Published in final edited form as:
Pharmacogenomics. 2014 February ; 15(2): 137–146. doi:10.2217/pgs.13.213.

Genome-wide Association and Pharmacological Profiling of 29 Anticancer Agents Using Lymphoblastoid Cell Lines

Chad C. Brown[1], **Tammy M. Havener**[2], **Marisa W. Medina**[3], **John R. Jack**[1], **Ronald M. Krauss**[3], **Howard L. McLeod**[2], and **Alison A. Motsinger-Reif**[1,2,*]

[1]Bioinformatics Research Center, Department of Statistics, North Carolina State University, Raleigh, NC, 27607, USA

[2]Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[3]Children's Hospital Oakland Research Institute, Oakland, CA, 94609, USA

Abstract

Aims—Association mapping with lymphoblastoid cell lines (LCLs) is a promising approach in pharmacogenomics research, and in the current study we utilize this model to perform association mapping for 29 chemotherapy drugs.

Materials and Methods—Currently, we use LCLs to perform genome-wide association mapping of the cytotoxic response of 520 European Americans to 29 different anticancer drugs, the largest LCL study to date. A novel association approach using a multivariate analysis of covariance design was employed with the software program MAGWAS, testing for differences in the dose-response profiles between genotypes without making assumptions about the response curve or the biological mode of association. Additionally, by classifying 25 of the 29 drugs into 8 families according to structural and mechanistic relationships, MAGWAS was used to test for associations that were shared across each drug family. Finally, a unique algorithm using multivariate responses and multiple linear regressions across pairs of response curves was used for unsupervised clustering of drugs.

Results—Among the single drug studies, suggestive associations were obtained for 18 loci, 12 within/near genes. Three of these, MED12L, CHN2 and MGMT, have been previously implicated in cancer pharmacogenomics. The drug family associations resulted in 4 additional suggestive loci (3 contained within/near genes). One of these genes, HDAC4, associated with the DNA alkylating agents, shows possible clinical interactions with temozolomide. For the drug clustering analysis, 18 of 25 drugs clustered into the appropriate family.

Conclusions—This study demonstrates the utility of LCLs for identifying genes having clinical importance in drug response, for assigning unclassified agents to specific drug families, and proposes new candidate genes for follow-up in a large number of chemotherapy drugs.

[*] motsinger@stat.ncsu.edu.

Keywords

genome-wide association study; lymphoblastoid cell lines; chemotherapeutics

Introduction

The identification of genetic variants that affect drug response (pharmacogenomics) is a high priority in human genetics, and has yielded a large number of translational successes [1]. While progress in this field has been encouraging, there are limitations in mapping dose response outcomes that impede discovery of new genes/variants associated with drug response. First, pharmacogenomic studies are often nested within clinical trials. Although this is the most direct approach for assessing associations with clinical outcomes, it is subject to a number of limitations, including small sample sizes (often restricting testing to a small set of genetic variants determined a priori), suboptimal study designs (since trials are rarely designed for genetic aims), confounding effects from variability in patient treatment and ethical restrictions in experimental manipulation [2]. These restrictions also influence the availability of patient cohorts for replication assessment.

These limitations have prompted the use of lymphoblastoid cell lines (LCLs) as an in vitro model of dose response for cytotoxic chemotherapy drugs [3-5]. There are a number of advantages that motivate the use of LCLs as a model of cytotoxic response. First, because LCLs can be derived from healthy individuals, there is no limitation to sample size, allowing studies to be powered for genome-wide associations. Additionally, family-based samples can be used as a powerful tool for estimating heritability and detecting genetic associations. Moreover, because the drug exposure is controlled, there is no confounding due to imprecise measures of exposure, concomitant medications, etc. Finally, with the use of robotics for high-throughput phenotyping, this model can facilitate rapid discovery of genetic associations for response to a large number of drugs.

Previous studies with LCLs have shown promising results, with success in both linkage and association studies [6-11]. For example, it has been used to assess the heritability of a number of dose-response outcomes [4, 12], and to successfully detect genetic associations of clinical relevance, with notable instances being in head and neck cancer [13] and temozolomide response [10]. Additionally, similarity in dose response has been seen across drugs within a single drug family [14], reinforcing the potential of the model to categorize responses as a function of mechanism of action. The use of the LCL model system in pharmacogenomics has been reviewed elsewhere [5].

In a previous study [12], we used LCL dose response assays in family-based cell lines, derived from the Centre d’Etude du Polymorphisme Humain (CEPH) reference population [15], to assess the heritability of differential dose response for a range of FDA-approved cytotoxic chemotherapy drugs. The drugs were originally selected for this study because they were commonly prescribed in modern clinical practice, did not require in vivo activation, were not hormone therapies, and were soluble at the concentrations necessary to generate a dynamic range of cytotoxic activity. This study showed that dose response was highly heritable for many drugs, and was modestly heritable for all drugs. This was an important first step, as establishing heritability of a given trait is a necessary (but not sufficient) step in gene mapping. In this first study, we performed linkage analysis to identify regions that may be associated with the dose response outcomes, but as the study was underpowered and the linkage peaks identified were generally very broad, we have adopted an association approach for fine mapping.

In addition, simulation studies have demonstrated that modeling the vector of responses across drug concentrations jointly can be more powerful than previously used methods in the detection of differences in dose-response profiles between genotypes [16]. Here, we report the results from applying this method to real data in genome-wide association mapping for a cohort of LCLs exposed to 29 cancer drugs. In addition, similarity of dose-response profiles in LCLs could be used to classify drugs according to known mechanisms of action using no a priori knowledge, demonstrating the possibility of using LCLs to assign novel cancer agents to known drug classes.

Materials and Methods

Phenotype and Genotype Data Collection

520 Epstein-Barr virus immortalized LCLs were derived from unrelated Caucasian participants of the Cholesterol and Pharmacogenetics (CAP) clinical trial, as described in detail in [18]. Alamar blue assays were used to assess the drug response for all cell lines, across all drugs. All of the drugs (including dasatinib and sunitinib) demonstrated cell killing in this system. The system would not necessarily be a predictor of isolated b-call toxicity in a patient, but rather is used as a cell autonomous system for evaluating drug effect. LCLs were cultured at 37°C and under 5% CO2 using RPMI medium 1640 containing 2 mM L-glutamine (Gibco) and 15% fetal bovine serum (Sigma), with no media antibiotics. Each cell line was seeded on two 384 well plates, 4000 cells/well, with each plate having only one LCL. These two plate formats were for separate experiments, with LCL viability measurements taken after exposure 14 anticancer drugs on the first plate format and 15 on the second. With each plate format, LCLs were exposed to six concentrations for each drug. All drugs were assayed on all cell lines except for 2 drugs due to technical problems (299 cell lines for 5- fluorouracil and 221 for nilotinib). All drugs and concentrations used in the current study are given in Supplementary Table 1. To estimate laboratory reproducibility, a subset of 18 and 19 LCLs were each assayed twice on the first and second plate formats, respectively, where each replicate was performed on a different laboratory day.

In addition, every plate also included controls for background fluorescence signal and drug vehicle. Background signal was estimated from viability readings for LCLs exposed to a lethal dose of 10% dimethyl sulfoxide (DMSO). Drug vehicle effects were assessed by exposing cells to vehicle only (water or DMSO at 0.01, 0.1, 1 and 2 percent). Every observation for each plate (for both controls and drug exposures) was performed in 2X2 quadruplicates. Handling of the resulting 413,568 wells was automated using a Tecan EVO150 (Tecan Group Ltd) with a 96 head MultiChannel Arm™. All plates were incubated for 72 hours, dyed with Alamar Blue (Biosource International) and incubated for another 24 hours. After incubation, an Infinite F200 microplate reader with Connect Stacker (Tecan

Group Ltd) and iControl software (Version 1.6) was used to measure fluorescence intensity at EX535nm and EM595nm. The resulting raw fluorescence units (RFUs) are proportional to the concentration of living cells in each well.

All individuals in the CAP study were previously genotyped for either 314,621 or 620,901 SNPs, using HumanHap300 bead chip or HumanQuad610 bead chip platforms, respectively, as previously described [17, 18]. These markers were used to impute 2.5 million SNPs from HapMap Release 22, using the Caucasian CEPH reference population and the software program MACH [19], as previously described [17].

A five-stage quality control (QC) pipeline was implemented on cytotoxicity data. This pipeline is based on common sources of laboratory confounding, and is described in more detail in the accompanying Supporting Information. In addition, several variables were used as covariates to correct for potential sources of confounding, including LCL growth rate, temperature during fluorescence intensity measurements, and experimental batch. These are explained in more detail in the Supporting Information. While there was a range in reproducibility and quality across drugs (as expected in such a high throughput experiment). The correlations between replicates for each drug is listed in Supplemental Table 2. Additionally, measures of reproducibility are shown in the Supplemental information. Raw data and quality control filtered data are both included in the supplement for transparency (Supplemental Figures 1-4).

Genotypic quality control

The software program PLINK was used to filter out SNPs whose genotyping rate was below 90%, whose minor allele frequency was below 0.05 or whose p-value from a Hardy-Weinberg test for equilibrium was below 10^{-5} [20]. These filtering steps left, 2,100,684 SNPs available for association analysis. The possibility for substructure in the filtered genetic data was assessed using principal component analysis (PCA). PLINK was used to prune SNPs in high linkage disequilibrium (LD) prior to the principle components analysis [21]. Specifically, a window of 50 SNPs, a step size of 5 SNPs and a pairwise r-squared threshold of 0.7 was used. This resulted in approximately 81% of the remaining SNPs being removed for PCA, leaving a total of 395,033 SNPs available for principal component analysis. PCA was performed on the remaining SNPs using the software package EIGENSTRAT [22]. The LD pruning was used only for the PCA analysis, and was not used as a filter for association analysis. Association analysis was performed on the complete set of 2,100,684 SNPs that passed QC.

Supplemental Figure 5 shows a scatter plot of the first two principal components (PCs), revealing two distinct, non-overlapping main groups, each containing over 230 individuals, with three smaller groups containing between 1 and 9 individuals. The two larger groups may be due to differences in imputation quality between the 314k and 620k chip sets. To confirm this, the PCA was repeated on the un-imputed data (only the markers that were directly genotyped), and the results are shown in Supplemental Figure 6. These results confirm that there is not evidence of population substructure, or strong clusters in the un-imputed data. Because we are using the imputed markers in the association mapping, we used the components from the PCA with the imputed markers for the association analysis (as confounding could occur based on the different imputation quality and allele frequencies). The meaning of the smaller, “outlier” groups is not known, but these outliers were removed from all subsequent association analysis and PCA was run on the remaining individuals. Using a multivariate analysis of covariance (MANCOVA) design, with the vector of mean viabilities at each drug concentration as the response, the first, second and third PCs were found to be significant ($p < 0.05$) in 24, 6 and 1 of the 29 drugs, respectively, so these PCs were used as covariates in association and in clustering analysis. While included too many components can be a concern (in both masking true signals and over-parameterizing the model), the significance test is a relatively standard approach for choosing covariates, and this resulted in a relatively small number of components to include in the model. Additionally, since these components were significant across the vast majority of drugs tested, it is likely that they represent confounding and not true signal.

Association analysis

Association mapping was performed using a multivariate analysis of covariance (MANCOVA) design, with the rationale that modeling the vector of normalized responses jointly provides more information than a single summary measure, such as half-maximal inhibitory concentrations (IC50). Simulation studies have shown this method to be both robust to differences in dose-response profiles between genotypes and powerful in the detection of true biological signals [16, 23]. The model used in association for any drug *d* at and SNP *s* is:

$$\begin{aligned} \boldsymbol{Y}_{ij} &= \mathbf{X}_{ij}\boldsymbol{\beta}+\boldsymbol{\mu}_{\mathbf{i}}+\mathbf{e}_{ij} \\ \mathbf{e}_{ij} &\sim \mathrm{N}_p\left(0, \boldsymbol{\Sigma}\right), \end{aligned} \quad (1)$$

where Y_i is the vector of normalized responses (across the six concentrations for *d*) for the j^{th} j individual having genotype *i* on *s*, X_{ij} is the matrix of covariates for the first two PCs, temperature, growth rate, and experimental batch, and μ_i is the vector of parameters modeling the effects of genotype *i* of *s* on *d*. As pointed out in Choy et al. 2008 [3], confounders such as growth rate need to be accounted for in analysis, and we found in the current study that growth rate is a significant covariate. A summary of the association analysis for the included covariates is found in Supplemental Table 3. Also, $Np\ (0, \Sigma)$ is the multivariate normal distribution, for vectors of length $p = 6$ and with mean 0 and variance Σ. Significance of estimates for μ_i were assessed using Pillai’s trace [24]. Because association tests rely on large sample asymptotic theory, only those loci that had at least 20 individuals in each genotype group were retained for association mapping. This left 1,278,133 SNPs for all drugs except 5-fluorouracil (971,593) and nilotinib (783,013).

Association tests were also performed for each drug family, as described in Table 1. For this, the mean normalized viability across each dose-response curve was calculated for every LCL and every drug. In this way, the model used for association of drug family *d* at an SNP *s* also uses Equation 1, where Y_{ij} now represents the vector of mean normalized viabilities (across all drugs in family *d*) for the j^{th} individual having genotype *i* on *s*. All other variables from Equation 1 remain the same, and p now equals the number of drugs in family *d*.

Traditional p-value cut-offs for genome-wide significant and genome-wide suggestive levels of association were used, motivated by standard practices in the field based on the effective number of markers in the genome [25].

Drug clustering

Distance metrics between each pair of drugs were calculated from their vectors of normalized viabilities, as in the association analysis. Specifically, the distance between drugs *a* and *b* was calculated by first fitting:

$$\boldsymbol{Y}_{ai}=\boldsymbol{Y}_{bi\gamma}+X_i,\beta,$$

where Y_{ai} and Y_{bi} are the vectors of normalized viabilities for the i^{th} LCL for drugs *a* and *b*, respectively, and X_i is the corresponding matrix of covariates used in association mapping (the first two PCs, experimental batch, temperature and growth rate). The coefficient of partial determination (partial r-squared) of Y_{bi} in predicting Y_{ai} after controlling for X_i was calculated for all possible pairs (*a*,*b*). Distance between drugs *a* and *b* was estimated as:

$$d\left(a,b\right)=1-\frac{1}{2}\left[r^2\left(Y_a,Y_b|X\right)+r^2\left(Y_b,Y_a|X\right)\right]$$

where r2 (Ya , Yb |X) (or r2 (Yb , Ya |X)) is the partial r-squared for *a* regressed on *b* (or *b* regressed on *a*), after controlling for covariates *X*. Using this method, it was not possible to include both 5-fluorouracil and nilotinib, since each cell line was exposed to exactly one of these agents. For this reason, and because nilotinib had lower laboratory replicability (see Supplemental Table 2), nilotinib was removed from clustering. The other 28 drugs were clustered using the distance metric described above, using no a priori knowledge of drug family. Clustering was performed using the matrix of distance metrics between all pairs, described above, and the “hclust” function, with the argument “method=ward” from the R statistical package [26]. To test whether this clustering was better than expected by chance, permutation testing was performed, with 10,000 permuted samples created. For the permutation, the tree is fixed and the drug name labels were permuted. Then the probability of getting at least as many clustered drugs as was observed in each category by chance is calculated. Probabilities less than 0.05 were considered statistically significant.

Consistency Analysis with Prior Linkage Results

While a direct validation/replication cohort is not readily available, to evaluate the replication potential and overall stability of the results presented here, the overlap of the association signals (at the suggestive level) from this study with the linkage peak from the study that established the heritability of the drug response was evaluated [12]. The percentage of suggestive variants from the current study that were located in a suggestive linkage peak in the Peters et al. paper was calculated. Details of the linkage analysis can be found in [12]. A SNP was considered to be within a linkage peak if the gene was within the linkage peak as defined as described in the previous paper (using Ensembl to map genes to the chromosomal locations).

Results

Genome-wide SNP associations

Manhattan plots illustrating the association mapping results for each of the 29 drugs are given in Figure 1. In total, 18 SNPs reached a suggestive level of significance at $p < 10^{-6}$ and are summarized in Table 2. Nine of these were located within a gene, and three were within 100kbp of a gene (2 downstream and 1 upstream). In addition, Manhattan plots illustrating the association mapping results for each of the 8 identified drug classes are given in Figure 2. Of the four loci having a suggestive association ($p < 10^{-6}$) with the vector of mean normalized viabilities across a drug class, three occurred within or just upstream of a gene (Table 3).

Drug clustering

Of the 28 different drugs investigated for clustering, 25 can be classified as members of one of eight families, according to their similarity in structure and/or putative mechanism of action (see Table 3). Clustering results are given in Figure 3. Drugs were considered to be clustered together if they either were grouped in the same dendrogram branch (most families), or if two drugs from neighboring branches were closer to each other than either were to any other drug (temozolomide and mitomycin only). Overall, 18 of the 25 drugs (72%) belonging to a family clustered very well according to that family. However, the accuracy of the clustering algorithm differed substantially between drug families. Every tubulin binding agent (either of the vinca alkaloid or taxol class), both tyrosine kinase inhibitors, both platinum agents, both DNA alkylating agents, 4 out of 5 nucleosides and 3 out of 5 anthracyclines clustered together. On the other hand, the fluoropyrimidines (floxuridine and 5-fluorouracil) and the podophyllotoxins (etoposide and teniposide) did not cluster well. The results of the permutation testing reveal that these clustering results are stronger than expected by chance for several of the drug classes (Nucleosides p<0.0002950753; DNA alkylating agents p< 0.04761905; Platinum agents p< 0.04761905; TK inhibitors p< 0.04761905; Anthracyclines p< 0.003958079; Tubulin binding agents 5.087505e-05).

Consistency with Prior Linkage Results

Our results show that 11 out of the 18 (61%) previous signals overlap with linkage peaks from the previous study [12]. This is a substantial number, indicating that a number of these signals are potentially reproducible. Referencing Table 2, the following SNPs were in linkage peaks: 1, 2, 3, 7, 8, 10, 12, 13, 15, 17, 18. Surprisingly, the MGMT SNP was not within a linkage peak, though this may reflect differences in power and assumptions between the linkage and association studies.

Discussion

In the current study, we present the results of a large set of genome-wide association studies performed using the LCL model system. We performed mapping for 29 drugs that had previously shown modest to high heritability, and discovered a number of potential associations. Importantly, the results recapitulated some genetic associations with known clinical relevance (including MGMT associations). This supports the potential of the model system to rapidly identify genes of clinical importance. Of high interest, novel associations were also seen in the results. While the clinical and functional relevance should be investigated in future studies, there are a number of lines of evidence that support the potential biological relevance of these findings. To our knowledge, this is the largest LCL mapping study completed to date, and the overall quality of the data in such a high throughput experiment was as high as previous, smaller studies [6-11].

Across all studies, 18 SNPs were found to have suggestive association ($p < 10^{-6}$) for at least one anticancer agent. Two thirds of these SNPs occurred within or nearby a gene, and many of these genes have demonstrated functional relevance in previous studies. The strongest association was between temozolomide and rs531572 (p < 10–15), located within the gene O-6-methylguanine-DNA methyltransferase (MGMT). A separate publication [10] describes this result in more detail, including MGMT 's known ability to reduce the toxicity of temozolomide, other polymorphisms within MGMT known to impact the clinical efficacy of temozolomide, and a biologically meaningful relationship between rs531572 and MGMT expression in LCLs.

Other associations also had meaningful clinical importance supported in the literature. Locus rs12637988, associated with cytarabine response ($p < 10^{-6}$), is located within the genes mediator complex subunit 12-like (MED12L) and purinergic receptor P2Y, G-protein coupled, 12 (P2RY12). MED12L previously has been associated with toxicity for LCLs from an African cohort exposed to another chemotherapy agent (carboplatin) [6]. Locus rs2270311 is associated with response to 5-fluorouracil ($p < 10^{-6}$), and is located within chimerin 2 (CHN2). Significant differences in the expression of CHN2 have been found between colon cancer cells having different levels of 5-fluorouracil resistance [27]. Relevant functions were less apparent for the other associated genes in Table 1.

The analysis across the drug families (Table 2) showed that locus rs7581424 is associated ($p < 10^{-6}$) with response to the alkylating class (temozolomide and mitomycin), and is upstream of the gene histone deacetylase 4 (HDAC4). Possible interaction with DNA alkylating agents may be especially interesting, as HDAC inhibitors are currently being investigated in stage II clinical trials as agents to be used in combination with temozolomide as therapy for glioblastoma patients [28]. Interestingly, there is no overlap in the genes associated with the drug families and the individual drugs. This may indicate that there are both private and shared signals between/amongst the drugs (though some of this may be due to false positive findings from each either analysis).

These positive findings were obtained from a MANCOVA model. All analysis was performed using the freely available software MAGWAS [16]. This model requires minimal modeling assumptions, unlike traditional methods, which often assume the dose-response curves follow a known form. In addition, this model makes no assumptions regarding the differences in the dose-response curves between genotypes, in contrast to assuming that differences are due to half-inhibitory concentrations or the slope of the non- linear curve. In addition, this multivariate response approach allows for pooling information across dose points. This advantage was exploited with the association across drug families, where information was pooled across drugs within a family. This allows for detection of weaker signals that are not significant for each individually.

Because LCL dose response data may include a number of confounders [3], the choice of appropriate QC techniques and association methods is vital for minimizing erroneous inference and providing unbiased estimates in genetic mapping. Previous simulation studies have demonstrated how a MANCOVA approach has superior power to detect signals, when evaluated across a diverse set of differences in dose response curves between genotypes [17, 24].

In addition to identifying several important genes through association mapping, similarities in LCL cytotoxicity profiles could be used to cluster drugs into appropriate groups based on mechanisms of action, using no a priori knowledge, with high accuracy. This is consistent with similar efforts using LCLs from multigeneration human families, that assessed structure-function relationships within a chemical structural drug class [9] and across a number of clinically used medications [12]. The ease of collecting LCL dose-response data makes this clustering particularly attractive, since the mechanism of action of novel anti-cancer agents may be discovered quickly and cheaply by testing together with other agents that have well-understood biologic effects. Furthermore, since LCL response essentially measures cytotoxicity, the clustering methods described here should be readily available for toxicologic analysis. In this way, the types of toxicities for novel chemicals may be discovered from screening together with other, better characterized chemicals.

As with any study, there are a number of limitations that need to be kept in mind in interpreting these results. While these are exciting new candidates, additional follow-up studies are needed. The large-scale nature of these experiments mean that concerns with multiple testing should be considered. With the large number of drugs tested, it is clear that the study is highly underpowered to consider more conservative cut-offs for significance that account for all drugs tests. To try to address the reproducibility of these findings we showed that there was substantial over-representation with these signals and a previous linkage study, but true replication and validation studies are an important next step. In this way, these results should be viewed as hypothesis generating, as opposed to hypothesis testing. Additionally, the LCL model system, like any model system, is inherently limited in it’s direct clinical relevance. The LCL drug response does not fully reflect the full complexity of the clinical drug response. Excellent discussions of the limitations are reviewed in [5].

Conclusions

Overall, the application of large scale discovery in LCL systems has promise for both discovery of genes of putative mechanistic interest and for assessing distinct features of agents across chemical structure in order to further the quest for rational drug therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH NCI RO1CA161608 from the National Cancer Institute.

References

1. Ritchie MD: The success of pharmacogenomics in moving genetic association studies from bench to bedside: study design and implementation of precision medicine in the post-GWAS era. Human genetics. 2012; 131(10):1615–1626. [PubMed: 22923055]
2. Welsh M, Mangravite L, Medina MW, et al. Pharmacogenomic discovery using cell-based models. Pharmacological reviews. 2009; 61(4):413–429. [PubMed: 20038569]
3. Choy E, Yelensky R, Bonakdar S, et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS genetics. 2008; 4(11):e1000287. [PubMed: 19043577]
4. Stark AL, Zhang W, Mi S, et al. Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. The pharmacogenomics journal. 2010; 10(6):505–512. [PubMed: 20142840]
5. Wheeler HE, Dolan ME. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. Pharmacogenomics. 2012; 13(1):55–70. [PubMed: 22176622]
6. Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. Molecular cancer therapeutics. 2008; 7(9):3038–3046. [PubMed: 18765826]
7. Tan XL, Moyer AM, Fridley BL, et al. Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. Clinical cancer research : an official journal of the American Association for Cancer Research. 2011; 17(17):5801–5811. [PubMed: 21775533]
8. Duan S, Bleibel WK, Huang RS, et al. Mapping genes that contribute to daunorubicin-induced cytotoxicity. Cancer research. 2007; 67(11):5425–5433. [PubMed: 17545624]
9. Watson VG, Motsinger-Reif A, Hardison NE, et al. Identification and replication of loci involved in camptothecin-induced cytotoxicity using CEPH pedigrees. PloS one. 2011; 6(5):e17561. [PubMed: 21573211]
10. Brown CC, Havener TM, Medina MW, et al. A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. Pharmacogenetics and genomics. 2012; 22(11):796–802. [PubMed: 23047291]
11. Li L, Fridley B, Kalari K, et al. Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression. Cancer research. 2008; 68(17):7050–7058. [PubMed: 18757419]
12. Peters EJ, Motsinger-Reif A, Havener TM, et al. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. Pharmacogenomics. 2011; 12(10):1407–1415. [PubMed: 22008047]
13. Ziliak D, O'donnell PH, Im HK, et al. Germline polymorphisms discovered via a cell-based, genome-wide approach predict platinum response in head and neck cancers. Translational research : the journal of laboratory and clinical medicine. 2011; 157(5):265–272. [PubMed: 21497773]
14. Watson VG, Hardison NE, Harris T, Motsinger-Reif A, Mcleod HL. Genomic profiling in CEPH cell lines distinguishes between the camptothecins and indenoisoquinolines. Molecular cancer therapeutics. 2011; 10(10):1839–1845. [PubMed: 21750217]
15. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics. 1990; 6(3):575–577. [PubMed: 2184120]
16. Brown CC, Havener TM, Medina MW, Krauss RM, Mcleod HL, Motsinger-Reif AA. Multivariate methods and software for association mapping in dose-response genome-wide association studies. BioData mining. 2012; 5(1):21. [PubMed: 23234571]

17. Medina MW, Gao F, Ruan W, Rotter JI, Krauss RM. Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin. Circulation. 2008; 118(4):355–362. [PubMed: 18559695]

18. Barber MJ, Mangravite LM, Hyde CL, et al. Genome-wide association of lipid-lowering response to statins in combined study populations. PloS one. 2010; 5(3):e9763. [PubMed: 20339536]

19. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annual review of genomics and human genetics. 2009; 10:387–406.

20. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3):559–575. [PubMed: 17701901]

21. Laurie CC, Doheny KF, Mirel DB, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genetic epidemiology. 2010; 34(6):591–602. [PubMed: 20718045]

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006; 38(8):904–909. [PubMed: 16862161]

23. Brown C, Havener TM, Everitt L, Mcleod H, Motsinger-Reif AA. A comparison of association methods for cytotoxicity mapping in pharmacogenomics. Frontiers in genetics. 2011; 2:86. [PubMed: 22303380]

24. Pillai K. Some new test criteria in multivariate analysis. The Annals of Mathematical Statistics. 1955; 26:4.

25. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human genetics. 2012; 131(5):747–756. [PubMed: 22143225]

26. Team RCD. R: A Language and Environment for Statistical computing. 2011.

27. Schmidt WM, Kalipciyan M, Dornstauder E, et al. Dissecting progressive stages of 5-fluorouracil resistance in vitro using RNA expression profiling. International journal of cancer. Journal international du cancer. 2004; 112(2):200–212. [PubMed: 15352031]

28. Shabason JE, Tofilon PJ, Camphausen K. Grand rounds at the National Institutes of Health: HDAC inhibitors as radiation modifiers, from bench to clinic. Journal of cellular and molecular medicine. 2011; 15(12):2735–2744. [PubMed: 21362133]

Summary Points

- Genome-wide association mapping in LCLs from unrelated Caucasians in 29 anti-cancer agents reveal a number of novel associations.

- Among the single drug studies, suggestive associations were obtained for 18 loci, 12 within/near genes.

- Three of these gene associations, MED12L, CHN2 and MGMT, have been previously implicated in cancer pharmacogenomics, reinforcing the potential of the LCL model to reveal clinically useful genetic associations.

- The drug family association analysis resulted in 4 additional suggestive loci (3 contained within/near genes).

- Cluster analysis was performed to evaluate the potential of the cell line responses to be similar due to similarity in the drug classes.

- The drug clustering results show that 18 of 25 drugs clustered into the appropriate family.

- This study demonstrates the utility of LCLs for identifying genes having clinical importance in drug response, for assigning unclassified agents to specific drug families, and proposes new candidate genes for follow-up in a large number of chemotherapy drugs.

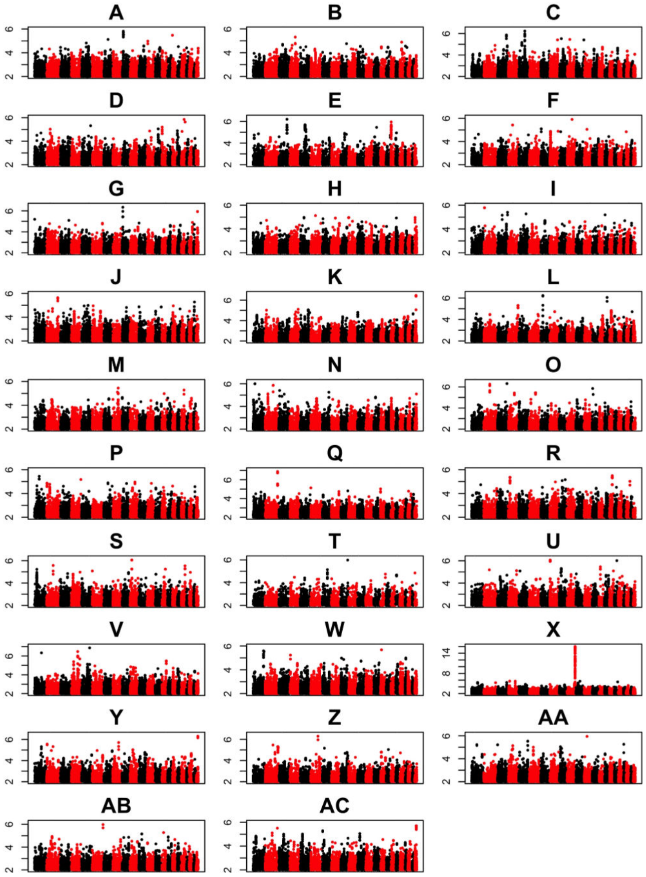

Figure 1.

Manhattan plots for chemotherapy drugs.

SNPs with – log10 (p) > 6 are indicated with red vertical lines, and also summarized in Table 1. Panel A represents arsenic trioxide, Panel B represents azacitidine, Panel C represents carboplatin, Panel D represents cladribine, Panel E represents cytarabine, Panel F represents dasatinib, Panel G represents daunorubicin, Panel H represents docetaxel, Panel I represents doxorubicin, Panel J represents epirubicin, Panel K represents etoposide, Panel L represents 5-fluorouracil, Panel M represents floxuridine, Panel N represents fludarabine, Panel O represents gemcitabine, Panel P represents hydroxyurea, Panel Q rep- resents idarubicin, Panel R represents mitomycin, Panel S represents mitoxantrone, Panel T represents nilotinib, Panel U represents oxaliplatin, Panel V represents paclitaxel, Panel W represents sunitinib, Panel X represents temozolomide, Panel Y represents teniposide, Panel Z represents topotecan, Panel AA represents vinblastine, Panel AB represents vincristine, Panel AC represents vinorelbine.

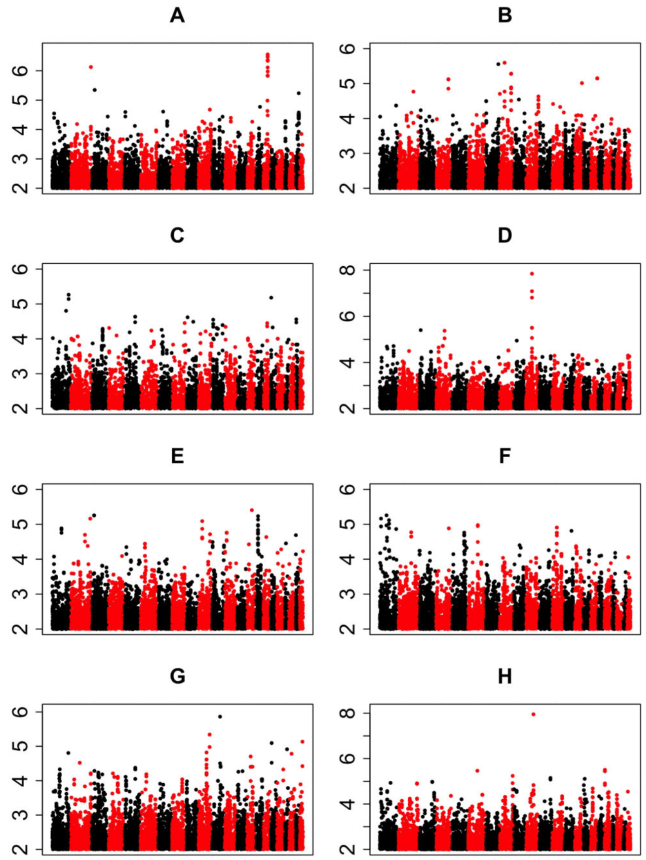

Figure 2.

Manhattan plots for families of drugs.

SNPs with – log10 (p) > 6 are indicated with red ver cal lines, and also summarized in Table 2. Panel A represents DNA alkylating agents, Panel B represents anthracyclines, Panel C represents fluoropyrimidines, Panel D represents platinum agents, Panel E represents nucleosides, Panel F represents epipodophylotoxins, Panel G represents tubulin binding agents, Panel H represents TK inhibitors.

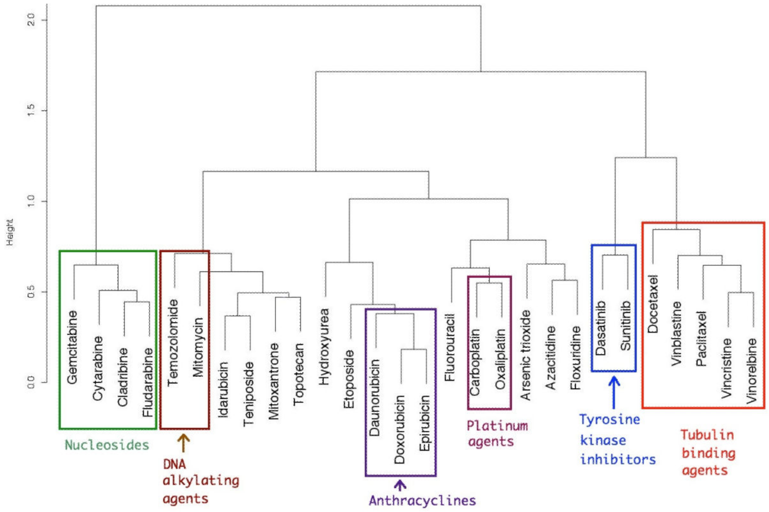

Figure 3.

Drug clustering using partial r-squared for viabilities between drugs, controlling for laboratory covariates. Clustered groups are given in colored rectangles.

Table 1

Drug family membership for 25 anticancer agents.

	Drug Class	**Drugs**				
1	Nucleosides	gemcitabine	cytarabine	cladaribine	fludaribine	azacitidine
2	TK inhibitors	dasatinib	sunitinib			
3	Tubulin binding	docetaxol	pacitaxol	vinblastine	vincristine	vinorelbine
4	DNA alkylating	mitomycin	temozolomide			
5	Platinum agents	carboplatin	oxaliplatin			
6	Anthracyclines	daunorubicin	doxorubicin	epirubicin	idaxubicin	mitoxantrone
7	Fluoropyrimidines	floxuridine	5-fluorouracil			
8	Epipodophylotoxins	etoposide	teniposide			

TK stands for tyrosine kinase.

Table 2

Single nucleotide polymorphisms (SNPs) most associated with drug response, for each drug separately.

	Drug	Chrom.	rsID	$-\log_{10}(p)$	Gene(s) nearby
1	Carboplatin	5	rs1982901	6.23	None
2	Cytarabine	3	rs12637988	6.21	MED12L / P2RY12
3	Daunorubicin	9	rs7867736	6.34	None
4	Etoposide	22	rs2076112	6.47	PLA2G6
5	Fluorouracil	7	rs2270311	6.23	CHN2
6	Fluorouracil	15	rs10152957	6.05	MEGF11
7	Gemcitabine	2	rs4851774	6.26	FHL2
8	Gemcitabine	3	rs513659	6.31	None
9	Idarubicin	2	rs7582313	6.87	None
10	Mitoxantrone	10	rs7068798	6.05	^{d}C10orf67
11	Oxaliplatin	8	rs2897377	6.07	CSMD1
12	Oxaliplatin	17	rs1808918	6.01	^{d}GNA13
13	Paclitaxel	1	rs1338990	6.33	None
14	Paclitaxel	4	rs306005	6.47	SPATA5
15	Paclitaxel	5	rs31878	6.86	None
16	Temozolomide	10	rs531572	15.48	MGMT
17	Teniposide	22	rs8138023	6.30	^{u}NUP50
18	Topotecan	6	rs11966294	6.29	DDO

Superscripts *u* and *d* indicate that genes are located within 100kpb upstream or downstream of the SNP, respectively.

Table 3

Single nucleotide polymorphisms (SNPs) most associated with drug response, for drug family.

	Drug Class	Chrom.	rsID	$-\log_{10}(p)$	Gene(s) nearby
1	DNA Alkylating Agents	2	rs7581424	6.13	^{u}HDAC4
2	DNA Alkylating Agents	16	rs11639947	6.55	NFAT5
3	Platinum Agents	10	rs10821910	7.84	C10orf107
4	TK Inhibitors	10	rs10762827	7.95	None

Superscripts u and d indicate that genes are located within 100kpb upstream or downstream of the SNP, respectively. TK stands for tyrosine kinase.