Pei-Chieng Cha · Ryo Yamada · Akihiro Sekine
Yusuke Nakamura · Chong-Lek Koh

Inference from the relationships between linkage disequilibrium and allele frequency distributions of 240 candidate SNPs in 109 drug-related genes in four Asian populations

Received: 21 June 2004 / Accepted: 23 July 2004 / Published online: 11 September 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract The extensive nucleotide diversity in drug-related genes predisposes individuals to different drug responses and is a major problem in current clinical practice and drug development. Striking allelic frequency differences exist in these genes between populations. In this study, we genotyped 240 sites known to be polymorphic in the Japanese population in each of 270 unrelated healthy individuals comprising 90 each of Malaysian Malays, Indians, and Chinese. These sites are distributed in 109 genes that are drug related, such as genes encoding drug-metabolizing enzymes and drug transporters. Allele frequency and linkage disequilibrium distributions of these sites were determined and compared. They were also compared with similar data of 752 Japanese. Extensive similarities in allele frequency and linkage disequilibrium distributions were observed among Japanese, Malaysian Chinese, and Malays. However, significant differences were observed between Japanese and Malaysian Chinese with

Malaysian Indians. These four populations were grouped into two genetic clusters of different ancestries. However, a higher correlation was found between Malaysian Malays and Indians, indicating the existence of extensive admixture between them. The results also imply the possible and rational use of existing single nucleotide polymorphism databases as references to assist future pharmacogenetic studies involving populations of similar ancestry.

Keywords Allelic frequency differences · Single nucleotide polymorphisms · Linkage disequilibrium distributions · Drug-related genes · Population ancestry

P.-C. Cha (✉) · C.-L. Koh
Institute of Biological Sciences (Genetics and Molecular Biology),
Faculty of Science, University of Malaya,
50603, Lembah Pantai, Kuala Lumpur,
Malaysia
E-mail: genescha@yahoo.com
Tel.: +60-3-79675942
Fax: +60-3-79675837

R. Yamada
Laboratory for Rheumatic Diseases,
SNP Research Center (SRC),
Institute of Physical and Chemical Research (RIKEN),
Yokohama, Japan

A. Sekine
Laboratory for Genotyping, SRC, RIKEN,
Yokohama, Japan

Y. Nakamura
Research Group for Personalized Medicine, SRC, RIKEN,
Yokohama, Japan

Y. Nakamura
Laboratory of Molecular Medicine,
Human Genome Center, Institute of Medical Science,
University of Tokyo, Tokyo, Japan

Introduction

Single nucleotide polymorphism (SNP) is the most common form of genetic variation in the human genome. SNPs mirror preceding mutations that were theoretically occasional and not recurrent. Hence, two individuals sharing a variant allele are more likely to be descended from a more recent common ancestor. Moreover, since particular alleles at adjacent loci tend to be co-inherited, for tightly linked loci, this might lead to a nonrandom association among SNPs, termed “linkage disequilibrium” (LD), along chromosomes such that the presence of one variant predicts the presence of other variants (Ardlie et al. 2002; Weiss and Clark 2002). Consequently, the study of the pattern of LD enables us to gain new insight about ancestry of the examined population since LD mapping relies on detection of recombinationally conserved regions around an ancestral mutation (Reich et al. 2001). It is now possible to determine SNPs and linkage disequilibrium of SNPs in different individuals’ genomes, paving the way to the practice of personalized medicine (Kleyn and Vesell 1998; Nebert 1999; McCarthy and Hilfiker 2000; The International SNP Map Working Group 2001).

It is well known that all drugs have the potential of causing different responses from different individuals, and this is a major obstacle in drug development and clinical practice (Meyer 2000). Previous pharmacogenetic studies have shown the existence of striking differences in allelic frequency distributions between populations in genes known to encode drug-metabolizing enzymes (DMEs), drug receptors, drug transporters, and ion channels (Kleyn and Vesell 1998; Pirmohamed and Park 2001; Xie et al. 2001), suggesting the urgency in understanding the ethnical background underlying these genes so as to accelerate the development of personalized medicine.

Many published pharmacogenetic studies (Meyer 1994; Ainoon et al. 1999; Hustert et al. 2001; Iwai et al. 2001; Xie et al. 2001; Ainoon et al. 2001; Chang et al. 2002; Chowbay et al. 2003; Tang et al. 2002; Kroetz et al. 2003; Tan et al. 2003) focused on individual genes and limited mutations. There is restricted information on allele frequency and LD distributions of these genes on a genome-wide scale. A pioneer study by Goddard et al. (2000) elucidated the allele frequency and LD distributions of 114 SNPs in five human populations. However, hitherto there is little information about people living in Asia.

In this study, we genotyped 240 sites known to be polymorphic in the Japanese population and located in 109 drug-related or candidate drug-related genes in each of 270 unrelated healthy individuals comprising 90 each of Malays, Indians, and Chinese living in Malaysia. Allele frequency and linkage disequilibrium distributions of these 240 candidate SNPs were determined, analyzed, and compared among themselves and with genotyping data of 752 unrelated healthy Japanese comprising 449 males and 303 females (Kamatani et al. 2004; Unpublished data of the Laboratory for SNP Genotyping, SRC, RIKEN). The implication of the similarities and differences in the number of polymorphic loci, as well as allele frequency and intermarker LD distributions, within and between populations of different ancestry for future pharmacogenetic studies was interpreted.

Materials and methods

Population samples

Blood samples were obtained with consent during various blood donation campaigns held in Kuala Lumpur and the Klang Valley from 270 unrelated healthy individuals (90 comprising 30 females and 60 males each of Malays, Indians, and Chinese) living in Malaysia. Each donor was given a short interview before the donation to ensure their ethnicity as well as family history (not descendant of intermarriage for at least three generations). Although the samples were collected in the capital city of Malaysia, the donors actually originated from various states in Malaysia.

Markers selection

For each individual, 240 sites known to be polymorphic in the Japanese population were screened. These candidate SNPs are located in 109 genes that are drug related, such as genes that encode DMEs and drug transporters, and are distributed on 20 human chromosomes excluding the sex chromosomes (Appendix). For each gene, a total of one to four sites were genotyped. Allele frequencies and genotype frequencies were both evaluated for these 240 sites. For those SNPs located within the same gene, pairwise intermarker LD was estimated through standard measurements (LD coefficient, D' , and difference in proportions, d). LD measurements were not estimated for all possible SNP pairs located on the same chromosome but in different genes.

Genotyping methods

Genomic DNA was extracted from white blood cells by using the Perfect gDNA Blood Mini Kit (Eppendorf, Germany). For each DNA sample, multiplex PCR was performed with 40 ng of genomic DNA in a total reaction volume of 20 μ l. In each multiplex reaction, up to 96 pairs of primers were used for amplification with Ex-Taq DNA polymerase (Takara Bio Inc, Japan) in the presence of TaqStart Antibody (Clontech Laboratories, BD Biosciences, USA), which reduces the formation of primer-dimer (Ohnishi et al. 2001). PCR was carried out with an initial denaturation step of 94°C for 2 min followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 2 min. These cycling reactions were followed by a final extension at 72°C for 7 min. All primer sequences used were the same as those developed for PCR amplifications in the Japanese Single Nucleotide Polymorphisms (JSNP) discovery project (Hirakawa et al. 2001; Haga et al. 2002) and provided by RIKEN. Information about these SNPs can be found on the following Web page, <http://snp.ims.u-tokyo.ac.jp/>. All SNPs were genotyped by the invader assay that combines structure-specific cleavage enzymes with universal fluorescent resonance energy transfer (FRET) system (Ryan et al. 1999; Mein et al. 2000). Allele-specific oligonucleotide pairs were designed and supplied by Third Wave Technologies and provided by RIKEN. FRET probes were labeled with FAM or VIC corresponding to each allele. Signal intensity was indicated as the ratio of FAM or VIC to ROX, an internal reference. Each total reaction volume of 5 μ l contained 0.25 μ l of FRET probe, 0.25 μ l of signal buffer, 0.25 μ l of structure-specific cleavage enzyme, 0.5 μ l of allele-specific probes, 1.75 μ l of distilled water, and 2 μ l of PCR product, which was diluted one in 20. Each PCR-invader reaction was set up in a 96-well format. Samples were incubated in GeneAmp PCR system 9700 (Applied Biosystems, USA) with an initial denaturation at 94°C for 5 min followed by incubation at 63°C for 15 min (Ohnishi et al. 2001). Signal was

detected at the end point with the ABIPRISM 7700 Sequence Detection System.

Statistical analyses

Genotype and allele frequencies for each marker were calculated and tested with the standard *Chi*-square test of Hardy–Weinberg equilibrium (HWE) (Weir 1996) at each locus. SNPs showing deviation from the HWE were eliminated from the following LD calculations and analyses. Haplotype frequencies for each locus were estimated by using the Expectation–Maximization (EM) algorithm (Slatkin and Excoffier 1996). Pairwise inter-marker LD was expressed by D' and d ; $D' = D/D_{\max}$ (Devlin and Risch 1995) and represents LD coefficient whereas d is the difference in proportions as mentioned in Nei and Li (1980) and $d = \pi_{11}/\pi_{.1} - \pi_{12}/\pi_{.2}$, where π_{ij} is the frequency of haplotypes with allele i at the first marker and allele j at the second, and $\pi_{.j}$ is the frequency of haplotypes with either allele at the first marker and allele j at the second. The measures of LD were not calculated when only one allele was observed for at least one SNP in the pair.

Results

Population differences in number of polymorphic loci

Of the 240 sites screened in each individual of the three Malaysian populations investigated, we found that some of these sites were nonpolymorphic in certain Malaysian populations. In Chinese, 226 sites were polymorphic while in Malays and Indians, 222 sites were polymorphic. On the whole, these three populations and the Japanese population shared 216 out of the 240 sites investigated. However, even though the Japanese population has the greatest variability of SNPs, most of them are of low minor-allele frequencies.

Departures from the Hardy–Weinberg equilibrium

Of the 240 sites genotyped, eight SNPs in each Malaysian population showed p values that were statistically

significant at a critical level of 0.01, similar to the Japanese population in which seven markers were found to show violations from the HWE. Most of these violations were private to the corresponding population. Only three out of the 27 deviations were observed in more than one population. This Hardy–Weinberg disequilibrium (HWD) might indicate the existence of population structure in these subject populations. Allele frequencies of all genotyped SNPs are summarized in Appendix with SNPs showing deviation from HWE marked with double asterisks (**).

Differences in minor-allele frequencies according to SNP locations

We eliminated the 27 SNPs that were not in HWE for the following analyses since the subsequent EM-algorithm-based haplotype estimation required HWE assumption (Weir 1996). Among the remaining 213 SNPs, 152 are in the introns of genes and 61 in the exonic regions. Of the 61 exonic SNPs, 33 are in the coding regions. Among them, 16 caused synonymous mutations, 16 resulted in nonsynonymous mutations, and the remaining one caused a nonsense mutation.

On the whole, intronic SNPs were found to have higher minor-allele frequencies. The average minor-allele frequencies for the intronic SNPs of the four populations ranged between 0.224 and 0.230. In contrast, the average minor-allele frequencies of the exonic SNPs ranged between 0.188 and 0.198. Similarly, SNPs that caused synonymous mutations had greater minor-allele frequencies than those that resulted in nonsynonymous mutations; the average minor-allele frequencies of the former and latter SNPs were 0.228 and 0.151, respectively. These observations agreed with the assumption that there might be stronger selection pressure acting on the exonic regions of the genes, especially against nonsynonymous mutations. Generally, we found that minor-allele frequency distributions were similar in the four populations (Table 1), suggesting that differences in minor-allele frequencies according to their locations in genes are independent of ethnical factor.

When these SNPs were resolved into six groups according to their minor-allele frequencies (Table 2), we found that 54% of them had minor-allele frequencies of

Table 1 Average minor-allele frequencies of single nucleotide polymorphisms (SNPs) according to SNP locations and types of mutations

	Japanese	Malays	Indians	Chinese	Average
Intronic SNPs	0.227	0.229	0.224	0.230	0.227
Exonic SNPs	0.188	0.196	0.198	0.192	0.193
Noncoding exonic SNPs	0.201	0.203	0.208	0.198	0.203
Synonymous mutations	0.220	0.229	0.248	0.216	0.228
Nonsynonymous mutations	0.138	0.160	0.140	0.167	0.151
Nonsense mutations	0.008	0.000	0.006	0.006	0.005
Transition, A/G (T/C)	0.206	0.209	0.205	0.210	0.208
Transversion_1, A/C (T/G)	0.229	0.236	0.245	0.234	0.236
Transversion_2, A/T	0.274	0.295	0.310	0.276	0.289
Transversion_3, C/G	0.234	0.239	0.228	0.236	0.234
Total SNPs	0.215	0.220	0.216	0.219	0.218

Table 2 Minor-allele frequency distributions of single nucleotide polymorphisms (SNPs) in Hardy–Weinberg equilibrium (HWE)

	All SNPs in HWE				SNPs with minor-allele frequencies of 0.2–0.5 in the Japanese population			
	Japanese	Malays	Indians	Chinese	Japanese	Malays	Indians	Chinese
$q=0$	0	18	17	14	0	0	0	0
$0 < q \leq 0.1$	50	40	41	41	0	1	13	0
$0.1 < q \leq 0.2$	49	48	49	42	0	23	29	15
$0.2 < q \leq 0.3$	55	41	33	47	55	33	15	32
$0.3 < q \leq 0.4$	31	35	38	43	31	28	27	42
$0.4 < q \leq 0.5$	28	31	35	26	28	29	30	25
Total	213	213	213	213	114	114	114	114

greater than 0.2 in both Japanese and Malaysian Chinese. On the other hand, for Malaysian Malays and Indians, 50% of these SNPs had minor-allele frequencies of more than 0.2. Since a previous study (Kruglyak 1997) revealed that markers with minor-allele frequencies of 0.2–0.5 have equivalent information content whereas markers with minor-allele frequencies below 0.2 are less informative, we further eliminated those SNPs with minor-allele frequencies of less than 0.2 in the Japanese population. As a result, 114 SNPs remained,

and we observed that 79% and 87% of them had minor-allele frequencies of 0.2–0.5 in Malaysian Malays and Chinese, respectively. In contrast, only 63% of the remaining SNPs in Malaysian Indians had minor-allele frequencies of 0.2–0.5.

Population differences in allele frequency distributions

When minor-allele frequencies of these SNPs were arranged in an upstream pattern and resolved into histograms (Fig. 1), striking patterns were observed. These population-specific histograms show at a glance that there was a greater similarity between the minor-allele frequency distributions in Japanese and Malaysian Chinese. With some exceptions observed for some of the markers, the minor-allele frequency distribution in Malaysian Malays resembled the minor-allele frequency

Fig. 1 Minor-allele frequencies of 213 single nucleotide polymorphisms (SNPs) that remained after elimination of SNPs showing deviations from the Hardy–Weinberg equilibrium (HWE). SNPs were arranged in an upstream pattern according to their minor-allele frequencies detected in the Japanese population. The minor-allele frequency distributions in **a** Japanese, **d** Malaysian Chinese, and **b** Malaysian Malays showed great similarities. **c** Malaysian Indians showed minor-allele frequency distribution distinctly different from the other three populations

distributions in Japanese and Malaysian Chinese. However, the minor-allele frequency distribution in Malaysian Indians was distinctively different from the other three populations under investigation. Similar observation was also denoted by pairwise correlation graphs comparing allele frequencies of each marker in these four populations (Appendix).

In general, allele frequencies were highly correlated among Japanese, Malaysian Malays, and Chinese. In contrast, both comparisons involving Japanese and Malaysian Chinese with Malaysian Indians showed results of distinct difference. However, correlation of allele frequencies between Malay and Indian populations was comparatively high. Besides, we also performed the standard *Chi*-square test for pairwise comparisons of allele and genotype frequencies between each population. Results were shown in the Appendix. Significant differences between populations were each indicated with an asterisk (*).

Linkage disequilibrium and distances between SNPs

By considering only those SNPs shared by the four populations and eliminating those SNPs showing deviation from the HWE, we calculated the LD coefficient, D' , for each SNP pair located in the same gene. Plots of D' versus physical distance between two SNPs in the same pair (Fig. 2) show that D' decayed rapidly with an increase in distance. We observed a dramatic drop in D' when physical distance between SNP pairs increased to 50–100 kb. D' value for markers 50–100 kb apart was about half that for markers 10–50 kb apart. This finding is similar to that reported by Dunning et al. (2000) though at a different extent. However, we also found closely located markers that were not in high LD. Similar results were observed for all four populations, suggesting that the relationship between distance and D' is a universal characteristic, which is ethnic-independent. Another interesting observation was the sudden increase

in LD for markers more than 100 kb apart. This phenomenon has been observed in all populations under investigation and may reflect false LD. Therefore, the 11 SNP pairs with physical distance between SNPs of more than 100 kb have been further eliminated from the following study.

Differences in linkage disequilibrium distributions

After SNP pairs with inter-SNP distance of more than 100 kb were removed, 101 SNP pairs were left for further analysis. The d values were calculated for different SNP pairs in individuals of each population and then resolved into plots of correlation, as shown by the six graphs in the lower triangle in Fig. 3. Subsequently, 38 SNP pairs were removed by using a minor allele frequency cutoff at 0.1. Similarly, d values for the remaining 63 SNP pairs were resolved into plots of correlation and were represented by six graphs in the upper triangle in Fig. 3. There was an apparent increase in the coefficient of determination values, R^2 , for each correlation graph when those SNPs with minor allele frequencies of less than 0.1 were eliminated from the plots.

Similarly, all correlation graphs that compared Japanese and Malaysian Chinese, Malaysian Malays and Chinese, Malaysian Malays and Indians, and Japanese and Malaysian Malays showed higher similarity regardless of minor-allele frequency cutoff. In contrast, both comparisons that involved Japanese and Malaysian Chinese with Malaysian Indians showed comparatively lower similarity. Again, correlation between Malays and Indians remained high.

Discussion

The fact that not all investigated sites were polymorphic in the three Malaysian populations may not imply that

Fig. 2 Relationship between LD coefficient (D') and physical distance separating two single nucleotide polymorphisms (SNPs) in a pair

Fig. 3 Pairwise comparisons of difference in proportions (d) values among four investigated populations after elimination of single nucleotide polymorphisms (SNPs) in Hardy–Weinberg disequilibrium (HWD) and SNPs with an inter-SNP distance of more than 100 kb. Each of six graphs in the lower triangle represents correlation of d between two populations by using 101 SNP pairs without minor-allele frequency cutoff. Each of six graphs in the upper triangle represents correlation of d between two populations by using the remaining 63 SNP pairs after elimination of SNPs with minor-allele frequencies of less than 0.1. The coefficient of determination value for each correlation, R^2 , is shown at the lower right corner of each plot

the Japanese population has more variation or has higher antiquity compared with other populations (Stephens et al. 2001). In fact, this may most possibly result from the ascertainment bias during SNP discovery. Since the investigated sites were previously discovered by using only 48 individuals of Japanese descent (Iida et al. 2001a,b,c, 2002a,b,c; Saito et al. 2001a,b, 2002a,b,c,d, 2003; Unpublished data of Laboratory for Genotyping, SRC) and then scored in 752 Japanese individuals (Kamatani et al. 2004; Unpublished data of Laboratory for Genotyping, SRC), more often than not, they tend to have higher population frequencies (Clark et al. 2003). Furthermore, since they were genotyped but not resequenced in the other three Malaysian populations, there is a tendency to introduce a bias toward lower frequencies in other populations (Weiss and Clark 2002).

Because DMEs are evolutionarily very old enzymes that existed before the divergence of eubacteria from eukaryotes (Nebert 1997), those seemingly population-specific SNPs might have been newly arisen, at least after populations' divergence (Kalow 2001). In addition to that, since genetic drift usually accumulates with time, this interpopulation variability might also indicate that these SNPs represent ancient polymorphisms with one of their alleles fixed while another drifted away and no longer existing in a particular population (Cargill et al. 1999; Halushka et al. 1999).

However, though different in the number of polymorphic sites, the extent of interpopulation variation is limited. This is reasonable, as DMEs are responsible for critical life functions (Ingelman-Sundberg et al. 1999). Therefore, genes encoding them might confer limited changes since their emergence because they might be subjected to higher constraints of selection pressures,

especially in the exonic regions. This assumption is supported by our observation that for all four populations studied, exonic SNPs, especially those causing nonsynonymous mutations, always have lower minor-allele frequencies compared with intronic SNPs. This observation is also compatible with previous reports (Cargill et al. 1999; Halushka et al. 1999; Goddard et al. 2000).

When we confined our subjected SNPs to those with minor-allele frequencies of greater than 0.2 in the Japanese population, we observed that about 80% of them had minor-allele frequencies of 0.2–0.5 in Malaysian Malays and Chinese. Since these SNPs have equivalent information content in association studies (Kruglyak 1997), and it is laborious and cost consuming to establish population-specific SNP databases at the present, this observation suggested the possibility and rational of using currently existing JSNP database as a reference to accelerate future pharmacogenetic studies involving Malaysian Malays and Chinese.

Population-specific histograms and pairwise correlation graphs of allele frequencies of these SNPs for the four subject populations (graphs not shown; the coefficients of determination values, R^2 , for each correlation graph were shown in Appendix) suggested classifying them into two distinctive genetic clusters of different ancestries. The first group consists of three populations: Japanese, Malaysian Malays, and Chinese, while the second comprises Malaysian Indians. In both cases, greater similarities have been observed between Japanese, Malaysian Chinese, and Malays. Correlations between Japanese and Malaysian Chinese with Indians remained low all the time.

Generally, allele frequencies among populations are correlated overall, suggesting that these SNPs originated

from an identical ancestral haplotype. But with time, these populations deviated not only from the ancestral condition but also from each other under the pressure of mutation, random genetic drift, and natural selection. Similarly, the number of common alleles shared by these populations gradually declined. Hence, various degrees of correlation of allele frequency observed for pairwise comparisons are good indicators to infer the degree of drift from the point when two groups separated.

The result goes along well with evidences from recent study of paleontology as well as anthropology, which suggested that Japanese [with a Northeast Asian origin (Nei 1995)], Malaysian Malays [the oldest indigenous people who inhabit the Malay Peninsula and are believed to be originated from the Northwestern part of Yunnan in China (Skeat 1902), and a distant offshoot of the Mongolian stock (Winstedt 1961)], and Chinese [descendants of immigrants from the Southern part of China who settled in number in the Malay Peninsula in the nineteenth century (Ryan 1971)] shared a more recent origin (Nei 1985), while Malaysian Indians, who originated from India, belong to another distinct genetic group: Caucasoid. The result was further supported by the recent study of ancient remains found in the Southern part of China, Minatogawa in Okinawa, Japan, and Niah Cave in Borneo, Malaysia, which showed that there were genetic contacts between people living in these areas (Bodmer and Cavalli-Sforza 1976; Cavalli-Sforza et al. 1994; Etler 1996).

Though the observation that Malaysian Malays also shared a high similarity with Malaysian Indians may render the above classification obscure, it can be explained by the histories of these populations in the Peninsula of Malaysia. The geographic location of Malaysia, which is halfway on the sea route between China and India, has established her political and cultural contacts with both of these countries over the past 2,000 years (Ryan 1971). Although Indians only settled en mass in the Peninsula of Malaysia in the nineteenth century (Nagata 1979), Malaysia has received her early civilization and two great religions, Hinduism and Islam, from India. Apart from the evidences of early Indian settlements in some of the states in Malaysia, the unremitting traffic of traders between the Archipelago and India has resulted in frequent intermarriages between the older-day Indians and Malays along the whole of the west coast of the peninsula (Ryan 1971).

In contrast, early contacts between the Malay Peninsula and China were largely confined to trades. Differences in cultures rendered limited intermarriages between Chinese and Malays. Apart from that, migration of citizens from China was discouraged by the Manchu government (1641–1911). Noticeable increase in emigration only came in the mid-nineteenth century (Nagata 1979) and the majority of the emigrants came from the southern provinces of Kwangtung and Fukien, which were nearest to Southeast Asia and most opposed to Manchu rule (Ryan 1971).

In fact, our preliminary analyses of the genotyping data by using the program STRUCTURE (Pritchard et al. 2000) revealed that Japanese, Malaysian Malays, and Chinese had similar genetic structures, with Malays genetically heterogeneous and showing substantial similarities with Malaysian Indians. On the other hand, Malaysian Indians were clustered into another distinct genetic group similar with that reported by Rosenberg et al. (2002). Consistent with their population histories, the admixture event between Malays and Indians, as mentioned above, is believed to have occurred quite some time ago between older-day Malays and Indians, with contemporary Malays representing the descendants of this admixture. However, present-day Malaysian Indians are descendants of Indian emigrants during the nineteenth century. This is supported by the fact that all donors in the present study were interviewed before blood donation to ensure that they were not descendants of mixed marriages for at least the nearest two generations.

Another plausible explanation in elucidating the substantial similarities shared between Malaysian Malay and Indian populations may be related with ecogenetic factors. Although genetic variations observed between populations were created by spontaneous mutations and passed on to the next generation under the chance event of random genetic drift, both of these events per se can hardly explain the existence of substantial allele frequency differences between populations (Nebert 1997). Although the higher generation turnover of the Indian population owing to early age at marriage may leave some effects, with the relatively low rates of occurrence of both events mentioned above and the long generation time for human population, other forces such as migration and natural selection play a more important role in shaping the distribution patterns of these genetic variations in different populations with regard to the environment.

For instance, the differences observed across various populations may reflect the different impacts that ecogenetic factors have on people from different geographic regions. In this case, the highest selection pressure is believed to be of dietary origin (Ingelman-Sundberg et al. 1999). It is possible for striking ethnic differences in DME polymorphisms to have arisen from selective pressures caused by tribal differences in diet or exposure to other environmental agents/signals in a span of 200–2,000 years that would represent 10–50 generations for the human population (Nebert 1997). This might also suggest the possibility of the existence of balanced polymorphisms, which are unappreciated at the moment.

Here we presented two sets of correlation graphs to compare the d values of the four populations by using SNP pairs with and without minor-allele frequency cutoff. Since SNPs of lower minor-allele frequencies are in general those of younger age (Watterson and Guess 1977), applying minor-allele frequency cutoff is likened to eliminating recent SNPs of younger age. Though this

may also risk losing some ancient SNPs with one of their alleles fixed in the population, Halushka et al. (1999) found that only 5% of the rare SNPs that they investigated represent ancient allele that is being lost through evolution.

Increasing coefficient of determination values, R^2 , for all correlation graphs without SNPs with minor-allele frequency of less than 0.1 further supports the single-origin hypothesis of human populations (Lewin 1987) and reflects that those young variants are shared by all populations and arose prior to population divergence, their fate in evolution varies as different environmental factors (cultural and habitual) act on them in different populations. Their elimination in this study therefore provides higher stringency to the study of population ancestry.

On the whole, inference from correlations of inter-marker LD among these populations further supports the results that we obtained from the allele frequency distributions both on population ancestries and the single-origin hypothesis of the investigated populations. The results also showed that the genetic distances suggested by the coefficient of determination values for correlations of allele frequency as well as LD index of different populations seemed to agree with the historical facts of their population origins.

In conclusion, we have shown here the existence of extensive similarities between populations of similar ancestry, such as the number of polymorphic loci shared, allele frequency distributions, as well as LD index, despite some minor variations that might be due to extensive interactions among genetic (population ancestry), environmental (diet), and habitual (like smoking and alcohol consumption) factors, which will

require further investigations. We have also shown that since most of those shared SNPs are of equivalent information content, and because it is laborious and cost consuming to establish population-specific SNP databases by using present-day technologies, our observation suggested the possibility and rational of using existing SNP database as a reference to hasten future pharmacogenetic studies involving populations of similar ancestry.

For instance, a genetic study on a clinically significant variant in response to a drug by using the Japanese samples will allow us to make a better and more relevant prediction on Malays and Chinese compared with Indians. Therefore, a better understanding in knowledge of pharmacanthropology not only promises to reduce the time and costs associated with producing effective, marketable, and competitive new drugs (Kleyn and Vesell 1998) but also allows the development of drugs *a priori* that will avoid metabolic pathways with adverse genetic variability as well as by modifying drug selection/dosing in patients with clinically significant genetic variation.

Acknowledgements We thank the Laboratory for Genotyping, SRC, RIKEN, Japan, for unpublished data and the Blood Transfusion Unit and Blood Bank of the University of Malaya Medical Centre, Kuala Lumpur, Malaysia, for permission to collect blood samples. We also thank Hong-Mae Nan, Shigeru Otsubo, Keiko Yamazaki, Fumihiro Akiyama, Ching-Ching Ng, Tatsuhiko Tsunoda, Takuya Kitamoto, Aritoshi Iida, and Susumu Saito for stimulating discussion and the two anonymous reviewers for helpful comments. This work was supported by Research for the Future Program grant #00L01402 from the Japan Society for the Promotion of Science, Japan, and research grants F0151/2004A from the University of Malaya and IRPA TD 36-02-03-6006 from the Ministry of Science, Technology, and Innovation, Malaysia.

Appendix

Tables 3, 4.

Table 3 Allele frequencies of 240 SNPs and results of pairwise comparisons between different populations

SNP	Gene	Allele frequency				Significant at 0.01 level for pairwise comparisons between populations					
		Japanese	Malays	Indians	Chinese	Japanese-Malays	Japanese-Indians	Japanese-Chinese	Malays-Indians	Malays-Chinese	Indians-Chinese
1	ABCA4 ^a	0.41	0.42	0.40	0.37						
2		1.00	0.99	1.00	1.00						
3		0.11	0.09	0.12	0.09						
4	ABCA7 ^a	0.63	0.84	0.84	0.68	*	*			*	*
5		0.56	0.43	0.42	0.58	*	*			*	*
6		0.56	0.44	0.42	0.58	*	*			*	*
7	ABCB1 ^b	0.02	0.00	0.03	0.01						
8		0.88	0.80	0.82	0.90	*	*				
9		0.61	0.56	0.65	0.62						
10	ABCB10 ^b	0.86	0.87	0.82	0.81						
11	ABCB4 ^b	0.70	0.80	0.88**	0.67	*	*			*	*
12		0.25	0.24	0.13	0.30		*				*
13	ABCB8 ^b	0.08	0.09	0.03	0.11				*		*
14		0.92	0.91	0.97	0.91				*		
15	ABCB9 ^b	0.05	0.05	0.14	0.06		*		*		
16		0.29	0.39	0.41	0.31	*	*				
17		0.30	0.41	0.42	0.31	*	*				
18	ABCC1 ^c	0.65	0.72	0.66	0.61					*	
19		0.60	0.61	0.56	0.69						
20		0.76	0.85	0.66	0.83	*	*		*		*
21	ABCC2 ^c	0.80	0.79	0.60	0.74		*		*		*
22		0.78	0.79	0.64	0.74		*		*		
23	ABCC3 ^c	0.77	0.76	0.86	0.76		*		*		
24		0.81	0.73	0.79	0.84	*					
25		0.96	0.92	0.88	0.96	*	*				*
26	ABCC4 ^c	0.38	0.23	0.34	0.22	*		*			
27		0.54	0.50	0.51	0.54						
28	ABCC5 ^c	0.65	0.64	0.81**	0.56		*		*		*
29		0.77	0.82	0.82**	0.81		*				
30		0.33**	0.24**	0.20	0.21**	*	*	*			
31	ABCC7/CFTR ^c	0.30	0.38	0.39	0.34		*				
32		0.73	0.74	0.83	0.73		*				
33	ABCC8 ^c	0.64	0.82	0.81	0.63	*	*			*	*
34		0.61	0.59	0.64	0.63						
35		0.85	0.83	0.63**	0.86		*		*		*
36	ABCC9 ^c	0.69	0.60	0.53	0.74		*		*		*
37		0.84	0.81	0.69	0.83		*				*
38	ABCD3 ^a	0.02	0.00	0.03	0.01						
39		0.77	0.85	0.79	0.79						
40	ABCD4 ^a	0.00	0.03	0.01	0.01	*					
41		0.02	0.04	0.06	0.05						
42	ABCE1 ^a	1.00	1.00	1.00	1.00						
43		0.94	0.99	1.00	0.99	*	*	*			
44	ABCF1 ^a	0.28	0.20	0.13	0.33		*			*	*
45	ABCG1 ^a	0.24	0.16	0.14	0.15		*	*			
46		0.38	0.54	0.48	0.48**	*		*			
47		0.38	0.54	0.46	0.50**	*		*			
48	ABCG2 ^a	0.10	0.27	0.22**	0.24	*	*	*			*
49		0.13	0.19	0.11	0.16						
50		0.15	0.32	0.38	0.28	*	*	*			
51	ADH1 ^d	0.93	0.83	0.65	0.89	*	*		*		*
52		0.92	0.94	0.97	0.92**						
53	ADH3 ^d	0.11	0.37	0.41	0.25	*	*	*			*
54		0.11	0.65	0.52	0.25	*	*	*		*	*
55	ALDH1A1 ^c	0.47	0.60	0.61	0.46	*	*			*	*
56		0.60	0.48	0.46	0.56	*	*				
57		0.60	0.48	0.46	0.55	*	*				
58	ALDH1A2 ^c	0.71	0.71	0.81	0.66		*				*
59		0.67	0.70	0.70	0.54**			*		*	*

Table 3 (Continued)

SNP	Gene	Allele frequency				Significant at 0.01 level for pairwise comparisons between populations					
		Japanese	Malays	Indians	Chinese	Japanese-Malays	Japanese-Indians	Japanese-Chinese	Malays-Indians	Malays-Chinese	Indians-Chinese
60	ALDH1A3 ^c	0.45	0.44	0.14	0.42		*		*		*
61		0.93	0.99	0.99	0.99	*	*	*			
62		0.55**	0.56	0.88	0.55		*		*		*
63	ALDH1B1 ^c	0.19	0.21	0.22	0.26**			*			
64		0.26	0.31**	0.33	0.21	*				*	*
65		0.99	1.00	1.00	1.00						
66		0.32	0.26	0.15	0.31		*				*
67	ALDH1L1 ^c	0.24	0.19	0.11	0.20		*				
68		0.24	0.18	0.11	0.20		*				
69		0.43	0.38	0.59	0.35		*		*		*
70	ALDH2 ^c	0.84	0.77	0.72	0.78		*				
71		0.84	0.77	0.73	0.77		*				
72		0.24	0.01	0.01	0.28	*	*			*	*
73	ALDH3A1 ^c	0.89	0.92	1.00	0.91		*		*		*
74		0.11	0.07	0.01	0.08		*		*		*
75	ALDH3A2 ^c	0.42	0.56	0.44	0.52	*		*	*		
76		0.41	0.59	0.44	0.52	*		*	*		
77	ALDH3B1 ^c	0.75	0.81	0.84	0.79		*				
78	ALDH3B2 ^c	0.19	0.46	0.09	0.21	*	*		*	*	*
79		0.02	0.03	0.03	0.01						
80	ALDH5A1 ^c	0.02	0.00	0.02	0.01						
81		0.03	0.07	0.13	0.06	*	*	*			
82	ALDH6A1 ^c	0.67	0.72	0.43	0.75		*		*		*
83		0.00	0.00	0.00	0.00						
84		0.23	0.19	0.37	0.18		*		*		*
85	ALDH8A1 ^c	0.23**	0.21	0.20**	0.27						
86		0.73	0.71	0.70	0.74						
87	ALDH9A1 ^c	0.02	0.04	0.15	0.02	*	*		*		*
88		0.98	1.00	0.98	0.98						
89	CES1 ^c	0.24**	0.29	0.43	0.21		*		*		*
90		0.58**	0.58	0.75	0.53		*		*		*
91		0.86	0.99	0.99	0.90	*	*			*	*
92	CES2 ^c	0.77	0.85	0.99	0.86		*	*	*		*
93		0.21	0.24	0.05	0.25		*		*		*
94	CHST1 ^f	0.44	0.58	0.53	0.53	*	*				
95		0.40	0.26	0.10	0.24	*	*	*	*		*
96	CHST5 ^g	0.75	0.72	0.74	0.78						
97		0.78	0.79	0.89	0.82		*				
98	CST ^g	0.77	0.83	0.67	0.71		*		*	*	
99		0.35	0.37	0.21	0.36		*		*		*
100		0.15	0.13	0.07	0.11		*				
101	CYP1A1 ^c	0.16	0.05	0.00	0.13	*	*		*	*	*
102		0.78	0.74	0.90	0.72		*		*		*
103	CYP1A2 ^c	0.96	0.94	0.94	0.88			*			
104		0.97	0.95	0.94	0.88		*	*			
105		0.21	0.16	0.11	0.12		*	*			
106	CYP1B1 ^c	0.10	0.36	0.46	0.21	*	*	*		*	*
107		0.87	0.93	0.91	0.94			*			
108		0.88	0.94	0.93	0.94						
109	CYP27A1 ^c	0.94	0.95**	0.96	0.87	*		*		*	*
110		0.94	0.96**	0.95	0.87	*		*		*	
111		0.16	0.17	0.31	0.22		*		*		
112	CYP2A13 ^c	0.93	0.86	0.98**	0.93	*	*		*		*
113		0.99	1.00	0.99	0.99						
114	CYP2B6 ^c	0.05	0.08	0.05	0.07						
115		0.17	0.66	0.73	0.48	*	*	*		*	*
116	CYP2E ^c	0.45	0.64	0.60	0.57	*	*	*			
117		0.44	0.56	0.49	0.45	*				*	
118	CYP2S1 ^c	0.71	0.72	0.80	0.77						
119		0.28	0.21	0.09	0.19		*		*		*
120	CYP3A4 ^h	0.24	0.44	0.32	0.26	*				*	
121		0.21	0.24	0.24	0.20						
122	CYP3A43 ^h	0.01	0.02**	0.00	0.02**	*		*			
123		1.00	1.00	1.00	1.00						

Table 3 (Continued)

SNP	Gene	Allele frequency				Significant at 0.01 level for pairwise comparisons between populations					
		Japanese	Malays	Indians	Chinese	Japanese-Malays	Japanese-Indians	Japanese-Chinese	Malays-Indians	Malays-Chinese	Indians-Chinese
124	CYP3A5 ^h	1.00	1.00	1.00	1.00						
125		0.01	0.02	0.01	0.00						
126	CYP3A7 ^h	0.25	0.38**	0.34	0.23	*	*			*	
127		0.25	0.38	0.37	0.24	*	*			*	*
128	CYP4B1 ^h	0.23	0.13	0.10	0.13	*	*	*			
129		0.17	0.28	0.22	0.28	*	*	*			
130		0.17	0.38	0.36	0.28	*	*	*			
131	CYP4F2 ^h	0.87	0.94	0.79	0.94	*	*	*	*		*
132		0.13	0.06	0.22	0.06	*	*	*	*		*
133	CYP4F3 ^h	0.62	0.71	0.62	0.62**					*	
134		0.14	0.07	0.18	0.13				*		
135		0.35	0.35	0.59	0.42		*		*		*
136	CYP4F8 ^h	0.55	0.58	0.77	0.61		*		*		*
137		0.31	0.24	0.41	0.28		*		*		*
138		0.69	0.78	0.60	0.73				*		*
139	CYP5A1/TBXAS1 ^c	0.92	0.93**	0.85	0.92		*		*		
140	CYP7A1 ^e	0.06	0.02	0.05	0.07						
141	EPHX1 ⁱ	0.16	0.16	0.11	0.16						
142		0.83	0.83	0.89	0.84						
143	EPHX2 ⁱ	0.72	0.71	0.84	0.81		*		*		
144	GSTA1 ^d	0.88	0.83	0.86	0.92					*	
145		0.88	0.80	0.69	0.92	*	*			*	*
146	GSTA4 ^d	0.05	0.14	0.02	0.14	*		*	*		*
147		0.06	0.04	0.08	0.05						
148	GSTM2 ^d	0.04**	0.03	0.15	0.02		*		*		*
149		0.01	0.00	0.00	0.00						
150	GSTM3 ^d	0.25	0.34	0.55	0.32	*	*		*		*
151		0.24	0.33	0.56	0.31	*	*		*		*
152	GSTM4 ^d	0.98	0.96	0.88	0.97	*	*		*		*
153		0.04	0.02	0.12	0.02		*		*		*
154	GSTP1 ^d	0.14	0.28	0.27	0.21	*	*	*			
155		0.14	0.27	0.21	0.20	*	*	*			
156	GSTZ1 ^d	0.52	0.63	0.84	0.53	*	*		*		*
157		0.03	0.09	0.18	0.04	*	*				*
158		0.48	0.37	0.16	0.46	*	*		*		*
159	GZMB ^c	0.80	0.58	0.57	0.72	*	*	*		*	*
160		0.80	0.58	0.57	0.72	*	*			*	*
161		0.81	0.61	0.58	0.66	*	*	*			
162	HEP27 ^d	0.01	0.00	0.00	0.00						
163		0.99	1.00	1.00	1.00						
164	MGST1L1 ^d	0.15	0.15	0.20	0.19						
165		0.21	0.16	0.18	0.21						
166	MGST2 ^d	0.55	0.64	0.78	0.63		*		*		*
167		0.56	0.46	0.29	0.61	*	*		*	*	*
168	MGST3 ^d	0.17	0.28	0.36	0.21	*	*				*
169		0.76	0.68	0.58	0.69		*				
170	NDUFA10 ^j	0.35	0.37	0.15	0.41		*		*		*
171		0.36	0.37	0.16	0.42		*		*		*
172	NDUFA2 ^d	0.50	0.45	0.47	0.56						
173	NDUFA3 ^d	0.28	0.19	0.07	0.34	*	*		*	*	*
174		1.00	1.00	1.00	1.00						
175	NDUFA5 ^d	0.70	0.61	0.76	0.67				*		
176	NDUFA6 ^d	0.17**	0.14	0.35	0.17		*		*		*
177		0.82	0.84	0.66	0.82		*		*		*
178		0.18	0.16	0.38	0.17		*		*		*
179	NDUFA8 ^d	0.24	0.28	0.44	0.25		*		*		*
180		0.60	0.49	0.39	0.67	*	*			*	*
181	NDUFB7 ^d	0.72	0.74	0.48	0.64		*		*		*
182		0.00	0.00	0.00	0.00						
183		0.89	0.88	0.77	0.89		*		*		*
184	NDUFS1 ^d	0.74	0.79	0.70	0.71						
185	NDUFS3 ^d	0.44	0.23	0.09	0.29	*	*	*	*		*
186		0.01	0.00	0.00	0.00						
187	NDUFS4 ^d	0.58	0.59	0.66	0.61						
188		0.29	0.26	0.21	0.19			*			
189		0.70	0.73	0.75	0.78			*			

Table 3 (Continued)

SNP	Gene	Allele frequency				Significant at 0.01 level for pairwise comparisons between populations					
		Japanese	Malays	Indians	Chinese	Japanese-Malays	Japanese-Indians	Japanese-Chinese	Malays-Indians	Malays-Chinese	Indians-Chinese
190	NDUFS8 ^d	0.76	0.80	0.85	0.80						
191		0.77	0.80	0.86	0.80		*				
192	NQO2 ^g	0.38	0.37	0.24	0.33		*				
193		0.40	0.45	0.29	0.37		*		*		
194	NTCP1/SLC10A1 ^k	0.12	0.23	0.09	0.12	*			*	*	
195		0.09	0.04	0.07	0.13					*	
196	OAT1/SLC22A6 ^l	0.43	0.28	0.15	0.41	*	*		*	*	*
197		0.59	0.72	0.84	0.60	*	*		*		*
198	OAT2/SLC22A7 ^l	0.07	0.04	0.07	0.04						
199		0.34	0.44	0.34	0.38						
200		0.93	0.95	0.87	0.94						
201	OAT3/SLC22A8 ^l	0.23	0.23	0.19	0.24						
202		0.23	0.16	0.10	0.26		*				*
203	OATP1/SLCO1C1 ^l	0.30	0.40	0.30	0.40	*		*			
204		0.30	0.39	0.29	0.41			*			
205	OCT1/SLC22A1 ^k	0.60	0.72	0.84	0.65	*	*		*		*
206		0.30	0.34	0.38	0.34						
207		0.69	0.64	0.61	0.63						
208		0.15	0.20	0.20	0.17						
209	OCT2/SLC22A2 ^k	0.17	0.18	0.28	0.16		*				*
210		0.83	0.81	0.75	0.79		*				
211		0.04	0.07	0.11	0.04		*				
212	OCTN1/SLC22A4 ^k	0.34	0.19**	0.04	0.34	*	*		*	*	*
213		0.33	0.15	0.04	0.32	*	*		*	*	*
214	OCTN2/SLC22A5 ^k	0.65	0.52	0.38	0.63	*	*		*	*	*
215		0.66	0.84	0.95	0.67	*	*		*	*	*
216		0.04	0.17	0.28	0.12	*	*	*	*	*	*
217	PEMT ^m	0.48	0.49	0.33	0.47		*		*		*
218		0.59	0.62	0.66	0.53						
219	PEPT1/SLC15A1 ^k	0.09	0.16	0.20	0.10	*	*				*
220		0.38	0.31	0.32	0.42						
221		0.38	0.33	0.38	0.38						
222	STE/SULT1E1 ^g	0.91	0.98	0.97**	0.95	*	*				
223	SULT1A2 ^g	0.88	0.96	0.83	0.96	*	*	*	*		*
224	SULT2A1 ^g	0.79	0.85	0.91	0.78		*				*
225	SULT1C1 ^g	0.94	0.86	0.65	0.94	*	*		*		*
226	SULT2B1 ^g	0.28	0.22	0.18	0.19		*	*			
227		0.25	0.20	0.41	0.25		*		*		*
228	SULTX3 ^g	0.07	0.14	0.28	0.12	*	*		*		*
229		0.95	0.91	0.76	0.94		*		*		*
230	TAP2 ^b	0.57	0.42	0.51	0.49	*			*		
231		0.61	0.67	0.61	0.63						
232	TGM1 ^c	0.98	1.00	1.00	1.00						
233		0.89	0.91	0.87	0.90						
234		0.17	0.09	0.17	0.12	*					
235	UGT1A1/UGT1 ^j	0.04	0.00	0.00	0.04	*				*	*
236		0.77	0.64	0.42	0.67	*	*	*	*		*
237		0.89	0.85	0.60	0.89	*	*		*		*
238	UGT2A1 ⁿ	0.53	0.56	0.47	0.50						
239	UGT2B15 ⁿ	0.79	0.70	0.90	0.67	*	*	*	*		*
240	UGT8 ⁿ	0.13	0.10	0.05	0.13		*				*
Total						92	143	48	83	40	104
Coefficient of determination, R^2 , for correlation allele frequency graphs						0.88	0.78	0.95	0.89	0.94	0.83

*Marker significantly different between two tested populations ($p \leq 0.01$)

**Marker in Hardy–Weinberg disequilibrium ($p \leq 0.01$)

^aIida et al. 2002b

^bSaito et al. 2002a

^cSaito et al. 2002b

^dIida et al. 2001b

^eSaito et al. 2003

^fIida et al. 2002a

^gIida et al. 2001a

^hSaito et al. 2002c

ⁱSaito et al. 2001a

^jUnpublished data of the Laboratory of SNP Genotyping, SRC, RIKEN

^kSaito et al. 2002d

^lIida et al. 2001c

^mSaito et al. 2001b

ⁿIida et al. 2002c

Table 4 Abbreviations used and related functions of genes

Gene symbol	Gene name	Example
ABC	ATP-binding cassette	ABCC5 (ATP-binding cassette, subfamily C, member 5) Resistance to thiopurine anticancer drugs, 6-mercaptopurine and thioguanine, and the anti-HIV drug 9-(2-phosphonylmethoxyethyl) adenine
ADH	Alcohol dehydrogenase	ADH1 (Alcohol dehydrogenase 1A (class I), alpha polypeptide) High activity for ethanol oxidation Catabolism of ethanol
ALDH	Aldehyde dehydrogenase	ALDH6A1 (Aldehyde dehydrogenase 6 family, member A1) Plays a role in the valine and pyrimidine catabolic pathways Cataclysm of malonate and methylmalonate semialdehydes to acetyl-CoA and propionyl-CoA
	Carboxylesterase	CES1 [Carboxylesterase 1 (monocyte/macrophage serine esterase 1)] Hydrolysis of aromatic and aliphatic esters Cellular cholesterol esterification Detoxification in the lung Protection of the central nervous system from ester or amide compounds
CHST	Carbohydrate sulfotransferase	CHST1 [Carbohydrate (keratan sulfate Gal-6) sulfotransferase 1] Generation of optimal L-selectin ligands in vascular endothelial cells at sites of inflammation
CST	Cerebroside (3'-phosphoadenylylsulfate: galactosylceramide 3') sulfotransferase	CST [Cerebroside (3'-phosphoadenylylsulfate: galactosylceramide 3') sulfotransferase] A cysteine proteinase inhibitor found in saliva, tears, urine, and seminal fluid
CYP	Cytochrome P450	CYP2A13 (Cytochrome P450, family 2, subfamily A, polypeptide 13) Metabolism of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, a major nitrosamine specific to tobacco
EPHX	Epoxide hydrolase	EPHX1 [Epoxide hydrolase 1, microsomal (xenobiotic)] Activation and detoxification of exogenous chemicals such as polycyclic aromatic hydrocarbons
GST	Glutathione S-transferase	GSTA1 (Glutathione S-transferase A1) Metabolism of bilirubin and certain anticancer drugs in the liver Protecting the cells from reactive oxygen species and the products of peroxidation
GZMB	Granzyme	GZMB [Granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)] Target cell lysis in cell-mediated immune responses Potent apoptosis-inducing serine protease of cytotoxic lymphocytes
HEP	Short-chain dehydrogenase/reductase	HEP27 [Dehydrogenase/reductase (SDR family) member 2] Gene with protein product, function known or inferred
MGST	Microsomal glutathione S-transferase	MGST3 (Microsomal glutathione S-transferase 3) Cataclysm of the conjugation of leukotriene A4 Reduced glutathione to produce leukotriene C4
MGST1L1	Prostaglandin E synthase	MGST1L1 (prostaglandin E synthase) Metabolism of eicosanoid and glutathione
NDUF	NADH dehydrogenase	NDUFS1 [NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75 kDa (NADH-coenzyme Q reductase)] It may form part of the active site crevice where NADH is oxidized
NQO	NAD(P)H dehydrogenas	NQO2 [NAD(P)H dehydrogenase, quinone 2] Gene with protein product, function known or inferred
SLC	Solute carrier family	OCTN2/SLC22A5 [Solute carrier family 22 (organic cation transporter), member 5] An organic cation transporter A sodium-dependent high affinity carnitine transporter Active cellular uptake of carnitine
PEMT	Phosphatidylethanolamine N-methyltransferase	PEMT (Phosphatidylethanolamine N-methyltransferase) Conversion of phosphatidylethanolamine to phosphatidylcholine by sequential methylation in the liver
STE	Sulfotransferase, estrogen-preferring	STE (Sulfotransferase, estrogen-preferring) Transfers a sulfo moiety to and from estrone, which may control levels of estrogen receptors
SULT	Sulfotransferase family	SULT2B1 (Sulfotransferase family, cytosolic, 2B, phenol-preferring, member 1) Sulfates dehydroepiandrosterone
TAP	Transporter for ATP-binding cassette	TAP2 (Transporter 2, ATP-binding cassette, sub-family B) Antigen presentation Transportation of peptides from the cytoplasm to the endoplasmic reticulum

Table 4 (Continued)

Gene symbol	Gene name	Example
TGM	Transglutaminase	TGM1 [Transglutaminase 1 (K polypeptide epidermal type I, protein-glutamine-gamma-glutamyltransferase)]
UGT	UDP glycosyltransferase	Gene with protein product, function known or inferred UGT1A1 (UDP glycosyltransferase 1 family, polypeptide A1) The preferred substrate of this enzyme is bilirubin, although it also has moderate activity with simple phenols, flavones, and C18 steroids

References

- Ainoon O, Joyce J, Boo NY, Cheong SK, Zainal, ZA, Hamidah NH (1999) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Chinese. *Hum Mut* 14:352–359
- Ainoon O, Yu YH, Amir Muhriz AL, Boo NY, Cheong SK, Hamidah NH (2001) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Malays. *Hum Mut* 21:101–109. DOI 10.1002/humu.9103
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299–309
- Bodmer WF, Cavalli-Sforza LL (1976) Genetics, evolution, and man. Freeman, San Francisco
- Cargill M, Alshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chang JG, Lee LS, Chen CM, Shih MC, Wu MC, Tsai FJ, Liang DC (2002) Molecular analysis of thiopurine S-methyltransferase alleles in south-east Asian populations. *Pharmacogenetics* 12:191–195
- Chowbay B, Cumaraswamy S, Cheung YB, Zhou QY, Edmund Lee JD (2003) Genetic polymorphisms in MDR1 and CYP3A4 genes in Asians and the influence of MDR1 haplotypes on cyclosporine disposition in heart transplant recipients. *Pharmacogenetics* 13:89–95
- Clark AG, Nielsen R, Signorovitch J, Matise TC, Glanowski S, Heil J, Winn-Deen ES, Holden AL, Lai E (2003) Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. *Am J Hum Genet* 73:285–300
- Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29:311–322
- Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, Xu CF, Dawson E, Rhodes S, Ueda S, Eric L, Luben RN, Van Rensburg EJ, Mannerman A, Kataja V, Rennart G, Dunham I, Purvis I, Easton D, Ponder BAJ (2000) The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am J Hum Genet* 67:1544–1554
- Etler DA (1996) The fossil evidence for human evolution in Asia. *Annu Rev Anthropol* 25:275–301
- Goddard KAB, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am J Hum Genet* 66:216–234
- Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project; identification of 190562 genetic variations in the human genome. *J Hum Genet* 47:605–610. DOI 10.1007/s100380200092
- Halushka MK, Fan JB, Bentley K, Linda H, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22:239–247
- Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y (2001) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 30:158–162
- Hustert E, Haberl M, Burk O, Wolbold R, He YQ, Klein K, Nuessler AC, Neuhaus P, Klattig J, Eiselt R, Koch I, Zibat A, Brockmoller J, Halpert JR, Zanger UM, Wojnowski L (2001) The genetic determinants of the CYP3A5 polymorphism. *Pharmacogenetics* 11:773–779
- Iida A, Sekine A, Saito S, Kitamura Y, Kitamoto A, Osawa S, Mishima C, Nakamura Y (2001a) Catalog of 320 single nucleotide polymorphisms (SNPs) in 20 quinone oxidoreductase and sulfotransferase genes. *J Hum Genet* 46:225–240
- Iida A, Saito S, Sekine A, Kitamoto A, Kitamura Y, Mishima C, Osawa S, Kondo K, Harigae S, Nakamura Y (2001b) Catalog of 434 single-nucleotide polymorphisms (SNPs) in genes of the alcohol dehydrogenase, glutathione S-transferase, and nicotinamide adenine dinucleotide, reduced (NADH) ubiquinone oxidoreductase families. *J Hum Genet* 46:385–407. DOI 10.1007/s100380170058
- Iida A, Saito S, Sekine A, Mishima C, Kondo K, Kitamura Y, Harigae S, Osawa S, Nakamura Y (2001c) Catalog of 258 single-nucleotide polymorphisms (SNPs) in genes encoding three organic anion transporters, three organic anion-transporting polypeptides, and three NADH: ubiquinone oxidoreductase flavoproteins. *J Hum Genet* 46:668–683. DOI 10.1007/s100380170019
- Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2002a) Catalog of 77 single-nucleotide polymorphisms (SNPs) in the carbohydrate sulfotransferase 1 (CHST1) and carbohydrate sulfotransferase 3 (CHST3) genes. *J Hum Genet* 47:14–19. DOI 10.1007/s10038-002-8650-9
- Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2002b) Catalog of 605 single-nucleotide polymorphisms (SNPs) among 13 genes encoding human ATP-binding cassette transporters: ABCA4, ABCA7, ABCA8, ABCD1, ABCD3, ABCD4, ABCE1, ABCF1, ABCG1, ABCG2, ABCG4, ABCG5, and ABCG8. *J Hum Genet* 47:285–310. DOI 10.1007/s100380200041
- Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2002c) Catalog of 86 single-nucleotide polymorphisms (SNPs) in three uridine diphosphate glycosyltransferase genes: UGT2A1, UGT2B15, and UGT8. *J Hum Genet* 47:505–510. DOI 10.1007/s100380200075
- Ingelman-Sundberg M, Oscarson M, McLellan RA (1999) Polymorphic human P450 enzymes: an opportunity for individualized drug treatment. *Trends Pharmacol Sci* 20:342–349
- Iwai K, Hirono A, Matsuoka H, Kawamoto F, Horie T, Lin K, Tantular IS, Dachlan YP, Notopuro H, Hidayah NI, Salim AM, Fujii H, Miwa S, Ishii A (2001) Distributions of glucose-6-phosphate dehydrogenase mutations in Southeast Asia. *Hum Genet* 108:445–449. DOI 10.1007/s004390100527
- Kalow W (2001) Perspectives in pharmacogenetics. *Arch Pathol Lab Med* 125:77–80
- Kamatani N, Sekine A, Kitamoto T, Iida A, Saito S, Kogame A, Inoue E, Kawamoto M, Harigai M, Nakamura Y (2004) Large-scale single-nucleotide polymorphism (SNP) and analyses, using dense SNP maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs

- within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. *Am J Hum Genet* 75:190–203
- Kleyn PW, Vesell ES (1998) Genetic variation as a guide to drug development. *Science* 281:1820–1821
- Kroetz DL, Pauli-Magnus C, Hodges LM, Huang CC, Kawamoto M, Johns SJ, Stryke D, Ferrin TE, DeYoung J, Taylor T, Carlson EJ, Herskowitz I, Giacomini KM, Clark AG on behalf of the Pharmacogenetics of Membrane Transporters Investigators (2003) Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. *Pharmacogenetics* 13:481–494
- Kruglyak L (1997) The use of a genetic map of biallelic markers in linkage studies. *Nat Genet* 17:21–24
- Lewin R (1987) Africa: cradle of modern humans. *Science* 237:1292–1295
- McCarthy JJ, Hilfiker R (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat Biotechnol* 18:505–508
- Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Philips MS, Jarvis N, Law S, de Arruda M, Todd JA (2000) Evaluation of single nucleotide polymorphism typing with Invader on PCR amplicons and its automation. *Genome Res* 10:330–343
- Meyer UA (1994) Pharmacogenetics: the slow, the rapid, and the ultrarapid. *Proc Natl Acad Sci USA* 91:1983–1984
- Meyer UA (2000) Pharmacogenetics and adverse drug reactions. *Lancet* 356:1667–1775
- Nagata J (1979) Malaysian mosaic: perspectives from a poly-ethnic society. University of Columbia Press, Taiwan
- Nebert DW (1997) Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? *Am J Hum Genet* 60:265–271
- Nebert DW (1999) Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clin Genet* 56:247–258
- Nei M (1985) Human Evolution at the molecular level. In: Ohta T, Aoki K (eds) *Population genetics and molecular evolution*. Japan Sci. Soc./Springer, Tokyo/Berlin, pp 41–64
- Nei M (1995) The origins of human populations: genetic, linguistic, and archaeological data. In: Brenner S, Hanihara K (eds) *The origin and past of modern humans as reviewed from DNA*. World Scientific, Singapore, pp 71–91
- Nei M, Li WH (1980) Non-random association between electrophoresis and inversion chromosomes in finite populations. *Genet Res* 35:65–83
- Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y (2001) A high-throughput SNP typing system for genome-wide association studies. *J Hum Genet* 46:471–477. DOI 10.1007/s100380170047
- Pirmohamed M, Park BK (2001) Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 22:298–305
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotyping data. *Genetics* 155:945–959
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* 411:199–204
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381–2385
- Ryan NJ (1971) *The cultural heritage of Malaya*. Longman, Malaysia
- Ryan D, Nuccie B, Arvan D (1999) Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous Invader microtiter plate assay. *Mol Diagn* 4:135–144
- Saito S, Iida A, Sekine A, Eguchi C, Miura Y, Nakamura Y (2001) Seventy genetic variations in human microsomal and soluble epoxide hydrolase genes (EPHX1 and EPHX2) in the Japanese population. *J Hum Genet* 46:325–329. DOI 10.1007/s100380170067
- Saito S, Iida A, Sekine A, Miura Y, Sakamoto T, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y (2001) Identification of 197 genetic variations in six human methyltransferase genes in the Japanese population. *J Hum Genet* 46:529–537. DOI 10.1007/s100380170035
- Saito S, Iida A, Sekine A, Miura Y, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y (2002) Three hundred twenty-six genetic variations in genes encoding nine members of ATP-binding cassette, subfamily B (ABCB/MDR/TAP), in the Japanese population. *J Hum Genet* 47:38–50. DOI 10.1007/s10038-002-8653-6
- Saito S, Iida A, Sekine A, Ogawa C, Kawauchi S, Higuchi S, Ohno M, Nakamura Y (2002) Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABCC/MRP/CFTR). *J Hum Genet* 47:147–171. DOI 10.1007/s100380200018
- Saito S, Iida A, Sekine A, Ogawa C, Kawauchi S, Higuchi S, Ohno M, Nakamura Y (2002) 906 variations among 27 genes encoding cytochrome P450 (CYP) enzymes and aldehyde dehydrogenases (ALDHs) in the Japanese population. *J Hum Genet* 47:419–444. DOI 10.1007/s100380200061
- Saito S, Iida A, Sekine A, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y (2002) Catalog of 238 variations among six human genes encoding solute carriers (hSLCs) in the Japanese population. *J Hum Genet* 47:576–584. DOI 10.1007/s100380200088
- Saito S, Iida A, Sekine A, Kawauchi S, Higuchi S, Ogawa C, Nakamura Y (2003) Catalog of 680 variations among eight cytochrome p450 (CYP) genes, nine esterase genes, and two other genes in the Japanese population. *J Hum Genet* 48:249–270
- Skeat WW (1902) The wild tribes of the Malay Peninsula. *J Anthropol Inst Great Br Ireland* 32:124–141
- Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. *Heredity* 76:377–383
- Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang RH, Messer CJ, Chew A, Han JH, Duan JC, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windermuth A, Xu CB, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489–493
- Tan JHH, Low PS, Tan YS, Tong MC, Saha N, Yang HY, Heng CK (2003) ABCA1 gene polymorphisms and their associations with coronary artery disease and plasma lipids in males from three ethnic populations in Singapore. *Hum Genet* 113:106–117. DOI 10.1007/s00439-003-0943-3
- Tang K, Ngoi SM, Gwee PC, John Chua MZ, Edmund Lee JD, Samuel Chong S, Cefroline Lee JL (2002) Distinct haplotype profiles and strong linkage disequilibrium at the MDR1 multidrug transporter gene locus in three ethnic Asian populations. *Pharmacogenetics* 12:437–450
- The International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933
- Watterson GA, Guess HA (1977) Is the most frequent allele the oldest? *Theor Popul Biol* 11:141–160
- Weir BS (1996) *Genetic data analysis II: methods for discrete population genetic data*. Sinauer Associates, Canada
- Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18:19–24
- Winstedt R (1961) *The Malays: a cultural history*. Lowe and Brydone, Great Britain
- Xie HG, Kim RB, Wood AJJ, Stein CM (2001) Molecular basis of the ethnic differences in drug disposition and response. *Annu Rev Pharmacol Toxicol* 41:815–850