

Characterization of Single Nucleotide Polymorphisms in 13 Members of the ABC Drug Transporter Genes in Three Different Populations

Zihua Wang^{1,2}, Samuel S. Chong⁵ and Caroline G.L. Lee^{*1,3,4}

Departments of ¹Biochemistry, and ⁵Pediatrics & Obstetrics/Gynecology, ²Graduate Programme in Bioengineering, National University of Singapore, Singapore

³Division of Medical Sciences, Humphrey Oei Institute of Cancer Research, National Cancer Center, Singapore

⁴DUKE-NUS Graduate Medical School, Singapore

Abstract: To evaluate the usefulness of the two public databases, HapMap and Perlegen, in facilitating studies associating polymorphisms in these genes with drug response, we examined 111 single-nucleotide-polymorphisms from 13 ABC-transporter genes in Singaporean-Chinese, European-Americans and African-Americans. We found that genotype data from the HapMap/Perlegen databases are generally transferable to different sampling of the same population and to similar populations residing elsewhere. However, not all ABC-transporter family genes are amenable to SNP-tagging due to the low tagging-efficiency of low-LD genes resulting in negligible cost-savings. Hence, alternative approaches may have to be explored for low-LD ABC genes.

Keywords: SNP, haplotype, LD, ABC transporter genes, HapMap, Perlegen.

INTRODUCTION

The ATP-binding cassette (ABC) superfamily of transporters represents one of the largest families of membrane proteins in both prokaryotic and eukaryotic organisms [1]. Best known as mediators of resistance to anti-cancer drugs, these transporters may have evolved to protect the cells from external insults through the recognition and energy-dependent removal of diverse drugs and naturally toxic agents. Hence, these ABC transporters are likely to play important roles in determining drug response.

At least 48 members belonging to this superfamily of transporters have been identified through sequence homology and grouped into seven subfamilies ranging from A to G. Functions attributed to these transporters include the extrusion of pharmacologically relevant compounds; export of physiological substrates (e.g. cholesterol (ABCA1)) and the regulation of chloride transport (e.g. ABCC7/CFTR) [1]. Although the functions or the physiological/pharmacological substrates of many of the ABC transporters remain unknown, it is likely that they may also be drug exporters and play important roles in regulating drug response due to their high sequence homology with known drug exporter members of this superfamily. Polymorphisms in members of this superfamily have been variously associated with differences in protein expression, function as well as drug disposition and response [2, 3]. Hence, genetic variations in the ABC transporter genes are likely to contribute to inter-individual differences in drug response.

Two publicly available databases HapMap (<http://www.hapmap.org/>) and Perlegen (<http://genome.perlegen.com/>) examined genome-wide single nucleotide polymorphisms

(SNPs) in several populations. The HapMap project aims to systematize the indirect approach through a genomewide study of haplotype-block structure in four major populations namely Chinese, Japanese, Caucasian and Yoruba and genome-wide tagging SNPs using either pairwise or multi-marker r^2 strategy can be identified [4, 5]. HapMap genotyped already known SNPs from public databases at a density of approximately one SNP per 1 kb of DNA from unrelated individuals in 4 different populations, namely, 45 Japanese from Tokyo, 45 Chinese from Beijing (CHB), 60 US residents (parents of trio data) with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH/CEU) and 60 Yoruba people (parents of trio) of Ibadan (YRI). Approximately 1.6 million SNPs from 24 Han Chinese from Los Angeles (CHLA), 24 European-American (EA) and 23 African-American (AA) was successfully genotyped in the Perlegen project. The SNPs genotyped in the Perlegen project were either reported in public databases or identified through their array-based re-sequencing of 24 human samples of diverse ancestry [6, 7].

Although there were previous reports examining the transferability of data from the HapMap or the Perlegen projects to similar populations residing in other regions of the world, these studies either focus on a single population, such as Spanish [8-12] or they focus on the portability of Tag SNPs to the similar populations or cross populations. Notably, Johansson *et al.* reported that the transferability of common tagSNPs to capture random sequence variation between such as CEU of HapMap and Sami are seriously limited [13].

Another database known as PharmGKB (the Pharmacogenetics and Pharmacogenomics Knowledge Base <http://www.pharmgkb.org/>) focusing on variation in drug response genes is also publicly available. However, SNPs genotyped in this database are primarily exonic or within the flanking intronic regions resulting in large gaps between some SNPs,

*Address correspondence to this author at the Department of Biochemistry, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore; Tel: 65-6436-8353; Fax: 65-6779-1453; E-mail: bchleec@nus.edu.sg

thus, making this database less useful for the identification of tSNPs.

To facilitate studies associating ABC transporter genes with drug response, we examined 111 SNPs in 13 ABC transporter genes known to play some roles in the transport of physiological substrates or drugs (Supplementary Table 1) in three different populations that are similar to populations examined by the HapMap/Perlegen databases. The three populations examined in this study are the Singaporean-Chinese originating from Southern China (CHS) as well as the European-Americans (EA) and African-Americans (AA) from the United States of America (USA) which are similar to the CHB, CEU and YRI in HapMap and the CHLA, EA and AA in Perlegen. We compared the SNP allele frequencies, the haplotype and LD profiles as well as tSNPs in the selected genes of our three populations with similar populations from the Perlegen and HapMap databases.

MATERIALS AND METHODS

Study Populations

The study populations examined in this study are the CHS, EA and AA. These populations are similar to the populations examined in the HapMap project except that the CHB examined in the HapMap project resides in China instead of Singapore and the YRI population in HapMap is the African population residing in Yoruba, Africa instead of the USA. Only the CEU population in HapMap and the EA in this study reside in the same country and will thus represent a different sampling of the same population. The EA population that we examined represents a re-sampling of the same population examined in the Perlegen database while the CHS population examined in this study represents similar population residing in a different region from the CHLA examined in the Perlegen database.

The genomic DNA of the EA and AA populations were purchased from the Coriell Cell Repositories (Camden, NJ). Anonymized samples for the CHS population were obtained from previously archived genomic DNAs extracted from unselected cord blood samples discarded after clinical newborn screening for glucose-6-phosphate dehydrogenase deficiency from the National University Hospital, one of two major hospitals in Singapore. National University of Singapore Institutional Review Board (NUS-IRB Reference Code: 04-126E) has granted ethical approval for this study. A total of 136 unrelated individuals were examined including 40 CHS, 46 EA and 50 AA.

Genes and SNPs Selection

We chose to examine 111 SNPs within the genes of 13 members of ABC superfamily known to play some roles in the transport of physiological substrates or pharmacologically relevant drugs (Supplementary Table 1). SNPs from each gene were from two sources: first, the SNPs databases: the Single Nucleotide Polymorphism database (dbSNP) from the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), the SNP Consortium LTD (<http://snp.cshl.org>) and Japanese SNP database (<http://snp.ims.u-tokyo.ac.jp>); and second, SNPs reported in the literature [14-16]. To characterize a set of SNPs that would be informative in further association studies for each drug transporter gene, SNPs were selected based on the

following criteria: 1) SNPs should be relatively evenly spaced throughout the genomic region of each gene; 2) maximum distance between consecutive two SNPs should be less than 30 kb; 3) SNPs should preferably be common in human populations (minor allele frequency (MAF) $\geq 5\%$).

Genotyping Method

These 111 SNPs were genotyped using the arrayed primer extension (APEX) assay as previously described [17] (Asper Ltd., Tartu, Estonia). SNPs that were not successfully assayed by APEX were then genotyped either using TaqMan[®] SNP Genotyping Assays or multiplex minisequencing [18].

Determination of Allele Frequency, Haplotype and Linkage Disequilibrium (LD) Profiles

To comprehensively characterize haplotype and LD profiles in these ABC drug transporters, we compared our data with those of HapMap (Release #20) (<http://www.hapmap.org>) and Perlegen (<http://genome.perlegen.com/>) databases. We downloaded the SNP genotype data of the 13 ABC transporter genes from these two databases. HapMap database contains the genotypes of 45 CHB, 60 CEU and 60 YRI samples, respectively. Perlegen database documents the genotypes of 24 CHLA, 24 EA and 23 AA samples, respectively. Data sets that were similar to our study, with respect to the number and distributions of SNPs in each gene (Table 1 and Fig. (1)) were then extracted from these databases and allele frequencies, LD and haplotype frequencies were determined. Pairwise comparison of population allele frequency difference and distribution were tested by Fisher's exact test and Pearson correlation analysis, respectively.

Haplotype frequencies and LD between SNP pairs were estimated as described previously [19]. Half length of LD ($LD_{0.5}$) is defined as the distance at which the average $|D'|$ or r^2 is 0.5.

Identification of tSNPs

The haplotype r^2 method was employed for the identification of tSNPs as r^2 measure of LD is a more relevant measure for association studies and haplotype r^2 may be more efficient in identifying tSNPs compared to the pairwise r^2 method [20]. The multiple-marker haplotype r^2 criteria (Criteria 11 haplotype r^2 in the software), of the "TagIT" program was utilized to identify tSNPs for each gene locus in the three populations which was reported to be useful for association studies based on tSNPs [21]. All genotyped SNPs of the gene with $MAF \geq 5\%$ were examined and the minimum number of tSNPs that satisfied the haplotype r^2 threshold value ≥ 0.85 was identified.

RESULTS AND DISCUSSION

Allele Frequencies of SNPs at the ABC Transporter Gene Loci in Similar Populations Residing in Different Places Do Not Differ Significantly

A total of 111 SNPs from 13 genes of the ABC transporters were genotyped in 136 unrelated individuals representing 3 populations, namely CHS, EA and AA (Table 1). Six SNPs could not be successfully genotyped in all three populations using the APEX method and were then genotyped using TaqMan[®] SNP Genotyping Assays or multiplex minise-

quencing [18]. Four additional SNPs were not successfully genotyped only in the CHS population. These 4 SNPs were thus not included in the analyses of the data from the CHS resulting in the number of SNPs examined for the CHS population (107) being lower than the other two populations (111) (Table 2). The total genomic region that was examined for the 13 genes was approximately 1.38Mb with each gene spanning between 25 and 320 kb in length (Table 1). The average gaps between two SNPs in each of the 13 transporter genes ranges from 2.78 to 20 kb in length (Table 1).

Table 1. Summary of SNPs in the ABC Transporter Genes that were Examined

Genes	Length (kb)	No. of SNPs	Average Inter-SNP Gap(kb)	Chromosome Position
ABCA2	92	8	11.50	9q34.3
ABCB1/4	320	17	18.82	7q21.12
ABCB2/3	25	9	2.78	6p21.3
ABCC1/6	261	19	13.74	16p13.12
ABCC2	70	10	7.00	10q24.2
ABCC3	62	9	6.89	17q21.33
ABCC4	280	14	20.00	13q32.1
ABCC5	100	7	14.29	3q27.1
ABCG1	90	8	11.25	21q22.3
ABCG2	83	10	8.30	4q22
Total	1383	111	-	-

Table 2. Profiles of SNPs at the ABC Transporter Gene Loci in the Three Populations

Population		CHS	EA	AA
Total No. of SNPs		107	111	111
MAF=0	No. of SNPs	15	14	12
	%	14.02	12.61	10.81
MAF≥0.05	No. of SNPs	89	87	89
	%	83.18	78.38	80.18
MAF≥0.1	No. of SNPs	85	74	77
	%	79.44	66.67	69.37
Inter-SNP Gaps	No. of SNPs	96	100	100
	Average size (kb)	14.39	13.81	13.81
	Gaps <30 kb (%)	91.67	92.00	92.00

As evident in Fig. (1), the SNPs examined are distributed relatively evenly across the genes. As shown in Table 2, there are a total of 96-100 inter-SNP intervals encompassing the 13 genes in the 3 populations. The average distance be-

tween two adjacent SNPs was ~14 kb in the 3 populations. Greater than 90% of these transporter genes genomic regions were covered by inter-SNP gaps of less than 30 kb (Table 2). More than 75% and 65% of the SNPs examined had MAFs of greater 5% and 10%, respectively, in the 3 populations (Table 2). Less than 15% of the SNPs examined in this study were monomorphic (Table 2).

The three populations examined in this study are the Singaporean-Chinese originating from Southern China (CHS) as well as the European-Americans (EA) and African-Americans (AA) from the United States of America (USA) which are similar to the CHB, CEU and YRI in HapMap and the CHLA, EA and AA in Perlegen. Similar to this study, the EA and AA samples examined by the Perlegen database were also obtained from the Coriell Repository. Of the 24 EA samples examined by Perlegen, only one sample was identical to the samples examined in this study. This sample was thus not included when we compared data from our study with those from Perlegen. With the removal of the single common sample, the EA population examined in this study represents a different sampling of the EA population in the Perlegen database as well as the CEU population from HapMap as all three EA/CEU populations reside in the USA. All of the AA samples examined in the Perlegen database were also examined in this study hence we did not compare the AA data from Perlegen with our study. SNPs from identical samples from Perlegen that were examined in this study generally has similar genotypes. The AA population from this study thus represents a similar population residing in a different country compared to the YRI population in the HapMap database. The CHS population in this study is similar to the CHB/CHLA populations in the HapMap and Perlegen databases in that they originated from the Han Chinese but they are not exactly the same since they have been residing in a different country (namely, Singapore). The transferability of the SNP data at the ABC transporter gene loci from the HapMap or Perlegen databases to different sampling of the same population or to similar populations residing elsewhere in the world would be noteworthy as it has important implications in studies associating these genes with drug response or complex diseases.

We thus evaluated how the allele frequencies of SNPs in the ABC transporters in the 3 populations examined in this study compare with those from similar populations in the HapMap and Perlegen databases (Fig. 2, Supplementary Table 2). More than sixty-three percent of the SNPs examined in this study were also found in the HapMap database while only <38% were common with the Perlegen database (Supplementary Table 2A). Less than 17% and 12% of these common SNPs in the HapMap and Perlegen databases, respectively, had MAFs of less 5% (Supplementary Table 2A). Of the polymorphic SNPs that were common between our study and those of HapMap or Perlegen, we found that in the CH (CHS/CHB/CHLA) population, approximately two times more SNPs in our data differed by more than 10% from those of the HapMap (22.1%) and Perlegen (20.0%) than between the two databases themselves (11.1%) (Supplementary Table 2A). On the other hand, in the AA/YRI population, although this population from our study is different from the YRI population, reported in the HapMap database, who reside in Africa, we found greater number of SNPs that

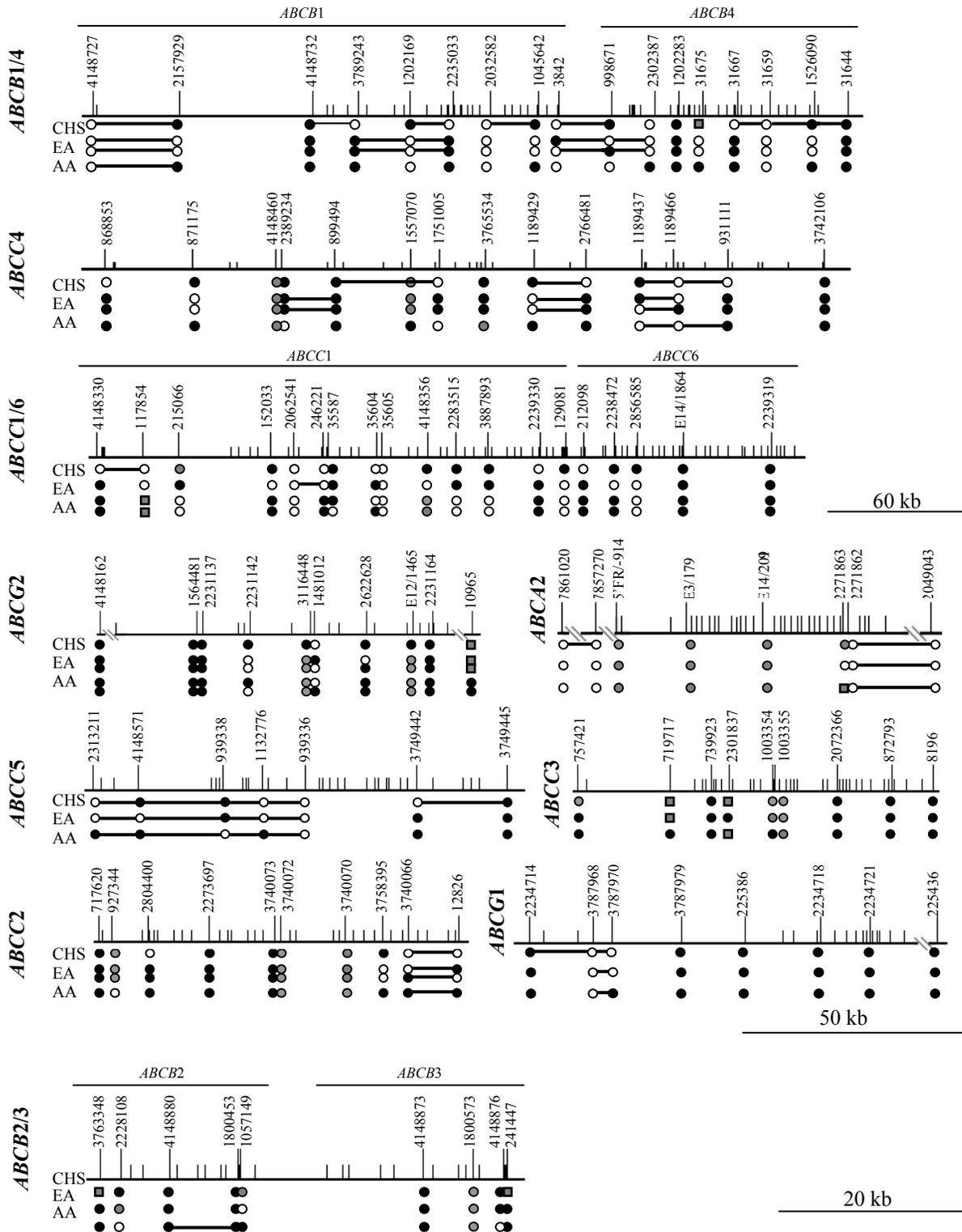


Fig. (1). Graphical representation of the SNPs, LD blocks and tSNPs at the ABC transporter genes. Due to great differences in the genomic size of the various ABC genes, the 13 genes were grouped into three panels with different scaling of the genes as indicated. Top panel includes the *ABC1/4*, *ABCC4* and *ABCC1/6* genes; the bottom panel comprises the *ABC2/3* genes while the middle panel contains the other genes. For each gene, the middle horizontal line represents the genomic region. Two types of vertical lines are drawn above the horizontal line with the location scaled to the physical position in the genes: the shorter lines represent the exons while the longer lines represent the SNPs that were genotyped in this study (with the dbSNP ID or “rs” number indicated wherever applicable). The LD blocks in which $r^2 \geq 0.3$ and tagging SNPs of the genes in the three populations are also shown. The black, white, gray circles and boxes below the horizontal line represent tagging SNPs, non tagging SNPs, monomorphic SNPs and SNPs without genotypic data or with too many missing data, respectively. For some populations of some genes, e.g. EA of *ABC1/4*, there are two rows representing two alternative sets of tagging SNPs with the same maximum haplotype r^2 value. Lines drawn between the circles represent LD blocks, where $r^2 \geq 0.3$.

differed by >10% between the HapMap and Perlegen databases (20.6%) than between our database and the HapMap database (15.5%) (Supplementary Table 2A). Although the EA/CEU population from our study and the HapMap/Perlegen databases are all residing in the USA, more SNPs differed by >10% between Perlegen and our study/HapMap (26.2% and 18.9%) than between our study and HapMap (12.3%). The percentages of SNPs with significantly different allele frequencies were less than 8.5% for all pairwise comparisons (Fisher's exact p-value < 0.01) [9] (Supplementary Table 2B). The distribution of allele frequencies in the respective populations between our study and HapMap/Perlegen databases was very similar as evident by the high Pearson correlation (≥ 0.94) for all comparisons. Hence, we can infer from the above that SNP allele frequencies at the ABC gene loci do not differ significantly in similar populations whether they are residing in the same country or not.

LD is Strongest at the *ABCC5* Locus but Weakest at the *ABCC3* Locus

LD is the non-random association between the alleles at one locus with those at adjacent loci. Alleles that are in LD

are more highly correlated than would be expected of them if they are segregating independently in a population [22]. Several factors can have an impact on LD, including mutation, recombination, genetic drift, admixture or migration and natural selection, etc. One useful application that can be derived from understanding the LD of a gene loci is the ability to select a subset of highly informative SNPs known as tagging SNPs (tSNPs) based on the LD profile within the gene locus to represent the most common variations segregating at the gene locus which would facilitate more efficient association studies [23]. Hence, understanding the LD profiles of the ABC transporter genes may facilitate rational approaches to studies associating these genes with functional changes including drug response.

Except for *ABCA2* where there were too few SNPs of $MAF \geq 5\%$, the LD profiles of SNPs with $MAF \geq 5\%$ at the 12 ABC transporter gene loci in the three populations were analyzed. $r^2 \geq 0.3$ is defined as relatively high LD. As shown in Fig. (1), the LD profiles differed greatly amongst the different ABC transporter genes and the different populations.

All three populations had one LD block that contains at least 5 SNPs encompassing greater than 50 kb of genomic

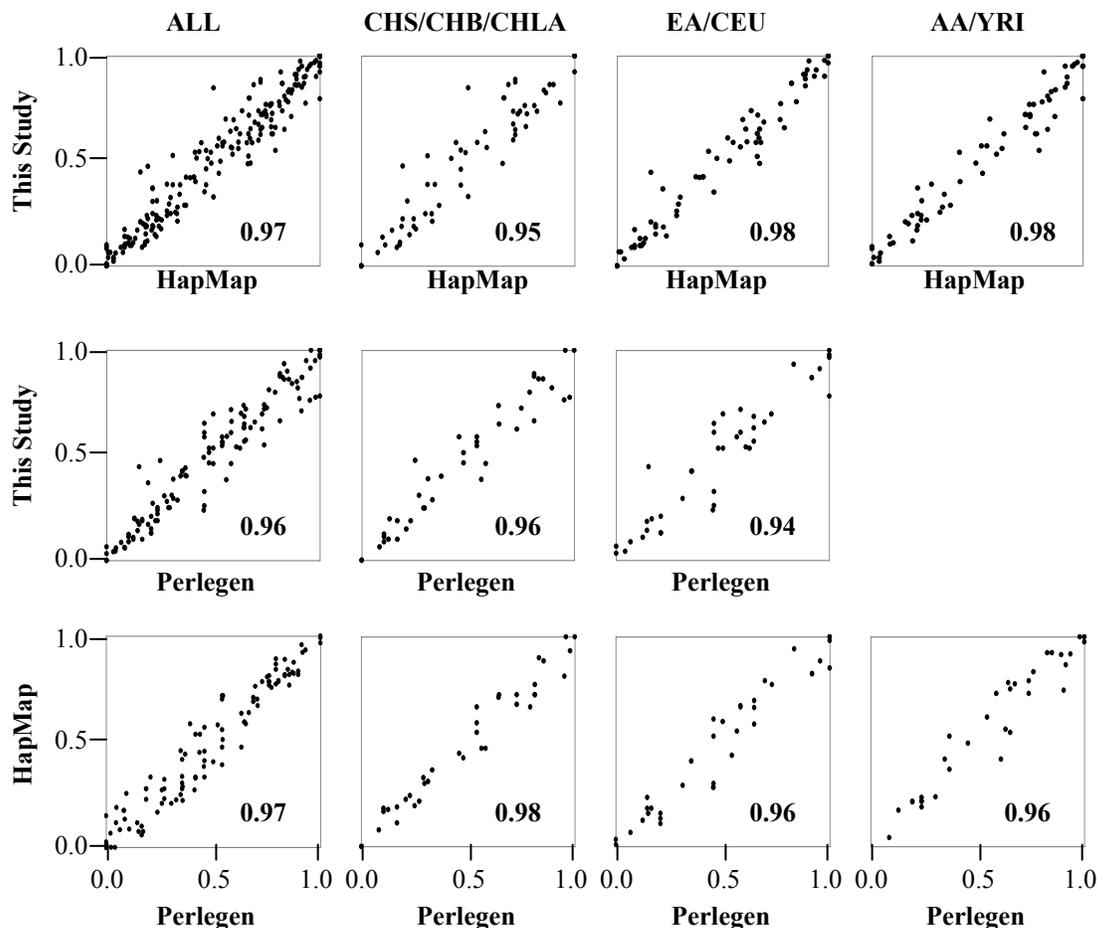


Fig. (2). Pairwise comparison of population allele frequencies of SNPs at the ABC transporter loci between this study, HapMap and Perlegen databases. Scatter plots of pairwise comparison of the distribution of allele frequencies of SNPs at the ABC transporter gene loci between CHS/CHB/CHLA, EA/CEU and AA/YRI amongst the three datasets (this study, HapMap and Perlegen). Pearson correlation coefficient is shown at the bottom right corner of each plot.

DNA (Fig. 1). The CHS population had the greatest percentage of SNPs (>40% vs <30%) residing in r^2 LD blocks comprising 2-4 SNPs while the AA population had the greatest percentage of SNPs (>75%) that did not reside in any LD block. These observations that the AA population displayed the weakest LD compared to the CHS or EA populations were consistent with previous reports on some members of the ABC transporter genes [19, 24, 25] as well as on a genome-wide scale [6].

ABCC5 showed the strongest r^2 LD of the 13 ABC genes with at least 5 SNPs residing in one LD block and the other two SNPs in another LD block in the CHS population (Fig. 1). Similar strong LD at the *ABCC5* gene was reported earlier [24]. *ABCB1/4* gene loci displayed the second strongest r^2 LD profiles (Fig. 1). Interestingly, no r^2 LD ≥ 0.3 was observed for *ABCC3* and *ABCG2*. Overall, r^2 LD at the 13 ABC transporter gene loci was generally weak except for *ABCC5*, *ABCB1/4* and perhaps *ABCC4* (Fig. 1). These observations have important implications for association studies. It suggests that for genes (e.g. *ABCC5*, *ABCB1/4* or *ABCC4*) which reside within genomic regions of strong LD, where several SNPs have high associative power, the identification of tSNPs may be useful to facilitate efficient association studies as fewer SNPs need to be examined. On the other hand, for most of the other ABC genes (e.g. *ABCC3*), which reside in genomic regions of low LD, more SNPs need to be examined initially in order to establish an association with functional differences and it may thus not be practical to utilize tSNPs to facilitate efficient association studies. Hence, alternative approaches may need to be explored to facilitate efficient association studies for such genes.

The LD and Haplotype Profiles at the ABC Transporter Gene Loci in Similar Populations Residing in Different Places are Comparable

We next evaluated whether the LD and haplotype profiles of the ABC transporter gene loci were comparable in different sampling of the same populations or in similar populations residing in different places. Pair-wise LD between SNPs was evaluated using 2 statistical measures, $|D'|$ and r^2 [26]. Haplotype profile was examined by determining the number of haplotypes that could constitute greater than 60% of the total chromosomes at each gene locus in the different populations.

As shown in Table 3, although there is variability in the LD (especially $|D'|$) and haplotype profiles amongst the three different datasets (this study, HapMap and Perlegen), the trends observed are reasonably similar. The high variability observed especially in the LD profiles could perhaps be due to SNP choice and small sample size which LD measures are particularly sensitive to [27]. Nonetheless, consistent with earlier reports [19, 24, 25, 27], r^2 LD was generally lower than $|D'|$ LD across all genes, all populations and all datasets (Table 3). Also consistent with earlier observations, the AA/YRI population generally had lower LD and higher haplotype diversity, as more haplotypes were required to account for 60% of their chromosomes compared to the CHS/CHLA/CHB or EA/CEU populations in all 3 datasets. Also generally consistent across the three datasets was the observation that *ABCC5* had very strong LD and required

fewer haplotypes to constitute 60% of the total chromosomes (Table 3). On the other hand, *ABCC3* and *ABCG1* gene loci had very low LD across all 3 datasets.

Hence, consistent across all 3 datasets were the observations of strong LD for the *ABCC5* and *ABCB1/4* gene loci and weak LD for the *ABCC3* and *ABCG1* gene loci suggesting that less SNPs may be needed for studies associating *ABCC5* or *ABCB1/4* but not *ABCC3* or *ABCG1* with drug response or disease susceptibility.

Fewer Tagging SNPs (tSNPs) are Required for the *ABCC5* Gene Locus which has Stronger LD

The identification of tSNPs to represent other SNPs (including those as yet unidentified) will facilitate efficient association studies. Currently, there are several methods to select a subset of tSNPs that will capture most of the information of the entire set of SNPs present in the genomic region [21, 23, 28]. These methods can be generally classified into two groups: (i) tSNPs based on haplotype diversity – a subset of SNPs are selected to capture the original haplotype diversity present in the set of SNP that were genotyped; (ii) tSNPs based on association – a subset of SNPs are selected that are highly associated with other SNPs in the genomic region including SNPs that are not genotyped or identified. This second method for selecting tSNPs is our method of choice since it has direct relevance in association studies. We thus utilized the TagIT algorithm which selects tSNPs based on the haplotype r^2 criteria [21] to select for a subset of tSNPs from all polymorphic SNPs in the 12 ABC transporter genes except *ABCA2* which was mainly monomorphic.

Fig. (1) displays the identity of the tagging SNPs (as black circles) in each gene for the 3 populations examined, while Supplementary Table 3 gives a summary of the profiles of tagging SNPs in the 12 ABC transporter genes in the 3 populations. The AA population which had the weakest LD and greatest haplotype diversity (Table 3) had the weakest tagging efficiency and shortest intervals between tSNPs of the 3 populations (Supplementary Table 3). Also evident from our results are the observations that in populations that generally show high LD, genes with higher LDs have longer intervals between tSNPs and better tagging efficiency. For example in the CHS and EA populations, tagging efficiency at *ABCC5* and *ABCB1/4* gene loci exceeded 2-fold. On the other hand, for *ABCC3* gene loci where the LD was very weak, tagging efficiency was 1. Our results thus suggest that not all the ABC transporter gene loci are amenable to tagging SNP strategy to facilitate efficient studies associating polymorphisms in these genes with drug response and alternative approaches have to be explored.

Tagging Profiles of SNPs at the ABC Transporter Gene Loci are Comparable Across the Three Datasets

To compare the tagging profiles of SNPs at the ABC transporter gene loci, we selected the same or nearby polymorphic SNPs with MAF>0 from the HapMap and Perlegen databases for determination of tSNPs for gene loci *ABCB1/4*, *ABCC3*, and *ABCC5* using the same haplotype r^2 approach. *ABCB1/4* and *ABCC5* gene loci were selected to represent loci with high LD and good tagging efficiency while *ABCC3*

Table 3. Comparison of LD, Haplotype Profiles of the ABC Transporters in this Study Versus HapMap/Perlegen Databases

Data	Genes	No. of SNPs*			Half Length of D'			Half Length of r ²			No. of Haplotypes**		
		CH ***	EA/ CEU	AA/ YRI	CH	EA/ CEU	AA/ YRI	CH	EA/ CEU	AA/ YRI	CH	EA/ CEU	AA/ YRI
This Study	<i>ABCB1/4</i>	16	16	15	113.7	135.1	96.6	14.5	9.8	2.9	11	8	17
	<i>ABCB2/3</i>	5	6	6	16.8	13.7	13.4	0.9	0.0	1.3	3	5	4
	<i>ABCC1/6</i>	18	18	17	45.6	43.6	41.2	3.4	5.8	3.9	16	22	22
	<i>ABCC2</i>	7	6	5	177.7	125.7	57.8	5.5	14.4	6.7	3	3	4
	<i>ABCC3</i>	4	4	7	24.3	22.9	27.0	2.3	4.5	1.6	3	4	5
	<i>ABCC4</i>	12	10	12	73.8	63.6	64.9	15.7	17.2	9.3	11	7	16
	<i>ABCC5</i>	7	7	7	778.9	839.5	209.9	28.1	58.7	21.8	2	2	2
	<i>ABCG1</i>	8	8	8	18.6	14.3	17.9	5.3	3.5	2.5	9	12	8
	<i>ABCG2</i>	8	7	8	50.7	127.9	34.6	4.1	5.4	6.8	5	2	5
	Average				156.2	157.3	66.1	9.5	14.2	6.3	7	8	10
HapMap	<i>ABCB1/4</i>	17	17	17	165.3	297.3	96.2	16.5	17.1	8.6	19	7	21
	<i>ABCB2/3</i>	9	9	9	25.9	9.5	38.0	3.6	2.0	2.2	3	3	4
	<i>ABCC1/6</i>	19	19	19	50.2	44.7	86.6	8.9	9.7	6.8	18	17	18
	<i>ABCC2</i>	10	10	10	2078.0	1036.7	59.0	46.3	19.7	6.8	3	3	3
	<i>ABCC3</i>	8	8	8	22.7	34.2	24.1	0.5	0.8	0.1	5	6	8
	<i>ABCC4</i>	14	14	14	60.3	78.8	47.7	9.6	2.9	3.2	13	14	21
	<i>ABCC5</i>	8	8	8	1482.3	1669.5	208.6	8.5	25.0	11.1	3	2	3
	<i>ABCG1</i>	8	8	8	12.5	21.1	18.6	4.3	3.6	0.1	8	11	8
	<i>ABCG2</i>	8	7	8	55.3	181.7	43.4	8.7	11.2	2.4	4	5	4
	Average				439.2	374.8	69.2	11.9	10.2	4.6	8	8	10
Perlegen	<i>ABCB1/4</i>	18	18	18	107.1	109.1	90.5	27.5	23.2	13.7	5	4	12
	<i>ABCB2/3</i>	7	8	8	26.9	30.7	21.1	2.7	2.9	3.2	3	3	4
	<i>ABCC1/6</i>	19	19	19	154.5	119.6	87.7	15.9	8.3	9.9	8	9	9
	<i>ABCC2</i>	10	9	10	78.7	125.2	163.4	28.3	23.9	13.1	3	2	3
	<i>ABCC3</i>	8	9	9	36.3	43.4	28.1	1.0	1.4	1.0	4	5	6
	<i>ABCC4</i>	14	14	14	66.8	89.7	95.0	8.7	10.7	1.7	12	9	14
	<i>ABCC5</i>	8	8	8	808.6	793.9	317.1	30.7	48.4	25.1	2	2	3
	<i>ABCG1</i>	8	8	8	22.1	33.1	27.1	5.9	8.0	3.1	7	6	7
	<i>ABCG2</i>	9	9	9	61.9	142.2	38.6	5.6	0.3	2.0	4	3	7
	Average				152.6	165.2	96.5	13.6	14.1	8.1	5	5	7

Note: *The numbers of SNPs with MAF≥0.05 at each gene locus in three populations. ** This represents the number of haplotypes that can account for >60% of total chromosomes at each gene locus in the respective populations. *** CH represents CHS, CHB and CHLA in our study, HapMap and Perlegen, respectively.

Table 4. Comparison of the Profiles of tSNPs Amongst the Three Datasets

Genes	Populations	This Study			HapMap			Perlegen		
		a	b	c	a	b	c	a	b	c
<i>ABCB1/4</i>	CH	16	8	2.0	17	8	2.1	18	7	2.6
	EA/CEU	17	7	2.4	17	7	2.4	18	5	3.6
	AA/YRI	17	11	1.5	17	7	2.4	18	7	2.6
<i>ABCC3</i>	CH	4	4	1.0	8	7	1.1	8	6	1.3
	EA/CEU	6	6	1.0	8	6	1.3	9	6	1.5
	AA/YRI	7	7	1.0	8	7	1.1	9	6	1.5
<i>ABCC5</i>	CH	7	3	2.3	8	4	2.0	8	2	4.0
	EA/CEU	7	3	2.3	8	3	2.7	8	2	4.0
	AA/YRI	7	5	1.4	8	5	1.6	8	5	1.6

Note: a, number of SNPs (MAF>0); b, number of tSNPs; c, tagging efficiency.

was selected to represent locus with low LD and poor tagging efficiency. As evident in Table 4, while the tagging efficiencies of the various genes in the 3 datasets were different, the trends were similar with *ABCC5* and *ABCB1/4* having good tagging efficiency and *ABCC3* having poor tagging efficiency for all 3 datasets. The tagging efficiency from our dataset seemed to be more similar to the HapMap dataset while the tagging efficiency from the Perlegen dataset seemed to be the highest amongst the three datasets.

CONCLUSION

To address the issue of whether genotype data at the ABC transporter gene loci from the HapMap and Perlegen databases are transferable to a different sampling of the same population residing in the same location as well as to a similar population residing in a different region of the world, we examined 111 SNPs from 13 genes of the ABC transporters in 136 unrelated individuals representing 3 populations, namely CHS, EA and AA. The CHS that we studied resides in Singapore and would represent a population that is similar to that examined in the HapMap (CHB) and Perlegen (CHLA) databases but residing in a different region of the world. The EA population in this study resides in the USA and would thus represent a different sampling of the same population examined in the HapMap (CEU population residing in the USA) as well as the Perlegen (EA residing in the USA) databases. The AA population in this study resides in USA and would thus represent a similar population as the population studied in the HapMap (YRI from Africa) database but residing in a different region of the world. We observed high similarity of allele frequencies and similar trends of LD and haplotype profiles across all 3 datasets for the respective ABC transporter genes in the respective populations. The AA/YRI population generally had the weakest LD and greatest haplotype diversity across all three datasets. The *ABCC5* and *ABCB1/4* gene loci had the strongest LD while the *ABCC3* gene locus had the weakest LD. Additionally, it was also observed across all three datasets that the tagging

efficiency of genes with strong LD e.g. *ABCC5* was considerably better than genes with weak LD e.g. *ABCC3*.

Hence our study shows that the genotype data from the HapMap and Perlegen databases are generally transferable to different sampling of the same populations as well as to similar populations residing elsewhere in the world. Importantly, our study also highlights an important limitation of utilizing SNP tagging to reduce genotyping cost in association studies. We demonstrated that not all genes were amenable to SNP tagging. For example for genes with weak LD, e.g. *ABCC3*, the tagging efficiency is so low that there will not be much cost savings utilizing tagging SNP strategy for these gene loci. Hence, other strategies have to be explored to facilitate efficient study associating these low LD genes with phenotype like drug response or complex disease susceptibility. These strategies may include utilizing statistical and computational approaches to identify SNPs that are likely to have evolutionary and/or biological functional significance. For example, the functional consequences of non-synonymous SNPs from the HapMap data can be predicted using computational approaches. Additionally, we have recently demonstrated the feasibility of utilizing genomic signatures of recent positive selection in the major ABC transporter genes to identify functionally important SNPs [25, 19, 29]. These SNPs are likely to be positively selected during the adaption of modern man to the selective pressures in the recent human evolution [30]. Hence, these SNPs are likely to be functionally significant and have the potential of facilitating efficient association studies.

ACKNOWLEDGEMENTS

This study is supported by a BioMedical Research Council Grant (04/1/21/19/310) to S. S. Chong and C. G. L. Lee through the National University of Singapore as well as a block grant from the National Medical Research Council (NMRC), Singapore to C.G.L.Lee at the National Cancer Center, and an Academic Research Fund – Ministry of Education Tier 2 grant to C.G. L. Lee through the National University of Singapore.

Supplementary Table 1. Characteristics of Selected ABC Superfamily Drug Transporters

Systematic Name	Common Name	Tissue	Physiological Substrates	Drugs	Ref.
ABCA2	ABCA2	Brain, monocytes	Steroids derivatives, lipids	Estramustine	[1, 2]
ABCB1	PGP/ MDR1	Intestine, liver, kidney, placenta, BBB*	Phospholipids, neutral and cationic organic compounds	Anthracyclines, vinca alkaloids, epipodophyllotoxins, antibiotics and others	[3-5]
ABCB2	TAP1	Intestine, prostate, respiratory tract, breast	Peptides	Mitoxantrone, epipodophyllotoxins	[6, 7]
ABCB3	TAP2	Intestine, prostate, respiratory tract, breast	Peptides	Mitoxantrone, epipodophyllotoxins	[6, 7]
ABCB4	MDR2	Liver	Phosphatidylcholine	Paclitaxel, vinca alkaloids	[8, 9]
ABCC1	MRP1	Most tissues	Glutathion-, and other conjugates, organic anions, leukotrienes	Anthracyclines, vinca alkaloids, epipodophyllotoxins, methotrexate	[10-12]
ABCC2	MRP2/ cMOAT	Liver, kidney, intestine	Glutathion-, and other conjugates, organic anions, leukotriene C4	Platin-drugs, anthracyclines, vinca alkaloids, epipodophyllotoxins, camptothecins, methotrexate	[13-15]
ABCC3	MRP3	Pancreas, kidney, intestine, liver, adrenal glands	Glucuronides, bile salts, peptides	Vinca alkaloids, epipodophyllotoxins, methotrexate	[16, 17]
ABCC4	MRP4	Prostate, testis, ovary, intestine, pancreas, lung	Organic anions, nucleotide analogues	Nucleotide analogues, methotrexate	[18-20]
ABCC5	MRP5	Most tissues	Organic anions, cyclic nucleotides	Nucleotide analogues	[21, 22]
ABCC6	MRP6	Liver, kidney	Anionic cyclic	Etoposide, teniposide, doxorubicin, daunorubicin	[23, 24]
ABCG1	White/ ABC8	Ubiquitous	Sterols, lipids	doxorubicin	[25, 26]
ABCG2	MXR/ BCRP/ ABC-P,	Placenta, intestine, breast, liver	Prazosin	Mitoxantrone, anthracyclines, camptothecins, topotecan	[27, 28]

- Vulevic, B., Chen, Z., Boyd, J.T., Davis, W., Jr., Walsh, E.S., Belinsky, M.G. and Tew, K.D. (2001) Cloning and characterization of human adenosine 5'-triphosphate-binding cassette, subfamily A, transporter 2 (ABCA2). *Cancer Res*, **61**, 3339-47.
- Laing, N.M., Belinsky, M.G., Kruh, G.D., Bell, D.W., Boyd, J.T., Barone, L., Testa, J.R. and Tew, K.D. (1998) Amplification of the ATP-binding cassette 2 transporter gene is functionally linked with enhanced efflux of estramustine in ovarian carcinoma cells. *Cancer Res*, **58**, 1332-7.
- Kartner, N., Riordan, J.R. and Ling, V. (1983) Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines. *Science*, **221**, 1285-8.
- Kartner, N., Evernden-Porelle, D., Bradley, G. and Ling, V. (1985) Detection of P-glycoprotein in multidrug-resistant cell lines by monoclonal antibodies. *Nature*, **316**, 820-3.
- Ueda, K., Yoshida, A. and Amachi, T. (1999) Recent progress in P-glycoprotein research. *Anticancer Drug Des*, **14**, 115-21.
- Izquierdo, M.A., Neeffjes, J.J., Mathari, A.E., Flens, M.J., Scheffer, G.L. and Scheper, R.J. (1996) Overexpression of the ABC transporter TAP in multidrug-resistant human cancer cell lines. *Br J Cancer*, **74**, 1961-7.
- Lage, H., Perltz, C., Abele, R., Tampe, R., Dietel, M., Schadendorf, D. and Sinha, P. (2001) Enhanced expression of human ABC-transporter tap is associated with cellular resistance to mitoxantrone. *FEBS Lett*, **503**, 179-84.
- Gottesman, M.M., Fojo, T. and Bates, S.E. (2002) Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature Rev Cancer*, **2**, 48-58.
- Ruetz, S. and Gros, P. (1994) Phosphatidylcholine translocase: a physiological role for the mdr2 gene. *Cell*, **77**, 1071-81.
- Cole, S.P., Bhardwaj, G., Gerlach, J.H., Mackie, J.E., Grant, C.E., Almqvist, K.C., Stewart, A.J., Kurz, E.U., Duncan, A.M. and Deeley, R.G. (1992) Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science*, **258**, 1650-4.
- Borst, P., Evers, R., Kool, M. and Wijnholds, J. (1999) The multidrug resistance protein family. *Biochim Biophys Acta*, **1461**, 347-57.
- Jedlitschky, G., Leier, I., Buchholz, U., Barnouin, K., Kurz, G. and Keppler, D. (1996) Transport of glutathione, glucuronate, and sulfate conjugates by the MRP gene-encoded conjugate export pump. *Cancer Res*, **56**, 988-94.
- Konig, J., Nies, A.T., Cui, Y., Leier, I. and Keppler, D. (1999) Conjugate export pumps of the multidrug resistance protein (MRP) family: localization, substrate specificity, and MRP2-mediated drug resistance. *Biochim Biophys Acta*, **1461**, 377-94.
- Taniguchi, K., Wada, M., Kohno, K., Nakamura, T., Kawabe, T., Kawakami, M., Kagotani, K., Okumura, K., Akiyama, S. and Kuwano, M. (1996) A human canalicular multispecific organic anion transporter (cMOAT) gene is overexpressed in cisplatin-resistant human cancer cell lines with decreased drug accumulation. *Cancer Res*, **56**, 4124-9.
- Cui, Y., Konig, J., Buchholz, J.K., Spring, H., Leier, I. and Keppler, D. (1999) Drug resistance and ATP-dependent conjugate transport mediated by the apical multidrug resistance protein, MRP2, permanently expressed in human and canine cells. *Mol Pharmacol*, **55**, 929-37.
- Kool, M., van der Linden, M., de Haas, M., Scheffer, G.L., de Vree, J.M., Smith, A.J., Jansen, G., Peters, G.J., Ponne, N., Scheper, R.J. et al. (1999) MRP3, an organic anion transporter able to transport anti-cancer drugs. *Proc Natl Acad Sci U S A*, **96**, 6914-9.
- Zeng, H., Bain, L.J., Belinsky, M.G. and Kruh, G.D. (1999) Expression of multidrug resistance protein-3 (multispecific organic anion transporter-D) in human embryonic kidney 293 cells confers resistance to anticancer agents. *Cancer Res*, **59**, 5964-7.
- Schuetz, J.D., Connelly, M.C., Sun, D., Paibir, S.G., Flynn, P.M., Srinivas, R.V., Kumar, A. and Fridland, A. (1999) MRP4: A previously unidentified factor in resistance to nucleoside-based antiviral drugs. *Nat Med*, **5**, 1048-51.
- Chen, Z.S., Lee, K. and Kruh, G.D. (2001) Transport of cyclic nucleotides and estradiol 17-beta-D-glucuronide by multidrug resistance protein 4. Resistance to 6-mercaptopurine and 6-thioguanine. *J Biol Chem*, **276**, 33747-54.
- Chen, Z.S., Lee, K., Walther, S., Raftogianis, R.B., Kuwano, M., Zeng, H. and Kruh, G.D. (2002) Analysis of methotrexate and folate transport by multidrug resistance protein 4 (ABCC4): MRP4 is a component of the methotrexate efflux system. *Cancer Res*, **62**, 3144-50.
- Wijnholds, J., Mol, C.A., van Deemter, L., de Haas, M., Scheffer, G.L., Baas, F., Beijnen, J.H., Scheper, R.J., Hulse, S., De Clercq, E. et al. (2000) Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. *Proc Natl Acad Sci U S A*, **97**, 7476-81.
- Jedlitschky, G., Burchell, B. and Keppler, D. (2000) The multidrug resistance protein 5 functions as an ATP-dependent export pump for cyclic nucleotides. *J Biol Chem*, **275**, 30069-74.
- Kool, M., van der Linden, M., de Haas, M., Baas, F. and Borst, P. (1999) Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. *Cancer Res*, **59**, 175-82.
- Belinsky, M.G., Chen, Z.S., Shchaveleva, I., Zeng, H. and Kruh, G.D. (2002) Characterization of the drug resistance and transport properties of multidrug resistance protein 6 (MRP6, ABCC6). *Cancer Res*, **62**, 6172-7.
- Klucken, J., Buchler, C., Orso, E., Kaminski, W.E., Porsch-Ozcurumez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M. et al. (2000) ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport. *Proc Natl Acad Sci U S A*, **97**, 817-22.
- Huang, Y., Anderle, P., Bussey, K.J., Barbacioru, C., Shankavaram, U., Dai, Z., Reinhold, W.C., Papp, A., Weinstein, J.N. and Sadee, W. (2004) Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer Res*, **64**, 4294-301.
- Komatani, H., Kotani, H., Hara, Y., Nakagawa, R., Matsumoto, M., Arakawa, H. and Nishimura, S. (2001) Identification of breast cancer resistant protein/mitoxantrone resistance/placenta-specific, ATP-binding cassette transporter as a transporter of NB-506 and J-107088, topoisomerase I inhibitors with an indolocarbazole structure. *Cancer Res*, **61**, 2827-32.
- Maliepaard, M., Scheffer, G.L., Faneyte, I.F., van Gastelen, M.A., Pijnenborg, A.C., Schinkel, A.H., van De Vijver, M.J., Scheper, R.J. and Schellens, J.H. (2001) Subcellular localization and distribution of the breast cancer resistance protein transporter in normal human tissues. *Cancer Res*, **61**, 3458-64.

Supplementary Table 2. Comparison of Allele Frequencies of SNPs, at the ABC Transporter Gene Loci, Between this Study and the HapMap/Perlegen Databases

A				
Populations		CH*	EA/CEU	AA/YRI
This Study	No. of SNPs	107	111	111
	No. of SNPs with MAF <5%	16	22	19
This Study vs HapMap	No. of common SNPs (% of total SNPs in this study)	68 (63.6%)	73 (65.8%)	71 (64.0%)
	No. of SNPs with MAF<5% (% of total common SNPs)	9 (13.2%)	10 (13.7%)	12 (16.9%)
	No. of SNPs with >10% difference of allele frequencies (% of common polymorphic SNPs)	15 (22.1%)	9 (12.3%)	11 (15.5%)
This Study vs Perlegen	No. of common SNPs (% of total SNPs in this study)	40 (37.4%)	42 (37.8%)	-
	No. of SNPs with MAF<5% (% of total common SNPs)	3 (7.5%)	5 (11.9%)	-
	No. of SNPs with >10% difference of allele frequencies (% of common polymorphic SNPs)	8 (20%)	11 (26.2%)	-
HapMap vs Perlegen	No. of common SNPs	36	37	34
	No. of SNPs with MAF<5% (% of total common SNPs)	3 (8.3%)	5 (13.5%)	4 (11.8%)
	No. of SNPs with >10% difference of allele frequencies (% of common polymorphic SNPs)	4 (11.1%)	7 (18.9%)	7 (20.6%)

B												
	This Study				HapMap				Perlegen			
	CHS	EA	AA	All	CHB	CAU	YRI	All	CHLA	EA	AA	All
This Study	-	-	-	-	5.9%	4.1%	8.3%	6.1%	2.5%	4.8%	-	3.7%
HapMap	7.02	4.47	5.10	5.50	-	-	-	-	0.0%	2.7%	0.0%	0.9%
Perlegen	6.15	7.65	-	6.92	5.07	6.63	6.13	5.95	-	-	-	-

*CH represents CHS, CHB and CHLA in our study, HapMap and Perlegen, respectively.

Table B - Cells in the top right triangular section indicate the percentage of SNPs with significantly different allele frequencies (Fisher's exact p-value < 0.01), whereas those in the lower triangular section contain the average differences in allele frequencies between comparable SNPs between the respective data sets.

Supplementary Table 3. Profile of tSNPs in the Three Populations

Genes	Populations	No. of SNP (MAF>0)	No. of tSNPs	Average Length/tSNPs (kb)	Tagging Efficiency*
ABCB1/4	CHS	16	8	40	2.0
	EA	17	7	45.7	2.4
	AA	17	11	29.1	1.5
ABCB2/3	CHS	5	5	5	1.0
	EA	7	6	4.2	1.2
	AA	8	7	3.6	1.1
ABCC1/6	CHS	18	10	26.1	1.8
	EA	19	10	26.1	1.9
	AA	17	9	29	1.9
ABCC2	CHS	7	4	17.5	1.8
	EA	7	5	14	1.4
	AA	8	7	10	1.1
ABCC3	CHS	4	4	15.5	1.0
	EA	6	6	10.3	1.0
	AA	7	7	8.9	1.0
ABCC4	CHS	12	7	40	1.7
	EA	12	9	31.1	1.3
	AA	12	8	35	1.5
ABCC5	CHS	7	3	33.3	2.3
	EA	7	3	33.3	2.3
	AA	7	5	20	1.4
ABCG1	CHS	8	6	15	1.3
	EA	8	6	15	1.3
	AA	8	7	12.9	1.1
ABCG2	CHS	9	8	7.5	1.1
	EA	7	5	13	1.4
	AA	8	7	11.9	1.1
Total	CHS	89	55	23.1	1.6
	EA	87	55	23.1	1.6
	AA	89	65	19.9	1.4

*Tagging efficiency is defined as the total number of SNPs genotyped divided by the number of tSNPs selected (Ke, *et al.* (2005) Human Molecular Genetics 14:2757-2767).

REFERENCES

- Dean, M.; Rzhetsky, A.; Allikmets, R. *Genome Res.*, **2001**, *11*, 1156-66.
- Cervenak, J.; Andrikovics, H.; Ozvegy-Laczka, C.; Tordai, A.; Nemet, K.; Varadi, A.; Sarkadi, B. *Cancer Lett.*, **2006**, *234*, 62-72.
- Lee, C.G.L.; Chong, S.S.; Lee, J.D. *Curr. Pharmacogenomics*, **2004**, *2*, 1-11.
- The International HapMap Consortium. *Nature*, **2005**, *437*, 1299-320.
- de Bakker, P.I.; Yelensky, R.; Pe'er, I.; Gabriel, S.B.; Daly, M.J.; Altshuler, D. *Nat. Genet.*, **2005**, *37*, 1217-23.
- Hinds, D.A.; Stuve, L.L.; Nilsen, G.B.; Halperin, E.; Eskin, E.; Ballinger, D.G.; Frazer, K.A.; Cox, D.R. *Science*, **2005**, *307*, 1072-9.
- Patil, N.; Berno, A.J.; Hinds, D.A.; Barrett, W.A.; Doshi, J.M.; Hacker, C.R.; Kautzer, C.R.; Lee, D.H.; Marjoribanks, C.; McDonough, D.P.; Nguyen, B.T.; Norris, M.C.; Sheehan, J.B.; Shen, N.; Stern, D.; Stokowski, R.P.; Thomas, D.J.; Trulson, M.O.; Vyas, K.R.; Frazer, K.A.; Fodor, S.P.; Cox, D.R. *Science*, **2001**, *294*, 1719-23.
- Ribas, G.; Gonzalez-Neira, A.; Salas, A.; Milne, R.L.; Vega, A.; Carracedo, B.; Gonzalez, E.; Barroso, E.; Fernandez, L.P.; Yankilevich, P.; Robledo, M.; Carracedo, A.; Benitez, J. *Hum. Genet.*, **2006**, *118*, 669-79.
- Marvelle, A.F.; Lange, L.A.; Qin, L.; Wang, Y.; Lange, E.M.; Adair, L.S.; Mohlke, K.L. *J. Hum. Genet.*, **2007** [Epub ahead of print].
- Mahasirimongkol, S.; Chantratita, W.; Promso, S.; Pasomsab, E.; Jinawath, N.; Jongjaroenprasert, W.; Lulitanond, V.; Krittayaposisitpot, P.; Tongsim, S.; Sawanpanyalert, P.; Kamatani, N.; Nakamura, Y.; Sura, T. *J. Hum. Genet.*, **2006**, *51*, 896-904.
- Willer, C.J.; Scott, L.J.; Bonnycastle, L.L.; Jackson, A.U.; Chines, P.; Pruim, R.; Bark, C.W.; Tsai, Y.Y.; Pugh, E.W.; Doheny, K.F.; Kinnunen, L.; Mohlke, K.L.; Valle, T.T.; Bergman, R.N.;

- Tuomilehto, J.; Collins, F.S.; Boehnke, M. *Genet. Epidemiol.*, **2006**, *30*, 180-90.
- [12] Montpetit, A.; Nelis, M.; Laflamme, P.; Magi, R.; Ke, X.; Rimm M.; Cardon, L.; Hudson, T.J.; Metspalu, A. *PLoS Genet.*, **2006**, *2*, e27.
- [13] Johansson, A.; Vavrouch-Nilsson, V.; Cox, D.R.; Frazer, K.A.; Gyllensten, U. *Hum. Genet.*, **2007**, *122*, 141-50.
- [14] Iida, A.; Saito, S.; Sekine, A.; Mishima, C.; Kitamura, Y.; Kondo, K.; Harigae, S.; Osawa, S.; Nakamura, Y. *J. Hum. Genet.*, **2002**, *47*, 285-310.
- [15] Ito, S.; Ieiri, I.; Tanabe, M.; Suzuki, A.; Higuchi, S.; Otsubo, K. *Pharmacogenetics*, **2001**, *11*, 175-84.
- [16] Saito, S.; Iida, A.; Sekine, A.; Miura, Y.; Ogawa, C.; Kawachi, S.; Higuchi, S.; Nakamura, Y. *J. Hum. Genet.*, **2002**, *47*, 147-71.
- [17] Tonissson, N.; Zernant, J.; Kurg, A.; Pavel, H.; Slavin, G.; Roomere, H.; Mejel, A.; Hainaut, P.; Metspalu, A. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 5503-8.
- [18] Gwee, P.C.; Tang, K.; Chua, J.M.; Lee, E.J.; Chong, S.S.; Lee, C.G. *Clin. Chem.*, **2003**, *49*, 672-6.
- [19] Tang, K.; Wong, L.P.; Lee, E.J.; Chong, S.S.; Lee, C.G. *Hum. Mol. Genet.*, **2004**, *13*, 783-97.
- [20] Goldstein, D.B.; Ahmadi, K.R.; Weale, M.E.; Wood, N.W. *Trends Genet.*, **2003**, *19*, 615-22.
- [21] Ahmadi, K.R.; Weale, M.E.; Xue, Z.Y.; Soranzo, N.; Yarnall, D.P.; Briley, J.D.; Maruyama, Y.; Kobayashi, M.; Wood, N.W.; Spurr, N.K.; Burns, D.K.; Roses, A.D.; Saunders, A.M.; Goldstein, D.B. *Nat. Genet.*, **2005**, *37*, 84-9.
- [22] Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; *3*: 299-309.
- [23] Johnson, G.C.; Esposito, L.; Barratt, B.J.; Smith, A.N.; Heward, J.; Di Genova, G.; Ueda, H.; Cordell, H.J.; Eaves, I.A.; Dudbridge, F.; Twells, R.C.; Payne, F.; Hughes, W.; Nutland, S.; Stevens, H.; Carr, P.; Tuomilehto-Wolf, E.; Tuomilehto, J.; Gough, S.C.; Clayton, D.G.; Todd, J.A. *Nat. Genet.*, **2001**, *29*, 233-7.
- [24] Gwee, P.C.; Tang, K.; Sew, P.H.; Lee, E.J.; Chong, S.S.; Lee, C.G. *Pharmacogenet. Genom.*, **2005**, *15*, 91-104.
- [25] Wang, Z.; Wang, B.; Tang, K.; Lee, E.J.; Chong, S.S.; Lee, C.G. *Hum. Mol. Genet.*, **2005**, *14*, 2075-87.
- [26] Shifman, S.; Kuypers, J.; Kokoris, M.; Yakir, B.; Darvasi, A. Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.*, **2003**, *12*, 771-6.
- [27] Weiss, K.M.; Clark, A.G. *Trends Genet.*, **2002**, *18*, 19-24.
- [28] Zeggini, E.; Rayner, W.; Morris, A.P.; Hattersley, A.T.; Walker, M.; Hitman GA, Deloukas, P.; Cardon, L.R.; McCarthy, M.I. *Nat. Genet.*, **2005**, *37*, 1320-2.
- [29] Wang, Z.; Wang, J.; Tantoso, E.; Wang, B.; Tai, A.Y.; Ooi, L.L.; Chong, S.S.; Lee, C.G. *Hum. Mol. Genet.*, **2007**, *16*, 1367-80.
- [30] Voight, B.F.; Kudravalli, S.; Wen, X.; Pritchard, J.K. *PLoS Biol.*, **2006**, *4*, e72.