Iranian Journal of Basic Medical Sciences

ijbms.mums.ac.ir

IJBMS

UGT1A1 gene linkage analysis: application of polymorphic markers rs4148326/rs4124874 in the Iranian population

Zakiye Nadeali [1], Sadeq Vallian [1*]

[1] Division of Genetics, Department of Biology, Faculty of Sciences, University of Isfahan, Isfahan, Iran

ARTICLE INFO

Article type:
Original article

Article history:
Received: Apr 10, 2016
Accepted: Apr 13, 2017

Keywords:
Genotyping
Linkage analysis
Molecular diagnostic
Polymorphic markers
UGT1A1 enzyme

ABSTRACT

Objective(s): Mutations in the UGT1A1 gene are responsible for hyperbilirubinemia syndromes including Crigler-Najjar type 1 and 2 and Gilbert syndrome. In view of the genetic heterogeneity and involvement of large numbers of the disease causing mutations, the application of polymorphic markers in the UGTA1 gene could be useful in molecular diagnosis of the disease.

Materials and Methods: In the present study, two polymorphic markers including rs4148326 and rs4124874 in the UGT1A1 gene region were characterized. The markers were selected using bioinformatics analysis of the UGT1A1 gene region and genotyped in 212 unrelated healthy individuals and 13 family trios in the Iranian population using Tetra-Primer ARMS PCR technique. The allele frequency and population status of the alleles were estimated using GENEPOP, FBAT, PowerMarker and Arlequin software.

Results: The results indicated that in the case of rs4148326 marker, allele frequency for T and C allele was 66.04% and 33.96%, respectively. For rs4124874 marker, allele frequency for G and T alleles was 39.4% and 60.6%, respectively. The values of heterozygosity index for the markers examined were 64.1 for rs4148326 and 72.1 for rs4124874, respectively. The haplotype estimation analysis of the markers resulted in three informative haplotypes with frequencies ≥0.05. Moreover, the results suggested the presence of linkage disequilibrium between two markers.

Conclusion: Altogether, the data suggested that rs4148326 and rs4124874 could be introduced as informative markers for molecular diagnosis of Crigler-Najjar type 1 and 2 and Gilbert syndrome in the Iranian population.

► *Please cite this article as:*
Nadeali Z, Vallian S. UGT1A1 gene linkage analysis: application of polymorphic markers rs4148326/rs4124874 in the Iranian population. Iran J Basic Med Sci 2017; 20:880-885. doi: 10.22038/IJBMS.2017.9109

Introduction

Uridine diphosphate glucuronosyl transferase 1A1 (UGT1A1) catalyzes the conjugation of insoluble bilirubin with glucuronic acid in hepatic cells, which is necessary for efficient biliary removal of serum bilirubin (1, 2). Deficiency of the enzyme activity results in the accumulation of unconjugated bilirubin in serum, causing hyperbilirubinemia (3, 4). Mutations in the *UGT1A1* gene are responsible for hyperbilirubinemia syndromes including Gilbert syndrome (GS) and Crigler–Najjar (CN) type 1 and 2 (CN-1, CN-2) (5, 6). CN-1 is a severe and lethal form of hereditary unconju-gated hyperbilirubinemia, with autosomal recessive inheritance manner. CN-2 and GS are the moderate and mild forms of hereditary unconjugated hyper-bilirubinemia, respectively (7-9). CN type 1 and 2 and GS result from the combination of several mutations of variable severity at a single locus.

Therefore, the phenotype variability of these syndromes is, at least in part, due to genetic heterogeneity. Genetic counseling as well as molecular diagnosis, including antenatal diagnosis is proposed in families at risk for CN and GS (8).

To date, over 130 different mutations in the *UGT1A1* gene have been recorded in the human genome mutation database (HGMD) (10). Molecular diagnosis of the syndrome(s) has been essentially based on direct mutation analysis of the *UGT1A1* gene region (11). However, due to the large number of mutations associated with the disease as well as the genetic heterogeneity of the mutations, direct analysis is mostly time-consuming and costly procedure, especially in developing countries. Alternatively, indirect analysis of mutations using linkage analysis could be considered as a practical approach in families with an affected individual.

Several polymorphic markers in the *UGT1A1* gene region have been introduced. However, the application of the markers is population-dependent.

*Corresponding author: Sadeq Vallian. Division of Genetics, Department of Biology, Faculty of Science, University of Isfahan, Isfahan, Iran. Tel/Fax: +98-31-37932456; email: svallian@sci.ui.ac.ir

Table 1. Primer sets for genotyping of rs4148326 and rs4124874

Primer	Sequence (5'→3')	
	rs4148326	rs4124874
IF	CAGGGTGTTCTTGCTACAAACCAAAAcAc	GCTGGCCAAGGGTAGAGTTCAaTG
OR	ATCCTCCCCACCACCATGCTTCA	TGTCCAAGCTCATTCCTCCTCTC
OF	AAGTAAGCCATTTACCAACGCTCAG	TCTTTGCTTTGATAAATTGTGGGGC
IR	AGGATGCTGGTCACCCTAGaTG	CACCATGTGGGTCATCTGTGACTTtAa

IF: Inner Forward; OR: Outer reverse; OF: Outer forward; IR: Inner reverse

Therefore, the suitability of each marker should be first determined in the population under study, before further application in linkage and carrier-ship investigations (12).

To our knowledge, there is no study performed on the molecular markers linked to the *UGT1A1* gene region by means of linkage study for molecular diagnosis of these syndromes. In the present study, using bioinformatics analysis of the *UGT1A1* gene region, among the markers present in the gene region, two single nucleotide polymorphisms (SNP), rs4148326 and rs4124874, located at 5′ region of the gene (respectively in intron 1 and the promoter region) were selected (13-15). The genetic structure as well as haplotype frequency of the markers was investigated in the Iranian population.

Materials and Methods

Bioinformatics studies and SNP selection

An *in silico* analysis was performed on the polymorphic markers in the *UGT1A1* gene region, which were reported in databases dbSNP, ALFRED, UCSC Genome Browser, SNPper and openSNP. The markers with high minor allele frequency (MAF) and heterozygosity including rs4124874 and rs4148326 were selected. MAF and heterozygosity rate are two important characteristics of genetic markers for their informativeness in linkage analysis (16).

DNA sample

Peripheral blood samples were collected from 225 individuals consists of 212 unrelated healthy persons and 13 family trios. The trios family was consisted of both parents and one child with healthy status, which were selected randomly. Unrelated healthy individuals and both parents in family group were sex- and age- matched (aged between 25 and 39). The children in the family group were 5-11 years old. All subjects were from the central province of Iranian population with Fars ethnicity.

Genomic DNA was extracted from peripheral blood mononuclear cells by salting out procedure (17).

SNPs genotyping

The selected rs4124874 and rs4148326 markers were genotyped using tetra-primer ARMS-PCR technique (18). This method combines two allele-specific inner primers and two outer primers in a single reaction and encompasses deliberate mismatches at position-2 from the 3' end of inner primers. Genotyping by tetra-primer ARMS-PCR requires only a single PCR amplification followed by electrophoresis for the determination of genotypes (18, 19).

The optimal selection of primers can be achieved in an automated way using a program, which evaluates candidate primers for a given sequence. In this study, the Primer1 software was used for tetra-primer ARMS-PCR primer design (20, 21). Primer sets are shown in Table 1.

The PCR reactions were carried out in reactions of 25 µl total volume containing 100 ng of template DNA, 15 pmol of each inner primer, 3 pmol of each outer primer, 75 mM MgCl2, 150 µM dNTP, 2.5 µl 10× buffer and 2.5 U Taq DNA polymerase. Initial denaturation was performed at 94 °C for 5 min, followed by 30 cycles including 30 sec denaturation at 94 °C; different annealing temperature depending on the marker used (64 °C for rs4148326 and 66 °C for rs4124874) for 50 sec, extension at 72 °C for 50 sec; followed by a 10 min final extension at 72 °C. The PCR products were separated on 2% agarose gel electrophoresis and analyzed.

Statistical analysis

Allele frequency as well as the observed and expected heterozygosity was determined using the GENEPOP software (22). The haplotype frequency was estimated using Arlequin (23) and FBAT software (24) on the data obtained from unrelated individuals and family trios, respectively. The program Arlequin implements expectation-maximization (EM) algorithm (23).

In order to provide a better description of the obtained results from the haplotype estimation, the estimation of linkage disequilibrium (LD) for 212 unrelated individuals was also performed using the PowerMarker computer program (25). LD was measured by using the standardized D'(26). D' is the LD relative to its maximum value for a given set of allelic frequencies for the pair of sites (27). In this sense, D' is a normalized value of LD. D' ranges between 0 and 1. When D'=1, the LD was completed and indicated LD. When D' was estimated zero, it indicated the presence of LD between two loci. Given that D' alone could not provide great prediction in the likely performance for multiallelic markers; therefore, Chi-square ($\chi 2$), which is another multiallelic LD statistics, was used (28). Chi-square provides a direct assessment of the strength

Table 2. The frequency distribution and the expected and observed heterozygosity frequency of rs4148326 and Rs4124874 markers in the Iranian population

Marker	Minnor allele frequency	Major allele frequency	Ho	He
rs4148326	C allele 33.96	T allele 66.04	64.15	44.96
rs4124874	T allele 39.4	G allele 60.6	72.1	47.86

Ho: Observed heterozygosity, He: Expected heterozygosity

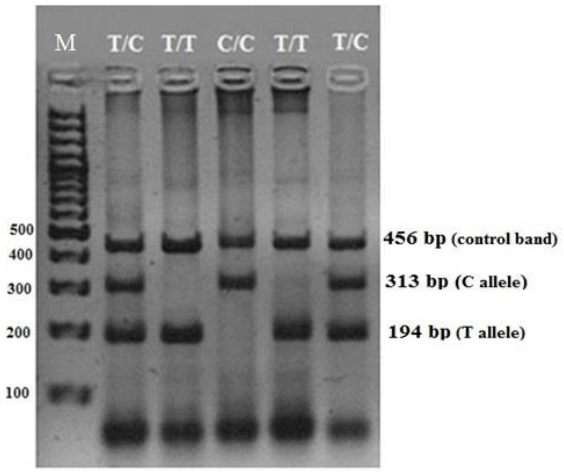

Figure 1. Genotyping of *rs4148326* marker in the Iranian population PCR products of five samples were shown. The heterozygous (TC), homozygous (TT) and (CC) genotype were indicated above each lane. Each product size is also shown at the right side of picture

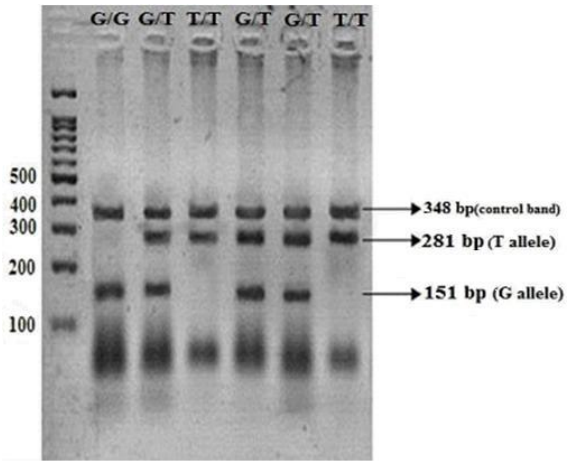

Figure 2. Genotyping of *rs4124874* marker in Iranian population PCR products of six samples were shown. The heterozygous (GT), homozygous (GG) and (TT) genotype were indicated above each lane. Each product size is also shown at the right side of picture

of the association between alleles located at the two loci. The estimation of D' and $\chi 2$ was performed using PowerMarker computer program. The null hypothesis of random association between pairs of alleles at the two loci (D'= 0) was assessed by $\chi 2$ value. The $\chi 2$ value for two pairwise marker is compared with $\chi 2$ obtained from the chi-square chart ($P \leq 0.05$).

Haplotypes construction for the markers was carried out through the PowerMarker computer program, which has been based on EM algorithm and calculates the most likely haplotypes with respect to the provided genotypic data (29).

HapMap analysis

The genotyping data were obtained from the HapMap database (30). Then the genotyping data from the Iranian population (present study) were compared with those of the other populations.

Results

In the present study, two markers in the *UGT1A1* gene region including rs4148326 and rs4124874 were genotyped in the Iranian population (Figure 1, 2). The allelic frequency and the expected and observed hetero-zygosity of the markers were estimated by GENEPOP software based on the identified unrelated hetero-zygous individuals. As presented in Table 2, the MAF for rs4148326 (C allele) was 33.96, and for rs4124874 (T allele) was 39.4. Both markers showed a relatively high degree of allelic heterozygosity (64.1 for rs4148326 and 72.1 for rs4124874, respectively) in the studied population (Table 2).

The haplotype frequency was estimated using Arlequin and FBAT software. As shown in Table 3, among the four possible haplotypes identified by

Table 3. Analysis of haplotypes frequency of *rs4148326 and*rs4124874 markers in the Iranian population using Arlequin and FBAT softwares and informative state of each haplotype

Index	Haplotype	Frequency		Informativeness
	rs4148326-rs4124874	Unrelated	Family trios	(+/-)
1	T- T	0.36024	0.311	+/+
2	C- G	0.30599	0.292	+/+
3	T- G	0.30014	0.285	+/+
4	C- T	0.03363	0.112	-/+

The symbols + and - in the table show the informative and non informative state of each haplotype in unrelated individuals and family trios,respectively

Table 4. Analysis of D' and χ2 values for pairing of the rs4148326 and rs4124874 markers in the *UGT1A1* locus in the Iranian population

Pairing of markers	D′	*P* value	χ^2
rs4124874-*rs4148326*	0.75	0.000	79.87

Arlequin software for the markers, three haplotypes were considered as informative haplotypes with frequencies ≥5% in the Iranian population.

Haplotype estimation of families was performed by FBAT software, which resulted in four informative haplotypes with frequencies ≥5% among four possible haplotypes. Three haplotypes including T-T, C-G and T-G were estimated as informative haplotypes using both computer programs (Table 3). The data suggested that a combination of rs4124874 and rs4148326 could be introduced as an informative haplotype.

Given the presence of relatively high proportion of informative haplotypes in the population under the study, the LD pattern for the pairing of these two markers was assessed by calculating average D' and χ2 using PowerMarker program (25). As shown in Table 3, the results of estimation for D' values for two possible pairing of markers were higher than zero. The calculated χ2 values were also higher than χ2 values obtained from the chi-square table (P<0.05) (Table 4) (31).

Discussion

CN syndrome type 1 and 2 and GS result from mutations in *UGT1A1*gene (32). One of the clinical manifestations of these diseases is neonatal jaundice, which is also very common in newborn infants. Even though it is often a natural and transitional condition, some infants develop severe hyperbilirubinemia in which unconjugated bilirubin in the serum may cross the blood-brain-barrier causing encephalopathy or kernicterus (33). To date, numerous polymorphisms of *UGT1A1* gene have been identified in association with hyperbilirubinemia diseases of CN-1, CN-2, and GS. Thus, the *UGT1A1* gene markers may be used as biomarkers pertaining to examine individualized disease risk. Due to the greater power and more appropriate state of using haplotypes for genotype-phenotype correlations than individual SNPs, it is useful to perform haplotype analyses (34, 35). Even though each marker could be assessed separately, it is more informative to analyze them in groups. When multiple markers in a unique chromosome were used to determine their association with a disease, analyzing of haplotype could be more informative than independent analyses of individual markers (36). Linkage analysis using polymorphic markers linked to *UGT1A1* gene region could facilitate carrier detection and prenatal diagnosis of CN-1, CN-2, and GS in families with an affected individual. To our knowledge, molecular diagnosis of the syndrome has been based on direct mutation analysis.

Table 5. Comparison of minor allele frequency (MAF) of the rs4148326 and rs4124874 markers between the Iranian population and data from populations presented in UCSC genome browser database

Population	rs4148326 C allele frequency	rs4124874 T allele frequency
IRI	33.96	39.4
CEU	43.94	56.06
CHB	30.36	69.64
CHD	34.12	65.88
YRI	63.47	10.24
GIH	60.23	39.77
ASW	54.22	20.73
JPT	33.14	66.86
LWK	69.44	11.11
MEX	49.35	48.70
MKK	69.88	15.59
TSI	42.61	57.39
Average	52.11	38.78

CEU (northern and western Europe), CHB (Han Chinese in Beijing, China), CHD (Chinese Ancestry in Metropolitan Denver, CO, US), YRI (Yoruba in Ibadan, Nigeria), GIH (Gujarati Indians in Houston, TX), ASW (African Ancestry in SouthWestern United States), JPT (Japanese in Tokyo, Japan), LWK (Luhya in Webuye, Kenya), MEX (Mexican Ancestery in Los Angeles, CA, US), Masai in Kinyawa, Kenya MKK and TSI (Toscani in Italia)

In the present study, among the markers present in the *UGT1* gene complex region, the intra-genic marker, rs4148326 and extra-genic marker, rs4124874 were investigated. The allele and haplotype frequency of the markers were estimated in the Iranian population. The results showed that both markers were in high allelic frequency. As shown in Table 5, the MAF of the SNP markers was compared with the data from other populations. The observed heterozygosity for the markers that was calculated based on the identified unrelated heterozygous individuals indicated a relatively high heterozygosity for these markers in the Iranian population. In view of the high level of MAF and observed heterozygosity, rs4148326 and s4124874 SNPs could be suggested as informative markers in the Iranian population when compared to other populations. Therefore, these markers could be considered as the first markers to be introduced in carrier detection and prenatal diagnosis of the CN-1, CN-2 and GS in the Iranian population.

Our study samples consisted of unrelated healthy individuals and family trios. Combining data when data are collected through different study designs, such as family trios and unrelated samples, achieves more accuracy and is cost-effective compared with analyzing each data separately (37). Estimation of haplotype frequency of the markers in families and unrelated individuals showed three out of four haplotypes to be common in the Iranian population with frequencies ≥5% using both the FBAT and Arlequin software. These result indicated that the estimated frequencies of common haplotypes did not differ significantly using phase-known versus phase-unknown data. Furthermore, the estimation of D' and χ2 for the pair of markers showed that they were in strong LD.

Therefore, a large percentage of the chromosomes in this population may display common haplotypes for these two markers.

Conclusion

Altogether, the results from the present study showed that the degree of heterozygosity for rs4148326 and s4124874 markers were high in the Iranian population. Our results also demonstrated the ethnic specificity in degree of heterozygosity and informative haplotypes of the markers. These findings suggested the presence of genetic diversity at these two markers in the *UGT1A1* gene region. These markers could be considered in families with at least one affected child for carrier detection and prenatal diagnosis with a high degree of accuracy regardless of the nature of mutation. The data also suggested that the combination of rs4148326-s4124874 could be considered as an informative tool in performing carrier and prenatal diagnosis of *UGT1A1* gene mutations in the Iranian families with the CN type 1 and 2 syndrome as well as GS.

Acknowledgment

The results described in this paper were part of student thesis. The work was supported by research funds from the Department of Research of University of Isfahan, Isfahan, Iran.

References

1. Borlak J, Thum T, Landt O, Erb K, Hermann R. Molecular diagnosis of a familial nonhemolytic hyperbilirubinemia (Gilbert's syndrome) in healthy subjects. Hepatology 2000; 32:792-795.
2. Sampietro M, Iolascon A. Molecular pathology of Crigler-Najjar type I and II and Gilbert's syndromes. Haematologica 1999; 84:150-157.
3. Kadakol A, Ghosh SS, Sappal BS, Sharma G, Chowdhury JR, Chowdhury NR. Genetic lesions of bilirubin uridine-diphosphoglucuronate glucuronosyltransferase (UGT1A1) causing Crigler-Najjar and Gilbert syndromes: Correlation of genotype to phenotype. Hum Mutat 2000; 16:297-306.
4. Zhou Y, Wang S-n, Li H, Zha W, Wang X, Liu Y, *et al.* Association of UGT1A1 variants and hyperbilirubinemia in breast-fed full-term Chinese infants. PloS One 2014; 9:e104251.
5. Canu G, Minucci A, Zuppi C, Capoluongo E. Gilbert and Crigler najjar syndromes: An update of the UDP-glucuronosyltransferase 1A1 gene mutation database. Blood Cells Mol Dis 2013; 50:273-280.
6. Costa E. Hematologically important mutations: Bilirubin UDP-glucuronosyltransferase gene mutations in Gilbert and Crigler–Najjar syndromes. Blood Cells Mol Dis 2006;36:77-80.
7. Moghrabi N, Clarke DJ, Boxer M, Burchell B. Identification of an A-to-G missense mutation in exon 2 of the UGT1 gene complex that causes Crigler-Najjar syndrome type 2. Genomics 1993; 18:171-173.
8. Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de Boer A, Oostra BA, *et al.* The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. New Engl J Med 1995; 333:1171-1175.
9. Seppen J, Bosma P, Roy Chowdhury J. Discrimination between Crigler-Najjar syndromes type I and II by expression of mutant bilirubin-UDP-glucuronosyltransferase. J Clin Invest 1994; 94:2385.
10. Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 2014; 133:1-9.
11. Francoual J, Trioche P, Mokrani C, Seboui H, Khrouf N, Chalas J, *et al.* Prenatal diagnosis of Crigler–Najjar syndrome type I by single-strand conformation polymorphism (SSCP). Prenatal Diag 2002; 22:914-916.
12. Kaniwa N, Kurose K, Jinno H, Tanaka-Kagawa T, Saito Y, Saeki M, *et al.* Racial variability in haplotype frequencies of UGT1A1 and glucuronidation activity of a novel single nucleotide polymorphism 686C> T (P229L) found in an African-American. Drug Metab Dispos 2005; 33:458-465.
13. Milton JN, Sebastiani P, Solovieff N, Hartley SW, Bhatnagar P, Arking DE, *et al.* A genome-wide association study of total bilirubin and cholelithiasis risk in sickle cell anemia. PloS One 2012; 7:e34741.
14. Kringen MK, Piehler AP, Grimholt RM, Opdal MS, Haug KBF, Urdal P. Serum bilirubin concentration in healthy adult north-europeans is strictly controlled by the UGT1A1 TA-repeat variants. PloS One 2014; 9:e90248.
15. Horsfall LJ, Hardy R, Wong A, Kuh D, Swallow DM. Genetic variation underlying common hereditary hyperbilirubinaemia (Gilbert's syndrome) and respiratory health in the 1946 British birth cohort. J Hepatol 2014; 61:1344-1351.
16. Jones AG, Ardren WR. Methods of parentage analysis in natural populations. Mol Ecol 2003; 12:2511–2523.
17. Miller S, Dykes D, Polesky H. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988; 16:1215.
18. Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res 2001; 29:e88-e.
19. Chiapparino E, Lee D, Donini P. Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. Genome 2004; 47:414-420.
20. Collins A, Ke X. Primer1: primer design web service for tetra-primer ARMS-PCR. Open Bioinform J 2012; 6:55-58.
21. Newton C, Graham A, Heptinstall L, Powell S, Summers C, Kalsheker N, *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 1989; 17:2503-2516.
22. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 1995; 86:248-249.
23. Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis. User manual ver2; 2000; 2:2496-2497.
24. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 2000; 50:211-223.
25. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005; 21:2128-2129.
26. Lewontin R. The interaction of selection and linkage. II. Optimum models. Genetics 1964; 50:757.

27. Zhao JH. 2LD, GENECOUNTING and HAP: Computer programs for linkage disequilibrium analysis. Bioinformatics 2004; 20:1325-1326.
28. Zhao H, Nettleton D, Soller M, Dekkers J. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet Res 2005; 86:77-87.
29. Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. Nat Rev Genet 2011; 12:703-714.
30. Karolchik D, Hinrichs AS, Kent WJ. The UCSC genome browser. Curr Protoc Bioinformatics 2009; 33:1-.4.
31. Fazeli Z, Vallian S. Estimation haplotype frequency of BglII/EcoRI/VNTR markers at the PAH gene region in Iranian population. Int J Hum Genet 2009; 9:115-121.
32. Kraemer D, Scheurlen M. [Gilbert disease and type I and II Crigler-Najjar syndrome due to mutations in the same UGT1A1 gene locus]. Med Klin (Munich) 2002; 97:528-532.
33. Morioka I, Morikawa S, Yusoff S, Harahap ISK, Nishimura N, Yokoyama N, *et al.* Genetic disorders associated with neonatal jaundice. East J Med 2013; 15:155-162.
34. Yang J, Cai L, Huang H, Liu B, Wu Q. Genetic variations and haplotype diversity of the UGT1 gene cluster in the Chinese population. PloS One 2012; 7:e33988.
35. Judson R, Stephens JC, Windemuth A. The predictive power of haplotypes in clinical response. Pharmacogenomics 2000; 1:15-26.
36. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. Annu Rev Med 2005; 56:303-320.
37. Wen SH, Tsai MY. Haplotype association analysis of combining unrelated case-control and triads with consideration of population stratification. Front Genet 2014; 5:103.