ORIGINAL ARTICLE

Robustness of single-base extension against mismatches at the site of primer attachment in a clinical assay
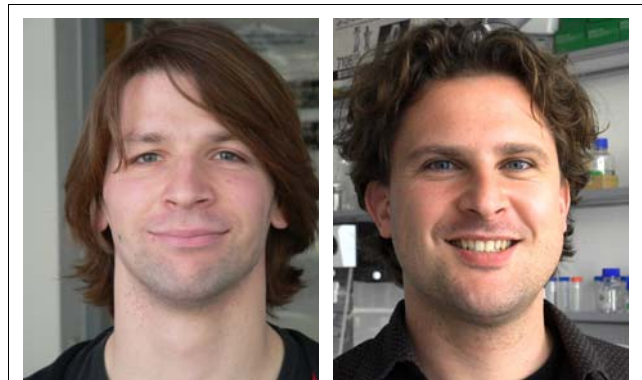
**Holger Kirsten · Daniel Teupser · Jana Weissfuss ·
Grit Wolfram · Frank Emmrich · Peter Ahnert**

Abstract DNA genotyping is important for epidemiological and clinical studies and diagnosis for individuals. Genotyping error can strongly influence the outcome of such investigations. One possible reason for genotyping error is additional DNA sequence variation, which can lead to allelic dropout. Based on a published study where allelic dropout occurred in genotyping the cholesteryl ester transfer protein TaqIB polymorphism by a TaqMan-based method, we investigated the susceptibility of the single-base extension (SBE)-based GenoSNIP method to additional sequence variation at the primer attachment site. SBE genotyping was applied to 147 patient samples with known alleles and to synthetic SBE templates. Variables were positions of nucleotide mismatches, yield of SBE reactions, primer design, and ratio of alleles in the template. No allelic dropout occurred when genotyping the TaqIB polymorphism regardless of the reported nucleotide mismatch. Yields of SBE assays critical for allelic dropout were decreased in the presence of the reported nucleotide mismatch depending on SBE assay design. In a systematic mutation scan, only the position immediately adjacent to the polymorphism caused allelic dropout under standard conditions. Depending on SBE assay design, changes in allelic

HOLGER KIRSTEN
received his Diploma in biochemistry from the University of Leipzig, Germany. He is presently a Ph.D. candidate in biochemistry and human genetics of the University of Leipzig, Center for Biotechnology and Biomedicine. His research interests include human genetics of complex diseases, genotyping technology, and concepts of bioinformatics software.

PETER AHNERT
received a Ph.D. in biochemistry from The Ohio State University in Columbus, OH. He leads a research group at the Center for Biotechnology and Biomedicine of the University of Leipzig, Germany and holds an appointment at the Fraunhofer Institute for Cell Therapy and Immunology. His main research interests include the genetics of complex diseases and application of findings to molecular diagnostics.

H. Kirsten · J. Weissfuss · G. Wolfram · F. Emmrich · P. Ahnert (✉)
IKIT/BBZ, Faculty of Medicine,
Center for Biotechnology and Biomedicine,
University of Leipzig,
Johannisallee 30,
04103 Leipzig, Germany
e-mail: ahnertpe@gmx.net

D. Teupser
Institute of Laboratory Medicine, Clinical Chemistry,
and Molecular Diagnostics, University Hospital Leipzig,
Liebigstr. 27,
04103 Leipzig, Germany

ratio due to a nucleotide mismatch were similar in appearance to changes due to sample mixture or copy number variation. In conclusion, we found the SBE genotyping assays to be relatively robust against interfering DNA variations. The importance of appropriate design and validation of assays, especially in regard to critical yields and potentially interfering nucleotide mismatches, should be emphasized particularly in clinical settings. Care should be taken when interpreting observed changes in the allelic ratio,

which could be caused by nucleotide mismatches, sample mixtures, or copy number variation.

Introduction

Genotyping error can have a profound effect on the outcome of genetic studies. For a review, see Pompanon et al. [1]. As an example, an error rate as small as 0.5% can drastically influence the result of a linkage study [2]. Various methods have been developed to account for genotyping error [3], but they have severe limitations in smaller cohorts and cannot be applied to diagnostic genotyping of single individuals as required in the emerging field of personalized medicine.

Genotyping errors can arise in any step of a genotyping procedure. Several strategies have been developed to minimize the error rate of genotyping procedures [4]. However, sample-specific factors may still compromise genotyping accuracy. A specific type of genotyping error that may be caused by sample properties is allelic loss. In this study, we define allelic loss as a genotyping event where one of the present alleles is not detected. This is particularly problematic in heterozygous samples, as a true heterozygote will appear as a good quality homozygote with no indication of a problem.

Allelic loss may result from different reasons. One is insufficient or unequal polymerase chain reaction (PCR) amplification of one or more alleles. Another reason for allelic loss can be copy number variation. In such a case, measured allelic ratios can be changed dramatically due to the greater amount of template for one allele [5]. This may lead to mistyping. All genotyping methods relying on an initial amplification of genetic material by PCR are expected to be similarly affected by these two reasons for allelic loss.

A third reason for allelic loss can be additional sequence variation adjacent to the DNA variation of interest. These variants may prevent correct binding of probes or enzymes and therefore potentially interfere with genotyping, leading to a decrease or disappearance of the allele-specific signal. In contrast to the first two reasons, the third reason may impact various genotyping techniques differently, depending on their principle of allele recognition.

Single-base variations are the most abundant type of variation in the human genome [6].

Therefore, they are especially prone to interfere with genotyping. Within this report, a potentially interfering single nucleotide change near a polymorphism of interest will be called "additional sequence variation" or "nucleotide mismatch." An overview of the genotyping accuracy of different methods in the context of interference of an additional sequence variation is shown in Table 1. Allelic loss was reported for TaqMan, allele-specific PCR, Invader, and Hybridization. No mistyping was found with probe melting curve analysis and single-base extension (SBE). In these studies, a deeper investigation of the dependence of mistyping on reaction parameters was not reported.

To address concerns about potential allelic loss in SBE-based genotyping, we applied the method "GenoSNIP" [17]. In this technique, additional sequence variation under the hybridized SBE primer may interfere with allele-specific primer extension by the polymerase. GenoSNIP is a mass spectrometry-based SBE approach. In this technique, even very small signals can be unambiguously assigned to certain genotypes because expected signals can be exactly calculated. We genotyped the pharmacogenetically important TaqIB polymorphism (rs708272) in the gene for cholesteryl ester transfer protein (CETP). The AA genotype is strongly associated with higher high-density lipoprotein-cholesterol

Table 1 Nucleotide mismatches and genotyping error reported in the literature

Genotyping method	Genotype of the sample	Distance from nucleotide mismatch to allele of interest (bp)	Genotyping error found	Reference
TaqMan	Homozygous	9	No	[7]
	Heterozygous	6, 9	Yes, allelic loss	[8, 7]
Probe melting	Homozygous	1, 2, 3, 4, 5, 7	No	[9–11]
	Heterozygous	1, 2, 7	No	[10, 12, 13]
Hybridization	Homozygous	8, 1	No	[14]
	Heterozygous	2, 5	Yes, allelic loss	[14, 15]
Allele-specific PCR	Heterozygous	5	Yes, allelic loss	[15]
Restriction digest	Homozygous	1	Yes, allelic loss	[16]
Invader	Heterozygous	5	Yes, allelic loss	[15]
Primer extension	Heterozygous	5	No	[15]

Note that heterozygous DNA was not investigated in every study so that allelic loss leading to genotyping error in heterozygous DNA could not always be found by default.

levels and a 20% lower risk of coronary artery disease in comparison to the GG genotype [18]. In addition, one study identified the AA genotype as a predictor for benefit of statin therapy for patients with significant coronary artery disease [19]. An assay for the TaqIB polymorphism is offered and routinely applied by the clinical chemistry center of our university hospital. However, in some samples, an additional sequence variation (rs5030708) occurred unexpectedly adjacent to the TaqIB polymorphism on the chromosomes carrying the Taq1B G allele. The interfering additional sequence variation rs5030708 lead to false genotype calling with TaqMan technology [7]. Using the same well-characterized sample set and synthetic SBE templates, we investigated the dependence of allelic loss on different variables. These were the position of nucleotide mismatches, yield of SBE reactions, primer design, and ratio of alleles in the template.

Materials and methods

Patients and DNA preparation

Three groups of 48, 49, and 50 patients carrying the g/g, g/a, and a/a alleles of the CETP TaqIB polymorphism, respectively, were selected for this study (CETP TaqIB cohort). Three patients with the g/g genotype and three patients with the g/a TaqIB genotype were heterozygous for a second variation, rs5030708. DNA preparation is described elsewhere [7]. Genotypes of the CETP TaqIB polymorphism and the occurrence of the additional single-base variant rs5030708 were already determined by Taq-Man, probe melting curve analysis, restriction digests, and, occasionally, DNA-sequencing in both directions [7]. All patients had given informed consent to participate in the study.

Genotyping by single-base extension (GenoSNIP assay)

Polymerase chain reaction

A 312-bp fragment was amplified by PCR using the primers PCRfor and PCRrev (Table 2). Reaction conditions were as follows: 0.2 μM primer (MWG AG, Ebersberg, Germany); 0.2 mM dNTP (Carl Roth, Karlsruhe, Germany); 2.5 mM MgCl$_2$; 0.8 U Hot FIREPol (both Solis Biodyne, Tartu, Estonia); V_{total}=20 μl. After 15-min initial denaturation at 95°C, 40 cycles at 92°C for 45 s, 58°C for 45 s, and 72°C for 45 s followed. The same PCR product from each sample was used for both primer extension assays.

SAP and ExoI digest

Eight-microliter PCR product was digested with 0.3-U shrimp alkaline phosphatase (SAP; Amersham, Buckinghamshire, England) and 0.2-U exonuclease I (ExoI; NEB, Ipswich, MA, USA) for 1 h at 37°C. Enzymes were deactivated for 20 min at 80°C.

Primer extension

In the extension reaction, 6.6-pmol extension primer was used (Biotez GmbH, Berlin, Germany). For assay SBE-PL8, primer PL8 was used, and for assay SBE-PL16, primer PL16 (Table 2). Conditions were 0.13 mM ddNTPs (ddC, ddA, ddG and ddT in assays with PL16 and ddC, ddA, ddG in assays with PL8; Carl Roth), 6.25 mM MgCl$_2$, and 1 U TERMIPol (both Solis Biodyne). All products

Table 2 SBE primers (sense) and synthetic SBE templates (antisense) used in the study

Name	Sequence	Comment
PCRfor	5′-aggtatagggatttgtgtttgt-3′	PCR primer
PCRrev	5′-gaacttggcgatactctatttt-3′	PCR primer
PL8	5′-bioacccagaatcacLgggggttc-3′	SBE primer
PL16	5′-bioacccLgaatcactggggttc-3′	SBE primer
MM_0	5′-accctaact*c*gaaccccagtgattctgggtctcagacaaa-3′	SBE template, allele G, no nucleotide mismatch
MM_1	5′-accctaact*cc*aaccccagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 1 bp away
MM_2	5′-accctaact*cgt*accccagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 2 bp away
MM_3	5′-accctaact*cga*gccccagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 3 bp away
MM_4	5′-accctaact*cgaa*gcccagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 4 bp away
MM_5	5′-accctaact*cgaac*gccagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 5 bp away
MM_6	5′-accctaact*cgaacc*gcagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 6 bp away
MM_7	5′-accctaact*cgaaccc*gagtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 7 bp away
MM_8	5′-accctaact*cgaacccc*tgtgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 8 bp away
MM_9	5′-accctaact*cgaaccccaa*tgattctgggtctcagacaaa-3′	SBE template, allele G, nucleotide mismatch 9 bp away
HET	5′-accctaact*t*gaaccccagtgattctgggtctcagacaaa-3′	SBE template, allele A, no nucleotide mismatch

Variable positions are written in italic. *L* Photocleavable linker [17], *bio* biotin

from the SAP/ExoI reaction were used. Primer extension was done as follows: initial denaturation for 4 min at 94°C followed by 35 cycles at 94°C for 10 s, 55°C for 30 s, and 72°C for 10 s.

Streptavidin purification, MALDI-TOF mass spectrometry, and genotype calling

The PureDisk robot-supported variant of GenoSNIP (Bruker Daltonics GmbH, Bremen, Germany) and subsequent matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry were done as recommended by the manufacturer with minor modifications. Briefly, primer extension products were bound to streptavidin plates, purified by subsequent washing, and released into the elution buffer by cleaving the linker (L) (Table 2) by UV-irradiation. In contrast to the original procedure, an additional washing step with deionized water was done after the use of buffer WP2. Deionized water was also used instead of elution buffer. If necessary, a further purification step using Clean Resin (Sequenom, San Diego, USA) was added. Laser-desorbed positive ions were analyzed in the linear mode of an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics GmbH) with 20 and 18.7 kV on the sample target and the conversion electrode, respectively. External calibration was performed using an oligonucleotide mixture of known molecular masses. MALDI-TOF mass spectra resulted from averaging five spectra, each consisting of 50 laser shots. All MALDI-TOF mass spectrometry measurements were done in four replicates for every sample. Peak picking was done with the software Genotools™ 2.0 (Bruker Daltonics GmbH). The software "AGS" (Automated Genotyping System, contact: ahnert@uni-leipzig.de) was used in final genotype calling. Briefly, it uses peak information from Genotools™ 2.0, filters unspecific signal based on signal intensities and known mass differences, and can take into account fourfold spotting of one sample. Yield was calculated using the signal to noise ratios (SNR) of the measured peaks. It was defined as:

$$\text{Yield} = 100\% \times (\text{SNR}_{\text{allele a}} + \text{SNR}_{\text{allele b}}) / (\text{SNR}_{\text{allele a}} + \text{SNR}_{\text{allele b}} + \text{SNR}_{\text{unextented primer}})$$

Determination of critical yield

To investigate the effect of PCR efficiency on the yield of SBE reactions, PCR products were diluted from 1:1 to 1:512 in 1× PCR buffer and genotyped as described above. PCR products were obtained from heterozygous individuals with and without the DNA variation leading to a nucleotide mismatch 9 bp away from the TaqIB polymorphism.

Effect of the position of the nucleotide mismatch on allelic loss

Oligonucleotides HET and MM_0 to MM_9 (Table 2) were synthesized and high-purity salt-free (HPSF) purified by MWG AG. The number in the name indicates the position of the nucleotide mismatch; HET represents a second allele to simulate heterozygous DNA at position 10 (counting from 5′, see Table 2). DNA was quantified using the optical density at 260 nm, corrected by the optical density at 320 nm. DNA purity was assessed by the (OD260–OD320)/(OD280–OD320) ratio. To find reaction conditions where the average yield of SBE reactions is similar to those obtained when PCR products are used as templates, the amount of synthetic SBE template was varied between 0.07 and 20 pmol. The amount of genotyping primer PL8 was 5 pmol. Based on this, in all further experiments, synthetic SBE template was used at concentrations between 0.07 and 0.13 pmol. The primer concentration then was 6.6 pmol. Genotyping was done as described above. All experiments were done in duplicates and each analyzed in fourfold redundant MALDI-TOF mass spectrometry as described above. Significance was calculated applying Kruskal–Wallis analysis of variance on ranks in combination with Dunn's method as implemented in SigmaStat 3.1.

Model system for copy number variation

DNA from a homozygous A/A and a heterozygous A/G individual was adjusted to the same concentrations (10 ng/µl) using the intercalating dye Picogreen (Molecular Probes, Eugene, OR, USA) as recommended by the manufacturer. Mixtures consisting of heterozygous DNA and homozygous DNA were pooled in a ratio of 1:0, 1:1, and 1:2. Total DNA amounts were 10 ng, 20 ng, and 30 ng, respectively. Genotyping was done as described above. For graphical representation, eight normalized MALDI-TOF spectra were averaged.

Database analysis

DbSNP build 125 was used (http://www.ncbi.nlm.nih.gov/projects/SNP/) to identify all adjacent single nucleotide polymorphisms (SNPs) with one and with two directly neighboring SNPs. Chromosome reports for mitochondrial,

autosomal, and sex chromosomes were downloaded. SNPs were filtered for mapping to a single position in the genome, mapping to the reference assembly, not withdrawn, total number of chromosome hits during mapping=1, and validation information. Statistics were calculated using Java™ 1.4, Microsoft Excel 2000, and Microsoft Access 2000 softwares.
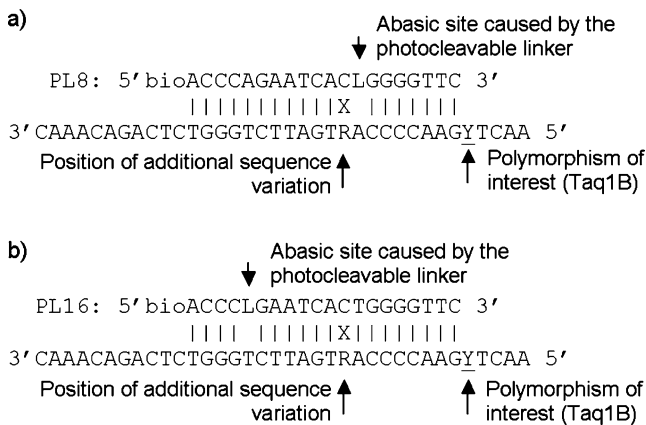
Results

The goal of this study was to assess the robustness of an SBE-based genotyping method in the context of an additional single nucleotide sequence variation present under the SBE primer.

The work focuses on genotyping the pharmacologically relevant TaqIB polymorphism in the CETP gene. In a previously published study, genotyping error due to allelic loss was reported for TaqMan analysis [7]. The reason for wrong genotype calling was a previously undetected polymorphism on one chromosome leading to a nucleotide mismatch 9 bp away from the TaqIB polymorphism.

The first aim was to investigate whether or not the SBE technique "GenoSNIP" is also prone to this error. We applied (double blind) two different genotyping assay designs with different positions of the photocleavable linker in the genotyping primer. The position of the linker of the first primer was exactly next to the nucleotide mismatch, the position of the linker of the second primer was eight bases in 5′ direction of the nucleotide mismatch (Fig. 1).

In both assays, no false genotype calling occurred under standard conditions. Yields in assays SBE-PL8 and SBE-PL16 were 47%–100% and 79%–100%, respectively. However, assay SBE-PL8 performed better than SBE-PL16. For assay SBE-PL8, four clearly distinct areas
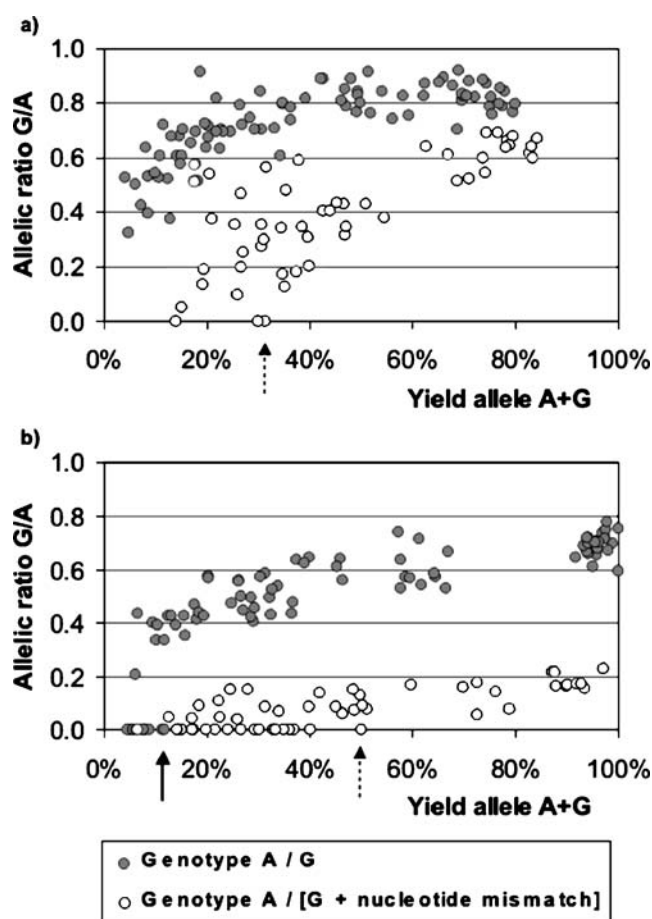
representing all three genotypes and the negative control can be seen (Fig. 2a). In contrast, in the scatter plot for assay SBE-PL16, an additional area can be seen (Fig. 2b). It is formed by signals from heterozygous DNA where the nucleotide mismatch leads to a decrease in the average allelic ratio from 0.6 to 0.2 (data calculated from values plotted in Fig. 2b).

The second aim was to assess whether allelic loss might be dependent on the SBE yield. A lower yield decreases product signals. For heterozygous genotypes with unequal allelic signals, allelic loss might occur. We made a series of dilutions of PCR products to produce a range of SBE yields more extreme than encountered in this genotyping study. This was done for samples with and without the potentially interfering nucleotide mis-
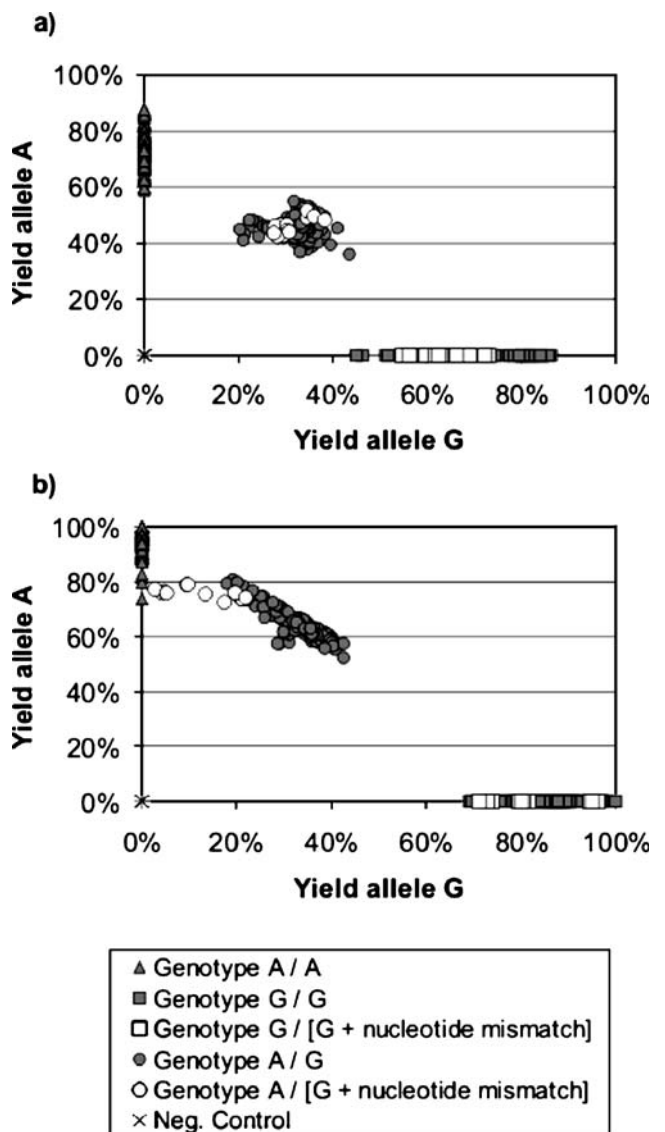
Fig. 2 Results of genotyping the CETP TaqIB polymorphism with two different SBE assays. **a** Result for assay SBE-PL8. **b** Result for assay SBE-PL16
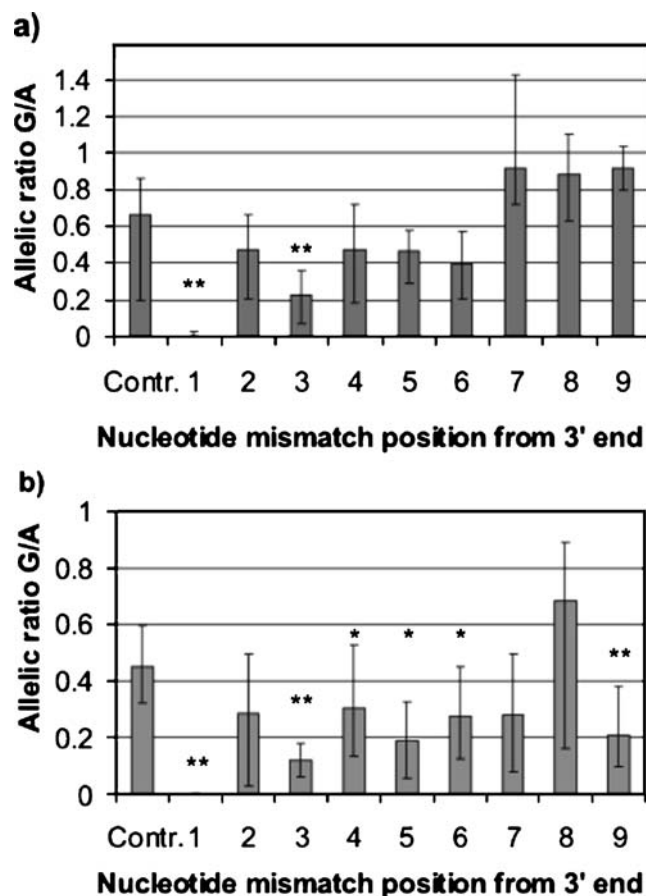

Fig. 1 Scheme of the first and second genotyping primer annealed with a template. The Taq1B polymorphism is assigned as *Y*. **a** Assay SBE-PL8 with primer PL8 (linker position 8). **b** Assay SBE-PL16 with primer PL16 (linker position 16)

Fig. 3 Determination of critical yields. **a** Result for assay SBE-PL8. **b** Result for assay SBE-PL16. Observed critical yield is indicated by a *dashed arrow* (nucleotide mismatch present) or by a *solid arrow* (nucleotide mismatch not present)

Fig. 4 Impact of nucleotide mismatch position on the allelic ratio. **a** Result for assay SBE-PL8. **b** Result for assay SBE-PL16. *Contr.* represents an assay with no nucleotide mismatch. *Error bars* range from the smallest to the largest observed value. Significant differences to the control are indicated by *asterisks* (*$p<0.05$, **$p<0.01$)

match rs5030708 (Fig. 3). We defined the term critical yield as the yield where, for the first time, the allele having normally the lower intensity is below the detection limit, but the other allele is still found.

If no nucleotide mismatch was present, in assay SBE-PL16, a critical yield (12%) was observed; critical yield of SBE-PL8 was below 4% (lowest dilution tested). When the nucleotide mismatch 9 bp away from the polymorphism of interest was present, critical yields were 50% for SBE-PL16 and 30% for SBE-PL8.

The third aim was to explore the effect of the position of the nucleotide mismatch on allelic loss. We designed synthetic SBE templates producing nucleotide mismatches between positions 1 and 9 (Table 2). The choice of mismatch types with strongest impact on polymerase elongation was made according to Ayyadevara et al. [20]. Yields of assays SBE-PL8 and SBE-PL16 with synthetic SBE templates were similar to yields observed for genotyping the original CETP TaqIB cohort. The position at the very 3′ end of the primer was most susceptible to the effect of the nucleotide mismatch—here, allelic loss

occurred for both primers (Fig. 4). All other nucleotide mismatch positions allowed correct genotyping; we never observed allelic loss on these positions. However, the influence of nucleotide mismatches at positions 2 to 7 and at position 9 was larger for SBE-PL16 than for SBE-PL8. For the latter, a nucleotide mismatch at positions 7 to 9 did not appear to influence the allelic ratio.

As the final aim, we wanted to analyze whether signal changes due to nucleotide mismatches interfering with the assay may be similar in appearance with signal changes due to other reasons. Using mixtures of homozygous and heterozygous samples as template, we simulated the situation of copy number variation. The mixtures represent the situations of genotyping either three or five copies of the allele adenine together with one copy of the allele guanine in the same genotyping reaction. The spectra resulting from genotyping are shown in Fig. 5. The nucleotide mismatch interfering with the assay decreased the allelic ratio in a similar way as copy number variation would do. The experiment also shows that sample mixtures may produce a similarly altered allelic ratio.
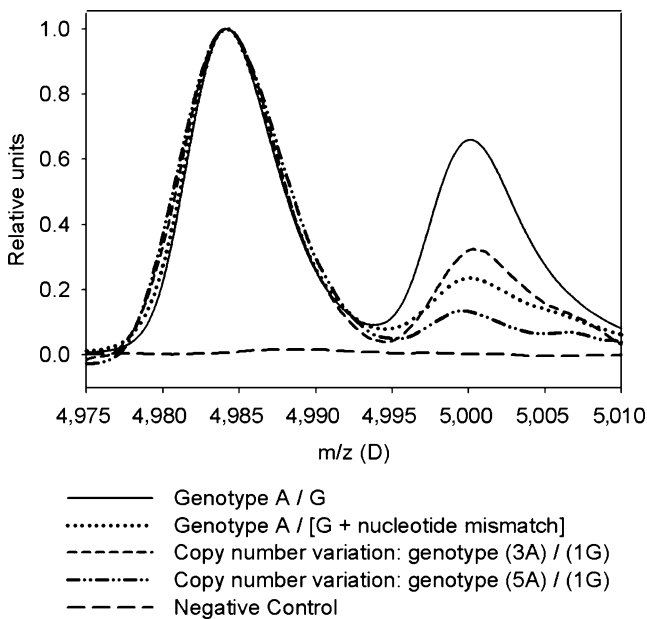
Fig. 5 Other reasons for an altered allelic ratio: effect of nucleotide mismatch interference and copy number variation (simulated by a sample mixture) on genotyping heterozygous DNA. The adenine and guanine alleles are expected to have masses of 4,984 D and 5,000 D, respectively. All observed masses were in the range expected from the accuracy of the measurements. The samples "genotype (3A)/(1G)" and "genotype (5A)/(1G)" indicate samples having one copy of the allele guanine and either three or five copies of the allele adenine, respectively

Discussion

The aim of this study was to address concerns about allelic loss in SBE-based genotyping in the context of potentially interfering single-base variations (called "nucleotide mismatches" or "additional sequence variation") as illustrated by the example of the TaqIB (rs708272) polymorphism in the gene for CETP.

Robustness of SBE: comparison to TaqMan and two primer designs

Genotyping error was reported previously using TaqMan and found to be due to a previously undetected nucleotide mismatch 9 bp away from the TaqIB polymorphism [7]. In all samples tested, this nucleotide mismatch occurred on only one chromosome. When we applied SBE genotyping to the same samples, no allelic loss occurred for two different assay designs (Fig. 1). In TaqMan, allele discrimination results from different hybridization behavior of the probe between the two alleles of the target. This results from an intended mismatch between the probe and one of the alleles. Therefore, probe design has to be sensitive to this intended mismatch but will likely be sensitive to previously undetected nucleotide mismatches as well. In

SBE-based methods, allele discrimination is achieved by target-specific extension of a primer. However, in the case of an interfering nucleotide mismatch on one chromosome only, there will be different hybridization behavior between the SBE primer and the two chromosomes. This may lead to genotyping problems in samples heterozygous for the SNP of interest. We speculate, if the SBE reaction for the template–primer pair with nucleotide mismatch is out-competed by the SBE reaction for the pair without nucleotide mismatch, allelic loss will occur. Hybridization of primer and template is a crucial factor in primer extension by DNA polymerases. Previous studies on polymerases suggest that next to primer attachment to the target in general, the 3′ terminal bases of the primer have the most impact on extension efficiency. We investigated the influence of assay-design-dependent primer hybridization on SBE genotyping of the TaqIB polymorphism.

Assays were designed to resemble two situations: In assay SBE-PL8, the placement of the linker together with the nucleotide mismatch formed a single two-base mismatch region between primer and template (Fig. 1a). No effect of the nucleotide mismatch on allelic loss or allelic ratio was observed. This is consistent with the notion that the two situations (perfect hybridization for 12 and 7 bp without nucleotide mismatch vs 11 and 7 bp with nucleotide mismatch) are very similar and likely sufficient for general primer attachment as well as for perfect hybridization at the 3′ end of the primer.

In the design of SBE-PL16, two disjoined mismatches were created, resulting in short regions of perfect hybridization (4 and 15 bp without nucleotide mismatch vs 4, 6, 8 bp in the presence of the nucleotide mismatch, Fig. 1b). In this study, the nucleotide mismatch led to a clear difference in clustering of allelic ratios in the scatter plot, but heterozygotes could still be assigned correctly (Fig. 2b). In this case, general hybridization of the primer to the template arises mostly from the 15-bp stretch. When this is disrupted, general hybridization most likely decreases, leading to less efficiency of the SBE reaction. This is in agreement with personal observations that hybridization decreases dramatically for stretches below 9 bp (data not shown). A similar observation was made in an SBE-based study of the INSR A431C polymorphism [15]. There, a previously undetected nucleotide mismatch 5 bp away from the polymorphism of interest also caused an altered allelic ratio. These observations show that, to a certain extent, an altered ratio of SBE products in heterozygotes can reveal adjacent nucleotide mismatches. We conclude that heterozygous samples with unusual allelic ratios should be sequenced to control for additional sequence variation. If an additional sequence variation interferes with SBE primer hybridization, a new assay design might be considered. The new assay design should strive to minimize the fragmen-

tation of continuous hybridization in the presence of the nucleotide mismatch.

Other reasons for an altered allelic ratio exist. The allelic ratio of the sample with interfering adjacent nucleotide mismatch was just in between the observed allelic ratios for the two simulated copy number variations. Figure 5 shows that the effect of an adjacent nucleotide mismatch on the allelic ratio sometimes may be similar to the effect of a copy number variation. Copy number variation or sample mixtures can be excluded by sequencing. Reasons for unusual allelic ratios may also be found at the level of the PCR.

Critical yield for allelic loss

Although the replication of genotyping the CETP TaqIB cohort with two differently designed SBE assays did not lead to incorrect genotypes, changes of allelic ratios depending on the presence of the nucleotide mismatch and the assay design were observed. Therefore, we wondered whether there might be a critical yield of the SBE reaction for the detection of heterozygous DNA. We defined the term critical yield as the yield where, for the first time, the allele having normally the lower intensity is below the detection limit, but the other allele is still found. This situation would lead to false genotyping. Therefore, a lower critical yield indicates a more robust genotyping assay. In the presence of the nucleotide mismatch, for the two assay designs, SBE-PL8 and SBE-PL16, we found critical yields of 30% and 50%, respectively. It seems important to note that even without a nucleotide mismatch in assay SBE-PL16, allelic dropout occurred at SBE yields of 12% and below. We conclude that it might be valuable to assess genotyping assays by varying the SBE yield for a heterozygous sample to judge whether all product signals are reliably detectable. This is especially important if the assay is part of a multiplex where signal/noise ratios are normally lower than in singleplexes and to ensure valid genotype calls even if a previously undetected nucleotide mismatch may interfere with the assay.

Role of nucleotide mismatch positions

To investigate the influence of the position of a nucleotide mismatch on SBE efficiency, we varied its location between the first and the ninth position counting from the 3′ ends of the genotyping primers for assays SBE-PL8 and SBE-PL16. We expected that these positions might influence the interaction of the polymerase with the primer–template complex. To maximize the effect of the nucleotide mismatches, we used nucleotide exchanges known to most strongly decreased PCR reaction efficiency [20]. In accordance with this study, a nucleotide mismatch at the ultimate 3′ end of the genotyping primer had the most impeding

effect for both assay designs (Fig. 4). This is in agreement with a very stringent requirement of the polymerase for a perfect match at the 3′ end in PCR [20]. All other nucleotide mismatches further away from the 3′ end of the primers did not lead to allelic loss in any single experiment. Despite the absence of allelic loss, a rather significant decrease in nucleotide incorporation was seen for mismatch position 3 for both assay designs. Polymerases make contact with the template and primer strands for about the first eight bases counting from the 3′ end of the primer toward its 5′ end [21]. It could be hypothesized that these contacts may be important for the ternary complex, ultimately contributing to the formation of the restricted binding pocket for the incoming nucleotide. In structures of the open and closed complexes of Taq DNA Pol I, direct contacts of amino acid side chains with both the primer and template bases at position 1 are observed. Starting from position 3, further contacts between the polymerase and the primer bases are observed [21]. Nucleotide mismatches directly adjacent to or at the site of the photocleavable linker (assay SBE-PL8 with nucleotide mismatch positions 7 to 9) had no discernable effect, in agreement with the discussion of the two assay designs. This illustrates an advantage of the GenoSNIP assay: For known sequence alterations, placing the photocleavable linker directly at the site of the potential nucleotide mismatch can eliminate mismatch effects. However, a linker at the first to fifth positions from the 3′ end of the SBE primer might result in genotyping problems (T. Wenzel, Bruker Daltonics, personal communication). Alternatively, primers with degenerate nucleotides and—as reported for probe melting assays [22]—primers with an additional intended mismatch can be considered to overcome the impact of the additional sequence variation.

Figure 4 illustrates that an altered allelic ratio in the products of an SBE reaction can, in some instances, reveal additional sequence variation (e.g., nucleotide mismatch position 3). In contrast, when the interference of the additional sequence variation leads to allelic loss (nucleotide mismatch position 1), this will only be noticeable if Mendelian inheritance is conflicted or the real genotype is known from other studies. In some cases, additional sequence variation does not result in an altered allelic ratio or allelic loss (e.g., nucleotide mismatch position 8). In such cases, not only the additional sequence variation is not detected but also has no consequence for genotyping accuracy.

Database analysis

To assess the relevance of the finding that a nucleotide mismatch directly next to a variation of interest may most severely impede genotyping by SBE, we carried out a

survey in dbSNP (build 125). We found more than 62,000 SNPs or 1.3% of 4.7 million validated SNPs with a directly neighboring second SNP. It clearly is not a very rare event that two polymorphisms are immediately adjacent to each other. Prominent clinical examples of this situation are the apolipoprotein 100 G10699A mutation [9], the prothrombin G20210A mutation, and the factor V G1691A mutation [23]. The impact of nucleotide mismatches on SBE reactions likely is sequence-specific, and somatic mutations may add to the problem. Nevertheless, our data suggest that assay design should always take into account the latest SNP database information to decrease the risk of assay interference caused by nearby SNPs. A very robust assay format might be a duplex SBE reaction with two primers genotyping from the 3′ and the 5′ side of the SNP of interest. In dbSNP, only 258 SNPs or 0.005% of all investigated validated SNPs were found to have two neighboring variations on both sides.

In conclusion, we found SBE to be relatively robust to additional sequence variations, which might interfere with the investigated SBE assays. If sufficient care is taken in assay design and validation, SBE-based methods should be well suitable for genotyping in clinical settings.

References

1. Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859
2. Walters K (2005) The effect of genotyping error in sib-pair genomewide linkage scans depends crucially upon the method of analysis. J Hum Genet 50:329–337
3. Sobel E, Papp JC, Lange K (2002) Detection and integration of genotyping errors in statistical genetics. Am J Hum Genet 70:496–508
4. Bonin A, Bellemain E, Bronken EP, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. Mol Ecol 13:3261–3273
5. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ (2004) Complex SNP-related sequence variation in segmental genome duplications. Nat Genet 36:861–866
6. Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8:1229–1231
7. Teupser D, Rupprecht W, Lohse P, Thiery J (2001) Fluorescence-based detection of the CETP TaqIB polymorphism: false positives with the TaqMan-based exonuclease assay attributable to a previously unknown gene variant. Clin Chem 47:852–857
8. Freedman ML, Penney KL, Stram DO, Riley S, McKean-Cowdin R, Le Marchand L, Altshuler D, Haiman CA (2005) A haplotype-based case-control study of BRCA1 and sporadic breast cancer risk. Cancer Res 65:7516–7522
9. von Ahsen N, Oellerich M, Armstrong VW, Schutz E (1999) Application of a thermodynamic nearest-neighbor model to estimate nucleic acid stability and optimize probe design: prediction of melting points of multiple mutations of apolipoprotein B-3500 and factor V with a hybridization probe genotyping assay on the LightCycler. Clin Chem 45:2094–2101
10. Phillips M, Meadows CA, Huang MY, Millson A, Lyon E (2000) Simultaneous detection of C282Y and H63D hemochromatosis mutations by dual-color probes. Mol Diagn 5:107–116
11. Bernard PS, Ajioka RS, Kushner JP, Wittwer CT (1998) Homogeneous multiplex genotyping of hemochromatosis mutations with fluorescent hybridization probes. Am J Pathol 153:1055–1061
12. Warshawsky I, Hren C, Sercia L, Shadrach B, Deitcher SR, Newton E, Kottke-Marchant K (2002) Detection of a novel point mutation of the prothrombin gene at position 20209. Diagn Mol Pathol 11:152–156
13. Lyondagger E, Millsondagger A, Phan T, Wittwer CT (1998) Detection and identification of base alterations within the region of factor V Leiden by fluorescent melting curves. Mol Diagn 3:203–209
14. Erali M, Schmidt B, Lyon E, Wittwer C (2003) Evaluation of electronic microarrays for genotyping factor V, factor II, and MTHFR. Clin Chem 49:732–739
15. Hawkins JR, Khripin Y, Valdes AM, Weaver TA (2002) Miniaturized sealed-tube allele-specific PCR. Human Mutat 19:543–553
16. Liebman HA, Sutherland D, Bacon R, McGehee W (1996) Evaluation of a tissue factor dependent factor V assay to detect factor V Leiden: demonstration of high sensitivity and specificity for a generally applicable assay for activated protein C resistance. Br J Haematol 95:550–553
17. Wenzel T, Elssner T, Fahr K, Bimmler J, Richter S, Thomas I, Kostrzewa M (2003) Genosnip: SNP genotyping by MALDI-TOF MS using photocleavable oligonucleotides. Nucleosides Nucleotides Nucleic Acids 22:1579–1581
18. Boekholdt SM, Sacks FM, Jukema JW, Shepherd J, Freeman DJ, McMahon AD, Cambien F, Nicaud V, de Grooth GJ, Talmud PJ, Humphries SE, Miller GJ, Eiriksdottir G, Gudnason V, Kauma H, Kakko S, Savolainen MJ, Arca M, Montali A, Liu S, Lanz HJ, Zwinderman AH, Kuivenhoven JA, Kastelein JJ (2005) Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. Circulation 111:278–287
19. Carlquist JF, Muhlestein JB, Horne BD, Hart NI, Bair TL, Molhuizen HO, Anderson JL (2003) The cholesteryl ester transfer protein Taq1B gene polymorphism predicts clinical benefit of statin therapy in patients with significant coronary artery disease. Am Heart J 146:1007–1014
20. Ayyadevara S, Thaden JJ, Shmookler Reis RJ (2000) Discrimination of primer 3′-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction. Anal Biochem 284:11–18
21. Li Y, Korolev S, Waksman G (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. EMBO J 17:7514–7525
22. Margraf RL, Mao R, Wittwer CT (2006) Masking selected sequence variation by incorporating mismatches into melting analysis probes. Human Mutat 27:269–278
23. Lyon E (2005) Discovering rare variants by use of melting temperature shifts seen in melting curve analysis. Clin Chem 51:1331–1332