Electronic Supplementary Material (ESI) for Analytical Methods.

Electronic Supplementary Information (ESI)

Toehold Probe-based Interrogation for Haplotype Phasing of Long Nucleic Acid Strands

Xinyu Zhuang [a], Henson L Lee Yu [b], and I-Ming Hsing [a,b,*]

Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

*Email: kehising@ust.hk

Table S1 sequence information and clam structure

Probe A F(4th)	/5Alex647N/TCCAACATAAGACATTCAGGTCGCGCCCGTTGAAGAAGTC**C**TGC
Probe A Q(4th)	GGCGCGACCTGAATGTCTTATGTTGGA/3IAbRQSp/
Probe B F	/5`ATTO488N/ACCTACATCCTCGGCTGCCAGGTCGGGAATATTCCAGGG**T**TTTC GCAAAAAA
Probe B Q	CCGACCTGGCAGCCGAGGATGTAGGT/3`IABkFQ/
Sink P	GGCGCGACCTGAATGTCTTATGTTGGA/3`Inverted dT/
Sink C	TCCAACATAAGACATTCAGGTCGCGCCCGTTGAAGAAGTC**T**TGC/3`Inverted dT/
Reverse primer	/5Phos/CTTGTTCAGGTCGCGCCCGT
Forward primer	GCG GCA CTC TGG CCT CTG ATT G
Probe A F(1st)	/5ATTO488N/ACCTACATCCTCGGCTGCCAGGTCGGGAATATTCCAGGGTTTTCG CAAAAAA
Probe A Q(1st)	CCGACCTGGCAGCCGAGGATGTAGGT/3IABkFQ/
Probe A F(2nd)	/5Alex647N/TCCAACATAAGACATCAGGTCGCGCCCGTTGAAGAAGTCCTGCA
Probe A Q(2nd)	GGGCGCGACCTGATGTCTTATGTTGGA/3IAbRQSp/
Probe A F(3rd)	/5Alex647N/TCCAACATAAGACATTCAGGTCGCGCCCGTTGAAGAAGTCCTG
Probe A Q(3rd)	GCGCGACCTGAATGTCTT ATG TTG GA/3IAbRQSp/
Sink P(2nd)	GGGCGCGACCTGATGTCTTCTGTAGGC/3InvdT/
Sink C(2nd)	GCCTACAGAAGACATCAGGTCGCGCCCGTTGAAGAAGTCTTGCAG /3InvdT/
Sink P(3rd)	GCGCGACCTGAATGTCACATGTTGGT/3InvdT/
Sink C(3rd)	ACCAACATGTGACATTCAGGTCGCGCCCGTTGAAGAAGTCTT G/3InvdT/
TD	GCGGCACTCTGGCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA**A**CCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGC TCGGTGATTGGCTCAGAAGGGAAAAGGCGGGTCTCCGTGACGACTTATAAAAGCCCAGGGGCAAGCGGTCCGGATAACGGCTAGCCTGAGGAGCT GCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTT TCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCCAGCCCCCAATCTCAGAGCGGAGCCGACAGAGAGCAGGGAACCGG CATGGCCAAAGCCGCGGCGATCGGCATCGACCTGGGCACCACCTACTCCTGCGTGGGGGTGTTCCAACACGGCAAGGTGGAGATCATCGCCAACGA CCAGGGCAACCGCACCACCCCCAGCTACGTGGCCTTCACGGACACCGAGCGGCTCATCGGGGATGCGGCCAAGAACCAGGTGGCGCTGAACCCGCA GAACACCGTGTTTGACGCGAAGCGGCTGATTGGCCGCAAGTTCGGCGACCCGGTGGTGCAGTCGGACATGAAGCACTGGCCTTTCCAGGTGATCAA CGACGGAGACAAGCCCAAGGTGCAGGTGAGCTACAAGGGGGAGACCAAGGCATTCTACCCCGAGGAGATCTCGTCCATGGTGCTGACCAAGATGA AGGAGATCGCCGAGGCGTACCTGGGCTACCCGGTGACCAACGCGGTGATCACCGTGCCGGCCTACTTCAACGACTCGCAGCGCCAGGCCACCAAGG ATGCGGGTGTGATCGCGGGGCTCAACGTGCTGCGGATCATCAACGAGCCCACGGCCGCCGCCATCGCCTACGGCCTGGACAGAACGGGCAAGGGG GAGCGCAACGTGCTCATCTTTGACCTGGGCGGGGGCACCTTCGACGTGTCCATCCTGACGATCGACGACGGCATCTTCGAGGTGAAGGCCACGGCC GGGGACACCCACCTGGGTGGGGAGGACTTTGACAACAGGCTGGTGAACCACTTCGTGGAGGAGTTCAAGAGAAAACACAAGAAGGACATCAGCCA GAACAAGCGAGCCGTGAGGCGGCTGCGCACCGCCTGCGAGAGGGCCAAGAGGACCCTGTCGTCCAGCACCCAGGCCAGCCTGGAGATCGACTCCC TGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTTCGAGGAGCTGTGCTCCGACCTGTTCCGAAGCACCCTGGAGCCCGTGGAGAA GGCTCTGCGCGACGCCAAGCTGGACAAGGCCCAGATTCACGACCTGGTCCTGGTCGGGGGCTCCACCCGCATCCCCAAGGTGCAGAAGCTGCTGCA **G**GACTTCTTCAACGGGCGCGACCTGAACAAG
TA	GCGGCACTCTGGCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA**C**CCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGC TCGGTGATTGGCTCAGAAGGGAAAAGGCGGGTCTCCGTGACGACTTATAAAAGCCCAGGGGCAAGCGGTCCGGATAACGGCTAGCCTGAGGAGCT GCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTT TCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCCAGCCCCCAATCTCAGAGCGGAGCCGACAGAGAGCAGGGAACCGG CATGGCCAAAGCCGCGGCGATCGGCATCGACCTGGGCACCACCTACTCCTGCGTGGGGGTGTTCCAACACGGCAAGGTGGAGATCATCGCCAACGA CCAGGGCAACCGCACCACCCCCAGCTACGTGGCCTTCACGGACACCGAGCGGCTCATCGGGGATGCGGCCAAGAACCAGGTGGCGCTGAACCCGCA GAACACCGTGTTTGACGCGAAGCGGCTGATTGGCCGCAAGTTCGGCGACCCGGTGGTGCAGTCGGACATGAAGCACTGGCCTTTCCAGGTGATCAA CGACGGAGACAAGCCCAAGGTGCAGGTGAGCTACAAGGGGGAGACCAAGGCATTCTACCCCGAGGAGATCTCGTCCATGGTGCTGACCAAGATGA AGGAGATCGCCGAGGCGTACCTGGGCTACCCGGTGACCAACGCGGTGATCACCGTGCCGGCCTACTTCAACGACTCGCAGCGCCAGGCCACCAAGG ATGCGGGTGTGATCGCGGGGCTCAACGTGCTGCGGATCATCAACGAGCCCACGGCCGCCGCCATCGCCTACGGCCTGGACAGAACGGGCAAGGGG GAGCGCAACGTGCTCATCTTTGACCTGGGCGGGGGCACCTTCGACGTGTCCATCCTGACGATCGACGACGGCATCTTCGAGGTGAAGGCCACGGCC GGGGACACCCACCTGGGTGGGGAGGACTTTGACAACAGGCTGGTGAACCACTTCGTGGAGGAGTTCAAGAGAAAACACAAGAAGGACATCAGCCA GAACAAGCGAGCCGTGAGGCGGCTGCGCACCGCCTGCGAGAGGGCCAAGAGGACCCTGTCGTCCAGCACCCAGGCCAGCCTGGAGATCGACTCCC TGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTTCGAGGAGCTGTGCTCCGACCTGTTCCGAAGCACCCTGGAGCCCGTGGAGAA GGCTCTGCGCGACGCCAAGCTGGACAAGGCCCAGATTCACGACCTGGTCCTGGTCGGGGGCTCCACCCGCATCCCCAAGGTGCAGAAGCTGCTGCA **G**GACTTCTTCAACGGGCGCGACCTGAACAAG
TB	GCGGCACTCTGGCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA**A**CCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGC TCGGTGATTGGCTCAGAAGGGAAAAGGCGGGTCTCCGTGACGACTTATAAAAGCCCAGGGGCAAGCGGTCCGGATAACGGCTAGCCTGAGGAGCT GCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTT TCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCCAGCCCCCAATCTCAGAGCGGAGCCGACAGAGAGCAGGGAACCGG CATGGCCAAAGCCGCGGCGATCGGCATCGACCTGGGCACCACCTACTCCTGCGTGGGGGTGTTCCAACACGGCAAGGTGGAGATCATCGCCAACGA CCAGGGCAACCGCACCACCCCCAGCTACGTGGCCTTCACGGACACCGAGCGGCTCATCGGGGATGCGGCCAAGAACCAGGTGGCGCTGAACCCGCA GAACACCGTGTTTGACGCGAAGCGGCTGATTGGCCGCAAGTTCGGCGACCCGGTGGTGCAGTCGGACATGAAGCACTGGCCTTTCCAGGTGATCAA CGACGGAGACAAGCCCAAGGTGCAGGTGAGCTACAAGGGGGAGACCAAGGCATTCTACCCCGAGGAGATCTCGTCCATGGTGCTGACCAAGATGA AGGAGATCGCCGAGGCGTACCTGGGCTACCCGGTGACCAACGCGGTGATCACCGTGCCGGCCTACTTCAACGACTCGCAGCGCCAGGCCACCAAGG ATGCGGGTGTGATCGCGGGGCTCAACGTGCTGCGGATCATCAACGAGCCCACGGCCGCCGCCATCGCCTACGGCCTGGACAGAACGGGCAAGGGG GAGCGCAACGTGCTCATCTTTGACCTGGGCGGGGGCACCTTCGACGTGTCCATCCTGACGATCGACGACGGCATCTTCGAGGTGAAGGCCACGGCC GGGGACACCCACCTGGGTGGGGAGGACTTTGACAACAGGCTGGTGAACCACTTCGTGGAGGAGTTCAAGAGAAAACACAAGAAGGACATCAGCCA GAACAAGCGAGCCGTGAGGCGGCTGCGCACCGCCTGCGAGAGGGCCAAGAGGACCCTGTCGTCCAGCACCCAGGCCAGCCTGGAGATCGACTCCC TGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTTCGAGGAGCTGTGCTCCGACCTGTTCCGAAGCACCCTGGAGCCCGTGGAGAA GGCTCTGCGCGACGCCAAGCTGGACAAGGCCCAGATTCACGACCTGGTCCTGGTCGGGGGCTCCACCCGCATCCCCAAGGTGCAGAAGCTGCTGCA **A**GACTTCTTCAACGGGCGCGACCTGAACAAG
WT	GCGGCACTCTGGCCTCTGATTGGTCCAAGGAAGGCTGGGGGGCAGGACGGGAGGCGAAA**C**CCCTGGAATATTCCCGACCTGGCAGCCTCATCGAGC TCGGTGATTGGCTCAGAAGGGAAAAGGCGGGTCTCCGTGACGACTTATAAAAGCCCAGGGGCAAGCGGTCCGGATAACGGCTAGCCTGAGGAGCT GCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTT TCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCCAGCCCCCAATCTCAGAGCGGAGCCGACAGAGAGCAGGGAACCGG CATGGCCAAAGCCGCGGCGATCGGCATCGACCTGGGCACCACCTACTCCTGCGTGGGGGTGTTCCAACACGGCAAGGTGGAGATCATCGCCAACGA CCAGGGCAACCGCACCACCCCCAGCTACGTGGCCTTCACGGACACCGAGCGGCTCATCGGGGATGCGGCCAAGAACCAGGTGGCGCTGAACCCGCA GAACACCGTGTTTGACGCGAAGCGGCTGATTGGCCGCAAGTTCGGCGACCCGGTGGTGCAGTCGGACATGAAGCACTGGCCTTTCCAGGTGATCAA CGACGGAGACAAGCCCAAGGTGCAGGTGAGCTACAAGGGGGAGACCAAGGCATTCTACCCCGAGGAGATCTCGTCCATGGTGCTGACCAAGATGA AGGAGATCGCCGAGGCGTACCTGGGCTACCCGGTGACCAACGCGGTGATCACCGTGCCGGCCTACTTCAACGACTCGCAGCGCCAGGCCACCAAGG ATGCGGGTGTGATCGCGGGGCTCAACGTGCTGCGGATCATCAACGAGCCCACGGCCGCCGCCATCGCCTACGGCCTGGACAGAACGGGCAAGGGG GAGCGCAACGTGCTCATCTTTGACCTGGGCGGGGGCACCTTCGACGTGTCCATCCTGACGATCGACGACGGCATCTTCGAGGTGAAGGCCACGGCC GGGGACACCCACCTGGGTGGGGAGGACTTTGACAACAGGCTGGTGAACCACTTCGTGGAGGAGTTCAAGAGAAAACACAAGAAGGACATCAGCCA GAACAAGCGAGCCGTGAGGCGGCTGCGCACCGCCTGCGAGAGGGCCAAGAGGACCCTGTCGTCCAGCACCCAGGCCAGCCTGGAGATCGACTCCC TGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTTCGAGGAGCTGTGCTCCGACCTGTTCCGAAGCACCCTGGAGCCCGTGGAGAA GGCTCTGCGCGACGCCAAGCTGGACAAGGCCCAGATTCACGACCTGGTCCTGGTCGGGGGCTCCACCCGCATCCCCAAGGTGCAGAAGCTGCTGCA **A**GACTTCTTCAACGGGCGCGACCTGAACAAG

*Underline is primer binding region and the green one is SNP B site and red is SNP A.

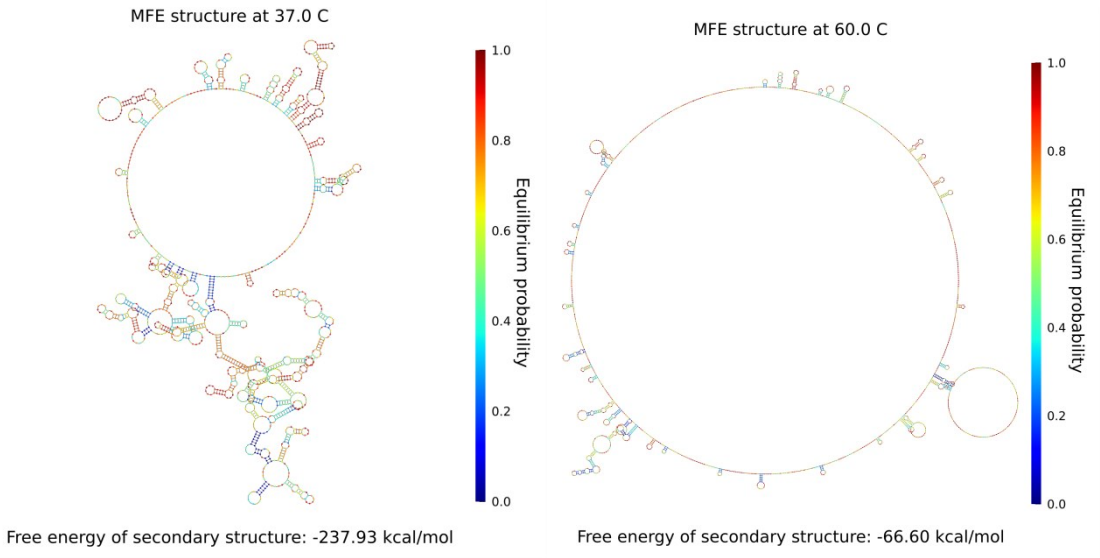


Figure S1 Possible secondary structures of the target (TD) sequence at 37.0℃ and 60 ℃ by NUPACK, respectively.

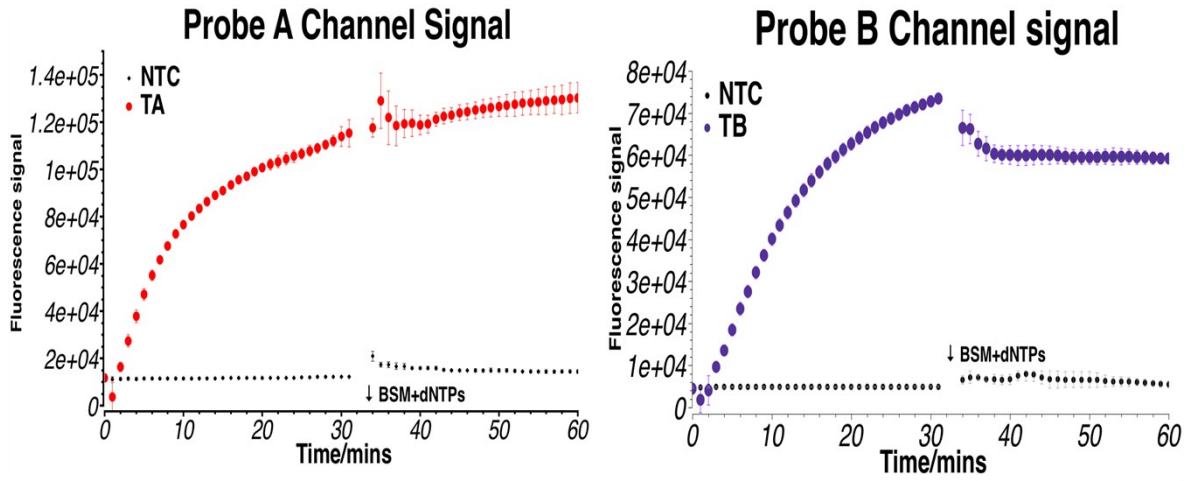


Figure S2 stability experiment of the probe A, probe B and sink probe. NTC represents negative control(without targets, only have probe A, probe B and sink probe.

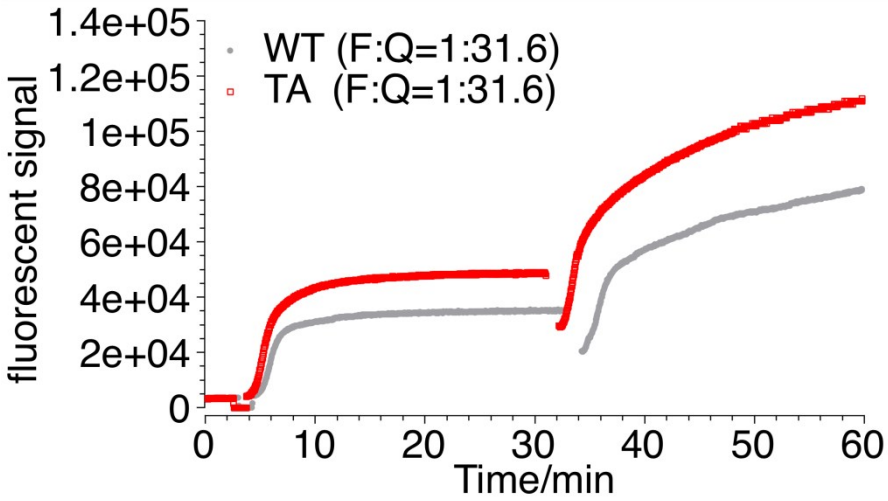


Figure S3 Fluorescence signals of 20nM WT and TA incubated with 20nM probe A, 20nM probe B , 1uM dNTPs and 8 units bsm in A channel. The fluorophore to quencher strands ratio of toehold probe A is 31.6:1.

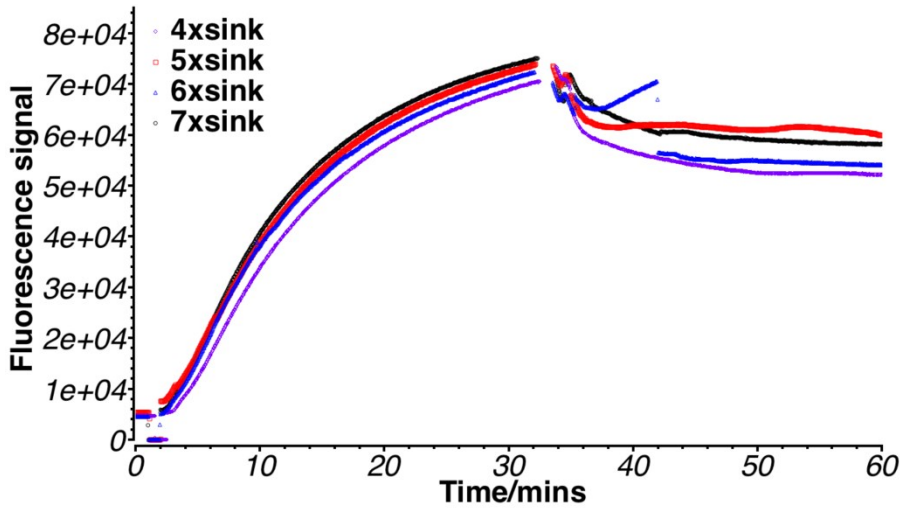


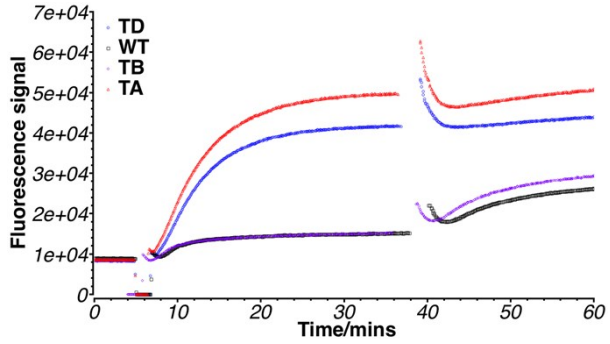
Figure S4 Fluorescence measurement of 20nM TB in 80nM(4x), 100nM(5x), 120nM(6x) and 140nM(7x) sink system in B channel, respectively. TB in 5x sink system was observed with lowest and stable drop compared to 4x, 6x and 7x.

Table S2 ΔGrxn and ΔΔG calculation by NUPACK of four kinds of different-design probe A. The Numbers in parentheses represent nucleotide numbers near SNP A sites of probe A fluorophore strand.

	TD ΔGrxn	TB ΔGrxn	ΔΔG
1st probe A(27)	-4.11	-0.4	3.71
2nd probe A(4)	-2.21	-0.29	1.92
3rd probe A(2)	-4.14	-1.8	2.34
4th probe A(3)	-3.98	-0.79	3.19

a)

b)



c)

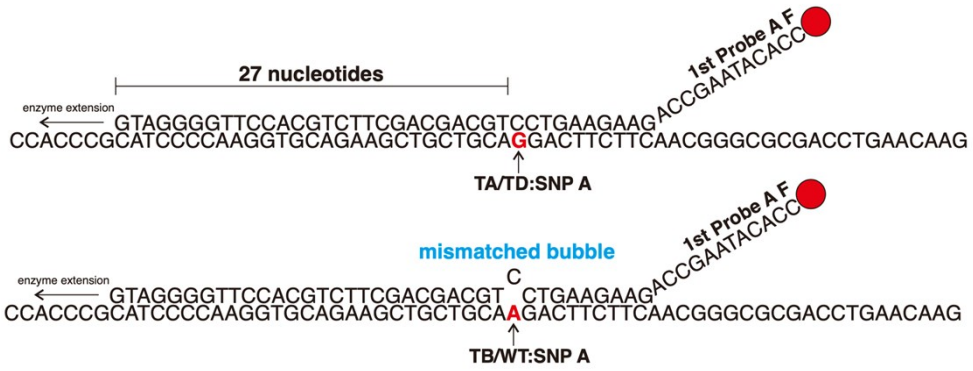


Figure S5 a) Fluorescence measurement of 20nM four haplotypes hybridized with 20nM 1st probe A, 20nM probe B, 1uM dNTPs and 8 units bsm polymerase in A channel. b) Discrimination comparison towards TD and TB in B channel among 1st, 2nd, 3rd, 4th probe A. The result of 1st probe A was obtained without sink while others were with their related sink molecules. c) illustration of structure of 1st probe A binding to four targets in SNP A sites.

The 1st probe A with the highest ΔΔG(table S2) was observed had lowest discrimination ability toward TB and TD in B channel though made a high distinguish between TB and TD in A channel , this can be explained by the hybridization structure when 1st probe A bound to WT/TB, due to the sequence with 27 nucleotides near the mismatched bubble(figure S3 c) is stable enough for the enzyme extension compared to 2, 3, 4 nucleotides of 3rd probe A, 4th probe A and 2nd probe A, thus stroking equilibrium that lead to small difference in TB and TD in B signal channel (figure S4 b) which even cannot be solved by sink. The kinetic process of 1st probe A reacted with four haplotypes(figure S4 a) was consistent with above hypothesis. To inhibit this effect, 4th probe A with short nucleotides and large discrimination was chosen for final experiments.

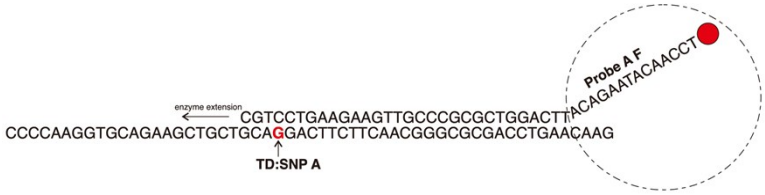


Figure S6 The formed clamp structure by from 5`overhang region of toehold probe A and 3`overhang of the template can prohibit polymerase extension which will result in quenching of probe A fluorescence signal.

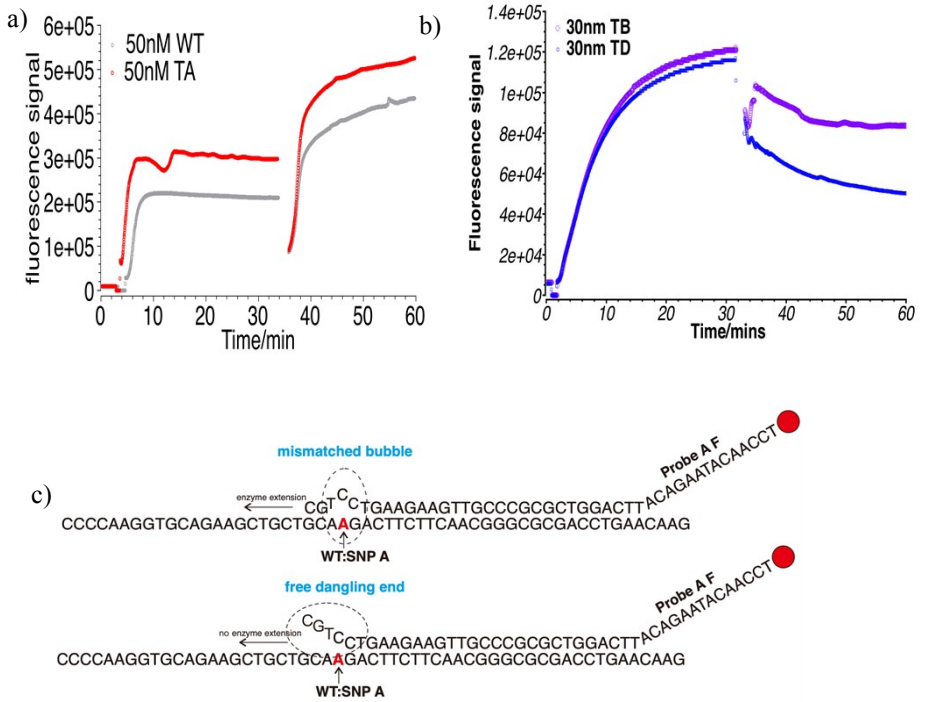


Figure S7 Fluorescence signals of a) 50nM TB and TD incubated with 50nM probe A, 50nM probe B, 250nM sink, 2.5uM dNTPs and 20 units bsm in A channel; b) 30nM TB and TD incubated with 30nM probe A, 30nM probe B, 150nM sink, 1.5uM dNTPs and 12 units bsm in B channel; c)two possible states in SNP A site : mismatched bubble and free dangling end when probe A binds to target WT.

After enzyme addition, probe A signal increase was found both in WT and TA under 50 and 60 nM targets that suggests more and more probe A bound to target WT and TA which is consistent with the suddenly probe B signal drop evidence at second stage in 30nM(shown in figure S6 abc). This phenomenon demonstrated that the reaction rate to the right of following

equilibrium reaction can be intensified due to the rate of FT consumption by polymerase also increased when the concentration exceeded 20nM.Another interesting point that we found was the signal of WT also had the same growing trend, theoretically, this structure is less likely to form mismatch bubbles than the free dangling end structure(shown in figure S6 d)because two bases hybridization usually is unstable. However, the chance for forming mismatched bubble improved a lot with the increase of target and probe concentration thus WT can also be extended by polymerase that resulted in disequilibration.

$FQ + T \leftrightharpoons FT + Q$