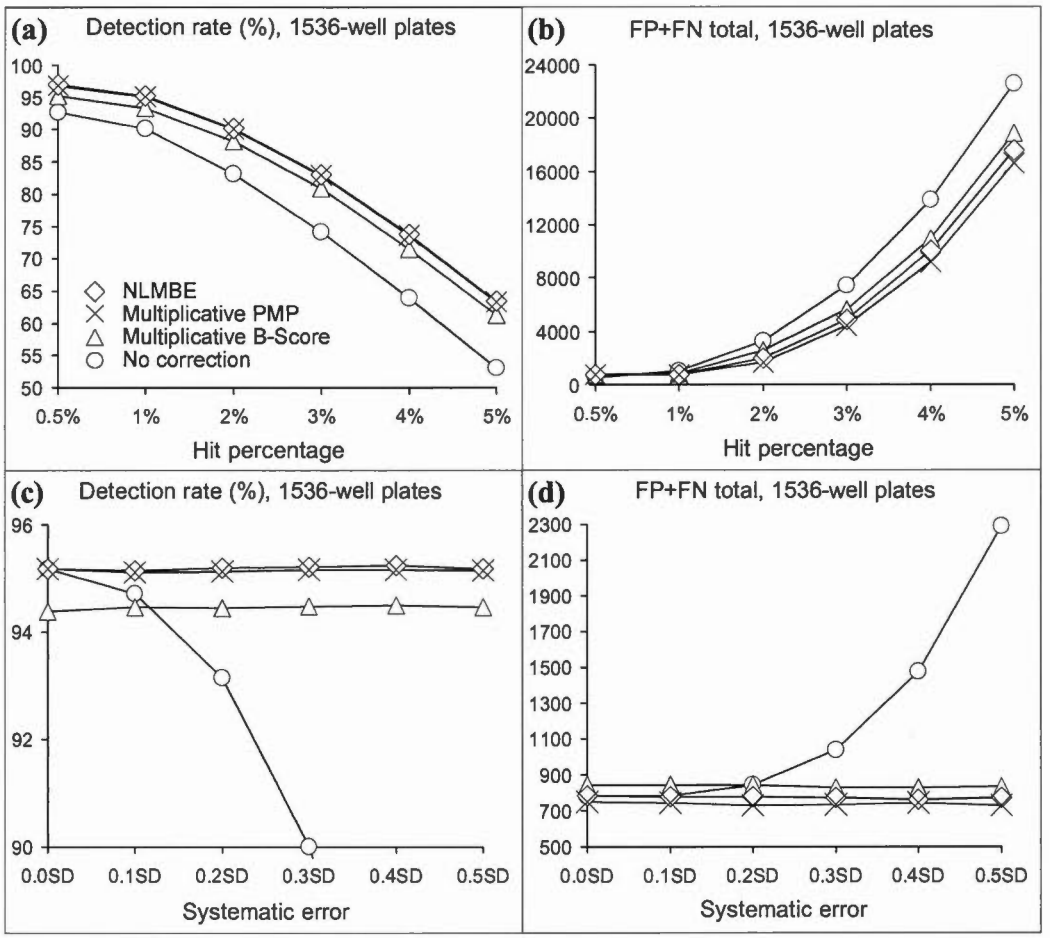


Figure 2.4 True positive rate, and combined false positive and false negative rate for 1536-well plate assays. Figure 2.2 panel description applies here.

Moreover, we also conducted simulations with higher hit rates, i.e., up to 20% (see Figure 2.5), which may occur in secondary screening. As in this case the plate's (the row's or the column's) mean can be heavily affected by outliers, and because the values of hits can be viewed as outliers, we used the median instead of the mean in all our calculations within the mPMP, NLMBE and multiplicative B-score methods when working with secondary screening data. We also used the median instead of the mean in the formula defining the hit selection threshold (i.e., *Hits values* $\leq$ *Median* -

$C \times SD$ for inhibition assays). These changes allowed our new methods, mainly mPMP and NLMBE, to achieve good correction and hit selection results, especially for large plates (Figure 2.5e and f).

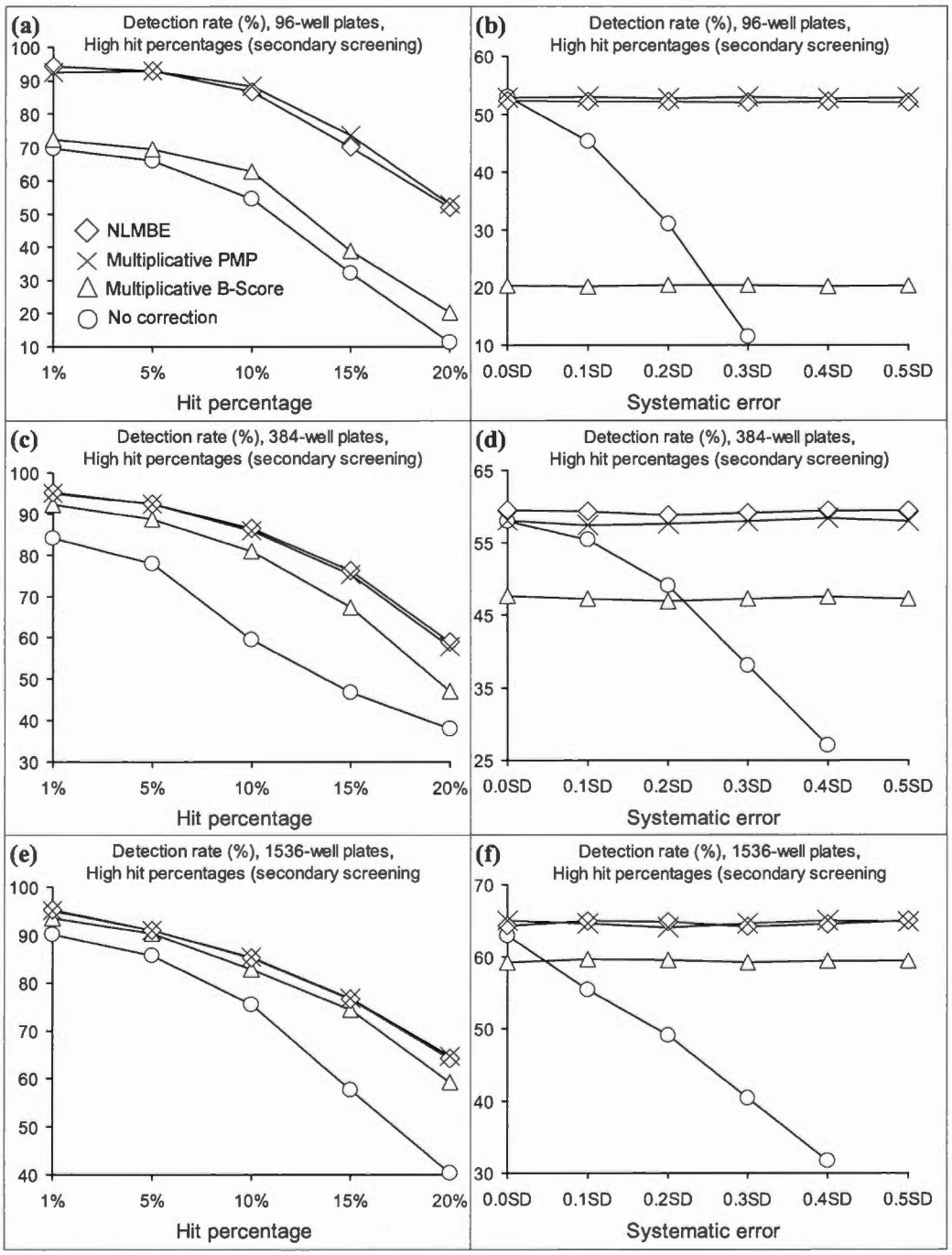


Figure 2.5 True positive rate for 96, 384 and 1536-well plate assays and high hit percentages (1% to 20%) obtained under the condition that a maximum of 4 rows and

4 columns of each plate were affected by spatial bias. Panels (a, c and e) present the results for datasets with the fixed standard deviation of spatial bias, equal to $0.3SD$. Panels (b, d and f) present the results for datasets with the fixed hit percentage rate of 20%.

We also carried out a simulation involving different plate layouts and the proposed mPMP bias correction method. Precisely, we compared hit detection results obtained for 384-well plates (a) without controls (i.e., all of the plate's wells comprised regular screening samples), (b) with the control layout corresponding to Figure 2.1a in (Mpindi et al. 2015) (i.e., layout based on placing controls in column 1 and 24), and (c) with the control layout corresponding to Figure 2.1b in (Mpindi et al. 2015) (i.e., layout based on randomly scattering controls across the entire plate). The hit rate in this simulation varied from 1 to 20%. The positions of controls (but not their values) were taken into account in this simulation during the computation of the method's parameters. Even though superior results were obtained for plates with no controls, followed by plates with the scattered control layout, and, finally, by plates with controls located in the first and last columns, the results of this simulation, presented in Supplementary Figure 2.1, suggest that the control layout has no major impact on the performance of the multiplicative PMP method. The size of the plate remains a more important factor in this context.

2.4.2 Analysis of the RNAi HIV HTS assay

We applied the introduced mPMP algorithm to correct the RNAi HIV data generated during a genome-wide siRNA screen, which was aimed at studying HIV-host interactions. This screen was used by Carralot and colleagues (Carralot et al. 2012) to validate their diffusion correction model. To identify host factors involved in the interactions with HIV, an RNAi screening of human cells infected by HIV-1 and

transfected with a genome-spanning siRNA library was carried out (for more details, see Supplementary Material 3 in Carralot et al. (2012). Carralot et al. (2012) showed that this screen was affected by multiplicative spatial bias, which was evident as edge effects. The original screen consisted of 68 plates with of size (16 × 24). Because columns 1 to 4 of all plates contained only non-target elements, their measurements were excluded from our analysis. Thus, our experiments were carried out using the 320 remaining measurements of each plate. The entire tested dataset is available at the following URL address: www.info2.uqam.ca/~makarenkov_v/HTS/downloads/RNAi_HIV_68.zip and the hit counts per well location represented in Figure 2.6a-f are available in Supplementary Tables ST1-ST6.

As the RNAi HIV screen is an inhibition assay, the hits correspond to low values of measurements. Consider Plate 7 of this assay (see Figure 2.1b or Figure 2.6g). First, we applied the Mann–Whitney U test to identify biased rows and columns present in this plate. The presence of spatial bias was detected in rows A to C, E, and M to P as well as in columns 23 and 24 of Plate 7. The Kolmogorov-Smirnov test's p-value for the additive PMP (aPMP) model was 0.0025, while the p-value for the multiplicative PMP (mPMP) model was 0.2196. The null hypothesis, H_0, here is that both unbiased raw measurements and corrected measurements come from the same distribution. Thus, the aPMP method provided strong evidence against the null hypothesis, while the mPMP method did not, at the selected significance level α (α was equal to 0.05 in our study). This result suggests that spatial bias in Plate 7 follows a multiplicative model.

Table 2.1 Number of hits found in the RNAi HIV assay using: raw data (No Correction), data corrected by diffusion model, data corrected by additive B-score, data corrected only assay-wise, data corrected only plate-wise by mPMP, data corrected assay-wise and then plate-wise by mPMP, and data corrected plate-wise by mPMP and then assay-wise. The selected thresholds, μ-1.348σ, μ-1.293σ, μ-1.255σ and μ-1.219σ, correspond to 1%, 2%, 3% and 4.13% of hits, respectively.

Number of hits found by the methods for four different hit selection thresholds	μ-1.348σ	μ-1.293σ	μ-1.255σ	μ-1.219σ
No Correction (raw data)	671	855	1036	1239
Diffusion model	677	874	1058	1270
B-score	1271	1460	1595	1719
Assay-wise correction	851	1069	1215	1399
Plate-wise correction (mPMP)	708	889	1050	1247
Assay+plate-wise correction	944	1170	1341	1517
Plate+assay-wise correction	750	920	1079	1284

Using the RNAi HIV experimental data, we compared the performances of seven data correction methods in terms of the number of hits (Table 2.1) and the data homogeneity, studied within Plate 7 (Figure 2.6g-l) and within the overall hit distribution surface representing the number of hit counts per well location (Figure 2.6a-f and Table 2.2). The seven compared methods were as follows: No Correction, diffusion model removing multiplicative plate-specific bias (Carralot et al. 2012), conventional (additive) B-score (Brideau et al. 2003) assay-wise correction by Well Correction (Makarenkov et al. 2007); this procedure removes both additive and plate-specific biases across a given well location), plate-wise correction by multiplicative PMP, plate-wise correction by multiplicative PMP followed by assay-wise correction using Z-score normalization, and assay-wise correction by Z-score followed by plate-wise correction by multiplicative PMP. Our computations were carried out for four different hit selection thresholds consisting of 1%, 2%, 3% and 4.13% of hits (the last threshold was selected following (Carralot et al. 2012)). Table 2.1 shows that the conventional B-score correction drastically overestimates the number of detected hits compared to raw data. In contrast, the diffusion model, plate-wise correction by

mPMP and the combined plate and assay-wise correction present hit totals close to that of raw data (see Table 2.1 and Figure 2.6).

The application of the diffusion model to the RNAi HIV data led to a partial correction of the multiplicative edge effect affecting both Plate 7 and the assay's hit distribution surface (Figure 2.6h and b). However, this correction was not enough to pass the χ^2 goodness of fit test for three of the four selected hit selection thresholds (Table 2.2). This test can be used in HTS to assess the deviation of the hit distribution surface from the expected (i.e., plane) surface (Makarenkov et al. 2007). The additive B-score technique removed an important part of the original edge effect at the expense of a significant increase in the number of detected hits (Table 2.1) and an inverse edge effect due to overfitting which can be observed within the corrected hit distribution surface (Figure 2.6c). Even though the B-score method was able to remove spatial bias from rows C and O and column 23 of Plate 7 (Figure 2.6i), it was by far the worst method in terms of the χ^2 goodness of fit test used to assess the homogeneity of the hit surface (Table 2.2). Assay-specific bias correction improved the uniformity of the hit count surface by removing from it the patterns of edge effect. The corrected hit surface passed the χ^2 goodness of fit test for all four hit selection thresholds.

Table 2.2 χ^2 goodness of fit statistic (given for α= 0.01) for the hit distribution surfaces of the RNAi HIV assay computed after the application of the following data correction methods: No Correction, diffusion model, additive B-score, assay-wise correction only, plate-wise correction only by mPMP, assay-wise correction followed by plate-wise correction by mPMP, and plate-wise correction by mPMP followed by assay-wise correction. The selected thresholds, μ-1.348σ, μ-1.293σ, μ-1.255σ and μ-1.219σ, correspond to 1%, 2%, 3% and 4.13% of hits, respectively.

χ^2-goodness-of-fit of hit distribution surfaces given for four hit selection thresholds	μ-1.348σ	μ-1.293σ	μ-1.255σ	μ-1.219σ
Critical value	*380.68*	*380.68*	*380.68*	*380.68*
No correction (raw data)	366.26	408.91	477.51	529.91
Diffusion model	369.97	421.64	480.30	537.12
B-score	670.90	644.99	629.35	639.77
Assay-wise correction	338.38	299.31	279.12	268.71
Plate-wise correction (mPMP)	315.28	329.45	345.81	368.40
Assay+plate-wise correction	300.07	282.85	280.95	268.21
Plate+assay-wise correction	302.16	282.09	280.18	261.67

However, an inverse edge effect pattern, similar to that introduced by B-score, can be observed on the corrected hit distribution surface (Figure 2.6d). Assay-wise correction was also unable to correct a strong edge effect present in row A of Plate 7; this edge effect was apparently much more significant within Plate 7 than within the rest of the plates of this assay. Plate-specific bias correction via multiplicative PMP better corrected Plate 7's edge effects (Figure 2.6k) , but still conserved the outlines of the original edge effect in rows O and P (Figure 2.6e). All hit count surfaces computed after the plate-wise correction were also successful in passing the χ^2 goodness of fit test (Table 2.2).

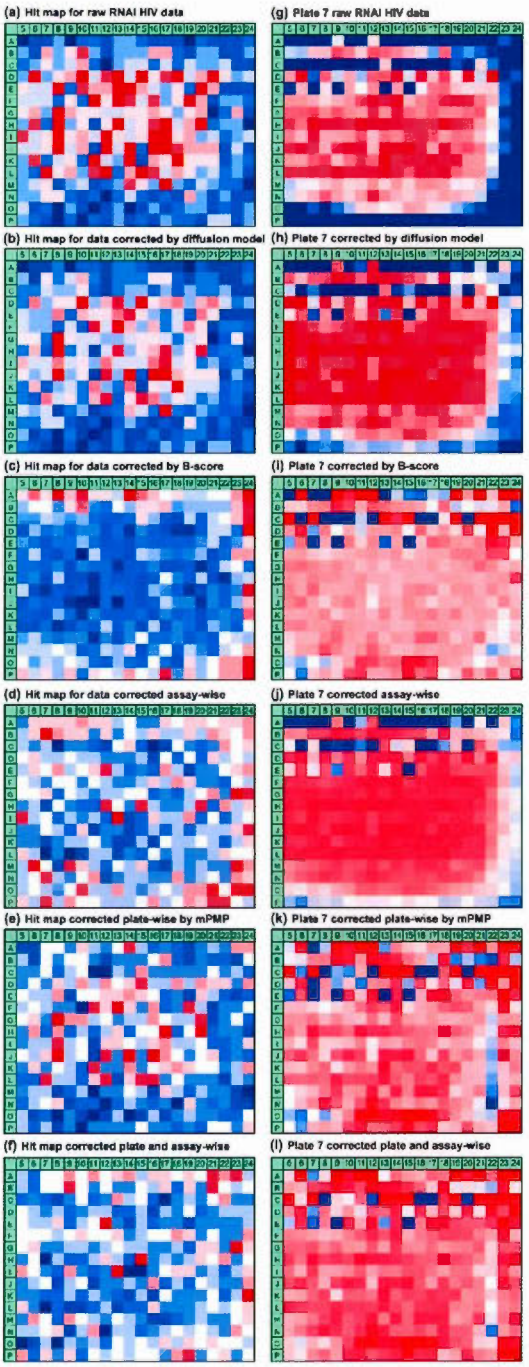


Figure 2.6 Hit distribution surfaces and Plate 7 heatmaps for the following types of RNAi HIV data: (a,g) raw data, (b,h) data corrected by the diffusion model, (c,i) data

corrected by additive B-score, (d,j) data corrected only assay-wise, (e,k) data corrected only plate-wise using mPMP, and (f,l) data corrected both plate and assay-wise. The results for the hit selection threshold of μ-1.219σ are depicted. Figure 2.1 description applies here.

As both plate and assay-wise corrections passed the χ^2 goodness of fit test for all four hit selection thresholds, they can be combined to obtain a more powerful bias correction technique. However, the order of their application is important as the results in Tables 2.1 and 2.2 suggest. While the results regarding the homogeneity of the hit distribution surface, reported in Table 2.2, give no advantage to one of the two methods, those regarding the total number of hits, reported in Table 2.1, show that plate-wise correction should precede assay-wise correction in the case of the RNAi HIV data. In fact, if the assay-wise correction precedes the plate-wise correction, a clear overestimation of the number of detected hits can be observed.

It is worth noting that the most suitable order of application of the plate and assay-wise corrections depends on the data only. For example, for a 5-plate assay presented in Supplementary Tables ST7-ST11 (see Supplementary Materials), the assay-wise correction (Supplementary Figures SF2 to SF6) should precede the plate-wise correction (Supplementary Figures SF7 to SF11) because otherwise the assay-specific bias present in well locations (A,12), (B,12), (C,12), (D,12) and (E,12) cannot be effectively recognized (see Supplementary Tables ST7 to ST11). Further validation of selected hits should be conducted through the structure-activity relationships (SAR) analysis and the subsequent clinical trials.

2.4.3 Analysis of ChemBank data

We first examined 100 experimental assays from ChemBank (Seiler et al. 2008), which is the most complete public small-molecule assay database, in order to

determine the dominant type of plate-specific spatial bias affecting the four available HTS screening categories, including HTS-homogeneous, HTS-microorganism, HTS-cell-based and HTS-gene-expression assays (25 assays per screening category were considered). Figure 2.7 shows the proportion of assays, per screening category, affected by either additive, or multiplicative, or undetermined type of spatial bias. The algorithm presented in Section 2.2.5 was used to identify the most appropriate bias model for each plate examined. An assay was declared to be affected by additive bias when it had more plates affected by additive than multiplicative or undetermined (see Step 4 of our algorithm in Section 2.2.5) type of spatial bias according to our algorithm. Similar considerations were applied for assays affected by multiplicative and undetermined types of bias. Our results demonstrate that the dominant type of plate-specific spatial bias varies across screening categories. For example, additive bias has been dominant in homogeneous (52%) and microorganism assays (64%), while multiplicative bias has been dominant in cell-based (52%) and gene expression assays (48%).

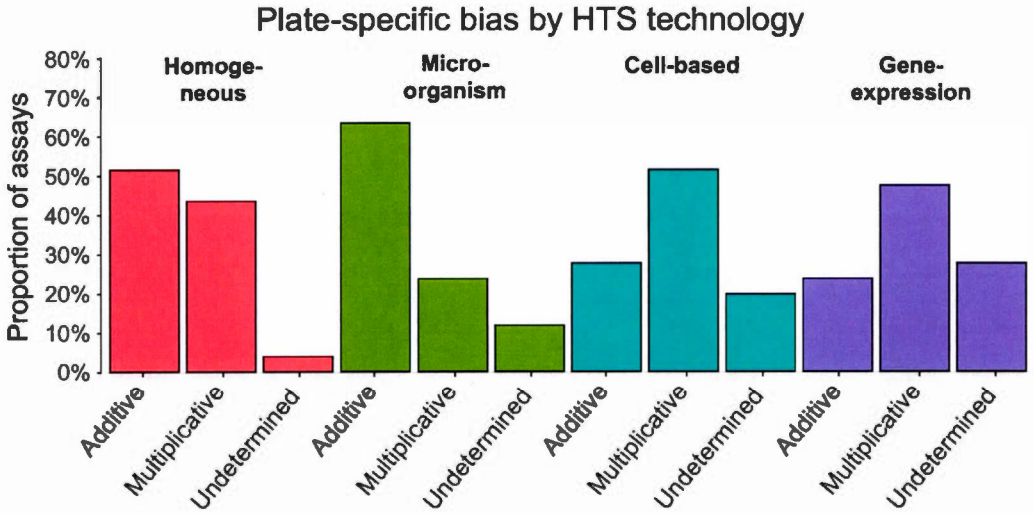


Figure 2.7 Plate-specific bias detected across the four HTS screening categories available in ChemBank. 100 HTS assays (25 per screening category) were analyzed (see Supplementary Table ST12 for the ChemBank IDs of the assays). The control wells were removed from all screens prior to bias detection. The proportion of assays per screening category, affected by additive bias, by multiplicative bias, and by an undetermined type of bias is reported. No assays containing no biased plates at all were found in this experiment.

We also investigated the presence of spatial bias in a comprehensive gene expression dataset (L1000) produced by the Library of Integrated Network-based Cellular Signatures (LINCS) (Duan et al. 2014; Lachmann et al. 2016). The considered L1000 mRNA profiling assay contains gene expression measurements for tens of thousands of distinct cell perturbations. This is the largest 384-well gene expression measurement assay ever performed (Lachmann et al. 2016). We examined the measurements from the Normalized L1000 dataset (available from the LINCS consortium at: http://lincsportal.ccs.miami.edu/datasets/#/view/LDS-1233). The dataset we analyzed included 7 816 microplates. It contained all the 15 available cell lines (A-375, A549, BT-20, HA1E, HCC515, Hep G2, HME1, Hs 578T, HT-29,

LNCaP, MCF10A, MCF7, MDA-MB-231, PC-3 and SK-BR-3). In total, our analysis was carried out on 3 001 344 different gene expression profiles. The Normalized L1000 dataset contains direct measurements for 978 genes (landmark genes), while for 22 000 more genes only the transcript amounts are available (Duan et al. 2014). Our analysis concerned these 978 landmark genes. Lachmann et al. (2016) have recently investigated the gene expression profiles from the L1000 dataset and found that the vast majority of the tested plates were affected by spatial bias. In our analysis, we found that spatial bias affected 6 957 out of 7 816 (89.01%) of the tested microplates from the Normalized L1000 dataset, compared to 96.36% biased microplates found by Lachmann et al. (2016) using an algorithm which combines spatial autocorrelation detection and principal component analysis. Precisely, we established that 25.67% of the assay plates were affected by additive bias, 44.65% by multiplicative bias, 18.69% by an undetermined type of bias and 10.99% contained no bias (Figure 2.8).

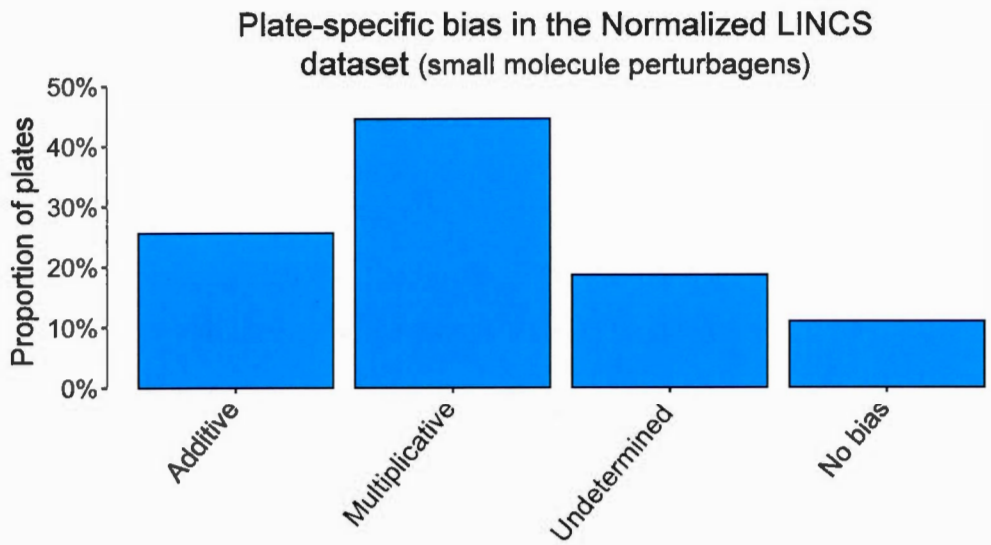


Figure 2.8 Plate-specific bias detected across 7 816 plates, including 3 001 344 gene expression profiles, from the Normalized L1000 mRNA profiling assay. The control wells were removed from all screens prior to bias detection. The proportion of plates affected by additive bias, by multiplicative bias, by an undetermined type of bias, as well as of those having no spatial bias at all, is reported. When the Mann-Whitney *U* test detected no any biased row or column in a given plate, the plate was reported as containing no spatial bias.

2.5 Discussion

In this paper, we described three novel methods, called Non-Linear Multiplicative Bias Elimination (NLMBE), multiplicative Partial Median Polish (mPMP) and multiplicative B-score, for removing multiplicative spatial bias from experimental screening data. The performances of the new methods were assessed in simulations, which confirmed that both NLMBE and mPMP outperformed the multiplicative B-score technique, which was prone to generating false positive hits. The NLMBE method yielded slightly better performances than mPMP in terms of the true positive

rate, while mPMP was better than NLMBE in terms of the combined false positive and false negative rate. Taking into account that mPMP converges much faster than NLMBE, the former is recommended for correcting multiplicative spatial bias in HTS assays. The proposed NLMBE and mPMP methods correct only the measurements of rows and columns of a given plate in which spatial bias was detected by the Mann-Whitney *U* test. This is the main advantage of these methods, compared to B-score (additive or multiplicative) and other data correction techniques that modify all the measurements of a given plate even though spatial bias is present in only a few of them. This property of the new methods allows us to address efficiently the over-fitting issue.

Moreover, we presented a general bias correction protocol, which can be used by HTS researchers to remove both assay and plate-specific spatial biases. The plate-specific part of this protocol includes a new algorithm, based on the use of the additive (Dragiev et al. 2012) and multiplicative PMP methods and the Kolmogorov-Smirnov two-sample test to identify the most appropriate (i.e., additive or multiplicative) spatial bias model for a given plate. We also propose to carry out the Mann–Whitney *U* test to detect the presence of both assay and plate-specific spatial biases. Importantly, the presented bias correction methods should be used cautiously since the application of error correction techniques on error-free data can introduce an additional bias that negatively affects the hit selection process (see for example the results of the multiplicative B-score method on error-free data in Figures 2.2-2.5). Thus, the application of spatial bias correction methods should be supported by statistical tests. Finally, we showed that the discussed methods for removing multiplicative spatial bias and the introduced general data correction protocol are effective in detecting and cleaning experimental data generated by screening technologies. For example, after analyzing the ChemBank data, we were able to determine that the additive type of spatial bias is dominant in homogeneous and microorganism HTS screens, while cell-based and gene-expression HTS assays are

mostly affected by multiplicative spatial bias. Clearly, the screening category has a direct impact on the nature of spatial bias (additive vs. multiplicative; see Figures. 2.7-2.8). In the future, it would be interesting to conduct some additional experiments in order to establish whether the type of spatial bias also depends on some technical and environmental factors which can affect experimental screening campaigns, such as reader and pipette malfunctioning, unintended variations in compound concentration associated with agent evaporation, or temperature, lighting and air flow fluctuations.

Mpindi et al. (2015) showed the importance of QC metrics (e.g., Z'-factor, (Zhang et al. 1999)) and per-plate data visualization for identifying systematic errors in experimental HTS, especially for data with a high hit rate. It would interesting to compare in the future the performances of Z'-factor and the χ^2 goodness of fit test, which was used in this paper to assess the deviation of the hit distribution surface from a plane surface. The advantage of the χ^2 goodness of fit test is that it can be carried out when the control measurement information is unavailable for data at hand (as was the case of the RNAi HIV assay analyzed in Section 2.2), but Z'-factor can provide a better indication of the presence of spatial bias for secondary screens involving high hit rates. The presented methods and data correction protocol have been implemented in the *AssayCorrector* package (AssayCorrector), which is freely available at CRAN.

2.6 Supplementary material

Table ST1 Hit distribution for the *raw RNAi HIV dataset* computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6a of the main manuscript. Columns 1 to 4 were discarded because they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	0	1	1	3	0	2	2	1	3	3	2	1	1	1	0	2	1	1	1	2
B	3	1	6	3	5	7	3	1	3	3	3	6	6	2	1	3	3	1	1	3
C	2	3	5	3	4	6	2	0	3	4	4	2	4	2	4	3	3	4	1	1
D	3	8	3	5	6	10	4	7	9	4	10	5	7	8	4	6	8	5	3	4
E	5	4	4	8	2	7	6	6	8	8	6	4	3	5	2	8	6	1	1	1
F	4	4	4	2	5	2	9	5	10	7	4	7	3	5	8	6	1	3	3	2
G	5	6	3	13	6	5	5	4	4	5	5	6	5	3	5	6	0	2	6	1
H	3	7	2	10	3	5	1	5	9	4	5	7	9	4	6	9	2	3	3	0
I	7	4	5	8	5	6	4	8	5	4	5	8	9	5	6	3	4	2	3	4
J	4	4	5	9	4	6	1	8	4	12	2	10	10	7	6	6	6	1	4	1
K	3	7	2	4	5	2	8	4	3	8	1	6	5	10	5	5	5	1	3	2
L	3	2	8	5	6	3	7	9	2	6	9	7	4	6	5	3	4	2	3	1
M	3	4	5	1	3	4	1	2	3	7	5	3	10	3	2	4	3	2	2	0
N	3	3	5	5	2	1	2	0	2	4	2	1	2	1	6	3	1	3	1	1
O	4	1	2	0	3	1	0	1	3	0	1	3	1	2	3	3	1	2	1	3
P	3	2	1	0	2	0	2	1	3	1	2	1	3	2	2	1	4	4	1	3

Table ST2 Hit distribution for the RNAi HIV dataset *corrected by the diffusion model* and computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6b of the main manuscript. Columns 1 to 4 were discarded because they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	0	1	1	3	0	2	2	1	3	3	2	1	1	1	0	2	1	1	1	2
B	4	1	6	3	5	7	3	1	3	3	3	6	6	2	1	3	3	1	1	3
C	3	3	5	4	4	6	2	1	3	4	4	2	4	2	4	3	3	4	1	1
D	3	8	4	6	6	10	4	7	9	4	10	6	7	8	4	7	8	5	3	4
E	5	4	4	8	2	7	6	6	8	9	6	4	3	5	2	8	6	1	1	1
F	5	4	4	2	5	2	9	5	11	7	4	7	3	5	8	6	1	3	3	2
G	6	6	3	14	6	5	5	4	4	5	6	6	5	3	5	6	0	2	6	1
H	3	7	2	10	3	5	1	5	9	4	6	7	9	4	6	9	2	3	3	0
I	7	5	5	8	5	6	4	8	5	5	5	8	9	5	6	3	4	2	3	4
J	6	4	5	11	4	6	1	9	4	12	2	10	10	7	6	6	6	1	4	1
K	3	7	2	4	5	3	8	4	3	8	1	6	5	10	5	5	5	2	3	2
L	3	2	8	5	6	3	7	9	2	6	9	8	4	6	5	3	4	2	3	1
M	5	4	5	1	3	4	1	2	3	7	5	3	10	3	2	4	3	2	2	0
N	3	3	5	5	2	1	2	0	2	4	2	1	2	1	6	3	1	3	1	1
O	4	1	2	0	3	1	0	1	3	1	1	3	1	2	3	3	1	2	1	3
P	3	2	1	0	2	0	2	1	3	1	2	1	4	2	4	1	4	4	1	3

Table ST3 Hit distribution for the RNAi HIV dataset corrected by *B-score* and computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6c of the main manuscript. Columns 1 to 4 were discarded because they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	9	12	9	14	10	13	9	9	7	16	12	10	8	9	7	7	8	8	11	26
B	6	8	8	5	9	12	5	4	8	6	9	9	10	5	6	6	6	6	10	13
C	5	5	7	6	5	6	4	3	3	4	9	6	3	2	7	3	7	8	10	13
D	2	6	4	5	3	5	5	2	4	1	5	3	2	1	1	3	7	7	4	15
E	5	4	2	4	4	6	4	2	3	2	4	3	3	3	3	5	7	3	5	9
F	4	7	3	2	4	0	7	3	3	2	7	6	5	4	5	6	2	4	11	11
G	4	4	3	5	2	5	4	4	1	3	3	5	2	2	3	1	3	7	8	4
H	4	5	2	7	2	2	1	4	3	3	2	3	8	5	3	7	2	1	6	7
I	7	2	1	3	2	3	1	4	4	3	2	6	5	3	1	2	5	3	5	14
J	2	4	3	3	1	0	0	4	0	3	2	3	7	5	5	1	5	3	6	10
K	3	3	2	6	4	1	5	3	3	3	4	2	5	3	4	2	4	5	4	12
L	2	2	4	1	2	4	3	7	3	5	2	7	2	5	4	2	6	5	3	8
M	5	7	5	4	5	5	1	5	2	5	8	4	4	5	3	4	3	8	7	6
N	7	7	7	8	7	3	4	7	6	9	2	5	8	3	7	5	3	7	8	11
O	8	5	7	6	8	7	4	3	6	4	3	8	4	2	9	6	8	11	11	14
P	10	7	9	6	10	4	9	8	7	11	8	7	13	5	10	6	8	11	11	14

Table ST4 Hit distribution for the RNAi HIV dataset *corrected assay-wise* and computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6d of the main manuscript. Columns 1 to 4 were discarded because they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	3	6	6	6	6	7	6	6	4	5	3	3	6	5	6	6	5	6	6	7
B	4	6	9	7	6	6	3	6	3	5	4	6	4	4	6	4	7	6	5	7
C	4	3	4	0	3	3	2	1	5	6	2	0	1	5	3	2	3	7	5	6
D	5	6	4	4	5	4	4	3	3	3	4	1	5	2	5	2	6	4	5	4
E	4	6	5	7	4	5	6	8	4	2	3	2	7	4	1	3	3	5	6	7
F	5	4	3	5	5	3	5	4	3	3	4	6	4	7	6	3	3	2	5	4
G	5	5	6	2	5	4	3	4	4	5	5	2	3	2	4	5	7	3	9	8
H	3	4	1	5	3	2	2	8	5	3	8	4	3	3	4	8	2	7	4	6
I	8	4	5	2	4	3	4	2	10	1	2	6	4	4	7	2	5	2	4	5
J	3	3	5	4	2	6	1	3	3	1	2	4	2	4	2	4	4	4	6	8
K	3	3	1	3	4	4	5	3	1	1	4	1	5	4	4	5	4	2	4	6
L	6	4	0	3	0	1	4	4	1	3	2	5	3	4	3	3	4	6	5	5
M	4	9	5	1	8	6	4	2	5	4	6	2	1	4	3	4	3	3	9	7
N	4	5	8	4	3	5	7	1	3	4	4	3	3	7	3	4	5	3	7	6
O	4	5	4	3	5	8	4	4	7	4	3	4	5	4	7	7	9	5	9	9
P	6	4	8	4	7	5	5	4	4	4	5	7	4	5	7	6	8	8	5	4

Table ST5 Hit distribution for the RNAi HIV dataset *corrected plate-wise by multiplicative PMP* and computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6e of the main manuscript. Columns 1 to 4 were discarded because they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	1	1	5	5	3	6	4	3	5	8	4	2	3	2	1	3	4	3	7	7
B	4	1	5	3	5	5	3	1	3	3	3	5	7	2	1	3	2	2	1	5
C	3	3	5	2	4	5	0	0	4	3	7	3	3	3	3	3	5	3	4	2
D	3	5	3	6	4	7	4	7	5	6	8	4	4	7	4	7	6	6	3	7
E	5	4	4	7	1	5	5	6	6	7	6	5	2	3	1	6	4	2	1	2
F	4	5	4	3	3	0	8	6	9	6	3	7	2	3	8	4	1	3	3	5
G	5	6	3	10	7	7	4	5	3	4	4	7	5	4	4	8	1	3	5	2
H	3	7	2	7	2	4	2	4	6	2	5	5	11	2	5	9	2	3	3	1
I	3	4	4	8	4	5	1	6	6	5	4	6	7	4	4	5	2	2	3	6
J	5	4	6	9	4	5	1	8	3	9	3	10	7	5	3	5	6	2	3	2
K	3	6	3	5	4	2	8	3	3	7	3	5	6	7	6	5	5	3	4	7
L	2	2	8	4	4	2	6	8	1	6	8	8	3	5	3	3	3	1	4	3
M	3	4	4	1	2	4	0	3	3	5	4	1	6	4	2	4	3	3	1	0
N	2	4	5	6	1	2	2	0	2	4	3	2	3	0	6	3	2	3	1	1
O	4	1	2	0	3	1	0	1	4	0	1	4	1	3	4	3	3	2	2	5
P	4	2	4	1	3	0	3	3	3	3	5	3	3	2	3	2	5	5	3	6

Table ST6 Hit distribution of the RNAi HIV dataset *corrected plate-wise by multiplicative PMP and then assay-wise* and computed for the μ-1.219*SD* (4.13%) threshold. This hit distribution surface is presented in Figure 2.6f of the main manuscript. Columns 1 to 4 were discarded since they did not contain target compounds.

	5	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	4	2	5	5	7	5	7	5	6	7	4	6	6	5	6	5	8	7	8	4
B	4	2	5	5	8	5	3	5	3	3	5	4	6	2	6	3	8	5	4	5
C	2	3	2	1	3	5	2	3	7	4	5	2	0	3	4	3	5	7	3	3
D	5	4	3	4	2	3	4	4	2	2	4	1	4	3	5	2	3	4	4	4
E	5	4	4	6	2	4	5	9	3	1	4	2	3	3	1	3	3	3	3	4
F	5	5	5	4	5	6	5	5	6	3	4	5	5	7	6	3	5	3	3	7
G	5	6	3	3	4	4	3	4	5	4	4	3	3	4	2	6	6	3	9	6
H	3	5	0	4	2	2	4	6	4	2	9	3	4	3	4	6	5	7	3	5
I	7	5	3	2	7	3	2	3	9	0	0	3	5	6	6	3	4	2	4	6
J	3	4	5	1	4	5	1	3	3	4	3	7	2	4	2	5	4	3	3	7
K	3	3	2	2	4	2	4	3	2	1	4	1	5	3	3	3	5	2	4	10
L	6	3	0	2	1	1	2	4	1	2	2	6	3	3	3	4	5	4	5	6
M	3	7	5	2	7	5	4	3	5	4	3	1	1	2	3	4	3	3	5	6
N	4	4	5	6	1	4	7	0	4	4	2	3	4	5	4	5	4	3	7	5
O	0	1	4	0	4	3	4	2	5	4	5	4	6	4	6	5	6	5	5	4
P	7	2	6	3	6	2	5	4	5	4	5	5	3	7	8	3	5	6	3	5

Supplementary tables ST7-ST11 contain synthetic data discussed in Section 2.2 of the main manuscript. This is a 5-plate assay used to illustrate the case where the assay-wise correction should precede the plate-wise correction because otherwise the assay-specific bias present in well locations (A,12), (B,12), (C,12), (D,12) and (E,12) cannot be effectively recognized (see also Supplementary figures SF2 to SF11).

Table ST7 Plate 1 raw measurements of the synthetic dataset presented in Section 2.2 of the main manuscript.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20.010	20.096	20.016	20.061	20.012	20.067	20.053	20.060	20.053	20.099	20.074	40.054
B	20.025	20.018	20.064	20.047	20.015	20.045	20.089	20.073	20.042	20.036	20.096	40.086
C	20.001	20.024	20.093	20.082	20.047	20.034	20.026	20.030	20.075	20.005	20.034	40.035
D	20.070	20.070	20.055	20.015	20.089	20.019	20.026	20.008	20.094	20.009	20.026	40.090
E	20.075	20.059	20.008	20.060	20.071	20.017	20.016	20.041	20.025	20.046	20.002	40.003
F	20.065	20.086	20.024	20.033	20.097	20.057	20.097	20.058	20.087	20.029	20.099	20.084
G	20.070	20.006	20.085	20.060	20.046	20.096	20.018	20.006	20.044	20.015	20.023	20.002
H	20.076	20.046	20.039	20.049	20.031	20.062	20.002	20.044	20.095	20.064	20.090	20.069

Table ST8 Plate 2 raw measurements of the synthetic dataset presented in Section 2.2 of the main manuscript.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20.011	20.072	20.090	20.058	20.074	20.006	20.030	20.002	20.046	20.019	20.081	40.094
B	20.068	20.033	20.005	20.095	20.023	20.047	20.008	20.088	20.068	20.087	20.005	40.001
C	20.029	20.037	20.028	20.030	20.070	20.093	20.074	20.005	20.086	20.004	20.059	40.035
D	20.035	20.038	20.066	20.071	20.034	20.082	20.074	20.032	20.013	20.019	20.075	40.010
E	20.016	20.004	20.036	20.067	20.054	20.029	20.049	20.010	20.086	20.055	20.089	40.000
F	20.049	20.022	20.018	20.085	20.020	20.003	20.009	20.037	20.003	20.033	20.087	20.023
G	20.001	20.064	20.063	20.041	20.037	20.075	20.019	20.055	20.012	20.083	20.054	20.013
H	20.018	20.070	20.010	20.009	20.031	20.007	20.064	20.054	20.035	20.050	20.017	20.083

Table ST9 Plate 3 raw measurements of the synthetic dataset presented in Section 2.2 of the main manuscript.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20.060	20.019	20.052	20.030	20.096	20.065	20.056	20.065	20.088	20.058	20.003	40.013
B	20.023	20.035	20.008	20.081	20.020	20.013	20.023	20.029	20.042	20.017	20.099	40.028
C	20.091	20.026	20.024	20.053	20.090	20.000	20.097	20.041	20.088	20.071	20.008	40.061
D	20.092	20.023	20.073	20.088	20.058	20.079	20.046	20.092	20.097	20.041	20.088	40.066
E	20.064	20.046	20.046	20.010	20.073	20.038	20.055	20.028	20.036	20.031	20.080	40.077
F	20.070	20.069	20.016	20.063	20.013	20.025	20.005	20.080	20.059	20.094	20.093	20.024
G	20.066	20.059	20.055	20.058	20.031	20.020	20.000	20.014	20.075	20.018	20.039	20.031
H	20.013	20.055	20.027	20.041	20.073	20.075	20.073	20.001	20.001	20.050	20.032	20.039

Table ST10 Plate 4 raw measurements of the synthetic dataset presented in Section 2.2 of the main manuscript.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20.003	20.055	20.066	20.059	20.005	20.047	20.026	20.099	20.021	20.004	20.041	40.008
B	20.086	20.037	20.009	20.072	20.011	20.003	20.054	20.098	20.044	20.018	20.036	40.065
C	20.084	20.069	20.061	20.042	20.057	20.067	20.094	20.065	20.044	20.041	20.066	40.036
D	20.088	20.076	20.013	20.013	20.046	20.016	20.046	20.050	20.086	20.035	20.016	40.002
E	20.055	20.098	20.092	20.099	20.027	20.050	20.094	20.020	20.065	20.009	20.049	40.002
F	20.012	20.048	20.045	20.006	20.067	20.091	20.039	20.043	20.040	20.053	20.092	20.029
G	20.017	20.015	20.015	20.016	20.014	20.060	20.067	20.048	20.018	20.093	20.040	20.072
H	20.046	20.057	20.074	20.012	20.071	20.088	20.021	20.028	20.037	20.021	20.017	20.002

Table ST11 Plate 5 raw measurements of the synthetic dataset presented in Section 2.2 of the main manuscript.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20.033	20.045	20.007	20.042	20.093	20.092	20.062	20.026	20.001	20.028	20.080	40.037
B	20.092	20.060	20.039	20.076	20.003	20.018	20.034	20.079	20.026	20.088	20.055	40.010
C	20.072	20.000	20.012	20.045	20.042	20.065	20.040	20.090	20.016	20.095	20.035	40.073
D	20.049	20.055	20.063	20.002	20.057	20.091	20.068	20.096	20.005	20.020	20.035	40.078
E	20.087	20.099	20.084	20.061	20.059	20.089	20.006	20.051	20.064	20.073	20.096	40.013
F	20.025	20.032	20.018	20.043	20.075	20.077	20.090	20.095	20.032	20.005	20.001	20.064
G	20.004	20.015	20.094	20.034	20.021	20.071	20.001	20.099	20.059	20.070	20.031	20.067
H	20.004	20.058	20.076	20.014	20.045	20.088	20.066	20.060	20.002	20.028	20.089	20.064

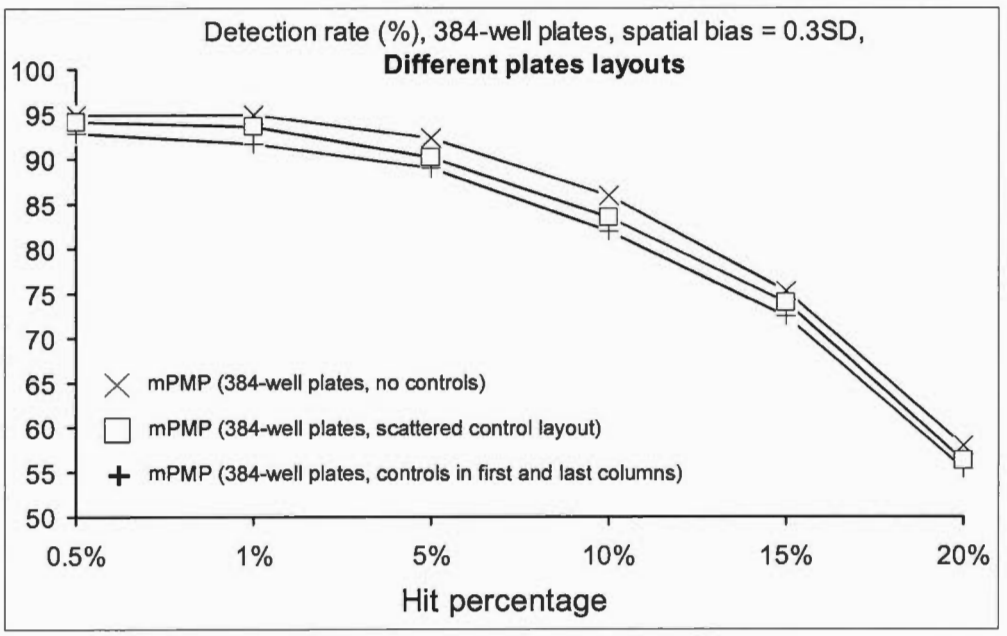
Table ST12 Set of 100 ChemBank assays analyzed in our bias detection simulation described in Section 2.3 of the main manuscript (see also Figure 2.7 in the main text); 25 assays per HTS screening category (Cell-based, Homogeneous, Microorganism and Gene expression) were considered.

High-throughput screening (Cell-based) – 25 assays	
1	AdipocyteDifferentiation1_OilRedO(913.0191)
2	AdipocyteDifferentiation2_NileRed(1015.0001)
3	AdipocyteDifferentiation2_NileRed(1015.0032)
4	AdipocyteDifferentiation2_NileRed(1015.0034)
5	AnnotationDevelopment_BrdUCytoblot(900.0001)
6	AnnotationDevelopment_BrdUCytoblot(900.0002)
7	AnnotationDevelopment_BrdUCytoblot(900.0021)
8	AnnotationDevelopment_BrdUCytoblot(900.0022)
9	AnnotationDevelopment_EthD1Staining(900.0005)
10	AnnotationDevelopment_JC1MitoDye(900.0013)
11	AR-NcoRBindingAssay_raw{Lux()}(268.0159)
12	AR-NcoRBindingAssay_raw{Lux()}(268.0173)
13	AR-NcoRBindingAssay_raw{Lux()}(268.0217)
14	AR-NcoRBindingAssay_user{AvgLux()}(268.0221)
15	BreastCancerCellProfiling_CellTiterGlo(915.0248)
16	BreastCancerCellProfiling_JC1MitoDye(915.0244)
17	CellularAutofluorescence_CpdAutofluor(908.0049)
18	CellularAutofluorescence_CpdAutofluor(908.0125)
19	CellViabilityProfiling_CellTiterGlo(1019.0001)
20	DeacetylaseInhibition_AcLysCytoblot(1027.0002)
21	EndothelialCellProfiling1_Calcein-AM(910.0153)
22	FacioscapulohumeralMD_Calc(E1-E2)(1026.0010)
23	FacioscapulohumeralMD_Calc(E1-E2)(1026.0011)
24	FacioscapulohumeralMD_LuxReporter(1026.0003)
25	FacioscapulohumeralMD_LuxReporter(1026.0019)
High-throughput screening (Homogeneous) – 25 assays	
1	ActinPolymerization_raw{FI()}(144.0030)
2	ActinPolymerization_user{Fold()}(144.0031)
3	AdipocyteDifferentiation1_OilRedO(913.0190)
4	AnnotationDevelopment_BrdUCytoblot(900.0020)
5	BRAF_HRPCytoblot(1110.0001)
6	BRAF_HRPCytoblot(1110.0002)
7	BRAF_HRPCytoblot(1110.0003)
8	BRAF_HRPCytoblot(1110.0005)
9	CellularAutofluorescence_CpdAutofluor(908.0050)
10	CMVPolymeraseBindingAssay_raw{Pol(P)}(299.0552)
11	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0543)
12	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0549)
13	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0553)

14	CREBReporterAssay_LacZReporter(1029.0010)
15	CyclinReporterGeneCdh1_user{Fold()}(219.0091)
16	DihydroorotateDehydrogenase_Calc(E1-E2)(1021.0033)
17	DihydroorotateDehydrogenase_EnzCoupledColor(1021.0001)
18	DihydroorotateDehydrogenase_EnzCoupledColor(1021.0013)
19	EColiFilamentation2006_OpticalDensity(1038.0010)
20	GlycanaseActivity_raw{Pol(P)}(295.0495)
21	HoxDNA-BindingAssay_FluorOligo(1031.0002)
22	HoxDNA-BindingAssay_FluorOligo(1031.0008)
23	HoxDNA-BindingAssay_FluorOligo(1031.0010)
24	KinaseInhibitorModifiers_BrdUCytoblot(901.0010)
25	TrypanothioneReductase_EnzCoupledColor(1017.0020)
High-throughput screening (Microorganism) – 25 assays	
1	ABAggregationInhibitors_OpticalDensity(1103.0009)
2	AntibacterialAssay_FluorProtein(1106.0016)
3	AntibacterialAssay_FluorProtein(1106.0027)
4	AspulvinoneUpregulation_MetabColor(1022.0007)
5	BiofilmFormationAssay_BacTiterGlo(1059.0006)
6	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0609)
7	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0613)
8	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0625)
9	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0657)
10	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0737)
11	ClathrinDependentMembraneTrafficking_user{Fold(Mut)}(310.0733)
12	EColiFilamentation2006_OpticalDensity(1038.0001)
13	EColiFilamentation2006_OpticalDensity(1038.0002)
14	EColiFilamentation2006_OpticalDensity(1038.0012)
15	EColiFilamentation2006_OpticalDensity(1038.0014)
16	EColiFilamentation2006_OpticalDensity(1038.0016)
17	EColiFilamentation2006_OpticalDensity(1038.0022)
18	EColiFilamentation2006_OpticalDensity(1038.0023)
19	EColiFilamentation2006_OpticalDensity(1038.0024)
20	PDERegulators_OpticalDensity(1091.0043)
21	PhosphatidylinositolKinase_OpticalDensity(1000.0008)
22	PhosphatidylinositolKinase_OpticalDensity(1000.0027)
23	PSACAntagonistScreen_OpticalDensity(1035.0016)
24	SulfurAssimilation_user{Inh(BioB)}(130.0018)
25	SulfurAssimilation_user{Inh(CysH)}(130.0020)
High-throughput screening (Gene expression) – 25 assays	
1	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0001)
2	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0003)
3	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0007)
4	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0009)
5	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0010)
6	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0012)
7	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0014)
8	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0016)

9	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0017)
10	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0019)
11	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0021)
12	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0028)
13	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0030)
14	GE-HTSApoptosis_GeneExprHTS(1055.0005)
15	GE-HTSApoptosis_GeneExprHTS(1055.0015)
16	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0083)
17	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0089)
18	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0092)
19	MetabolismCellProfiling_GeneExprHTS(1020.0071)
20	MetabolismCellProfiling_GeneExprHTS(1020.0078)
21	NeuroblastomaDifferentiation_GeneExprHTS(1149.0007)
22	NeuroblastomaDifferentiation_GeneExprHTS(1149.0016)
23	NeuroblastomaDifferentiation_GeneExprHTS(1149.0039)
24	NeuroblastomaDifferentiation_GeneExprHTS(1149.0046)
25	NeuroblastomaDifferentiation_GeneExprHTS(1149.0047)

Figure SF1 presents simulation results for 384-well plates with different control layouts (see the main text for more detail). Spatial bias was set to 0.3SD. The following plate layouts: (a) (16x24)-well plate with no controls, (b) scattered control layout (with 22 control wells - as shown in Figure 2.1b in (Mpindi et al. 2015) and (c) layout with 32 control wells, located in first and last columns of the plate see (Mpindi et al. 2015), were compared.



Supplementary figures SF2-SF11 present the corrected plate measurement maps for synthetic data discussed in Section 2.2 of the main manuscript (Supplementary Tables ST7-ST11 contain the raw data for this synthetic 5-plate assay). Supplementary figures SF2-SF6 present the plate maps for the data corrected first assay-wise, and then plate-wise (using aPMP). Supplementary figures SF7-SF11 present the plate maps for the data corrected first plate-wise (using aPMP), and then assay-wise.

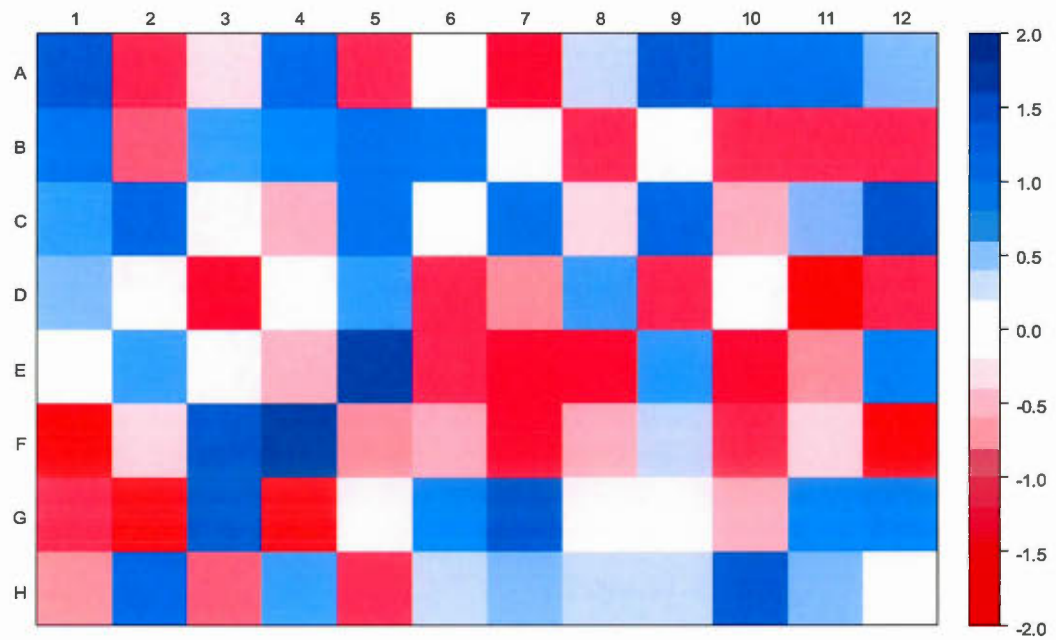


Figure SF2 Plate 1 data corrected first assay-wise, and then plate-wise (using aPMP).

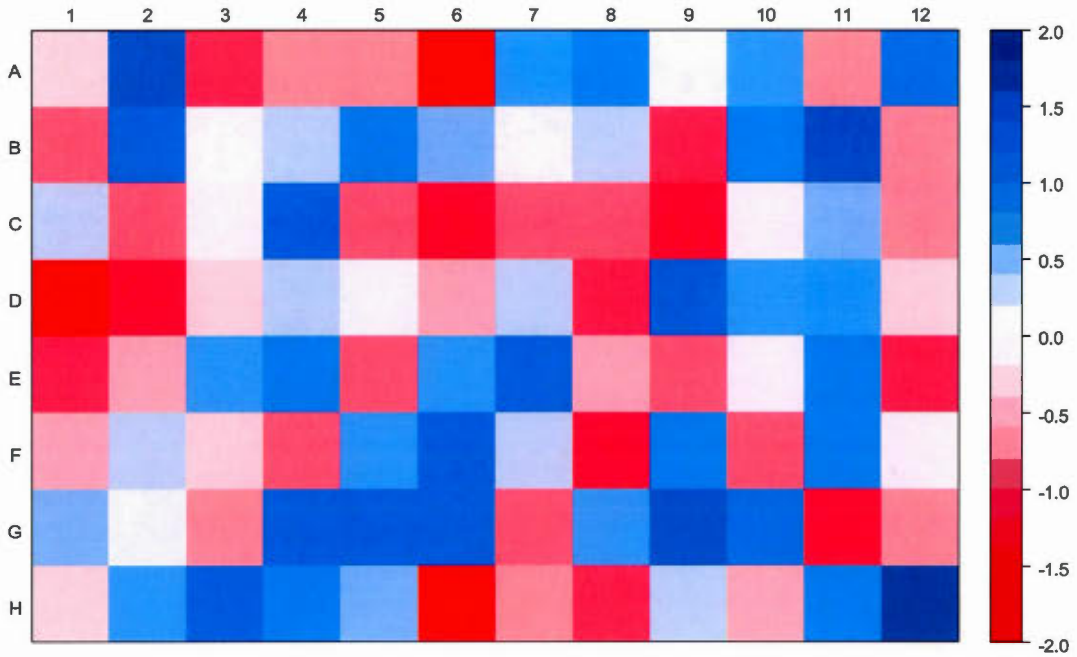


Figure SF3 Plate 2 data corrected first assay-wise, and then plate-wise (using aPMP).

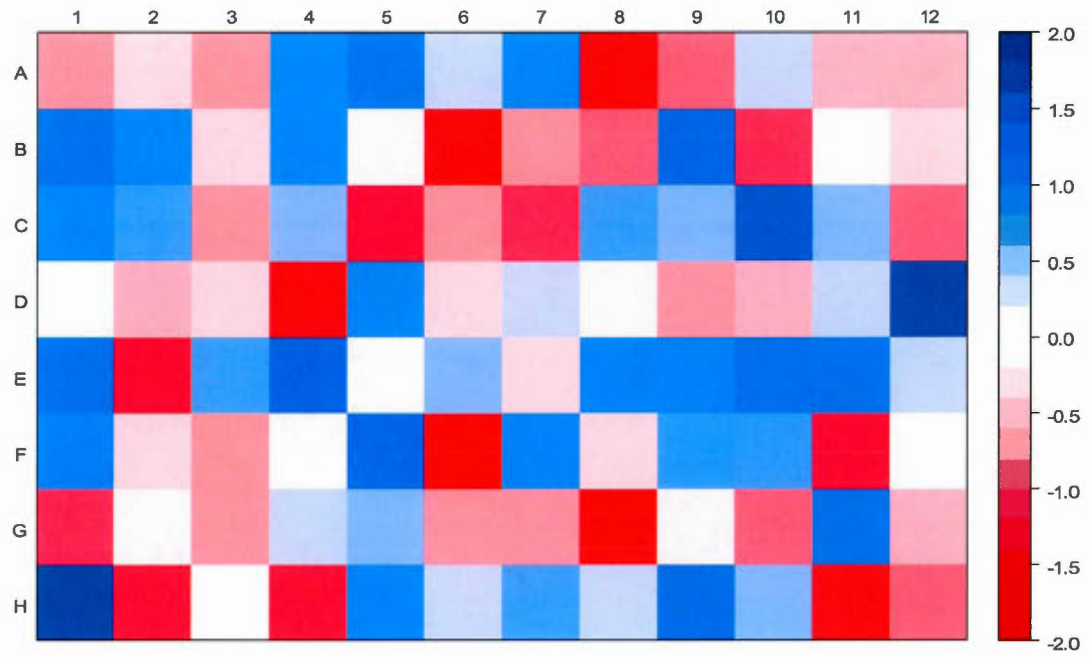


Figure SF4 Plate 3 data corrected first assay-wise, and then plate-wise (using aPMP)

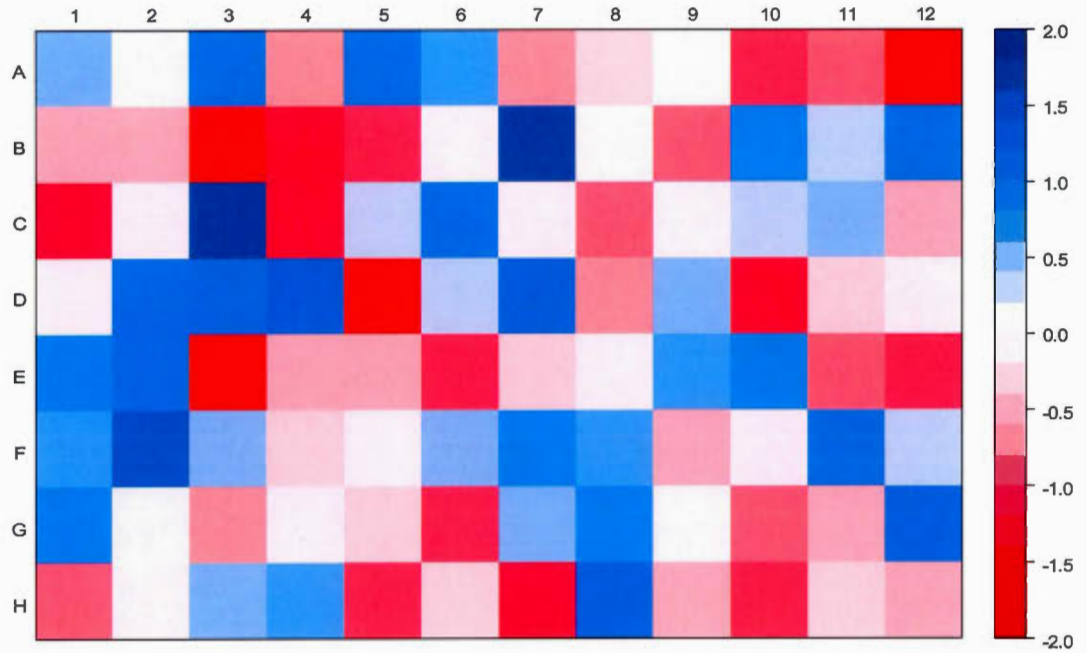


Figure SF5 Plate 4 data corrected first assay-wise, and then plate-wise (using aPMP).

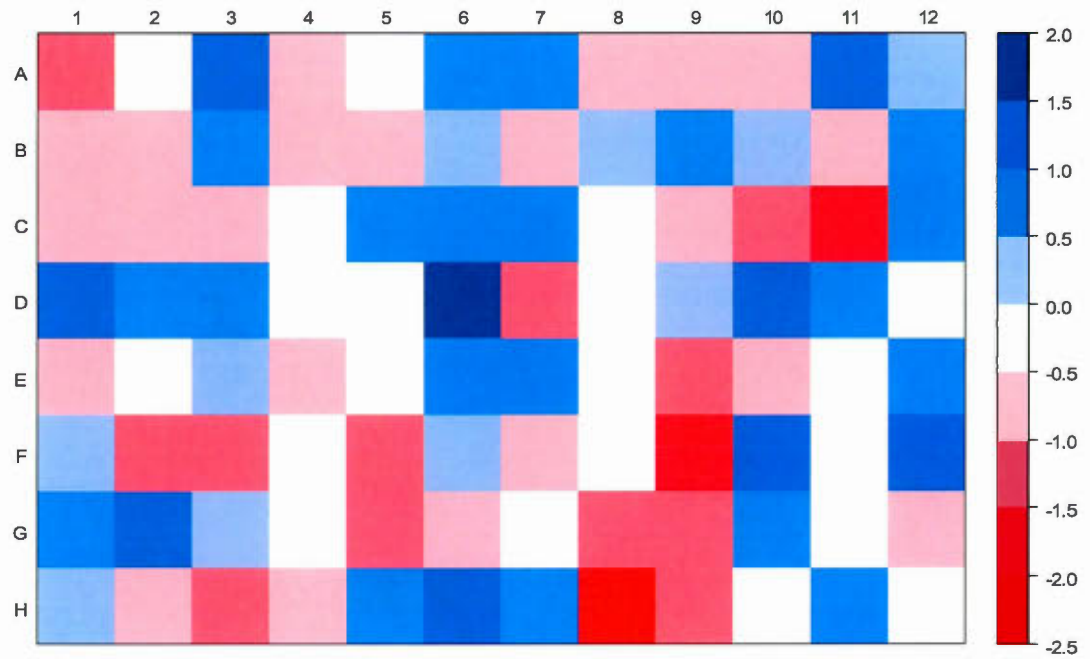


Figure SF6 Plate 5 data corrected first assay-wise, and then plate-wise (using aPMP).

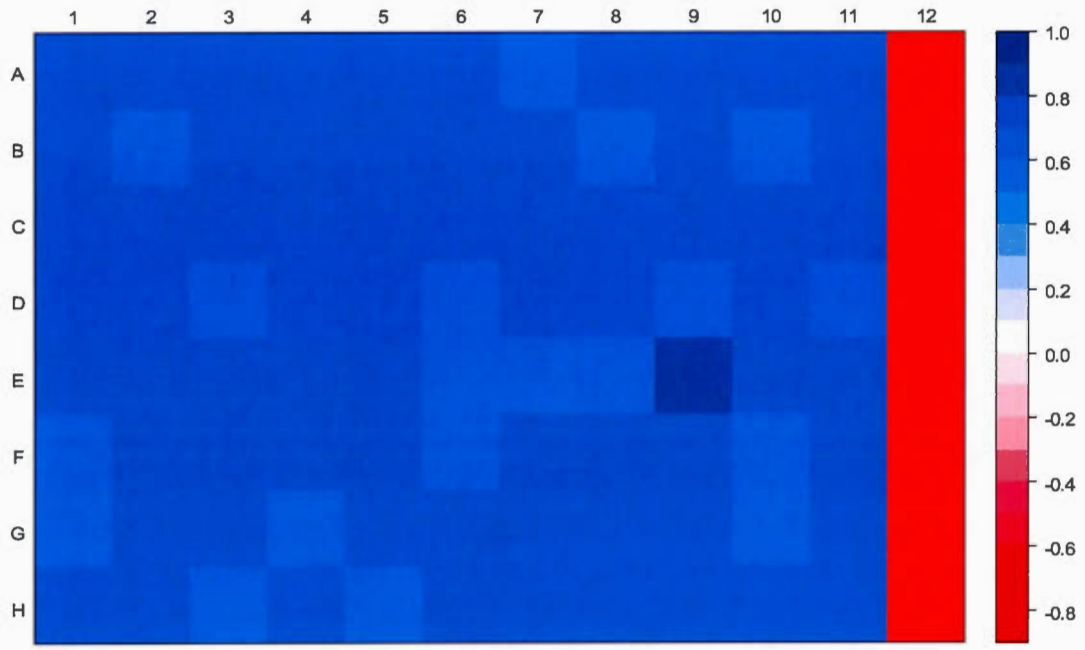


Figure SF7 Plate 1 data corrected first plate-wise (using aPMP), and then assay-wise.

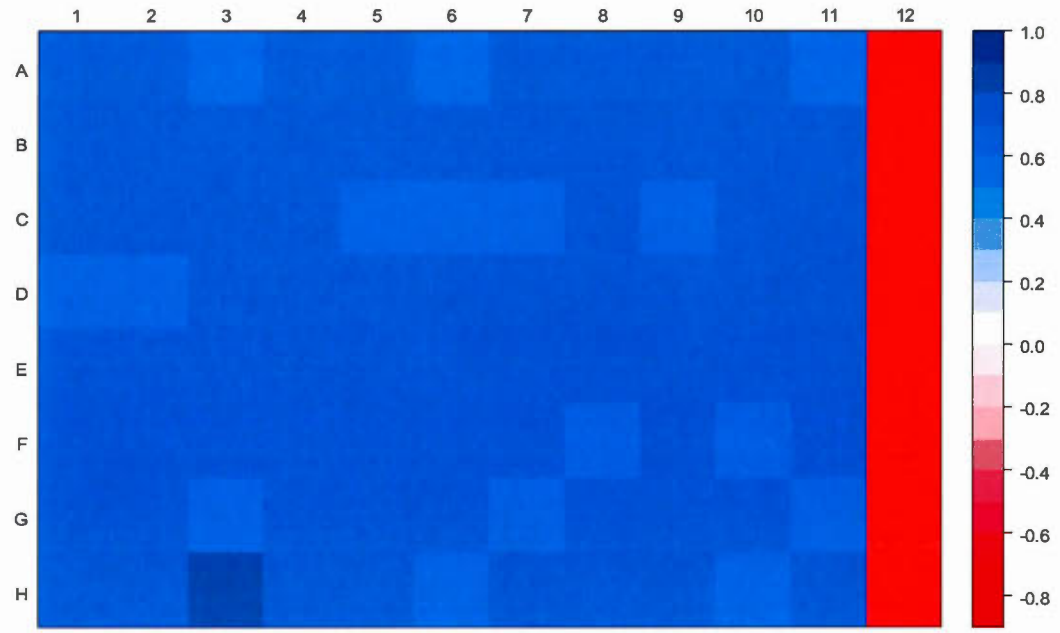


Figure SF8 Plate 2 data corrected first plate-wise (using aPMP), and then assay-wise.

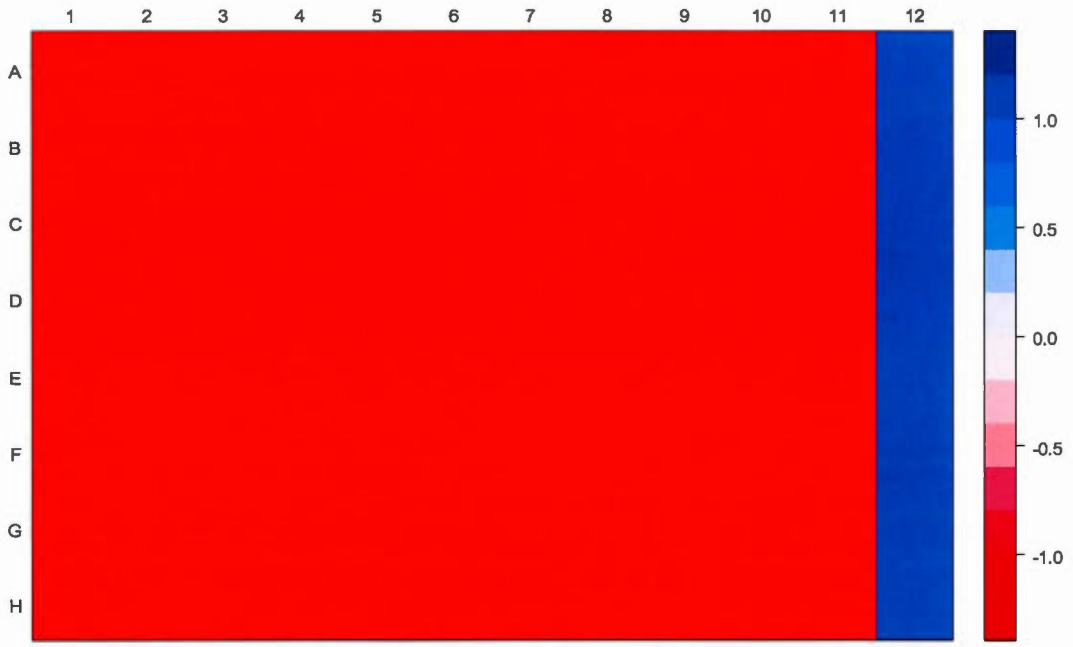


Figure SF9 Plate 3 data corrected first plate-wise (using aPMP), and then assay-wise.

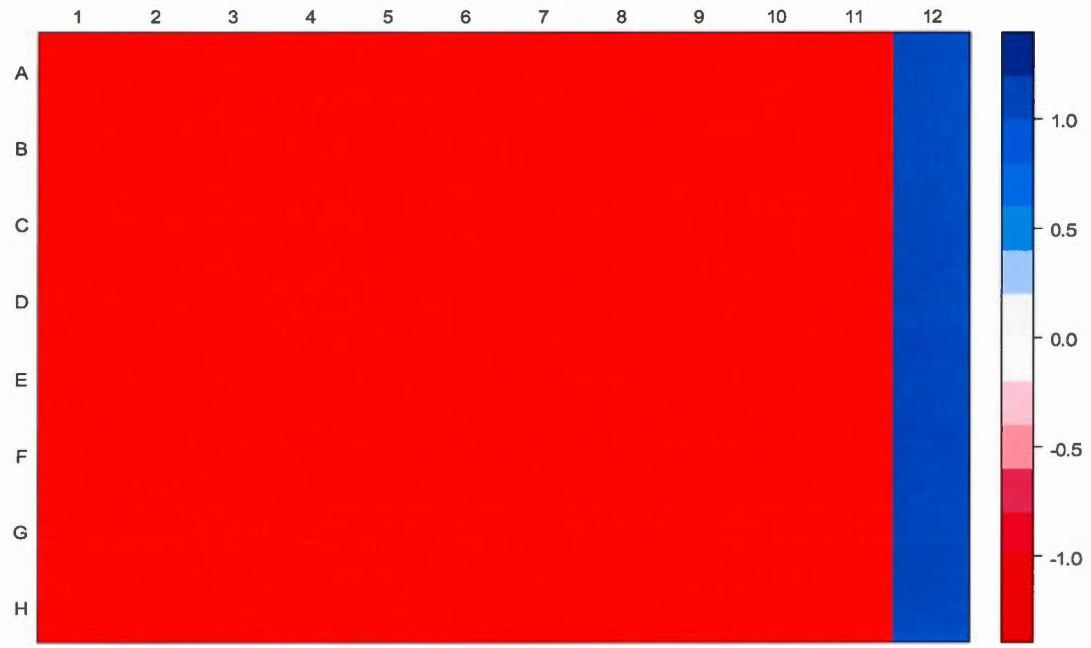


Figure SF10 Plate 4 data corrected first plate-wise (using aPMP), and then assay-wise.

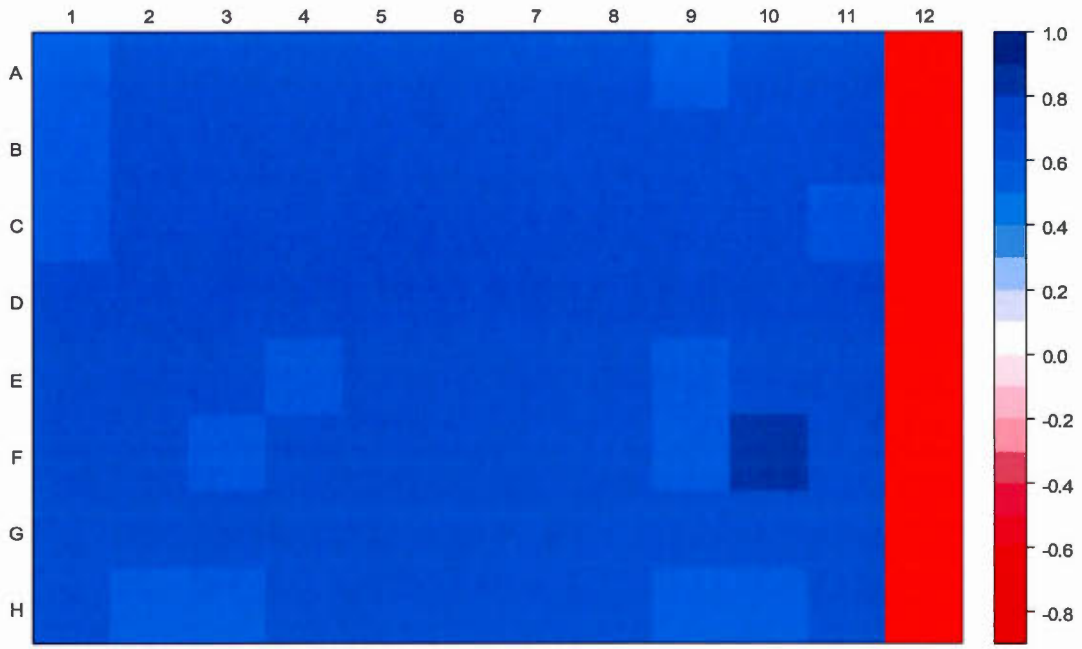


Figure SF11 Plate 5 data corrected first plate-wise (using aPMP), and then assay-wise

CHAPTER III

IDENTIFICATION AND CORRECTION OF ADDITIVE AND MULTIPLICATIVE SPATIAL BIASES IN EXPERIMENTAL HIGH-THROUGHPUT SCREENING

This chapter is a reproduction of the following article: Iurie Caraus, Bogdan Mazoure, Robert Nadon and Vladimir Makarenkov. "Identification and correction of additive and multiplicative spatial biases in experimental high-throughput screening", DISC-17-0095, *Revue SLAS Discovery*, 2017 (submitted).

3.1 Abstract

Data generated by high-throughput screening technologies are prone to spatial bias. Traditionally, bias correction methods used in high-throughput screening assume either a simple additive or, more recently, a simple multiplicative spatial bias model. However, these models do not always provide an accurate correction of measurements in wells located at the intersection of rows and columns affected by spatial bias. The measurements in these wells depend on the nature of interaction between the involved biases. Here, we propose two novel additive and two novel multiplicative spatial bias models accounting for different types of bias interactions. We describe a statistical procedure which allows for detecting and removing different types of additive and multiplicative spatial biases from multiwell plates. We show how this procedure can be applied by analyzing data generated by the four high-throughput screening technologies (homogeneous, microorganism, cell-based and gene expression HTS), the three high-content screening technologies (area, intensity and cell-count HCS) and the only small-molecule microarray technology available in the ChemBank small-molecule screening database. The proposed

methods are included in the *AssayCorrector* program (AssayCorrector), implemented in R and available on CRAN.

Keywords: Anderson-Darling test, Cramer-von-Mises test, data correction, high-content screening, high-throughput screening, Mann-Whitney *U* test, partial mean polish, small-molecule microarray, spatial bias

3.2 Introduction

The amount of data generated by high-throughput screening technologies that can be used to identify active compounds, such as small molecules, siRNAs or gene expression profiles, has exploded in recent years. This breakthrough has been triggered by a significant drop in the cost of screening technologies (Lachmann et al. 2016; Montgomery et al. 2011). However, data generated by high-throughput screening technologies are often subject to different types of spatial bias which negatively influence the outcomes of high-throughput screening campaigns (Caraus et al. 2015; Carralot et al. 2012). Spatial bias is usually caused by environmental (e.g., irregular changes in the temperature, incubation time, lighting and air flow) or technical (e.g., pipette and reader effects) factors (Heyse 2002; Kevorkov and Makarenkov 2005; Makarenkov et al. 2007). Plate-specific spatial bias (i.e., plate-specific systematic error) is evident as an under or over-estimation of measurements in certain rows and columns of a given plate. Often spatial bias affects the edges of a given plate, causing so-called edge effect. The intersection of rows and columns affected by spatial biases can be of particular interest because the resulting activity measurements depend on the interaction between these biases.

Plate-specific spatial bias in high-throughput screening has been traditionally assumed to fit the following additive model (Malo et al. 2006) (Equation 3.1):

$$x'_{ijp} = x_{ijp} + r_{ip} + c_{jp} + \varepsilon_{ijp}, \quad (3.1)$$

where x'_{ijp} is the biased measurement value in well (i,j) of plate p, x_{ijp} is the unbiased measurement value in this well, r_{ip} is the bias affecting row i of plate p, c_{jp} is the bias affecting column j of plate p and ε_{ijp} is the random error in well (i,j) of plate p.

Recent studies have suggested that plate-specific spatial bias can also be of multiplicative kind (Caraus et al. 2017) (Equation 3.2):

$$x'_{ijp} = x_{ijp} \times r_{ip} \times c_{jp} + \varepsilon_{ijp}, \quad (3.2)$$

Correction methods based on different statistical procedures have been developed to remove systematic error using either the classical additive or classical multiplicative spatial bias models. For instance, the additive bias model was used in the following bias correction techniques: B-score (Brideau et al. 2003), R-score (Wu et al. 2008), additive Partial Mean Polish (Dragiev et al. 2012) (aPMP) and SPAWN (Murie et al. 2013). The Diffusion model (Carralot et al. 2012) and the multiplicative PMP (mPMP) method (Caraus et al. 2017) are among a few multiplicative bias models. However, the issue of correcting measurements located at the intersection of biased rows and columns has not been addressed so far.

In this paper, we introduce two additive and two multiplicative bias models, proposing different treatments of measurements located at the intersections of biased rows and columns, and a new statistical procedure, based on the use of the Anderson-Darling (Anderson and Darling 1952) or the Cramer-von-Mises (Cramér 1928) goodness-of-fit tests, which allows one to select the most appropriate spatial bias model for a given plate. This procedure includes modeling bias effects by using traditional additive and multiplicative bias equations (Equations 3.1-3.2) as well as the arithmetic and geometric means to describe possible interactions between row and column spatial biases. The Mann-Whitney U test is used in our procedure in order to identify rows and columns of a given plate which are affected by spatial bias. A new

variant of the Partial Mean Polish algorithm (Caraus et al. 2017; Dragiev et al. 2012) will be used to remove systematic error from biased plate measurements. The discussed statistical procedure will be applied to analyze experimental data generated by high-throughput (HTS), high-content (HCS) and small-molecule microarray (SMM) screening technologies, publicly available at ChemBank (Seiler et al. 2008).

3.3 Methods

In this section, we present the statistical methods we propose to carry out in order to detect and remove additive and multiplicative spatial biases from multiwell plates used in screening technologies. Spatial bias present in rows and columns of a given plate is first detected using the non-parametric Mann-Whitney *U* test. Then, the most appropriate additive or multiplicative bias model will be determined and the corresponding bias removal algorithm will be carried out. Six spatial bias models will be described here, including four new models. The Cramer-von-Mises (CVM) and Anderson-Darling (AD) tests will be used to assess the goodness-of-fit between raw and corrected screening measurements.

3.3.1 Spatial bias detection in multiwell plates

The non-parametric Mann-Whitney *U* test was applied to identify rows and columns of a given plate which are affected by spatial bias. In contrast to the *t*-test, which was used for example in Dragiev et al. (2012), the Mann-Whitney *U* test makes no assumptions about the underlying distribution (e.g., normality), is at least 86.4% asymptotically as efficient as the *t*-test (Hodges Jr and Lehmann 1956), and is robust to outlying observations.

Our spatial bias detection algorithm proceeds by assessing the difference in the compound activity values between a given row or column (Sample 1) and the rest of

the plate's measurements (Sample 2). Once all rows and columns have been assessed in turn, the row or column with the smallest statistically significant p-value is added to the set of biased locations (i.e., rows and columns). The process continues either until convergence, i.e., until no more significant p-values are found, or until a fixed proportion of rows and columns containing spatial bias has been found.

3.3.2 Correction of plate-specific bias

Plate-specific biases affecting rows and columns of a given plate p with m rows and n columns can fit either the additive (Models 1-3 below) or the multiplicative (Models 4-6 below) bias model, which can be described by the following equations (with $1 \le i \le m$ and $1 \le j \le n$):

Model 1 – Additive model

$$\textit{Model: } x'_{ijp} = x_{ijp} + r_{ip} + c_{jp} + \varepsilon_{ijp}; \tag{3.3}$$

Correction: $\hat{x}^{(0)}_{ijp} = x'_{ijp}$ (*initial condition – same for all models*);

$$\hat{x}^{(k+1)}_{ijp} = \hat{x}^{(k)}_{ijp} - (\mu^{(k)}_{ip} - \mu_p) - (\mu^{(k)}_{jp} - \mu_p) \text{ (iterative formula).}$$

Model 2 – Additive model with arithmetic mean at the intersection

$$\textit{Model: } x'_{ijp} = x_{ijp} + r_{ip} + \varepsilon_{ijp} \text{ (when bias is present in row } i \text{ and absent in column } j); \tag{3.4}$$

$$x'_{ijp} = x_{ijp} + c_{jp} + \varepsilon_{ijp} \text{ (when bias is absent in row } i \text{ and present in column } j);$$

$$x'_{ijp} = x_{ijp} + \frac{r_{ip} + c_{jp}}{2} + \varepsilon_{ijp} \text{ (when bias is present in both row } i \text{ and column } j);$$

$$\textit{Correction: } \hat{x}^{(k+1)}_{ijp} = \hat{x}^{(k)}_{ijp} - (\mu_p - \frac{\mu^{(k)}_{ip} + \mu^{(k)}_{jp}}{2}) \text{ (iterative formula).}$$

Model 3 – Additive model with multiplicative interaction of biases at the intersection

$$\textit{Model: } x'_{ijp} = x_{ijp} + r_{ip} + \varepsilon_{ijp} \text{ (when bias is present in row } i \text{ and absent in column } j); \quad (3.5)$$

$$x'_{ijp} = x_{ijp} + c_{jp} + \varepsilon_{ijp} \text{ (when bias is absent in row } i \text{ and present in column } j);$$

$$x'_{ijp} = x_{ijp} + r_{ip} \times c_{jp} + \varepsilon_{ijp} \text{ (when bias is present in both row } i \text{ and column } j);$$

$$\textit{Correction: } \hat{x}_{ijp}^{(k+1)} = \hat{x}_{ijp}^{(k)} - (\mu_{ip}^{(k)} - \mu_p)(\mu_{jp}^{(k)} - \mu_p) \text{ (iterative formula).}$$

Model 4 – Multiplicative model

$$\textit{Model: } x'_{ijp} = x_{ijp} \times r_{ip} \times c_{jp} + \varepsilon_{ijp}; \quad (3.6)$$

$$\textit{Correction: } \hat{x}_{ijp}^{(k+1)} = \frac{\mu_p^2 \hat{x}_{ijp}^{(k)}}{\left|\mu_{ip}^{(k)} \mu_{jp}^{(k)}\right|} \text{ (iterative formula).}$$

Model 5 – Multiplicative model with arithmetic mean at the intersection

$$\textit{Model: } x'_{ijp} = x_{ijp} \times r_{ip} + \varepsilon_{ijp} \text{ (when bias is present in row } i \text{ and absent in column } j); \quad (3.7)$$

$$x'_{ijp} = x_{ijp} \times c_{jp} + \varepsilon_{ijp} \text{ (when bias is absent in row } i \text{ and present in column } j);$$

$$x'_{ijp} = x_{ijp} \times \left(\frac{r_{ip} + c_{jp}}{2}\right) + \varepsilon_{ijp} \text{ (when bias is present in both row } i \text{ and column } j);$$

$$\textit{Correction: } \hat{x}_{ijp}^{(k+1)} = \frac{2\left|\mu_p\right| \hat{x}_{ijp}^{(k)}}{\left|\mu_{ip}^{(k)}\right| + \left|\mu_{jp}^{(k)}\right|} \text{ (iterative formula).}$$

Model 6 – Multiplicative model with geometric mean at the intersection

$$\textit{Model: } x'_{ijp} = x_{ijp} \times r_{ip} + \varepsilon_{ijp} \text{ (when bias is present in row } i \text{ and absent in column } j); \quad (3.8)$$

$$x'_{ijp} = x_{ijp} \times c_{jp} + \varepsilon_{ijp} \text{ (when bias is absent in row } i \text{ and present in column } j);$$

$$x'_{ijp} = x_{ijp} \times \sqrt{r_{ip} \times c_{jp}} + \varepsilon_{ijp} \text{ (when bias is present in both row } i \text{ and column } j);$$

$$\textit{Correction: } \hat{x}_{ijp}^{(k+1)} = \frac{\mu_p \hat{x}_{ijp}^{(k)}}{\sqrt{\left|\mu_{ip}^{(k)} \mu_{jp}^{(k)}\right|}} \text{ (iterative formula),}$$

where x'_{ijp} is the biased (i.e., observed) raw measurement in well (i,j) of plate p, x_{ijp} is the unbiased activity measurement in well (i,j) of plate p, $\hat{x}_{ijp}$ is the estimated measurement in well (i,j) of plate p, k is the iteration number, r_{ip} is the value of systematic error (i.e., spatial bias) affecting row i of plate p, c_{jp} is the value of systematic error affecting column j of plate p, μ_p is the mean of *unbiased measurements* of plate p, μ_{ip} is the mean of measurements in row i of plate p, μ_{jp} is the mean of measurements in column j of plate p, and μ_{jp} is the random error affecting well (i,j) of plate p. The random error ε_{ijp} is assumed to be small compared to the values of spatial biases c_{jp} and r_{ip}. The impact of random errors on the hit selection process can be minimized by using compounds' replicates (Caraus et al. 2015; Malo et al. 2006). In the additive models (Models 1-3), unbiased rows and columns have $r_{ip} = 0$ and $c_{jp} = 0$, respectively, while in the multiplicative models (Models 4-6) unbiased rows and columns have $r_{ip} = 1$ and $c_{jp} = 1$, respectively.

Model 1 is the traditional additive model, which is assumed in many bias correction methods (Brideau et al. 2003; Dragiev et al. 2012; Murie et al. 2013). Model 2 is an additive model, which assumes that the row and column biases at the intersections are combined through their arithmetic mean. Model 3 is an additive model, which

assumes that the row and column biases at the intersections interact in a multiplicative fashion. Model 4 is a recently introduced basic multiplicative bias model. Model 5 is a multiplicative model, which assumes that the intersection effects of row and column biases are combined through their arithmetic mean. Model 6 is a multiplicative model, which assumes that the row and column biases at the intersection are combined through their geometric mean. Models 2 and 3 can be viewed as extensions of the *additive* Partial Mean Polish model (Model 1), while Models 5 and 6 can be viewed as extensions of the *multiplicative* Partial Mean Polish model (Model 4). Partial mean polish algorithms(Caraus et al. 2017; Dragiev et al. 2012) are variations of Tukey's median polish (Tukey 1977), which iteratively removes spatial bias from the rows and columns affected.

3.3.3 Model selection for removing plate-specific spatial bias

In this section we present a data processing protocol which can be used to identify the most appropriate spatial bias model for a given plate. If the presence of spatial bias in a given plate has been detected using the Mann-Whitney U test, then we first partition all wells of the plate into: (i) wells with no bias detected and (ii) wells located in biased rows and columns. Control wells should be excluded from all computations. Wells affected by spatial bias can then be corrected using the six iterative methods (Models 1-6) discussed above. Our procedure generates seven sets of data: a set of unbiased wells and six sets of biased wells corrected by each of the six methods. The Anderson-Darling or Cramer-von-Mises test can then be used on the set of unbiased wells and on each of the six corrected sets of wells in order to assess goodness-of-fit of the corrections performed. The resulting p-values can then be used to determine the most appropriate bias model (among Models 1 to 6) for the data at hand. In this work, the significance level α was set to 0.01 for all the three statistical tests being performed (Mann-Whitney U test, Anderson-Darling test and Cramer-von-Mises

test). It is worth noting that both the Anderson-Darling and Cramer-von-Mises tests can be used in the algorithm below, but the Anderson-Darling tests is preferred due to its higher power (Razali and Wah 2011).

Our algorithm proceeds as follows:

i. Perform the Mann–Whitney *U* test on each individual plate of the assay (i.e., plate-wise correction) to identify biased rows and columns.

For each plate containing biased rows and/or columns, perform:

ii. Apply each of the six additive and multiplicative PMP algorithms corresponding to Models (1-6) discussed in this section (see the related iterative formulas 3 to 8);

iii. Apply the Anderson-Darling or Cramer-von-Mises two-sample test on the co*rrected* plates after the application of the six versions of the PMP algorithm. Compute the corresponding p-values (3 additive and 3 multiplicative);

iv. If all additive and all multiplicative p-values are higher than the selected significance level α, then the bias model for this plate is the model with the highest p-value (low confidence);

v. If all additive p-values are lower than the selected significance level α and all multiplicative p-values are higher than α,, then the bias model for this plate is multiplicative (high confidence);

vi. If all multiplicative p-values are lower than the selected significance level α and all additive p-values are higher than α, then the bias model for this plate is additive (high confidence);

vii. If none of the conditions specified in Steps iv to vi apply, then the bias model is undefined;

viii. If the bias model has been identified (i.e., it is not undefined), then correct the plate measurements using the corresponding bias correction procedure.

3.3.4 Traditional and modified Partial Mean Polish (PMP) algorithms

In this section we present the details of the additive and multiplicative PMP algorithms corresponding to Models 1 to 6 discussed above. The main advantage of these algorithms compared to the popular *B*-score method of Brideau et al. (2003) is that they do not reduce the original measurements to residuals and do not modify the unbiased measurements.

Our generalized PMP algorithm proceeds as follows:

1. Let $R = \{r_1, r_2, ..., r_s \mid 0 \le s < m\}$ and $C = \{c_1, c_2, ..., c_t \mid 0 \le t < n\}$ be the sets of biased rows and columns of plate $p(m \times n)$, respectively. Calculate the mean, μ_p, of all unbiased measurements of p:

$$\mu_p = \frac{1}{(m-s)(n-t)} \sum_{i \notin R, j \notin C} x_{ijp} \tag{3.9}$$

2. For each biased row $i: 1 \le i \le s$, calculate the mean value, $\mu_{r_i p}$, of row r_i: $\mu_{r_i p} = \frac{1}{n} \sum_{j=1}^{n} x_{r_i j p}$, and the estimate of the row bias, $\hat{e}_{r_i p}$. For each biased column $j: 1 \le j \le t$, calculate the mean value, $\mu_{c_j p}$, of column c_j: $\mu_{c_j p} = \frac{1}{m} \sum_{i=1}^{m} x_{i c_j p}$, and the estimate of the column error, $\hat{e}_{c_j p}$. Use the following equations to calculate the error estimates:

$\hat{e}_{r_i p} = \mu_{r_i p} - \mu_p$ and $\hat{e}_{c_j p} = \mu_{c_j p} - \mu_p$ (Model 1);

$\hat{e}_{r_i p} = \frac{\mu_p - \mu_{r_i p}}{2}$ and $\hat{e}_{c_j p} = \frac{\mu_p - \mu_{c_j p}}{2}$ (Model 2);

$\hat{e}_{r_i p} = \mu_{r_i p} - \mu_p$ and $\hat{e}_{c_j p} = \mu_{c_j p} - \mu_p$ (Model 3);

$\hat{e}_{r_i p} = \left| \frac{\mu_{r_i p}}{\mu_p} \right|$ and $\hat{e}_{c_j p} = \left| \frac{\mu_{c_j p}}{\mu_p} \right|$ (Model 4);

$$\hat{e}_{\eta p} = \left|\frac{\mu_{\eta p}}{\mu_p}\right| \text{ and } \hat{e}_{cjp} = \left|\frac{\mu_{cjp}}{\mu_p}\right| \quad \text{(Model 5);}$$

$$\hat{e}_{\eta p} = \sqrt{\left|\frac{\mu_{\eta p}}{\mu_p}\right|} \text{ and } \hat{e}_{cjp} = \sqrt{\left|\frac{\mu_{cjp}}{\mu_p}\right|} \quad \text{(Model 6).}$$

3. For all rows and columns affected by spatial bias, iteratively adjust their measurements using the error estimates determined in Step 2, i.e., for all $i : 1 \le i \le s$ and $j : 1 \le j \le t$, proceed as follows:

$$\hat{x}_{ijp} = \hat{x}_{ijp} - \left(\hat{e}_{\eta p} + \hat{e}_{cjp}\right) \quad \text{(Model 1);}$$

$$\hat{x}_{ijp} = \hat{x}_{ijp} - \left(\hat{e}_{\eta p} + \hat{e}_{cjp}\right) \quad \text{(Model 2);}$$

$$\hat{x}_{ijp} = \hat{x}_{ijp} - \hat{e}_{\eta p}\hat{e}_{cjp} \quad \text{(Model 3);}$$

$$\hat{x}_{ijp} = \frac{\hat{x}_{ijp}}{\hat{e}_{\eta p}\hat{e}_{cjp}} \quad \text{(Model 4);}$$

$$\hat{x}_{ijp} = \frac{2\hat{x}_{ijp}}{\hat{e}_{\eta p} + \hat{e}_{cjp}} \quad \text{(Model 5);}$$

$$\hat{x}_{ijp} = \frac{\hat{x}_{ijp}}{\hat{e}_{\eta p}\hat{e}_{cjp}} \quad \text{(Model 6).}$$

4. If $\sum_{i=1}^{s} |\hat{e}_{\eta p}| + \sum_{j=1}^{t} |\hat{e}_{cjp}| > \varepsilon$, then go to Step 2, otherwise stop the algorithm.

Here ε is a small fixed positive threshold.

3.3.5 Choice of statistical test to determine the bias model

The two following statistical tests, used in Step iii of the above-presented bias correction protocol, were examined in order to assess the goodness-of-fit between raw and corrected measurements: Cramer-von-Mises (CVM) test and Anderson-Darling (AD) test. The AD test is based on a quadratic empirical distribution function (EDF) statistic. Importantly, the AD test has been shown to be the most powerful

among the EDF tests (Arshad et al. 2003). The Cramer-von-Mises test is a special case of AD which puts less weight on the tails of the distribution. It has been shown that the power of the Anderson-Darling test is higher than that of the Kolmogorov-Smirnov (Kolmogorov 1933), probability-plot, L moments and chi-square (χ^2) tests (Razali and Wah 2011). Moreover, the Anderson-Darling test is generally more powerful than the Cramer-von-Mises test (Laio 2004). These conclusions mirror those formulated by Stephens (Stephens 1986).

3.4 Results and Discussion

To assess the extent of plate-specific bias in the HTS, HCS and SMM technologies, we examined 175 experimental assays from the ChemBank screening repository. ChemBank (Seiler et al. 2008) is a public small-molecule screening database created by the Broad Institute's Chemical Biology Program, which provides life scientists access to biomedically relevant screening data and tools. Here we considered all the eight screening categories available in ChemBank: HTS (homogeneous), HTS (microorganism), HTS (cell-based), HTS (gene expression), HCS (area), HCS (intensity), HCS (count) and SMM. Among the 175 examined assays we considered: 25 assays of each HTS category, 8 assays of HCS (area - all available non-empty assays of this type in April 2017), 18 assays of HCS (intensity - all available non-empty assays of this type in April 2017), 24 assays of HCS (cell count - all available non-empty assays of this type in April 2017) and 25 assays of SMM. The ChemBank IDs of these assays are presented in Supplementary Table 3.2 (see Supplementary Material).

First, we calculated the proportion of plates with at least one row and at least one column under or over-estimation effect (i.e., having at least one intersection of biased rows and columns). After analyzing a total of 2241 plates from the selected 175 assays, we found that plates with intersections of biased rows and columns (51.9%)

were slightly more frequent than plates without such intersections (48.1%). Note that the difference between Models 1 to 3 and between Models 4 to 6 occurs only in the presence of intersections. Models 1 to 6 discussed in the Methods section allow us to take into account the complex nature of interactions between row and column spatial biases.

Second, we assessed the distribution of the number of biased rows and columns per plate in the selected set of ChemBank assays. Figure 3.1 shows the distribution of the number of rows (Fig. 3.1a) and columns (Fig. 3.1b) per plate affected by spatial bias, computed over the 2441 plates of the examined ChemBank assays. The Mann-Whitney *U* test was carried out to detect the presence of spatial bias within each plate. The presented results suggest that plates' columns tend to be more biased than plates' rows in experimental small-molecule screens.

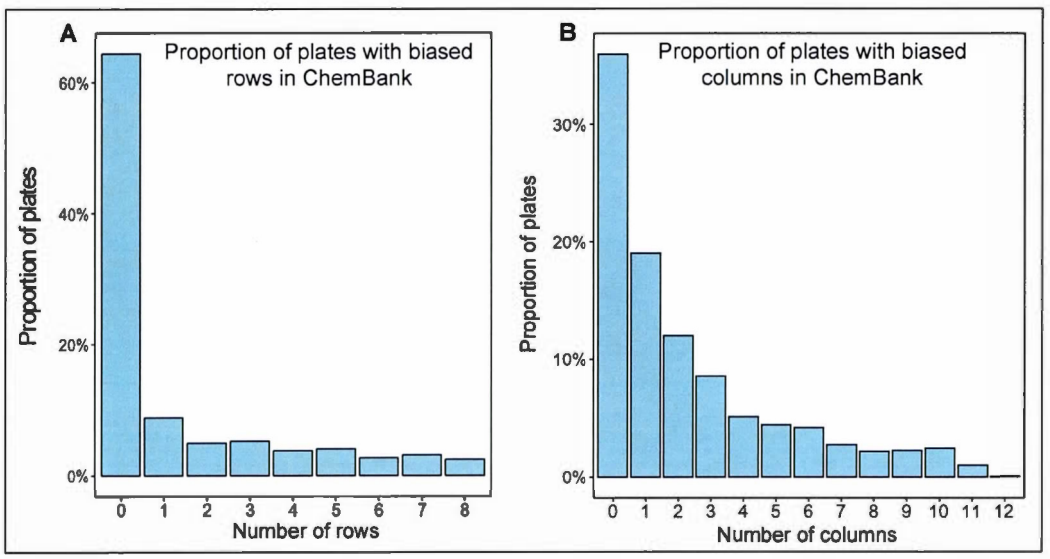


Figure 3.1 Proportion of plates with the indicated number of rows (A) affected by spatial bias and number of columns (B) affected by spatial bias. The non-parametric Mann-Whitney *U* test was used to detect the presence of spatial bias within each plate.

Third, we carried out the correction of biased rows and columns following the procedure described in the Methods section. Plate-specific error correction was applied on a plate-by-plate basis. It was applied only on plates in which the presence of spatial bias was detected by the Mann-Whitney *U* test. If the presence of spatial bias was identified within a given plate, then the detected biased measurements were corrected using the most appropriate additive or multiplicative bias correction model according to our algorithm (see the Methods section). Evidence from the Anderson-Darling and Cramer-von-Mises tests was used to select the best fitting model.

Figure 3.2 presents the model selection frequencies for our six spatial bias models (Models 1 to 6) across the eight screening categories. Here, the Anderson-Darling test was used to assess the goodness-of-fit of the corrected measurements to the unbiased measurements. According to the Anderson-Darling test, HTS data tend to contain more spatial bias than HCS or SMM data. Moreover, HTS data followed an undefined bias model more frequently than HCS or SMM data (23.7%, 5.8% and 16.3%, respectively).

Table 3.1 Contingency table of the number of plates affected by additive, multiplicative and undefined type of spatial bias, for the HTS, HCS and SMM technologies.

	HTS	**HCS**	**SMM**
Undefined	362	24	115
Additive	509	83	259
Multiplicative	352	88	58

Table 3.1 shows that the number of plates affected by additive, multiplicative and undefined type of spatial bias differed across the three technologies, as suggested by the χ^2 test of independence (χ^2 (4) = 97.32; $p = 5\text{x}10^{-4}$). Post-hoc tests showed that the proportion of HTS plates affected by undefined bias was significantly higher than

in HCS ($\chi^2(1) = 25.38$; $p = 4.7\text{x}10^{-7}$), but not in SMM (χ^2 (1) = 1.38; $p = 2.4\text{x}10^{-1}$) plates. The proportion of plates in SMM assays affected by undefined bias was also higher than in HCS assays (p-value = $6.50\text{x}10^{-5}$, χ^2=15.95). The two-by-two contingency tables were constructed by combining respective bias model and screening technologies plate counts.

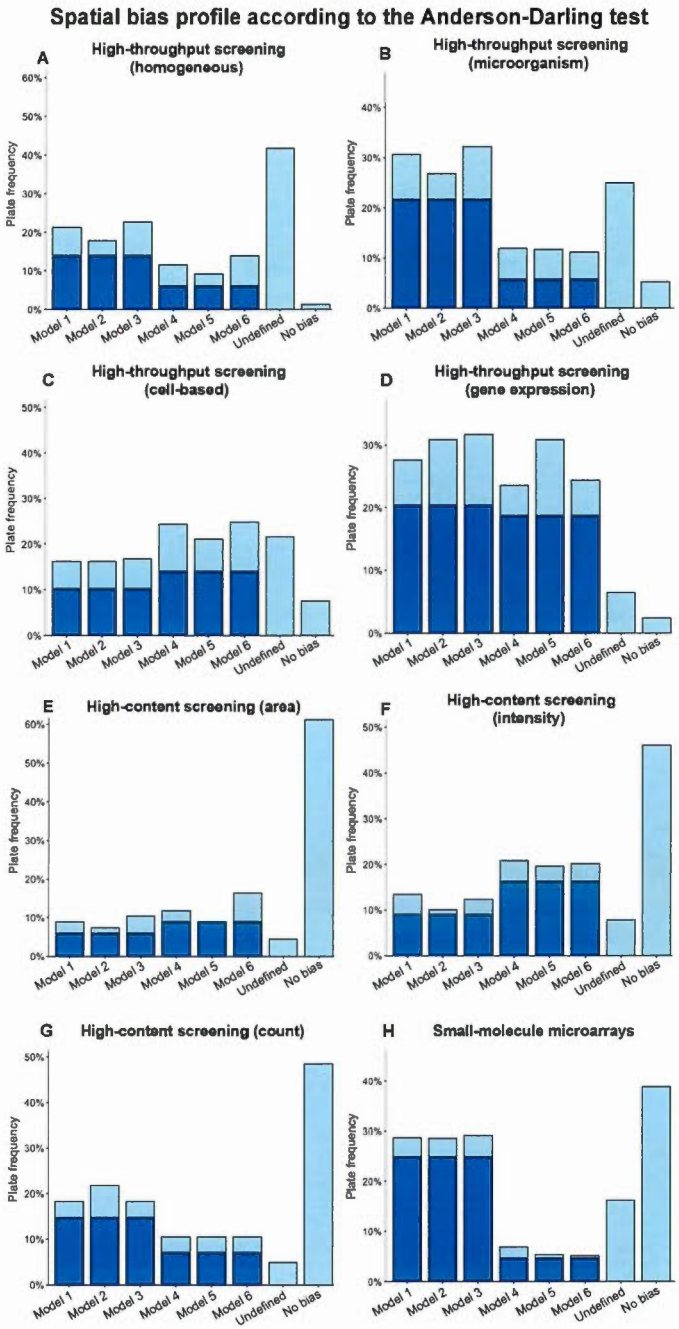


Figure 3.2 Spatial bias model frequency based on evidence obtained from the Anderson-Darling test for high-throughput screening data (A)-(D), high-content

screening data (E)-(G) and small-molecule microarrays data (H). Control wells were excluded from all computations. Darker portions of bars account for plates without intersections of rows and columns affected by spatial bias; in this case any additive model (Models 1-3) can be applied when spatial bias was classified as additive and any multiplicative model (Models 4-6) can be applied when spatial bias was classified as multiplicative. Lighter portions of bars corresponding to Models 1 to 6 account for plates with intersections of rows and columns affected by spatial bias (in this case, the model yielding the largest *p*-value should be applied).

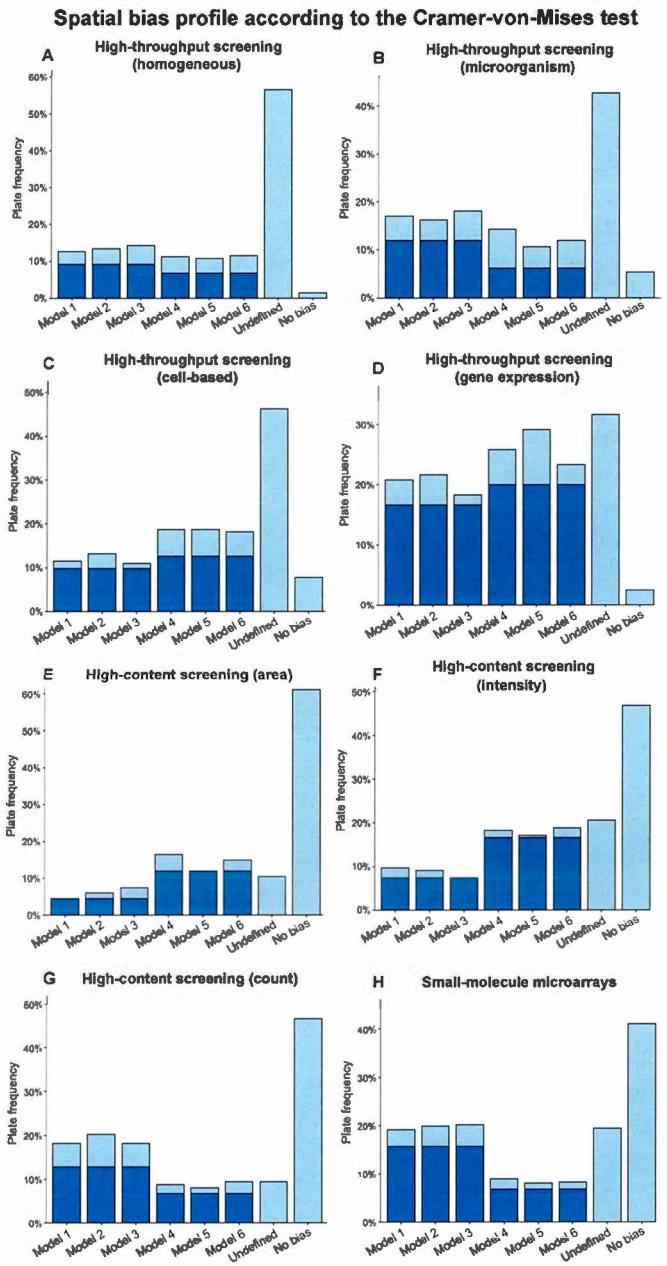


Figure 3.3 Spatial bias model frequency based on evidence obtained from the Cramer-von-Mises test for high-throughput screening data (A)-(D), high-content screening data (E)-(G) and small-molecule microarrays data (H). Control wells were

excluded from all computations. Darker portions of bars account for plates without intersections of rows and columns affected by spatial bias; in this case any additive model (Models 1-3) can be applied when spatial bias was classified as additive and any multiplicative model (Models 4-6) can be applied when spatial bias was classified as multiplicative. Lighter portions of bars corresponding to Models 1 to 6 account for plates with intersections of rows and columns affected by spatial bias (in this case, the model yielding the largest *p*-value should be applied).

Figure 3.3 depicts the bias model selection frequencies according to the Cramer-von-Mises test obtained for data of the eight considered screening categories. Similarly to the AD test, evidence from the CVM test indicates that HTS data tend to contain more biased plates that do not correspond to any of the six bias models presented in our study (i.e., the bias model for HTS data was classified as undefined more frequently than for HCS and SMM data). In particular, the bias model was identified as undefined in 44.3% of HTS plates (computed over the four HTS screening categories), compared to 13.5% in HCS (computed over the three HTS screening categories) and 19.5% in SMM.

Figures 3.2 and 3.3 show that only 1.4% of HTS (homogeneous) plates, 5.2% of HTS (microorganism) plates, 7.6% of HTS (cell-based) plates and 2.4% of HTS (gene expression) plates were unbiased, contrary to 61.2% of unbiased HCS (area) plates, 46.1 % of unbiased HCS (intensity) plates, 48.6% of unbiased HCS (cell count) plates and 38.9% of unbiased SMM plates. All *p*-values obtained by applying the McNemar test on two-by-two contingency tables (screening technologies versus bias model types) were below $1x10^{-3}$. The number of unbiased plates, determined by the Mann-Whitney *U* test, was same for both figures.

Let us now examine in more detail the results presented in Figures 3.2 and 3.3. The McNemar test was used to test the differences in proportions mentioned below.

- For HTS homogeneous data (Figs. 3.2A and 3.3A), spatial bias was found in 98.6% of plates. Here, the CVM test found the same quantity of plates corresponding to an undefined bias model as the AD test (56.6% compared to 41.8%, χ^2 (1) = 0.16; p = 6.85x10^{-1}). Both the CVM and AD tests found that both multiplicative and additive bias proportions were not different for data of the HTS homogeneous category (χ^2 (1) = 2.07; p = 1.50x10^{-1}).

- For HTS microorganism data (Figs. 3.2B and 3.3B), spatial bias was found in 94.6% of plates. The AD test identified much less plates with undefined spatial bias than the CVM test (25.0% compared to 42.7%, χ^2(1) = 51.51; p = 7.11x10^{-13}). Both the AD and CVM tests suggest that the additive bias models are more appropriate than the multiplicative ones for data of the HTS microorganism category (χ^2(1) = 11.73; p = 6.14x10^{-4}).

- For HTS cell-based data (Figs. 3.2C and 3.3C), spatial bias was found in 92.3% of plates. The CVM test suggested that 46.2% of plates have an undefined type of bias, compared to 21.6% of plates according to the AD test (χ^2(1) = 15.72; p = 7.34x10^{-5}). Both tests found that the multiplicative spatial bias model is more adequate for plates of the HTS cell-based category than the additive one (χ^2(1) = 0.53; p = 4.65x10^{-1}).

- For HTS gene expression data (Figs. 3.2D and 3.3D), spatial bias was found in 97.5% of plates. Here, the AD test suggested that 6.5% of plates are affected by an undefined type of bias, compared to 31.7% of plates according to the CVM test (χ^2(1) = 37.75; p = 8.03x10^{-10}). The AD and CVM tests did not find a significant difference in proportions of HTS gene expression plates affected by multiplicative or additive type of spatial bias (χ^2(1) = 2.50; p = 1.14x10^{-1}).

- For HCS area data (Figs. 3.2E and 3.3E), spatial bias was found in 38.8% of plates. Both the CVM and AD tests determined that an undefined bias model is rather rare for this type of data (10.4% and 4.5%, respectively, χ^2(1) = 44.17; p = 3.01x10^{-}

[11]). Here, both statistical tests found no significant difference between the multiplicative and additive spatial bias models ($\chi^2(1) = 0.22$; $p = 6.38\text{x}10^{-1}$).

- For HCS intensity data (Figs. 3.2F and 3.3F), spatial bias was found in 92.1% of plates. The CVM test identified that the bias model was undefined for 20.6% of HCS intensity plates, compared to 7.9% found by the AD test ($\chi^2(1) = 80.65$; $p = 2.70\text{x}10^{-19}$). Both the CVM and AD tests did not provide enough evidence that the additive and multiplicative bias models were present in equal proportions in plates generated by the HCS intensity technology (χ^2 (1)= 0.09; p=$7.69\text{x}10^{-1}$).

- For HCS count data (Figs. 3.2G and 3.3G), spatial bias was found in 51.4% of plates. Both the AD and CVM tests indicated that only a small proportion of plates follow an undefined spatial bias model (4.9% and 9.5%, respectively, $\chi^2(1) = 96.64$; $p = 8.30\text{x}10^{-23}$). No significant difference in proportions of plates affected by additive or multiplicative bias was found here (χ^2 (1) = 0.74; $p = 3.88\text{x}10^{-1}$).

- For small-molecule microarrays data (Figs. 3.2H and 3.3H), spatial bias was found in 58.8% of plates. Here, the CVM and AD tests determined that 19.5% and 16.3% of plates, respectively, follow an undefined bias model ($\chi^2(1) = 294.35$; $p = 5.61\text{x}10^{-66}$). Both tests also suggested that the small-molecule microarrays data tend to be affected by additive bias more frequently than by multiplicative one (χ^2 (1) = 9.1; $p = 2.56\text{x}10^{-3}$).

In this work, we described two new additive and two new multiplicative spatial bias models along with a general bias detection and removal procedure. We presented evidence which suggests that all the six bias models (Models 1 to 6) considered in this study are relevant for the analysis of experimental high-throughput, high-content and small-molecule microarrays data. One of the challenges in spatial bias model selection consists of minimizing the number of plates with undefined type of spatial bias. The AD test generally outperformed the CVM test in distinguishing between the additive and multiplicative bias models, which resulted in a lower number of plates

where spatial bias was classified as undefined. Overall, the AD and CVM tests were in agreement on the spatial bias model that should be used to correct the measurements of a given plate. However, because the AD test is generally more powerful than CVM (Laio 2004) and because it classifies fewer plates as being undefined, it can be recommended for analysis of experimental high-throughput screening assays. We also discovered that data generated by high-throughput screening (HTS) technologies are generally more prone to spatial bias than data from high-content screening (HCS) and small-molecules microarrays (SMM). While HCS and SMM data are more likely to follow one of the six considered bias models, the bias model in HTS was more frequently classified as undefined leaving more challenges for future investigations.

3.5 Supplementary material

Table 3.2 Set of 175 ChemBank assays examined in our plate-specific bias detection simulation (see Figs. 3.1-3.3). Note that only 8 non-empty HCS Area, 18 non-empty HCS Intensity and 24 non-empty HCS Cell count assays were available in ChemBank (as of April 14th, 2017). For all other screening categories, 25 assays per data category were examined.

High-throughput screening (Cell-based) – 25 assays	
1	AdipocyteDifferentiation1_OilRedO(913.0191)
2	AdipocyteDifferentiation2_NileRed(1015.0001)
3	AdipocyteDifferentiation2_NileRed(1015.0032)
4	AdipocyteDifferentiation2_NileRed(1015.0034)
5	AnnotationDevelopment_BrdUCytoblot(900.0001)
6	AnnotationDevelopment_BrdUCytoblot(900.0002)
7	AnnotationDevelopment_BrdUCytoblot(900.0021)
8	AnnotationDevelopment_BrdUCytoblot(900.0022)
9	AnnotationDevelopment_EthD1Staining(900.0005)

10	AnnotationDevelopment_JC1MitoDye(900.0013)
11	AR-NcoRBindingAssay_raw{Lux()}(268.0159)
12	AR-NcoRBindingAssay_raw{Lux()}(268.0173)
13	AR-NcoRBindingAssay_raw{Lux()}(268.0217)
14	AR-NcoRBindingAssay_user{AvgLux()}(268.0221)
15	BreastCancerCellProfiling_CellTiterGlo(915.0248)
16	BreastCancerCellProfiling_JC1MitoDye(915.0244)
17	CellularAutofluorescence_CpdAutofluor(908.0049)
18	CellularAutofluorescence_CpdAutofluor(908.0125)
19	CellViabilityProfiling_CellTiterGlo(1019.0001)
20	DeacetylaseInhibition_AcLysCytoblot(1027.0002)
21	EndothelialCellProfiling1_Calcein-AM(910.0153)
22	FacioscapulohumeralMD_Calc(E1-E2)(1026.0010)
23	FacioscapulohumeralMD_Calc(E1-E2)(1026.0011)
24	FacioscapulohumeralMD_LuxReporter(1026.0003)
25	FacioscapulohumeralMD_LuxReporter(1026.0019)
High-throughput screening (Homogeneous) – 25 assays	
1	ActinPolymerization_raw{FI()}(144.0030)
2	ActinPolymerization_user{Fold()}(144.0031)
3	AdipocyteDifferentiation1_OilRedO(913.0190)
4	AnnotationDevelopment_BrdUCytoblot(900.0020)
5	BRAF_HRPCytoblot(1110.0001)
6	BRAF_HRPCytoblot(1110.0002)
7	BRAF_HRPCytoblot(1110.0003)
8	BRAF_HRPCytoblot(1110.0005)
9	CellularAutofluorescence_CpdAutofluor(908.0050)
10	CMVPolymeraseBindingAssay_raw{Pol(P)}(299.0552)
11	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0543)
12	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0549)
13	CMVPolymeraseBindingAssay_raw{Pol(s)}(299.0553)

14	CREBReporterAssay_LacZReporter(1029.0010)
15	CyclinReporterGeneCdh1_user{Fold()}(219.0091)
16	DihydroorotateDehydrogenase_Calc(E1-E2)(1021.0033)
17	DihydroorotateDehydrogenase_EnzCoupledColor(1021.0001)
18	DihydroorotateDehydrogenase_EnzCoupledColor(1021.0013)
19	EColiFilamentation2006_OpticalDensity(1038.0010)
20	GlycanaseActivity_raw{Pol(P)}(295.0495)
21	HoxDNA-BindingAssay_FluorOligo(1031.0002)
22	HoxDNA-BindingAssay_FluorOligo(1031.0008)
23	HoxDNA-BindingAssay_FluorOligo(1031.0010)
24	KinaseInhibitorModifiers_BrdUCytoblot(901.0010)
25	TrypanothioneReductase_EnzCoupledColor(1017.0020)
High-throughput screening (Microorganism) – 25 assays	
1	ABAggregationInhibitors_OpticalDensity(1103.0009)
2	AntibacterialAssay_FluorProtein(1106.0016)
3	AntibacterialAssay_FluorProtein(1106.0027)
4	AspulvinoneUpregulation_MetabColor(1022.0007)
5	BiofilmFormationAssay_BacTiterGlo(1059.0006)
6	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0609)
7	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0613)
8	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0625)
9	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0657)
10	ClathrinDependentMembraneTrafficking_raw{Abs(Mut)}(310.0737)
11	ClathrinDependentMembraneTrafficking_user{Fold(Mut)}(310.0733)
12	EColiFilamentation2006_OpticalDensity(1038.0001)
13	EColiFilamentation2006_OpticalDensity(1038.0002)
14	EColiFilamentation2006_OpticalDensity(1038.0012)
15	EColiFilamentation2006_OpticalDensity(1038.0014)
16	EColiFilamentation2006_OpticalDensity(1038.0016)
17	EColiFilamentation2006_OpticalDensity(1038.0022)

18	EColiFilamentation2006_OpticalDensity(1038.0023)
19	EColiFilamentation2006_OpticalDensity(1038.0024)
20	PDERegulators_OpticalDensity(1091.0043)
21	PhosphatidylinositolKinase_OpticalDensity(1000.0008)
22	PhosphatidylinositolKinase_OpticalDensity(1000.0027)
23	PSACAntagonistScreen_OpticalDensity(1035.0016)
24	SulfurAssimilation_user{Inh(BioB)}(130.0018)
25	SulfurAssimilation_user{Inh(CysH)}(130.0020)
High-throughput screening (Gene expression) – 25 assays	
1	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0001)
2	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0003)
3	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0007)
4	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0009)
5	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0010)
6	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0012)
7	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0014)
8	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0016)
9	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0017)
10	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0019)
11	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0021)
12	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0028)
13	AndrogenSignalingGE-HTS_GeneExprHTS(1004.0030)
14	GE-HTSApoptosis_GeneExprHTS(1055.0005)
15	GE-HTSApoptosis_GeneExprHTS(1055.0015)
16	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0083)
17	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0089)
18	GE-HTSNotch1Inhibition_GeneExprHTS(1131.0092)
19	MetabolismCellProfiling_GeneExprHTS(1020.0071)
20	MetabolismCellProfiling_GeneExprHTS(1020.0078)
21	NeuroblastomaDifferentiation_GeneExprHTS(1149.0007)

22	NeuroblastomaDifferentiation_GeneExprHTS(1149.0016)
23	NeuroblastomaDifferentiation_GeneExprHTS(1149.0039)
24	NeuroblastomaDifferentiation_GeneExprHTS(1149.0046)
25	NeuroblastomaDifferentiation_GeneExprHTS(1149.0047)
High-content screening (Area) – 8 assays	
1	Autophagy_AvVesicleArea(1050.0019)
2	Autophagy_AvVesicleArea(1050.0064)
3	Autophagy_AvVesicleArea(1050.0077)
4	Autophagy_AvVesicleArea(1050.0116)
5	Autophagy_VesicleAreaPerCell(1050.0016)
6	Autophagy_VesicleAreaPerCell(1050.0079)
7	Autophagy_VesicleTotalArea(1050.0015)
8	Autophagy_VesicleTotalArea(1050.0084)
High-content screening (Intensity) – 18 assays	
1	Autophagy_EGFPVes7(1050.0076)
2	Autophagy_VesicleAvInt(1050.0018)
3	Autophagy_VesicleAvInt(1050.0067)
4	Autophagy_VesicleAvInt(1050.0080)
5	Autophagy_VesicleInt(1050.0017)
6	Autophagy_VesicleInt(1050.0083)
7	BetaCatenin_AvgIn(1152.0003)
8	BetaCatenin_AvgIn(1152.0007)
9	BetaCatenin_PosCellsW2_Int(1152.0015)
10	BetaCatenin_PosCellsW2_Int(1152.0019)
11	BetaCatenin_W2AvgIntPosNuc(1152.0002)
12	DNADamageImagingScreen_pChk1AvNucInt(1037.0008)
13	DNADamageImagingScreen_pChk1AvNucInt(1037.0010)
14	KLF2Regulators_KLF2-GFP_AvInt(1085.0006)
15	KLF2Regulators_KLF2-GFP_Int(1085.0005)
16	ProteinDegradationInhibition_MODCInt(1053.0006)

17	ProteinDegradationInhibition_MODCInt(1053.0036)
18	ProteinDegradationInhibition_MODCInt(1053.0044)
High-content screening (Cell count) – 24 assays	
1	Autophagy_CellCount(1050.0007)
2	Autophagy_CellCount(1050.0065)
3	Autophagy_CellCount(1050.0078)
4	Autophagy_CellCount(1050.0117)
5	Autophagy_CellCountSupra10Punctae(1050.0010)
6	Autophagy_CellCountSupra15Punctae(1050.0011)
7	Autophagy_CellCountSupra20Punctae(1050.0012)
8	Autophagy_CellCountSupra5Punctae(1050.0008)
9	Autophagy_CellCountSupra7Punctae(1050.0009)
10	Autophagy_VesicleCount(1050.0014)
11	Autophagy_VesicleCount(1050.0081)
12	Autophagy_VesicleCountPerCell(1050.0013)
13	Autophagy_VesicleCountPerCell(1050.0082)
14	KLF2Regulators_CellCount(1085.0003)
15	StemCellChemicalBiology_CellCount(1032.0652)
16	StemCellChemicalBiology_CellCount(1032.0657)
17	StemCellChemicalBiology_CellCount(1032.0660)
18	StemCellChemicalBiology_CellCount(1032.0663)
19	StemCellChemicalBiology_CellCount(1032.0666)
20	StemCellChemicalBiology_LiveCells(1032.0653)
21	StemCellChemicalBiology_LiveCellsPerCellCount(1032.0654)
22	StemCellChemicalBiology_Sox17PosPerCellCount(1032.0659)
23	StemCellChemicalBiology_Sox17PosPerCellCount(1032.0662)
24	StemCellChemicalBiology_Sox17PosPerCellCount(1032.0668)
Small-molecule microarrays – 25 assays	
1	Abeta40SMM_AutoSNR(1115.0001)
2	Abeta40SMM_AutoSNR(1115.0002)

3	Abeta40SMM_AutoSNR(1115.0003)
4	Abeta40SMM_AutoSNR(1115.0004)
5	CFTRSMM_ManualSNR(1098.0001)
6	CFTRSMM_ManualSNR(1098.0002)
7	CFTRSMM_ManualSNR(1098.0004)
8	CFTRSMM_ManualSNR(1098.0005)
9	CFTRSMM_ManualSNR(1098.0006)
10	DHODHSMM_ManualSNR(1089.0001)
11	EBNA1SMM_ManualSNR(1159.0001)
12	HIV-1NefSMM_ManualSNR(1150.0001)
13	HIV-1NefSMM_ManualSNR(1150.0002)
14	HPVE7SMM_ManualSNR(1049.0001)
15	LRP130_ManualSNR(1140.0001)
16	LRP130_ManualSNR(1140.0002)
17	LRP130_ManualSNR(1140.0003)
18	LRP130_ManualSNR(1140.0007)
19	MaleGermCellSMM_ManualSNR(1154.0009)
20	MaleGermCellSMM_ManualSNR(1154.0015)
21	NeuroSMM_ManualSNR(1069.0001)
22	PETLigandSMM_ManualSNR(1153.0001)
23	PETLigandSMM_ManualSNR(1153.0012)
24	SMMDIV06Annotation_AutoSNR(1066.0013)
25	Transcription Factor Profile_AutoSNR(1125.0075)

CONCLUSION

High-Throughput Screening (HTS) and High Content Screening (HCS) are popular screening technologies which are widely used in the modern pharmaceutical industry. HTS/HCS technologies allow scientists to rapidly assess biological activity of a large number of candidate samples in order to detect a small number of active features (for example, small molecules or small interfering RNAs). HTS/HCS is usually the first step in modern drug discovery campaigns. This process is applied to screen a large number of chemical compounds. The samples that satisfy particular activity criteria are detected and retained for further testing. We should mention that the HTS/HCS technologies may also be used to study fundamental biochemical processes.

Today HTS/HCS assays are applicable in both the pharmaceutical and academic contexts. This thesis research is at the interface between the computational, statistical, and life sciences. The impact of systematic errors (i.e., spatial biases) on the results of the hit identification process, an important problem in the quantification of HTS/HCS assays (Malo et al. 2006, Makarenkov et al. 2007), was examined in this work. New methods for systematic error detection and elimination were presented and tested in simulations. The current knowledge of how systematic error affects HTS/HCS raw data was reviewed. We described the different types and causes of systematic errors typical for HTS and HCS technologies, as well as the existing normalization and bias correction methods. We adapted two well-known statistical tests, i.e., Welch's *t*-test (Welch 1947) and the Mann-Whitney *U* test (Wilcoxon 1945), for spatial bias detection in experimental HTS/HCS assays. We should mention that the spatial bias elimination methods should be used only if the presence of this bias has been demonstrated by statistical tests. We presented a comprehensive data pre-processing and correction protocol intended for experimental HTS/HCS assays. This protocol is of a general nature. Thus, the academic and industrial scientists involved in the

analysis of current or next generation high-throughput screening data can use our protocol in their screening experiences. It is important to note that most of the existing bias correction methods are designed to minimize the impact of the additive type of spatial bias (Dragiev et al. 2012, Caraus et al. 2017). It is also known that HTS RNAi primary screens can be affected by multiplicative spatial bias (Carralot et al. 2012). However, the problem of eliminating multiplicative spatial bias has not been studied in detail. In this thesis, the ability to detect the presence of multiplicative spatial bias in HTS/HCS data was investigated by identifying the bias-affected rows and columns of all plates of a given assay, as well as by determining the bias-affected well locations (i.e., well positions scanned across all plates of a given assay). The possibility to detect the additive, multiplicative or mixed (i.e., including both additive and multiplicative bias interactions) types of spatial biases was studied in different practical situations. We first presented three new algorithms for correcting the multiplicative type of spatial bias. Our algorithms correct the values situated in plates' rows and columns containing multiplicative bias without modifying the rest of the (unbiased) data. We showed that our new algorithms can efficiently eliminate multiplicative bias from experimental HTS/HCS assays. We implemented the new algorithms and the related data correction protocol, allowing for removing both plate and assay-specific biases as well as both additive and multiplicative biases, in the *AssayCorrector* R package (the URL address of the package is: https://cran.r-project.org/web/packages/AssayCorrector/index.html). The proposed protocol can be readily used by researches working in HTS/HCS technologies. Furthermore, we studied the performances of the Anderson-Darling test (Anderson 1952) and the Cramer-von-Mises test (Cramér 1928) in terms of distinguishing between the additive and multiplicative spatial bias models.

In Chapter I we presented the different types of systematic error common for HTS/HCS technologies such as batch effect, edge effect and spatial biases specific for a given screening technology (e.g., non-uniformity bias). Using experimental HTS

data (Harvard's 164-plate assay; Helm et al. 2003), we presented an example of edge effect. To eliminate the presented edge-effect, we applied the well-known B-score method (Brideau et al. 2003). Similarly, an example of non-uniformity bias for a (96-well x 4-field) HCS assay was given. We then described simple data normalization techniques used in HTS/HCS (Malo et al. 2006). These methods allow for making the results comparable over all plates of an HTS/HCS assay. We also discussed some useful tests for systematic error detection in experimental HTS/HCS campaigns, including Welch's *t*-test (Welch 1947), which can be applied in the case of samples with different sizes and different variances, and the Mann-Whitney *U* test (Wilcoxon 1945). We examined the presence of systematic error in experimental HTS data from the popular ChemBank database (Seiler et al. 2008); in total 735 plates aimed at the inhibition of *Escherichia coli* were analyzed. One of the goals our study was to assess the proportion of rows and columns affected by systematic error in a large variety of publically available HTS assays. The Welch's *t*-test with five different values of the α parameter was carried out (i.e., with α = 0.01, 0.025, 0.05, 0.075 and 0.1). We calculated the average row and column systematic error rates. Our experiments indicated that the row and column systematic error was very frequent in HTS data (i.e., at least 30% of rows and columns in the raw data and 20% of rows and columns in the background-subtracted data were affected by systematic bias, see Figure 2.4). In addition, we studied how systematic error influences the hit selection process in HTS. We found that at least 50% of raw hit distribution surfaces and 65% of background-subtracted hit distribution surfaces were affected by systematic error, making it difficult to identify true hits. In the second chapter we also highlighted the advantage of sample randomization and underlined the importance of using replicates in experimental HTS/HCS campaigns. Furthermore, we provided guidance concerning which normalization and/or correction methods should used in different experimental contexts. Finally, we presented a broad spectrum data preprocessing protocol. This general protocol can be successfully applied for detecting and

eliminating spatial biases in experimental HTS/HCS data before the hit identification process.

In Chapter II we presented three novel algorithms for multiplicative bias elimination in HTS/HCS technologies. These algorithms are particularly well suited for RNAi primary screens, in which the number of active features is typically very small and the spatial bias is multiplicative (Carralot et al. 2012). The presence of bias was identified and visualized using the hit distribution surface of the assay (Makarenkov et al. 2007). Such a surface can be computed by determining the number of selected active features (i.e., hits) for each well location. Here, we applied the Mann-Whitney *U* test to identify rows and columns of a given plate or of a calculated hit distribution surface that are affected by systematic error (Caraus et al. 2017). We compared our new data correction algorithms to the basic No Correction procedure using a well-defined simulation protocol. The first algorithm, called Non-Linear Multiplicative Bias Elimination (NLMBE), solves a system of nonlinear equations in which the unknowns correspond to the systematic biases affecting rows and columns of a given plate. We applied the Levenberg-Marquardt method (Moré 1978) to solve this system of nonlinear equations. The second algorithm, called Multiplicative PMP (mPMP), is based on an iterative partial mean polish procedure in which the biased plate measurements of a given plate are iteratively adjusted (row and column-wise) using the mean of the unbiased samples of the same plate. It is worth noting that if a large number of hits or outliers are expected for a given plate (i.e., > 10% of the data), then the means of the plate's measurements should be replaced by medians in order to obtain more robust parameter estimates. Our third algorithm, called Multiplicative B-scores, uses the same principle that the traditional B-score method (Brideau et al. 2003). This new technique is based on a 2-way median polish procedure in which the subtractions are replaced by the divisions. To estimate the performance of the proposed algorithms, we conducted simulations with artificially generated screening data. We randomly generated 1000 HTS assays for each of the following plate sizes:

96-, 384-, and 1536-well plates, adding to them hits (different hit percentages were generated) and spatial biases (biases of different amplitude were considered). The true positive rate and the combined false positive and false negative rates were calculated. The results of our simulations suggest that the NLMBE, Multiplicative PMP and Multiplicative B-score algorithms clearly outperformed the No Correction method in all cases. The results presented in Figures 2.2-2.5 suggest that the NLMBE and the Multiplicative PMP algorithms were the best methods for correcting the multiplicative type of spatial bias. We recommend using the Multiplicative PMP algorithm for correcting multiplicative spatial bias because it converges much faster than NLMBE while providing similar data correction results.

Moreover, a general data correction protocol was elaborated in the third chapter. This protocol can be used to eliminate both additive and multiplicative systematic errors; it is publicly available in the *AssayCorrector* package. This protocol allows one to detect *plate-specific spatial biases* by identifying the bias-affected rows and columns within all plates of the assay (following the results of the Mann-Whitney *U* test applied row and column-wise, respectively) as well as *assay-specific spatial biases* by identifying the bias-affected well locations (following the results of the Mann-Whitney *U* test applied well-wise). We propose to correct plate-specific spatial bias using either the Additive or Multiplicative PMP (Partial Mean Polish) algorithm (the most appropriate spatial bias model and the corresponding algorithm can be either specified by the user or determined by the program following the results of the Kolmogorov-Smirnov two-sample test, the Anderson-Darling test or the Cramer-von-Mises). The assay-specific spatial bias can be corrected by carrying out robust Z-scores within each plate of the assay and then traditional Z-scores across all of its well locations. The usefulness of our new algorithms and protocol were confirmed by their application to publicly available data from ChemBank (Seiler et al. 2008) . Summing up the results, we can conclude that the additive type of spatial bias is

prevalent in homogeneous and microorganism HTS screens, whereas multiplicative bias is usually dominant in cell-based and gene-expression HTS assays.

In Chapter III, four new spatial bias elimination models accounting for different types of bias interactions were presented. A new statistical procedure using the Anderson-Darling and Cramer-von-Mises goodness-of-fit tests has been described. This procedure can be carried out by HTS/HCS researchers to determine the most effective spatial bias models for a given plate. We showed how this procedure can be used in practice by examining data generated by: the four high-throughput screening technologies (homogeneous, microorganism, cell-based and gene expression HTS), the three high-content screening technologies (area, intensity and cell-count HCS) and a single small-molecule microarray technology, available in ChemBank.

The methodology presented in this thesis is designed to minimize the impact of both plate-specific (additive or multiplicative) and assay-specific spatial biases. A plate-specific bias means that the observed bias patterns appear within a given plate only and may be different for different plates of the assay. An assay-specific bias consists of a bias pattern that appears within all plates of a given assay. A typical example of an assay-specific bias is the case in which a single well location (i.e., measurements taken across all plates of a given assay and corresponding to a fixed well position) gives a very high or a very low reading. An assay-specific bias can be removed by applying the traditional Z-score normalization plate-wise, first, and well location-wise, second. The traditional Z-score normalization allows one to remove both additive and multiplicative biases when applied to the measurements of a given well location. Obviously, the presence of spatial bias in this well location should be primarily confirmed by an appropriate statistical test (e.g., by the Mann-Whitney U test).

The main advantage of our method presented in Chapter III, compared to the No Correction strategy and the B-score method, is that it copes well with both additive and multiplicative spatial biases by taking into account complex interactions between them. This was confirmed by our simulation study. Obviously, any bias correction method has to be applied with caution. For instance, a high-throughput screening practitioner should always verify that all tested samples are randomly distributed within given HTS/HCS/SMM plates. If the randomization condition does not hold, some areas of these plates can correctly give very high or very low readings. In this case, the application of bias correction methods may be rather damaging as an unwanted bias may be introduced into the data at hand.

The source code of the main methods developed in the framework of this doctoral project is presented in the Appendix section of this thesis. Overall, the new statistical methods presented in this thesis contribute to the reduction of systematic error in HTS/HCS screens as well as to quality amelioration of the results of the HTS/HCS hit selection procedures. It is very important to note that the presented bias correction methods should be applied judiciously, i.e., after the presence of spatial bias was confirmed by statistical tests, because the application of these techniques (e.g., of B-score) on error-free data or on data including another type of bias (i.e., when the multiplicative bias is corrected using additive bias correction methods) can introduce a supplementary bias into data in hand.

In Chapters II and III of this thesis we considered new models for correcting either the additive or multiplicative type of spatial bias in HTS/HCS assays. Moreover, the new models presented in Chapter III assume different types of bias interaction (additive or multiplicative) on the intersections of the biased rows and columns of a given plate. In the future, it would be interesting to consider a more general model that assumes different values, and thus uses different variables, to account for the influence of additive and multiplicative biases. This general model, in which the

additive and multiplicative biases are completely independent, could be described by the following non-linear equation: $\hat{x}_{ijp} = x_{ijp} \times R_{ip}^{M} \times C_{jp}^{M} + R_{ip}^{A} + C_{jp}^{A} + \varepsilon_{ijp}$, where $\hat{x}_{ijp}$ is the resulting (biased) measurement value in well (i,j) of plate p, x_{ijp} is the original error-free measurement, R_{ip}^{M} is the multiplicative bias affecting row i of plate p, C_{jp}^{M} is the multiplicative bias affecting column j of plate p, R_{ip}^{A} is the additive bias affecting row i of plate p, C_{jp}^{A} is the additive bias affecting column j of plate p and ε_{ijp} is the random error in well (i,j) of plate p. This model leads to a particular system of non-linear equation for each plate of the assay. Some additional work should be done for elaborating a new efficient algorithm for determining and then removing both additive and multiplicative spatial biases defined by this general model.

Our new methods have been designed to reduce multiplicative and additive spatial biases which are characteristic for HTS/HCS screens. However, these approaches are not powerful enough to completely eliminate errors in experimental HTS/HCS technologies, e.g., in siRNA screens. One of the main limitations of siRNA screens is a high rate of off-target effects (Birmingham et al. 2009). These effects can appear due to partial complementarity between the targeted siRNA and random mRNA transcripts of non-targeted genes. It is well known that many false positive hits appear due to off-target effects. Thus, future developments could be focused on the elaboration of new statistical methods for predicting, detecting and reducing off-target effects in siRNA screens and on a possible interaction between these effects and systematic biases.

APPENDIX A

SOURCE CODES

I have implemented three new error correction methods to eliminate the multiplicative bias:

- Multiplicative B-score method,
- Non-Linear Multiplicative Bias Elimination method,
- Multiplicative PMP method.

The Multiplicative B-score method is an extension of the well-known (additive) B-score algorithm (Brideau et al. 2003). This method is used to eliminate plate-specific multiplicative spatial biases by reducing the original data to residuals. In order to detect the locations of rows and columns affected by multiplicative spatial bias we have implemented Mann-Whitney *U* test. The results of the Mann-Whitney *U* test are transferred to both the Non-Linear Multiplicative Bias Elimination and Multiplicative PMP methods. The presented bias correction methods and the Mann-Whitney *U* test were used to produce the simulation and real data experimental results for my second article (Caraus et al. 2017, *Bioinformatics*).

APPENDIX B

MANN-WHITNEY U TEST

```
//This C# function performs the Mann Whitney U test for all rows
//and columns of a given plate in order to detect
//the biased rows and columns within it
protected bool MannWhitneyUtest (Prj2Task T, Plate PL, int p, List<List<int>>
    MTestRows, List<List<int>> MTestColumns, ref TestResult MTest, LogFile
    PLLog)
  {
    bool Res = false;
    int Rows = PL.Rows();
    int Columns = PL.Columns();
    int i, j;
    List<int> MRows = new List<int>();
    List<int> MCols = new List<int>();
    int RowsTP = 0;
    int RowsFP = 0;
    int RowsTN = 0;
    int RowsFN = 0;
    int ColsTP = 0;
```

```
int ColsFP = 0;
int ColsTN = 0;
int ColsFN = 0;
StringBuilder SB = null;
if (PLLog != null)
{
   PLLog.WriteLine();
   SB = new StringBuilder();
}

// Mann-Whitney U test by rows
double CV1, CV2;
double lefttail;
double righttail;
List<double> RowItems;
List<double> NotRowItems;
for (i = 0; i < Rows; i++)
{
   RowItems = new List<double>();
   NotRowItems = new List<double>();

   int k;
```

```
for (k = 0; k < Rows; k++)
{
  if (k == i)
  {
    for (j = 0; j < Columns; j++)
      RowItems.Add(PL[k, j]);
  }
  else
  {
    for (j = 0; j < Columns; j++) NotRowItems.Add(PL[k, j]);
  }
}
double[] Row = RowItems.Cast<double>().ToArray();
double[] NotRow = NotRowItems.Cast<double>().ToArray();
alglib.mannwhitneyutest(Row, Row.Length, NotRow, NotRow.Length, out
CV1, out lefttail, out righttail);
bool RealError = T.IsTrueRowError(p, i);

if (T.Alpha <= CV1)
{
  // true - no error
  if (RealError)
```

```
            RowsFN++;
        else
            RowsTN++;
    }
    else
    {
        Res = true;
        // false - there's error detected by the test
        MRows.Add(i);
        if (RealError)
            RowsTP++;
        else
            RowsFP++;
        if (PLLog != null)
        {
            if (MRows.Count > 1) SB.Append(", ");
            SB.Append(i.ToString());
            SB.Append("(" + (RealError ? "TRUE" : "FALSE") + ")");
        }
    }
}
MTestRows.Add(TRows);
```

```
MTest.FN += RowsFN;
MTest.FP += RowsFP;
MTest.TN += RowsTN;
MTest.TP += RowsTP;
if (PLLog != null)
{
   SB.AppendFormat("     [TP:{0}  FP:{1}  FN:{2}  TN:{3}]",  RowsTP,
RowsFP, RowsFN, RowsTN);
   PLLog.WriteLine("Error detected in " + (RowsTP + RowsFP).ToString() +
" rows: " + SB.ToString());
   SB.Length = 0;
}


// Mann-Whitney U test by columns
List<double> ColumnItems;
List<double> NotColumnItems;
for (j = 0; j < Columns; j++)
{
   ColumnItems = new List<double>();
   NotColumnItems = new List<double>();
   int k;
```

```
for (k = 0; k < Columns; k++)
{
  if (k == j)
  {
    for (i = 0; i < Rows; i++) ColumnItems.Add(PL[i, k]);
  }
  else
  {
    for (i = 0; i < Rows; i++) NotColumnItems.Add(PL[i, k]);
  }
}
double[] Col = ColumnItems.Cast<double>().ToArray();
double[] NotCol = NotColumnItems.Cast<double>().ToArray();
alglib.mannwhitneyutest(Col, Col.Length, NotCol, NotCol.Length, out
CV2, out lefttail, out righttail);
bool RealError = T.IsTrueColumnError(p, j);
if (T.Alpha <= CV2)
{
  // true - no error
  if (RealError)
    ColsFN++;
  else
```

```
            ColsTN++;
        }
        else
        {
            // false - there's error detected by the test
            Res = true;
            MCols.Add(j);
            if (RealError)
                ColsTP++;
            else
                ColsFP++;
            if (PLLog != null)
            {
                if (MCols.Count > 1) SB.Append("  ");
                SB.Append(j.ToString());
                SB.Append("(" + (RealError ? "TRUE" : "FALSE") + ")");
            }
        }

    }
    if (PLLog != null)
    {
```

```
    SB.AppendFormat("  [TP:{0} FP:{1} FN:{2} TN:{3}]", ColsTP, ColsFP,
  ColsFN, ColsTN);
    PLLog.WriteLine("Error detected in " + (ColsTP + ColsFP).ToString() + "
  columns: " + SB.ToString());
    PLLog.WriteLine();
    PLLog.Flush();
  }
  MTestColumns.Add(MCols);
  MTest.FN += ColsFN;
  MTest.FP += ColsFP;
  MTest.TN += ColsTN;
  MTest.TP += ColsTP;
  return Res;
}
```

APPENDIX C

MULTIPLICATIVE B-SCORE METHOD

```
//This C# function performs the Multiplicative B-score method
//applied to all measurements of a given plate
public void BScorePlateMulti(Prj2Task T, int p, bool Print)
    {
      double BSCORE_EPSILON = 0.05;
      double BSCORE_EPS_PERC = 0.01;
      int BSCORE_MAX_ITERATIONS = 50;
      bool PrintIter = false;
      Plate PL = T.BScoreDS[p];
      int MaxIterations = BSCORE_MAX_ITERATIONS;
      int Rows = PL.wells.GetLength(0);
      int Columns = PL.wells.GetLength(1);
      double[] MRow = new double[Rows];
      double[] MCol = new double[Columns];
      double[] R = new double[Rows];
      double[] C = new double[Columns];
      Array.Clear(R, 0, Rows);
      Array.Clear(C, 0, Columns);
```

```
int i, j, k;
int Iteration = BSCORE_MAX_ITERATIONS;
double OldSum = 0.0;
bool converge = false;

do
{
   Array.Clear(R, 0, Rows);
   Array.Clear(C, 0, Columns);

   // Computing the median of the rows
   for (i = 0; i < Rows; i++)
      R[i] += MRow[i] = G.MedianInPlace(PL.Row(i));

   for (i = 0; i < Rows; i++)
      for (j = 0; j < Columns; j++)
         PL.wells[i, j] /= MRow[i];

   double RMed = G.Median(MRow);
   for (i = 0; i < Rows; i++) R[i] /= RMed;

   // Computing the median of the columns
```

```
    for (j = 0; j < Columns; j++)
       C[j] += MCol[j] = G.MedianInPlace(PL.Column(j));
    double WellSum = 0.0;

    for (i = 0; i < Rows; i++)
       for (j = 0; j < Columns; j++)
       {
          PL.wells[i, j] /= MCol[j];
          WellSum += Math.Abs(PL.wells[i, j]);
       }
    double CMed = G.Median(MCol);

    for (j = 0; j < Columns; j++)
         C[j] /= CMed;

    converge = (Math.Abs(WellSum - Rows * Columns) <
                  BSCORE_EPSILON) ||
                 (Math.Abs(WellSum - OldSum) <
                  BSCORE_EPS_PERC*WellSum);
    OldSum = WellSum;
  }
 while (--Iteration > 0 && !converge);
```

```
    double[] Resid = new double[Rows * Columns];
    for (k = i = 0; i < Rows; i++)
       for (j = 0; j < Columns; j++)
          Resid[k++] = PL.wells[i, j];

    double ResMed = G.MedianInPlace(Resid);
    for (i = 0; i < Resid.Length; i++)
             Resid[i] = Math.Abs(Resid[i] - ResMed);
    double MAD = G.MedianInPlace(Resid);

    // the following has been added for compatibility with HTS Corrector
    MAD *= 1.4826;
    if (MAD > 0.0001)
       for (i = 0; i < Rows; i++)
          for (j = 0; j < Columns; j++)
             PL.wells[i, j] /= MAD;
}
```

APPENDIX D

NON-LINEAR MULTIPLICATIVE BIAS ELIMINATION

```
//This C# function performs the Non-Linear Multiplicative Bias Elimination
// method applied only to the bias-affected rows and columns of a given plate
    public void NLMBE(Prj2Task TT, Dataset DS, string Label, int p,
                List<int> ERows, List<int> EColumns, bool Print)
    {
      Plate PL = DS[p];
      int Rows = TT.Rows;
      int Columns = TT.Columns;
      V my_object = new V();
      my_object.ERows = ERows;
      my_object.wells = PL;
      my_object.EColumns = EColumns;
      my_object.Columns = Columns;
      my_object.Rows = Rows;
      int NR = ERows != null ? ERows.Count : 0;
      int NC = EColumns != null ? EColumns.Count : 0;
      int N = NR + NC;

      // Is there any row or column affected by systematic error?
```

```
if (N == 0) return;
bool[] RFlag = new bool[Rows];
bool[] CFlag = new bool[Columns];
Array.Clear(RFlag, 0, Rows);
Array.Clear(CFlag, 0, Columns);
foreach (int r in ERows) RFlag[r] = true;
foreach (int c in EColumns) CFlag[c] = true;
double Mu = 0.0;
int i, j;

for (i = 0; i < Rows; i++)
{
   if (RFlag[i]) continue;
   for (j = 0; j < Columns; j++)
   {
      if (CFlag[j]) continue;
      Mu += PL[i, j];
   }
}
Mu /= (Rows - NR) * (Columns - NC);
double[] RMu = new double[Rows];
double[] CMu = new double[Columns];
```

```
Array.Clear(RMu, 0, Rows);
Array.Clear(CMu, 0, Columns);

//Computing the means of the rows to be corrected
for (i = 0; i < Rows; i++)
{
   if (!RFlag[i]) continue;
   for (j = 0; j < Columns; j++)
   {
      RMu[i] += PL[i, j];
   }
   RMu[i] /= Columns;
}

//Computing the means of the columns to be corrected
for (j = 0; j < Columns; j++)
{
   if (!CFlag[j]) continue;
   for (i = 0; i < Rows; i++)
   {
      CMu[j] += PL[i, j];
   }
```

```
    CMu[j] /= Rows;
}

  //eliminates row and column effects
if ((NR != 0) && (NC != 0))
{
    int mm = Rows + Columns;  // number of f(i) equations
    double[] x = new double[mm]; //initial values
    for (i = 0; i < mm; i++)
       x[i] = 1;
    x[0] = 2;
    x[1] = 2;
    double diffstep = 0.0001;
    double epsg = 0.0001;
    double epsf = 0;
    double epsx = 0;
    int maxits = 0; // unlimited iterations

  //Nonlinear least-squares optimization using function vector only
  //does not require to provide the partial derivatives.
   alglib.minlmstate state;
   alglib.minlmreport rep;
```

```
                // create optimizer
alglib.minlmcreatev(mm, x, diffstep, out state);
                //conditions
alglib.minlmsetcond(state, epsg, epsf, epsx, maxits);
                //optimization
alglib.minlmoptimize(state, function1_fvec, null, my_object);
                //Optimization result
alglib.minlmresults(state, out x, out rep);
System.Console.WriteLine("{0}", rep.terminationtype);
System.Console.WriteLine("{0}", alglib.ap.format(x, 5));
for (i = 0; i < NR; i++)
{
   for (j = 0; j < Columns; j++)
   {
      PL[ERows[i], j] = PL[ERows[i], j] * x[i];
   }
}
for (j = 0; j < NC; j++)
{
   for (i = 0; i < Rows; i++)
   {
      PL[i, EColumns[j]] = PL[i, EColumns[j]] * x[j + NR];
```

```
      }
   }
} //eliminate row effects
else if (NR != 0 && NC == 0)
{

   for (i = 0; i < Rows; i++)
   {
      for (j = 0; j < Columns; j++)
      {
         if (RFlag[i])
         {
            PL[i, j] = Mu * PL[i, j] / RMu[i];
         }
      }
   }
} //eliminate column effects
else if (NR == 0 && NC != 0)
{

   for (j = 0; j < Columns; j++)
   {
```

```
            for (i = 0; i < Rows; i++)
            {
               if (CFlag[j])
               {
                  PL[i, j] = Mu * PL[i, j] / CMu[j];
               }
            }
         }

      }
   }

   //general form functions
   public void function1_fvec(double[] x, double[] fi, object obj)
   {
      int i, j;
      List<int> ERows = new List<int>();
      List<int> EColumns = new List<int>();
      V my_object = (V)obj;
      ERows = my_object.ERows;
      EColumns = my_object.EColumns;
```

```
int Columns = my_object.Columns;
int Rows = my_object.Rows;
double[,] Wells = new double[Rows, Columns];

for (i = 0; i < Rows; i++)
{
   for (j = 0; j < Columns; j++)
   {
      Wells[i, j] = my_object.wells[i, j];
   }
}
int NR = ERows != null ? ERows.Count : 0;
int NC = EColumns != null ? EColumns.Count : 0;
int N = NR + NC;
if (N == 0) return;
bool[] RFlag = new bool[Rows];
bool[] CFlag = new bool[Columns];
Array.Clear(RFlag, 0, Rows);
Array.Clear(CFlag, 0, Columns);
foreach (int r in ERows) RFlag[r] = true;
foreach (int c in EColumns) CFlag[c] = true;
double Mu = 0;
```

```
for (i = 0; i < Rows; i++)
{
   if (RFlag[i]) continue;
   for (j = 0; j < Columns; j++)
   {
      if (CFlag[j]) continue;
      Mu += Wells[i, j];
   }
}

Mu /= (Rows - NR) * (Columns - NC);

double[] f_row = new double[Rows];
for (i = 0; i < Rows; i++)
   f_row[i] = Mu * Columns;

double[] f_col = new double[Columns];
for (i = 0; i < Columns; i++)
   f_col[i] = Mu * Rows;

double[] rr = new double[Rows];
for (i = 0; i < Rows; i++)
```

```
    rr[i] = 1;
for (i = 0; i < NR; i++)
    rr[ERows[i]] = x[i];

double[] cc = new double[Columns];

for (i = 0; i < Columns; i++)
    cc[i] = 1;
for (i = 0; i < NC; i++)
    cc[EColumns[i]] = x[i + NR];
double[,] Rdiag = new double[Rows, Rows];
Rdiag = diag(rr);
double[,] Cdiag = new double[Columns, Columns];
Cdiag = diag(cc);

double[] rb = new double[Rows];
alglib.rmatrixmv(Rows, Columns, Wells, 0, 0, 0, cc, 0, ref rb, 0);
double[] rw = new double[Rows];
alglib.rmatrixmv(Rows, Rows, Rdiag, 0, 0, 0, rb, 0, ref rw, 0);
for (i = 0; i < Rows; i++)
    fi[i] = rw[i] - f_row[i];
double[,] BTransp = new double[Columns, Rows];
```

```
        alglib.rmatrixtranspose(Rows, Columns, Wells, 0, 0, ref BTransp, 0, 0);

        double[] cb = new double[Columns];
        alglib.rmatrixmv(Columns, Rows, BTransp, 0, 0, 0, rr, 0, ref cb, 0);

        double[] rc = new double[Columns];


        alglib.rmatrixmv(Columns, Columns, Cdiag, 0, 0, 0, cb, 0, ref rc, 0);
        for (i = 0; i < Columns; i++)
            fi[i + Rows] = rc[i] - f_col[i];

    }

    public class V
    {
        /*Global Variables*/
        public List<int> ERows;
        public List<int> EColumns;
        public Plate wells;
        public int Columns;
```

```
    public int Rows;
}
//function calculate the diagonal matrix
public double[,] diag(double[] C)
{
    double[,] CMu = new double[C.Length, C.Length];

    Array.Clear(CMu, 0, C.Length);
    for (int i = 0; i < C.Length; i++)
        CMu[i, i] = C[i];

    return CMu;
}
```

APPENDIX E

MULTIPLICATIVE PMP METHOD

```
//This C# function performs the Multiplicative PMP method
//applied only to the bias-affected rows and columns of a given plate
    public void mPMP(Prj2Task TT, Dataset DS, string Label, int p, List<int>
      ERows, List<int> EColumns, bool Print)
    {
      int Rows = TT.Rows;
      int Columns = TT.Columns;
      int NR = ERows.Count;
      int NC = EColumns.Count;
      int N = NR + NC;
      Plate PL = DS[p];
      int PMP_MAX_ITERATIONS = 200;
      double PMP_EPSILON = 0.001;
      // Is there any row column affected by systematic error?
      if (N == 0) return;
      bool[] RFlag = new bool[Rows];
      bool[] CFlag = new bool[Columns];
      Array.Clear(RFlag, 0, Rows);
      Array.Clear(CFlag, 0, Columns);
```

```
foreach (int r in ERows) RFlag[r] = true;
foreach (int c in EColumns) CFlag[c] = true;
double Mu = 0.0;
int i, j;
double[] Non_Error_Matrix = new double[(Rows-NR)*(Columns-NC)];
Array.Clear(Non_Error_Matrix, 0, (Rows-NR) * (Columns-NC));

int k = 0;
for (i = 0; i < Rows; i++)
{
   if (RFlag[i]) continue;
   for (j = 0; j < Columns; j++)
   {
      if (CFlag[j]) continue;
      Non_Error_Matrix[k] = PL[i, j];k++;
                }
}
alglib.samplemedian(Non_Error_Matrix, out Mu);

//Computing the median of the Matrix
double[] RMu = new double[Rows];
```

```
double[] CMu = new double[Columns];
double[] R_copy = new double[Rows];
double[] C_copy = new double[Columns];
Array.Clear(R_copy, 0, Rows);
Array.Clear(C_copy, 0, Columns);
Array.Clear(RMu, 0, Rows);
Array.Clear(CMu, 0, Columns);
int Loop = 1;
double Converge = 0.0;
double Diff = 0.0;
do
{
   Diff = 0.0;
   Converge = 0.0;
      //Computing the median of the rows to be corrected
      for (i = 0; i < Rows; i++)
      {
         if (RFlag[i])
         {
            Array.Clear(C_copy, 0, Columns);

            RMu[i]=G.MedianInPlace(PL.Row(i));
```

```
      }
    }
   //Computing the median of the columns to be corrected
  for (j = 0; j < Columns; j++)
  {
    k = 0;
    if (CFlag[j])
    {
      CMu[j] = G.MedianInPlace(PL.Column(j));
    }
  }

  //elimination of the row effect
  for (i = 0; i < Rows; i++)
  {
    for (j = 0; j < Columns; j++)
    {
      if (RFlag[i])
      {
        Diff = Mu - RMu[i];
        Converge += Math.Abs(Diff);
        PL[i, j] = Mu * PL[i, j] / RMu[i];
```

```
            }
          }
        }

        //elimination of the column effect
        for (j = 0; j < Columns; j++)
        {
          for (i = 0; i < Rows; i++)
          {
            if (CFlag[j])
            {
              PL[i, j] = Mu * PL[i, j] / CMu[j];
              Diff = Mu - CMu[j];
              Converge += Math.Abs(Diff);
            }
          }
        }
      }
      while (Converge > PMP_EPSILON && Loop++ <
      PMP_MAX_ITERATIONS);
    }
```

REFERENCES

Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.-C., Marquez, M., Klibanov, A.M., Griffiths, A.D. et Weitz, D.A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences, 107*(9), 4004-4009.

Amberkar, S., Kiani, N.A., Bartenschlager, R., Alvisi, G. et Kaderali, L. (2013). High-throughput RNA interference screens integrative analysis: Towards a comprehensive understanding of the virus-host interplay. *World J Virol, 2*(2), 18-31.

Anderson, T.W. et Darling, D.A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193-212.

Armknecht, S., Boutros, M., Kiger, A., Nybakken, K., Mathey-Prevot, B. et Perrimon, N. (2005). High-Throughput RNA Interference Screens in Drosophila Tissue Culture Cells. Dans David, R. E. et J. R. John (dir.), *Methods in Enzymology* (Vol. Volume 392, p. 55-73) : Academic Press.

Arshad, M., Rasool, M. et Ahmad, M. (2003). Anderson Darling and modified Anderson Darling tests for generalized Pareto distribution. *Pakistan Journal of Applied Sciences, 3*(2), 85-88.

AssayCorrector. *https://cran.r-project.org/web/packages/AssayCorrector/index.html*.

Auer, P.L. et Doerge, R.W. (2010). Statistical Design and Analysis of RNA Sequencing Data. *Genetics, 185*(2), 405-416.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J. et Bandyopadhyay, S. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods, 7*(12), 1017-1024.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K. et Karpilow, J. (2006). 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature methods, 3*(3), 199-204.

Birmingham, A., Selfors, L.M., Forster, T., Wrobel, D., Kennedy, C.J., Shanks, E., Santoyo-Lopez, J., Dunican, D.J., Long, A. et Kelleher, D. (2009). Statistical methods for analysis of high-throughput RNA interference screens. *Nature methods, 6*(8), 569-575.

Blainey, P., Krzywinski, M. et Altman, N. (2014). Points of significance: replication. *Nature methods, 11*(9), 879-880.

Boutros, M. et Ahringer, J. (2008). The art and design of genetic screens: RNA interference. *Nat Rev Genet, 9*(7), 554 - 566.

Boutros, M., Brás, L.P. et Huber, W. (2006). Analysis of cell-based RNAi screens. *Genome biology, 7*(7), R66.

Box, G.E. (2006). *Improving almost anything: Ideas and essays*. (Vol. 629) : Wiley-Interscience.

Box, G.E., Hunter, J.S. et Hunter, W.G. (2005). *Statistics for experimenters: design, innovation, and discovery*. (Vol. 2) : Wiley-Interscience New York.

Bray, M.-A., Fraser, A.N., Hasaka, T.P. et Carpenter, A.E. (2012). Workflow and metrics for image quality control in large-scale high-content screens. *Journal of biomolecular screening, 17*(2), 266-274.

Breitling, R., Armengaud, P., Amtmann, A. et Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters, 573*(1-3), 83-92.

Brewer, G.J. (2009). Drug development for orphan diseases in the context of personalized medicine. *Translational Research, 154*(6), 314-322.

Brideau, C., Gunter, B., Pikounis, B. et Liaw, A. (2003). Improved statistical methods for hit selection in high-throughput screening. *Journal of biomolecular screening, 8*(6), 634-647.

Brown, C.M. (2007). Fluorescence microscopy - avoiding the pitfalls. *Journal of Cell Science, 120*(10), 1703-1705.

Brown, J.C. et Song, C. (2000). High quality cDNA libraries for discovery and validation of novel drug targets. *Emerging Therapeutic Targets, 4*(1), 113-120.

Buchser, W., Collins, M., Garyantes, T., Guha, R., Haney, S., Lemmon, V., Li, Z. et Trask, O.J. (2014). Assay development guidelines for image-based high content screening, high content analysis and high content imaging.

Buehler, E., Khan, A.A., Marine, S., Rajaram, M., Bahl, A., Burchard, J. et Ferrer, M. (2012). siRNA off-target effects in genome-wide screens identify signaling pathway members. *Scientific reports, 2*

Bushway, P.J., Azimi, B., Heynen-Genel, S., Price, J.H. et Mercola, M. (2010). Hybrid median filter background estimator for correcting distortions in microtiter plate data. *Assay and drug development technologies, 8*(2), 238-250.

Caraus, I., Alsuwailem, A. A., Nadon, R., & Makarenkov, V. (2015). Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions.*Briefings in bioinformatics*, *16*(6), 974-986.

Caraus, I., Mazoure, B., Nadon, R. et Makarenkov, V. (2017). Detecting and removing multiplicative spatial bias in high-throughput screening technologies. *Bioinformatics*

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. et Hayashizaki, Y. (2000). Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome research, 10*(10), 1617-1630.

Carralot, J.-P., Ogier, A., Boese, A., Genovesio, A., Brodin, P., Sommer, P. et Dorval, T. (2012). A novel specific edge effect correction method for RNA interference screenings. *Bioinformatics, 28*(2), 261-268.

Chan, L.L., Lidstone, E.A., Finch, K.E., Heeres, J.T., Hergenrother, P.J. et Cunningham, B.T. (2009). A method for identifying small-molecule aggregators using photonic crystal biosensor microplates. *Journal of the Association for Laboratory Automation, 14*(6), 348-359.

Chen, X., Liu, P. et Chou, H.-H. (2013). Whole-genome thermodynamic analysis reduces siRNA off-target effects. *PloS one, 8*(3), e58326.

Chiao, E., Leonard, J., Dickinson, K. et Baker, J.C. (2005). High-throughput functional screen of mouse gastrula cDNA libraries reveals new components of endoderm and mesoderm specification. *Genome Research, 15*(1), 44-53.

Cooley, J.W. et Tukey, J.W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation, 19*(90), 297-301.

Cram101. (2012). e-Study Guide for: Polymers: Chemistry and Physics of Modern Materials: Chemistry, Materials sciences. *Kindle Edition*

Cramér, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal, 1928*(1), 13-74.

D'Agostino, R.B. et Stephens, M.A. (1986). *Goodness-of-fit Technicques*. (Vol. 68) : CRC press.

Das, S., Ghosal, S., Chakrabarti, J. et Kozak, K. (2013). SeedSeq: off-target transcriptome database. *BioMed research international, 2013*

Douglas Zhang, X., Yang, X.C., Chung, N., Gates, A., Stec, E., Kunapuli, P., J Holder, D., Ferrer, M. et S Espeseth, A. (2006). Robust statistical methods for hit selection in RNA interference high-throughput screening experiments.

Dove, A. (2003). Screening for content--the evolution of high throughput. *Nature biotechnology, 21*(8), 859.

Dragiev, P., Nadon, R. et Makarenkov, V. (2011). Systematic error detection in experimental high-throughput screening. *BMC bioinformatics, 12*(1), 25.

Dragiev, P., Nadon, R. et Makarenkov, V. (2012). Two effective methods for correcting experimental high-throughput screening data. *Bioinformatics, 28*(13), 1775-1782.

Duan, Q., Flynn, C., Niepel, M., Hafner, M., Muhlich, J.L., Fernandez, N.F., Rouillard, A.D., Tan, C.M., Chen, E.Y. et Golub, T.R. (2014). LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic acids research*, gku476.

Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G. et Hacohen, N. (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nature methods, 3*(10), 777-779.

Elowe, N.H., Blanchard, J.E., Cechetto, J.D. et Brown, E.D. (2005). Experimental screening of dihydrofolate reductase yields a "test set" of 50,000 small molecules for a computational data-mining and docking competition. *Journal of biomolecular screening, 10*(7), 653-657.

Fisher, R.A. (1925). *Statistical methods for research workers.* : Genesis Publishing Pvt Ltd.

Fosså, A., Alsøe, L., Crameri, R., Funderud, S., Gaudernack, G. et Smeland, E.B. (2004). Serological cloning of cancer/testis antigens expressed in prostate cancer using cDNA phage surface display. *Cancer Immunology, Immunotherapy, 53*(5), 431-438.

Gagarin, A., Makarenkov, V. et Zentilli, P. (2006). Using clustering techniques to improve hit selection in high-throughput screening. *Journal of biomolecular screening, 11*(8), 903-914.

Gibbons, J.D. et Chakraborti, S. (2011). *Nonparametric statistical inference.* : Springer.

Gilat A. (2014). MATLAB: An Introduction with Applications, 5 edition(Solutions Manual), Wiley. 416.

Giuliano, K.A., Haskins, J.R. et Taylor, D.L. (2003). Advances in high content screening for drug discovery. *Assay and drug development technologies, 1*(4), 565-577.

Gosai, S.J., Kwak, J.H., Luke, C.J., Long, O.S., King, D.E., Kovatch, K.J., Johnston, P.A., Shun, T.Y., Lazo, J.S. et Perlmutter, D.H. (2010). Automated high-content live animal drug screening using C. elegans expressing the aggregation prone serpin α1-antitrypsin Z. *PloS one, 5*(11), e15460.

Hall, N.S. (2007). RA Fisher and his advocacy of randomization. *Journal of the History of Biology, 40*(2), 295-325.

Harmon, D.E., Davis, A.J., Castillo, C. et Mecsas, J. (2010). Identification and characterization of small-molecule inhibitors of Yop translocation in Yersinia pseudotuberculosis. *Antimicrobial agents and chemotherapy, 54*(8), 3241-3254.

Hays, W.L. (1994). Statistics Fort Worth. *TX: Harcourt Brace Jovanovich*

Helm, J.S., Hu, Y., Chen, L., Gross, B. et Walker, S. (2003). Identification of active-site inhibitors of MurG using a generalizable, high-throughput glycosyltransferase screen. *Journal of the American Chemical Society, 125*(37), 11168-11169.

Heyse, S. (2002). *Comprehensive analysis of high-throughput screening data. International Symposium on Biomedical Optics*, Actes du colloque, 2002, : International Society for Optics and Photonics.

Hill, A., LaPan, P., Li, Y. et Haney, S. (2007). Impact of image segmentation on high-content screening data quality for SK-BR-3 cells. *BMC Bioinformatics, 8*(1), 340.

Hodges Jr, J.L. et Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. *The Annals of Mathematical Statistics*, 324-335.

Honma, K., Ochiya, T., Nagahara, S., Sano, A., Yamamoto, H., Hirai, K., Aso, Y. et Terada, M. (2001). Atelocollagen-based gene transfer in cells allows high-throughput screening of gene functions. *Biochemical and biophysical research communications, 289*(5), 1075-1081.

Inglese, J., Shamu, C.E. et Guy, R.K. (2007). Reporting data from high-throughput screening of small-molecule libraries. [10.1038/nchembio0807-438]. *Nat Chem Biol, 3*(8), 438-441.

Janzen, W.P. et Bernasconi, P. (2009). High throughput screening. Methods and protocols, Preface. *Methods in molecular biology (Clifton, NJ), 565*

Kaiser, J. (2008). Industrial-style screening meets academic biology. *Science, 321*(5890), 764-766.

Kelley, B. (2003). Automated Detection of Systematic Errors in Array Experiments. *Journal of the Association for Laboratory Automation, 8*(2), 24-26.

Kenakin, T. (2011). Functional selectivity and biased receptor signaling. *Journal of Pharmacology and Experimental Therapeutics, 336*(2), 296-302.

Kevorkov, D. et Makarenkov, V. (2005). Statistical analysis of systematic errors in high-throughput screening. *Journal of biomolecular screening, 10*(6), 557-567.

Kim, S.S., Peng, L.F., Lin, W., Choe, W.-H., Sakamoto, N., Schreiber, S.L. et Chung, R.T. (2007). A cell-based, high-throughput screen for small molecule regulators of hepatitis C virus replication. *Gastroenterology, 132*(1), 311-320.

Knapp, B. et Kaderali, L. (2012). *Statistical Analysis and Processing of Cellular Assays* : Technische Universität Dresden, Germany, iConcept Press.

Knapp, B., Rebhan, I., Kumar, A., Matula, P., Kiani, N.A., Binder, M., Erfle, H., Rohr, K., Eils, R. et Bartenschlager, R. (2011). Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC bioinformatics, 12*(1), 485.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Attuari, 4, 83-91. *Kolmogorov834Giorn. Ist. Ital. Attuari*

König, R., Chiang, C.-y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U. et Zhou, Y. (2007). A probability-based approach for the analysis of large-scale RNAi screens. *Nature methods, 4*(10), 847-849.

Kopczynski, C.C., Noordermeer, J.N., Serano, T.L., Chen, W.-Y., Pendleton, J.D., Lewis, S., Goodman, C.S. et Rubin, G.M. (1998). A high throughput screen to identify secreted and transmembrane proteins involved in Drosophila embryogenesis. *Proceedings of the National Academy of Sciences, 95*(17), 9973-9978.

Korn, K. et Krausz, E. (2007). Cell-based high-content screening of small-molecule libraries. *Current Opinion in Chemical Biology, 11*(5), 503-510.

Kozak, K., Agrawal, A., Machuy, N. et Csucs, G. (2009). Data Mining Techniques in High Content Screening: A Survey. *J Comput Sci Syst Biol, 2*, 219-239.

Koziol, J.A. (2010). The rank product method with two samples. *FEBS letters, 584*(21), 4481-4484.

Lachmann, A., Giorgi, F.M., Alvarez, M.J. et Califano, A. (2016). Detection and removal of spatial bias in multiwell assays. *Bioinformatics, 32*(13), 1959-1965.

Laio, F. (2004). Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research, 40*(9)

Lazo, J.S., Brady, L.S. et Dingledine, R. (2007). Building a Pharmacological Lexicon: Small Molecule Discovery in Academia. *Molecular Pharmacology, 72*(1), 1-7.

Lee, M.-L.T., Kuo, F.C., Whitmore, G. et Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of the National Academy of Sciences, 97*(18), 9834-9839.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K. et Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. [10.1038/nrg2825]. *Nat Rev Genet, 11*(10), 733-739.

Lo, E., Soleilhac, E., Martinez, A., Lafanechère, L. et Nadon, R. (2012). Intensity quantile estimation and mapping—a novel algorithm for the correction of image non-uniformity bias in HCS data. *Bioinformatics, 28*(20), 2632-2639.

Macarrón, R. et Hertzberg, R.P. (2011). Design and implementation of high throughput screening assays. *Molecular biotechnology, 47*(3), 270-285.

Makarenkov, V., Kevorkov, D., Zentilli, P., Gagarin, A., Malo, N. et Nadon, R. (2006). HTS-Corrector: software for the statistical analysis and correction of experimental high-throughput screening data. *Bioinformatics, 22*(11), 1408-1409.

Makarenkov, V., Zentilli, P., Kevorkov, D., Gagarin, A., Malo, N. et Nadon, R. (2007). An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics, 23*(13), 1648-1657.

Malo, N., Hanley, J.A., Carlile, G., Liu, J., Pelletier, J., Thomas, D. et Nadon, R. (2010). Experimental design and statistical methods for improved hit detection in high-throughput screening. *Journal of biomolecular screening, 15*(8), 990-1000.

Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J. et Nadon, R. (2006). Statistical practice in high-throughput screening data analysis. *Nature biotechnology, 24*(2), 167-175.

Menke, K.C. (2002). Unit automation in high throughput screening. *High Throughput Screening: Methods and Protocols*, 195-211.

Mohr, S., Bakal, C. et Perrimon, N. (2010). Genomic screening with RNAi: results and challenges. *Annu Rev Biochem, 79*, 37 - 64.

Montgomery, S.B., Lappalainen, T., Gutierrez-Arcelus, M. et Dermitzakis, E.T. (2011). Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS genetics, 7*(7), e1002144.

Moré, J.J. (1978). The Levenberg-Marquardt algorithm: implementation and theory *Numerical analysis* (p. 105-116) : Springer.

Mpindi, J.-P., Potdar, S., Bychkov, D., Saarela, J., Saeed, K., Wennerberg, K., Aittokallio, T., Östling, P. et Kallioniemi, O. (2015). Impact of normalization methods on high-throughput screening data with high hit-rates and drug testing with dose-response data. *Bioinformatics*, btv455.

Müller, P., Kuttenkeuler, D., Gesellchen, V., Zeidler, M.P. et Boutros, M. (2005). Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature, 436*(7052), 871-875.

Murie, C., Barette, C., Button, J., Lafanechère, L. et Nadon, R. (2015). Improving detection of rare biological events in high-throughput screens. *Journal of biomolecular screening, 20*(2), 230-241.

Murie, C., Barette, C., Lafanechere, L. et Nadon, R. (2013). Control-Plate Regression (CPR) Normalization for High-Throughput Screens with Many Active Features. *J Biomol Screen, 19*(5), 661-671.

Murie, C., Barette, C., Lafanechère, L. et Nadon, R. (2013). Single assay-wide variance experimental (SAVE) design for high-throughput screening. *Bioinformatics*, btt538.

Nadon, R. et Shoemaker, J. (2002). Statistical issues with microarrays: processing and analysis. *TRENDS in Genetics, 18*(5), 265-271.

Nelson, R.M.a.Y., J.D.Conference, I. U. (dir.). (2004). *Introduction to High-Throughput Screening for Drug Discovery. Inc., San Diego,CA*, Actes du colloque.

Noah, J.W. (2010). New developments and emerging trends in high-throughput screening methods for lead compound identification. *Int J High Throughput Screening, 1*, 141-149.

Ogier, A. et Dorval, T. (2012). HCS-Analyzer: open source software for high-content screening data correction and analysis. *Bioinformatics, 28*(14), 1945-1946.

Okiyoneda, T. et Lukacs, G.L. (2012). *Fixing cystic fibrosis by correcting CFTR domain assembly* : Rockefeller University Press.

Peter, R. et Roy, A. (2011). A roadmap for achieving self-sustainability of academic high throughput screening core facilities. *Drug Discovery, 12*, 59.

Preuss, J., Hedrick, M., Sergienko, E., Pinkerton, A., Mangravita-Novo, A., Smith, L., Marx, C., Fischer, E., Jortzik, E. et Rahlfs, S. (2012). High-throughput screening for small-molecule inhibitors of plasmodium falciparum glucose-6-phosphate dehydrogenase 6-phosphogluconolactonase. *Journal of biomolecular screening, 17*(6), 738-751.

Ramadan, N., Flockhart, I., Booker, M., Perrimon, N. et Mathey-Prevot, B. (2007). Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. *Nature protocols, 2*(9), 2245-2264.

Razali, N.M. et Wah, Y.B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics, 2*(1), 21-33.

Rieber, N., Knapp, B., Eils, R. et Kaderali, L. (2009). RNAither, an automated pipeline for the statistical analysis of high-throughput RNAi screens. *Bioinformatics, 25*(5), 678-679.

Seiler, K.P., George, G.A., Happ, M.P., Bodycombe, N.E., Carrinski, H.A., Norton, S., Brudz, S., Sullivan, J.P., Muhlich, J. et Serrano, M. (2008). ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic acids research, 36*(suppl 1), D351-D359.

Sharma, S. et Rao, A. (2009). RNAi screening: tips and techniques. *Nature immunology, 10*(8), 799-804.

Shelat, A.A. et Guy, R.K. (2007). The interdependence between screening methods and screening libraries. *Current Opinion in Chemical Biology, 11*(3), 244-251.

Shun, T.Y., Lazo, J.S., Sharlow, E.R. et Johnston, P.A. (2011). Identifying Actives from HTS Data Sets Practical Approaches for the Selection of an Appropriate HTS Data-Processing Method and Quality Control Review. *Journal of biomolecular screening, 16*(1), 1-14.

Silber, B.M. (2010). Driving drug discovery: the fundamental role of academic labs. *Science Translational Medicine, 2*(30), 30cm16-30cm16.

Sirois, S., Hatzakis, G., Wei, D., Du, Q. et Chou, K.-C. (2005). Assessment of chemical libraries for their druggability. *Computational biology and chemistry, 29*(1), 55-67.

Smellie, A., Wilson, C.J. et Ng, S.C. (2006). Visualization and interpretation of high content screening data. *Journal of chemical information and modeling, 46*(1), 201-207.

Snijder, B. et Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. [10.1038/nrm3044]. *Nat Rev Mol Cell Biol, 12*(2), 119-125.

Snijder, B., Sacher, R., Rämö, P., Damm, E.-M., Liberali, P. et Pelkmans, L. (2009). Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature, 461*(7263), 520-523.

Soneson, C., Gerster, S. et Delorenzi, M. (2014). Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PloS one, 9*(6), e100335.

Stephens, M.A. (1986). Tests based on EDF statistics. *Goodness-of-fit Techniques, 68*, 97-193.

Tukey, J.W. (1977). Exploratory data analysis.

Vassal, E., Barette, C., Fonrose, X., Dupont, R., Sans-Soleilhac, E. et Lafanechère, L. (2006). Miniaturization and validation of a sensitive multiparametric cell-based assay for the concomitant detection of microtubule-destabilizing and microtubule-stabilizing agents. *Journal of Biomolecular Screening, 11*(4), 377-389.

Venables, W.N., Ripley, B.D. et Venables, W. (1994). *Modern applied statistics with S-PLUS*. (Vol. 250) : Springer-verlag New York.

Verdugo, R.A., Deschepper, C.F., Muñoz, G., Pomp, D. et Churchill, G.A. (2009). Importance of randomization in microarray experimental designs with Illumina platforms. *Nucleic acids research, 37*(17), 5610-5618.

Wan, K.H., Yu, C., George, R.A., Carlson, J.W., Hoskins, R.A., Svirskas, R., Stapleton, M. et Celniker, S.E. (2006). High-throughput plasmid cDNA library screening. *Nature protocols, 1*(2), 624-632.

Welch, B.L. (1947). The generalization ofstudent's' problem when several different population variances are involved. *Biometrika, 34*(1/2), 28-35.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin, 1*(6), 80-83.

Willard, F.S., Bueno, A.B. et Sloop, K.W. (2012). Small molecule drug discovery at the glucagon-like peptide-1 receptor. *Experimental diabetes research, 2012*

Wootten, D., Savage, E.E., Willard, F.S., Bueno, A.B., Sloop, K.W., Christopoulos, A. et Sexton, P.M. (2013). Differential activation and modulation of the glucagon-like peptide-1 receptor by small molecule ligands. *Molecular pharmacology, 83*(4), 822-834.

Wright, G.W. et Simon, R.M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics, 19*(18), 2448-2455.

Wu, Z., Liu, D. et Sui, Y. (2008). Quantitative assessment of hit detection and confirmation in single and duplicate high-throughput screenings. *Journal of biomolecular screening, 13*(2), 159-167.

Zanella, F., Lorens, J.B. et Link, W. (2010). High content screening: seeing is believing. *Trends in Biotechnology, 28*(5), 237-245.

Zhang, J.-H., Chung, T.D. et Oldenburg, K.R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of biomolecular screening, 4*(2), 67-73.

Zhang, X.D. (2007). A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics, 89*(4), 552-561.

Zhang, X.D. (2008). Novel analytic criteria and effective plate designs for quality control in genome-scale RNAi screens. *Journal of biomolecular screening, 13*(5), 363-377.

Zhang, X.D., Espeseth, A.S., Johnson, E., Chin, J., Gates, A., Mitnaul, L.J., Marine, S.D., Tian, J., Stec, E.M. et Kunapuli, P. (2008). Integrating experimental and analytic approaches to improve data quality in genome-wide RNAi screens. *Journal of biomolecular screening, 13*(5):378-389.