Research article

Nucleotide sequence analyses of the MRPI gene in four populations suggest negative selection on its coding region

Zihua Wang1,2, Pui-Hoon Sew3, Helen Ambrose4, Stephen Ryan5, Samuel S Chong6,7, Edmund JD Lee8 and Caroline GL Lee*1,3

Address: 1Department of Biochemistry, National University of Singapore, Singapore, 2Graduate Programme in Bioengineering, National University of Singapore, Singapore, 3Division of Medical Sciences, National Cancer Center, Singapore, 4AstraZeneca, Alderley Park, Macclesfield, UK, 5AstraZeneca, Wilmington, DE, USA, 6Departments of Pediatrics & Obstetrics/Gynecology, Singapore, 7Departments of Pediatrics and Gynecology & Obstetrics, and McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA and 8Department of Pharmacology, National University of Singapore, Singapore

Email: Zihua Wang - g0202483@nus.edu.sg; Pui-Hoon Sew - bchsph@nus.edu.sg; Helen Ambrose - Helen.Ambrose@astraZeneca.com; Stephen Ryan - Stephen.Ryan@astraZeneca.com; Samuel S Chong - paecs@nus.edu.sg; Edmund JD Lee - phcelee@nus.edu.sg; Caroline GL Lee* - bchleec@nus.edu.sg

* Corresponding author

Abstract

**Background:** The MRPI gene encodes the 190 kDa multidrug resistance-associated protein 1 (MRPI/ABCC1) and effluxes diverse drugs and xenobiotics. Sequence variations within this gene might account for differences in drug response in different individuals. To facilitate association studies of this gene with diseases and/or drug response, exons and flanking introns of MRPI were screened for polymorphisms in 142 DNA samples from four different populations.

**Results:** Seventy-one polymorphisms, including 60 biallelic single nucleotide polymorphisms (SNPs), ten insertions/deletions (indel) and one short tandem repeat (STR) were identified. Thirty-four of these polymorphisms have not been previously reported. Interestingly, the STR polymorphism at the 5' untranslated region (5'UTR) occurs at high but different frequencies in the different populations. Frequencies of common polymorphisms in our populations were comparable to those of similar populations in HAPMAP or Perlegen. Nucleotide diversity indices indicated that the coding region of MRPI may have undergone negative selection or recent population expansion. SNPs E10/1299 G>T (R433S) and E16/2012 G>T (G671V) which occur at low frequency in only one or two of four populations examined were predicted to be functionally deleterious and hence are likely to be under negative selection.

**Conclusion:** Through in silico approaches, we identified two rare SNPs that are potentially negatively selected. These SNPs may be useful for studies associating this gene with rare events including adverse drug reactions.

**Background**

The development of drug resistance poses a serious limitation to the effective treatment of cancer. Although several different drug resistance mechanisms have been described, members of the ABC transporter superfamily have generated great interest because of their contribution...
### Table 1: Primers and PCR conditions for amplifying and sequencing MRP1

| Exons | Primer Direction | Length (bp) | Primer Sequence (5’ – 3’) | Amplicon Length(bp) | Additional Primer For Sequencing | Tm  |
|-------|------------------|-------------|----------------------------|---------------------|-----------------------------------|-----|
| Exon 1 | Forward          | 22          | GCATTTGAAAAGTGGTCGCAGG     | 688                 |                                   | 59  |
|        | Reverse          | 20          | TCCGCAGGAACTGAGTCCAC      |                     |                                   |     |
| Exon 2 | Forward          | 20          | GCAAGAAGCACCACATACCT      | 510                 |                                   | 60  |
|        | Reverse          | 20          | AGAAGAAGGAACTTACGAGC      |                     |                                   |     |
| Exon 3 | Forward          | 19          | GCATGTTGACCCAGCAAAAC      | 501                 |                                   | 60  |
|        | Reverse          | 19          | CTCAGCTCTGATTCAGCCTC      |                     |                                   |     |
| Exon 4 | Forward          | 21          | AGCTGCTGCATTAAGAAATAG     | 571                 |                                   | 61  |
|        | Reverse          | 20          | GACCTTCTCACACACAGCCAGAG   |                     |                                   |     |
| Exon 5 | Forward          | 19          | CCCAGCCGCCAGAATGTGAT      | 358                 |                                   | 58  |
|        | Reverse          | 18          | CCCAGCCGACATCTAAGC        |                     |                                   |     |
| Exon 6 | Forward          | 22          | TGTGTGATTTGTCGTCAGAT      | 383                 |                                   | 60  |
|        | Reverse          | 20          | GAGCTGAGAGATTTCTTACC      |                     |                                   |     |
| Exon 7 | Forward          | 20          | TCCCCATAGCTTTTGTATGC      | 574                 |                                   | 61  |
|        | Reverse          | 19          | TACCCCATTTTTCGACAGAA      |                     |                                   | 59  |
| Exon 8 | Forward          | 20          | AGAAGACGCTTTAAAGCCTTAGT   | 591                 |                                   | 59  |
|        | Reverse          | 19          | TATGACCTCACCTTCAGGACAG    |                     |                                   |     |
| Exon 9 | Forward          | 19          | CGTGTTCCTCTATGCATTC       | 602                 |                                   | 59  |
|        | Reverse          | 18          | CCTGCCACCTAAGGCAC         |                     |                                   |     |
| Exon 10 | Forward       | 18          | TCTCTGGGCAAGCAAGATAG      | 439                 |                                   | 61  |
|        | Reverse         | 18          | TGAAACCACGGCCGAGAAT       |                     |                                   |     |
| Exon 11 | Forward       | 19          | GCTGGGGAAGAAAGGAGGT       | 452                 |                                   | 61  |
|        | Reverse         | 19          | TGAGTGAACACTGGGAGGCA      |                     |                                   |     |
| Exon 12 | Forward       | 20          | TAAATAGACGCGTGAAATAGGAG   | 743                 | ATGAAGAGCAAGAAGACATCG (forward)  | 61  |
|        | Reverse         | 20          | AAGTAATCTCTCTTGTCCCTC     |                     |                                   |     |
| Exon 13 | Forward       | 21          | GTCTGTTGGTTTATTCAGTTC     | 523                 |                                   | 61  |
|        | Reverse         | 20          | CTCTGCTTCTACAGGAGCAG      |                     |                                   |     |
| Exon 14 | Forward       | 20          | TGGAAATTACCTTTTTGTG       | 627                 |                                   | 60  |
|        | Reverse         | 19          | GTCGAAAACCTTGGGAAAGG      |                     |                                   |     |
| Exon 15 | Forward       | 21          | TTAAAGGAGGAATAAGCTAGT     | 475                 |                                   | 60  |
|        | Reverse         | 19          | TGATCTGCAACCACATGTG       |                     |                                   |     |
| Exon 16 | Forward       | 20          | TTGGACTAGATCTTCTAGGT      | 463                 |                                   | 60  |
|        | Reverse         | 19          | CCAAAATCTGCTCCTTCTAG      |                     |                                   |     |
| Exon 17 | Forward       | 21          | GTGGCGACCTGTGTGCTCTCAG    | 441                 |                                   | 61  |
|        | Reverse         | 20          | AGTGAGACCTGAGGCCACACC     |                     |                                   |     |
| Exon 18 | Forward       | 18          | CGATTTGAGTCTGCTTCAA       | 596                 | TACCCATTACCAACAATCG (reverse)    | 61  |
|        | Reverse         | 18          | TTTTCCGACCCCTGTACC        |                     |                                   |     |
| Exon 19 | Forward       | 18          | GAGTTTTGCCGCCACAGGT       | 462                 |                                   | 61  |
|        | Reverse         | 19          | GTGTGTTTCCCGACATGTG       |                     |                                   |     |
| Exon 20 | Forward       | 18          | TATGCCCTTTTCCCTCAGT       | 758                 |                                   | 61  |
|        | Reverse         | 18          | GAGGAAAGACCCACACCA        |                     |                                   |     |
| Exon 21 | Forward       | 19          | CGTGGACGGTACATGTTG        | 504                 |                                   | 60  |
|        | Reverse         | 20          | GGGTGACCTGGATGAAGTCA      |                     |                                   |     |
| Exon 22 | Forward       | 20          | GTTGAATAGCTAAAGGGGAG      | 599                 |                                   | 61  |
|        | Reverse         | 19          | GTGAAATGGGCAACACCTGG      |                     |                                   |     |
| Exon 23 | Forward       | 20          | ATGCTGGTTTGCTACATATT     | 514                 |                                   | 58  |
|        | Reverse         | 20          | TTTTAGAATACACGTGTATA      |                     |                                   |     |
| Exon 24 | Forward       | 20          | AGCCTGAGCTATTATGTTCA      | 727                 |                                   | 61  |
|        | Reverse         | 19          | ACTCAATAGCAGATCGGTG       |                     |                                   |     |
| Exon 25 | Forward       | 19          | CCCATTTGAGCATTTTTGA       | 1442                | AGATGCAGCTAAAGGACGTTCC*         | 61  |
|        | Reverse         | 20          | AAAGACGTGGAACGCTTAA       |                     | GTAGCTGATGGGTTTGCG**           |     |
| Exon 26 | Forward       | 19          | CCCATTTGAGCATTTTTGA       | 1442                | AGATGCAGCTAAAGGACGTTCC*         | 61  |
|        | Reverse         | 20          | AAAGACGTGGAACGCTTAA       |                     | GTAGCTGATGGGTTTGCG**           |     |
| Exon 27 | Forward       | 22          | AACTTCAATGTCCTGAGGCCC     | 693                 |                                   | 61  |
|        | Reverse         | 20          | TACTGGCCACAGTCCACCTC      |                     |                                   |     |
| Exon 28 | Forward       | 18          | CGATCTTCTCTTTTTGAGG       | 493                 |                                   | 61  |
|        | Reverse         | 20          | GTAGGCGCAAAATGAATTTT      |                     |                                   |     |
| Exon 29 | Forward       | 18          | AGAAGACGCTTCTGATACCC      | 405                 |                                   | 61  |
|        | Reverse         | 19          | GGCAAAATGCTCCTAGG         |                     |                                   |     |
| Exon 30 | Forward       | 19          | CTCCCCAAAGTGCCTAGATT      | 706                 |                                   | 61  |
|        | Reverse         | 19          | GAAATGCTTGAACCCCGA        |                     |                                   |     |
| Exon 31 | Forward       | 21          | TCTGAATGTAATGGAACAGTG     | 1677                |                                   | 61  |
|        | Reverse         | 21          | GGAAGACACATGTAACCTAA      |                     |                                   |     |

*reverse primer for Exon 25, **forward primer for Exon 26
Table 2: Frequencies of single nucleotide polymorphisms (SNPs) identified at MRPI

| No.* | SNP ID** | Region | Protein residue | dbSNP (rs) | Population | n | Allele frequency (%) | No. | SNP ID | Region | Protein residue | dbSNP (rs) | Population | n | Allele frequency (%) |
|------|----------|--------|----------------|-----------|------------|---|----------------------|------|--------|--------|----------------|-----------|------------|---|----------------------|
|      |          |        |                |           |            |   |                      |      |        |        |                |           |            |   |                      |
Table 2: Frequencies of single nucleotide polymorphisms (SNPs) identified at MRPI (Continued)

| No.  | SNP ID | Region | Protein residue | dbSNP (rs) | Population | n | Allele frequency (%) | No.  | SNP ID | Region | Protein residue | dbSNP (rs) | Population | n | Allele frequency (%) |
|------|--------|--------|-----------------|------------|------------|---|----------------------|------|--------|--------|------------------|------------|------------|---|------------------|
| ML   | 35     |        |                 |            | 98.57      | 1.43 |                      | IN   | 35     |        |                 |            | 11.43      | 88.57 |                  |
| CAU  | 36     |        |                 |            | 84.29      | 15.71 |                      | CAU  | 36     |        |                 |            | 13.89      | 86.11 |                  |

SNPi7 13/-66 T>C Intron 3 - rs4148337

| ML   | 35     | 44.29  | 55.71 | SNPi17 111/-48 C>T Intron 11 - rs3765129 |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 48.57  | 51.43 |                                           |
| CAU  | 35     | 74.29  | 25.71 |                                           |

SNPi8 17/31 G>T Intron 7 - rs8187850

| ML   | 34     | 100.00 | 0.00  | SNPi19 112/-143 C>T Intron 12 - -        |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 98.57  | 1.43  |                                           |
| CAU  | 36     | 98.61  | 1.39  |                                           |

SNPi9 17/54 C>A Intron 7 - rs903880

| ML   | 34     | 2.94   | 97.06 | SNPi12 112/-85 G>A Intron 12 - rs4148348  |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 20.00  | 80.00 |                                           |
| CAU  | 36     | 13.89  | 86.11 |                                           |

SNPi10 17/64 C>G Intron 7 - rs246232

| ML   | 34     | 72.06  | 27.94 | SNPi11 112/-37 A>G Intron 12 - rs35604    |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 65.71  | 34.29 |                                           |
| CAU  | 36     | 80.00  | 20.00 |                                           |

SNPi22 113/33 113/33 Intron 13 - -

| ML   | 35     | 44.29  | 55.71 | SNPi1  E21/2793 C>T Exon 21 Thr>Thr       |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 98.57  | 1.43  |                                           |
| CAU  | 36     | 98.61  | 1.39  |                                           |

| ML   | 35     | 27.14  | 72.86 |                                           |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 11.43  | 88.57 |                                           |
| CAU  | 36     | 13.89  | 86.11 |                                           |

SNPe2 E6/651 G>A Exon 6 Ser>Ser -

| ML   | 35     | 0.00   | 100.00|                                           |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 2.86   | 97.14 |                                           |
| CAU  | 36     | 0.00   | 100.00|                                           |

SNPe8 E13/1704 C>T Exon 13 Tyr>Tyr rs8187858

| ML   | 35     | 100.00 | 0.00  | SNPe9 E13/1684 C>T Exon 13 Leu>Leu        |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 11.43  | 88.57 |                                           |
| CAU  | 36     | 13.89  | 86.11 |                                           |

SNPe21 E13/1704 C>T Exon 13 Tyr>Tyr rs8187858

| ML   | 35     | 100.00 | 0.00  | SNPe9 E13/1684 C>T Exon 13 Leu>Leu        |
|------|--------|--------|-------|-------------------------------------------|
| IN   | 35     | 11.43  | 88.57 |                                           |
| CAU  | 36     | 13.89  | 86.11 |                                           |

No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%) No. SNP ID Region Protein residue dbSNP (rs) Population n Allele frequency (%)
| SNP | E14/1898 G>A | Exon 14 | Arg>Gln | rs11075296 | CH 36 0.00 100.00 | ML 35 0.00 100.00 | IN 35 1.43 98.57 | CAU 36 0.00 100.00 |
|-----|--------------|---------|---------|-------------|------------------|------------------|------------------|------------------|
| SNP | E16/2001 C>T | Exon 16 | Ser>Ser | rs8187863    | CH 36 100.00 0.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 98.61 1.39 |
| SNP | E16/2007 C>T | Exon 16 | Pro>Pro | rs2301666    | CH 36 100.00 0.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | E16/2012 G>T | Exon 16 | Gly>Val | -            | CH 36 100.00 0.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | E17/2168 G>A | Exon 17 | Arg>Gln | rs4148356    | CH 36 0.00 100.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | E17/2217 C>T | Exon 17 | Ser>Ser | -            | CH 36 100.00 0.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | E17/2268 G>A | Exon 17 | Gly>Gly | -            | CH 36 100.00 0.00 | ML 35 100.00 0.00 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | I17/-168 A>G | Intron 17 | - | rs3887893    | CH 36 100.00 0.00 | ML 35 52.86 47.14 | IN 35 47.14 52.86 | CAU 36 43.06 56.94 |
| SNP | I24/-43 C>T | Intron 22 | - | rs3887893    | CH 36 91.43 8.57  | ML 35 91.43 8.57  | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | I25/103 C>T | Intron 24 | - | rs3887893    | CH 36 93.06 6.94  | ML 35 93.06 6.94  | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | I25/114 C>T | Intron 25 | - | rs3887893    | CH 36 1.39 98.61 | ML 35 1.39 98.61 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
| SNP | I25/146 G>A | Intron 26 | - | rs3887893    | CH 36 1.39 98.61 | ML 35 1.39 98.61 | IN 35 100.00 0.00 | CAU 36 100.00 0.00 |
Table 2: Frequencies of single nucleotide polymorphisms (SNPs) identified at MRPI (Continued)

| SNP   | Mutation | Intron | rsID     | C     | G     | CH   | ML   | IN   | CAU  | C     | T     |
|-------|----------|--------|----------|-------|-------|------|------|------|------|-------|-------|
| SNPi24| 118/-30 G>C | Intron 18 | rs2074087 | 15.28 | 84.72 |      |      |      |      |       |       |
|       |          |        | ML       | 12.86 | 87.14 |      |      |      |      |       |       |
|       |          |        | IN       | 18.06 | 81.94 |      |      |      |      |       |       |
| SNPi25| 118/-27 G>A | Intron 18 | rs28363997 | 0.00  | 100.00 |      |      |      |      |       |       |
|       |          |        | ML       | 0.00  | 100.00 |      |      |      |      |       |       |
|       |          |        | IN       | 0.00  | 100.00 |      |      |      |      |       |       |
| SNPi26| 119/-175 C>T | Intron 19 | rs4148369 | 91.67 | 8.33  |      |      |      |      |       |       |
|       |          |        | ML       | 91.43 | 8.57  |      |      |      |      |       |       |
|       |          |        | IN       | 81.43 | 18.57 |      |      |      |      |       |       |
| SNPe1  | E20/2651 C>T | Exon 20 | Thr>Met |       |       |      |      |      |      |       |       |
|       |          |        | ML       | 98.57 | 1.43  |      |      |      |      |       |       |
|       |          |        | IN       | 97.14 | 2.86  |      |      |      |      |       |       |
| SNPi27| 120/-59 C>T | Intron 20 |       | 100.00 | 0.00  |      |      |      |      |       |       |
|       |          |        | ML       | 97.14 | 2.86  |      |      |      |      |       |       |
|       |          |        | IN       | 97.14 | 2.86  |      |      |      |      |       |       |
|       |          |        | CAU      | 97.14 | 2.86  |      |      |      |      |       |       |

* The numbers of SNPs are defined as SNP(SNP locates the region)(In the certain region, the consequent number of the SNP). For example, SNPi1 means the SNP is located in the untranslated region (UTR) and the first SNP in the UTR.
** SNP ID for these SNPs are defined as (Region)/(position number)(major allele)(minor allele). Region shows whether SNP is in an exon (E) or intron (I). Position number of exonic SNP is the mRNA nucleotide position using the translation start site as position 1. For intronic SNP, the position in the genomic sequence is either upstream (-) of using the 5' boundary of the immediate downstream exon as position -1, or downstream (+) using 3' boundary of the immediate upstream exon as position 1. The same principle is employed to the indel and STR. The position is specific to the first nucleotide in the genomic sequence. (reference mRNA ID:NM_004996.2).
*** CH, ML, IN and CAU represent Chinese, Malay, Indian and Caucasian, respectively.
to multidrug resistance of tumor[1,2]. The 170 kDa P-glycoprotein, encoded by the MDR1 gene, was the first member of this family to be described [3]. Subsequently, the 190-kDa multidrug resistance-associated protein-1 (MRP1/ABCC1) was isolated from a multidrug resistance lung cancer cell line that does not express MDR1 [4]. Both these transporters have been implicated in the resistance of various cancers to chemotherapy. Although MRP1 is only 18% identical to MDR1 at the amino acid level, it transports several similar drugs as MDR1 including doxorubicin, vincristine and colchicine. However, while drugs transported by MDR1 are usually neutral or cationic, drugs effluxed by MRP1 are anionic, frequently conjugated with glutathione and other anions, or are co-transported with glutathione [2]. MRP1 has also been implicated to play important roles in cellular anti-oxidative defense and inflammation [5,6]. MRP1 is located on chromosome 16 at band 13.1 and spans approximately 200 kb. It contains 31 exons and encodes 1531 amino acids. The MRP1 protein is predicted to comprise three membrane spanning domains (MSDs) and two nucleotide binding domains (NBDs) [4-7].

Genetic polymorphisms in MDR1 have been associated with differences in MDR1 expression and function as well as drug response and disease susceptibilities [8-10]. SNPs within MDR1 that have been associated with functional

| No.  | indel ID         | Region      | Amino acid change | dbSNP (rs) | Population | n | frequency (%) |
|------|------------------|-------------|-------------------|------------|------------|---|---------------|
| Indel1 | 5'UTR/ -74 14bp indel | 5'UTR       | -                 | -         | CH 36   | 100.00 0.00  
|       |                  |             |                   |            | ML 35   | 100.00 0.00  
|       |                  |             |                   |            | IN 35   | 97.14 2.86   
|       |                  |             |                   |            | CAU 36  | 100.00 0.00  
| Indel2 | 12/17 G indel   | Intron 2    | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 0.00 100.00  
|       |                  |             |                   |            | IN 35   | 0.00 100.00  
|       |                  |             |                   |            | CAU 36  | 1.39 98.61   
| Indel3 | 13/ -70 T indel | Intron 3    | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 0.00 100.00  
|       |                  |             |                   |            | IN 35   | 0.00 100.00  
|       |                  |             |                   |            | CAU 36  | 1.39 98.61   
| Indel4 | 15/-25 12 bp indel | Intron 5    | -                 | rs3830390 | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 0.00 100.00  
|       |                  |             |                   |            | IN 35   | 1.43 98.57   
|       |                  |             |                   |            | CAU 36  | 0.00 100.00  
| Indel5 | 112/ -119 G indel | Intron 12   | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 1.43 98.57   
|       |                  |             |                   |            | IN 35   | 0.00 100.00  
|       |                  |             |                   |            | CAU 36  | 0.00 100.00  
| Indel6 | 116/7 6bp indel | Intron 16   | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 1.43 98.57   
|       |                  |             |                   |            | IN 35   | 0.00 100.00  
|       |                  |             |                   |            | CAU 36  | 0.00 100.00  
| Indel7 | 117/ -182 5bp indel | Intron 17  | -                 | -         | CH 30   | 81.70 18.30  
|       |                  |             |                   |            | ML 35   | 74.29 25.71  
|       |                  |             |                   |            | IN 29   | 86.20 13.80  
|       |                  |             |                   |            | CAU 30  | 80.00 20.00  
| Indel8 | 118/ -38 AT indel | Intron 18   | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 2.86 97.14   
|       |                  |             |                   |            | IN 35   | 0.00 100.00  
|       |                  |             |                   |            | CAU 36  | 1.39 98.61   
| Indel9 | 124/49 C indel  | Intron 24   | -                 | -         | CH 36   | 0.00 100.00  
|       |                  |             |                   |            | ML 35   | 0.00 100.00  
|       |                  |             |                   |            | IN 35   | 1.43 98.57   
|       |                  |             |                   |            | CAU 36  | 0.00 100.00  
| Indel10 | 129/-13 T indel | Intron 29   | -                 | rs4148379 | CH 36   | 45.83 54.17  
|        |                   |             |                   |            | ML 35   | 41.43 58.57  
|        |                   |             |                   |            | IN 35   | 61.43 38.57  
|        |                   |             |                   |            | CAU 36  | 75.00 25.00  


differences were found to demonstrate evidence of recent positive selection [11]. However, less is known about the polymorphisms within MRP1. Although numerous SNPs have been identified within this gene, most of these studies were performed on a single population which was primarily either Chinese, Japanese or Caucasians in origin. Thus far, no association have been observed between the few SNPs at MRP1 and functional differences [12,17-19] possibly because neither the functionally important SNP nor SNPs in LD with the functional SNP were examined. These studies which examined only a few of the many SNPs within MRP1, without knowledge of the functional SNP nor the LD or haplotype profile in that population may not have been powerful enough to identify any association. Recently, we found evidence of genomic signatures of recent positive selection in a SNP at the 5' flanking region (5'FR) of MRP1 in a Caucasian population and demonstrated that this SNP altered MRP1 promoter activity [20].

In the present study, we sequenced all the exons as well as the 5' and 3' flanking regions of MRP1 to comprehensively scan for polymorphisms in 142 DNA samples from four different populations, namely, the Chinese, Malays, Indians and Caucasians. Nucleotide diversity of the exonic polymorphisms was determined and the functional effects of the non-synonymous SNPs were predicted using three programs, SIFT, PolyPhen and PANTHER. We found that SNPs E10/1299G>T, which resulted in arginine-serine substitution at amino acid position 433 (R433S) and E16/2012G>T, which resulted in glycine-valine substitution at amino acid position 671 (G671V), may potentially adversely affect the function of MRP1. While these two SNPs, which have low minor allele frequencies (<3%), may not be useful for studies associating this gene with common diseases/drug response, it may, nonetheless, be useful for studies associating this gene with rare events including adverse drug reactions (ADRs).

Results and discussion
Profile of polymorphisms within MRP1 in the different populations
De novo sequencing of approximately 18 kb of genomic DNA at MRP1, including all the 31 exons as well as flanking regions, was performed in 142 healthy individuals from four different populations to identify polymorphisms at MRP1 in the different populations. A total of 71 polymorphisms were identified including 60 bi-allelic SNPs, ten indels and one short tandem repeat (STR) (Figure 1, Tables 2, 3, 4). An examination of currently reported SNPs in the dbSNP Build 125 database [21] and published reports [12-17,22] revealed that 26 SNPs and 8 indels were not previously reported and hence represent novel polymorphisms.

Nineteen of the 60 SNPs identified were found in all four populations while 28 of these SNPs were population-specific with 22 of these population-specific SNPs occurring only once out of 284 chromosomes examined (singletons). While the STR and two indels were polymorphic in all the populations examined, the other seven indels were population-specific of which six were singletons.

None of the indels or STR identified occurred within exons (Fig. 1, Tables 3 and 4). Eighteen of the 60 bi-allelic SNPs were found in exonic regions, six of which resulted in non-synonymous change (Fig. 1, Table 2). These results suggest that polymorphisms at MRP1 are largely conservative since less than 10% of these polymorphisms (6/71) presented as non-synonymous changes which are potentially capable of disrupting the MRP1 protein structure/function. Nonetheless, it is possible for synonymous or intronic SNPs to affect MRP1 expression or function through the alteration of the mRNA transcript stability or

| Trinucleotide repeats | Region | Amino acid change | dbSNP (rs) | GCC repeat No. | frequency (%)* |
|-----------------------|--------|-------------------|------------|---------------|----------------|
| 5'UTR/-118 GCC repeats | 5'UTR | -                 | -          | 7             | 18.57          |
|                       |        |                   |            | 8             | 1.43           |
|                       |        |                   |            | 9             | 5.56           |
|                       |        |                   |            | 10            | 1.39           |
|                       |        |                   |            | 11            | 25.00          |
|                       |        |                   |            | 12            | 9.72           |
|                       |        |                   |            | 13            | 50.00          |
|                       |        |                   |            | 14            | 4.17           |
|                       |        |                   |            | 15            | 2.78           |
|                       |        |                   |            | 16            | 1.39           |

* Repeat frequencies in bold represent GCC repeats that have relative frequencies (≥10%).
folding [23] thereby affecting downstream splicing[24,25], processing[26], translational control [27] or regulation [28]. Additionally, polymorphisms at the 5'UTR/promoter and 3'UTR may influence promoter activity and hence gene expression or mRNA transcript stability.

Interestingly, although no polymorphisms were identified at the 3'UTR region (exon 31), four polymorphisms, including the STR (Table 4) and one indel (Table 3) were found to reside at the 5'UTR/reported core promoter region [29] of MRP1. Three of these promoter polymorphisms were novel but population-specific with SNPs 5'UTR/-46C>T and 5'UTR/-51C>T occurring only in the Malay population and the polymorphism 5'UTR/-74 14 bp indel occurring only in the Indian population. Insertion/deletion polymorphisms in promoter regions have been correlated with the modulation of the expression of genes (e.g. matrix metalloproteinase I gene [30]). It is thus possible that the 14 bp indel polymorphism in the Indian population may influence the promoter activity and hence the expression of MRP1.

The STR polymorphism found at the 5'UTR/promoter region of MRP1 is a GCC trinucleotide repeat and 7–16 of such repeats were observed in the four populations (Table 4). The most commonly occurring GCC repeat number in all the four populations was 13 which occurred at a frequency of 50.00%, 51.43%, 47.14% and 56.94% in the Chinese, Malay, Indian and Caucasian populations, respectively. Eleven GCC repeats also occurred at high frequencies (≥ 20%) in most of the populations except the
### Table 5: Comparisons of polymorphisms identified in this study with those reported in HapMap/Perlegen

#### A

| HAPMAP | Perlegen |
|--------|----------|
| CHB | CEPH | CHB | EA |
| # common polymorphisms | 19 | 19 | 14 | 14 |
| This Study # non-common polymorphisms found in HapMap/Perlegen but not in this study | 25 (22) | 23 (22) | 1 (0) | 1 (0) |
| polymorphisms found in this study but not in HapMap/Perlegen | 41 (26) | 41 (26) | 46 (28) | 46 (26) |

(\()\text{ refers to number of monomorphic SNPs in that category in that population}

#### B

| SNP ID | dbSNP ID | Popu | N | Our data | N | HapMap | n | Perlegen |
|--------|----------|------|---|----------|---|--------|---|----------|
| 12/26 G>A | rs8187843 | CH | 36 | 1.00 | 100.00 | 45 | 0.00 | 100.00 |
| | | CAU | 35 | 7.14 | 92.86 | 60 | 5.83 | 94.17 |
| 17/31 G>T | rs8187850 | CH | 35 | 100.00 | 0.00 | 45 | 100.00 | 0.00 |
| | | CAU | 36 | 98.61 | 1.39 | 60 | 100.00 | 0.00 |
| 17/54 C>A | rs903880 | CH | 34 | 1.47 | 98.53 | 45 | 4.44 | 95.56 |
| | | CAU | 36 | 13.89 | 86.11 | 60 | 25.00 | 75.00 |
| 17/64 C>G | rs246232 | CH | 34 | 61.76 | 38.24 | 44 | 55.68 | 44.32 |
| | | CAU | 35 | 80.00 | 20.00 | 60 | 66.67 | 33.33 |
| 8/825 T>C | rs246221 | CH | 35 | 31.44 | 68.56 | 45 | 23.33 | 76.67 |
| | | CAU | 36 | 98.57 | 1.43 | 60 | 95.71 | 4.29 |
| 9/1062 T>C | rs35587 | CH | 34 | 38.51 | 61.49 | 45 | 46.67 | 53.33 |
| | | CAU | 36 | 31.94 | 68.06 | 60 | 28.33 | 71.67 |
| 111/-78 G>A | rs35595 | CH | 36 | 20.83 | 79.17 | 45 | 31.11 | 68.89 |
| | | CAU | 36 | 13.89 | 86.11 | 60 | 13.33 | 86.67 |
| 111/-48 C>T | rs3765129 | CH | 36 | 88.89 | 11.11 | 45 | 88.89 | 11.11 |
| | | CAU | 36 | 90.28 | 9.72 | 60 | 85.34 | 14.66 |
| 112/56 C>T | rs17241551 | CH | 35 | 98.57 | 1.43 | 45 | 95.56 | 4.44 |
| | | CAU | 36 | 97.22 | 2.78 | 60 | 90.83 | 9.17 |
| 112/-85 G>A | rs4148348 | CH | 36 | 8.33 | 91.67 | 45 | 4.44 | 95.56 |
| | | CAU | 36 | 8.33 | 91.67 | 60 | 7.50 | 92.50 |
| 112/-37 A>G | rs35604 | CH | 36 | 82.33 | 16.67 | 45 | 72.22 | 27.78 |
| | | CAU | 36 | 79.17 | 20.83 | 60 | 82.50 | 17.50 |
| 13/1684 C>T | rs35605 | CH | 36 | 83.33 | 16.67 | 45 | 72.22 | 27.78 |
| | | CAU | 36 | 80.56 | 19.44 | 60 | 82.50 | 17.50 |
| 13/1704 C>T | rs8187858 | CH | 36 | 100.00 | 0.00 | 45 | 100.00 | 0.00 |
Caucasians. Interestingly, while seven GCC repeats occurred at relatively high frequencies in the Indian and Caucasian populations (≥12%), this number of repeats was not observed in either the Chinese or Malay population. These observations highlight the differences in the distribution of the number of the MRP1 promoter GCC repeats in the different populations with the Indians and Caucasians being more similar to each other than to the Chinese and Malays. The number of STR repeats residing within or close to promoters has been found to modulate the promoter activity of genes [31-34]. Interestingly, differences in the CGG and GCC trinucleotide repeats at the 5'UTR/promoter region of the Fragile X mental retardation genes (FMR1 and FMR2, respectively) have been associated with differences in the methylation status of the promoter and expression of the genes [35]. Hence this common polymorphism at the 5'UTR/promoter region of MRP1 with distinctly different distribution of repeat numbers in the different population may have potential functional significance.

Comparison of polymorphisms identified in this study with those reported in the HapMap and Perlegen databases

Two publicly available databases HapMap [36] and Perlegen [37] examined genome-wide polymorphisms (including polymorphisms at the MRP1 gene) in several populations. HapMap genotyped already known SNPs from public databases at a density of approximately one SNP per 5 kb of DNA in 4 different populations namely, 45 Japanese from Tokyo, 45 Chinese from Beijing (CHB), 60 US residents with northern and western European ancestry by the Centre d'Étude du Polymorphisme Humain (CEPH) and 60 Yoruba people of Ibadan (YRI). Approximately 1.6 million SNPs from 24 Han Chinese (CH), 24 European American (EA) and 23 African American (AA) were successfully genotyped in the Perlegen project. The SNPs genotyped in the Perlegen project were either reported in public databases or identified through their array-based re-sequencing of 24 human samples of diverse ancestry [38,39]. We thus compared the polymorphisms at the MRP1 gene that we identified through de
The extent of variation at nucleotide diversity at fewer samples being examined in the Perlegen database, especially for the Chinese population probably due to the database (Table 5C). Interestingly, significant difference from our study and those from the HapMap or Perlegen between allele frequencies in the respective populations. Paired T-test revealed no significant difference (P > 0.05) similar frequencies in similar populations (Table 5B). The HapMap and Perlegen projects were found to have polymorphisms in our study that was also genotyped in the two databases (Table 5A). On the other hand, 41/41 and 46/46 polymorphisms that we identified were not examined in either the HapMap or the Perlegen project, respectively. While many of these polymorphisms were of low frequencies or were monomorphic in the two populations that were similar to the HapMap/Perlegen polymorphisms, 8 of these polymorphisms were found to be of relatively high frequencies (>5%) in at least one of the two populations. Some of the low frequency polymorphisms represent novel SNPs identified in this study.

Polymorphisms in our study that was also genotyped in the HapMap and Perlegen projects were found to have similar frequencies in similar populations (Table 5B). Paired T-test revealed no significant difference (P > 0.05) between allele frequencies in the respective populations from our study and those from the HapMap or Perlegen database (Table 5C). Interestingly, significant difference (P < 0.05) was observed between data obtained from the HapMap database and those from the Perlegen database especially for the Chinese population probably due to fewer samples being examined in the Perlegen database.

**Nucleotide diversity at MRPI**

The extent of variation at MRPI was evaluated using two conventional measures of nucleotide diversity: \( \pi \), the average heterozygosity per site and \( \theta \), the population mutation parameter[40]. Tajima’s D statistic was also calculated to assess deviation from the neutral mutation model[41]. A positive Tajima’s D value for a single gene is indicative of positive heterozygote advantage while a negative Tajima’s D value for an individual gene suggests selection of a specific allele over the alternative allele(s)[42]. However, when a negative Tajima’s D value is observed in most of the genes that were examined in a particular population, it is suggestive of a recent expansion in that population[42].

With all the exonic regions sequenced, the above nucleotide diversity statistics were determined for non-synonymous versus synonymous SNPs at MRPI (Table 6). The \( \theta \) value for synonymous SNPs at MRPI was found to be 16.15 \( \times 10^4 \) (Table 6) which was comparable to mean \( \theta \) values of other reported genes including 24 transporter genes (20.14 ± 4.10) \( \times 10^4 \)[43], 75 candidate genes associated with blood pressure homeostasis (15.1 ± 3.6) \( \times 10^4 \)[44] but slightly higher than the mean \( \theta \) values of 106 random genes (10.03 ± 2.52) \( \times 10^4 \)[45]. However, the \( \theta \) value for non-synonymous SNPs (11.73 ± 4.14) at MRPI was much higher than mean \( \theta \) values of the other reports (3.39 ± 0.90 to 5.7 ± 1.4) \( \times 10^4 \)[43-45] probably due to the small size of the MRPI exons in which the non-synonymous SNPs reside. Interestingly, while the \( \pi \) of synonymous SNPs (\( \pi_s \)) at MRPI (12.62) was comparable to the reported mean \( \pi_s \) values in other genes (9.73 ± 4.86 to 10.67 ± 5.07) \( \times 10^4 \)[43,45], the \( \pi_m \) at MRPI (0.94) was much lower than the mean reported \( \pi_m \) values for the other genes (2.20 ± 1.12 to 2.75 ± 1.31) \( \times 10^4 \)[43,45]. This low \( \pi_m \) at MRPI was also reported previously [43] with the reported \( \pi_m \) value (0.15) being much lower than the present observation (0.94). Notably, the \( \pi_m/\pi_s \) at MRPI was less than 1 (0.0743 in this study and 0.0110 in the previous study [43]), suggesting that this gene is likely to be under selective pressure. Importantly, the \( \theta \) values for both synonymous and non-synonymous SNPs were greater than the corresponding \( \pi \) values, resulting in negative Tajima’s D statistic which suggests that the coding region of MRPI may have undergone negative selection or population expansion. It is more likely that MRPI gene have undergone negative selection since the average total nucleotide diversity in the MRPI gene (\( \pi_{\text{total}} \) (0.25)) was found to be greater than the amino acid diversity (\( \pi_m \)) (0.94) [43].

**SNPs E10/1299G>T and E16/2012 G>T are potentially deleterious**

As nucleotide diversity statistics suggest that the coding region of MRPI may be under negative selection, we thus further analyzed the exonic SNPs at MRPI to evaluate if any of these SNPs may have deleterious effects on MRPI structure/function.

Exonic SNPs, particularly non-synonymous SNPs, have the potential to alter the secondary/tertiary structure of proteins and/or affect the protein function. A total of 18 exonic SNPs were identified at this gene locus of which five have not been previously reported (Fig. 2A, C). Most of the exonic SNPs occurred at low frequencies (<5%) in only one or two populations. While at least 30% of the synonymous SNPs at MRPI occurred at greater than 5% frequency in all the four populations examined, all of the non-synonymous SNPs occurred at less than 3% in only one or at most two populations (Fig. 2C). This observation highlights the conservation of exonic polymorphisms at MRPI and suggests that altering the non-synonymous SNPs may have a deleterious effect and are
Figure 2
Profiles of the 18 exonic SNPs in MRPI. (A) The distribution of all the exonic SNPs on the mRNA of MRPI is presented. (B) The topological model of MRPI protein secondary structure is predicted using the SOSUI program and the positions of the SNPs on the topological image are displayed using the TOPO2 program. Approximate locations of predicted individual transmembrane helices, membrane spanning domains and nucleotide binding domains are indicated as TM 1-17, MSD 1-3 and NBD1-2, respectively. The six nonsynonymous SNPs are highlighted in black while the 12 synonymous SNPs are colored green. Consensus sequences for Walker A and B are highlighted in red. The six nonsynonymous SNPs predicted by PANTHER to be potentially deleterious are highlighted by large black dots. (C) Table showing in silico prediction of functional significance of exonic SNPs and their frequencies in the different populations. SNPs that have previously been utilized for association/functional studies are also presented. For PANTHER prediction, if the subPSEC score is lower than -3.5, it can be interpreted that the amino acid change could have high probability of deleterious functional effect. Note: Underlined SNPs represent SNPs that have not been previously reported. * The six nonsynonymous SNPs are highlighted with single asterisks. ** This SNP is predicted to have effect on the function of the protein.
likely to be selected against, resulting in their low frequencies.

To assess if any of the non-synonymous SNPs at MRP1 have potentially damaging effect on the protein structure/function, the location of these six SNPs were displayed on the MRP1 protein topological image using the SOSUI and TOPO2 programs. As evident in figure 2B, none of the non-synonymous SNPs reside in the transmembrane regions although four of these SNPs reside near or within the nucleotide binding domain (NBD) of the MRP1 protein. Nonetheless, SNPe1 (SNP e2/218 G>T) and SNPe6 (SNP e10/1299 G>T) reside near the transmembrane region, while SNPe12 (SNP E16/2012 G>T) reside on a conserved glycine residue near the conserved Walker A consensus motif of the NBD [12], suggesting that these SNPs may have functional significance. SNP e2/218 G>T was only found at less than 3% in the Chinese and Malay populations while SNP E10/1299 G>T occurred at less than 2% in the Caucasian population only and SNP E16/2012 G>T occurred at less than 3% in the Indian and Caucasian populations (Fig. 2C). The SNP frequencies of SNP E10/1299 G>T and SNP E16/2012 G>T in the Caucasian population were comparable to a previous report [12].

Three different algorithms, SIFT [46], Polymorphism Phenotyping (PolyPhen) [47] and PANTHER [48] were then utilized to predict the functional significance of the six non-synonymous SNPs. SIFT predicts the effect of amino acid substitutions based on the assumption that the important amino acid will be conserved in the protein family [46]. PolyPhen predicts the effect of the amino acid variant on the function or structure of the protein based on current knowledge of protein structure, interactions and evolution [47] while the PANTHER program predicts the effect of an amino acid substitution on the protein's function using amino acid substitution scores derived from an alignment of related protein sequences and statistics from hidden Markov models [48].

Interestingly, SNP E10/1299 G>T, which is located near the transmembrane domain, was predicted to be potentially deleterious by the PANTHER but not the SIFT or PolyPhen algorithms. This SNP was reported to affect the ability of MRP1 to confer drug resistance as well as to transport organic anions [49] suggesting that the PANTHER program may be more accurate in predicting the functional impact of polymorphisms than SIFT or PolyPhen. This observation is similar to a previous report that utilizes both bioinformatics and biochemical approaches to compare the accuracy of the PolyPhen and PANTHER programs in predicting functionally deleterious polymorphisms in the ABCA1 gene [50]. They found that the PANTHER software is significantly (P < 0.05) more accurate in its prediction of the functional consequence of nonsynonymous SNPs. They also reported that the PANTHER program is capable of correctly predicting the functional impact of greater than 94% of the polymorphisms examined while PolyPhen is only ~88% accurate in predicting the functional impact of polymorphisms [50].

Significantly, all of the three different algorithms predicted that SNP E16/2012 G>T, which resides close to Walker A and results in G671V substitution, was likely to have a potentially deleterious effect on protein function (Fig. 2C). The significance of this polymorphism has also been demonstrated previously by Conrad et al. [12] who reported that the mRNA expression of peripheral lymphocytes from individuals carrying the SNP E16/2012 G>T polymorphism was lower than the average expression level. The lower expression of the MRP1 G671V transcript is suggestive of greater accumulation of MRP1 drugs in the cells which may lead to adverse drug reactions. Curiously, that report also found that the G671V polymorphism did not affect the transport of MRP1 substrates including leukotriene C4, 17β-estradiol, 17β-(D)-glucuronide and estrone sulfate by membrane vesicles prepared from transiently transfected HEK293T cells [12].

Recently, the same group also reported similar MRP1 protein expression levels and transport properties in human embryonic kidney cells were transfected with MRP1 constructs carrying either glycine or valine at amino acid position 671 [51]. The observation that the G671V polymorphism did not affect MRP1 protein expression or transport ability of some MRP1 substrates in vitro [12,51] does not rule out the possibility of functional significance of this polymorphism in vivo especially since the same group reported decreased transcript expression in individuals carrying this polymorphism. It is still possible that this polymorphism affect the transport of other MRP1 substrates that has not been examined. It is also possible

Table 6: Nucleotide diversity at MRP1

| sequence section | length(bp) | SNPs | θ* | π* | πNS/πS | Tajima's D |
|------------------|-----------|------|----|----|--------|-----------|
| Synonymous       | 1194      | 12   | 16.15 | 12.62 | 0.0743 | -0.0209   |
| Nonsynonymous    | 822       | 6    | 11.73 | 0.94  | -0.0735|

*Values of θ and π are listed as value × 10^-4
that although the SNP E16/2012 G>T polymorphism does not affect MRP1 transport ability, it may affect other yet-to-be-examined functional properties of the protein (e.g. drug resistance capability or cellular anti-oxidative defense or inflammation). It has been reported that an artificial mutation E1089Q created in MRP1 markedly affected the ability of MRP1 protein to confer resistance without affecting its ability to transport organic anions [52]. Hence, the SNP E16/2012 G>T polymorphism warrants further investigation.

Hence, the bioinformatics approach may be useful in facilitating the prediction of potentially functionally significant polymorphism so that future research may be directed to characterizing these polymorphisms.

**Functional implications of polymorphisms at MRP1**

The current detailed characterization of polymorphisms at MRP1 in four different ethnic populations highlights several characteristics about this gene that may facilitate more rational approaches to studies associating this gene with functional changes. We have previously reported that the diverse haplotypes and weak LD across MRP1 [20] could perhaps provide an explanation for the failure of previous studies to detect association between polymorphisms in this gene and functional differences [12,17-19] and highlight the importance of fully characterizing the LD and haplotype profiles of the gene before embarking on association studies. Its LD and haplotype architecture suggest that it may be necessary to identify alternative approaches for association studies of this gene as it may not be feasible to utilize tag SNPs. A possible approach is to identify polymorphisms with potential functional significance before performing association studies, possibly by identifying those polymorphisms that may have been subjected to selection pressures.

We recently identified a high frequency SNP at the 5' flanking promoter region of MRP1 that demonstrated evidence of recent positive selection and affected the promoter activity of MRP1 [20]. In this report, through the sequencing of the MRP1 exonic and flanking regions, we identified a GCC-trinucleotide multi-allelic STR polymorphism residing within the 5'UTR/promoter region of MRP1 that was found at relatively high frequencies in all populations examined. Notably, the frequency distribution of the different number of STR alleles in the different population was found to be different (Table 4). Although it was previously reported that the 5'UTR/promoter region contains the GCC-triplet repeat that is absent in the rodent sequence and 7, 13 and/or 14 of these repeats were observed in different cell lines and PBMC from a single individual [7,53], no reports have yet examined the variation of this polymorphism in the different ethnic populations. This STR is approximately 296 bp from the SNP that we previously reported to show evidence of recent positive selection [20]. Given that the selection is recent it may be expected that it would be in strong LD with the positively selected SNP. Since STRs within/near promoters have been implicated to affect promoter activity and expression levels of the gene, it would be worthwhile to further examine the effect that this polymorphism together with the positively selected SNP have in influencing promoter activity and hence expression of MRP1.

Interestingly, while the promoter region of MRP1 may be under recent positive selective pressure [20], in this study we also found that the coding region of this gene may have undergone negative selection pressure as suggested by nucleotide diversity indices. Two coding SNPs E16/2012 G>T and E10/1299 G>T have been predicted by either the PANTHER program or all three programs (SIFT, PolyPhen and PANTHER) to have a deleterious effect on the structure/function of the protein. The significance of these SNPs for general association studies may be limited since these SNP occurs at very low frequencies (<3%) in only one or two of the four populations examined. Nonetheless, this SNP may be associated with rare events including ADR.

**Conclusion**

In summary, based on the "common disease-common variant" hypothesis, the previously reported common polymorphism within the promoter of MRP1 that showed evidence of recent positive selection [20] would be useful for association studies of common diseases/drug response. Nonetheless, the rare exonic SNP(s) in this gene that we demonstrate here to be likely to be under negative selection pressure may be useful for studies associating this gene with rare phenotypes including ADR, which has been listed as the top five leading causes of death in Western countries [54].

**Methods**

**Study population**

The populations examined include individuals residing in Singapore from the following ethnic groups: 36 Caucasians and Chinese as well as 35 Malays and Indians. Race and ethnic group were declared by the volunteers to be true to three generations. Informed consent from the volunteers and ethical approval from the National University Hospital and the Changi General Hospital Institutional Review Boards were obtained.

**PCR and DNA sequencing**

The MRP1 genomic DNA (NT_0101393.13) sequence was obtained from GenBank [55] and used as the reference sequence. For the sequencing of 31 exons of MRP1, 30 pairs of primers (see Table 1) were designed using Vector NTI 7.0 software and utilized to amplify these exons. The...
The PCR reaction was performed in a 10 μl reaction volume containing 40 ng of genomic DNA template from the above-mentioned samples, 5 μl 2 × PCR master mix buffer (Qiagen, Valencia, CA, USA), with or without 1 μl Q-solution (depending on the GC content of the amplicon) as well as 0.20 μM of sense and anti-sense primers. PCR was carried out in a GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA) with the following thermal cycling conditions: an initial denaturation at 94°C for 15 min followed by 35 cycles at 94°C for 30 sec, temperature for the optimal annealing of each amplicon as specified (Table 1) for 90 sec, and extension at 72°C for 60 sec. This was then followed by a final elongation step at 72°C for 10 min. The PCR products obtained were then treated with exonuclease I and shrimp alkaline phosphatase (SAP, United States Biochemical). Sequencing reactions were performed using ABI PRISM Big Dye Terminator (V3.0) kit and the conditions for the sequencing reactions were (for all the exons except exon 1): 94°C for 15 min followed by 30 cycles at 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. Due to the high GC content of exon 1, the sequencing conditions were modified as follows: 94°C for 15 min followed by 35 cycles at 98°C for 30 sec, 48°C for 10 sec and 60°C for 5 min. The final product was resolved by automated capillary electrophoresis on an ABI PRISM 3700™ DNA analyzer (Applied Biosystems). The DNA sequence of each exon obtained experimentally was then aligned against the reference sequence (NT_0101393.13) using the Vector NTI 7.0 software to identify the polymorphic sites. Polymorphisms identified were verified through bi-directional re-sequencing of all samples whose chromatograms do not clearly display the polymorphism as well as randomly selected samples whose chromatograms clearly show the polymorphism.

**Population genetic parameters**

Two common parameters of nucleotide diversity were calculated: the neutral parameter (θ) which is the estimate of population mutation parameter based on the number of polymorphic sites in the samples [40] and nucleotide diversity (π) which is the direct estimate of heterozygosity per site, or the average proportion of nucleotides that differ between any randomly sampled pair of sequences [40]. Each of these two parameters was calculated for synonymous and nonsynonymous SNP sites. Tajima’s D statistic was also calculated to assess deviations from the neutral mutation model [41].

**In Silico characterization of polymorphisms in exons**

The programs Sorting Intolerant From Tolerant (SIFT) [46], Polymorphism Phenotyping (PolyPhen) [47] and PANTHER [48] were utilized to evaluate the potential effect of amino acid substitutions resulting from the polymorphisms. Since the position of the non-synonymous polymorphic amino acid residue on the MRP1 protein may provide important clues with regards to its potential functionality, the SOSUI program [56] was utilized to predict the topology of the MRP1 protein and the TOPO2 program [57] was used to display the location of the SNPs on the MRP1 protein topological image.

**Authors’ contributions**

ZW contributed to the design of the experiments, data analysis and the write-up of the manuscript. PHS carried out the sequencing experiments. HA and SR contributed to the critical review of the draft manuscript. SSC and EJDL contributed to the conception and critical review of the draft manuscript. CGL (corresponding author) conceived the study, and contributed to the design of experiments, coordination, critical evaluation of the data and analyses as well as the final writing of the manuscript. All the authors have given the final approval of the version to be published.

**Acknowledgements**

This study was sponsored by AstraZeneca.

**References**

1. Cole SP, Deeleay RG: Multidrug resistance mediated by the ATP-binding cassette transporter protein MRP. *Blood* 1998, 92:931-940.

2. Bosorst P, Evers R, Kool M, Wijnholds J: The multidrug resistance protein family. *Biochim Biophys Acta* 1999, 1461:347-357.

3. Chen CJ, Chin JE, Udala K, Clark DP, Pistan I, Gottesman MM, Roninson IB: Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. *Cell* 1986, 47:381-389.

4. Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, Stewart AJ, Kurz EU, Duncan AM, Deeleay RG: Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* 1992, 258:1650-1654.

5. Leslie EP, Deeleay RG, Cole SP: Toxicological relevance of the multidrug resistance protein gene MRP1 (MRP1) and related transporters. *Toxicology* 2001, 167:3-23.

6. Hipfner DR, Deeleay RG, Cole SP: Structural, mechanistic and clinical aspects of MRP1. *Biochim Biophys Acta* 1999, 1461:359-376.

7. Grant CE, Kurz EU, Cole SP, Deeleay RG: Analysis of the intron-exon organization of the human multidrug-resistance protein gene (MRP) and alternative splicing of its mRNA. *Genomics* 1997, 45:368-378.

8. Fellay J, Marzolini C, Meaden ER, Back DJ, Bucin T, Chave JP, Decoster A, Furrer H, Opravil M, Pantealeo G, Retelska D, Ruiz L, Schinkel AH, Vernazza P, Eap CB, Telenti A: Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 2002, 359:30-36.

9. Lee CGL, Chong SS, Lee EJ: Pharmacogenetics of the Human MDR1 multidrug transporter. *Curr Pharmaceut Genomics* 2004, 2:1-11.

10. Hoffmeyer S, Burki O, van Richter O, Arnold HP, Brockmoller J, Johne A, Cescorbi I, Gerloff T, Roops I, Eichelbaum M, Brinkmann U: Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc Natl Acad Sci USA* 2000, 97:3473-3478.

11. Tang K, Wong LP, Lee EJ, Chong SS, Lee CGL: Genomic evidence for recent positive selection at the human MDR1 gene locus. *Hum Mol Genet* 2004.
12. Conrad S, Kauffmann HM, Ito K, Deeley RG, Cole SP, Schrenk D: Identification of human multidrug resistance protein 1 (MRP1) mutations and characterization of a G617V substitution. J Hum Genet 2001, 46:656-663.

13. Ito S, Ieiri I, Tanabe M, Suzuki A, Higuchi S, Otsubo K: Polymorphism of the ABC transporter genes, MDRI, MRPI and MRPII/cMOAT, in healthy Japanese subjects. Pharmacogenetics 2001, 11:175-184.

14. Saito S, Iida A, Sekine A, Miura Y, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y: Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABC/C/MPR/CFTR). J Hum Genet 2002, 47:147-171.

15. Perdue J, Germain DP: Identification of novel polymorphisms in the p53 and MRPI (ABCC1) genes at locus 16p13.1 and exclusion of both genes as responsible for pseudoxanthoma elasticum. Hum Mutat 2001, 17:74-75.

16. Wang H, Hao B, Zhou K, Chen X, Wu S, Zhou G, Zhu Y, He F: Linkage analysis of two ABC transporter genes (ABCC1 and ABCG2) in Chinese populations: implications for pharmacogenomic association studies. Ann Hum Genet 2004, 68:563-573.

17. Oselin K, Mrakiszewicz PSH, Gaivkivitch E, Pakhlk R, Roots I: Frequency of MRPI genetic polymorphisms and their functional significance in Caucasians: detection of a novel mutation G816A in the human MRPI gene. Eur J Clin Pharmacol 2003, 59:347-350.

18. Mathijssen RH, Marsh S, Karlsson MO, Xie Z, Baker SD, Verweij J, Sparreboom A, McLeod HLI: Irinotecan pathway: genotype analysis to predict pharmacokinetics. Clin Cancer Res 2003, 9:3246-3253.

19. Moriya Y, Nakamura T, Horinouchi M, Sakaeda T, Tamura T, Aoyama N, Shirakawa T, Gotot A, Fujimoto S, Matsuo M, Kasuga M, Oikumura K: Effects of polymorphisms of MDRI, MRPI, and MRPII genes on their mRNA expression levels in duodenal enterocytes of healthy Japanese subjects. Biol Pharm Bull 2002, 25:1356-1359.

20. Wang Z, Wang B, Tang K, Lee EJ, Chong SS, Lee CG: A functional polymorphism within the MRPI gene locus identified through its genomic signature of positive selection. Hum Mutat 2005, 14:2075-2087.

21. dbSNP database homepage. [http://www.ncbi.nlm.nih.gov/SNP/]

22. Le Sidero Y, Urban Z, Tedde R, Cassar K, Becchieli B, Quaglini S, Pasqualli-Ronchetti I, Pope FM, Richards A, Terry S, Bercoff L, de Paepe A, Boyd CD: Mutations in a gene encoding an ABC transporter cause pseudoxanthoma elasticum. Nat Genet 2000, 25:223-227.

23. Sheth LX, Basilion JP, Stanton VP Jr: Single-nucleotide polymorphisms can cause different structural folds of mRNA. Proc Natl Acad Sci USA 1999, 96:7871-7876.

24. Coleman TP, Roesser JR: RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. Biochemistry 1998, 37:15941-15950.

25. Ardile K, Liu-Cordero SN, Eberle MA, Barrett J, Winchester E, Lander ES, Kruglyak L: Genomic signature of positive selection: implications for pharmacogenomic association studies. Biomed Central Commu 2003, 3:901-907.

26. Borrmann L, Seebeck B, Rogalla P, Bullerdiek J: Human HMGAA2 promoter is coregulated by a polymorphic dinucleotide (C/T)-repeat. Oncogene 2001, 20:756-760.

27. Rothenburg S, Koch-Nolte F, Rich A, Haag F: A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity. Proc Natl Acad Sci USA 2001, 98:8985-8990.

28. Wang B, Ren J, Ooi LL, Chong SS, Lee CG: Dinucleotide repeats negatively modulate the promoter activity of Cyr61 and is unstable in hepatocellular carcinoma patients. Oncogene 2005, 24:3999-4008.

29. Cummings CJ, Zhangbi HY: Fourteen and counting: unraveling the nucleotide repeat story. Hum Mutat 2006, 22:909-916.

30. International HapMap Project [http://www.hapmap.org/]

31. Hartl DL: A primer of population genetics. 3rd edition. Sunderland, Mass.: Sinauer Associates; 2005.

32. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989, 123:585-595.

33. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar A, Zhang G, Neidell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF: Haplotype variation and linkage disequilibrium in 313 human genes. Science 2001, 293:489-493.

34. Leaemb MK, Huang CC, DeYoung J, Carlson EJ, Taylor TR, de la Cruz M, Johns J, Stynke D, Kawamoto M, Urban TJ, Kozicz F, Ferrar TE, Clark AG, Rich N, Hershkovitz I, Giacomini KM: Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. Proc Natl Acad Sci USA 2003, 100:5896-5901.

35. Halfa MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravariti A: Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 1999, 22:239-247.

36. Cargill M, Altshuler D, Ireland J, Sklar P, Ardile K, Patil N, Shaw N, Lange CR, Lim EP, Kalyanaraman N, Nenm J, Ziaura L, Frieldland R, Rolf A, Warrington J, Lipshutz R, Daley QQ: Lander ES: Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 1999, 22:231-238.

37. Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003, 31:3812-3814.

38. Suyayae S, Ramensky V, Kich I, Lathe W 3rd, Konard SR, Aker B, P: Prediction of deleterious human alleles. Hum Mol Genet 2001, 10:591-597.

39. Thomas PD, Keijariwal A: Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci USA 2004, 101:15398-15403.

40. Conrad S, Kauffmann HM, Ito K, Leslie EM, Deeley RG, Schrenk D, Cole SP: A naturally occurring mutation in MRPI results in a selective decrease in organic anion transport and an increased doxorubicin resistance. Pharmacogenetics 2002, 12:321-330.

41. Brunham LR, Singaraja RR, Paue TD, Keijariwal A, Thomas PD, Hayden MR: Accurate Prediction of the Functional Significance of Single Nucleotide Polymorphisms and Mutations in the ABCA1 Gene. PloS Genet 2005, 1:e83.

42. Letourneau IJ, Deeley RG, Cole SP: Functional characterization of non-synonymous single nucleotide polymorphisms in the...
gene encoding human multidrug resistance protein 1 (MRP1/ABCC1). Pharmacogenet Genomics 2005, 15:647-657.
52. Zhang DW, Cole SP, Deeley RG: Identification of an amino acid residue in multidrug resistance protein 1 critical for conferring resistance to anthracyclines. J Biol Chem 2001, 276:13231-13239.
53. Muredda M, Nunoya K, Burtch-Wright RA, Kurz EU, Cole SP, Deeley RG: Cloning and Characterization of the Murine and Rat mrrp1 Promoter Regions. Mol Pharmacol 2003, 64:1259-1269.
54. Lazarou J, Pomeranz BH, Corey PN: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998, 279:1200-1205.
55. GenBank database [http://www.ncbi.nlm.nih.gov/Genbank/]
56. SOSUI server [http://sosui.proteome.bio.tuat.ac.jp/sosui_submit.html]
57. TOPO2 server [http://www.sacs.ucsf.edu/TOPO-run/wtopo.pl]