A High-Performing and Cost-Effective SNP Genotyping Method Using rhPCR and Universal Reporters

Kristin Beltz*, Daniel Tsang*, Junzhou Wang*, Scott Rose*, Yun Bao*, Yu Wang*, Katelyn Larkin, Susan Rupp, Daniela Schrepfer, Krishnalekha Datta, Keith Gunderson, Chris Sailor, Scott Hansen, Joseph Dobosy, Lynette Lewis, Aurita Menezes, Joseph Walder, Mark Behlke, Caifu Chen

Genetic Analysis R&D, Integrated DNA Technologies, Coralville, IA, USA
Email: *srose@idtdna.com

Abstract

We have developed a novel dual enzyme chemistry called rhAmp® SNP genotyping based on RNase H2-dependent PCR (rhPCR) that provides high signal and specificity for SNP analysis. rhAmp SNP genotyping combines a unique two-enzyme system with 3’ end blocked DNA-RNA hybrid primers to interrogate SNP loci. Activation of the blocked primers occurs upon hybridization to its perfectly matched target, which eliminates or greatly reduces primer dimers. A thermostable hot-start RNase H2 cleaves the primer immediately 5’ of the ribose sugar, releasing the blocking group and allowing primer extension. PCR specificity is further improved with the use of a mutant Taq DNA polymerase, resulting in improved allelic discrimination. Signal generation is obtained using a universal reporter system which requires only two reporter probes for any bi-allelic SNP. 1000 randomly selected SNPs were chosen to validate the 95% design rate of the design pipeline. A subsampling of 130 human SNP targets was tested and achieved a 98% call rate, and 99% call accuracy. rhAmp SNP genotyping assays are compatible with various qPCR instruments including QuantStudio™ 7 Flex, CFX384™, Intelliqube®, and Biomark HD™. In comparison to TaqMan®, rhAmp SNP genotyping assays show higher signal (Rn) and greater cluster separation, resulting in more reliable SNP genotyping performance. The rhAmp SNP genotyping solution is suited for high-throughput SNP genotyping applications in humans and plants.

Keywords

SNP Genotyping, RNase H2, rhPCR, rhAmp SNP Genotyping, Universal Reporter

*These authors contributed equally to this work.
1. Introduction

Genetic diversity at the single nucleotide polymorphism (SNP) level has been exploited in many branches of science and medicine. SNPs have been used to determine an individual’s susceptibility to a variety of disease states ranging from breast cancer to cardiovascular diseases. Additionally, polymorphisms can be used to predict a drug’s activity within a tissue. In pharmacogenetics, polymorphisms in genes encoding drug metabolism enzymes (DME) are used to help predict a drug’s potency because the absorption, distribution, metabolism and excretion (ADME) of specific compounds can all be affected by genetic variations. Known polymorphisms in those genes are known as ADME SNPs. Additionally, SNPs are also commonly used as molecular markers in breeding and genetic research.

Multiple different chemistries or assays have been developed for SNP genotyping including mass spectrometry, oligonucleotide arrays, single stranded conformational polymorphism, and sequencing. Some of the most widely used technologies tend to be centered on a fluorescent based PCR assay (i.e. 5' nuclease, Molecular Beacons, Scorpion primers, KASP, and Invader) [1]. With all these methods, the allele-specific discrimination ability is solely based on a single factor, either probe hybridization or primer extension.

Here we present a new SNP genotyping method, rhAmp SNP, that combines the allele-specific hybridization and extension into a single assay. rhAmp SNP genotyping relies on two enzymes, RNase H2 (an endoribonuclease) and a mutant Taq DNA polymerase with enhanced allelic discrimination. The rhAmp SNP genotyping primers contain a single RNA base and are 3' end blocked. In addition, each of the allele-specific primers has a unique universal tail that incorporates sequence into the amplicon that is complementary to the reporter sequences. Rapid de-blocking of the primers by RNase H2 only occurs upon formation of a perfectly matched heteroduplex between the blocked RNA-containing primer and DNA target. Once de-blocked by RNase H2, the newly activated 3’ end of the primer lays over the SNP site, and is interrogated by a mutant Taq DNA polymerase for allele-specific PCR. The major advantages of the rhAmp SNP genotyping chemistry include enhanced allelic discrimination by rhPCR, and high signal generation and cost reduction by the universal reporter system.

2. Materials and Methods

2.1. SNPs, Genomic DNA Samples and Extraction

**SNP targets:** A total of 130 SNPs were randomly selected from dbSNP Build 144 (https://www.ncbi.nlm.nih.gov/projects/SNP/, Supplemental Table S1) for most of the analyses in this study, covering the types of SNPs found in the human genome. A set of 1000 randomly selected common bi-allelic SNPs were used to evaluate the pipeline design rate.

**Synthetic templates:** Synthetic gBlocks® Gene Fragments (Integrated DNA Technologies,
Genomic DNA samples: Human genomic DNA samples were purchased from the Coriell Institute for Medical Research. A total of 136 unique samples in two sets were obtained, representative of the three original HapMap populations: Yoruba from Ibadan (YRI), Han Chinese from Beijing, China (CHB), and CEPH Utah residents (CEU) (Supplemental Table S2). DNA samples were quantified using a qPCR method targeting the RNase P (RPPH1) gene.

2.2. rhAmp SNP Genotyping

SNP genotyping was carried out using rhAmp SNP Assays, rhAmp Genotyping Master Mix, and rhAmp Reporter Mix with or without a passive reference dye (www.idtdna.com/rhAmp-Genotyping). Genotyping reactions were performed typically using 3 ng of dried down genomic DNA from either 46 (human SNP panel) or 90 (ADME SNP panel) samples (Coriell Institute) in 5 µL reactions. Unless otherwise noted, reactions were run on the QuantStudio 7 Flex instrument and analysis was performed using the QuantStudio Real-Time PCR Software (Thermo Fisher Scientific, CA). For DNA input testing, the gDNA was reduced from 3 ng to 125 pg per reaction. rhAmp SNP genotyping reactions were run with a thermal cycling profile of 95°C for 10 minutes, followed by 40 cycles at 95°C for 10 seconds, 60°C for 30 seconds, and 68°C for 20 seconds per the published protocol (www.idtdna.com/rhAmp-SNP-protocol). The same thermal cycling program was used on the CFX384 (BioRad, CA) platform, and reactions contained rhAmp SNP genotyping master mix at a final 0.85x concentration, rhAmp SNP assay and reporter mix with passive reference dye both at final 1X concentration, and Coriell gDNA input at 1.6 ng per reaction. With the Biomark HD (Fluidigm, CA) platform, 23 gDNA samples from a subset of 90 (ADME SNP panel) at 150 ng per inlet was used according to the manufacturer’s recommendation. A modified protocol without additional HotStar Taq polymerase was performed. The thermal mix protocol was modified to the following: 37°C for 2 minutes, 45°C for 10 minutes, then 25°C for 10 minutes. The reactions were run with the thermal cycling at 95°C for 10 minutes, followed by 40 cycles at 95°C for 10 seconds, 60°C for 30 seconds, and 68°C for 20 seconds. The call rate was defined as the percent of samples with an assigned SNP genotype call, and the call accuracy was the percent of called samples with the correct genotype assigned. The reported call rate and call accu-
racy were determined using auto-calls assigned by the QuantStudio Real-Time PCR Software to each of 46 samples for 130 assays.

2.3. TaqMan SNP Genotyping

SNP genotyping was carried out using TaqMan Genotyping Master Mix (Thermo Fisher Scientific) and either TaqMan SNP Genotyping Assays or TaqMan Drug Metabolism Genotyping Assays (Thermo Fisher Scientific) (assay ID numbers listed in Supplemental Table S3). SNPs were assessed using 3 ng dried down genomic DNA samples in 5 µL reactions. TaqMan SNP Genotyping Assay reactions were run with the following thermal cycling profile: 95˚C for 10 minutes, followed by 40 cycles at 95˚C for 15 seconds and 60˚C for 1 minute. TaqMan Drug Metabolism Genotyping Assay reactions were run with the following thermal cycling profile: 95˚C for 10 minutes, followed by 50 cycles at 95˚C for 15 seconds and 60˚C for 90 seconds. Thermal cycling profiles were based on the manufacturer recommended protocols. Cycling and allelic discrimination analysis was performed using the QuantStudio 7 Flex instrument and QuantStudio Real-Time PCR Software Autocaller (Thermo Fisher).

2.4. Assay Design

Universal reporter system: Of 10 million candidate sequences screened, a small set of universal primers and probes that met stringent thermodynamic properties (identical length, GC%, Tm, etc.) and had high sequence specificity against the human genome were selected. We further narrowed the probe set so that they did not form dimers and the pairwise edit distance was at least 4 for minimal interference among probes. A pair of the universal reporter probes were selected based on empirical testing (data not shown).

rhAmp SNP assay design pipeline: We have developed a sophisticated assay design pipeline that performs the following steps for each target SNP: 1) design all possible allele-specific primers (ASP), locus-specific primers (LSP) and their combinations; 2) evaluate the thermodynamic properties of the primers and amplicons, such as length, percent GC, Tm, sequence complexity, folding, etc.; 3) assess the impact of overlapping SNPs and repeats; 4) analyze the dimerization of assay primers, including self- and hetero-dimers; 5) evaluate primer compatibility, i.e. having similar Tm and length, particularly among ASPs; 6) determine the assay specificity against the human genome. Finally, based on the best combination of all the above parameters, an assay with the best overall quality is selected.

3. Results

3.1. Description of rhAmp SNP Genotyping

The schematic describing rhAmp SNP genotyping is shown in Figure 1. Each rhAmp SNP assay is comprised of one locus-specific primer (LSP) and two tailed allele-specific primers (ASPs). Each primer has a single RNA base and a 3’ end
Figure 1. Schematic of the rhAmp SNP genotyping system. (A) SNP allele C template DNA is shown as a red oval on a solid black line. Solid lines with blocking group (X) at the 3’ ends indicate a 3’ end-blocked locus-specific primer (LSP) and two 3’ end-blocked allele-specific primers (ASPs), with green and blue lines at the 5’ end indicating allele-specific tail sequences. One ASP contains the DNA base complementary to the target SNP reference allele, and the other ASP contains the DNA base complementary to the alternative allele, shown as a red oval and yellow oval, respectively. The blue triangle with R symbol in the ASPs and LSP represents the RNA base. Blue round shapes represent RNase H2. (B) RNase H2 recognizes the perfectly matched RNA-DNA heteroduplex and cleaves the primers to release the blocked 3’ end, thereby activating the primers for PCR, while the mismatched ASP remains blocked. The purple water drop shape represents mutant Taq DNA polymerase, and dashed black lines represent extension by the polymerase. (C) Extension from the ASP incorporates the allele-specific tail sequence into the ampli-con, followed by extension of the LSP to form the complement strand. A green circle and black circle represent a fluorophore and quencher, respectively, and are connected by a linker sequence (green line), together representing a dual-labeled probe. The probe and universal primer (black arrow) anneal to the strand complementary to the incorporated allele tail sequence. During PCR, the 5’ nuclease activity of the Taq polymerase degrades the probe, releasing the fluorophore and generating fluorescence signal.
blocking modification. The RNA base in the ASP is located immediately down-stream of the target SNP location (Figure 1(A)).

The blocked primers are efficiently activated by the RNase H2 enzyme only in the presence of a perfectly matched RNA-DNA heteroduplex. Once the RNase H2 cleaves the primer, the 3' end is available for extension by the mutant Taq DNA polymerase (Figure 1(B)). Each ASP has a unique 5' tail sequence that becomes incorporated into the allele-specific amplicon. The tail region adds the necessary sequences for the binding of a universal primer and fluorescent probe (Figure 1(C)). The assays are designed such that the reference allele gives a signal in the FAM dye channel, and the alternative allele signal is detected in the VIC® dye channel. The alternate allele fluorophore is Yakima Yellow®, which has a very similar excitation and emission spectra as VIC dye. With such a similar excitation and emission profile, no spectral re-calibration is required.

3.2. Comparison of Wild-Type vs. Mutant Taq DNA Polymerase

While the RNase H2 enzyme exhibits significant differences in the rate of enzymatic cleavage of an RNA base in the presence or absence of a mismatch RNA-DNA heteroduplex [2], the rhAmp SNP assay also requires a DNA polymerase to complete allele-specific PCR. A novel mutant Taq DNA polymerase was developed to function effectively in the same buffer as the RNase H2 enzyme and provide improved allelic discrimination in comparison to the wild-type Taq DNA polymerase. Figure 2 shows the performance difference in the allelic discrimination (AD) plot between two master mixes, one containing the wild type Taq DNA polymerase (Figure 2(A)) and the other with mutant Taq DNA polymerase (Figure 2(B)), with an example assay designed to detect SNP rs2269829. A total of 46 different human gDNA samples were tested at 3 ng of gDNA input per reaction. With wild-type Taq DNA polymerase, the three genotype clusters can be automatically called, however, the cluster separation is relatively poor, likely due to ASP mispriming, and some of the no template control replicates drift from the origin. Replacing the Taq with the mutant version improves the separation, tightness, and angles of the genotype clusters.

3.3. Performance of rhAmp SNP Genotyping Assays

Of 1000 common human SNPs randomly selected from the public dbSNP database, rhAmp SNP assay designs were successfully generated for 950 targets (95%). Functional test results from a randomly selected subset of 130 assays tested with 46 genomic DNA samples (human SNP panel) are summarized in Table 1. Two examples of rhAmp SNP genotyping assays targeting human SNPs, rs6068816 and rs4148946 are shown in Figure 3. Overall performance of rhAmp SNP assays shows 98% call rate and 99% accuracy (concordance with published genotypes in the NCBI database) while maintaining a high assay design rate.
Figure 2. A novel mutant Taq DNA polymerase improves allelic discrimination. rhAmp SNP genotyping master mixes formulated with either wild type (A) or mutant (B) Taq DNA polymerase were compared in a rhAmp SNP genotyping assay targeting a human SNP (rs2269829) in the presence of 3 ng gDNA from 46 individuals. With improved specificity at the SNP site, reactions containing mutant Taq DNA polymerase result in lower non-specific signal and greater cluster angle separation.

Table 1. Performance summary of rhAmp SNP assays. Design rate is based on 1,000 randomly selected common bi-allelic SNPs from NCBI. Call rate and call accuracy performance testing was generated for 130 targets using 46 Coriell gDNA samples.

| Metric                  | Percent |
|-------------------------|---------|
| Design rate             | 95      |
| Average call rate       | 98      |
| Average call accuracy   | 99      |

Effect of DNA input on rhAmp SNP genotyping

DNA input can vary due to factors such as limited biological samples, purification loss, and other causes. Low sample input may impact the quality of SNP genotyping calls due to lower fluorescent signals. Therefore, it is important to determine the DNA input range that can provide robust genotyping calls. We tested the lower limit of input DNA for the rhAmp SNP system (Figure 4). A rhAmp SNP assay targeting a human SNP (rs4657751) was tested in the presence of gDNA at 0 (no template control or NTC), 0.125 (blue), 0.5 (green) and 3 (orange) ng per PCR reaction on the QuantStudio 7 Flex instrument. Genotypes for 46 individual gDNA samples are accurately auto-called with 0.5 and 3 ng gDNA and can be manually called with 0.125 ng gDNA. For some assays tested, accurate genotypes could be manually called using sample input as low as 25 pg
Figure 3. Examples of rhAmp SNP assays targeting two human SNPs, rs6068816 (A) and rs4148946 (B), each with reference allele C and alternate allele T. Post-PCR read data was collected and normalized reporter signal (Rn) for allele 1 (FAM) and allele 2 (Yakima Yellow) is plotted along X- and Y-axes, respectively. The allelic discrimination plot displays three distinct genotype clusters including homozygous for the reference allele C/C (red), heterozygous C/T (green) and homozygous for the alternate allele T/T (blue). The assays were run on 46 human gDNA samples in 5 μL reactions.

Figure 4. Effect of gDNA input on rhAmp SNP genotyping. Allelic discrimination plot for a rhAmp SNP assay targeting a human SNP (rs4657751) is shown in the presence of gDNA at 0 (no template control), 0.125 (blue), 0.5 (green) and 3 (orange) ng per reaction. Genotypes for 46 individual gDNA samples are accurately auto-called with the QuantStudio Real-Time PCR Software in reactions containing 0.5 and 3 ng gDNA, and can be manually called with 0.125 ng gDNA.
Compatibility of rhAmp SNP genotyping with various qPCR platforms

All rhAmp SNP assays in this study were initially evaluated on the QuantStudio 7 Flex (Thermo Fisher) and CFX384 (BioRad) instruments using 3 ng genomic DNA or 1000 copies gBlocks Gene Fragments. To demonstrate the compatibility of small volume platforms, some of the assays were tested on the IntelliQube (LGC Group) and Biomark HD system (Fluidigm). An example of allelic discrimination plots for a human ADME SNP rs7668258 in the UGT2B7 gene is shown on four different platforms: QuantStudio 7 Flex (Figure 5(A)), CFX384 (Figure 5(B)), IntelliQube (Figure 5(C)) and Biomark HD (Figure 5(D)). For the QuantStudio 7 Flex and CFX384, the assay was tested with 3 ng gDNA from 90 individuals and a reaction volume of 5 µL. The same gDNA samples were tested on the IntelliQube with 1.6 ng per reaction and reaction volume of 1.6 µL. For the Biomark HD, the assay was tested with 23 Coriell gDNA samples from a subset of the 90 ADME SNP gDNA panel samples, each at 150 ng per inlet. As noted in the methods, the thermal mix protocol is modified to achieve maximum assay sensitivity and specificity. rhAmp SNP assays perform well on these four major SNP genotyping platforms, with similar cluster angles and concordant genotyping calls achieved across all samples tested.

3.4. TaqMan vs. rhAmp SNP Genotyping

A robust SNP genotype call requires relatively high fluorescent signal, large cluster angle separation and tight clusters. In a study of 18 target SNPs, rhAmp SNP assays generated, on average, at least two-fold higher signal than TaqMan for both alleles (Supplemental Figure S1). The allelic discrimination plot and raw fluorescence of an ADME SNP rs776746 assayed by rhAmp SNP and TaqMan genotyping chemistries was compared (Figure 6). The rhAmp SNP assay achieves higher signal for all samples (Figure 6(A)), and more uniform signal for both alleles in heterozygote samples (Figure 6(B)), resulting in a heterozygote cluster angle closer to the ideal 45 degrees.

4. Discussion

SNP genotyping using target-specific fluorogenic probes or allele-specific primers, such as TaqMan [3] and KASP assays [4], are two commonly used methods in medicine and agriculture. A new method, rhAmp SNP genotyping, which is based on allele-specific rhPCR combined with a universal fluorogenic reporter system, was first reported by Broccanello et al. [5]. This new SNP genotyping method has several advantages over both TaqMan and KASP.

First, the rhAmp SNP genotyping chemistry is a dual enzyme system with improved specificity of the allele-specific PCR, resulting in better genotyping cluster angles and separation (Figure 2). Dobosy and his colleagues [2] compared specificity of allele-specific PCR using unmodified PCR primers to that of 3’ blocked rhPCR primers. Of 12 mismatched combinations, the mid-range dis-
crimination in ΔCq (mismatched Cq – matched Cq) averaged 7.4 for unmodified primers and 10.9 for 3’ blocked primers, suggesting that rhPCR is at least 5-times better in allelic discrimination than standard PCR [2]. Secondly, the rhAmp SNP genotyping allele-specific and locus-specific primers are inactive.

**Figure 5.** Compatibility of rhAmp SNP genotyping with various qPCR platforms. Allelic discrimination plots for a rhAmp SNP assay targeting human ADME SNP rs7668258 are displayed with genotyping auto-calls assigned to samples run on the QuantStudio 7 Flex (A), CFX384 (B), IntelliQube (C) and Biomark HD (D). Allele 1 and Allele 2 calls indicate homozygous genotypes. Allele1/Allele2 call indicates heterozygous genotype.
Figure 6. Comparison of rhAmp and TaqMan SNP genotyping. Assays targeting human ADME SNP rs776746 in the CYP3A5 gene were tested with 3 ng gDNA from 90 individuals. Allelic discrimination plot for all samples show higher signal and better cluster separation for the rhAmp assay (A) and multicomponent amplification curves for 5 samples with heterozygote genotype show higher and more uniform signal for the rhAmp SNP assay (B).

Initially due to the 3’ end blocking group. Only upon hybridization to its perfectly matched target is the 3’ blocking group readily removed by RNase H2 cleavage. Therefore, primer dimers are eliminated or significantly reduced in the absence of target DNA (data not shown). High assay design rate (Table 1) is achieved in part due to relaxed rules checking primer to primer interactions. Thirdly, the rhAmp SNP genotyping method eliminates or reduces non-specific signal from universal reporters often caused by primer dimers. Some KASP assays can generate false genotyping clusters in the absence of targets or NTC wells, often due to non-specific interactions [6]. Non-specific interactions between TaqMan probes and PCR primers also cause higher NTC signal in some TaqMan
SNP genotyping assays [7]. Next, reduced primer dimers and the use of universal reporters with the rhAmp SNP genotyping system make multiplex SNP genotyping possible. Simultaneous detection of two SNPs in a single PCR reaction will reduce the assay cost and increase the throughput (unpublished data). Finally, use of the universal reporter system for any SNP or any species not only generates high fluorescent signals (Figure 6) but also makes SNP genotyping cost-effective.

5. Conclusion

In conclusion, we have successfully applied a new rhPCR method for SNP genotyping from purified gDNA on existing qPCR instruments such as the QuantStudio 7 Flex, CFX384, IntelliQube, and Biomark HD. This method provides a high-performance and cost-effective SNP genotyping solution for biomedical and agricultural applications.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kim, S. and Misra, A. (2007) SNP Genotyping: Technologies and Biomedical Applications. Annual Review of Biomedical Engineering, 9, 289-320. https://doi.org/10.1146/annurev.bioeng.9.060906.152037

[2] Dobosy, J.R., Rose, S.D., Beltz, K.R., Rupp, S.M., Powers, K.M., Behlke, M.A. and Walder, J.A. (2011) RNase H-Dependent PCR (rhPCR): Improved Specificity and Single Nucleotide Polymorphism Detection Using Blocked Cleavable Primers. BMC Biotechnology, 11, 80. https://doi.org/10.1186/1472-6750-11-80

[3] Livak, K.J. (1999) Allelic Discrimination Using Fluorogenic Probes and the 5’ Nuclease Assay. Genetic Analysis: Biomolecular Engineering, 14, 143-149. https://doi.org/10.1016/S1050-3862(98)00019-9

[4] Neelam, K., Brown-Guedira, G. and Huang, L. (2013) Development and Validation of a Breeder-Friendly KASPar Marker for Wheat Leaf Rust Resistance Locus Lr21. Molecular Breeding, 31, 233-237. https://doi.org/10.1007/s11032-012-9773-0

[5] Broccanello, C., Chiodi, C., Funk, A., McGrath, J.M., Panella, L. and Stevanato, P. (2018) Comparison of Three PCR-Based Assays for SNP Genotyping in Plants. Plant Methods, 14, 28. https://doi.org/10.1186/s13007-018-0295-6

[6] Semagn, K., Babu, R., Hearne S. and Olsen, M. (2014) Single Nucleotide Polymorphism Genotyping Using Kompetitive Allele Specific PCR (KASP): Overview of the Technology and Its Application in Crop Improvement. Molecular Breeding, 33, 1-14. https://doi.org/10.1007/s11032-013-9917-x

[7] Woodward, J. (2014) Bi- Allelic SNP Genotyping Using the TaqMan® Assay. Methods in Molecular Biology, 1145, 67-74. https://doi.org/10.1007/978-1-4939-0446-4_6
**Supplemental**

**Figure S1.** Signal-to-noise comparison between rhAmp and TaqMan SNP genotyping. Assays targeting 18 human SNPs were tested with 3 ng of Coriell gDNA from 46 individuals in 5 µL reactions with post-PCR read data collected and analyzed on the QuantStudio 7 Flex Real-Time PCR System software (Thermo Fisher). Signal to noise is determined for each allele by calculating the distance of the homozygous cluster from the no template control (NTC) (average assay NTC signal subtracted from the on-target allele signal from each homozygous sample). Compared to Taqman, rhAmp SNP genotyping results in higher cluster to NTC distance for both alleles in all 18 assays (A), with greater than 2-fold higher average cluster to NTC distance across the 18 assays (B).

| SNP Genotyping Assay | Homozygous cluster to NTC distance (Rn) | Homozygous cluster to NTC distance (Rn) |
|---------------------|----------------------------------------|----------------------------------------|
|                     | Allele 1 Average | Allele 1 Standard Deviation | Allele 2 Average | Allele 2 Standard Deviation |
| rhAmp               | 4.4            | 0.6                        | 3.6            | 0.3                        |
| Taqman              | 1.6            | 0.4                        | 1.7            | 0.5                        |
**Table S1.** List of human SNPs from dbSNP Build 144 used in this study.

| No. | dbSNP rsID | Polymorphism | No. | dbSNP rsID | Polymorphism | No. | dbSNP rsID | Polymorphism |
|-----|------------|--------------|-----|------------|--------------|-----|------------|--------------|
| 1   | rs1013087  | T/C          | 46  | rs2045434  | T/C          | 91  | rs6908755  | C/T          |
| 2   | rs10280568 | C/A          | 47  | rs2283270  | C/T          | 92  | rs6918402  | G/A          |
| 3   | rs10422074 | T/C          | 48  | rs2286163  | G/A          | 93  | rs6976501  | C/T          |
| 4   | rs1049434  | T/A          | 49  | rs2289221  | T/C          | 94  | rs6977778  | C/T          |
| 5   | rs11131763 | C/G          | 50  | rs2707507  | C/T          | 95  | rs6987931  | G/T          |
| 6   | rs11203651 | C/T          | 51  | rs2744290  | T/A          | 96  | rs7000593  | C/T          |
| 7   | rs11243026 | G/T          | 52  | rs27843    | T/C          | 97  | rs7141006  | G/A          |
| 8   | rs1135093  | A/G          | 53  | rs2823211  | A/T          | 98  | rs7228817  | G/A          |
| 9   | rs1149736  | C/T          | 54  | rs2851069  | T/C          | 99  | rs7260544  | A/C          |
| 10  | rs11726362 | A/T          | 55  | rs28626972 | G/A          | 100 | rs7304811  | G/T          |
| 11  | rs11835373 | T/C          | 56  | rs35512316 | T/C          | 101 | rs7306322  | G/A          |
| 12  | rs11870691 | T/C          | 57  | rs3809671  | A/G          | 102 | rs7330299  | C/T          |
| 13  | rs11909987 | G/T          | 58  | rs3825505  | A/G          | 103 | rs7395097  | T/C          |
| 14  | rs11910635 | A/G          | 59  | rs3850713  | A/G          | 104 | rs7428372  | T/G          |
| 15  | rs11972082 | A/G          | 60  | rs41521249 | G/A          | 105 | rs7518139  | A/G          |
| 16  | rs11986143 | T/C          | 61  | rs4388130  | T/C          | 106 | rs7591813  | G/A          |
| 17  | rs11996643 | C/T          | 62  | rs4591363  | G/A          | 107 | rs7669896  | G/A          |
| 18  | rs12054049 | C/A          | 63  | rs4646406  | T/A          | 108 | rs7687806  | T/C          |
| 19  | rs12141503 | G/A          | 64  | rs4653334  | G/T          | 109 | rs7704188  | T/C          |
| 20  | rs1217974  | A/G          | 65  | rs4867403  | G/A          | 110 | rs7717092  | G/A          |
| 21  | rs12197997 | T/C          | 66  | rs4907732  | A/G          | 111 | rs7743013  | A/G          |
| 22  | rs12279214 | C/A          | 67  | rs4911456  | G/A          | 112 | rs7753012  | T/G          |
| 23  | rs12372167 | G/C          | 68  | rs4971525  | T/C          | 113 | rs7760561  | A/G          |
| 24  | rs12420422 | C/T          | 69  | rs5019104  | A/G          | 114 | rs7871600  | C/T          |
| 25  | rs12507217 | C/T          | 70  | rs5030740  | C/T          | 115 | rs7939374  | G/A          |
| 26  | rs12537914 | G/A          | 71  | rs540719   | T/C          | 116 | rs7949609  | C/T          |
| 27  | rs12568373 | G/A          | 72  | rs6016306  | G/A          | 117 | rs7966055  | A/G          |
| 28  | rs12574217 | C/T          | 73  | rs6054484  | C/G          | 118 | rs8030470  | C/T          |
| 29  | rs12576497 | C/T          | 74  | rs6063733  | C/T          | 119 | rs8095910  | A/G          |
| 30  | rs12587509 | T/C          | 75  | rs6080387  | C/T          | 120 | rs8104441  | A/G          |
| 31  | rs12593964 | A/C          | 76  | rs6115865  | G/A          | 121 | rs9346650  | G/A          |
| 32  | rs12597401 | A/G          | 77  | rs6432323  | A/G          | 122 | rs9369584  | C/T          |
| 33  | rs12600458 | G/A          | 78  | rs6441819  | A/G          | 123 | rs9372649  | T/C          |
| 34  | rs1269628  | A/G          | 79  | rs6452030  | T/C          | 124 | rs9482300  | T/C          |
| 35  | rs12933889 | T/C          | 80  | rs6452908  | T/C          | 125 | rs9500864  | G/A          |
| 36  | rs12992152 | G/A          | 81  | rs6553586  | C/T          | 126 | rs9506510  | G/A          |
Continued

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 37 | rs13071497 | A/G | 82 | rs6559515 | C/G | 127 | rs9551052 | C/T |
| 38 | rs13275593 | G/A | 83 | rs6602481 | C/T | 128 | rs9573940 | G/T |
| 39 | rs1347188  | T/C | 84 | rs6693646 | C/A | 129 | rs9788745 | A/G |
| 40 | rs1415105  | C/T | 85 | rs6710520 | A/G | 130 | rs9926533 | C/T |
| 41 | rs1505022  | T/C | 86 | rs6722027 | T/C |
| 42 | rs1507710  | G/A | 87 | rs6740491 | C/A |
| 43 | rs17121237 | G/A | 88 | rs6830950 | A/G |
| 44 | rs1823762  | C/T | 89 | rs6843867 | G/A |
| 45 | rs1992631  | A/G | 90 | rs6873402 | C/T |

**Table S2.** List of human genomic DNA samples used in the study. DNA was obtained from the Coriell Institute for Medical Research and includes samples derived from the three original HapMap populations: Yoruba from Ibadan (YRI), Han Chinese from Beijing, China (CHB) and CEPH Utah residents (CEU).

|   | 46 human SNP gDNA panel |   | 90 ADME SNP gDNA panel |
|---|---|---|---|
| NA10853 | NA18558 | NA10859 | NA17123 | NA17202 | NA17230 |
| NA10855 | NA18572 | NA14474 | NA17124 | NA17203 | NA17231 |
| NA10856 | NA18576 | NA14476 | NA17125 | NA17204 | NA17232 |
| NA11829 | NA18582 | NA17102 | NA17126 | NA17205 | NA17235 |
| NA11830 | NA18603 | NA17103 | NA17127 | NA17206 | NA17237 |
| NA11831 | NA18611 | NA17104 | NA17128 | NA17207 | NA17239 |
| NA11832 | NA18620 | NA17105 | NA17129 | NA17208 | NA17240 |
| NA11839 | NA18623 | NA17107 | NA17130 | NA17209 | NA17241 |
| NA11843 | NA18633 | NA17108 | NA17131 | NA17210 | NA17242 |
| NA11882 | NA18637 | NA17109 | NA17132 | NA17211 | NA17245 |
| NA12376 | NA18854 | NA17110 | NA17134 | NA17212 | NA17247 |
| NA12383 | NA18859 | NA17111 | NA17136 | NA17213 | NA17251 |
| NA12399 | NA19093 | NA17112 | NA17137 | NA17214 | NA17253 |
| NA12400 | NA19098 | NA17113 | NA17139 | NA17215 | NA17254 |
| NA12489 | NA19099 | NA17114 | NA17140 | NA17216 | NA17255 |
| NA12546 | NA19101 | NA17115 | NA17144 | NA17217 | NA17258 |
| NA18506 | NA19102 | NA17116 | NA17147 | NA17220 | NA17259 |
| NA18522 | NA19130 | NA17117 | NA17148 | NA17221 | NA17260 |
| NA18526 | NA19131 | NA17118 | NA17149 | NA17223 | NA17261 |
| NA18529 | NA19141 | NA17119 | NA17155 | NA17225 | NA17262 |
| NA18537 | NA19172 | NA17120 | NA17188 | NA17226 | NA17263 |
| NA18545 | NA19192 | NA17121 | NA17194 | NA17227 |
| NA18552 | NA19208 | NA17122 | NA17201 | NA17228 |
Table S3. List of rhAmp SNP and TaqMan SNP genotyping assay IDs used in this study.

| dbSNP ID   | Target gene | rhAmp SNP Design ID | Taqman Assay ID |
|------------|-------------|---------------------|-----------------|
| rs2273697  | ABCC2       | Hs.ADME.rs2273697.A.1 | C__22272980_20  |
| rs717620   | ABCC2       | Hs.ADME.rs717620.T.1  | C___2814642_10  |
| rs762551   | CYP1A2      | Hs.ADME.rs762551.A.1  | C___8881221_40  |
| rs776746   | CYP3A5      | Hs.ADME.rs776746.C.1  | C__26201809_30  |
| rs1143671  | SLC15A2     | Hs.ADME.rs1143671.T.1 | C___385930_60   |
| rs1143672  | SLC15A2     | Hs.ADME.rs1143672.A.1 | C__7504282_30   |
| rs6068816  | CYP2A1      | Hs.GT.rs6068816.T.1  | C__25620091_20  |
| rs2295475  | XDH         | Hs.GT.rs2295475.A.1  | C___25473873_30 |
| rs4148946  | CHST3       | Hs.GT.rs4148946.T.1  | C__1904468_20   |
| rs3749442  | ABCC5       | Hs.GT.rs3749442.A.1  | C___479128_40   |
| rs1049434  | SLC16A1     | Hs.GT.rs1049434.T.1  | C__2017662_30   |
| rs2269829  | PON1        | Hs.GT.rs2269829.G.1  | C__2548961_1    |
| rs13126239 | ARAP2       | Hs.GT.rs13126239.A.1 | C__1485150_10   |
| rs4073846  | MAST2       | Hs.GT.rs4073846.G.1  | C___37954_10    |
| rs332433   | Intergenic  | Hs.GT.rs332433.G.1   | C__7951370_10   |
| rs1799865  | CCR2        | Hs.GT.rs1799865.C.1  | C__2610509_30   |
| rs495529   | MIR3658, UCK2 | Hs.GT.rs495529.G.1   | C___19642_1    |
| rs4657751  | Intergenic  | Hs.GT.rs4657751.A.1  | C__29085526_10  |