

RESEARCH

Open Access

Copy number polymorphisms and anticancer pharmacogenomics

Eric R Gamazon¹, R Stephanie Huang², M Eileen Dolan^{2*} and Nancy J Cox^{1,3*}

Abstract

Background: Recent studies have investigated the contribution of copy number variants (CNVs) to disease susceptibility in a multitude of complex disorders, including systemic lupus erythematosus, Crohn's disease, and various neurodevelopmental disorders. Relatively few CNV studies, however, have been conducted on pharmacologic phenotypes even though these structural variants are likely to play an important role. We developed a genome-wide method to identify CNVs that contribute to heterogeneity in drug response, focusing on drugs that are widely used in anticancer treatment regimens.

Results: We conducted a comprehensive genome-wide study of CNVs from population-scale array-based and sequencing-based surveys by analyzing their effect on cellular sensitivity to platinating agents and topoisomerase II inhibitors. We identified extensive CNV regions associated with cellular sensitivity to functionally diverse chemotherapeutics, supporting the hypothesis that variation in copy number contributes to variation in drug response. Interestingly, although single nucleotide polymorphisms (SNPs) tag some of the CNVs associated with drug sensitivity, several of the most significant CNV-drug associations are independent of SNPs; consequently, they represent genetic variations that have not been previously interrogated by SNP studies of pharmacologic phenotypes.

Conclusions: Our findings demonstrate that pharmacogenomic studies may greatly benefit from the study of CNVs as expression quantitative trait loci, thus contributing broadly to our understanding of the complex traits genetics of CNVs. We also extend our PACdb resource, a database that makes available to the scientific community relationships between genetic variation, gene expression, and sensitivity to various drugs in cell-based models.

Background

Copy number variants (CNVs) have received considerable attention in recent years as studies have implicated them in a wide range of complex human phenotypes, including susceptibility to HIV-1/AIDS [1], Crohn's disease [2], and various autoimmune disorders. The systematic assessment of their role in the etiology of complex disease has been predicated on improvements in genotyping technologies (including SNP-based genotyping arrays and clone-based comparative genomic hybridization) and on advances in algorithms for copy number analysis [3]. Genome-wide surveys of CNVs

[4,5] have sought to produce a comprehensive map to enable disease association studies, but a recent comprehensive study reports a somewhat disappointing finding that CNVs are likely to make a relatively minor contribution to the genetic basis of complex traits [6], particularly disease susceptibility.

While the study of the contribution of CNVs to drug response has lagged behind the investigation of their contribution to disease risk, there have been some notable findings coming out of candidate gene approaches. The gene *CYP2D6* encodes an enzyme to which the metabolism of a large number of drugs, such as antidepressants, neuroleptics, analgetics and anticancer drugs, is attributed. It has been demonstrated that *CYP2D6* may occur in CNVs of 0 to 13 copies [7]. Studies have shown that copy number for this gene affects the plasma levels of the active metabolite of tamoxifen, namely endoxifen, so that ultra-rapid metabolizers who carry

* Correspondence: edolan@medicine.bsd.uchicago.edu; ncox@medicine.bsd.uchicago.edu

¹Section of Genetic Medicine, Department of Medicine, University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA

²Section of Hematology/Oncology, Department of Medicine, University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA

Full list of author information is available at the end of the article

more than two copies of the gene show much higher levels of endoxifen than those who carry the regular copy number for the gene [8]. Higher *CYP2D6* activity due to gene amplification has also been shown to predispose to life-threatening opioid intoxication [9]. Another drug metabolizing cytochrome P450 gene, *CYP2A6*, also occurs in variable copy number. *CYP2A6* encodes an enzyme that metabolizes several drugs, including nicotine and its metabolite cotinine. Increased *CYP2A6* activity has been shown to be responsible for increased risk for nicotine addiction [10] and for tobacco-related cancers. The *SULT* family of Phase II conjugating enzymes, particularly that encoded by *SULT1A1*, has been the subject of extensive pharmacogenetic studies that show the importance of CNVs as a genetic source of variability in the metabolic activity of these enzymes. *SULT* pharmacogenomic studies [11] have highlighted CNV-based mechanisms that lead to increased risk for chemical carcinogenesis and adverse drug reactions. Glutathione S-transferase (*GST*), also a phase II family of conjugation enzymes, plays an important role in the detoxification of drugs. Studies have shown that homozygous deletion of *GSTM1* is correlated with increased cancer risk and with better treatment outcome [12,13]. These findings and related developments highlight the necessity of incorporating copy number analysis in elucidating the genetic underpinnings of drug response.

The recently released catalog [4] from an extensive survey of copy number regions assayed in cell lines from the International HapMap project and the subsequent study of genomic structural variants based on whole genome DNA sequencing data (the 1000 Genomes Project) [14] allow for new pharmacogenomic discoveries and for deep insights into the genetic basis of pharmacologic phenotypes, which to date has largely been based on studies of SNPs [15]. In whole-genome studies using lymphoblastoid cell lines (LCLs), cellular sensitivity to drug [16] as well as gene expression phenotypes [17] have been shown to be heritable [18] and to include a significant genetic component. Although many CNV pharmacogenetic studies have focused on pharmacokinetic genes, we chose to evaluate pharmacodynamic genes using an LCL-based model. Studies in our laboratory have generated a rich resource of pharmacologic data [19] on a wide array of chemotherapeutic agents using the HapMap cell lines, enabling us to conduct a systematic analysis of the role of CNVs for a variety of anticancer drugs.

Results

Genome-wide association studies

LCLs from unrelated CEU samples were phenotyped for cellular sensitivity to the four chemotherapeutic drugs

included in our study: carboplatin [20], cisplatin [21], daunorubicin [22], and etoposide [16]. We conducted genome-wide association scans using drug IC_{50} as a quantitative trait.

A total of 5,238 CNVs from an array-based study [4] were evaluated in genome-wide association studies (GWAS) against cellular sensitivity drug phenotypes. Of these CNVs, 77% are deletions (0, 1, or 2 copy number), 16% are amplifications (2, 3, or 4 copy number), and the remainder are multi-allelic (greater than 3 diploid copy number genotypes) [4]. At the nominally significant threshold of $P < 0.05$, we identified 67 CNVs associated with carboplatin IC_{50} , 70 CNVs with cisplatin IC_{50} , 73 CNVs with daunorubicin IC_{50} , and 113 CNVs with etoposide IC_{50} .

Genomic characterization of drug susceptibility-associated CNVs

We further evaluated the genomic characteristics of these drug susceptibility-associated CNVs for their size and type (deletion versus amplification). In general, there is little (Pearson) correlation between the size of a CNV and its association with cellular sensitivity to carboplatin ($r = 0.020$), cisplatin ($r = 0.008$), daunorubicin ($r = 0.054$) and etoposide ($r = 0.024$). We did, however, observe that the top CNVs associated ($P < 0.05$) with IC_{50} for daunorubicin are significantly smaller (average size of 10.6 kb) than expected (average size of 14 kb) from the full set of CNVs included in our study; etoposide-associated CNVs are, in contrast, close to expectation (average size of 13.4 kb). The CNVs associated with carboplatin and cisplatin IC_{50} (average size of 11.2 kb and 11.4 kb, respectively) are significantly smaller than expected.

Sixty-two of the 67 carboplatin-associated CNVs ($P < 0.05$) are biallelic (that is, deletions or amplifications); the remaining five CNVs are multi-allelic CNVs (that is, defined as having more than three CNV genotypes). Deletions are significantly more frequent (85%) than duplications among the carboplatin-associated CNVs. Similarly, all but 4 of the 70 cisplatin-associated CNVs ($P < 0.05$) are biallelic. The top cisplatin-associated CNVs are significantly more likely to be deletions (87%) than duplications.

All but 9 etoposide-associated CNVs ($P < 0.05$) are biallelic; 69 out of the 73 daunorubicin-associated CNVs ($P < 0.05$) are biallelic. Nearly 82% of the daunorubicin-associated CNVs and 82% of the etoposide-associated CNVs are deletions.

Drug susceptibility-associated CNVs are enriched for expression quantitative trait loci

We observed that no exons overlap the genomic regions defined by the top associated CNVs for each anticancer

drug included in our study, suggesting that these CNVs do not act to disrupt coding sequence. We thus hypothesized that these CNVs act to influence gene regulation. We evaluated the functional import of the drug susceptibility-associated CNVs by considering their effect on global gene expression. At an expression association threshold of $P < 0.0001$, 60% ($N = 40$) of the top CNVs associated with carboplatin ($P < 0.05$) were found to be expression quantitative trait loci (eQTLs). Interestingly, two of the top carboplatin-associated CNVs (*CNVR3882.1* on chromosome 8 and *CNVR666.1* on chromosome 2) predict the expression of *SELL*. We found that *SELL* expression level is also significantly correlated with carboplatin IC₅₀ ($P = 0.02$) in the CEU samples. We identified several target genes of carboplatin-associated CNVs (as eQTLs) whose expression levels were significantly correlated (after multiple testing correction [23], false discovery rate (FDR) < 0.05) with carboplatin IC₅₀, including *PHGDH*, *MYO1B*, *TGFBR2*, and *PRF1*. Similarly, nearly 56% ($N = 39$) of the cisplatin-associated CNVs ($P < 0.05$) were associated with the transcript level of genes as eQTLs. We found a target gene, *MAST4*, for two cisplatin-associated CNVs (*CNVR2968.1* on chromosome 6 and *CNVR7881.1* on chromosome 20). *MAST4* trends toward significance ($P = 0.06$) with cisplatin IC₅₀ in the CEU samples. A target gene (*C4A* at $P = 8.2 \times 10^{-6}$) for a cisplatin-associated CNV eQTL (*CNVR4748.1* on chromosome 10) was found to be significantly correlated (after multiple testing correction [23], FDR < 0.05) with cisplatin IC₅₀.

Restricting our analysis to biallelic CNVs, we found, through simulations, that the top CNVs, for each platinating agent, are significantly enriched for eQTLs relative to frequency-matched SNPs (enrichment $P < 0.05$). The eQTL enrichment holds at a lower P -value threshold ($P = 10^{-6}$) used to define an eQTL, showing the robustness of our observation to the definition of eQTL. See Materials and methods for details on the simulation procedure.

Of the top CNVs associated with etoposide IC₅₀ ($P < 0.05$), 76% ($N = 86$) were found to be eQTLs. Of these CNV eQTLs, eight share *UBAI* as a target gene ($P < 9.7 \times 10^{-5}$). Two target genes (*PLEKHG6* at $P = 1.3 \times 10^{-6}$ and *WSB2* at $P = 8.0 \times 10^{-5}$) for etoposide-associated CNV eQTLs (*CNVR4784.1* on chromosome 10 and *CNVR1874.1* on chromosome 4, respectively) were found to be significantly correlated with etoposide IC₅₀ (FDR < 0.01) [23]. Nearly 52% ($N = 38$ of the top CNVs associated with daunorubicin IC₅₀ ($P < 0.05$) were eQTLs. We identified two daunorubicin-associated CNVs (*CNVR479.2* and *CNVR332.1* on chromosome 1) predicting the expression of *HIST1H4A* ($P < 6.7 \times 10^{-5}$); we also found the expression level of this gene to be correlated ($P = 3.7 \times 10^{-8}$) with daunorubicin IC₅₀ in

the CEU samples. We identified several target genes (including *PAPLN* at $P = 3.3 \times 10^{-5}$ and *KLF12* at $P = 6.1 \times 10^{-5}$) for daunorubicin-associated CNV eQTLs (*CNVR2616.1* on chromosome 5 and *CNVR948.1* on chromosome 2, respectively) whose expression levels were significantly correlated (after multiple testing correction [23], FDR < 0.05) with daunorubicin IC₅₀.

As in the case of the platinating agents, we found, through simulations, that the top CNVs for each topoisomerase II inhibitor are more likely to be eQTLs than frequency-matched SNPs (enrichment $P < 0.05$).

Functional characterization of transcripts *cis*-regulated by deletions from whole genome sequencing data

Given the observed high proportion of deletions among CNVs associated with cellular sensitivity to chemotherapeutic agents, we sought additional functional support for the role of CNVs as transcriptional regulators from whole genome sequencing data coming out of the 1000 Genomes project, which characterized the CNV deletions with Gencode/ENCODE transcripts [14]. The resulting enlarged catalog of CNVs (with an initial focus on deletions) included CNVs of size 50 bp or larger mapped at single nucleotide resolution. We identified 376 transcripts to which CNV deletions were annotated [14] (by Gencode/ENCODE) as influencing (*cis*-regulating) transcription and/or translation. We proceeded to test the 376 transcripts for their role in predicting cellular sensitivity to chemotherapeutics. At $P < 0.05$, we found 21 transcript correlations with carboplatin, 15 with cisplatin, 23 with daunorubicin, and 21 with etoposide (see Table 1). Three transcripts (*MOXD1*, *RCCI1*, *SULF2*) were significant after multiple testing adjustment ($p_{adj} < 0.05$, Bonferroni). Remarkably, the three transcripts were the only CNV deletions associated with all four agents at the nominal $P < 0.05$ threshold (Figure 1).

Drug susceptibility-associated CNVs are independent of drug susceptibility-associated SNPs

We investigated to what extent the CNVs associated with cellular sensitivity to chemotherapeutic agents may already be interrogated by SNP-based GWAS through linkage disequilibrium [6]. We found that the top CNV (*CNVR1616.1*) associated with carboplatin IC₅₀ ($P = 5 \times 10^{-4}$) is not well-tagged by SNPs. Indeed, the best proxy SNP for this CNV on chromosome 3 is *rs967422* (at $r^2 = 0.075$). We found that the same CNV is also associated with cisplatin IC₅₀ ($P = 6.5 \times 10^{-3}$). Another cisplatin-associated CNV ($P = 5.5 \times 10^{-3}$), *CNVR7870.1*, is also not well-tagged; the best proxy SNP, *rs915049*, tags the CNV at a low $r^2 = 0.11$. In each case, the best proxy SNP, in contrast to the 'tagged' drug susceptibility-associated CNV, shows no evidence of being associated with

Table 1 Nominally significant ($P < 0.05$) gene expression correlations with cellular sensitivity to chemotherapeutic agents

Gene	Drug	P-value	Chromosome
<i>AKT1S1</i>	Carboplatin	0.04497	19
<i>AKT1S1</i>	Daunorubicin	0.01149	19
<i>AKT1S1</i>	Etoposide	0.00610	19
<i>ALKBH8</i>	Daunorubicin	0.01151	11
<i>AMY1A</i>	Etoposide	0.03933	1
<i>AMY2A</i>	Etoposide	0.03933	1
<i>ANKLE1</i>	Carboplatin	0.03017	19
<i>ANKLE1</i>	Daunorubicin	0.01706	19
<i>ANKRD36B</i>	Daunorubicin	0.02380	2
<i>BCLAF1</i>	Daunorubicin	0.01498	6
<i>C13orf3</i>	Carboplatin	0.00830	13
<i>C13orf3</i>	Cisplatin	0.01980	13
<i>C13orf3</i>	Daunorubicin	0.00572	13
<i>C18orf1</i>	Cisplatin	0.00398	18
<i>CAB39L</i>	Etoposide	0.04179	13
<i>DIS3L2</i>	Daunorubicin	0.03077	2
<i>DNAJC5</i>	Carboplatin	0.00166	20
<i>DNAJC5</i>	Cisplatin	0.04058	20
<i>FBRS</i>	Carboplatin	0.00661	16
<i>FGFR4</i>	Carboplatin	0.00556	5
<i>FKBP14</i>	Carboplatin	0.03912	7
<i>FLI1</i>	Carboplatin	0.04735	11
<i>FLI1</i>	Daunorubicin	0.00314	11
<i>FLI1</i>	Etoposide	0.04230	11
<i>GALNT1</i>	Daunorubicin	0.00065	18
<i>GPR107</i>	Etoposide	0.00674	9
<i>GPR137</i>	Carboplatin	0.00616	11
<i>GPR137</i>	Cisplatin	0.00828	11
<i>GPR137</i>	Etoposide	0.01493	11
<i>GPR144</i>	Etoposide	0.04526	9
<i>GSPT1</i>	Carboplatin	0.04002	16
<i>GSPT1</i>	Cisplatin	0.01429	16
<i>HLA-DQA1</i>	Daunorubicin	0.00946	6
<i>HLA-DQA1</i>	Etoposide	0.00022	6
<i>IGLV3-21</i>	Carboplatin	0.04800	22
<i>MOXD1</i>	Carboplatin	0.00002	6
<i>MOXD1</i>	Cisplatin	0.00002	6
<i>MOXD1</i>	Daunorubicin	0.03454	6
<i>MOXD1</i>	Etoposide	0.00007	6
<i>MTA1</i>	Etoposide	0.01401	14
<i>NCOA1</i>	Daunorubicin	0.04278	2
<i>NEK6</i>	Carboplatin	0.04050	9
<i>NUB1</i>	Etoposide	0.02671	7
<i>PPP1R3B</i>	Cisplatin	0.03039	8
<i>PTP4A2</i>	Carboplatin	0.01930	1
<i>PTP4A2</i>	Cisplatin	0.01182	1
<i>PTP4A2</i>	Etoposide	0.03348	1
<i>RCC1</i>	Carboplatin	0.01252	1
<i>RCC1</i>	Cisplatin	0.01722	1

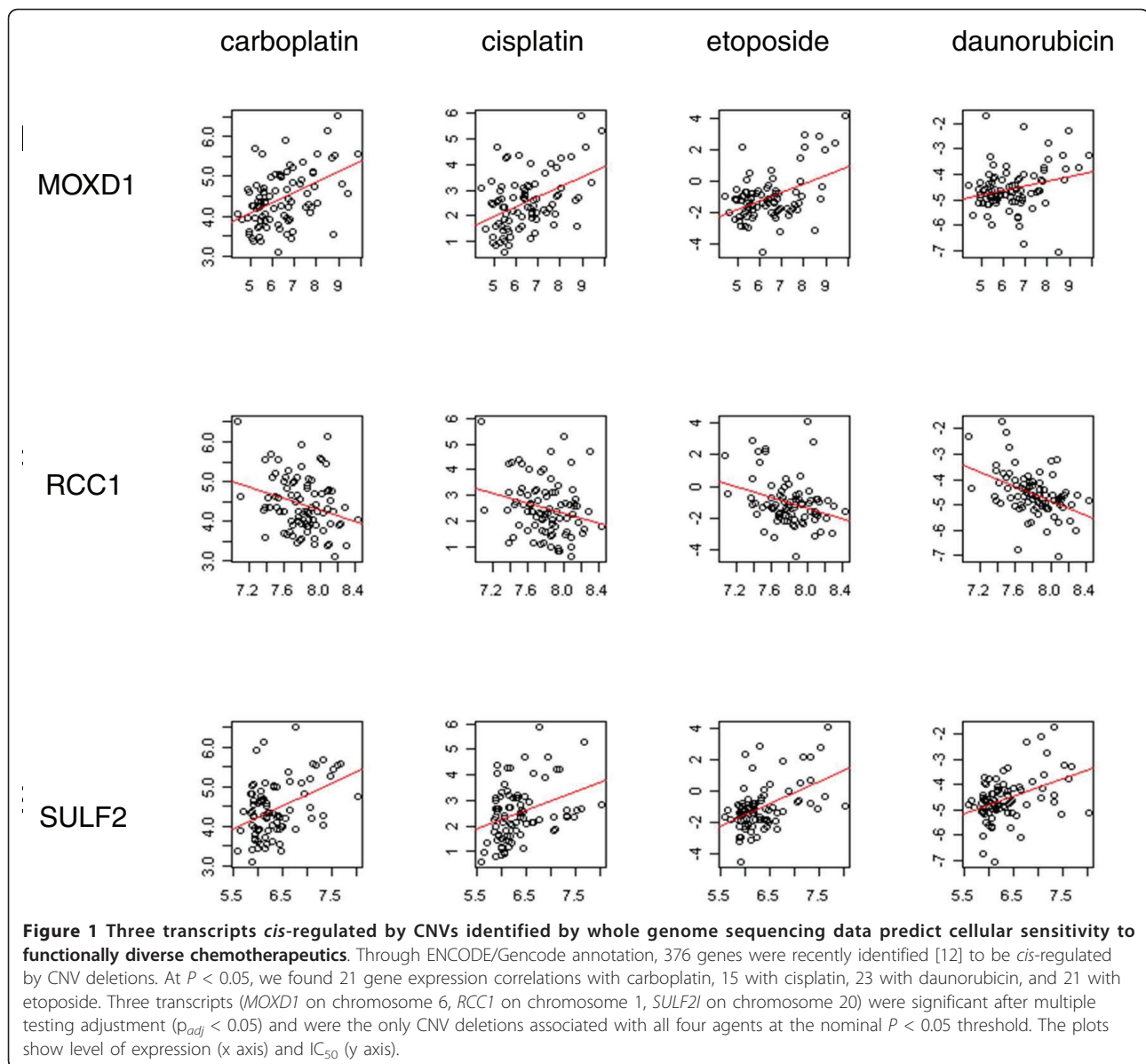
Table 1 Nominally significant ($P < 0.05$) gene expression correlations with cellular sensitivity to chemotherapeutic agents (Continued)

<i>RCC1</i>	Daunorubicin	0.00002	1
<i>RCC1</i>	Etoposide	0.01358	1
<i>RNF103</i>	Daunorubicin	0.02655	2
<i>RPN2</i>	Daunorubicin	0.01206	20
<i>SLC10A7</i>	Daunorubicin	0.00481	4
<i>SLC4A8</i>	Cisplatin	0.02189	12
<i>SLC4A8</i>	Etoposide	0.00248	12
<i>SMPD4</i>	Daunorubicin	0.02446	2
<i>SNRK</i>	Carboplatin	0.00396	3
<i>SNRK</i>	Cisplatin	0.00479	3
<i>SNRK</i>	Daunorubicin	0.02429	3
<i>SULF2</i>	Carboplatin	0.00002	20
<i>SULF2</i>	Cisplatin	0.00053	20
<i>SULF2</i>	Daunorubicin	0.00014	20
<i>SULF2</i>	Etoposide	0.00000	20
<i>TBC1D22A</i>	Cisplatin	0.03807	22
<i>TBL1XR1</i>	Carboplatin	0.04028	3
<i>TFDP1</i>	Cisplatin	0.04967	13
<i>TNK2</i>	Carboplatin	0.01901	3
<i>TNK2</i>	Daunorubicin	0.00431	3
<i>TNK2</i>	Etoposide	0.01159	3
<i>TYRO3</i>	Etoposide	0.02840	15
<i>UPK3B</i>	Cisplatin	0.02839	7
<i>WDR90</i>	Daunorubicin	0.02575	16
<i>WDR90</i>	Etoposide	0.00111	16
<i>ZDHC4</i>	Daunorubicin	0.02463	7
<i>ZFAND2A</i>	Daunorubicin	0.00227	7
<i>ZFAND2A</i>	Etoposide	0.03567	7
<i>ZNF277</i>	Etoposide	0.04521	7
<i>ZNF331</i>	Carboplatin	0.02122	19
<i>ZNF664</i>	Carboplatin	0.04722	12

Transcripts are *cis*-regulated by CNVs identified from whole-genome sequencing.

cellular sensitivity to the drug even at the nominal threshold of $P = 0.05$.

In the case of the topoisomerase II inhibitors, of the CNVs showing association with both etoposide and daunorubicin ($P < 0.05$), we found two - *CNVR7205.1* and *CNVR3293.1* - that are only modestly tagged (by *rs563079* at $r^2 = 0.77$ and *rs17166803* also at $r^2 = 0.77$, respectively). Neither *rs563079* nor *rs17166803* is associated with etoposide or daunorubicin IC_{50} . In contrast, *CNVR2930.1*, which is one of two etoposide-associated CNVs predicting the expression of *CCND1* (expression $P = 2.4 \times 10^{-7}$), is perfectly tagged ($r^2 = 1$) by *rs9500270*. We identified a daunorubicin-associated CNV (*CNVR2766.1*; $P = 3.7 \times 10^{-3}$) for which the best proxy SNP, *rs10484327*, tags the CNV at only $r^2 = 0.11$.



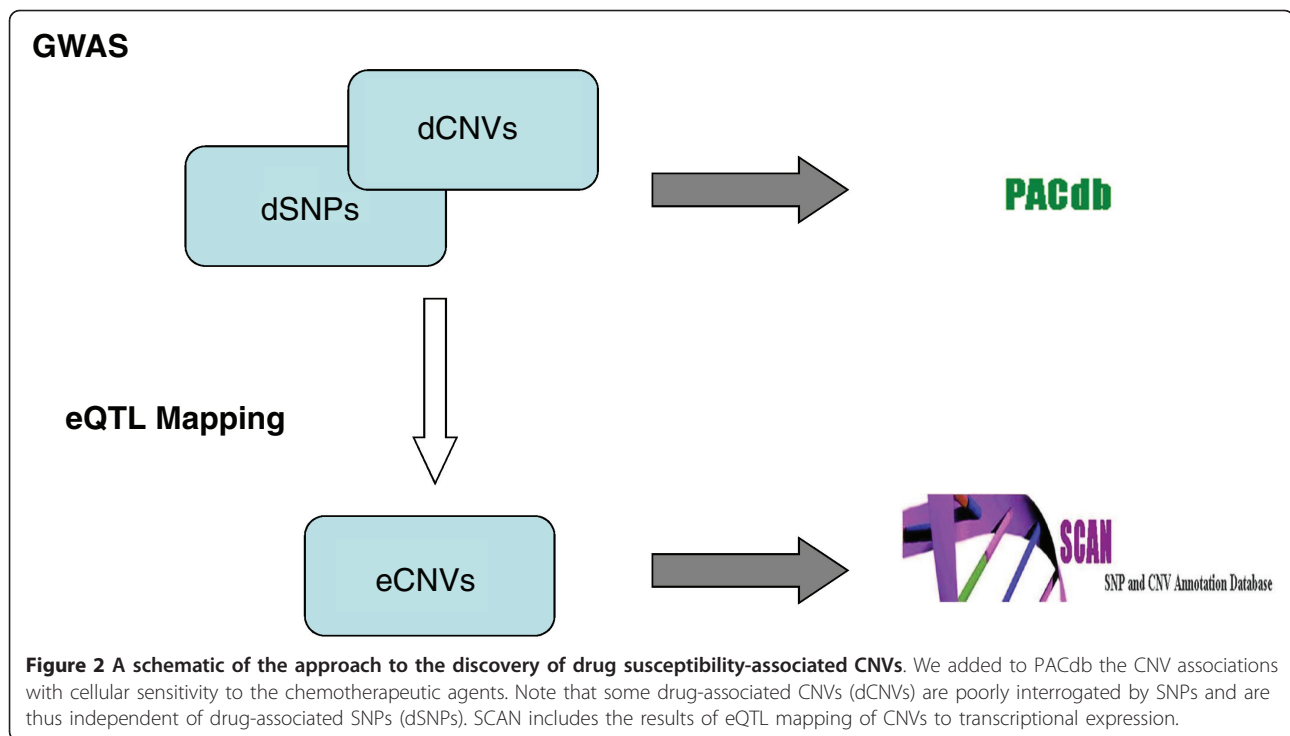
PACdb: a database for cell-based pharmacogenomics

PACdb [24] is a large-scale, publicly available genomic database, which to date holds the results of our SNP-based GWAS on the following chemotherapeutic agents: carboplatin, cisplatin, etoposide, daunorubicin, and cytarabine. PACdb implements a structured repository for incorporating other datasets, including information on other drugs, gene expression profiling, and cellular phenotypes. GWAS were initially conducted using SNP genotype data made available by the International HapMap project. We expanded PACdb to include the results of our CNV-based GWAS on carboplatin, cisplatin, etoposide, and daunorubicin. Furthermore, the results of eQTL mapping of HapMap CNVs to transcriptional expression are made available in the eQTL

repository SCAN. Figure 2 shows a schematic diagram of our approach to the discovery of CNVs associated with sensitivity to drug and to the identification of such CNVs that act as eQTLs; it also illustrates the genomic resources we have made publicly available to the scientific community.

CNVs and drug classes

We evaluated to what extent the top CNV associations for a given drug would overlap with the top CNV associations for another drug belonging to the same chemotherapeutic drug class, defined in terms of mechanism of action. At the suggestive threshold of $P < 0.05$, of the CNVs showing association with carboplatin IC_{50} , 16% ($n = 11$) were also associated with cisplatin



IC₅₀. Thus, we see a significant overlap ($P = 7.7 \times 10^{-10}$) between the (two) sets of CNVs associated with cellular sensitivity to the platinating agents. Figure 3 illustrates a duplication (*CNVR7826_full* on chromosome 20) that is associated with sensitivity to carboplatin (Figure 3a) and to cisplatin (Figure 3b); note that the observed genotype associations with the platinums have concordant direction. Furthermore, the CNV is an eQTL predicting the expression of *GSR* ($P = 4.67 \times 10^{-5}$) and *SPARC* ($P = 4.70 \times 10^{-5}$). Remarkably, the expression levels of these target mRNAs, *GSR* ($P = 0.045$) and *SPARC* ($P = 0.004$), are correlated with carboplatin IC₅₀; similarly, *GSR* ($P = 0.005$) and *SPARC* ($P = 0.005$) are correlated with cisplatin IC₅₀. Glutathione reductase (*GSR*) has been implicated in several studies of platinum sensitivity [25,26].

In the case of the topoisomerase II inhibitors, 12% of the etoposide-associated CNVs were found to associate with daunorubicin IC₅₀, and the observed overlap is still quite significant ($P = 2.7 \times 10^{-10}$). The slightly greater percentage of overlap for the platinating agents is not due to higher phenotypic correlation (platinating agents ($r = 0.52$) versus topoisomerase II inhibitors ($r = 0.69$)).

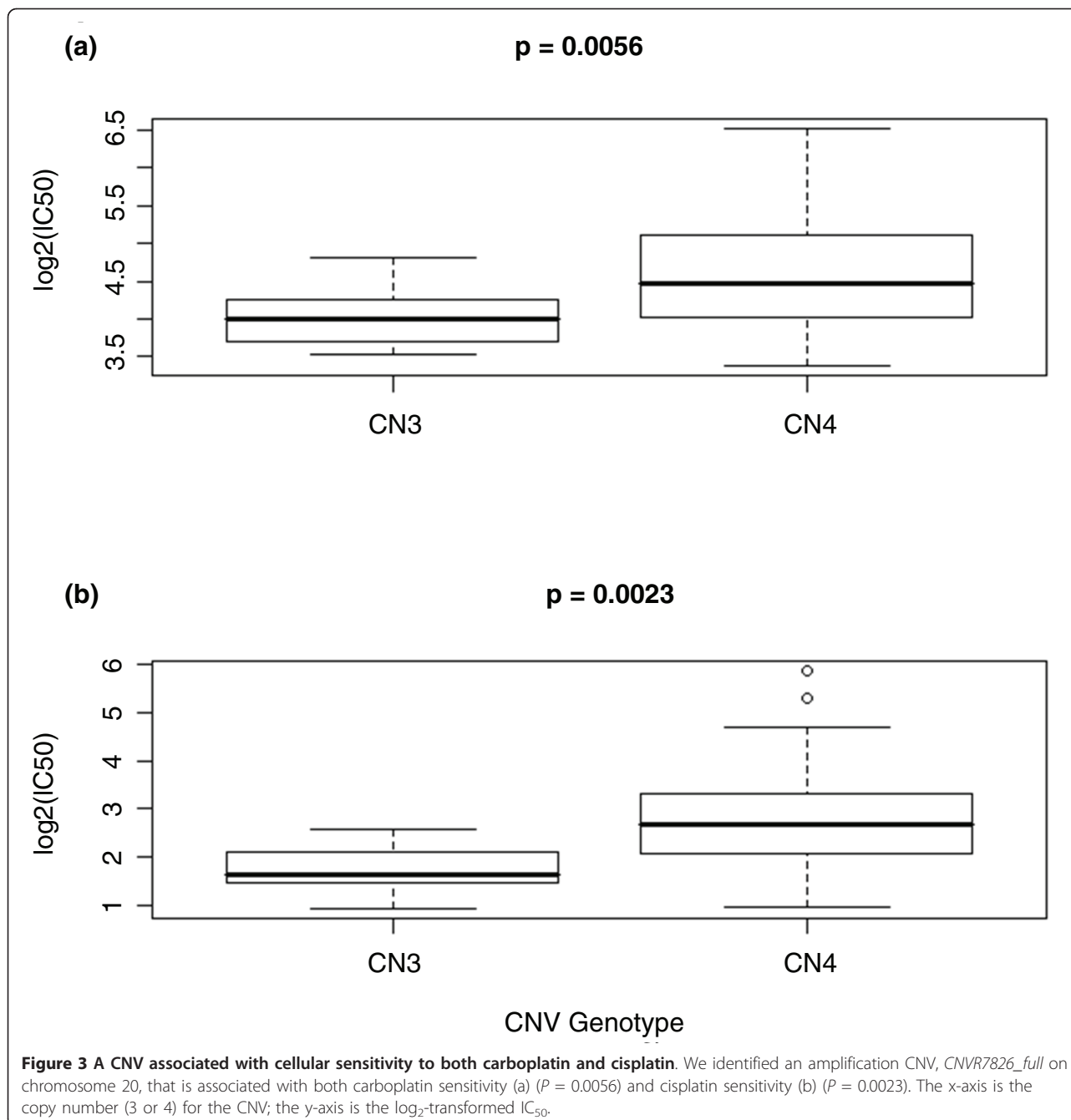
Real-time PCR validation

We sought additional experimental support for the genes targeted by multiple CNVs associated with drug susceptibility. We identified two etoposide-associated CNV eQTLs that share *CCND1* as a target gene (expression $P = 2.4 \times 10^{-7}$). The over-expression of

CCND1 has been shown to be associated with the up-regulation of the *GST-π* gene, increasing the sensitivity of a cancer cell line to etoposide [27]. We found *CCND1* expression to be significantly correlated with etoposide IC₅₀ ($P = 7.8 \times 10^{-6}$) in the CEU samples. After multiple testing correction, the gene remained significant (q-value = 0.0027). We subsequently conducted functional validation of the role of *CCND1* expression in altering sensitivity to etoposide by performing real-time quantitative-PCR assays in an independent set of 52 CEPH LCLs (Figure 4; see Table S1 in Additional file 1 for the real-time PCR data on *CCND1*). Consistent with the direction of effect in the CEU samples, increased *CCND1* mRNA levels resulted in increased IC₅₀ ($P = 0.05$) in the validation set. Thus, increasing *CCND1* expression confers resistance to etoposide.

Discussion

Understanding in a comprehensive manner the genetic risk factors contributing to variation in drug response is a crucial component of the realization of personalized medicine. The drugs evaluated in our study are widely used in the treatment of many cancer types, including ovarian, colorectal, testicular, and lung; all are associated with particular toxicities and resistance. Although SNPs have long been used in association studies to elucidate the effect of genetic polymorphisms on drug response, CNVs have been relatively understudied. Recent



genome-wide surveys of CNVs have now established that these structural variants are a common phenomenon in the human genome [5]. With rapid advances in methods that facilitate their assay and analysis, variation in copy number for genes encoding drug metabolizing enzymes has been increasingly implicated for their dramatic consequences on responsiveness to drugs. Such CNVs have been observed to alter gene dosage and are thus likely to play an important role in determining drug efficacy or toxicity.

In this study, we set out to utilize recent developments in the assay of CNVs in recent population-scale projects, including an extensive comparative genomic hybridization-based catalog of CNVs [4] and a map of structural variants based on whole genome DNA sequencing data (the 1000 Genomes Project) [14], in order to evaluate the role of CNVs in cellular sensitivity to chemotherapeutic agents. The cell lines for the samples express a sizable part of the genome [28], thus enabling the investigation of genes represented in

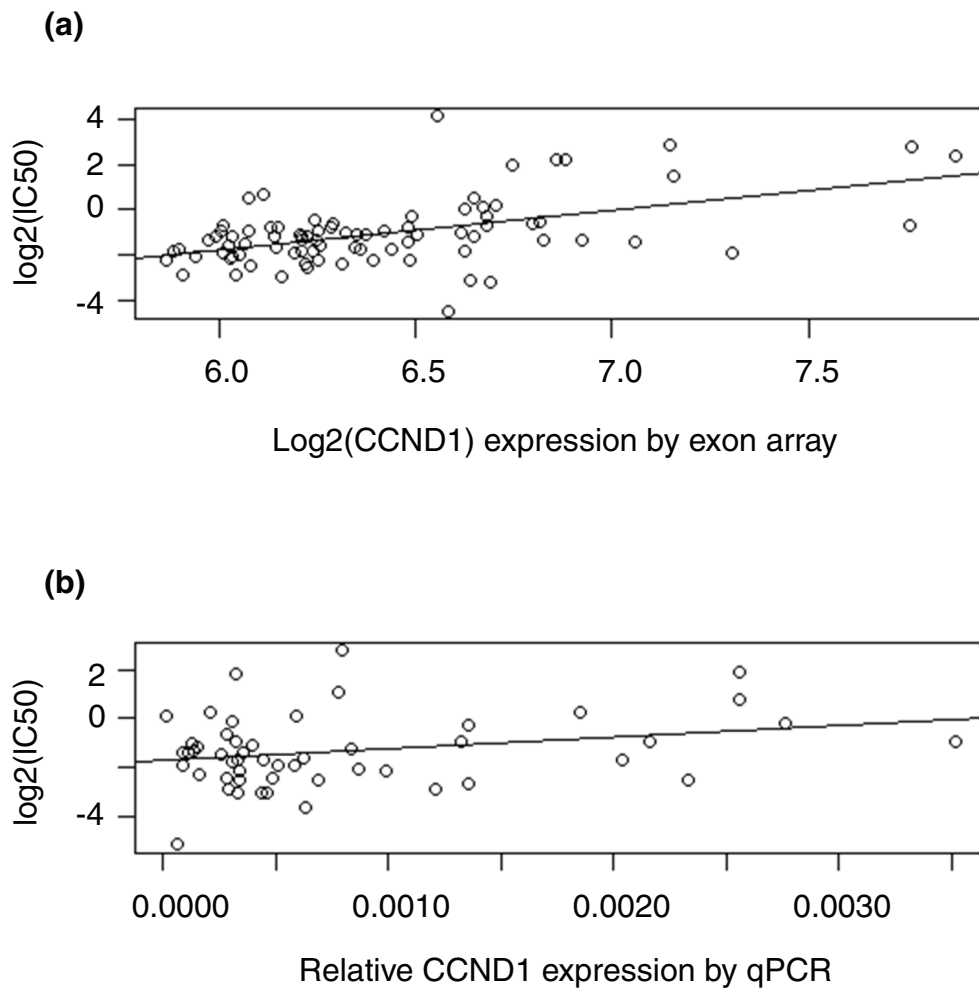


Figure 4 The relationship between $CCND1$ expression and etoposide IC_{50} in the discovery set and in an independent replication set of 52 CEPH LCLs. (a) Discovery set; (b) independent replication set of 52 CEPH LCLs. Panel (b) describes the expression- IC_{50} relationship using real-time expression; the relationship is consistent with that in the discovery set of CEU samples.

biologically relevant pathways. While the cancer genome is clearly necessary for understanding chemotherapeutic response, the importance of germline genetic variation in drug sensitivity has also been consistently demonstrated [15].

For each drug included in our study, we found that the top associated CNVs are more likely to act as eQTLs and predict transcript levels than minor allele frequency (MAF)-matched SNPs. The overlap of the drug susceptibility-associated CNVs with expression-associated CNVs (eCNVs) is greater than is expected, based on simulation studies. Consistent with a previous report [6], CNVs associated with cellular sensitivity to drug treatment are not likely to overlap exons, suggesting that they act not to disrupt coding sequence but to regulate gene expression. The high proportion of eQTLs among the CNVs associated with cellular sensitivity to each of the drugs further supports the hypothesis that

these CNVs mediate their phenotypic consequences through their effect on the transcriptome. Genome-wide studies of pharmacologic phenotypes, such as response to antineoplastic agents, may benefit from studies of CNVs as eQTLs.

This study, to our knowledge, is the first comprehensive genome-wide study of the effect of CNVs, from the most extensive array-based and sequencing-based surveys of these structural variants, on pharmacologic phenotypes. In contrast to a recent disease susceptibility study that concluded that most CNVs that are well-typed have been indirectly explored by SNP studies [6], we found a number of CNVs associated with drug sensitivity that are independent of SNPs. These CNVs therefore constitute novel genetic variations that have not been previously interrogated by SNP-based GWAS of pharmacologic phenotypes. Our discovery of drug susceptibility-associated variations, in the form of CNVs,

that are independent of previous SNP findings and that show evidence for altering gene expression as eQTLs, suggests that CNVs should be included in comprehensive pharmacogenomic studies.

Candidate pharmacogenetic studies on drug metabolism-related genes, namely *CYP2D6*, *CYP2A6*, *SULT1A1* and *GSTM1*, have documented the effect of CNVs on gene activity. Our results strongly support the necessity of integrating both SNP and CNV data to tighten the genotype-phenotype gap in pharmacogenetic studies. While the functional validation we conducted in this study may not allow robust predictions, the functional characterization of the effect of *CCND1* mRNA level on cellular sensitivity to etoposide underscores the importance of considering the role of the transcripts that are the targets of drug susceptibility-associated CNVs (acting as eQTLs) in conferring drug susceptibility.

We found a significant overlap ($P = 7.7 \times 10^{-10}$) between the CNVs associated with cisplatin and carboplatin. Platinating agents share a similar mechanism of therapeutic action and interact with DNA to form inter-strand and intrastrand cross-links, leading to cytotoxic DNA lesions and eventually apoptosis-induced cell death. Our findings strongly support the hypothesis that CNV-based mechanisms play a crucial role in determining platinum sensitivity. Particularly, we identified a duplication that is associated with cellular sensitivity to both carboplatin and cisplatin. Furthermore, the CNV predicts the expression of glutathione reductase (*GSR*), a gene that has been the subject of several studies on cisplatin sensitivity [26,30]. The glutathione pathway is involved in the metabolism of platinum compounds, which are subject to inactivation by glutathione conjugation [27].

A significant level of overlap is also observed with the topoisomerase II inhibitors. Daunorubicin is a DNA intercalator that indirectly interacts with topo II while etoposide binds directly to the enzyme. We identified 14 CNVs associated with both etoposide and daunorubicin at $P < 0.05$. The extent of overlap between the platinating agents (as well as between the topoisomerase II inhibitors) is significantly higher than the level of overlap across drug classes (7%).

There is a general caveat to our findings concerning the set of CNVs included in this analysis. The CNVs tested for association with cellular sensitivity to drugs may be biased towards genotypeable variants; consequently, many highly complex regions may have been excluded. Furthermore, our study makes no assertions about low frequency variants. Nevertheless, our findings represent the most comprehensive study of the effect of common CNVs, from the most extensive map of these variants available, on chemotherapeutic susceptibility to a wide array of drugs.

Finally, we provide the results of our genome-wide study of CNVs and sensitivity to chemotherapeutic agents in a publicly available online database, PACdb. Analysis results on our cell-based model are easy to query, which should allow investigators to utilize the resource as a discovery platform or as a validation tool for clinical observations.

Conclusions

Our study identified CNVs that predict cellular sensitivity to an array of chemotherapeutic agents of heterogeneous molecular therapeutic action. Importantly, several of the most significant CNV-drug associations are independent of SNPs; thus, these CNVs provide genetic variations that have not been previously explored by SNP-based GWAS of pharmacologic phenotypes. Furthermore, our findings show that pharmacogenomic studies may be greatly enhanced by studies of CNVs as eQTLs. Target genes of CNVs, especially those associated with multiple independent CNVs associated with drug response, provide robust gene expression signatures of chemotherapeutic susceptibility.

Materials and methods

In vitro cellular sensitivity to chemotherapeutic agents

We obtained unrelated HapMap phase II CEU (Utah residents with ancestry from northern and western Europe) samples from Coriell Institute for Medical Research (Camden, NJ, USA). Cell lines were maintained in RPMI 1640 media supplemented with 15% fetal bovine serum (Hyclone, Logan, UT, USA) and 1% l-glutamine. The cell lines were passaged three times per week at a concentration of 350,000 cells/ml at 37°C in a 95% humidified 5% CO₂ atmosphere. Cellular sensitivity to drugs was measured in these cell lines with increasing concentrations of drug (from carboplatin, cisplatin, daunorubicin, and etoposide). Cell growth inhibition was evaluated using the alamarBlue™ assay (BioSource International Inc., Camarillo, CA, USA), as previously described [21]. IC₅₀ (the concentration required to inhibit 50% of cell growth) was determined by curve fitting of percent cell survival against concentrations of the drug.

A catalog of CNVs

Recent population-based surveys have mapped thousands of CNVs with increasing resolution. Such surveys have opened up approaches for modeling the relationship between structural variation and complex traits. Efforts to catalog these CNVs are necessary to clarify the functional impact of these variants. Here we utilize the recent comprehensive survey of CNVs [4] larger than 1 kb in size in the human genome, assayed in human LCLs from CEU (Utah residents with ancestry from northern and western Europe) samples. To

investigate further the effect of deletions and to confirm our findings on the role of drug-associated CNVs as eQTLs, we studied the relationship between *cis*-regulated transcripts (from Gencode/ENCODE functional annotation) and cellular sensitivity to chemotherapeutics from a recent comprehensive study based on whole genome sequencing data of the 1000 Genomes Project [14], which mapped CNVs of 50 bp or larger in size at nucleotide resolution.

Association analysis of CNVs or transcript levels with cellular sensitivity to drugs

For each CNV, genotypes were tested for association with cellular sensitivity to each of the drugs separately. Linear regression was performed between the copy number genotype (as the independent variable) and \log_2 -transformed IC_{50} (as the dependent variable). Analysis was done in the statistical computing and graphics software R; the *lm* function was used to fit linear models.

Similarly, to examine the relationship between transcript level and drug-induced cellular sensitivity, a linear model was constructed for each drug, as previously described [19], between \log_2 -transformed gene expression and \log_2 -transformed IC_{50} . Generally, for multiple testing adjustment, FDRs were calculated using the *q*-value [23] package in R. Unless otherwise stated, an FDR cutoff <0.05 was used as the statistical significance threshold.

Mapping CNVs as expression quantitative trait loci

SCAN [29] is an online database that makes publicly available the results of our eQTL studies, initially on single base polymorphisms. Global mRNA expression was assayed using the Affymetrix GeneChip Human Exon 1.0 ST Array [30]. To map CNVs as genomic loci influencing the transcriptome, we conducted linear regression on over 13,000 transcript clusters and the set of CNVs identified in the HapMap LCLs [31].

Simulation studies

We performed simulations to evaluate enrichment for eQTLs among the CNVs associated with cellular sensitivity to the drugs included in our study. To empirically generate the null distribution, we randomly generated sets of SNPs of matching minor allele frequency as the original list of CNVs (see Figure S1 in Additional file 2 for MAF distribution of the biallelic CNVs included in our study), as previously described [32]. To enable us to perform simulations conditional on MAF, we constructed non-overlapping MAF bins, each of width 0.05, using the MAFs of the SNPs in the HapMap CEU samples. The null sets were drawn from the combined platform SNPs (Affymetrix 6.0 and Illumina 1M) as well as from the

entire set of HapMap CEU SNPs. The observed count is then compared to the empirically generated distribution to get an empirical *P*-value for the enrichment.

Functional validation

We obtained 52 unrelated non-HapMap CEPH (Centre d'Etude du Polymorphisme Humain) samples (independent of the discovery cohort consisting of HapMap CEU samples) from Coriell Institute for Medical Research. Cellular sensitivity to etoposide phenotype was quantified as described above with increasing concentrations of etoposide treatment (0.02 μ M, 0.1 μ M, 0.5 μ M, and 2.5 μ M for 72 hours). IC_{50} was determined for each cell line. *CCND1* mRNA levels were evaluated using a real-time quantitative PCR assay in the samples using TaqMan Gene Expression Assays (Applied Biosystems, Foster City, CA, USA) on the Applied Biosystems 7500 real-time PCR system. Primer/probes were obtained from Applied Biosystems. The human beta 2M (*huB₂M*, beta-2 microglobulin; NM_004048; Applied Biosystems catalog number 4326319E) was used as endogenous control. Relative quantification of gene expression utilized the $2^{-\Delta\Delta Ct}$ method [33].

Additional material

Additional file 1: Real-time PCR data on *CCND1*. The table lists the real-time PCR values for *CCND1* expression, as measured in the independent set of 52 cell lines.

Additional file 2: The minor allele frequency distribution of the biallelic CNVs included in our study. The plot is a histogram of the minor allele frequency of the biallelic CNVs that were evaluated in our study.

Abbreviations

bp: base pair; CNV: copy number variant; eQTLs: expression quantitative trait loci; FDR: false discovery rate; GST: Glutathione S-transferase; GWAS: genome-wide association studies; LCL: lymphoblastoid cell line; MAF: minor allele frequency; PCR: polymerase chain reaction; SNP: single nucleotide polymorphism.

Acknowledgements

This work was funded through Pharmacogenomics of Anticancer Agents Research (PAAR; U01 GM61393), ENDGAME (Enhancing Development of Genome-wide Association Methods) initiative (U01 HL084715), the Genotype-Tissue Expression project (GTEx) (R01 MH090937), Rare Variants and Complex Human Phenotypes (U01HG005773), the University of Chicago Breast Cancer SPORE (P50 CA125183), Specialized Center of Research Grant from the Leukemia and Lymphoma Society and the University of Chicago DRTC (Diabetes Research and Training Center; P60 DK20595). RSH received support from NIH/NIGMS grant K08GM089941, NIH/NCI grant R21 CA139278, University of Chicago Cancer Center Support Grant (#P30 CA14599), and Breast Cancer SPORE Career Development Award. The authors are grateful to Anuar Konkashbaev and Wasim Bleibel for excellent technical assistance. We acknowledge the Wellcome Trust Case Control Consortium for making available data about SNP tagging of common CNVs [34].

Author details

¹Section of Genetic Medicine, Department of Medicine, University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA. ²Section of

Hematology/Oncology, Department of Medicine, University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA. ³Department of Human Genetics, University of Chicago, 920 East 58th Street, CLSC 5th floor, Chicago, IL 60637, USA.

Authors' contributions

ERG conceived the study. ERG, RSH, MED, and NJC designed the experiments and the analyses and wrote the manuscript. ERG analyzed the data. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 22 March 2011 Revised: 8 May 2011 Accepted: 25 May 2011
Published: 25 May 2011

References

- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK: **The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.** *Science* 2005, **307**:1434-1440.
- McCarroll SA, Huett A, Kuballa P, Cholewicki SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ: **Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease.** *Nat Genet* 2008, **40**:1107-1112.
- McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40**:1166-1174.
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2010, **464**:704-712.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, *et al*: **Global variation in copy number in the human genome.** *Nature* 2006, **444**:444-454.
- Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatos E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, *et al*: **Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.** *Nature* 2010, **464**:713-720.
- Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjöqvist F, Ingelman-Sundberg M: **Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine.** *Proc Natl Acad Sci USA* 1993, **90**:11825-11829.
- Goetz MP, Knox SK, Suman VJ, Rae JM, Safgren SL, Ames MM, Visscher DW, Reynolds C, Couch FJ, Lingle WL, Weinsilboum RM, Fritcher EG, Nibbe AM, Desta Z, Nguyen A, Flockhart DA, Perez EA, Ingle JN: **The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen.** *Breast Cancer Res Treat* 2007, **101**:113-121.
- Gasche Y, Daali Y, Fathi M, Chiappe A, Cottini S, Dayer P, Desmeules J: **Codeine intoxication associated with ultrarapid CYP2D6 metabolism.** *N Engl J Med* 2004, **351**:2827-2831.
- Malaiyandi V, Sellers EM, Tyndale RF: **Implications of CYP2A6 genetic variation for smoking behaviors and nicotine dependence.** *Clin Pharmacol Ther* 2005, **77**:145-158.
- Hebbring SJ, Moyer AM, Weinsilboum RM: **Sulfotransferase gene copy number variation: pharmacogenetics and function.** *Cytogenet Genome Res* 2008, **123**:205-210.
- Nagle CM, Chenevix-Trench G, Spurdle AB, Webb PM: **The role of glutathione-S-transferase polymorphisms in ovarian cancer survival.** *Eur J Cancer* 2007, **43**:283-290.
- Müller P, Asher N, Heled M, Cohen SB, Risch A, Rund D: **Polymorphisms in transporter and phase II metabolism genes as potential modifiers of the predisposition to and treatment outcome of de novo acute myeloid leukemia in Israeli ethnic groups.** *Leuk Res* 2008, **32**:919-929.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, *et al*: **Mapping copy number variation by population-scale genome sequencing.** *Nature* 2011, **470**:59-65.
- Welsh M, Mangravite L, Medina MW, Tantisira K, Zhang W, Huang RS, McLeod H, Dolan ME: **Pharmacogenomic discovery using cell-based models.** *Pharmacol Rev* 2009, **61**:413-429.
- Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME: **A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity.** *Proc Natl Acad Sci USA* 2007, **104**:9758-9763.
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG: **Genetic analysis of genome-wide variation in human gene expression.** *Nature* 2004, **430**:743-747.
- Dolan ME, Newbold KG, Nagasubramanian R, Wu X, Ratain MJ, Cook EH Jr, Badner JA: **Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity.** *Cancer Res* 2004, **64**:4353-4356.
- Huang RS, Ratain MJ: **Pharmacogenetics and pharmacogenomics of anticancer agents.** *CA Cancer J Clin* 2009, **59**:42-55.
- Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME: **Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans.** *Mol Cancer Ther* 2008, **7**:3038-3046.
- Huang RS, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, Schweitzer AC, Blume JE, Dolan ME: **Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genome-wide approach.** *Am J Hum Genet* 2007, **81**:427-437.
- Huang RS, Duan S, Kistner EO, Bleibel WK, Delaney SM, Fackenthal DL, Das S, Dolan ME: **Genetic variants contributing to daunorubicin-induced cytotoxicity.** *Cancer Res* 2008, **68**:3161-3168.
- Storey JD, Tibshirani R: **Statistical significance for genome-wide studies.** *Proc Natl Acad Sci USA* 2003, **100**:9440-9445.
- Gamazon ER, Duan S, Zhang W, Huang RS, Kistner EO, Dolan ME, Cox NJ: **PACdb: a database for cell-based pharmacogenomics.** *Pharmacogenet Genomics* 2010, **20**:269-273.
- Ogawa J, Iwazaki M, Inoue H, Koide S, Shohtsu A: **Immunohistochemical study of glutathione-related enzymes and proliferative antigens in lung cancer. Relation to cisplatin sensitivity.** *Cancer* 1993, **71**:2204-2209.
- Moyer AM, Sun Z, Batzler AJ, Li L, Schaid DJ, Yang P, Weinsilboum RM: **Glutathione pathway genetic polymorphisms and lung cancer survival after platinum-based chemotherapy.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:811-821.
- Bennaceur-Griscelli A, Bosq J, Koscielny S, Lefrère F, Turhan A, Brousse N, Hermine O, Ribrag V: **High level of glutathione-S-transferase pi expression in mantle cell lymphomas.** *Clin Cancer Res* 2004, **10**:3029-3034.
- Cheung VG, Conlin LK, Weber TM, Arcafo M, Jen KY, Morley M, Spielman RS: **Natural variation in human gene expression assessed in lymphoblastoid cells.** *Nat Genet* 2003, **33**:422-425.
- Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ: **SCAN: SNP and copy number annotation.** *Bioinformatics* 2010, **26**:259-262.
- Duan S, Huang RS, Zhang W, Bleibel WK, Roe CA, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME: **Genetic architecture of transcript-level variation in humans.** *Am J Hum Genet* 2008, **82**:1101-1113.
- Gamazon ER, Nicolae DL, Cox NJ: **A study of CNVs as trait-associated polymorphisms and as expression quantitative trait loci.** *PLoS Genet* 2011, **7**:e1001292.
- Gamazon ER, Huang RS, Cox NJ, Dolan ME: **Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci.** *Proc Natl Acad Sci U S A* 2010, **107**:9287-9292.

33. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* 2001, **25**:402-408.
34. Wellcome Trust Case Control Consortium: Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls - supplementary information. [http://www.wtccc.org.uk/wtcccplus_cnv/supplemental.shtml].

doi:10.1186/gb-2011-12-5-r46

Cite this article as: Gamazon *et al.*: Copy number polymorphisms and anticancer pharmacogenomics. *Genome Biology* 2011 **12**:R46.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

