# A discovery study of daunorubicin induced cardiotoxicity in a sample of acute myeloid leukemia patients prioritizes P450 oxidoreductase polymorphisms as a potential risk factor

**Joanna M. Lubieniecka[1,2]\*, Jinko Graham[2], Daniel Heffner[3], Randy Mottus[1], Ronald Reid[4], Donna Hogge[3,5], Tom A. Grigliatti[1] and Wayne K. Riggs[4]**

[1] Department of Zoology, Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada
[2] Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada
[3] Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[4] Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
[5] Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada

Anthracyclines are very effective chemotherapeutic agents; however, their use is hampered by the treatment-induced cardiotoxicity. Genetic variants that help define patient's sensitivity to anthracyclines will greatly improve the design of optimal chemotherapeutic regimens. However, identification of such variants is hampered by the lack of analytical approaches that address the complex, multi-genic character of anthracycline induced cardiotoxicity (AIC). Here, using a multi-SNP based approach, we examined 60 genes coding for proteins involved in drug metabolism and efflux and identified the *P450 oxidoreductase* (*POR*) gene to be most strongly associated with daunorubicin induced cardiotoxicity in a population of acute myeloid leukemia (AML) patients (FDR adjusted *p*-value of 0.15). In this sample of cancer patients, variation in the *POR* gene is estimated to account for some 11.6% of the variability in the drop of left ventricular ejection fraction (LVEF) after daunorubicin treatment, compared to the estimated 13.2% accounted for by the cumulative dose and ethnicity. In *post-hoc* analysis, this association was driven by 3 SNPs—the rs2868177, rs13240755, and rs4732513—through their linear interaction with cumulative daunorubicin dose. The unadjusted odds ratios (ORs) and confidence intervals (CIs) for rs2868177 and rs13240755 were estimated to be 1.89 (95% CI: 0.7435–4.819; $p = 0.1756$) and 3.18 (95% CI: 1.223–8.27; $p = 0.01376$), respectively. Although the contribution of *POR* variants is expected to be overestimated due to the multiple testing performed in this small pilot study, given that cumulative anthracycline dose is virtually the only factor used clinically to predict the risk of cardiotoxicity, the contribution that genetic analyses of *POR* can make to the assessment of this risk is worthy of follow up in future investigations.

**Keywords: anthracycline induced cardiotoxicity, daunorubicin, acute myeloid leukemia, P450 oxidoreductase variants, complex trait, global test, adverse drug reactions**

## INTRODUCTION

Anthracyclines are very effective chemotherapeutic agents used in the treatment of a wide variety of leukemias, lymphomas, and solid tumors. Due to their high efficacy and broad-spectrum application, these agents are being used despite the associated side effects. The most adverse side effect of anthracycline therapy is cardiotoxicity, which often leads to chronic cardiomyopathy and in some cases congestive heart failure and death (Sawyer et al., 2010). In addition to limiting the dose of the drug administered, several other strategies, including liposomal encapsulation, modification of anthracycline structure, and concurrent administration of iron chelators, have been used as attempts to lessen the cardiotoxic effect of anthracycline treatment (Sawyer et al., 2010). Despite these efforts, chronic cardiomyopathy remains as a life-long threat with the risk for chronic heart failure ranging from 3 to 26% depending on the dose of the drug (Yeh and Bickford, 2009).

The proposed risk factors for anthracycline induced cardiotoxicity (AIC) include cumulative dose, age, gender, pre-existent medical conditions, and combination therapy (Menna et al., 2012). Among these risk factors, the strongest is the cumulative dose of the drug. Consequently, physicians and regulators have set limits on the maximum anthracycline dose that can be used. However, the variability in sensitivity to anthracyclines is quite broad and some patients experience cardiotoxicity with doses as low as 200 mg/m$^2$, while others tolerate doses as high as 1000 mg/m$^2$ (Yeh and Bickford, 2009). These observations suggest that there is no truly "safe" anthracycline dose and

bio-markers are needed to identify patients vulnerable to AIC, as well as those that can tolerate much higher doses of the drugs with little risk, and thus could benefit from more aggressive treatment (Estey, 2012; Menna et al., 2012).

The range of variability among patients in their sensitivity to anthracyclines cannot be fully explained by the known risk factors or pharmacokinetic parameters, and suggests there is an indispensible genetic component (Duan et al., 2007). Genetic contribution to the variability in sensitivity is further supported by the relatively high heritability of anthracycline cytotoxicity ($h^2 = 0.18$–0.63) (Duan et al., 2007; Peters et al., 2011). A number of predisposing genetic variants have been reported; however, these do not explain all of the variability in risk (Wojnowski et al., 2005; Blanco et al., 2008, 2012; Rajic et al., 2009; Visscher et al., 2012). The limited success of genetic studies searching for variants that would explain all of the observed variability in AIC suggests that AIC is a complex phenotype with many genes, environmental factors, and gene-environment interactions contributing to its development (Duan et al., 2007). For such complex traits many genes (even hundreds) contribute to the final phenotype and variants at each of the loci will generally have a modest effect. Therefore, they are difficult to identify, and will often be missed by the standard single-variant based association tests (De Bakker et al., 2005; Kraft and Hunter, 2009; De Los Campos et al., 2010; Yang et al., 2010; Stringer et al., 2011). Moreover, single-variant based association tests ignore the possibility of synergistic and gene-environment interactions, which often results in missing potentially important loci (Ober and Vercelli, 2011; Wheeler et al., 2012). Increasing the size of the study sample will increase the power to detect variants that have modest effects. However, even a large (e.g., 10,000 subjects) study sample will have low power to detect any specific marker with the least stringent significance threshold ($p = 10^{-7}$) in a genome-wide association study of thousands of SNPs (Kraft and Hunter, 2009). Therefore, in realistically sized study samples, the search for genetic variants of complex, polygenic traits requires a multi-variant based approach that allows for potential gene-environment interactions to identify genetic contributors of small or modest effect (De Bakker et al., 2005; Ober and Vercelli, 2011; Wheeler et al., 2012).

The exact causes of AIC are unclear; however, data accumulated to date indicates that anthracycline metabolism, either through generation of free radicals or toxic metabolites, or a combination of both, is a significant contributor to the drugs' cardiotoxicity (Sawyer et al., 2010; Menna et al., 2012). Therefore, we selected 60 genes coding for proteins involved in drug metabolism and efflux as candidates for a pilot project focused on identifying candidate markers of AIC in a sample of acute myeloid leukemia (AML) patients undergoing daunorubicin (DAUN) based treatment. Using a multi-variant based approach that allows for potential gene-anthracycline dose interactions, we identify genetic variants in the *POR* gene as potential markers of AIC among AML patients undergoing DAUN treatment.

## MATERIALS AND METHODS

### PATIENTS AND CLINICAL OUTCOME DEFINITION
Peripheral blood samples were obtained from 286 AML patients after informed consent and with approval from the Clinical

Research Ethics Board of the University of British Columbia. The patient samples were obtained over an 8–10 years period and the samples and patient data were examined retrospectively. All patients received DAUN or mitoxantrone (MITOX) in combination with cytarabine for initial remission induction and subsequent consolidation therapy. For each patient cardiac function was monitored by left ventricular ejection fraction (LVEF) measurements determined with radionuclide ventriculogram (RVG) scans and/or echocardiograms. This patient population has been described in more detail elsewhere (Kim et al., 2012).

### PATIENT SELECTION
The aim of this study was to identify genetic variants that render some patients vulnerable to the cardiotoxic side effect of anthracycline treatment. Since objective measurement of cardiotoxicity due to treatment (and not other causes) can only be done if cardiac function is assessed before and after treatment, only patients that had their LVEF measurements done before and after (less than a year after last dose of anthracycline) anthracycline treatment were used in subsequent statistical analyses (114 patients).

Percentage drop in LVEF was used as a measure of decrease in cardiac function after treatment. LVEF is a measure commonly used to monitor cardiac function in patients undergoing anthracycline treatment and allows for analysis of cardiotoxicity as a quantitative trait (Ewer and Ewer, 2009). A genetic statistical analysis of such traits is typically more powerful than an analysis of binary traits because it allows for discernment of differences in unaffected individuals. In addition, this approach eliminates the often imprecise process of binning individuals into subjective categories (such as cases and controls) (De Los Campos et al., 2010; Stringer et al., 2011).

Furthermore, to keep the treatment uniform across the study sample, patients undergoing MITOX based therapy were also removed, leaving 91 patients for the statistical analyses. A detailed description of this sample population is presented in **Table 1**.

### GENOTYPING
DNA samples from all patients were genotyped for haplotype tagging (ht) and non-synonymous (ns) SNPs in 35 cytochrome p450s (CYPs), 23 oxidoreductases (reductases), and two ABC transporter genes (ABCB1 and ABCG2). The genes were selected based on their documented or possible role in anthracycline metabolism (reductases) or efflux (ABC transporter genes). The CYP genes were also added due to their involvement in redox cycling, which is also implicated in AIC. The gene sequence used for SNP selection included 5 kb of gene flanking region upstream from the start codon and 10kb downstream from the stop codon.

**Table 1 | Characteristics of the AML study sample.**

| Number of patients | | Females | Males | Total |
|---|---|---|---|---|
| | | 43 | 48 | 91 |
| Age | Average | 48.2 | 48.7 | 48.4 |
| | Range | 19–74 | 20–72 | 19–74 |
| DAUN cumulative dose [mg/m²] | Average | 293.1 | 260.4 | 275.9 |
| | Range | 50–630 | 60–540 | 50–630 |

The htSNPs were selected from the HapMap database, release #28 (http://hapmap.ncbi.nlm.nih.gov/) using the Tagger tool and the following cut-offs: linkage disequilibrium (LD) value ($r^2$) = 0.8 and the minor allele frequency (MAF) cutoff = 1% (De Bakker et al., 2005). Using these criteria a total of 921 SNPs were selected to be genotyped. Genotyping was performed using the Sequenom genotyping system (Sequenom Inc., San Diego, CA). For quality control, SNPs with call rate of less than 95%, those with minor allele frequencies less than 1% or not in Hardy-Weinberg equilibrium ($p > 0.001$) were removed from the raw dataset, leaving a total of 465 SNPs for statistical analysis. All 91 patients had call rates of at least 90% across the 465 SNPs.

## STATISTICAL ANALYSIS

All statistical analyses were performed in the R programming environment (R-Development-Core-Team, 2011). Genotypes were coded as 0, 1, and 2 according to the number of copies or "dosage" of the minor allele. Sporadically missing genotypes were imputed with the posterior expected dosage of the minor allele, as estimated by *BEAGLE v3.3.1* (Browning and Browning, 2009).

As expected, the distribution of percent drop in LVEF was highly skewed, with most patients experiencing no or very little change but a few experiencing drops over 10%. To stabilize variance and ensure validity of permutation tests, the percent drop in LVEF was transformed to the logarithmic scale (base 2), after shifting up values by 1.

Potential confounding variables included age, gender, and cumulative DAUN dose. An additional source of potential confounding is population stratification due to ethnic differences. In this study ethnicity was measured by the top three principal components (PCs) of genetic variation (Price et al., 2006). To capture ethnicity as completely as possible, additional genotype data for 195 SNPs in 27 oxidative stress response genes (generated for the patients through a separate study) were added to the genotype set used to define the PCs. To reduce the number of potential confounding variables, forward stepwise selection of these variables in a linear regression model of transformed LVEF drop was applied, based on predictive ability as measured by Akaike's information criterion (Venables and Ripley, 2002). Subsequent genetic analyses were adjusted for the predictive variables only. This base model was used to estimate the proportion of variability in the transformed LVEF drop explained by the potential confounding variables, so that it could be compared to the proportion of variability explained by SNPs in the set of significant candidate genes. Candidate genes explaining a proportion of the variability in LVEF drop that is comparable to variables such as dose or ethnicity that are suspected to influence cardiotoxicity are compelling candidates for further investigation.

To test for association of each SNP, a linear model was fitted to the log-transformed LVEF percent drop that adjusts for dose, the first two genetic PCs, the SNP additive "main effect" and the SNP's linear interaction with dose. To summarize the joint significance of the SNP main effect and interaction, a likelihood ratio statistic was calculated. The exact *p*-value for this test statistic was approximated by 2000 permutation replicates. To control the expected proportion of false discoveries among rejected hypotheses (the false discovery rate or FDR), the nominal *p*-values were corrected for multiple testing of SNPs (Benjamini

and Hochberg, 1995). The Manhattan plot of the nominal *p*-values was done using the *Haploview* program (Barrett, 2009).

Testing genes rather than individual SNPs for association with cardiotoxicity reduces the multiple-testing penalty, enhancing power in this pilot study. Therefore, each of the 60 candidate metabolic genes was tested for association with cardiotoxicity using a global test of its SNP effects and their linear interactions with dose, as implemented in the Bioconductor package *globaltest* (Goeman et al., 2006). This test was chosen because comparative simulation studies of gene-based tests demonstrate that it is well-powered under a variety of alternative hypotheses (Pan, 2009). The global test is motivated by a linear mixed model which adjusts for the fixed effects of confounding variables but specifies random effects for SNPs and their interactions with dose. To control the FDR, the nominal *p*-values were corrected for multiple testing of genes (Benjamini and Hochberg, 1995). All *p*-values for gene-based associations were obtained from permutation tests based on 200,000 permutation replicates.

In *post-hoc* analyses, gene-based associations significant at an FDR level of 20% were further examined in order to understand which SNPs within a gene were driving the association. Specifically, the global test statistics were decomposed into the contributions from each SNP main effect and interaction with dose in the mixed model, using the inheritance procedure as implemented in the *covariates()* function of the *globaltest* package (Goeman and Finos, 2012). For SNPs within a gene, a standard level of 5% was applied to control the family-wise error rate (FWER) and to minimize the number of false-positive SNPs or SNP-by-dose interactions incorrectly identified as drivers of the gene-based association. When decomposing a gene-based association, SNPs were identified as important if they were within an associated gene (at FDR level 20%) and were also significant within the gene (at FWER level 5% in the *post-hoc* analysis), either through their main effect or through their interaction with dose.

To estimate the proportion of variability in LVEF drop explained by the candidate metabolic genes, important SNPs in associated genes were included in an ordinary (fixed effects) regression model with log-transformed LVEF values as the response and terms for dose, ethnicity, SNPs and linear interactions of the SNPs with dose. The proportion of variability in the log-transformed LVEF drop explained by the SNPs in candidate genes was then estimated from the fitted model, without adjusting for multiple testing.

Pairwise LD between significant SNPs and their haplotype phase were estimated using the PLINK package (Purcell et al., 2007).

## RESULTS

Initially 286 patients were genotyped in this blinded, retrospective study. However, subsequent analysis of clinical data identified variability in assessment of cardiac function. Since the outcome under investigation was the cardiotoxic effect of anthracycline treatment and not general cardiac health of the patients, only patients with LVEF measurements done before and after treatment were included in statistical analyses (114 patients). From this sample only those patients treated with the anthracycline DAUN were retained for the statistical analysis (91 patients). Although this patient selection reduced the sample size, it had the

benefit of making the sample much more uniform and therefore less prone to bias from hidden confounding.

To reduce the dimensionality of the data set, we first determined if the known potential confounding variables (DAUN dose, patient age, gender, and ethnicity) contributed to the variance in LVEF drop. Since information on ethnicity was not available we applied PC analysis using 660 SNPs in 89 genes to determine the three major axes of genetic variation. These axes were taken as a summary of the underlying large-scale population structure and used to correct for population stratification due to ethnic differences. Forward selection of the potential confounding variables found that anthracycline cumulative dose and the top two genetic PCs are predictive of LVEF drop. The potential confounding factors, age, and gender, as well as the third genetic PC were not predictive of LVEF drop and were therefore excluded from further analyses. Anthracycline cumulative dose and the top two genetic PCs accounted for an estimated 13.2% of the variation in the log-transformed LVEF drop.

Initially we considered 921 SNPs in 60 metabolic and efflux genes for their effect on % drop in LVEF (Supplemental Table 1). After eliminating SNPs that did not pass QC filters (see Materials and Methods), a total of 465 SNPs remained for tests of association. Tests for the effect of individual SNPs on drop in LVEF identified 25 SNPs at a nominal $p$-value of = 0.05 (**Table 2**). As could be expected, given the small number of patients relative to the number of genetic variants, none of these SNPs survived a correction for multiple comparisons at FDR level of 20%. This suggests that if any of the analyzed SNPs has an effect on LVEF individually, it is probably too small to be detected in the small sample of patients. However, a Manhattan plot of the nominal $p$-values for the single SNP association tests (**Figure 1**) suggests that some genes are "enriched" for SNPs with low $p$-values. This was especially true for the *POR* and *HSD17B2*, where more than three SNPs in each gene were found to be associated with LVEF drop at a nominal $p$-value of less than 0.05. Some of these SNPs, such as the four SNPs in *HSD17B2*, were found to be in strong LD ($r^2 > 0.8$) explaining their concurrent association with LVEF drop (**Table 2**). However, others were in weak LD ($<0.6$) suggesting that their concurrent association with the outcome is due to factors other than linkage and flags the gene as an interesting locus.

Another limitation of the Manhattan plot is that the graphic downplays the gene size, as well as the differences in the number of SNPs between genes. Therefore, we applied a multi-SNP based approach in which each of the 60 genes was tested for association with drop in LVEF using a global test of its SNPs and their linear interactions with dose (Goeman et al., 2006). This test statistic adjusts for gene size, number of SNPs, and LD between the SNPs. The unadjusted $p$-values for the 60 genes are shown in the quantile-quantile plot (**Figure 2A**), in which observed $p$-values are plotted against $p$-values expected under the null hypothesis (no genetic association). The FDR-adjusted $p$-values for the 60 genes are displayed in **Figure 2B**, with the strongest signal detected for the *POR* gene (FDR adjusted $p$-value = 0.15). After accounting for the effects of anthracycline dose and ethnicity, the proportion of variability in the LVEF drop attributable to POR is estimated to be 11.6% in *post-hoc* analyses.
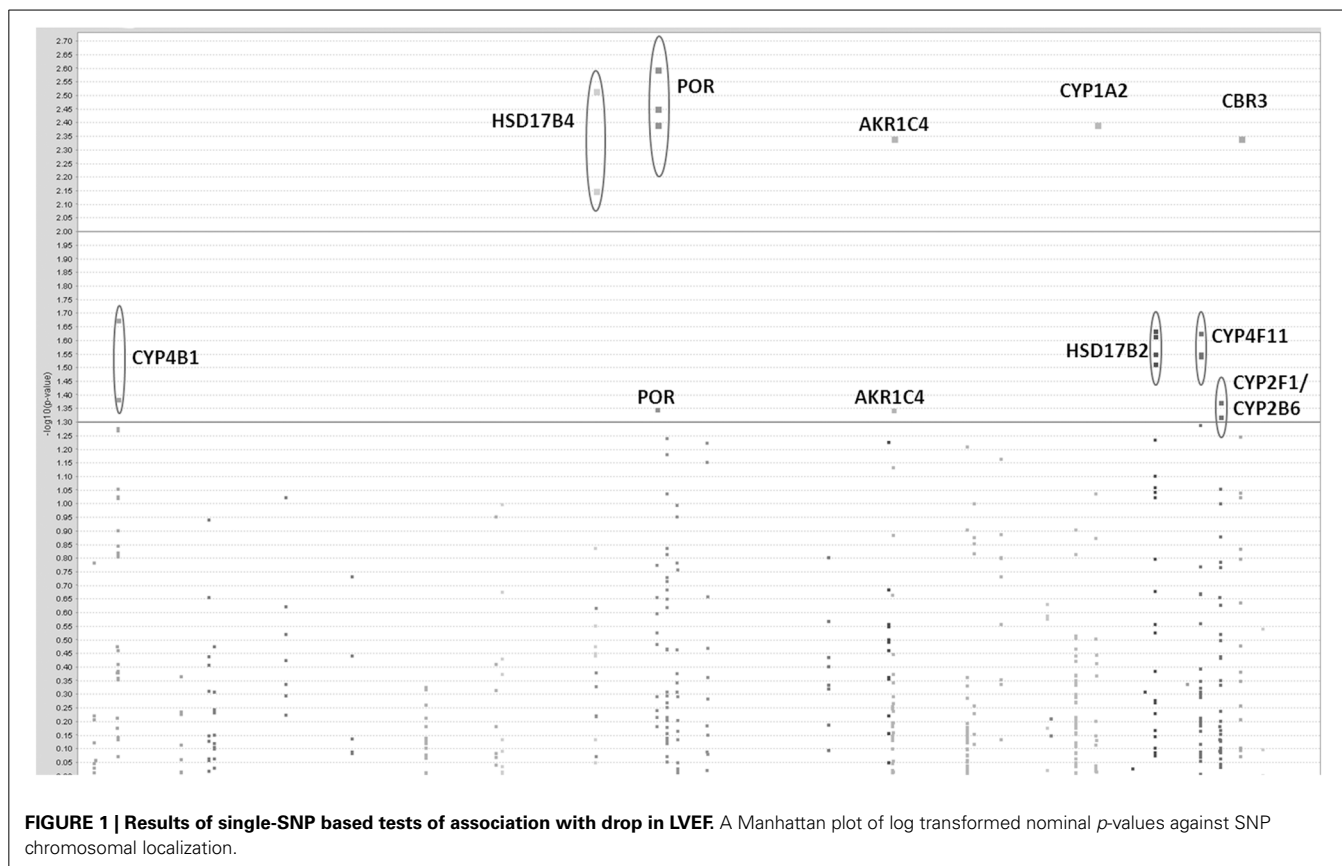
**Table 2 | Top 25 SNP identified through single SNP tests of association.**

| Chr | SNP | $r^2$ | Gene symbol | $p$-value |
|---|---|---|---|---|
| 1 | rs837400 | 0.253 | CYP4B1 | 0.021 |
| 1 | rs4646495 | | CYP4B1 | 0.041 |
| 5 | rs257970 | **0.912** | HSD17B4 | 0.003 |
| 5 | rs2636968 | | HSD17B4 | 0.007 |
| 7 | rs2868177 | 0.524 | POR | 0.0025 |
| 7 | rs13240755 | **0.948** | POR | 0.0035 |
| 7 | rs4732513 | | POR | 0.004 |
| 7 | rs6953065 | 0.258 | POR | 0.0445 |
| 10 | rs7083869 | 0.322 | AKR1C4 | 0.0045 |
| 10 | rs2151896 | | AKR1C4 | 0.045 |
| 15 | rs2069522 | | CYP1A2 | 0.004 |
| 15 | rs2069526 | **1** | CYP1A2 | 0.004 |
| 15 | rs4646427 | | CYP1A2 | 0.004 |
| 16 | rs16956248 | | HSD17B2 | 0.023 |
| 16 | rs13333826 | **1** | HSD17B2 | 0.024 |
| 16 | rs7196087 | | HSD17B2 | 0.028 |
| 16 | rs2955159 | | HSD17B2 | 0.028 |
| 16 | rs2966245 | 0.004 | HSD17B2 | 0.0305 |
| 19 | rs8112732 | 0.405 | CYP4F11 | 0.0235 |
| 19 | rs12610962 | **1** | CYP4F11 | 0.028 |
| 19 | rs2072270 | | CYP4F11 | 0.0285 |
| 19 | rs11086012 | 0.359 | CYP4F11 | 0.051 |
| 19 | rs1709115 | na | CYP2F1 | 0.042 |
| 19 | rs7255904 | na | CYP2B6 | 0.0475 |
| 21 | rs10483032 | na | CBR3 | 0.0045 |

*The p-values have not been corrected for multiple comparisons. $r^2$—squared correlation coefficient, a measure of linkage disequilibrium (LD). For genes with more than 2 SNPs the $r^2$ value refers to LD between the SNP and a reference SNP. $r^2 > 0.8$.*

Next, the overall association of LVEF drop with *POR* was decomposed into the contributions of each SNP main effect and its interaction with dose. Three of the 13 SNPs in *POR*—rs2868177, rs13240755, and rs4732513—were significant drivers of the association at 5% FWER level, through their linear interaction with cumulative DAUN dose. Given that the SNPs rs13240755 and rs4732513 are in strong LD ($r^2 = 0.948$; $D' = 1.000$) only rs13240755 and rs2868177 were included in further analyses.

To estimate the effect of the two *POR* SNPs on the risk of AIC in *post-hoc* analysis, the patient population was split into high–risk (at least 15% drop in LVEF) and low–risk (less than 15% drop in LVEF) groups. The allele frequency distributions at both loci were significantly different between the two risk groups at 5% level (**Figure 3**). The difference in allele frequency was especially pronounced at rs13240755, where the minor allele (G) was almost

**FIGURE 1 | Results of single-SNP based tests of association with drop in LVEF.** A Manhattan plot of log transformed nominal *p*-values against SNP chromosomal localization.

twice as frequent in the high risk (0.60) as compared to low risk group (0.32). For rs2868177, the frequency of the minor allele (G) was also higher in the high risk group (0.50) than it was in low risk group (0.35); however the difference was not as large as for the other SNP. The unadjusted odds ratios (ORs) for rs2868177 and rs13240755 were estimated to be 1.89 (95% CI: 0.7435–4.819; $p = 0.1756$) and 3.18 (95% CI: 1.223–8.27; $p = 0.01376$), respectively.

We caution that these *post-hoc* estimates of effect sizes, variability proportions and ORs do not adjust for the multiple testing and so are descriptive only.

## DISCUSSION AND CONCLUSIONS

The use of anthracyclines at optimal efficacy is hampered by their cumulative dose-dependent adverse side effects, the most serious of which is the AIC that with time can result in anthracycline associated cardiomyopathy (AAC) leading to congestive heart failure. Each anthracycline treatment in addition to killing cancer cells also kills cardiomyocytes, resulting in AIC (De Angelis et al., 2010; Sawyer et al., 2010). Due to the limited regenerative capacity of the heart, cumulative toxicity eventually surpasses a threshold of damage. This triggers a more generic process of ventricular remodeling which leads to AAC (Sawyer et al., 2010). Therefore, careful monitoring and reduction of the risk of AIC during treatment will minimize the threat of developing AAC, which can occur many months or even years later. Biomarkers of the risk of acute AIC could thus greatly aid in optimizing treatment strategies before treatment begins. The relatively high heritability of

sensitivity of normal cells to anthracyclines indicates that the inter-patient variability in risk of AIC should at least partially be explained by common genetic variants (Duan et al., 2007; Peters et al., 2011).
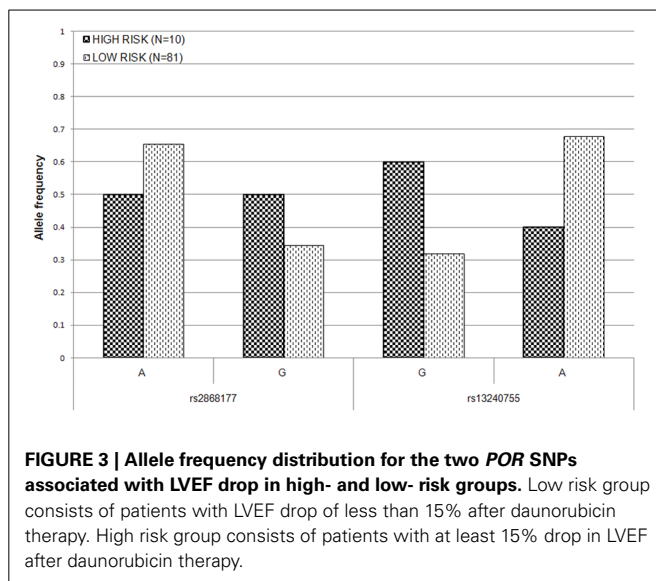
The exact mechanism of AIC is still unclear; however experimental evidence suggests that accumulation of toxic metabolites generated during the drug's metabolism significantly contributes to the observed cardiotoxicity (Menna et al., 2012; Visscher et al., 2012). Therefore, we tested SNPs in 60 genes coding for proteins involved in drug metabolism and efflux (Supplemental Table1) for association with AIC in a sample of AML patients undergoing daunorubicin based chemotherapy.

Using the standard single SNP based association test, none of the analyzed SNPs was found to be significantly associated with AIC. However, given that the cardiotoxicity due to anthracycline treatment is a complex, polygenic phenotype it is expected to be weakly associated with single underlying genetic contributors. Such weak associations are missed by the typical "single-SNP-at-a-time" analysis due to the high cost of adjustment for multiple comparisons that are required to control the false discovery rates. In fact, in exploratory studies of multifactorial phenotypes, correction for multiple comparisons has been criticized as an overly stringent requirement demanding unrealistic sample sizes (Miaskowski et al., 2013). On the other hand, as these corrections help to limit the chance of false positive results, we believe that they should be performed, but with relaxed significance thresholds, particularly for complex traits involving a

**FIGURE 2 | Results of gene-based tests of association with drop in LVEF. (A)** Quantile-quantile plot of nominal *p*-values for the 60 genes included in analysis—the observed *p*-values are plotted against *p*-values expected under the null hypothesis. **(B)** Benjamini-Hochberg false-discovery rates (FDRs) for the 60 analyzed genes. The strongest signal was detected for the POR gene (FDR adjusted *p*-value = 0.15).



**FIGURE 3 | Allele frequency distribution for the two *POR* SNPs associated with LVEF drop in high- and low- risk groups.** Low risk group consists of patients with LVEF drop of less than 15% after daunorubicin therapy. High risk group consists of patients with at least 15% drop in LVEF after daunorubicin therapy.

large number of contributing factors of small effect. Indeed, it has been shown that a set of SNPs with significance value ($P < 0.5$) much below the accepted GWAS threshold ($P < 5 \times 10^{-8}$) was a better predictor of a complex outcome than a single SNP with $P < 5 \times 10^{-8}$, despite the fact that many of the SNPs in the set were probably false positive associations (Purcell et al., 2009). This suggests that both the number of undetected (weak) variants as well as their joint effect on the outcome is substantial (Amin et al., 2009).

To minimize the number of missed associations, we used a multi-SNP analysis to identify genes associated with the percent drop in LVEF due to DAUN treatment. This analysis identified P450 oxidoreductase (*POR, CPR, CYPOR, P450R*) as the gene most strongly associated with DAUN induced LVEF drop (FDR-adjusted *p*-value 15%). The human *POR* gene has 16 exons and exons 2–16 code for the 677 amino acid POR protein (NCBI NP_000932.2). The single copy of the 50 kb *POR* gene is located on chromosome 7 (7q11.23). Interestingly, this chromosome was recently found to harbor a region with significant linkage to daunorubicin cytotoxicity (Peters et al., 2011). The *POR* gene encodes a flavoprotein that transfers electrons from NADPH to various proteins, including the cytochrome P450 enzymes. The anthracyclines, as well as other quinone containing compounds, are converted to semiquinone radical forms by microsomes through a one electron reduction catalyzed by POR (Bachur et al., 1979). This bioactivation step stabilizes cross-linking of the drugs with DNA which is thought to significantly increase their cyto-toxicity (Kostrzewa-Nowak et al., 2005). At the same time, the anthracycline bioactivation step generates free radicals that may further contribute to more widespread anthracycline induced cel-lular damage (Gille and Nohl, 1997). There is also evidence that CYP-mediated oxidative metabolism of anthracyclines may con-tribute to the drug induced cardiotoxicity through generation of oxidative stress and/or disruption of arachidonic acid metabolism (Zordoky et al., 2010).

The *POR* gene is highly polymorphic and a number of *POR* genetic variants were shown to have substrate dependent effects on cytochrome P450 mediated drug metabolism (Gomes et al., 2009; Miller et al., 2011). The three SNPs found here to drive the *POR* association with DAUN induced drop in LVEF are all located in the intronic region of the gene and thus have not been assessed for their effect on POR enzyme function. Interestingly, the SNP rs2868177 has been found to be associ-ated with inter-individual variability in the levels of warfarin maintenance dose, suggesting that it may affect POR mediated metabolism in general (Zhang et al., 2011). The mechanism underlying this association is beyond the scope of this study. We can speculate that these intronic SNPs could be tagging

rare, unidentified mutations that affect the protein sequence. Alternatively, as has been shown for other risk associated SNPs in non-coding regions, they could themselves disturb elements that regulate gene expression or features that modify processing (Jensen et al., 2009; Ott et al., 2009; Cowper-Sal Lari et al., 2012). Whichever the case may be, the association of such SNPs with disease risk is not that unusual. In fact, over 70% of risk associated SNPs in the National Genome Research Institute GWAS do not disturb the protein coding sequence (Abecasis et al., 2010).

In *post-hoc* analyses of our study sample, *POR* is estimated to account for some 11.6% of the variability in the drop in LVEF after DAUN treatment, over and above the estimated 13.2% accounted for by cumulative dose and ethnicity. Given that cumulative anthracycline dose is virtually the only factor used clinically to predict the risk of developing cardiotoxicity as a result of anthracycline treatment, the contribution that genetic analyses of *POR* can make to the assessment of this risk is worthy of follow up. The SNP rs13240755 was the most significant driver of the *POR* association with AIC, with individuals carrying the G allele having increased risk of cardiotoxicity (unadjusted OR = 3.18; 95% CI: 1.223–8.27; $p = 0.01376$). These point and interval estimates of the OR are expected to be biased away from one as a result of the multiple genes queried (Ioannidis, 2008). Thus, validation of these results in an independent study is required. Once validated, the *POR* alleles can be included in a panel with other genetic and non-genetic risk factors, to predict AIC. Given that anthracyclines are used in the treatment of a wide variety of cancers including HER2 positive breast cancer, hematological malignancies, soft tissue sarcomas, and many types of carcinomas (http://www.bccancer.bc.ca/HPI/DrugDatabase/DrugIndexPro/Doxorubicin.htm), such a panel of risk factors should improve the clinical outcome of a substantial number of cancer patients. Additionally, this personalized approach to cancer treatment should reduce the cost of health care by eliminating inappropriate treatments (that is, those cases in which the patient derives no therapeutic benefit) and reducing number of conditions caused by adverse drug reactions.

A limitation of this study is the small sample size. However, unfortunate, this candidly is a reflection of the challenges of pharmacogenomic studies in cancer patient populations undergoing heterogeneous treatments (Wheeler et al., 2012; McLeod, 2013). Even in the case of more common cancers, objective characterization of drug response phenotype is faced with lack of uniform treatments and the tumor molecular heterogeneity. The need for relatively homogeneous patient populations for accurate assessment of the genetic effect on drug response will result in a decrease of the study population size. For very common forms of cancer, relatively large, uniform populations can be potentially recruited through a prospective study; however, for less common cancers such approaches would take many years to complete and are impossible for any single research group. The impracticality of randomized, controlled, prospective trials for clinical implementation and a need for a more efficient approach has been recently discussed in more detail elsewhere (Hattersley and McCarthy, 2005; Wheeler et al., 2012; McLeod, 2013; Miaskowski et al., 2013). We believe that publication of preliminary discoveries, even of marginal statistical significance, can facilitate the translation of potential biomarkers into clinics. Such preliminary candidates can be included by other research groups in validation studies. Furthermore, there may be similar findings discovered at an independent research center that are not being published because of similar strict requirements but in fact represent an independent validation sample, as has been the case with a SNP associated with methotrexate clearance (Trevino et al., 2009; Lopez-Lopez et al., 2011). Although the *POR* SNP association with DAUN induced cardiotoxicity has not been reported by other research groups, the SNP has been found to be associated with levels of warfarin maintenance dose (Zhang et al., 2011), and thus may represent a general marker of drug clearance. This speculation is supported by the essential role of POR for the function of the cytochrome P450 system.

Lastly, our study provides additional support for the recent observations which suggest that identification of a comprehensive list of markers for complex traits requires more than a simple increase in the size of study sample, especially since the required sample size for classic genetic association studies of multifactorial phenotypes is estimated to have unrealistic magnitude (Purcell et al., 2009; Yang et al., 2010; Ober and Vercelli, 2011; Han et al., 2012; Urbach et al., 2012; Wheeler et al., 2012). We strongly support the need for modification of the experimental approach and application of statistical models that take into account the multifactorial (genetic and environmental) character of such traits, including interaction between them. As the cost of generating data plummets there is a continually increasing need for more sophisticated approaches to genome wide analyses. The future of personalized medicine now depends on careful attention to study design, and the development and application of more powerful analytic approaches to the resulting data (Urbach et al., 2012; Wheeler et al., 2012).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2013.00231/abstract

## REFERENCES

Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534

Amin, N., Van Duijn, C. M., and Janssens, A. C. (2009). Genetic scoring analysis: a way forward in genome wide association studies? *Eur. J. Epidemiol.* 24, 585–587. doi: 10.1007/s10654-009-9387-y

Bachur, N. R., Gordon, S. L., Gee, M. V., and Kon, H. (1979). NADPH cytochrome P-450 reductase activation of quinone anticancer agents to free radicals. *Proc. Natl. Acad. Sci. U.S.A.* 76, 954–957. doi: 10.1073/pnas.76.2.954

Barrett, J. C. (2009). Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.* 2009, pdb ip71. doi: 10.1101/pdb.ip71

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B (Methodol.)* 57, 289–300.

Blanco, J. G., Leisenring, W. M., Gonzalez-Covarrubias, V. M., Kawashima, T. I., Davies, S. M., Relling, M. V., et al. (2008). Genetic polymorphisms in the carbonyl reductase 3 gene CBR3 and the NAD(P)H:quinone oxidoreductase 1 gene NQO1 in patients who developed anthracycline-related congestive heart failure after childhood cancer. *Cancer* 112, 2789–2795. doi: 10.1002/cncr.23534

Blanco, J. G., Sun, C. L., Landier, W., Chen, L., Esparza-Duran, D., Leisenring, W., et al. (2012). Anthracycline-related cardiomyopathy after childhood cancer: role of polymorphisms in carbonyl reductase genes–a report from the Children's Oncology Group. *J. Clin. Oncol.* 30, 1415–1421. doi: 10.1200/JCO.2011.34.8987

Browning, B. L., and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. doi: 10.1016/j.ajhg.2009.01.005

Cowper-Sal Lari, R., Zhang, X., Wright, J. B., Bailey, S. D., Cole, M. D., Eeckhoute, J., et al. (2012). Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat. Genet.* 44, 1191–1198. doi: 10.1038/ng.2416

De Angelis, A., Piegari, E., Cappetta, D., Marino, L., Filippelli, A., Berrino, L., et al. (2010). Anthracycline cardiomyopathy is mediated by depletion of the cardiac stem cell pool and is rescued by restoration of progenitor cell function. *Circulation* 121, 276–292. doi: 10.1161/CIRCULATIONAHA.109.895771

De Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nat. Genet.* 37, 1217–1223. doi: 10.1038/ng1669

De Los Campos, G., Gianola, D., and Allison, D. B. (2010). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* 11, 880–886. doi: 10.1038/nrg2898

Duan, S., Bleibel, W. K., Huang, R. S., Shukla, S. J., Wu, X., Badner, J. A., et al. (2007). Mapping genes that contribute to daunorubicin-induced cytotoxicity. *Cancer Res.* 67, 5425–5433. doi: 10.1158/0008-5472.CAN-06-4431

Estey, E. H. (2012). Acute myeloid leukemia: 2012 update on diagnosis, risk stratification, and management. *Am. J. Hematol.* 87, 89–99. doi: 10.1002/ajh.22246

Ewer, S. M., and Ewer, M. S. (2009). Anthracycline cardiotoxicity: why are we still interested? *Oncology (Williston Park)* 23, 234–235, 239.

Gille, L., and Nohl, H. (1997). Analyses of the molecular mechanism of adriamycin-induced cardiotoxicity. *Free Radic. Biol. Med.* 23, 775–782. doi: 10.1016/S0891-5849(97)00025-7

Goeman, J. J., and Finos, L. (2012). The inheritance procedure: multiple testing of tree-structured hypotheses. *Stat. Appl. Genet. Mol. Biol.* 11:11. doi: 10.1515/1544-6115.1554

Goeman, J. J., Van De Geer, S. A., and Van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soci. Series B (Stat. Methodol.)* 68, 477–493. doi: 10.1111/j.1467-9868.2006.00551.x

Gomes, A. M., Winter, S., Klein, K., Turpeinen, M., Schaeffeler, E., Schwab, M., et al. (2009). Pharmacogenomics of human liver cytochrome P450 oxidoreductase: multifactorial analysis and impact on microsomal drug oxidation. *Pharmacogenomics* 10, 579–599. doi: 10.2217/pgs.09.7

Han, X., Li, Y., Huang, J., Zhang, Y., Holford, T., Lan, Q., et al. (2012). Identification of predictive pathways for non-hodgkin lymphoma prognosis. *Cancer Inform.* 9, 281–292. doi: 10.4137/CIN.S6315

Hattersley, A. T., and McCarthy, M. I. (2005). What makes a good genetic association study? *Lancet* 366, 1315–1323. doi: 10.1016/S0140-6736(05)67531-9

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology* 19, 640–648. doi: 10.1097/EDE.0b013e31818131e7

Jensen, C. J., Oldfield, B. J., and Rubio, J. P. (2009). Splicing, cis genetic variation and disease. *Biochem. Soc. Trans.* 37, 1311–1315. doi: 10.1042/BST0371311

Kim, H. P., Bernard, L., Berkowitz, J., Nitta, J., and Hogge, D. E. (2012). Flow cytometry-based assessment of mitoxantrone efflux from leukemic blasts varies with response to induction chemotherapy in acute myeloid leukemia. *Cytometry B. Clin. Cytom.* 82, 283–294. doi: 10.1002/cyto.b.21028

Kostrzewa-Nowak, D., Paine, M. J., Wolf, C. R., and Tarasiuk, J. (2005). The role of bioreductive activation of doxorubicin in cytotoxic activity against leukaemia HL60-sensitive cell line and its multidrug-resistant sublines. *Br. J. Cancer* 93, 89–97. doi: 10.1038/sj.bjc.6602639

Kraft, P., and Hunter, D. J. (2009). Genetic risk prediction–are we there yet? *N. Engl. J. Med.* 360, 1701–1703. doi: 10.1056/NEJMp0810107

Lopez-Lopez, E., Martin-Guerrero, I., Ballesteros, J., Pinan, M. A., Garcia-Miguel, P., Navajas, A., et al. (2011). Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia. *Pediatr. Blood Cancer* 57, 612–619. doi: 10.1002/pbc.23074

McLeod, H. L. (2013). Cancer pharmacogenomics: early promise, but concerted effort needed. *Science* 339, 1563–1566. doi: 10.1126/science.1234139

Menna, P., Paz, O. G., Chello, M., Covino, E., Salvatorelli, E., and Minotti, G. (2012). Anthracycline cardiotoxicity. *Expert Opin. Drug Saf.* 11(Suppl. 1), S21–S36. doi: 10.1517/14740338.2011.589834

Miaskowski, C., Dodd, M., Paul, S. M., West, C., Hamolsky, D., Abrams, G., et al. (2013). Lymphatic and angiogenic candidate genes predict the development of secondary lymphedema following breast cancer surgery. *PLoS ONE* 8:e60164. doi: 10.1371/journal.pone.0060164

Miller, W. L., Agrawal, V., Sandee, D., Tee, M. K., Huang, N., Choi, J. H., et al. (2011). Consequences of POR mutations and polymorphisms. *Mol. Cell. Endocrinol.* 336, 174–179. doi: 10.1016/j.mce.2010.10.022

Ober, C., and Vercelli, D. (2011). Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet.* 27, 107–115. doi: 10.1016/j.tig.2010.12.004

Ott, C. J., Blackledge, N. P., Leir, S. H., and Harris, A. (2009). Novel regulatory mechanisms for the CFTR gene. *Biochem. Soc. Trans.* 37, 843–848. doi: 10.1042/BST0370843

Pan, W. (2009). Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet. Epidemiol.* 33, 497–507. doi: 10.1002/gepi.20402

Peters, E. J., Motsinger-Reif, A., Havener, T. M., Everitt, L., Hardison, N. E., Watson, V. G., et al. (2011). Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics* 12, 1407–1415. doi: 10.2217/pgs.11.92

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752. doi: 10.1038/nature08185

Rajic, V., Aplenc, R., Debeljak, M., Prestor, V. V., Karas-Kuzelicki, N., Mlinaric-Rascan, I., et al. (2009). Influence of the polymorphism in candidate genes on late cardiac damage in patients treated due to acute leukemia in childhood. *Leuk. Lymphoma* 50, 1693–1698. doi: 10.1080/10428190903177212

R-Development-Core-Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna: F.F.S. Computing.

Sawyer, D. B., Peng, X., Chen, B., Pentassuglia, L., and Lim, C. C. (2010). Mechanisms of anthracycline cardiac injury: can we identify strategies for cardioprotection? *Prog. Cardiovasc. Dis.* 53, 105–113. doi: 10.1016/j.pcad.2010.06.007

Stringer, S., Wray, N. R., Kahn, R. S., and Derks, E. M. (2011). Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS ONE* 6:e27964. doi: 10.1371/journal.pone.0027964

Trevino, L. R., Shimasaki, N., Yang, W., Panetta, J. C., Cheng, C., Pei, D., et al. (2009). Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J. Clin. Oncol.* 27, 5972–5978. doi: 10.1200/JCO.2008.20.4156

Urbach, D., Lupien, M., Karagas, M. R., and Moore, J. H. (2012). Cancer heterogeneity: origins and implications for genetic association studies. *Trends Genet.* 28, 538–543. doi: 10.1016/j.tig.2012.07.001

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S.* 4th Edn. New York, NY: Springer.

Visscher, H., Ross, C. J., Rassekh, S. R., Barhdadi, A., Dube, M. P., Al-Saloos, H., et al. (2012). Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J. Clin. Oncol.* 30, 1422–1428. doi: 10.1200/JCO.2010.34.3467

Wheeler, H. E., Maitland, M. L., Dolan, M. E., Cox, N. J., and Ratain, M. J. (2012). Cancer pharmacogenomics: strategies and challenges. *Nat. Rev. Genet.* 14, 23–34. doi: 10.1038/nrg3352

Wojnowski, L., Kulle, B., Schirmer, M., Schluter, G., Schmidt, A., Rosenberger, A., et al. (2005). NAD(P)H oxidase and multidrug resistance protein genetic polymorphisms are associated with doxorubicin-induced cardiotoxicity. *Circulation* 112, 3754–3762. doi: 10.1161/CIRCULATIONAHA.105.576850

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi: 10.1038/ng.608

Yeh, E. T., and Bickford, C. L. (2009). Cardiovascular complications of cancer therapy: incidence, pathogenesis, diagnosis, and management. *J. Am. Coll. Cardiol.* 53, 2231–2247. doi: 10.1016/j.jacc.2009.02.050

Zhang, X., Li, L., Ding, X., and Kaminsky, L. S. (2011). Identification of cytochrome P450 oxidoreductase gene variants that are significantly associated with the interindividual variations in warfarin maintenance dose. *Drug Metab. Dispos.* 39, 1433–1439. doi: 10.1124/dmd.111.038836

Zordoky, B. N., Anwar-Mohamed, A., Aboutabl, M. E., and El-Kadi, A. O. (2010). Acute doxorubicin cardiotoxicity alters cardiac cytochrome P450 expression and arachidonic acid metabolism in rats. *Toxicol. Appl. Pharmacol.* 242, 38–46. doi: 10.1016/j.taap.2009.09.012