Health Biotechnology and Biopharma (2019), 3(3): 12-49

Original Research Article

A novel approach for analysis of shared genetic architecture of drug response based on summary statistics from genome-wide association studies

Mohammad Afzal [1,*], Soroush Sardari [2], Shomaila Sikandar [3], Imran Afzal [3]

[1]Consultant for Health Research and M&E, Islamabad, Pakistan
[2]Drug Design and Bioinformatics Unit, Medical Biotechnology Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran
[3]Department of Biology, Lahore Garrison University, Lahore, Pakistan

***Corresponding author:**. Mohammad Afzal, Consultant for Health Research and M&E, Islamabad, Pakistan.
Email: cmafzal@gmail.com*

DOI: 10.22034/HBB.2019.21

Received: November 2, 2019; **Accepted:** November 29, 2019

ABSTRACT

The main objective of this study is to gain an exploratory sense of the relationships between the phenotypes (response to drugs) employing the genome-wide association studies (GWAS) summary statistics for searching underlying contribution of pleiotropy across the associated SNPs to this relationship. The findings in this study show that the effect size should be weighted appropriately, depending on the nature of assumption made for conducting phenotype clustering based on GWAS summary. Thus the methods used were helpful for exploring workable clusters of related drugs on the basis of the effect size of the shared associated SNPs when weighted with SNP-sharing strength. The presence of strong correlation between the weighted fixed effect size and the principal components provides a possibility of using principal component regression for predicting the effect size of association of a drug response with SNPs. Overall, the phenotypic pattern of drugs and associated SNPs revealed by this study are expected to help in future for understanding aetiological basis of different drugs by highlighting relevant biological pathways.

***Keywords*:** Genome-wide association studies, cross-phenotype associations, effect size

INTRODUCTION

For last several decades researchers have reported variability in response to drugs between different human subjects. Historically, in 510 B.C. Pythagoras observed that eating fava beans caused severe sickness and death of some consumers [1]. It was two millennia later when a defect in the glucose-6-phosphate dehydrogenase enzyme was discovered and it was associated with haemolytic anaemia following the consumption of fresh fava beans as well as exposure to some drugs such as aspirin, phenacetin or primaquine [1,2].

Many factors affect response to a drug, including individual's genetic makeup, age, BMI, food/beverage intake, smoking, dietary supplements, reaction with other drugs, drug storage, tolerance and resistance to drug, physical condition (disease, fatigue, trauma) [3]. Some of the factors like tobacco use or alcohol intake can be controlled but factors like age, health condition (including other diseases) and genetic factors cannot be changed [4]. Practically, the response to drugs is determined by three factors: i) the phenotype (usually the disease), ii) the responsiveness of tissues and iii) the drug concentration at the site of action (reflected in the plasma). The relative response to drugs does not only differ between drugs but also within a drug [5].

Response to drugs is affected by genetic variants in few known ways: i) by affecting the influx and efflux of drug transporters within and outside of a cell; ii) altering drug metabolizing enzymes that control the chemical conversions of chemotherapeutics for their degradation and ultimate excretion; iii) how well a drug works as intended by altering the amount of the target protein or by delivering therapy, and iv) affecting allergic response to certain drugs [6].

In the last two decades, Genome-wide association studies (GWAS) have provided powerful tools for exploring the impact of single nucleotide polymorphism (SNP) on individual drug response and for the investigating targets and effects of drugs [7,8]. Therefore, the use of the phrase 'response to drug' or shortly 'drug response' (DR) throughout this study refers to combined effect of all the determining factors reported in GWAS.

Pharmacogenomics GWAS (PGx GWAS) mostly deals with changes in response due to "off-target" genes diverting from the intended target, genetic changes in in the downstream pathway of a drug, or changes in pharmacodynamics or pharmacokinetics [12]. Although only little more than 10% of

GWAS have focused on PGx GWAS [13,14], scientific community has ever increasing interest in predicting targets identification of drugs and rely more and more on genetic evidence. One of the major contributions of PGx GWAS is that drugs with genetically supported targets identified by GWAS have higher likelihood to receive clinical approval [9-11]. A publically available online database, the Pharmacogenomics Knowledgebase, responsible for summarizing regarding the impact of human genetic variation on drug response, to date reported more than 1,100 genetic loci having significant association with response to drug [15]. However, clinical implications of the associated variants are very small which may be affected by their linkage disequilibrium with the causal variants [16,17].

PGx GWAS is further constrained due to rarity of drug related phenotype and collection of sufficient samples [18,13,19]. Therefore, sample size required for successful PGx GWAS to obtain association with drug response phenotypes, necessitates concerted international collaboration, similar to those for complex human diseases. Consequently, various networks and consortia have been formed by GWAS researchers to undertake after aggressive data sharing across same or related diseases or traits [20]. For example, an international consortium, PGRN-RIKEN, has conducted scores of PGx GWAS and supported the largest number of PGx studies since 2008 which are reported in the GWAS catalog [18]. Likewise, international consortia like the International Serious Adverse Events Consortium [21], EUDRAGENE [22], and others [23,24] mediate the process of collective sampling. Despite the sampling challenges, the available samples have identified genetic risk factors for some adverse drug reactions (ADRs) [2,18,25,26].

Extensive pleiotropy among the common variants associated with complex traits has been reveled [27,28]. The studies involving pleiotropic associations are called cross-phenotype associations (CPA), i.e., testing association of multiple traits simultaneously influenced by many genetic variants, such as single nucleotide polymorphisms (SNPs). There is a growing evidence that more power can be gained by CPA associations [29-33,34].

Considering growing evidence for widespread pleiotropy among pharmacogenes [9,27], the main objective of this study is to gain an exploratory sense of the relationships between the phenotypes (response to drugs) employing the GWAS summary statistics and searching underlying contribution of pleiotropy across the

associated SNPs to this relationship [35]. We specifically treat GWAS binary phenotype (response to drug) as dependent and SNP effect size (SNP score) as independent variables. In the remainder of this article we frequently use the shortened terms 'drug response' and 'SNP score' for brevity.

MATERIALS AND METHODS

Methods of data collection and analyses used in this study are described along with the background and justification as below.

GWAS Data

Most of the multi-variant methods are employed to analyze GWAS summary data collected from databases available in public domain, e.g., the National Human Genome Research Institute (NHGRI) and The European Bioinformatics Institute (EBI) database, the NHGRI-EBI GWAS Catalog [36], with particular reference to the minor allele frequency (MAF), effect size, standard errors, p-values for associated variants (mostly SNPs) analyzed in a GWAS [29,33]. The GWAS Catalog is a manually curated GWAS genetic evidence database widely used by researchers to identify new candidate loci to predict disease risk, identify rate of success in drug targets and undertake trait analysis. As of 2019-04-21, the GWAS Catalog recorded 3955 publications and identifying 136287 statistically significant associations between SNP risk alleles and the complex traits (at p-value threshold of 5×10^{-8}). Summaries collected from GWAS Catalog have provided useful results GWAS meta-analysis methods and useful results [20,29,30,33,34,36-40,41,42,51,52,53,54].

For this study summary statistics related to response to drugs were collected from the GWAS Catalog. The data were downloaded as the tab-separated values (TSV) GWAS Catalog file (gwas_catalog_v1.0.2-associations_e96_r2019-09-24.tsv). The data file was converted to MS Excel for tabulation and filtered according to the fields related to the phenotype (response to drug), namely, 'study', disease/trait' and 'mapped trait' along with the fields addressing sample size, strongest SNP-risk allele, mapped gene, risk allele frequency, odds ratio/beta-coefficient, standard error of the effect size and p-value for all associated SNPs. To make sure the GWAS chosen for this study are well-powered, we only included the SNPs that had at least 6 hits on the genome-wide significant level of $p \leq 10^{-6}$.

While downloading a data set with all SNPs related to multiple traits, there is always a possibility of a particular SNP being reported more than once in GWAS catalog due to contribution of multiple studies [55].

Therefore, we developed an Excel Macro to explore a unique set of SNPs associated with response to drugs. The data were then filtered according to a set comprising 6762 SNPs strongly associated with response to 128 drug. This set of SNPs was assumed as a random sample from a population SNPs strongly associated with drug response. All drugs used for analysis in this study are enlisted in Table 1 with their abbreviations.

Since values of odds ratio/beta-coefficient values were not available for all the short listed 6762 SNPs, the data were further filtered out to according to these statistics for 5663 SNPs associated with response to 113 drugs. This set of 5663 SNPs was also treated as a random sample from a population SNPs strongly associated with 113 drugs.

For defining variables use in this study, we explain dependent vector Y. Each $Y_j$denotes the drug response (binary variable, 1 = response, 0 = no response), where $j = 1, 2, 3, \dots, k$ (number of drugs). The explanatory vector X denotes SNP score represented by X_i, where $i = 1, 2, 3, \dots, n$ (number of SNPs). The input to the model are, therefore, the 'SNP scores' (a set of effect size estimates for each SNP) on each of the phenotypes (response to drug) in the sample. Each explanatory vector for a drug is, therefore, a component of the matrix $M = n \times k$, used as the feature vector in this study. All feature vectors of M are assumed to be independent due to linkage disequilibrium (LD)-based pruning in GWAS Catalog [36]. Since index SNPs likely map to the same genetic locus in GWAS Catalog, we assume a linear polygenic structure of M between Y and X and describe as below [56].

Let $y_j = (y_{j1}, \dots, y_{jn})$ denote the vector of measures on the j^{th} drug for n SNPs.

We assume that each drug response is modeled by linear regression denoted as $y_j = x_{ij}\beta_j + \varepsilon_j$. We also assume that all y_j's and x_i's are centered, so intercepts can be ignored. For simplicity, we assume absence of any other fixed effect in the model and ignore adjusting covariates, or other drug specific covariates can be equivalently observed and adjusted for. By stacking vectors, the model can be expressed as $Y = \boldsymbol{X}\beta + \varepsilon$, where Y is a vector of drug response , $\boldsymbol{X}$ is a matrix of effect size (fixed or random), β is coefficient vector of effect size and ε is residual [56-59],. The null hypothesis is H_0: β = 0 and the alternative hypothesis H_1 is that at least one of the elements of β ≠ 0. A standard method to test β = 0 is the Wald test statistic [37,60].

The sample sizes in GWAS Catalog are available only as text under the field "Initial Sample Size". As the description of samples in the text (number of cases, controls, population, etc.) is quite consistent, an Excel Macro was developed to extract the sample sizes.

Various effect size (ES) measures for SNP versus drug response association are employed in the statistical analysis [61-63]. Odds ratio (OR) is often referred to as standardized measures of effect. It is also possible to convert OR to other measure of Pearson correlation coefficient (r) and Cohen's d and z score (z) [64-67], while the z scores can be worked out with available sample sizes and p values.

Although different effect size measures are related their values are not directly comparable due to different scales (metrics) used in the calculation. For example, OR measures the effect when the phenotype (dependent variable) is binary, i.e., measures the difference between the mean values of two groups. Instead, a correlation coefficient is a measure of the relationship between the state of the phenotype and the outcome [64]. For conducting meat-analysis dichotomous data, GWAS investigators prefer odds ratio and the log odds ratio (L_{or}) over correlation coefficient and Cohen's d [68]

Most conversion from OR to other measures of effect size are mediated through log odds ratio (L_{or}) [64-67], since log odds ratio (L_{or}) has an approximately normal distribution:

$$L_{or} \sim \mathcal{N}\ (\log(OR), \sigma^2)$$

Parallel with odds ratio, Cramér's (phi) is also used to measure effect size and it is a probabilistic measure like point correlation, i.e., [0, 1] [65]:

$$\emptyset \equiv \sqrt{\frac{\chi^2}{N\times(k-1)}} \qquad (1)$$

Where, χ^2 is chi-square value, and N is the sample size and k the smaller of the vertical or horizontal categories of the contingency table [61,65]. When association analysis is based on 2 × 2 table ($k = 1$), the equation simplifies to

$$\emptyset = \sqrt{\frac{\chi^2}{N}} \qquad (2)$$

Statistical Methods for GWAS Meta-Analysis

In the ideal situation, when genotyping results of all studies together are available, a "mega-analysis can be undertaken to directly combine data from the different populations. In this case meta-analysis would be redundant [69]. However, individual-level data for a GWAS are often not available as GWAS investigators may be reluctant or

unavailable to share the raw data. Besides, difficulty to retrieve, cost or time involved to procure the data may cause further constraints. This is particularly true for the individual level data for most of the PGx GWAS [13,18,19]. Fortunately, the earlier perception that mega-analysis involving individual-level data is statistically more powerful than meta-analysis based on GWAS summary, is no longer considered effective [53,20,70]. For conducting statistical analysis, meta-analyzed GWAS summary results prove equally efficient as mega-analysis of individual studies provided same modeling assumptions are used both types of analyses [20,71].

A researcher would need convert individual studies into a common metric before undertaking statistical meta-analysis to synthesize results from different GWAS estimates [64,72]. Meta-analysis of GWAS in principle combines values of association from individual studies using appropriate weights followed by normalization for different sample sizes [30,40,73]. For clustering and estimation of genetic correlation from GWAS summary statistics for a polygenic traits, it is usually assumed that the meta-analyzed effect size value of a SNP embodies the influence of all SNPs in LD with that SNP [38,39,74]. Selecting GWAS for analysis necessitates filtering summary statistics on the basis that statistical power related to the index SNPs having desired genome-wide significant (usually ranging from $p \leq 10^{-6}$ to $p \leq 10^{-8}$). Further filtering may involve colocalization of multiple index SNPs using appropriate methods [55,75-78,45-51,79]. With the availability of large numbers of GWAS summary statistics there is a growing trend to conduct meta-analysis to explore pleiotropy and explicate the genetic relationship among traits [19,30,35,80].

As the collected genetic variants come from a large number of GWAS with different sample sizes, meta-analysis of all individual studies is usually undertaken for determination of combined effect size using normalization of sample sizes and appropriate weights, often employing specific analysis tool sets [36,40,52,68,71]. Selecting GWAS for meta-analysis necessitate filtering summary statistics on the basis that statistical power related to the index SNPs with desired genome-wide significance (usually ranging from $p \leq 10^{-8}$ to $p \leq 10^{-6}$). Further filtering may involve colocalization of multiple index SNPs using appropriate methods [45,48,51,55,75-78,79].

Meta-analysis of GWAS datasets requires a number of pieces of information related to epidemiological study design (cohort, case-

control, cross-sectional studies or clinical trials), quality checks, analytical methods, population structure, genotyping methods, amputation techniques, strand and build of the human genome [81].

In this study the odds ratio (OR) and Cramér's are considered the effect size (t_i) attributed to a strongest SNP-risk allele associated with response to the drug. It is assumed that t_i is determined as

$$t_i = \mu + \varepsilon_i$$

where μ is population mean and ε_i is the within-study error [69,70,82].

Most of the studies concerning clustering and genetic correlation of GWAS data are based on colocalization of genomic regions. Besides, larger GWAS are often weighted more than the smaller ones [69,82]. It is also argued that an association of a variant to a pair of traits does not establish that the variant is causal for both the traits due to possibility linkage disequilibrium of distinct variants [78,83]. Thus, effect size values, whether assessed directly or through or meta-analysis, with or without involving colocalization, may incorporate a lot of noise due underlying assumptions and rounding errors [45,48,51,75,78,79]. On the other hand direct use of a feature/variable, e.g., effect size summary, can prove equally noisy due to the coalition process. This necessitated preprocessing the GWAS summary statistics to de-noise the input data in this study as below.

To circumvent the noise in the input data, we hypothesize in this study that weighting effect size of SNPs with the strength of sharing can be more efficient for clustering or correlation analysis of the phenotype. We assume that if a set of drugs is associated with a set of SNPs, the effect size of each SNP should be weighted by the number of drugs associated with it ($l =$ size of drug group). However, the more the number of SNPs influencing a drug group ($\eta =$ size of the drug group associated SNPs) the more will be the combined effect. Therefore, taking into account the influence of the drug group (l_b, $where\ b \in \{1, 2, 3\}$ and corresponding size of the drug group associated SNPs ($\eta_{l_b}, where\ l_b = 1,2,3, \dots, \eta$), we define strength of sharing (S) as follows:

$$S = \frac{l_b}{\eta_{l_b}} \qquad (3)$$

Both measures of effect size in this study, OR and ϕ, are weighted with a factor of m, henceforth referred to as SNP-sharing strength (SSS). For brevity in the remainder of this article, the effect size weighted as above is referred to τ_{ij}:

$\tau_{ij} = S * OR_{ij}$ or $\tau_{ij} = S * \phi_{ij}$ (4)

Where $i = 1, 2, 3, \ldots, n$ (number of SNPs) and $j = 1, 2, 3, \ldots, k$ (number of drugs).

The principle adapted in different methods for meta-analyses is that of bivariate meta-analysis for handling the correlation between outcomes, both binary and to continuous outcomes [73,84]. Other than the standard weighting used in meta-analysis, most of existing methods employ different weighting schemes to test the effect of combination of variants on single or multiple traits [35,39,74,85]. The performance of these methods depends on the selection and application of weights. For example, large weights on rare variants and small on common variants can conceal certain characteristics of complex diseases and traits as these are caused by both types of variants. Therefore, optimal weights have been suggested to overcome these limitation [35].

Inverse variance weighting is normally considered optimal in the meta-analysis [86]. The practicality of such weighting in GWAS meta-analysis is well documented for initial screenings and discovery [73,87]. The optimal weights from score statistics representing small effect sizes essentially have power comparable to the inverse variance weighting [73,88]. Normally, weights are derived from the sample sizes as follows [30,66]:

$$w_{ij} = \sqrt{n_j}, \quad (5)$$

Where, for example, n_j is the of the sample size of j[th] drug in this study. In this way larger studies are weighted more than the smaller ones. GWAS investigators who derive weights from sample size argue that a variance is dependent on the scale of measurement. However, this limitation of inverse variance weighting can be overcome by using the effect size (τ_{ij}) and sample size (n_j) in calculation of variance (V_{ij}) as follows [162-165]69,70,82,89:

$$V_{ij} = \frac{{\tau_{ij}}^2}{n_j * {\tau_{ij}}^2}, \quad (6)$$

therefore,

$$w_{ij} = \frac{1}{V_{ij}} = \frac{n_j * {\tau_{ij}}^2}{{\tau_{ij}}^2} \quad (7)$$

For fixed effect (FE) model our effect summary for a pair of drug is:

$$\bar{\tau} = \frac{\Sigma(w_a * \tau_a)}{\Sigma w_a}, \quad (8)$$

and the standard error is:

$$SE_{\bar{\tau}} = \sqrt{\frac{1}{\Sigma w_a}}, \quad (9)$$

Where $a = 1, 2, 3, \ldots, s,\ and\ s$ is number SNPs shared by a pair of drugs.

For random effect (RE) model we assume that variability is not only due to random error but also to variability in the population of effects, i.e., number of shared SNPs in drug pair (s). Therefore, the weight for each shared SNP in Equation 8 is adjusted with a constant (λ)[82,90].

$$\lambda = \left(\frac{Q-(s-1)}{\Sigma w_a - \frac{\Sigma(w_a)^2}{\Sigma w_a}} \right) \qquad (10)$$

where the Q measures heterogeneity of effect among s SNPs,

$$Q = \Sigma(w_a - {\tau_a}^2) - \left(\frac{\Sigma(w_a * \tau_a)^2}{\Sigma w_a} \right) \qquad (11)$$

For RE model, new weight (w_λ) for each shared SNP for a drug is calculated as:

$$w_{\lambda_j} = \frac{1}{V_{ij} + \lambda_j} \qquad (12)$$

Final random effect summary is calculated as in Equations 8 and 9 by replacing Σw_a with w_{λ_a}.

Hierarchical Clustering (HC) and Principal Component Analysis (PCA)

Extensively observed overlaps of association of diseases/traits to genetic variants have been considered reflecting shared etiology between those diseases/traits [78,83]. Clustering of phenotypes based on shared genetic loci identified in sets of publicly available GWAS catalogue data has revealed groups of diseases and traits with related clinical characteristics [75]. Recently, scores of statistical and bioinformatics methods have been developed to investigate shared genetic variants across phenotypes [78,80,91,92].

Small sample size and diffident individual effect of variants often render GWAS findings non-reproducible [93,94]. Besides, high degrees of freedom involved in the statistical tests are an issue even in multi-locus analysis methods use for testing multiple SNPs. Therefore, a huge challenge for GWAS is to detect SNPs with weak individual effects that might have strong epistatic effect to influence the outcome of the disease or trait. Currently, multi-locus and multi-trait analysis methods used for testing multiple SNPs mainly focus on reducing number of degrees of freedom ensuing from [17,93,95-98]: (i) integrating statistical significance for all individual SNPs into a combined test statistic; or (ii) a combined signal directly emanating from all SNPs. However, the calculation methods of combined test statistics conceal the epistatic effect and thus obliterate relationships among SNPs in functional regions of interest [93].

Depending on the biological aspect under consideration, multi-trait influence of a genetic variant can be interpreted in two ways: (i) the effect of variant (E) may be through independent physiological processes on phenotypes P_1 and P_2 and may be represented as $P_1 \leftarrow E \rightarrow P_2$; or (ii) a phenotype may be influenced by the variant via other phenotype as $E \rightarrow P_1 \rightarrow P_2$ or $E \rightarrow P_2 \rightarrow P_1$ [80,99]. The increasing number of GWAS has demonstrated existence of pleiotropy that may cause genetic relationships among multiple traits (biological pleiotropy) or linkage disequilibrium (LD) between genes for multiple traits (spurious pleiotropy) [19,35,78,83,100,101]. Similar pleiotropic associations have also been reported for response to drugs, drug-metabolizing enzymes, and drug transporters [27].

Generally, pleiotropic traits are phenotypically or genetically correlated. According to NHGRI, pleiotropy exists in 17% of trait-associated genes and 5% of trait-associated single nucleotide polymorphisms (SNPs) [102,103]. Evidence shows that a combined analysis of two or more pleiotropic traits can enhance GWAS detection power. Hierarchical clustering and principal component analysis (PCA) are two potential approaches to search for pleiotropy in multiple-trait GWAS [103,104].

Recent association studies involving multiple traits, so-called phenome-wide association studies (PheWAS), have many advantages. PheWAS elucidate the complex interrelated networks among phenotypes and by underpinning their genetic basis spell out the underlying pleiotropy [95,96,105,106].

Two main multi-trait association pleiotropic patterns have been reported in GWAS: (i) association of more than one phenotype with a single SNP within a functional region, e.g, a coding region; (ii) different phenotype associations with two different SNPs within a single functional region [95,100,107]. Some studies have demonstrated that pleiotropic regions tend to be longer and they are within a gene, the pleiotropic regions are mostly exonic [95,102].

The main aim of this study is to identify clusters of phenotypes (response to drugs) that share multiple strongest SNP-risk allele in common. We employed principal component analysis (PCA) and hierarchical cluster analysis (HCA) to explore groups of drugs that show evidence of a genetic relationship between the identified cluster of drugs. It is assumed that clustering of the SNP-drug matrix will result in clusters of drugs with similar genetic associations. The

drug association clusters can thus be seen as non-overlapping sets of drugs, or overlapping sets of SNPs.

Clustering of phenotype-SNP association is generally based on a feature vector of a genomics track [39,55,108]. Typically, a SNP is treated as a nominal categorical or ordinal variable. As both categorical and ordinal values can create a monotone relationship between SNP association statuses, their use in clustering becomes inappropriate. Therefore, dummy variable coding is often considered. [144]. As categorical variable in association studies is binary coding strategy has little effect on the subsequent analysis, for example clustering of original or GWAS summary data. For this purpose, GWAS investigators mostly employ GWAS profile matrix of SNP-phenotype association [39,55,108,109].

For conducting efficient hierarchical cluster analysis (HCA) and principal component analysis (PCA), we define three SNP-drug matrices ($\boldsymbol{X}, \boldsymbol{\Phi}$ and $\mathbf{OR}$) in which each row represents a strongly SNP and each column represents a phenotype (response to drug) and each entry as follows:

$$\boldsymbol{X}_{ij} = \begin{cases} 1 & if\ SNP\ i\ is\ associated\ with\ phenotype\ j, and\ i \neq j \\ 0 & otherwise \end{cases}$$

For $\boldsymbol{\Phi}_{ij}$ and $\boldsymbol{OR}_{ij}$ each entry represents the point estimate of the i-th SNP with j-th drug.

$$\boldsymbol{\Phi}_{ij} = \begin{cases} S * \phi_{ij}\ if\ SNP\ i\ is\ associated\ with\ phenotype\ j, and \\ 0\ otherwise \end{cases}$$

$$\boldsymbol{OR}_{ij} = \begin{cases} S * OR_{ij}\ if\ SNP\ i\ is\ associated\ with\ phenotype\ j, a \\ 0\ otherwise \end{cases}$$

where $S * \phi\ and\ S *$ OR are as defined in Equation 4.

The matrix $\boldsymbol{X}, \boldsymbol{\Phi}$ and $\mathbf{OR}$ were subjected to HC and PCA using IBM SPSS Statistics.

RESULTS

We performed analysis of GWAS data on 6762 SNPs strongly associated with response to 128 drug based on GWAS summary statistics collected from the GWAS Catalog. For this purpose drugs versus SNPs data matrices were subjected to meta-analysis, clustering and PCA as explained in Materials ad Methods. Only two measures of effect size were used in this study, namely, Odds ratio (OR) and Cramér's. The former was directly collected and the latter was estimated from the available p-values, confidence interval (CI) and sample size from the GWAS Catalog.

We defined the SNP-sharing strength (SSS) in this study and hypothesized that the weighting effect size of SNPs with SSS can be more efficient for clustering of the drug response and identify SNPs sub-sets from a set of 6762 SNPs that have similar effects over a cluster of drugs. Table 1 shows the distribution of SNPs shared by drugs in a set of 6762 SNPs strongly associated with response to 128 drug. Only 52 (41%) drugs shared SNPs. Out of 6762 SNPs, 101 (1%) are shared by 3 drugs, 386 (6%) are shared by 2 drugs while 6275 (93%) are singly associated with a drug without any sharing.

Meta-analysis summary of weighted fixed effect size (OR) of 211 strongest SNP-risk alleles associated with response to 40 drugs (Table 4) shows a strong positive correlation between fixed and random weighted effect size ($p = 4.00 \times 10^{-40}$). Therefore, most of the analysis is focused on weighted fixed effect size (OR).

The results of preliminary HCA and PCA analysis showed that the SNP-drug matrices using both non-weighted and SSS-weighted Log OR and Cramér's generated inconsequential clustering of drugs. Only SSS-weighted OR yielded meaningful clustering and revealed several SNP-drug response connections. Therefore, we focus only on the results based on analysis of SSS-weighted OR matrices.

Meta-analysis of SSS-weighted OR in the 5663×113 SNP-drug helped in short-listing 40 drugs sharing 211 SNPs. The resulting 211×40 SNP-drug was subjected to PCA and PCA.

Three linkage measures were tried in this study, viz., between groups linkage, single linkage, and complete linkage. All these measures returned almost similar clustering of the drugs. Between groups linkage provided a natural compatibility between various linkage approaches and offered a more reliable between cluster measures. [110]. As there is no built-in method for further partitioning dense blobs of clusters in SPSS, this was performed by shortening data matrix according to the composition of a dense cluster for further clustering analysis. All resulting sub-clusters were assembled together into a single dendrogram (Figure 1), keeping the scale of the original cluster. The resulting dendrogram is depicted in Figure 1. The clusters of drugs in Figure 1 and the distribution of shared SNPs in Table 1 are further summarized in Table 2 and diagrammatically depicted in Figure 3.

The dendrogram in Figure 1 was cut where the difference was most significant and thus six clusters of drugs are identified. The first

three clusters of drugs are almost equal size together comprise 34 drugs (85%) and a variable number of sub-clusters. The HCA clustering is closely resembles the clustering of the drugs plotted against the first three principal components (Figure 4).

The values of fixed effect size resulting from the meta-analysis and corresponding number of shared SNPs for pairs of drugs are presented as heatmap in Figure 2 in which all the drugs are arrayed according to the clusters in Figure 1. In another heatmap (Figure 5) the same array of drug clusters is presented with the weighted OR values of all 211 SNPs arranged in the ascending order of the data file obtained from from GWAS Catalog file (gwas_catalog_v1.0.2-associations_e96_r2019-09-24.tsv).

The first cluster consists of 11 drugs (28%) and 4 distinct sub-clusters of antidepressant, antihyperthyroid, antineoplastic and antihypertensive drugs. The compactness of the linkage in Cluster-1 is also evident in Figures 2 summarizing the result of meta-analysis where the corresponding fixed effect size values are along the diagonal. Most of the drugs in Cluster-1 also share a larger number of SNPs with other drugs as shown in Table 2 and Figure 3.

The composition of Cluster 1 is supported by interactions among the clustered drugs reported in the literature. For example, interactions between antidepressants and antihypertensive has been documented [111]. Tricyclic antidepressants, antipsychotic phenothiazines and some other psychoactive drugs can cause iatrogenic effects in thyroid [112]. Thyroidectomy in SHR exhibits antihypertensive effect and associated changes in heart performance [113]. There is evidence of interactions between idiosyncratic effect of antidepressants (ADs) and antineoplastic agents [114].

The Cluster-2 is little larger than the Cluster-1 and 3, and mainly comprises cytotoxic & adjuvant medicines linked with a small cluster of antibiotic and interferon. The skin lesions of cancer patients treated with cetuximab, an epidermal growth factor receptor (EGFR), have shown type I interferon signaling [115]. Close clustering of cetuximab and Capox-b in Cluster-2 is supported with the reported improved activity and tolerability in first-line treatment of metastatic colorectal cancer [116]. Likewise, close clustering of antibiotic and bleomycin and their linkage with the cluster of diverse anti-cancer drugs is compatible with clinical and pre-clinical demonstration of beneficial effect of the adjuvant antimicrobial therapy in cancer treatment [117].

In the Cluster-3 is a cluster of anticancer drugs with the exception of a sub-cluster of two thiazide diuretics. This thiazide cluster is closely linked with a singular sub-cluster comprising topoisomerase inhibitor. Patients treated with combination chemotherapy and thiazide diuretics against metastatic breast cancer have shown an augmentation of granulocytopenia as compared with only chemotherapy treatment, revealing an interaction between the diuretic and the cytotoxic agents [118].

The last three clusters comprising antipsychotics, antiviral/antituberculotic and antidepressants, respectively, with each cluster consisting of two drug, are all isolated in the dendrogram (Figure 1). The Cluster-6 of the two antidepressants, citalopram and escitalopram, is one of the most isolated.

The clustering of drugs in Figure 1 and the distribution of shared SNPs in Table 2 are further summarized in Table 3a and Table 3b with addition of shared SNP composition of drug clusters and corresponding mapped gene composition, respectively.

Tables 3a shows that Cluster 1 &2 do not share their within cluster shared SNPs (WCSS). However, Cluster 1 &2 strongly share the WCSSs with those of Cluster 3. However, only Sub-cluster 1c of the Cytotoxic and adjuvant medicines, shares 13 WCSS with Cluster 3. The scenario of within cluster shared mapped genes (WCSG) is different (Table 3b) where Sub-cluster 1c shares 9 WCSGs with Cluster 3. In Cluster 2 only Sub-cluster 2d shares all its 42 WCSSs with Cluster 3.

Therefore, Cluster 3, comprising mostly cytotoxic and adjuvant medicines is not only a visible connecting link between Clusters 1 & 2 (Figures 1 & 4), but also due to sharing of the most of its WCSSs with Clusters 1 & 2 the exception of those in Sub-cluster 3b consisting of thiazide diuretics. The Cluster 3 also has large inter sub-cluster sharing of WCSSs and WCSG not found in other clusters (Table 3a&b). The diversity of SNPs sharing in Cluster 3 is also evident in Figure 3.

Table 1. Distribution of SNPs shared by drugs in a set of 6762 SNPs strongly associated with response to 128 drug as reported in GWAS catalog [11]

No	Drug		Number of drugs sharing associated SNP together		
			3-drugs	2-drugs	No sharing
1	5' fluorouracil	5' Fl	11	4	0
2	Abacavir	Aba	0	0	27
3	Acetaminophen	Ace	0	0	5
4	Allopurinol	All	0	0	26
5	Angiotensin-converting enzyme inhibitor	Ang	0	0	79
6	Anthracycline-based chemotherapy	ABC	0	6	3
7	Antibiotic	Abi	0	3	0
8	Anticoagulant	Aco	0	0	34
9	Anticonvulsant	Anv	0	0	18
10	Antidepressant	ADT	6	23	96
11	Antihypertensive drug	AHD	3	0	0
12	Antimetabolite	Ame	0	6	0
13	Antimicrotubule agent	AMA	0	12	7
14	Antineoplastic agent	AND	7	94	81
15	Antipsychotic drug	ASD	0	2	105
16	Antithyroid drug	ATD	0	1	2
17	Anti-tuberculosis drugs	Ant	0	0	3
18	Aripiprazole	Ari	0	2	0
19	Aromatase inhibitor	Aro	0	0	2
20	Beta blocker	Bet	3	6	64
21	Bevacizumab	Bev	0	0	16
22	Bleomycin	Ble	0	3	0
23	Bortezomib	Bor	0	0	4
24	Bronchodilator	Bro	0	5	3471
25	Bupropion	Bup	2	4	16
26	Calcium channel blocker	Cal	3	0	3
27	Camptothecin	Cam	0	3	0
28	Candesartan	Can	0	0	25
29	Capox-b	Cap	0	4	0
30	Carbamazepine	Cmp	0	0	1
31	Carboplatin	Car	7	9	36
32	Cetuximab	Cet	0	4	8
33	Cholinesterase inhibitor	Cho	0	0	3
34	Cisplatin	Cis	0	21	41
35	Citalopram	Cit	2	30	0
36	Clopidogrel	Cld	0	0	28
37	Clozapine	Clz	0	0	73
38	Cold medicine	Col	0	0	4
39	Combination chemotherapy	Com	0	0	58
40	Corticosteroid	Cor	0	17	39
41	Cyclophosphamide	Cyc	11	3	14
42	Cytokine	Cyt	0	95	2
43	Cytosine arabinoside	Csa	0	0	22
44	Dabigatran etexilate	Dab	0	0	2
45	Dalcetrapib	Dal	0	0	1
46	Darapladib	Dar	0	0	23
47	Diclofenac	Dic	0	0	4
48	Docetaxel	Doc	0	7	16
49	Doxorubicin	Dox	8	3	0
50	Duloxetine	Dul	0	0	2
51	Efavirenz	Efa	0	3	36
52	Endocrine therapy	End	0	0	2
53	Epirubicin	Epi	3	5	0
54	Escitalopram	Esc	0	17	0
55	Etoposide	Eto	0	4	0
56	Fenofibrate	Fen	74	0	104
57	Fluoroquinolones	Flq	0	0	4
58	Flupirtine	Flu	0	0	1
59	Gemcitabine	Gem	0	5	15
60	Glucocorticoid	Glu	0	0	26
61	Haloperidol	Hal	0	0	8
62	Heparin	Hep	0	0	5
63	Homoharringtonine	Hom	0	0	10
64	Hydrochlorothiazide	Hyd	0	77	0
65	Iloperidone	Ilo	0	0	6
66	Immunochemotherapy	Imm	0	0	2
67	Interferon	Inf	0	10	80
68	Interferon beta	Inb	0	10	0
69	Irinotecan	Iri	0	0	9
70	Lamotrigine	Lam	0	0	27
71	Lapatinib	Lap	0	0	3
72	Lithium ion	Lit	0	0	42
73	Mepolizumab	Mep	0	0	2
74	Mercaptopurine	Mer	0	4	0
75	Metformin	Met	0	0	15
76	Methotrexate	Mtr	0	0	24
77	Methylphenidate	Mpd	0	0	7
78	Montelukast	Mon	0	0	2
79	Morphine	Mor	0	0	1
80	Mtor inhibitor	Mto	0	0	13
81	Nitrofurantoin	Nit	0	0	4
82	Non-steroidal anti-inflammatory	Non	0	0	14
83	Olanzapine	Ola	0	0	15
84	Opiate	Opt	0	0	4
85	Opioid	Opd	0	0	55
86	Paclitaxel	Pac	7	33	63
87	Paliperidone	Pal	0	0	249
88	Pazopanib	Paz	0	0	1
89	Perphenazine	Per	0	0	16
90	Phenytoin	Phe	0	0	1
91	Platelet aggregation inhibitor	PAI	0	0	1
92	Platinum based chemotherapy	PBC	0	11	14
93	Platinum-based neoadjuvant chemotherapy	PBN	0	0	3
94	Protease inhibitor	Pro	0	0	9
95	Quetiapine	Que	0	0	18
96	Ranibizumab	Ran	0	0	4
97	Reverse transcriptase inhibitor	Rev	0	0	46
98	Ribavirin	Rib	0	0	1
99	Rifampicin	Rif	0	3	2
100	Risperidone	Ris	0	0	20
101	Selective serotonin reuptake inhibitor	Sel	0	0	84
102	Simvastatin	Sim	74	7	0
103	Statin	Sta	74	7	83
104	Sulfasalazine	Sus	0	0	32
105	Sulfonylurea	Sun	0	0	49
106	Synacthen	Syn	0	10	0
107	Tamoxifen	Tmf	0	0	2
108	Tamsulosin	Tml	0	0	4
109	Taxane	Tax	0	0	3
110	Temozolomide	Tem	0	0	1
111	Tenofovir	Ten	0	0	305
112	Terbinafine	Ter	0	0	1
113	Tetracyclic antidepressant	Tet	4	0	0
114	Thiazide	Thz	0	83	9
115	Thioamide	Thm	0	1	9
116	Thiopurine	Thp	0	4	23
117	Tnf antagonist	Tnf	0	0	60
118	Topoisomerase inhibitor	Top	0	6	0
119	Trastuzumab	Tra	0	6	0
120	Triamcinolone acetonide	TAA	0	0	1
121	Tricyclic antidepressant	TCA	4	5	0
122	Triptolide	Trp	0	0	14
123	Vaccine	Vac	0	93	91
124	Vancomycin	Van	0	0	1
125	Venlafaxine	Ven	0	1	17
126	Vincristine	Vin	0	0	5
127	Zileuton	Zil	0	0	21
128	Ziprasidone	Zip	0	0	12

Table 2. Distribution of SNPs shared by drugs in a set of 8762 SNPs strongly associated with response to 40 drugs arrayed according to the clustering in Figure 1

Cluster	Sub-Cluster	Drug		Number of drugs sharing associated SNP		
				3-drugs	2-drugs	No sharing
1	1a	Tetracyclic antidepressant	Tet	4	0	0
		Tricyclic antidepressant	TCA	4	5	0
		Antidepressant	ADT	6	23	96
	1b	Antithyroid drug	ATD	0	1	2
		Thioamide	Thm	0	1	9
	1c	5' fluorouracil	5' Fl	11	4	0
		Cyclophosphamide	Cyc	11	3	14
		Doxorubicin	Dox	8	3	0
	1d	Beta blocker	Bet	3	6	64
		Calcium channel blocker	Cal	3	0	3
		Antihypertensive drug	AHD	3	0	0
2	2a	Capox-b	Cap	0	4	0
		Cetuximab	Cet	0	4	8
	2b	Interferon	Inf	0	10	80
		Interferon beta	Inb	0	10	0
	2c	Antibiotic	Abi	0	3	0
		Bleomycin	Ble	0	3	0
	2d	Antimicrotubule agent	AMA	0	12	7
		Platinum based chemotherapy	PBC	0	11	14
		Cisplatin	Cis	0	21	41
		Antimetabolite	Ame	0	6	0
		Anthracycline-based chemotherapy	ABC	0	6	3
		Docetaxel	Doc	0	7	16
3	3a	Topoisomerase inhibitor	Top	0	6	0
	3b	Hydrochlorothiazide	Hyd	0	77	0
		Thiazide	Thz	0	83	9
	3c	Antineoplastic agent	AND	7	94	81
		Paclitaxel	Pac	7	33	63
		Carboplatin	Car	7	9	36
	3d	Trastuzumab	Tra	0	6	0
		Epirubicin	Epi	3	5	0
	3e	Gemcitabine	Gem	0	5	15
		Etoposide	Eto	0	4	0
		Camptothecin	Cam	0	3	0
4	4	Antipsychotic drug	ASD	0	2	105
		Aripiprazole	Ari	0	2	0
5	5	Efavirenz	Efa	0	3	36
		Rifampicin	Rif	0	3	2
6	6	Citalopram	Cit	2	30	0
		Escitalopram	Esc	0	17	0

Table 3a. Shared SNP composition of the clusters and sub-clusters of 40 drugs according to the clustering in Figure 1. Inter cluster sharing of SNPs is shown in red and blue, intra-cluster sharing is shown in green

Cluster	Sub-cluster	Cluster label	Shared SNPs	
1	Tet TCA ADT	1a	4	rs6737205-A rs9830388-A rs11867129-T rs2291477-C
	ATD Thm	1b	1	rs185386680-?
	5' Fl Cyc Dox	1c	21	rs3885907-G rs4474258-A rs7089227-G rs594206-A rs672203-A rs6475600-C rs6845621-C rs10818894-G rs10112481-A rs637644-G rs11644424-T rs2519974-T rs10922438-T rs3745571-T rs11857176-A rs4380275-G rs2512987-T rs6740660-G rs1567482-G rs10488226-A rs6706693-A
	Bet Cal AHD	1d	3	rs113296370-C rs4833103-? rs11124945-G
		1	29	
2	Cap Cet	2a	4	rs2936519-A rs2073016-T rs4377367-C rs885036-A
	Inf Inb	2b	3	rs2205986-? rs522308-T rs2454138-A
	Abi Ble	2c	3	rs8093763-A rs4662834-A rs708547-A
	AMA PBC Cis Ame ABC Doc	2d	42	rs4235898-A rs12882718-T rs9386485-T rs6817170-G rs4591358-C rs11651483-C rs4771859-G rs12145418-T rs12935229-A rs1043763-T rs8022296-G rs33428-G rs12589282-G rs1895302-C rs16825455-T rs10253216-T rs11944965-T rs2406342-T rs6077251-T rs11774576-A rs4627050-G rs12142335-A rs4149639-C rs1654260-A rs2505059-G rs12615435-T rs10040979-G rs1367448-C rs7228133-C rs8127977-A rs894734-G rs9580312-G rs2055011-C rs12582168-C rs488248-T rs4727963-C rs3747851-T rs12660691-A rs1756650-G rs4553720-T rs837841-T rs6961860-G
		2	52	
3	Top	3a	5	rs1035147-T rs10074959-T rs7494275-C rs480409-G rs1632744-T
	Hyd Thz	3b	59	rs11077614-G rs6870564-A rs427576-C rs1974942-A rs1669070-T rs1511453-A rs10989824-T rs7762018-A rs890749-C rs6859974-T rs1551678-C rs17560407-G rs2274788-C rs11216831-A rs2980003-T rs12303914-A rs11648716-G rs10943724-A rs17183114-T rs4722750-T rs1559759-A rs225675-G rs7077606-T rs7801534-G rs7964748-G rs12507634-G rs11785622-T rs10778699-A rs11599315-T rs13402330-A rs7965364-T rs16951120-T rs4808260-C rs2944755-A rs6493487-G rs17668704-A rs12279250-C rs2686586-T rs12455924-T rs12307997-G rs10152811-A rs10508921-T rs12127679-T rs9420790-A rs10985375-A rs2613514-G rs2116940-G rs11810574-G rs7203315-T rs878131-A rs166881-C rs9788333-G rs2460504-C rs7147996-A rs17133858-T rs12904863-C rs12498735-T rs2099077-A rs6928289-G
	AND Pac Car	3c	101	rs7989332-G rs3822735-G rs11071200-T rs3845905-G rs12446319-A rs4235898-A rs12882718-T rs9386485-T rs6817170-G rs4591358-C rs3138083-A rs4666360-C rs4886670-A rs792975-T rs1623879-G rs936229-G rs1714746-G rs1277203-A rs11651483-C rs4771859-G rs12145418-T rs12935229-A rs1043763-T rs8022296-G rs922106-T rs2444896-T rs3786094-C rs2519974-T rs10922438-T rs3745571-T rs33428-G rs12589282-G rs1895302-C rs16825455-T rs10253216-T rs11944965-T rs2406342-T rs6077251-T rs11774576-A rs4627050-G rs12142335-A rs4149639-C rs1654260-A rs2505059-G rs12615435-T rs10040979-G rs1367448-C rs1857176-A rs4380275-G rs2512987-T rs1923834-G rs908366-A rs2964475-C rs1553091-G rs7228133-C rs8127977-A rs894734-G rs9580312-G rs2055011-C rs12582168-C rs6740660-G rs1567482-G rs10488226-A rs6706693-A rs9961113-C rs9609078-T rs2547917-A rs12900463-C rs6863418-A rs488248-T rs4727963-C rs3747851-T rs12660691-A rs1756650-G rs4553720-T rs1035147-T rs10074959-T rs7494275-C rs480409-G rs17318866-G rs17027130-C rs303386-A rs6039763-A rs2506991-G rs12987465-A rs10785877-T rs6429703-T rs12310399-A rs995834-C rs922107-G rs1425132-T rs1476029-A rs1632744-T rs837841-T rs6961860-G rs55756123-? rs10117876-? rs4305714-? rs707557-? rs77679196-? rs7698718-?
	Tra Epi	3d	13	rs3885907-G rs4474258-A rs7089227-G rs1923834-G rs908366-A rs2964475-C rs1553091-G rs55756123-? rs10117876-? rs4305714-? rs707557-? rs77679196-? rs7698718-?
	Gem Eto Cam	3e	11	rs9961113-C rs9609078-T rs2547917-A rs12900463-C rs6863418-A rs17318866-G rs17027130-C rs303386-A rs6039763-A rs2506991-G rs12987465-A
		3	163	
4	ASD Ari	4	2	rs17022006-G rs68057166-G
5	Efa Rif	5	3	rs4842407-C rs11012476-A rs251891-A
6	Cit Esc	6	17	rs111365677-? rs142484554-? rs188843168-? rs6546604-? rs34177316-? rs201921722-? rs80278479-? rs6063349-? rs142641502-? rs201569130-? rs1548076-? rs182377406-? rs74860738-? rs143934587-? rs189864513-? rs11924809-? rs117375960-?

Table 3b. Mapped gene composition of the clusters and sub-clusters of 40 drugs according to the clustering in Figure 1. Inter cluster sharing of mapped genes is shown in red and blue, intra-cluster sharing is shown in green

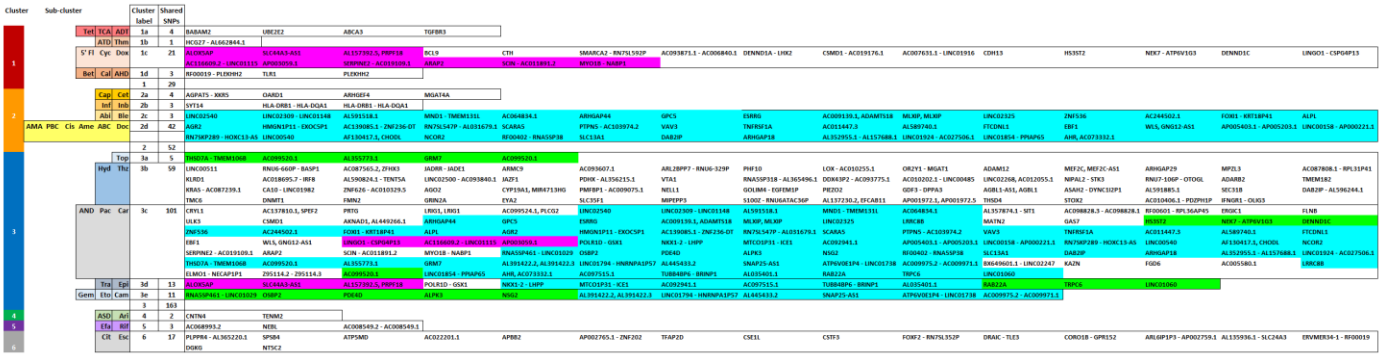


Table 4. Meta-analysis summary of weighted fixed effect size (OR) of 211 strongest SNP-risk alleles associated with response to 40 drugs. The drugs are arrayed according to the clustering in Figure 1. Where K is the number of studies representing a strongly associated SNP; Q is statistics for heterogeneity measure among studies; I^2 is statistics to quantify heterogeneity (% of the total variability in the set of effect sizes due to true heterogeneity); ES is effect size (weighted OR); SE is standard error; and .CI is confidence interval

Cluster	Sub-Cluster	Drug		K	Q	I^2	ES Fixed	Es_{es} Fixed	Lower CI (fixed)	Upper CI (fixed)	ES Random	Es_{es} Random	Lower CI (random)	Upper CI (random)
1	1a	Tetracyclic antidepressant	Tet	4	2665913.74	100.00	13.03	0.01	13.02	13.05	0.01	0.01	0.00	0.02
		Tricyclic antidepressant	TCA	4	2665913.74	100.00	13.03	0.01	13.02	13.05	0.01	0.01	0.00	0.02
		Antidepressant	ADT	76	7020192.27	100.00	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00
	1b	Antithyroid drug	ATD	3	34563.48	99.99	4.01	0.05	3.91	4.10	0.11	0.03	0.04	0.17
		Thioamide	Thm	10	38760.91	99.98	0.64	0.01	0.63	0.66	0.03	0.01	0.01	0.05
	1c	5' fluorouracil	5' Fl	15	2212.67	99.37	1.09	0.01	1.07	1.12	0.04	0.01	0.02	0.07
		Cyclophosphamide	Cyc	14	1876.57	99.31	1.05	0.02	1.02	1.09	0.06	0.02	0.03	0.10
		Doxorubicin	Dox	11	1901.65	99.47	2.18	0.04	2.09	2.26	0.14	0.04	0.06	0.22
	1d	Beta blocker	Bet	73	14259.09	99.50	0.09	0.00	0.08	0.09	0.01	0.00	0.01	0.02
		Calcium channel blocker	Cal	6	11741.59	99.96	0.11	0.00	0.11	0.11	0.00	0.00	0.00	0.01
		Antihypertensive drug	AHD	3	551.68	99.64	6.90	0.06	6.77	7.03	0.11	0.06	-0.01	0.24
2	2a	Capox-b	Cap	4	26.72	88.77	0.89	0.02	0.85	0.93	0.04	0.02	0.00	0.08
		Cetuximab	Cet	12	1534.26	99.28	0.92	0.01	0.89	0.95	0.05	0.01	0.02	0.08
	2b	Interferon	Inf	90	2324.16	96.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Interferon beta	Inb	10	1154.14	99.22	0.11	0.00	0.10	0.12	0.01	0.00	0.00	0.02
	2c	Antibiotic	Abi	3	26.39	92.42	0.08	0.01	0.07	0.10	0.01	0.01	0.00	0.02
		Bleomycin	Ble	3	26.39	92.42	0.08	0.01	0.07	0.10	0.01	0.01	0.00	0.02
	2d	Antimicrotubule agent	AMA	19	87.79	79.50	0.25	0.00	0.24	0.26	0.02	0.00	0.01	0.02
		Platinum based chemotherapy	PBC	15	997.04	98.60	0.30	0.01	0.29	0.31	0.02	0.01	0.01	0.03
		Cisplatin	Cis	41	461.62	91.33	0.21	0.00	0.20	0.22	0.02	0.00	0.01	0.03
		Antimetabolite	Ame	6	63.21	92.09	0.47	0.01	0.46	0.48	0.01	0.01	0.00	0.02
		Anthracycline-based chemotherapy	ABC	9	702.22	98.86	0.70	0.01	0.67	0.72	0.03	0.01	0.01	0.06
		Docetaxel	Doc	23	786.34	97.20	0.22	0.00	0.21	0.23	0.02	0.00	0.01	0.03
3	3a	Topoisomerase inhibitor	Top	6	419.74	98.81	0.62	0.02	0.59	0.66	0.05	0.02	0.01	0.08
	3b	Hydrochlorothiazide	Hyd	77	15030.74	99.49	0.06	0.00	0.05	0.06	0.01	0.00	0.00	0.01
		Thiazide	Thz	85	16901.81	99.50	0.05	0.00	0.05	0.05	0.01	0.00	0.00	0.01
	3c	Antineoplastic agent	AND	125	4125.68	96.99	0.04	0.00	0.04	0.04	0.01	0.00	0.01	0.01
		Paclitaxel	Pac	84	3930.24	97.89	0.08	0.00	0.08	0.09	0.02	0.00	0.01	0.02
		Carboplatin	Car	34	3077.07	98.93	0.29	0.01	0.28	0.30	0.03	0.01	0.02	0.05
	3d	Trastuzumab	Tra	6	2604.62	99.81	0.90	0.01	0.88	0.93	0.03	0.01	0.00	0.06
		Epirubicin	Epi	8	740.46	99.05	1.09	0.02	1.05	1.12	0.05	0.02	0.01	0.09
	3e	Gemcitabine	Gem	10	1704.18	99.47	0.76	0.02	0.73	0.79	0.05	0.02	0.02	0.08
		Etoposide	Eto	3	83.35	97.60	2.48	0.09	2.30	2.67	0.16	0.09	-0.02	0.34
		Camptothecin	Cam	3	423.06	99.53	2.15	0.06	2.04	2.27	0.10	0.06	-0.01	0.21
4	4	Antipsychotic drug	ASD	86	1393895.19	99.99	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00
		Aripiprazole	Ari	2	30.43	96.71	6.62	0.09	6.44	6.80	0.13	0.09	-0.05	0.31
5	5	Efavirenz	Efa	39	3654.19	98.96	0.06	0.00	0.06	0.07	0.01	0.00	0.00	0.01
		Rifampicin	Rif	5	1623.75	99.75	1.92	0.04	1.84	1.99	0.08	0.04	0.00	0.15
6	6	Citalopram	Cit	17	3833564.65	100.00	0.31	0.00	0.31	0.31	0.00	0.00	0.00	0.00
		Escitalopram	Esc	17	3833564.65	100.00	0.31	0.00	0.31	0.31	0.00	0.00	0.00	0.00

In Table 2 and Figure 3, out of the SNPs individually shared by drugs, a positive correlation (p=0.0533) can be seen between the number of SNPs that are not shared by drugs and those shared between two drugs. It can be inferred from this trend that more a drug has strongly associated SNPs the more likely it shares SNPs with one more drug and less so with any additional drug. This correlation is more significant pronounced in drugs of Cluster-1 (p=0.0002) as compared with Cluster 2 (p=0.0444) and Cluster-3, being the least in Cluster-3 (p=0.0798).

A significant positive correlation was found between the weighted fixed effect size vector of the 40 drugs with the corresponding vectors of first two principal components, with p = 0.0009 and p = 0.0113, respectively, but insignificant correlation with that of 3rd component (Table 4, Figure 4). The weighted random effect size vector had no significant correlation any of the principal components. This finding supports our assumption to treat SNPs included in this study as a random sample from a single population of SNPs strongly associated with the studied drugs. The presence of strong correlation between the weighted fixed effect size and the principal components provides a possibility of using principal component regression for predicting the effect size of association of a drug response with SNPs [127].

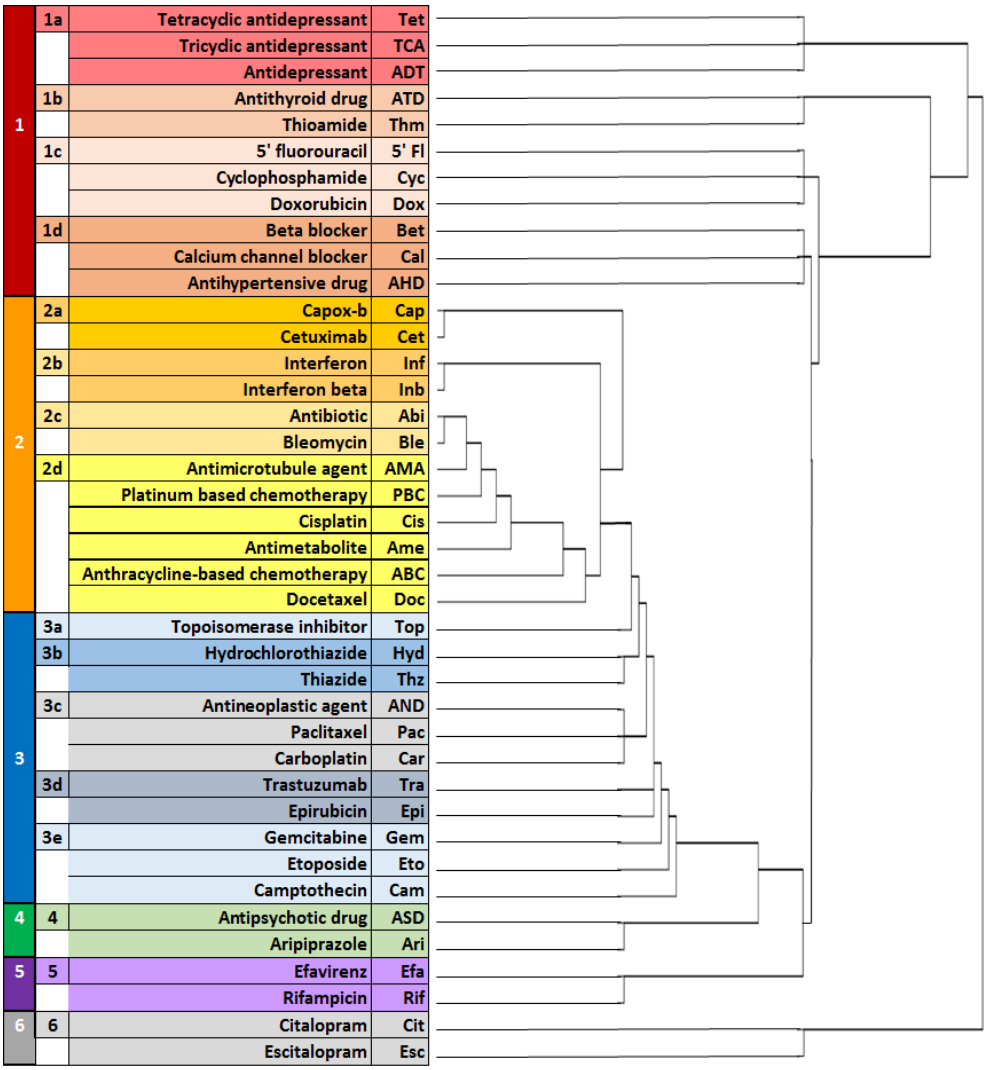


Figure 1. Dendrogram showing between cluster grouping of 40 drugs on the basis of weighted fixed effect size (OR) of 211 associated strongest SNP-risk alleles.

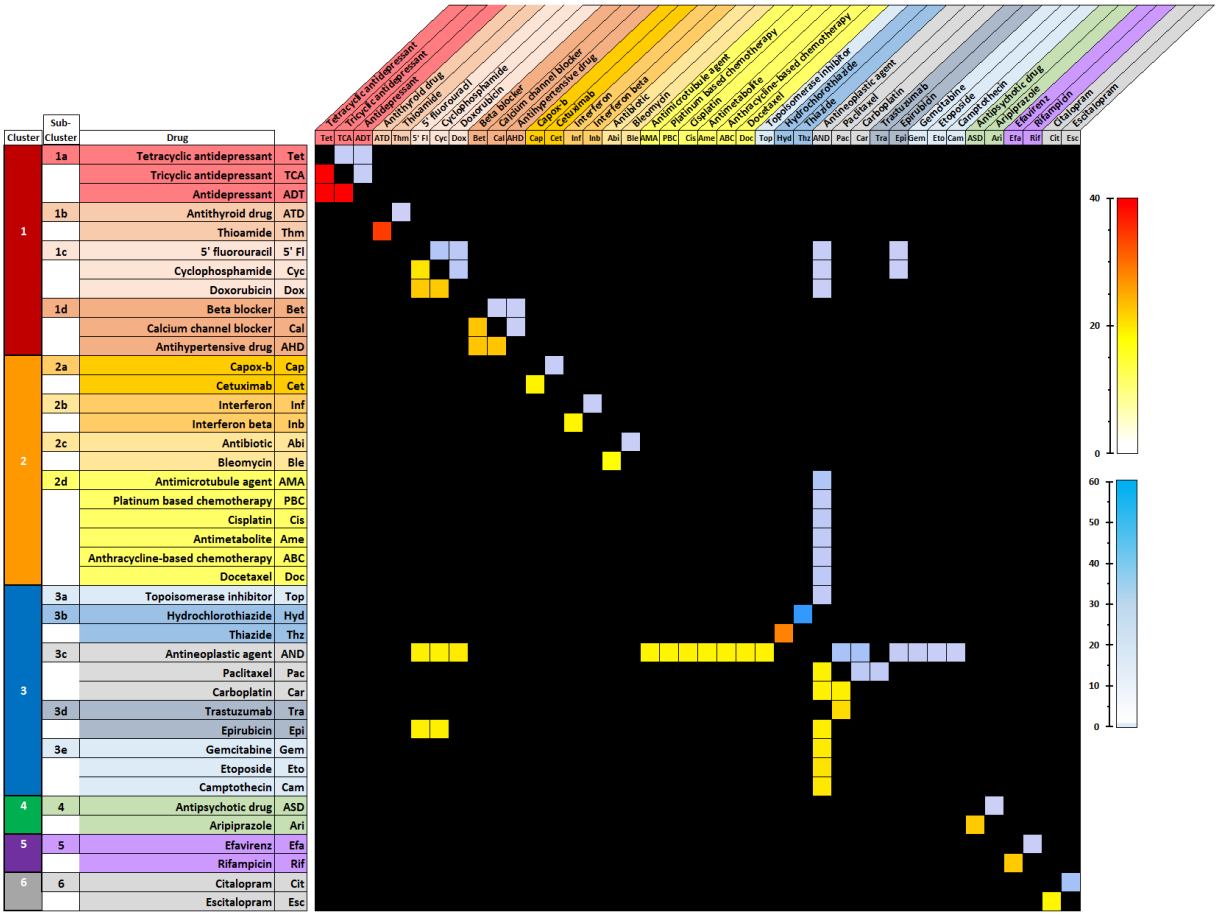


Figure 2. Heatmap showing weighted fixed effect size (OR) meta-analyzed on the basis of 211 associated strongest SNP-risk alleles between 40 drugs (values below the diagonal). values above the diagonal indicate number of shared SNPs between the drugs. The darker colors representing higher values.

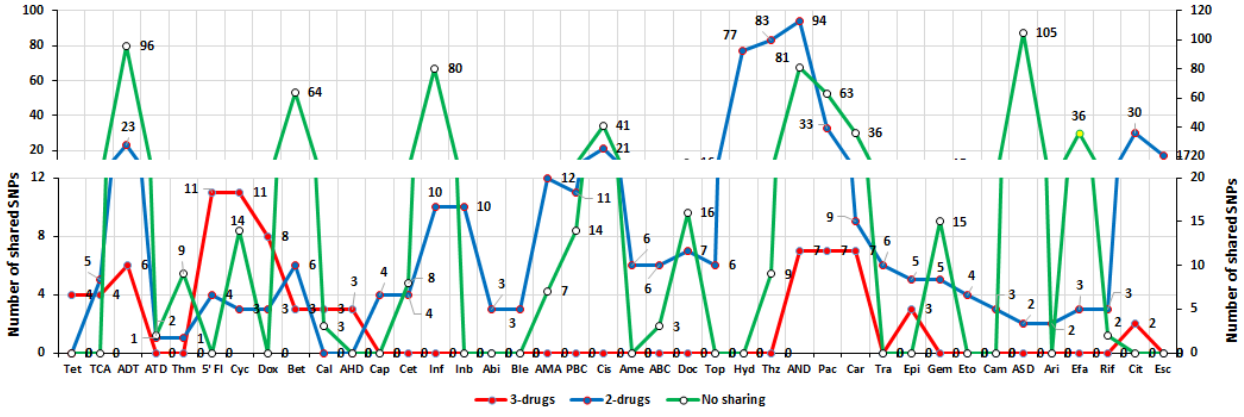
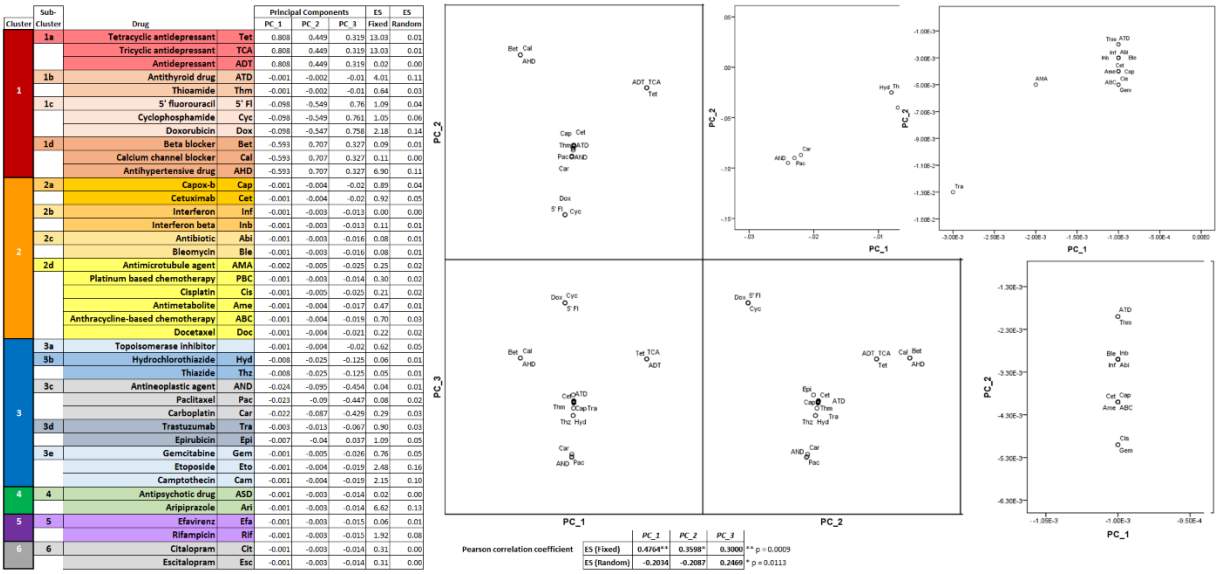


Figure 3. Distribution of SNPs shared by drugs in a set of 211 SNPs strongly associated with response to 40 drugs arrayed according to the clustering in Figure 1. a positive correlation (p=0.0533) can be seen between the number of SNPs not shared by drugs and those shared between two drugs. It can be inferred from this trend that more a drug has strongly associated SNPs the more likely it shares SNPs with one more drug and less so with any additional drug. This correlation is more significant pronounced in drugs of Cluster-1 (p=0.0002) as compared with Cluster 2 (p=0.0444) and Cluster-3, where it is the least (p=0.0798).



Cluster	Sub-Cluster	Drug		Principal Components PC_1	PC_2	PC_3	ES Fixed	ES Random
1	1a	Tetracyclic antidepressant	Tet	0.808	0.449	0.319	13.03	0.01
		Tricyclic antidepressant	TCA	0.808	0.449	0.319	13.03	0.01
		Antidepressant	ADT	0.808	0.449	0.319	0.02	0.00
	1b	Antithyroid drug	ATD	-0.001	-0.002	-0.01	4.01	0.11
		Thioamide	Thm	-0.001	-0.002	-0.01	0.64	0.03
	1c	5' fluorouracil	5' Fl	-0.098	-0.549	0.76	1.09	0.04
		Cyclophosphamide	Cyc	-0.098	-0.549	0.761	1.05	0.06
		Doxorubicin	Dox	-0.098	-0.547	0.758	2.18	0.14
	1d	Beta blocker	Bet	-0.593	0.707	0.327	0.09	0.01
		Calcium channel blocker	Cal	-0.593	0.707	0.327	0.11	0.00
		Antihypertensive drug	AHD	-0.593	0.707	0.327	6.90	0.11
2	2a	Capox-b	Cap	-0.001	-0.004	-0.02	0.89	0.04
		Cetuximab	Cet	-0.001	-0.004	-0.02	0.92	0.05
	2b	Interferon	Inf	-0.001	-0.003	-0.013	0.00	0.00
		Interferon beta	Inb	-0.001	-0.003	-0.013	0.11	0.01
	2c	Antibiotic	Abi	-0.001	-0.003	-0.016	0.08	0.01
		Bleomycin	Ble	-0.001	-0.003	-0.016	0.08	0.01
	2d	Antimicrotubule agent	AMA	-0.002	-0.005	-0.025	0.25	0.02
		Platinum based chemotherapy	PBC	-0.001	-0.003	-0.014	0.30	0.02
		Cisplatin	Cis	-0.001	-0.005	-0.025	0.21	0.02
		Antimetabolite	Ame	-0.001	-0.004	-0.017	0.47	0.01
		Anthracycline-based chemotherapy	ABC	-0.001	-0.004	-0.019	0.70	0.03
		Docetaxel	Doc	-0.001	-0.004	-0.021	0.22	0.02
3	3a	Topoisomerase inhibitor		-0.001	-0.004	-0.02	0.62	0.05
	3b	Hydrochlorothiazide	Hyd	-0.008	-0.025	-0.125	0.06	0.01
		Thiazide	Thz	-0.008	-0.025	-0.125	0.05	0.01
	3c	Antineoplastic agent	AND	-0.024	-0.095	-0.454	0.04	0.01
		Paclitaxel	Pac	-0.023	-0.09	-0.447	0.08	0.02
		Carboplatin	Car	-0.022	-0.087	-0.429	0.29	0.03
	3d	Trastuzumab	Tra	-0.003	-0.013	-0.067	0.90	0.03
		Epirubicin	Epi	-0.007	-0.04	0.037	1.09	0.05
	3e	Gemcitabine	Gem	-0.001	-0.005	-0.026	0.76	0.05
		Etoposide	Eto	-0.001	-0.004	-0.019	2.48	0.16
		Camptothecin	Cam	-0.001	-0.004	-0.019	2.15	0.10
4	4	Antipsychotic drug	ASD	-0.001	-0.003	-0.014	0.02	0.00
		Aripiprazole	Ari	-0.001	-0.003	-0.014	6.62	0.13
5	5	Efavirenz	Efa	-0.001	-0.003	-0.015	0.06	0.01
		Rifampicin	Rif	-0.001	-0.003	-0.015	1.92	0.08
6	6	Citalopram	Clt	-0.001	-0.003	-0.014	0.31	0.00
		Escitalopram	Esc	-0.001	-0.003	-0.014	0.31	0.00

Figure 4. Clustering of the drugs plotted against the first three principal components obtained by PCA of 40 drugs on the basis of weighted fixed effect size (OR) of 211 associated strongest SNP-risk. Correlation between the weighted fixed effect size vector of the 40 drugs (from Table 4) with the corresponding vectors of first two principal components is shown at bottom right.

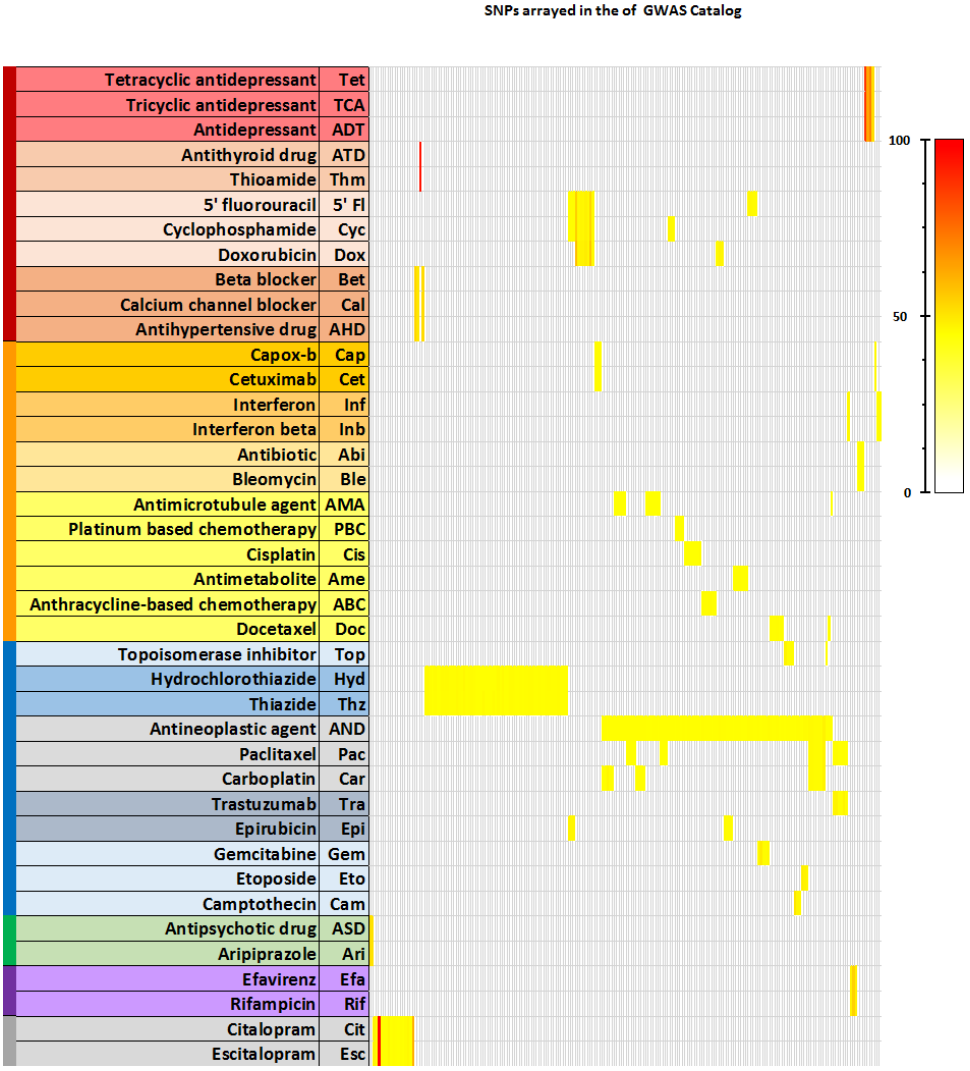


Figure 5. Heatmap showing weighted effect size (OR) of 211 associated strongest SNP-risk alleles between 40 drugs. The SNPs are arrayed in the ascending order of the data file obtained from GWAS Catalog file (gwas_catalog_v1.0.2-associations_e96_r2019-09-24.tsv) [11].

DISCUSSION

The methods used in this study were helpful for exploring workable clusters of related drugs on the basis of the effect size of the shared associated SNPs when weighted with SNP-sharing strength (SSS), a weighting factor defined for this study. Simple GWAS profile matrices representing SNP-phenotype associations provided a rather pointless clustering of the presently studied drugs. The reason could be mere binary representation of the association that lowers the degree of variation use for evaluation of similarity between phenotype. Effect size in GWAS summary statistics has been used mostly by investigators for is for testing power of association and clustering of SNPs rather than phenotypes [80]. The use of standardized mean differences by many researchers for meta-analysis on dichotomous data was proven incorrect. While odds ratio (OR) and Log OR always provided better results for computing effect size [68]. The findings in this study show that even the effect size based on OR should be weighted appropriately, depending on the nature of assumption made for conducting phenotype clustering based on GWAS summary.

Clustering of GWAS results have been used to discover phenotypic patterns of cells and tissues, including sensitivity to drugs, as well as to detect artifacts of experimental conditions [119-121]. PCA-based multiple-trait GWAS prove robust method for explaining heritability [122,126-129]. As compared with multivariate model, PCA takes much less time and has been widely used in pleiotropic mapping [122]. However, one of the limitations of PCA-based GWAS is that it can only be applied when the data on all traits and the variants in the sample are available. Fortunately, this limitation has not been faced in this study.

Exploring a meaningful pattern of genome-wide correlations would require identification of SNPs subsets shaping genetic architecture differentially and giving insight into the etiological mechanism underlying the phenotype (disease, response to a drug, etc.) that cannot be detected by standard GWAS [123]. Numerous tools are available to conduct CPA analyses with varying plusses and limitations [107,123]. Most of the multi-variant methods are employed to analyze GWAS summary data available in public domain [29,33], as we did in this study.

Some studies have reported larger effect sizes for pharmacogenomic phenotypes due to genetic variants as compared with other phenotypes, such as complex disease and traits [124,125]. Other studies suggested that effect size difference were not influenced by the over-estimated effect size due to smaller sample sizes in pharmacogenomic studies [124]. The use of standardized mean differences by many researchers for meta-analysis on dichotomous data was proven incorrect. While odds ratio (OR) and Log OR always provided better results for computing effect size [68]. However, simple GWAS profile matrices representing SNP-phenotype associations produce a rather pointless clustering of phenotype. The main reason is the binary representation of the association that lowers the degree of variation needed for evaluation of similarity between phenotype. Since the collected genetic variants come from a large number of GWAS with different sample sizes, meta-analysis of all individual studies is usually undertaken for determination of combined effect size using normalization of sample sizes and appropriate weights, often employing specific analysis tool sets [36,40,52,68,71]. The methods used in this study helped exploring workable clusters of related drugs on the basis of the effect size of the shared associated SNPs when weighted with SNP-sharing strength (SSS), a factor defined for this study.

Effect size in GWAS summary statistics has been used mostly by investigators for is for testing power of association and clustering of SNPs rather than phenotypes [80]. Meta-analysis of th,e weighted OR of all shared SNPs for each pair of drugs and their use for clustering the phenotype (response to drug) not only provides a meaningful phenotypic pattern of drugs but also provided knowledge about the inter- and intra-cluster distribution associated SNPs. The methods used in this study helped exploring workable clusters of related drugs on the basis of the effect size of the shared associated SNPs when weighted with SNP-sharing strength (SSS), a factor defined for this study.

The phenotypic pattern of drugs and associated SNPs revealed by the methods used in this study are expected to help in future for understanding aetiological basis of different drugs by highlighting relevant biological pathways.

CONCLUSION

The findings in this study show that the effect size should be weighted appropriately, depending on the nature of

assumption made for conducting phenotype clustering based on GWAS summary. Thus methods used in this study are helpful for exploring workable clusters of related drugs on the basis of the effect size of the shared associated SNPs when weighted with SNP-sharing strength.

In sets of SNPs individually shared by drugs, a trend based on positive correlation between the number of SNPs that are not shared by drugs and those shared between two drugs, suggest that more a drug has strongly associated SNPs the more likely it shares SNPs with one more drug, but less so with any additional drug. The presence of strong correlation between the weighted fixed effect size and the principal components provides a possibility of using principal component regression for predicting the effect size of association of a drug response with SNPs.

Overall, the phenotypic pattern of drugs and associated SNPs revealed by this study are expected to help in future for understanding aetiological basis of different drugs by highlighting relevant biological pathways.

ACKNOWLEDGMENT

Data used in this study are managed and reserved by the National Human Genome Research Institute (NHGRI) and The European Bioinformatics Institute (EBI) database, the NHGRI-EBI GWAS Catalog. We gratefully acknowledge the GWAS Catalog.

REFERENCES

[1]**.** Alving AS, Carson PE, Flanagan CL, Ickes CE. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science,* 1956; 124(3220): 484–85.

[2]. Maggo SD, Savage RL, Kennedy MA. Impact of new genomic technologies on understanding adverse drug reactions. *Clin Pharma*. 2016; 55(4): 419-36.

[3]. Alomar MJ. Factors affecting the development of adverse drug reactions. *Saudi Pharm J*. 2014; 22(2): 83-94.

[4]. Gonzalez FJ, Skoda RC, Kimura S, Umeno M, Zanger UM, Nebert DW, et al. Characterization of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature.* 1988; 331(6155): 442–46. doi:10.1038/331442a0.

[5]. Rawlins MD. Variability in response to drugs. *Br Med J.* 1974; 4(5936): 91–94. doi:10.1136/bmj.4.5936.91.

[6]**.** PharmGKB. Types of pharmacogenes. https://www.pharmgkb.org/page/typesOfPg

[7]. Zhou K, Pearson ER. Insights from genome-wide association studies of drug response. *Annu Rev Pharmacol Toxicol*.

2013; 53: 299–310. doi:10.1146/annurev-pharmtox-011112-140237.

[8]. Robinson JR, Denny JC, Roden DM, Van Driest SL. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. *Clin Transl Sci.* 2018; 11(2): 112–122. doi:10.1111/cts.12522

[9]. O'Mara TA, Glubb D, Batra J. Editorial: Establishing genetic pleiotropy to identify common pharmacological agents for common diseases. *Frontiers in Pharm.* 2019; 10: 1038.

[10]. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR. The support of human genetic evidence for approved drug indications. *Nat Genetics*. 2015; 47(8): 856.

[11]. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. BioRxiv. 2019 Jan 1:513945.

[12]. Campion DP, Dowell FJ. Translating Pharmacogenetics and Pharmacogenomics to the Clinic: Progress in Human and Veterinary Medicine. Front Vet Sci. 2019; 6:22. Published 2019. doi:10.3389/fvets.2019.00022

[13]. Schärfe CPI, Tremmel R, Schwab M, Kohlbacher O, Marks DS. Genetic variation in human drug-related genes. Genome Med. 2017;9(1):117. Published 2017 Dec 22. doi:10.1186/s13073-017-0502-5

[14]. Wright GEB, Carleton B, Hayden MR, Ross CJD. The global spectrum of protein-coding pharmacogenomic diversity. Pharmacogenomics J. 2016

[15]. Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. Methods Mol Biol. 2013; 1015:311–320. doi:10.1007/978-1-62703-435-7_20

[16]. Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics. 2009; 10:327–334.

[17]. Choi J, Tantisira KG, Duan QL. Whole genome sequencing identifies high-impact variants in well-known pharmacogenomic genes. bioRxiv. 2018: 368225.

[18]. Giacomini KM, Yee SW, Mushiroda T, Weinshilboum RM, Ratain MJ, Kubo M. Genome-wide association studies of drug response and toxicity: an opportunity for genome medicine. Nature Reviews Drug Discovery. 2017; 16(1):70.

[19]. Motsinger-Reif AA, Jorgenson E, Relling MV, et al. Genome-wide association studies in pharmacogenomics: successes and lessons. Pharmacogenet Genomics. 2013; 23(8):383–394. doi:10.1097/FPC.0b013e32833d7b45

[20]. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. Genet Epidemiol. 2010; 34(1):60–66. doi:10.1002/gepi.20435

[21]. Holden AL, Contreras JL, John S, Nelson MR. The international serious adverse events consortium. Nat Rev Drug Discov. 2014; 13(11):795–6. doi:10.1038/nrd4441.

[22]. Molokhia M, McKeigue P. EUDRAGENE: European collaboration to establish a case-control DNA collection for studying the genetic basis of adverse drug reactions. Pharmacogenomics. 2006; 7(4):633–8. doi:10.2217/14622416.7.4.633.

[23]. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008; 84(3):362–9. doi:10.1038/clpt.2008.89.

[24]. Ross CJ, Visscher H, Sistonen J, Brunham LR, Pussegoda K, Loo TT, et al. The Canadian Pharmacogenomics Network for Drug Safety: a model for safety pharmacology. Thyroid. 2010; 20(7):681–7. doi:10.1089/thy.2010.1642.

[25]. Search Collaborative Group, Link E, Parish S, Armitage J, Bowman L, Heath S, et al. SLCO1B1 variants and statin-induced myopathy–a genomewide study. N Engl J Med. 2008; 359(8):789–99. doi:10.1056/NEJMoa0801936.

[26]. Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A, et al. HLA-B*5701 genotype is a major determinant of druginduced liver injury due to flucloxacillin. Nat Genet. 2009; 41(7):816–9. doi:10.1038/ng.379.

[27]. Oetjens MT, Bush WS, Denny JC, Birdwell K, Kodaman N, Verma A, Dilks HH, Pendergrass SA, Ritchie MD, Crawford DC. Evidence for extensive pleiotropy among pharmacogenes. Pharmacogenomics. 2016; 17(8):853-66.

[28]. Zhu W, Zhang H. Why do we test multiple traits in genetic association studies?. Journal of the Korean Statistical Society. 2009; 38(1):1-0.

[29]. Guo B, Wu B. Integrate multiple traits to detect novel trait–gene association using GWAS summary data with an adaptive test approach. Bioinformatics. 2018.

[30]. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, Smith JA, Yanek LR, Sun YV, Edwards TL, Chen W. Meta-analysis of correlated traits via summary

statistics from GWASs with an application in hypertension. The American Journal of Human Genetics. 2015; 96(1):21-36.

[31]. Broadaway KA, Cutler DJ, Duncan R, Moore JL, Ware EB, Jhun MA, Bielak LF, Zhao W, Smith JA, Peyser PA, Kardia SL. A statistical approach for testing cross-phenotype effects of rare variants. The American Journal of Human Genetics. 2016; 98(3):525-40.

[32]. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nature Reviews Genetics. 2017; 18(2):117.

[33]. Stephens M. A unified framework for association analysis with multiple related phenotypes. PloS one. 2013; 8(7):e65245.

[34]. Li X, Zhu X. Cross-Phenotype Association Analysis Using Summary Statistics from GWAS. Methods Mol Biol. 2017; 1666: 455–467. doi:10.1007/978-1-4939-7274-6_22

[35]. Zhang Z, Ma P, Li Q, Xiao Q, Sun H, Olasege BS, Wang Q, Pan Y. Exploring the genetic correlation between growth and immunity based on summary statistics of genome-wide association studies. Frontiers in genetics. 2018; 9: 393.

[36]. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*, 2019; 47(D1): 1005–1012.

[37]. Pontiroli A, Simonet P, Frostegard A, Vogel TM, Monier JM. Fate of transgenic plant DNA in the environment. *Env Biosaf Res*, 2007; 6: 15-35. doi: 10.1051/ebr:2007037.

[38]. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. 2015; 47(3):291.

[39]. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan L, Perry JR, Patterson N, Robinson EB, Daly MJ. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015; 47(11):1236.

[40]. Naz M, Younesi E, Hofmann-Apitius M (2017) Systematic Analysis of GWAS Data Reveals Genomic Hotspots for Shared Mechanisms between Neurodegenerative Diseases. J Alzheimers Dis Parkinsonism 7: 368. DOI: 10.4172/2161-0460.1000368

[41]. Dai M, Wan X, Peng H, Wang Y, Liu Y, Liu J, Xu Z, Yang C. Joint analysis of individual-level and summary-level GWAS data by leveraging pleiotropy. arXiv preprint arXiv:1804.11011. 2018.

[42]. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS genetics. 2014; 10(5):1004383.

[43]. Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK. Identification of a Brazil-nut allergen in transgenic soybeans. *New Engl J Med*, 1996; 334: 688-92.

[44]. Prescott VE, Campbell PM, Moore A. Transgenic expression of bean alpha-amylase inhibitor in peas results in altered structure and immunogenicity. *J Agric Food Chem*, 2005; 53: 9023–30.

[45]. Kanduri C, Bock C, Gundersen S, Hovig E, Sandve GK. Colocalization analyses of genomic elements: approaches, recommendations and challenges. Bioinformatics. 2018; 35(9):1615-24.

[46]. Bohme H, Aulrich K, Daenicke R, Flachowsky G. Genetically modified feeds in animal nutrition. 2nd communication: Glufosinate tolerant sugar beet (roots and silage) and maize grains for ruminants and pigs. *Arch Anim Nutrit*, 2001; (54) 197-207.

[47]. Finamore A, Roselli M, Britti S. Intestinal and peripheral immune response to MON810 maize ingestion in weaning and old mice. *J Agric Food Chem*, 2008; 56: 11533–39.

[48]. Dozmorov MG. Epigenomic annotation-based interpretation of genomic data: from enrichment analysis to machine learning. Bioinformatics. 2017; 33(20):3323-30.

[49]. Demeke, Tigst, and David Dobnik. Critical assessment of digital PCR for the detection and quantification of genetically modified organisms. *Anal Bioanal Chem*, 2018; 410: 4039-50.

[50]. Eriksson, S. Invited review: Breeding and ethical perspectives on genetically modified and genome edited cattle. *J Dairy Sci*, 2018; 101: 1-17.

[51]. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, Pickrell J, Jaffe AE, CommonMind Consortium, Pasaniuc B, Roussos P. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2018; 34(15):2538-45.

[52]. Holland D, Wang Y, Thompson WK, Schork A, Chen CH, Lo MT, Witoelar A, Werge T, O'donovan M, Andreassen OA, Dale AM. Estimating effect sizes and expected replication probabilities from GWAS summary statistics. Frontiers in genetics. 2016; 7:15.

[53]. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin

K, Hayes BJ, Goddard ME. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLoS genetics. 2014; 10(3):1004198.

[54]. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genetics research. 2009; 91(1): 47-60.

[55]. Horn JH. Tools for cluster analysis of data from genome-wide association studies - role of data representation and similarity measures (Master's thesis). Department of Informatics, University of OSLO. 2016.

[56]. Zhao B, Zhu H. On genetic correlation estimation with summary statistics from genome-wide association studies. arXiv preprint arXiv:1903.01301. 2019.

[57]. Yang C, Li C, Wang Q, Chung D, Zhao H. Implications of pleiotropy: challenges and opportunities for mining Big Data in biomedicine. Frontiers in genetics. 2015; 6:229.

[58]. Schaid DJ, Tong X, Larrabee B, Kennedy RB, Poland GA, Sinnwell JP. Statistical methods for testing genetic pleiotropy. Genetics. 2016 Oct 1;204(2):483-97.

[59]. Deng Y, Pan W. Testing genetic pleiotropy with GWAS summary statistics for marginal and conditional analyses. Genetics. 2017; 207(4): 1285-99.

[60]. Wei LJ, Johnson WE. Combining dependent tests with incomplete repeated measurements. Biometrika. 1985; 72(2): 359-64.

[61]. Rosenberg MS. A generalized formula for converting chi-square tests to effect sizes for meta-analysis. *PloS one.* 2010; 5(4):10059.

[62]. Rosenberg MS, Adams DC, Gurevitch J (2000) MetaWin. Statistical Software for Meta-Analysis. 2.0 ed. Sunderland, Massachusetts: Sinauer Associates.

[63]. Hedges LV, Olkin I (1985) Statistical Methods for Meta-Analysis. San Diego, CA: Academic Press, Inc. 369 p.

[64]. Campbell Collaboration. Converting between effect sizes Version 1.1. Campbell Methods Series: Policy Note 3. December 2016.

[65]. Wallis SA. z-squared: The origin and application of χ^2. Journal of Quantitative Linguistics. 2013; 20(4): 350-78.

[66]. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Converting Among Effect Sizes-Chapter 7. Michael Borenstein. John Wiley & Sons, Ltd. 2009; ISBN: 978-0-470-05724-7

[67]. Ialongo C. Understanding the effect size and its measures. *Biochem Med.* 2016;

26(2): 150–63. doi:10.11613/BM.2016.015

[68]. Haddock CK, Rindskopf D, Shadish WR. Using odds ratios as effect sizes for meta-analysis of dichotomous data: a primer on methods and issues. Psychological Methods. 1998; 3(3):339.

[69]. Borenstein M, Hedges L, Rothstein H. Meta-analysis fixed effect vs. random effects. 2007.

[70]. Begum F, Ghosh D, Tseng GC, Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic acids research. 2012; 19; 40(9): 3777-84.

[71. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Jul 8; 26(17):2190-1.

[72]. Kim HY. Statistical notes for clinical researchers: effect size. Restorative dentistry & endodontics. 2015 Nov 1; 40(4): 328-31.

[73]. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nature Reviews Genetics. 2013; 14(6): 379.

[74]. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'connell JR, Mangino M, Mägi R. Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics. 2011; 19(7):807.

[75]. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, Hatan M. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature.* 2015; 518(7539): 337.

[76]. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, Oskarsson S. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016; 533(7604):539.

[77]. Fowler T, Zammit S, Owen MJ, Rasmussen F. A population-based study of shared genetic variation between premorbid IQ and psychosis among male twin pairs and sibling pairs from Sweden. Archives of general psychiatry. 2012; 69(5): 460-66.

[78]. Fortune MD, Guo H, Burren O, Schofield E, Walker NM and Ban M. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nat. Genet*, 2015; 47, 839–846.

[79]. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, Lovering RC, Tajuddin SM, Winkler TW, Graff M, Kavousi M. GWAS and

colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nature communications. 2018; 9(1):5141.

[80]. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48(7):709-17.

[81]. Chyou PH. A simple and robust way of concluding meta-analysis results using reported P values, standardized effect sizes, or other statistics. Clinical medicine & research. 2012; 10(4):219-23.

[82]. Neyeloff JL, Fuchs SC, Moreira LB. Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. BMC research notes. 2012; 5(1):52.

[83]. Cotsapas C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011; 7

[84]. Bagos PG. A unification of multivariate methods for meta-analysis of genetic association studies. Statistical applications in genetics and molecular biology. 2008; 7(1).

[85]. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A, Vollenweider P, Stirnadel H, Sundseth SS, Lai E, Burns DK. Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score—the CoLaus Study. Diabetologia. 2009; 52(4):600-8.

[86]. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954; 10(1):101-29.

[87]. Pereira TV, Patsopoulos NA, Salanti G, Ioannidis JP. Discovery properties of genome-wide association signals from cumulatively combined data sets. American journal of epidemiology. 2009; 170(10):1197-206.

[88]. Zhou B, Shi J, Whittemore AS. Optimal methods for meta-analysis of genome-wide association studies. Genetic epidemiology. 2011; 35(7):581-91.

[89]. Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. The American Journal of Human Genetics. 2009 Dec 11; 85(6):862-72.

[90]. Shi J, Lee S. A novel random effect model for GWAS meta-analysis and its application to trans-ethnic meta-analysis. *Biometrics*. 2016; 72(3):945-54.

[91]. Shu L, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y et al. 2017 Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet.* 13, 1007040. [PubMed: 28957322]

[92]. Arneson D, Zhang Y, Yang X, Narayanan M. Shared mechanisms among neurodegenerative diseases: from genetic factors to gene networks. J Genet. 2018; 97(3): 795-806.

[93]. Lu M, Lee HS, Hadley D, Huang JZ, Qian X. Supervised categorical principal component analysis for genome-wide association analyses. *InBMC genomics* 2014; 15(1): 10.

[94]. Ioannidis J, Castaldi P, Evangelou E: A compendium of genome-wide associations for cancer: Critical synopsis and reappraisal. Journal of the National Cancer Institute 2010, 102:846-858.

[95]. Weighill D, Jones P, Bleker C, et al. Multi-Phenotype Association Decomposition: Unraveling Complex Gene-Phenotype Relationships. Front Genet. 2019; 10: 417. 10. doi:10.3389/fgene.2019.00417

[96]. Pendergrass SA, Verma A, Okula A, Hall MA, Crawford DC, Ritchie MD. Phenome-wide association studies: embracing complexity for discovery. Human heredity. 2015; 79(3-4): 111-23.

[97]. Levine ME, Langfelder P, Horvath S. A Weighted SNP Correlation Network Method for Estimating Polygenic Risk Scores. *Methods Mol Biol*. 2017; 1613: 277–290. doi:10.1007/978-1-4939-7027-8_10

[98]. Gupta M, Cheung CL, Hsu YH, et al. Identification of homogeneous genetic architecture of multiple genetically correlated traits by block clustering of genome-wide associations. *J Bone Miner Res.* 2011; 26(6): 1261–71. doi:10.1002/jbmr.333

[99]. Momen M, Ayatollahi Mehrgardi A, Amiri Roudbar M, Kranis A, Mercuri Pinto R, Valente B, Morota G, Rosa GJ, Gianola D. Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models. Frontiers in genetics. 2018; 9:455.

[100]. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 2013; 14(7): 483–495.

[101]. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics. 2017; 101(1): 5-22.

[102]. Sivakumaran S, Agakov F, Theodoratou E et al. Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. 2011; 89(5), 607–618.

[103]. Zhang W, Gao X, Shi X, Zhu B, Wang Z, Gao H, Xu L, Zhang L, Li J, Chen Y. PCA-Based Multiple-Trait GWAS Analysis: A Powerful Model for

Exploring Pleiotropy. Animals. 2018; 8(12):239.

[104]. Guinot F, Szafranski M, Ambroise C, Samson F. Learning the optimal scale for GWAS through hierarchical SNP aggregation. BMC bioinformatics. 2018; 19(1): 459.

[105]. Pendergrass SA, Brown-Gentry K, Dudek S et al. Phenomewide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS Genet. 2013;9(1), e1003087.

[106]. Pendergrass SA, Brown-Gentry K, Dudek SM et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype–phenotype relationships and pleiotropy discovery. Genet. Epidemiol. 2011;35(5):410–422.

[107]. Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits. *Open Biol.* 2017; 7(11):170125.

[108]. Rydbeck H, Sandve GK, Ferkingstad E, Simovski B, Rye M, Hovig E. ClusTrack: feature extraction and similarity measures for clustering of genome-wide data sets. *PloS one*. 2015; 10(4):0123261.

[109]. Chen H. Clustering and Network Analysis with Single Nucleotide Polymorphism (SNP) (Doctoral dissertation, The Graduate School, Stony Brook University: Stony Brook, NY.); 2011

[110]. Yim O, Ramdeen KT. Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *Quantitative methods psychol*. 2015; 11(1): 8-21.

[111]. Coelho PV, Brum CD. Interactions between antidepressants and antihypertensive and glucose lowering drugs among patients in the HIPERDIA Program, Coronel Fabriciano, Minas Gerais State, Brazil. *Cadernos de saude publica*. 2009; 25: 2229-36.

[112]. Sauvage MF, Marquet P, Rousseau A, Raby C, Buxeraud J, Lachâtre G. Relationship between psychotropic drugs and thyroid function: a review. Toxicology and applied pharmacology. 1998; 149(2): 127-35.

[113]. Rodgers RL, McNEILL JH. Antihypertensive effect of thyroidectomy in SHR: associated changes in heart performance. *Am J Physiol-Heart and Circulatory Physiol*. 1986; 250(4): 600-605.

[114]. Miguel C, Albuquerque E. Drug interaction in psycho-oncology: antidepressants and antineoplastics. Pharmacology. 2011; 88(5-6): 333-39.

[115]. Carbone ML, Lulli D, Passarelli F, Pastore S. Topical Plant Polyphenols Prevent Type I Interferon Signaling in the Skin and Suppress Contact Hypersensitivity. *Int J Mol Sci*. 2018; 19(9): 2652. doi:10.3390/ijms19092652

[116]. Borner M, Koeberle D, Von Moos R, Saletti P, Rauch D, Hess V, Trojan A, Helbling D, Pestalozzi B, Caspar C, Ruhstaller T. Adding cetuximab to capecitabine plus oxaliplatin (XELOX) in first-line treatment of metastatic colorectal cancer: a randomized phase II trial of the Swiss Group for Clinical Cancer Research SAKK. *Annals of oncology*. 2008; 17; 19(7): 1288-92.

[117]. Alibek K, Bekmurzayeva A, Mussabekova A, Sultankulov B. Using antimicrobial adjuvant therapy in cancer treatment: a review. *Infect Agent Cancer*. 2012; 7(1): 33. doi:10.1186/1750-9378-7-33

[118]. Orr LE. Potentiation of myelosuppression from cancer chemotherapy and thiazide diuretics. Drug intelligence & clinical pharmacy. 1981; 15(12):967-70.

[119]. Ben-Hur A, Guyon I. Detecting stable clusters using principal component analysis. InFunctional genomics 2003 (pp. 159-182). Humana press.

[120]. Getz G, Levine E, Domany E. Coupled two-way clustering analysis of gene microarray data. *Proceed Nation Academy Sci*. 2000; 24; 97(22): 12079-84.

[121]. Shamir R, Sharan R. 1 1 algorithmic approaches to clustering gene expression data. *Curr Topics Comput M Biol*. 2002; 269.

[122]. Aschard H, Vilhjálmsson BJ, Greliche N, Morange PE, Trégouët DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am J Human Genetics*. 2014; 1; 94(5): 662-76.

[123]. Lam M, Hill WD, Trampush JW, Yu J, Knowles E, Davies G, Stahl E, Huckins L, Liewald DC, Djurovic S, Melle I. Pleiotropic meta-analysis of cognition, education, and schizophrenia differentiates roles of early neurodevelopmental and adult synaptic pathways. *BioRxiv*. 2019; 1: 519967.

[124]. Maranville JC, Cox NJ. Pharmacogenomic variants have larger effect sizes than genetic variants associated with other dichotomous complex traits. *Pharma J*, 2016; 16(4): 388.

[125]. Chhibber A, Kroetz DL, Tantisira KG, et al. Genomic architecture of pharmacological efficacy and adverse

events. *Pharmacogenomics.* 2014; 15(16): 2025–48. doi:10.2217/pgs.14.144.

[126]. Klei L, Luca D, Devlin B, Roeder K. Pleiotropy and principal components of heritability combine to increase power for association analysis. Genetic Epidemiology. *Inter Genetic Epidemiol Society*. 2008; 32(1): 9-19.

[127]. Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. PloS one. 2014; 9(4): 93766.

[128]. Price AL, Patterson NJ, Plenge RM, Weinblatt ME. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 2006: 904–909.

[129]. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2: 190.

[130]. Park SH. Collinearity and optimal restrictions on regression parameters for estimating responses. *Technometrics.* 1981; 23 (3): 289–95.