Published in final edited form as:

Pharmacogenet Genomics. 2008 June ; 18(6): 545–549. doi:10.1097/FPC.0b013e3282fe1745.

Identification of genetic variants and gene expression relationships associated with pharmacogenes in humans

Rong Stephanie Huang^{a,*}, Shiwei Duan^{a,*}, Emily O. Kistner^c, Wei Zhang^a, Wasim K. Bleibel^a, Nancy J. Cox^d, and M. Eileen Dolan^a

^aSection of Hematology-Oncology, University of Chicago, Chicago, Illinois, USA

^bSection of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA

^cDepartment of Health Studies, University of Chicago, Chicago, Illinois, USA

Abstract

Objectives—The very important pharmacogenes (VIPs) were selected by Pharmacogenetic Research Network (National Institutes of Health-PGRN) owing to their significant effects on drug treatment both at the pharmacokinetic and pharmacodynamic levels. Our objective was to identify single nucleotide polymorphisms (SNPs) that potentially affected the expression of these genes or potential SNP–gene interactions involved to improve our understanding of genetic effects on drug therapy.

Basic methods—Gene expression was evaluated in 176 International HapMap lymphoblastoid cell lines derived from CEU (CEPH, Utah residents with ancestry from northern and western Europe; $n = 87$) and YRI (Yoruba in Ibadan, Nigeria; $n = 89$) using Affymetrix GeneChip Human Exon 1.0 ST arrays (Affymetrix Laboratory, Affymetrix Inc., Santa Clara, California, USA) with interrogation of greater than 17000 human genes. Genome-wide association was performed between over two million publicly available HapMap SNPs and gene expression.

Main results—The expression of two PGRN-VIPs (*GSTT1* and *GSTM1*) are significantly associated with SNPs within 2.5 Mb of the genes; whereas the expression of three and ten PGRN-VIPs are significantly associated with distant-acting SNPs in CEU and YRI, respectively. In addition, three and four PGRN-VIPs harbor SNPs that are distantly associated with other gene expressions in CEU and YRI, respectively.

Principal conclusion—Using this information, one may identify genetic variants that are significantly associated with the expression of any set of genes of interest; or evaluate potential gene–gene interaction through SNP expression relationships.

Keywords

exon array; expression quantitative trait loci; gene expression; glutathione-S-transferase; pharmacogenes

Correspondence to M. Eileen Dolan, 5841 S. Maryland Ave. Box MC2115, University of Chicago, Chicago, IL 60637, USA Tel: + 1 773 702 4441; fax: + 1 773 702 0963; e-mail: edolan@medicine.bsd.uchicago.edu.

*R. Stephanie Huang and Shiwei Duan contributed equally.

Supplementary data

Supplementary data are available at *The Pharmacogenetics and Genomics Journal* Online (www.jpharmacogenetics.com).

Introduction

Genetics has been shown to play a role in human variation of many complex traits, including gene expression [1]. Our laboratory has recently examined gene expression in 176 International HapMap lymphoblastoid cell lines (LCLs) derived from CEU (CEPH, Utah residents with ancestry from northern and western Europe) and YRI (Yoruba in Ibadan, Nigeria) using Affymetrix GeneChip Human Exon 1.0 ST array (exon array), which interrogates more than 17 000 human genes [2]. Over two million publicly available HapMap single nucleotide polymorphisms (SNPs) (www.hapmap.org) [3] and gene expression (GenBank accession no: GSE7761) are available for these cell lines. In this report, we demonstrate the utility of this information using the very important pharmacogenes (VIPs) selected by the Pharmacogenetic Research Network (PGRN-VIPs). The PGRN-VIPs are a collection of genes, although not comprehensive, that have been shown to play a significant role in drug treatment both at the pharmacokinetic and pharmacodynamic levels (<http://www.pharmgkb.org/search/annotatedGene/index.jsp>). Genetic polymorphisms that affect the expression and/or activity of these PGRN-VIP products are likely to impact the pharmacokinetics and pharmacodynamics of drug therapy. Our objective was to identify SNPs that potentially affected the expression of these genes or potential gene–gene interactions involved in samples from populations of CEU and YRI. This study is aimed to improve our understanding of genetic effects on drug therapy.

Methods

Cell lines

Epstein-Barr virus-transformed B-LCLs derived from 30 CEU trios (HAPMAPPT01) and 30 YRI trios (HAPMAPPT03) were purchased from the Coriell Institute for Medical Research (Camden, New Jersey, USA). Cell lines were maintained and diluted as described [4].

Genotype and gene expression association analysis

SNP genotypes were downloaded from the International HapMap database (<http://www.HapMap.org>) (release 22). Only 2 098 437 and 2 286 186 SNPs that passed Mendelian error checks and have a minor allele frequency greater than 5% in CEU and YRI, respectively, were included in the analysis. The mRNA expression of 176 LCLs (87 CEU and 89 YRI) was assessed through exon arrays as described previously [2]. The quantitative transmission disequilibrium test (QTDT) was performed to identify any genotype–gene expression association on the whole-genome scale using QTDT software [5] (<http://www.sph.umich.edu/csg/abecasis/QTDT>). Given the complex linkage disequilibrium (LD) existence between SNPs, Bonferroni-corrected $P < 0.05$ with 550 000 tag SNPs was considered statistically significant. The 550 000 tag SNPs were selected based on Illumina HumanHap550 BeadChip design using HapMap SNP data [6] (http://www.illumina.com/downloads/HUMANHAP550_TechBull.pdf).

Evaluation of very important pharmacogenes selected by Pharmacogenetic Research Network

To examine whether the PGRN-VIPs were probed on the exon array, the gene symbols or Entrez gene IDs representing all 63 PGRN-VIPs were queried through Affymetrix NetAffx database (https://www.affymetrix.com/analysis/netaffx/xmlquery_ex.affx). Transcript cluster IDs that represent PGRN-VIPs on the exon array were also obtained through NetAffx. An average signal density greater than 25th quantile of all transcript clusters expression, which translates to mean $\log_2$ -transformed mRNA expression signals greater than 5.34, was defined as expressed in LCLs in this study. To identify differentially expressed PGRN-VIPs in CEU and YRI, a *t*-test-based Westfall–Young (W–Y) approach [7] that adjusts for the correlations

among trio members and a general linear model using a Toeplitz correlation matrix for the trio structure were independently performed. The former is a standard pooled variance t statistic. As gene expression from individuals in the same trio may be correlated, trios were permuted between the CEU and YRI samples. The W–Y approach (10 000 permutations) was then used to compute simultaneous P values that control the overall or family-wise error rate. This is equivalent to assuming that the trios are independent and membership is defined at the trio-level. The permutation-adjusted P value ($P_c < 0.1$) was considered statistically significant. The permutation-adjusted, one-sided P values were calculated using the software Permax 2.2, which was provided as a contributory library by Robert Gray in the R statistical package [8]. Details of the latter approach have been described elsewhere [2].

The false discovery rates less than 0.05 were considered to be statistically significant for this general linear model. Utilizing results of the genome-wide association (GWA) analysis between genotype and gene expression, we evaluated whether there were potential SNP regulators of PGRN-VIP expression. We also examined potential gene–gene interaction through PGRN-VIPs harboring SNPs that regulate expression of other genes. The LD of significant SNPs within each population was evaluated using Haploview version 3.32. A general linear model was constructed to estimate the association between multiple SNPs and gene expression as described [2].

Results and discussion

Sixty-three PGRN-VIPs (22 well-annotated) are currently listed on the PGRN website (<http://www.pharmgkb.org/search/annotatedGene/index.jsp>) as the first set of genes of focus for the PGRN. Baseline gene expression was evaluated in 176 HapMap LCLs (87 CEU and 89 YRI) using exon arrays [2]. A total of 61 of the 63 PGRN-VIPs were probed on this array. Among them, 39 genes (43 transcript clusters) were expressed in CEU and YRI LCLs (defined as mean $\log_2$ -transformed expression signals greater than 25th quantile of all transcript clusters expression) (Table 1) with large interindividual gene expression variation within each sample set. Two independent statistical approaches were used to identify differentially expressed PGRN-VIPs: a t -test-based W–Y approach [7] that adjusts for the correlations among trio members and a general linear model using a Toeplitz correlation matrix for the trio structure. Of the 39 PGRN-VIPs expressed in LCLs, seven showed significant expression difference between the CEU and YRI samples ($P_c < 0.1$ after W–Y approach and $P_c < 0.05$ after linear regression corrected by 43 transcript clusters. Table 1; Fig. 1). The seven differentially expressed genes are *ABCC1*, *ADRB1*, *BRCA1*, *GSTM1*, *GSTP1*, *TGFB1*, and *VDR*.

GWA was performed between over two million publicly available HapMap SNPs (www.hapmap.org) [3] and gene expression in LCLs to identify the relationship between genetic variants and gene expression for the PGRN-VIPs. As a result of the observed difference in gene expression, the QTDT-GWA analysis was performed separately in both the CEU and YRI populations with 2 098 437 and 2 286 186 SNPs, respectively. We identified genome-wide SNP regulators for five and 12 PGRN-VIP expressions in the CEU and YRI samples, respectively ($P_c < 0.05$ after Bonferroni correction by 550 000 tag SNPs). Significant SNP regulators were found common in CEU and YRI samples for five genes, including *ALOX5*, *GSTT1*, *GSTM1*, *HMGCR*, and *G6PD*. In addition, we found three and four PGRN-VIPs that harbor SNP regulators that are significantly associated with other gene expressions in either CEU or YRI, respectively. Specifically, SNPs located in *DRD2* (two SNPs), *ESR1* (one SNP), and *KCNQ1* (one SNP) genes were significantly associated with the expression of *EDG2*, *C20orf23*, and *KIAA1407* genes in CEU; whereas SNPs located in *ESR1* (four SNPs), *KCNQ1* (one SNP), and *EGFR* (one SNP) genes, were significantly associated with the expression of *PLEC1*, *ALPI*, and *SHARPIN* in YRI, respectively. Additionally in YRI, two

SNPs in *DPYD* gene are associated with the expression of both *IL1B* and *CBS* expression (Supplementary Table 1).

Many studies of genetic effects on gene expression have focused on SNPs located in or near a gene. Our genome-wide approach examined not only the local-acting relationships [gene expression associated with SNP(s) within 2.5 Mb on the same chromosome], but also the distant-acting relationships [gene expression associated with SNP(s) on different chromosome (s) or more than 2.5 Mb away on the same chromosome]. Of all the SNPs that significantly associated with the PGRN-VIP expression, the majority displayed a distant-acting relationship with their target genes. Specifically, we identified local-acting SNPs for *GSTT1* and *GSTM1* expression, which showed greater *P* values for association as compared with distant-acting SNPs. The more significant *P* value associated with local-acting SNPs are consistent with the possibility that local regulation provides more direct (i.e. less variable) genetic control than distant regulation. Local-acting effects are more likely to be functional. Alternatively, it could be that the proportion of false positive SNP regulators is greater among distant-acting effects [9]. Nevertheless, the distant-acting SNPs suggested potential novel genetic regulators of gene expression.

Both *GSTT1* and *GSTM1* are members of glutathione S-transferases family. The phase II detoxification enzymes *GSTT1* and *GSTM1* play a role in cellular protection from environmental and oxidative stress. The overexpression of GST isozymes has also been shown to be associated with chemotherapeutic resistance [10]. We identified local-regulating elements significantly associated with mRNA expression levels of *GSTT1* in CEU (35 SNPs) and YRI (one SNP). The 35 SNPs are located in proximity to *GSTT1* gene on chromosome 22 and exhibit high LD (Supplementary Table 1). A general linear model was constructed with SNP genotypes as independent variables and *GSTT1* expression as the dependent variable. A backward selection model was performed. Specifically, seven representative SNPs (rs5760176, rs5760147, rs140188, rs4822458, rs6003959, rs4461358, and rs4820571) of the 35 were significant predictors of *GSTT1* expression. Computing a weighted sum of R^2 from each group of unrelated individuals, the multivariate model including all seven SNPs gives an overall estimate of $R^2 = 0.79$, suggesting that these seven SNP genotypes account for 79% of the sample variation in *GSTT1* expression.

We also identified local-regulating elements that significantly associated with *GSTM1* expression in CEU (two SNPs) and YRI (two SNPs). One of the SNPs (rs366631) was significantly associated with *GSTM1* expression in both CEU and YRI samples. Further examination of the difference in *GSTM1* expression in CEU and YRI revealed a potential genetic contribution. The TT genotype is associated with lower *GSTM1* expression and the observed expression difference between CEU and YRI samples is likely because of the allele frequency difference between the two samples (Fixation index; $F_{st} = 0.11$; Fig. 2). This finding is in agreement with previous findings that genetic polymorphisms may contribute to observed sample differences in gene expression [11,12,13].

Common gene deletions have been reported for both *GSTT1* [14] and *GSTM1* [15]. Our gene expression data support the potential *GSTM1* and *GSTT1* gene deletion given the large expression variation ($\log_2$ -transformed signal density range from 5 to 11 for *GSTM1* and from 4.8 to 9 for *GSTT1*). Both *GSTM1* and *GSTT1* expression show a bimodal frequency distribution with expression intensity for the lower expression group close to limit of detection for expression suggesting gene deletion (Supplementary Figs 1 and 2). On the basis of these observations, it is likely that SNPs we identified through our GWA represent 'markers' of *GSTM1* and *GSTT1* deletion.

In summary, we evaluated the relationship between genetic variants and expression using a whole-genome approach with a group of pharmacogenes. We identified two PGRN-VIPs (*GSTT1* and *GSTM1*) whose expression is significantly associated with SNPs within 2.5 Mb of the genes. In addition, we identified distant-acting elements that were significantly associated with the expression of three and 10 PGRN-VIPs in CEU and YRI, respectively. Three and four PGRN-VIPs harbor SNPs that distantly regulate other gene expressions in CEU and YRI, respectively. Although the exact mechanism for these SNP effects on gene expression is not yet clear, this genome-wide approach identified some previously unknown/unrecognized genetic effects on PGRN-VIP expression. Given the important role that these PGRN-VIPs play in drug therapy, these SNPs may impact drug therapy as well.

The dense genotyping and gene expression data we have made publicly available for the HapMap cell lines allow systematic GWA analysis that would not be possible otherwise. We recognize that there are likely transcriptional, pretranslational, and/or posttranslational modification differences in various tissues. Obtaining panels of tissues from large numbers of individuals remains an important limitation for this type of study [1]. Furthermore, there is a potential cross-hybridization issue when evaluating a gene (e.g. *SULT1A1*, *CYP2A6*, and *CYP2D6*) that shares high homology within gene family. Gene level expression could be affected depending on the number of probe sets that cross-hybridize. Many studies have shown that mRNA gene expression can be used to predict chemotherapeutic drug response [16-18]. Our study explored potential genetic markers for gene expression variation. As DNA genotyping is clinically more practical than gene expression measurements, the identification of genetic determinants for genes whose expression plays a role in drug response/toxicity can be clinically useful as pharmacogenetic markers. Using PGRN-VIPs as an example, we illustrated the utility of genome-wide genotype and expression association findings. However, one may easily utilize this information on any gene or group of genes of interest to identify potential genetic determinants of gene expression and to explore the genetic contributions to drug therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors are grateful for excellent technical support provided by Dr Jeong-Ah Kang in maintaining the cell lines and T.A. Clark, T.X. Chen, A.C. Schweitzer, J.E. Blume (Expression Research, Affymetrix Laboratory, Affymetrix Inc., Santa Clara, CA 94706) for their contribution in generating exon array data. This Pharmacogenetics of Anticancer Agents Research (PAAR) Group (<http://pharmacogenetics.org>) study was supported by NIH/NIGMS grant UO1GM61393 and UO1GM61374 (<http://pharmgkb.org/>).

References

1. Pastinen T, Ge B, Hudson TJ. Influence of human genome polymorphism on gene expression. *Hum Mol Genet* 2006;15:R9–R16. [PubMed: 16651375]
2. Huang RS, Duan S, Shukla SJ, Kistner EO, Clark TA, Chen TX, et al. Identification of genetic variants contributing to cisplatin-induced cytotoxicity using a genome-wide approach. *Am J Hum Genet* 2007;81:427–437. [PubMed: 17701890]
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]
4. Huang RS, Kistner EO, Bleibel WK, Shukla SJ, Dolan ME. Effect of population and gender on chemotherapeutic agent-induced cytotoxicity. *Mol Cancer Ther* 2007;6:31–36. [PubMed: 17237264]
5. Abecasis G, Cardon L, Cookson W. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000;66:279–292. [PubMed: 10631157]

6. Carlson C, Eberle M, Rieder M, Qian Y, Kruglyak L, Mickerson D. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120. [PubMed: 14681826]
7. Westfall, PH.; Young, SS. *Resampling-based multiple testing: examples and methods for P-value adjustment.* Wiley; New York: 1993.
8. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2005.
9. de Koning D-J, Haley CS. Genetical genomics in humans and model organisms. *Trends Genet* 2005;21:377–381. [PubMed: 15908034]
10. McIlwain C, Townsend D, Tew K. Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* 2006;25:1639–1648. [PubMed: 16550164]
11. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007;39:226–231. [PubMed: 17206142]
12. Storey J, Madeoy J, Strout J, Murfel M, Ronald J, Akey J. Gene-expression variation within and among human populations. *Am J Hum Genet* 2007;80:502–509. [PubMed: 17273971]
13. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008;82:631–640. [PubMed: 18313023]
14. Garte S, Gaspari L, Alexandrie A, Ambrosone C, Autrup H, Autrup J, et al. Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 2001;10:1239–1248. [PubMed: 11751440]
15. Smith G, Stanley L, Sim E, Strange R, Wolf C. Metabolic polymorphisms and cancer susceptibility. *Cancer Surv* 1995;25:27–65. [PubMed: 8718512]
16. Breit S, Stanulla M, Flohr T, Schrappe M, Ludwig W-D, Tolle G, et al. Activating NOTCH1 mutations predict favorable early treatment response and long-term outcome in childhood precursor T-cell lymphoblastic leukemia. *Blood* 2006;108:1151–1157. [PubMed: 16614245]
17. Chung Y, Kim T, Kim D, Namkoong H, Kim H, Ha S, et al. Gene expression signatures associated with the resistance to imatinib. *Leukemia* 2006;20:1542–1550. [PubMed: 16855633]
18. Potti A, Dressman H, Bild A, Riedel R, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 2006;12:1294–1300. [PubMed: 17057710]

Fig. 1.

The differentially expressed pharmacogenetic research network-very important pharmacogenes between CEU (open) and YRI (grey). CEU ($n = 87$), YRI ($n = 89$). Significant expression differences were indicated by both the Westfall–Young permutation approach ($P_c < 0.1$) and the general linear model (false discovery rate less than 0.05). CEU, CEPH, Utah residents with ancestry from northern and western Europe; YRI, Yoruba in Ibadan.

Fig. 2.

Significant local association between rs366631 genotype and *GSTM1* gene expression. (a) CEU samples represented by open squares; (b) YRI samples represented by closed circles. TT genotype is associated with lower *GSTM1* expression in both samples. The higher C allele frequency in YRI is likely contributing to the higher *GSTM1* expression in YRI. The number of cell lines per genotype is shown directly above the genotype. *P* values for association of genotype with expression are $P = 3 \times 10^{-21}$ and $P = 1 \times 10^{-22}$ for CEU and YRI, respectively. CEU, CEPH, Utah residents with ancestry from northern and western Europe; YRI, Yoruba in Ibadan.

Table 1

Summary of PGRN-VIPs expressed in LCLs

PGRN-VIP gene symbol	Differentially expressed in CEU and YRI	Mean log ₂ (gene expression)		Regulated by SNPs	Host SNP regulators
		CEU	YRI		
<i>ABCC1</i>		9.03	8.90		
<i>ADRB1</i>	+	8.32	8.51	+	
<i>ADRB2</i>	+	5.92	5.93		
<i>AHR</i>		8.75	8.67		
<i>ALOX5</i>		8.98	8.98	+	
<i>BRCA1</i>		7.49	7.66	+	
<i>COMT</i>	+	9.01	8.91		
<i>CRHR1</i>		5.84	5.81	+	
<i>CYP1A1</i>		5.59	5.55		
<i>CYP1A2</i>		6.31	6.31		
<i>CYP2A6</i>		5.55	5.56	+	
<i>CYP2D6</i>		6.47	6.43		
<i>DRD2</i>		5.72	5.68		+
<i>ESR1</i>		5.56	5.54		+
<i>ESR2</i>		5.45	5.40		
<i>G6PD</i>		8.83	8.88	+	
<i>GNB3</i>		6.18	6.17	+	
<i>GSTM1</i>		7.36	8.53		
<i>GSTP1</i>	+	11.67	11.51	+	
<i>GSTT1</i>	+	7.25	6.96		
<i>HLA-B</i>		12.88	12.91		
<i>HMGCR</i>		9.58	9.74	+	
<i>KCNE1</i>		5.44	5.37		
<i>KCNH2</i>		6.28	6.20		
<i>KCNH1</i>		5.46	5.43		
<i>KCNQ1</i>		6.59	6.53		
<i>MTHFR</i>		7.20	7.21		+
<i>NQO1</i>		6.32	6.47	+	
<i>P2RY1</i>		6.83	6.83		
<i>RYR1</i>		5.71	5.66		
<i>SCN5A</i>		5.57	5.53		
<i>SLC19A1</i>		8.08	7.98		
<i>SULT1A1</i>		6.06	6.05		
<i>TGFB1</i>		9.45	9.31		
<i>TNF</i>	+	7.99	7.86		
<i>TPMT</i>		7.02	7.16		
<i>TYMS</i>		10.05	9.97		
<i>VDR</i>	+	6.96	6.54		
<i>VKORC1</i>		6.97	6.92		

Among the 63 PGRN-VIPs, *CHE2* and *HLA-DQB3* are not probed on exon array and *ABCB1*, *ACE*, *ABCC2*, *ADH1A*, *ADH1B*, *ADH1C*, *AGTR1*, *ALDH1A1*, *CYP2C19*, *CYP2C9*, *CYP3A4*, *CYP3A5*, *DPYD*, *EGFR*, *F5*, *HTR2A*, *KCNE2*, *NAT2*, *P2RY12*, *SLCO1B1*, *UGT1A1*, and *UGT2B7* are not expressed in LCLs (based on average signal density greater than 5.34).
CEU, CEPH, Utah residents with ancestry from northern and western Europe; LCLs, lymphoblastoid cell lines; PGRN-VIPs, Pharmacogenetic Research Network-Very Important Pharmacogenes; SNPs, single nucleotide polymorphisms; YRI, Yoruba in Ibadan, Nigeria.