Universal probe-based intermediate primer-triggered qPCR (UPIP-qPCR) for SNP genotyping

Baowei Li[1], Xiaodan Hao[1], Yanran Liu[1], Wei Wu[2], Ying Liu[1], Jianxun Wang[1], Yin Wang[1], and Peifeng Li[1]

[1]Qingdao University
[2]Qingdao Agricultural University

September 11, 2020

Abstract

The detection and identification of single nucleotide polymorphism (SNP) is an important basis for the evaluation of individual medicine and the judgement of disease susceptibility. At present, SNP genotyping technology includes the Sanger sequencing, TaqMan probe quantitative polymerase chain reaction (qPCR), amplification-refractory mutation system (ARMS)-PCR, Kompetitive allele specific PCR (KASP), and next-generation sequencing (NGS), etc. However, there are some disadvantages such as high cost of development and detection, long detection period and easily occurring false-positive results to these technologies. Focusing on these limitations, we proposed a new SNP detection method named as universal probe-based intermediate primer-triggered qPCR (UPIP-qPCR). In this method, only two types of fluorescence-labeled probes were used for all SNP genotyping, thus, greatly reducing the cost of development and detection for SNP genotyping. In the amplification process, unlabeled intermediate primers with template specific recognition function were able to trigger probe hydrolysis and specific signal release. The sensitivity of UPIP-qPCR in SNP genotyping was 0.01 ng, the call rate was more than 99.1%, and the accuracy was more than 99.9%. This novel technology gave rise to the low cost and high accuracy of SNP genotyping, and provided a new and reliable SNP genotyping method for the development of precision medicine.

Universal probe-based intermediate primer-triggered qPCR (UPIP-qPCR) for SNP genotyping

Baowei Li[*1], Xiaodan Hao[1], Yanran Liu[2], Wei Wu[3], Ying Liu[1], Jianxun Wang[2], Yin Wang[1], and Peifeng Li[*1]

[1] Institute for Translational Medicine, Qingdao University, Qingdao, 266021 China

[2] School of Basic Medicine, Qingdao University, Qingdao, 266021 China

[3] College of Food Science and Engineering, Qingdao Agricultural University, Qingdao, 266109 China

[*] Correspondence: Baowei Li: lbwwbl@126.com; Peifeng Li: peifli@qdu.edu.cn; Tel.: +86-0532-8299-1791

Grant numbers

This work was funded by the National Natural Science Foundation of China (31430041), the Major Research Program of the National Natural Science Foundation of China (91849209) and China Postdoctoral Science Foundation (2018M632613)

Abstract

The detection and identification of single nucleotide polymorphism (SNP) is an important basis for the evaluation of individual medicine and the judgement of disease susceptibility. At present, SNP genotyping technology includes the Sanger sequencing, TaqMan probe quantitative polymerase chain reaction (qPCR), amplification-refractory mutation system (ARMS)-PCR, Kompetitive allele specific PCR (KASP), and next-generation sequencing (NGS), etc. However, there are some disadvantages such as high cost of development and detection, long detection period and easily occurring false-positive results to these technologies. Focusing on these limitations, we proposed a new SNP detection method named as universal probe-based intermediate primer-triggered qPCR (UPIP-qPCR). In this method, only two types of fluorescence-labeled probes were used for all SNP genotyping, thus, greatly reducing the cost of development and detection for SNP genotyping. In the amplification process, unlabeled intermediate primers with template specific recognition function were able to trigger probe hydrolysis and specific signal release. The sensitivity of UPIP-qPCR in SNP genotyping was 0.01 ng, the call rate was more than 99.1%, and the accuracy was more than 99.9%. This novel technology gave rise to the low cost and high accuracy of SNP genotyping, and provided a new and reliable SNP genotyping method for the development of precision medicine.

Key words

UPIP-qPCR, Intermediate primer, Universal probe, SNP genotyping, Sanger sequencing

Introduction

Single-nucleotide polymorphism (SNP) genotyping is an important basis for individual medicine and disease susceptibility (Ho, Schierding, Wake, Saffery, & O'Sullivan, 2019; Li et al., 2020; Scheen, 2016; Tanner & Tyndale, 2017). Currently, SNP genotyping technology mainly includes first-generation sequencing, TaqMan probe quantitative polymerase chain reaction (qPCR), amplification-refractory mutation system (ARMS)-PCR, Kompetitive allele specific PCR (KASP), nucleic acid mass spectrometry, and next-generation sequencing (NGS), etc (Matsuda, 2017; Nielsen, Paul, Albrechtsen, & Song, 2011; J. Perkel, 2008; You, Yang, Peng, Xu, & Wang, 2018). However, these technologies have limitations such as high cost of development and detection, long testing duration, and false positive tendencies, etc (Kim & Misra, 2007; J. Perkel, 2008).

The Sanger sequencing, is a DNA sequencing technology established by Sanger et al. in 1977 (Sanger, Nicklen, & Coulson, 1977). This method possesses high accuracy and is the gold standard for SNP detection (Wu, Wu, & Liang, 2017). Due to its tedious operation, expensive reagents and equipment, long reporting period, and low throughput, the clinical application of Sanger sequencing is greatly reduced. TaqMan probe-based qPCR is the most widely used SNP detection technology in clinical diagnostics (Chu et al., 2017; Dooms, Chango, & Abdel-Nour, 2014; Lee, Connell, & Bloch, 1993; Sanjuan-Jimenez, Colmenero, & Morata, 2017; L. Zhang et al., 2013). This technology has advantages such as low detection cost, short of detection period, and high accuracy. However, the high cost of probe synthesis and the complex process of probe optimization lead to the high cost and long period of detection kit development, which limited its application in large-scale SNPs genotyping.

ARMS-PCR is an amplification-refractory mutation system (Newton, Graham, et al., 1989; Newton, Heptinstall, et al., 1989). It uses two-tube reactions with different 3'-end-specific allele primers, when exceeding a certain number of cycles, non-complementary allele primers with initially low amplification efficiency would obtain products accumulation, causing false-positive results thus greatly limiting the application of this method in clinical detections (Ferrie et al., 1992). Mass spectrometry was used to detect SNPs for the first time in 1997 (Braun, Little, & Koster, 1997). This technology has a high accuracy, but the miss rate is relatively high, and the instrument used is very expensive, making it difficult for clinical promotion and application.

The qPCR technology that utilizes a universal template probe was first invented by Zhang et al. in 2003 (Y.

Zhang et al., 2003), and the KASP technology based on single labelled universal primers was invented by Chunlin et al in 2014 (He, Holme, & Anthony, 2014).These two technologies lack specific-sequence primers similar to the TaqMan probe and usually result in false-positive results, which make them difficult to be used for clinical examinations (Ertiro et al., 2015).

The NGS technology is characterized by high-throughput sequencing, and can detect millions of DNA sequences simultaneously. It has advantages in massive unknow gene mutation screenings, such as tumor gene screening. However, due to the long reaction period, high cost of reagents and expensive machines, the NGS technology is not suitable for genotyping small numbers of known SNPs (Cirulli et al., 2015; J. M. Perkel, 2017). Similarly, a highly integrated gene chip-based system can detect nearly one million SNPs at one time. However, this type of gene chip system is expensive and has a long detection period, which is not suitable for the needs of clinical individualized SNP genotyping (J. Perkel, 2008; Ragoussis, 2006).

In view of the limitations of above technologies, we established a new and improved SNP detection method, named universal-probe intermediate primer-triggered qPCR (UPIP-qPCR). Generally, in the development of any type of SNP genotyping kits, two types of fluorescence probes are used, and intermediate primers are introduced to guarantee specificity. Due to the lack of fluorophore and quencher labeling in intermediate primers, the cost and duration of research and development for genotyping kits have been greatly reduced, giving rise to the need for low cost and high accuracy of SNP genotyping.

Materials and Methods

1. PCR and Sanger Sequencing

The nucleotide sequence containing specific SNPs were acquired from the dbSNP database of National Center for Biotechnology Information (NCBI, Bethesda, MD, USA) (https://www.ncbi.nlm.nih.gov/snp/). The primers flanking the SNP site (Fla-primers) and sequencing primers (Seq-primers) were designed using Primer Premier version 5 software. Sequences of Fla-primers and Seq-primers were listed in Table S1. Human genomic DNAs were extracted from whole blood samples via Rapid Blood Genomic DNA Isolation Kit according to the manufacturer's instructions (Sangon Biotech, Shanghai, China). Polymerase chain reactions (PCR) were performed to amplify 151–921 bp products containing targeted SNPs. For each SNP, the total PCR volume was 40 μL, containing 1× EmeraldAmp PCR Master Mix (TaKaRa, Shiga, Japan), genomic DNA (80 ng) and Fla-primers-F & R (500 nM/each). The PCR was programmed as follows: initial denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 95 °C for 25 s, annealing at 62 °C for 30 s, elongation at 72 °C for 1 min, and final elongation at 72 °C for 5 min. PCR products were purified by gel extraction and used as templates in the dideoxy chain-termination PCR system using BigDye Terminator v3.1 Kit (Applied Biosystems, Foster City, CA, USA), the nucleotide sequences of the PCR products containing specific SNPs were obtained by the 3730xl DNA analyzer (Applied Biosystems).

2. UPIP-qPCR

UPIP-qPCR consisted of two separated reactions. The first stage (stage I) of UPIP-qPCR was a general PCR. For each SNP, the total stage I PCR volume was 20 μL, containing genomic DNA (20 ng), upstream site-specific primers and downstream specific primer (100 nM/each), ddH_2O, dNTP, PCR buffer and hot-start DNA polymerase according the manufacturer's instructions (TaKaRa Taq Hot Start Version). In this stage, the reaction conditions were: initial denaturation at 95 °C for 3 min, followed by 10–22 cycles of denaturation at 98 °C for 5 s, annealing at 67 °C for 25 s, elongation at 72 °C for 20 s, and final elongation at 72 °C for 1 min. PCR products were diluted 10 times with ddH_2O and used as templates in the second stage reaction. The stage I PCR was carried out using the T100^{TM} Thermal Cycler PCR machine (Bio-Rad, Hercules, CA, USA).

The second stage (stage II) of UPIP-qPCR was a qPCR. For each SNP, the total stage II PCR volume was 20 μL, containing template DNA (2 μL), intermediate primer (500 nM), upstream universal primer (500 nM) and universal probes (300 nM), and ddH_2O, dNTP, PCR buffer and hot-start DNA polymerase according

the manufacturer's instructions (TaKaRa Taq Hot Start Version). In this stage, the reaction conditions were: initial denaturation at 95 °C for 3 min, followed by 35–40 cycles of denaturation at 98 °C for 5 s, annealing at 49 °C for 25 s, elongation at 72 °C for 1 s and fluorescent signals were obtained via plate reading. The stage II qPCR was carried out on the CFX96 Touch qPCR machine (Bio-Rad). Sequences of primers for UPIP-qPCR were listed in Table 1. Sequences of FAM- and HEX-labeled universal probes and universal primer were listed in Table 2.

3. Optimization of UPIP-qPCR

In the optimization process of UPIP-qPCR, the upstream site-specific primers and downstream specific primer concentration in the stage I reaction system were set up in three gradients of 100 nM, 200 nM and 500 nM, stage I PCR reactions were set up in four different cycles of 10, 14, 18 and 22, and the PCR products were set up to four different dilutions of 0, 5, 10 and 20 times prior to use as templates in stage II reactions. The categories of reagent components in the stage I reaction system were not changed and is described in "2. UPIP-qPCR" of the methods. There were no changes in the primer and probe concentrations, reagent components, and reaction conditions in this optimization process, as described in "2. UPIP-qPCR".

4. Range Setting of Intermediate Primer's Position

In the second stage of UPIP-qPCR, under the fixed amplification condition of extension at 72 for 1 s, six intermediate primers at different positions were set for the ALDH2 rs671 site, which were at a distance of 0, 30, 60, 90, 120, 150 bases from the 3'-end of intermediate primers to the SNP site. This was done to detect the effects due to the position of the intermediate primers on the generation of amplification curves and typical signals. The primer sequences are listed in Table S2. In this experiment, only the intermediate primers added in the stage II reaction system were different, and there were no changes in the other factors such as the reaction volume, reagent composition and reaction conditions, as described in "2. UPIP-qPCR" of the methods.

5. Sensitivity Analysis

The genomic DNA samples of the three genotypes of ALDH2 rs671 were used for UPIP-qPCR sensitivity analysis, and ddH_2O was used NTC. In the first stage of UPIP-qPCR, the concentration gradients of genomic DNA were 100 ng, 10 ng, 1 ng, 0.1 ng, 0.01 ng, 0.003 ng and 0.001 ng per 10 μL reaction system. In the first stage reaction system, the primer concentration was 100 nM/each and there were 18 amplification cycles. Other reaction parameters and conditions were identical to the stage I reactions described in "2. UPIP-qPCR" of the methods. There were 40 amplification cycles in the second stage and the other reaction parameters and conditions were identical to the stage II reactions described in "2. UPIP-qPCR". TaqMan probe-qPCR was set as control method for the sensitivity analysis, the concentration gradients of genomic DNA were the same as UPIP-qPCR. The total volume of TaqMan probe-qPCR was 10μL, containing forward and reverse primers (500nM/each), probes (400nM/each), genomic DNA (1μL) and ddH_2O, dNTP, PCR buffer and hot-start DNA polymerase according the manufacturer's instructions (TaKaRa Taq Hot Start Version). The TaqMan probe-qPCR reaction conditions were: initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturizing at 98 °C for 5 s, annealing at 62 °C for 25 s, elongating at 72 °C for 30 s and followed by plate reading. Sequences of primers and probes for TaqMan probe-qPCR were listed in Table S3.

6. Analysis of Call Rate and Accuracy for UPIP-qPCR

UPIP-qPCR was used to detect the genotypes of 224 human genomic DNA samples at rs671 (G > A), rs1057910 (A > C) and rs9923231 (C > T) loci. The reaction system and detailed thermal cycle parameters can be found in "2. UPIP-qPCR" of the methods. In each experiment, DNA samples of three known genotypes with specific SNP loci were used as controls, while ddH_2O was used as NTC. Reactions for each DNA sample of each SNP were repeated for three tubes, and only two or more tubes had the same fluorescence signals,

then the result would be considered effective. The number of effective results with the total number of samples were compared to obtain the call rate of the UPIP-qPCR method, and the genotyping results were compared with the Sanger sequencing results of same samples to obtain the accuracy rate of UPIP-qPCR method. The UPIP-qPCR genotyping primers and Sanger sequencing primers for rs671 (G > A), rs1057910 (A > C) and rs9923231 (C > T) are shown in Table 1 and Table S1, respectively. TaqMan probe-qPCR was set as control method for the analysis of call rate and accuracy. The concentration of genomic DNA was 10ng/μL in TaqMan probe-qPCR system. The reagent components and reaction conditions of TaqMan-qPCR were the same as that described in "5. Sensitivity analysis". Sequences of primers and probes for TaqMan probe-qPCR were listed in Table S3.

7. Wide Applicability Analysis of UPIP-qPCR in SNP Genotyping

Based on UPIP-qPCR, we designed primers (Table 1) and genotyped 13 different SNPs, including rs10234411 (A > T), rs4961 (G > T), rs1801253 (G > C), rs1801131 (A > C), rs1801133 (C > T), rs394 (A > G), rs1045642 (T > C), rs3918290 (G > A), rs55886062 (A > C), rs1695 (A > G), rs25487 (A > G), rs35305379 (TTTA > TTTTA) and rs34481414 (ACTACAAT > ACAAT). These SNPs covered all SNP mutation types. The system volumes of the first and second stage UPIP-qPCR reactions were all 10 μL, and the cycle numbers of the first and second stage reactions were 10 and 40, respectively. For the reagent composition of the system, please refer to "2. UP- qPCR" of the methods. The reaction conditions of stage I were: initial denaturation at 95 °C for 3 min, followed by 10 cycles of denaturation at 95 °C for 20 s, annealing at 67 °C for 30 s (touchdown -0.4 °C/cycle), elongation at 72 °C for 30 s, and followed by 10 cycles of denaturation at 95 °C for 20 s, annealing at 63 °C for 30 s, elongation at 72 °C for 30 s, and final elongation at 72 °C for 2 min. Please refer to "2. UPIP-qPCR" for the reaction conditions of stage II.

8. Preparation of DNA Standards Containing Specific SNPs

Positive control samples were artificially constructed for all genotypes of all SNPs involved in this study. SNP loci were designed in the primers of positive control. As an internal reference, primers of human genome GAPDH was also designed to construct the positive control, however, the product was only wild-type. Primer sequences are shown in Table S4. For each genotype of an SNP, the total PCR volume was 50 μL, containing 1× EmeraldAmp PCR Master Mix (TaKaRa), genomic DNA (100 ng) and Control-wtF and/or Control-mutF & Control-R (500 nM /each). The PCR was programmed as follows: initial denaturation at 95 °C for 5 min, followed by 36 cycles of denaturation at 95 °C for 25 s, annealing at 65 °C for 30 s, elongation at 72 °C for 1 min, and final elongation at 72 °C for 5 min. PCR products were purified by gel extraction and re-dissolved in 50 μL ddH_2O. DNA concentration was determined by Nanodrop 2000 (Thermo Fisher Scientific, USA) and diluted to 0.1 pg/μL as a working solution stored at 4 . The equal volume mixture of wild-type positive control working solution of each SNP and the wild-type products of GAPDH formed the wild-type positive control substance of the multiplex PCR. Similarly, the mutant and heterozygous positive control substance of multiplex PCR were also obtained by this method.

9. Multiplex UPIP-qPCR

In the stage I reactions, 16 types of SNPs genotyping primers and GAPDH primers were mixed in proportion to form 17 multi-primer working solutions. Although the concentration of upstream and downstream primers of a specific SNP was equal, the final concentration in the reaction system of each SNP were not equal. rs671, rs1057910, rs9923231, rs10234411, rs4961, rs1801131, rs1801133, rs394, rs3918290, rs55886062 and rs1695 primers had a final concentration of 12.96 nM; rs1045642, rs25487, rs35305379 and rs34481414 primers had a final concentration of 38.87 nM; and rs1801253 primer had a final concentration of 77.74nM and the GAPDH primer had a final concentration of 4.28 nM. In order for effective amplification, the ratio of these four concentrations was 3:9:18:1. Seventeen multi-primer working solutions were prepared by mixing 2 μM original primers of each SNP according to the volume ratio of the above proportion. The total volume of multiplex PCR system was 20 μL, including 10 μL 2× buffer, 0.4 μL multiplex DNA polymerase (Vazyme, PM101), 2μL DNA (10 ng/μL), 7.6 μL of 17 multi-primer working solution. The reaction conditions of

multiplex PCR were: initial denaturation at 95°C for 5 min, followed by 10 cycles of denaturation at 95 °C for 20 s, annealing at 67 °C for 30 s (touchdown -0.4 °C/cycle), elongation at 72 °C for 30 s, and followed by 10 cycles of denaturation at 95 °C for 20 s, annealing at 63 °C for 30 s, elongation at 72 °C for 30 s, and final elongation at 72 °C for 5 min. The first stage products were digested by exonuclease I (Exo I, Takara) to remove redundant primers prior to use as templates for the second stage reactions. The Exo I reaction system consisted of 2 μL products of stage I, 2 μL 10× buffer, 1 μL Exo I (5 U/μL), 15 μL ddH_2O. The Exo I reaction conditions were 37 for 30 min and 85 for 15 min. The second stage were single reactions, please refer to "2. UPIP-qPCR" of methods for the detailed reaction conditions. The template used in the first stage reactions was human genome DNA sample No. 1, ddH_2O was used as NTC. Correctness was defined by comparing the experimental results with that of the Sanger sequencing.

10. DAN Microarray Assay

The microarrays integrated 20 types of probes with five duplicates, these probes included intermediate primers of 16 types of SNPs, complementary sequences of FAM and HEX single-labeled primers as positive reference, intermediate primers of GAPDH as internal reference, and amino modified 14-poly deoxythymine (NH_2-dT_{14}) as negative control probe. The NH_2-dT_{14} was added to each 5'-terminal of all these primers prior to use. Probes were injected on an aldehyde modified glass slide with a concentration of 10 μM/each. See Table 5 for sequences of these probes. The microarrays were stored at 4 or directly used after fixation overnight at room temperature (Microarrays were constructed by Qingdao OE Biotech).

The preparation process of PCR products for hybridization was divided into three stages. The first stage was a multiplex PCR. Please see "9. Multiplex UPIP-qPCR" of methods for the detailed reaction system and parameters. A universal reverse connector, 5'-TGGGAGCTGAGGGCGA-3', was added to the 5'-terminal of the downstream specific primer of each SNP, and the primer with the same sequence was called the universal reverse primer. In the first stage, normal human genome DNA (No. 1 & No. 2) were used as the template DNA of the experimental group, and the template DNA of the control groups was divided into three types, namely, the mixture of wild type, heterozygous and mutant positive control DNA of 16 SNP sites, and the wild type positive control DNA of GAPDH, respectively. ddH_2O was used as the blank control template.

The second stage was a product treatment process, i.e. using Exo I to digest the products of the first stage which would be used as templates in the third stage reactions. Please refer to "9. Multiplex UPIP-qPCR" of methods for the detailed reaction system and parameters of this digestion process. The third stage was the fluorescence labeling PCR reaction. FAM and HEX single-labeled primers and universal reverse primer (Table 2) were combined to amplify the templates processed by Exo I to obtain sufficient DNA fragments for microarray hybridization. The total stage III PCR volume was 30 μL, containing 4 μL template DNA, 667 nM FAM single-labeled primer, 667 nM HEX single-labeled primer, 667 nM universal reverse primer and ddH_2O, dNTP, PCR buffer and hot-start DNA polymerase according the manufacturer's instructions (TaKaRa Taq Hot Start Version). In this stage, the reaction conditions were: initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 25 s, annealing at 52 °C for 30 s, elongation at 72 °C for 20 s.

The microarrays were placed in a pre-hybridizing solution (25% formamide, 5× SSC, 0.1% SDS, 0.5% BSA) at 42 for 5 min in water bath. Products of the third stage were denatured at 95 for 5 min, cooled on ice for 2 min, then mixed with equal volume (30 μL) hybridizing buffer (50% formamide, 10× SSC, 0.2% SDS) to form hybridizing solution. 20 μL of hybridizing solution was aliquoted into the microarrays, covered with coverslips, and incubated at 40 for 16 h–20 h. Following hybridization, microarrays were transferred into cleaning solution A (0.1× SSC, 0.1% SDS) for 5 min at room temperature (18–25°C), then transferred into cleaning solution B (0.1× SSC) for 5 min, and dehydrated with 100% alcohol for 10 s. After drying at room temperature (18–25°C), the images of FAM and HEX signals were captured using a confocal microscope (Leica TCS SP8, Leica, Wetzlar, Germany). By merging the FAM and HEX signal images of the same field, the genotyping results of the samples could be obtained. Specifically, dots with only FAM signals (green) were homozygous wild type, those with only HEX signals (red) were homozygous mutant, and those with

both signals (yellow) were heterozygous.

11. Data Analyses

One-way ANOVA tests were used in data comparation between groups. Differences were statistically significant when the values of $P< 0.05$. Differences were statistically more significant when the values of $P < 0.01$. Data analyses were performed using SPSS 16.0 software.

Results

1. Principle of UPIP-qPCR Method

UPIP-qPCR is a universal probe-based and intermediate primer-triggered qPCR with two-stage reactions. The reactions of UPIP-qPCR are divided into two stages. The first stage consists of a general PCR reaction within 30 min, aiming to obtain a certain amount of DNA fragments containing specific SNP sites. In this stage, the primer set includes upstream site-specific primers and the downstream specific primer, and the reaction conditions are similar to general PCR (details in methods section). The second stage consists of a qPCR reaction within 60 min, which uses the PCR products of the first stage as templates to obtain the corresponding fluorescence signals of the alleles. In this stage, the primer set includes a intermediate primer, upstream universal primer and universal probes, and the reaction conditions are similar to that of general qPCR (details in methods section).

The differences between the reaction parameters of the two stages are the annealing temperature and the number of cycles. The first stage has a high annealing temperature of 67 , and generally carries out 10–22 cycles, and its products serve as the templates of the second stage. The second stage has a low annealing temperature of 49 , and generally carries out 35–40 cycles to complete the signal acquisition and SNP genotyping. The reaction process is shown in Figure 1A. Due to the two-stage amplification process in UPIP-qPCR, the intermediate primer used in the second stage is able to identify the correct DNA fragments derived from the first stage. This shows that the fluorescence signals triggered by the intermediate primer possesses a high specificity, which is key to the reliability of this technology. Despite the real-time detection of its fluorescence signals on a qPCR instrument, the second stage reactions can also be carried out on an ordinary PCR instrument. Following the completion of the reactions, end-point scanning can be performed by the qPCR instrument to achieve signal collection and genotyping, which can greatly improve the efficiency of qPCR instrument and indirectly reduce the cost of equipment. As a new genotyping method, its time and cost consuming are less than that of TaqMan probe-qPCR and Sanger sequencing, and its operation complexity is Moderate (Table 3).

The characteristics of the primers and probes involved in UPIP-qPCR are described as follows:

The upstream site-specific primer is divided into two parts: (1) the 5'-end, which is a universal connector composed of the same sequence of 14-base upstream universal primers and the reverse complementary sequence of 18-base universal probe, and (2) the 3'-end, which is the upstream specific primer sequence combining with the template DNA, and whose 3'-terminal base is designed according to the specific SNP genotype (Figure 1A). As most of the SNPs are dimorphic and only few are polymorphic, there are generally two upstream site-specific primers in a reaction system to distinguish the different base types.

The downstream specific primer combines to the template DNA to complete the first stage of the PCR reaction together with upstream site-specific primers, and provides the template for the second stage reaction (Figure 1A).

The universal primer is a fixed-sequence primer with –14 bases, and a sequence identical to the 5'-terminal segment in the upstream specific primer. The universal primer acts as the upstream primer, and cooperates with the intermediate primer to complete the reaction of the second stage of UPIP-qPCR (Figure 1A).

The intermediate primer, likes the downstream nested primer, is complementary to the template between the upstream site-specific primer and the downstream specific primer; the 5'-3' direction is the same as that of the downstream specific primer. The primer plays a role in targeting the correct template and triggering specific amplification and signal release in the second stage reaction. Intermediate primers are fundamental in the specificity of this technology (Figure 1A).

This universal probe, like the TaqMan probe, is designed based on the principle of FRET and is composed of a fluorophore at one end and a quencher at the other with its sequence reverse complementary to the middle part of the upstream site-specific primer. In the second stage, DNA polymerases containing 5'-3' exonuclease activity initiate the hydrolysis of universal probes and release fluorescence signals (Figure 1A). In this reaction system, two types of universal probes with different sequences are designed and tagged with different fluorescent labels to differentiate between two alleles signals of an SNP. In the reaction system of this study, to facilitate the genotyping of SNPs, FAM signals were set to represent wild-type alleles and HEX signals were set to represent mutant alleles.

2. Feasibility Verification of UPIP-qPCR in SNP Genotyping

According to the experimental design, we used *ALDH2* rs671 (G > A) and *CYP2E1* rs2031920 (C > T) loci as candidate SNPs to verify the feasibility of UPIP-qPCR and optimize this technology. First, we designed the Sanger sequencing related primers (Table S1) for rs671 and rs2031920, and obtained homozygous wild-, heterozygous, and homozygous mutant- types of human genome DNA positive standards (Figure S1). Then, a pair of upstream site-specific primers were designed, with G and A as the 3'-terminal bases, respectively (Table 1). The 5'-terminal of each upstream specific primer contained a universal connector (Table 2). Concomitantly, a downstream specific primer and a intermediate primer were designed (Table 1). FAM-labeled universal probes were used to display the signal of the G allele, and HEX-labeled universal probes were used to display the signal of the A allele (Table 2). The first stage of the reaction was carried out on an ordinary PCR instrument, with a total of 10 cycles, at an annealing temperature of 67 , to complete the initial amplification of DNA fragments containing G/A sites. In the second stage, the signals were collected in real-time as the reaction was carried out on a qPCR instrument with 40 cycles, at an annealing temperature of 49 . Following the reaction, the corresponding genotypes were identified according to the final fluorescence category and intensity (relative fluorescence units, RFU). The results showed a typical S-type amplification curve with exponential growth and that the amplification signals of the three positive DNA standards were specific. The amplification signals of the GG genotype were only FAM-positive, the GA genotype were both FAM- and HEX- positive, the AA genotype were only HEX-positive, and the no-template control group (NTC) had no false positive signals, showing the accuracy of the genotyping results (Figure 1B). These results show that UPIP-qPCR was feasible for SNPs genotyping.

Based on the feasibility of UPIP-qPCR, we optimized the primer concentration, amplification cycle number and template dilution ratio of the first stage reaction, to ensure the stability and reliability of UPIP-qPCR technology, and also to reduce the detection cost and duration as much as possible. The final concentration of PCR primers is generally 500nM/each. In the optimization process, we set up three gradients of 100 nM, 200 nM and 500 nM for the upstream and downstream specific primers, set at 10-, 14-, 18- and 22-different cycles for the PCR reaction, and set to 0-, 5-, 10- and 20-times dilutions for the PCR products, to identify the effects of these experimental factors on the curves and results of the second stage reaction. The results showed that only the products of the 100 nM/each primer concentration and 10 cycles amplification could be directly used as templates for the second stage without dilution (1 μL template/10 μL system). In all other conditions, the products should be diluted before used as templates (Figure S2A–S2D).

To further shorten the detection period, the duration of the elongation step at 72 in the second stage was set to 1s, and signal acquisition was directly conducted. A suitable distance, theoretically, the closer the better, between the 3'-end of the intermediate primer and the SNP site was required. Based on the genotyping experiment of *ALDH2* rs671 (G > A), we designed five intermediate primers with a base distance of 0, 30, 60, 90 and 120 from their 3'-terminal to the SNP site, and observed the effect of the different intermediate

primers on the amplification efficiency and genotyping results, based on which we formulated the optimal design principle of intermediate primers. The results showed that the intermediate primers with the base distance of 0, 30 and 60 had good amplification efficiency (Figure S3A) and acquired accurate genotyping results for all three different genotypic DNA samples (Figure S3B); however, the RFU value of reactions with a base distance of 0, and 30 was significantly higher than that of 60 bases ($p < 0.01$) (Figure S3C–S3E). Thus, to ensure the difference in the RFU values, the intermediate primers should be designed within a distance of 30 bases from its 3'-terminal to the SNP site.

3. UPIP-qPCR Presented High Sensitivity in SNP Genotyping Detections

The sensitivity of UPIP-qPCR was analyzed using a primer concentration of 100 nM/each and 18 cycles of amplification in the first stage reaction; human genomic DNA samples of three genotypes of rs671 with different concentrations were used. The results showed that most of the DNA samples with different concentrations presented typical S-type curves, and the appearance order of these curves were positively correlated with the decrease in concentration gradient (Figure 2A, 2C & 2E). The accuracies of all the concentrations of three genotypic genomic DNA were 100%, and although there were good call rates in the high concentration samples, it was not ideal in the low concentration samples (Figure 2B, 2D & 2F). Specifically, the call rates of all three genotypes were 100% at five concentration gradients from 100 ng/10 μL to 0.01 ng/10 μL, and the copy number gradients from 33,000 genomic DNA per 10 μL to 3 genomic DNA per 10 μL. The other call rates were: 100% for GG and AA samples with a concentration of 0.003 ng/10 μL, 72.22% for GA samples with a concentration of 0.003 ng/10 μL, 27.78%, 11.11% and 22.22% for GG, GA and AA samples with a concentration of 0.001 ng/10 μL, respectively (Figure 2B, 2D & 2F). The above data showed that the concentration of 0.01 ng/10 μL was the highest sensitivity of UPIP-qPCR, namely, every 10 μL reaction system containing three copies of genomic DNA can obtain reliable genotyping results.

TaqMan probe-qPCR was performed to comparing the sensitivity between these two methods. The results showed that DNA samples with high concentrations presented typical S-type curves, (Figure S4A, S4C & S4E). The accuracies in three genotypic genomic DNA with concentrations from 100 ng/10 μL to 0.01 ng/10 μL were 100% (Figure S4B, S4D & S4F). Call rates in three genotypic genomic DNA with concentrations from 100 ng/10 μL to 0.1 ng/10 μL were100%, but were <100% or even 0% in concentrations from 0.03 ng/10 μL to 0.01 ng/10 μL (Figure S4B, S4D & S4F). These data showed that the concentration of 0.1 ng/10 μL was the highest sensitivity of TaqMan probe-qPCR, namely, every 10 μL reaction system containing 33 copies of genomic DNA can obtain reliable genotyping results by TaqMan probe-qPCR method.

4. UPIP-qPCR Possessed High Call Rate and Accuracy

In this study, the genotypes of rs671, rs1057910 and rs9923231 in 224 DNA samples were detected by UPIP-qPCR, TaqMan probe-qPCR and Sanger sequencing. The UPIP-qPCR scatter plots showed good signal differentiation. Each scatter diagram contained three repeats of positive standards and NTC. The yellow dots represented the wild-type genotype, the green triangles represented the heterozygous genotype, and the blue squares represented the mutant genotype (Figure S5A). The TaqMan probe-qPCR scatter plots also showed good signal differentiation in genotyping detections of these three SNPs (Figure S5B). The call rates of rs671, rs1057910 and rs9923231 generated by UPIP-qPCR were 99.11%, 100% and 100% respectively, which were all higher than that of TaqMan probe-qPCR (Table 4). Compared with Sanger sequencing results, the accuracies of UPIP-qPCR and TaqMan probe-qPCR were all 100% in the detection of these three SNPs (Table 4).

5. UPIP-qPCR Has Wide Applicability of Identifying All Kinds of Variations in SNPs

The variation types of SNPs include point mutations and base insertion/deletion mutations (InDels). There are six types of nucleotide alterations in point mutations, and these include interchanges between A-G, A-C, A-T, G-C, G-T and C-T. InDels are defined as an increase or decrease in one or more nucleotides in the DNA sequence. In addition to the A-G, A-C and C-T mutation types involved in the above experiments,

we also selected SNPs of other mutation types to test the wide adaptability of UPIP-qPCR. These genes and SNP sites were, *ABCB1*rs10234411 (A > T), *ADD1* rs4961 (G > T), *ADRB1* rs1801253 (G > C), *MTHFR* rs1801131 (A > C), *MTHFR* rs1801133 (C > T),*MTRR* rs394 (A > G), *ABCB1* rs1045642 (T > C), *DPYD* rs3918290 (G > A),*DPYD* rs55886062 (A > C), *GSTP1* rs1695 (A > G), *XRCC1* rs25487 (A > G), *APC*rs35305379 (TTTA > TTTTA), *APC* rs34481414 (ACTACAAT > ACAAT). These loci are closely related to individual medicine, and included all six types of point mutations and InDels. We have obtained three genotypes of human genomic DNA standards mostly by Sanger sequencing. The positive DNA standards of several gene loci related to tumor drug metabolism, including rs1801253, rs1045642, rs3918290, rs55886062, rs1695, and rs25487 were constructed artificially, as the homozygous mutant DNA of these loci could not be found in the limited normal population. The results showed that UPIP-qPCR well identified these SNPs with specific amplification signals and typical S-type amplified curves (Figure 3A). Different genotypes were clustered in an obvious manner in the scatter diagram, and results of genotyping were accurate (Figure 3B). These results suggested that UPIP-qPCR can be widely used in the genotyping of different variations of SNPs.

To reduce the complexity of the reaction and the consumption of template DNA, we mixed the primers involved in the first stage reaction of 16 SNPs in one tube to perform a multiplex PCR. The product was diluted ten times and digested by Exo I enzyme before used as the template of the second stage reaction. The results showed that specific amplification signals and typical amplification curves were obtained in the second stage reaction from the DNA sample (DNA No. 1) (Figure S6A), NTC had no false positive signals (Figure S6B), and the genotyping results were consistent with that of Sanger sequencing (Table S5).

6. DNA Microarrays Based on Intermediate Primers Were Feasible for SNP Genotyping

As the intermediate primers are able to specifically recognize the template DNA, 20*5 dot DNA microarrays were made (Figure 4A) with 16 types of SNP intermediate primers, positive primers, negative primers, and internal reference (GAPDH) primers as probes, and the hybridization method was utilized to achieve SNP genotyping. Through multiplex PCR, and single-labeled fluorescent probes plus universal reverse primer-PCR, we prepared DNA templates for hybridization, and adopted an overnight hybridization method. The results showed that the positive reference DNA of the three genotypes could obtain specific and accurate hybridization signals. The hybridization signals of homozygous wild-type, heterozygous, and homozygous mutant DNA were FAM-positive and HEX-negative (green dots), FAM- and HEX-positive (yellow dots), FAM-negative and HEX-positive (red dots), respectively. (Figure 4B). There were no signals on the blank control microarray which used water as template, except for the FAM- and HEX-positive control dots (Figure 4B). FAM (green), HEX (red), FAM & HEX (yellow) and no signal (black background) were displayed respectively at the positions of FAM-positive, HEX-positive, GAPDH and negative reference probes fixed on each microarray, and was consistent with the expected results (Figure 4B). The hybridization signals of the two types of human genomic DNA (No. 1 & No. 2) were also specific, and consistent with Sanger sequencing results (Figure 4B & Table S6). Using this method, 16 SNPs can be genotyped for one DNA sample by one microarray at the same time, thus increasing the detection throughput.

Discussion

In the UPIP-qPCR, we introduced intermediate primers and universal probes to show signal specificity and reduced cost, which makes the method more suitable for clinical detections. UPIP-qPCR adopted a two-stage system requiring two reaction tubes. In the first stage, the product containing specific SNPs were generated and used as templates for the second stage qPCR, and genotyping results can be obtained following several thermal cycles and real-time signal acquisition in the second stage. In addition, we tried a single-tube reaction to obtain genotyping signals, however, the NTC often presented false-positive signals due to the interference of high annealing temperature primers in the first stage, making it difficult to obtain accurate genotyping results. Thus, it is necessary to separate the two stages of the reactions. These characteristics, especially the use of intermediate primers and the separated two-stage reactions, account for the differences

between UPIP-qPCR and KASP (Broccanello et al., 2018; He et al., 2014).

Intermediate primers were designed to ensure the accuracy of the signal. This was because these primers would only combine with the template DNA to generate specific signals, when the products of the first stage were of the correct DNA segments. In principle, the closer the 3'-end of the intermediate primer to the SNP locus, the higher the amplification efficiency, and so we suggest that the intermediate primers should be designed [?]30 bases distance from their 3'-end to the SNPs loci. In the multiplex UPIP-qPCR, due to the diversity of template DNA from the first stage reactions, it generated non-specific signals in the second stage of qPCR reactions. In this situation, the position of intermediate primers should be adjusted in order to remove the non-specific signals of the multiplex UPIP-qPCR.

Although two-stage reactions were required, the sensitivity of UPIP-qPCR can reach three copies per 10 μL reaction system, which could make up for this operational defect. The call rate, which needed to be based on the premise that the genomic DNA should have high purity with the OD_{260}/OD_{280} ratio between 1.75 and 2.0, could reach 100%; a ratio lower than 1.75, implies a missed detection. At this time, the DNA would require re-extracted for reliable results.

According to the experimental data, the first stage reaction of UPIP-qPCR can adopt either single PCR or multiplex PCR. In the clinical genotyping detection of a specific SNP or several SNP loci, the corresponding experimental method can be selected flexibly according to the actual requirements. A single PCR can be adopted when the number of SNPs detected is low and the amount of DNA available is sufficient, otherwise a multiplex PCR can be adopted when the number of SNPs detected is high and the amount of DNA available is limited. For the single PCR, a final concentration of primers in the first stage is suggested to be 100 nM/each with a thermal cycle range between 10–22, and the product can be used in the second stage following a 10 times dilution. For the multiplex PCR, we suggest a final primer concentration of 4 nM–40 nM/each in the first stage with 22 thermal cycles, and the products should be diluted 10 times and treated with Exo I prior to use in the second stage. At present, there has been only one type of intermediate primer used in each reaction tube in the second stage, that is, only one SNP can be genotyped in one tube reaction. In the future, with further studies, it is possible to introduce more types of fluorescence-labeled universal probes, so that two or more SNPs can be genotyped in a single-tube reaction of the second stage.

The microarray hybridization experiment with intermediate primers as probes provided a basis for the development of a new SNP screening method with higher throughput and high accuracy, which would further reduce the price of SNP genotyping. In addition, intermediate primers can also be fixed in multi-hole fluorescence quantitative microfluidic reaction plates, such as Thermofisher QuantStudio 12K. In this case, only the universal primer, universal probes and the stage I multiplex PCR products need be added to the reaction system, so that in addition to high-throughput SNP genotyping, high-throughput detection of copy number variation (CNV) can also be carried out, thus extending the application range of UPIP-qPCR. Although, we are looking forward to the development of a high-throughput real-time planar fluorescent qPCR and hybridization-sequencing technology based on intermediate primers, achieving accurate results and analysis would be a tough challenge.

In summary, the UPIP-qPCR developed in this study is a novel SNP genotyping technology with low cost, fast detection, and high accuracy, which will help to reduce the cost of clinical detection, reduce the burden of quizzes, and promote the development of precision medicine.

Acknowledgments

We would like to acknowledge Dr. Yin Wang and Prof. Peifeng Li for the revision of the manuscript and guidance for the development of UPIP-qPCR.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

Braun, A., Little, D. P., & Koster, H. (1997). Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem, 43* (7), 1151-1158. doi:10.1093/clinchem/43.7.1151

Broccanello, C., Chiodi, C., Funk, A., McGrath, J. M., Panella, L., & Stevanato, P. (2018). Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods, 14* , 28. doi:10.1186/s13007-018-0295-6

Chu, M. K. M., Wong, E. Y. L., Wong, S. C. C., Alikian, M., Gale, R. P., Apperley, J. F., & Foroni, L. (2017). Molecular techniques for the personalised management of patients with chronic myeloid leukaemia. *Expert Rev Mol Diagn, 11* , 4-20. doi:10.1016/j.bdq.2017.01.001

Cirulli, E. T., Lasseigne, B. N., Petrovski, S., Sapp, P. C., Dion, P. A., Leblond, C. S., . . . Goldstein, D. B. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science, 347* (6229), 1436-1441. doi:10.1126/science.aaa3650

Dooms, M., Chango, A., & Abdel-Nour, A. (2014). [Quantitative PCR (qPCR) and the guide to good practices MIQE: adapting and relevance in the clinical biology context]. *Ann Biol Clin (Paris), 72* (3), 265-269. doi:10.1684/abc.2014.0955

Ertiro, B. T., Ogugo, V., Worku, M., Das, B., Olsen, M., Labuschagne, M., & Semagn, K. (2015). Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genomics, 16* , 908. doi:10.1186/s12864-015-2180-2

Ferrie, R. M., Schwarz, M. J., Robertson, N. H., Vaudin, S., Super, M., Malone, G., & Little, S. (1992). Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *Am J Hum Genet, 51* (2), 251-262.

He, C., Holme, J., & Anthony, J. (2014). SNP genotyping: the KASP assay. *Methods Mol Biol, 1145* , 75-86. doi:10.1007/978-1-4939-0446-4_7

Ho, D. S. W., Schierding, W., Wake, M., Saffery, R., & O'Sullivan, J. (2019). Machine Learning SNP Based Prediction for Precision Medicine. *Front Genet, 10* , 267. doi:10.3389/fgene.2019.00267

Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng, 9* , 289-320. doi:10.1146/annurev.bioeng.9.060906.152037

Lee, L. G., Connell, C. R., & Bloch, W. (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res, 21* (16), 3761-3766. doi:10.1093/nar/21.16.3761

Li, B., Wang, X., Hao, X., Liu, Y., Wang, Y., Shan, C., . . . Li, P. (2020). A novel c.2179T>C mutation blocked the intracellular transport of PHEX protein and caused X-linked hypophosphatemic rickets in a Chinese family. *Mol Genet Genomic Med, n/a* (n/a), e1262. doi:10.1002/mgg3.1262

Matsuda, K. (2017). PCR-Based Detection Methods for Single-Nucleotide Polymorphism or Mutation: Real-Time PCR and Its Substantial Contribution Toward Technological Refinement. *Adv Clin Chem, 80* , 45-72. doi:10.1016/bs.acc.2016.11.002

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., . . . Markham, A. F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res, 17* (7), 2503-2516. doi:10.1093/nar/17.7.2503

Newton, C. R., Heptinstall, L. E., Summers, C., Super, M., Schwarz, M., Anwar, R., . . . Markham, A. F. (1989). Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet, 2* (8678-8679), 1481-1483. doi:10.1016/s0140-6736(89)92931-0

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet, 12* (6), 443-451. doi:10.1038/nrg2986

Perkel, J. (2008). SNP genotyping: six technologies that keyed a revolution. *Nat Methods, 5* , 447. doi:10.1038/nmeth0508-447

Perkel, J. M. (2017). Single-cell sequencing made simple. *Nature, 547* (7661), 125-126. doi:10.1038/547125a

Ragoussis, J. (2006). Genotyping technologies for all. *Drug Discov Today Technol, 3* (2), 115-122. doi:10.1016/j.ddtec.2006.06.013

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74* (12), 5463-5467. doi:10.1073/pnas.74.12.5463

Sanjuan-Jimenez, R., Colmenero, J. D., & Morata, P. (2017). Lessons learned with molecular methods targeting the BCSP-31 membrane protein for diagnosis of human brucellosis. *Clin Chim Acta, 469* , 1-9. doi:10.1016/j.cca.2017.03.014

Scheen, A. J. (2016). Precision medicine: The future in diabetes care?*Diabetes Res Clin Pract, 117* , 12-21. doi:10.1016/j.diabres.2016.04.033

Tanner, J. A., & Tyndale, R. F. (2017). Variation in CYP2A6 Activity and Personalized Medicine. *J Pers Med, 7* (4). doi:10.3390/jpm7040018

Wu, H., Wu, X., & Liang, Z. (2017). Impact of germline and somatic BRCA1/2 mutations: tumor spectrum and detection platforms. *Gene Ther, 24* (10), 601-609. doi:10.1038/gt.2017.73

You, Q., Yang, X., Peng, Z., Xu, L., & Wang, J. (2018). Development and Applications of a High Throughput Genotyping Tool for Polyploid Crops: Single Nucleotide Polymorphism (SNP) Array. *Front Plant Sci, 9* , 104. doi:10.3389/fpls.2018.00104

Zhang, L., Cui, G., Li, Z., Wang, H., Ding, H., & Wang, D. W. (2013). Comparison of high-resolution melting analysis, TaqMan Allelic discrimination assay, and sanger sequencing for Clopidogrel efficacy genotyping in routine molecular diagnostics. *J Mol Diagn, 15* (5), 600-606. doi:10.1016/j.jmoldx.2013.04.005

Zhang, Y., Zhang, D., Li, W., Chen, J., Peng, Y., & Cao, W. (2003). A novel real-time quantitative PCR method using attached universal template probe. *Nucleic Acids Res, 31* (20), e123. doi:10.1093/nar/gng123

Table 1 Primers for UPIP-qPCR

Gene	Rs. No.
ALDH2	rs671
CYP2E1	rs2031920
CYP2C9	rs1057910
VKORC1	rs9923231
ABCB1	rs10234411

Table 1 Primers for UPIP-

Table 1 Primers for UPIP-qPCR	**Table 1** Primers for UPIP-
ADD1	rs4961
ADRB1	rs1801253
MTHFR	rs1801131
MTHFR	rs1801133
MTRR	rs394
ABCB1	rs1045642
DPYD	rs3918290
DPYD	rs55886062
GSTP1	rs1695
XRCC1	rs25487
APC	rs35305379
APC	rs34481414
GAPDH	

Table 1 Primers for UPIP-qPCR

UPq: UPIP-qPCR; WT: wild type; Mut: mutant; Red colored letters are allele specific bases.

Table 1 Primers for UPIP-

UPq: UPIP-qPCR; WT: wi

Table 2 Communal Primers and Probes	**Table 2** Communal Primers and Probes
Name	Sequence (5'-3')
FAM universal connector	CTGTCCTCGGCACGCAGGGAAGGTGGTAGGTG
HEX universal connector	CTGTCCTCGGCACGGTGATTTGGTGGGAGGAG
FAM universal probe	FAM-CACCTACCACCTTCCCTG-BHQ1
HEX universal probe	HEX-CTCCTCCCACCAAATCAC-BHQ1
Universal primer	CTGTCCTCGGCACG
FAM-only probe	FAM-CAGGGAAGGTGGTAGGTG
HEX-only probe	HEX-AGTGATTTGGTGGGAGGAG
Universal reverse connector/primer	TGGGAGCTGAGGGCGA

Table 3 Characteristics of Three Methods for SNP Genotyping	**Table 3** Characteristics of Three Methods for SNP Genotyping	**Table 3** Characteristics of Three Methods for SNP Genotyping	**Table 3** Characteristics of Three Methods for SNP Genotyping
	UPIP-qPCR	TaqMan	Sanger
Reagent cost for kit development ($/SNP)	35-70	500-1000	10-20
Reagent cost for detection ($/SNP)	˜0.3	˜1	˜3
Detection period (hours/SNP)	1.5-2	1.5-2	12-24
Operation complexity	Moderate (Two stages)	Low (One stages)	High (Five stages)

Table 4 Call Rate and Accuracy of UPIP-qPCR and TaqMan Probe-qPCR	**Table 4** Call Rate and Accuracy of UPIP-q
SNPs	Call rate (n=224)
	UPIP-qPCR
rs671	99.11%
rs1057910	100%
rs9923231	100%

Table 5 Amino Modified Probes for Microarrays	**Table 5** Amino Modified Probes for Microarrays	**Table 5** Amino Mo
No.	Name	Rs. No.
1	FAM-Cont	
2	HEX-Cont	
3	ALDH2	rs671
4	CYP2C9	rs1057910
5	VKORC1	rs9923231

Table 5 Amino Modified Probes for Microarrays	**Table 5** Amino Modified Probes for Microarrays	**Table 5** Amino Mo
6	ABCB1	rs10234411
7	ADD1	rs4961
8	ADRB1	rs1801253
9	MTHFR	rs1801131
10	MTHFR	rs1801133
11	MTRR	rs394
12	ABCB1	rs1045642
13	DPYD	rs3918290
14	DPYD	rs55886062
15	GSTP1	rs1695
16	XRCC1	rs25487
17	APC	rs35305379
18	APC	rs34481414
19	GAPDH	
20	Negative-Cont	

Figure 1 Principle and application of UPIP-qPCR for SNP genotyping. The reactions of UPIP-qPCR are divided into two stages (A). The first stage is a general PCR reaction aiming to obtain a certain amount of DNA fragments containing specific SNP sites (A left panel). The second stage is a quantitative PCR (qPCR) reaction, which uses the PCR products of the first stage as templates to obtain the corresponding fluorescence signals of alleles (A right panel). UPIP-qPCR was designed for SNP genotyping, and initially used to identifying the SNP genotypes of ALDH2 rs671 and CYP2E1 rs2031920 (B). All three genotypes of each SNP samples were successfully yielded accurate results by UPIP-qPCR. The homozygous WT genotypes presented FAM-only signals (blue curves) and were shown in the lower right area of the genotyping scatter diagrams with yellow dots, the heterozygous genotypes presented FAM & HEX signals (blue curves & green curves) and were shown in the middle area of the genotyping scatter diagrams with green triangles, and the homozygous mutant genotypes presented HEX-only signals (green curves) and were shown in the upper left area of the genotyping scatter diagrams with blue squares. No-template control group (NTC) presented no signals and were shown in the lower left area of the genotyping scatter diagrams with black rhombuses. Each sample was detected with three duplicates in one experiment, and the experiments were repeated more than three times.

Figure 2 Sensitivity analysis of UPIP-qPCR. Three genotypic DNA samples of ALDH2 rs671 with concentrations from 100ng/10μL to 0.003ng/10μL showed typical S-type curves, and the appearance order of the curves were positively correlated with the decrease in concentration gradient (A, C & E). The call rates were 100% in all three genotypic samples with concentrations from 100ng/10μL to 0.01ng/10μL, but <100% in concentrations of 0.003ng/10μLand 0.001ng/10μL (B, D & F). The accuracies of all concentrations of three genotypic genomic DNA were all 100% (B, D & F). FAM and HEX signal curves of the same concentration in genotype GA were indicated by gray ellipse. Each sample was detected with nine duplicates in one reaction, and the experiments were repeated four times. Bars show *SD* ($n = 36$).

Figure 3 UPIP-qPCR possessed wide applicability of identifying all kinds of variations in SNPs. UPIP-qPCR presented specific amplification signals and typical S-type curves in the genotype detection of 13 different SNPs with all kinds of variations, including interchanges between A-G, A-C,A-T,G-C,G-T, C-T, and base insertion/deletion mutations (A). All three different genotypes of these SNPs were obviously clustered in the scatter diagrams, and results of genotyping were correct compared to Sanger sequencing(B). In each scatter diagram, WT-, heterogeneous-, and mutant-genotypes were represented by yellow dots, green triangles and blue squares, respectively. Each sample was detected with three duplicates at one experiment, and the experiments were repeated more than three times.

Figure 4 DNA microarrays for genotyping. DNA microarrays was made of 20 types of probes with 5

repeated dots, the order of probes was marked by circled numbers, and the SNPs and sequences related to these probes were listed in Table 3 with the same No., briefly, are FAM positive control probes, HEX positive controls, ˜ probes of 16 kinds of SNPs, internal control GAPDH probes, negative control NH_2-dT_{14} probes (A). The hybridization signals of the two types of human genomic DNA (No.1 & No.2) were both specific and correct (B). The positive control DNA was able to obtain specific and accurate hybridization signals. The hybridization signals of wild-type DNA were FAM-positive and HEX-negative (green dots), the hybridization signals of heterozygous DNA were FAM- and HEX- double positive (yellow dots), and the hybridization signals of mutant DNA were FAM-negative and HEX-positive (red dots) (B). There were no signals on the blank control microarray with water as template, except for FAM- and HEX- positive control probes (B). FAM (green), HEX (red), FAM & HEX (yellow) and no signal (black background) were displayed respectively at the positions of , , and (B). WT: wild type; Mut: mutant; Hetero: Heterozygous; Cont: control.

Figure 1

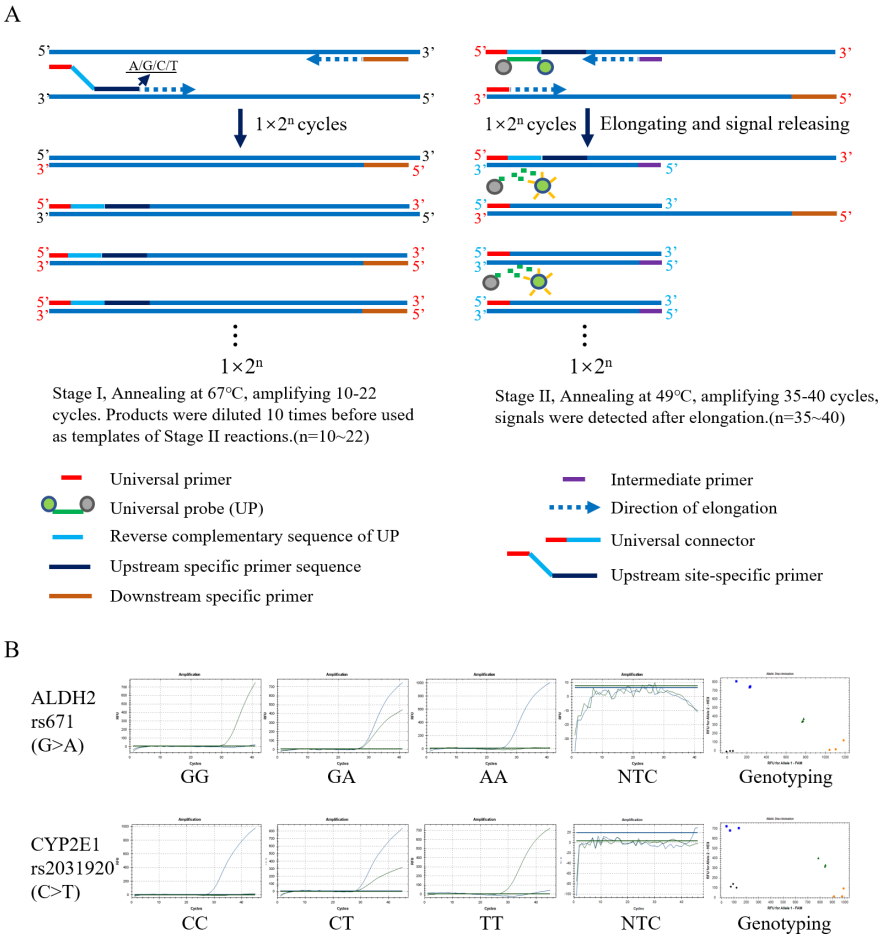

Figure 2

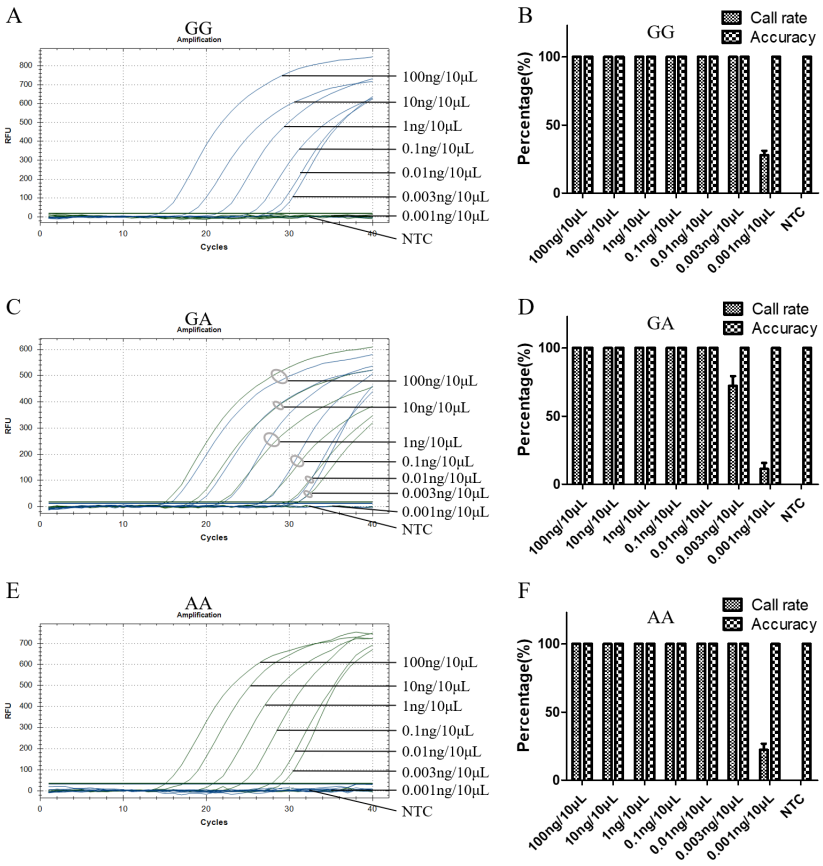

Figure 3

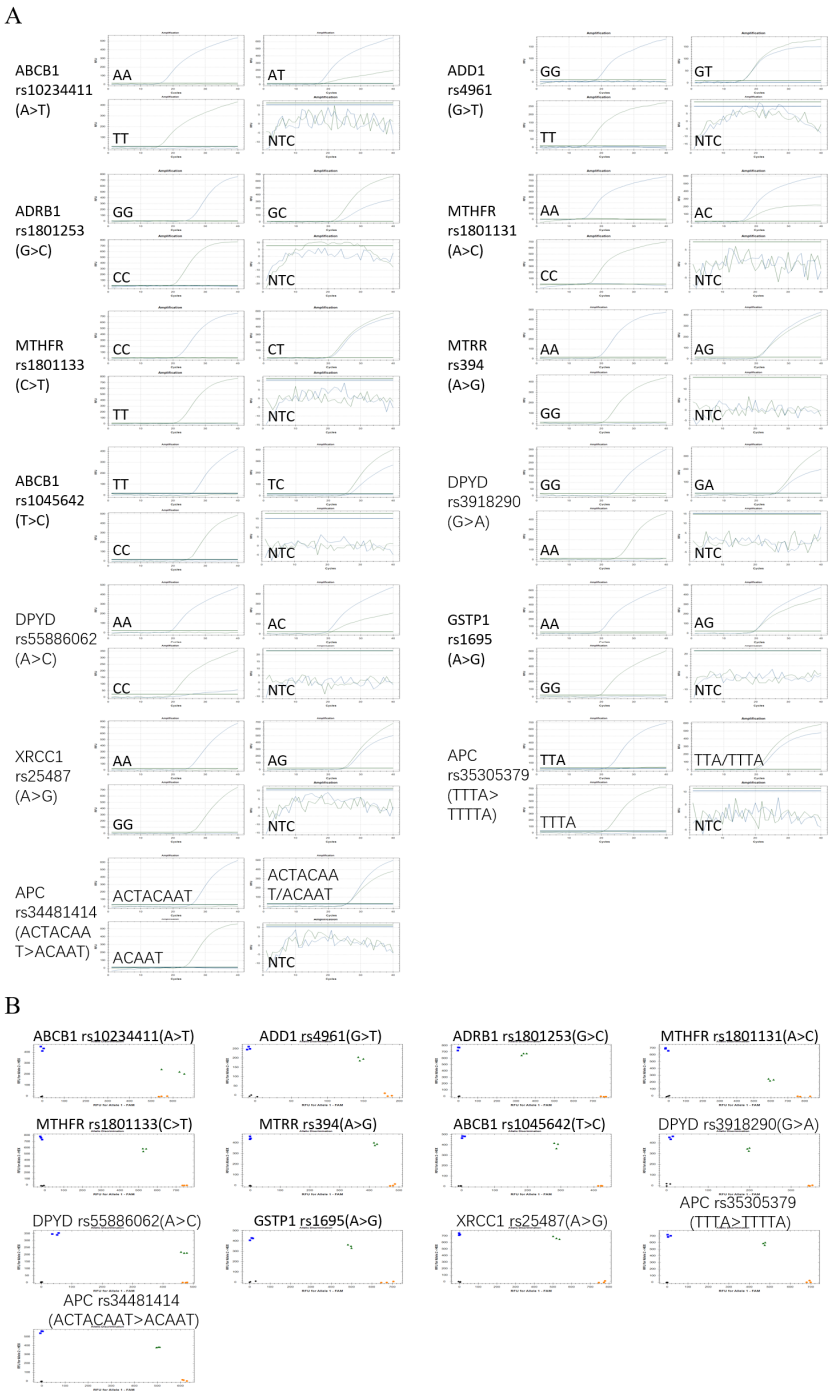

Figure 4