**Supplementary Information**

# Conditional Displacement Hybridization Assay for Multiple SNP Phasing

Tsz Wing Fan†, Henson L. Lee Yu† and I-Ming Hsing †‡*

†

†Department of Chemical and Biomolecular Engineering and ‡Division of Biomedical Engineering, The Hong Kong University of Science and Technology, Hong Kong

*Email: kehsing@ust.hk

**Abstract:** We describe control experiments showing that the changes in the fluorescence signal is solely due to the presence and absence of SNPs, and not due to other matrix effects (S1). Also, the results of the optimization of the concentration used and SNP position are shown in S2 and S3, respectively. S4 explains in greater detail why an observed drop in green fluorescence when the enzyme is added. S5 tabulates the sequences used in this experiment.

## S1. The effect of dNTP, polymerase, and BSM buffer to the red and green fluorescence signal

In addition to the fluorescence measurement at template concentrations of 100 nM using the plate reader, we also conducted test for 50 nM of the target DNA, as shown in **Figure S1A**. Three separate experimental trials were carried out for each haplotype. From which, the average fluorescence and standard deviation were used to construct the curves and error bars shown on the graph. Results were in agreement with the amount and phase of SNPs present on the template, i.e., templates containing SNP A (i.e. TD, TA) gave a much higher red fluorescence signal and those that contain SNP B (i.e. TD, TB) gave a much higher green signal during the first 30 minutes of incubation. Furthermore, after the addition of enzymatic mixture, a significant drop in red fluorescence was seen only for TD, and not for TA, which means that the phase information can be determined in the diplotype setting (i.e., when pairs of the target strands are used). **Figure S2** shows an agarose gel visualization of the products when the different target strands reacted with PA and PB, and after 30 minutes of incubation with BSM polymerase. The bands prior to the addition of enzyme show that TD formed a product with either the fluorophore strand of PA (PA_F), PB_F, or both and therefore showed the highest intensity, while the WT showed the lowest intensity because the bands are only generated due to non-specific binding to any of the PA_F or PB_F strand (see scheme **Figure S6B**). The lower intensity of TA + PA_F compared to TB + PB_F may be attributed to the combined fluorescence of the green fluorophore labelled onto PB and the chelating agent for gel visualization. After the addition of the enzyme, the non-zero, but significantly lower, green fluorescence signals for TA and WT were confirmed by the significantly thinner bands in the 165-bp position that corresponds to the polymerase product. In the case for TA, there was still a visible band slightly higher than the 100 bp molecular ladder corresponding to a partial double stranded target-PA_F structure, that could not be extended by the BSM polymerase because of the poly A tail. This band was also present to a smaller extent for WT, which explains the non-zero red fluorescence signal generated in such case.

To prove that the drop in fluorescence (red channel) was not caused by the system components, we also conducted tests by adding either the enzymatic buffer, polymerase, or dNTP alone at the 31st minute, but keeping all other conditions the same. As shown in **Figure S1C**, negligible change in fluorescence was seen for all these cases, and only in the presence of both polymerase and dNTP will the drop happen in TD (and in TB, but to a lesser extent). Similar control experiments for the green fluorescence channel were also shown here (**Figure S1B**) although their fluorescence signal after 30 minutes would not be

used to interpret the phase information. The change of the green fluorescence after the enzyme addition (i.e., a sharp drop followed by a gradual increase in signal) will be discussed in detail in a latter section.

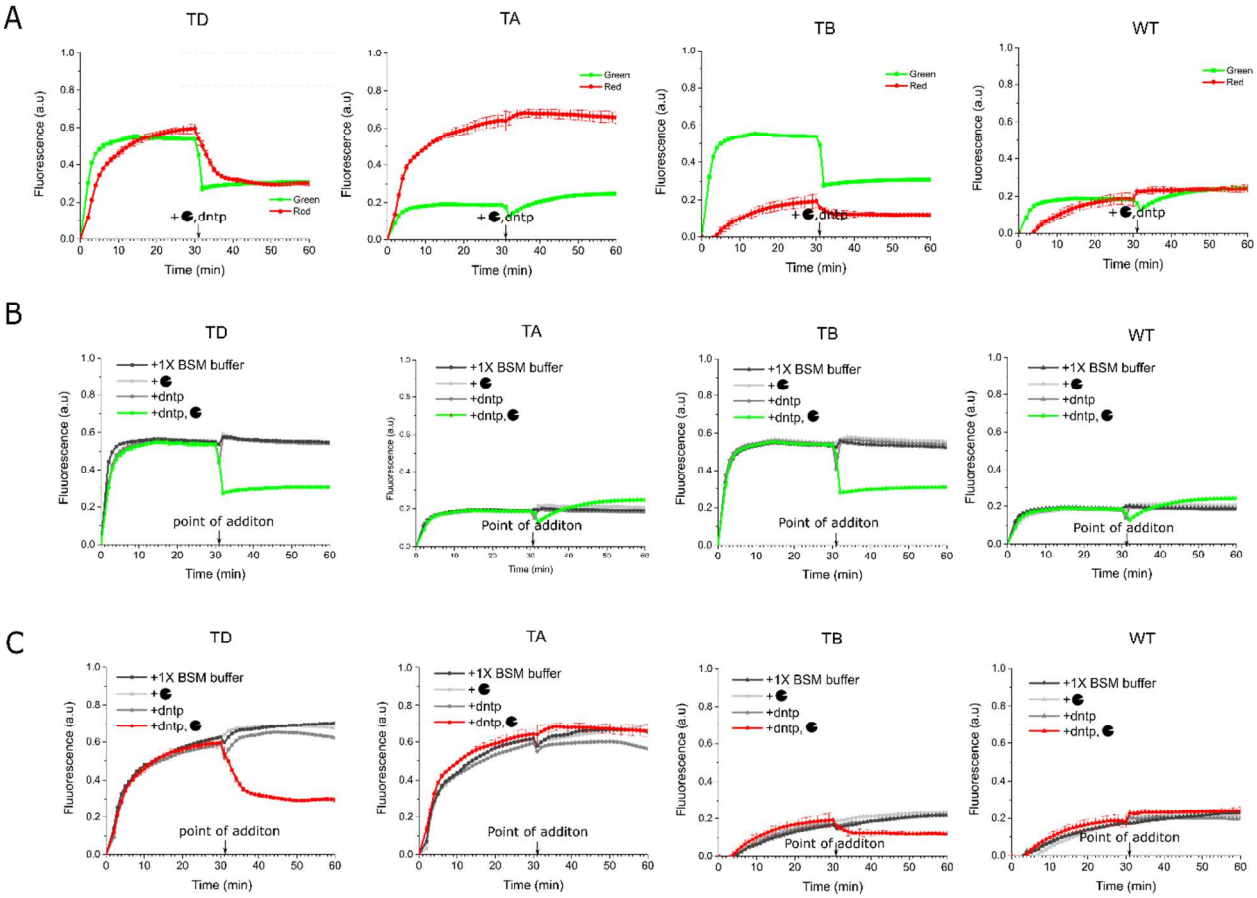



**Figure S1**. A) The fluorescence readouts of the four haplotypes (TD, TA, TB and WT) at concentrations of 50 nM measured by the microplate reader. Template (50 nM) and probes (100 nM) were incubated for the first 30 minutes; while enzymatic mixture was added at the 31st minute (indicated by the arrow) and the resultant mixture were incubated for another 30 minutes of reactions. Curves showed the average taken from three experimental trials and the error bars on the curves indicated their standard deviation. B) and C) were the control experiments for the green and red fluorescence channels respectively at the same template and probe concentrations, but either with the enzymatic buffer, BSM polymerase or dNTP alone, showing that the change in fluorescence solely occurred when both dNTP and polymerase were present. Fluorescence values shown here were all normalized by setting 0 as the background given by the probes alone, while 1 was the maximum fluorescence given by the TD/TB (green channel) and TD/TA (red channel) at 100 nM before enzyme addition obtained from the curves in Figure 2A.

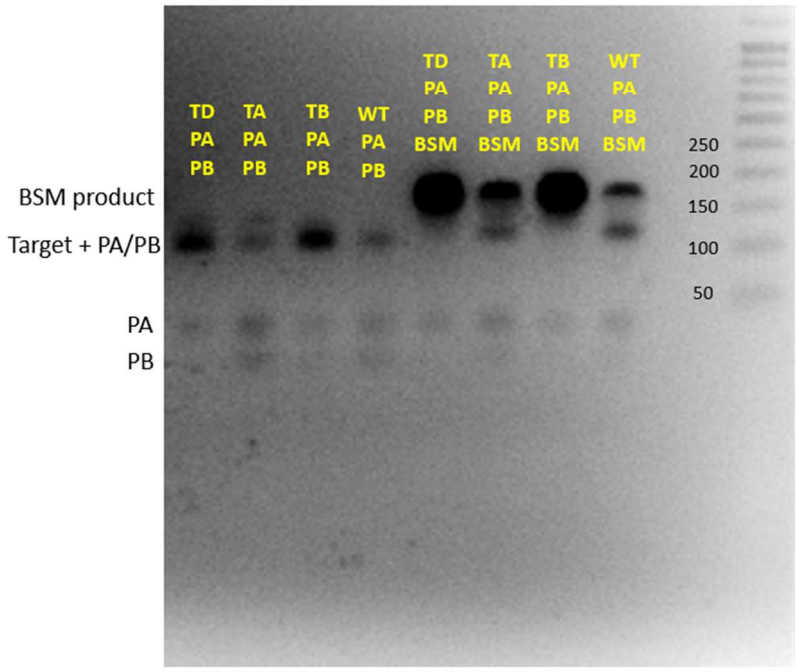



**Figure S2.** Visualization of the product before (lanes 1-4) and after (lanes 5-8) the addition of enzymes, and a 50-bp molecular ladder in lane 10. The product in lanes 1-4 were from the addition of 100 nM each of the target, PA, and PB, and incubated for 30 minutes. Lanes 5-8 were the result of the similar mixture but with 5 nmol of dNTP and 0.8 U of BSM polymerase for an additional 30 minutes. Brightness was adjusted to show the bands corresponding to the unhybridized probes (PA and PB). Lanes 1-4 verified the formation of the target-fluorophore strand complex, and lanes 5-8 confirmed that PB preferentially binds to TD and TB, forming the 165-bp product more favorably than TA and WT. The band in the ~120bp position after the enzyme addition corresponded to the partial double-stranded TA + PA_F complex.

## S2. Haplotype testing at different template concentrations

**Figure S3** shows the real-time fluorescence measurement for the four templates (or four different haplotype combinations of the two SNPs) at concentrations 0.25x, 0.5x, 1x, and 2x relative to the probes. The two fluorescence channel readouts correctly reflected the relative amount and the type of SNPs present in each of the samples during the first 25 minutes, or the first stage of reactions when the probes (PA and PB) were incubated with the respective templates. For example, in TD and TB where SNP B was present, they showed a relatively higher green fluorescence than in TA or WT; while in TD and TA, where SNP A was present, they gave a higher red fluorescence than in TB/WT. The fluorescence signals in the presence of the corresponding SNPs also increased with the increase in template concentrations. In the second stage of reactions, the addition of polymerase caused a significant drop for TD, but none

or minimal drop in TA at all concentrations tested. This demonstrated the capability of CDHA in obtaining phase information at template concentrations up to 2 times thatof the probes. As seen from the graphs, both hybridization and enzymatic reactions reached equilibrium soon after 10 minutes the reactions took place. This allowed us to take the average of the saturated fluorescence values for comparison. For example, the fluorescence values (of the red and green channels) at the $11^{th}$ – $25^{th}$ minutes were used for SNP quantification; while the red fluorescence values at $36^{th}$ -$50^{th}$ minutes were used for measuring the signal drop which was calculated by the difference to the average red fluorescence at $11^{th}$-$25^{th}$ time points to infer the phase of SNPs. Average fluorescence values were shown in Figure 2C of the main manuscript.

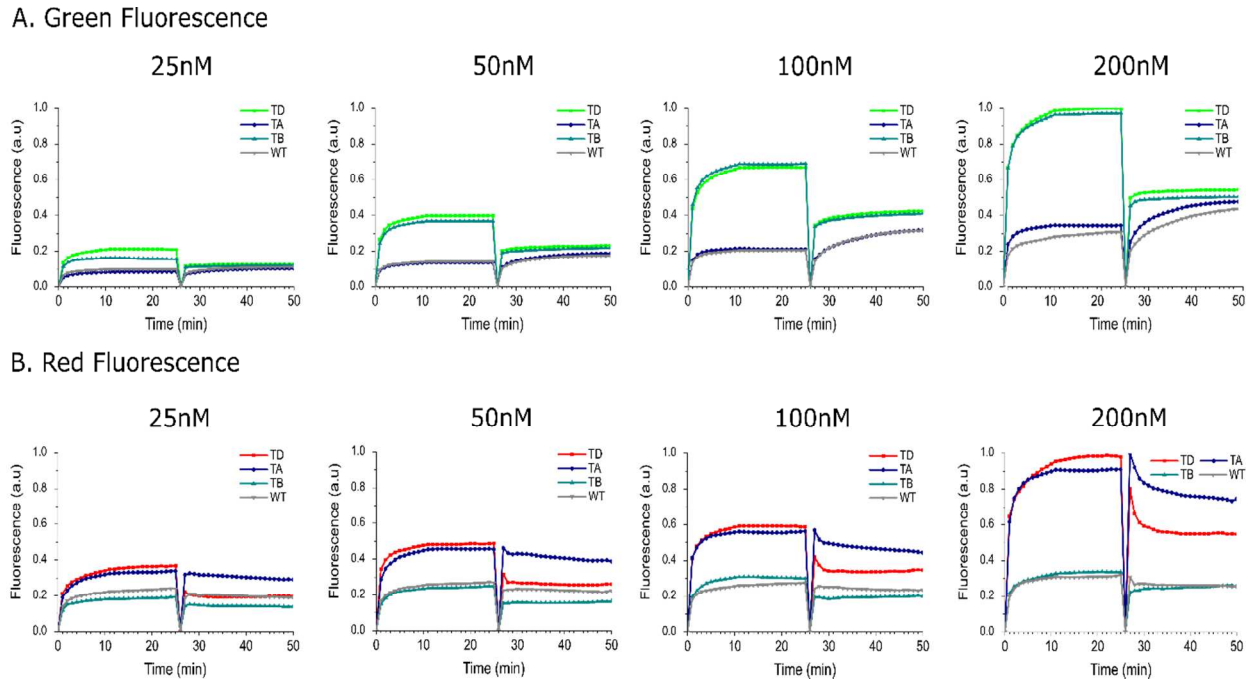


**Figure S3.** Real-time spectrofluorometric measurement of the four haplotypes (TD, TA, TB, and WT) at different target concentrations (25 nM, 50 nM, 100 nM, and 200 nM) for A) green and B) red fluorescence channels. 100 nM probes were allowed to incubate with the respective haplotype at concentrations 0.25x, 0.5x, 1x, or 2x that of the probes for 25 minutes. Afterwards, at the 26th minute, the cuvette was taken out from the chamber (as seen by the drop in fluorescence) and the enzymatic mixture (BSM polymerase and dNTP) was added. Reactions was taken for another 25 minutes. Curves shown here were normalized by taking the average fluorescence given by the probes alone (the negative control) as zero; while the maximum fluorescence given by 200 nM template (i.e. TD/TB for green channel; TD/TA for red channel) was set as one.

**S3. Comparison of Probe B with SNP identification sites on toehold and branch migration region**

The design of Probe B is important as its fluorophore strand, once hybridized on the template, acts as a primer for the polymerase to extend. This extension displaces any red fluorophore strand hybridized beyond which would then re-associate with the quencher strand in solution and cause the conditional drop in red fluorescence. Therefore, probe B should be designed such that only in the presence of SNP B will the PB_F hybridize onto the template in order to minimize any undesired priming and extension upon the addition of the polymerase.

We tested two different types of probe B, one identifying SNP through the toehold region (T2-T8) and another through the branch migration region (B2) (**Figure S4**). Results showed that the two probes performed very differently and led to a different degree of drop in the red fluorescence signal after enzyme addition. As shown in **Figure S5**, using the probe B where the SNP was interrogated through the toehold region (T2) caused a significant drop in red fluorescence only for TD, but not for TA. In contrast, using the probe B where the SNP recognition site was designed in the branch migration region (B1), the red fluorescence dropped by almost the same amount for both TA and TD. This discrepancy implied that an unfavorable structure was formed between the probe and template when the SNP recognition site was placed at the branch migration region, which resulted in an undesired extension, making it not possible to differentiate TA from TD.

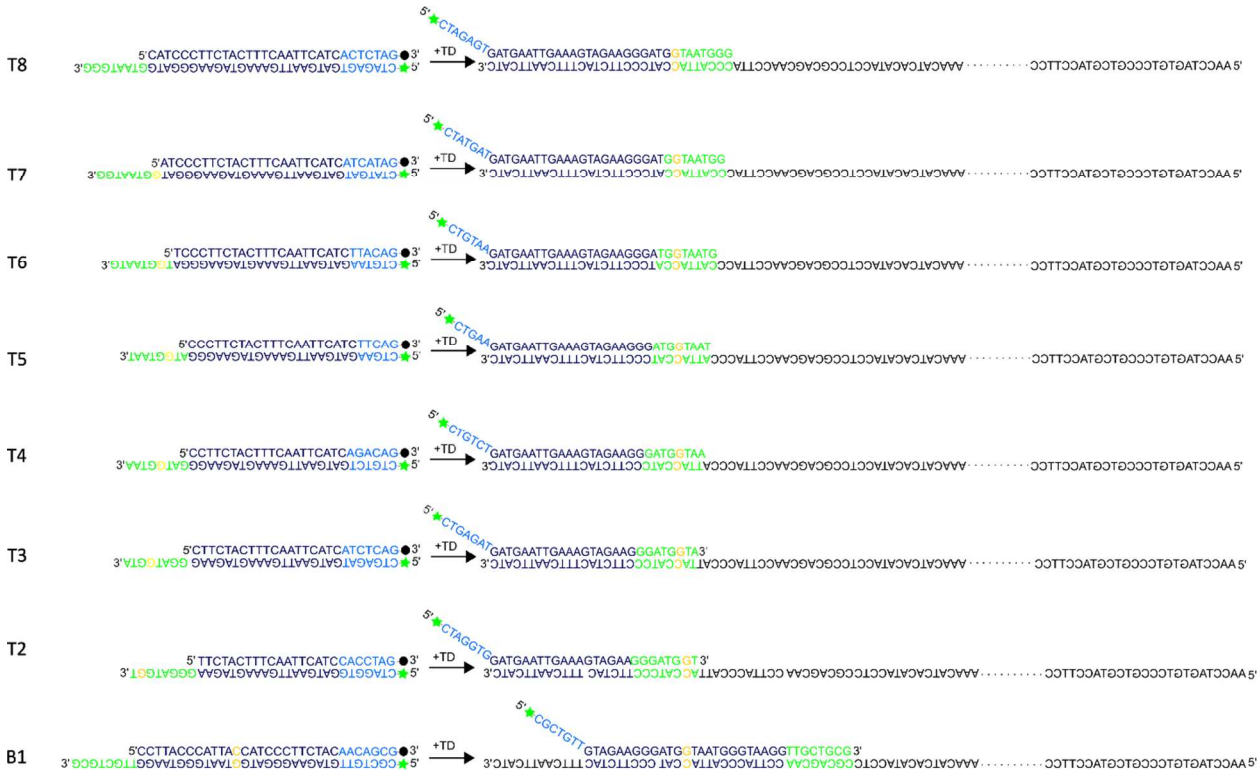
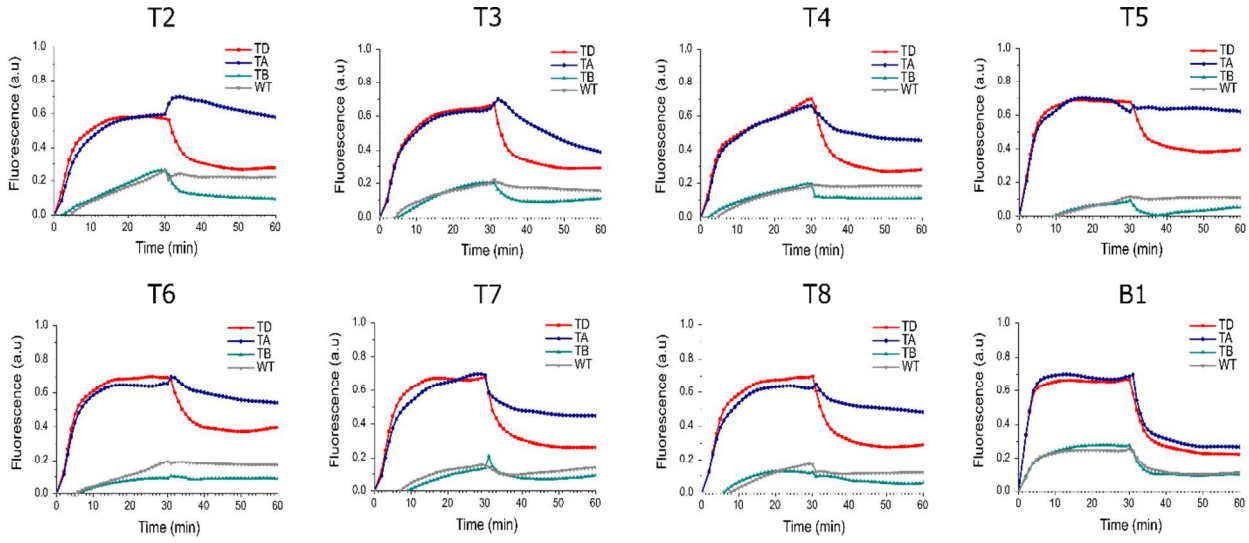



**Figure S4.** The eight different designs of Probe B and their binding with TD. Probe B was designed with the SNP recognition site either placed at the toehold (T2-T8) or the branch migration region (B2) of the probe to compare their performance for SNP phase recording. T and B stands for toehold and branch migration region respectively; while the number behind T represents the position of SNP recognition site counting from the 3' end of the toehold.




**FigureS 5.** The red fluorescence measurement of the eight different probe B, with SNP identification occurs either at the toehold (T) or the branch migration (B1). The number after 'T' was the SNP position

counting from the 3' end. In all cases, probes A and B (100 nM each) were first incubated with the templates (50 nM) for 30 minutes. Then enzymatic mixture was added at the 31st minute, and measurement was taken for another 30 minutes of reactions. In this set of experiments, all probe B used were not fluorophore- or quencher-labelled.

These experimental results coincide with the thermodynamic calculations made in Nupack, an online software which computes the Gibbs free energy of possible structures and the partitioning of these structures at equilibrium based on the particular set of input sequences. In the analysis, for each set of probes, we input the sequence of fluorophore and quencher strands of probe B and the template strand (either TD or TA) at 50 nM, 100 nM, and 50 nM respectively. For simplicity, these strands are referred to as F, Q, and T correspondingly. By simulating the possibility of forming a maximum of a three-stranded complex at 1 M $Na^+$ and 37$^oC$, both probes exhibited significant formation of FT (the desired product) in the case of TD, while negligible amount for TA. As both probes were designed to have similar reaction Gibbs free energies, their amount of FT at equilibrium was also similar. However, B1, where SNP identification was at the branch migration region, formed an undesired structure with TA through toehold association (see **Figure S6**), and their amount was similar to the FT formed in the case of TD (see **Table S1**). This was likely to happen because the toehold of probe B, which was 8 nt long and exactly complementary to the template, was long enough to associate onto the template. Additionally, this binding was also thermodynamically favorable to lower the system's enthalpy. The formation of FQT can act as a primer similar to FT that causes any red fluorophore strand attached afterwards to dissociate and quench after rehybridizing with the quencher strand in solution. By considering the equilibrium partition of FT and FQT formed in the case of B1, the sum of these products, or the structures that could cause the polymerase to prime were similar for TD and TA, reasonably explaining their similar drop in red fluorescence after the addition of enzymes. In contrast to B1, T1 shows negligible formation of FQT in the case of TA, owing to the thermodynamic barrier imposed by the mismatch. Proposed mechanisms accounting for the difference in red fluorescence signal after enzyme addition between B1 and T2 were shown in **Figure S7**.

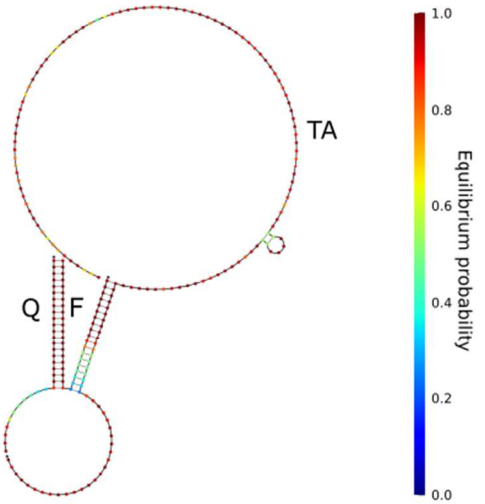



**Figure S6.** The structure of FQT formed between B1 and Ta from Nupack computation

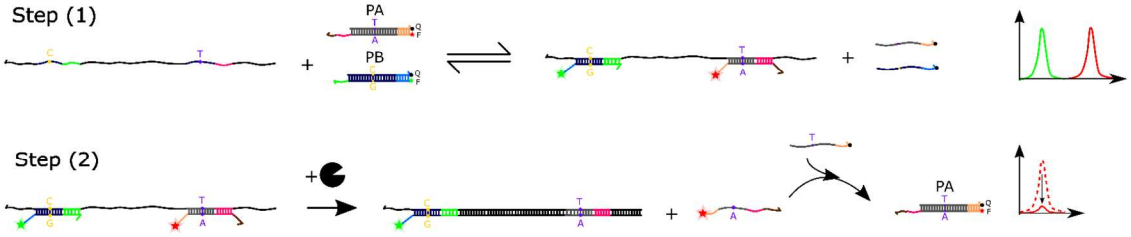
**Table S1.** Free energies and equilibrium concentrations of FT and FQT of different probe B while interacting with TD/ TA as calculated by Nupack.

| | FQ | T | FT | Q | $\Delta G_{reaction}$ | Equilibrium Conc (nM) | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | FT | FQT | FT+FQT |
| B1 | | | | | | | | |
| TD | -43.3 | -4.57 | -50.43 | -0.46 | -3.02 | 31 | 14 | 45 |
| TA | -43.3 | -4.67 | -47.53 | -0.46 | -0.02 | 9.6 | 39 | 48.6 |
| T8 | | | | | | | | |
| TD | -38.04 | -4.57 | -45.34 | -0.3 | -3.03 | 43 | 1.2 | 44.2 |
| TA | -38.04 | -4.67 | -42.44 | -0.3 | -0.03 | 17 | 0.45 | 17.45 |
| T7 | | | | | | | | |
| TD | -35.87 | -4.57 | -43.13 | -0.37 | -3.06 | 44 | 0.31 | 44.31 |
| TA | -35.87 | -4.67 | -40.22 | -0.37 | -0.05 | 17 | 0.096 | 17.096 |
| T6 | | | | | | | | |
| TD | -33.95 | -4.57 | -41.3 | -0.28 | -3.06 | 45 | / | 45 |
| TA | -33.95 | -4.67 | -38.41 | -0.28 | -0.07 | 17 | / | 17 |
| T5 | | | | | | | | |
| TD | -32.4 | -4.57 | -39.76 | -0.21 | -3 | 44 | / | 44 |
| TA | -32.4 | -4.67 | -36.89 | -0.21 | -0.03 | 17 | / | 17 |
| T4 | | | | | | | | |
| TD | -32.01 | -4.57 | -39.02 | -0.62 | -3.06 | 44 | 19 | 63 |
| TA | -32.01 | -4.67 | -36.22 | -0.62 | -0.16 | 27 | 0.089 | 27.089 |
| T3 | | | | | | | | |
| TD | -30.89 | -4.57 | -38.24 | -0.24 | -3.02 | 43 | 20 | 63 |
| TA | -30.89 | -4.57 | -35.63 | -0.24 | -0.41 | 15 | 0.56 | 15.56 |
| T2 | | | | | | | | |
| TD | -30.24 | -4.57 | -37.6 | -0.22 | -3.01 | 39 | 22 | 61 |
| TA | -30.24 | -4.67 | -35.26 | -0.22 | -0.57 | 5.9 | 2.8 | 8.7 |

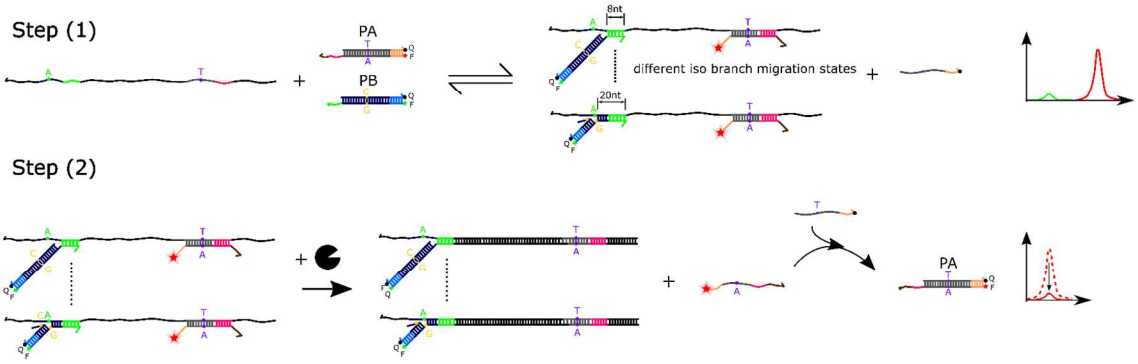
ΔG reaction was calculated by subtracting the free energies of the products (FT and Q) to that of the reactants (FQ and T)

## A. Recognition Site at Branch Migration (PB_B1)

TD:

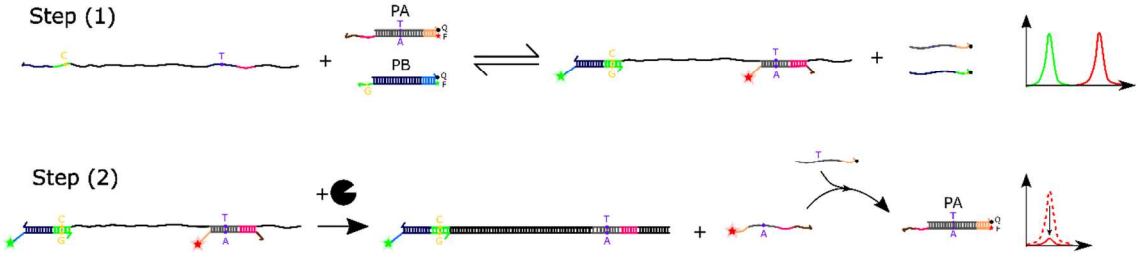



TA:




## B. Recognition Site at the Toehold (PB_T2)

TD:




TA:

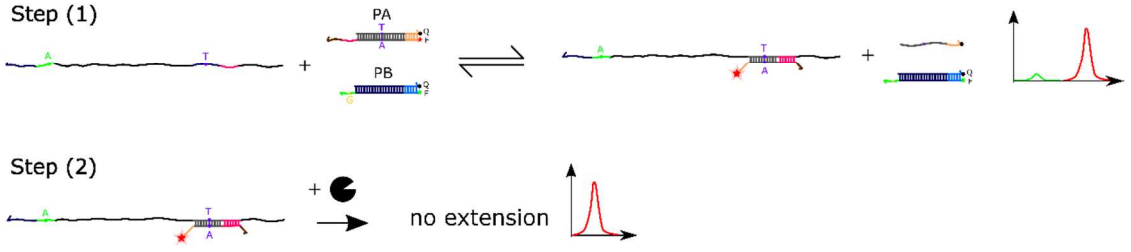


**Figure S7.** Proposed mechanism for the difference in red fluorescence obtained in A) B1 and B) T2 after enzyme addition. In the case for B2, the undesired association of B1 on TA through the 8-nt complementary region on the toehold (or even up to 20 nt till the mismatched site) caused unfavorable extension by the polymerase and resulted in a drop in red fluorescence signal in TA comparable to TD. When the recognition site was designed at the toehold region, this unfavorable binding was prohibited which allowed the fluorescence drop to occur only if both SNPs were present on the same template strand.

The performance of seven different positions of the SNP recognition site in probes B were then compared. All the SNP recognition sites were found at the toehold, but at different toehold positions (see **Figure S4**) They were designed such that the reaction Gibbs free energies between the probes and TD were similar, which were around -3 kcal/mol (in the absence of green fluorophore quencher pair). **Figure S5** shows that all of them (T2 to T8) demonstrated much higher drop in red fluorescence for TD than TA after the addition of enzyme. It is important to note that the PB probes tested here were not attached with a fluorophore or quencher molecule and this could increase the overall hybridization Gibbs free energy by about ~3-4 kcal/mol, which would subsequently lower their specificity in identifying the target SNP compared to the labelled ones. It was expected that with more specific hybridization of the green fluorophore strand on the template, more stringent conditions would be set for the polymerase to prime and extend. Therefore, the difference in the drop of red fluorescence would intuitively be stronger in the probes PB which were conjugated with the fluorophores and quenchers compared to the results shown here. **Table S1** also provides additional information about the equilibrium concentrations of TF and FQT for the probes tested. The ability to classify TA and TD through the drop in red fluorescence signal at different toehold positions demonstrated the flexibility in the design of probes for detection. For the succeeding experiments, T2 was chosen as the probe B, and all the results shown in the main manuscript use this probe.

## S4. Explanation for the signal change in green fluorescence channel after enzyme addition

Although the green fluorescence signal after the enzyme addition would not be used to confer any SNP or phase data, it is interesting that the results consistently showed that the green fluorescence would drop sharply, followed by a gradual increase in green fluorescence (as seen in the case using T2 in **Figure S3**). We believe that the initial sharp drop in green fluorescence during the first few minutes was due to the quenching of the green fluorophore due to the undesired extension on the 3' end of the template strand. This happened because T2, which was used as the final probe B in the rest of the experiments,

would create a 3’ recessed end on the template (or 5’ protruding end for the green fluorophore strand) once the fluorophore strand was attached to it. This 3’ recessed end would then act as a primer for BSM polymerase to extend a few bases yielding a double-stranded DNA product. Since the probes were designed using the GC clamp strategy to increase the probe stability, the green fluorophore molecule was adjacent to cytosine. Therefore, the extension of the template would result to having a guanine molecule near the fluorophore. As previously reported, guanine can act as a quencher and can lower the fluorescence signal of a fluorophore (see **Figure S8** for the proposed mechanism). This hypothesis was further supported by noting that this drop in green fluorescence was proportional to the saturated fluorescence in the first stage of the reaction (for example, much higher drop for TD and TB than that of TA and WT) as seen in Figure 2A. It is reasonable because with more of the hybridized structures between probe B and template formed, more of them will be extended that quenched the fluorophore. Also, as only 8 bases were needed to be incorporated, this explains why the fluorescence could drop sharply within just a few minutes (compared to the time needed for the drop in red fluorescence for TD which took around ten minutes to reach saturation because more than 100 nt needs to be incorporated onto the strand). We also proved this experimentally by comparing the green fluorescence of the reactions using T2 to that of B1 which would not form any 3’ recessive end after hybridization (see **Figure S9**). Clearly, after the addition of enzymes, B1 did not show any drop in green fluorescence as compared to T2, but a gradual increase in fluorescence afterwards similar to T2. This gradual increase in fluorescence was due to the consumption of the product formed between the green fluorophore strand and template after the fluorophore strand got extended by the enzyme. In **Figure S2**, the agarose gel also showed a decrease in the intensity on the position corresponding to PB. This was because the product removal distorted the equilibrium previously developed in the first 30 minutes of hybridization reactions as predicted by the Le Châtelier’s principle and pushed the reactions to produce more of the fluorophore- template complex that generated additional fluorescence signal.

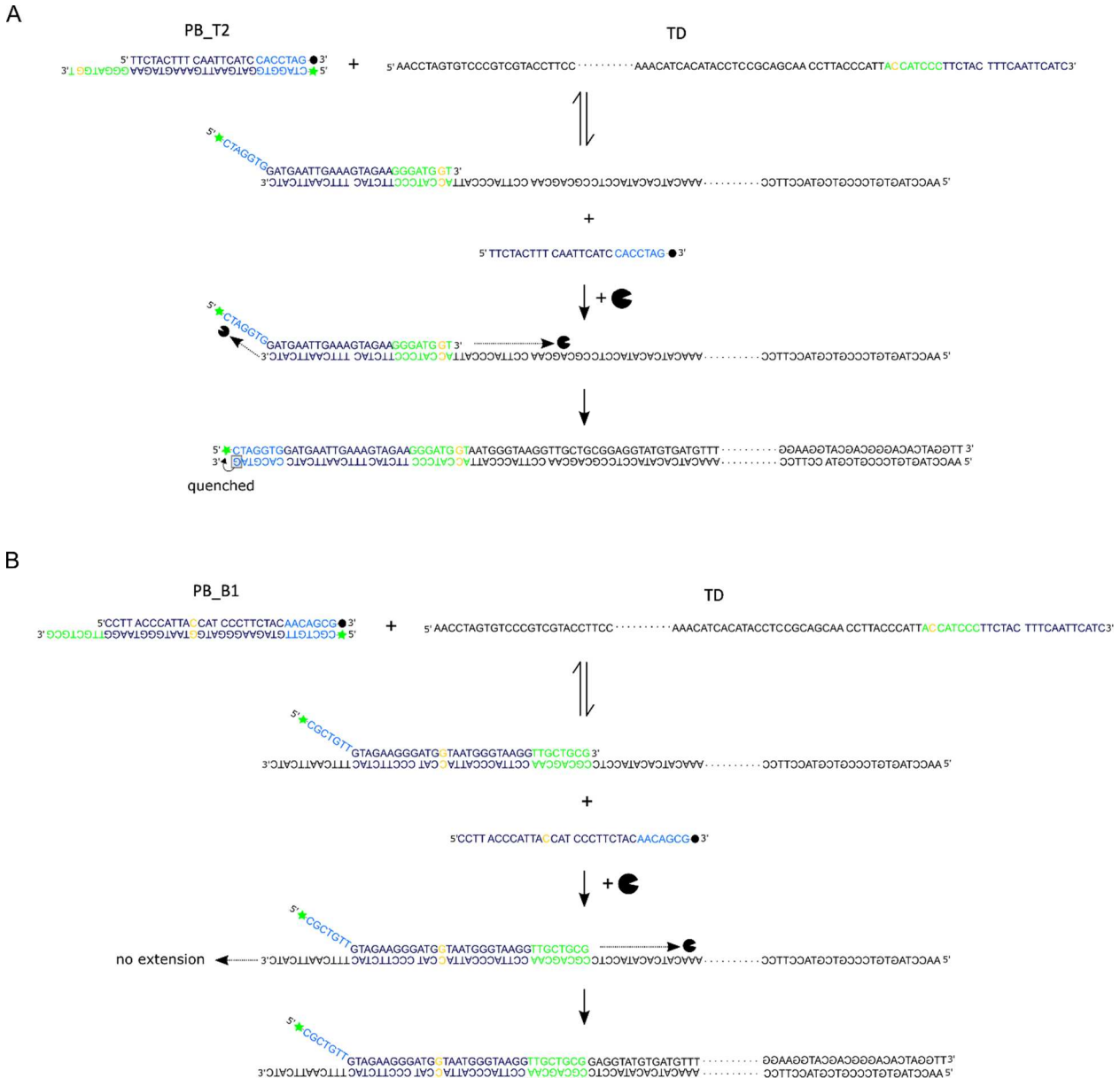



**Figure S8.** Proposed mechanism for the sharp drop in green fluorescence signal in the case of T2 after polymerase addition. A) In the case of T2, as a 3' recessed end will be generated after the probe is hybridized on the template. This allows the polymerase to prime and exert its 3' to 5' polymerase activity to form complementary nucleotides for both 3' ends. As a guanine was added as the last base after the extension in the blue region which is in close proximity to the green fluorophore, it will then quench the fluorophore from fluorescence emission. B) In the design of B1, a 3' overhang (the blue region) of B1_F after probe hybridization will prevent further polymerase extension or drop of green fluorescence signal.

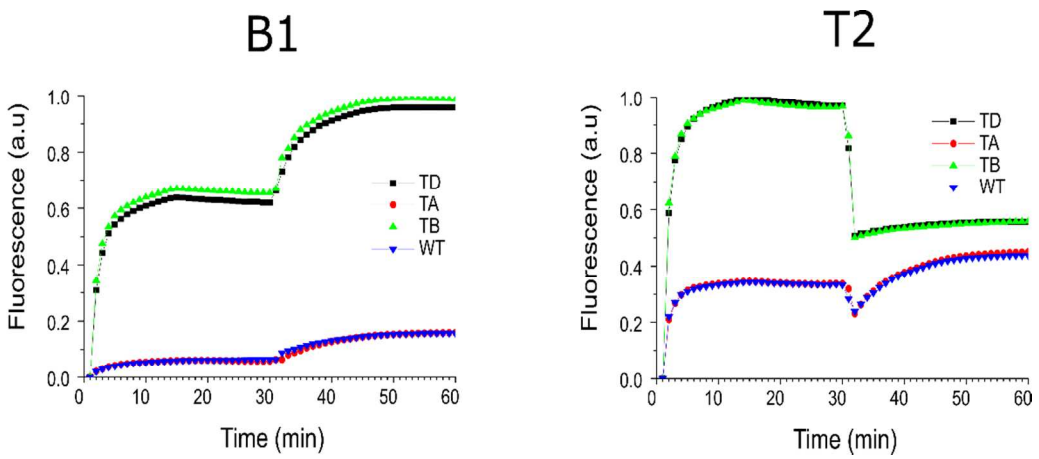



**Figure S9.** The green fluorescence measurement for probe B designed with SNP identification at the branch migration (B1) or the toehold region (T2, which is also the one shown throughout the main manuscript). Probes (100 nM) and templates (50 nM) were incubated for 30 minutes first and enzyme was added afterwards for another 30 minutes of reactions.

## S5. Sequence design for the experiments

All sequences demonstrated in this paper are shown in **Table S2 and S3**. The probes were designed such that the overall hybridization Gibbs free energies between the probes and the correct SNP were close to, but slightly lower than, zero to increase specificity. As the free energy contribution of the fluorophore-quencher pair is unknown and usually in the range of 3-4 kcal/mol, we designed the probes by Nupack such that their reaction energies for the correct target were less than -3 kcal/mol to provide some allowance for the enthalpy gain caused by the fluorophore-quencher pair.

**Table S2.** Sequences for the results in main figures

| Template | |
|---|---|
| TD | AACCTAGTGTCCCGTCGTACCTTCCATGTTCCTACCCTCAGTTTCTAAATCCTCACCCTACTACCACTA CCTCTAATACACATAAAGCCACAGCCATAAACATCACATACCTCCGCAGCAACCTTACCCATTACCAT CCCTTCTACTTTCAATTCATC |
| TA | AACCTAGTGTCCCGTCGTACCTTCCATGTTCCTACCCTCAGTTTCTAAATCCTCACCCTACTACCACTA CCTCTAATACACATAAAGCCACAGCCATAAACATCACATACCTCCGCAGCAACCTTACCCATTAACAT CCCTTCTACTTTCAATTCATC |

| | |
|---|---|
| TB | AACCTAGTGTCCCGTCGTACCTTCCATGTTCCGACCCTCAGTTTCTAAATCCTCACCCTACTACCACTA CCTCTAATACACATAAAGCCACAGCCATAAACATCACATACCTCCGCAGCAACCTTACCCATTACCAT CCCTTCTACTTTCAATTCATC |
| WT | AACCTAGTGTCCCGTCGTACCTTCCATGTTCCGACCCTCAGTTTCTAAATCCTCACCCTACTACCACTA CCTCTAATACACATAAAGCCACAGCCATAAACATCACATACCTCCGCAGCAACCTTACCCATTAACAT CCCTTCTACTTTCAATTCATC |
| Probe A (PA) | |
| PA_F | /5ATTO647NN/CT TGG CTG AAA CTG AGG GTA GGA ACA TGG AAG GTA CGA CGA AAA A*A*A |
| PA_Q | CTT CCA TGT TCC TAC CCT CAG TTT CAG CCA *A*G/3IAbRQSp/ |
| Probe B (PB or T2) | |
| PB_F | /5ATTO488N/CTAGGTG GATGAATTGAAA GTAGAA GGGATG*G*T |
| PB_Q | TTCTACTTTCAATTCATC CACCT*A*G/3IABkFQ/ |

Base underlined represent the SNP sites, * indicated nucleotide linked through phosphorothioate modification

**Table S3.** Sequences for probe B comparison

| | |
|---|---|
| B1 | |
| B1_F | CGCTGTTGTAGAAGGGATGGTAATGGGTAAGGTTGCTGCG |
| B1_Q | CCTTACCCATTACCATCCCTTCTACAACAGCG |
| B1 with fluorophore and quencher | |
| B1_F | /5ATTO488N/CG CTG TTG TAG AAG GGA TGG TAA TGG GTA AGG TTG CTG *C*G |
| B1_Q | CCT TAC CCA TTA CCA TCC CTT CTA CAA CAG *C*G/3IABkFQ/ |
| T2 | |
| T2_F | CTAGGTG GATGAATTGAAA GTAGAA GGGATGGT |
| T2_Q | TTCTACTTTCAATTCATC CACCTAG |
| T3 | |
| T3_F | CTGAGAT GATGAATTGAAA GTAGAAG GGATGGTA |
| T3_Q | CTTCTACTTTCAATTCATC ATCTCAG |
| T4 | |
| T4_F | CTGTCT GATGAATTGAAA GTAGAAGG GATGGTAA |
| T4_Q | CCTTCTACTTTCAATTCATC AGACAG |
| T5 | |
| T5_F | CTGAA GATGAATTGAAA GTAGAAGGG ATGGTAAT |
| T5_Q | CCCTTCTACTTTCAATTCATC TTCAG |
| T6 | |
| T6_F | CTGTAA GATGAATTGAAA GTAGAAGGGA TGGTAATG |

| | |
|---|---|
| T6_Q | TCCCTTCTACTTTCAATTCATC TTACAG |
| T7 | |
| T7_F | CTATGAT GATGAATTGAAA GTAGAAGGGAT GGTAATGG |
| T7_Q | ATCCCTTCTACTTTCAATTCATC ATCATAG |
| T8 | |
| T8_F | CTAGAGT GATGAATTGAAA GTAGAAGGGATG GTAATGGG |
| T8_Q | CATCCCTTCTACTTTCAATTCATC ACTCTAG |

* indicates nucleotide under phosphothioate modification