Electronic Supplementary Material (ESI) for Chemical Communications

Supporting Information for "Quantitative genotyping of single nucleotide polymorphism by single-molecule multi-color fluorescence resonance energy transfer"

Hye Ran Koh, Kyu Young Han, Jiwon Jung, and Seong Keun Kim*

Department of Chemistry and WCU Department of Biophysics and Chemical Biology
Seoul National University, Seoul 151-747, Korea

1. Sample preparation

All oligonucleotides were purchased in HPLC-purified form from Integrated DNA Technologies (Coraville, IA) and all fluorescent dyes with N-hydroxysuccinimide (NHS) ester group from Molecular Probes (Eugene, OR). The sequence of oligonucleotides and the dye labeling position are given in Table S1. The designed oligonucleotide with a primary aliphatic amine modifier was labeled with a fluorescent dye through the reaction of the dye's NHS ester group with the amine group of the oligonucleotide in its labeling position. The dye-labeled oligonucleotides were purified by gel electrophoresis using 18% denaturing polyacrylamide/5M urea. We obtained a labeling efficiency of 95% or higher by measuring the absorbance of the oligonucleotides and comparing it against that of the dye. For hybridization, the reference and probe DNAs were added to the target DNA containing a SNP site in an annealing buffer (25 mM Tris-HCl pH 8.0, 250 mM NaCl) by incubating the mixture at 95 °C for 2 min and slowly cooling it down to room temperature. For 3c-ALEX SNP genotyping, the annealed sample was

diluted in single-molecule buffer (50 mM Tris-HCl pH 7.0, 1 mM mercaptoethylamine, 100 μg/ml BSA, 5 % glycerol) to 50 pM and incubated for 5 min at 35 ℃.

2. Single-molecule data acquisition and analysis

We detect one freely-diffusing molecule at a time by using a tightly focused laser light with a confocal scheme and a low concentration of sample (50 ~ 100 pM), which means that one can prepare the sample simply by dilution without any special sample preparation procedure such as a surface immobilization used in a chip-based SNP genotyping assay. Three lasers of different colors were used to excite three different fluorophores in our sample in a cyclically alternating way and the resulting fluorescence at different wavelengths was detected by three avalanche photodiodes. We obtained nine independent channels of fluorescence burst signals from three lasers and three detectors with 600 μs binning, from which FRET efficiency (E_{XY}) and color-pair stoichiometry (S_{XY}) were calculated. The details of instrumentation, data acquisition, and data analysis for ALEX have been described in our previous paper[1].

3. Probe Design for an optimal single-base mismatch discrimination

The working principle of our SNP genotyping scheme is based on the discrimination between a fully complementary DNA duplex and one with a single base mismatch. Both base paring and

base stacking should be considered in designing an optimal probe DNA for maximum allele discrimination. First of all, the mismatch site with frustrated pi-stacking between adjacent bases should be in the middle of the probe DNA rather than its either end to ensure the maximum effect of energetic instability resulting from a single base mismatch.[2, 3] The length of the oligonucleotide is important as well, since the loss of the hydrogen bonding and pi-stacking by single base mismatch would affect the relative stability and target specificity differently. The relative instability due to single base mismatch is greater for a shorter probe but the specificity of probe binding to the target DNA suffers, whereas a longer oligonucleotide has higher target specificity but lower allele discrimination[3]. The optimal length of the oligonucleotide depends on many factors such as a base sequence and an incubation temperature, but a typical length of 10 to 15 bases was employed in our study. Likewise, the optimal length of the reference DNA was also determined by trial and error in order to achieve the maximum difference in FRET efficiency with the probe DNA labeled at the 5' end from the one at 3' end. In case its own binding to the target DNA is not strong enough for efficient hybridization, we used a longer reference DNA but maintained the dye position at the optimal value by using internal labeling instead of end labeling.

4. SNP discrimination based on the color-pair sorting capability of 3c-ALEX

Two different alleles (T and G) at a given locus were discriminated thru 2-D *S* diagram based on the color-pair sorting capability of 3c-ALET. For genotyping of the binary T414G SNP, the target DNAs were hybridized with a reference DNA labeled red and their respective allele-specific probe DNAs labeled blue (for detecting T) or green (for G) at the SNP site. The 2-D *S* diagram for target G should show bursts centered around $S_{GR} \sim 0.5$ for the "G-R" pair between the red reference and the green probe for the allele G. Likewise, the 2-D *S* diagram for target T should show bursts around $S_{BR} \sim 0.5$ for the "B-R" pair between the red reference and the blue probe for the allele T. The actual bursts for a 1:1 mixture of target G and target T yield the 2-D *S* diagram of Fig. S1 (a), which indeed shows both the G-R pair and the B-R pair at $S \sim 0.5$. The actual number of bursts for target G is about 1.5 times that for target T for the 1:1 mixture (Figs. S1 (b) and (c)), which simply reflects the stronger binding of the GC base pair over the AT pair. In order to investigate if this difference is mainly from the stronger hydrogen bonding of GC than AT, we checked for possible artifacts causing different burst counts that may result from different photophysical properties of fluorophores, but any such artifact was deemed insignificant when we compared the burst counts of donor-only species and donor-acceptor species between the G-R and B-R dye pairs. The G-R dye pair showed only 5% higher quenching effect than the B-R dye pair (data not shown). A correction factor for this

difference between GC and AT base pairs was thus introduced to make our measurement of the heteroplasmy rate quantitatively correct.

Reference

1. N. K. Lee, A. N. Kapanidis, H. R. Koh, Y. Korlann, S. O. Ho, Y. Kim, N. Gassman, S. K. Kim and S. Weiss, *Biophysical Journal*, 2007, **92**, 303-312.
2. T. Naiser, J. Kayser, T. Mai, W. Michel and A. Ott, *Bmc Bioinformatics*, 2008, **9**, 509-520.
3. Y. You, B. G. Moreira, M. A. Behlke and R. Owczarzy, *Nucleic Acids Res*, 2006, **34**, e60.

Table S1. Sequences of oligonucleotides for 3c-ALEX SNP genotyping.

	Random sequence	T414G sequence
Target	ACGTACTCTATGCCCGCGTAAGTCGTA NTGCCGTATTAGCCTGTA (46 mer) (N = A, T, G, or C)	AACCAGATTTCAAATTTTATCTTTTGGC GGNATGCACTTTTAACAG (46 mer) (N = A, T, G or C)
reference	ACGCGGGCAT /TMR/ (10 mer)	AAAAGATAAAA/TMR/TTG (15 mer)
		AAAAGATAAAA/Al647/TTG
probe	/Al647/TACGGCAATACGACT (15 mer) /Al488/TACGGCACTACGACT TACGGCAGTACGACT /Al647/ TACGGCATTACGACT /Al488/	/Al647/TGCATACCGCC (11 mer) /Al488/TGCATCCCGCC TGCATGCCGCC/Al647/ TGCATTCCGCC /Al488/
		/Al488/TGCATACCGCC /TMR/TGCATCCCGCC

The underlined base denotes the base at the SNP site or its complementary site.

The labeling position of each dye can be adjustable.

Figure S1. Binary SNP genotyping scheme using the "color-pair sorting" scheme via a 2-D *S* diagram. (a) 2-D *S* diagram obtained by 3c-ALEX SNP genotyping assay using a 1:1 mixture of target G (red circle) and target T (blue circle). (b) Selected bursts from Fig. S1(a), displaying the number of target G and (c) target T in the mixture. The ratio of target G to target T is 1077 / 719 ~ 1.5, reflecting the difference in the binding affinity between the GC and AT base pairs.