



# Integrative identification of the pathogenic role of a novel G6PD missense mutation c.697G>C

Hongyang Zhang<sup>1,2,#</sup>, Danyi Peng<sup>1,2,3,#</sup>, Yi Shu<sup>1,2</sup>, Dan Zhu<sup>1,2</sup>, Weiwei Hu<sup>4</sup>, Chaowen Yu<sup>1,2</sup>, Juan Zhang<sup>1,2</sup>, Shan Liu<sup>1,2</sup>, Kexing Wan<sup>1,2</sup>, Zhaojian Yuan<sup>1,2</sup>, Hao Liu<sup>1,2</sup>, Dongjuan Wang<sup>1,2</sup>, Tingting Jiang<sup>1,2</sup>, Jie Yu<sup>1,5</sup>, Penghui Zhang<sup>1,6</sup>, Lin Zou<sup>1,2</sup>

<sup>1</sup>Center of Clinical Molecular Medicine & Newborn Screening Center, Ministry of Education Key Laboratory of Child Development and Disorders; National Clinical Research Center for Child Health and Disorders (Chongqing), China International Science and Technology Cooperation base of Child development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China; <sup>2</sup>Chongqing Engineering Research Center of Stem Cell Therapy, Chongqing, China; <sup>3</sup>Department of Respiratory Medicine, Children's Hospital of Chongqing Medical University, Chongqing, China; <sup>4</sup>Department of respiratory and critical care medicine, First Affiliated Hospital of Chongqing Medical University, Chongqing, China; <sup>5</sup>Department of Hematology, The Children's Hospital of Chongqing Medical University, Chongqing, China; <sup>6</sup>Center of Clinical Laboratory, The Children's Hospital of Chongqing Medical University, Chongqing, China

**Contributions:** (I) Conception and design: L Zou, P Zhang; (II) Administrative support: L Zou, Y Shu, S Liu; (III) Provision of study materials or patients: L Zou, D Peng, D Wang, C Yu, T Jiang, P Zhang, J Yu; (IV) Collection and assembly of data: H Zhang, D Peng, W Hu, D Zhu, K Wan, J Zhang, Z Yuan, H Liu; (V) Data analysis and interpretation: L Zou, H Zhang, D Peng, Y Shu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Lin Zou, MD, PhD. Center of Clinical Molecular Medicine, Children's Hospital of Chongqing Medical University, 136 Zhongshan Er Road, Chongqing 400014, China. Email: zoulin74@126.com.

**Background:** Glucose-6-phosphate dehydrogenase (G6PD) deficiency is a hereditary disease caused by pathogenic mutations of *G6PD*. While most of the pathogenic variants of *G6PD* have been annotated, hemolysis of unknown etiology but analogous to that in G6PD deficiency persists, implying the existence of undocumented pathogenic variants. In our previous study, we reported four novel *G6PD* variants in China, for which the pathogenicity remains to be verified.

**Methods:** The variants were verified by exogenous expression in HEK-293 cells, and their functions were predicted by PolyPhen-2 and SIFT. The CRISPR/Cas9 system was exploited to edit the *G6PD* c.697G>C variant in HEK-293 cells and K562 cells. The expression of *G6PD* was detected by quantitative PCR (qPCR) and western blotting. The cell growth capacity was detected by the CCK-8 assay and crystal violet staining. The G6PD enzyme activity was reflected by the G6P/6PG ratio test. The apoptosis of cells was detected by Annexin V-APC/7-AAD staining. The secondary and crystallographic structures were denoted according to the literature and PyMOL software. The G6PD protein was purified from lysis of transformed *Escherichia coli* (*E. coli*) cell with Ni-charged Resin Column. The enzymatic activity was detected at different temperatures.

**Results:** The G6PD activity of exogenous *G6PD* c.697G>C in HEK-293 cells was significantly lower than that of wild type (WT) *G6PD*, a finding that was consistent with the observation in clinical samples. The functional predictions conducted by different algorithms indicated the damage role of the *G6PD* c.697G>C variant in its enzymatic activity. We recapitulated the *G6PD* c.697G>C variant both in HEK-293 cells and K562 cells by adapting the CRISPR/Cas9 strategy. Using distinct cell lines expressing the *G6PD* c.697G>C variant endogenously, we confirmed the deteriorative role of the *G6PD* c.697G>C variant in its enzymatic activity. Regarding the secondary and crystallographic structure, we found a mutated amino acid approaching the structural NADP<sup>+</sup> binding site. Finally, we demonstrated the c.697G>C variant compromised the thermal stability of G6PD protein.

**Conclusions:** Our data delineated the pathogenic role of *G6PD* c.697G>C variant for G6PD deficiency,

implying the wide usage of CRISPR/Cas9 for genetic disease research.

**Keywords:** Glucose-6-phosphate dehydrogenase deficiency (G6PD deficiency); pathogenic variant; novel missense mutation; genome editing; CRISPR/Cas9

Submitted May 14, 2020. Accepted for publication Nov 09, 2020.

doi: 10.21037/atm-20-3941

View this article at: <http://dx.doi.org/10.21037/atm-20-3941>

## Introduction

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is the most common enzymopathy, threatening almost 400 million people worldwide. Etiologically, this X chromosome-linked hereditary genetic defect is caused by mutations in *G6PD*, resulting in protein variants with different levels of enzymatic activity, and is associated with several biochemical and clinical phenotypes. G6PD is a key regulatory enzyme of the pentose phosphate pathway and provides reducing power to all cells. For human red blood cells, the pentose phosphate pathway is the only source of NADPH. Therefore, G6PD deficiency leads to a distinct degree of hemolysis, clinically presented as neonatal jaundice, acute hemolytic anemia, and favism (1,2). The World Health Organization (WHO) has stratified patients with G6PD deficiency into four classes based on clinical presentation and G6PD activity in patient blood samples: Class I (<10% activity and CNSHA), Class II (<10% activity and hemolytic episodes), Class III (on 10–60% activity and hemolytic episodes), and Class IV (on 60–150% activity and no clinical manifestations) (3). Studies on G6PD deficiency and *G6PD* genetic mutations are warranted.

To prevent severe symptoms caused by G6PD deficiency, the most beneficial and economic strategy is newborn screening (4). Recently, more than 200 mutations of *G6PD* were detected and characterized, indicating an imbalanced distribution in individuals (5). According to our previous study, the 11 most common mutations were responsible for approximately 97% of the clinical manifestations and were tested routinely. Little attention was given to rare mutations, particularly unverified mutations. This ambiguity risked individuals being in unknown danger. After large-scale screening, we detected four novel missense variants of *G6PD* (6). The impact of those variants on the function of G6PD needs to be determined.

The CRISPR/Cas9 system is an effective means of introducing targeted mutations at specific sites in the

genome. Guided by sgRNA, Cas9 can be programmed to induce DNA double-strand breaks (DSBs) in the targeted genomic locus. Combined with the homologous recombinant, we can edit the genomic DNA with high efficiency (7,8). Unlike expressing the exogenous gene using expression vectors, CRISPR/Cas9 enables the delineation of the function of somatic mutations based on the endogenous genome and is more rigorous than other strategies (9,10). In this study, we verified the pathogenic role of the four newly detected *G6PD* variants based on exogenous gene vectors. After screening, we focused on *G6PD* c.697G>C, which induced a significant decrease in its enzyme activity. We exploited CRISPR/Cas9 to establish two *G6PD* c.697G>C variant cell lines, further confirming its deleterious role in its enzymatic activity. Finally, we identified the mutated amino acid position by integrating the crystallographic data and demonstrated the mutation compromised thermal stability of its protein.

We have presented the study in accordance with the MDAR reporting checklist (available at <http://dx.doi.org/10.21037/atm-20-3941>).

## Methods

### *Clinical samples*

Information on the clinical samples involved in this study was obtained from participants in our previous study (6), which was approved by the Research Ethics Committees of Children's Hospital of Chongqing Medical University, and was registered in the Chinese Clinical Trial Registry (ChiCTR-SOC-17014057). Before the collection and use of the clinical samples, their guardians were provided with detailed information about the benefits and risks of the study. Written informed consent forms were signed by the guardians according to the Declaration of Helsinki (as revised in 2013). The completed consent forms were kept on file.

### *Cell culture and chemicals*

HEK-293 cells (CLS Cat# 300192/p777\_HEK293, RRID:CVCL\_0045) and human erythrocytic leukemia K562 cells (NCI-DTP Cat# K-562, RRID:CVCL\_0004) were obtained from American Type Culture Collection (ATCC, Manassas, USA). HEK-293 cells were maintained in complete Dulbecco's Modified Eagle Medium (DMEM) (Gibco, USA), while K562 cells were maintained in RPMI-1640 supplemented with 10% fetal bovine serum (Invitrogen, USA), L-glutamine and penicillin/streptomycin. Primaquine was purchased from Med Chem Express (MCE, USA).

### *Total RNA isolation and touchdown-qPCR analysis (qPCR)*

Total RNA was extracted using TRIzol Total RNA Isolation Reagent (Invitrogen) according to the manufacturer's protocol. To generate the RT products for mRNA quantification, random hexamers were annealed to RNA at 70 °C for 5 min, which was then reverse transcribed using M-MuLV Reverse Transcriptase (NEB, USA). The resultant RT products were used as qPCR templates. The qPCR program was carried out as previously reported (11). *GAPDH* was used as a reference gene. All the samples were normalized to *GAPDH* expression using the  $2^{-\Delta\Delta Ct}$  method. The qPCR primer sequences are listed in [Table S1](#).

### *Immunoblotting*

Western blotting procedures were performed as previously described (12). Briefly, all the protein samples for immunoblotting were subjected to SDS-PAGE and transferred to Polyvinylidene fluoride (PVDF) membranes, which were blocked and incubated overnight with primary antibodies for G6PD (CST, Cat# 12263) and GAPDH (Proteintech Cat# 10494-1-AP, RRID:AB\_2263076). After washing the membranes, they were incubated with Horseradish peroxidase (HRP)-conjugated Affinipure Goat Anti-Rabbit IgG (H+L) (Proteintech, Cat# SA00001-2). Immune-reactive signals were detected using the ECL Kit (Millipore, USA).

### *G6PD catalytic activity analysis*

The activity of G6PD was detected by DBS fluorescence

testing (Darui Biotech, China) and NBT G6P/6PG testing (Micky Med, China), and the experiments were carried out according to the manufacturer's instructions.

### *DNA sequencing*

DNA was extracted from cells using the TIANamp® Genomic DNA Kit (Tiangen Biotech, China) according to the manufacturer's protocol. The *G6PD* gene was amplified using PCR, and the products were Sanger sequenced at BGI (BGI, China). Chromas (Technelysium Pty. Ltd, Australia) was used for data analysis. The PCR primer sequences are listed in [Table S1](#).

### *Crystal violet staining*

After primaquine treatment, the cells were washed with phosphate-buffered saline and stained with 0.5% crystal violet/formalin solution at room temperature for 30 min. The stained cells were washed with tap water and air-dried before the acquisition of macrographic images.

### *CCK8 and cell apoptosis analysis*

For cell growth analysis, cells were seeded in 96-well plates, FBS starved for 24 h, and then cultured in complete medium. The CCK-8 Reagent (MCE, USA) was added to each well, followed by incubation at 37 °C for 150 min and absorbance measurement at 450 nm. Each assay condition was performed in triplicate.

For the apoptosis assay, the cells were stained with Annexin V-PE and 7-AAD according to the manufacturer's instructions (BD Biosciences, USA). The stained cells were subjected to FACS analysis using LSR Fortessa (BD Biosciences, USA) and analyzed using FlowJo software as described.

### *Plasmid construction and cell transfection*

sgRNAs targeting human *G6PD* (RefSeq NM\_001042351.3) were designed using the Zhang laboratory CRISPR design tool (<http://crispr.mit.edu>) and then were subcloned into plasmid px458-sgRNA (RRID:Addgene\_48138) as previously described. The sgRNA sequences are listed in [Table S1](#). The 600-nt *G6PD* mutation templates (PAGE purified, BGI) were designed using a homologous genomic

flanking sequence centered around the predicted CRISPR/Cas9 cleavage site and containing the c.697G>C variant (13). Next, the templates were subcloned into pSEH-Donor (Kindly provided by Prof. Tong-Chuan He) (14).

G6PD expression vectors were constructed as follows: The coding regions of human G6PD were PCR amplified from cDNA templates. Then the G6PD was subcloned into pcDNA3.1 vector (GeneChem, China) and pET-33B(+) vector (Novagen, USA). The mutations were introduced using the QuickChange kit (Agilent, USA).

For HEK-293 cells, subconfluent cells seeded in T-25 cell culture flasks were transfected with a pDNA-PEI mixture of 5 µg of px458-sgRNA vector, 5 µg of pSEH-Donor and 30 µg of PEI (Polysciences Inc., USA) under the serum-free condition for 3 h. For K562 cells,  $3 \times 10^6$  cells suspended in 300 µL of RPMI-1640 media, 8 µg of px458-sgRNA and 8 µg of pSEH-Donor were added to the cell suspension and incubated at room temperature for 15 min. Electroporation was performed as previously described (15).

#### ***Expression and purification of the His-tagged G6PD protein***

The recombinant G6PD protein was purified according to the previous reports (16,17). In brief, the pET-33b-G6PD-His plasmids were transfected into the *Escherichia coli* (E. coli) BL21(DE3)pLysS competent cells (Promega, USA). After incubation in Luria-Bertani (LB) medium for 5 hours, the expression of G6PD was induced with isopropyl-1-thio-β-D-galactopyranoside (IPTG) (Sangon Biotech, China) 1.0 mM for 18 h at 37 °C. Then the cells were resuspended in lysis buffer and disrupted by sonication. After removing the cell debris by centrifuge, the resident supernatant was fractionated with 25% and 50% ammonium sulfate respectively. The resident protein was resolved with PBS buffer. The His-tagged G6PD protein was isolated with Ni-charged Resin Column (GenScript, China) according to the manufacturer's instructions. Finally, the G6PD enzymes were dialysed by centrifuging in Centricon YM-50 filtration tubes (Millipore, USA) and aliquoted for stocking at -80 °C.

#### ***Stability studies***

The purified G6PD proteins were kept on ice unless required and then incubated at various temperatures for 30 min and subsequently cooled to 4 °C. Residual activity was tested at 25 °C in triplicate.

#### ***Pathogenicity prediction***

SIFT (SIFT, RRID:SCR\_012813) and PolyPhen-2 (PolyPhen: Polymorphism Phenotyping, RRID:SCR\_013189) were used to predict the possible harmful influence of point mutations to G6PD. The amino acid sequence of G6PD protein was downloaded from the Protein Data Bank (Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), RRID:SCR\_012820), and the protein structure was analyzed by PyMOL (PyMOL, RRID:SCR\_000305).

#### ***Statistical analysis***

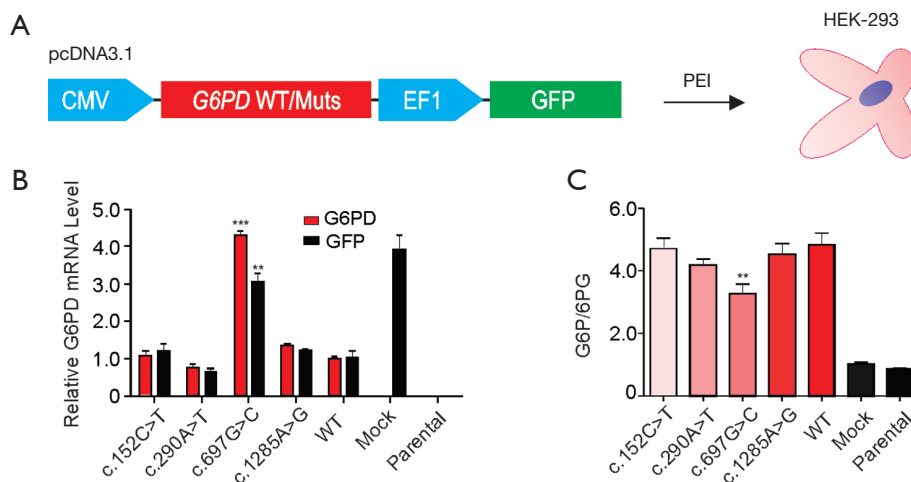
All the quantitative experiments were performed in triplicate and/or repeated in three independent experiments. The data was presented as means ± SD. Statistical significance between the treatment and control groups was determined by one-way analysis of variance and Student's *t*-test. A value of  $P < 0.05$  was considered significant.

## **Results**

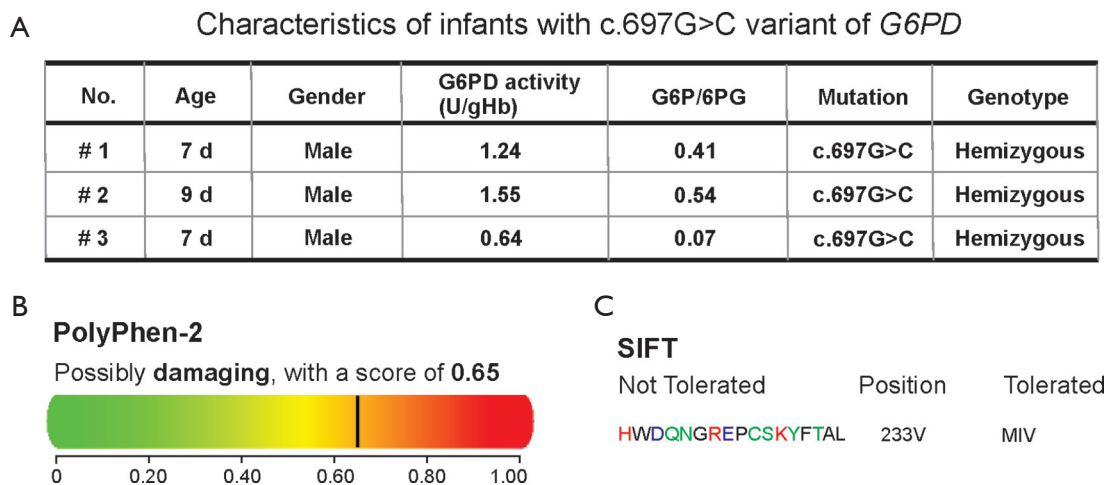
#### ***The G6PD c.697G>C variant was a potential pathogenic mutation***

In our previous study, we profiled the prevalence map of the G6PD variants in China based on newborn screening data from 2013 to 2017. Four new missense variants of G6PD, c.152C>T, c.290A>T, c.697G>C and c.1285A>G, were documented (6). To determine the role of those variants, we constructed GFP-tagged vectors to express wild-type G6PD, as well as its variants, in HEK-293 cells exogenously to test G6PD activity (Figure 1A). All the G6PD variants were expressed effectively in the transfected HEK-293 cells (Figure 1B); only the cells harboring the c.697G>C variant showed significantly lower G6PD activity than the observation of wild-type control (Figure 1C), indicating a damaging effect on the activity of this variant.

Based on the experimental data, we retrospectively reviewed the clinical characteristics of infants harboring those variants (Figure 2A, Figure S1A). The tests demonstrated all the blood samples derived from the infants with those novel mutations have an obvious decline on their G6PD activity, especially those harboring the c.697G>C variant (Figure 2A). According to the G6P/6PG ratio and WHO criteria (3), we stratified the G6PD c.697G>C variant into Class III G6PD deficiency. In addition to the G6PD



**Figure 1** The exogenous *G6PD* c.697G>C variant shows inferior enzymatic activity in HEK-293 cells. (A) HEK-293 cells with overexpression of wild-type *G6PD* or the four indicated variants of *G6PD*. (B) Relative mRNA level of *G6PD* and *GFP* in HEK-293 cells with or without overexpression of the indicated *G6PD* gene. Mock: HEK-293 cells transfected with the pcDNA3.1 vector; Parental: HEK-293 cells. (C) Ratio of *G6P/6PG* detected in HEK-293 cells with overexpression of the indicated *G6PD* gene. \*,  $P<0.05$ ; \*\*,  $P<0.01$ ; \*\*\*,  $P<0.001$ .



**Figure 2** The c.697G>C variant of *G6PD* gene is a potential pathogenic mutation. (A) Clinical manifestation of three patients with the c.697G>C variant. For the *G6PD* activity, the cut-off values were 2.6 (U/gHb) for males and 3.3 (U/gHb) for females. For the *G6P/6PG* ratio, the cut-off value for healthy people was 1.0-1.6. The predictive role of the c.697G>C variant was determined using the PolyPhen-2 (B) and SIFT (C) databases. Regarding the data from the SIFT database, the nonpolar, uncharged polar, basic, and acidic amino acids are marked in black, green, red, and blue, respectively.

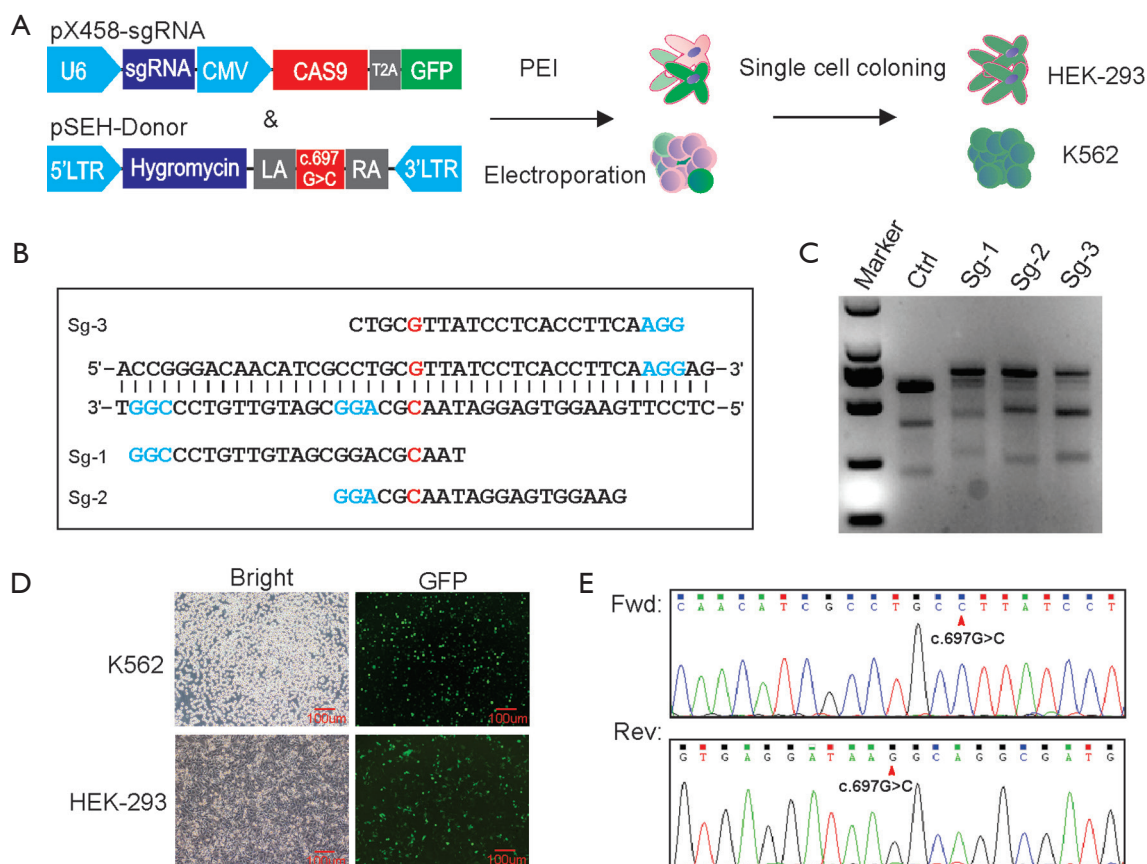
activity results, we also predicted the effect of the variants with different algorithms: SIFT, which uses evolutionary conservation, and PolyPhen-2, a machine learning method that combines chemical similarity, sequence information, and 3D structural information (Figure 2B,C, Figure S1B,C). Both algorithms indicated a damaging role of the *G6PD*

c.697G>C variant on its protein function (Figure 2B,C).

#### Confirming the pathogenic role of *G6PD* c.697G>C with endogenous gene editing

The recent availability of gene editing has provided an

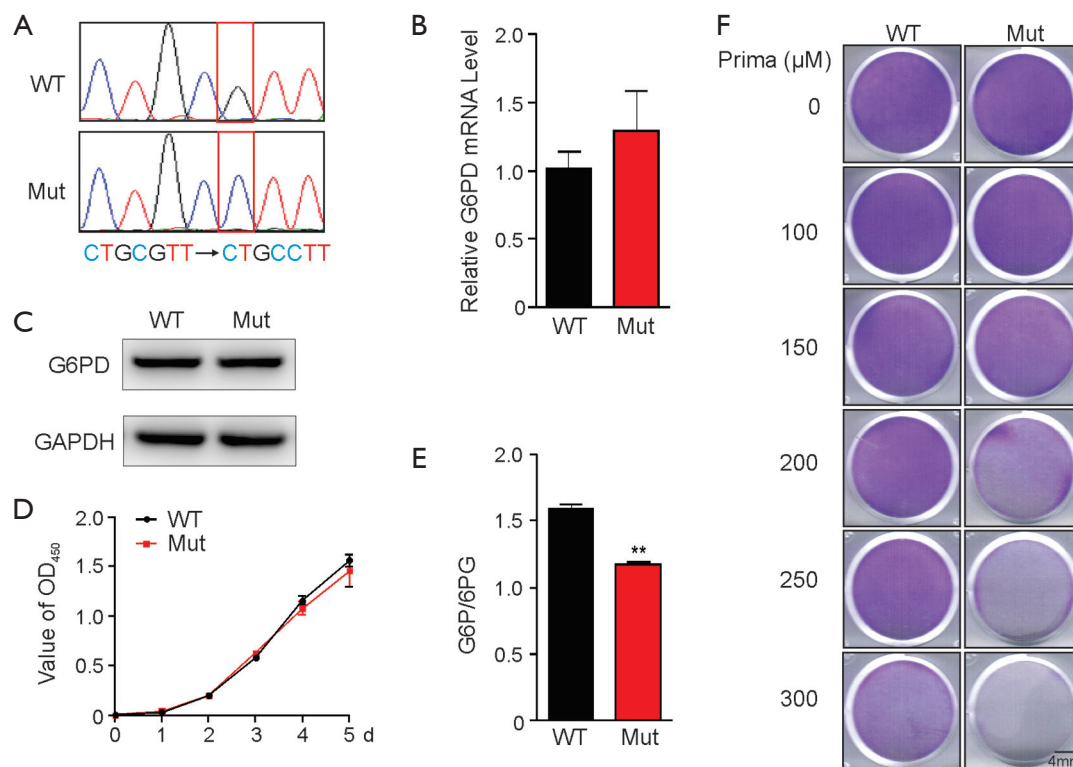




**Figure 3** CRISPR/Cas9-mediated c.697G>C of *G6PD* in HEK-293 and K562 cells. (A) Schematic representation of the conversion of c.697G>C at *G6PD* loci by CRISPR/Cas9 in HEK-293 and K562 cells. (B) The sgRNAs designed for creating the DBS around the c.697G of *G6PD*. (C) The Sg-3 showed a prior cutting efficiency via T7E1 digestion. Marker: DL2000 marker; Ctrl: positive control. (D) The transfection efficiency reflected by percentage of GFP positive cells. (E) The cells harbor c.697G>C variants was verified by Sanger sequencing.

excellent toolbox to delineate the pathogenic mutations detected in clinical samples. To verify the deleterious role of c.697G>C rigorously, we designed single guide RNAs (sgRNAs) targeting genomic DNA around the c.697G of *G6PD* and a homologous arm to convert the wild-type *G6PD* to the c.697G>C variant based on the CRISPR/Cas9 strategy (Figure 3A). We screened three different sgRNAs to create double-strand breaks (DSBs) around the c.697G of *G6PD* (Figure 3B) and selected Sg-3 for subsequent tests for its high efficiency (Figure 3C). PEI was used to transfect HEK-293 cells. After isolating and subcloning the GFP<sup>+</sup> single cells, we obtained several *G6PD* c.697G>C variant HEK-293 colonies (Figure 3D,E). Next, we mixed them to produce HEK-293 cells harboring *G6PD* c.697G>C variant (Figure 4A). We checked the expression level of HEK-

293 cells with *G6PD* c.697G>C variant and its parental control cells by qPCR and Western blotting. Neither cells showed significant alterations (Figure 4B,C). Next, HEK-293 cells with *G6PD* c.697G>C variant or wild type (WT) form were subconfluently seeded to test the growth capacity in routine culture medium. Few differences were observed by the CCK-8 assay (Figure 4D). Thus, the *G6PD* c.697G>C variant did not affect the growth of HEK-293 cells under conventional conditions. To confirm the effect of this variant on *G6PD* enzyme activity, we tested the G6P/6PG ratio. A less extensive decline was observed than that in patients' blood cells, and a significant decrease in the G6P/6PG ratio was observed in the *G6PD* c.697G>C variant HEK-293 cell line (Figure 4E). Moreover, we found HEK-293 cells with *G6PD* c.697G>C variant were more



**Figure 4** The endogenous c.697G>C of *G6PD* variant damages activity of protein in HEK-293 cells. (A) The c.697G>C variant of *G6PD* mediated by CRISPR/Cas9 in HEK-293 cells. Relative mRNA (B) and protein (C) level of the *G6PD* gene in cells with WT *G6PD* or the variant. (D) The results of CCK8 assay to show the effect of the variant on the cell growth of HEK-293 cells. (E) The ratio of G6P/6PG to demonstrate the effect of the variant on activity of protein in HEK-293 cells. (F) HEK-293 cell with wild type *G6PD* or the c.697G>C variant were treated with different dose of primaquine, stained the cell with crystal violet, and then scanned the plate with scanister. \*\*,  $P < 0.01$ . Scale bar, 4 mm.

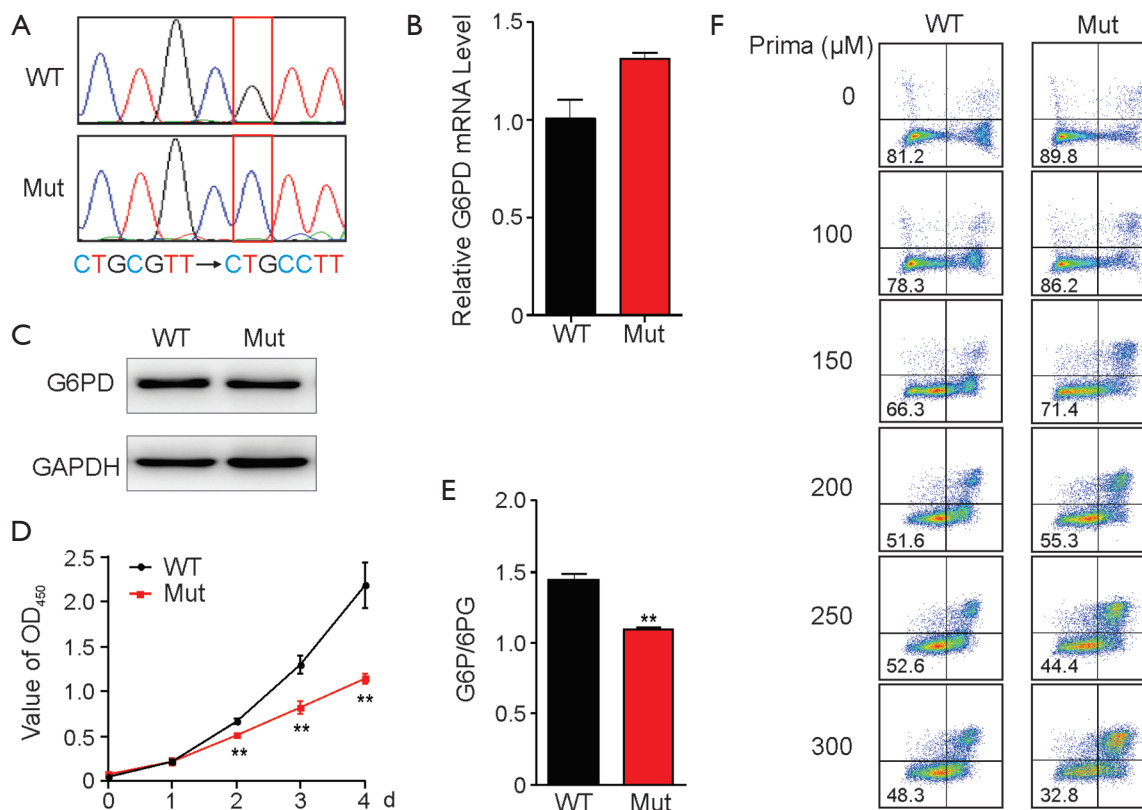
vulnerable to oxygen (18), as demonstrated by crystal violet staining (Figure 4F). These data confirmed the deleterious role of c.697G>C in *G6PD* activity.

We then expressed the *G6PD* c.697G>C variant in the erythroblastic cell line K562 cells (19) by adapting the CRISPR/Cas9 strategy. Because of the difficulty of transfecting with PEI, we adopted the electroporation method for K562 transfection (Figure 3D). After pooling the *G6PD* c.697G>C variant subcolonies, we obtained K562 cells with *G6PD* c.697G>C homozygous variant (Figure 5A). Consistent with the observation in HEK-293 cells, the *G6PD* c.697G>C variant in K562 cells did not interfere with gene expression (Figure 5B,C). Nevertheless, the growth capacity of the K562 cell was curbed when c.697G>C mutation occurred (Figure 5D). The decrease in the G6P/6PG ratio was also observed in the cells harboring the *G6PD* c.697G>C variant (Figure 5E). To test the effect

of oxygen on the viability of the two types of K562 cells, we treated the cells with different doses of primaquine, which could produce a mass of reactive oxygen. The apoptosis of K562 cells with *G6PD* c.697G>C variant was much more severe than that in K562 cells with wild-type *G6PD*, as demonstrated by Annexin-V staining (Figure 5F), although the two cell types had both undergone apoptosis in medium with high doses of primaquine.

#### ***The c.697G>C variant compromised the thermal stability of G6PD protein***

Considering that the clinical phenotype of the *G6PD* variants is largely determined by a trade-off between protein stability and catalytic activity, we aligned the mutated site to the locus of *G6PD* to delineate the function of the *G6PD* c.697G>C variant. The c.697G>C variant was located within



**Figure 5** The endogenous c.697G>C of *G6PD* variant damages activity of protein in K562 cells. (A) The c.697G>C variant of *G6PD* mediated by CRISPR/Cas9 in HEK-293 cells. Relative mRNA (B) and protein (C) level of the *G6PD* gene in K562 cells with WT *G6PD* or the variant. (D) CCK8 assay to show the effect of the variant on the cell growth of K562 cells. (E) The ratio of G6P/6PG to demonstrate the effect of the variant on activity of protein in K562 cells. (F) K562 cell with wild type or the c.697G>C variant of *G6PD* were treated with different dose of primaquine, and then the apoptosis of cells was detected with Annexin V staining. \*\*,  $P < 0.01$ .

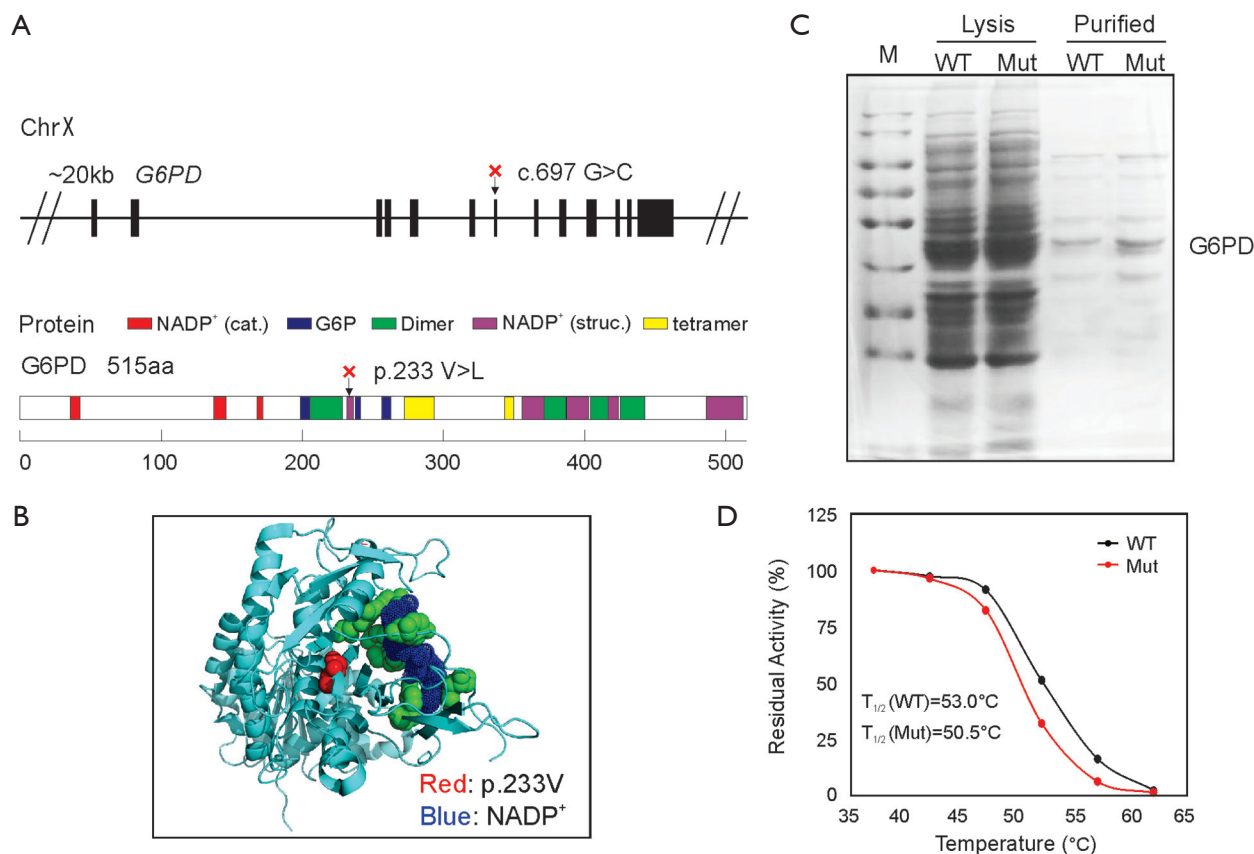
exon 7 of *G6PD*, leading to the p.Val233Leu missense mutation of the G6PD protein. We denoted the functional domain of G6PD according to previous reports, which found that the mutation occurred within the structural NADP<sup>+</sup> binding domain (Figure 6A). Furthermore, the crystallographic structure of human G6PD (PDB.2BH9) showed that p.233Val was located within the nine-stranded antiparallel  $\beta$ -sheet and approached to the amino acids that bind to structural NADP<sup>+</sup> (Figure 6B). This information indicated that p.Val233Leu of G6PD protein might interfere with its capacity to bind to structural NADP<sup>+</sup>, a function that is crucial to maintain the thermostability of the protein (17,20,21). To confirm this hypothesis, we constructed inducible G6PD expression plasmids and purified His-tagged G6PD protein from transformed *E. coli* lysis. After incubating at different temperatures, the enzymatic activity of those G6PD variants was tested. The

results showed that the WT form of G6PD was more stable than that of the *G6PD* c.697G>C variant (Figure 6C,D).

## Discussion

G6PD deficiency is the most common X chromosome-linked hereditary genetic defect. As a consequence of the functional damage of the protein, the clinical manifestation of this deficiency ranges from chronic non-spherocytic hemolytic anemia (CNSHA), acute hemolytic anemia, neonatal jaundice, favism to no obvious clinical symptoms, depending on the position of the mutation of *G6PD* (1,2,22). For males, the phenotype of the pathogenic *G6PD* variants is obvious because there is only one allele of the X chromosome. The activity of G6PD in blood could reflect the function of a *G6PD* variant and the diagnosis and management are easy. For females, because of the genetic





**Figure 6** The c.697G>C variant of *G6PD* compromises thermal stability of protein. (A) Genomic arrangement and domain structural features of *G6PD*. The mutated loci of *G6PD* are marked with a red cross. (B) Crystallography of human *G6PD*. *G6PD* p.233Val is marked in red, and the coupled NADP<sup>+</sup> is marked in blue. (C) SDS-PAGE of purified *G6PD* enzymes. M: molecular weight marker; lane 2–3, total cell lysis derived from *E. coli* cell transformed with indicated plasmid; lane 4–5, purified *G6PD*. (D) Thermal inactivation assays of WT *G6PD* and the variant after incubation at the indicated temperature. The  $T_{1/2}$ , the temperature at which the enzyme loses 50% of its original activity.

mosaics of the X-chromosome, heterozygous females with pathogenic variants randomly express one *G6PD* allele, causing a hemolytic risk (23). Depending on the activity of *G6PD*, it is difficult to diagnose *G6PD* deficiency in females, making genetic analysis necessary. Therefore, pathogenic annotation of the *G6PD* variants is very important.

Exogenous expression of protein variants using expression vectors is the conventional means to verify the function of gene mutations. Simultaneously expressing the variants and wild-type protein in cells to check the diversity of their function hinted at a pathogenesis role. Despite being widely used, several drawbacks exist, particularly the difficulty of normalizing the diversity of their function to their expression (24–26). CRISPR/Cas9 enables the

editing of genomic DNA directly, excluding the impact of the aforementioned issues. Additionally, CRISPR/Cas9 makes it possible to delineate the role of gene mutations in the noncoding region (27,28). Thus, for the first time, we adapted the CRISPR/Cas9 strategy to reveal a novel pathogenic mutation of *G6PD*, making a paradigm for further somatic mutation research.

To delineate the role of the c.697G>C of *G6PD* variant, two cell lines were edited by the CRISPR/Cas9 strategy, the nontumorous cell line HEK-293 and the erythroblastic leukemia cell line K562. For *G6PD* activity, the two cell lines were quite similar, showing a significant decline. Interestingly, the growth capacity of K562 cells expressing the pathogenic *G6PD* was inferior to that observed in K562 cells with wild-type *G6PD*, indicating that the increased

redox in leukemia cells might curb their growth capacity; similar results were reported by others (29-31). The findings indicate that G6PD could be a targetable molecule in leukemia for further research.

By exploiting the CRISPR/Cas9 strategy, we revealed the pathogenic role of the G6PD c.697G>C variant and stratified this variant into the Class III category, but the mechanism of this phenotype is still vague. Several characteristics determine the activity of G6PD protein, including the catalytic activity, thermostability and folding of the protein. Active human G6PD exists in a rapid dimer-tetramer equilibrium, which is affected by ionic strength and pH value. The mutants affecting the binding with structural NADP<sup>+</sup> displayed decreased thermostability and high susceptibility to chymotrypsin digestion. Recently, an integrative analysis based on the structural and statistical data highlighted the role of the trade-off between catalytic activity and protein stability (32). As shown by crystallography, p.233Val approached amino acids binding with structural NADP<sup>+</sup>, implying the thermostability of this variant might be damaged. Furthermore, we purified His-tagged recombination G6PD protein expressed by *E. coli* cells and revealed the mutation compromised the thermal stability of G6PD.

## Acknowledgments

The authors would like to acknowledge technical support provided by Prof. Tong-Chuan He of the Molecular Oncology Laboratory, The University of Chicago Medical Center.

**Funding:** This work was supported by Natural Scientific Foundation of China (Grant Number: #81870126 and #81570142 to LZ, #81900190 to YS), and Chongqing Municipal Social People's Livelihood Technology Innovation Project (Grant Number: #cstc2016shmszx130032). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Footnote

**Reporting Checklist:** The authors have completed the MDAR reporting checklist. Available at: <http://dx.doi.org/10.21037/atm-20-3941>

**Data Sharing Statement:** available at: <http://dx.doi.org/10.21037/atm-20-3941>

**Peer Review File:** Available at: <http://dx.doi.org/10.21037/atm-20-3941>

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-20-3941>). The authors have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was approved by the Research Ethics Committees of Children's Hospital of Chongqing Medical University, and registered on Chinese Clinical Trial Registry (NO.: ChiCTR-SOC-17014057). Written informed consent forms were signed by the guardians according to the Declaration of Helsinki (as revised in 2013). The completed consent forms were kept on file.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Cappellini MD, Fiorelli G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* 2008;371:64-74.
2. Luzzatto L, Arese P. Favism and Glucose-6-Phosphate Dehydrogenase Deficiency. *N Engl J Med* 2018;378:60-71.
3. Glucose-6-phosphate dehydrogenase deficiency. WHO Working Group. *Bull World Health Organ* 1989;67:601-11.
4. Mak CM, Lee HC, Chan AY, et al. Inborn errors of metabolism and expanded newborn screening: review and update. *Crit Rev Clin Lab Sci* 2013;50:142-62.
5. Gómez-Manzo S, Marcial-Quino J, Vanoye-Carlo A, et al. Glucose-6-Phosphate Dehydrogenase: Update and Analysis of New Mutations around the World. *Int J Mol Sci* 2016;17:2069.
6. Liu Z, Yu C, Li Q, et al. Chinese newborn screening for the incidence of G6PD deficiency and variant of G6PD

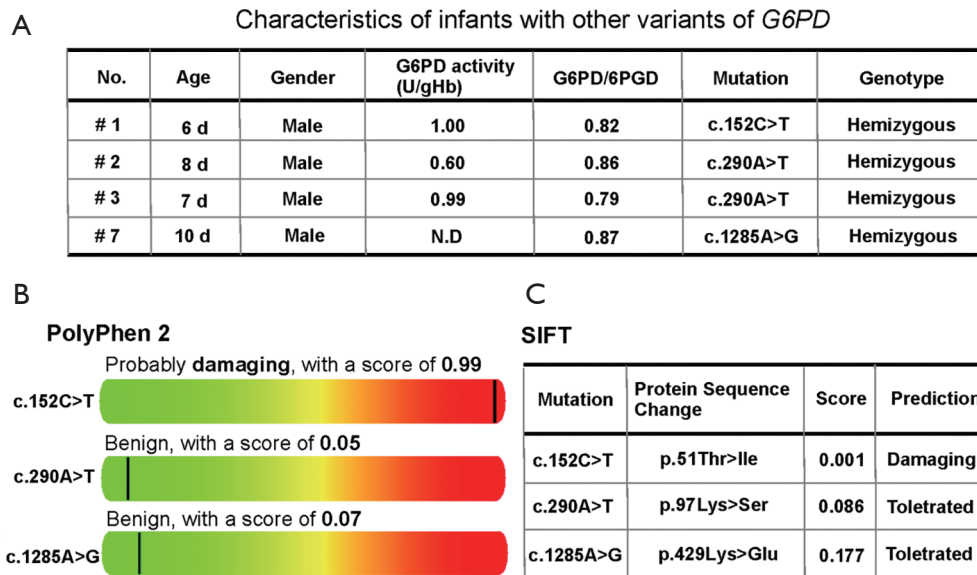
- gene from 2013 to 2017. *Hum Mutat* 2020;41:212-21.
7. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;157:1262-78.
  8. Jiang W, Bikard D, Cox D, et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013;31:233-9.
  9. Rath A, Mishra A, Ferreira VD, et al. Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Hum Mutat* 2019;40:2044-56.
  10. Zhou X, Wang L, Du Y, et al. Efficient Generation of Gene-Modified Pigs Harboring Precise Orthologous Human Mutation via CRISPR/Cas9-Induced Homology-Directed Repair in Zygotes. *Hum Mutat* 2016;37:110-8.
  11. Zhang Q, Wang J, Deng F, et al. TqPCR: A Touchdown qPCR Assay with Significantly Improved Detection Sensitivity and Amplification Efficiency of SYBR Green qPCR. *PLoS One* 2015;10:e0132666.
  12. Shu Y, Wang Y, Lv WQ, et al. ARRB1-Promoted NOTCH1 Degradation Is Suppressed by OncomiR miR-223 in T-cell Acute Lymphoblastic Leukemia. *Cancer Res* 2020;80:988-98.
  13. Zhang JP, Li XL, Li GH, et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol* 2017;18:35.
  14. Shu Y, Wu K, Zeng Z, et al. A Simplified System to Express Circularized Inhibitors of miRNA for Stable and Potent Suppression of miRNA Functions. *Mol Ther Nucleic Acids* 2018;13:556-67.
  15. Delgado-Cañedo A, Santos DG, Chies JA, et al. Optimization of an electroporation protocol using the K562 cell line as a model: role of cell cycle phase and cytoplasmic DNAses. *Cytotechnology* 2006;51:141-8.
  16. Gómez-Manzo S, Terrón-Hernández J, de la Mora-de la Mora I, et al. Cloning, expression, purification and characterization of his-tagged human glucose-6-phosphate dehydrogenase: a simplified method for protein yield. *Protein J* 2013;32:585-92.
  17. Gómez-Manzo S, Terrón-Hernández J, De la Mora-De la Mora I, et al. The stability of G6PD is affected by mutations with different clinical phenotypes. *Int J Mol Sci* 2014;15:21179-201.
  18. Howes RE, Battle KE, Satyagraha AW, et al. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Adv Parasitol* 2013;81:133-201.
  19. Kanjee U, Grüning C, Chaand M, et al. CRISPR/Cas9 knockouts reveal genetic interaction between strain-transcendent erythrocyte determinants of *Plasmodium falciparum* invasion. *Proc Natl Acad Sci U S A* 2017;114:E9356-65.
  20. Au SW, Gover S, Lam VM, et al. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure* 2000;8:293-303.
  21. Kotaka M, Gover S, Vandeputte-Rutten L, et al. Structural studies of glucose-6-phosphate and NADP+ binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2005;61:495-504.
  22. Vulliamy TJ, D'Urso M, Battistuzzi G, et al. Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proc Natl Acad Sci U S A* 1988;85:5171-75.
  23. Minucci A, Giardina B, Zuppi C, et al. Glucose-6-phosphate dehydrogenase laboratory assay: How, when, and why?. *IUBMB Life* 2009;61:27-34.
  24. Choi H, Andersen JP, Molday RS. Expression and functional characterization of missense mutations in ATP8A2 linked to severe neurological disorders. *Hum Mutat* 2019;40:2353-64.
  25. Quartier A, Courraud J, Thi Ha T, et al. Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Hum Mutat* 2019;40:2021-32.
  26. Rodríguez-Jiménez C, Pernía O, Mostaza J, et al. Functional analysis of new variants at the low-density lipoprotein receptor associated with familial hypercholesterolemia. *Hum Mutat* 2019;40:1181-90.
  27. Mansour MR, Abraham BJ, Anders L, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 2014;346:1373-77.
  28. Wright JB, Sanjana NE. CRISPR Screens to Discover Functional Noncoding Elements. *Trends Genet* 2016;32:526-9.
  29. Elf S, Lin R, Xia S, et al. Targeting 6-phosphogluconate dehydrogenase in the oxidative PPP sensitizes leukemia cells to antimalarial agent dihydroartemisinin. *Oncogene* 2017;36:254-62.
  30. Gregory MA, D'Alessandro A, Alvarez-Calderon F, et al. ATM/G6PD-driven redox metabolism promotes FLT3 inhibitor resistance in acute myeloid leukemia. *Proc Natl Acad Sci U S A* 2016;113:E6669-78.
  31. Poulain L, Sujobert P, Zylbersztejn F, et al. High mTORC1 activity drives glycolysis addiction and

- sensitivity to G6PD inhibition in acute myeloid leukemia cells. *Leukemia* 2017;31:2326-35.
32. Cunningham AD, Colavin A, Huang KC, et al. Coupling

between Protein Stability and Catalytic Activity Determines Pathogenicity of G6PD Variants. *Cell Rep* 2017;18:2592-9.

**Cite this article as:** Zhang H, Peng D, Shu Y, Zhu D, Hu W, Yu C, Zhang J, Liu S, Wan K, Yuan Z, Liu H, Wang D, Jiang T, Yu J, Zhang P, Zou L. Integrative identification of the pathogenic role of a novel G6PD missense mutation c.697G>C. *Ann Transl Med* 2021;9(3):194. doi: 10.21037/atm-20-3941





**Figure S1** Characteristics of three novel *G6PD* variants. (A) Clinical feature of patients with the other three novel *G6PD* variants. For the *G6PD* activity, the cut-off values were 2.6 (U/gHb) for males and 3.3 (U/gHb) for females. For the *G6P/6PG* ratio, the cut-off value for healthy people was 1.0–1.6. The predictive role of the c.697G>C variant was determined using the PolyPhen-2 (B) and SIFT (C) databases. Regarding the data from the SIFT database, the nonpolar, uncharged polar, basic, and acidic amino acids are marked in black, green, red, and blue, respectively.

**Table 1** List of Oligonucleotide & Antibody

Primer	Sequence	Use
G6PD-QF	GCCCGAAAACACCTTCATCG	Real Time PCR
G6PD-QR	GCAAGGCCAGGTAGAAGAGG	
GAPDH-QF	CAGCGACACCCACTCCTCCACCTT	
GAPDH-QR	CATGAGGTCCACCACCCTGTTGCT	
697-F	CTCTGATCCTCACTCCCCGA	Amplify the Genome Segment of c.697G>C both for T7E1 Assay and DNA Sequencing
697-R	TCATAAAACCGTGGGGTGCT	
Sg-1 F	cacc GTAACGCAGGCGATGTTGTCC	SgRNAs for <i>G6PD</i> Gene c.697G>C Site
Sg-1 R	aaac CGGACAACATCGCCTGCGTTA	
Sg-2 F	cacc CTTGAAGGTGAGGATAACGC	
Sg-2 R	aaac GCGTTATCCTCACCTTCAAG	
Sg-3 F	cacc CTGCGTTATCCTCACCTTCA	
Sg-3 R	aaac TGAAGGTGAGGATAACGCAG	
c.697G>C Donor	CGTGAGGACCAGATCTACCGCATCGACACTACCTGGGCAAGGA GATGGTGCAGAACCTCATGGTGCTGAGGTGGGGCCAAGCCTGG GCCGGGGACCAGGTGGGGTGGTACTCAGGAGCCTCACCT GGCCACTGCCTCCCGAGGACGAATTCCTCCAGAACTCAGAC AAGGGTGACCCCTCACATGTGGCCCTGCACCACAGAGGCCA AGGTCAGTTCCTCCACCTTGCCCTCCCTGCAGATTTGCCAACA GGATCTTCGGCCCCATCTGGAACCGGGACAACATCGCCTGCCTT ATCCTCACCTTCAAGGAGCCCTTTGGCACTGAGGGTCGCGGGG GCTATTTCCGATGAATTTGGGATCATCCGGTGAGAGCTTCTCTC TCCTGGGAGGCTGGCACAGGGTGGCAGAGCCAGTCACCCTGCA GGGCTACTCTCCCTATCTTGGGGAGCTCCTCCTCACCTGCA GTTCAAACCTAAGTGTCTGAGCTATCAGACCGGGCTGGAAAGG GCTGGACCCCTACACAGCCAAGCACCCACCGTTTTATGATTCA GTGATAGCATCACCATGTCTTCTCTGATTTAAG	Homology Arm for CRISPR/Cas9
Antibody	Source	Cat.No.
G6PD (D5D2) Rabbit mAb	Cell Signaling Technology	#12263
HRP-conjugated Affinipure Goat Anti-Rabbit IgG(H+L)	Proteintech	SA00001-2