



OPEN ACCESS

EDITED BY

Ni Zhao,
Johns Hopkins University, United States

REVIEWED BY

Judong Shen,
Merck & Co., Inc., United States
Chixiang Chen,
University of Maryland, Baltimore,
United States

*CORRESPONDENCE

Benjamin Haibe-Kains,
bhaibeka@uhnresearch.ca

SPECIALTY SECTION

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Genetics

RECEIVED 24 August 2022

ACCEPTED 25 November 2022

PUBLISHED 04 January 2023

CITATION

Abbas-Aghababazadeh F, Xu W and
Haibe-Kains B (2023), The impact of
violating the independence assumption
in meta-analysis on
biomarker discovery.
Front. Genet. 13:1027345.
doi: 10.3389/fgene.2022.1027345

COPYRIGHT

© 2023 Abbas-Aghababazadeh, Xu and
Haibe-Kains. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The impact of violating the independence assumption in meta-analysis on biomarker discovery

Farnoosh Abbas-Aghababazadeh¹, Wei Xu^{1,3} and
Benjamin Haibe-Kains^{1,2,3,4,5*}

¹Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada, ²Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada, ³Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, ⁴Ontario Institute for Cancer Research, Toronto, ON, Canada, ⁵Department of Computer Science, University of Toronto, Toronto, ON, Canada

With rapid advancements in high-throughput sequencing technologies, massive amounts of “-omics” data are now available in almost every biomedical field. Due to variance in biological models and analytic methods, findings from clinical and biological studies are often not generalizable when tested in independent cohorts. Meta-analysis, a set of statistical tools to integrate independent studies addressing similar research questions, has been proposed to improve the accuracy and robustness of new biological insights. However, it is common practice among biomarker discovery studies using preclinical pharmacogenomic data to borrow molecular profiles of cancer cell lines from one study to another, creating dependence across studies. The impact of violating the independence assumption in meta-analyses is largely unknown. In this study, we review and compare different meta-analyses to estimate variations across studies along with biomarker discoveries using preclinical pharmacogenomics data. We further evaluate the performance of conventional meta-analysis where the dependence of the effects was ignored *via* simulation studies. Results show that, as the number of non-independent effects increased, relative mean squared error and lower coverage probability increased. Additionally, we also assess potential bias in the estimation of effects for established meta-analysis approaches when data are duplicated and the assumption of independence is violated. Using pharmacogenomics biomarker discovery, we find that treating dependent studies as independent can substantially increase the bias of meta-analyses. Importantly, we show that violating the independence assumption decreases the generalizability of the biomarker discovery process and increases false positive results, a key challenge in precision oncology.

KEYWORDS

meta-analysis, non-independent effects, gene expression, pharmacogenomics, biomarker

1 Introduction

Patient response to anticancer drugs varies widely, and the genomic-make-up of a patient's tumor is a major factor contributing to this variation. The field of pharmacogenomics studies the influence of genomic variation on individualized drug response (Iorio et al., 2016). Due to the limited access and high cost of clinical samples, cancer cell lines are frequently used to investigate disease biology and its relationship to drug response for biomarker discovery (Basu et al., 2013; Iorio et al., 2016).

Due to the complexity of generating pharmacogenomic datasets, the reproducibility of preclinical data and the findings from high-throughput profiling studies in cancer cell lines has been extensively investigated (Haibe-Kains et al., 2013; Haverty et al., 2016; Safikhani et al., 2016). Technological improvements in high-throughput drug screening enable the generation of large-scale pharmacogenomic datasets, which provide a remarkable opportunity for the identification of biomarkers predictive of drug response. However, biomarkers often fail to generalize across independent studies. This is due to the complexity of biological systems, the use of different experimental protocols, and the application of various technology platforms for both molecular profiling and data processing methods (Haibe-Kains et al., 2013; Haverty et al., 2016; Dempster et al., 2019). Moreover, the large number of features and relatively small number of cell lines can lead to findings that are not generalizable and show bias when assessed in independent cohorts (Tseng et al., 2012; Hatzis et al., 2014; Sweeney et al., 2017).

To address these issues, meta-analyses can be performed to integrate independent studies to identify more reliable biomarkers by increasing statistical power and reducing false positives (Sweeney et al., 2017; Abbas-Aghababazadeh et al., 2020; Borenstein et al., 2021). In recent years, several meta-analysis methods have been proposed such as combining p -values (Fisher et al., 1934; Stouffer et al., 1949; Won et al., 2009), combining effect estimates (Borenstein et al., 2021), and rankings (Hong et al., 2006). The strengths and limitations of meta-analyses are evaluated particularly with respect to their ability to assess variation across studies or heterogeneity (e.g., platform variability, inconsistent annotation, various methods for data processing, cell lines heterogeneity, and laboratory-specific effects) beyond within-study variation (e.g., experimental designs and populations of interest) (Veroniki et al., 2016). As a result, the selection of an optimal meta-analysis method depends considerably on the available data structure and the hypothesis setting to achieve the underlying biological goal (Chang et al., 2013).

Combining the p -values from multiple independent studies has the benefit of simplicity and extensibility to different kinds of outcome variables (Tseng et al., 2012; Chang et al., 2013). However, this can only be performed under the parametric assumption where the p -values are uniformly distributed

under the null hypothesis, while not accounting for the data heterogeneity and direction of effect sizes, which represents a major limitation of this method (Marot et al., 2009). Approaches that combine effects including fixed- and random-effects (FE & RE) models are widely used to achieve a broad inferential basis for evaluations of effects (Cochran, 1954; Borenstein et al., 2021). Under the FE model, we assume that there is one true effect that underlies all the studies in the analysis, and that all differences in observed effects are due to sampling error. In contrast, the RE model incorporates the variability of the effects across studies in addition to the within-study variability using a two-stage hierarchical process. Several approaches have been proposed to combine individual study results into an overall estimate of effect using the inverse-variance strategy (Egger et al., 2008).

Assessing heterogeneity is a critical issue in meta-analysis because different models may lead to different estimates of overall effect and different standard errors. Several approaches have been suggested that vary in popularity and complexity for how best to carry out such combinations (Veroniki et al., 2016; Guolo and Varin, 2017; Langan et al., 2019). Cochran (Cochran, 1954) proposed a Q test to determine the heterogeneity across studies, however its statistical power depends on the number of studies and sample size (Whitehead and Whitehead, 1991; Turner et al., 2012; Hoaglin, 2016; Bodnar et al., 2017). Higgins and Thompson proposed better statistic I^2 to describe heterogeneity that reflects the proportion of total variance that is attributed to heterogeneity (Higgins et al., 2003). Meta-analysis allows us to quantify heterogeneity between studies, but precisely estimating the between-study heterogeneity is challenging. This is especially true if the number of studies included is small. A Bayesian approach was proposed to capture the uncertainty in the estimation of the between-study variance by incorporating prior knowledge (Sutton and Abrams, 2001; Bodnar et al., 2017; Röver, 2017).

Traditional meta-analysis procedures make a crucial assumption: effects are independent. When this assumption is violated, conclusions based on meta-analyses can be misleading and will bias the overall effect estimate along with inflating Type I error. In real-world applications, there are various sources of effect dependencies within and across studies. For instance, a meta-analysis uses more than one outcome measure for the same sample of participants, two treatment groups with the same control group, duplicate full or partial data or samples, and effects reported by the same research group (Becker, 2000; Wood, 2008; Lin and Sullivan, 2009; Hedges et al., 2010; Cheung, 2014; Scammacca et al., 2014; Van den Noortgate et al., 2015; Cheung, 2019; Cooper et al., 2019; Liu et al., 2019; Wilson, 2019; Liu and Xie, 2020; Luo et al., 2020; Borenstein et al., 2021). The dependence of the effects must be resolved in a way that permits each study to contribute a single independent effect to the meta-analysis or modeled dependence in order to avoid threats to the validity of the meta-analysis results. Reflecting on duplicate data problems, detecting potential duplicate studies

and options for the correction of duplication were discussed, while the issue of prevention of duplication has not yet been addressed (Wood, 2008). In general, different strategies have been proposed to handle dependency including ignoring dependence, avoiding dependence [e.g., averaging effects (Wood, 2008) or shifting unit-of-analysis (Cooper et al., 2019)], modeling dependence [e.g., robust variance estimation (Hedges et al., 2010), multivariate meta-analysis (Becker, 2000), multilevel meta-analysis (Cheung, 2014; Van den Noortgate et al., 2015; Cheung, 2019)], and determining the covariance of effects across studies for overlap samples (Lin and Sullivan, 2009; Luo et al., 2020). Moreover, the Cauchy combination test Liu et al. (2019); Liu and Xie (2020) and the harmonic mean p -value Wilson (2019) were proposed as robust choices to combine p -values under arbitrary dependency structures to control type I error rate and improve power of analysis.

Several preclinical pharmacogenomic meta-analyses have been performed in an effort to discover predictive biomarkers with consistent evidence across multiple research laboratories (Cohen et al., 2011; Safikhani et al., 2016, 2017; Haverty et al., 2016; Ding et al., 2018; Jaiswal et al., 2021; Xia et al., 2022). However, virtually all existing datasets suffer from missing observations due to limitations of the experimental techniques, insufficient resolution, and sequencing costs which prevent complete profiling of cancer samples at the genomic and pharmacological levels (Choi and Pavelka, 2012; Basu et al., 2013; Muir et al., 2016; Li et al., 2017). Notably, the lack of either molecular (Basu et al., 2013) or pharmacological data [e.g., drug response Li et al. (2017)] in a given study prevents its use for biomarker discovery. Hence, investigators often attempt to make datasets more complete by borrowing a subset of the data (e.g., molecular data, drug response data, or combination of both) from one study and duplicating them in another study (Consortium of Drug Sensitivity in Cancer Consortium, 2015; Safikhani et al., 2016; Li et al., 2017; Ding et al., 2018; Jia et al., 2021; Xia et al., 2022). For instance, the re-analysis of drug response consistency by the Cancer Cell Line Encyclopedia and Genomics of Drug Sensitivity in Cancer, investigators duplicated the gene mutation, copy number alteration and mRNA expression when assessing the generalizability of biomarkers predictive of drug response (gene-drug associations) (Consortium of Drug Sensitivity in Cancer Consortium, 2015; Safikhani et al., 2016). In applied research, detection of non-independent effects or duplicate study effects along with modeling dependency remain challenging. When study effect estimates are non-independent, conclusions based on the conventional meta-analyses will bias the aggregated effects and can be misleading or even wrong (Wood, 2008; Cheung, 2019).

In this study, we reviewed and compared the performance of frequentist and Bayesian meta-analysis approaches to assess gene-drug associations or biomarker discovery using independent large-scale breast cancer and pan-cancer pharmacogenomic datasets. We found that changes in the

number and size of studies along with the type of meta-analysis methods can affect the identification of statistically significant gene-drug associations. We further conducted simulation studies to assess the performance of including non-independent studies or effects in a traditional meta-analysis approach. Results showed that as the number of non-independent effects increased, higher relative mean squared error and lower coverage probability were observed. In addition, we aimed to evaluate the bias of avoiding the dependence of effects in traditional meta-analyses *via* preclinical pharmacogenomic data. To do so, we showed how increases in the number of duplicated studies can impact the bias of estimated overall effect and the identification of gene-drug associations. The results indicate that by increasing the number of dependent studies, bias of estimated overall effect may increase and genes with lower similarity of measured expression across studies denote higher bias.

2 Materials and Methods

2.1 Data types and sources

We used transcriptomic (RNA-Sequencing and gene expression microarray) and drug response data from pharmacogenomic cancer cell line sensitivity screenings, including the Cancer Cell Line Encyclopedia (CCLE: Broad-Novartis) (Barretina et al., 2012), the Genomics of Drug Sensitivity in Cancer (GDSC: Wellcome Trust Sanger Institute) (Garnett et al., 2012; Yang et al., 2012), the Genentech Cell Line Screening Initiative (gCSI) (Haverty et al., 2016), the Cancer Therapeutics Response Portal (CTRP: Broad Institute) (Basu et al., 2013), Oregon Health and Science University breast cancer screen (GRAY) (Heiser et al., 2012), and University Health Network Breast Cancer Screen (UHNBreast) (Marcotte et al., 2016; Safikhani et al., 2017) (Table 1).

The cell lines' gene expression profiles were generated using RNA sequencing (RNA-seq) in all datasets except for CTRP and GDSC datasets where the Affymetrix microarrays were used (Table 1). Molecular information was obtained from the *PharmacoGx* R package along with details on data processing (Safikhani et al., 2016; Smirnov et al., 2016). Cell line drug response data, in the form of area above the curve (AAC) recomputed information, was also obtained from the *PharmacoGx* R package (Table 1) (Smirnov et al., 2016).

In this study, we performed a breast cancer-specific and a pan-cancer analysis to assess the generalization of our results. For the breast cancer analysis, we included seven independent studies, while the pan-cancer analysis excluded the UHNBreast and GRAY datasets as they included only breast cancer cell lines (Figure 1). Pan-cancer data was filtered for tissue types containing at least 10 cell lines with available expression

TABLE 1 Studies used in meta-analyses and corresponding drug response versions, number of cell lines, and drugs along with molecular data type. Different versions of GDSC study refer to drug response assays along with updated cell lines and drugs. (*) represents the CCLE microarray data. Transcriptomic and drug response data were obtained from the *PharmacGx* R package.

Dataset	Drug sensitivity	No. of cell lines	No. of drugs	Molecular data
CCLE Barretina et al. (2012)	2015	1,094	24	RNA-Seq
CTRP Basu et al. (2013)	2015	887	544	Microarray*
GDSC1 Garnett et al. (2012)	2020 (v1-8.2)	1,060	343	Microarray
GDSC2 Garnett et al. (2012)	2020 (v2-8.2)	1,104	190	Microarray
gCSI Haverty et al. (2016)	2017	747	16	RNA-Seq
GRAY Heiser et al. (2012)	2017	74	107	RNA-Seq
UHNBreast Safikhani et al. (2017)	2019	84	8	RNA-Seq

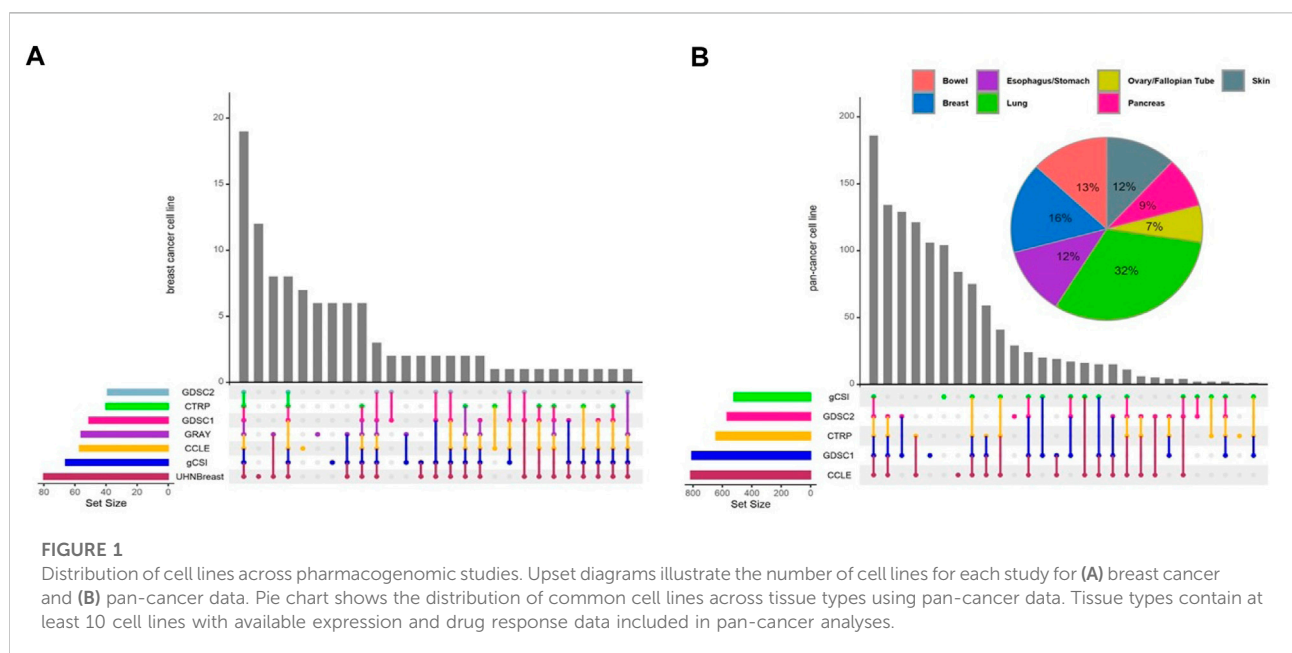
and drug response data (Supplementary Figure S1). Breast cancer meta-analyses consisted of 11,198 genes and 19 cell lines shared between studies. To analyze pan-cancer data, we selected 11,340 common genes and 168 cell lines between studies. Common anticancer drugs such as Erlotinib, Lapatinib, and Paclitaxel across studies were used to evaluate gene-drug associations and discover biomarkers. Notably, Lapatinib and Erlotinib are a potent, oral, reversible, dual inhibitor of epidermal

growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2 or ERBB2) for breast cancer and a few other solid tumors (Schlessinger, 2000; Geyer et al., 2006; Medina and Goodin, 2008; Iqbal et al., 2011). Additionally, Paclitaxel is a widely used drug for various solid tumors and breast cancer treatment, however, resistance occurs frequently and the evasion mechanisms remain unclear (Volk-Draper et al., 2014; Weaver, 2014).

2.2 Data pre-processing

The drug dose-response curves were summarized into AAC values using a 3-parameter logistic model as implemented in *PharmacGx* (Smirnov et al., 2016). The AAC values range between 0 and 1, with AAC close to 1 indicating the drug sensitive cell lines and AAC close to 0 indicating the drug resistant cell lines. Missing drug response data (AAC values) were imputed within each study to simplify further analyses. Imputation was completed *via* multiple imputation by classification and regression trees (MI-CART) (Burgette and Reiter, 2010; Burgette and Reiter, 2010) as implemented in *mice* R package (Buuren and Groothuis-Oudshoorn, 2010).

For individual breast cancer and pan-cancer data analyses, we focused on genes and drugs that were measured in all studies. To avoid scaling issues and make comparisons between the estimated effects across studies, the z-score transformation (mean centered and variance scaled) was considered across cell lines using raw gene expression values and imputed AAC values. For a given gene, Pearson correlation was applied to assess the similarity of measured expression across studies.



2.3 Meta-analysis methods

Let us consider G matched genes across K studies. For a given study k ($k = 1, \dots, K$), denote by $x_{g|lk}$ the standardized gene expression intensity of gene g ($g = 1, \dots, G$) and cell line l ($l = 1, \dots, L_k$), where L_k represents the number of cell lines in study k . Let y_{lk} be the standardized imputed drug response variable of cell line l and study k . Under the breast cancer data, a standardized linear regression model was applied to assess the association of gene expression and imputed drug response, while adjusting for the tissue types to evaluate pan-cancer data (Peterson and Brown, 2005). For each individual gene-drug association analysis, estimated effect, standard error of estimated effect, and p -value were obtained. To identify a broad inferential basis for evaluation of estimated effects, meta-analysis techniques were applied to integrate results of separate independent studies including combining p -values and combining estimated effects (Supplementary Figure S2A).

2.3.1 Combination of p -values

Combining the p -values from multiple independent studies offers the benefits of simplicity and extensibility to different kinds of outcome variables (Tseng et al., 2012). Fisher's method (Fisher et al., 1934) and Stouffer's method (Stouffer et al., 1949) are widely used to combine p -value results from different independent studies.

For each gene, Fisher's method sums up the log transformed p -values obtained from individual study (Fisher et al., 1934). The combined Fisher's statistic $\chi^2_{\text{Fisher}} = -2 \sum_k \log(p_k)$ follows a chi-squared χ^2 distribution with $2K$ degrees of freedom under the null hypothesis.

For each gene, Stouffer's method sums the inverse normal transformed p -values (Stouffer et al., 1949). The combined Stouffer's statistic $T_{\text{Stouffer}} = \sum_k z_k / \sqrt{K}$; where $z_k = \Phi^{-1}(p_k)$ and Φ is a standard normal cumulative distribution function; it follows a standard normal distribution under the null hypothesis.

Major limitations of such classical methods are that they do not account for the data heterogeneity and direction of effects. Additionally, such methods are performed under the parametric assumption that p -values are uniformly distributed under the null hypothesis, though in practice such an assumption is not always satisfied (Marot et al., 2009).

2.3.2 Combination of estimated effects

Let β_k and s_k be the estimated effect and its standard error in study k , respectively. The most popular parametric statistical methods for combining effects are based on fixed- and random-effects (FE & RE) models (Cochran, 1954; Borenstein et al., 2021). The FE model assumes that all studies in meta-analysis share a single true effect β (i.e., average effect across all studies) is specified as

$$\beta_k = \beta + \epsilon_k,$$

where ϵ_k is assumed to be independently distributed as $\epsilon_k \sim N(0, \sigma_k^2)$ and within-study variance σ_k^2 . Under the FE model, estimated effects are assumed to be homogeneous across studies and all differences in observed effects are due to sampling error or within-study variability, while in practice such an assumption is questionable. One of the challenges in meta-analysis is that true effects are different across studies. Hence, the RE model was proposed to incorporate the variability of the estimated effects across studies in addition to the within-study variability using a two-stage hierarchical process as follows (DerSimonian and Laird, 1986; Brockwell and Gordon, 2001; Viechtbauer, 2005; Bodnar et al., 2017)

$$\beta_k = \beta + \eta_k + \epsilon_k,$$

where η_k indicates the error accounting for the between-study variability. Assume error η_k is independently distributed as $\eta_k \sim N(0, \tau^2)$, where τ^2 is the between-study or heterogeneity variance of η_k around β . The marginal distribution of β_k follows the normal distribution as $\beta_k \sim N(\beta, \tau^2 + \sigma_k^2)$, where η_k and ϵ_k are independent. A common assumption is that each study is based on studies large enough to consider the within-study variance as known and equal to the estimate provided by each study (i.e., s_k^2).

The goal is to estimate the overall effect β . Following the summarized assumptions in Supplementary Table S1, several approaches have been suggested for combining individual study results into an overall estimate of effect applying an inverse variance scheme (Egger et al., 2008). Under the FE model, the most flexible and widely used inverse variance-weighted average approach was proposed to estimate the overall effect as

$$\hat{\beta} = \frac{\sum_{k=1}^K w_k \beta_k}{\sum_{k=1}^K w_k}, \quad w_k = \frac{1}{s_k^2}.$$

However, under the RE model, the proposed inverse variance approach is considered to estimate the overall effect where $w_k = 1/(\tau^2 + s_k^2)$ and the estimate of overall effect depends on the between-study variance. w_k provides the uniformly minimum variance unbiased estimator of β . A Wald-type 95% confidence interval is estimated for the overall effect using the normal approximation (DerSimonian and Laird, 1986). The proposed RE meta-analysis approaches are typically performed in two stages, where first the heterogeneity parameter is estimated, and then the effect estimate is derived using the estimated heterogeneity. One of the most troublesome challenging aspects of meta-analysis is the determination of whether there is true heterogeneity, as it can influence the choice of the statistical method to combine estimated effects.

2.3.3 Measures of between-study heterogeneity

Under the RE model, the detection of heterogeneity requires testing the null hypothesis $\tau^2 = 0$, which corresponds to the FE

model. The homogeneity hypothesis is tested using the Q statistic (DerSimonian and Kacker, 2007) given by

$$Q = \sum_{k=1}^K (\beta_k - \hat{\beta})^2 w_k$$

which has an asymptotic chi-squared χ^2 distribution with $K-1$ degrees of freedom under the hypothesis of consistent or homogeneous association, and $\hat{\beta}$ and w_k are computed using the FE model analysis. The DerSimonian and Laird (DL) estimator DerSimonian and Laird (1986) is possibly the most frequently used approach to estimate the between-study heterogeneity *via* moments approach given by

$$\hat{\tau}^2 = \max \left\{ 0, \frac{Q - (K - 1)}{\sum_{k=1}^K w_k - \sum_{k=1}^K w_k^2 / \sum_{k=1}^K w_k} \right\}.$$

Note that the Q test suffers from low power when studies have small sample size (L_k) or are few in number (K) (Hoaglin, 2016). More importantly, the Q statistic and the estimate of between-study variance depend on the scale of effects. Hence, neither Q nor $\hat{\tau}^2$ can be used to compare degrees of heterogeneity between different meta-analyses (Higgins et al., 2003). Several approaches have been suggested over the years for how best to assess homogeneity across studies, and the corresponding research is ongoing.

The relative amount of between-study heterogeneity can be expressed in terms of the measure of $I^2 = \tau^2 / (\tau^2 + \bar{s}^2)$ that reflects the proportion of total variance that is attributed to heterogeneity (Higgins et al., 2003). \bar{s}^2 is required as an estimate of the within-study variances. Unlike the Q test, I^2 can be directly compared between meta-analyses with different numbers of studies and different types of outcome data and also does not inherently depend on the number of studies in the meta-analysis (Higgins et al., 2003). Due to the low power situation of meta-analyses, Higgins et al. (2003) suggested that we could tentatively assign a significant degree of heterogeneity (i.e., substantial heterogeneity) when $I^2 > 50\%$ and Q test p -value < 0.1 (Hoaglin, 2016).

Many alternatives to the DL approach have been proposed such as modifications of the method of moments, likelihood principle, model error variance, Bayes, and the non-parametric approaches with the aim of avoiding distributional assumptions (Veroniki et al., 2016; Guolo and Varin, 2017; Langan et al., 2019). In this study, we focus on six different heterogeneity variance estimators including moments estimators such as DerSimonian and Laird (DL) (DerSimonian and Laird, 1986), Paule and Mandel (PM) (Paule and Mandel, 1982), Hedges (HE) (Hedges and Olkin, 2014) and Hunter-Schmidt (HS) (Schmidt and Hunter, 2004), error variance estimator Sidik-Jonkman (SJ) (Sidik and Jonkman, 2007), and empirical Bayes (EB) estimator (Sutton and Abrams, 2001; Turner et al., 2012). A common problem with such estimators is that they frequently produce higher numbers for small meta-analyses, and lower for analyses

involving many studies which leads to inadequate heterogeneity and effect estimates along with challenges in the choice of FE or RE analysis methods (Turner et al., 2012; Bodnar et al., 2017; Langan et al., 2019).

To compare meta-analyses results, within each study, the effect size for each gene was measured. Meta-analyses for each gene were performed across studies using the *meta* (Schwarzer et al., 2019) and *metap* (Dewey, 2018) R packages. To correct for multiple testing, the Benjamini & Hochberg procedure was used to control false discovery rate (FDR) (Benjamini and Hochberg, 1995). The Jaccard coefficient index was applied to compare the similarity between the identified top-ranked genes associated with a drug using different independent meta-analyses. An UpSet plot was applied to visualize the number and the overlap of genes associated with drugs by considering different meta-analysis methods and volcano plot to highlight statistically significant gene-drug associations.

2.3.4 Bayesian approaches

Bayesian inference has been suggested in the context of meta-analysis to capture uncertainty in estimation of the heterogeneity by incorporating prior information (Sutton and Abrams, 2001; Bodnar et al., 2017; Röver, 2017). Compared to the earlier discussed approaches, Bayesian methods offer several potential advantages which include producing a distribution for the effect and heterogeneity, leading to credible intervals for overall effect and heterogeneity.

Within the explained two-stage hierarchical model (Supplementary Table S1), to infer the unknown hyper parameters β and τ , prior knowledge needs to be specified where the joint prior probability can be factored into independent marginals $p(\beta, \tau) = p(\beta) \times p(\tau)$. The range of reasonably specified priors is usually limited. Non-informative or weakly informative priors for the effect β are usually applied. However, informative normal prior constitutes the conditionally conjugate prior distribution for the effect along with being computationally convenient (Röver, 2017). In addition, for the between-study heterogeneity τ an informative prior is often appropriate especially when only a small number of studies is involved (Bodnar et al., 2017; Röver, 2017). Let

$$p(\beta, \tau) \propto \tau \sqrt{\sum_{k=1}^K \left(\frac{1}{\sigma_k^2 + \tau^2} \right)^2},$$

denote the non-informative uniform and Jeffreys priors for parameters β and τ , respectively Bodnar et al. (2017); Röver (2017). The Jeffreys prior's dependence on the standard errors σ_k implies that the prior information varies with the precision of the underlying data β_k . Posterior density may be accessed in quasi-analytical form and 95% central credible intervals derived from a posterior probability distribution can be constructed using the *bayesmeta* R package Röver (2017). We draw attention to this

Bayesian procedure by comparing its performance with commonly used classical DL meta-analysis procedure.

2.4 Assessing the impact of duplication of data across studies

Duplication in study effects has been defined as the estimated effect results from a complete replication of a particular study or from some subset of measured data (Wood, 2008; Cheung, 2019). The dependency in the meta-analysis is studies at the effects level, which can then be caused by duplication of input data (e.g., gene expression), output data (e.g., drug response measurements) or a combination of both (Supplementary Figure S2B).

Aggregating non-independent effects will bias the estimated overall effect in meta-analyses Wood (2008). For a given study k' ($k' = 1, \dots, K$), let $x_{gik'}^*$ be the missing gene expression intensity of gene g across cell lines $l = 1, \dots, L$, where L matched cell lines is considered across studies (i.e., $L_k = L, \forall k = 1, \dots, K$). Assume available measured expression data x_{gik} such that $k' \neq k, \forall k = 1, \dots, K$ be duplicated for the missing data (i.e., $x_{gik} \rightarrow x_{gik'}^*$) which leads to a lack of independence of effects across studies (i.e., $\beta_k \not\perp \beta_{k'}$). In each meta-analysis, $\Gamma(\gamma, K)$ denotes a set of all possible dependent studies by computing γ -combination of K studies, where $\gamma = 1, \dots, K-1$ is the number of missing expression data.

One of the considerable consequences of traditional meta-analyses of non-independent studies compared with independent ones is the significant differences in the estimate of overall effect (Wood, 2008). Hence, for a given gene and γ , to investigate the bias of ignoring the independence assumption of effects, the mean absolute deviation (MAD) metric is computed as

$$MAD = \frac{1}{|\Gamma|} \sum_{d \in \Gamma} |\hat{\beta}_d - \hat{\beta}|,$$

where $\hat{\beta}$ and $\hat{\beta}_d$ demonstrate the estimated overall effect by using the independent and non-independent studies, respectively. In addition, we evaluate the association between increases in the number of non-independent studies and the similarity of measured expression across studies. The Jaccard similarity index was applied to compare the similarity between the detected top-ranked genes associated with a drug over duplicated study effects. We evaluate whether there is an increasing trend in bias over the number of non-independent studies. Trend detection non-parametric Mann-Kendall (MK) (Mann, 1945; Pohlert, 2016) method is considered to test the null hypothesis H_0 : there is no monotonic trend in bias, versus the alternative hypothesis H_a of an increasing trend of bias when the number of studies increase. The MK trend test follows asymptotically a normal distribution where a lack of samples may decrease the power of analysis to detect trends.

2.5 Simulation study design

Most simulation studies currently explore the performance of various methods by focusing on measuring inconsistency in meta-analyses rather than in dealing with dependent effects across studies. Therefore, we conducted a simulation study to investigate the performance of the widely used RE meta-analysis approach (DL) in estimating the overall effect in which the dependence of the effects are ignored. We explored the effects of different parameters, varied in the simulations including the heterogeneity across studies (τ^2), number of studies (K), variation in studies (σ_k^2), correlation between random effects (ρ), and overall effects (β). When effects were correlated across studies, the traditional RE method (DL) was applied to assess whether and for which conditions lack of independence assumption across studies produced significantly different results. In this study, we focused on the overall effect estimate and on its standard error, because applied researchers are often mostly interested in testing overall effect.

2.5.1 Simulation method and performance measures

Parameter estimates and their performance measures are assessed in a $3(K) \times 4(\tau^2) \times 3(\sigma^2) \times 3(\beta) \times 4(\rho)$ factorial design (Table S2), where number of studies, heterogeneity across studies, variation within-studies, overall effects, and correlation between random effects are varied and each scenario is tested using a DL meta-analysis approach. Note that the within-study variance is assumed to be equal. Under the explained two-stage hierarchical model, for studies $1, \dots, K$ in each meta-analysis, true effects $(\beta_1, \dots, \beta_K)'$ are simulated from multivariate normal distribution $N_K(\beta \mathbf{I}_K, \Sigma)$, where \mathbf{I}_K is a $K \times K$ identity matrix and the symmetric variance-covariance matrix

$$\Sigma = \begin{bmatrix} \tau^2 + \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \dots & \rho_{1k}\sigma_1\sigma_k \\ \rho_{21}\sigma_2\sigma_1 & \tau^2 + \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \dots & \rho_{2k}\sigma_2\sigma_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k1}\sigma_k\sigma_1 & \rho_{k2}\sigma_k\sigma_2 & \rho_{k3}\sigma_k\sigma_3 & \dots & \tau^2 + \sigma_k^2 \end{bmatrix}$$

Parameter values are chosen to represent the range of values observed in the pharmacogenomic meta-analysis. The number of studies to be combined is manipulated at 3 levels (3, 7, and 10) to simulate a statistical combination of a range of meta-analytic preclinical pharmacogenomic studies. The Variation within studies represents that large number of cell lines within studies produces smaller within-study variance, which influences overall effect estimation.

Heterogeneity variance parameter values (τ^2) are defined such that the resulting meta-analyses vary through a wide range of levels of inconsistency between study effects. A between-study variance of 0.001 would signify that almost no true heterogeneity exists. The correlation selected to calculate the

covariances between study effects ranges between 0.1 and 0.7. To set one duplication scenario, the ρ_{12} had values 0.1 and 0.7, and to increase the number of duplications to two and three, ρ_{23} and ρ_{13} had values 0.2 and 0.6, respectively. Finally, the overall effects are chosen to be 0.2, 0.5, and 0.8 which are representing small, medium, and large gene-drug correlation association.

Simulating all combinations of parameter values leads to 432 scenarios for non-independent effects meta-analyses. The overall effect estimates were summarized across the 1,000 iterations. To evaluate the properties of the effect estimators, we estimated the relative mean squared error (MSE) and the coverage probability of the 95% confidence intervals by recording the percentage of replications where intervals included the true overall effect.

3 Results

3.1 Comparison of meta-analysis techniques: Independent studies

We performed gene-drug association analysis for each study using (adjusted) standardized linear regression model to obtain estimated effect and its standard error along with p -value per gene. Meta-analysis approaches were applied to integrate information across studies including Fisher and Stouffer methods to combine p -values and the RE model to integrate estimated effects across individual studies. Meta-analysis approaches differ on how they treat heterogeneity. Hence, various heterogeneity approaches such as DL, HS, PM, HE, SJ, and EB were compared (see Materials and Methods section).

To assess the performance of different meta-analysis approaches used to identify significant gene-drug associations using breast cancer and pan-cancer pharmacogenomic datasets, we compared the lists of genes significantly associated with drug response (FDR <0.05; Supplementary Figure S3). The p -value combination methods (Fisher and Stouffer) are almost the most conservative (least number of significant genes) meta-analyses. SJ is typically less conservative compared to p -value combining methods, while it is most conservative among the RE meta-analysis approaches. HS is always least conservative (greatest number of significant associated genes). In addition, a large percentage (~80–97%) of genes are found in common across the methods that are not most conservative. Other methods including HE, EB and PM demonstrate the similar pattern to DL method with more than 80% of commonly identified significant associated genes.

To compare the performance of different meta-analyses, we computed the Jaccard index to assess the similarity between the ranked lists of gene pairs obtained from different meta-analyses (Supplementary Figure S4). We find that the top 100 ranked gene-pairs are most similar with a range from 65% to 100%

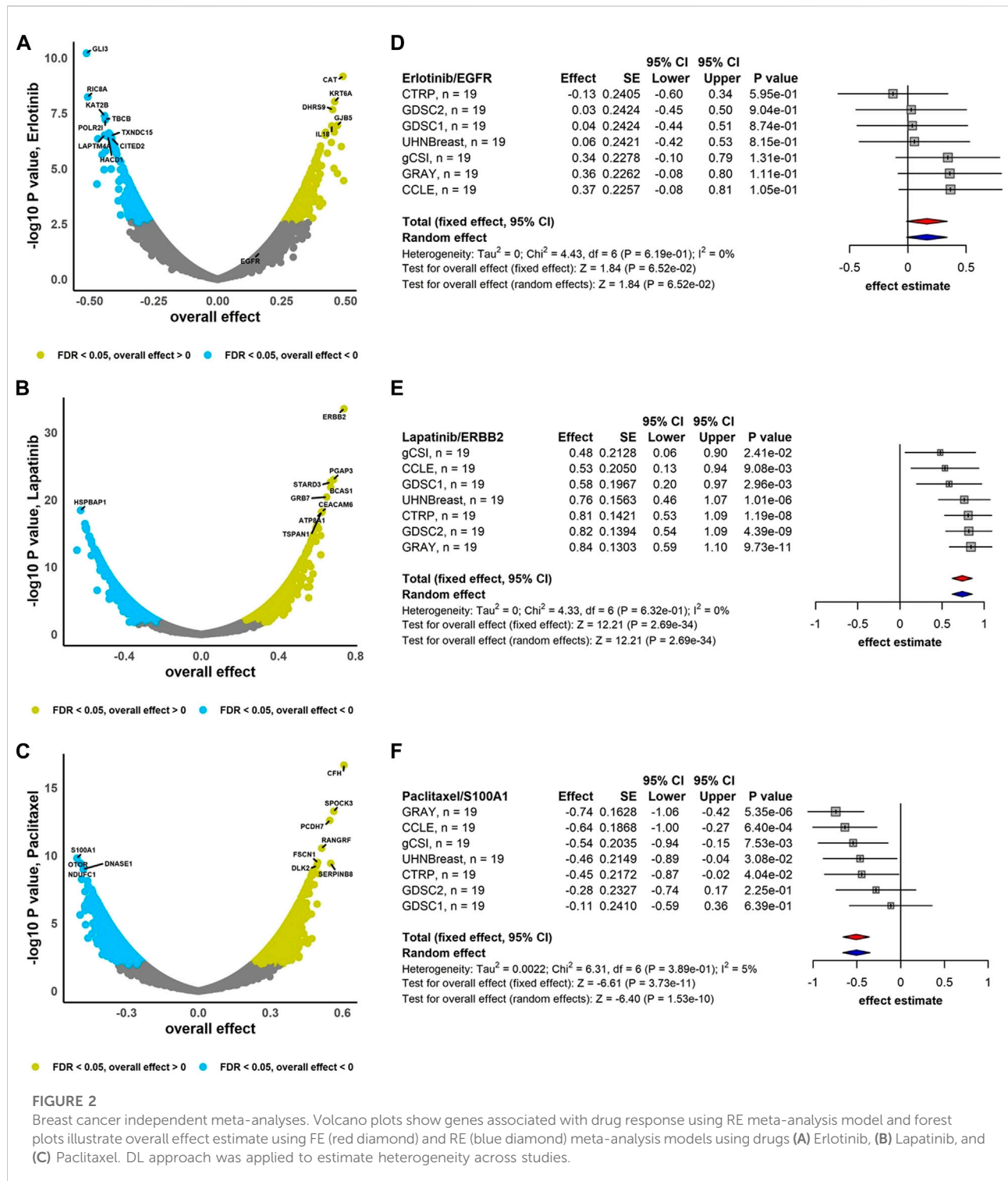
across combining effects approaches except for the error variance estimator SJ. This observation of similarity measures suggests that meta-analyses are not identical and there is considerable diversity between combining p -values and effects approaches. Different types of meta approaches had been reviewed and compared (Tseng et al., 2012; Sweeney et al., 2017). Hence, we chose the less conservative and most commonly used DL method to estimate heterogeneity for the rest of the meta-analyses.

3.2 Breast cancer biomarker discovery

We performed the association analysis between genes and drug response (AAC) using the RE meta-analysis model including the DL heterogeneity estimation approach to combine estimated effects (Pearson correlation coefficient) across breast cancer studies (Figure 2). Drugs Erlotinib and Lapatinib demonstrate that 60% and 62% of genes are negatively correlated with FDR <0.05, respectively, i.e., higher gene expression is associated with lower drug response, and therefore lower drug activity (Figures 2A,B). However, Paclitaxel with 59% of genes correlated positively with FDR <0.05 represents higher gene expression is associated with higher drug activity (Figure 2C). In addition, ERBB2 is highly sensitive to Lapatinib (overall effect = 0.74, 95% CI: 0.62 to 0.86, $p = 2.69\text{e-}34$), while Paclitaxel and S100A1 are negatively associated (overall effect = -0.51, 95% CI: -0.66 to -0.35, $p = 1.53\text{e-}10$) (Figure 2). Moreover, EGFR demonstrates less sensitivity to Erlotinib (overall effect = 0.16, 95% CI: -0.01 to 0.34, $p = 6.52\text{e-}02$) compared to the breast cancer meta-analyses where not limited to the common cell lines (overall effect = 0.26, 95% CI: 0.16 to 0.36, $p = 2.47\text{e-}07$) (Supplementary Figure S5). For breast cancer gene-drug association meta-analyses, around 8%–13% of genes have substantial estimated heterogeneity, where no greater than 1% of them are significantly associated with drugs. However, almost 10%–30% of genes are associated to drugs with non-substantial estimated heterogeneity (Supplementary Figure S6A).

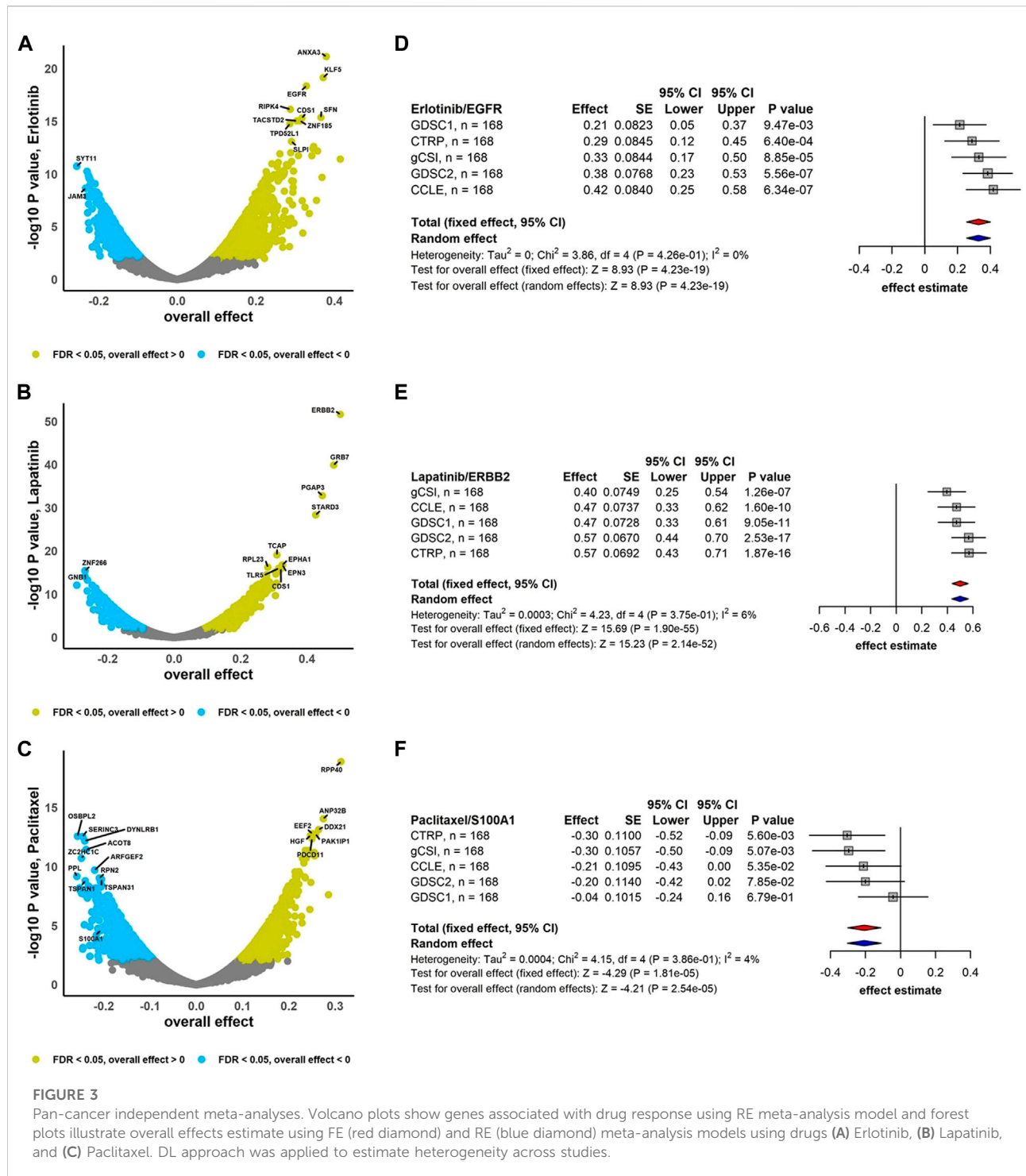
3.3 Pan-cancer biomarker discovery

We considered the integration of estimated effects to assess the gene-drug association across pan-cancer data using RE meta-analysis including the DL heterogeneity estimation approach (Figure 3). We obtained that 50% and 60% of genes are negatively associated (FDR <0.05) with Erlotinib and Lapatinib, respectively, i.e., higher gene expression was associated with lower drug activity, while 51% of genes are positively associated with drug Paclitaxel (Figure 3). EGFR and ERBB2 show considerable sensitivity to Erlotinib (overall effect = 0.33, 95% CI: 0.26 to 0.40, $p = 4.23\text{e-}19$) and Lapatinib (overall effect = 0.50, 95% CI: 0.44 to 0.56, $p = 2.14\text{e-}52$), respectively (Figure 3). However, S100A1 is significantly resistant to Paclitaxel (overall effect = -0.21, 95% CI:



-0.30 to -0.11, $p = 2.54e-05$) (Figure 3). Compared to the breast cancer meta-analysis, the pan-cancer gene-drug association meta-analyses contain less studies and more cell lines where almost 5%–13% of genes have substantial estimated

heterogeneity and around 1% or less of them are associated with drugs. Moreover, significant gene-drug associated with non-substantial estimated heterogeneity ranges between 16%–20% (Supplementary Figure S6B).



3.4 Comparison of frequentist and bayesian meta-analysis

To capture uncertainty in estimation of the heterogeneity, we applied Bayesian technique by incorporation Jeffreys prior information. We compared the performance of DL and

Bayesian approaches to estimate heterogeneity and overall effect along with 95% confidence or credible regions. DL and Bayesian procedures yield almost the same estimates for the overall effect (Supplementary Figure S7). For the breast cancer meta-analyses, the Bayesian credible interval is between 1.37 and 1.45 times wider than the DL interval, while as the number of

studies decreases using pan-cancer meta-analyses, the Bayesian interval gets wider from 1.66 to 1.68 times of DL interval. The conventional DL method detects no or low heterogeneity, while Bayesian estimate of I^2 ranges between 38% and 56%. In addition, across all genes, to compare the impact of Bayesian and DL methods, the length of 95% confidence and credible regions along with estimate of I^2 were computed (Supplementary Figures S8, S9). The median of length of Bayesian intervals is 1.33–1.40 times of DL interval using breast cancer data, while considering pan-cancer data, the median of length of Bayesian credible becomes around 1.52 to 1.54 times of DL interval across all drugs (Supplementary Figures S8, S9A). The ranges of median of I^2 Bayesian estimates are (36%–42%) and (47%–53%) using breast cancer and pan-cancer meta-analyses, respectively (Supplementary Figures S8, S9B). Our results are comparable to those reported by Bodnar et al. (2017) due to assuming the same priors.

3.5 Simulation study and non-independent effects

In traditional DL meta-analysis where the independence of effects was ignored, higher relative MSEs were observed as the number of non-independent effects increased (Figure 4, Supplementary Figures S10, S11). However, relative MSEs were similar across different numbers of non-independent effects in scenarios with small variation within and across studies and large numbers of studies. In scenarios with a small number of studies, relative MSEs were higher than scenarios with larger numbers of studies. In addition, as the variation within and across studies increased, the relative MSEs gradually rose. We also observed that large correlation across the effects can yield slightly higher relative MSE. As the overall effect increased, the relative MSEs had considerable decreases across all scenarios.

Coverage of the 95% confidence interval can differ by 5%–15% in all scenarios (Figure 4, Supplementary Figures S10, S11). As the number of non-independent effects increased, lower coverage was observed. Coverage had variations up to 10% between numbers of studies. Coverage varies between almost 90% and 96% when studies are homogeneous and up to two non-independent effects included and can be as low as 80% in scenarios with a small number of studies and three non-independent effects. In addition, large correlation across the effects can yield slightly lower coverage than smaller correlation. As the overall effect increased, the coverage decreased gradually in all scenarios.

The summary results indicated that the coverage of the 95% confidence intervals improved (or relative MSE decreased) as the number of studies increased, the variation within and across studies decreased, and had a smaller number of duplications across studies.

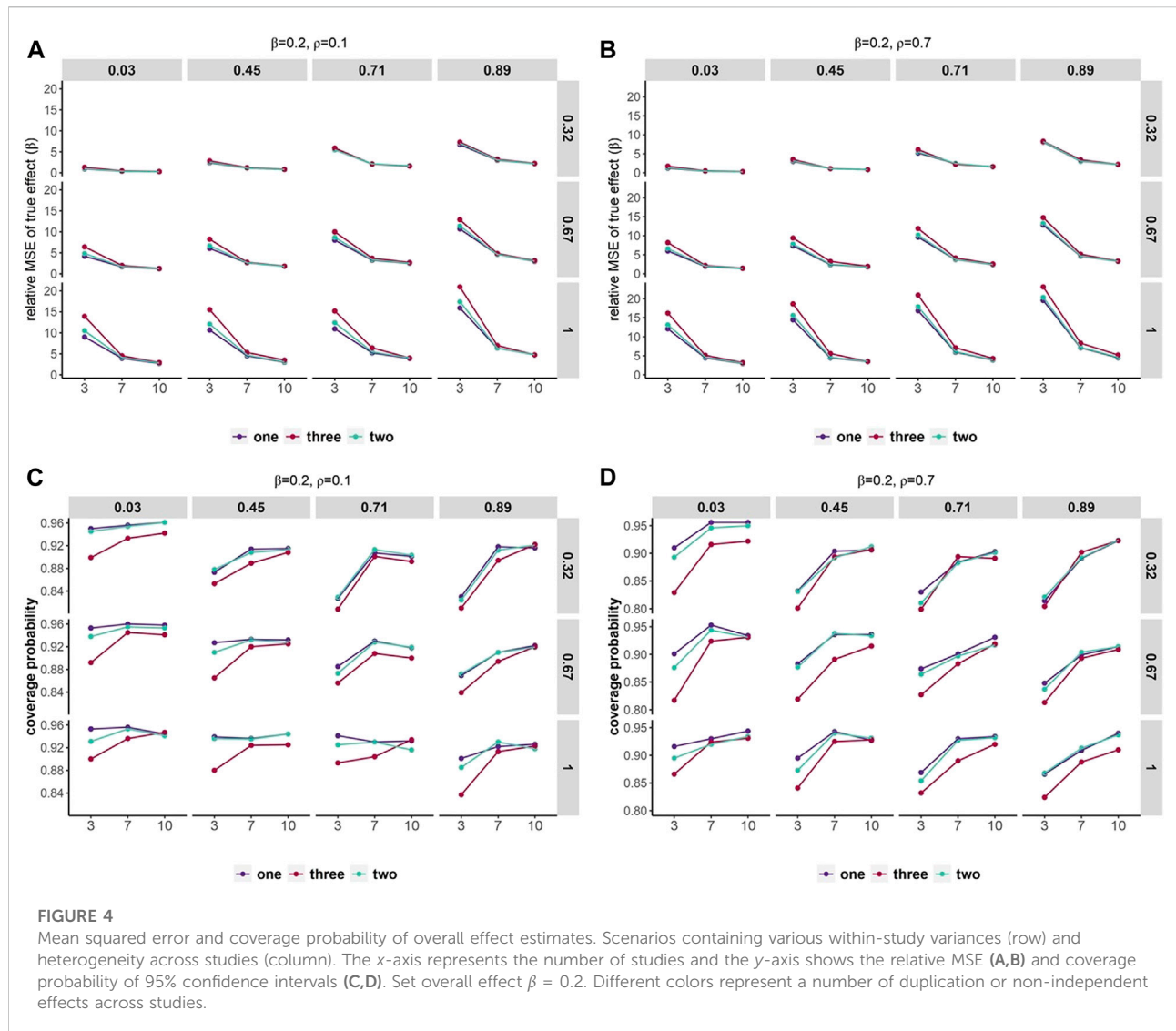
3.6 Duplicate study effects in a meta-analysis: Application in pharmacogenomic datasets

Duplication in study effects has been defined when estimated effect results from a complete replication of a particular study or from some subset of measured data. For instance, the generated expression duplicated in the study contains missing expression data for a given gene (Supplementary Figure S2B). To assess the bias of ignoring the dependence of the effects, we considered matched cell lines across non-independent studies for breast cancer and pan-cancer meta-analyses individually. Therefore, across whole genes, we considered all possible duplicated study effects and applied the RE model, including DL heterogeneity estimation approach to integrate the duplicate study effects, to estimate overall effect and heterogeneity. Reflecting on duplicate data problems, conclusions based on the traditional meta-analyses will bias the aggregated estimated effects and can increase the false positive results. Additionally, bias is determined using the mean distance between the estimated overall effect using non-independent studies and the overall effect by integrating independent studies (i.e., MAD). Additionally, we investigate the association of violating the independence assumption over the similarity of measured expression per gene across studies by applying the Pearson correlation. The median of estimated Pearson correlation can be classified as low $|r| < 0.3$, medium $0.3 \leq |r| < 0.7$ and high $|r| \geq 0.7$ (Mukaka, 2012).

3.6.1 Bias assessment in non-independent meta-analysis: Breast cancer and pan-cancer data

We assessed the increases in the number of duplication using breast cancer data and its impact on the bias of the estimated overall effect using the MAD across drugs and selected genes with estimated substantial and non-substantial heterogeneity (Figures 5A–C). The results indicate that increases in the number of duplicate study effects can considerably raise the bias of meta-estimates of effects. In addition, testing the trend of bias across all genes by growth in the number of non-independent studies denotes almost 97%–99% of genes following an increasing trend with p -value < 0.05 across drugs. Moreover, we obtain genes with higher median Pearson correlations that show on average less bias compared with genes that have low levels of median correlations across studies (Figure 5D).

Additionally, under the pan-cancer non-independent meta-analyses with more matched cell lines but fewer studies compared with breast cancer data, we obtain a similar increasing pattern in computed bias (Figures 6A–C). To evaluate the bias changes across whole genes when more duplicate study effects were added in meta-analyses, MK trend test results show almost 86%–93% of genes with p -value < 0.05 were detected across drugs. Moreover, similar to the breast cancer data, the average of bias across whole genes was reduced as the median of Pearson correlations increases (Figure 6D).



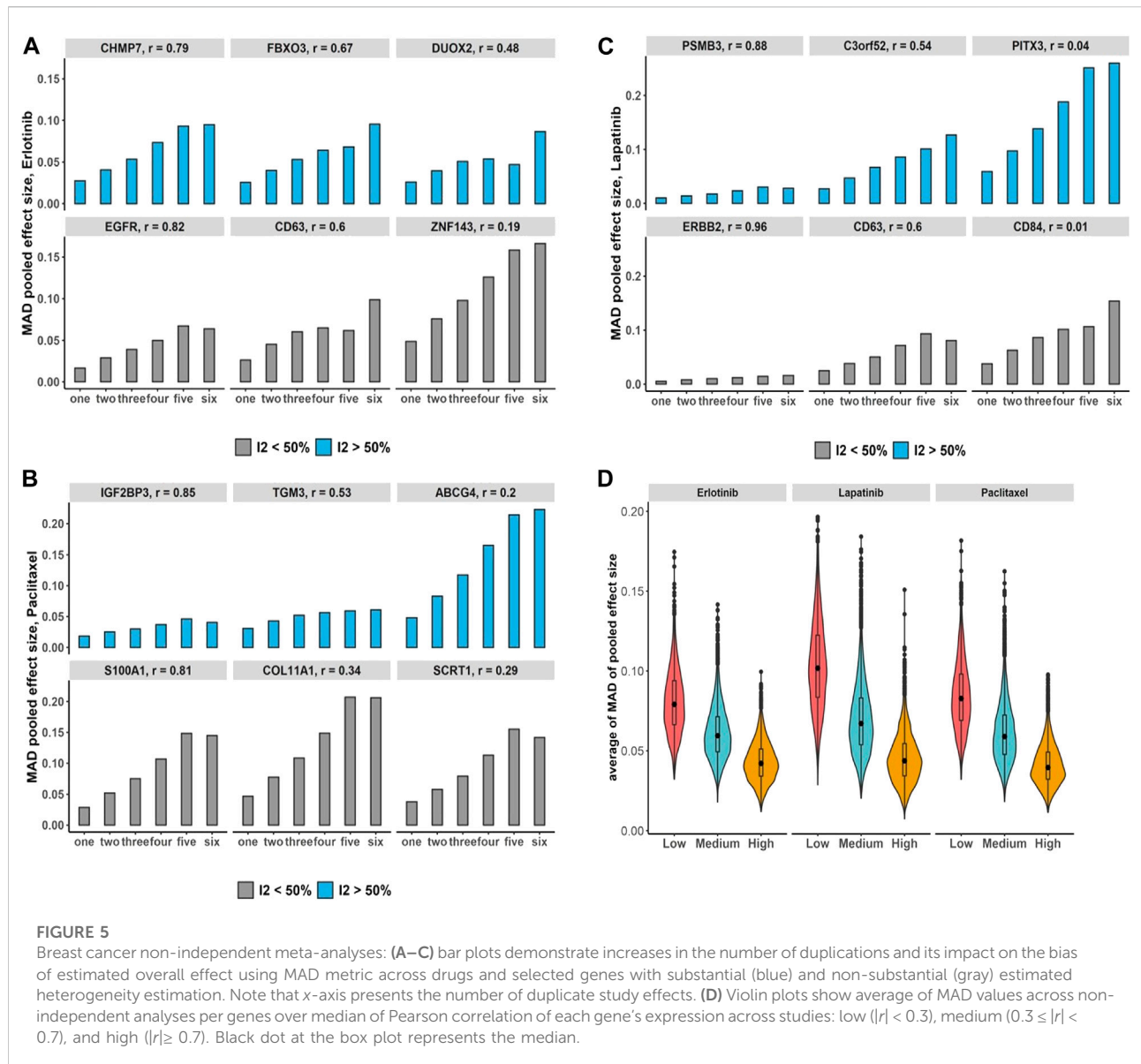
3.6.2 Non-independent meta-analysis and biomarker discovery

We evaluated whether violating the independence assumption in meta-analyses can impact on the differential expression analyses by considering Jaccard similarity index to compare the 100 top-ranked genes associated with drugs under non-independent and independent meta-analyses. The results illustrate that if we have more duplicate study effects, the overlap of detected expressed genes decreases and leads to a decrease in power of analyses (Figure 7).

3.6.3 Non-independent bayesian meta-analysis

We also investigated how duplicate study effects can bias the meta-analyses using the Jeffreys Bayesian approach and compared the results with traditional DL meta-analysis. Because Bayesian approaches take more time and are computationally

expensive, we only considered specific genes used by DL meta-analysis instead of including whole genes. We evaluated the pattern of changes in the number of duplicate study effects and the bias of the estimated overall effect using the Bayesian approach across drugs and selected genes. This process also considered estimated substantial and non-substantial estimated heterogeneity using breast cancer and pan-cancer data, respectively (Supplementary Figures S12, S13A–C). By increasing the number of duplicate study effects, bias of estimated overall effect will increase and genes with higher correlation across studies will indicate less bias. In addition, we observed no considerable trend for the average of 95% intervals using Bayesian and DL meta-analyses by increasing the number of duplicate study effects (Supplementary Figures S12, S13D). The results show that the median length of Bayesian intervals ranges almost between 2.4 and 2.5 times the DL interval using breast cancer data. When considering the pan-cancer data, the median

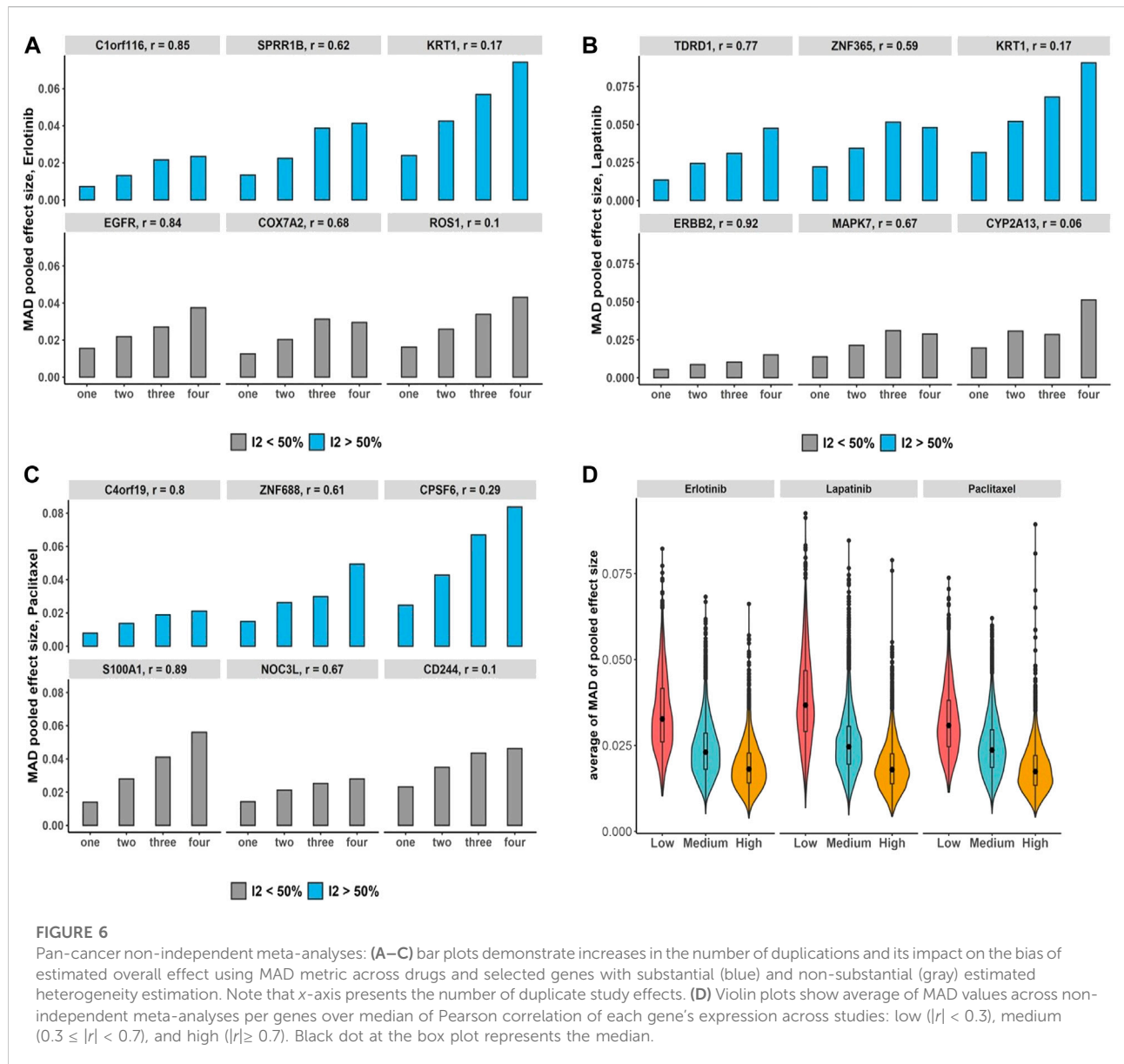


length of Bayesian interval becomes around 4.6 to 4.8 times of DL interval across all drugs.

4 Discussion

In this study, we reviewed and compared the performance of various traditional meta-analysis including frequentist and Bayesian approaches to improve the reproducibility of the identification of biomarkers using independent large-scale pharmacogenomic datasets. We observed that meta-analyses are not identical and there is considerable diversity between combining *p*-values and effect sizes approaches. We further

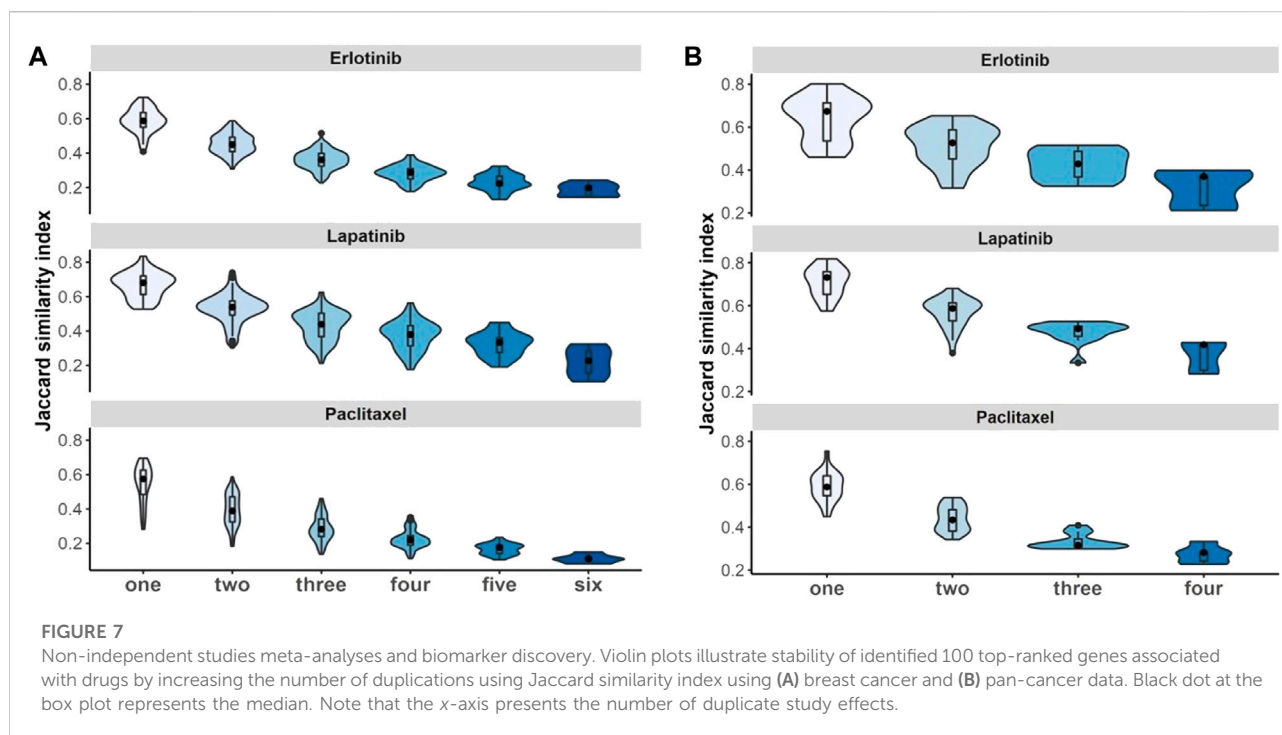
assessed the bias of including non-independent effects (or duplicate data) in a conventional meta-analysis. When effects are not independent, conclusions based on these conventional procedures will bias overall effect estimates and inflate the type I error rates (Becker, 2000; Wood, 2008; Lin and Sullivan, 2009; Hedges et al., 2010; Scammacca et al., 2014; Van den Noortgate et al., 2015; Cheung, 2019; Cooper et al., 2019; Liu et al., 2019; Wilson, 2019; Liu and Xie, 2020; Luo et al., 2020; Borenstein et al., 2021). We demonstrated how increases in the number of duplicated studies can impact the bias of overall estimate of effect and the identification of gene-drug associations. We also evaluated whether violating the independence assumption in meta-analyses can impact on the biomarker discovery.



Combining *p*-values and effects approaches are used to aggregate results from separate independent analyses. Although meta-analyses that combine *p*-values have widely been used before, they were not able to address the direction of effects and data heterogeneity (Marot et al., 2009). It was denoted that *p*-values combination methods (Fisher and Stouffer) are the most conservative and the HS RE model is the least conservative to identify genes associated with drugs. Among approaches for estimating between-study variance, SJ RE method is the most conservative method. Assessing heterogeneity is a critical issue in meta-analyses. Several statistical methods are routinely used to identify the statistical significance of heterogeneity which may lead to different estimates of overall effect and different standard errors

(Veroniki et al., 2016; Guolo and Varin, 2017; Langan et al., 2019). To address the uncertainty in the estimation of the heterogeneity in a RE model when studies have either small cell lines or are few in number, the Bayesian RE model was proposed (Bodnar et al., 2017). The results indicated that the 95% Bayesian confidence interval is considerably wider than the DL method, while both DL and Bayesian procedures yielded the same estimate for the overall effect. In addition, compared to the Bayesian approach, the DL method detected no heterogeneity which is comparable with reported results by Bodnar et al. (2017).

The most substantial outcomes of the non-independent studies are the significant differences in the estimates of the overall effect. From the results of the simulation study, as the



number of non-independent effects increased, higher relative MSE and lower-than-95% coverage probability were observed. However, almost equal relative MSEs across different numbers of non-independent effects were observed in scenarios with small variation within and across studies as well as large numbers of studies. Regarding the pharmacogenomics data analyses, the results indicated that by increasing the number of duplicate study effects, bias of overall effect will increase and genes with higher correlation denote less bias. In addition, when we have more duplicate study effects, the overlap of detected associated genes with drugs decreases and produces low power of analyses and more false positive findings.

Early meta-analytic researchers noted the non-independent study problem and suggested solutions. Doing nothing to correct dependent effects or accepting duplication inflate the Type I error and bias the estimated overall effect. Avoiding dependence by averaging effects (Wood, 2008) or shifting unit-of-analysis (Cooper et al., 2019) can reduce the variance between effects while informative differences get lost. Additionally, including the correlation between effects from the same study using multivariate approach, determining the covariance of effects across studies for overlap samples, and estimating the variance components (e.g., between-studies and within studies) were proposed to deal with non-independent effects in meta-analyses (Becker, 2000; Lin and Sullivan, 2009; Hedges et al., 2010; Van den Noortgate et al., 2015; Luo et al., 2020). The most

complex strategy is to model the dependence by proposing multilevel models where the correlation between effect sizes from the same study is not needed and separate estimates of the different variance components are estimated. However, when using multilevel models, it is very important to correctly specify all relevant random effects in the model (Cheung, 2014; Van den Noortgate et al., 2015; Cheung, 2019). Moreover, to combine p -values under arbitrary dependency structures, the Cauchy combination test (Liu et al., 2019; Liu and Xie, 2020) and the harmonic mean p -value (Wilson, 2019) were proposed. However, the proposed methods may be less powerful or even powerless under some conditions (Chen, 2022).

In conclusion, we should carefully define the inclusion and exclusion criteria and use these criteria to determine whether or not the studies or the effects should be included. We have to properly incorporate the dependence in a meta-analysis by including the covariance of non-independent effects or modeling the dependency. This research area remains challenging. Novel powerful and robust tests for combining non-independent effects are still highly desired.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://doi.org/10.24433/CO.2186077.v1>.

Author contributions

BH-K and FA-A designed the study. FA-A, BH-K, and WX analyzed data, interpreted data and wrote the manuscript.

Conflict of interest

BH-K is a shareholder and paid consultant for Code Ocean Inc. and is part of the SAB of the Break Through Cancer Foundation and the Consortium de Québécois recherche biopharmaceutique.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abbas-Aghababazadeh, F., Mo, Q., and Fridley, B. L. (2020). Statistical genomics in rare cancer. *Seminars cancer Biol.* 61, 1. doi:10.1016/j.semcancer.2019.08.021
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi:10.1038/nature11003
- Basu, A., Bodycombe, N. E., Cheah, J. H., Price, E. V., Liu, K., Schaefer, G. I., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154, 1151–1161. doi:10.1016/j.cell.2013.08.003
- Becker, B. J. (2000). “Multivariate meta-analysis,” in *Handbook of applied multivariate statistics and mathematical modeling*, 499–525.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
- Bodnar, O., Link, A., Arendacká, B., Possolo, A., and Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Stat. Med.* 36, 378–399. doi:10.1002/sim.7156
- Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2021). *Introduction to meta-analysis*. Oxford, United Kingdom: John Wiley & Sons.
- Brockwell, S. E., and Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Stat. Med.* 20, 825–840. doi:10.1002/sim.650
- Burgette, L. F., and Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *Am. J. Epidemiol.* 172, 1070–1076. doi:10.1093/aje/kwq260
- Burgette, L. F., and Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American journal of epidemiology* 172 (9), 1070–1076.
- Buuren, S. v., and Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in r. *J. Stat. Softw.* 1–68.
- Chang, L.-C., Lin, H.-M., Sibille, E., and Tseng, G. C. (2013). Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. *BMC Bioinforma.* 14, 368–415. doi:10.1186/1471-2105-14-368
- Chen, Z. (2022). Robust tests for combining p-values under arbitrary dependency structures. *Sci. Rep.* 12, 3158–8. doi:10.1038/s41598-022-07094-7
- Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychol. Rev.* 29, 387–396. doi:10.1007/s11065-019-09415-6
- Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol. Methods* 19, 211–229. doi:10.1037/a0032968
- Choi, H., and Pavelka, N. (2012). When one and one gives more than two: Challenges and opportunities of integrative omics. *Front. Genet.* 2, 105. doi:10.3389/fgene.2011.00105
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* 10, 101–129. doi:10.2307/3001666
- Cohen, A. L., Soldi, R., Zhang, H., Gustafson, A. M., Wilcox, R., Welm, B. E., et al. (2011). A pharmacogenomic method for individualized prediction of drug sensitivity. *Mol. Syst. Biol.* 7, 513. doi:10.1038/msb.2011.47
- Consortium of Drug Sensitivity in Cancer Consortium (2015). Pharmacogenomic agreement between two cancer cell line data sets. *Nature* 528, 84–87. doi:10.1038/nature15736
- Cooper, H., Hedges, L. V., and Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Dempster, J. M., Pacini, C., Pantel, S., Behan, F. M., Green, T., Krill-Burger, J., et al. (2019). Agreement between two large pan-cancer crispr-cas9 gene dependency data sets. *Nat. Commun.* 10, 5817–5914. doi:10.1038/s41467-019-13805-y
- DerSimonian, R., and Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemp. Clin. Trials* 28, 105–114. doi:10.1016/j.cct.2006.04.004
- DerSimonian, R., and Laird, N. (1986). Meta-analysis in clinical trials. *Control. Clin. Trials* 7, 177–188. doi:10.1016/0197-2456(86)90046-2
- Dewey, M. (2018). *Metap: meta-analysis of significance values*. R package version 1.
- Ding, M. Q., Chen, L., Cooper, G. F., Young, J. D., and Lu, X. (2018). Precision oncology beyond targeted therapy: Combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Mol. Cancer Res.* 16, 269–278. doi:10.1158/1541-7786.MCR-17-0378
- Egger, M., Davey-Smith, G., and Altman, D. (2008). *Systematic reviews in health care: meta-analysis in context*. Oxford, United Kingdom: John Wiley & Sons.
- Fisher, R. A., et al. (1934). *Statistical methods for research workers*. New York, NY: Springer
- Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483, 570–575. doi:10.1038/nature11005
- Geyer, C. E., Forster, J., Lindquist, D., Chan, S., Romieu, C. G., Pienkowski, T., et al. (2006). Lapatinib plus capecitabine for her2-positive advanced breast cancer. *N. Engl. J. Med.* 355, 2733–2743. doi:10.1056/NEJMoa064320
- Guolo, A., and Varin, C. (2017). Random-effects meta-analysis: The number of studies matters. *Stat. Methods Med. Res.* 26, 1500–1518. doi:10.1177/0962280215583568
- Haibe-Kains, B., El-Hachem, N., Birkbak, N. J., Jin, A. C., Beck, A. H., Aerts, H. J., et al. (2013). Inconsistency in large pharmacogenomic studies. *Nature* 504, 389–393. doi:10.1038/nature12831
- Hatzis, C., Bedard, P. L., Birkbak, N. J., Beck, A. H., Aerts, H. J., Stern, D. F., et al. (2014). Enhancing reproducibility in cancer drug screening: How do we move forward? *Cancer Res.* 74, 4016–4023. doi:10.1158/0008-5472.CAN-14-0725
- Haverty, P. M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., et al. (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* 533, 333–337. doi:10.1038/nature17987
- Hedges, L. V., and Olkin, I. (2014). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1027345/full#supplementary-material>

- Hedges, L. V., Tipton, E., and Johnson, M. C. (2010). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Res. Synth. Methods* 1, 164–165. doi:10.1002/jrsm.17
- Heiser, L. M., Sadanandam, A., Kuo, W.-L., Benz, S. C., Goldstein, T. C., Ng, S., et al. (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2724–2729. doi:10.1073/pnas.1018854108
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. doi:10.1136/bmj.327.7414.557
- Hoaglin, D. C. (2016). Misunderstandings about q and 'cochran's q test' in meta-analysis. *Stat. Med.* 35, 485–495. doi:10.1002/sim.6632
- Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L., and Chory, J. (2006). Rankprod: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825–2827. doi:10.1093/bioinformatics/btl476
- Schmidt, F. L., and Hunter, J. E. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. California, United States: Sage.
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754. doi:10.1016/j.cell.2016.06.017
- Iqbal, S., Goldman, B., Fenoglio-Preiser, C., Lenz, H., Zhang, W., Danenberg, K., et al. (2011). Southwest oncology group study s0413: A phase ii trial of lapatinib (gw572016) as first-line therapy in patients with advanced or metastatic gastric cancer. *Ann. Oncol.* 22, 2610–2615. doi:10.1093/annonc/mdr021
- Jaiswal, A., Gautam, P., Pietilä, E. A., Timonen, S., Nordström, N., Akimov, Y., et al. (2021). Multi-modal meta-analysis of cancer cell line omics profiles identifies ecdh1 as a novel breast tumor suppressor. *Mol. Syst. Biol.* 17, e9526. doi:10.15252/msb.20209526
- Jia, P., Hu, R., Pei, G., Dai, Y., Wang, Y.-Y., and Zhao, Z. (2021). Deep generative neural network for accurate drug response imputation. *Nat. Commun.* 12, 1740–1816. doi:10.1038/s41467-021-21997-5
- Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., et al. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* 10, 83–98. doi:10.1002/jrsm.1316
- Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., et al. (2017). Characterization of human cancer cell lines by reverse-phase protein arrays. *Cancer Cell* 31, 225–239. doi:10.1016/j.ccell.2017.01.005
- Lin, D.-Y., and Sullivan, P. F. (2009). Meta-analysis of genome-wide association studies with overlapping subjects. *Am. J. Hum. Genet.* 85, 862–872. doi:10.1016/j.ajhg.2009.11.001
- Liu, Y., Chen, S., Li, Z., Morrison, A. C., Boerwinkle, E., and Lin, X. (2019). Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am. J. Hum. Genet.* 104, 410–421. doi:10.1016/j.ajhg.2019.01.002
- Liu, Y., and Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* 115, 393–402. doi:10.1080/01621459.2018.1554485
- Luo, L., Shen, J., Zhang, H., Chhibber, A., Mehrotra, D. V., and Tang, Z.-Z. (2020). Multi-trait analysis of rare-variant association summary statistics using mtar. *Nat. Commun.* 11, 2850–2911. doi:10.1038/s41467-020-16591-0
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica* 13, 245–259. doi:10.2307/1907187
- Marcotte, R., Sayad, A., Brown, K. R., Sanchez-Garcia, F., Reimand, J., Haider, M., et al. (2016). Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell* 164, 293–309. doi:10.1016/j.cell.2015.11.062
- Marot, G., Foulley, J.-L., Mayer, C.-D., and Jaffrézic, F. (2009). Moderated effect size and p-value combinations for microarray meta-analyses. *Bioinformatics* 25, 2692–2699. doi:10.1093/bioinformatics/btp444
- Medina, P. J., and Goodin, S. (2008). Lapatinib: A dual inhibitor of human epidermal growth factor receptor tyrosine kinases. *Clin. Ther.* 30, 1426–1447. doi:10.1016/j.clinthera.2008.08.008
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., et al. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol.* 17, 53–59. doi:10.1186/s13059-016-0917-0
- Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* 24, 69. doi:10.4314/MMJ.V24I3
- Paule, R. C., and Mandel, J. (1982). Consensus values and weighting factors. *J. Res. Natl. Bur. Stand.* 87, 377–385. doi:10.6028/jres.087.022
- Peterson, R. A., and Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *J. Appl. Psychol.* 90, 175–181. doi:10.1037/0021-9010.90.1.175
- Pohlert, T. (2016). Non-parametric trend tests and change-point detection. *CC BY-ND* 4.
- Röver, C. (2017). *Bayesian random-effects meta-analysis using the bayesmeta r package*. *arXiv preprint arXiv:1711.08683*.
- Safikhani, Z., Smirnov, P., Freeman, M., El-Hachem, N., She, A., Rene, Q., et al. (2016). Revisiting inconsistency in large pharmacogenomic studies. *F1000Res.* 5, 2333. doi:10.12688/f1000research.9611.1
- Safikhani, Z., Smirnov, P., Thu, K. L., Silvester, J., El-Hachem, N., Quevedo, R., et al. (2017). Gene isoforms as expression-based biomarkers predictive of drug response *in vitro*. *Nat. Commun.* 8, 1126–1211. doi:10.1038/s41467-017-01153-8
- Scammacca, N., Roberts, G., and Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Rev. Educ. Res.* 84, 328–364. doi:10.3102/0034654313500826
- Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell* 103, 211–225. doi:10.1016/s0092-8674(00)00114-8
- Schwarzer, G., Balduzzi, S., Rütke, G., et al. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-based mental health* 22 (4), 153–160.
- Sidik, K., and Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Stat. Med.* 26, 1964–1981. doi:10.1002/sim.2688
- Smirnov, P., Safikhani, Z., El-Hachem, N., Wang, D., She, A., Olsen, C., et al. (2016). Pharmacogx: An r package for analysis of large pharmacogenomic datasets. *Bioinformatics* 32, 1244–1246. doi:10.1093/bioinformatics/btv723
- Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., and Williams, R. M., Jr (1949). *The American soldier: Adjustment during army life (studies in social psychology in world war ii, 1*. Princeton, NJ: Princeton University Press.
- Sutton, A. J., and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Stat. Methods Med. Res.* 10, 277–303. doi:10.1177/096228020101000404
- Sweeney, T. E., Haynes, W. A., Vallania, F., Ioannidis, J. P., and Khatri, P. (2017). Methods to increase reproducibility in differential gene expression via meta-analysis. *Nucleic Acids Res.* 45, e1. doi:10.1093/nar/gkw797
- Tseng, G. C., Ghosh, D., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res.* 40, 3785–3799. doi:10.1093/nar/gkr1265
- Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., and Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane database of systematic reviews. *Int. J. Epidemiol.* 41, 818–827. doi:10.1093/ije/dys041
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., and Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behav. Res. Methods* 47, 1274–1294. doi:10.3758/s13428-014-0527-2
- Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., et al. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* 7, 55–79. doi:10.1002/jrsm.1164
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Statistics* 30, 261–293. doi:10.3102/10769986030003261
- Volk-Draper, L., Hall, K., Griggs, C., Rajput, S., Kohio, P., DeNardo, D., et al. (2014). Paclitaxel therapy promotes breast cancer metastasis in a tlr4-dependent manner. *Cancer Res.* 74, 5421–5434. doi:10.1158/0008-5472.CAN-14-0067
- Weaver, B. A. (2014). How taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell.* 25, 2677–2681. doi:10.1091/mbc.E14-04-0916
- Whitehead, A., and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Stat. Med.* 10, 1665–1677. doi:10.1002/sim.4780101105
- Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc. Natl. Acad. Sci. U. S. A.* 116, 1195–1200. doi:10.1073/pnas.1814092116
- Won, S., Morris, N., Lu, Q., and Elston, R. C. (2009). Choosing an optimal method to combine p-values. *Stat. Med.* 28, 1537–1553. doi:10.1002/sim.3569
- Wood, J. (2008). Methodology for dealing with duplicate study effects in a meta-analysis. *Organ. Res. Methods* 11, 79–95. doi:10.1177/1094428106296638
- Xia, F., Allen, J., Balaprakash, P., Brettin, T., Garcia-Cardona, C., Clyde, A., et al. (2022). A cross-study analysis of drug response prediction in cancer cell lines. *Briefings in bioinformatics* 23 (1), bbab356.
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2012). Genomics of drug sensitivity in cancer (gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111