Multiplex Single-Molecule DNA Barcoding Using an Oligonucleotide Ligation Assay

Severins, Ivo; Szczepaniak, Malwina; Joo, Chirlmin

DOI
10.1016/j.bpj.2018.08.013

Publication date
2018

Document Version
Final published version

Published in
Biophysical Journal

Biophysical Journal
Article

Biophysical Society

Multiplex Single-Molecule DNA Barcoding Using an Oligonucleotide Ligation Assay

Ivo Severins,[1] Malwina Szczepaniak,[1,*] and Chirlmin Joo[1,*]
[1]Kavli Institute of NanoScience, Department of BioNanoScience, Delft University of Technology, Delft, The Netherlands

ABSTRACT Detection of specific nucleic acid sequences is invaluable in biological studies such as genetic disease diagnostics and genome profiling. Here, we developed a highly sensitive and specific detection method that combines an advanced oligonucleotide ligation assay with multicolor single-molecule fluorescence. We demonstrated that under our experimental conditions, 7-nucleotide long DNA barcodes have the optimal short length to ascertain specificity while being long enough for sufficient ligation. Using four spectrally separated fluorophores to label DNA barcodes, we simultaneously distinguished four DNA target sequences differing by only a single nucleotide. Our single-molecule approach will allow for accurate identification of low-abundance molecules without the need for target DNA preamplification.

INTRODUCTION

Personalized medicine relies on recognition of specific nucleic acid sequences. Variations in DNA sequences are associated with risks of developing diseases and with varying metabolic response to drugs or vaccines. Many of the currently available nucleic-acid-recognition methods rely on DNA sample amplification that involves polymerase chain reactions (1–5). Even though polymerase chain reactions revolutionized life sciences research, this method is not completely error free and may lead to a false positive detection of mutations (6). Thus, it has become necessary to develop a method to detect even single-nucleotide (nt) variations without the need to employ a step that could be a possible source of an erroneous readout. In addition, such a technique should be sensitive enough to detect low-abundance target molecules.

DNA hybridization and specificity of enzymatic ligation (7) of two complementary nucleic acid fragments constitute two basic elements of many bulk and single-molecule nucleic-acid-detection methods (for more exhaustive reviews of the methods used in single-nt genotyping, see (8–10)). For example, Tyagi et al. (11,12) designed molecular beacons that exploit the hybridization capability of DNA. These hairpin-shaped oligonucleotide probes, labeled with an internally quenched fluorophore, were used to distinguish four different target DNA strands in bulk in homogenous solution. In independent studies, Landegren et al. (13) and Alves et al. (14) combined the annealing property with the enzymatic ligation reaction, thus introducing the oligonucleotide ligation assay (OLA). This original ligation assay was later combined with a Förster resonance energy transfer (FRET) detection method (15). Similar to OLA, padlock probes proposed by Nilsson et al. (16), take advantage of both reactions (hybridization and ligation), creating circular DNA molecules catenated to the target sequence.

More recently, in an attempt to increase sensitivity and specificity without the need for enzymatic target amplification, researchers moved toward single-molecule methods. Castro et al. (17) combined DNA hybridization with laser-based single-molecule detection of single-copy genes in a complex genome. Also, molecular beacons were extensively studied, and their various versions were employed in the single-molecule single-nt polymorphism (SNP) genotyping techniques. Single-molecule FRET allowed for analysis of low-abundancy point mutations in K-ras oncogenes by development of reverse molecular beacons (18). Another variation of molecular beacons called "Smart Probes" was developed to minimize unwanted background signal, thus increasing identification sensitivity (19). Wang et al. (20) introduced molecular confinement via electrokinetic focusing, which, coupled with the original molecular beacons and a confocal fluorescence spectroscope, brought the limit of detection down to attomolar range. Similar

Submitted March 20, 2018, and accepted for publication August 6, 2018.
*Correspondence: malwina.szczepaniak@gmail.com or c.joo@tudelft.nl
Ivo Severins and Malwina Szczepaniak contributed equally to this work.
Malwina Szczepaniak's present address is DiabetOmics, Inc., Hillsboro, Oregon.
Editor: Antoine van Oijen.
https://doi.org/10.1016/j.bpj.2018.08.013

detection sensitivity, attributed to even lower background signal, was shown by Zhang et al. (21,22) for a quantum dot-FRET nanosensing platform. Most of the single-molecule techniques utilize fluorescence signal for detection. A distinct detection method (23) relies on the measurement of the extension of a DNA hairpin attached to a magnetic bead, which is controlled by a magnetic trap.

However, many of the existing single-molecule approaches use low-volume detection methods (e.g., confocal microscopy), which limit the resulting data yield. Moreover, they often apply relatively long DNA probes (15 nt or more) that, because of the higher binding affinity, can potentially lead to false positive detection. Most recently, a few techniques emerged that use shorter and therefore more specific probes. The method developed by Su et al. (24) can detect SNPs by measuring transient DNA binding of 9-nt fluorescently labeled oligonucleotide probes with a total internal reflection fluorescence (TIRF) microscopy setup, whereas Levesque et al. (25) report on RNA SNP in situ fluorescence detection using a "toe-hold" strategy. Ding et al. (23) employed a magnetic trap, a more specialized and less common detection method than fluorescence. These techniques, however, currently lack the ability to simultaneously distinguish and quantify multiple sequences.

Here, we introduce a highly specific single-molecule SNP detection technique that stably attaches a short DNA probe to its matching sequence. For this, we furthered the bulk OLA technique (13,14) by combining it with a single-molecule fluorescence detection scheme. Our method provides a highly specific and sensitive DNA barcoding tool with potential for multiplexing applications. The high specificity is assured by the use of a pair of very short 7-nt DNA barcodes (probes), which are ligated only if complete complementarity to two adjacent sites on a target DNA molecule is realized. The sensitivity, on the other hand, is provided by the single-molecule approach, which enables the detection of low-abundance targets. We show that our method can be applied to simultaneously distinguish at least four single-nt variants.

MATERIALS AND METHODS

Slide preparation

Microfluidic chambers used in all DNA barcoding experiments were prepared on polyethylene glycol (PEG)-coated quartz microscopy slides according to a previously published video protocol (26). As in the protocol, a fraction of the PEG was biotinylated to enable DNA immobilization. The quality of the slide surface was further improved by 10 min incubation with 5% (v/v) Tween-20 (Sigma-Aldrich, St. Louis, MO) in T50 buffer (10 mM Tris-HCl (pH 8.0), 50 mM NaCl), followed by an extensive wash with T50 (27).

Target DNA immobilization

As targets for barcoding, 60-nt-long single-stranded DNA molecules were used (Integrated DNA Technologies, Coralville, IA and ELLA Biotech, Martinsried, Germany), which were annealed to an 18-nt biotinylated anchor sequence (sequence information can be found in Table S1). Target strands were immobilized through streptavidin-biotin conjugation of the biotinylated anchor DNA and the biotinylated PEG coating. This was achieved by incubating the slide surface with 0.1 mg/mL streptavidin (Invitrogen, Carlsbad, CA) for 1 min, followed by a 3-min incubation with 38 pM target DNA; when multiple targets were used, the total target concentration remained 38 pM. Between incubations, the channels were flushed with T50.

Barcode preparation

Short DNA oligonucleotides (Integrated DNA Technologies and ELLA Biotech) were deployed as barcodes complementary to two neighboring sites on the target DNA. The barcodes were termed upstream if they were complementary to the 5′-site and downstream if they were complementary to the 3′-site on the target sequence (see Fig. 1). To enable ligation, all upstream barcodes had a phosphorylated 5′-end, while all barcodes contained one amine modification (an amino group on a six-carbon spacer arm) for fluorescent labeling, either at an internal thymine or at the 3′- or 5′-end. Barcodes were labeled with Alexa Fluor 488 (Invitrogen), Cy3, Cy5, or Cy7 (GE Healthcare, Little Chalfont, UK) via N-hydroxysuccinimide ester cross-linking. Free dye was removed through ethanol precipitation.

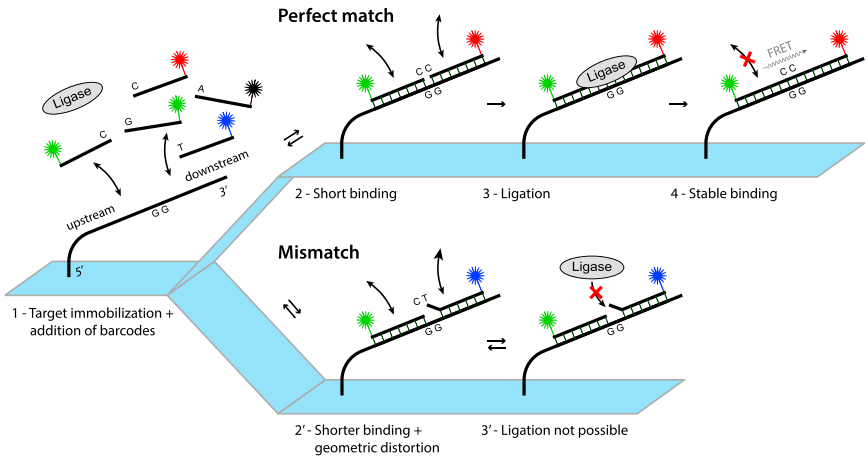

FIGURE 1 DNA barcoding experimental scheme. Target DNA strands are immobilized on a microscope slide, and dye-labeled barcodes are introduced together with T4 DNA ligase in the microfluidic chamber (1). Complementary barcodes bind transiently to the target site (2), whereas mismatched barcodes bind on an even shorter timescale (2′). Successful ligation is observed for the complementary barcodes (3) but not for the mismatched barcodes (3′). Ligation product shows stable binding to the target DNA (4), whereas mismatched barcodes dissociate and are washed away before imaging. To see this figure in color, go online.

Barcoding procedure

Immobilized target DNA was incubated with 50 nM of each upstream and 50 nM of each downstream barcode (independent of the number of different barcode sequences used) and 14 Weiss units/mL of T4 DNA ligase (Thermo Fisher Scientific, Waltham, MA) in freshly prepared ligation buffer (40 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.5 mM ATP) for 1 h at 25°C. Subsequently, the ligation buffer was replaced by imaging buffer (20 mM Tris-HCl (pH 8.0), 50 mM $MgCl_2$, 50 mM NaCl, 1 mM Trolox, 0.1 mg/mL glucose oxidase, 17 μL/mL catalase, 0.8% (w/v) glucose) to enhance the photostability of the dyes during imaging (28). For four-color imaging (see below), buffer components were dissolved in deuterium oxide instead of water to increase fluorophore brightness (29).

Although introduction of imaging buffer simultaneously removes free fluorophores and unbound barcodes from microfluidic chamber, this procedure is not a requirement for achieving a high signal/noise ratio, as the TIRF microscopy employed in this work ensures selective excitation.

Restriction reaction

Restriction of target-attached barcodes at the formed GGCC palindromic sequence was performed by incubation with 30 units/mL HaeIII (New England Biolabs, Ipswich, MA) in CutSmart buffer (20 mM Tris-acetate (pH 7.9), 50 mM potassium acetate, 10 mM magnesium acetate, 100 μg/mL bovine serum albumin; New England Biolabs) for 45 min at 25°C. Before the images were acquired, the restriction enzyme buffer was replaced by imaging buffer.

Single-molecule fluorescence microscopy

Image acquisition was performed using a prism-type TIRF microscopy setup. Alexa Fluor 488, Cy3, Cy5, and Cy7 fluorophores were excited with 473-nm (blue), 532-nm (green), 637-nm (red), and 730-nm (near-infrared) lasers (OBIS 473 LX 75 mW, Sapphire 532 LP 100 mW, OBIS 637 LX 140 mW, OBIS 730 LX 30 mW; Coherent, Santa Clara, CA), respectively. The laser beams were combined using dichroic mirrors with 523-, 544-, and 652-nm cutoff wavelengths (ZT514rdc, ZT532rdc, and ZT640rdc, respectively; Chroma, Bellows Falls, VT). Emitted fluorescence was collected using a 60× numerical aperture 1.2 water-immersion objective (Olympus, Tokyo, Japan) mounted on an inverted microscope (IX73; Olympus). The image was additionally magnified (2.5×) using two achromatic doublet lenses with 100- and 250-mm focal lengths (AC508-100-A-ML and AC508-250-A-ML, respectively; Thorlabs, Newton, NJ).

For two-color experiments, only the green laser was used. Scattered excitation light was blocked using a 488/532/635-nm triple-notch filter (NF01-488/532/635; Semrock, Rochester, NY), and the remaining signal was projected onto two halves of an electron-multiplying charge-coupled device (CCD) camera (iXon+ DU-897D; Andor Technology, Belfast, UK) using a 635-nm dichroic mirror (635dcxr; Chroma). For four-color experiments, alternating laser excitation (ALEX) (30) was performed using blue, green, red, and infrared lasers at a power ratio of 8:4:2:3; these values are proportional to the extinction coefficients of the corresponding dyes. Scattered light was blocked by a long-pass filter with 50% transmission at 482 nm (BLP01-473R-25; Semrock) and three individual notch filters with rejection peaks at 532, 633, and 730 nm (NF03-532E-25, NF03-633E-25, and the custom-made ZET730NF, respectively; Semrock). The signal was projected onto the same camera, now divided into four parts using 540-, 635-, and 740-nm dichroic mirrors (540dcxr, 635dcxr, and 740dcxr, respectively; Chroma).

Images were acquired with homemade software written in Visual C++ (Microsoft, Redmond, WA). For each experiment, 5–20 independent fields of view were imaged. It should be noted that CCD camera images shown in Figs. 2, *a–d*, S1, and S2 represent one-half of a full-sized field of view, whereas the CCD camera images in Fig. S3 represent only a quarter of a full-sized field of view.

Data analysis

Single molecules were localized in the acquired images by searching for fluorescence spots with a Gaussian profile, and after background subtraction from the single-molecule peaks (31), the intensity time traces were extracted using homemade scripts written in Interactive Data Language (ITT Visual Information Solutions, Boulder, CO). Time traces were further analyzed using MATLAB (The MathWorks, Natick, MA), in which the apparent FRET efficiency (E) and for four-color data also stoichiometry (S) were calculated:

$$E = \frac{I_{Aem}^{Dex}}{I_{Dem}^{Dex} + I_{Aem}^{Dex}}$$

and

$$S = \frac{I_{Total}^{Dex}}{I_{Total}^{Dex} + I_{Total}^{Aex}} = \frac{I_{Dem}^{Dex} + I_{Aem}^{Dex}}{I_{Dem}^{Dex} + I_{Aem}^{Dex} + I_{Dem}^{Aex} + I_{Aem}^{Aex}},$$

where I denotes the intensity of the donor or acceptor emission (*Dem* or *Aem*) upon donor or acceptor excitation (*Dex* or *Aex*).

In two-color experiments, FRET efficiency histograms were constructed, and only molecules that had FRET efficiency within a manually chosen window of 0.2 width were counted. For four-color data, because of higher complexity, more elaborate selection criteria were used.

E- and *S*-values in four-color experiments

It is important to take into consideration several factors that can influence the experimentally determined *E*- and *S*-values and cause their deviation from theoretical predictions. First, a nonzero background signal causes a slight shift of *E* and *S* from their theoretical values; this effect is most prominent for the donor-only and acceptor-only case. Second, the background signal can differ for each laser and each detection channel. Third, *E*- and *S*-values depend on the relative apparent dye intensities, determined, e.g., by dye quantum yield and detection efficiency. Fourth, a shift in *E*- and/or *S*-values may be caused by variations in focus; as refraction indices vary with wavelength, only one of the four dyes can be in focus, leaving the others slightly blurred with reduced intensity. Finally, cross-excitation and spectral bleed-through can contribute to the deviation of *E* and *S* from the expected values. In principle, one could computationally correct for many of the abovementioned factors (32,33); however, this procedure would have to be carefully applied. We found that only adjusting the ratio of different laser powers such that all dyes showed similar apparent intensities upon direct excitation was sufficient for a reliable distinction between the barcode pairs.

Barcode selection in four-color experiments

To identify the four different barcode pairs among all detected molecules in the four-color experiments, selection criteria based on intensity, stoichiometry, and FRET efficiency values were computationally established. First, the intensity range for each of the four dyes was determined by finding the number of peaks in the intensity histograms, followed by fitting those peaks with a sum of univariate Gaussian distributions, and finally by establishing the confidence intervals of the distributions (using MATLAB functions "findpeaks" and "fit"). The upper boundary for the background peak (lowest intensity) was determined using a 99% confidence interval, whereas for the boundaries of the other peaks, a 95% confidence interval was used. In some cases, in which because of FRET, more than one peak corresponded to a single fluorophore, the outermost boundaries given by those peaks were used. By comparing the intensity values of each detected single molecule with these boundaries, the presence or absence of each of the four

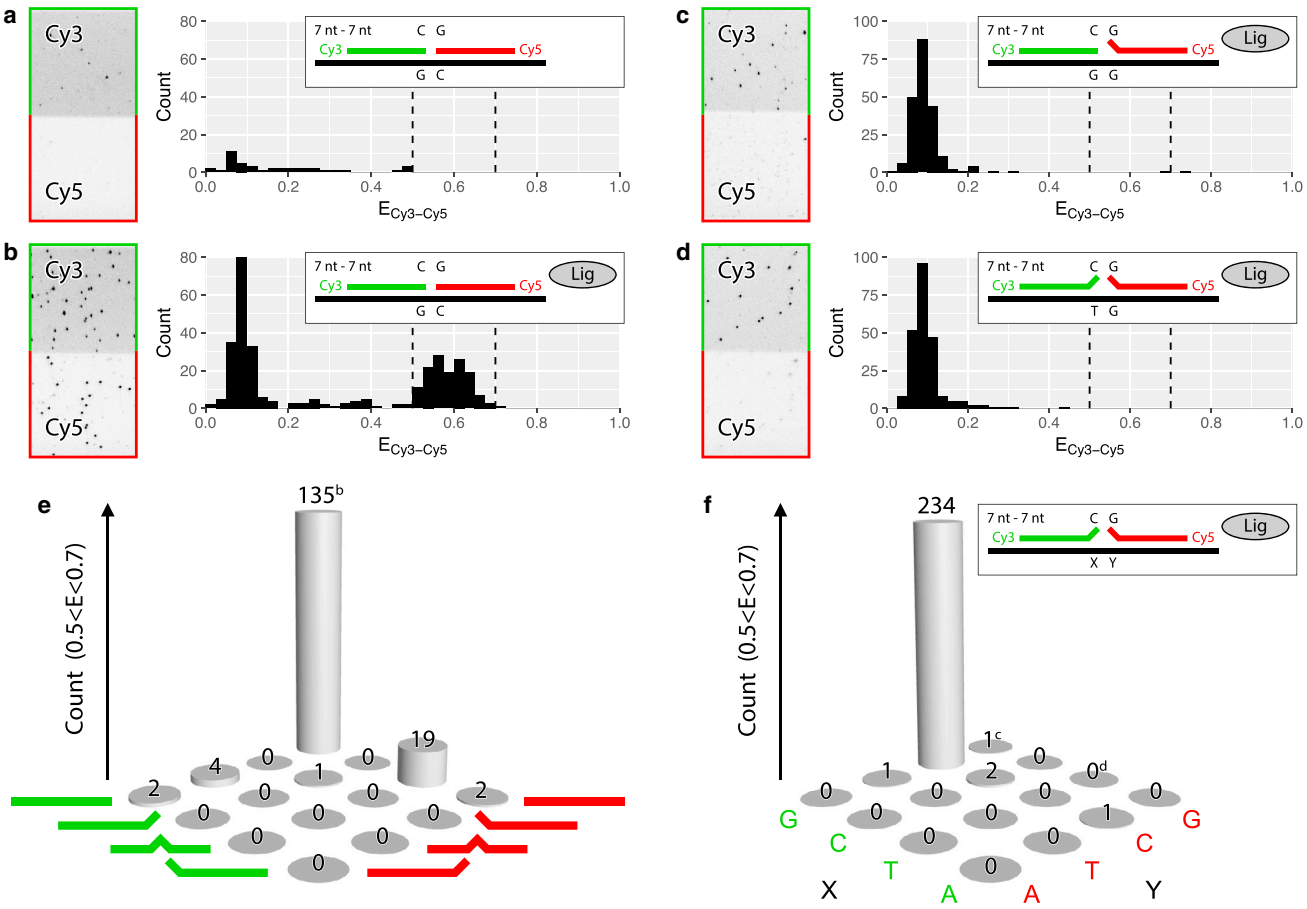

FIGURE 2 Barcode ligation enables detection of complementary target DNA. (*a* and *b*) Detection of DNA target with complementary barcodes in the absence (*a*) and presence (*b*) of ligase is shown. (*c* and *d*) Detection of DNA target with one (*c*) or two (*d*) mismatched barcodes in the presence of ligase is shown. (*e* and *f*) The effect of mismatch position within the barcode (at the 5′-end, center, or 3′-end) (*e*) and basepair identity (*f*) on the ligation efficiency is shown. Green and red lines correspond to upstream and downstream barcodes, respectively. Experiments in (*a*), (*b*), and (*e*) and (*c*), (*d*), and (*f*) are based on 5 and 15 fields of view, respectively. (*a*)–(*d*) show CCD camera images (*left*) of the Cy3 (*green*) and Cy5 (*red*) channels upon Cy3 excitation and experimental schemes and FRET efficiency (*E*) histograms (*right*), in which the dotted lines indicate the range of *E* for the target-barcode complexes ($0.5 < E < 0.7$). This range was used to determine the count in the bar plots (*e* and *f*); bars indicated with the letters "b," "c," and "d" were derived from the FRET histograms shown in (*b*)–(*d*), respectively. To see this figure in color, go online.

fluorophores was determined; the presence of double fluorophores was determined in a similar way. Second, *E-S* scatter plots were constructed for the FRET pairs that were allowed by our labeling scheme, i.e., for the A488-Cy3, Cy3-Cy5, and Cy3-Cy7 pairs. For clarity and for simplified fitting, each scatter plot only contained relevant single molecules, i.e., molecules with only one or two of the dyes in the pair and no other dyes above the background threshold (as determined by the earlier established intensity criteria). In the presence of a specific barcode pair, the corresponding *E-S* plot always showed three populations: donor-only, acceptor-only, and donor-acceptor. After constructing a two-dimensional histogram, these three populations were fitted with a sum of bivariate Gaussian distributions, and subsequently the 95% confidence interval of each population was established (using MATLAB functions "histcounts2," "kmeans," and "fit," in which "kmeans" was used to estimate the starting parameters for the fit). This resulted in an ellipsoid boundary for each population in the *E-S* scatter plot (Figs. 3, 4, S5, and S6). Selection of the donor-acceptor pairs was achieved by determining whether the combination of *E*- and *S*-values of each single molecule was located within the boundary of the donor-acceptor population. Finally, barcode selection was performed by combining the intensity criteria—having both dyes of the pair and no others present—with the criteria of FRET and stoichiometry. Selection of the molecules corresponding to the Cy3-Cy3 pair was based on intensity information only. Here, in addition to the background and the single-dye peak, the higher-intensity peak (at approximately twice the single-dye intensity) corresponding to two colocalized Cy3 fluorophores also had to be discerned.

In the four single-target experiments, each single experiment provided the selection criteria for only one of the four barcode pairs; therefore, the selection criteria of the four experiments were combined and the combination was subsequently applied to each experiment. Background and Cy3 intensity criteria could be extracted from more than one experiment; in this case, conservatively, the outermost bounds were used for selection, i.e., the largest background value and the lowest and highest Cy3 intensity (independently). In the four-target experiment, all selection criteria were determined simultaneously.

RESULTS AND DISCUSSION

Single-molecule OLA

In a conventional OLA experiment, two DNA probes hybridize to two immediately adjacent sequences on a denatured

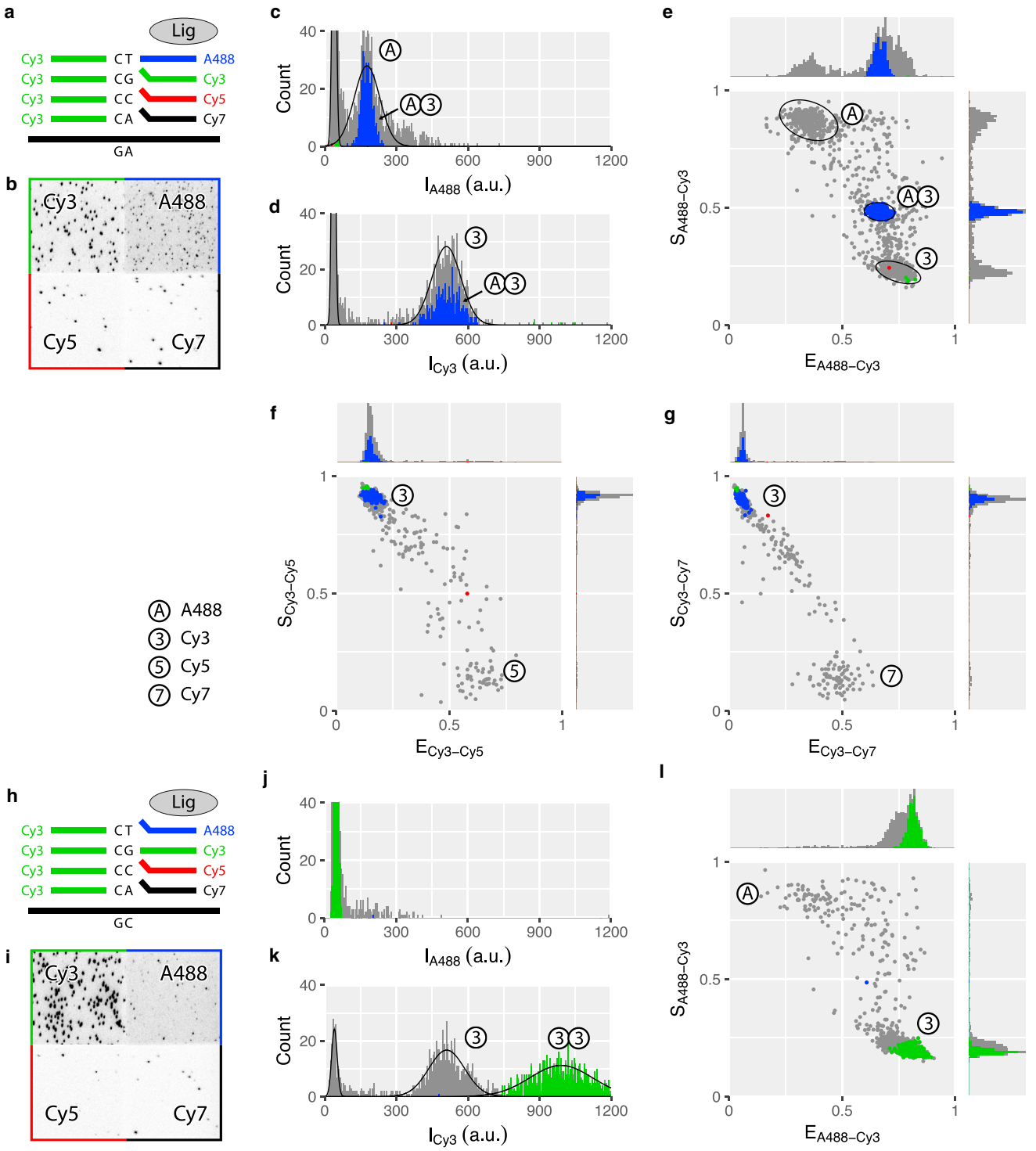

FIGURE 3 Single target complementary to one of four distinctively labeled barcode pairs. (*a* and *h*) The experimental scheme is shown. (*b* and *i*) A combined CCD camera image of the four channels (A488 (*blue*), Cy3 (*green*), Cy5 (*red*), and Cy7 (*black*)) upon respective direct excitation is shown. For the original CCD camera images used in (*b*), see Fig. S4. (*c* and *j*, *d* and *k*) Intensity (*I*) histograms for A488 and Cy3, respectively, are shown. Solid lines show fits of univariate Gaussian distributions. (*e–g* and *l*) FRET-stoichiometry (*E-S*) scatter plots for the A488-Cy3 (*e* and *l*), Cy3-Cy5 (*f*), and Cy3-Cy7 (*g*) fluorophore pairs are shown. In each plot, only relevant molecules are shown, i.e., with one or both of the two fluorophore intensities above background. Ellipses indicate the 95% confidence interval of fitted bivariate Gaussian distributions. Data are based on 20 fields of view. Molecules selected as A488-Cy3, Cy3-Cy3, Cy3-Cy5, and Cy3-Cy7 barcode pairs are indicated with blue, green, red, and black, respectively; nonselected molecules are shown in gray. The selection criteria were based on the combination of four experiments, each of them using a different target sequence; data in graphs that show fits were used for barcode selection. Single white circles indicate donor-only or acceptor-only populations, and pairs of white circles indicate donor-acceptor populations, i.e., barcode pairs. To see this figure in color, go online.

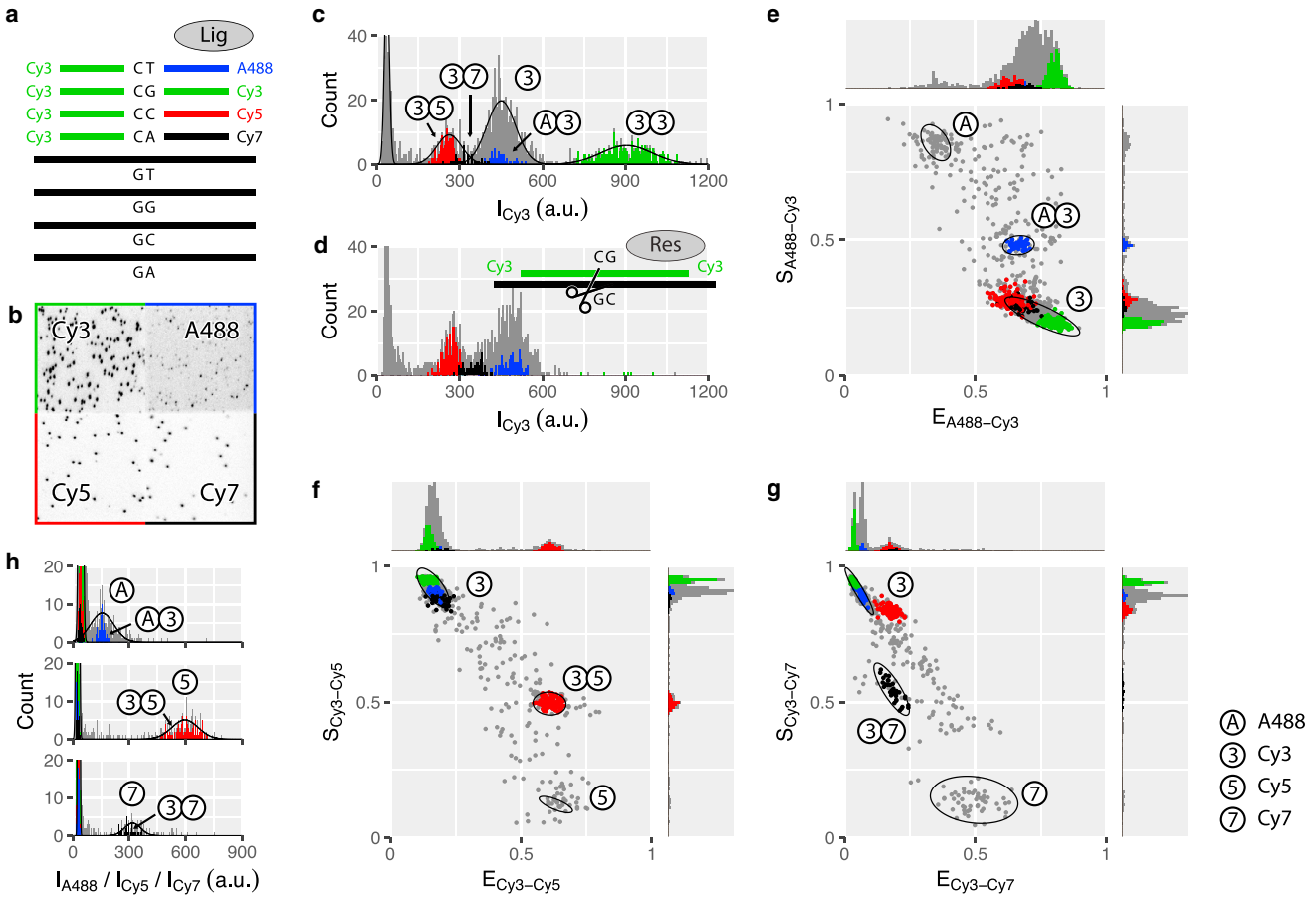

FIGURE 4 Four targets in mixture identified by four complementary barcode pairs. (*a*) The experimental scheme is shown. (*b*) A combined CCD camera image of the four channels (A488 (*blue*), Cy3 (*green*), Cy5 (*red*), and Cy7 (*black*)) upon respective direct excitation is shown. (*c* and *h*) Intensity (*I*) histograms for Cy3 (*c*) and A488, Cy5, and Cy7 (*h*) are shown. Solid lines show fits of univariate Gaussian distributions. (*d*) An intensity histogram for Cy3 after addition of a restriction enzyme specific to the bound Cy3-Cy3 barcode pair is shown. (*e–g*) FRET-stoichiometry (*E-S*) scatter plots for the A488-Cy3 (*e*), Cy3-Cy5 (*f*), and Cy3-Cy7 (*g*) fluorophore pairs are shown. In each plot, only relevant molecules are shown, i.e., with one or both of the two fluorophore intensities above background. Ellipses indicate the 95% confidence interval of fitted bivariate Gaussian distributions. Data are based on 20 fields of view. Molecules selected as A488-Cy3, Cy3-Cy3, Cy3-Cy5, and Cy3-Cy7 barcode pairs are indicated with blue, green, red, and black, respectively; nonselected molecules are shown in gray. The selection criteria were based on a single experiment with four different target sequences. Single white circles indicate donor-only or acceptor-only populations, and pairs of white circles indicate donor-acceptor populations, i.e., barcode pairs. To see this figure in color, go online.

target DNA (13,14). DNA ligase joins the 3′-end of one probe with the 5′-end of the other one only if nucleotides at the junction formed by the juxtaposed strands are correctly basepaired with the target DNA. The newly formed oligonucleotide is dehybridized from the target DNA and visualized on a separation gel either via autoradiography or fluorescence, provided one of the DNA probes was labeled with 32P or a fluorescent dye, respectively. Under certain conditions, however, this method could lead to false positive readout errors, as the length of the used probes (usually 15–20 nt) appears to be sufficient to lead to a stable probe binding to a target DNA even in the presence of a mismatch. Authors of the OLA technique reported such errors to depend on the ligase and, more importantly, salt concentration (13).

We set out to find an optimal length of the DNA probes, here called DNA barcodes, for our single-molecule fluorescence and FRET experiments. For ligation, T4 DNA ligase was used in standard conditions (25°C, 10 mM $MgCl_2$), and the GC content of the target site was ~50%. The ideal candidates would be long enough to simultaneously hybridize to the desired site on the target DNA while being sufficiently short to avoid unwanted hybridization to a mismatched target or to a nontargeted part of the sequence. Thus, we designed an assay in which two short DNA barcodes, each labeled with a fluorescent dye, are added with ligase to a microfluidic chamber with immobilized target DNA molecules (Fig. 1, steps 1–4). The barcodes are complementary to two adjacent sites on the target strand and should become ligated in the presence of the enzyme. After 1 h incubation time, the extent of ligation can be verified by monitoring FRET efficiency (*E*) between the two fluorophores upon excitation of the donor dye. *E* is defined here

as the ratio of the acceptor fluorescence intensity and the sum of the donor and acceptor fluorescence intensities. If both DNA barcodes are simultaneously hybridized to the target DNA, their fluorophores should exhibit FRET, with the FRET efficiency dependent on the labeling position (see Table S1 for detailed sequences and labeling positions).

We tested 15-nt DNA barcodes because their length falls in the size range of the original OLA constructs. The experimental scheme shown in Fig. 1 was used here in a two-color fashion, i.e., with only a single upstream and a single downstream barcode labeled with Cy3 and Cy5, respectively. The constructs showed stable binding even in the absence of ligase (as inferred from the observed signal in Cy3 and Cy5 channels in Fig. S1 *a*, as well as the presence of a FRET peak), proving undesirable for further studies. As our aim was to find a set of two DNA barcodes that only after ligation would form a stable duplex with the target DNA, we next evaluated various DNA probes shorter than 15 nt. Our single-molecule data show that under our experimental conditions, 7-nt barcodes only transiently hybridize to target sequences in the absence of ligase (Fig. 2 *a*) on a timescale shorter than the observation time. Despite such short, limiting interactions, in the presence of ligase we were able to capture stable binding events between a pair of 7-nt barcodes and the target DNA (Fig. 2 *b*). Conversely, replacing one of these DNA barcodes by a 6-nt probe led to a radical decrease in the ligation efficiency (cf. Fig. S1, *b* and *c*). This result is in accordance with the phenomenological rule of seven, postulated by Cisse et al. (34), which states that the annealing efficiency of two complementary single-stranded DNAs (ssDNAs) into a duplex drastically decreases when the number of contiguous nucleotides forming a duplex changes from seven to six. On the other hand, increasing the length of both barcodes to 8 nt resulted in an opposite effect—annealing efficiency and thereby ligation efficiency increased in comparison with the 7-nt barcodes (Fig. S2, *a–c*). However, this increase also facilitated ligation of 8-nt DNA barcodes with introduced single-point mismatches in their sequences, especially when the mismatch was located away from the ligation site. (Fig. S2, *d* and *e*). Therefore, based on these unwanted false positive readouts of mismatched 8-nt barcodes and too-low detection of 6-nt barcodes, a pair of 7-nt DNA barcodes was deemed a suitable choice for further studies.

Single-molecule OLA shows single-nucleotide specificity

To further assure the specificity of our assay, single-point mutations were introduced in both 7-nt DNA barcodes. DNA barcodes were mutated at either end (3′ or 5′) or in the center, at the third or fourth nt from the 5′-end (for detailed sequences and labeling positions, see Table S1). This resulted in a total of four probes per each upstream or downstream DNA barcode: one complementary barcode and three mutated ones. We performed 16 independent experiments in which an equimolar mixture of one upstream and one downstream DNA barcode was incubated in the presence of ligase in a microfluidic chamber with immobilized target DNA (Fig. 1). Counting the number of target DNA molecules with ligated products that showed FRET efficiencies between 0.5 and 0.7 (E range for both DNA barcodes complementary to the target strand, Fig. 2 *b*) shows only a few successful ligation events that involved one or two mismatched probes (Fig. 2 *e*), giving ∼10% sensitivity and 99.9% specificity (for a detailed description of the sensitivity and specificity, see the Supporting Materials and Methods). Our single-molecule data support our expectations (Fig. 1, steps 1–3′), based on earlier reports on the effect of the mutations at the exact ligation site on the efficiency of the enzymatic reaction (7,13).

However, we observed one exception: the combination of a complementary downstream DNA barcode with an upstream barcode with the "center" mutation, which presents 19 ligated products as compared to 135 for both barcodes being complementary to the target DNA. This mutation in the center of the upstream barcode (A was replaced by G) most likely led to the formation of a G-T wobble pair, which was postulated half a century ago by Crick (35) and was later shown in a crystal structure of A-DNA, B-DNA, and Z-DNA (36). Brown et al. also concluded that the G-T wobble can be easily accommodated in a DNA double helix without substantial perturbation of its overall conformation. Additionally, the G-T noncanonical pair was found to be the least destabilizing in NMR in vitro studies (37), whereas DNA mismatch-repair in vivo studies (38) reported on a G-T mismatch showing the highest repair efficiency among all tested mismatches. Thus, we conclude the higher number of ligated products to be a side effect of the specific mutation choice and expect only a negligent number of ligation events provided the introduced mutation would be of a different identity, e.g., if the A was replaced by a T.

Mutations introduced at either side of the ligation junction show the most disruptive effect on ligation efficiency (Fig. 2, *c* and *e*); the same effect can be seen when using 8-nt barcodes (see Fig. S2, *d* and *e*). Therefore, in our experimental design, we decided to probe the target DNA sequence at the ligation site: opposite to the 5′-end of the upstream barcode and 3′-end of the downstream barcode. To ascertain that none of the different mismatch pairs between target DNA and barcode, including potentially formed wobble pairs, give rise to false positive readouts, we measured the extent of ligation for all 16 possible sequences of the target DNA while keeping the set of DNA barcodes constant throughout all 16 experiments. Once again, only the two complementary barcodes became ligated and showed FRET, whereas introduction of even a single mismatch precluded efficient ligation (Fig. 2, *c*, *d*, and *f*). Notably, a potential G-T wobble pair also did not undergo ligation despite its rather minor effect on the global

conformation of DNA as compared to a G-C canonical base-pair (36). We attribute it to the position at which the wobble pair is present—at the ligation site—where it most likely causes greater steric hindrance, as opposed to a wobble pair further away from the ligation site (39,40). We thus showed that our assay is highly specific to even a single nucleotide mismatch at the ligation site.

To illustrate that our approach of ligating a pair of two 7-nt barcodes has a significant advantage over a simpler assay using a single 14-nt ssDNA probe, we added a single Cy3-labeled 14-nt ssDNA probe, either complementary to the target DNA or containing one or two mutations in its central part, to the microfluidic chamber with immobilized target DNA in the absence of ligase. Comparison of the number of bound molecules shows that the presence of a single mismatch in the middle of the 14-nt probe does not abolish its stable binding to the target sequence (Fig. S3, *a* and *b*). Moreover, even introduction of two mutations in the 14-nt probe does not completely prevent it from stably binding to the target DNA (Fig. S3 *c*). Therefore, a single 14-nt probe is not recommended as an alternative for ligating two 7-nt barcodes to perform accurate target recognition.

Four-color detection of four distinct barcode pairs

For the versatile use of our single-molecule method, we increased the number of spectrally distinct fluorophores from two to four by introducing Alexa Fluor 488 (henceforth called A488) and Cy7 to the previously used Cy3 and Cy5. These four fluorophores were attached to one of the four sequence variants of the downstream DNA barcode at its 5′-end while a single upstream barcode was labeled with Cy3 at its 3′-end. ALEX allowed us to cycle through four different laser excitation beams and, for each laser color, to simultaneously collect fluorescence signal in four spectrally separated channels on the CCD detector (Fig. S4).

We set out to characterize each dye pair separately. For this, we added a mixture of four downstream DNA barcodes, one upstream DNA barcode, and ligase to a microfluidic chamber with only one of the four complementary target DNA strands immobilized. Based on our two-color experiments (Fig. 2), we expected to observe only the ligation products formed by complementary DNA barcodes. For example, as in one chamber we immobilized target DNA with the GA sequence at the ligation site, we expected to detect mostly a downstream DNA barcode with T at the 3′-end, labeled with A488, and the upstream DNA barcode with C at its 5′-end, labeled with Cy3 (Fig. 3 *a*). Imaging indeed shows that most of the bound DNA barcodes are labeled with A488, Cy3, or both (Figs. 3, *b–d* and S5, *a* and *b*).

To identify target DNA molecules with two simultaneously bound barcodes, we used the detected fluorescence intensities to calculate stoichiometry and FRET efficiency, which were further combined into scatter plots (41) for each of the three possibly formed FRET pairs (Figs. 3, *e–g* and S5). In these scatter plots, FRET efficiency (E) is defined as the ratio of acceptor fluorescence intensity upon donor excitation and the sum of donor and acceptor fluorescence intensities upon donor excitation, whereas stoichiometry (S) is defined as the ratio of the sum of donor and acceptor fluorescence intensities upon donor excitation and the sum of donor and acceptor fluorescence intensities upon donor and acceptor excitation. Stoichiometry gives an indication of the presence of donor and acceptor dyes. In an ideal case without, e.g., bleed-through or cross-excitation between the channels, the donor-only population will theoretically have $E = 0$ and $S = 1$, the acceptor-only population that produces no FRET signal will have $S = 0$, and the donor-acceptor population will have $0 < E < 1$ (E-value depends on the inter-dye distance) and $0 < S < 1$ (with $S = 0.5$ when equal dye brightness is assumed). The experimental scatter plots (Figs. 3 *e* and S5) indeed show donor-only, acceptor-only, and donor-acceptor populations near the expected E- and S-values (see Materials and Methods and Hellenkamp et al. (33) for the causes of the E and S shift away from the theoretical values). Furthermore, the donor-acceptor populations are found only in the E-S plots of the expected dye pair given the sequence of immobilized target DNA (e.g., for the GA target sequence, donor-acceptor populations can be seen only in the A488-Cy3 E-S plot but not in the Cy3-Cy5 or Cy3-Cy7 E-S plots; see Fig. 3, *e–g*).

For the twin pair Cy3-Cy3, formed in the presence of the GC target sequence (Fig. 3, *h–l*), FRET efficiency and stoichiometry cannot be determined; therefore, identification of this pair had to be achieved using fluorescence intensity data alone. The Cy3 intensity histogram (Fig. 3 *k*) shows three peaks; the lowest intensity peak corresponds to the background signal, whereas the remaining two represent one Cy3-labeled barcode and two Cy3-labeled barcodes, respectively, as the latter peak has twice the intensity of the single Cy3 barcode peak. In addition, the intensity histograms of A488 (Fig. 3 *j*), Cy5, and Cy7 (Fig. S5 *b*) show no other signal than the background, further substantiating that barcodes do not bind to a mismatched target sequence. Furthermore, Fig. S5 *c* shows that the Cy3-Cy3 pair does not erroneously contribute to the donor-acceptor populations in the E-S plots of the remaining FRET dye pairs and therefore it can be reliably discriminated from these pairs.

We next quantified the bound DNA barcode pairs in each of the experiments with one of the four target molecules immobilized and four barcode pairs present in the microfluidic chamber. Selection criteria for each barcode pair were based on Gaussian fits of the intensity, FRET efficiency, and stoichiometry data (for details, see Materials and Methods). This procedure yielded 344 A488-Cy3, 681 Cy3-Cy3, 385 Cy3-Cy5, and 259 Cy3-Cy7 barcode pairs bound to their matching targets and only a few percent of the barcode pairs

bound to mismatching targets (Table S2), thus confirming the specificity of our method. The difference in the number of detected barcode pairs may be caused by the sequence at the ligation site, by the detection efficiency of the specific dyes, and/or by the used detection parameters (intensity, FRET, and stoichiometry).

Four-color scheme to distinguish four different DNA target sequences

As a proof of principle, we demonstrated detection of four different DNA sequences in mixture. We introduced the four barcode pairs into a microfluidic chamber with a mixture of the four different immobilized target sequences (Fig. 4 *a*). The camera image (Figs. 4 *b* and S6) shows bound molecules in all four detection channels. Intensity histograms and *E-S* plots (Figs. 4, *c* and *e–h* and S6, *b* and *c*) show populations well resembling those observed in the single-target experiments (cf. Figs. 3 and S5). The Cy3 intensity histogram (Fig. 4 *c*), in addition to the peaks representing single-Cy3 and double-Cy3 molecules, shows peaks of Cy3 dyes forming FRET pairs with A488, Cy5, and Cy7 (colored in *blue*, *red*, and *black*, respectively). To quantify the number of bound barcode pairs, we again applied selection criteria based on Gaussian fits (described in detail in the Materials and Methods). Using these criteria, 47 A488-Cy3, 226 Cy3-Cy3, 106 Cy3-Cy5, and 35 Cy3-Cy7 molecules were identified (Table S2).

To verify that the barcodes were correctly bound to their matching target molecules, a restriction enzyme (HaeIII) was applied that recognizes and specifically cuts the double-stranded GGCC sequence. In our experimental design, this sequence corresponds to the Cy3-Cy3 barcode pair bound to the GC target sequence. In the resulting Cy3 intensity histogram (Fig. 4 *d*), as well as the remaining intensity histograms and E-S plots (Fig. S7), the previously observed Cy3-Cy3 population disappeared, whereas the other populations remained unchanged (cf. Fig. 4 *h*). Quantitative analysis, upon applying the same selection criteria as used before adding the restriction enzyme, also shows a dramatic decrease of Cy3-Cy3 pairs from 226 to 6 without a substantial effect on the A488-Cy3, Cy3-Cy5, and Cy3-Cy7 pairs ("All" in Fig. 5 and Table S3). Additionally, when the same restriction enzyme was used in the single-target experiments, again only the Cy3-Cy3 population bound to the GC target sequence vanished, whereas other populations remained unaffected ("GA," "GC," "GG," and "GT" in Figs. 5 and S8 and Table S3).

CONCLUSIONS

Here, we showed that DNA barcoding by ligation of two fluorescently labeled 7-nt ssDNA barcodes, complementary to two neighboring sites on the target DNA strand, can be used to simultaneously distinguish at least four different DNA sequences differing by a single nucleotide.

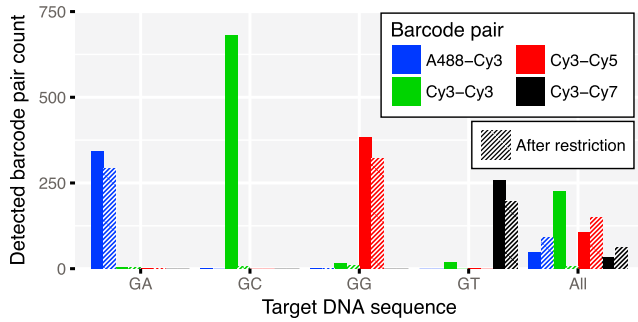

FIGURE 5 Enzymatic restriction confirms specificity of DNA barcoding. The number of barcode pairs detected in four-color single-target and four-target experiments is shown, indicated with the sequence at the ligation site ("GA," "GC," "GG," and "GT") and with "All," respectively. Hatched bars show barcode pair counts after the addition of a restriction enzyme specific to the bound Cy3-Cy3 barcode pair. To see this figure in color, go online.

In terms of kinetics, the method is reminiscent to kinetic proofreading (42), having multiple steps in which the substrate can dissociate and an energy-dependent ligation step that drives product formation. Because the specific dissociation rates are of prime importance to achieve high specificity, careful adjustment of barcode binding affinity is essential. Although many factors could have been varied, as a proof of principle, we chose to do a judicious selection of barcode length.

At our experimental conditions (25°C, 50% GC barcode content, 10 mM $MgCl_2$, 50 nM barcode concentration), we found that a length of 7 nt is large enough to allow for simultaneous hybridization of two barcodes, leading to an efficient ligation, but is sufficiently small to avoid hybridization and ligation of mismatching barcodes. Barcodes shorter than 7 nt are not readily ligated because of their low binding affinity to target DNA, whereas barcodes longer than 7 nt allow for their ligation even in the presence of mismatches, thus increasing the risk of detecting false positives.

The inability of the mismatched barcodes to be ligated is the result of a reduced binding affinity in combination with a geometrical distortion in the double helix. This gives the barcoding method its high specificity. The specific location of the mismatch within the barcode and the basepair identity do not appear to be of importance except when a G-T wobble pair is formed away from the ligation site. Barcodes with a mismatch directly adjacent to the ligation site are least likely to be ligated because of the additional spatial misalignment of the chemical groups otherwise involved in the ligation reaction.

Distinction of barcode pairs bound to different target sequences can be accomplished by fluorescently labeling each barcode, followed by detection of the different fluorophore combinations. Such detection can be achieved by employing an ALEX excitation scheme coupled with the separate

collection of the fluorescence signal from each dye. With the resulting intensity data and the calculated FRET efficiency and stoichiometry, populations of molecules corresponding to the four dye pairs can be visualized. Fitting of these populations with univariate and bivariate Gaussian distributions results in reliable selection criteria that enable computational identification of each dye pair and thereby recognition of four distinct target DNA sequences.

Our DNA barcoding technique exploits the inherent sensitivity of a single-molecule approach and therefore may be used as a method for SNP detection within low-abundance target molecules without the need for preamplification. Such SNP analysis would likely consist of the following steps: 1) extraction of genomic DNA, 2) DNA cleavage, 3) DNA melting and annealing to a biotinylated anchor DNA sequence, 4) immobilization on a quartz microscopy slide, 5) barcode ligation, and 6) imaging and analysis.

Furthermore, the multiplexing feature of the barcoding method can be used, for example, to detect multiple SNPs at once or to simultaneously perform multiple single-molecule experiments. To further increase the multiplexing potential, the number of distinguishable barcode pairs should increase. For this, one could take advantage of the remaining fluorophore combinations (e.g., A488-A488, A488-Cy5, etc.). This, together with expanding the set of the upstream barcodes, would enable the distinction of a total of 10 different barcode pairs and therefore 10 different target DNA sequences. Multiple labeling positions leading to various inter-dye distances and thus to different FRET values could be potentially explored. Even though such an expansion of barcode pairs would require an increase in the number of acquired images, it could potentially lead to a much higher number of simultaneously detected target DNA sequences.

SUPPORTING MATERIAL

Supporting Materials and Methods, eight figures, and three tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30-965-2.

AUTHOR CONTRIBUTIONS

Single-molecule experiments were designed by M.S. and C.J. and performed by M.S. and I.S.; data analysis was done by I.S.; I.S., M.S., and C.J. wrote the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

The authors thank Iason Katechis and Sungchul Kim for reading the manuscript and all members of C.J.'s group for helpful discussions.

This work was financially supported by grants from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) and European Research Council grant agreement no. 309509.

A preliminary version of this work, https://doi.org/10.1101/265215, was deposited in bioRxiv on February 13, 2018.

REFERENCES

1. Syvänen, A. C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* 2:930–942.
2. Kwok, P. Y., and X. Chen. 2003. Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.* 5:43–60.
3. Taylor, C. F., and G. R. Taylor. 2004. Current and emerging techniques for diagnostic mutation detection: an overview of methods for mutation detection. *Methods Mol. Med.* 92:9–44.
4. Andras, S. C., J. B. Power, …, M. R. Davey. 2001. Strategies for signal amplification in nucleic acid detection. *Mol. Biotechnol.* 19:29–44.
5. Yim, S. W., T. Kim, …, A. Reitmair. 2012. Four-color alternating-laser excitation single-molecule fluorescence spectroscopy for next-generation biodetection assays. *Clin. Chem.* 58:707–716.
6. Potapov, V., and J. L. Ong. 2017. Examining sources of error in PCR by single-molecule sequencing. *PLoS One.* 12:e0169774.
7. Wu, D. Y., and R. B. Wallace. 1989. Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene.* 76:245–254.
8. Chang, K., S. Deng, and M. Chen. 2015. Novel biosensing methodologies for improving the detection of single nucleotide polymorphism. *Biosens. Bioelectron.* 66:297–307.
9. Astakhova, K. 2014. Toward non-enzymatic ultrasensitive identification of single nucleotide polymorphisms by optical methods. *Chemosensors.* 2:193–206.
10. Ye, T., R. Tong, and Z. Gao. 2015. Genotyping of single nucleotide polymorphisms. *In* RNA and DNA Diagnostics. V. A. Erdmann, S. Jurga, and J. Barciszewski, eds. Springer International Publishing, pp. 123–144.
11. Tyagi, S., and F. R. Kramer. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14:303–308.
12. Tyagi, S., D. P. Bratu, and F. R. Kramer. 1998. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* 16:49–53.
13. Landegren, U., R. Kaiser, …, L. Hood. 1988. A ligase-mediated gene detection technique. *Science.* 241:1077–1080.
14. Alves, A. M., and F. J. Carr. 1988. Dot blot detection of point mutations with adjacently hybridising synthetic oligonucleotide probes. *Nucleic Acids Res.* 16:8723.
15. Chen, X., K. J. Livak, and P. Y. Kwok. 1998. A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res.* 8:549–556.
16. Nilsson, M., H. Malmgren, …, U. Landegren. 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science.* 265:2085–2088.
17. Castro, A., and J. G. Williams. 1997. Single-molecule detection of specific nucleic acid sequences in unamplified genomic DNA. *Anal. Chem.* 69:3915–3920.
18. Wabuyele, M. B., H. Farquar, …, F. Barany. 2003. Approaching real-time molecular diagnostics: single-pair fluorescence resonance energy transfer (spFRET) detection for the analysis of low abundant point mutations in K-ras oncogenes. *J. Am. Chem. Soc.* 125:6937–6945.
19. Knemeyer, J. P., N. Marmé, and M. Sauer. 2000. Probes for detection of specific DNA sequences at the single-molecule level. *Anal. Chem.* 72:3717–3724.
20. Wang, T. H., Y. Peng, …, C. M. Ho. 2005. Single-molecule tracing on a fluidic microchip for quantitative detection of low-abundance nucleic acids. *J. Am. Chem. Soc.* 127:5354–5359.
21. Zhang, C. Y., S. Y. Chao, and T. H. Wang. 2005. Comparative quantification of nucleic acids using single-molecule detection and molecular beacons. *Analyst (Lond.).* 130:483–488.
22. Zhang, C. Y., H. C. Yeh, …, T. H. Wang. 2005. Single-quantum-dot-based DNA nanosensor. *Nat. Mater.* 4:826–831.

23. Ding, F., M. Manosas, …, V. Croquette. 2012. Single-molecule mechanical identification and sequencing. *Nat. Methods.* 9:367–372.
24. Su, X., L. Li, …, C. Yu. 2017. Single-molecule counting of point mutations by transient DNA binding. *Sci. Rep.* 7:43824.
25. Levesque, M. J., P. Ginart, …, A. Raj. 2013. Visualizing SNVs to quantify allele-specific expression in single cells. *Nat. Methods.* 10:865–867.
26. Chandradoss, S. D., A. C. Haagsma, …, C. Joo. 2014. Surface passivation for single-molecule protein studies. *J. Vis. Exp*:86:e50549.
27. Pan, H., Y. Xia, …, W. Wang. 2015. A simple procedure to improve the surface passivation for single molecule fluorescence studies. *Phys. Biol.* 12:045006.
28. Rasnik, I., S. A. McKinney, and T. Ha. 2006. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods.* 3:891–893.
29. Klehs, K., C. Spahn, …, M. Heilemann. 2014. Increasing the brightness of cyanine fluorophores for single-molecule and superresolution imaging. *ChemPhysChem.* 15:637–641.
30. Hohlbein, J., T. D. Craggs, and T. Cordes. 2014. Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* 43:1156–1171.
31. Roy, R., S. Hohng, and T. Ha. 2008. A practical guide to single-molecule FRET. *Nat. Methods.* 5:507–516.
32. Lee, J., S. Lee, …, S. Hohng. 2010. Single-molecule four-color FRET. *Angew. Chem. Int.Engl.* 49:9922–9925.
33. Hellenkamp, B., S. Schmid, …, T. Hugel. 2018. Precision and accuracy of single-molecule FRET measurements – a worldwide benchmark study. *Nat. Methods.* 15:669–676.
34. Cisse, I. I., H. Kim, and T. Ha. 2012. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* 19:623–627.
35. Crick, F. H. 1966. Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* 19:548–555.
36. Brown, T., O. Kennard, …, D. Rabinovich. 1985. High-resolution structure of a DNA helix containing mismatched base pairs. *Nature.* 315:604–606.
37. Patel, D. J., S. A. Kozlowski, …, K. Itakura. 1984. Dynamics of DNA duplexes containing internal G.T, G.A, A.C, and T.C pairs: hydrogen exchange at and adjacent to mismatch sites. *Fed. Proc.* 43:2663–2670.
38. Kramer, B., W. Kramer, and H. J. Fritz. 1984. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell.* 38:879–887.
39. Pritchard, C. E., and E. M. Southern. 1997. Effects of base mismatches on joining of short oligodeoxynucleotides by DNA ligases. *Nucleic Acids Res.* 25:3403–3407.
40. Luo, J., D. E. Bergstrom, and F. Barany. 1996. Improving the fidelity of thermus thermophilus DNA ligase. *Nucleic Acids Res.* 24:3071–3078.
41. Kapanidis, A. N., N. K. Lee, …, S. Weiss. 2004. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. USA.* 101:8936–8941.
42. Hopfield, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA.* 71:4135–4139.