

ORIGINAL ARTICLE

3'-UTR variations and G6PD deficiency

Farahnaz Amini^{1,2} and Endom Ismail¹

The combination of two silent mutations, c.1311C>T in exon 11 and IVS11 T93C (glucose-6-phosphate dehydrogenase (G6PD) 1311T/93C), with unknown mechanism, have been reported in G6PD-deficient individuals in Asian populations including Malaysian aboriginal group, Negrito. Here, we report the screening of *G6PD* gene in 103 Negrito volunteers using denaturing high-performance liquid chromatography (dHPLC) and direct sequencing. A total of 48 individuals (46.6%) were G6PD deficient, 83.3% of these carried G6PD 1311T/93C with enzyme activity ranging from 1.8 to 4.8 U gHb⁻¹. Three novel single-nucleotide polymorphisms (SNPs), rs112950723, rs111485003 and rs1050757, were found in the *G6PD* 3'-untranslated region (UTR). Strong association was observed between haplotype 1311T/93C and rs1050757G, which is located inside the 35 bp AG-rich region. *In silico* analysis revealed that the transition of A to G at position rs1050757 makes significant changes in the *G6PD* mRNA secondary structure. Moreover, putative micro (mi)RNA target sites were identified in 3'-UTR of *G6PD* gene, two of these in the region encompassing rs1050757. It could be speculated that rs1050757 have a potential functional effect on the downregulation of mRNA and consequently G6PD deficiency either by affecting mRNA stability and translation or miRNA regulation process. This is the first report of biochemical association of an SNP in 3'-UTR of *G6PD* gene and the possible role of mRNA secondary structure.

Journal of Human Genetics (2013) 58, 189–194; doi:10.1038/jhgc.2012.155; published online 7 February 2013

Keywords: G6PD deficiency; miRNA; Negrito; 3'-UTR

INTRODUCTION

Genetic variations in the *G6PD* gene are responsible for G6PD deficiency in humans. Other than mutations that cause G6PD enzyme deficiency, few polymorphic sites have been found in *G6PD* exons and introns. Overall, seven silent mutations in exons^{1–3} and five polymorphic sites have been reported in intronic regions of *G6PD* gene.⁴ Also, about 186 ethnic reliant nucleotide variations in the *G6PD* gene have been reported.⁵ Most of these variants are single missense mutations, with the rest being either double or triple missense mutations or small in frame deletions.⁶ With only one exception, all these mutations alter the protein sequence of the G6PD enzyme by either amino-acid substitution, deletion or alternative splicing. The exception is for a combination of c.1311C>T in exon 11 with IVS11 T93C (designated here as G6PD 1311T/93C). This combination is a special G6PD variant where the carrier is deficient without any changes to the protein sequence of the G6PD enzyme. Interestingly, some individuals with genotype 1311T/93C show significant reduction in G6PD enzyme level,^{7–10} while some individuals with this genotype present normal G6PD activity.⁷

Although comprehensive studies have identified the molecular basis of G6PD deficiency worldwide, some pertinent questions remain to be addressed. For instance, several studies have reported deficient samples with normal nucleotide sequences in the coding regions of *G6PD* gene.^{7,11–14} Unfortunately, these uncharacterised-deficient

samples have never undergone further studies for the involvement of other possible mechanism(s) of gene expression and regulation. The importance of mRNA secondary structure, processing, regulation and also the role of regulatory mechanisms involving regulatory proteins or micro (mi)RNAs have not been extensively studied with regards to G6PD deficiency. The roles of the untranslated regions (UTRs) of the *G6PD* gene have also not received much attention.

In this study, we sought to determine if any single-nucleotide polymorphisms (SNP) in the coding regions, 5'-UTR and 3'-UTR of *G6PD* gene, have the potential association with G6PD 1311T/93C and to confirm if this is responsible for the enzyme deficiency in affected individuals.

SUBJECTS AND METHODS

This study was approved by the ethics committee of the University Kebangsaan Malaysia Hospital. All subjects gave their written informed consent.

In our previous study, we attempted to identify the molecular basis of G6PD deficiency in 25 deficient individuals from one of the Malaysia aborigine group called Negrito.¹⁵ Our earlier results showed that G6PD 1311T/93C is the most common G6PD variant in the sampled Negrito population. No other mutations were detected in the remaining exons or adjacent regions of the *G6PD* gene for subjects with G6PD 1311T/93C. In this study, blood was collected from 103 consenting volunteers (48 males and 55 females) from four

¹School of Biosciences and Biotechnology, Faculty of Science and Technology, UKM, Bangi, Selangor, Malaysia and ²School of Anti-aging, Aesthetic and Regenerative Medicine, Faculty of Medical Science, UCSI, Kuala Lumpur Campus (North Wing), Cheras, Kuala Lumpur, Malaysia

Correspondence: Dr E Ismail, School of Biosciences and Biotechnology, Faculty of Science and Technology, UKM, Bangi, Selangor 43600, Malaysia.

E-mail: eismail@ukm.my

Received 29 July 2012; revised 26 November 2012; accepted 4 December 2012; published online 7 February 2013

subethnic groups of Negrito i.e. Kintak, Lanoh, Jahai and Bateq. G6PD quantification test was undertaken using the G6PD Kit from RANDOX Laboratory (Antrim, UK) according to the manufacturer's instructions. Genomic DNA was extracted using the salting out method.¹⁶ The oligonucleotides used as primers were either designed by online primer-BLAST program or obtained from the published data.⁷ A denaturing high-performance liquid chromatography (dHPLC)-based assay was used to detect the presence of *G6PD* mutations according to the previously published studies.^{7,17} The 5'- and 3'-UTR regions of the *G6PD* gene in all of the study samples were amplified and sequenced in four different reactions: 3'-UTR1 primers 1F (5'-GAGCCCTGGGCACCCACCTC-3') and 1R (5'-TCTGTGGGC TGGAGTGA-3') (320 bp); 3'-UTR2 primers 2F (5'-TCACTCCAGCCCAACA GA-3') and 2R (5'-GGTCCTCAGGGAAGCAAA-3') (397 bp); 5'-UTR1 primers 3F (5'-AGGCGGGAAACCGGACAGT-3') and 3R (5'-GTCCCCCTT CGCTCTCGGGT-3') (574 bp); 5'-UTR2 primers 4F (5'-ACCCCGAGAGCG AAGGGAC-3') and 4R (5'-CGGCTGGGCATTGGGGAGTG-3') (330 bp).

The following web-based bioinformatic tools were implemented in this study. The RegRNA program (<http://regna.mbc.nctu.edu.tw/>)¹⁸ and MicroInspector (<http://bioinfo.uni-plovdiv.bg/microinspector/>)¹⁹ were utilized to identify the miRNA binding sites inside 3'-UTR of *G6PD* gene. Secondary structures of the full-length *G6PD* mRNA and also 3'-UTR were predicted using GeneBee (<http://www.genebee.msu.su/genebee.html>), mFold (<http://mobyle.pasteur.fr/cgi-bin/portal.py>)²⁰ and CLC Main Workbench 5.6.1 (<http://www.clcbio.com>) (CLC bio, Aarhus, Denmark).

RESULTS

Overall, 48 individuals (17 males and 31 females) out of 103 Negrito individuals were found to be G6PD-deficient indicating a prevalence of 46.6%. Accordingly, the enzyme activity was in the range 0.86–5.7 U gHb⁻¹ in deficient individuals (including severe and partial deficiency) and 7.8–10.6 U gHb⁻¹ in the normal group.

dHPLC and direct sequencing indicated that 40 out of 48 deficient individuals carry a combination of c.1311C>T and IVS11 T93C (*G6PD* 1311T/93C). No other mutations were found in the remaining *G6PD* gene exons or flanking introns. Data from *G6PD* activity test indicated that the range of enzyme activity in *G6PD* 1311T/93C carriers were between 10 and 60% of mean *G6PD* activity for normal individuals, prompting a categorization of the *G6PD* 1311T/93C in World Health Organization class III form of *G6PD* deficiency. None of the individuals with normal *G6PD* activity carry this combination. However, Viangchan variant (G871A) was detected in one male subject with haplotype 1311T/93C. One male subject and one female subject carried the Coimbra variant. The *G6PD* variant for five deficient individuals was not identified. The range of enzyme activity in uncharacterized individuals was from 1.2 to 5.7 U gHb⁻¹.

Although no SNPs were detected in the 5'-UTR of *G6PD* gene in our study samples, dHPLC and direct sequencing of 3'-UTR revealed the presence of three SNPs in the 3'-UTR region of *G6PD* gene (Figure 1 and Supplementary Figure 1). All three SNPs have been successfully reported to the dbSNP NCBI (<http://www.ncbi.nlm.nih.gov/projects/SNP>) and the following NCBI IDs have been obtained: rs112950723 (ss218178027) for SNP +272G to A, rs111485003 (ss218178028) for +304T to C and rs1050757 (ss218178024) for +357A to G. Table 1 shows the frequency of these SNPs among the sampled Negrito population. All the Negrito individuals who carry *G6PD* 1311T/93C and also the Viangchan carrier with genotype 1311T/93C had mutant allele rs1050757G. Interestingly, heterozygous females for *G6PD* 1311T/93C also were heterozygous for rs1050757. Moreover, it was found that in the individuals with *G6PD* 1311T/93C, those siblings who carry normal alleles at c.1311 and IVS11 93 also have normal allele rs1050757A. On the other hand, all those siblings or blood relatives who carry *G6PD* 1311T/93C also carry rs1050757G. In addition, other observed haplotypes; that is, 1311T/93T, 1311C/93T and 1311C/93C, were always found to be associated with rs1050757A.

To gain further insights into the role of rs112950723, rs111485003 and rs1050757 in *G6PD* deficiency, the potential effect of the sequence alterations on the mRNA secondary structure was evaluated. The secondary structures of the full-length normal mRNA and mutant counterparts; that is, mRNA containing mutant allele 1311T and mRNA containing mutant haplotypes 1311T/rs112950723A, 1311T/rs111485003C and 1311T/rs1050757G, were predicted and compared. Results indicated that nucleotide change at position 1311 alone does not affect the mRNA folding. In another words, the mRNA conformation corresponding to the mutant allele 1311T was the same as the normal allele 1311C, only that a marginal change in the minimum free energy was observed from $\Delta G = -918.8$ to -917.7 kcal mol⁻¹ (Figure 2). Subsequently, secondary structures of the full-length mRNA containing other mutant haplotypes were compared with the normal transcript. It was observed that mutant haplotypes of 1311T/rs112950723A and 1311T/rs111485003C have no effect on the mRNA folding, while significant alterations were predicted in the secondary structure of mutant transcript of 1311T/rs1050757G (Figures 2, 3 and 4). Notably, it was found that the internal base pairing in some regions of mutant mRNA led to extensively different stem-loop and internal loop structures. Specifically, the presence of different stem-loop conformation at the upstream of start codon in mutant haplotype compared with the

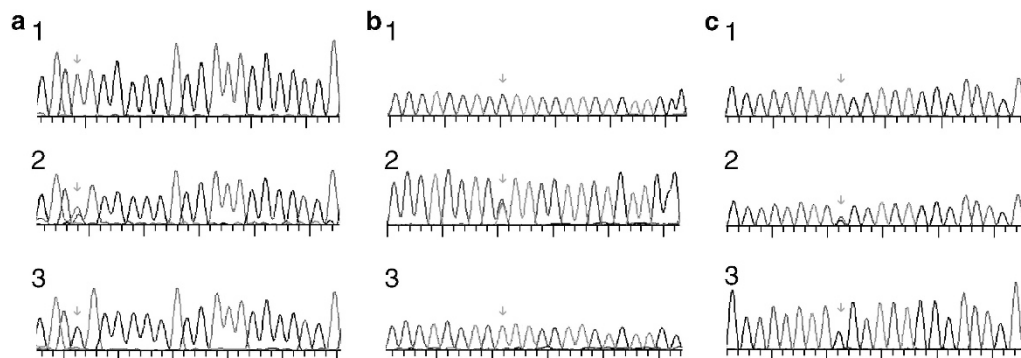


Figure 1 Partial nucleotide sequence of normal, heterozygote and homozygote female subjects, respectively, for forward strand of rs1050757 (a1, a2 and a3), reverse strand of rs112950723 (b1, b2 and b3) and reverse strand of rs111485003 (c1, c2 and c3). Arrows show position of each SNP. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

normal mRNA was remarkable (Figure 5). This conformation made a stable secondary structure in the vicinity of start codon in mutant mRNA.

In the next step, bioinformatic programs were implemented to detect putative miRNA target sites in the wild-type sequence of *G6PD* 3'-UTR. Accordingly, it was predicted that there are 152 putative target sites for 118 miRNAs in normal 3'-UTR. Additionally, 3'-UTR with mutant variant of rs1050757G was submitted to evaluate its ability to create or abolish miRNA target sites. Results indicated that transition of A to G creates two more miRNA binding sites in the regions encompassing rs1050757, namely hsa-miR-1238 and hsa-miR-877*. Notably, rs1050757G is located inside the "seed region" of these miRNAs. However, further analysis revealed that in normal mRNA only 1 out of 152 potential target sites are accessible to the three miRNAs (hsa-miR-206, hsa-miR-613 and hsa-miR-1) owing to the mRNA folding (Supplementary Figure 2). Moreover, transition of A

to G at position rs1050757 leads to accessibility of two more sites for miRNAs binding (hsa-miR-3173 and hsa-miR-149*).

DISCUSSION

The results revealed high prevalence of G6PD deficiency among the Negrito group as compared with other Southeast Asians and with other major groups in Malaysia (almost 10 times higher).²¹⁻²⁴

The presence of G6PD 1311T/93C/rs1050757G as predominant variant in *G6PD* gene among the Negrito group (83.3%) indicates that they are a very homogenous population with respect to the *G6PD* gene. This is contrary to the G6PD profile of other groups in Malaysia as they are very heterogenous for this gene. Ainoon *et al.*^{25,26} reported 11 different G6PD variants in Malay and 10 in Chinese-Malay. As the different evolutionary factors such as natural selection and genetic drift can affect pattern of human genetic diseases in the concept of diversity, then the highest ever reported frequency of G6PD 1311T/93C in Negrito might be due to these evolutionary factors and most probably because of genetic drift and inbreeding. Notably, available literatures show that with one exception, G6PD 1311T/93C has been reported only in deficient individuals in Asia.^{7-9,13,18,27} The exception is from Tunisia, in which 2 out of the 42 study samples carried G6PD 1311T/93C.³ Even so, a high degree of gene flow has been reported for Tunisian people.²⁸

The presence of three SNPs rs112950723, rs111485003 and rs1050757 in 3'-UTR of *G6PD* gene in G6PD-deficient individuals appears to be novel and has not been reported previously. However,

Table 1 Presence of each SNP in deficient and normal individuals in the Negrito population

Population	Phenotype	rs112950723A	rs111485003C	rs1050757G
Negrito	G6PD 1311T/93C	6	1	41
	Normal	0	3	0

Abbreviation: SNP, single-nucleotide polymorphism.

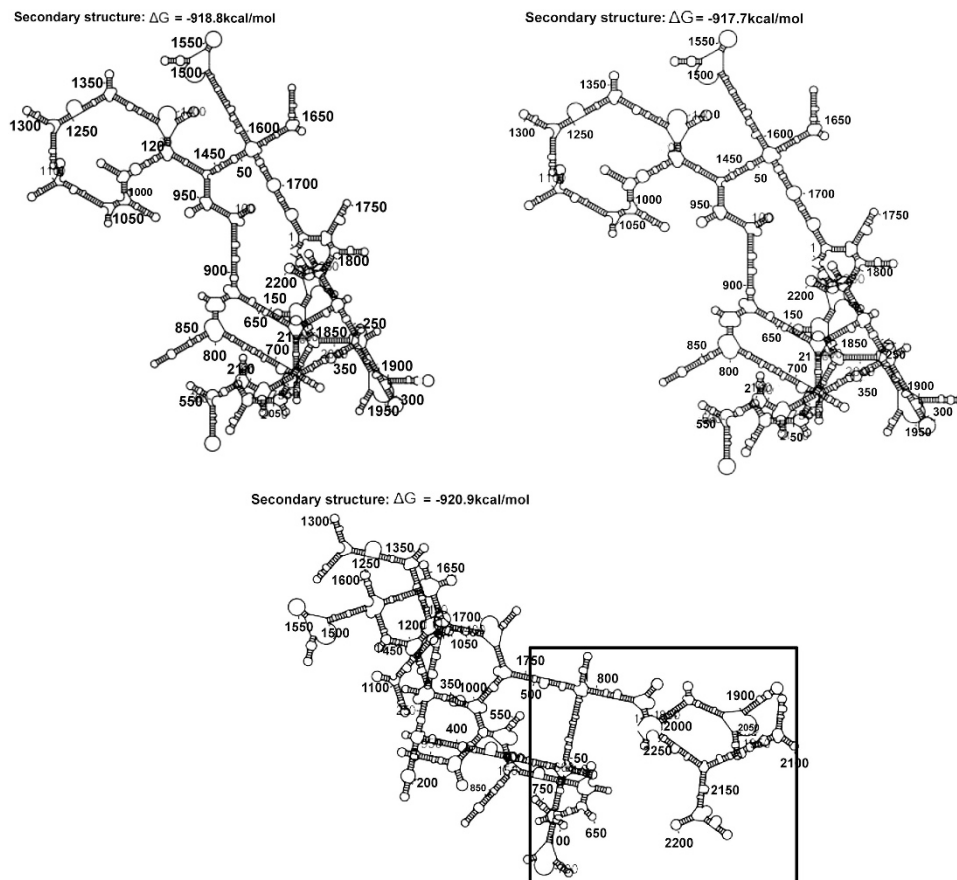


Figure 2 mRNA secondary structures as predicted by the CLC RNA Workbench software. Normal sequence (left panel, up) and sequence with mutant allele of 1311T (right panel, up) compared with the mRNA conformation in haplotype 1311T/rs1050757G (middle panel, down). The region with significant change in secondary structure is boxed in black. The Gibbs energy of formation for each folded RNA is shown.

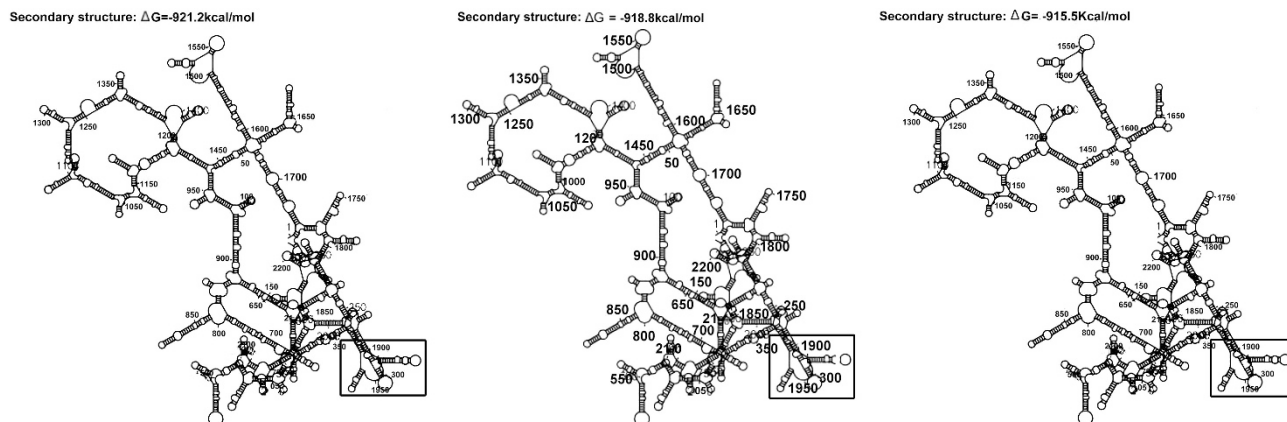


Figure 3 mRNA secondary structures for normal mRNA (middle) compared to transcripts with mutant haplotypes 1311T/rs111485003C (left) and 1311T/rs112950723A (right). The region encompassing rs111485003C and rs112950723A boxed in black (Figure 4 visualizes these boxes). These predicted structures by the CLC RNA WorkBench indicate that these substitutions do not affect the mRNA folding. Only marginally higher minimum free energy was observed for rs111485003C ($\Delta G = -921.2$), while it was slightly lower for rs112950723A ($\Delta G = -915.5$).

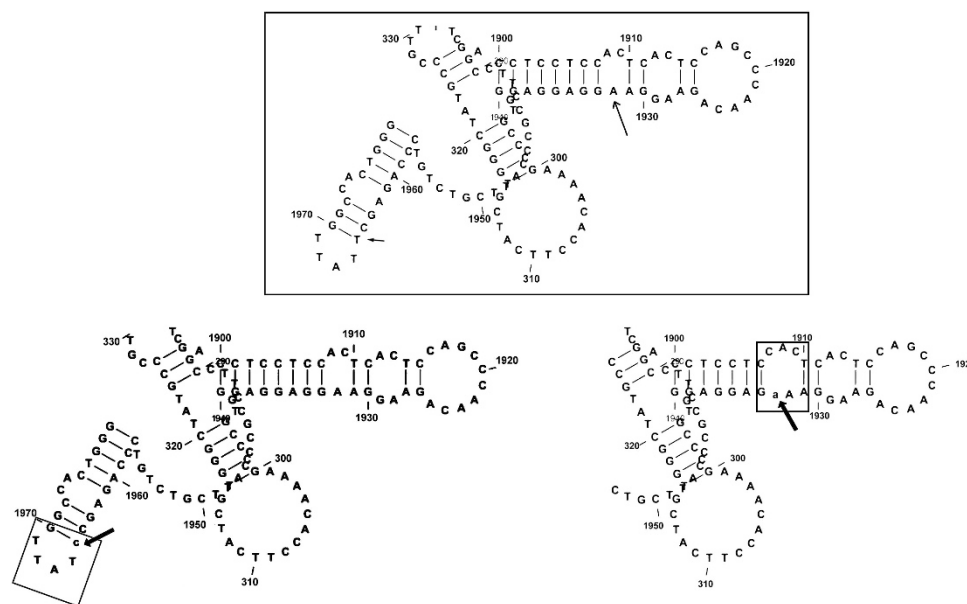


Figure 4 Effect of transition T to C at position rs111485003 (left) and G to A in rs112950723 (right). These changes do not affect the secondary structure of their respective transcripts even in their surrounding region. The upper panel shows the normal mRNA. Arrows show position of each SNP.

our comprehensive literature search revealed that only three reports had investigated changes in 3'-UTR of *G6PD* gene in their respective deficient populations. However, they did not find any SNP in the 3'-UTR.^{22,29,30} In this study, we evaluated the potential associations of these SNPs and enzyme deficiency among our studied samples. Accordingly, a strong association was observed between rs1050757G and *G6PD* 1311T/93C as this SNP was solely observed in individuals who carried *G6PD* 1311T/93C. We noticed that there are two AG-rich regions in 3'-UTR of human *G6PD* gene with 17 and 35 nucleotides length. rs112950723 is surrounded by 17bp AG-rich region (5'-AGAAGGAAGGAGGAGGG-3') and rs1050757 is located inside the 35bp AG-rich region (5'-AGGGTGGGAGGGAGGGACAAGGGGGGAGAAAGGGG-3'). Accumulating evidence on the importance of nucleotide-rich elements in 3'-UTR, such as AU-, C-, CU- and AG-rich element, have been implicated in mRNA stability through changes in mRNA secondary structure.^{31–33} Moreover, as the role of structure on RNA function is akin to that of protein, we thus

investigated the effect of each SNP on mRNA secondary structure. We observed that rs112950723 and rs111485003 have no effect on the mRNA folding (Figures 3 and 4), while rs1050757G caused significant alterations in the secondary structure of mutant transcript (Figures 2 and 5). This is in concurrence with a number of reports, which indicate that non-functional SNPs in a gene usually do not affect the mRNA secondary structure.^{34,35} Furthermore, data from a recent publication indicates that a majority of exonic mutations, which leads to amino-acid substitution, also cause some degree of alteration to the mRNA structure.³⁶ On the other hand, growing bodies of evidence suggest that mRNA structural rearrangement might be considered as a proposed mechanism for disease-associated mutations.³⁷ As well, it is believed that structural conformation of mRNA influences gene expression and regulation. Similarly, Shen *et al.*³⁸ demonstrated that synonymous mutation in coding regions of human alanyl tRNA synthetase and human replication protein A led to different mRNA folding structure. Moreover, experimental procedures using RNase

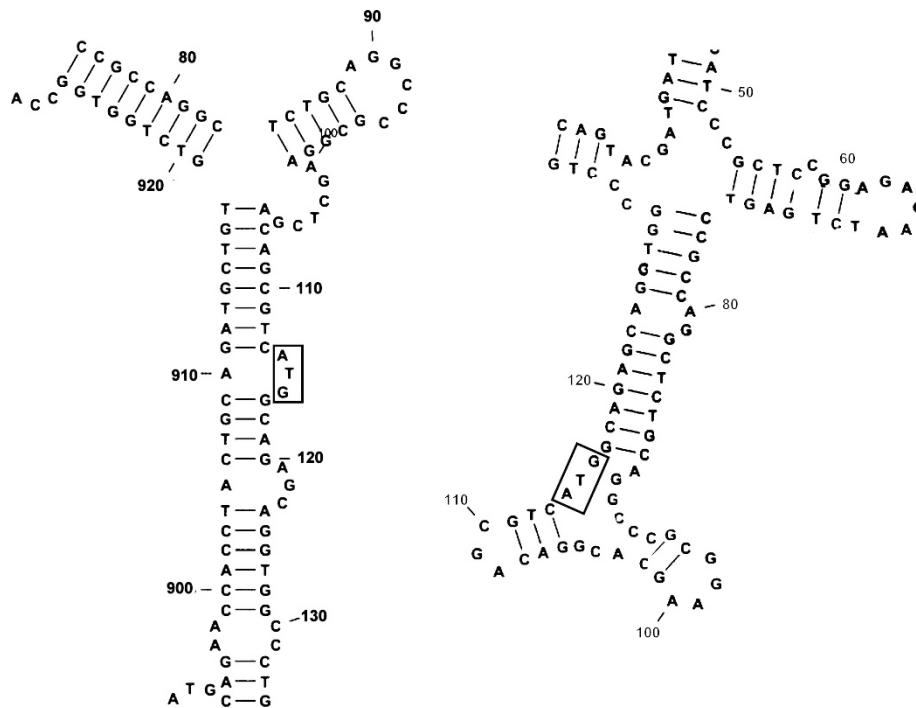


Figure 5 A comparison of the predicted secondary structures of 5'-UTR in the normal (left panel) and mutant (right panel) mRNA. In mutant transcript, a stable secondary structure in the vicinity of start codon (boxed region) was predicted by involvement of nucleotides 54–129.

assay show that both normal and mutant alleles in a synonymous SNP, which have not affected the secondary structure of human protrombin mRNA, display similar sensitivities to single- and double-strand-specific RNase.³⁹ Consistent with these data, in this study, it was assumed that detection of the allele 1311T alone, which neither changes amino-acid sequence nor mRNA secondary structure, is a neutral mutation probably due to its position in loop structure. However, the combination of mutant alleles 1311T and rs1050757G caused significant alterations in the secondary structure of *G6PD* mRNA. Even so, this is the first report on combination of synonymous mutations affecting secondary structure of mRNA in *G6PD* deficiency; there are other reports in prokaryotes and eukaryotes with similar effects. Duan *et al.*⁴⁰ studied functional effect(s) of six naturally silent mutations in the human dopamine receptor D2 (DRD2). They found that one of these silent mutations (C957T) altered the predicted mRNA secondary structure and led to a decrease in mRNA stability and translation, consequently changed dopamine-induced upregulation of DRD2 expression. Moreover, they have reported that another silent mutation, which did not cause any change in mRNA structure by itself, repealed the above effects when it was combined with C957T. This demonstrates that combinations of synonymous mutations may result in functional effects, which are significantly different from the effect of each isolated mutation. This is in parallel with the results from this study in which a combination of 1311T/93C/rs1050757G leads to *G6PD* deficiency.

We proposed two linked hypotheses to explain *G6PD* deficiency in genotype of 1311T/93C/rs1050757G and both of these hypotheses are related to mRNA secondary structure. Firstly, using predicted mRNA secondary structure, we observed that mRNA conformation in mutant haplotype 1311T/93C/rs1050757G made a stable secondary structure in the vicinity of start codon, which suggested a negative effect on mRNA translation and consequently low amount of *G6PD* enzyme (Figure 5). This conclusion is somewhat derived from other reports, which mentioned that stem and loop structures introduced

between the cap and the AUG codon could never facilitate initiation of translation.^{41,42} Moreover, the majority of eukaryotic mRNAs initiate translation according to the ribosome linear scanning model.⁴³ In this model, it is proposed that highly stable secondary structures, especially self-complementary sequences, which form stem-loop structures within mRNA, can inhibit translation.^{33,41,44} Secondly, the new conformation of mRNA placed some sequences in the open structures such as loop or arc; therefore, these sequences are no longer involved in the base pairing and they are accessible to post-transcriptional modulators such as miRNA. In other words, in the secondary structure of 3'-UTR with normal allele rs1050757A, a possible standard Watson-Crick paired duplex was observed in the majority of predicted miRNA binding sites. However, mutant allele G shows reshuffling of the base pairings resulting in a different predicted secondary structure in some regions (Figure 2). According to Kertesz *et al.*⁴⁵ accessibility of target sites is now known to be an even more accurate predictor for miRNA prediction. Most of the predicted miRNA sites by computational programs are inactive owing to their inaccessibility. Consistent with the above data, in this study, miRNA repression is considered as a possible mechanism to explain *G6PD* deficiency in individuals with haplotype 1311T/93C/rs1050757G. It is postulated that miRNA binding site is likely inaccessible in the wild-type mRNA because of its secondary structure, but in the mutant mRNA the target sites are accessible to miRNA(s) for perfect complementary of seed region. Notably, a number of miRNAs can induce mRNA degradation as well as translational repression, whereas the others act primarily on translation.⁴⁶ However, we were not able to predict which of these mechanisms may be involved in *G6PD* deficiency. In concurrence with our second hypothesis, there is a report on significant alteration of miRNA expression in erythrocytes with another type of hemolytic anemia.⁴⁷

In conclusion, high frequency and molecular homogeneity of *G6PD* deficiency among the sampled Negrito population is remarkable. In addition, to the best of our knowledge, this is the first report

on association of 3'-UTR variation and G6PD deficiency. We also demonstrated that rs1050757G may potentially change the secondary structure of G6PD mRNA and repress G6PD expression by affecting either translation or stability of mRNA with the possible involvement of miRNA. However, further experimental studies are necessary to reliably determine the role(s) of mRNA secondary structure on G6PD deficiency.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the Department of Orang Asli Affairs (JHEOA), all Negrito Orang Asli involved in this study and their Tok Batin and Tok Penghulu. We also thank MOSTI for financial support under ScienceFund Grant 02-01-02-SF0316. We would also like to express our gratitude to Dr Reza Farhour for his help with blood collections.

- D'Urso, M., Luzzatto, L., Perroni, L., Ciccocicola, A., Gentile, G., Peluso, I. *et al.* An extensive search for RFLP in the human glucose-6-phosphate dehydrogenase locus has revealed a silent mutation in the coding sequence. *Am. J. Hum. Genet.* **42**, 735–741 (1988).
- Pinto, F. M., Gonzalez, A. M., Hernandez, M., Larruga, J. M. & Cabrera, V. M. Sub-Saharan influence on the Canary Islands population deduced from G6PD gene sequence analysis. *Hum. Biol.* **68**, 517–522 (1996).
- Daoud, B. B., Mosbeh, I., Prehu, C., Chaouachi, D., Hafsia, R. & Abbes, S. Molecular characterization of erythrocyte glucose-6-phosphate dehydrogenase deficiency in Tunisia. *Pathol. Biol.* **56**, 260–267 (2008).
- Vulliamy, T. J., Othman, A., Town, M., Nathwani, A., Falusi, A. G., Mason, P. J. *et al.* Polymorphic sites in the African population detected by sequence analysis of the glucose-6-phosphate dehydrogenase gene outline the evolution of the variants A and A. *Proc. Natl Acad. Sci. USA* **88**, 8568–8571 (1991).
- Minucci, A., Giardina, B., Zuppi, C. & Capoluongo, E. Glucose-6-phosphate dehydrogenase laboratory assay: How, when, and why? *IUBMB Life* **61**, 27–34 (2009).
- Cappellini, M. D. & Fiorelli, G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74 (2008).
- Jiang, W., Yu, G., Liu, P., Geng, Q., Chen, L., Lin, Q. *et al.* Structure and function of glucose-6-phosphate dehydrogenase-deficient variants in Chinese population. *Hum. Genet.* **119**, 463–478 (2006).
- Laosombat, V., Sattayasevana, B., Chotsampancharoen, T. & Wongchanchailert, M. Glucose-6-phosphate dehydrogenase variants associated with favism in Thai children. *Int. J. Hematol.* **83**, 139–143 (2006).
- Wang, J., Luo, E., Hirai, M., Arai, M., Abdul-Manan, E., Mohamed-Isa, Z. *et al.* Nine different glucose-6-phosphate dehydrogenase (G6PD) variants in a Malaysian population with Malay, Chinese, Indian and Orang Asli (aboriginal Malaysian) backgrounds. *Acta Med. Okayama* **62**, 327–332 (2008).
- Yu, G. L., Jiang, W. Y., Du, C. S., Lin, Q. D., Chen, L. M., Tian, Q. H. *et al.* Complex mutations of 1311 C→T in exon 11 and 93 T→C in intron 11 in G6PD gene. *Zhonghua Xue Ye Xue Za Zhi* **25**, 610–612 (2004).
- Arambula, E., Aguilar, L. J. & Vaca, G. Glucose-6-phosphate dehydrogenase mutations and haplotypes in Mexican Mestizos. *Blood Cells Mol. Dis.* **26**, 387–394 (2000).
- Barisic, M., Korac, J., Pavlinac, I., Krzelj, V., Marusic, E., Vulliamy, T. *et al.* Characterization of G6PD deficiency in southern Croatia: description of a new variant, G6PD split. *J. Hum. Genet.* **50**, 547–549 (2005).
- Laosombat, V., Sattayasevana, B., Janejindamai, W., Viprasakit, V., Shirakawa, T., Nishiyama, K. *et al.* Molecular heterogeneity of glucose-6-phosphate dehydrogenase (G6PD) variants in the south of Thailand and identification of a novel variant (G6PD Songklanagarind). *Blood Cells Mol. Dis.* **34**, 191–196 (2005).
- Nuchprayoon, I., Louichareon, C. & Charoenvaj, W. Glucose-6-phosphate dehydrogenase mutations in Mon and Burmese of southern Myanmar. *J. Hum. Genet.* **53**, 48–54 (2008).
- Amini, F., Ismail, E. & Zilfalil, B. Prevalence and molecular study of G6PD deficiency in Malaysian Orang Asli. *Int. Med. J.* **41**, 351–354 (2011).
- Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
- Tsang, C.-P., Chien-Ling, H., Kowit-Yu, C., Iou-Jih, H. & Daniel Tsun-Yee, C. Rapid detection of glucose-6-phosphate dehydrogenase gene mutations by denaturing high-performance liquid chromatography. *Clin. Biochem.* **38**, 973–980 (2005).
- Huang, D. A., Wang, X. Y., Wang, Z., Zhou, D. F. & Cai, W. W. Molecular characterization of 71 cases of glucose-6-phosphate dehydrogenase deficiency in Hainan province. *Zhonghua Xue Ye Xue Za Zhi* **28**, 250–254 (2007).
- Rusinov, V., Baev, V., Minkov, I. N. & Tabler, M. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res.* **33**, W696–W700 (2005).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
- Ainon, O., Alawayah, A., Yu, Y. H., Cheong, S. K., Hamidah, N. H., Boo, N. Y. *et al.* Semiquantitative screening test for G6PD deficiency detects severe deficiency but misses a substantial proportion of partially-deficient females. *Southeast Asian J. Trop. Med. Public Health* **34**, 405–414 (2003).
- Hue, N. T., Charliou, J. P., Chau, T. T., Day, N., Farrar, J. J., Hien, T. T. *et al.* Glucose-6-phosphate dehydrogenase (G6PD) mutations and haemoglobinuria syndrome in the Vietnamese population. *Malar J.* **8**, 152 (2009).
- Iwai, K., Hirono, A., Matsuoka, H., Kawamoto, F., Horie, T., Lin, K. *et al.* Distribution of glucose-6-phosphate dehydrogenase mutations in Southeast Asia. *Hum. Genet.* **108**, 445–449 (2001).
- Matsuoka, H., Wang, J., Hirai, M., Arai, M., Yoshida, S., Kobayashi, T. *et al.* Glucose-6-phosphate dehydrogenase (G6PD) mutations in Myanmar: G6PD Mahidol (487G>A) is the most common variant in the Myanmar population. *J. Hum. Genet.* **49**, 544–547 (2004).
- Ainon, O., Boo, N. Y., Yu, Y. H., Cheong, S. K., Hamidah, H. N. & Lim, J. H. Complete molecular characterisation of glucose-6-phosphate dehydrogenase (G6PD) deficiency in a group of Malaysian Chinese neonates. *Malay J. Pathol.* **26**, 89–98 (2004).
- Ainon, O., Yu, Y. H., Amir Muhriz, A. L., Boo, N. Y., Cheong, S. K. & Hamidah, N. H. Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Malays. *Hum. Mutat.* **21**, 101 (2003).
- Moiz, B., Nasir, A., Moatter, T., Naqvi, Z. A. & Khurshid, M. Population study of 1311 C/T polymorphism of glucose 6 phosphate dehydrogenase gene in Pakistan—an analysis of 715 X-chromosomes. *BMC Genet.* **10**, 41 (2009).
- Manni, F., Leonardi, P., Patin, E., Berrebi, A., Khodjet el Khil, H., Skorecki, K. *et al.* A Y-chromosome portrait of the population of Jerba (Tunisia) to elucidate its complex demographic history. *Bull. Mém. Soc. Anthropol. Paris* **17**, 1–2 (2005).
- Karadsheh, N. S., Moses, L., Ismail, S. I., Devaney, J. M. & Hoffman, E. Molecular heterogeneity of glucose-6-phosphate dehydrogenase deficiency in Jordan. *Haematologica* **90**, 1693–1694 (2005).
- van Bruggen, R., Bautista, J. M., Petropoulou, T., de Boer, M., van Zwieten, R., Gomez-Gallego, F. *et al.* Deletion of leucine 61 in glucose-6-phosphate dehydrogenase leads to chronic nonspherocytic anemia, granulocyte dysfunction, and increased susceptibility to infections. *Blood* **100**, 1026–1030 (2002).
- Knowles, J. W., Erickson, L. M., Guy, V. K., Sigel, C. S., Wilder, J. C. & Maeda, N. Common variations in noncoding regions of the human natriuretic peptide receptor A gene have quantitative effects. *Hum. Genet.* **112**, 62–70 (2003).
- Peyvandi, F., Garagiola, I., Palla, R., Marziliano, N. & PM, M. Role of the 2 adenine (g.11293_11294insAA) insertion polymorphism in the 3' untranslated region of the factor VII (FVII) gene: molecular characterization of a patient with severe FVII deficiency. *Hum. Mutat.* **26**, 455–461 (2005).
- Wang, L. & Wessler, S. R. Role of mRNA secondary structure in translational repression of the maize transcriptional activator Lc1.2. *Plant Physiol.* **125**, 1380–1387 (2001).
- Jiang, Z., Michal, J. J., Tobey, D. J., Wang, Z., Macneil, M. D. & Magnuson, N. S. Comparative understanding of UTS2 and UTS2R genes for their involvement in type 2 diabetes mellitus. *Int. J. Biol. Sci.* **4**, 96–102 (2008).
- Chen, J. M., Ferec, C. & Cooper, D. N. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum. Genet.* **120**, 1–21 (2006).
- Johnson, A. D., Trumbower, H. & Sadee, W. RNA structures affected by single nucleotide polymorphisms in transcribed regions of the human Genome. *WebmedCentral Bioinform.* **2**, WMC001600 (2011).
- Halvorsen, M., Martin, J., Broadaway, S. & Laederach, A. Disease-associated mutations that alter the mRNA structural ensemble. *PLoS Genet.* **6**, e1001074 (2010).
- Shen, L. X., Basilion, J. P. & Stanton, V. P. Single-nucleotide polymorphisms can cause different structural folds of mRNA. *Proc. Natl Acad. Sci. USA* **96**, 7871–7876 (1999).
- Liu, X., Jiang, Y. & Russell, J. E. A potential regulatory role for mRNA secondary structures within the prothrombin 3'UTR. *Thromb. Res.* **126**, 130–136 (2010).
- Duan, J., Wainwright, M., Cameron, J., Saitou4, N., Sanders1, A., Gelernter, J. *et al.* Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* **12**, 205–216 (2003).
- Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 19867–19870 (1991).
- Sehgal, A., Hughes, B. T. & Espenshade, P. J. Oxygen-dependent, alternative promoter controls translation of tco1+ in fission yeast. *Nucleic Acids Res.* **36**, 2024–2031 (2008).
- Kozak, M. Features in the 5'-non-coding sequences of rabbit a and b-globin mRNAs that affect translational efficiency. *J. Mol. Biol.* **235**, 95–110 (1994).
- Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
- Lee, P. H. & Shatky, H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res.* **36**, D820–D824 (2008).
- Chen, S. Y., Wang, Y., Telen, M. J. & Chi, J. T. The genomic analysis of erythrocyte microRNA expression in sickle cell diseases. *PLoS One* **3**, e2360 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)