

# A Penalized Likelihood Approach for Investigating Gene and Gene-Drug Interactions in Pharmacogenetic Studies

Megan L. Neely<sup>1</sup>

Howard D. Bondell<sup>2</sup>

Jung-Ying Tzeng<sup>2,3\*</sup>

<sup>1</sup>Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina

<sup>2</sup>Department of Statistics, North Carolina State University, Raleigh, North Carolina

<sup>3</sup>Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina

\*Correspondence to: Jung-Ying Tzeng, Department of Statistics, Campus Box 7566, North  
Carolina State University, Raleigh, NC 27695.

Email: [jytzeng@stat.ncsu.edu](mailto:jytzeng@stat.ncsu.edu)

Tel: 919-513-2723

Fax: 919-515-7315

## SUMMARY

Pharmacogenetics investigates the relationship between heritable genetic variation and the variation in how individuals respond to drug therapies. Often, gene-drug interactions play a primary role in this response, and identifying these effects can aid in the development of individualized treatment regimes. Haplotypes can hold key information in understanding the association between genetic variation and drug response. However, the standard approach for haplotype-based association analysis does not directly address the research questions dictated by individualized medicine. A complimentary post-hoc analysis is required, and this post-hoc analysis is usually under powered after adjusting for multiplicity and may lead to seemingly contradictory conclusions. In this work, we propose a penalized likelihood approach that is able to overcome the drawbacks of the standard approach and yield the desired personalized output. We demonstrate the utility of our method by applying it to the Scottish Randomized Trial in Ovarian Cancer. We also conducted simulation studies and showed that the proposed penalized method has comparable or more power than the standard approach and maintains low Type I error rates for both binary and quantitative drug responses. The largest power gains are seen when haplotype frequency is low, the difference in effect size is small, or the true relationship among the drugs is more complex.

**Key Words:** haplotype; association analysis; multiple comparisons; penalized regression; pharmacogenetics; individualized medicine

## INTRODUCTION

Inter-individual variation in the efficacy and toxicity of drug therapies is common among patients. Although the differences among patients could be the result of many factors, pharmacogenetics focuses on the relationship between drug response and heritable genetic variation. Two research goals are often paramount when conducting pharmacogenetic studies: (1) For an individual, determine which drugs will perform 'best/worst' for them based on their genetic make-up, and (2) for a particular drug, determine the class of individuals that it is 'best/ill' suited for based on their genetic make-up. Tailoring drug therapies based on genetics is commonly referred to as individualized or personalized medicine. Similar to other genetic disease association studies, where the trait of interest is disease phenotypes rather than drug response, employing haplotype-based methods in pharmacogenetic studies is an attractive approach. Haplotype-based association analysis evaluates the joint effects of closely linked genetic markers on a trait of interest. When compared to its single-marker counterparts, this multi-marker approach can be more powerful to detect associations when the causal variants are not genotyped [de Bakker et al., 2005; Zaitlen et al., 2007], have low frequency [de Bakker et al., 2005; Schaid, 2004], or exhibit cis-acting effects [Clark, 2004; Schaid, 2004]. Moreover, for pharmacogenetics, haplotypes are expected to better capture many pharmacogenetic variants [Goldstein et al., 2003]. Diploidy information (i.e. the haplotype pair of an individual) can more precisely reflect the underlying genetic makeup of an individual for identifying suitable personalized healthcare.

The standard approach for performing haplotype-based analysis is to regress the drug response on the haplotypes, treatments, and their corresponding interactions and test the significance of the regression parameters [Balding, 2006]. However, determining whether

individual coefficient estimates are significantly different from zero does not directly address the research goals of individualized medicine. To establish which drugs have similar effects for a particular individual, or conversely, which individuals react similarly to a particular drug, a large number of pair-wise comparisons must be performed. Thus, the standard approach requires a complementary post-hoc analysis in order to produce the personalized output sought after in individualized medicine and pharmacogenetic research. However in practice, this post-hoc analysis is usually under-powered after adjusting for multiplicity. Furthermore, the pair-wise comparisons can yield contradictory conclusions about the effect structure. That is, when the pair-wise comparisons are translated into a grouping structure on the effects, the groups overlap rather than falling into distinct, interpretable subsets. These issues were encountered in an analysis of the Scottish Randomized Trial in Ovarian Cancer (SCOTROC) data [McWhinney-Glass et al., *submitted*]. The trial investigated the effects of two treatment regimes involving taxane/platinum-based chemotherapy. In a preliminary analysis, McWhinney-Glass et al. found that the gene *BCL2* was significantly associated with an increased risk of experiencing severe neurotoxicity. Focusing on this gene, the goal of our analysis was to investigate potential gene and gene-treatment interactions. Specifically, we wanted to determine if individuals responded differently to the assigned treatment based on their genetic make-up. Using the standard haplotype-based approach of regression and subsequent pair-wise testing of the effect estimates, we were able to detect significant differences after adjusting for multiplicity using the Benjamini-Hochberg procedure. However, in each case, the groupings led to contradictory conclusions about which genetic variants had similar effects for a given treatment. Such results are difficult to interpret clinically and prevent the development of coherent personalized

treatments. For individualized medicine to be clinically relevant, these issues need to be addressed and overcome.

In this work, we propose a penalized likelihood method that directly addresses the pharmacogenetic research goals stated above. The method builds off the regression framework of the standard haplotype-based approach, but is able to overcome the drawbacks of having to do a secondary post-hoc analysis and yields the desired personalized output directly. In the literature, using penalized likelihood methods to identify important haplotypic effects has become increasingly popular [Guo and Lin, 2009; Li et al., 2007; Li et al., 2010; Tzeng et al., 2010]. These methods introduce a penalty on the regression coefficients and shrink the coefficient estimates of non-important covariates toward zero. Recently, modifications of classic penalized methods have also been developed to perform haplotype-based analysis and attempt to address issues specific to this type of analysis. For example, Tanck and colleagues [Souverein et al., 2006, 2008; Tanck et al., 2003] use a modified version of Ridge regression [Hoerl and Kennard, 1970] to stabilize inference for rare haplotypes. Chen et al. [2009] develop an adaptive penalized likelihood framework to address the precision-efficiency tradeoff encountered in haplotype-based retrospective methods. Tzeng and Bondell [2010] modify the traditional adaptive least absolute shrinkage and selection operator (LASSO) [Tibshirani, 1996; Zou, 2006] to allow for effect comparisons between all pairs of distinct haplotypes, rather than with respect to an arbitrary baseline haplotype, during the estimation process.

In our penalized likelihood method, we place an  $L_1$ -penalty on all relevant pair-wise comparisons, which are dictated by the research questions, so that the estimates of group means with the same magnitude are collapsed to be equal. For example, if two drugs perform similarly for an individual, we would like our method to collapse their means and produce estimates that

are equal to reflect this inter-relationship. In this way, our method simultaneously achieves the effect estimation and pair-wise comparisons necessary to address the research questions stated above, which eliminates the need to perform any post-hoc analysis. In addition, our method is able to collapse the effects into an overall group structure without leading to any contradictory conclusions. As a result, the proposed method directly produces the personalized output desired when developing individualized treatments. Simulation studies show that the proposed penalized method has comparable or more power than the standard approach and maintains low Type I error rates for both binary and quantitative drug responses. The largest gain is seen when the haplotype frequency is low, the difference in effect sizes are small, and more complex effect structures are present. The utility and validity of the proposed method is further demonstrated when used to analyze the SCOTROC data. Our penalized method was able to avoid the contradictory results of the standard approach and provide findings that are not only significant, but are also easily interpretable in a clinical setting. In the remaining sections of this paper, we develop the proposed penalized method, describe the simulation study used to investigate its performance, present and interpret results of the simulation study, provide the real-data application of the method to SCOTROC data, and end with a discussion of the work's major findings and avenues for future extension.

## METHODS

Let  $(Y_i, G_i, E_i)$  represent the observed data for individual  $i$  in a sample of size  $n$ . Let  $Y_i$  be the phenotype of individual  $i$ , while  $G_i$  denotes the unphased genotype of individual  $i$  at  $m$  biallelic SNPs and  $E_i$  denotes any environmental covariates measured on individual  $i$ . Let  $H_i$

represent the vector of haplotype counts for individual  $i$ . Although researchers want to investigate the relationship between  $Y_i$  and  $H_i$ , they only have access to  $G_i$ ; therefore, the individual's haplotype pair, or diplotype, must be inferred from their unphased genotypes. Because the focus of this work is pharmacogenetic studies, the phenotype of interest is drug response and the environmental covariates of primary interest are drug therapies.

The relationship between the drug response and the covariates can be characterized by the conditional density function  $f(Y|H, E)$ . The standard approach for haplotype-based association analysis is regression, and generalized linear models (GLMs) provide a flexible and robust framework for the analysis. Under GLMs, the density of  $Y$  is given by

$$f(Y|H, E) = \exp\left\{\frac{Y\eta - b(\eta)}{a(\sigma)} + c(Y, \sigma)\right\},$$

where  $a$ ,  $b$ , and  $c$  are known functions,  $\sigma$  is a scale parameter, and  $\eta$  is the linear predictor. GLMs assume that there exists a monotonic function  $g(\cdot)$  that links the expected value of  $Y$ , denoted by  $\mu = E(Y|H, E)$ , to the linear predictor  $\eta$ . That is,

$$\eta = g(\mu) = \gamma_0 + \beta^T Z(H) + \delta^T E + \gamma^T Z(H) \otimes E,$$

where  $\gamma_0$  represents a global intercept,  $\beta$  is the vector of haplotype effects,  $\delta$  is the vector of drug effects,  $\gamma$  is the vector of haplotype-drug interaction effects,  $Z(H)$  is a vector-valued function of the vector of haplotype counts, and  $Z(H) \otimes E$  is the Kronecker product of the vectors  $Z(H)$  and  $E$ , the vector of drug indicators. In this work, we assumed an additive model on the haplotype

main and interaction effects. This implies that the function  $Z(\cdot)$  is the identity, and that the linear predictor can be written in the following ANOVA-like representation

$$\eta = g(\mu) = \gamma_0 + \beta_h + \beta_{h'} + \delta_j + \gamma_{hj} + \gamma_{h'j},$$

where the sums  $\sum_h \beta_h$ ,  $\sum_j \delta_j$ ,  $\sum_h \gamma_{hj}$ , and  $\sum_j \gamma_{h'j}$  must be set to zero to account for the overparameterization,  $h = 1, \dots, l$ , the number of distinct haplotypes, and  $j = 1, \dots, d$ , the number of drugs under study. The linear predictor above represents an individual with the haplotypes  $h$  and  $h'$ , or the diplotype  $hh'$ , who received drug  $j$  in the study. For quantitative responses, the distribution of  $Y$  is typically assumed to be normal with the identity link to lead to linear regression. For binary responses, the distribution of  $Y$  is typically assumed to be Bernoulli with the logit link to lead to logistic regression.

Recall the two research goals stated earlier: (1) For an individual, determine which drugs will perform ‘best/worst’ for them based on their genetic make-up and (2) For a particular drug, determine the class of individuals that it is ‘best/ill’ suited for based on their genetic make-up. To address the first goal, all pair-wise comparisons between drugs would need to be performed for each distinct “genetic make-up”. In the haplotype setting, an individual is uniquely identified by their diplotype. The appropriate hypothesis tests of drug differences for a diplotype  $hh'$  are thus given by  $H_0: \eta_{hh'j} - \eta_{hh'j'} = 0$  for  $j < j'$ . This hypothesis states that the means of drug  $j$  and  $j'$  do not differ for diplotype  $hh'$ . Substituting the linear expression of  $\eta$ , the tests can be written in terms of the regression parameters and are reduced to  $H_0: \delta_j + \gamma_{hj} + \gamma_{h'j} - \delta_{j'} - \gamma_{hj'} - \gamma_{h'j'} = 0$ . Following the idea of the penalized approach to ANOVA [Bondell and Reich, 2009], the

proposed penalized method estimates the haplotype, drug, and interaction effects as (for brevity, let  $\varphi = (\gamma_0, \beta, \delta, \gamma)^T$ )

$$\hat{\varphi}_\lambda = \operatorname{argmin}_\varphi \left\{ -\ell_n(\varphi, \xi; Y, G, E) + \lambda \sum_{hh'} \sum_{j < j'} w_{hh'jj'} |\delta_j + \gamma_{hj} + \gamma_{hj'} - \delta_{j'} - \gamma_{hj'} - \gamma_{hj'}| \right\}$$

subject to  $\sum_h \beta_h = \sum_j \delta_j = \sum_h \gamma_{hj} = \sum_j \gamma_{hj} = 0$ , where  $\ell_n(\varphi, \xi; Y, G, E)$  is the log-likelihood,  $\xi$  is a (possible) set of nuisance parameters (e.g. haplotype frequencies),  $\lambda$  is the non-negative regularization parameter that controls the amount of shrinkage, and  $w_{hh'jj'}$  are data-dependent weights. By placing an  $L_1$ -norm penalty on each of the pair-wise comparisons, the penalized method can set these differences in linear predictors to be exactly zero if the value of  $\lambda$  is large enough. It is this feature that allows the procedure to perform parameter estimation while simultaneously considering the overall drug effect structure for each diplotype. The linear predictors of drugs that perform similarly for a particular diplotype will be shrunk towards each other. As a result, the proposed method will be able to collapse some linear predictors to exact equality, and thus yield the overall drug effect structure for each diplotype directly, without leading to any contradictory groupings, which is the desired personalized output.

The proposed method places a different penalty on each pair-wise comparison through the use of adaptive weights that are inversely proportional to their effect size. Consequently, linear predictors that are similar receive larger penalties on their pair-wise comparisons and are more readily grouped together, while those with different effects remain separated. Zou [2006] suggests setting the weights as  $w_{hh'jj'} = |\tilde{\delta}_j + \tilde{\gamma}_{hj} + \tilde{\gamma}_{hj'} - \tilde{\delta}_{j'} - \tilde{\gamma}_{hj'} - \tilde{\gamma}_{hj'}|^{-\nu}$ , where  $\tilde{\varphi}$  is an initial root-n consistent estimator of  $\varphi$  and  $\nu > 0$  is an additional tuning parameter. In our

analysis, we chose  $\nu = 1$  and  $\tilde{\varphi}$  be the maximum likelihood estimate (MLE) of  $\varphi$  obtained from `haplo.glm` in R [Lake et al., 2003]. It has been shown that these adaptive weights have desirable asymptotic properties in this setting [Bondell and Reich, 2009]. The penalized solution ( $\hat{\varphi}_\lambda$ ) also depends on the value of  $\lambda$ . Many model selection criteria, such as Mallows's  $C_p$ , Akaike information criterion (AIC), Bayesian information criterion (BIC), and cross validation [Shao, 1997; Hastie, Tibshirani, and Friedman, 2009; Arlot and Celisse, 2010], can be used to determine the appropriate value of  $\lambda$  from an exhaustive grid search. Because the goal of this analysis is more aligned with selecting the true model structure than minimizing prediction error, we use BIC for tuning which can achieve consistent model selection [Yang, 2005]. BIC is defined as

$$BIC = -2\ell_n(\hat{\varphi}_\lambda, \hat{\xi}; Y, G, E) + df_\lambda \cdot \log(n)$$

where  $\ell_n(\hat{\varphi}_\lambda, \hat{\xi}; Y, G, E)$  is the log-likelihood evaluated at the estimated regression coefficients and maximized over  $\xi$  for a given  $\lambda$  and  $df_\lambda$  is the degrees of freedom based on  $\hat{\varphi}_\lambda$ . The quantity  $df_\lambda$  is given by  $\sum_{hh} \text{unique}_j\{\hat{\eta}_{hhj}\}$ , which equals the number of unique estimated linear predictors among all drugs within each homozygous diplotype. The  $\lambda$  that minimizes BIC is chosen as the regularization parameter, and its corresponding  $\hat{\varphi}_\lambda$  is the penalized estimate.

It should be noted that the data likelihood is a function of  $G$  and not  $H$ . Haplotypes are not directly observable and must be inferred from the genotype data during estimation. This requires an iterative procedure, such as the EM Algorithm [e.g. Lake et al., 2003], to optimize the likelihood. However, this can be computationally intensive when fitting the penalized method as the optimization must be done at every grid point  $\lambda$ . For computational convenience, the objective function created via the least squares approximation (LSA) method was used to

calculate the penalized solution. The LSA method replaces the objective function of the original penalized problem with a least squares objective function [Wang and Leng, 2007]. The method is motivated by a standard Taylor series expansion of  $-\ell_n(\varphi, \xi)$  about  $(\tilde{\varphi}, \tilde{\xi})$ , the function's unpenalized minimizer, and shows that the original penalized estimate has the exact same asymptotic distribution as the estimator given by

$$\hat{\varphi}_\lambda = \operatorname{argmin}_\varphi \left\{ (\varphi - \tilde{\varphi})^T \tilde{\Sigma}^{-1} (\varphi - \tilde{\varphi}) + \lambda \sum_{hh'} \sum_{j < j'} w_{hh'jj'} |\delta_j + \gamma_{hj} + \gamma_{hj'} - \delta_{j'} - \gamma_{h'j} - \gamma_{h'j'}| \right\}$$

where  $\tilde{\Sigma}$  is the estimated covariance matrix of  $\tilde{\varphi}$ . Because the underlying data likelihood is not quadratic in the regression coefficients, using the alternative least squares objective function greatly reduces the computational costs for finding the penalized solution [Weng and Lang, 2007]. Using the LSA method eliminates the need for an iterative procedure to optimize the data likelihood; it only requires one unpenalized fit of the original objective function and then a grid search to determine  $\lambda$ . The final estimate is again chosen by minimizing the BIC.

Employing similar logic, a penalty term can be developed to address the second research goal. Because its aim is to determine which genetic variants behave similarly for a given drug, all pair-wise comparisons among diplotypes for each drug need to be performed. The appropriate hypothesis tests of diplotype differences for a drug  $j$  are given by  $H_0: \eta_{hh'j} - \eta_{kk'j} = 0$  for  $hh' < kk'$ . Substituting the linear expression of  $\eta$ , the tests can be written in terms of the regression parameters and are reduced to  $H_0: \beta_h + \beta_{h'} + \gamma_{hj} + \gamma_{h'j} - \beta_k - \beta_{k'} - \gamma_{kj} - \gamma_{k'j} = 0$ , which results in a penalty term that is complementary to the one developed under the first

research goal. The penalty terms for the first and second research goals are given by, respectively,

$$P_1(\lambda, \varphi) = \lambda \sum_{hh'} \sum_{j < j'} w_{hh'jj'} |\delta_j + \gamma_{hj} + \gamma_{h'j} - \delta_{j'} - \gamma_{hj'} - \gamma_{h'j'}| \text{ and}$$

$$P_2(\lambda, \varphi) = \lambda \sum_j \sum_{hh' < kk'} w_{hh'jj'} |\beta_h + \beta_{h'} + \gamma_{hj} + \gamma_{h'j} - \beta_k - \beta_{k'} - \gamma_{kj} - \gamma_{k'j}|.$$

In this way, the penalized method can be written succinctly as

$$\hat{\varphi}_\lambda = \operatorname{argmin}_\varphi \{-\ell_n(\hat{\varphi}_\lambda, \hat{\xi}; Y, G, E) + s_1 P_1(\lambda, \varphi) + s_2 P_2(\lambda, \varphi)\}$$

subject to  $\sum_h \beta_h = \sum_j \delta_j = \sum_h \gamma_{hj} = \sum_j \gamma_{hj} = 0$ , where  $s_1$  and  $s_2$  take on the value 0 or 1 to indicate which penalty term should be used when fitting the model. Choosing the penalty term depends on the research goal, which dictates the desired directionality of the effects collapsing. Once this choice is made, fitting the penalized model proceeds in the manner described above.

## SIMULATION STUDIES

We performed simulation studies to examine the performance of the proposed penalized likelihood method. We compared the performance of our method against the standard approach of regression followed by a post-hoc analysis to address the research goals described in the

previous section. The standard regression analysis was performed using `haplo.glm` in R, the haplotype-based method of Lake et al [2003]. Two post-hoc analysis procedures were considered – the ‘unadjusted’ and the ‘FDR-adjusted’ methods. The ‘unadjusted’ method performs the pair-wise testing without adjusting for multiple comparisons. The ‘FDR-adjusted’ method uses the Benjamini-Hochberg [1995] adjustment procedure to control the false discovery rate (FDR) among the multiple comparisons. In our simulations, we controlled the FDR at 5%. We also considered, but did not report the results of Bonferroni adjustment that controls the family-wise error rate. The procedure was too conservative and its power was negligible when compared to the others.

## **SIMULATION SETTINGS**

Our simulation studies were based on a haplotype distribution (given in Table 3.1) studied by Lin and Huang [2008]. The distribution is based on the common haplotypes formed by five SNPs found on chromosome 18 in the CEU sample of the HapMap data. The distribution was normalized so that the haplotype frequencies summed to 1. Using this haplotype distribution, we considered two simulation studies – a quantitative drug response and a binary drug response. We generated data under the context of the first research goal: determine which drugs perform similarly for a given diplotype (i.e. individual). We considered 5 drugs, and the 5 haplotypes from the frequency distribution leads to 15 possible diploypes and a table of 75 possible linear predictors. To investigate the performance of the three methods, we varied the drug effect structure among the diploypes (given in Table 3.2). Specifically, we considered three increasingly complex drug effect structures – Null, One Best Drug (1BD) and Two Best Drugs (2BD). Under the Null structure, all drugs perform the same with respect to each other; the only

diplotype assigned to this structure is H5H5. Under the 1BD structure, there is one drug that outperforms the other four drugs, meaning it has a higher or lower linear predictor than the others. The effect size difference of the ‘best’ drug is denoted by  $\theta$  in Table 3.2. Examples of diplotypes assigned to this drug structure are H1H1 and H1H5. It is important to note that the magnitude of the effects for H1H1 is twice that of H1H5, a byproduct of the assumed additive haplotype-interaction model and the null structure placed on H5H5. As result, there are two 1BD drug structures, referred to as 1BD Full and 1BD Half respectively, which allowed us to examine the effect of effect magnitude as well as the effect of the drug structure on the performance of the methods. Under the 2BD structure, there are two drugs that outperform the other three drugs, but are equivalent to each other; an example of a diplotype assigned to this structure is H1H2. We set the values of  $n$  and  $\theta$  so that the power of identifying the effect structures fell within a reasonable range. For the quantitative response simulation, we set  $n = 500$  and  $\theta = \{1.5, 2.0\}$ . For the binary response simulation, we set  $n = 750$  and  $\theta = \{2.0, 2.5\}$ . We assumed a balanced design among the different drugs; that is, the same number of individuals assigned to each drug.

## COMPUTATIONAL DETAILS

We generated the data prospectively by first randomly selecting  $n$  diplotypes from the two trial multinomial distribution parameterized by the haplotype frequencies given in Table 3.1. Each diplotype was then randomly paired with a drug using a balanced design. Table 3.3 gives the specific values of the regression coefficients ( $\varphi = (\gamma_0, \beta, \delta, \gamma)^T$ ) used to induce the drug effect design given in Table 3.2 and uses H5 as the baseline haplotype and Drug5 as the baseline drug. Given Table 3.3, a quantitative drug response was generated under a normal distribution with a mean of  $\varphi^T X$ , where  $X = (H, E, H \otimes E)^T$ , and a standard deviation of  $\sigma_\varepsilon = 1$ . A binary

drug response was generated under a Bernoulli distribution with  $p = P(Y = 1|X) = \text{logit}^{-1}(\varphi^T X)$ .

For each simulation setting, 500 replicate data sets were generated. For each data set, analysis began by calculating the unpenalized MLEs of the regression coefficients ( $\tilde{\varphi}$ ) and their estimated covariance matrix ( $\tilde{\Sigma}$ ) using `haplo.glm` in R. To determine which drugs perform similarly (i.e. which pairs of drugs should be grouped together or not) for each diplotype, the tests  $H_0: \eta_{hhvj} - \eta_{hhvj'} = \delta_j + \gamma_{hj} + \gamma_{hj'} - \delta_{j'} - \gamma_{hj'} - \gamma_{hj'} = 0$  need to be performed for  $j < j'$  and for all diplotypes  $hh'$ . This leads to  $\binom{d}{2} \cdot \left[ l + \binom{l}{2} \right]$  tests, where  $\binom{d}{2}$  is the number of pair-wise comparisons between  $d$  drugs and  $\left[ l + \binom{l}{2} \right]$  is the number of possible diplotypes created from  $l$  haplotypes. In our simulation studies, there are 10 pair-wise comparisons between the 5 drugs and 15 diplotypes that can be created from the 5 haplotypes, leading to 150 total tests. For the unadjusted and the FDR-adjusted methods, the unpenalized MLEs and their covariance matrix were used directly to perform the tests using the asymptotic normal distribution of  $\eta_{hhvj} - \eta_{hhvj'}$ , described by Lake et al. [2003]. For the penalized method, the final penalized estimates of the regression coefficients ( $\hat{\varphi}$ ) were computed using the unpenalized MLEs and their covariance matrix in the penalized likelihood with  $s_1 = 1$  and  $s_2 = 0$ . Tuning was achieved by minimizing BIC. The significance of a test was then determined by computing all 75 linear predictors using the penalized estimate and checking which estimated predictors were identical within each diplotype. If the estimated linear predictors are identical for two drugs for a particular diplotype, then the null hypothesis for that pair of drugs would fail to be rejected, and we would conclude that those drugs perform similarly for that particular diplotype. The converse is true if the estimated linear predictors are not identical, and we would conclude that

those drugs performed differently for that particular diplotype. Based on the testing results, power measures were calculated over the 500 simulated runs to compare the performance of the three methods.

## SIMULATION RESULTS

We present power results, true negative rates (for diplotype H5H5) and true positive rates (for all other diplotypes), in Figures 3.1 – 3.3. Each figure contains the results of the quantitative and binary simulations (across columns) at both levels of the effect difference  $\theta$  (down rows). Because the power of each method decreased (increased) as diplotype frequency decreased (increased), for brevity, we only present the results for 7 diplotypes, which represent the other diplotypes with similar frequencies. The selected diplotypes are listed on the x-axis of each figure. For each drug effect structure, we selected a representative ‘low’ and ‘high’ frequency diplotype. What was considered a ‘low’ versus ‘high’ frequency diplotype was relative to the drug structure and is depicted in each figure as unshaded (low) versus shaded (high) point character. To compare the performance of the three procedures, we calculated the power to correctly identify the complete drug effect structure for a given diplotype, referred to as *correct structure power* and the power to identify just the best drug(s) for a given diplotype, referred to as *best drug power*. In each figure, the solid line with the square point characters represents our penalized likelihood method. The unadjusted method is represented by the dotted line with the triangle point characters, and the FDR-adjusted method is represented by the dashed line with the circle point characters.

## CORRECT STRUCTURE POWER

Figure 3.1 presents the correct structure power (CSP) results, which is the proportion of runs a method was able to correctly identify the inter-relationship among all drugs for a given diplotype. For all three methods, the CSP is lower for the binary drug response than the quantitative drug response (except in the Null drug structure), but the relative patterns among the different methods are similar across the two simulation studies. Under the Null drug structure (diplotype H5H5), the unadjusted method yields a low true negative rate, while the penalized and FDR-adjusted methods yield similar true negative rates. A low true negative rate translates into a high overall Type I error rate, where the Type I error rate for the unadjusted method ranges from 0.27 to 0.30 versus 0.02 to 0.08 and 0.02 and 0.14 for the penalized and FDR-adjusted methods, respectively. For the remaining diplotypes in Figure 3.1, the unadjusted method performs comparably or better than the penalized method (i.e. dotted line is about the same level or higher than the solid line). However, this gain in power comes at the expense of having a much higher Type I error rate than the penalized method and, hence, is not really comparable. On the other hand, the FDR-adjusted method maintains a similar Type I error rate as the penalized method, but its performance suffers as a result (i.e. the dashed line is about the same level or lower than the solid line). From these findings, it's apparent that the standard approaches must compromise between identifying true negatives and true positives, and the penalized method finds a better balance between identifying these true signals. As a result, the penalized method can uncover the underlying drug effect structure more accurately for a given diplotype. As expected, as the difference in effect size increases (i.e. power in top row of Figure 3.1 < power in bottom row of Figure 3.1) or as the diplotype frequency increases (i.e. unshaded power < shaded power) or as the effect magnitude increases (i.e. 1BD Half power < 1BD Full power), the CSP increases.

Thus, the power of a method to identify the correct drug structure for a particular diplotype depends on the underlying diplotype frequency, the effect size, and the complexity of the drug structure. The largest gain in performance of the penalized method occurs when the diplotype frequency is low, the effect size small, or the effect structure is more complex.

### **BEST DRUG POWER – ONE BEST DRUG STRUCTURE**

Figure 3.2 presents the best drug power (BDP) for the 1BD structure. Under this structure, we defined BDP as the proportion of runs a method was able to identify the single best drug as significantly better than all other drugs for a given diplotype. This power definition is less strict than the definition of correct structure power. A method only has to correctly evaluate 4 pair-wise tests among the drugs rather than all 10 for a given diplotype; the relationship between the non-best drugs is ignored. As a result, the BDP is higher than the CSP for a given diplotype with the 1BD structure across all simulation scenarios (i.e. power in Figure 3.1 < power in Figure 3.2). When comparing the BDP between the different methods, the power results in Figure 3.2 mirror those in Figure 3.1. Again, as the diplotype frequency increases (i.e. unshaded power < shaded power) or as the effect magnitude increases (i.e. 1BD Half power < 1BD Full power), the BDP increases. The BDP of the penalized method is comparable to or better than the BDP of the unadjusted and FDR-adjusted methods (i.e. solid line is about the same level or higher than the dotted or dashed lines). The penalized method's power gain over the two standard methods is more pronounced for BDP than for CSP (i.e. solid line – dotted (dashed) line in Figure 3.1 < solid line – dotted (dashed) line in Figure 3.2). This gain diminishes as diplotype frequency increases. These results again suggest that the penalized method outperforms the standard approaches and can more accurately determine which drug is the best

for a given diplotype among a group of treatments, even if the complete drug structure was not identified.

### **BEST DRUG POWER – TWO BEST DRUGS STRUCTURE**

Figure 3.3 presents the best drug powers (BDPs) for the 2BD structure. Under the 2BD structure, we defined two separate BDPs. The first is the proportion of runs a method was able to find the two best drugs as significantly better than all other drugs; this is referred to as *Both Found* BDP. Because this power definition ignores the relationship between the two best drugs, we also defined a second BDP for this drug structure, which is the proportion of runs a method was able to find the two best drugs as significantly better than all other drugs and equivalent to each other; this is referred to as *Both & Exact* BDP. These power definitions are less strict than the definition of CSP, but not as liberal as the BDP definition for 1BD drug structure. A method has to correctly evaluate 6 or 7 pair-wise tests, for the Both Found and the Both & Exact BDPs respectively, among the drugs rather than all 10 for a given diplotype; the relationship between the non-best drugs is ignored. As a result, the BDPs are higher than the CSP for a given diplotype with the 2BD structure across all simulation scenarios (i.e. power in Figure 3.1 < power in Figure 3.3). As expected, the Both BDP is less than the Both & Exact BDP for a given diplotype (i.e. unshaded (shaded) Both power < unshaded (shaded) Both & Exact power). For both BDPs, as the diplotype frequency increases, the power increases (i.e. unshaded power < shaded power for Both Found (Both & Exact)). When comparing the BDPs between the different methods, the penalized method has uniformly higher BDPs (i.e. solid line above the dotted and dashed lines). These results again suggest that the penalized method outperforms the standard approaches and can more accurately determine which drugs are the best for a given diplotype among a group of treatments, even if the complete drug structure was not identified.

## ANALYSIS OF SCOTROC DATA

The penalized method proposed in this work was applied to the pharmacogenetic data from the Scottish Randomized Trial in Ovarian Cancer (SCOTROC) [McWhinney-Glass et al., *submitted*]. The trial investigated the effects of two treatment regimes involving taxane/platinum-based chemotherapy: Treatment A refers to docetaxel-carboplatin and Treatment B refers to paclitaxel-carboplatin. In a preliminary analysis, McWhinney-Glass et al. found that the gene *BCL2* was significantly associated with an increased risk of experiencing severe neurotoxicity. Focusing on this gene, the goal of our analysis was to investigate potential gene and gene-treatment interactions. Specifically, we wanted to determine if individuals responded differently to the treatment they received based on their genetic make-up. In other words, will an individual have a higher or lower risk of experiencing severe neurotoxicity based on their diplotype when given a particular chemotherapy treatment? In practice, the post-hoc analysis required by the standard approach to address pharmacogenetic research goals is usually under powered after adjusting for multiplicity. Furthermore, the pair-wise comparisons can yield contradictory conclusions about the overall effect structure by producing overlapping groups that are hard to interpret clinically. These issues were encountered in our analysis of the SCOTROC data (see Figure 3.4 and Table 3.6).

In addition to collecting information on treatment outcomes and toxicities during the trial, each patient was genotyped at 1536 SNPs chosen from three categories of candidate genes – nerve related genes, inherited peripheral neuropathy genes, and drug related genes. Patients that had less 90% complete genotype information were excluded from the analysis. In addition, any patient with missing phenotype information was also excluded, resulting in a final sample size of 808 patients. SNPs that demonstrated significant deviations from HWE or were missing for more

than 10% of the patients were removed from the analysis, resulting in 1303 SNPs under consideration. We focused our analysis on the 105 SNPs genotyped within the *BCL2* gene. A sliding window analysis was used to investigate the existence of diplotype-treatment interactions. For each window, the resulting data was analyzed using the three procedures discussed in the Simulation Study section – the penalized method, the unadjusted method, and the FDR-adjusted method. All methods were performed assuming an additive haplotype-interaction model. For each method, all pair-wise comparisons between diplotypes were performed within each treatment. That is, all tests of the form  $H_0: \eta_{hh'j} - \eta_{kk'j} = \beta_h + \beta_{h'} + \gamma_{hj} + \gamma_{h'j} - \beta_k - \beta_{k'} - \gamma_{kj} - \gamma_{k'j} = 0$  were tested for all diplotypes  $hh' < kk'$  and for both treatments  $j$ . For the unadjusted and FDR-adjusted methods, the testing occurs after the regression parameters have been estimated. For the penalized method, the estimation and testing occur simultaneously and the penalized likelihood is fit using  $s_1 = 0$  and  $s_2 = 1$ . The sliding window analysis was performed using a window of size 6 SNPs, which resulted in 100 adjacent windows within the *BCL2* gene. Note that although the FDR-adjusted method adjusts for multiple testing within a particular window, it does not adjust for the multiplicity that results from the sliding window analysis. Hence, the error rate is not controlled overall.

Figure 3.4 presents the results of the sliding window analysis of the gene *BCL2* for all three methods. In each plot, the findings are broken down into three categories: (1) method found no significant differences between the diplotypes in either treatment (denoted by “No SDs” on the y-axis); (2) method found significant differences between the diplotypes within at least one of the treatments (denoted by “SDs” on the y-axis); and (3) method found significant differences between the diplotypes within at least one of the treatments and the diplotype groupings did not contradict each other (denoted by “SDs & No CGs” on the y-axis). The crosses (+) indicate the

findings of the penalized method for each window, the triangles ( $\Delta$ ) indicate the findings of the unadjusted method, and the circles ( $\circ$ ) indicate the findings for the FDR-adjusted method. The findings of the penalized method are compared to the unadjusted method in Figure 3.4a and the FDR-adjusted method in Figure 3.4b.

The standard methods are able to detect significant differences between diplotypes within a treatment (i.e. string of triangles (circles) at the “SDs” position in Figure 3.4a (3.4b)). However, in each case, the groupings led to contradictory conclusions about which diplotypes had similar effects for a given treatment (i.e. lack of triangles (circles) at the “SDs & No CGs” position in Figure 3.4a (3.4b)). Such results are difficult to interpret clinically. For example, both standard methods found that diplotypes H1H1 and H2H2 did not differ in their effects under Treatment A for Window 73 in the *BCL2* gene (i.e. the pair-wise test between these diplotypes was non-significant, indicated by – at the intersection of the H1H1 row (column) and H2H2 column (row) in Table 3.6). In addition, both standard methods found that diplotypes H2H2 and H4H4 had different effects under Treatment A (i.e. the pair-wise test between these diplotypes was significant, indicated by  $\Delta$  for the unadjusted method and  $\circ$  FDR-adjusted method at the intersection of the H2H2 row (column) and H4H4 column (row) in Table 3.6). From this it should follow that H1H1 and H4H4 have different effects under Treatment A, but instead both standard methods find that the diplotypes do not have significantly different effects (i.e. there is – at the intersection of the H1H1 row (column) and H4H4 column (row) in Table 3.6). The contradictory grouping of diplotype effects could result from the significant test of H2H2 vs. H4H4 being a false positive or the non-significant tests of H1H1 vs. H2H2 or H1H1 vs. H4H4 being false negatives. Either way, clinicians are left wondering how to interpret the contradictory groupings. For example, consider grouping only the homozygous diplotypes under Treatment A.

Based on the output of the standard approaches, two group structures would emerge: Group A = {H1H1, H2H2, H3H3} and Group B = {H1H1, H3H3, H4H4}. How should clinicians interpret the overlapping group structures when trying to determine which diplotypes incur a higher risk of severe neurotoxicity? From the significant test of H2H2 vs. H4H4, they can infer that patients with the diplotype H2H2 will incur a higher risk of severe neurotoxicity than patients with the diplotype H4H4 because its estimated odds are larger (1.155 vs. 0.314, respectively). But what about patients with diplotypes H1H1 and H3H3 – should clinicians consider these patients to have elevated risk because they are in Group A or not because they are in Group B? In this context, overlapping group structures prevent the development of coherent personalized treatments. For individualized medicine to be clinically relevant, these issues need to be resolved.

The penalized method is able to overcome these hurdles by combining the estimation and testing into a single procedure (i.e. the presence of plus marks at the “SDs & No CGs” position in Figure 3.4). Although the penalized method found significant differences in six windows (8, 31, 72, 73, 74, and 100), specific results are only presented for Window 73 because the method found significant differences for a string of adjacent windows surrounding Window 73 (see Figure 3.4). Table 3.4 lists the SNPs corresponding to this window and the resulting haplotypes; Tables 3.5 and 3.6 present the specific results of the penalized, unadjusted, and FDR-adjusted analyses for this window. The penalized method found three distinct groupings of diplotypes under Treatment A, while no significant differences were found among the diplotypes under Treatment B (i.e. three unique values for the estimated odds under Treatment A and only one under Treatment B in Table 3.5). It appears that H2 may be the causal haplotype, or at least in linkage disequilibrium (LD) with the causal variant. With each additional copy of H2, the risk of

a patient undergoing Treatment A experiencing severe neurotoxicity increases. Thus, patients with no copies of H2 have the lowest risk, with one copy have an elevated risk, and with two copies have the highest risk of experiencing the adverse outcome. The standard approaches detected very few of the same significant differences found by the penalized method (see Table 3.6). Neither approach was able to separate out all diplotypes involving H2, or even separate out the homozygous diplotype H2H2 from all other diplotypes not involving H2. This leads to contradictory groupings of the diplotypes, like the one described above, and making the output of the standard approaches hard to interpret clinically. This will never occur when using the penalized method. Again consider grouping only the homozygous diplotypes under Treatment A. Unlike the overlapping group structure formed by the standard approaches, two distinct groups emerge when using the penalized method: Group A = {H2H2} and Group B = {H1H1, H3H3, H4H4}. From these results, it is clear that patients with diplotype H2H2 incur a different, in this case elevated, risk of experiencing severe neurotoxicity than patients with the other homozygous diplotypes. This suggests that haplotype H2, and perhaps other variants in LD with it, should be investigated further. Because it automatically yields non-overlapping group structures, output from the penalized method (Table 3.5) is easily interpretable and can be used directly in the development and implementation of individualized medicine.

## DISCUSSION

Through the use of genetic personalization, pharmacogenetics aims to reduce variation in how individuals respond to drug therapies. By investigating genetic variants, and how they interact with each other and treatments, pharmacogenetics can uncover predispositions to adverse drug reactions or determine which among several alternative drugs is most effective for an

individual based on their genetic make-up. Haplotype-based association methods provide an attractive and flexible approach for addressing these research goals which require a complete understanding of the effect structure underlying the genetic variants. Many methods are available for investigating haplotypic effects, but most require a complementary post-hoc analysis to test the significance of pair-wise effect comparisons in order to appropriately address these research goals. As a result, these methods are typically under powered after adjusting for multiplicity and often lead to contradictory conclusions about the effect structure.

In this work, we introduce a penalized likelihood approach that is powerful and that can effectively address research goals pertinent in pharmacogenetics. Our method places an  $L_1$ -penalty on all pair-wise effect comparisons dictated by the research goals. By combining effect estimation and testing into a single procedure, the penalized method can bypass the issues associated with multiple testing. As a result, our penalized method directly produces personalized output that is easily interpretable and clinically relevant. When compared with standard haplotype-based approaches, the penalized method finds a better balance between identifying true negatives and true positives. Thus, the penalized method has comparable or more power, while maintaining reasonable Type I error rates, to identify the underlying effect structure for a particular drug or genetic variant. The largest gain in performance for the penalized method occurs when the diplotype frequency is low, the effect size small, or the effect structure is more complex.

In addition, the penalized method can be used to decipher the global signals from gene-based association studies of genetic variants that may influence drug response. Because gene-based methods consider all variants within a gene jointly, they are considered to be a promising tool for pharmacogenetic research. Drug response is believed to be influenced by both regulatory

and structural polymorphisms within in the gene and by a wide range of gene products [Goldstein et al., 2003]. Marker-based methods might miss these signals; however, because gene-based methods aggregate information across all variants within a gene, they can better capture these effects. The compromise is that gene-based methods assign a single p-value to all variants within a gene. Our penalized method can be used to characterize which variants may be responsible for the signal.

We investigated the performance of the penalized method under a prospective design, but the framework can easily be extended to consider a retrospective design. Using a prospective likelihood when a retrospective likelihood is dictated by the sampling scheme can be detrimental when performing haplotype-based association analyses [Koehler et al., 2010]. The penalized retrospective likelihood can be optimized directly through the use of an iterative procedure; however, we suggest taking advantage of the least square approximation to alleviate the computational burden. This approach requires an initial optimization of the unpenalized method to obtain effect estimates and their estimated covariance matrix; one example of freely available software for obtaining retrospective estimates is HAPSTAT [Lin et al., 2005]. Although this work focused on investigating gene-drug interactions, the penalized approach can also be used investigate gene-gene interactions by considering the diplotypes of another gene in place of the drug therapies. The penalized method can also be extended to simultaneously collapse effects across both factors being studied (i.e. taking  $s_1 = 1$  and  $s_2 = 1$ ) so that the personalized output can be compared across rows and down columns, rather than in one direction only.

## TABLES AND FIGURES

**Table 3.1:** Haplotype frequency distribution used in simulations

Hap ID	Haplotype	Frequency
H1	00001	0.225
H2	01111	0.149
H3	10000	0.139
H4	10001	0.058
H5	00000	0.429

**Table 3.2:** Drug effect structure for the linear predictors (by diplotype) used in simulations\*

	Drug1	Drug2	Drug3	Drug4	Drug5	Diplo Freq
<b>H1H1</b>	$c + \theta$	$c$	$c$	$c$	$c$	0.051
<b>H1H2</b>	$c + \theta/2$	$c + \theta/2$	$c$	$c$	$c$	0.067
<b>H1H3</b>	$c + \theta/2$	$c$	$c + \theta/2$	$c$	$c$	0.063
<b>H1H4</b>	$c + \theta/2$	$c$	$c$	$c + \theta/2$	$c$	0.026
<b>H1H5</b>	$c/2 + \theta/2$	$c/2$	$c/2$	$c/2$	$c/2$	0.193
<b>H2H2</b>	$c$	$c + \theta$	$c$	$c$	$c$	0.022
<b>H2H3</b>	$c$	$c + \theta/2$	$c + \theta/2$	$c$	$c$	0.042
<b>H2H4</b>	$c$	$c + \theta/2$	$c$	$c + \theta/2$	$c$	0.017
<b>H2H5</b>	$c/2$	$c/2 + \theta/2$	$c/2$	$c/2$	$c/2$	0.128
<b>H3H3</b>	$c$	$c$	$c + \theta$	$c$	$c$	0.019
<b>H3H4</b>	$c$	$c$	$c + \theta/2$	$c + \theta/2$	$c$	0.016
<b>H3H5</b>	$c/2$	$c/2$	$c/2 + \theta/2$	$c/2$	$c/2$	0.119
<b>H4H4</b>	$c$	$c$	$c$	$c + \theta$	$c$	0.003
<b>H4H5</b>	$c/2$	$c/2$	$c/2$	$c/2 + \theta/2$	$c/2$	0.050
<b>H5H5</b>	$\gamma_0$	$\gamma_0$	$\gamma_0$	$\gamma_0$	$\gamma_0$	0.184

\* We set  $\theta = \{1.5, 2.0\}$  and  $c = 1 + \gamma_0$  for the quantitative simulations with  $n = 500$ , and  $\theta = \{2.0, 2.5\}$  and  $c = 0.4 + \gamma_0$  for the binary simulations with  $n = 750$ .

**Table 3.3:** Specific values of the regression parameters ( $\varphi$ ) used to induce the effect structure of Table 3.2.

Effect	Parameter	Quantitative Simulation		Binary Simulation	
		$\theta = 1.5$	$\theta = 2.0$	$\theta = 2.0$	$\theta = 2.5$
Intercept	$\gamma_0$	0	0	-0.6	-0.6
H1	$\beta_1$	0.5	0.5	0.2	0.2
H2	$\beta_2$	0.5	0.5	0.2	0.2
H3	$\beta_3$	0.5	0.5	0.2	0.2
H4	$\beta_4$	0.5	0.5	0.2	0.2
Drug1	$\delta_1$	0	0	0	0
Drug2	$\delta_2$	0	0	0	0
Drug3	$\delta_3$	0	0	0	0
Drug4	$\delta_4$	0	0	0	0
H1*Drug1	$\gamma_{11}$	0.75	1	1	1.25
H2*Drug1	$\gamma_{21}$	0	0	0	0
H3*Drug1	$\gamma_{31}$	0	0	0	0
H4*Drug1	$\gamma_{41}$	0	0	0	0
H1*Drug2	$\gamma_{12}$	0	0	0	0
H2*Drug2	$\gamma_{22}$	0.75	1	1	1.25
H3*Drug2	$\gamma_{32}$	0	0	0	0
H4*Drug2	$\gamma_{42}$	0	0	0	0
H1*Drug3	$\gamma_{13}$	0	0	0	0
H2*Drug3	$\gamma_{23}$	0	0	0	0
H3*Drug3	$\gamma_{33}$	0.75	1	1	1.25
H4*Drug3	$\gamma_{43}$	0	0	0	0
H1*Drug4	$\gamma_{14}$	0	0	0	0
H2*Drug4	$\gamma_{24}$	0	0	0	0
H3*Drug4	$\gamma_{34}$	0	0	0	0
H4*Drug4	$\gamma_{44}$	0.75	1	1	1.25

**Table 3.4:** Haplotypes and corresponding SNPs found in Window 73 in *BCL2* gene

	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6	
SNP ID	RS4987821	RS4987825	RS4987828	RS4987835	RS4987839	RS4987844	Hap Freq
H1	0	0	0	1	1	0	0.098
H2	0	0	1	0	0	0	0.270
H3	1	1	0	0	0	0	0.073
H4	0	0	0	0	0	0	0.558
Minor/Major Allele	A/G	A/C	A/G	G/A	G/A	A/A	

**Table 3.5:** Results of penalized analysis for Window 73 in *BCL2* gene\*

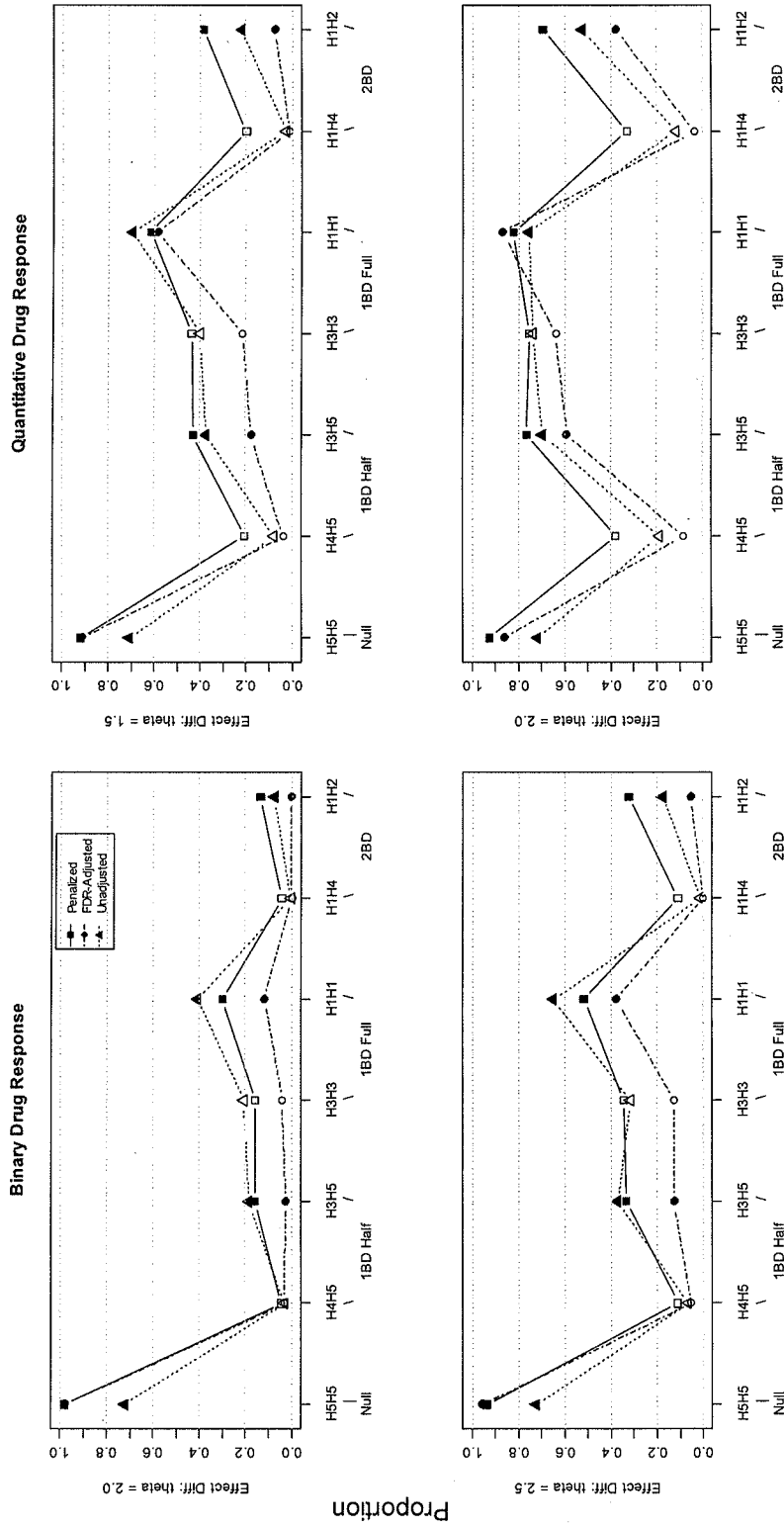
	Treatment A	Treatment B
H2H2	0.970	0.139
H1H2	0.589	0.139
H2H3	0.589	0.139
H2H4	0.589	0.139
H1H1	0.357	0.139
H1H3	0.357	0.139
H1H4	0.357	0.139
H3H3	0.357	0.139
H3H4	0.357	0.139
H4H4	0.357	0.139
<i>df</i>	3	1

\* Values given are the estimated odds of each cell

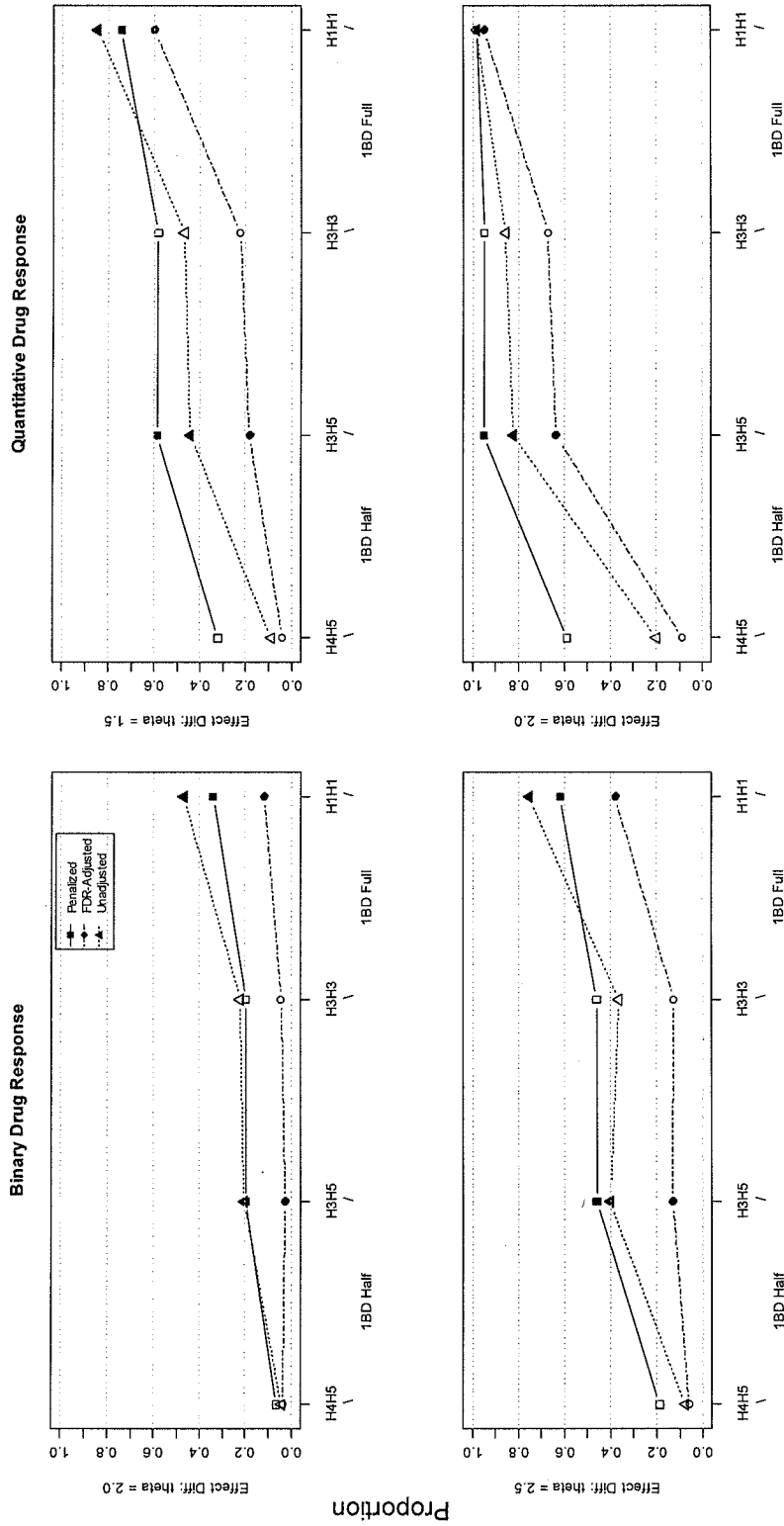
**Table 3.6:** Pair-wise testing results of Unadjusted and FDR-Adjusted analyses for Window 73 in *BCL2* gene under Treatment A\*

	H1H 1	H1H 2	H1H 3	H1H 4	H2H 2	H2H 3	H2H 4	H3H 3	H3H 4	H4H 4
H1H 1	-	-	-	-	-	-	-	-	-	-
H1H 2	-	-	-	Δ ○	-	-	-	-	-	-
H1H 3	-	-	-	-	Δ	-	-	-	-	-
H1H 4	-	Δ ○	-	-	Δ	-	-	-	-	-
H2H 2	-	-	Δ	Δ	-	-	Δ ○	-	Δ	Δ ○
H2H 3	-	-	-	-	-	-	-	-	Δ ○	-
H2H 4	-	-	-	-	Δ ○	-	-	-	-	Δ ○
H3H 3	-	-	-	-	-	-	-	-	-	-
H3H 4	-	-	-	-	Δ	Δ ○	-	-	-	-
H4H 4	-	-	-	-	Δ ○	-	Δ ○	-	-	-

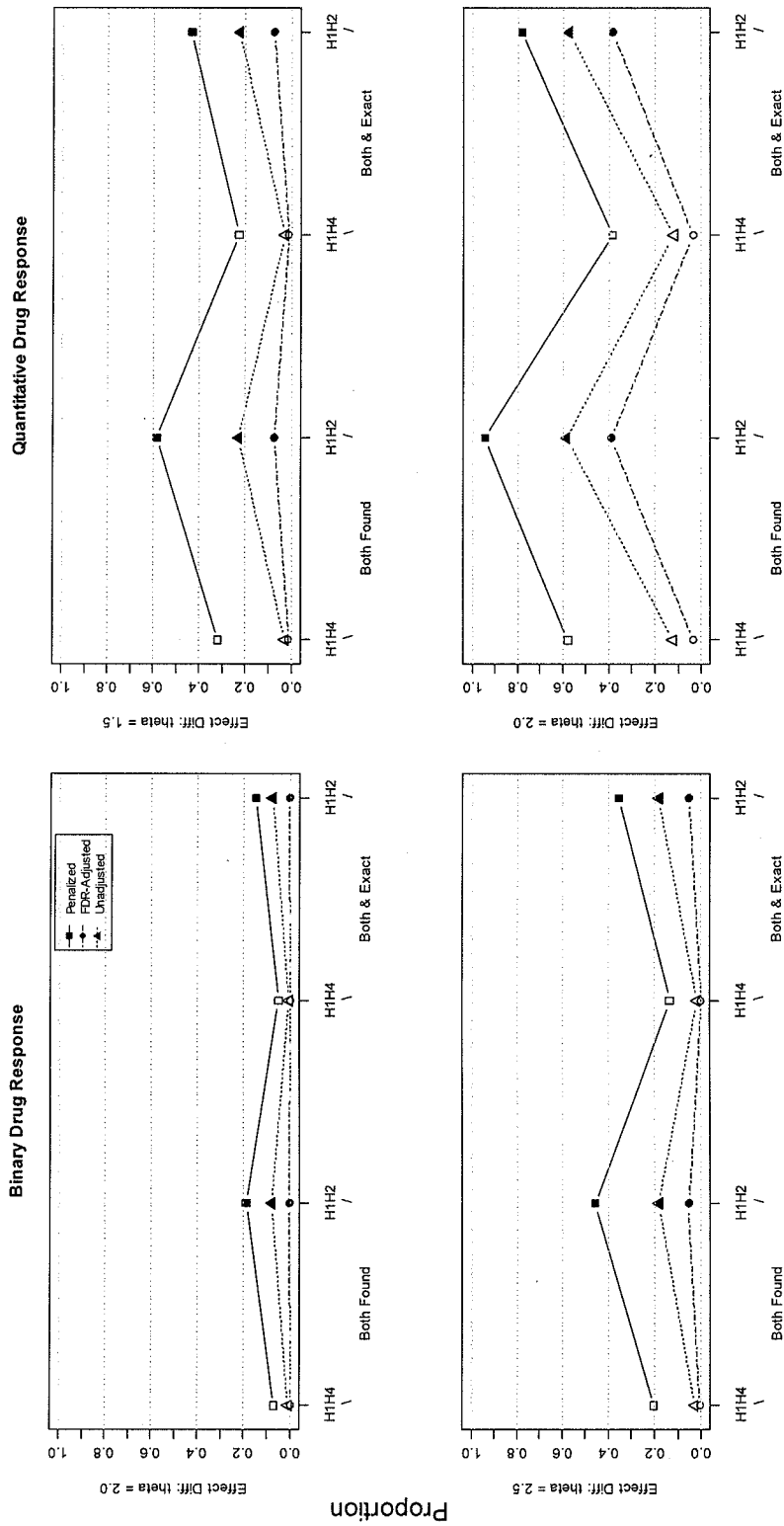
\* For a pair of diplotypes, - indicates the test was non-significant, Δ indicates the test was significant at  $\alpha = 0.05$  when using the Unadjusted method, and ○ indicates the test was significant at  $\alpha = 0.05$  when using the FDR-adjusted method.



**Figure 3.1: Correct Structure Power** – the proportion of runs the complete drug effect structure was correctly identified for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with  $\alpha$  set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the drug structure groups (Null, 1BD half, 1BD full, and 2BD; for definitions, see Simulation Studies section).

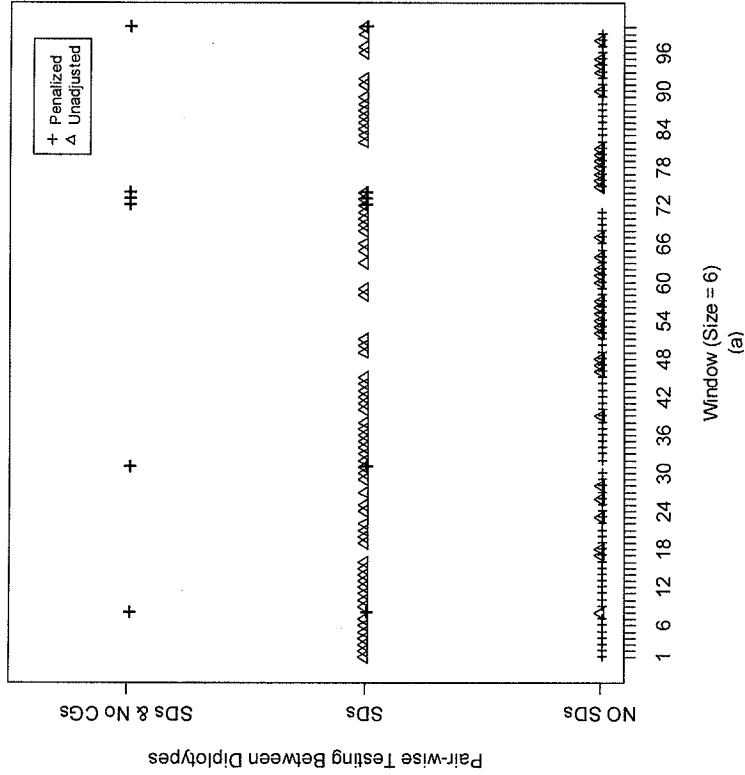


**Figure 3.2: Best Drug Power for 'One Best' Drug Effect Structure** – the proportion of runs the 'best' drug found significantly better than all other drugs for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with  $\alpha$  set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the drug structure groups (1BD half and 1BD full; for definitions, see Simulation Studies section).

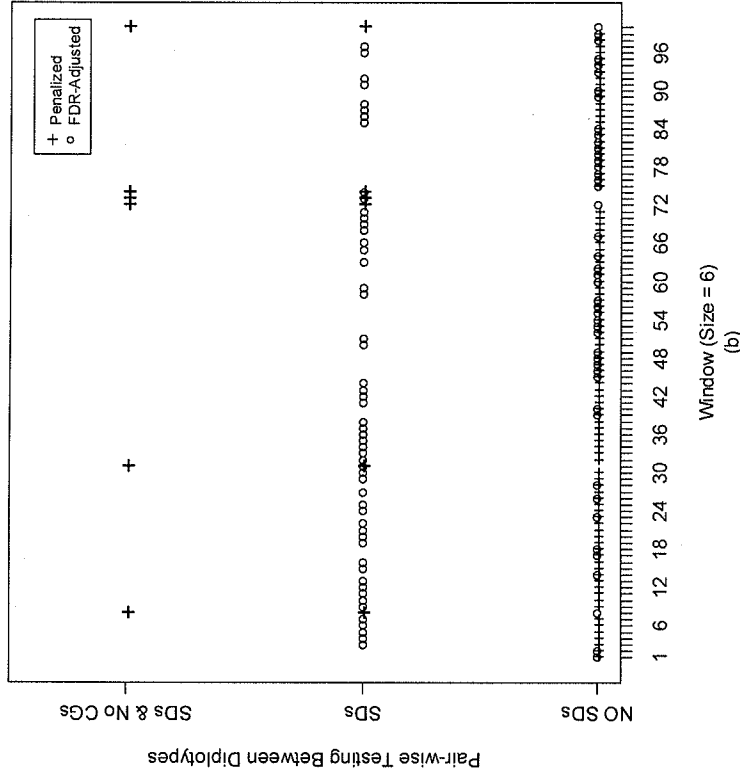


**Figure 3.3: Best Drug Powers for 'Two Best' Drugs Effect Structure** – the proportion of runs (1) the two 'best' drugs found significantly better than all other drugs (*Both Found*) and (2) the two 'best' drugs found significantly better than all other drugs and equivalent to each other (*Both & Exact*) for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with  $\alpha$  set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the 2BD drug structure group (for definition, see Simulation Studies section).

Sliding Window Analysis Results of BCL2 Gene:  
Penalized Method vs. Unadjusted Tests



Sliding Window Analysis Results of BCL2 Gene:  
Penalized Method vs. FDR-Adjusted Tests



**Figure 3.4: Results of analyzing SCOTROC data using the penalized vs. standard methods** – Results of the penalized analysis (+) are compared to the Unadjusted tests ( $\Delta$ ) with  $\alpha$  at 0.05 in Figure 4a and to the Benjamini-Hochberg adjusted tests (o) with FDR at 0.05 in Figure 4b. “NO SDs” on the y-axis indicates that a method found no significant differences (i.e.SD) between the diploypes within either treatment. “SDs” on the y-axis indicates that a method found significant differences between the diploypes within at least one of the treatments. “SDs & No CGs” on the y-axis indicates that a method found significant differences between the diploypes within at least one of the treatments and the diplotype groupings did not contradict each other (i.e. No CG).

## REFERENCES

- Arlot S, Celisse A. 2010. A survey of cross-validation procedures for model selection. *Stat Surveys* 4: 40 – 79.
- Balding DJ. 2006. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7: 781 – 791.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57: 289 – 300.
- Bondell HD, Reich BJ. 2009. Simultaneous factor selection and collapsing levels in ANOVA. *Biometrics* 65: 169 – 177.
- Clark AG. 2004. The role of haplotypes in candidate-gene studies. *Genet Epidemiol* 27: 321 – 333.
- de Bakker PW, Yelensky R, Pe'er I, Gabriel S, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. *Nat Genet* 37: 1217 – 1223.
- Goldstein DB, Tate SK, Sisodyia SM. 2003. Pharmacogenetics goes genomic. *Nat Rev Genet* 4: 937 – 947.
- Guo W, Lin S. 2009. Generalized linear modeling with regularization for detecting common disease rare haplotype association. *Genet Epidemiol* 33: 308 – 316.
- Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2<sup>nd</sup> Edition. Springer-Verlag.
- Hoerl DW, Kennard RW. 1970. Ridge regression: application to nonorthogonal problems. *Technometrics* 12: 55 – 68.
- Koehler ML, Bondell HD, Tzeng JY. 2010. Evaluating haplotype effects in case-control studies via penalized likelihood approaches: prospective or retrospective analysis? *Genet Epidemiol* 34: 892 – 911.
- Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ. 2003. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 55: 56 – 65.
- Li M, Romero R, Wenjiang JF, Cui Y. 2010. Mapping haplotype-haplotype interactions with adaptive LASSO. *BMC Genet* 11: 79.

- Li Y, Sung WK, Liu JJ. 2007. Association mapping via regularized regression analysis of single nucleotide polymorphism haplotypes in variable-sized sliding windows. *Am J Hum Genet* 80: 705 – 715.
- Lin DY, Huang BE. 2008. The use of inferred haplotypes in downstream analysis. *Am J Hum Genet* 80: 577 – 579.
- Lin DY, Zeng D, Milikan R. 2005. Maximum likelihood estimation of haplotype effects in genetic association studies. *J Am Stat Assoc* 101: 89 – 104.
- McWhinney-Glass S, Winham SJ, Revolla JY, Paul J, Brown R, Motsinger-Reif A, McLeod HL. 2011. Cumulative genetic risk predicts platinum/taxane-induced neurotoxicity. *Submitted*.
- Schaid DJ. 2004. Evaluating associations of haplotypes with traits. *Genet Epidemiol* 27: 348 – 364.
- Shao J. 1997. An asymptotic theory for linear model selection. *Stat Sinica* 7: 221 – 264.
- Tibshirani R. 1996. Regression shrinkage and selection via the LASSO. *J R Stat Soc* 58: 267 – 288.
- Tychonoff AN. 1963. Solution of incorrectly formulated problems and the regularization method. *Soviet Math* 4: 1035 – 1038.
- Tzeng JY, Lu W, Farmen MW, Liu Y, Sullivan PF. 2010. Haplotype-based pharmacogenetic analysis for longitudinal quantitative traits in the presence of dropout. *J Biopharm Stats* 20: 334 – 350.
- WengH, Leng C. 2007. Unified LASSO estimation by least squares approximation. *J Am Stat Assoc* 102: 1039 – 1048.
- Yang Y. 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika* 92: 937 – 950.
- Zaitlen M, Kang H, Eskin E, Halperin E. 2007. Leveraging the HapMap correlation structure in association studies. *Am J Hum Genet* 80: 683 – 691.
- Zou H. 2006. The adaptive LASSO and its oracle properties. *J Am Stat Assoc* 101: 1418 – 1429.