OPEN

The Pharmacogenomics Journal (2014) 14, 555–563

www.nature.com/tpj

ORIGINAL ARTICLE

Pharmacogenomic diversity in Singaporean populations and Europeans

LR Brunham[1,2,3,15], SL Chan[1,15], R Li[4], F Aminkeng[3], X Liu[4,5], WY Saw[4], RTH Ong[4], EN Pillai[4], BC Carleton[6,7], D Toh[8], SH Tan[8], SH Koo[9], EJD Lee[9], KS Chia[4,10,11], CJD Ross[3,6], MR Hayden[1,3], C Sung[8,12] and YY Teo[4,5,13,14]

Differences in the frequency of pharmacogenomic variants may influence inter-population variability in drug efficacy and risk of adverse drug reactions (ADRs). We investigated the diversity of ~4500 genetic variants in key drug-biotransformation and -response genes among three South East Asian populations compared with individuals of European ancestry. We compared rates of reported ADRs in these Asian populations to determine if the allelic differentiation corresponded to an excess of the associated ADR. We identified an excess of ADRs related to clopidogrel in Singaporean Chinese, consistent with a higher frequency of a known risk variant in *CYP2C19* in that population. We also observed an excess of ADRs related to platinum compounds in Singaporean CHS, despite a very low frequency of known ADR risk variants, suggesting the presence of additional genetic and non-genetic risk factors. Our results point to substantial diversity at specific pharmacogenomic loci that may contribute to inter-population variability in drug response phenotypes.

The Pharmacogenomics Journal (2014) **14,** 555–563; doi:10.1038/tpj.2014.22; published online 27 May 2014

INTRODUCTION

Many drugs vary in efficacy and risk for adverse drug reactions (ADRs) between different ancestral populations. One important source of this variability is the difference in frequency of alleles that modulate drug response.[1] In some instances, inter-population differences in drug response have been linked to specific pharmacogenomic variants. For instance, differences in the frequency of *VKORC1* variants account for differences in daily warfarin dose requirement between African-American, European and multiple South East Asian populations,[2,3] and differences in the frequency of a key variant in the *IFNL3* gene (formerly *IL28B*) appear to explain inter-population differences in the response rate for treatment of chronic hepatitis C virus infection.[4] Individuals of African ancestry have been shown to harbor higher frequencies of certain variants that predispose to ADRs related to cancer chemotherapeutics and drugs for HIV/AIDS and tuberculosis.[5] However, to a large extent, we lack comprehensive information regarding differences in pharmacologically important genetic variants in many world populations. This is particularly the case for Asian populations, which represent many of the world's most populous regions and yet have been under-represented in pharmacogenomics research as well as clinical trials of medications.[6]

Recently, progress has been made to define genetic diversity of drug-response alleles in global populations, including a report describing variation at ~1150 variants in individuals from 19 populations,[7] including those from China and Japan. Man *et al.*[8] previously explored the diversity of 165 drug-metabolizing enzyme and transporter variants in individuals of Han Chinese, Japanese and Korean ancestry, and identified many similarities, but also important differences between these Asian populations. Singapore is a multi-ethnic city–state comprising three major population groups, the Chinese, Malays and Indians, with a shared environment. Previous studies have documented important differences among Singaporean populations in the frequency of specific drug-response alleles, for example, variants in *CYP3A4*, *PXR*, *CAR*, *HNF4α*, *CBR3*, *AKR1C3*, *SLC28A1* and other genes associated with response and toxicity to chemotherapeutic agents,[9–15] *SLCO1B1* variants associated with statin exposure,[16] and *CYP2C19* variants associated with clopidogrel response.[17] These studies suggested that, despite the overall similarities between these population groups, they may harbor clinically important differences in the frequencies of specific drug-response alleles that contribute to inter-population differences in the risk of ADRs and drug efficacy.

Here, we explored the allelic diversity of drug-biotransformation and -response genes among individuals in the Singapore Genome Variation Project (SGVP) in comparison with individuals of European ancestry from the Canadian Pharmacogenomics Network for Drug Safety (CPNDS)[18] using a custom assay for pharmacogenomic variation in ~4500 markers. We made use of dense genome-wide genotyping of the SGVP populations to investigate

[1]Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore; [2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; [3]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; [4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore; [5]NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore; [6]Division of Translational Therapeutics, Department of Pediatrics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; [7]Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada; [8]Health Sciences Authority, Singapore; [9]Department of Pharmacology, National University of Singapore, Singapore; [10]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; [11]Center for Molecular Epidemiology, National University of Singapore, Singapore; [12]Duke-NUS Graduate Medical School, Singapore; [13]Life Sciences Institute, National University of Singapore, Singapore and [14]Department of Statistics and Applied Probability, National University of Singapore, Singapore. Correspondence: Professor LR Brunham, Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research and the National University of Singapore, 8A Biomedical Grove, Immunos L5, Biopolis S138648, Singapore.
E-mail: Liam_brunham@tlgm.a-star.edu.sg
[15]These authors contributed equally to this study.
Received 4 August 2013; revised 28 March 2014; accepted 15 April 2014; published online 27 May 2014

the haplotypic landscape around selected pharmacogenomic variants.[19] Finally, we tested the hypothesis that the observed allelic differentiation corresponds to inter-population differences in drug-response by examining rates of reported ADRs in the three Asian populations. Our results point to substantial diversity at specific variants both within and between these populations, which may explain some of the observed variability in drug-response phenotypes between these population groups.

MATERIALS AND METHODS

Study populations

The study populations consisted of the 268 individuals of Chinese, Malay and Indian ancestry from the SGVP,[19] designated CHS, MAS and INS, respectively, and 958 individuals of European ancestry designated EUR, recruited through the CPNDS.[18] Population group was self-reported for both SGVP and CPNDS populations on the basis of all four grandparents belonging to the same population, and was verified by principal component analysis. All samples were assessed for cryptic relatedness on the basis of excessive identity by state genotypes or identity by descent estimation for SGVP and EUR, respectively. For SGVP, the 268 individuals were those that passed QC, thus any individuals with non-concordant self-reported and genetically inferred ancestry, evidence of admixture or possible duplicates and relatives were removed. Among the EUR samples, no duplicates or related individuals were identified.

All participants, or their legal guardians in the case of minors, provided written informed consent. This study was approved by the Institutional Review Boards of the National University Hospital and National University of Singapore and the University of British Columbia.

ADME panel and genotyping

All samples were genotyped using a custom Illumina Infinium genotyping assay (Illumina, San Diego, CA, USA) designed to probe 4535 single-nucleotide polymorphisms (SNPs) in 359 key genes involved in drug absorption, distribution, metabolism and excretion (ADME). The design of this panel has been reported previously.[5,20] In brief, this assay was designed to capture variation in candidate genes selected for known or suspected involvement in drug biotransformation as well as drug transporters and drug targets. This assay includes both tagging SNPs, selected using the LDSelect algorithm[21] and Hapmap[22] as well as putative functional variants identified via public databases and literature review. Genotypes were called using the Illumina Genome Studio software package as previously described.[5,20]

Quality control

QC procedures were performed for SGVP and EUR populations separately. For each data set, duplicate samples with lower call rate and samples with call rates $<90\%$ were first removed. Subsequently, duplicate SNPs with lower call rate and SNPs with call rates $<90\%$ were removed. Any duplicate SNPs with $<90\%$ concordance were also removed. For SGVP samples, concordance of SNPs that overlap with the existing SGVP database was also checked and any samples or SNPs with $<95\%$ concordance were removed. For SNPs present on both the ADME panel and the SGVP database, the data set with higher call rate for that SNP was used. Deviation from Hardy–Weinberg equilibrium was assessed within each population using the χ^2 goodness-of-fit test and a threshold of $P<0.001$ to indicate deviation. SNPs with genotype data for both SGVP and EUR samples that passed QC procedures were used for analysis.

Statistical analysis

Variance in allele frequencies. The variance in minor allele frequency (MAF) at each SNP was compared across populations using Wright's fixation index (F_{ST}).[23] An F_{ST} value of >0.05 indicates a moderate degree of between-population differentiation, and >0.15 indicates a substantial degree of between-population differentiation.[23] Comparisons were made within SGVP populations, and between SGVP populations and EUR. For SNPs present in the SGVP database but not on the ADME panel, the MAF for EURs were obtained from the HapMap CEU population. For each F_{ST}, an empirical *P*-value was calculated as the proportion of SNPs with a F_{ST} of that value or greater in the SGVP and HapMap CEU databases. The empirical *P*-value therefore indicates how differentiated a SNP is with respect to the rest of the genome. For comparisons among the SGVP populations only, empirical *P*-values were calculated using 1 547 281 SNPs that were common across the three populations. For comparisons involving EUR, empirical *P*-values were calculated using 1 296 966 SNPs common between SGVP and HapMap CEU. The 95% confidence intervals around the MAFs were calculated using the Jeffrey's interval. Analyses were carried out using R version 2.15.2.[24]

Correlation with ADR reports from the HSA database. ADR reports submitted to the Health Sciences Authority (HSA) of Singapore between 1 January 2001 and 18 December 2013 were retrieved for drugs associated with clinically annotated SNPs for ADRs in the Pharmacogenetics and Pharmacogenomics Knowledge base (PharmGKB) (www.pharmgkb.org, data downloaded 5 February 2012)[25] (level 1A—2B evidence) that displayed a moderate-to-high degree of differentiation ($F_{ST}\geqslant0.05$) between the CHS, MAS and INS in our data set. We also included clopidogrel because its associated variant *CYP2C19* rs4986893 displayed a large inter-population frequency difference despite a lower F_{ST} value in our study and because other functional clopidogrel-associated SNPs have previously been reported to be highly differentiated in Singaporean populations.[17] We compared the actual proportion of ADR reports per population to the expected proportion based on the demographic composition of Singapore[26] using the χ^2 test after correction for multiple-hypothesis testing. ADR reports were further analyzed using the World Health Organization Adverse Reaction Terminology (WHO-ART) preferred terms.

Haplotype construction and variance in haplotype structure. To explore how well genotyped variants tag for possible functional variants, we examined gene regions containing highly differentiated clinically associated SNPs ($F_{ST}>0.05$), which are not known or are unlikely to be functional. The ADME genotype data for SGVP samples were pooled with the existing SGVP database and re-phased using BEAGLE version 3.3.2[27] by population. Haploview version 4.2[28] was used and the risk allele-carrying haplotypes with at least 1% frequency in each population were displayed diagrammatically for comparison using R. Data for CEU were downloaded from HapMap for haplotype construction. *TPMT* haplotypes were constructed following phasing of genotypes using BEAGLE. For the SGVP populations, the ADME genotype data were combined with SGVP data and phased again, and the *TPMT* SNPs that define *2, *3A, *3B and *3C haplotypes (rs1800460, rs1142345 and rs1800462) were extracted and the haplotype frequencies estimated by counting within each population. For EUR, these three SNPs were phased using BEAGLE and the haplotype frequencies obtained by counting. Haplotypes were identified as *1 if they had the common allele at these three SNPs.

The haplotype similarity index (HSI) was calculated as previously described[29] for the common set of SNPs that are polymorphic in at least one population and 10 kb of flanking region. Briefly, the HSI is the amount of variance explained by the first principal component of a similarity matrix representing differences between all the haplotypes in the same region. The HSI ranges from 0 to 1 with lower values indicating greater diversity.

RESULTS

Pharmacogenomic diversity across SGVP populations and EURs

Following QC procedures, our data set included genotypes for 3972 SNPs in 253 individuals from SGVP (88 CHS, 87 MAS and 78 INS) and 956 EURs. Importantly, our custom ADME panel provided genotype data for an additional 934 SNPs to the densely genotyped SGVP data set previously typed for 1.6 million sites,[19] including many more coding and potentially functional variants that may influence drug-response phenotypes (coding variants: 12.7% of new 934 SNPs vs 1.8% of SGVP SNPs, $P=5.2\times10^{-164}$), highlighting the additional information that can be derived from dedicated ADME genotyping assays beyond what is available from commercially available genome-wide panels. The MAF, F_{ST} and empirical *P*-values for the eight different population comparisons across the 3972 SNPs are shown in Supplementary Table 1.

Figure 1 shows the distribution of F_{ST} values within and between these populations. The three SGVP populations tended to be more similar among themselves than with EUR at these pharmacogenomic markers (Figure 1a). Within the SGVP populations, CHS and MAS tended to be more similar at these sites than either CHS and INS, or MAS and INS (Figure 1b). Similarly, EUR and

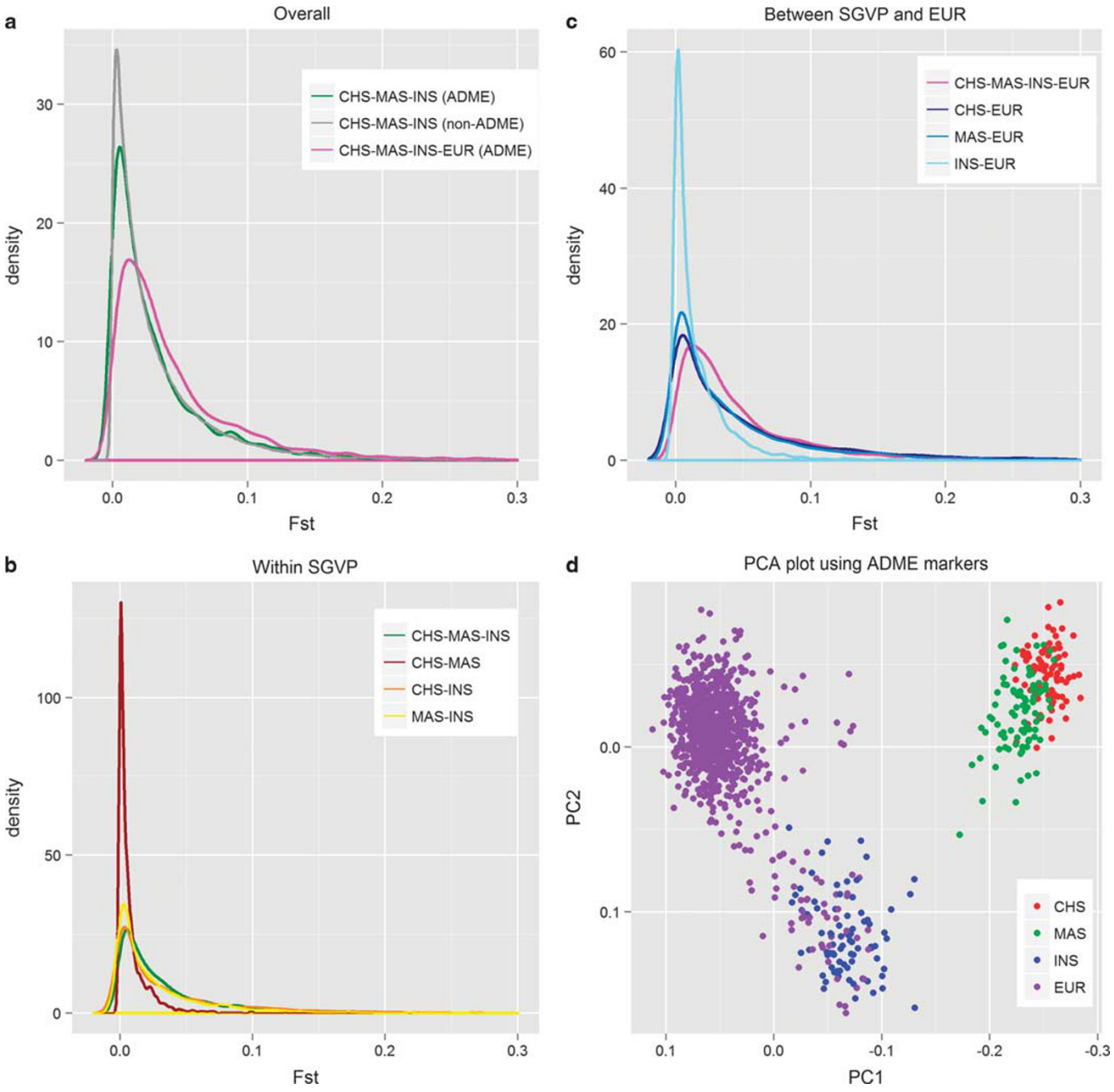

Figure 1. Distribution of F_{ST} values across populations. The distribution of F_{ST} values are shown for (**a**) the three Singapore Genome Variation Project (SGVP) populations (for both absorption, distribution, metabolism and excretion (ADME) and non-ADME single-nucleotide polymorphisms (SNPs)) and the SGVP populations and Europeans (EUR); (**b**) pairwise comparisons within the SGVP populations and across all three populations; and (**c**) pairwise comparisons between each SGVP population and EUR, and across all four populations. Curves that are shifted to the left indicate a lower degree of differentiation at the ADME variants. Within the SGVP populations (**b**), the Chinese (CHS) and Malay (MAS) populations are more similar at these ADME sites than are CHS and Indians (INS) or MAS and INS. Between SGVP and EUR (**c**), the INS and EUR show a lower degree of relative differentiation than do the CHS and EUR or MAS and EUR. Panel **d** shows a principal component analysis (PCA) plot of the four populations.

INS were more similar than either CHS and EUR, or MAS and EUR at these sites (Figure 1c). Figure 1d shows the principal component analysis plot of these populations based on ADME genotype data, reflecting a similar pattern of population structure compared with that determined using genome-wide data.[19] The median F_{ST} of these ADME markers in the CHS, MAS and INS populations is 0.0187 compared with 0.0186 for genome-wide non-ADME markers, indicating that the degree of differentiation of ADME markers is similar to that of the rest of the genome. Consistent with this, the distribution of F_{ST} in CHS, MAS and INS is similar for ADME compared with non-ADME markers (Figure 1a).

Pharmacogenomic diversity at clinically annotated markers

Table 1 shows the MAFs across the four populations for clinically annotated SNPs with a moderate to high level of evidence (levels

Table 1. Highly differentiated PharmGKB clinically annotated loci

Gene	*rsid*	*Functional annotation*	*Alleles (maj/ min)*[a]	*MAF (95% CI)*				F_{ST}	*Level of evidence*[b]	*Associated drugs*
				CHS	*MAS*	*INS*	*EUR*			
VKORC1	rs9923231[c]	Upstream of gene	C/T	0.906 (0.859, 0.941)	0.758 (0.692, 0.817)	0.127 (0.082, 0.183)	0.398 (0.336, 0.463)	0.375	1A	Warfarin
VKORC1	rs9934438[c]	Intronic	G/A	0.906 (0.859, 0.941)	0.761 (0.692, 0.817)	0.127 (0.082, 0.183)	0.398 (0.336, 0.463)	0.377	1B	Acenocoumarol, warfarin
VKORC1	rs7294	Flanking 3'UTR	C/T	0.102 (0.064, 0.153)	0.224 (0.167, 0.29)	0.776 (0.706, 0.836)	0.384 (0.268, 0.406)	0.276	1B	Warfarin
COMT	rs9332377	Intronic	C/T	0 (0, 0.014)	0.006 (0.001, 0.027)	0.205 (0.148, 0.274)	0.169 (0.132, 0.186)	0.101	1B	Cisplatin
GRIK4	rs1954787[c]	Intronic	C/T	0.161 (0.115, 0.218)	0.191 (0.138, 0.253)	0.386 (0.314, 0.461)	0.487 (0.422, 0.552)	0.086	1B	Citalopram
CYP3A5	rs776746	Intronic	C/T	0.290 (0.227, 0.360)	0.379 (0.310, 0.455)	0.253 (0.187, 0.322)	0.102 (0.081, 0.116)	0.053	1B	Tacrolimus
									2A	Sirolimus
SOD2	rs4880[c]	Coding V16A	A/G	0.104 (0.067, 0.153)	0.242 (0.183, 0.308)	0.53 (0.454, 0.605)	0.447 (0.383, 0.512)	0.127	2B	Cyclophosphamide
FDPS	rs2297480[c]	Intronic	T/G	0.628 (0.560, 0.696)	0.610 (0.539, 0.682)	0.289 (0.224, 0.361)	0.296 (0.240, 0.358)	0.108	2B	Alendronate, bisphosphonates, risedronate
ABCG2	rs2231142	Coding Q141K	G/T	0.358 (0.290, 0.431)	0.293 (0.229, 0.366)	0.070 (0.038, 0.119)	0.104 (0.083, 0.118)	0.091	2B	Rosuvastatin
ERCC1	rs11615	Coding N118N	A/G	0.753 (0.688, 0.815)	0.661 (0.588, 0.732)	0.545 (0.467, 0.622)	0.374 (0.262, 0.396)	0.082	2B	Carboplatin, cisplatin, oxaliplatin, platinum, platinum compounds
OPRM1	rs1799971[c]	Coding N40D	A/G	0.417 (0.349, 0.487)	0.444 (0.372, 0.517)	0.434 (0.360, 0.510)	0.155 (0.112, 0.206)	0.063	2B	Naloxone
ADD1	rs4961[c]	Coding G460W	G/T	0.406 (0.339, 0.477)	0.386 (0.318, 0.461)	0.169 (0.118, 0.231)	0.204 (0.155, 0.260)	0.054	2B	Furosemide, spironolactone

Abbreviations: CI, confidence interval; CHS, Chinese; EUR, European; INS, Indian; MAF, minor allele frequency; MAS, Malay; Maj/Min, major/minor; SGVP, Singapore Genome Variation Project; SNP, single-nucleotide polymorphism. [a]Minor allele frequency is with respect to EUR. [b]Clinical annotation levels of evidence from PharmGKB. [c]MAFs for these SNPs obtained from SGVP and HapMap CEU populations.

1A to 2B) in PharmGKB that display at least a moderate degree of differentiation ($F_{ST} \geqslant 0.05$) across the four populations. Among these variants, we observed significant differentiation at three SNPs in *VKORC1*: rs9923231 ($F_{ST} = 0.375$), rs9934438 ($F_{ST} = 0.377$) and rs7294 ($F_{ST} = 0.276$), all associated with warfarin dose requirement. We observed a higher frequency of the *CYP3A5* rs776746 'T' allele, associated with increased clearance of tacrolimus, among the three SGVP populations compared with EUR ($F_{ST} = 0.053$). *ABCG2* rs2231142, associated with increased statin exposure,[30] was more common in CHS compared with EUR (30.4% vs 10.8%, $F_{ST} = 0.091$). *COMT* rs9332377, associated with increased risk of cisplatin-induced ototoxicity,[20] is common in EUR but was extremely rare or absent in the CHS and MAS populations ($F_{ST} = 0.101$). Conversely, *ERCC1* rs11615, associated with a lower rate of cisplatin-induced nephrotoxicity,[31] was more common in CHS compared with EUR (75.3% vs 37.4%, $F_{ST} = 0.082$).

Differences at some of these sites, while not associated with large F_{ST} values, may still be clinically relevant. For instance, the rs4986893 variant in *CYP2C19* (*CYP2C19*3* allele), associated with decreased conversion of the pro-drug clopidogrel to its active metabolite, was >50-times more common in CHS compared with EUR (5.7% vs 0.1%), consistent with previous reports;[17] because of the low frequency of this variant the F_{ST} value did not exceed 0.05. The F_{ST} for variants in the *TPMT* gene, associated with metabolism of thiopurine compounds as well as cisplatin-induced ototoxicity, were also low but may nonetheless be informative. Consistent with previous reports, we observed that ~10% of EURs have a non–*1/*1 *TPMT* diplotype, indicating reduced activity of this enzyme.[32,33] In contrast, this was observed in <2% of CHS or MAS (Supplementary Table 2), suggesting that the diagnostic yield of pre-emptive genotyping before thiopurine administration is likely to be much lower in these two Asian populations. These two examples highlight that while the F_{ST} provides a useful initial measure of differentiation at particular SNPs, a low F_{ST} does not exclude the possibility of a clinically meaningful difference in allele frequency, in particular toward the lower end of the frequency spectrum.

Correlation of allele frequencies with the proportion of observed ADR reports by population

To determine if the differentiation we observed at important pharmacogenomic loci in Singaporean populations correlate with inter-population differences in ADR rates, we examined data from the HSA of Singapore, which acts as the national center that monitors and reviews ADR reports in Singapore. We examined all ADR reports from January 2001 to December 2013 for drugs associated with a clinically annotated ADR marker that displayed at least a moderate degree of differentiation among CHS–MAS–INS in our study ($F_{ST} \geqslant 0.05$). We also included clopidogrel, as its ADR-associated SNP, *CYP2C19* rs4986893, displayed a large inter-population difference in frequency despite a lower F_{ST} and because previous studies have reported that *CYP2C19* polymorphisms influence clopidogrel bioactivation in Singaporean populations.[17] This resulted in four drug-SNP sets: warfarin (*VKORC1* rs9923231, rs9934438 and rs7294), platinum compounds (*COMT* rs9332377), statins (*ABCG2* rs2231142) and clopidogrel (*CYP2C19* rs4986893). We then compared the observed number of ADRs related with these drugs reported per population to the expected number under a model of no inter-population differences, based on the population structure of Singapore from the Singapore 2005 General Household Survey,[26] to identify drugs with a significant excess or deficiency of ADRs reported in a particular population (Supplementary Table 3).

Of these four drugs examined, two displayed a statistically significant excess or deficiency of ADRs in one or more population compared with the model of no inter-population differences after correction for multiple testing: platinum compounds and

clopidogrel (Figure 2). We observed an excess of ADRs related to clopidogrel in the Chinese population, in keeping with the higher frequency of the low activity *CYP2C19*3* (rs4986893) allele in CHS. We also noted a statistically significant excess of ADRs associated with the platinum drugs in the Chinese population that was in the opposite direction of the *COMT* rs9332377 risk allele, being more rare in Chinese. We therefore considered other SNPs that have been reported to be associated with platinum-related ADRs to identify candidate variants that may contribute to the excess of ADRs in the Chinese population. The frequencies of *GSTP* rs1695 and *XRCC1* rs25487 mirrored the observed distribution of ADRs, with a higher frequency of the risk alleles in the CHS population, suggesting that these variants may contribute to the excess of platinum-related ADRs in the Chinese population (Supplementary Table 1).

For warfarin, we observed a trend toward fewer ADRs in the Indian population, including all ADRs, and the subset related to over-coagulation, although this difference did not reach statistical significance. Nevertheless, this trend is in keeping with the lower frequencies of the *VKORC1* rs9923231, rs9934438 and rs7294 variants in INS, all associated with increased sensitivity to warfarin.

Haplotype diversity

For many pharmacogenomic biomarkers, the underlying functional variant is unknown and the genotyped variant acts as a 'tagging' SNP for a presumed functional variant that is in linkage disequilibrium (LD). However, the predictive accuracy of such a marker depends on the degree of LD between the tagging marker and the functional SNP. As patterns of LD differ widely between populations,[34] a tagging SNP identified as being predictive of a drug-response phenotype in one population may not be accurate in other population groups. This is particularly the case for loci that are highly differentiated between populations.

We made use of the previous genome-wide genotyping of the SGVP cohort as well as HapMap CEU data to examine the haplotypic diversity around pharmacogenomic markers that displayed a high degree of differentiation and for which the underlying functional variant is unknown. These included variants in the *COMT*, *ERCC1*, *FDPS* and *GRIK4* genes. We also calculated the HSI[29] to quantify the degree of haplotypic diversity, with lower values indicating greater diversity. In the *COMT* gene region shown in Figure 3a, the risk variant was present on only one main haplotype in CEU, MAS and INS, with an additional low frequency haplotype in INS. This suggests that rs9332377 would be expected to tag a putative functional variant within this LD block similarly in CEU, MAS and INS. This variant is absent in CHS, making it a non-useful marker in that population. The HSI for the entire *COMT* region was 0.859, suggesting that there may be greater haplotypic diversity outside of the LD block that we examined; tagging by rs9332377 may thus be less reliable if the functional variant is distant to rs9332377. For *ERCC1* (Figure 3b) and *FDPS* (Figure 3c), we observed little haplotypic diversity (HSI for these loci were 0.940 and 0.988, respectively), with the dominant haplotype in each population being shared. This suggests that these SNPs would be expected to tag for nearby functional variant(s) similarly in the four populations. In contrast, the region surrounding the rs1954787 variant in *GRIK4* displayed substantial diversity with different predominant haplotypes in the different populations. The HSI for this locus was 0.687. Given this haplotypic diversity, rs1954787 may not be an adequate tagging SNP in different population groups. Collectively, these results suggest that the described SNPs are likely to provide information about nearby functional variants across different populations for *COMT*, *ERCC1* and *FDPS*, but not for *GRIK4* in which case a single tagging SNP may not be appropriate in different populations.

DISCUSSION

Here we investigated the allelic diversity of key pharmacogenomic genes in three major South East Asian populations compared with

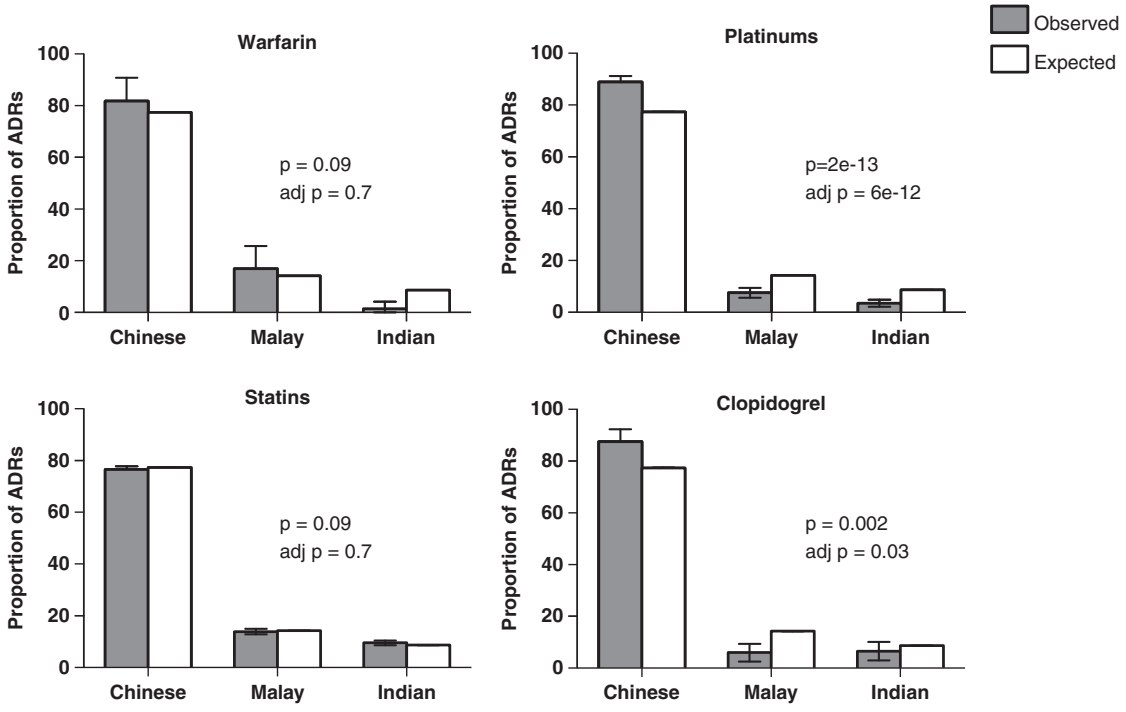

Figure 2. Observed proportions of adverse drug reaction (ADR) reports in the three major Singaporean ethnic groups and their expected proportions. The proportions of total ADR reports collected by the Health Sciences Authority (HSA) between 1 January 2001 and 18 December 2013 in the Chinese, Malays and Indians for drugs associated with moderate to highly differentiated PharmGKB clinically annotated single-nucleotide polymorphisms (SNPs) for ADRs are shown along with the expected proportion based on the demographic makeup of Singapore. Error bars represent the 95% confidence interval. *P*-values are shown for each drug category.

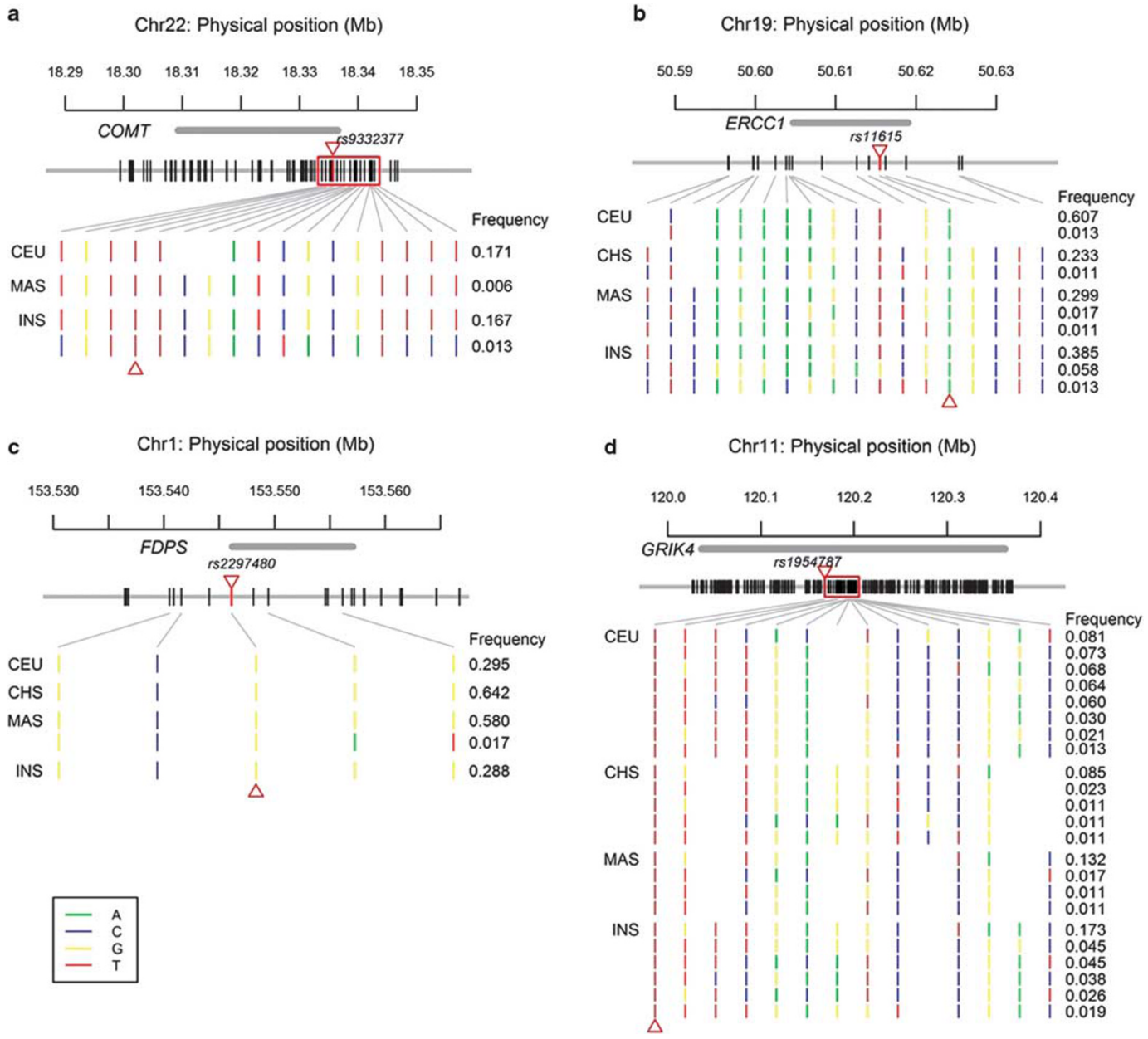

Figure 3. Haplotypic diversity in *COMT*, *ERCC1*, *FDPS* and *GRIK4*. SNPs genotyped in Singapore Genome Variation Project (SGVP), HapMap and the absorption, distribution, metabolism and excretion (ADME) genotyping panel are shown in black vertical lines with respect to their genomic positions. In panel **a**, the haplotypes carrying the risk allele of *COMT* rs9332377 (T allele), which is associated with cisplatin-induced ototoxicity, are shown. Only the linkage disequilibrium (LD) block in which rs9332377 is located is shown. The Chinese population is not shown as there were no haplotypes containing the T allele observed in that population. In panel **b**, the haplotypes carrying the risk allele of *ERCC1* rs11615 (A allele), which is associated with a higher risk of toxicity with platinum compounds, are shown. In panel **c**, haplotypes carrying the risk allele of *FDPS* rs2297480 (G allele), which is associated with decreased bisphosphonate response, are shown. In panel **d**, the haplotypes carrying the risk allele of *GRIK4* rs1954787 (T allele), which is associated with decreased citalopram response, are shown. Regions were chosen such that haplotypes with frequency $>1\%$ accounted for at least 90% of all haplotypes. *GRIK4* rs1954787 is located between two LD blocks and no region spanning both the left and right of the SNP contained haplotypes common enough to represent graphically. Only the region to the right of the SNP is shown but similar diversity exists on the left (data not shown). For all genes, only SNPs with minor allele frequency (MAF) $\geqslant 0.05$ were used for haplotype construction and only haplotypes with frequencies of at least 1% are shown, except for the *COMT* rs9332377 haplotype in Malay (MAS), which is the only haplotype in that population that carries the risk allele (**a**). SNPs that contain information in only one population are not shown. For *COMT* and *GRIK4*, the red boxes indicate the gene regions for which the haplotypes are shown. CHS, Chinese; INS, Indians.

Europeans, providing insight into the genetic basis of observed inter-population differences in drug-response and laying a groundwork for future pharmacogenomics research in these populations. Our findings extend previous studies examining population differences in frequency of ADME variants[35] to include many more pharmacologically relevant markers, including a much larger number of potentially actionable PharmGKB clinically annotated genes. To our knowledge, this study represents the

most comprehensive survey to-date of pharmacogenomic variants in these three South East Asian populations.

We used data on ADR reporting to investigate whether highly differentiated ADME SNPs in the Singaporean population were correlated with an excess of ADRs for the associated drug. This analysis revealed a statistically significant excess of ADRs related to clopidogrel and platinum compounds in the Singaporean Chinese population, and a non-statistically significant reduction in warfarin-related ADRs in the Singaporean Indian population. The excess of clopidogrel-related ADRs in the Singaporean Chinese is in accordance with the higher rate of the risk *CYP2C19*3* allele in CHS. Similarly, the lower than expected rate of warfarin-related ADRs in the Singaporean Indian population mirrors the much lower frequency of *VKORC1* variants in INS and is in agreement with previous data showing that *VKCOR1* haplotypes explain interethnic differences in warfarin dose requirement in Singapore.[3] Some contradictory signals were also observed; for instance, clopidogrel-related ADRs were lower than expected in the Malay population despite a relatively high frequency of the *CYP2C19*3* allele, pointing to other genetic or non-genetic factors that also influence these ADR rates.

Although these data provide initial empirical evidence for inter-population differences in ADR rates in Singaporean population, several limitations are noteworthy. First, ADR reporting data are from voluntary reports rather than active surveillance systems. Although Singapore has the highest per capita rate of spontaneous ADR reporting in the world,[36] such data sets are nonetheless incomplete, often lacking information on drug dose, and capture an unknown percentage of all ADRs.[37] Second, ADRs are reported from the total population, rather than the total population receiving a particular medication. As such, inter-ethnic differences in disease prevalence, medication use or adherence could bias these results. Finally, for platinum compounds and clopidogrel, many of the reported ADRs were cutaneous reactions (Supplementary Table 3) for which the role of pharmacogenomic variants is less established.

Platinum-based compounds such as cisplatin and carboplatin are associated with significant ADRs, including both hematological toxicity and ototoxicity. Individuals of Asian ancestry appear to be more sensitive to both the efficacy and toxicities of these agents.[38–42] We observed an excess of ADRs related to platinum compounds in the Singapore Chinese. The *COMT* rs9332377 variant has been associated with an increased risk of cisplatin-induced ototoxicity in a predominantly European pediatric cohort.[20] This variant is common in EUR and INS (16.9% and 20.5%, respectively) but nearly absent in CHS and MAS (0.0% and 0.6%, respectively). The *TPMT* rs12201199 variant, which has also been reported to be associated with cisplatin-induced ototoxicity in pediatric populations of European ancestry[20,43] was similarly much less common in the Asian populations. These data indicate that these pharmacogenomic risk loci are unlikely to contribute to the risk of toxicity to platinum compounds in individuals of Chinese ancestry and suggest the presence of additional risk variants in these populations that remain to be discovered.

Individuals of Asian ancestry are often considered to require a lower starting dose of statins than non-Asians. Indeed, equivalent dosing of rosuvastatin leads to a higher plasma drug level in many Asian populations, including Japanese, Chinese, Malay and Indian, compared with Europeans.[16,44] The *SLCO1B1* gene product mediates uptake of a variety of statins into hepatocytes and a nonsynonymous variant in this gene, rs4149056, is a key pharmacogenomic biomarker of simvastatin-associated myopathy.[45–47] We observed a low degree of differentiation at this SNP among the populations studied, suggesting that this marker is unlikely to explain inter-ethnic differences in statin response. In contrast, we observed significant differentiation of the non-synonymous rs2231142 variant in the *ABCG2* gene, being more common in CHS and MAS than in EUR. This variant impacts the pharmacokinetics of both rosuvastatin and atorvastatin, and is associated with higher area under the plasma concentration-time curve of these drugs.[30] This variant, together with other genetic or non-genetic factors, is therefore a candidate for explaining the higher risk of statin-related ADRs in specific Asian populations.

We were able to harness previous genome-wide genotyping of the SGVP populations to supplement our ADME genotypes and investigate the haplotypic landscape surrounding pharmacologically important SNPs. For three of the four highly differentiated SNPs (rs9332377 in *COMT*, rs11615 in *ERCC1* and rs2297480 in *FDPS*), the regional diversity was well reflected in a single marker. This implies that although the functional SNP is uncertain, it is likely to be tagged by the clinically annotated SNPs across these different population groups. However, this was not the case with rs1954787 in *GRIK4*. In this case, the risk allele is found on different haplotypes in the four populations indicating that knowledge of the underlying causal variant would be required to accurately predict drug-response in multiple populations. These data highlight the importance of identifying the causal variants that underlie pharmacogenetic associations in order to optimize their use as predictive markers across diverse populations in which background haplotypes may differ.

Our findings provide direction for future pharmacogenomic studies in Singaporean and South East Asian populations by pointing to specific variants that may contribute to drug-response phenotypes in these populations and may help to prioritize variants for clinical implementation in specific populations. For example, whereas published clinical practice guidelines consider primarily the **2* allele in *CYP2C19*,[48] our results showing a high frequency of the **3* allele in the Singaporean CHS population and an excess of clopidogrel ADRs in that population suggest that testing for **3* may be warranted in this population. Conversely, we report that loss-of-function haplotypes in *TPMT* are very rare in the Singaporean CHS and MAS populations, suggesting that testing for these variants is likely to identify few carriers. Our results may also inform the study of the cost-effectiveness of pharmacogenomic testing for specific variants in specific populations based on knowledge of a variant's allele frequency, its effect size, the prevalence of the disease being treated and other factors. Such an analysis was recently reported for the *HLA-B*1502* allele associated with Stevens–Johnson syndrome, indicating that testing for this variant before administration of carbamazepine is most cost-effective in Singaporean Chinese and Malay, and less so in Singapore Indian.[49] Our data may also contribute to an assessment of the public health impact of pharmacogenomic testing by means of the population attributable risk[50] that incorporates knowledge of both a variant's effect size and its allele frequency to quantify the benefit of pharmacogenomic testing to reduce the burden of an associated ADR.

In summary, we have surveyed variation in key genes involved in drug-biotransformation and -response in three major South East Asian populations compared with Europeans. We observed significant inter-population differences in many of these variants, potentially contributing to observed inter-populations differences in drug-response. We have made the SGVP ADME genotype data publically available on the SGVP web portal (http://www.statgen.nus.edu.sg/~SGVP). These data expand our understanding of the population diversity of the genetic basis of drug-response and form a foundation for future pharmacogenomic investigations in these populations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This study was supported by the Biomedical Research Council (BMRC) of the Agency for Science, Technology and Research (A*STAR), the National University of Singapore, the Health Sciences Authority of Singapore and the Canadian Institutes of Health

Research. MRH is a Canada Research Chair in Human Genetics. CJDR is supported by a CIHR New Investigator Award. LRB received support from the Clinician Investigator Program at the University of British Columbia.

REFERENCES

1 Yasuda SU, Zhang L, Huang SM. The role of ethnicity in variability in response to drugs: focus on clinical pharmacology studies. *Clin Pharmacol Ther* 2008; **84**: 417–423.
2 Limdi NA, Wadelius M, Cavallari L, Eriksson N, Crawford DC, Lee MT *et al.* Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* 2010; **115**: 3827–3834.
3 Lee SC, Ng SS, Oldenburg J, Chong PY, Rost S, Guo JY *et al.* Interethnic variability of warfarin maintenance requirement is explained by VKORC1 genotype in an Asian population. *Clin Pharmacol Ther* 2006; **79**: 197–205.
4 Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ *et al.* Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 2009; **461**: 399–401.
5 Aminkeng F, Ross CJ, Rassekh SR, Brunham LR, Sistonen J, Dube MP *et al.* Higher frequency of genetic variants conferring increased risk for ADRs for commonly used drugs treating cancer, AIDS and tuberculosis in persons of African descent. *Pharmacogenomics J* 2013; **14**: 160–170.
6 Thiers FA, Sinskey AJ, Berndt ER. Trends in the globalization of clinical trials. *Nat Rev Drug Disc* 2008; **7**: 13–14.
7 Ramos E, Doumatey A, Elkahloun AG, Shriner D, Huang H, Chen G *et al.* Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J* advance online publication, 9 July 2013; doi:10.1038/tpj.2013.24 (e-pub ahead of print).
8 Man M, Farmen M, Dumaual C, Teng CH, Moser B, Irie S *et al.* Genetic variation in metabolizing enzyme and transporter genes: comprehensive assessment in 3 major East Asian subpopulations with comparison to Caucasians and Africans. *J Clin Pharmacol* 2010; **50**: 929–940.
9 Voon PJ, Yap HL, Ma CY, Lu F, Wong AL, Sapari NS *et al.* Correlation of aldo-ketoreductase (AKR) 1C3 genetic variant with doxorubicin pharmacodynamics in Asian breast cancer patients. *Br J Clin Pharmacol* 2013; **75**: 1497–1505.
10 Loh M, Chua D, Yao Y, Soo RA, Garrett K, Zeps N *et al.* Can population differences in chemotherapy outcomes be inferred from differences in pharmacogenetic frequencies? *Pharmacogenomics J* 2012; **13**: 423–429.
11 Wong AL, Yap HL, Yeo WL, Soong R, Ng SS, Wang LZ *et al.* Gemcitabine and platinum pathway pharmacogenetics in Asian breast cancer patients. *Cancer Genomics Proteomics* 2011; **8**: 255–259.
12 Sandanaraj E, Lal S, Selvarajan V, Ooi LL, Wong ZW, Wong NS *et al.* PXR pharmacogenetics: association of haplotypes with hepatic CYP3A4 and ABCB1 messenger RNA expression and doxorubicin clearance in Asian breast cancer patients. *Clin Cancer Res* 2008; **14**: 7116–7126.
13 Fan L, Goh BC, Wong CI, Sukri N, Lim SE, Tan SH *et al.* Genotype of human carbonyl reductase CBR3 correlates with doxorubicin disposition and toxicity. *Pharmacogenet Genomics* 2008; **18**: 621–631.
14 Soo RA, Wang LZ, Ng SS, Chong PY, Yong WP, Lee SC *et al.* Distribution of gemcitabine pathway genotypes in ethnic Asians and their association with outcome in non-small cell lung cancer patients. *Lung Cancer* 2009; **63**: 121–127.
15 Hor SY, Lee SC, Wong CI, Lim YW, Lim RC, Wang LZ *et al.* PXR, CAR and HNF4alpha genotypes and their association with pharmacokinetics and pharmacodynamics of docetaxel and doxorubicin in Asian patients. *Pharmacogenomics J* 2008; **8**: 139–146.
16 Lee E, Ryan S, Birmingham B, Zalikowski J, March R, Ambrose H *et al.* Rosuvastatin pharmacokinetics and pharmacogenetics in white and Asian subjects residing in the same environment. *Clin Pharmacol Ther* 2005; **78**: 330–341.
17 Chan MY, Tan K, Tan HC, Huan PT, Li B, Phua QH *et al.* CYP2C19 and PON1 polymorphisms regulating clopidogrel bioactivation in Chinese, Malay and Indian subjects. *Pharmacogenomics* 2012; **13**: 533–542.
18 Carleton B, Poole R, Smith M, Leeder J, Ghannadan R, Ross C *et al.* Adverse drug reaction active surveillance: developing a national network in Canada's children's hospitals. *Pharmacoepidemiol Drug Saf* 2009; **18**: 713–721.
19 Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E *et al.* Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res* 2009; **19**: 2154–2162.
20 Ross CJ, Katzov-Eckert H, Dube MP, Brooks B, Rassekh SR, Barhdadi A *et al.* Genetic variants in TPMT and COMT are associated with hearing loss in children receiving cisplatin chemotherapy. *Nat Genet* 2009; **41**: 1345–1349.
21 Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; **74**: 106–120.
22 International HapMap Consortium. The International HapMap Project. *Nature* 2003; **426**: 789–796.
23 Wright S. *Evolution and the Genetics of Populations, Volume 4. Variability Within and Among Natural Populations, vol. 4*. University of Chicago Press: Chicago, 1978.
24 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.
25 PharmGKB The pharmacogenetics and pharmacogenomics knowledge base.
26 Department of Statistics, Singapore. Statistics Singapore-General Household Survey 2005 Statistical Release 1: Socio-Demographic and Economic Characteristics, 2013.
27 Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. *Am J Hum Genet* 2011; **88**: 173–182.
28 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
29 Liu X, RT-H Ong, Pillai EN, Elzein AM, Small KS, Clark TG *et al.* Detecting and Characterizing Genomic Signatures of Positive Selection in Global Populations. *Am J Hum Genet* 2013; **92**: 866–881.
30 Keskitalo JE, Zolk O, Fromm MF, Kurkinen KJ, Neuvonen PJ, Niemi M. ABCG2 polymorphism markedly affects the pharmacokinetics of atorvastatin and rosuvastatin. *Clin Pharmacol Ther* 2009; **86**: 197–203.
31 Tzvetkov MV, Behrens G, O'Brien VP, Hohloch K, Brockmoller J, Benohr P. Pharmacogenetic analyses of cisplatin-induced nephrotoxicity indicate a renoprotective effect of ERCC1 polymorphisms. *Pharmacogenomics* 2011; **12**: 1417–1427.
32 Relling MV, Hancock ML, Rivera GK, Sandlund JT, Ribeiro RC, Krynetski EY *et al.* Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst* 1999; **91**: 2001–2008.
33 Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin Pharmacol Ther* 2011; **89**: 387–391.
34 Goldstein DB. Islands of linkage disequilibrium. *Nat Genet* 2001; **29**: 109–111.
35 Chen J, Teo YY, Toh DS, Sung C. Interethnic comparisons of important pharmacology genes using SNP databases: potential application to drug regulatory assessments. *Pharmacogenomics* 2010; **11**: 1077–1094.
36 Biriell C. VigiBase - 8 million ICSRs. *Uppsala Reports* 2013; **61**: 6–7.
37 Mittmann N, Knowles SR, Gomez M, Fish JS, Cartotto R, Shear NH. Evaluation of the extent of under-reporting of serious adverse drug reactions: the case of toxic epidermal necrolysis. *Drug Saf* 2004; **27**: 477–487.
38 O'Donnell PH, Gamazon E, Zhang W, Stark AL, Kistner-Griffin EO, Stephanie Huang R *et al.* Population differences in platinum toxicity as a means to identify novel genetic susceptibility variants. *Pharmacogenet Genomics* 2010; **20**: 327–337.
39 Millward MJ, Boyer MJ, Lehnert M, Clarke S, Rischin D, Goh BC *et al.* Docetaxel and carboplatin is an active regimen in advanced non-small-cell lung cancer: a phase II study in Caucasian and Asian patients. *Ann Oncol* 2003; **14**: 449–454.
40 Paulino AC, Lobo M, Teh BS, Okcu MF, South M, Butler EB *et al.* Ototoxicity after intensity-modulated radiation therapy and cisplatin-based chemotherapy in children with medulloblastoma. *Int J Radiat Oncol Biol Phys* 2010; **78**: 1445–1450.
41 Kwong DL, Wei WI, Sham JS, Ho WK, Yuen PW, Chua DT *et al.* Sensorineural hearing loss in patients treated for nasopharyngeal carcinoma: a prospective study of the effect of radiation and cisplatin treatment. *Int J Radiat Oncol Biol Phys* 1996; **36**: 281–289.
42 Chan SH, Ng WT, Kam KL, Lee MC, Choi CW, Yau TK *et al.* Sensorineural hearing loss after treatment of nasopharyngeal carcinoma: a longitudinal analysis. *Int J Radiat Oncol Biol Phys* 2009; **73**: 1335–1342.
43 Pussegoda K, Ross CJ, Visscher H, Yazdanpanah M, Brooks B, Rassekh SR *et al.* Replication of TPMT and ABCC3 genetic variants highly associated with cisplatin-induced hearing loss in children. *Clin Pharmacol Ther* 2013; **94**: 243–251.
44 Kim KT, Birmingham BK, Azumaya CT, Chen Y, Schneck D, Zalikowski J. Increased systemic exposure to rosuvastatin in Asian subjects residing in the United States compared to Caucasian subjects. *Clin Pharmacol Ther* 2008; **83**: S14–S14.
45 Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F *et al.* SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med* 2008; **359**: 789–799.
46 Voora D, Shah SH, Spasojevic I, Ali S, Reed CR, Salisbury BA *et al.* The SLCO1B1*5 genetic variant is associated with statin-induced side effects. *J Am Coll Cardiol* 2009; **54**: 1609–1616.
47 Brunham LR, Lansberg PJ, Zhang L, Miao F, Carter C, Hovingh GK *et al.* Differential effect of the rs4149056 variant in SLCO1B1 on myopathy associated with simvastatin and atorvastatin. *Pharmacogenomics J* 2011; **12**: 233–237.
48 Scott SA, Sangkuhl K, Gardner EE, Stein CM, Hulot JS, Johnson JA *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-

2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin Pharmacol Ther* 2011; **90**: 328–332.
49 Dong D, Sung C, Finkelstein EA. Cost-effectiveness of HLA-B*1502 genotyping in adult patients with newly diagnosed epilepsy in Singapore. *Neurology* 2012; **79**: 1259–1267.
50 Chan SL, Suo C, Chia KS, Teo YY. The population attributable fraction as a measure of the impact of warfarin pharmacogenetic testing. *Pharmacogenomics* 2012; **13**: 1247–1256.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)