Single molecule screening of disease DNA without amplification

by

Ji-young Lee

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Analytical Chemistry

Program of Study Committee:
Edward S. Yeung, Major Professor
Robert S. Houk
Victor S.-Y. Lin
Mei Hong
Yeon-Kyun Shin

Iowa State University

Ames, Iowa

2006

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION 1

Dissertation Organization 1
General Introduction 1
References 5

CHAPTER 2. MOLECULAR IDENTIFICATION OF NUCLEIC ACID WITH CAPILLARY ELECTROPHORESIS AND SINGLE MOLECULE SPECTROSCOPY 10

Introduction 10

1. Single-molecule spectroscopy for molecular identification in capillary electrophoresis 13
Experimental Section 13
Results and Discussion 15

2. Single molecule detection of mRNA with Fluorescence resonance energy transfer (FRET) 17
Experimental Section 18
Results and Discussion 19

Conclusions 22
References 24

CHAPTER 3. QUANTITATIVE SINGLE MOLECULE SCREENING OF HUMAN PAPILLOMA VIRAL DNA 36

Introduction 36

1. Quantitative single molecule screening of HPV-16 DNA in flow system 39
Experimental Section 39
Results and Discussion 41

2. Quantitative single molecule screening of HPV-16 DNA using surface hybridization 46
Experimental Section 46
Results and Discussion 49

Conclusions 55
References 57

CHAPTER 4. GENERAL CONCLUSIONS 76

medicinal studies is bigger than ever, providing ultimate sensitivity and selectivity to trace and identify low copy number target bio-molecules. In the new millennium, more strategies to identify and quantify single molecule nucleic acid and proteins were continuously developed. New devices were elaborated[37-41] to facilitate individual molecule's detection and counting. Novel fluorescence labeling agents other than organic dyes were also successfully utilized in single biomolecule imaging.[42,43] The single molecule detection technique was also applied in common tasks in genetics such as single nucleotide polymorphism (SNP) detection[44-49], mRNA quantification and gene expression study[50,51], sequencing[52-55] and DNA mapping[56,57], suggesting its huge potential as a standard detection method. Physical properties of single biomolecules also continued to be investigated intensely[58-62].

Since the ultimate detection limit for any analytical assay will be counting targets one by one[63], the importance of single molecule detection was appreciated in many fields where an ultrasensitive detection technique is crucial, such as clinical screening. Sensitive detection methods mean less pathogen or abnormal targets are required to be detected, and lead to the possibility of earlier diagnosis. When they are discovered at an early stage, diseases can be treated faster, and the chance to achieve a complete cure is higher.

With its sensitivity usually higher than conventional bulk detection methods by at least 10-100 times, single molecule detection is a good candidate for a next-generation disease screening method. Many attempts have been made to detect specific sequence DNA and proteins of clinical interest at the single molecule level[64]. They successfully demonstrated detection limits lower than conventional bulk analyses, without amplification.

CHAPTER 1. GENERAL INTRODUCTION

Dissertation Organization

This dissertation begins with a general introduction. The following chapters are presented as two complete scientific manuscripts, each followed by cited literature, tables and figures. General conclusions summarize the work and provide some perspective for future research.

General Introduction

Ever since the first optical detection of small single molecules in the late 1980s[1,2], the study of the single molecule has bloomed over a decade, enriching the world with new information from the nanoscopic world. Early studies extensively explored individual molecule's physical properties such as diffusion[3,4], interactions of individual molecules[5-10], surface adsorption and interaction[11], conformational dynamics[12-19] and single molecule scale reaction[20-23]. The information from individual molecules, not scrambled by ensemble average, provided a whole new scope in science.

Soon after the single molecule detection technique emerged, it was also applied to biological and clinical fields. Research in areas such as high throughput DNA screening and single-molecule immunoassay[24], sizing and sorting of DNA[25-29], and single molecule imaging in living cells[30,31] were carried out. Some have elucidated single molecule-level mechanisms in bulk analytical separation procedures, such as gel electrophoresis[32,33], liquid chromatography and capillary electrophoresis[34-36]. Now the technique's contribution in genetic and

In practical clinical screening, however, the low copy number target molecules are mixed with an overwhelming excess of similar non-targets. Cancer cells with inherited mutation, for example, contain one or two copies of mutation in a cell, and the chance to find the DNA fragment that carries the mutated site in the entire genome mixture is nearly one out of a million. The ambiguity caused by non-targets is one of the biggest obstacles in biological and clinical analysis. Thorough identification, not just detection, is required. With single molecule detection, the information can be achieved from individual target molecules, apart from its background-generating large amount of non-targets, once proper labeling strategy is incorporated.

As much as sensitivity and selectivity, accuracy is another important aspect that a good clinical screening method should possess. One should detect or quantify the target with the least error. In many assays employed in clinical screening, the target molecule or the signal is often amplified to help detection at lower target concentration. In polymerase chain reaction (PCR), for example, a designated sequence of DNA is selectively amplified in repeated temperature cycles. It certainly helps in detecting low copy number pathogen or abnormal DNA with high selectivity and sensitivity, but its problems are also well known. Since amplification is very effective, a small contamination can be amplified as well and lead to an error. Amplification is exponential rather than linear, which makes quantification of the sample complicated. The efficiency of the reaction itself depends on many factors such as enzyme activity, temperature and metal ion concentration that should be controlled accurately altogether. In the mean time, however, single molecule detection can skip these

steps, avoiding problems caused in the course of the amplification procedure and hence provides less error-prone detection.

Bio-polymers such as DNA and proteins are large in size and can have multiple fluorophores in one molecule, which facilitates single molecule fluorescence imaging. With imaging, a large number of molecules can be analyzed at the same time in a frame of a movie. This makes the screening even faster with higher throughput. However, faster and more sensitive detection can further suggest something beyond just lower detection limit or higher throughput; sample preparation steps can be minimized as well. It takes less than a second to image and identify molecules in one movie frame, and the detection procedure can be completed before the sample-destroying reaction may proceed too far, eliminating the necessity of a clean up procedure. The ability to image molecules clearly with a high signal-to-noise ratio means the presence of impurities can be tolerated to some extent. Eliminating one preparation procedure can be important in situations where many samples have to be analyzed, or short sample handling time is crucial.

Here, we present the potential of single molecule imaging as a clinical screening method. DNA targets were detected and identified with electrophoretic mobility and a combination of probe colors. Along with the purified target-only sample, real pathogen DNA of human papillomavirus (HPV) type-16 in human genomic DNA was detected and quantified. Strategies to make the target appear more distinctively and selectively were exploited using fluorescence resonance energy transfer (FRET) with multiple color dyes, surface hybridization with washing steps to eliminate excessive non-target molecules, and a combination of two different phase probes (one tethered on the surface, the other in free

solution). They were tested with cell cultures and actual clinical samples and proved to be accurate, meaning our single molecule detection can indeed be a good clinical analysis method.

References

1. W. E. Moerner and L. Kador, *Phys. Rev. Lett.* **1989,** *62,* 2535.
2. E. Betzig and R. J. Chichester, *Science* **1993**, *262*, 1422.
3. T. H. Schmidt, G. J. Schutx, W. Baumgartner, H. J. Gruber and H. Schindler, *Proc. Natl. Acad. Sci, U.S.A.* **1996**, *93*, 2926.
4. X. Xu and E. S. Yeung, *Science* **1997**, *275*, 1106.
5. T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin and S. Weiss, *Proc. Natl. acad. Sci. U. S. A.* **1996**, *93*, 6264.
6. Y. Harada, T. Funatsu, K. Murakami Y. Nonoyama, A. Ishihama and T. Yanagida, *Biophys. J.* **1999**, *76*, 709.
7. A. D. Mehta, J. T. Finerand J, A. Spudich, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 7927.
8. Y. Sako, S. Minoghchi and T. Yanagida, *Nat. Cell Biol.* **2000**, *2*, 168.
9. S. Trabesinger, G. J. Schultz, H. J. Gruber, H. Schindler, T. Schmidt, *Anal. Chem.* **1999**, *71*, 2779.
10. J. Y. Ye, Y. Yamane, M. Yamauchi, H. Nakatsuka and M. Ishikawa, *Chem. Phys. Lett.* **2000**, *320*, 607.
11. X.-H. Xu and E. S. Yeung, *Science* **1998**, *281*, 1650.
12. C. Eggeling, J. R. Fries, L. Brand, R. Gunther, C. A. Seidel, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1556.

13. L. Edman, U. Mets and R. Rigler, *Proc Natl Acad Sci USA,* **1996**, *93*, 6710.

14. T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz and S. Weiss, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 893.

15. T. T. Perkins, S. R. Quake, D. E. Smith and S. Chu, *Science,* **1994,** *264*, 822.

16. S. R. Quake, H. Babcock and S. Chus, *Nature,* **1997**, *388*, 151.

17. S. Wennmalm, L. Edman and R. Rigler, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10641.

18. P. Schwille, S. Kummer, A. A. Heikal, W. E. Moerner and W. W. Webb, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 151.

19. X. S. Xie and R. C. Dunn, *Science,* **1994**, *265,* 361.

20. M. M. Collionson and R. M. Wightman. *Science* **1995**, *268*, 1883.

21. M. O. Vlad and J. Ross, *Chem. Phys. Chem.*, **2004**, 1671.

22. T. Ha, *Biochem.* **2004**, *3* (14), 4055.

23. F. L. H. Brown, *Phys. Rev. Lett.* **2003**, *90*(2), 028302.

24. M. R. Shortreed, H. Li, W.-H. Li and E. S. Yeung, *Anal. Chem.* **2000**, *72*, 2879.

25. A. Castro, F. R. Fairfield, E. B. Shera, *Anal. Chem.* **1993**, *65*, 849.

26. A. Van Orden, R. A. Keller and W. P. Ambrose, *Anal. Chem.* **2000**, *72*, 37.

27. H. P. Chou, C. Spence, A. Scherer and S. Quake, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11.

28. J. T. Petty, M. E. Johnson, P. M. Goodwin, J. C. Martin, J. H. Jett and R. A. Keller, *Anal. Chem.* **1995**, *67*, 1755.

29. J. J. Zheng, E. S. Yeung, *Anal. Chem.* **2003**, *75*, 3675.

30. Y. Sako, S. Minoguchi and T. Yanagida, *Nat. cell boil.* **2000**, *2*, 168.

31. H. Murakoshi, R. Iino., T. Kobayashi, T. Fujiwara, C. Ohshima, A. Yoshimura, A. Kusumi, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7317.

32. S. B. Smith, P. K. Aldridge, and J. B Callis, *Science* **1989,** *243,* 203.

33. R. M. Dickson, D. J. Norris, Y.-L. Tzeng and W. E. Moerner *Science* **1996**, *274*, 966.

34. S. H. Kang, M. R. Shortreed, and E. S. Yeung, *Anal. Chem.* **2001**, *73*, 1091.

35. S. H. Kang and E. S. Yeung, *Anal. Chem.* **2002**, *74*, 6334.

36. H.-W. Li, H. –Y. Park, M. D. Porter and E. S. Yeung, *Anal. Chem.* **2005**, *77*, 3256.

37. A. Crut, B. Geron-Landre, I. Bonnet, S. Bonneau and P. Desbiolles, *Nucl Acids Res.* **2005**, *33*, e98.

38. K. Dörre, J. Stephan and M. Eigen. *Single Mol.* **2001**, *2,* 165.

39. A. Ros, W. Hellmich., T. Duong and D. Anselmetti, *J. Biotech*, **2004**, **112**, 65.

40. T.-H. Wang, Y. Peng, C. Zhang, P. K. Wong, and C.-M. Ho, *J. Am. Chem. Soc.* **2005**, *127,* 5354.

41. N. K. Lee, A. N. Kapanidis., Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright,. And S. Weiss, *Biophs. J.* **2005**, *88*, 2939.

42. W. C. Chan and S. Nie, *Science* **1998**, *281*(5385), 2016.

43. M. Han, X. Ga., J. Z. Su, and S. Nie, *Nat. Biotech.* **2001**,*19*, 631.

44. H. –C. Yeh, Y.-P. Ho, I. –M. Shih and T.-H.Wang, *Nuc. Acids Res.* **2006**, *34*, e35.

45. C.-Y. Zhang, H.-C. Yeh, M. T. Kuroki and T. H. Wang, *Nat. Mat.* **2005**, *4*, 826.

46. B. Nie, M. R. Shortfreed, and L. M. Smith, *Anal. Chem.* **2005**, *77*, 6594.

47. C. R. Twist, M. K. Wilson, J. J. Rowland and D. B. Kell, *Anal. Biochem.* **2004**, *327*, 35.

48. P. M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D. C. Thomas, D. C. Ward, *Nat. Genet.* **1998**, 19.

49. N. Marme, A. Friedrich, M. Muller, O. Nolte, J. Wolfrum, J. D. Hoheisel, M. Sauer. J. -P. Knemeyer. *Nucl. Acids Res.* **2006**, *34*, e90.

50. A. Camacho, K. Korn., M. Damonda, J. -F. Cajot, E. Litborn, B. Liao, P. Thyberg, H. Winter, A. Honegger, P. Gardellin.and R. Rigler *J. Biotech.* **2004**, *107*, 107.

51. K. Korn, P. Gardellin, B. Liao, M. Amacker, A. Bergstroém, H. Bjoérkman, A. Camacho, S. Doérhoèfer, K. Doérre1,; J. Enstroém, T. Ericson, T. Favez, M. Goésch, A. Honegger, S. Jaccoud, M. Lapczyna, E. Litborn, P. Thyberg, H. Winter and R. Rigler, *Nucl. Acids Res.* **2003**, *31*, e89.

52. R. Lebofsky and A. Bensimon. *Brief. in funct. genomics and proteomics,* **2003**, *1*, 385.

53. Y. Astier, O. Braha. and H. Bayley, *J. Am. Chem. Soc.* **2006**, *128*, 1705.

54. X. Michalet, R. Ekong, F. Fougerousse, S. Rousseaux, C. Schurra, N. Hornigold, M. van Slegtenhorst, J. Wolfe, S. Povey, J. S. Beckmann, A. Bensimon, *Science* **1997**, *277*, 1518.

55. I. Braslavsky, B. Hebert., E. Kartalov, and S. R. Quake *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3960.

56. E. Y. Chan, N. M. Goncalves, R. A. Haeusler, A. J. Hatch, J. W. Larson, A. M. Maletta., G. R. Yantz, E. D. Carstea,. M. Fuchs, G. G. Wong, S. R. Gullans, R. Gilmanshin, *Genome Res.* **2004**, *14*, 1137.

57. S. Zhou, A. Kile, M. Bechner, M. Place, E. Kvikstad, W. Deng, J. Wei, J. Severin, R. Runnheim, C. Churas, D. Forrest, E. T. Dimalanta, C. Lamers, V. Burland, F. R. Blattner, D. C. Schwartz, *J. Bacteriol.* **2004**, *186*, 7773.

58. K. A. Schmidt, C. V. Henkel, G. Rozenberg and H. P Spaink, *Nucl . Acids Res.* **2004**, *32*, 4962.

59. N. H. Green, P. M. Williams, O. Wahab, M. C. Davies, C. J. Roberts, S. J. B. Tendler, and S. Allen, *Biophys. J.* **2004**, *86*, 3811.

60. V. Namasivayam, R. G. Larson, D. T. Burke and M. A. Burns, *Anal. Chem.* **2003**, *75*, 4188.

61. H. Yan, H.-W. Li, and E. S. Yeung *J. Phys. Chem. B* **2005**, *109*, 8820.

62. M. Singh-Zocchi, S. Dixit, V. Ivanov, and G. Zocchi, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100,* 7605.

63. T. Anazawa, H. Matsunaga and E. S. Yeung, *Anal. Chem.* **2002**, *74*, 5033.

64. J. Li, J. -y. Lee and E. S. Yeung, *Anal. Chem.* **2006**, *78*, 6490.

CHAPTER 2. MOLECULAR IDENTIFICATION OF NUCLEIC ACID WITH CAPILLARY ELECTROPHORESIS AND SINGLE MOLECULE SPECTROSCOPY*

Introduction

Single molecule detection has been successfully applied to various genetic studies: small abnormality detection in bio-molecules[1-4], quantification of gene expression[5,6], DNA and protein binding mechanisms and enzyme kinetics studies[7,8]. Mutation or single nucleotide polymorphism (SNP) detection was one of the most actively studied areas using single molecules, suggesting its further application as a clinical screening method.

In genetic and clinical studies dealing with nucleic acids, using sequence specificity is the most common way of differentiation. Restriction enzymes that recognize and cleave certain sequences of nucleic acid are often used to detect sequence difference or mutation (restriction fragment length polymorphism, RFLP[9]). A single nucleotide difference can eliminate or create a cutting site of a restriction enzyme and result in different size fragments from wild type digestion. When the mutation site does not have an enzyme cutting site, PCR amplification was used to add the site near mutation and amplify[10]. Once a different size fragment is generated, gel electrophoresis is typically used for separation and identification. However, conventional slab-gel electrophoresis is slow and insensitive. A single molecule scale substitute will provide a good answer to solve these problems. Several attempts have

* Reprinted in part from Journal of Chromatography A, 1053 (2004), Ji-young Lee; Hung-Wing Li; Edward S. Yeung, Single-molecule spectroscopy for molecular identification in capillary electrophoresis, 173–179, Copyright(2004), with permission from Elsevier

been made to scale down restriction fingerprinting and gel electrophoresis[11-13], but often the preparation and detection were still complicated, and the number of molecules analyzed at one time was low.

Hybridization of labeled probe molecules to target is another strategy in genetics. Nucleic acid's ability to hybridize only with a complementary sequence makes a selective assay. This strategy was widely adopted in single molecule detection as well, where no bulk amount is needed to collect enough signals from targets. Probes labeled with a single fluorophore or more efficient quantum dot are hybridized to targets and create a unique signal, distinctive from unbound probes or non-targets. The FRET system was popularly employed to make the target hybrid look different as a well as a coincidence of two spectra.

In this study, a new method of measuring the electophoretic mobility of individual molecules was tried. The single-molecule imaging procedure can measure the electrophoretic mobilities of many distinct molecules in a fraction of a second. The molecules are not separated as in conventional electrophoresis, but are simply identified and counted with close to 100% efficiency while they pass through the detection area (Figure 1). Virtually all electrophoresis-based analysis protocols from slab gels to capillary electrophoresis should be adaptable to single-molecule detection.

We have also developed a high-speed high-throughput single-molecule imaging technique for identifying molecules that are labeled with different dyes in free solution based on differences in their fluorescence emission spectra[14]. Recording the emission spectra of individual molecules is considered challenging, since only a limited number of photons can be collected from a molecule before photobleaching sets in[14]. Further, instrumental setup that

are employed in most single molecule spectroscopy investigations are complicated, involving filters, dichroic mirrors, and multiple detectors. However, in our system, simple insertion of a transmission grating in the fluorescence pathway allowed recording of the entire spectrum, rather than selected wavelengths through optical filters, without additional optics or detectors. The complete emission spectra of many individual molecules are imaged in the millisecond time scale and recorded simultaneously while they are freely moving in the observation region. This technique can be used for screening single molecules for disease markers and for monitoring individual molecular interactions at a rate of thousands of molecules per second.

We combine the best features of electrophoretic and spectroscopic discrimination among single molecules. In addition to size-based identification, the spectral information from individual molecules was collected while it moved along the flow in a capillary under electrophoresis. At high signal-to-noise (S/N) conditions, electrophoresis provides confident identification of the molecules. At low S/N, the spectroscopic information allows identification even when the electrophoretic mobilities are similar.

Beta-actin mRNA of ~600 nt in length was detected with single molecule detection in capillaries as well. The mRNA was probed by hybridization to fluorescently labeled cDNA, taking advantage of FRET. It was stained with Alexa Fluor 488 dye and hybridized to a synthetic cDNA probe which was over-labeled with Alexa Fluor 532 to be invisible. After hybridization, energy transfer made the hybrid bright, with only Alexa Fluor 532 dye emission. Taking advantage of the sensitive and fast detection of single molecule spectroscopy, the target mRNA-cDNA hybrid was further imaged in the presence of bio-

matrices such as blood, plasma and saliva. Its sensitivity allowed up to 50% of bio-matrix in the sample, suggesting less purification steps in sample preparation.

I. Single-molecule spectroscopy for molecular identification in capillary electrophoresis

Experimental Section

Buffer solutions. Polyoxyethylene-6-cetylether (C16E6) is a monomeric non-ionic surfactant used as the sieving matrix for the separation of DNA fragments in capillary electrophoresis[15]. 0.25% (w/w) C16E6 (Sigma, St. Louis, MO, USA) prepared in 10 mM Gly-Gly buffer (Sigma) was used in all experiments. Gly-Gly was dissolved in ultrapure water and adjusted to pH 8.2 by dropwise addition of 2 M of NaOH. The buffer solution was filtered with a 0.2 μm filter. C16E6 was dissolved in Gly-Gly buffer at 90 °C and cooled to room temperature with constant stirring at medium speed. The C16E6 solution was also filtered with a 0.2 μm filter before use.

DNA samples. For electrophoresis and spectroscopic experiments, M13mp18 (7249 bp) and M13KE (7222 bp) DNA were purchased from New England Biolabs. Both DNA were cut with Acc65I and BsmI restriction enzymes to generate 4497 and 2752 bp DNA fragments for M13mp18 and 7088 bp and 135 bp for M13KE. The resulting fragments were also labeled fluorescently; M13mp18 fragments were labeled with YOYO-I and M13KE

fragments with POPO-III. Both YOYO-I and POPO-III dyes were used at a ratio of one dye molecule per five base pairs according to the manufacturer's instructions. All samples were allowed to incubate for 5 min before they were further diluted. Lambda DNA was prepared at a concentration of 200 pM and diluted to 0.2 pM. M13mp18 and M13KE DNA samples were prepared at 1.4 nM and diluted to 2.8 pM in 0.25% (w/w) C16E6 solution prior to the experiments.

Instrument setup. A 27 cm long square capillary (75 μm i.d. × 365 μm o.d.; Polymicro Technologies, Phoenix, AZ, USA) was used for all electrophoresis experiments, The capillary was mounted on a Bakalite holder, which was placed on a Zeiss Axioskop upright microscope. The light source and detection system were very similar to those used in previous studies done in our group[16,17]. A Uniphase (San Jose, CA, USA) argon ion laser operating at 488 nm with 12.2 mW output power was used as the excitation source. The laser beam was focused by a 2.5 cm focal length cylindrical lens (Edmund Industrial Optics, Barrington, NJ, USA) so that the focal point was at the center of the capillary. Fluorescence from single DNA molecules was collected with a Zeiss 20× Fluar (0.75 NA) microscope objective at right angle to the incident laser beam. Two 488 nm holographic notch filters (Keiser Optical, Ann Arbor, MI, USA; HNFP) with optical density >6 were placed between the microscope objective and a Pentamax 512-EFT/1EIA intensified charge-coupled device camera (ICCD, Princeton Instruments, Princeton, NJ, USA). The ICCD collected images with 30 ms exposure time at 2 Hz. The laser beam was chopped synchronously to the ICCD image collection rate with a Uniblitz mechanical shutter (LS2Z2, Vincent Associates, Rochester, NY, USA) and a driver (model T132, Vincent Associates) to reduce

photobleaching. A transmission grating with 70 grooves/mm (Edmund) was mounted in front of the ICCD camera to disperse fluorescence light from the molecules. The distance between the transmission grating and the ICCD was set to be 4 cm so that the zeroth-order and first-order images of the 75 μm capillary could not overlap each other.

Electrophoresis conditions. The capillary was equilibrated with 0.25% C16E6 in 50 mM Gly-Gly solution (pH 8.2) for 15 min before use. The entire capillary was filled with sample solution with a syringe 5 min before irradiation by the excitation beam. Before each electrophoresis experiment, the heights of the buffer reservoirs were adjusted while molecular movement was monitored until no hydrodynamic flow existed in the capillary. Electric field, 1 kV/27.0 cm d.c., was applied with a Spellman high-voltage d.c. supply (Hauppauge, NY, USA). Data collection started 10 s after applying voltage. Between electrophoresis experiments, the capillary was washed with 0.25% C16E6 in Gly-Gly solution for 5 min.

Results and Discussion

Fluorescence imaging with transmission grating. A fluorescence image from single DNA molecules in a capillary dispersed with the transmission grating is shown in Figure 2. The capillary was filled with a 1:1 mixture of YOYO-I-labeled Lambda DNA and POPO-III-labeled Lambda DNA, both dissolved in 0.25% (w/w) C16E6 in 50 mM Gly-Gly buffer (pH 8.2). Since POPO-III has an absorption maximum at 534 nm and has an efficiency of 26% at 488 nm, POPO-III labeled DNA was not as bright as YOYO-I-labeled DNA that has an absorption maximum at 491 nm. With the distance of 4 cm between the transmission grating

and the ICCD, the non-dispersed zeroth-order image (left half) of the 75 μm capillary is imaged simultaneously with the dispersed first-order image (right half).

The zeroth-order image is identical to that without transmission grating and can be used to calculate the mobility of individual DNA in the sieving matrix under electric field, just as in capillary gel electrophoresis. The zeroth-order capillary image was also used as a reference point to determine the dispersion distance between zeroth-order and first-order images. The emission maximum at 509 nm for YOYO-I and 570 nm for POPO-III resulted in dispersion distances of 1.425 mm for YOYO-I and 1.597 mm for POPO-III at this setting. Counting the number of pixels between the highest intensity pixels in the zeroth-order and corresponding first-order images allows the molecules to be distinguished from each other.

DNA fragments after restriction enzyme reaction. M13KE is a derivative of M13mp19 and is different from M13mp18 by 26 bp. This small difference, however, creates different cutting sites for Acc65 I enzyme. BsmI cuts both at the same location to open the circular DNAs. As a result, M13KE cleaves into 7088 and 135 bp fragments, while M13mp18 DNA gives 4497 and 2752 bp fragments after cutting with Acc65I and BsmI restriction enzymes. These two DNA mimic a situation with normal DNA and a DNA with mutation caused by disease.

Electrophoretic and spectroscopic data. Figure 3 is a histogram of mobility of POPO-III-labeled M13KE and YOYO-I-labeled M13mp18 when the laser power is not high enough to provide adequate S/N ratio. The low S/N condition provides a stringent test of our discrimination scheme. The low laser power is required to avoid photobleaching over the series of successive images. The x-axis is the number of pixels traveled by the molecules in 1

second and the y-axis is the number of molecules. Picking molecules in these movies and calculation of their mobilities were done with the in-house software described in the image analysis section. Only the zeroth-order image area was subjected to the calculation of mobilities. Since the smallest DNA fragment that can be seen in this setup is ~800 bp[18], the 135 bp fragment from M13KE is not shown. As shown in Figure 3, the mobility differences found for 7088 bp M13KE and 4497 bp and 2752 bp M13mp18 were not distinct. With S/N = 2, the 2752 bp fragments were not picked up by the software. Table 1 shows that lower S/N from the less efficient emitter, POPO-III-labeled M13KE, resulted in a broad range of measured mobilities (with a standard deviation of almost 50%), which even covers the entire M13mp18 mobility region. So, it is not possible to identify molecules simply by mobility.

Figure 4 is the wavelength dispersion distribution of POPO-III-labeled M13KE and YOYO-I-labeled M13mp18. Higher laser powers can be used since only one image is required for spectral determination. From Table 1, the average dispersion of POPO-III is 87.9 pixels with a standard deviation of 2.16 and that of YOYO-I is 79.5 pixels with a standard deviation of 1.78. The difference between the two averages is 8.43, which is 4.3 times the average of the two standard deviations. With more than four times the standard deviation, the two peaks are >99% separated from each other[19]. In other words, when DNA molecules are labeled with different dye colors, their identification can be done with greater than 99% accuracy.

II. Single molecule detection of mRNA with fluorescence resonance energy transfer (FRET)

Experimental Section

Reagents. Nuclease-free water (Ambion, Austin, TX) was used for the preparation of all reagents. Digoxigenin-labeled human actin RNA probe (588 nt) was purchased from Roche (Indianapolis, IN) and labeled using a Ulysis Alexa Fluor 488 nucleic acid labeling kit (Molecular Probes). Human beta-actin cDNA was prepared by PCR amplification using a Human beta-Actin Control Amplimer Set (BD Biosciences Clontech, Palo Alto, CA) and SuperTaq Plus polymerase (Ambion, Austin, TX). The positive control in the set was used as a template. The resulting size (~800 bp) and concentration (0.1μg/ul) was estimated from the agarose gel electrophoresis result. The PCR product went through a cleaning procedure using a Qiaquick PCR purification kit (Qiagen, Valencia, CA) and ethanol precipitation. Ulysis Alexa Fluor 532 (Molecular Probes) was used to label cDNA. The labeling reaction for Alexa Fluor 488 was carried out according to the manufacturer's instructions. For cDNA labeling, five times higher amount of dye was used along with a longer reaction time (~30 min) at 80 °C than that suggested in the manual.

After the labeling reaction, the samples were further cleaned with a Qiaquick PCR purification kit one more time to eliminate excess labeling reagents. Assuming 100% yield for every cleaning and precipitation step, the concentration of the samples was 20 μg/ml for cDNA and 5 μg/ml for mRNA after the final clean-up procedure. For single component experiments, cDNA was diluted 5000-fold and mRNA 100-fold.

The hybridization reaction mixture was prepared with 10 μl cDNA and 20 μl mRNA mixed with 100 μl DIG Easy Hyb solution (Roche, Indianapolis, IN). The mixture was kept at 95 °C for 10 min for denaturation and 50 °C for 16 h for hybridization. The hybrid sample was

diluted 2×10^4 times with Nuclease-free water before imaging. For the control sample, the same concentration mixture of mRNA and cDNA as in the hybrid sample was prepared in DIG Easy Hyb solution and did not undergo hybridization steps.

Bio-matrix preparation. Two 5 ml tubes of whole blood were donated by a healthy donor. Plasma was prepared from one tube of blood, by applying centrifugation at ~28,800 g to separate blood cells from plasma. Half of the plasma was photobleached under a UV lamp overnight. Whole blood and plasma were diluted with deionized water just prior to mixing with the mRNA-cDNA hybrid samples, and the mixture was further diluted with deionized water.

Instrumental setup. The instrument setup is similar to that used in previous research[16,17]. A 30 cm long 75 μm i.d. and 365 μm o.d. square capillary was filled with sample before the image collection was started. A 25 mW 488 nm Ar-ion laser beam was focused with a 2.5 focal length cylinder lens and illuminated from a side of the capillary.

Fluorescence emission was collected at right angles to the excitation beam by a Zeiss 20× Fluar (0.75 NA) microscope objective connected to an intensified CCD (ICCD, Princeton Instruments, Princeton, NJ). A transmission grating of 70 grooves/mm (Edmund Industrial Optics, Barrington, NJ) was placed ~3 cm before the CCD in the collection beam path to split the light from samples into zero- and first-order images. The ICCD collected images with a 150-ms exposure for the cDNA and hybrid sample and 100 ms for mRNA, both at 2 Hz.

Results and Discussion

Table 2 shows the average dispersion distance and relative fluorescence intensity of three

species imaged in this study. With ~3 cm distance between the transmission grating and ICCD, the dispersion of fluorescence from samplse between zero- and first-order images were 64.6, 68.7, and 68.2 pixels for mRNA with Alexa Fluor 488, cDNA with Alexa Fluor 532, and hybrid, respectively. Figure 5 shows single molecule images of three species in a capillary. At the imaging condition of cDNA and hybrid with 150 ms exposure of 25 mW laser power, mRNA labeled with Alexa Fluor 488 had signal saturation and was imaged with 100 ms instead. Signals from hybrid and Alexa Fluor 532 labeled cDNA were weaker than that of mRNA and imaged at 150 ms exposure. cDNA molecules appeared very dim and it was hard to recognize them individually. With five times' the excessive amount of Alexa Fluor 532 than suggested by the manufacturer, cDNA molecules were stained to be 'invisible'. With 488 nm excitation light, the fluorescence efficiency for Alexa Fluor 532 is lower than maximum, but it is also believed that over-labeling resulted in shorter distances between covalently-bound dye molecules and caused quenching. The importance of the invisible probe is that it does not contribute to the background fluorescence much and hence allows good signal-to-noise ratio even at high probe concentration. However, when the cDNA was denatured and hybridized to the mRNA target, energy transfer took place to enhance the fluorescence of Alexa Fluor 532. Fluorescence from the hybrid was higher than that from cDNA by 2 to 2.5 times. The concentration used for single species movies was higher than the starting material concentration for the hybrid; 100 times for mRNA and 10 times for cDNA. For the control experiment, mRNA was mixed with cDNA at the same ratio as in the hybrid, but without the hybridization process. Overall, a smaller number of molecules appeared than in the hybrid movie, and their dispersion distance was ~ 65 pixels, indicating that they are all mRNA molecules. There was no molecule with ~ 68 pixel

dispersion distance that has comparable intensity with hybrids found in the control movie.

Figure 6 shows an overlaid histogram of three species. The difference in dispersion distances between Alexa Fluor 488 labeled mRNA and the hybrid is 3.6 pixels and it is 2.53 times average standard deviation, which should mean more than 98.8%[19] of molecules can be identified from each other in this system.

The hybrid was further imaged in the presence of bodily fluid to estimate effects of crude sample preparation. Table 3 shows the maximum percentage of whole blood, plasma and saliva contents, and their signal and noise levels.

Employing deionized water to dilute blood and plasma helped rupture the blood cells by creating hypotonic conditions. To minimize possible destruction of target with enzymes in bio-matrices, the hybrid was added right before imaging. It took no more than 1.5 min from the addition of hybrid to bio-matrix to the start of a movie. Within the time period of making a movie (~20 sec, 40 frames/movie with 2 Hz imaging frequency), the loss of sample was not noticeable. With the detection procedure within 2 min, it was not necessary to deactivate or eliminate nucleases in bio-matrices. If an adequate labeling reaction is employed and can be completed before losing the target dramatically, the whole assay can be done without complicated and precautious purification steps.

With fluorescence from protein and blood cell materials, the background level was raised in the presence of these bio-matrices (Figure 7). Whole blood and unbleached plasma allowed the hybrid's imaging with their contents of up to 8%. When plasma was photobleached overnight under a UV lamp to lower auto-fluorescence, our detection system could tolerate

up to 50% of its content to image beta-actin mRNA-cDNA hybrid. Saliva could be present to 50%, but it contained bright particles that could be mistakable for hybrids in monotonous imaging. The dispersion distance of this particle was 63 pixels, which indicated it did not have Alexa Fluor 488 or Alexa Fluor 532. They were thought to be cheek cells or cell debris from the oral cavity.

Conclusions

Single molecule DNA was imaged and identified in a thin capillary, using two criteria: electrophoretic mobility and fluorescence spectrum. In the identification of molecules with electrophoretic mobility, molecules were observed in three consecutive frames and the mobility information was obtained before the sample was separated as in conventional bulk capillary electrophoresis. Multiple fluorescence dyes were used to add spectroscopic information of the molecules. The fluorescence spectrum of individual DNA molecules was recorded while it passed through the detection window during electophoresis. No complicated instrumental addition was necessary. A transmission grating inserted in the fluorescence collection pathway dispersed the light from DNA molecules which appeared in part of the frame of the image collected. It took only one frame to obtain spectrum from each molecule with this system, and high laser power could be used to extract all available photons out of the molecules. With conditions where photobleaching of dyes is of concern and electrophoretic mobility measurements are limited by it, spectral information can be added to identify molecules with >98.8% accuracy.

Hybridization of fluorescently labeled probe to target mRNA was performed and detected in the instrument set-up. Beta-actin mRNA was stained with shorter wavelength emitting dye

(Alexa Fluor 488) and probe cDNA was stained with longer wavelength Alexa Fluor 532. The probe cDNA was deliberately over-labeled to give out the least fluorescence with the detection condition. When proper hybridization takes place, energy from the Alexa Fluor 488 (donor) transferred to the Alexa Fluor 532 molecule in probe strand (acceptor) and the hybrid molecule emitted acceptor dye fluorescence with enhanced intensity. The negative control showed no molecule with Alexa Fluor 532 emission.

The beta-actin mRNA-cDNA hybrid was further imaged with bio-matrices added to the sample. Whole blood, blood plasma, and saliva ware tested. Up to 8% for whole blood and plasma, and up to 50% of photobleached plasma and saliva was tolerated before the signal-to-noise ratio dropped below 3. Bright light scattering from impurities such as cells or cell debris could be ruled out because its spectrum differed from the target molecules'.

Our strategy of single molecule nucleic acid detection provided a fast and accurate way to identify individual analytes. Taking less than a second to make a frame of a movie of the target molecules, along with the presence of many molecules in one frame, it provides a high-throughput method to screen DNA more efficiently.

Acknowledgments

We thank Gang Xue for developing the in-house image analysis software. E.S.Y. thanks the Robert Allen Wright Endowment for Excellence for support. The Ames Laboratory is operated for the US Department of Energy by Iowa State University under Contract No. W-7405-Eng-82. This work was supported by the Director of Science, Office of Basic Energy Sciences, Division of Chemical Sciences and by the National Institutes of Health.

References

1. H. –C. Yeh, Y.-P. H., I. –M. Shih and T. –H. Wangl, *Nuc. Acids Res.* **2006**, *34*, e35
2. C. –Y. Zhang, H.-C. Yeh., M. T. Kuroki and T. –H. Wang, *Nat.Mat.* **2005**, *4*, 826.
3. T. Hirose, T. Otani, H. Muramatsu and A. Tanaka, *Photochemistry and Photobiology.* **2002**, *76*, 123.
4. C. R. Twist, M. K. Winson., J. J. Rowland and D. B. Kell, *Anal.l Biochem.* **2004**, *327*, 35.
5. K. Korn, P. Gardellin, B. Liao, M. Amacker, A. Bergstroém, H. Bjoérkman, A. Camacho, S. Doérhoèfer, K. Doérrel,; J. Enstroém, T. Ericson, T. Favez, M. Goésch, A. Honegger, S. Jaccoud, M. Lapczyna, E. Litborn, P. Thyberg, H. Winter and R. Rigler, *Nucl. Acids Res.* **2003**, *31*, e89.
6. A. Camacho, K. Korn., M. Damonda, J.-F. Cajot, E. Litborn, B. Liao, P. Thyberg, H. Winter, A. Honegger, P. Gardellin.and R. Rigler *J. Biotech.* **2004**, *107*, 107.
7. B. Schafer, H. Genmeinhardt and K. O. Greulich, *Angew. Chem. Int. Ed.* **2001**, *40*, 4663.
8. H. –W. Li and E. S. Yeung, *Anal. Chem.* **2005**, *77*, 4374.
9. C. Pourzand and P. Cerutti, *Mutation Res.* **1993**, *288*, 113.
10. B. L. Parsons and R. H. Heflich, *Mutation Research,* **1997**, *387*, 97.
11. J. R. Taylor, M. M. Fang and S. Nie, *Anal. Chem.* **2000**, *72*, 1979.
12. M. M. Ferris, T. M. Yoshida, B. L. Marrone and R. A. Keller, *Anal..Biochem.* **2005**, *337*, 278.
13. B. Geron-Landre, T. Roulon, P. Desbiolles1 and C. Escude, *Nuc. Acids Res.* **2003**, **31,** e125.
14. Y. Ma, M.R. Shortreed, and E.S. Yeung, *Anal. Chem.* **2000**, *72*, 4640.

15. W. Wei and E.. S. Yeung *Anal. Chem.* **2001**, *73*, 1776.

16. J. Zheng, and E. S Yeung, *Austral. J. Chem.* **2003**, *56*, 149.

17. J. Zheng, and E. S Yeung, *Anal. Chem.* **2002**, *74*, 4536.

18. T. Anazawa, H. Matsunaga and E. S. Yeung, *Anal. Chem.* **2002**, *74*, 5033.

19. P.R. Bevington, D.K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, **1992**.

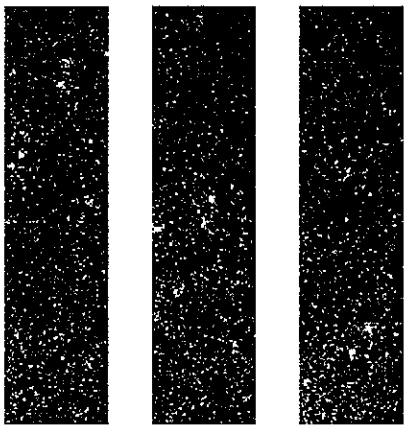

Figure 1. Images of single molecules passing through the observation region in a 75μm capillary under electric field in three consecutive frames of a movie.

Figure 2. Fluorescence emission from single molecules dispersed with a 70 grooves/mm transmission grating. 1:1 mixture of YOYO-I-labeled lambda DNA and POPO-III-labeled lambda DNA were filled in a 75-μm square capillary. Left, undispersed zeroth-order image. Each spot corresponds to one molecule. Right, dispersed first-order image spaced from the zero-order according to the wavelength. Each horizontal streak corresponds to a dispersed spectrum.

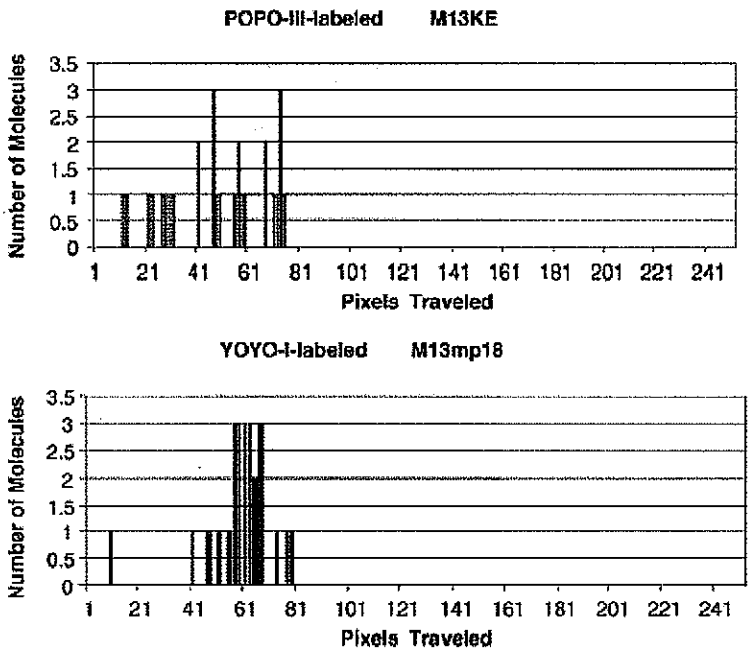

Figure 3. Histograms of DNA mobilities for POPO-III-labeled M13KE restriction enzyme digest and YOYO-I-labeled M13mp18 restriction enzyme digest.

Table 1. Discrimination between DNA by electrophoresis and by spectroscopy

	Electrophoretic mobility		Wavelength dispersion	
	YOYO-I-labeled M13mp18 (pixels)	POPO-III-labeled M13KE (pixels)	YOYO-I-labeled M13mp18 (pixels)	POPO-III-labeled M13KE (pixels)
Average	60.4	46.6	79.5	87.9
Standard deviation	8.58	21.3	1.78	2.16

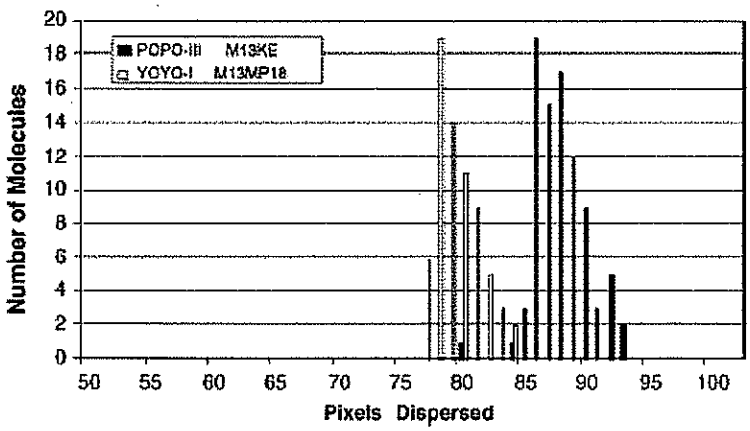

Figure 4. Distribution of fluorescence maxima from POPO-III-labeled M13KE (7088 bp) and YOYO-I-labeled M13mp18 (4497 and 2752 bp) fragments.

Table 2. Average dispersion distance and relative fluorescence signal intensity

	Alexa Fluor 488 labeled mRNA	Alexa Fluor 532 labeled cDNA	Hybrid
Average dispersion distance (pixels)	64.6	68.7	68.2
Standard deviation (pixels)	1.30	2.88	1.56
Relative signal intensity	6	1	2

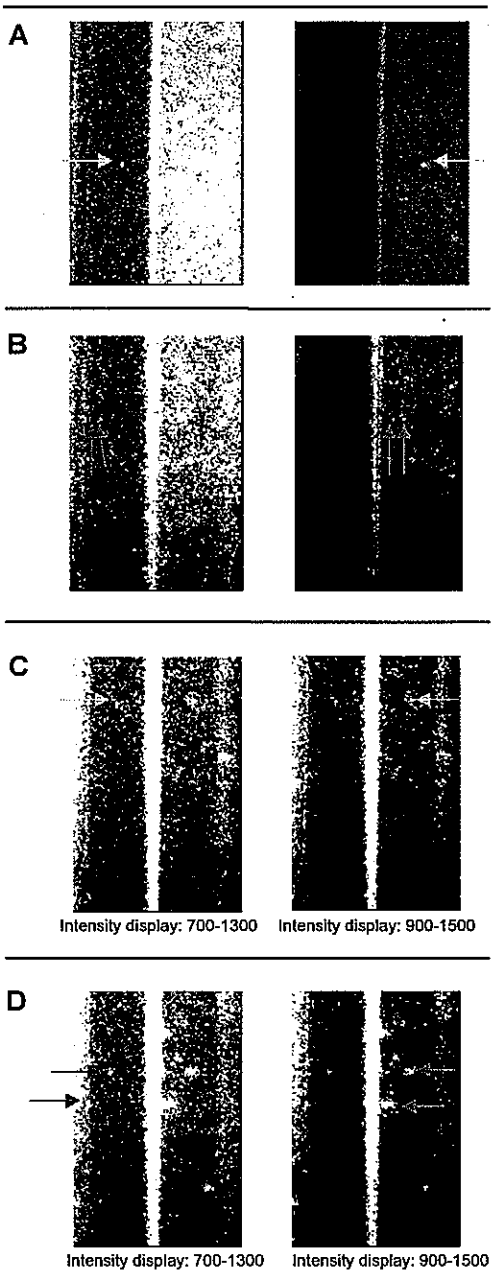

Figure 5. Single molecule images of mRNA, cDNA, and hybrids (A) mRNA stained with Alexa Fluor 488, dispersion distance 65 pixels (B) cDNA stained with Alexa Fluor 532, dispersion distance 68 pixels (C) Control (RNA mixed with cDNA without hybridization) dispersion distance 66 pixels (mRNA). There was no molecule with ~ 68 pixel dispersion (D) Hybrid, dispersion distance 68-69 pixels. Left and right images are the same, with only different contrast settings to clearly visualize the molecules.

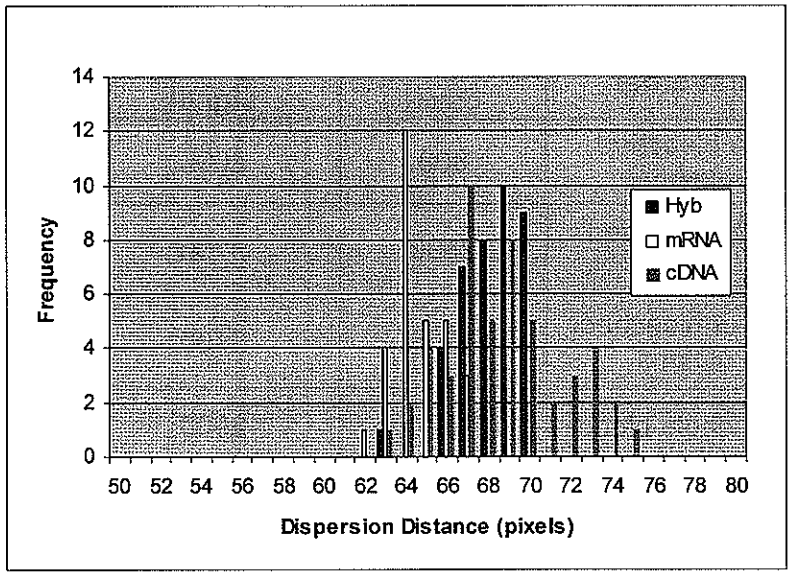

Figure 6. Overlayed histogram of three individual samples. The mRNA and cDNA/hybrid peaks are separated more than 98.8%, with a difference of the average dispersion distance of 3.6 pixels. This difference is 2.53 times of the average standard deviation.

Table 3. Signal and noise level of hybrid molecule images in the presence of bio-matrices at their optimal exposure

first order	Background Level	Bright spot signal	Exposure time (ms)
8% bleached plasma	800	2000	100
50% bleached plasma	2300	3700	70
8% unbleached plasma	1600	2700	100
8% whole blood	1800	2900	100
50% Saliva	2000	3500	150
water	500	1500	100

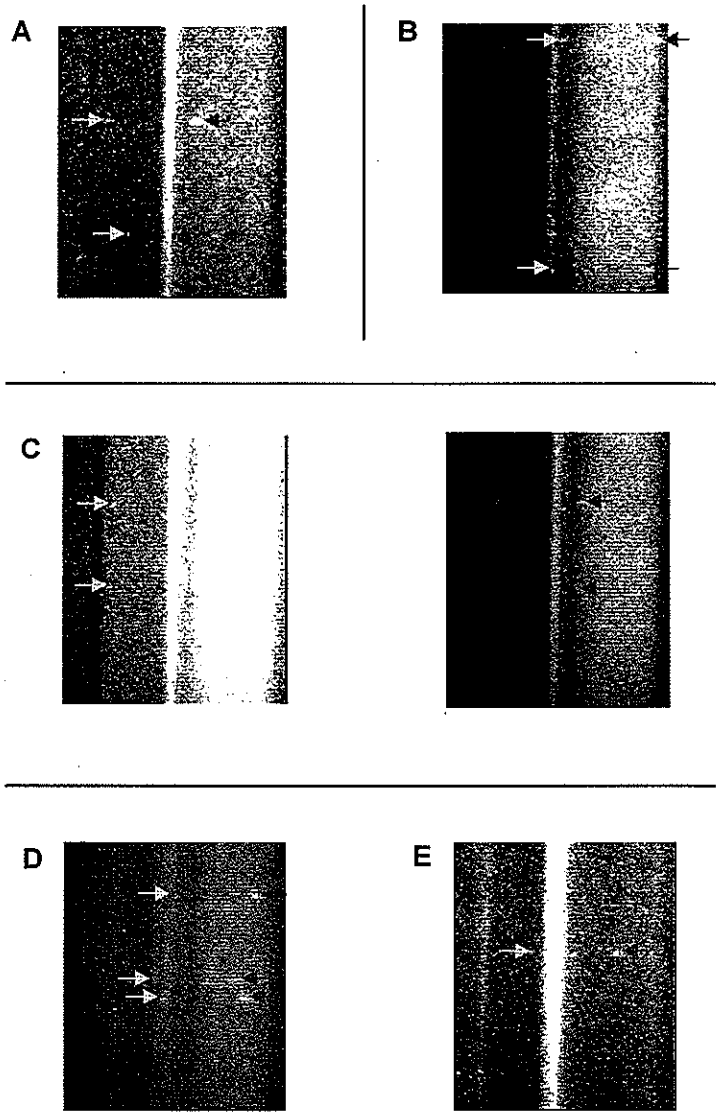

Figure 7. Single molecule images in added bio-matrices (A) 8% photobleached plasma (B) 50% photobleached plasma (C) 8% whole blood (D) 50% saliva (E) A particle in saliva, 63 pixels in separation.

CHAPTER 3. QUANTITATIVE SINGLE MOLECULE SCREENING OF HUMAN PAPILLOMA VIRAL DNA*

Introduction

Targeting DNA abnormalities has been one of the most popular strategies in clinical screening. Many diseases such as virus/bacteria infections, hereditary genetic problems, and acquired mutations can be detected with DNA identification. However, compared to the size of the human genome, the amount of target abnormality is usually very small. The human genome consists of 3 billion base pairs of DNA and for mutation that can cause diseases such as cancer, it can be one nucleotide difference or several copies of certain genes. For inherited mutations, its copy number in a cell is only one or two, while in acquired mutations the number is even lower; generally, by the time the patient is diagnosed with the disease, less than 50% of cells possess the mutation[1]. The virus inserts part of its genome into the human's and use the human cell as a factory to make many replicas of itself, so the viral load in cells varies according to the stages of its infection. However, for an early stage which is often used for an early diagnosis, the viral DNA in a cell should be less than 10^3 copies[2]. To be an effective analytical method in clinical screening, one must have high selectivity as well as sensitivity. A strategy to make the target stand out with less false positives or background is required. This can be achieved with reinforced identification criteria for targets, getting rid of

---

* Reprinted in part with permission from Analytical Chemistry, **2006,** *78*, 6490-6496.

non-target species through washing, or both. A combination of discrimination methods such as multiple probes or colors for labeling can be employed as reinforced criteria.

In this study, small and hence invisible probes were used to hybridize and probe a 800 bp part of the target DNA. A second color dye was later introduced to ensure the identification of the target and reduce false positives. In another strategy, a solid substrate was used to capture the target molecules. This heterogeneous hybridization provides many advantages such as washing and reducing non-target molecules, as well as 100% detection efficiency.

To test the applicability and performance of the single molecule imaging technique in real clinical samples, a study of human papillomavirus (HPV) and cervical cancer was carried out. Cervical cancer is the second most common cancer among woman with over 50% mortality rate[3,4], killing more than 300,000 women per year. HPV is considered to be the primary cause of this cancer. Although this virus is very common and generally suppressed by the body's immune system before it causes problems, certain high-risk subtypes of HPV such as HPV-16 and HPV-18 can cause cell changes that may develop into cancer.

The major reason for cervical cancer's high mortality rate is that it does not show early-warning symptoms or physical changes until it is very advanced. Early detection of HPV could decrease the incidence of cervical cancer and increase the curability. Although the threshold of the viral load to the cancer stage is not exactly known, it has been established that there is a positive relationship between viral load and cervical cancer risk[5,6]. For early detection, high sensitivity is the key. Currently, the simple, inexpensive Papanicolaou smear (Pap smear) screening is the most widely used method. It certainly helps in reducing the incidence of cervical cancer, but it suffers from a significant false-negative rate of 20-40%[7]

as a result of inherent screening errors[8]. HPV testing based on viral DNA detection has drawn increasing attention and been suggested as a supplement to the Pap test[9,10]. Several amplification-based technologies have been developed for viral nucleic acid screening[11].

We demonstrate a new quantitative single-molecule method for detecting HPV DNA without amplification. The assay is based on the hybridization between HPV DNA and fluorescently labeled complementary DNA probes. HPV type 16 was chosen to be our target DNA, which alone accounts for ~50% of the cancer cases[12], among more than 100 types of HPV. The length of HPV-16 DNA is 7904 base pairs (bp), and its oncogene E6-E7, which is 800 bp in size, is expressed in all HPV-containing tumors and derived cell lines. This gene can be found in both episomal and integrated forms in human chromosomes of the cells it infects[13], thereby making it suitable for recognition.

In flow system experiments, the viral DNA was probed by eight 100-nt complementary single-stranded (ss)-DNA. The individual probe molecules have fluorescence intensities below the threshold at the detection condition for hybrid molecules. The identification criteria were further reinforced by adding a second color dye. The additional staining was done by simply mixing an intercalating dye to the hybrid, but it lowered false positives to zero. In surface hybridization experiments, the entire human genomic DNA containing HPV was stained and hybridized to probe molecules tethered on a glass surface. To minimize ambiguity from false positives, another probe was also added. In this dual-probe system, only the second probe was stained and detected. For both experiments, added criteria reduced false positives greatly without compromising the sensitivity too much. This technique not only allows the detection and quantification of viral DNA with a low detection limit but also

offers a universal approach to other DNA sequences by simply changing the sequence-specific probes.

1. Quantitative single molecule screening of HPV-16 DNA in flow system

Experimental Section

Sample preparation. Eight 100-nt ss-DNA probes were designed according to the sequences of the E6/E7 region of HPV-16 from Genbank (K02718) (Table 1) and synthesized by Integrated DNA Technologies (Coralville, IA). The probes were labeled individually with a Ulysis Alexa Fluor 532 nucleic acid labeling kit (Molecular Probes, Eugene, OR) according to the manufacturer's instructions. After the staining reaction, the probes were submitted to a cleanup procedure with micro biospin columns (Bio-Rad Laboratories, Philadelphia, PA) to eliminate excess labeling reagents. All eight probes were mixed together for a subsequent hybridization step.

In hybridization samples, the total amount of HPV-16 DNA (Maxim Biotech, San Francisco, CA) and female human genomic DNA (Promega, Madison, IL) was kept at 1 μg, while the ratio of them was varied to achieve a series of HPV-16 contents from 10^{-4} to 10%. The probe solution was mixed with the DNA so that the concentration of each probe was 8 times that of HPV-16 DNA. Finally, hybridization buffer (Maxim Biotech) was added to make a total volume of 30 μL. The samples were kept at 95 °C for 10 min for denaturation and at 42 °C

overnight for hybridization in a thermal cycler (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA). The hybridization samples were diluted with nuclease-free water (Ambion) to proper concentrations right before imaging (see Table 2). In dual-color detection, the samples were further labeled with YOYO-3 (Molecular Probes) at a ratio of one dye molecule per five bases.

Cell line extracts. Cervical cancer-derived cell lines CaSki and SiHa were purchased from ATCC (Manassas, VA). Normal Pap test specimens were donated by a healthy adult female. DNA from these specimens were extracted by the DNeasy tissue kit (Qiagen, Valencia, CA) and used without further purification. The DNA extraction yield is ~80% according to the manufacturer's handbook. The DNA concentration was calculated from UV absorption at 260 nm. The amount of extracted DNA from each Pap sample was 8 μg.

Imaging System. The imaging system was similar to the previous apparatus used in our group[14,15], with some modifications (Figure 1). A 532-nm solid-state continuous-wave laser (45 mW, μ-Green model 4611, Uniphase, San Jose, CA) was used as an excitation source. A Uniblitz mechanical shutter (model LS2Z2, Vincent Associates, Rochester, NY) and a driver (model T132, Vincent Associates) were synchronized to the Pentamax 512-EFT/1EIA intensified charge-coupled device (ICCD) camera (Princeton Instruments, Princeton, NJ). The shutter allowed the laser beam to pass through only when the ICCD was on to reduce photobleaching. After the mechanical shutter, the laser beam was expanded by the first cylindrical lens (focal length, 200 mm; Thorlabs, Newton, NJ) in the *Y*-direction. The expanded beam was then focused into the capillary at normal angle as a thin sheet by the second cylindrical lens (focal length, 25 mm; Edmund Industrial Optics, Barrington, NJ).

The image of the laser focusing plane is schematically shown in the enlarged part in Figure 1. Only molecules inside the laser focusing plane could be excited and detected. The illuminated volume was ~0.2 nL, implying a detection efficiency of 13% considering the illuminated portion compared to the total thickness of the capillary bore. Hybridization samples were injected into a 30-cm-long square capillary (75 μm i.d., 365 μm o.d.; Polymicro Technologies, Phoenix, AZ), and the flow was driven hydrodynamically.

Fluorescence from single molecules was collected by a Zeiss 20×/0.75 NA Plan-Apochromat microscope objective lens. A 532-nm, long-pass edge filter (Semrock, Rochester, NY) and a 532-nm holographic notch filter (OD >6, Kaiser Optical, Ann Arbor, MI) were placed between the objective lens and the ICCD to cut off scattering from the excitation beam. A transmission grating with 70 grooves/mm (Edmund) was mounted in front of the ICCD camera to disperse fluorescence collected from the molecules. The distance between the transmission grating and the ICCD was set to be 6.5 cm so that the zeroth-order and first-order images of the 75-μm-i.d. capillary would not overlap each other. The sampling frequency was 2 Hz, with the shutter driver set to 40-ms exposure and 460-ms delay. Images were obtained with WinView software provided by Princeton Instruments.

Results and Discussion

Single-color detection. The eight oligo nucleotide probes aimed for E6/E7 genes of HPV were hybridized and imaged with the single molecule detection setup shown in Figure 1. The smallest DNA fragment size to be imaged with this setup is ~400 nt, so the individual probe does not appear as a molecule in the movies. For a meaningful quantitative comparison, the imaged volume was kept constant from run to run. Since the volume probed per detection

experiment (per movie) was proportional to the flow rate which was driven by hydrodynamic force, the heights of the sample reservoirs at the two ends of the capillary were fixed to ensure reproducibility. As shown in Figure 1, the sample volume of each frame is 75 × 75 × 300 μm^3, which is 1.7 nL. To make sure that the spots are molecules instead of random noise bursts of the ICCD, the molecules were tracked for five consecutive frames. A total of 1.7 nL of liquid flowed through the capillary and was imaged in five frames. With a 7500-frame movie, the total volume that flowed through the detection area was 2.6 μL.

To evaluate quantification, HPV-16 DNA was added to commercial human genomic DNA. The total amount of viral and human genomic DNA in a vial was kept constant, while the ratio between them was varied to achieve a series of HPV-16 percentages from 0 to 10% (7 × 10^5 copy/cell). After hybridization, the samples were diluted to suitable concentrations for counting individual molecules. The dilution factor for each sample was designed (Table 2) to make sure that molecules did not overlap with each other in the high-concentration samples, and the number of hybridized molecules was higher than three times that of false-positive counts in the low-concentration samples. A flow system was selected to allow screening of a much larger volume of sample than that within the detection window. Note that only a small fraction of the total sample passed through the observation region per count period. After the movies were recorded, the molecules were counted manually while fast-forwarding the movies. The numbers were further normalized considering the movie length and dilution factor. Figure 2B shows a good linear relationship over a range of 6 orders of magnitude between HPV-16 DNA percentage and the number of molecules detected. At low concentrations, the precision is limited by counting statistics.

To test the compatibility of the current Pap test sampling method with this assay, we extracted DNA from normal Pap smear specimens. About 8 μg of total DNA could be obtained from each specimen. Viral DNA was added to the extracted DNA matrix in the same way as in the commercial human genomic DNA experiments (Table 2). In Figure 2B, the red data points for the extracted DNA matched well with the blue ones for the standard samples. The results also show that the background levels of the two matrices were similar, indicating that DNA extracts from Pap test specimens did not have biological impurities that would interfere with single-molecule detection.

This single-molecule assay was further tested for cell lines with known viral load. According to the supplier, American Type Culture Collection (ATCC), the HPV-16 viral loads in cervical cancer-derived cell lines CaSki and SiHa are ~600 and 1-2 copies/cell, respectively. The average number of molecules detected with our system was 8200 molecules for CaSki and 49 for SiHa per experiment. The detected numbers were converted to the HPV-16 DNA percentages based on the flow rate and the sample volume, and the DNA content of each cell is calculated by dividing the extracted DNA amount by the cell population. The calculated viral loads were 540 copies/cell for CaSki and 2.4 copies/cell for SiHa. The agreement with the numbers from ATCC means that this quantitative assay is well-behaved.

For the false positives, the number of molecules measured for the 0% HPV-16 DNA sample (inset of Figure 2B), there are three possible sources: aggregation of unbound probes, non-sequence-specific hybridization with human genomic DNA, and random noise of the ICCD camera. The first factor could be neglected as no counts were found in control experiments that contained probes only. Cross-hybridization could be minimized by optimization of

hybridization conditions, such as temperature and ionic strength of solution. When two random noise spikes accidentally showed up in the images with reasonable dispersion distance, they would be counted as false positives as well. These could be greatly reduced by the dual-color detection mode described below. The ratio of the number of detected HPV-16 DNA molecules to that of false positives is the signal-to-noise ratio (S/N). The detection limit of HPV-16 DNA is 0.0001% (0.7 copy/cell) when S/N equals three. This sensitivity is extremely high among the quantitative assays of nucleic acids. Moreover, this was achieved without amplification.

Dual-color detection. A second dye was added to the single-color detection scheme to improve the selectivity. Dual-color detection involves three spots in the movie (one zeroth-order, and two first order spots) for identification, and hence has less possibility of false positives from ICCD noise. The chances of three random noise spikes accidentally appearing with the correct dispersion distance is negligible.

The intercalating dye YOYO-3 (612/631) was added after hybridization of Alexa Fluor 532-labeled probes to the target so that it stains the whole genome. YOYO-3 was chosen as the second dye according to three criteria. First, it has emission maximum at 631 nm, which is easily distinguished from that of Alexa Fluor 532 (555 nm). Second, its absorption maximum (612 nm) is far away from the excitation wavelength (532 nm) so that it is very dim without energy transfer from Alexa Fluor 532. Third, its fluorescent intensity is comparable to the Alexa Fluor 532 donor when partial FRET occurs. These features allowed us to identify the dual-color molecules conveniently.

Figure 3A shows the hybridized molecules labeled with YOYO-3 as an enhancement to the

single-color system. When excited by a 532-nm laser, fluorescence from both Alexa Fluor 532 and YOYO-3 were detected. A typical dual-color image of pure HPV-16 DNA hybridization sample is shown in Figure 3B. The left peak in the first-order image corresponds to the emission of Alexa Fluor 532 at 555 nm while the right peak corresponds to the emission of YOYO-3 at 631 nm. The average dispersion of Alexa Fluor 532 was 106.6 ± 0.9 pixels and that of YOYO-3 was 121.8 ± 1.2 pixels. The difference between the two averages was 15.2 pixels, which was 14.4 times the average of the two standard deviations. That is, the two peaks are clearly separated from each other. The statistical information was collected in the same way as for single-color detection. There were 90 molecules counted in the experiment, in which 77 molecules had two peaks and 13 molecules had one. All of the single-color molecules were associated with the dispersion distance of YOYO-3, which represented the target DNA bound with too small of a number of probes. The hybrids with at least five probes clearly appeared as bright spots in our setup (data not shown).

In the presence of bulky human genomic DNA as the matrix for two-color DNA detection, YOYO-3 made them appear as bright single-color molecules. The background level was increased, and it was not practical to seek out the dual-color hybridized molecules among the single-color matrix, especially when the HPV-16 DNA content was low. This problem resulted in poor sensitivity, and the detection limit was only 1% (7×10^3 copies/cell) HPV-16 DNA. To solve this problem, we used a restriction endonuclease to digest human genomic DNA into small pieces, with an average length of 4 kb. *Hin*dIII worked well for cutting human genomic DNA while preserving the E6-E7 region of HPV-16 DNA uncut. There was 1 order of magnitude improvement in detection limit, that is, down to 0.1% (7×10^2

copies/cell) with this digestion step. In the negative control experiment, where the sample contained everything except the target DNA, no dual-color signal was detected. It is worth pointing out that although the sensitivity could not compete with single-color detection, the lack of false positives was a significant advantage for dual-color detection.

2. Quantitative single molecule screening of HPV-16 DNA using surface hybridization

Experimental Section

Slide pre-treatment. Custom-made cover glass coated with poly-L-lysine (22 × 40 mm) was purchased from TeleChem International (Sunnyvale, CA). The surface probe was 100-nt oligo nucleotide of 661-760th base region of HPV-16, in E7 gene (IDT, Corralville, IA). The probe was dissolved in 10 mM Gly-Gly solution (pH 8.2) to a concentration of 0.02 μg/μl. 5μl of this solution was pipetted onto a poly-L-lysine coated cover glass ("Slide"), sandwiched with clean 22 × 22 mm cover glass, and kept at room temperature for 1 h. The top cover glass was removed before UV-crosslinking in StrataLinker UV-crosslinker (Stratagene, La Jolla, CA) at 150 mJ. The slides were dipped in deionized water 2-3 times and dried under a nitrogen stream. To deactivate unused functional groups on poly-L-lysine, the slides were immersed in a solution of 3 g succinic anhydride (Fluka) dissolved in 170 ml 1-methyl-2-pyrrolidinone (Aldrich), buffered with 7.5ml 1M sodium borate (pH 8.3, Fluka) for 20 min. After a gentle washing in 200 ml deionized water for 5 min, slides were

immersed in 95 °C deionized water for 2 min, dipped 5-10 times in 95% ethanol and blow-dried with nitrogen.

DNA preparation. HPV-16 genomic DNA (Maxim Biotech) and human female genomic DNA (Promega) was digested with three restriction enzymes: HaeIII, NcoI, and BstNI (New England Biolabs, Ipswich, MA). After digestion, both DNA were cleaned with a QiaQuick PCR purification kit (Qiagen) and redissolved in 10 mM Tris-Cl. For the single-probe detection experiment, both HPV and human genomic DNA were stained with a Ulysis Alexa Fluor 532 kit, and purified with a QiaQuick PCR purification kit. At the final elution step of purification, Maxim Bio Hybridization II solution (Maxim Biotech) was used to redissolve DNA. HPV DNA was added to the human genomic DNA according to the concentration ratios shown in Table 3.

Second probe preparation for dual-probe detection. For dual-probe detection, the second probe was made to accompany a 1 kbp DNA fragment stained with Alexa Fluor 532 dye as a fluorescence tag. The 1 kbp tail sequence was from pBR322 vector, so that the sequence would not match either HPV or human genomic DNA. The probe molecules were synthesized by PCR, using one scorpion primer that has 50-nt HPV probing sequence on the 5'-end of the primer for clockwise PCR (Table 1, probing part is underlined) and one regular counter-clockwise primer for pBR322 vector (Table 1). These primers ware also custom-made by IDT. SuperTaq PCR kit (Ambion) was used for PCR. After purification with a QiaQuick PCR purification kit, the product was submitted to gel electrophoresis with 2% agarose E-gels (Invitrogen). The main DNA band was excised and DNA was extracted with a QiaQuick gel extraction kit. After quantification with UV absorption measurement, 1

μg of probe was submitted to a labeling reaction with a Ulysis Alexa Fluor 532 kit. All DNA were dissolved in Corning Pronto! long oligo hybridization solution. Each slide was loaded with 70 μl solution containing 5 μg human genomic DNA, spiked HPV, and stained second probes.

Cell line extracts. Two cell lines with known HPV infection (CaSki, SiHa) and one normal cell line (C-33A) were purchased from ATCC, as described in the flow system experiment. The cells were directly submitted to DNA extraction using a DNeasy kit without culture. Obtained DNA was digested with three restriction enzymes, HaeIII, NcoI, and BstNI. After cutting, the DNA was purified with a QiaQuick PCR purification kit. For single-probe experiments, this cut DNA was stained with Ulysis Alexa Fluor 532 kit and purified. For dual-probe experiments, unstained DNA was added along with a second probe onto each slide. Each slide was loaded with DNA extracted from 10^6 cells and dissolved in 70 μl Corning Pronto! long oligo hybridization solution.

Loading, hybridization and washing. At least 30 min before sample loading, the slide and hybridization chamber (Corning) were pre-heated to 55-60 °C. The sample mixture was kept at 95 °C for 10 min, and then at 72 °C until it was loaded on the slide. On each slide, a 19 × 6 mm silicon gasket perfusion chamber (Molecular Probes) was added to accommodate the hybridization solution. For each slide, 5 μg of human genomic DNA and spiked HPV was loaded in 70 μl total volume. Slides with samples were enclosed in Corning Pronto! hybridization chambers, and kept in a 55 °C oven for ~16 h. After hybridization, the slides were submitted to a series of washing with wash solutions from a Corning Pronto! hybridization kit, following the manufacturer's instruction. After the

washing, the slides were dried under a gentle nitrogen stream and kept in a dark box. Right before imaging, the slide was hydrated with 5 µl 10 mM Gly-Gly buffer (pH. 8.2) and covered with a clean 22 × 22 mm cover glass.

Imaging system. To image molecules on the glass surface, evanescent field layer illumination was used (Figure 4). The slide was located on top of a right-angle fused silica prism (Melles Griot, Carlsbad, CA), with emulsion oil in between. A 532-nm solid-state continuous wave laser (45 mW, µ-Green model 4611, Uniphase, San Jose, CA) was used again as the excitation source. The laser was modulated with a Uniblitz mechanical shutter (model LS2Z2, Vincent Associates, Rochester, NY) and a driver (model T132, Vincent Associates), which were synchronized to the Pentamax 512-EFT/1EIA intensified charge-coupled device (ICCD) camera (Princeton Instruments, Princeton, NJ). Fluorescence from individual hybridized molecules was collected by a Zeiss 100×/1.2 NA Plan-Neofluar microscope objective lens. Two 532-nm long-pass edge filters (Semrock, Rochester, NY) were placed between the objective lens and the ICCD to cut off scattering from the excitation beam. The sampling frequency was 2 Hz, with 20-ms exposure. Images were obtained with WinView software provided by Princeton Instruments.

Results and Discussion

Single-probe detection. In this assay, both female human genomic DNA and HPV-16 DNA were stained with Alexa Fluor 532 dye. Since DNA fragments larger than 1 kbp are likely to precipitate out when stained with the dye, both DNA were cut with restriction enzymes. Nco I and Hae III cut near the E6/E7 gene of HPV-16 and make the target fragment 1001 bp in size. BstN I enzyme was used to chop up human genomic DNA smaller than 1 kbp, without

affecting the size of the target (Table 4). Figure 5 is a schematic diagram of two surface hybridization strategies. After overnight hybridization, the slide was washed, dried and submitted to imaging. Evanescent field illumination was used to excite surface-hybridized molecules. The incident laser beam angle was set to create a maximum evanescent field depth. With an oil type ×100/1.2 N.A microscope objective lens, the imaging area was 0.0132 mm^2.

When the hybridization was carried out at 42 °C, non-specific adsorption of DNA on the surface was very high, making counting almost impossible. A higher temperature for hybridization (55 °C) helped reduce non-specific adsorption down to 128 ± 20.8 counts per frame, but could not eliminate it totally. Further increase in temperature to 60 °C reduced the number somewhat, but also lowered counted numbers in the HPV sample, especially near the detection limit, making it undistinguishable from the background (data not shown).

Figure 6 shows typical images of single probe hybridization. In the negative control sample (Fig. 6 A), non-specific adsorption was high, increasing the background signal. This non-specific adsorption also caused large aggregation spots on the surface and a background signal different from frame to frame in a single movie. For image analysis, the background signal average and sigma value of each frame was obtained with WinView software and the display range of movie was set in between (average+2σ) and (average+5σ) during the counting.

When the imaging area is 0.0132 mm^2, the expected number of the target hybrid molecules was 60 for 10^{-4}% HPV, 600 for 10^{-3}% HPV and so on. Considering the net count (each concentration's molecule count subtracted by the background count) compared to the

expected population in the imaging area, the hybridization efficiency of this assay is between 50-60% in average for 10^{-4} - 10^{-3} % HPV samples (Table 5). The maximum number of molecules that can be imaged in a 512 × 512 pixel camera is 2500, so a 10^{-2}% HPV content (70 copies/cell) sample should be virtually our high-end detection limit. However, the actual molecule count increase slows dramatically after 10^{-3}% HPV, and 10^{-2}% HPV shows 398 ± 101 counts, only 10% higher counts for 10^{-3}% sample. The equilibrium hybridization isotherm is of Langmuir form, and in general cases it reflects the electrostatic penalty incurred because each hybridization event increases the charge of the probe layer[16-18].

Although many studies that involve surface hybridization such as microarray still claim good results with a linear dynamic range of two orders of magnitude, it is not wide enough, especially for viral DNA quantification where the target DNA amount can be higher than the inherited or acquired mutation – the HPV viral load can go as high as 10^3 copies/cell at the starting stage of cancer[2]. Reducing the background will be essential to widen the linear dynamic range.

The average count in negative control is considered as background, and its standard deviation is noise. When the S/N ratio is 3, the detection limit is 10^{-4}% HPV content (0.7 copy/cell), equal to that of the flow system. In molar concentration, this HPV content is ~16 fM in a loaded volume of 70 µl and at least 10 times lower detection limit than that of a conventional microarray.

As in flow detection system experiments, cell line extracts with known HPV-16 contents were used with the single probe surface hybridization method. In each slide, DNA extract from 10^6 cells was loaded. A human cell contains 6 pg of DNA, and with ~80% yield of

extraction and staining procedures, the actual DNA load per slide is considered to be ~5 ug, the same as in purified commercial DNA experiments. The counted number of each cell line sample is shown in Table 6. For SiHa cells, calculated copy number was ~0.7 copy/cell, which was slightly lower than the known value of 1-2 copies/cell. A normal cell line without HPV-16 (C-33A) showed the number counted per frame very close to that of negative control. The known viral contents for the CaSki cell line is ~600 copies/cell, and is beyond our high-end detection limit. When the molecule number was counted, it was 393 ± 138 counts per frame, which is close to the saturation value of our system.

Dual-probe detection. Instead of staining the entire human genome and HPV, a second probe with a 1 kbp tail was added to the assay to lower down background caused by non-specific adsorption (Figure 5B). The second probe has a 50 nt 'probing' part, which is complementary to the 121-170th base region of the HPV genome, located in the E6 gene. The surface probe targets the 661-760th base region of HPV, so there is ~490 nt gap between the two probes, as shown in Figure 7. A 1 kbp tail is attached to the 3' end of the probing part via two C-18 spacers. The tail was made by PCR amplifying a part of pBR322 DNA, between BamH I and Sty I restriction enzyme cutting sites. The PCR products were run with gel electophoresis, and the exact 1 kbp band was excised and extracted to ensure size uniformity. In dual-probe hybridization, only the second probe was stained with Alexa Fluor 532 dye. An equal amount of the second probe was added to all samples (0.025 μg/slide), regardless of the HPV DNA contents. Extra care was used during the washing steps, and blow drying was omitted. Instead, slides sat for 30 minutes at room temperature.

Figure 8 shows the frame captures from a hybrid detection movie. With a 1 kbp DNA tail

labeled with Alexa Fluor 532 dye, the signal from the target clearly stands out with a signal-to-noise ratio >10. The amount of DNA stained with fluorescent dye in dual-probe detection is 0.025 μg per slide and it is 1/200 of the stained DNA used in single-probe mode. Compared to single-probe experiment results, non-specific adsorption is far less (2.4 counts per frame). The average molecule counts for each concentration from 0.7 copy/cell to 7,000 copies/cell are shown in Table 7. When plotted in a log-log relationship, the viral contents and the number of molecules counted in a frame have a linear relationship with a slope of 0.4838 (Figure 9). The assay has nearly 4 orders of linear dynamic range, and when the average molecule count in negative control sample and standard deviation were considered as background and noise respectively, the detection limit of this assay is 2.36 copies/cell.

It is thought that the hybridization of target molecules changes the surface environment, and causes deviations from the simple Langmuir isotherm. In dual probe hybridization mode, there are two methods that hybridization takes place in at the same time:

(1) target + surface probe, and then second probe + target on the surface

(2) target + second probe, and then this hybrid + surface probe.

In scheme (2), it is similar to hybridization in the single probe experiment, except that the target is now 2 kbp hybridizing to the surface probe. In scheme (1), however, the target's hybridization increases the surface charge density of that area, whose electrostatic repulsion reduces the second probe's hybridization chance[18]. As a result, when both schemes take place simultaneously, the number of 3-component hybridization is lower than expected. At low target concentrations this effect can be neglected, but at high concentrations it becomes important.

Cell line extracts were also used with the dual-probe scheme, to test the performance with a real sample. In the same way as the single probe experiments, three cell line extracts were loaded on slides with 10^6 cells of DNA per slide concentration. The detected HPV contents for CaSki (600 copies/cell HPV-16), SiHa (1-2 copies/cell) and C-33A normal cell line experiments were calculated using the equation of the log-log plot (Figure 9) with average molecule counts from each movie. The result values were 815 copies/cell for CaSki and 1.22 copies/cell for SiHa (Table 6). The molecule count for C-33A was 3.80 counts per frame, which falls within the range of the negative control (2.54±1.40 counts per frame). The actual average count for CaSki was lower than that of 700 copies/cell sample, but the calculated value with our experiment showed a slightly higher number. The viral content of the SiHa cell line was lower than the detection limit determined by the calibration curve (2.36 copies/cell), but the counted number of 6.09 ± 3.25 counts per frame was slightly higher than that of the 10^{-4}% sample and the calculated number is close to the known value. All three results showed good agreement with known viral loads and results from other assays of our flow system and single probe surface hybridization system.

Although the performance of our surface hybridization method cannot be directly compared to that of conventional microarrays, where up to 10^5 genes can be analyzed in a single experiment, the detection limit and dynamic range of our surface experiments showed at least 10 times improvement in both categories[19]. The genomic DNA or cDNA labeling procedure that is often employed in microarrays involves amplification either by PCR (>1000×) or using random hexamer primers (> 7-10×) . Since the amplification makes the original amount or ratio of the sequence in the cell meaningless, the sample set of DNA is usually run together with a control set without mutation to compare in microarray and the ratio between

their fluorescence intensities is obtained as data. With single molecule detection, the actual molecule count is obtained and it can be directly multiplied or divided with a dilution factor to obtain the target nucleic acid quantity in the original cell. Furthermore, without this amplification step, single molecule detection can provide a less error-prone result. When a transmission grating is incorporated to recognize the spectrum of individual molecules, multiple color dyes can be used to compare the target amounts from several cells from different origins, simultaneously. Application of single molecule assay and detection strategy can make conventional microarray an even more efficient and meaningful method to analyze genomes.

Conclusions

Four DNA screening systems have been tried, detected in a flow and on a surface, each with one and two probes. For the flow detection system with a one-color probe, the assay is sensitive with a detection limit of 0.7 copy/cell and has a wide linear dynamic range over 6 orders of magnitude. For the surface hybridization method, detection limits were 0.7 copies/cell and 2.4 copy/cell for single- and dual-probe modes, respectively, which is as sensitive as the flow system. With linear log-log relationship, the dual probe method had a linear dynamic range of 4 orders of magnitude. It is reliable in quantification since no amplification is involved. Although PCR methods have enabled detection at the level of 0.05 copy/cell[20], they are prone to false positives, with problems of contamination, enzyme inhibition and other concerns.

The probe design is a key step in these methods. In flow detection experiments, the single unbound probes are invisible at the selected threshold but become visible when at least five

of them hybridized to the target. The digital counting approach avoids the need to remove the unbound probes or add quenchers after hybridization to decrease background. Dual-color detection could achieve zero false-positive counts, with unmistakable and distinct two-peak spectra only with target molecules. In surface hybridization experiments, the number of false positives was dramatically cut by avoiding staining of the entire genome and adding second probe with a fluorescent tail instead.

For all four methods, the whole procedure involved little more than DNA extraction, hybridization, and imaging. Moreover, this technique can be directly coupled with current sampling methods such as Pap smears. Since this screening technology does not require other HPV-16 DNA specific features (e.g., antibodies) except for the probe sequence, it is readily applicable to any DNA detection and quantification by changing the probes.

In comparison with photonburst detection[21], imaging allows substantially higher throughput in molecule counting. Up to 2500 molecules can be screened in 20 ms. The transmission grating offers simple, accurate, and sensitive spectral analysis when compared to using multiple optical filters or multiple lasers. The detection efficiency of the flow system can be further increased to near 100% with radial focusing of molecules inside the capillary when an electric field is applied across the capillary along with hydrodynamic flow[14]. When a solid surface was used to capture target DNA, 100% detection efficiency is possible by scanning the sample area. Furthermore, the detection can be done later, with an absence of non-targets and excessive probes after washing.

Acknowledgments

We thank Dr. Mary Jo Schmerr for help with cellular DNA extraction. E.S.Y. thanks the Robert Allen Wright Endowment for Excellence for support. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract W-7405-Eng-82. This work was supported by the Director of Science, Office of Basic Energy Sciences, Division of Chemical Sciences and by the National Institutes of Health.

References

1. A. Makoto, T. Ochiya, H. Toriyama-Baba, T. Ota, T. Sekiya, M. Terada and H. Tsuda *Carcinogenesis* **2000**, *21*, 243.
2. D. C. Swan, R. A. Tucker, G. Tortolero-Luna, M. F. Mitchell, L. Wideroff, E. R. Unger, R. A. Nisenbaum, W. C. Reeves and J. P. Icenogle, *J. Clin. Microbiol.* **1999**, *37*, 1030.
3. J. H. Jett, R. A. Keller, J. C. Martin, B. L. Marrone, R. K. Moyzis, R. L. Ratliff, N. K. Seitzinger, E. B. Shera and C. C. Stewart, *J. Biomol. Struct. Dyn.* **1989**, *7*, 301.
4. L. M. Davis, F. R. Fairfield, C. A. Harger, J. H. Jett, R. A. Keller, J. H. Hahn, L. A. Krakowski, B. L. Marrone, J. C. Martin, H. L. Nutter, R. L. Ratliff, E. B. Shera, D. J. Simpson and S. A. Soper, *Genet. Anal.: Biomol. Eng.* **1991**, *8*, 1.
5. A. M. Josefsson, P. K. E. Magnusson, N. Ylitalo, P. Sorensen, P. Qwarforth- Tubbin, P. K. Andersen, M. Melbye, H. O. Adami and U. B. Gyllensten, *Lancet* **2000**, *355*, 2189.
6. N. Ylitalo, P. Sorensen, A. M. Josefsson, P. K. E. Magnusson, P. K. Andersen, J. Ponten, H. O. Adami, U. B. Gyllensten and M. Melbye, *Lancet* **2000**, *355*, 2194.
7. N. S. Larsen, *J. Nat. Cancer Inst.* **1994**, *86*, 6.
8. L. G. Koss, *J. Am. Med. Assoc.* **1989**, *261*, 737-743.

9. C. J. Meijer, T. J. M. Helmerhorst, L. Rozendaal, J. C. van der Linden, F. J. Voorhorst, J. M. M. Walboomers, *Histopathology* **1998**, *33*, 83.

10. J. Cuzick, *J. Am. Med. Assoc.* **2000**, *283*, 108.

11. M. T. Sandri, P. Lantati, E. Benini, P. Dell'Orto, L. Zorzino, F. M. Carozzi, P. Maisoneuve, R. Passerini, M. Salvatici, C. Casadio, S. Boveri, and M. Sideri, *J. Clin. Microbio.* **2006**, *44*, 2141.

12. F. X. Bosch, M. M. Manos, N. Munoz, M. Sherman, A. M. Jansen, J. Peto, M. H. Schiffman, V. Moreno, R. Kurman, K. V. Shah, E. Alihonou, S. Bayo, H. C. Mokhtar, S. Chicareon, A. Daudt, E. Delosrios, P. Ghadirian, J. N. Kitinya, M. Koulibaly, C. Ngelangel, L. M. P. Tintore, J. L. Riosdalenz, Sarjadi; A. Schneider, L. Tafur, A. R. Teyssie, P. A. Rolon, M. Torroella, A. V. Tapia, H. R. Wabinga, W. Zatonski, B. Sylla, P. Vizcaino, D. Magnin, J. Kaldor, C. Greer and C. J. Wheeler, *Nat. Cancer Inst.* **1995**, *87*, 796.

13. K. B. Choo, C. C. Pan and S. H. Han, *Virology* **1987**, *161*, 259.

14. J. J. Zheng, E. S. Yeung, *Anal. Chem.* **2002**, *74*, 4536.

15. J. Y. Lee, H. W. Li, E. S. Yeung, *J. Chromatogr. A* **2004**, *1053*, 173.

16. A. Halperin, A. Buhot and E. B. Zhulina, *Clin. Chem.* **2004**, *50*, 2254.

17. A. Halperin, A. Buhot, and E. B. Zhulina, *Biophys. J.* **2004**, *86*, 718.

18. A. Vainrub and B. Montgomery Pettitt, *Phys. Rev. E* **2002**, 66, 041905.

19. A. E. Oostlandera, G. A. Meijerb and B. Ylstraa, *Clin. Genet.* **2004**, 66, 488.

20. K. W. Hart, O. M. Williams, N. Thelwell, A. N. Fiander, T. Brown, L. K. Borysiewicz and C. M. Gelder, *J. Clin. Microbiol.* **2001**, *39*, 3204.

21. L. A. Neely, S. Patel, J. Garver, M. Gallo, M. Hackett, S. McLaughlin, M. Nadel, J. Harris, S. Gullans and J. Rooke, *Nat. Methods* **2006**, *3*, 41.

Table 1. Probe sequences

Flow system detection	Fragment 1	ttagtataaa agcagacatt ttatgcacca aaagagaact gcaatgtttc aggacccaca ggagcgaccc agaaagttac cacagttatg acagagctg
	Fragment 2	caaacaacta tacatgatat aatattagaa tgtgtgtact gcaagcaaca gttactgcga cgtgaggtat atgactttgc ttttcgggat ttatgcatag
	Fragment 3	tatatagaga tgggaatcca tatgctgtat gtgataaatg tttaaagttt tattctaaaa ttagtgagta tagacattat tgttatagtt tgtatggaac
	Fragment 4	aacattagaa cagcaataca acaaaccgtt gtgtgatttg ttaattaggt gtattaactg tcaaaagcca ctgtgtcctg aagaaaagca aagacatctg
	Fragment 5	gacaaaaagc aaagattcca taatataagg ggtcggtgga ccggtcgatg tatgtcttgt tgcagatcat caagaacacg tagagaaacc cagctgtaat
	Fragment 6	catgcatgga gatacaccta cattgcatga atatatgtta gatttgcaac cagagacaac tgatctctac tgttatgagc aattaaatga cagctcagag
	Fragment 7	gaggaggatg aaatagatgg tccagctgga caagcagaac cggacagagc ccattacaat attgtaacct tttgttgcaa gtgtgactct acgcttcggt
	Fragment 8	tgtgcgtaca aagcacacac gtagacattc gtactttgga agacctgtta atgggcacac taggaattgt gtgccccatc tgttctcaga aaccataatc
Surface hybridization	Suface Probe	5'- GAG GAG GAT GAA ATA GAT GGT CCA GCT GGA CAA GCA GAA CCG GAC AGA GCC CAT TAC AAT ATT GTA ACC TTT TGT TGC AAG TGT GAC TCT ACG CTT CGG T -3'
	2nd probe Clock-wise PCR Primer	5'- GGA GCG ACC CAG AAA GTT ACC ACA GTT ATG CAC AGA GCT GCA AAC AAC TA -(Spacer C-18)-(Spacer C-18)- TAC TTG GAG CCA CTA TCG ACT ACG C -3'
	2nd probe Counter-clock-wise PCR Primer	5'-GCT GGA GAT GGC GGA CG-3'

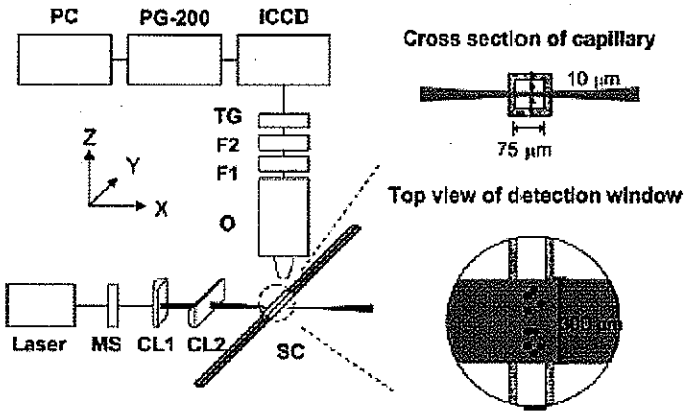

Figure 1. Schematic illustration of the experimental setup: laser, 532-nm continuous-wave solid-state laser; MS, mechanical shutter; CL1, cylindrical lens, FL) 200 mm; CL2, cylindrical lens, FL) 25 mm; SC, square capillary; O, objective lens, 20x/NA 0.75; F1, 532-nm long-pass edge filter; F2, 532-nm holographic notch filter; TG, transmission grating; ICCD, intensified CCD camera; PG-200, function generator for controlling the ICCD; PC, personal computer for data acquisition and analysis. The inset shows the excitation and detection regions.

Table 2. Contents of hybridization samples and the corresponding dilution factors for imaging

HPV-16 DNA (μg)	HPV-16 DNA (%)	human genomic DNA (μg)	probes (μL)[a]	volume (μL)	dilution factor
0.1	10	0.9	1	30	5000
0.05	5	0.95	5(10)	30	5000
0.01	1	0.99	1(10)	30	1000
0.005	0.5	0.995	5(100)	30	500
0.001	0.1	0.999	1(100)	30	100
0.0005	0.05	0.9995	5(1000)	30	100
0.0001	0.01	0.9999	1(1000)	30	50
0.00005	0.005	0.99995	5(10000)	30	20
0.00001	0.001	0.99999	1(10000)	30	20
0.000005	0.0005	0.999995	5(100000)	30	10
0.000001	0.0001	0.999999	1(100000)	30	10
0	0	1	1(100000)	30	10

a The values in parentheses are the dilution factors for the probes; e.g. for 5% HPV-16 DNA, the original probe solution was diluted 10 times and then 5 μL of the diluted probe solution was added into the sample.

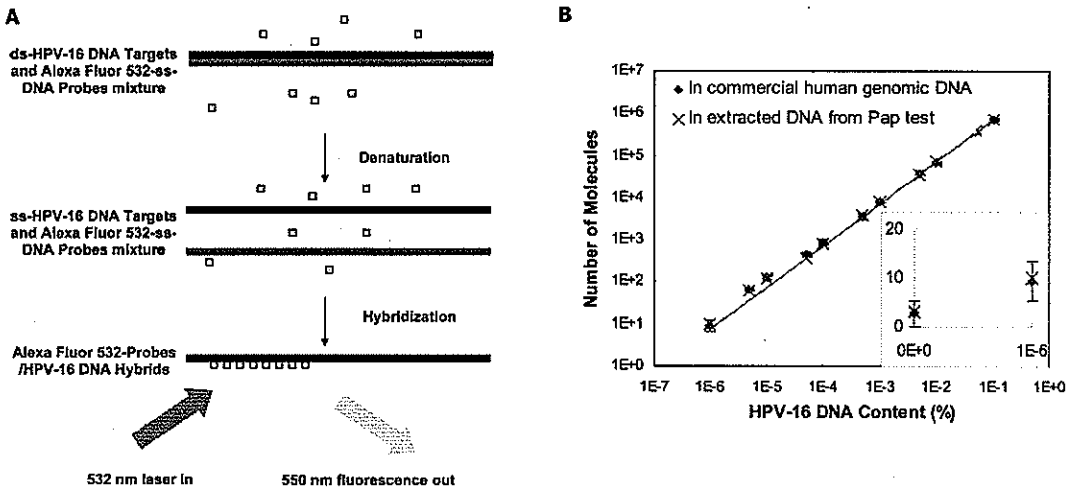

Figure 2. Single-molecule detection of HPV-16 DNA in single-color mode. (A) Schematic diagram of single-color mode. After heat denaturing, HPV-16 DNA was hybridized to the 8 complementary oligonucleotide probes labeled with Alexa Fluor 532. The 550-nm fluorescence images of the hybrids were recorded by ICCD camera when excited by 532-nm laser. (B) Standard curve of the quantification of HPV-16 DNA. Both axes are logarithmic scales. The blue points depict the averages over three experiments with commercial human genomic DNA, and the error bars are the standard deviations (Table 2) of these three sets; the red points are data from one experiment with extracted DNA from the Pap smear sample. The inset shows the number of counts in negative control and the low-concentration region.

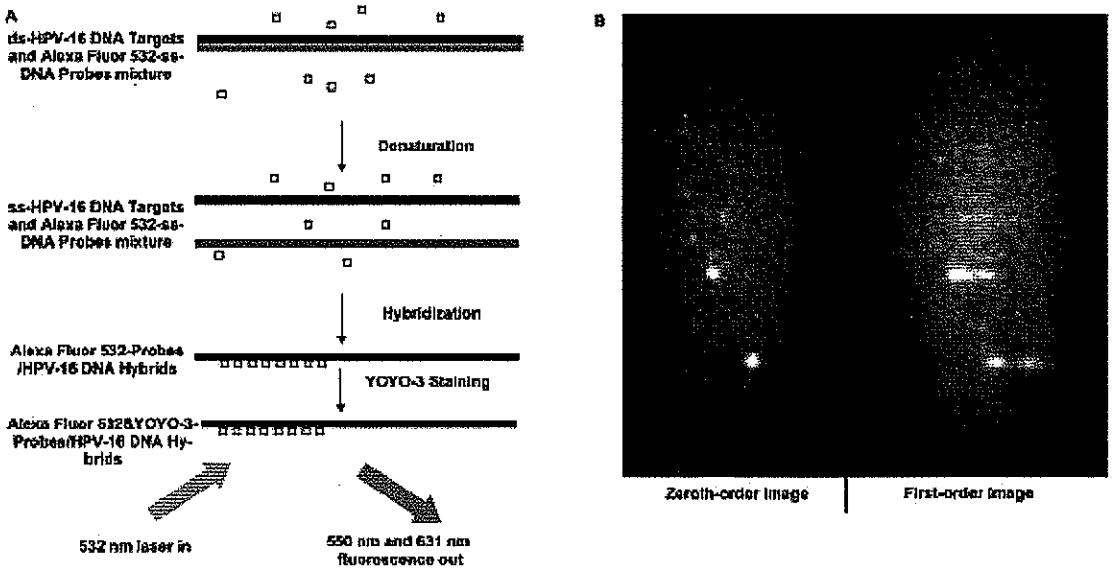

Figure 3. Single-molecule detection of HPV-16 DNA in dual-color mode. (A) Schematic drawing of dual-color imaging. After heat denaturing, HPV-16 DNA was hybridized to the complementary oligonucleotide probes labeled with Alexa Fluor 532. The hybrids were then labeled with YOYO-3. Both 550- and 631-nm fluorescence images of the hybrids were recorded by the ICCD camera when excited by 532-nm laser. (B) Fluorescence images of two Alexa Fluor 532 + YOYO-3 colabeled probes/HPV-16 DNA hybrid molecule.

Table 3. Contents of hybridization mixtures loaded on a slide

HPV-16 DNA (μg)	HPV-16 copy number per cell	HPV-16 DNA	Human genomic DNA (μg)	Volume loaded (μL)
$5x10^{-3}$	700	0.1%	5	70
$5x10^{-4}$	70	0.01%	5	70
$5x10^{-5}$	7	0.001%	5	70
$5x10^{-6}$	0.7	0.0001%	5	70
0	0	0	5	70

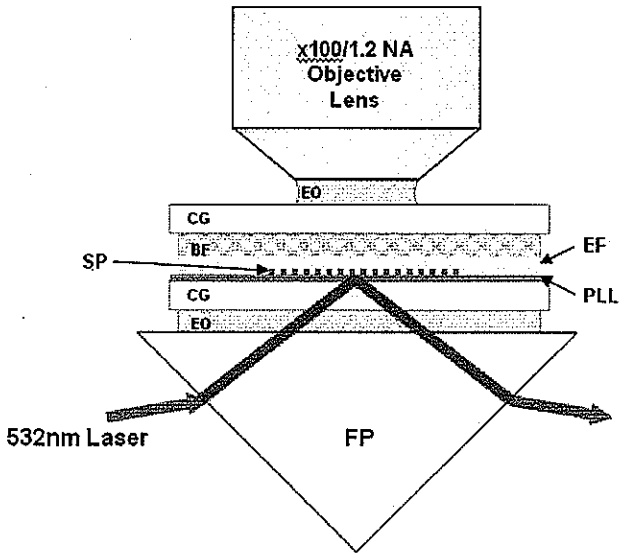

Figure 4. Schematic diagram of evanescent field layer illumination system: FP, fused silica right angle prism; EO, emulsion oil; CG, cover glass; PLL, poly-L-lysine coating; SP, sample layer of hybridized molecules on surface; EF, evanescent field layer; BF, buffer solution.

Table 4. Restriction enzyme cutting sites for HPV-16 DNA

Restriction enzyme	NcoI	HaeIII	BstN I
Recognition sequence	C/CATGG	GG/CC	CC/WGG
Cutting site in HPV-16 genome (7,904bp)	863/867	1953/1953 2591/2591 2890/2890 3400/3400 4468/4468 4530/4530 5107/5107 5661/5661 5937/5937 5992/5992 6589/6589 7062/7062 7432/7432 7766/7766	4884/4885 5529/5530 6181/6182 6865/6866 6903/6904 7429/7430

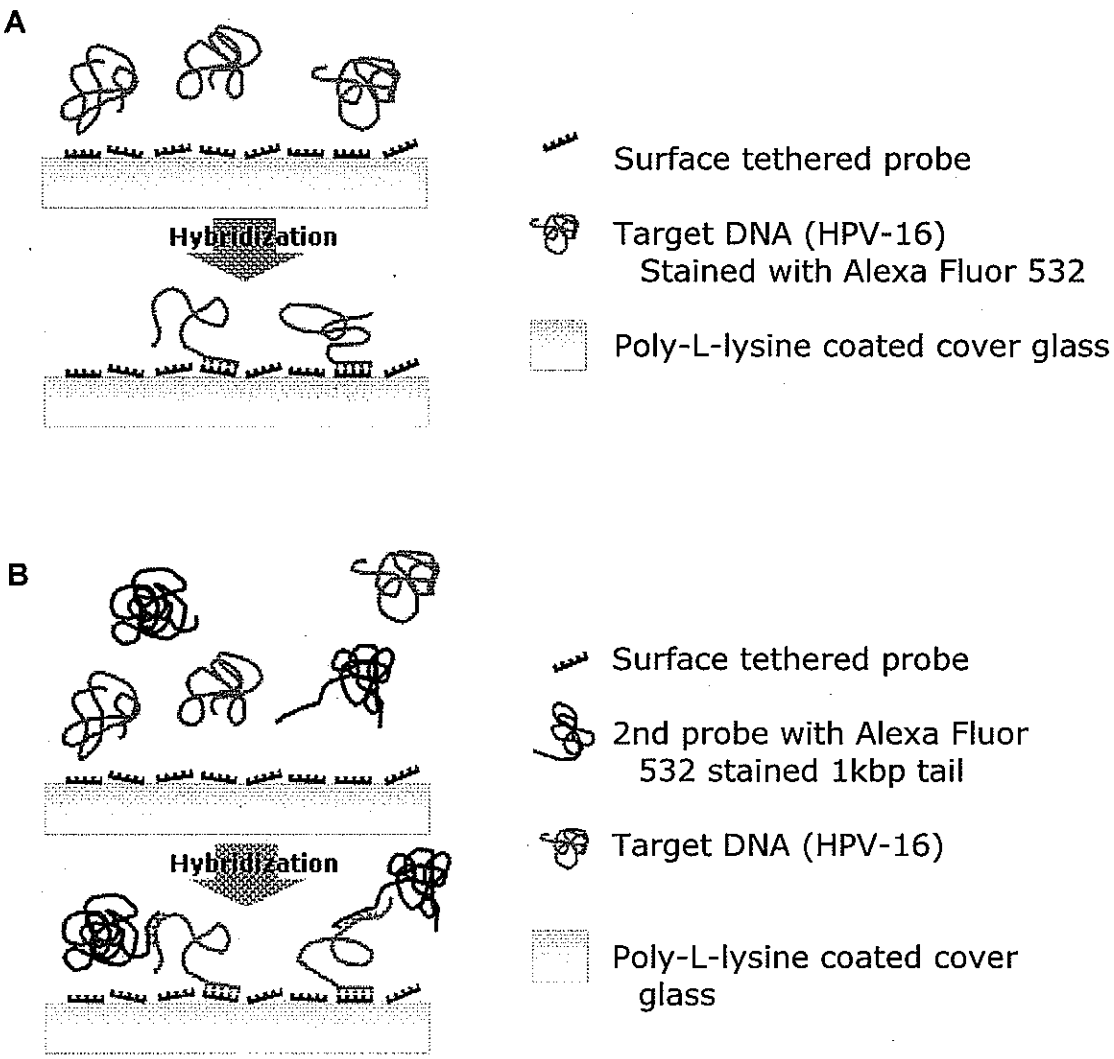

Figure 5. Two surface hybridization modes. (A) Single probe hybridization, (B) Dual probe hybridization.

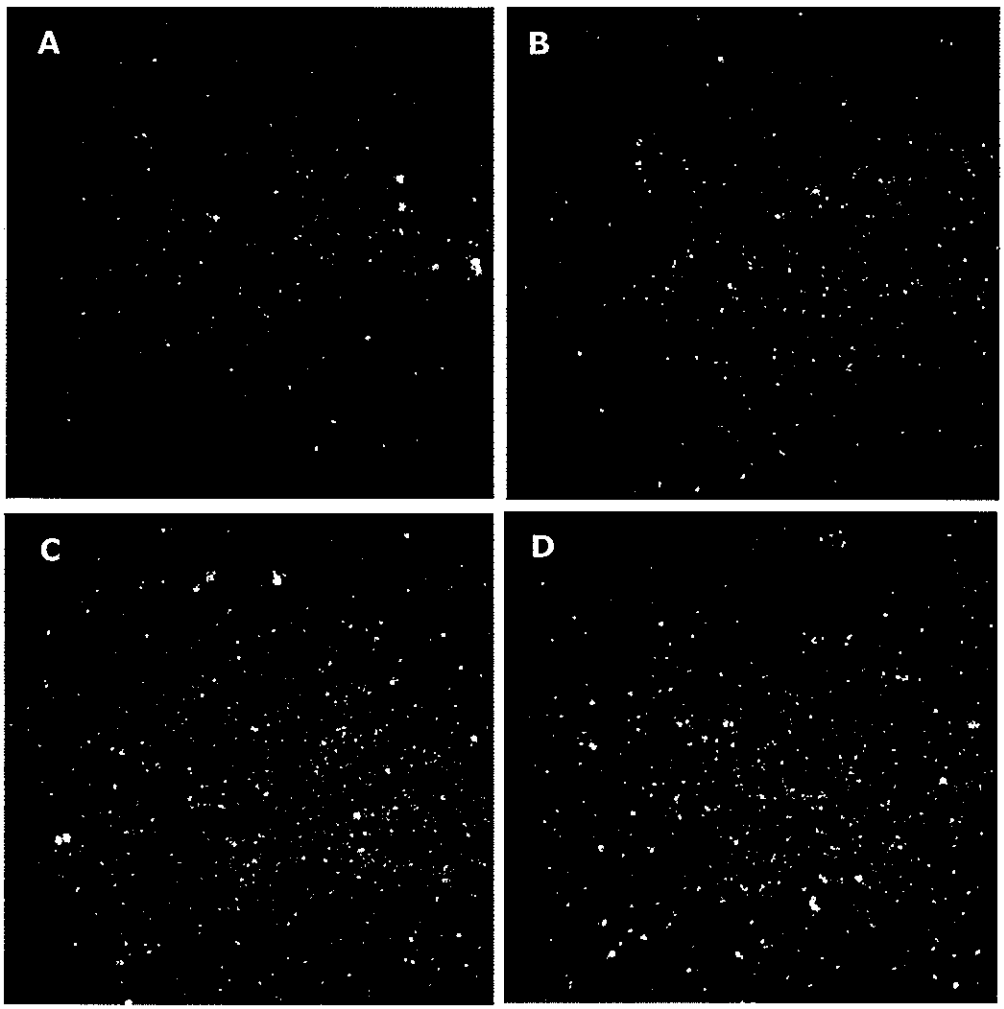

Figure 6. Single molecule images of single probe surface hybridization (A) 0% HPV DNA, (B) 10^{-4}% HPV DNA, (C) 10^{-3}% HPV DNA, (D) 10^{-2}% HPV DNA. Total amount of 5 μg human genomic DNA with added HPV-16 DNA was loaded on a glass slide that has 100nt oligo nucleotide probes on the surface. All slides were hybridized at 55 °C for 16 h. Display range: (Background Average+2σ) – (Background Average+5σ)

Table 5. Number of molecules observed in single probe hybridization assay

	0 % HPV DNA (Negative Control))	10^{-4} %HPV DNA (0.7 copy/cell)	10^{-3} %HPV DNA (7 copy/cell)	10^{-2} %HPV DNA (70 copy/cell)
Number counted	128	177	369	398
Standard Deviation	20.8	19.3	77.0	101.4
Net Count* (Expected number**)	0 (0)	49 (60)	241 (600)	270 (6000)

**Net Count* = (Number counted in each concentration experiment) – (Number counted in Negative control)

***Expected number* = Number of molecules in one frame expected, when hybridization efficiency is 100%

Table 6. Average and standard deviation of counted molecules in cell-line extract hybridization, and calculated HPV-16 DNA contents per cell

	Single probe hybridization			Dual probe hybridization		
Cell Line	CaSki (600/cell)	SiHa (1-2/cell)	C-33A (0/cell)	CaSki (600/cell)	SiHa (1-2/cell)	C-33A (0/cell)
Average count	393	177	128	117.3	6.08	3.80
Std Dev	138	22.7	27.9	16.5	3.25	1.80
Calculated HPV-16 contents (copy/cell)	≥ 70	0.7	Less than detection limit (~ 0)	815	Less than detection limit (1.22)*	Less than detection limit (0.212)*

*Calculated values using standard curve equation shown in Figure 9.

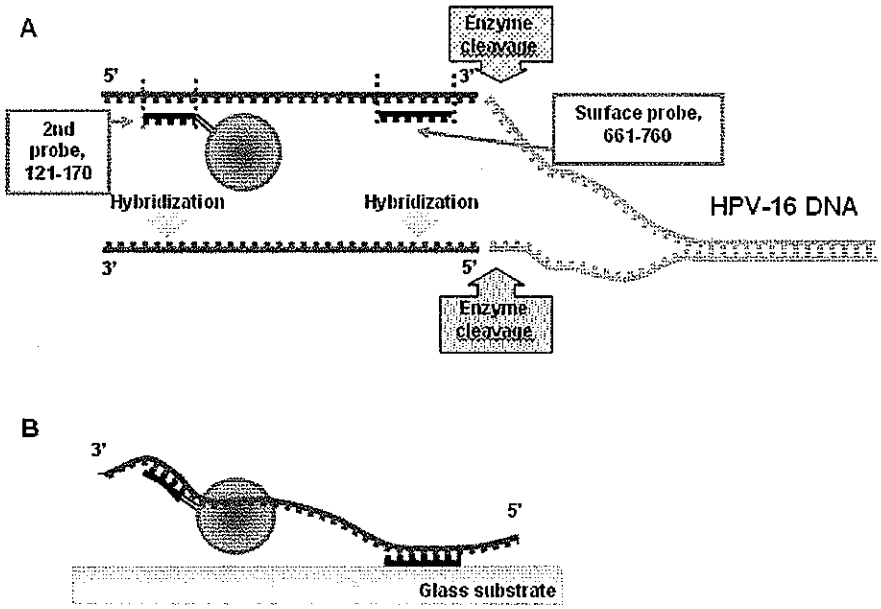

Figure 7. (A) Location of probes, (B) HPV-16 DNA fragment properly hybridized with two probes.

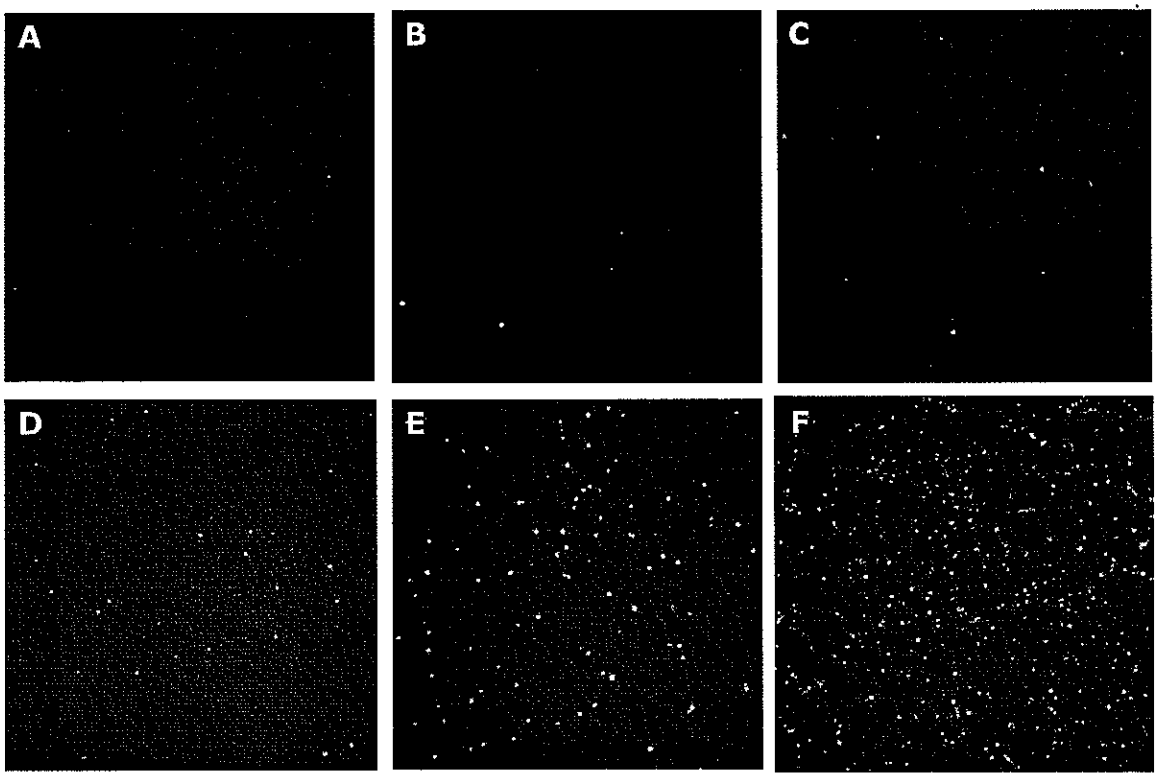

Figure 8. Frame captures from dual probe hybridization detection movies. Each dot represents one probe molecule (second probe with 1 kbp tail), stained with Alexa Fluor 532 dye. (A) 0% HPV DNA, (B) 10^{-4}% HPV DNA, (C) 10^{-3}% HPV DNA, (D) 10^{-2}% HPV DNA, T (E) 10^{-1} % HPV DNA, (F) 1% HPV DNA. Display range: 67 – 4095.

Table 7. Single molecule counts per a frame of a movie, average and standard deviation

	0 % HPV DNA (Negative Control)	10^{-4} %HPV DNA (0.7 copy/cell)	10^{-3} %HPV DNA (7 copy/cell)	10^{-2} %HPV DNA (70 copy/cell)	10^{-1} %HPV DNA (700copy/cell)	1 %HPV DNA (7000copy/cell)
Number counted	2.67	5.84	11.9	34.6	141	446
Standard Deviation	1.75	2.49	3.44	5.35	27.5	40.1

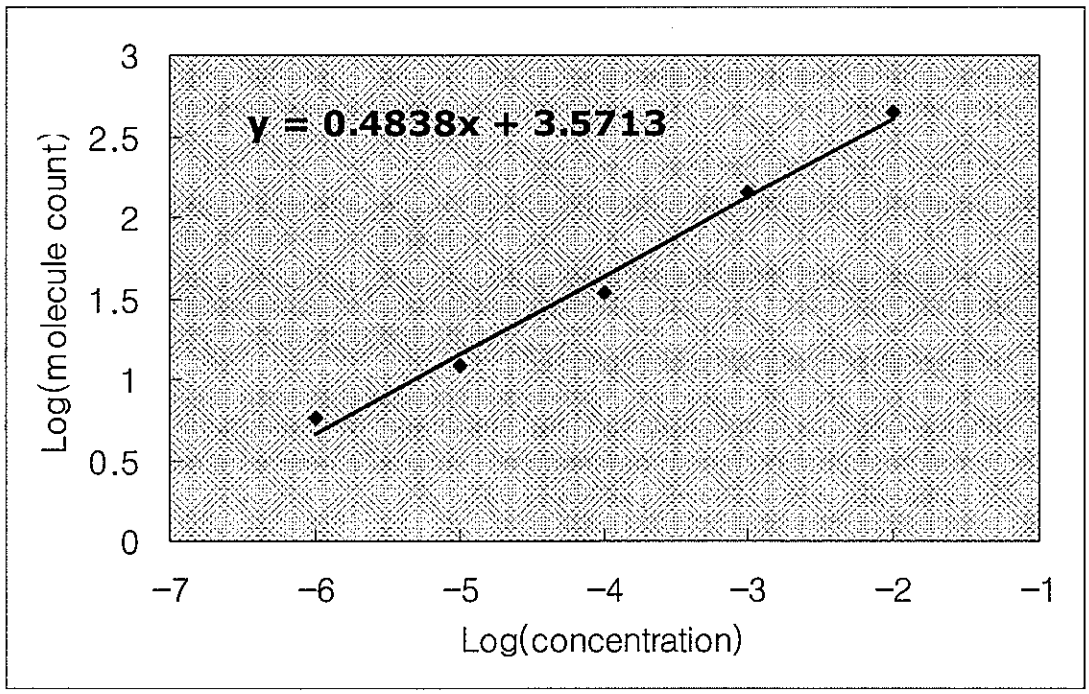

Figure 9. Standard curve of the quantification of HPV-16 DNA with dual probe surface hybridization method. Both axes are shown in logarithmic scales.

CHAPTER 4. GENERAL CONCLUSIONS

The potential of single molecule detection as an analysis tool in biological and medical fields is well recognized today. This fast evolving technique will provide fundamental sensitivity to pick up individual pathogen molecules, and therefore contribute to a more accurate diagnosis and a better chance for a complete cure. Many studies are being carried out to successfully apply this technique in real screening fields.

In this dissertation, several attempts are shown that have been made to test and refine the application of the single molecule technique as a clinical screening method. A basic applicability was tested with a 100% target content sample, using electrophoretic mobility and multiple colors as identification tools. Both electrophorestic and spectral information of individual molecule were collected within a second, while the molecule travels along the flow in a capillary. Insertion of a transmission grating made the recording of the whole spectrum of a dye-stained molecule possible without adding complicated instrumental components. Collecting two kinds of information simultaneously and combining them allowed more thorough identification, up to 98.8% accuracy.

Probing mRNA molecules with fluorescently labeled cDNA via hybridization was also carried out. The spectral differences among target, probe, and hybrid were interpreted in terms of dispersion distances after transmission grating, and used for the identification of each molecule. The probes were designed to have the least background when they are free, but have strong fluorescence after hybridization via fluorescence resonance energy transfer.

The mRNA-cDNA hybrids were further imaged in whole blood, plasma, and saliva, to test how far a crude preparation can be tolerated. Imaging was possible with up to 50% of clear bio-matrix contents, suggesting a simple lysis and dilution would be sufficient for imaging for some cells.

Real pathogen DNA of human papillomavirus (HPV) type-16 in human genomic DNA was probed with fluorescently-labeled probe molecules and imaged. When only the probes were stained and hybridized in a vial, it had 6 orders of magnitude dynamic range with a detection limit of ~0.7 copy/cell. A second dye was added to lower the false positive levels. Although there was a sacrifice of two orders of magnitude in detection limit, the number of false positives was reduced to zero.

HPV-16 DNA was also hybridized and detected on surface-tethered probes. When the entire human genomic DNA and HPV was labeled and hybridized, the detection limit was similar to that of one-color assay detected in capillary. However, non-specific adsorption was high, and the dynamic range was narrow because of saturation of the surface and electrostatic repulsion between hybridized targets on the surface. The second probe was introduced to lower non-specific adsorption, and the strategy succeeded in 4 orders of magnitude linear dynamic range in a log-log plot, along with 2.4 copies/cell detection limit.

DNA extracts of cell lines that contained a known copy number of HPV-16 DNA were tested with the four strategies described above. The calculated numbers from observed molecule counts matched the known values. Results from the Pap test sample with added HPV DNA were similar to those of purified DNA, suggesting our method is compatible with the conventional Pap test sample collection method.

Further optimization will be needed before this single molecule level detection and identification can actually be used in a real clinical lab, but it has good potential and applicability. Improvement such as automated imaging and scanning, more accurate data processing software as well as sensitive camera, should help increase the efficiency and throughput.