Figure S7. Structural effect of the three-way DNA junction on the performance of Angler real time PCR. The PCR experiments were conducted using the tetranucleotide probe Pb4 and one of the reverse primers Pr4-Pr8 which were originally designed for the FRET probes Pb2-Pb6 of variable length (see Figure 3A). Application of these reverse primers with the tetranucleotide probe Pb4 changes the structure of the three-way DNA junction as illustrated for each of the curves (structures A-F). FAM is 6-fluorescein (marked yellow), and Q is a BHQ1 quenching dye (marked brown). The duplex-stabilizing nucleoside analogs, 2-amino deoxyadenosine and 5-propynyl deoxyuridine in the probe anchoring segment are shown by blue italic font. The nucleotides within the junction which are not involved in the base pairing are marked red. The rest of the reaction components, their concentrations and PCR cycle profile were identical to the experiments of Figure 3A, except for one important difference. The forward primer (Pr11) used in this study had the same sequence 5’-TGCGAGACATCGCATCCTGAAGCTGAG-3’ as primer Pr3 (Fig.1). However, the guanosines and thymidine (shown in bold font) within the cleavage enhancing flap sequence (underlined) were replaced with the duplex-distabilizing modifications, respectively deoxynosine and deoxyuridine. Each real time curve of this figure is an average of four identical reactions.

P1. Structural flexibility in the design of the Angler three-way junction.
Application of the reverse primers Pr4-Pr8 with the tetranucleotide probe Pb4 changes the nucleotide distance between the probe-target and probe-directing duplexes in folded Angler amplicons (structure VI, Fig.2). As illustrated in Figure S7, the gradual decrease in the number of the primer applied from Pr7 to Pr4 increases the number of unpaired nucleotides coupling the discussed duplexes in junctions B-F from 0 to 3. According to the results of Figure S7, this structural change has little, if any effect on the Angler assay performance. The structure A represents an extreme case when the probe-binding and probe-directing sites overlap by one nucleotide. Although it is not clear which of the two proposed A junctions is more stable, the negative effect of this design on the Angler assay is obvious. Structure C (Fig. S7) is the junction design used throughout the study. Similar to the Snake system’s response (27), the signal improvement of the fluorescence curve Pr11+Pb4+Pr6 compared to the corresponding curve in Fig.3A was anticipated because of the nucleoside modifications in the forward primer Pr11. The data of Figure S7 and the conclusions made need to be treated as preliminary because this experiment was done only once.
**Figure S8.** Shown are the Taqman and Angler system designs used in this project for the detection of the rs4680 SNP variation in 96 human genomic DNA samples (Figure S9). The PCR primers and FRET-probes are aligned with the target sequence in the 5'→3' orientation as indicated. The flap sequences in the Angler primers participating in the amplicon folding are slightly outlined and underlined. The 5'-terminal probes' sequences are marked in red. The A and T nucleotides marked by italic font in the Angler probe 1 were modified with 2-amino deoxyadenosine, 5-propynyl deoxyuridine, respectively. The polymorphic nucleotides in the target and nucleotides. Ideally the probes hybridize and generate a signal only with the amplicons of their full complementarity – which would be the case of absolute polymorphism discrimination. Unfortunately, the absolute discrimination is very difficult to achieve in the detection of polymorphic variations as small as single nucleotide polymorphisms.

**P2. Design of the primers and probes for detection of polymorphic variations.**

Unless the target nucleic acids are very short, like in the case of micro RNAs, the design of the primers and probes for quantitative real time PCR can be made for any sequence within the detected nucleic acid as long as this sequence is highly characteristic for the target. The numerous opportunities in the sequence selection simplify the primer and probe design for qPCR. Detection of the polymorphic variations which are uniquely positioned within the nucleic acid of interest is more difficult in this respect and requires observing a number of rules. The primers have to bind to the conserved target sequences surrounding the polymorphism of interest. As illustrated in Figure S8, the resulting amplicon needs to be kept as short as possible to ensure high PCR yield, but leaving enough space for the probe binding site in Taqman and also the amplicon folding sequences in Angler. Regardless of the technology applied (Taqman, Molecular Beacons, Angler, etc.), the FRET probes must incorporate the polymorphic nucleotides. Ideally the probes hybridize and generate a signal only with the amplicons of their full complementarity – which would be the case of absolute polymorphism discrimination. Unfortunately, the absolute discrimination is very difficult to achieve in the detection of polymorphic variations as small as single nucleotide polymorphisms.
(SNPs). The methods relying on the use of the relatively long probes (>20-25-mers), for example Taqman and Molecular Beacons, are usually the least discriminatory (18). The preferred location of the SNP polymorphism within the probe sequences which provides the best discrimination depends on the particular technology. For example, in the Taqman, Molecular Beacons and Scorpion methods which solely rely on the thermodynamic discrimination (18), the best position of the SNP nucleotide is near the center of the probe (Fig. S8). The hybridization properties of the Taqman probes may change especially in the cases of A↔G or T↔C variations. It is preferable that the probes have the same hybridization properties, which explains the three nucleotide length reduction of the G-detecting Taqman probe of Fig.S8 vs. the A-detecting probe. The same length adjustment was made in the Snake design (Fig. S9). Angler is different in this aspect, and probes have always the same length.

The SNP discrimination in Angler and its predecessor Snake approach is mainly based on the 5′-nuclease cleavage specificity, and the SNP nucleotides are preferably located second from the 5′-end of the probes (18). Due to the extreme shortness of the FRET probes in Angler, this also happens to be the center of the probe. The powerful combination of the thermodynamic (hybridization) and enzymatic (cleavage) factors explains the Angler assay superiority in target specificity and SNP discrimination (see Fig. S9 and paragraph P3). Figure S8 illustrates the most common case of the polymorphism with only two sequence variants present. Actually, the PCR technologies used in SNP genotyping can detect simultaneously more than two sequence variants as long as the probes corresponding to each of the variations can be labeled with different and instrument-distinguishable fluorescent dyes. Using combinations of excitation and emission filters, the real time instruments can presently detect simultaneously up to four-six different dyes.

During PCR, the SNP-detecting probes hybridize to their complementary amplicons and generate fluorescent signal in the respective dye channels. At the end of PCR, the fluorescence of each individual reaction in the DNA sample pool is measured and mapped out in the form of a scatter plot diagram. Examples of the diagram can be seen in Figures S9 (top row) and S10. The individual reaction data usually appear as three sample groups, two homozygous for each of the SNP variations and one heterozygous. It can be only two groups in the case of low frequency polymorphism (e.g. rs1799864 and rs16902147 in Fig. S10). The assignment of every individual sample to one of these groups determines the sample genotype. Confidence in SNP genotyping depends on the separation of the sample groups from each other, and the factors that are discussed in paragraph P3.
Fluorescein, and Yakima Yellow (Glen Research). BHQ1 is a quenching dye from Biosearch Technologies, and magnesium chloride concentration was increased to 5 mM. The reaction composition of the Snake Inc. The reaction composition and PCR profile used in the Angler experiments were the same as described respectively, abbreviations of the fluorescent dyes 6-fluorescein, 6-(4, 7, 2', 4', 5', 7'-hexachloro) fluorescein, and Yakima Yellow (Glen Research). BHQ1 is a quenching dye from Biosearch Technologies, Inc. The reaction composition and PCR profile used in the Angler experiments were the same as described in Figure 5 and Table 1 (bottom row). The Taqman reaction composition is provided in Table 1, but the magnesium chloride concentration was increased to 5 mM. The reaction composition of the Snake reaction was identical to the asymmetric Angler format (Table 1), but the enzyme concentration was reduced to 0.04 Units/µL. Other experimental details can be found in the MATERIAL AND METHODS section.

Figure S9. Detection of the rs4680 SNP variation in 96 individual human DNA samples (Sigma-Aldrich®, Panel 1) by three different technologies: Angler, Taqman and Snake. The top row of the three diagrams represents the genotyping charts for the indicated PCR methods. The fluorescence was measured at cycle 40 and plotted as a function of the signals in two reporting dye channels for every one of the 96 reactions. Black crosses mark the statistical centers of the homozygous sample groups, i.e. mean fluorescence values of the sample groups in both ordinates. Each of the three diagram rows at the bottom collectively shows 96 blue (FAM) and 96 red (YY or HEX) curves provided by real time PCR of the Angler, Taqman and Snake assays. All 96 curves of each color are distributed between three diagrams such that every diagram contains only the curves assigned to a specific SNP genotype shown on the top of the Angler diagrams. The number in parentheses indicates the number of the curves of each color in the particular diagram.

‘No target control’ reactions are not shown. The sequences of the primers and probes used in the Angler assay are listed in Table 2. The oligonucleotides used in the Taqman and Snake methods are shown in this figure. The SNP-discriminating nucleotides in the probes are underlined. FAM, HEX and YY are, respectively, abbreviations of the fluorescent dyes 6-fluorescein, 6-(4, 7, 2', 4', 5', 7'-hexachloro) fluorescein, and Yakima Yellow (Glen Research). BHQ1 is a quenching dye from Biosearch Technologies, Inc. The reaction composition and PCR profile used in the Angler experiments were the same as described in Figure 5 and Table 1 (bottom row). The Taqman reaction composition is provided in Table 1, but the magnesium chloride concentration was increased to 5 mM. The reaction composition of the Snake reaction was identical to the asymmetric Angler format (Table 1), but the enzyme concentration was reduced to 0.04 Units/µL. Other experimental details can be found in the MATERIAL AND METHODS section.
P3. Factors affecting the confidence in SNP genotyping.

Two main factors affect the confidence in SNP genotyping. First is the ‘compactness’ of the genotyping groups determined by the signal variations between the individual reactions. There are many reasons for the signal variations within the same genotyping group. For example, it can be due to inaccuracy in dispensing of the target DNA and other reaction components, amount and quality of the individual DNAs loaded into the PCR reactions, presence of PCR inhibitors in tested samples, fluctuations in time and temperature cycle profiles for some of the instruments, presence of other polymorphic variations within the target amplicons including those which are outside of the primers’ and probes’ binding site, etc. The second and perhaps most influential factor affecting the confidence in SNP genotyping is the distance between the genotyping groups. Unlike the first factor (group compactness), this second factor is mainly controlled by the detection technology applied, e.g. Taqman, Molecular Beacon, Scorpion, Snake, Angler, etc. The distance between the genotyping groups, in turn, depends on (i) the ability of a particular technology to discriminate SNPs and (ii) the signal strength provided by the technology.

**Discrimination.** As illustrated in the Figure S9 (top diagram row), the SNP discrimination can be measured by the angle between the centers of the homozygous groups (mean fluorescence values) and the coordinate datum (X=Y=0). In the case of the absolute discrimination, this angle is 90°. The Angler assay is one of the best technologies with respect to the SNP discrimination. For example, analysis of the SNP scatter plots of Figure S10 indicates that the technology is commonly reaching (rs10505477 and rs2070673) or closely approaching (rs2246945, rs3843549, rs9643226, rs134428, and rs1946234) the theoretical maximum in SNP discrimination. Actually, the rs4680 polymorphism shown in Figure S9 is least discriminated by the Angler assay in the pool of 10 SNP tested (Fig.S10). However, even in this case the technology excels, with its competitors providing the discrimination angle of 78° vs. 60° in Snake and Taqman.

**Fluorescent signal.** The fluorescent signal generated by a particular real time method is yet another factor separating the SNP genotyping groups. The three top diagrams in Figure S9 illustrate the signal effect. The Snake technology was shown to supersede all other methods in the signal productivity, commonly resulting in complete probe cleavage (18). Although, in comparison to the Angler assay, the Snake system turned to be list discriminatory in detection of rs4680 SNP (Figure S9), but in this case unambiguous genotyping was ensured by the strong signal distancing and separating the genotyping groups from each other. Low fluorescent signal combined with poor SNP discrimination can be a problem for making proper genotyping calls. Especially problematic are the data outliers which are commonly difficult to assign to any specific genotyping group because of their intergroup location. An example of this potentially ambiguous data point or outlier is shown in the Taqman diagram of Figure S9.

Appearance of the outliers requires repeating the experiment as many times as necessary to reinstate the confidence in genotyping and therefore it can be time consuming and costly. The data points somewhat deviated from the main clusters can be seen in the Snake and Angler scatter plots (Fig. S9), but they are not problematic for proper assignment because of the good separation between the genotyping groups.
Figure S10. Results of genotyping of 96 individual human DNA samples on 10 single nucleotide variations (SNPs). Real time curves of these genotyping experiments are shown in Figure S11. The scatter plots display the individual reaction fluorescence values at the end of PCR (cycle 40) in two reporting dye channels. The right column represents the original 96-well plate format of Human Random Control DNA Panel from Sigma-Aldrich® (Panel 1, cat.№ HRC1-1EA). The SNP numbers as well as the abbreviated probe structures are shown on the top left side of each diagram row. The SNP-discriminating nucleotides are underlined. The nucleotides shown in bold font represent base modifications, 2-amino deoxyadenosine and 5-propynyl deoxuridine. **FAM, HEX** and **YY** are, respectively, abbreviations of the fluorescent dyes 6-fluorescein, 6-(4, 7, 2', 4', 5', 7'-hexachloro) fluorescein, and Yakima Yellow (Glen Research). **Q** is a BHQ1 quenching dye (Biosearch Technologies, inc.) whereas **R** is a 3'-anchoring sequence. Complete sequences of the primers and probes used in each individual SNP case are listed in Table 2 and Table S3. Details of the PCR experiments can be found in the MATERIAL AND METHODS section. The homozygous (homo) and heterozygous (hetero) results are marked by a specific color, blue and yellow for homozygous and green for the heterozygous samples. **NTC** means ‘no target control’ reactions shown as black dots. The diagram A, B and C are the same shown in Figure 5.
Figure S11

rs2246945

C/C (6)  
A/C (44)  
A/A (46)

rs3843549

G/G (2)  
G/A (24)  
A/A (70)

rs16902147

C/C (0)  
C/T (3)  
T/T (93)

rs10505477

A/A (30)  
A/G (44)  
G/G (22)

rs2070673

T/T (68)  
T/A (25)  
A/A (3)

rs4680

A/A (24)  
A/G (58)  
G/G (14)
Figure S11. Real time curves obtained during genotyping by the Angler assay of 96 individual human DNA samples on 10 single nucleotide variations (SNPs). The fluorescence values of these curves at cycle 40 were used in preparing the scatter plots of Figure S10. Every row comprises three diagrams and collectively shows 96 blue (FAM) and 96 red (YY or HEX) curves for each genotyped SNP. The SNP numbers are located on the left side of the diagram rows. All 96 curves of each color are distributed between three diagrams such that every diagram contains only the curves assigned to a specific SNP genotype. The SNP genotypes are shown on the top of each diagram. The number in parentheses indicates the number of the curves of each color which also corresponds to the number of the individual DNA samples of the specific genotype. NTC means ‘no target control’ reactions (total 10) shown as green curves. The NTC curves in the left diagram of the row represent reading in the HEX or Cy3 channels of the instrument whereas the right diagram shows the NTC curves obtained in the FAM channel. FAM (blue), HEX and YY (red) are, respectively, abbreviations of the fluorescent dyes 6-fluorescein, 6-(4, 7, 2', 4', 5', 7'-hexachloro) fluorescein, and Yakima Yellow (Glen Research). Complete sequences of the primers and probes used in each individual SNP case are listed in Table 2 and Table S3. The reaction composition and PCR profile used in the Angler experiments were the same as described in Figure 5 and Table 1 (bottom row). Details of the PCR experiments can be found in the MATERIAL AND METHODS section.
### Table S3. Structures of primers and probes used in the SNP-genotyping study of Figure S10*

| SNP      | PCR function | Oligonucleotide sequence and modifications** |
|----------|--------------|--------------------------------------------|
| rs2246945| F primer     | 5'-ATGAAATTATAGTGTGTTGGAATGCATGGAACATTT-3' |
|          | R primer     | 5'-CGGCCTTGTATTTGCGCAATACCTGCTGTACAGGCC-3' |
|          | Probe 1      | 5'-(FAM)-GAC-(BHQ1)-ATTACGGCGCCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(YY)-GCAA-(BHQ1)-ATTACGGCGCCGp-O(CH$_2$)$_2$-OH-3' |
| rs3843549| F primer     | 5'-AGGAGGATATGAAACTCTCT-3'                     |
|          | R primer     | 5'-CGGGCGGCTTTTATACCTGCTGCTGCCT-3'           |
|          | Probe 1      | 5'-(FAM)-AGAT-(BHQ1)-TAAAGGCAGGCCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(YY)-AATAT-(BHQ1)-TAAAGGCAGGCCGp-O(CH$_2$)$_2$-OH-3' |
| rs16902147| F primer     | 5'-AATTTTCCTTTAAGCATACAGCATATTT-3'            |
|          | R primer     | 5'-CGGCCGCTTATATAATGCTGACCTGTCCCTCCTG-3'     |
|          | Probe 1      | 5'-(FAM)-ATCA-(BHQ1)-ATAACGGCGCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(YY)-AATC-(BHQ1)-ATAACGGCGCGp-O(CH$_2$)$_2$-OH-3' |
| rs13428  | F primer     | 5'-TTGTTGAAAAAGACACTCTGTTGACACTTAAATCCTAGG-3' |
|          | R primer     | 5'-CGGCCGCTTTAAACGGCTGTAAAATGCTGACAGGCCAAG-3' |
|          | Probe 1      | 5'-(FAM)-GCTC-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(HEX)-CGTC-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
| rs1799864| F primer     | 5'-TACCCAGATCTGCGCTACTCTGCTTGTTT-3'          |
|          | R primer     | 5'-CGGCCGCTTTAAATGCTGACAGGCCAAG-3'           |
|          | Probe 1      | 5'-(FAM)-AATC-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(HEX)-CGTC-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
| rs1946234| F primer     | 5'-TATAGCTGCGCTACTCTGCTTGTTT-3'              |
|          | R primer     | 5'-CGGCCGCTTTAAATGCTGACAGGCCAAG-3'           |
|          | Probe 1      | 5'-(FAM)-CAGT-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(HEX)-CGTC-(BHQ1)-TTATCGCCGCCGp-O(CH$_2$)$_2$-OH-3' |
| rs2070673| F primer     | 5'-ACTGCGCCCTCCTGCTGTTGCTT-3'                |
|          | R primer     | 5'-CGGCCGCTTTAAATGCTGACAGGCCAAG-3'           |
|          | Probe 1      | 5'-(FAM)-GACG-(BHQ1)-ATAACGGCGCGp-O(CH$_2$)$_2$-OH-3' |
|          | Probe 2      | 5'-(HEX)-GTCT-(BHQ1)-ATAACGGCGCGp-O(CH$_2$)$_2$-OH-3' |

*The primer and probe sequences used in the detection of the other three SNPs are provided in Table 2.
**The cleavage-enhancing flap sequences in forward (F) primers and the probe-directing sequences in reverse (R) PCR primers are underlined. The 2-amino deoxyadenosine and 5-propynyl deoxyuridine duplex-stabilizing modifications used in the design of the A/T-rich Angler probes are marked by bold italic font. **FAM** is 6-fluorescein, **HEX** is 6-(4', 7', 2', 4', 5', 7'-hexachlorofluorescein), and **Y** is Yakima Yellow fluorescent dyes (Glen Research). **BHQ1** is Black Hole Quencher 1 from Biosearch Technologies, Inc. In all cases, the second nucleotide (underlined) from the 5'-end of the FRET probe was the SNP detecting one. All Angler probes incorporated a 1,3-propanediol moiety conjugated to the 3'-terminal phosphate to prevent the accidental extension during PCR. Marked by brown color is the same probe used in two different assays.
Table S4. SNP allelic frequencies determined in present and other studies

| SNP     | Allelic frequencies (Anger assay) | Allelic frequencies determined in other genotyping studies* | Allelic frequencies |
|---------|----------------------------------|----------------------------------------------------------|---------------------|
|         |                                  | Project | Method | Allelic frequencies | Caucasian | Multiracial |
|---------|----------------------------------|---------|---------|---------------------|-----------|-------------|
| rs2246945 | A(0.708) C(0.292) | HAPMAP | Taqman | A(0.729) C(0.271) | A(0.679) C(0.321) |
|         |                                  | SNP500  | Taqman | A(0.741) C(0.259) | A(0.658) C(0.342) |
| rs3843549 | A(0.854) G(0.146) | Pop1    | Sequencing | — | A(0.787) G(0.213)** |
| rs16902147 | T(0.984) C(0.016) | Pop1    | Sequencing | — | T(0.936) C(0.064)** |
| rs10505477 | A(0.542) G(0.458) | HAPMAP | Taqman | A(0.449) G(0.551) | A(0.541) G(0.459) |
|         |                                  | SNP500  | Taqman | A(0.444) G(0.556) | A(0.551) G(0.449)** |
|         |                                  | Pop1    | Sequencing | — | A(0.487) G(0.513)** |
| rs9643226 | C(0.135) G(0.865) | Pop1    | Sequencing | — | C(0.167) G(0.833)** |
| rs13428   | C(0.625) G(0.375) | SNP500  | Taqman | C(0.532) G(0.468) | C(0.559) G(0.441) |
|         |                                  | SNP500  | Sequencing | C(0.532) G(0.468) | C(0.559) G(0.441) |
| rs1799864 | A(0.052) G(0.948) | SNP500  | Unknown*** | A(0.129) G(0.871)** | A(0.152) G(0.848)*** |
|         |                                  | HAPMAP  | Unknown*** | A(0.108) G(0.892)** | — |
| rs4680   | A(0.552) G(0.448) | HAPMAP  | Taqman | A(0.517) G(0.483) | A(0.336) G(0.664) |
|         |                                  | SNP500  | Taqman | A(0.467) G(0.533) | A(0.391) G(0.609) |
|         |                                  | HDP     | Taqman | A(0.477) G(0.523) | A(0.396) G(0.604) |
| rs1946234 | A(0.844) G(0.156) | SNP500  | Taqman | A(0.903) G(0.097) | A(0.855) G(0.145) |
|         |                                  | SNP500  | Sequencing | A(0.893) G(0.107) | A(0.853) G(0.147) |
| rs2070673 | A(0.161) T(0.839) | SNP500  | Taqman | A(0.145) T(0.855) | A(0.377) T(0.623) |
|         |                                  | SNP500  | Sequencing | A(0.145) T(0.855) | A(0.377) T(0.623) |

*Allelic frequencies of the other genotyping studies were taken from a National Cancer Institute web site: [http://variantgps.nci.nih.gov/cgseq/pages/sequenceSubmit.do?method=sequence&regionId=1](http://variantgps.nci.nih.gov/cgseq/pages/sequenceSubmit.do?method=sequence&regionId=1) (19).

** The information about racial origin of the Pop1 sample pool was not available.

***These data were obtained from dbSNP web site: [http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1799864](http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1799864)
Figure S12. Shown are fragments of the targeted sequences in the human genome incorporating the detected SNP variations (bold in square brackets). Other polymorphic variations are marked by red font. The PCR primers are aligned with the target sequence in 5'→3' orientation as indicated. The cleavage enhancing and probe directing flanking sequences are slightly outlined. R are the corresponding probe luring sequences which are not shown here. Complete sequences of the PCR primers can be found in Tables 2 and S3. The nucleotide marked by blue color (rs16902147) is a sequence ‘divider’ that is used for the purpose explained in paragraph P6 (below).
**P4. Design of the 3’-anchoring tails for a complete inventory of tetranucleotide FRET probes (256 variations).**

The selection of the probe tail sequences should follow simple and straightforward rules. First, the tails should be as short as possible. Thus, the preference should be given to the 9-10-mer G/C-rich oligonucleotides which form duplexes with melting temperatures in a range of ~55-62 °C. Second, in order to be applicable for SNP detection, all four tetranucleotide probes of an identical sequence except the second nucleotide from the 5’ end should have the same anchoring tail sequence. Each of the 4 variations at the second nucleotide should be labeled with one of four instrument-distinguishable dyes. One benefit of following these rules for the Angler assay could be its application for multiplex detection which should be at least as good as other conventional technologies. Taking this into account, each of the 64 SNP-detecting probe subfamilies should carry the 3’-anchoring tails of different sequences, in order to prevent the different probe competition for binding to the same 5’-flap luring sequence of the amplicons. In theory, a nonanucleotide comprising only G and C nucleotides has \(2^9 = 512\) sequence variations. It may not look difficult to select the required 64 tail sequences out of this 512 member pool. However, the extremely G and C-rich sequences as well as self-complementary oligonucleotides need to be avoided. Moreover, it would be beneficial for the assay to exclude the cross-complementarity between the tails of the different probes. In this regard, the tail selection can be simplified by random insertion of a single A or T nucleotide throughout the core G/C-rich sequence. Such an insertion does not affect much the tail hybridization properties, but greatly increases the tail sequence variability. Examples of the proposed tail sequences are shown in Table S5. Since the tail sequences are G/C-rich, the individual tail assignment for each subfamily of four SNP-detecting probes should be started from the G/C-rich probes. The obvious goal is to avoid complementarity between the probe and its 3’-tail, both of which are G/C-rich. This intermolecular complementarity can lead to the probe ‘self-cleavage’ during PCR generating undesirable fluorescence background. Fortunately, there are no problems in the tail assignment for the A/T-rich and also mixed sequence probes.
Table S5. Examples of 3’-tail sequences that can be used in design of the FRET probes.

| Sequence 5’→3’ | T_m, °C | Sequence 5’→3’ | T_m, °C | Sequence 5’→3’ | T_m, °C |
|---------------|---------|---------------|---------|---------------|---------|
| CGGCAGCCCCC   | 53      | CGCCGGCGGC    | 55      | CGGCGGCGC     | 53      |
| CAGGCGGCCCC   | 53      | CAGCGGGGCGC   | 55      | CTGGCGGCGGC   | 53      |
| CGACGGGCCCC   | 54      | CGACGGGCGC    | 55      | CGTAGCGGCGC   | 55      |
| CGGCAGGCCCC   | 53      | CGCAGGGCGC    | 55      | CGGTGCGGCGC   | 55      |
| CGGCCAGCCC    | 54      | CGGGAGGGCGC   | 55      | CGGCTGGGCGC   | 53      |
| CGGGCGACC     | 55      | CGGGGGAGGC    | 55      | CGGGCCTGGGC   | 55      |
| CGGGCGGCC     | 55      | CGGGGGGCAGC   | 55      | CGGGGGHYGC    | 55      |
| CGGGCGGCCAC   | 55      | CGGGGGGCAGC   | 55      | CGGGGGGGTC    | 53      |

*T_m* is an abbreviation for melting temperatures which were calculated and rounded to an integer for the fully complementary duplexes at 200 nM concentrations in 40 mM NaCl and 5 mM MgCl₂ using OligoAnalyzer 3.1 from Integrated DNA Technologies (https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/).

P5. Price range of the FRET probes obtained from market inventories and custom order.

Regardless of the recent progress in methods, instrumentation and overall oligonucleotide manufacturing, the FRET probes remain as one of the most expensive components of real time PCR. Many factors such as the nucleotide length, scale of synthesis, difficulties in purification, special modifications used in the production of certain probes, cost of the labeling dyes and intellectual property associated with some proprietary detection technologies, etc. contribute to the overall market prices of the FRET probes. The following price comparison reflects the minimum amounts guaranteed by the providers for the custom order of the least expensive Taqman probes commonly labeled with 6-fluorescein (FAM). For example, Integrated DNA Technologies, Inc. sells 5 nmole of a probe labeled with 5’-FAM and 3’-Iowa Black quencher for $229. The 5’-FAM or VIC labeled probes from Life Technologies (6 nmole amount) are more expensive ($263 per a probe), but these FRET probes are well known for their superior performance in real time PCR because of the 3’-MGB technology applied (15). As a result of ‘price war’ between the market providers, Biosearch Technologies, Inc. presently offers a 10 nmole amount of a FRET probe for as little as $95. However, this substantial price reduction applies only to the probes labeled with 5’-FAM and 3’-BHQ1 quencher. For any other dye combination, the FRET probe price for the same amount is more than doubled.

The biotechnology industry has been always under pressure to make nucleic acid detection less costly. Building up the oligonucleotide inventories and establishing fractional sale is a straightforward way to achieve that. However, this is possible only in special cases of relatively high demand from the research community to detect and analyze the same target sequences. For example, Life Technologies has presently
established an inventory of oligonucleotide components for detection of as many as 199,756 human genes. Each kit for detection of a human gene includes a MGB-Taqman probe (FAM or VIC labeled) and two primers in amounts sufficient to run 75 PCR reactions for $82. This represents a substantial price break for the low volume users. The savings can be greater, but taking into account the company expenses associated with the assays’ validation, costly MGB technology and the maintenance of such an enormously sized inventory, the kit price seems well justified. Integrated DNA Technologies, Inc. is presently establishing a competing inventory for the same purpose. The kits comprising 0.5 nmole of FAM-labeled probe and 1 nmole of primers can be obtained for $75.

The duplex-stabilizing MGB technology (15) allows significant reduction of the FRET probes’ length, which in turn enables the detection of the nucleic acid sequences as short as micro RNAs (~18-20-mers). An advantage of the micro RNA targets is the relatively low number of their sequence variations. Life Technologies established an inventory of kits for qPCR detection of 1,818 human micro RNAs. Each kit is sufficient to run 50 reverse transcription and 150 PCR reactions for $175. The kits includes an RT-primer, two PCR primers and a MGB-Taqman probe with all other necessary reagents like buffers, enzymes and nucleoside 5’-triphosphates.

Roche Molecular Diagnostics explores yet another duplex-stabilizing LNA technology (16). The company is currently marketing a ‘Universal ProbeLibrary’ comprised of 165, predominantly G/C-rich 8-9-mer Taqman probes labeled with 5’-FAM and containing insertions of LNA nucleotides for elevated hybrid stability. The detection system design starts from finding a complementary site within a sequence of interest for one of the FRET probes from the 165-member inventory. Once the site is found, then two PCR primers surrounding the site can be selected. Unfortunately, this 165 probe inventory can be used only for quantitative PCR analysis (qPCR) of fairly long sequences with a high probability of finding at least one binding site for the probes. It is useless for the detection of the relatively short targets like micro RNAs and does not cover the detection of polymorphic variations. Another disadvantage of this product is its market price. The LNA-stabilized FRET probes are presently sold in amounts of 2.5 nmole for $185. Actually, this price does not look much different from the custom order of the conventional Taqman probes (see above).

All discussed inventories enable the fractional sale and therefore the sensible price reduction of the FRET-probes for the majority of researchers who need limited amounts of nucleic acids analysis. However, many of the inventories are limited to humans and a few of the most common laboratory animals. The inventories for detection of the polymorphic variations are exceptionally rare and limited in size (e.g. 2,700 human SNPs, Life Technologies). By its definition, a true universal library of FRET probes, like the one enabled by the Angler assay, would contain a complementary component for the detection of all possible sequences and sequence variations (SNPs) in the entire kingdom of life including humans, animals, plants, microbes, viruses, etc.
P6. Main principles and rules for the probes’ and primers’ design in the Angler assay.

Since no Angler-designated software is presently available, the initial primer selection (only genomic parts without flaps) can be made using any other appropriate program for real time PCR. The targeted T\text{m} range is 65-67 °C in 40 mM NaCl, 5 mM MgCl\textsubscript{2}. The buffering component, 20 mM Tris-HCl (pH8) should be used in PCR, but not counted in the primers’ T\text{m} estimates. Preferably the amplicons should be as short as possible. However, the target sequence between the primers’ binding sites should be long enough to accommodate the binding sites for the probe, cleavage enhancing and probe directing flaps (at least ~25-35 nucleotides depending on the nucleotide content). A small overlap of ~2-3 nucleotides between the primer and flap binding sites is allowed. A longer overlap can lead to the primer self-extension during PCR. Examples of positioning of the primers’ and flaps’ binding sites within the SNP-detected targets are shown in Figure S12.

During the detection step, the Angler amplicons fold into the secondary structure V (Fig. 2). On one hand, this structure has to be stable at 56 °C to serve its purpose, i.e. capturing the FRET probe and directing it to the target-binding site. On the other hand, the amplicons should unfold during the following amplification step at 72 °C in order to speed up the primers’ annealing and extension. Apparently, there is an optimal range in the stem-loop structure stabilities which is primarily defined by the detection and amplification temperatures. No attempt was made in this study to determine the upper and lower stability limits. Nevertheless, the results of the present study have proved that the undertaken approach for the flap sequence selection is somewhere between those limits. Table S6 summarizes the hybridization properties of flap sequences used in the SNP study. The melting temperatures were calculated for corresponding hairpin structures and also for the duplexes assuming that the flap sequences are not linked to the primers. It is more difficult to accurately predict the hairpin melting temperatures in comparison to the corresponding duplex values. Therefore, in this study, the flap length adjustment was done based on the duplex stability estimates. In rare cases, the 5’-terminal nucleotides of the genomic part of the primer can be complementary to the target at the end of the flap binding site, i.e. become a part of the flap sequence. In order to avoid the potential flap over-stabilization, it is recommended to insert a single non-complementary nucleotide (divider) between the flap and genomic sequences of the primer. An example of this was found during design of the reverse primer for the detection of SNP rs16902147 (Figure S12).
Shown is only the flap sequence participating in the stem formation of the amplicon folding. Single uncomplementary 5’-nucleotides in the cleavage enhancing flaps are not shown. Full sequences of the primers can be found in Tables 2 and S3.

**Melting temperatures ($T_m$) of the corresponding hairpins were calculated for 40 mM NaCl and 5 mM MgCl$_2$ using the DNA folding program of the mfold web server (http://mfold.rna.albany.edu/?q=mfold/dna-folding-form). The probe luring flap sequences was ignored and unused in the calculations.

***These melting temperatures were calculated for the fully complementary duplexes at 200 nM concentrations in 40 mM NaCl and 5 mM MgCl$_2$ using OligoAnalyzer 3.1 from Integrated DNA Technologies (https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/).

| SNP        | Forward primer: cleavage enhancing flap | Reverse primer: probe directing flap |
|------------|----------------------------------------|-------------------------------------|
|            | Flap sequence* | Hairpin $T_m$ ** $^\circ$C | Duplex $T_m$ *** $^\circ$C | Flap sequence* | Hairpin $T_m$ ** $^\circ$C | Duplex $T_m$ *** $^\circ$C |
| rs2246945  | 5’…TGAATTATAGTGTTC… | 63 | 44 | 5’…TTGGGGCAAC… | 72 | 44 |
| rs3843549  | 5’…GCAAGATAATGA… | 60 | 39 | 5’…AATGACCTTTGTG… | 63 | 41 |
| rs16902147 | 5’…ACTATTCTTTTTA… | 59 | 36 | 5’…TAAAAATGTCATTC… | 66 | 41 |
| rs10505477 | 5’…GGAGATGAAG… | 61 | 34 | 5’…AGGCTTCCTG… | 71 | 41 |
| rs9643226  | 5’…TTTGATCTG… | 65 | 37 | 5’…TTCTAGGTCATAG… | 97 | 41 |
| rs13428    | 5’…TGTGAAAAAG… | 62 | 34 | 5’…ACGTCTCAAAT… | 59 | 40 |
| rs1799864  | 5’…ACAGCATGTAT… | 66 | 41 | 5’…TCATCTTAATAAC… | 58 | 40 |
| rs4680     | 5’…CCAGCGAAAAT… | 71 | 40 | 5’…AGGACACGGT… | 72 | 39 |
| rs1946234  | 5’…ATTAGTGATCTC… | 64 | 37 | 5’…CCATCATTACAA… | 62 | 39 |
| rs2070673  | 5’…CTGCCCCTTT… | 63 | 35 | 5’…TACCTCACCC… | 72 | 39 |

*Shown is only the flap sequence participating in the stem formation of the amplicon folding. Single uncomplementary 5’-nucleotides in the cleavage enhancing flaps are not shown. Full sequences of the primers can be found in Tables 2 and S3.

**Melting temperatures ($T_m$) of the corresponding hairpins were calculated for 40 mM NaCl and 5 mM MgCl$_2$ using the DNA folding program of the mfold web server (http://mfold.rna.albany.edu/?q=mfold/dna-folding-form). The probe luring flap sequences was ignored and unused in the calculations.

***These melting temperatures were calculated for the fully complementary duplexes at 200 nM concentrations in 40 mM NaCl and 5 mM MgCl$_2$ using OligoAnalyzer 3.1 from Integrated DNA Technologies (https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/).