The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Microchip Analyzer for Forensic Short Tandem Repeat Typing of Single Cells

Author(s): Tao Geng, Richard A. Mathies

Document No.: 247280

Date Received: July 2014

Award Number: 2009-DN-BX-K180

Microchip Analyzer for Forensic Short Tandem Repeat Typing of Single Cells

USDJ Office of Justice Award Number: 2009-DN-BX-K180

Tao Geng and Richard A. Mathies

Department of Chemistry, University of California, Berkeley CA 94720

June 17th, 2013

Abstract

Short tandem repeat (STR) typing at the single-cell level is a promising tool for human forensic identification when the biological evidence materials are comprised of mixtures of cells from multiple individuals at relatively low concentrations. Here we describe a novel single-cell STR typing method with high sensitivity, fidelity and throughput that combines microfluidic droplet generation with single-cell multiplex emulsion PCR. Individual cells are separately isolated within microdroplets that subsequently function as miniaturized reactors for PCR amplification, producing high quality STR profiles from single cells at high throughput.

In our method, a microfluidic droplet generator is constructed with soft lithography using polydimethylsiloxane (PDMS) layers. Millions of 1.5 nL nanoliter monodisperse agarose-in-oil microdroplets are produced using a flow-focusing channel geometry with a high generation rate of 444 droplets per second. An individual cell along with a microbead functionalized with multiplex primers for the STR targets are statistically encapsulated within the droplets. The beads serve as amplicon-binding substrates to maintain the monoclonality of STR analysis by preventing the cross-contamination of DNA products from different droplets and different cells. The unique thermo-responsive sol-gel switching property of agarose enables the gel droplets containing the individual cells to be flexibly processed for cell lysis, amplification, mechanical manipulation and long-term storage. Following lysis and digestion of the cell-containing droplets in a chemical lysis buffer containing sodium dodecyl sulfate (SDS) and proteinase K, genomic DNA is released from the cell but remains trapped in the porous agarose network. The gel

1

droplets are then equilibrated with PCR mixture and redispersed in the carrier oil by mechanical agitation to form a uniform emulsion of nanoliter reactors. Massively parallel single-cell emulsion PCR is then performed in a single PCR tube using a conventional thermocycler, during which STR loci information from an individual cell is transferred onto the microbead within a droplet. No droplet merging is observed using a silicon oil mixture containing 1% Triton X-100 with both bovine serum albumin (BSA) and Tween 80 in PCR mixture. Following PCR amplification, the beads are recovered from the droplets by removing the oil and melting the agarose to disrupt the droplets. To analyze the STR products immobilized on the beads, secondary PCR was carried out to transfer the STR information into free solutions. The beads are quantified and diluted to the stochastic limit in standard PCR tubes or 96-well plates to serve as the DNA templates for a secondary PCR amplification. Finally, the secondary-PCR products from single beads are detected using a conventional capillary electrophoresis (CE) system for fragment sizing analysis.

To explore the utility of this method for forensic DNA typing, a 9-plex STR system was developed with eight core STR loci including D3S1358, D5S818, D7S820, D8S1179, D13S317, D21S11, vWA, and TH01 plus a sex marker, amelogenin. The protocols for the microbead-based multiplex PCR were initially optimized both in bulk solutions and microdroplets using standard 9947A female and 9948 male genomic DNA to validate the method. Under optimized procedures, we could obtain complete STR profiles of standard GM09947 (female) and GM09948 (male) human lymphoblast cell lines starting from the droplets containing 0.15 cells and 0.9 beads per droplet. The results indicated the conservation of single-genome integrity within droplets during cell lysis as well as the successful transformation of STR information from cells to microbeads. The mixtures of two human lymphoblast cells GM09947 and GM09948 were tested with cell ratios of 1:1 and 2:1. Although mixed STR profiles were observed when the cells stuck to each other, STR profiles from single cells were selectively detected in both cases and the number of dual profiles was reduced with lower cellular concentration as expected. The ability of our method to detect multiple STR loci from single cells in a mixture with high-throughput enables it to be applicable to analyzing evidence samples involving low-abundance materials and multiple suspects thereby solving the classic mixture analysis problem.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents

Executive Summary

Short tandem repeat (STR) typing has become one of the most commonly used tools for human forensic identification relying upon the collection of homogeneous, high quality and concentrated genetic samples from a crime scene. For a majority of crimes, however, the biological evidence materials are often comprised of mixtures of cells, and hence DNA, from multiple individuals at relatively low concentrations. A primary problem compromising the STR analysis is that these complex biological materials generate mixed genotypes and thus lead to subsequent challenges in interpreting the results, especially if the number of contributors exceeds two. A more difficult situation occurs when the perpetrator cells are much rarer than the victim cells, resulting in preferential amplification of the victim DNA and the inability to detect the perpetrator genotype. Although various strategies have been developed to resolve the problems, these methods are limited due to low efficiency, low throughput, high possibility of sample cross-contamination, and lack of universality.

The state-of-the-art microfluidic technology now offers a promising strategy for rapid, high-throughput generation of highly monodisperse microdroplets which serve as miniaturized reactors for biological and chemical assays. Single cells and/or particles are able to be compartmentalized within the separate aqueous droplets surrounded by immiscible carrier oil, which dramatically reduces the risk of cross-contamination among different cells. Due to the controllable droplet size, shape and uniformity, the droplet content including the reagent composition and concentration can also be precisely tuned to provide a well-defined microenvironment for the individual cells. The low volume (femtoliter to nanoliter) of the droplets allows massively parallel handling of millions of independent reactions with ultrahigh-throughput and thus single-cell analyses of vast populations to probe cellular heterogeneity and detect low-frequency events. Therefore, performing emulsion PCR within droplets for forensic DNA analysis would be of particular interest in cases where only very small amounts of mixed evidence materials are available, owing to the ability to encapsulate single cells into discrete and well-defined microdroplets having identical amplification environment.

Here we present a novel microfluidic droplet-based method to perform forensic STR typing of single cells with high sensitivity, fidelity and throughput for confident identification of the

4

genetic fingerprints of each component. We have also published this method in "Single-cell forensic short tandem repeat typing within microfluidic droplets", *Analytical Chemistry*, 2014, 86(1): 703-712. Microbead-based emulsion PCR is initially employed to efficiently copy the STR targets within a single cell onto a single microbead encapsulated within a nanoliter droplet. Up to 100 attomoles of total STR amplicon is expected to be produced on each bead which should be sufficient for the high quality forensic analysis. These beads are subsequently recovered for performance of secondary PCR under statistically dilute conditions for the purpose of conventional capillary electrophoresis (CE) fragment sizing analysis. This secondary PCR ensures that all positive samples will produce sufficient DNA for a strong CE profile to be observed. The workflow of the method is illustrated in **Figure 1**. The whole analysis process can be accomplished in about 22 h, including 3.5 h working time and 18.5 h waiting time for cell lysis (~10 h), emulsion PCR (3.5 h), secondary PCR (3 h) and CE analysis (2 h).

To explore the utility of this method for forensic DNA typing, a 9-plex STR system was developed with eight core STR loci from the combined DNA index system (CODIS) including D3S1358, D5S818, D7S820, D8S1179, D13S317, D21S11, vWA, and TH01 plus a sex marker amelogenin. The protocols for the microbead-based multiplex PCR were first optimized both in bulk solutions and in microdroplets using standard 9947A female and 9948 male genomic DNA to validate the method. Individual human cells from GM09947 and GM09948 human lymphoblast cell lines were then typed with the method for overall optimization of system performance as well as evaluation of "stochastic effects" and PCR kinetics at the nanoliter volume scale. It is important to realize that the term "stochastic effects" has a very different meaning in our study as compared to conventional forensic analysis. Conventionally the term represents the statistical variation of template copy number for a discrete locus from a genomic DNA pool. In our case it represents the statistic variation in the number of whole cells (each with a full genome) within a given PCR droplet reaction. The selectivity and performance of the forensic analysis with single-cell resolution were tested on mixtures of different cells with varying ratios.
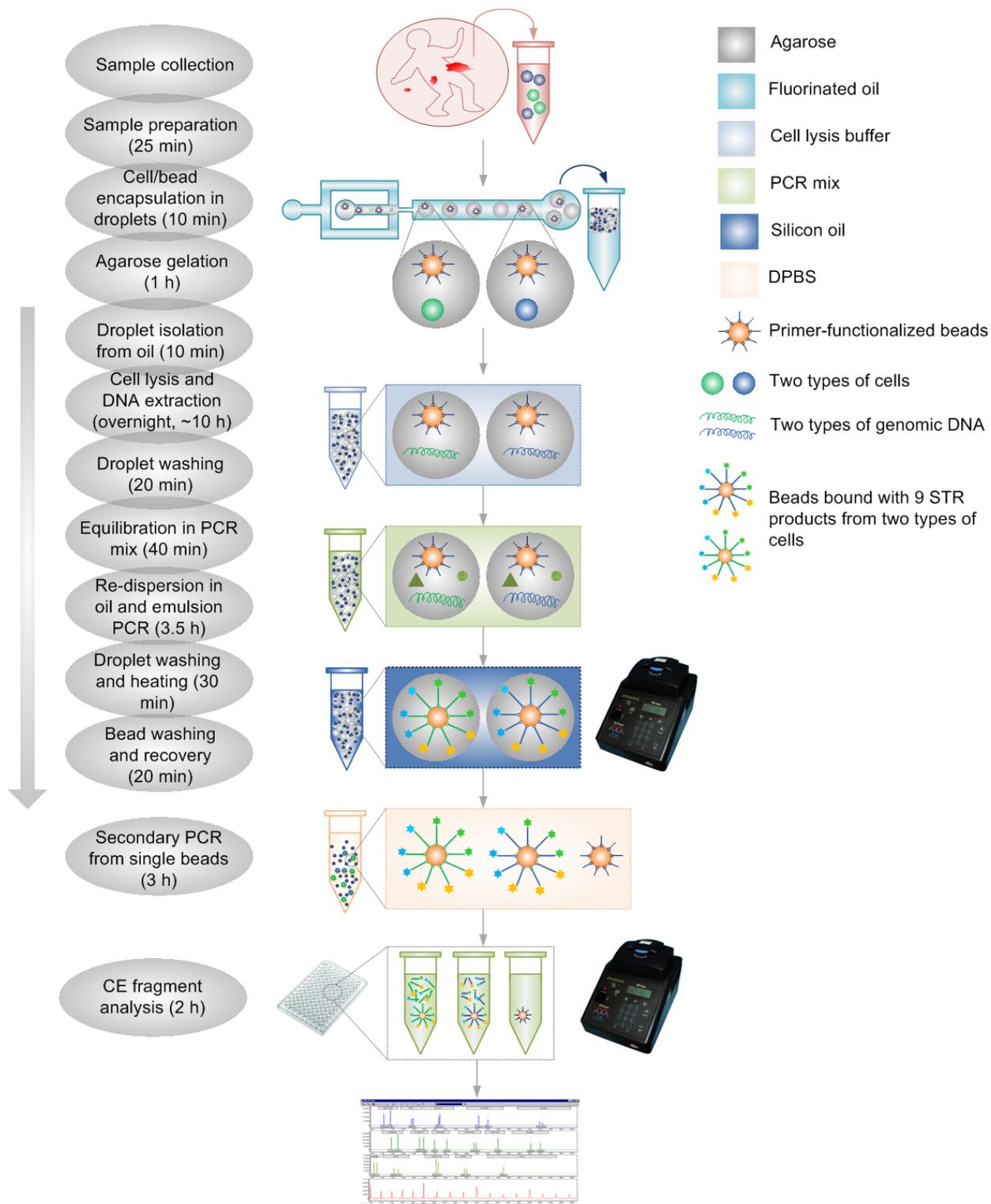
5

Figure 1. Workflow for high-throughput single-cell forensic STR typing illustrated using two types of cells. **(1)** Cell mixture is initially collected from crime scene, dispersed and suspended in appropriate buffer solutions. **(2)** Individual cells together with primer-functionalized microbeads are encapsulated within agarose microdroplets using a microfluidic chip. **(3)** The gelled droplets are incubated in cell lysis buffer to release genomic DNA. **(4)** The PCR

6

components are diffused into the gel droplets by equilibrating in PCR mixture. **(5)** After the droplets are re-dispersed in oil by mechanical agitation, emulsion PCR is performed on a thermal cycler. **(6)** After DNA amplification, beads are recovered by breaking the droplets and melting the agarose. **(7)** Secondary PCR is conducted starting from single beads in standard PCR microplates. **(8)** The STR products from single beads are processed using conventional capillary electrophoresis (CE) system for fragment sizing analysis. The total analysis time is about 22 h, including 3.5 h working time and 18.5 h waiting time for cell lysis (~10 h), emulsion PCR (3.5 h), secondary PCR (3 h) and CE analysis (2 h).

In our approach, a microfluidic droplet generator was constructed with standard soft lithography using the elastomeric polydimethylsiloxane (PDMS) due to the simple, fast, low-cost but reliable fabrication process. Surface silanization of the microchannel was performed with a fluorosilane reagent immediately after oxygen plasma exposure to increase the hydrophobic properties, thereby improving the stability of the water-in-oil droplets and prolonging the lifetime of the devices. Millions of uniform aqueous microdroplets could be rapidly produced based on a flow-focusing channel geometry where the dispersed (water) phase flowed in the central channel and the continuous (oil) phase flowed in two outside channels. The inner phase was broken into droplets inside of a small nozzle at the downstream of the channel junction owing to the hydrodynamic pressure and viscous shear stresses applied by the outer phase. The droplet size and generation frequency were dependent on the particular channel configuration, the physical properties of the fluids and the relative flow rates of the immiscible phases. The volumetric flow rates of two phases were independently modulated by two syringe pumps. Under optimized flow rates (e.g. 40 µL/min for the disperse phase and 100 µL/min for the continuous phase), a generation frequency of approximate 444 Hz could be achieved to produce 1.5 nL agarose droplets in Quantalife fluorinated oil (Bio-Rad). The ultra-low-gelling temperature agarose with a gelling point about 8-17 ºC and a remelting point of around 50 ºC was selected to facilitate the droplet generation. To further reduce agarose gelling, the whole setup including the droplet generator and the syringes was placed under a heated air stream (42-45 ºC). In addition to biocompatibility, the most significant advantage of agarose droplets is the ability to rapidly transform into microgels by simply cooling to below the gelling temperature and remain solid state unless the temperature rises above the remelting point. This feature is particularly important to subsequent mechanical manipulation and long-term storage of millions of microdroplets simultaneously while maintaining the single-genome fidelity of each compartmentalized cell.

7

Mechanical properties of the gelled microdroplets are adjusted by the concentration of the polysaccharides in the feed solutions. We found that 1.5% (w/v) of agarose provides sufficient strength without compromising the PCR efficiency.

Single cells (or desired copies of DNA molecules) along with primer-functionalized microbeads were stochastically encapsulated into agarose droplets following Poisson statistics. Thus, the average number of cells (or DNA molecules) and microbeads per droplet could be easily changed by varying their concentration suspended in the feed agarose solution. Typically, each droplet contained 0.9 beads but only 0.15 or 0.1 cells on average to ensure sufficient dilution of cells. The primer-functionalized microbeads coencapsulated in droplets acted as amplicon-binding substrates to maintain the monoclonality of genetic analysis by avoiding the cross-contamination of DNA products from different droplets after droplet disruption. The beads also facilitated downstream high-throughput manipulation and analysis. The 6% cross-linked *N*-hydroxysuccinimide (NHS)-activated Sepharose beads with 34 μm average diameter were chosen for their ability to carry sufficient multiplex STR amplicons for PCR-based CE analysis. The equimolar concentrations of 9 different primers were conjugated on beads including the reverse primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the forward primers for D3S1358, D5S818, vWA, and D7S820. Amelogenin was selected for sex-typing and sample quality evaluation. Primer sequences were designed based on the sequences used in Promega PowerPlex® 16 System. The primers were amine-functionalized so that they could be immobilized onto the beads *via* standard NHS ester reaction chemistry.

Encapsulation of cells in agarose droplets allowed robust and reproducible single-cell DNA extraction in parallel. After the gel droplets were isolated from the oil using a cell strainer with 40 μm nylon mesh (BD Biosciences) followed by extensive washes with nuclease-free water, the cell-containing droplets were incubated in cell lysis buffer containing 0.5% sodium dodecyl sulfate (SDS), 100 mM EDTA, 10 mM Tris-HCl, pH 8 and 0.1 mg/mL proteinase K at 37 ºC overnight to allow the genomic DNA liberation and the enzymatic protein digestion. This step was especially critical to PCR assays of mammalian cells, as it removed the vast majority of histones and other nuclear proteins that inhibit polymerase activity, thereby enhancing the PCR efficiency. Following cell lysis and digestion, the released high-molecular-weight DNA

8

remained trapped in the porous network of gelled agarose. DNA-containing microdroplets were then washed with 2% Tween 20 in water to reduce the potential PCR inhibitory effects of SDS. Afterwards, they were washed with 100% ethanol to inactivate remaining proteinase K, followed by several washes with water containing 0.02% Tween 20. Tween 20 was added to prevent the attachment of droplets on the tube wall which would result in sample loss. The agarose droplets not immediately used could be stored at 4 ºC in 100% ethanol for at least a week and washed with water just prior to emulsion PCR.

Prior to emulsion PCR, a certain amount of droplets were incubated in PCR mix for 30 min with occasional agitation to enable the transport of PCR components containing fluorescently labeled primers into the agarose matrix structure. Bovine serum albumin (4 µg/µL BSA) and 0.01% Tween 80 were included in the mixture to ensure the thermostability of the droplets during PCR cycling. Small amounts of primers that were already conjugated on microbeads for each locus to be amplified were also incorporated into the solution mixture in order to initiate the solid-phase PCR. These gel droplets were redispersed in sufficient volume of oil by mechanical vibration at a frequency of 17 Hz for 30 s using a Qiagen TissueLyser mixer. Uniform nanoliter reactors were produced in a standard PCR tube for massively parallel single-cell PCR using a conventional thermocycler. The solution fraction of the PCR mix was agitated into microemulsions less than 1 µm in diameter which potentially improved the stability of agarose emulsion during the amplification process. After melting during the hot start phase of PCR at 95 ºC, the agarose droplets remained liquid throughout the whole thermal cycling process, enhancing the mixing rate of reagent and amplicon within the nanoreactors. The composition of the carrier oil was a critical factor to the success of emulsion PCR. We used freshly prepared oil containing DC 5225C formulation aid (Dow Chemical), KF-7312J fluid (Shin-Etsu Silicones), AR20 silicone oil (Sigma-Aldrich), and Triton X-100 surfactant (Sigma-Aldrich) with a mass ratio of 40:30:30:1 to ensure the good performance of emulsion PCR. When using the Quatalife oil, droplet merging was observed after emulsion PCR.

Following PCR amplification, the agarose droplets were gelled and harvested by centrifugation. The residual oil was completely removed by washes with 100% isopropanol once, 100% ethanol once and Dulbecco's PBS (DPBS) containing 0.02% Tween 20 for 5 times. The microbeads

carrying multiplex STR products were recovered from the droplets through melting the agarose at 62 ºC for 20 min to disrupt the droplets, followed by extensive washes in DPBS to remove the DNA fragments not linked to the beads. To detect the STR products immobilized on the beads, secondary PCR was carried out to transfer the STR information into free solution. The beads were quantified and diluted at appropriate dilute statistical concentrations in standard PCR tubes or 96-well PCR plates to serve as the DNA templates for reamplification. Finally, the secondary-PCR products in free solution were processed using a conventional CE system (Applied Biosystems 3730XL DNA Analyzer) for fragment sizing analysis.

The efficiency of PCR performed on a solid phase surface is generally lower than that obtained from PCR in solution due to steric hindrance and charge repulsion. In particular, when multiple targets are involved in an amplification reaction, PCR yield will be low due to competition between each locus, and the amplification process for each locus is not fully balanced. Therefore, we first validated the microbead-based solid-phase PCR for STR typing with 9947A female and 9948 male standard genomic DNA in bulk solutions (40 pg/µL or ~13 copies/µL). The multiplex-primer microbeads (with a final concentration of 200 beads/µL) were incorporated into the PCR mixture containing 1X *Ampli*Taq® Gold PCR buffer, 1.5 mM MgCl$_2$, 200 µM each dNTP, 4 µg/µL BSA, 0.01% Tween 80, primer mixture and 0.2 U/µL *Ampli*Taq® Gold DNA polymerase. BSA and Tween 80 were added to be compatible with microdroplet-based emulsion PCR. DNA amplification was initiated with the presence of the primers that were not conjugated on beads (the forward primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the reverse primers for D3S1358, D5S818, vWA, and D7S820) and 1/10 amount of corresponding reverse-direction primers (the reverse primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the forward primers for D3S1358, D5S818, vWA, and D7S820) in the free solution. The primers not bound on beads were fluorescently labeled with 6-FAM (6-carboxyfluorescein), TAMRA (carboxytetramethylrhodamine) or JOE (6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein) dyes. The PCR thermal cycling conditions and primer concentrations were systematically optimized. The thermal cycling protocol was composed of initial activation of the AmpliTaq Gold DNA polymerase at 95 ºC for 10 min, followed by 10 cycles of 94 ºC for 1 min, 58 ºC for 1 min, 70 ºC for 1.5 min, 22 cycles of 90 ºC for 1 min, 58 ºC for 1 min, 70 ºC for 1.5 min, and a final extension step for 30 min at 60 ºC. The STR profiles of

10

both types of genomic DNA are in agreement with the well established locus information, and the peaks are well balanced under the optimized primer concentrations.

Subsequently, the protocol of the secondary PCR was investigated by reamplifying single beads conjugated with multiplex STR fragments. The microbeads from the first bulk PCR reaction were washed with DPBS, counted using a hemocytometer, and statistically diluted into standard PCR tubes. By detecting the DNA in the supernatants of each wash, we found that 8 washes could completely remove the PCR products suspended in the free solution and nonspecifically bound on the beads so that the results were solely induced by amplicons linked to the beads. The secondary PCR was accomplished using Promega Gold ST*R buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 160 μg/ml BSA, 200 μM each dNTP) and 0.1 U/μL AmpliTaq Gold DNA polymerase. Unlike the first-round of PCR, the concentrations of forward and reverse primers for each STR locus were the same. The amplification protocol involved 10 min hot start at 95 ºC, then 10 cycles of 94 ºC for 30 s, ramp at the rate of 0.5 ºC/s to 58 ºC, hold for 30 s, ramp at the rate of 0.3 ºC/s to 70 ºC, hold for 45 s, followed by 15 cycles of 90 ºC for 30 s, ramp at the rate of 0.5 ºC/s to 58 ºC, hold for 30 s, ramp at the rate of 0.3 ºC/s to 70 ºC, hold for 45 s, and a final extension step at 60 ºC for 30 min. 25 cycles were verified to be enough for CE detection. Totally 15 samples were tested for each kind of DNA, and two bead concentrations (0.8 and 0.15 beads/reaction) were used for beads with products from 9947A DNA to verify the statistics. The results show that in 15 samples, 7 and 2 samples were positive when the bead concentration was 0.8 and 0.15 beads/reaction, respectively, corresponding well to the theoretical value of 8.25 (55%) and 2.1 (14%) predicted by the Poisson distribution.

After establishing the two rounds of 9-plex PCR in bulk solution, we then translated the process into the microdroplet format. Single-molecule emulsion PCR was initially performed with 9947A female genomic DNA based on the optimized procedure. Two DNA concentrations were tested: 0.2 and 2 copy/droplet, while maintaining bead concentration at 0.9 beads/droplet. In the secondary PCR, single beads were diluted to a concentration of 0.15 beads/reaction in order to reduce the possibility of more than one bead in each well of a PCR plate to <1% (the possibility of the reactions involving beads is 14%). Twenty samples were tested. The results indicate that 3 and 2 samples were positive when involving 0.2 and 2 copies/droplet of DNA, respectively,

which was generally in agreement with the Poisson distribution (theoretically $20 \times 0.14 = 2.8$). No peaks were detected in the negative samples which is a critical and nontrivial result when performing single copy experiments. In the positive samples, each single bead exhibited distinct STR peak profiles, which suggested that the fragments of genomic DNA were randomly encapsulated in the nanoliter droplets and thus each bead was conjugated with amplicon from STR loci located in different chromosomes. When the DNA concentration was as low as 0.2 copies/droplet, fewer DNA fragments are encapsulated within each droplet, and thus only 2 to 3 STR loci were detected from each bead. When the DNA concentration increased to 2 copies/droplet, 6 to 7 STR loci could be observed per reaction.

Single-cell emulsion PCR was then conducted using two standard cell lines GM09947 (female) and GM09948 (male) human lymphoblast cells as models. Each agarose droplet encapsulated approximately 0.15 cells and 0.9 microbeads, which resulted in the statistical possibility that 14% of beads should be bound with STR amplicons (positive). Each positive bead was expected to contain all the products from the 9 STR loci. To detect the positive samples, 20 samples were tested in the secondary PCR with the bead concentration of 0.9 beads/reaction corresponding to 0.126 positive beads/reaction. The results show that 1 and 2 samples were positive for GM09947 and GM09948 cells, respectively, which was consistent with the theoretical calculation ($20 \times 0.118 = 2.36$). The profiles of the negative samples were also very clean. The STR profiles for the two cell types contained all the peaks of the 9 STR loci, and the two positive samples for GM09947 cells demonstrated similar pattern, indicating the conservation of single-genome fidelity within droplets during cell lysis as well as the successful transformation of STR information from cells to microbeads.

The mixtures of GM09947 (female) and GM09948 (male) human lymphoblast cells with different ratios were then tested to verify the selectivity and sensitivity of single-cell STR typing. To reduce the possibility of cell aggregation in the agarose feed solution, the cells were further diluted to achieve approximate 0.1 cells per droplet while the bead concentration remained unchanged (0.9 microbeads/droplet), predicting that 9.5% of beads should be positive. The reaction number of the secondary PCR was increased to 40 to improve counting statistics, and the bead concentration increased to more than 1 (1.5 and 2) beads/ reaction resulted in 0.1425

12

and 0.19 positive beads/reaction. No peaks appeared in the negative samples, which confirmed that negative beads did not affect the PCR amplification starting from positive beads. Therefore, the high bead concentration in secondary PCR ensured that more useful positive data could be obtained from fewer PCR reactions, and thus greatly reduced the consumption of PCR reagents and the cost of the method. When the cell ratio was 1:1, 5 samples were positive based on 1.5 bead/reaction (0.1425 positive bead/reaction), which closely agrees with the theoretically predicted Poisson distribution of 13%. Among the 5 positive samples, we could detect the complete STR profiles from a single GM09948 cell and a GM09947 cell. A mixed STR profile containing all the peaks from the two cell types was also detected, which was likely caused by cell sticking. Allelic drop-out of a TH01 peak was observed in the profile from a GM09947 cell, possibly due to the preferential amplification in the individual cell or genomic heterogeneity of cells in this cell culture. Similarly, the dropout of the TH01 allele also occurred in a mixed profile, though the heights of the peaks specific for GM09948 cells were much lower. When the GM09947 (female) to GM09948 (male) cell ratio increased to 2:1, 8 samples were positive based on 2 bead/reaction (0.19 positive bead/reaction), which is generally consistent with the theoretical prediction of 17%. Among the 8 positive samples, 3 single GM09947 cells, 1 single GM09948 cell and 4 mixed cells were detected. The heights of the peaks specific for GM09948 cells were much lower in 1 mixed profile. To further improve the reliability and accuracy, lower cell concentration will be tested to avoid mixed STR profiles generated by cell aggregates. More cells will be studied to explore the cell heterogeneity, and the primer concentrations will be further optimized if necessary to balance the peaks.

In summary, we have developed an agarose microfluidic droplet method to separately type single cells in a highly parallel manner. The expected profiles of 9 STR loci could be successfully detected from pure and mixed single cells (GM09947 and GM09948 human lymphoid cells) with high single-genome integrity. Improved sensitivity, resolution, reliability, robustness and speed of single-cell STR typing will lead to more accurate and faster results at crime laboratories in cases of evidence samples containing low amounts of cells or mixed cells. We envision that this novel technology will be applicable to analyzing real-world samples in the casework involving low-abundance evidence materials and multiple suspects and will open up many new novel uses of "touch evidence".

1. Introduction

Short tandem repeat (STR) typing is a powerful tool in modern forensics.[1-4] PCR-based amplification of multiple STR loci has become a gold standard for human forensic identification, relying upon the collection of homogeneous, high quality and concentrated genetic samples from a crime scene.[5,6] However, the biological evidence materials collected from a majority of crimes are often comprised of mixtures of cells, and hence DNA, from multiple individuals at relatively low concentrations. For instance, sexual assault casework typically involves mixed specimens from a sexual assailant and a victim. A primary problem facing the STR analysis is that these complex biological materials generate mixed genotypes and thus lead to subsequent challenges in interpreting the results, especially if the number of contributors exceeds two. A more difficult situation occurs when the perpetrator cells are much rarer than the victim cells, resulting in preferential amplification of the victim DNA and thus more ambiguous genotype inference. A variety of strategies have been developed to separate different cell populations prior to analysis to reduce the requirement for mixture interpretation, including differential extraction,[7-9] filtration,[10] fluorescence-activated cell sorting,[11] and microchip-based separation[12-14]. More recently, laser microdissection[15,16] and micromanipulation[17,18] are employed to analyze at the single-cell level. However, these methods are limited due to their low efficiency, low throughput, high possibility of sample cross-contamination, and lack of universality. In many cases, the same type of cells from distinct donors cannot be easily isolated from a mixed trace. Moreover, additional DNA loss is inevitable during sample preparation process, which is not applicable to analyzing minute amounts of cellular materials.

The state-of-the-art microfluidic technology offers a promising strategy for rapid, high-throughput generation of highly monodisperse microdroplets to serve as miniaturized reactors for biological and chemical assays.[19-21] Single cells and/or particles are able to be compartmentalized within the separate aqueous droplets surrounded by immiscible carrier oil, which dramatically reduces the risk of cross-contamination among different cells. Due to the controllable droplet size, shape and uniformity, the droplet content such as the reagent composition and concentration could also be precisely tuned to provide a well-defined microenvironment to the individual cells. The low volume (femtoliter to nanoliter) of the droplets allows massively parallel handling of millions of independent reactions with ultrahigh-

throughput and thus single-cell analyses of vast populations to probe cellular heterogeneity and detect low-frequency events. In addition, a myriad of techniques have been proposed to fuse, split, mix, store, and sort the microdroplets to facilitate the various assays on demand.[22] Our previous work established a high-throughput single-cell PCR process based on the use of uniform nanoliter microfluidic droplets.[23-25] Therefore, performing emulsion PCR within droplets for forensic DNA analysis would be of particular interest in cases where only very small amounts of mixed evidence materials are available, owing to the ability to encapsulate single cells into discrete and well-defined microdroplets having identical amplification environment.

Here we present a novel microfluidic droplet-based method to perform forensic STR typing of single cells with high sensitivity, fidelity and throughput for confident identification of the genetic fingerprints of each component.[26] (Please refer to the methods in "Single-cell forensic short tandem repeat typing within microfluidic droplets", *Analytical Chemistry*, 2014, 86(1): 703-712) In our method (**Figure 1**), a single cell suspension collected from a crime scene is encapsulated into microdroplets using a microfluidic device along with primer-functionalized microbeads. Microbead-based emulsion PCR is employed to efficiently transfer replicas of the STR targets from a single cell onto a single microbead encapsulated within a nanoliter droplet. Up to 100 attomoles of total STR amplicon is expected to be produced on each bead which should be sufficient for the high quality forensic analysis in principle.[23] These beads are subsequently recovered to conduct secondary PCR under statistically dilute conditions for the purpose of conventional capillary electrophoresis (CE) fragment sizing analysis. The total analysis time is about 22 h, including 3.5 h working time and 18.5 h waiting time for cell lysis (~10 h), emulsion PCR (3.5 h), secondary PCR (3 h) and CE analysis (2 h). To explore the utility of this method for forensic DNA typing, a 9-plex STR system was developed with eight core STR loci from the combined DNA index system (CODIS) including D3S1358, D5S818, D7S820, D8S1179, D13S317, D21S11, vWA, and TH01 plus a sex marker amelogenin. The protocols for the microbead-based multiplex PCR were first optimized both in bulk solutions and in microdroplets using standard 9947A female and 9948 male genomic DNA to validate the method. Individual human cells from GM09947 and GM09948 human lymphoblast cell lines were then typed with the method for overall optimization of system performance as well as evaluation of "stochastic effects" and PCR kinetics at the nanoliter volume scale. The selectivity

15

and performance of the forensic analysis with single-cell resolution were finally tested on mixtures of different cells with varying ratios.

2. Methods

2.1 Microfluidic device fabrication

The devices were fabricated with PDMS on glass using standard soft lithography technique. Briefly, a photomask (CAD/Art Services, Bandon, OR) was first created based on the microscale patterns designed by AutoCAD software and then printed on a high-resolution transparency. A master was manufactured by patterning a 130 μm thick layer of negative photoresist SU-8 2075 (MicroChem, Newton, MA) on a 4-inch silicon wafer. A layer (~5 mm thick) of degassed PDMS (Sylgard 184; Dow Corning, Midland, MI) prepolymer mixture with a mass ratio of 10:1 (base: curing agent) was poured onto the master and baked at 80 ℃ for 1 h. The cured PDMS replica was then peeled off and punched to produce access holes for inlets and outlets. Glass slides were cleaned with isopropanol and then blown dry. After the PDMS and a pre-cleaned glass slide were oxidized in a plasma cleaner (Tegal, Petaluma, CA), the PDMS was immediately brought into contact against the slide to form closed channels. After baking at 80 ℃ for 5 min, the microchannels were treated with a 0.1% solution of (heptadecafluoro-1,1,2,2-tetrahydrodecyl)dimethylchlorosilane (Gelest, Morrisville, PA) in 100% ethanol for 10 min to increase the surface hydrophobicity. The treated device was rinsed with 100% ethanol to remove excess solutions and baked at 100 °C overnight to ensure strong bonding between the PDMS and the glass.

2.2 Preparation of primer-functionalized beads

To avoid sample contamination, all reagents were handled in a UV-treated laminar flow hood (UVP, Upland, CA). All primers used in this study were purchased from IDT (Coralville, IA) and designed based on the sequences and fluorescence dye labeling used in Promega PowerPlex® 16 System (**Table 1**). Equimolar 5'-amine-modified primers with C12 spacers including the reverse primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the forward primers for D3S1358, D5S818, vWA, and D7S820 were initially combined together to a final total concentration of 2 mM (0.22 mM for each primer) in 20 μL water. Afterwards, approximate 0.1 g of packed microbeads were removed from a 1 mL HiTrap *N*-hydroxysuccinimide (NHS)-activated Sepharose HP column (Amersham Biosciences,

16

Piscataway, NJ). To completely remove the isopropanol where the beads were stored, the beads were immediately washed with ice-cold 0.1 N HCl for 3 times, ice-cold water once and ice-cold phosphate buffered saline (PBS, pH 7.4; Life Technologies, Grand Island, NY) for 3 times, and resuspended in 0.4 mL PBS. The primer mixture was then added to the bead solutions and incubated overnight at room temperature on a rotator with a low speed. Primer-bead complexes were washed with PBS for 5 times to remove any unbound primers. The beads were finally resuspended in nuclease-free water at a final concentration of 6×10^6 beads/mL and stored at 4 ℃ until use. The bead concentration was determined by counting with a hemacytometer (Hausser Scientific, Horsham, PA).

Table 1. Primer information for multiplex PCR.

Primer		Sequence and Dye Labeling	Cont. 1 (μM)	Cont. 2 (μM)
Amelogenin	F	**[TAMRA]**-CCCTGGGCTCTGTAAAGAA	0.18	0.18
	R	ATCAGAGCTTAAACTGGGAAGCTG	0.018	0.18
vWA	F	GCCCTAGTGGATGATAAGAATAATCAGTATGTG	0.025	0.25
	R	**[TAMRA]**-GGACAGATGATAAATACATAGGATGGATGG	0.25	0.25
D8S1179	F	**[TAMRA]**-ATTGCAACTTATATGTATTTTTGTATTTCATG	0.66	0.66
	R	ACCAAATTGTGTTCATGAGTATAGTTTC	0.066	0.66
TH01	F	**[FAM]**-GTGATTCCCATTGGCCTGTTC	0.2	0.2
	R	ATTCCTGTGGGCTGAAAAGCTC	0.02	0.2
D3S1358	F	ACTGCAGTCCAATCTGGGT	0.02	0.2
	R	**[FAM]**-ATGAAATCAACAGAGGCTTGC	0.2	0.2
D21S11	F	ATATGTGAGTCAATTCCCCAAG	0.062	0.62
	R	**[FAM]**-TGTATTAGTCAATGTTCTCCAGAGAC	0.62	0.62
D5S818	F	GGTGATTTTCCTCTTTGGTATCC	0.022	0.22
	R	**[JOE]**-AGCCACAGTTTACAACATTTGTATCT	0.22	0.22
D7S820	F	**[JOE]**-ATGTTGGTCAGGCTGACTATG	0.45	0.45
	R	GATTCCACATTTATCCTCATTGAC	0.045	0.45
D13S317	F	**[JOE]**-ATTACAGAAGTCTGGGATGTGGAGGA	0.1	0.1
	R	GGCAGCCCAAAAAGACAGA	0.01	0.1

Abbreviations F: Forward; R: Reverse; Cont. 1: Primer Concentrations used in the first-round of PCR (for bulk PCR) or emulsion PCR; Cont. 2: Primer Concentrations used in the secondary PCR

17

2.3 Cell culture and sample preparation

To avoid sample contamination, all reagents were handled in a UV-treated laminar flow hood. The GM09947 (female) and GM09948 (male) human lymphoblast cell lines (Coriell Institute for Medical Research, Camden, NJ) were grown in RPMI 1640 medium supplemented with 15% fetal bovine serum (FBS), 2 mM L-glutamine and 100 U/ml penicillin−100 mg/mL streptomycin at 37 ºC in a humidified incubator containing 5% CO_2. Once harvested, the cells were washed for three times in PBS by centrifuging at 300 g for 5 min, resuspended in PBS at a final concentration of 10^6 cells/mL, and incubated at 37 ºC prior to use. Meanwhile, 3% agarose solution, primer-functionalized microbead solution (6×10^6 beads/mL) and PBS were also heated to 60 ºC to avoid agarose gelation. For droplet generation, these solutions were combined with appropriate volumes to achieve a final concentration of 10^5 or 0.67×10^5 cells/mL and 6×10^5 beads/mL in 1.5% agarose. The 3% agarose solution was prepared by dissolving 3 g ultra-low-gelling temperature agarose (Type IX; Sigma-Aldrich, St. Louis, MO) in 100 mL PBS at 70 ºC for 5 h to remove bubbles and store at room temperature until use.

For experiments involving the GM9947A female and GM9948 male standard genomic DNA (Promega, Madison, WI), the DNA samples were diluted to desired concentrations using nuclease-free water. The DNA amount was calculated assuming 3 pg genomic DNA per copy.

2.4 Microdroplet generation

The microfluidic device was mounted on an inverted microscope to monitor droplet generation. The whole setup was continually heated using a heated air stream (42-45 ºC) to prevent agarose gelation in the device. The two sample inlets were connected to two independently controlled syringe pumps (PHD 2000 infusion pump; Harvard Apparatus, Holliston, MA) through polytetrafluoroethylene (PTFE) plastic tubing (Small Parts, Logansport, IN). Water-in-oil droplets were generated by injecting the molten agarose solution containing cells and beads as well as the Quantalife fluorinated oil (Bio-Rad, Hercules, CA) into the microchannel under optimized flow rates of 40 µL/min for agarose and 100 µL/min for oil. The agarose droplets were collected in 0.5 mL PCR tubes and immediately gelled at 4 ºC. After agarose gelation for at least 1 h, the droplets were isolated from the oil using a cell strainer with 40 µm nylon mesh (BD Biosciences, San Jose, CA), extensively washed with nuclease-free water, and resuspended in nuclease-free water.

18

2.5 Cell lysis and DNA purification

Cells were lysed by combining equal volume of droplet suspension and 2X cell lysis buffer [1% sodium dodecyl sulfate (SDS; Sigma-Aldrich), 200 mM EDTA (Life Technologies), 20 mM Tris-HCl (Life Technologies) and 0.2 mg/mL proteinase K (Roche Applied Science, Indianapolis, IN)] and incubated overnight at 37 ºC. Droplets were washed with 2% (w/v) Tween 20 (Sigma-Aldrich) to reduce the potential PCR inhibitory effects of SDS. Afterwards, they were washed with 100% ethanol to inactivate any remaining proteinase K, followed by 5 washes with water containing 0.02% (w/v) Tween 20. Tween 20 was added to prevent the attachment of droplets on the tube wall and hence reduce the sample loss. The gel droplets were finally washed with water twice and resuspended in nuclease-free water. Samples not immediately used could be stored at 4 ºC in 100% ethanol for at least one week and washed with water just prior to emulsion PCR.

2.6 Emulsion PCR

To avoid sample contamination, all reagents were handled in a UV-treated laminar flow hood. The emulsion PCR mixture was composed of 1X $AmpliTaq$® Gold PCR buffer, 1.5 mM $MgCl_2$, 200 μM dNTP, 4 μg/μL heat-inactivated BSA, 0.01% Tween 80, 0.2 U/μL $Ampli$Taq® Gold DNA polymerase, primer mixture for 9 STR targets and 17.22 μL agarose droplets containing purified single-cell genomes in each 50 μL PCR reaction. The primer mixture contained the fluorescent-dye labeled primers (the forward primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the reverse primers for D3S1358, D5S818, vWA, and D7S820) and 1/10 amount of corresponding reverse-direction primers (the reverse primers for amelogenin, TH01, D13S317, D21S11, and D8S1179 as well as the forward primers for D3S1358, D5S818, vWA, and D7S820). (**Table 1**) The droplets were incubated in the PCR cocktail in standard 0.5 mL PCR tubes for 30 min at 4 ºC with occasional agitation to enable the diffusion of PCR components into agarose matrix structure. The fresh carrier oil was prepared before each run of emulsion PCR containing DC 5225C formulation aid (Dow Chemical, Miland, MI), KF-7312J fluid (Shin-Etsu Silicones, Akron, OH), AR20 silicone oil (Sigma-Aldrich), and Triton X-100 surfactant (Sigma-Aldrich) with a mass ratio of 40:30:30:1. To redisperse the agarose droplets, 100 μL carrier oil were added and mechanically vibrated at a frequency of 17 Hz for 30 s using a TissueLyser mixer (Qiagen, Valencia, CA). Each tube contained 10 μL PCR mixture including droplets as well as 100 μL carrier oil to ensure uniform heating when fitting into the thermoblock

19

of PTC100 thermocycler (MJ Research, Waltham, MA). The thermal cycling condition was composed of initial activation of the AmpliTaq Gold DNA polymerase at 95 ºC for 10 min, followed by 10 cycles of 94 ºC for 1 min, 58 ºC for 1 min, 70 ºC for 1.5 min, 22 cycles of 90 ºC for 1 min, 58 ºC for 1 min, 70 ºC for 1.5 min, and a final extension step for 30 min at 60 ºC. The samples were then cooled to 4 ºC to enable agarose gelation.

2.7 Bead recovery

Following PCR amplification, the gelled agarose droplets were harvested by centrifuging at 250 g for 1 min. After aspirating the supernatant, the droplet pellet was sequentially washed with 100% isopropanol once, 100% ethanol once and Dulbecco's PBS (DPBS) containing 0.02% Tween 20 for 5 times to completely remove the residual oil. The microbeads carrying STR products were released from the droplets through melting the agarose at 62 ºC for 20 min to break the droplets. The beads were washed with 0.1% SDS by centrifuging at 400 g for 1 min to facilitate removal of BSA, followed by washed in DPBS for 8 times to remove the DNA fragments not linked to the beads. Finally, the beads were counted, resuspended in appropriate volume of nuclease-free water, and stored at 4 ºC for at least a week.

2.8 Secondary PCR and fragment sizing analysis

To detect the STR products immobilized on the beads, secondary PCR was carried out to transfer the STR information into free solution. The beads were diluted at appropriate concentrations in standard PCR tubes or 96-well PCR plates to serve as the DNA templates for reamplification. The secondary PCR was conducted in 12.5 µL reaction using PCR mixture that consisted of 1X Gold ST*R buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 160 µg/ml BSA, 200 µM each dNTP; Promega), 0.1 U/µL *AmpliTaq®* Gold DNA polymerase, primer mixture, nuclease-free water and bead solution. The primer mixture contained the same amounts of the fluorescent-dye labeled primers as their corresponding reverse-direction primers. The amplification protocol involved 10 min hot start at 95 ºC, then 10 cycles of 94 ºC for 30 s, ramp at the rate of 0.5 ºC/s to 58 ºC, hold for 30 s, ramp at the rate of 0.3 ºC/s to 70 ºC, hold for 45 s, followed by 15 cycles of 90 ºC for 30 s, ramp at the rate of 0.5 ºC/s to 58 ºC, hold for 30 s, ramp at the rate of 0.3 ºC/s to 70 ºC, hold for 45 s, and a final extension step at 60 ºC for 30 min.

The amplified products were processed for fragment sizing analysis on Applied Biosystems 3730XL DNA Analyzer (Life Technologies). To prepare the samples for CE analysis, 0.5 µL of

20

products in free solution were mixed with 9 µL of Hi-Di formamide (Life Technologies) and 0.5 µL of GeneScan 500 ROX size standard (Life Technologies), after the beads were settled on the bottom of PCR tubes or plates. The samples were denatured at 95 ºC for 3 min and immediately chilled on ice for 3 min just prior to loading into the instrument. The data were analyzed using Peak Scanner software.

3. Results

3.1 Microbead-based solid-phase PCR in bulk solutions

The efficiency of PCR performed on a solid phase surface is generally lower than that obtained from PCR in solution due to steric hindrance and charge repulsion.[24] In particular, when multiple targets are involved in an amplification reaction, PCR yield will be low due to competition between each locus, and the amplification process for each locus is not fully balanced. Therefore, we first validated the microbead-based solid-phase PCR for STR typing with 9947A female and 9948 male standard genomic DNA in bulk solutions (40 pg/µL or ~13 copies/µL). The multiplex-primer microbeads were incorporated into the PCR mixture with a final concentration of 200 beads/µL. BSA and Tween 80 were included to be compatible with microdroplet-based emulsion PCR.[25] DNA amplification was initiated with the presence of the fluorescent-dye labeled primers that were not conjugated on beads and 1/10 amount of corresponding reverse-direction primers in the free solution. The PCR thermal cycling conditions and primer concentrations were systematically optimized to balance the peaks. **Figure 2** demonstrates the STR profiles of both types of genomic DNA, which are in agreement with the well established locus information (as shown in **Table 2**). For instance, 9947A female DNA has only one 106-bp X-chromosome product for amelogenin, while 9948 male DNA has not only the 106-bp X-chromosome product but also a 112-bp Y-chromosome product. The peaks are well balanced under the optimized primer concentrations. No allele drop in and drop out is observed. For 9947A genomic DNA (**Figure 2a**), the heterozygous peak height ratios (the peak height of the weaker intensity allele peak over that of the stronger intensity allele peak) for D3S1358, vWA, TH01 and D7S820 are 99%, 94%, 92% and 82%. The percentage of stutter products for D3S1358-14, D5S818, vWA-17, TH01-8, D13S317, D21S11-30 and D8S1179 are 8%, 7%, 12%, 4%, 7%, 9% and 9%. For 9948 genomic DNA (**Figure 2b**), the heterozygous peak height ratios for amelogenin, D3S1358, D5S818, TH01, D21S11 and D8S1179 are 85%, 86%, 99%,

86%, 85% and 85%. The percentage of stutter products for D3S1358-15, D3S1358-17, D5S818-11, D5S818-13, vWA, D13S317, D21S11-29, D8S1179-12 and D7S820 are 10%, 10%, 8%, 11%, 12%, 7%, 9%, 8% and 7%. All stutter percentages are below 15%, so the stutters can be ignored as biological artifact of the samples. In **Figure 2a**, the inter-color balance is 40%. The intra-color balances for TAMRA, 6-FAM and JOE are 68.2%, 61.5% and 82.4%, respectively. In **Figure 2b**, the inter-color balance is 35.6%. The intra-color balances for TAMRA, 6-FAM and JOE are 57.4%, 65.2% and 85.2%, respectively.
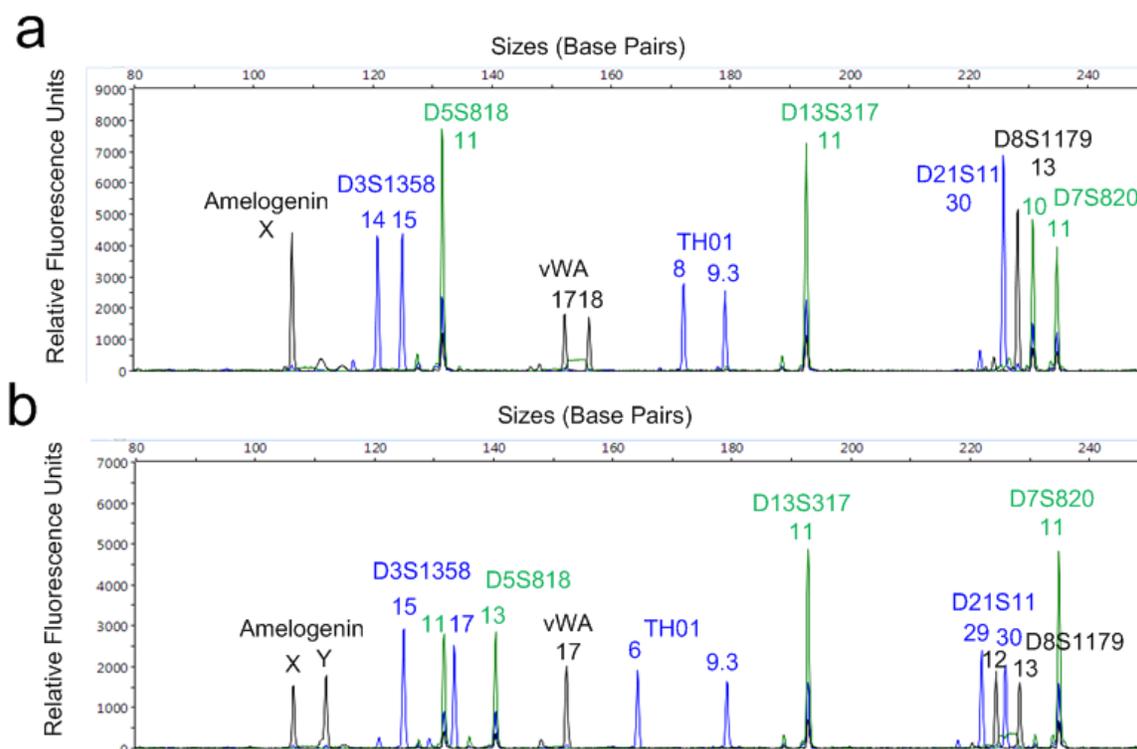

Figure 2. 9-plex STR profiles resulted from PCR amplification of 9947A female genomic DNA **(a)** and 9948 male genomic DNA **(b)** in bulk solutions containing 9-plex primer microbeads. The PCR is performed with 32 cycles from 40 pg/µL (~13 copies/µL) of the genomic DNA. The PCR products in free solution are processed by a conventional CE system for fragment sizing analysis. These traces illustrate the success in balancing the solid-phase 9-plex PCR involving primer-functionalized beads. The black, blue and green peaks are from products labeled with fluorescent dyes TAMRA, 6-FAM and JOE, respectively.

22

Table 2. Locus-specific information for 9947A female and 9948 male genomic DNA.

STR Locus	9947A female DNA		9948 male DNA	
	Repeat number	Amplicon size (bp)	Repeat number	Amplicon size (bp)
Amelogenin	X, X	106	X, Y	106, 112
vWA	17, 18	151, 155	17,17	151
D8S1179	13, 13	227	12, 13	223, 227
TH01	8, 9.3	172, 179	6, 9.3	164, 179
D3S1358	14, 15	123, 135	15, 17	127, 135
D21S11	30, 30	227	29, 30	223, 227
D5S818	11, 11	135	11, 13	135, 143
D7S820	10, 11	231, 235	11, 11	235
D13S317	11, 11	181	11, 11	181

Subsequently, the protocol of the secondary PCR was investigated by reamplifying single beads bound with multiplex STR fragments. The microbeads from the first bulk PCR reaction were washed with DPBS, counted using a hemocytometer, and statistically diluted into standard PCR tubes. By detecting the DNA in the supernatants of each wash, we found that 8 washes could completely remove the PCR products suspended in the free solution and nonspecifically bound on the beads so that the results were solely caused by amplicons linked to the beads (data not shown). The secondary PCR was accomplished using Promega Gold ST*R buffer without additional BSA and Tween 80. Unlike the first-round of PCR, the concentrations of forward and reverse primers for each STR locus were identical. 25 cycles of PCR were verified to be enough for CE detection. Totally 15 samples were tested for each kind of DNA, and two bead concentrations (0.8 and 0.15 beads/reaction) were used for beads with products from 9947A DNA to verify the statistics. The results show that in 15 samples, 7 and 2 samples were positive when bead concentration was 0.8 and 0.15 beads/reaction, respectively, corresponding well to the theoretical value of 8.25 (55%) and 2.1 (14%) predicted by the Poisson distribution. **Figure 3** presents typical data from the secondary PCR, indicating that the peaks were well balanced using optimized PCR protocols.
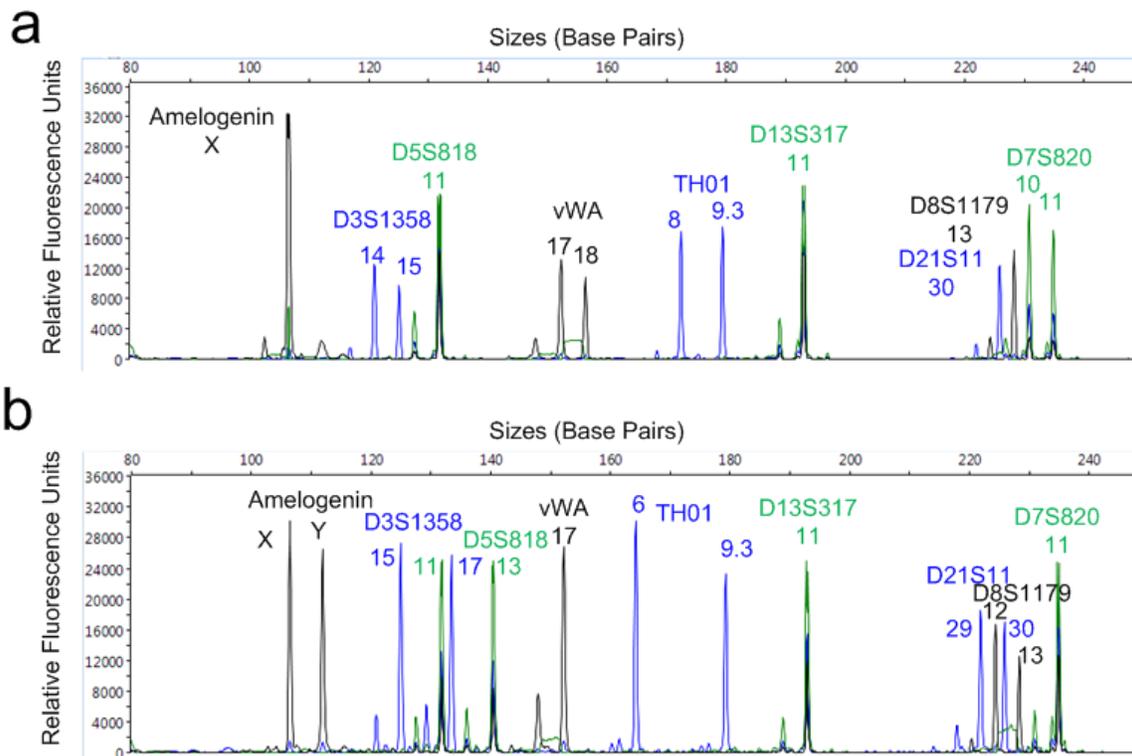
23

Figure 3. Representative 9-plex STR profiles resulted from secondary PCR amplification of single beads carrying STR products of 9947A female genomic DNA **(a)** and 9948 male genomic DNA **(b)**. The amplicon-bound beads are prepared by 32 cycles of PCR seeded with isolated DNA in bulk solution. The PCR starting from single beads is performed in bulk solution with 30 cycles. These traces illustrate the success in balancing the 9-plex PCR starting from a single-beads carrying STR products.

3.2 Agarose microdroplet-based emulsion PCR

After validating the two rounds of 9-plex PCR in bulk solutions, we optimized the experimental steps to translate the process to a microdroplet format. A microfluidic droplet generator was constructed with soft lithography technique using elastomeric PDMS due to the simple, fast, low-cost but reliable fabrication process. Surface silanization of the microchannel was performed with a fluorosilane reagent to increase the hydrophobic properties, thereby improving the stability of the water-in-oil droplets and prolonging the lifetime of the devices. **Figure 4a** shows aqueous microdroplets rapidly produced based on a flow-focusing channel geometry where the disperse (water) phase flowed in central channel and the continuous (oil) phase flowed in two outside channels. The inner phase was broken into droplets inside of a 125 μm wide nozzle at the downstream end of the channel junction owing to the hydrodynamic pressure and viscous shear

24

stresses applied by the outer phase.[27] The collected droplets were highly uniform (1.5 nL), as demonstrated in **Figure 4b**. Under optimized flow rates, agarose droplets could be generated with a frequency of approximate 444 Hz in Quantalife fluorinated oil.
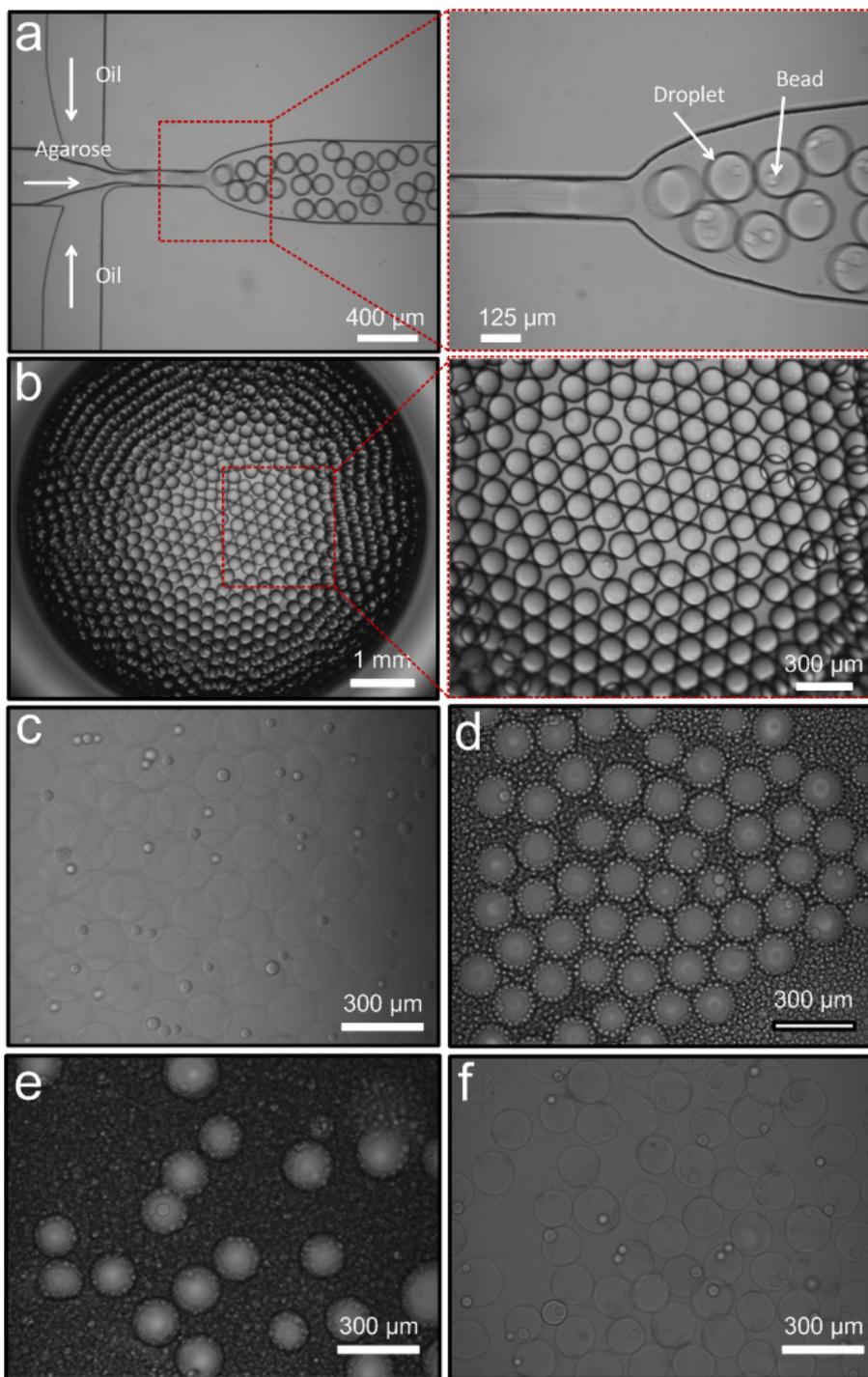

Figure 4. Agarose microdroplet generation and manipulation. **(a)** Microdroplet generation based on a flow-focusing structure of a PDMS/glass microfluidic device. Primer-functionalized beads and the desired number of cells (or

25

genomic DNA when desired) are encapsulated within the droplets. Quantalife oil and 3% agarose solution are infused into the channel using a syringe pump. **(b)** Uniform gel microdroplets suspended in Quantalife oil within a 96-well plate. **(c)** Uniform gel microdroplets suspended in water after cell lysis and washes. **(d)** Gel microdroplets dispersion in freshly prepared silicon oil mixture by mechanical vibration. **(e)** Droplets after 32 cycles of emulsion PCR in freshly prepared silicon oil mixture. No droplet merging is observed. **(f)** Droplets after emulsion PCR and washes by isopropanol, ethanol and PBS containing 0.02% (w/w) Tween 20. Bead leakage is rarely observed.

The ultra-low-gelling temperature agarose with a gelling point about 8-17 ℃ and a remelting point of around 50 ℃ was selected to facilitate the droplet generation at room temperature. In addition to biocompatibility, the most significant advantage of agarose droplets is the ability to rapidly transform into microgels by simply cooling at below gelling temperature and remain solid state unless the temperature rises above the remelting point. This feature is particularly important to various mechanical manipulation and long-term storage of millions of microdroplets simultaneously while maintaining the single-genome fidelity of each compartmentalized cell. **Figure 4c** shows the droplets isolated from the oil through filtering using a cell strainer with 40 μm nylon mesh and extensive washes with nuclease-free water. **Figure 4d** shows the droplets uniformly redispersed in oil by mechanical vibration using a Qiagen TissueLyser mixer. Mechanical properties of the gelled microdroplets are adjusted by the concentration of the polysaccharides in the feed solutions. We found that 1.5% (w/v) of agarose provides sufficient strength without comprising the PCR efficiency.

Single cells (or desired copies of DNA molecules) along with primer-functionalized microbeads were stochastically encapsulated into agarose droplets following Poisson statistics. Thus, the average number of cells (or DNA molecules) and microbeads per droplet could be easily changed by varying their concentration suspended in the feed agarose solution. Encapsulation of cells in agarose droplets allowed robust and reproducible single-cell DNA extraction in parallel. The liberation of genomic DNA and the enzymatic digestion of proteins could be accomplished by incubating the cell-containing droplets in SDS cell lysis buffer containing proteinase K at 37 ℃ overnight. This step was especially critical to PCR assays of mammalian cells, as it could remove the vast majority of histones and other nuclear proteins that inhibit polymerase activity, thereby enhancing the PCR efficiency.[25] The released high-molecular-weight DNA remained trapped in the porous network of gelled agarose. The primer-functionalized microbeads

26

coencapsulated in droplets acted as amplicon-bound substrates to maintain the monoclonality of genetic analysis by avoiding the cross-contamination of DNA products from different droplets after droplet disruption, and also facilitated downstream high-throughput manipulation and analysis.

Following cell lysis and digestion, a certain amount of droplets were incubated in PCR cocktail to enable the transport of PCR components into agarose matrix structure. BSA and Tween 80 were included in the mixture to ensure the thermostability of the droplets during PCR cycling. Small amounts of primers conjugated on microbeads were also incorporated into the mixture in order to initiate the solid-phase PCR. Uniform nanoliter reactors were produced in a standard PCR tube for massively parallel single-cell PCR using a conventional thermocycler. The solution fraction of PCR mixture was agitated into microemulsions less than 1 μm in diameter which potentially improved the stability of agarose emulsion during the amplification process.[28] After melting during the hot start phase of PCR at 95 ºC, the agarose droplets remained liquid state throughout the whole thermal cycling, enhancing the mixing rate of reagent and amplicon within the nanoreactors. The composition of the carrier oil was a critical factor to the success of emulsion PCR. As illustrated in **Figure 4e**, the droplets were thermostable in silicon oil mixture. After 30 cycles of PCR, the size and monodispersity did not change. In contrast, droplet merging was observed after PCR when using the Quatalife oil. After emulsion PCR, the droplets were gelled and washed to remove oil. There was no bead leakage out of the droplets during the washes, as shown in **Figure 4f**. Afterwards, the agrose droplets were heated at 62 ºC for 20 min to melt the agarose and release beads. The beads were then thoroughly washed for secondary PCR detection.

3.3 Single-molecule PCR in microdroplets

Single-molecule emulsion PCR was initially performed from 9947A female genomic DNA based on the optimized procedure. Two DNA concentrations were tested: 0.2 and 2 copy/droplet, while maintaining bead concentration at 0.9 beads/droplet. In the secondary PCR, single beads were diluted to a concentration of 0.15 beads/reaction in order to reduce the possibility of more than one bead in each well of a PCR plate (<1%) while making the possibility of the reactions involving beads be 14%. Twenty samples were tested. The results shown in **Figure 5** and **6**

indicate that 3 and 2 samples were positive when involving 0.2 and 2 copies/droplet of DNA, respectively, which was generally in agreement with the Poisson distribution (theoretically $20 \times 0.14 = 2.8$). The profiles of the negative samples were very clean. In the positive samples, each single bead exhibited distinct STR peak profiles, which suggested that the fragments of genomic DNA were randomly encapsulated in the nanoliter droplets and thus each bead was conjugated with amplicon from STR loci located in different chromosomes. When the DNA concentration was as low as 0.2 copies/droplet, fewer DNA fragments are encapsulated within each droplet, and thus only 2 to 3 STR loci were detected from each bead. **(Figure 5)** When the DNA concentration increased to 2 copies/droplet, 6 to 7 STR loci could be observed per reaction. **(Figure 6)**
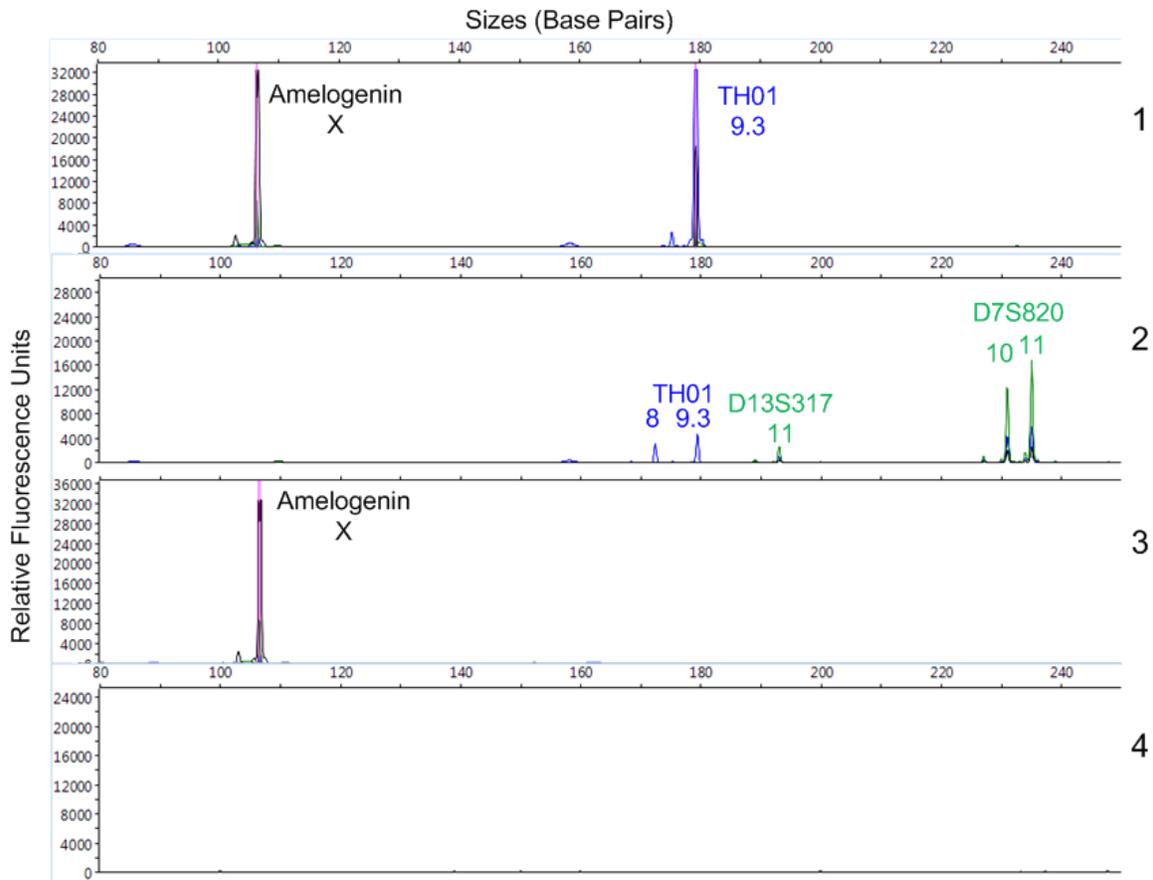

Figure 5. Single-molecule STR profiles resulted from microdroplets containing 0.2 copies of 9947A female DNA and 0.9 beads per droplet on average. 32 cycles of emulsion PCR are performed. Under the statistically dilute conditions it is expected that each STR locus will appear randomly. Totally 20 samples are tested during secondary PCR (30 cycles) based on 0.15 beads per reaction on average. There are 3 analyses that are positive (shown in 1, 2 and 3) and 17 analyses that are null results with no input beads carrying STR products, which is consistent with the theoretical value of 14% predicted by Poisson distribution. Panel 4 shows a typical null result.
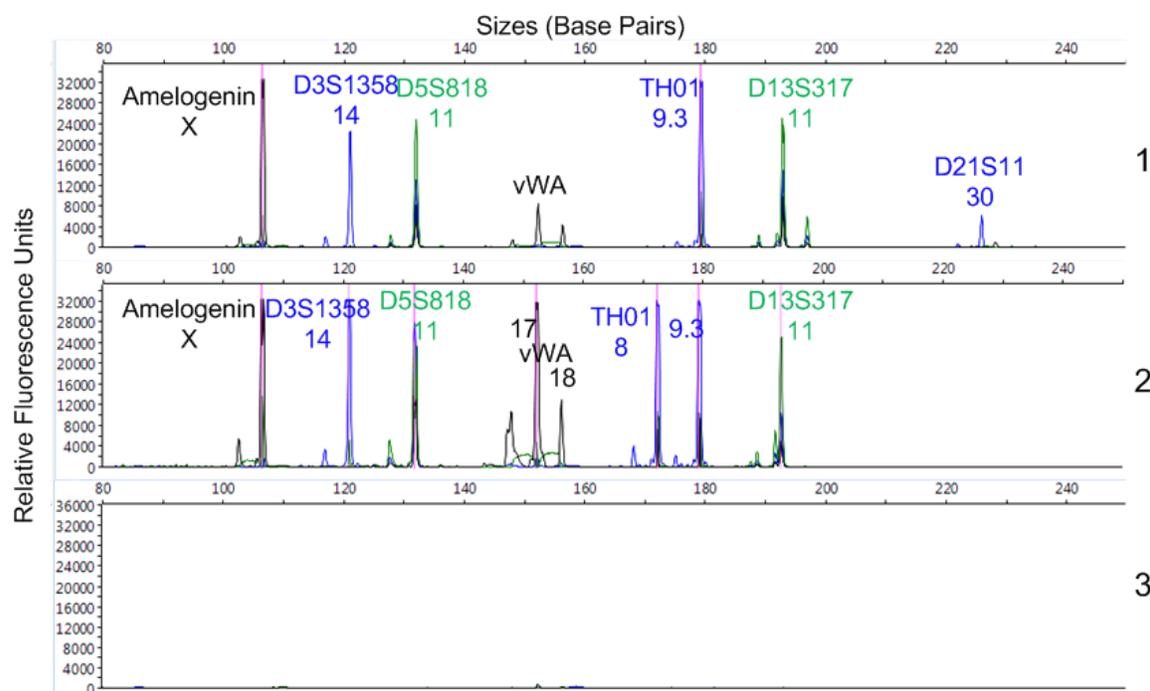
Figure 6. Single-molecule STR profiles resulted from microdroplets containing 2 copies of 9947A female DNA and 0.9 beads per droplet on average. 32 cycles of emulsion PCR are performed. Under the statistically dilute conditions it is expected that each STR locus will appear randomly and almost all beads are conjugated with STR products. Totally 20 samples are tested during secondary PCR (30 cycles) based on 0.15 beads per reaction on average. There are 2 analyses that are positive (shown in 1 and 2) and 18 analyses that are null results with no input beads, which is consistent with the theoretical value of 14% predicted by Poisson distribution. Panel 3 shows a typical null result.

3.4 Single-cell STR typing in microdroplets

Single-cell emulsion PCR was then conducted using two standard cell lines GM09947 (female) and GM09948 (male) human lymphoblast cells as models. Each agarose droplet encapsulated approximate 0.15 cells and 0.9 microbeads, which resulted in the statistical possibility that 14% of beads should be bound with STR amplicons (positive). Each positive bead was expected to contain all the products from the 9 STR loci. To detect the positive samples, 20 samples were tested in the secondary PCR with the bead concentration of 0.9 beads/reaction corresponding to 0.126 positive beads/reaction. The results demonstrated in **Figure 7** and **8** show that 1 and 2 samples were positive for GM09947 and GM09948 cells, respectively, which was consistent with the theoretical calculation ($20 \times 0.118 = 2.36$). The profiles of the negative samples were also very clean showing that background contamination and premature lysis of cells is not occuring. The STR profiles for the two cell types contained all the peaks of the 9 STR loci, and

29

the two positive samples for GM09947 cells demonstrated similar pattern, indicating the conservation of single-genome fidelity within droplets during cell lysis as well as the successful transformation of STR information from cells to microbeads.
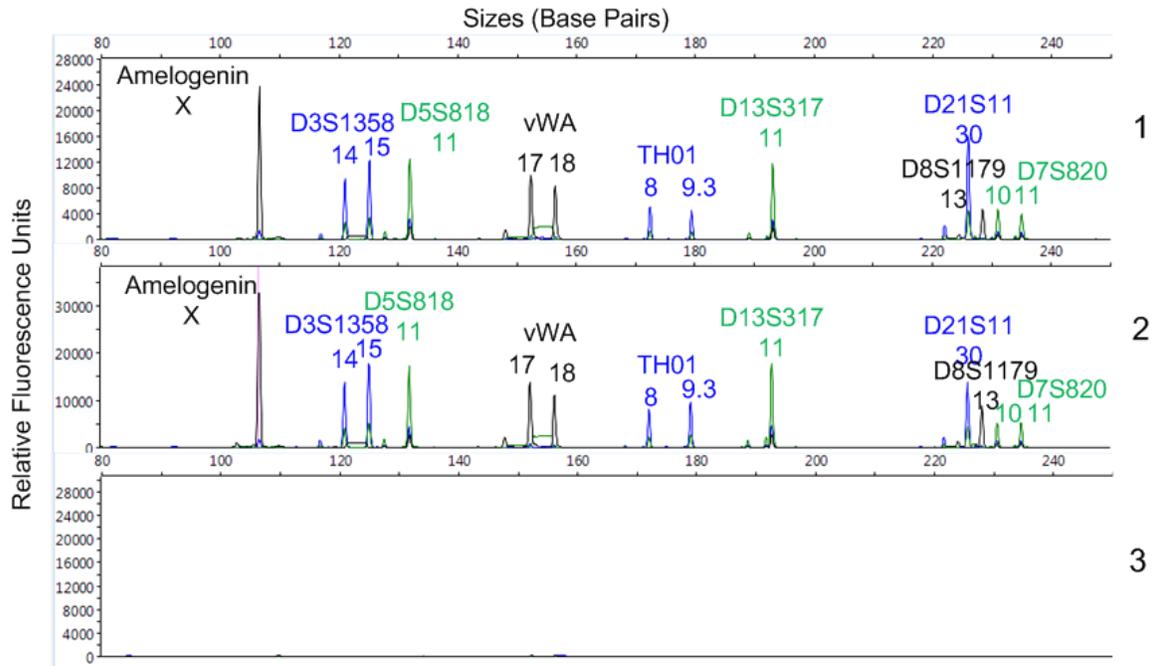

Figure 7. Single-cell STR profiles resulted from microdroplets containing 0.15 GM09947 human (female) lymphoid cells and 0.9 beads per droplet on average. 30 cycles of emulsion PCR are performed. Under the statistically dilute conditions it is expected that approximately 14% beads are bound with all 9 STR products. Totally 20 samples are tested during secondary PCR (25 cycles) based on 0.9 beads per reaction (corresponding to 0.126 positive beads per reaction) on average. There are 2 analyses that are positive (shown in 1 and 2) and 18 analyses that are null results, which is consistent with the theoretical value of 12% predicted by Poisson distribution. Panel 3 shows a typical null result.
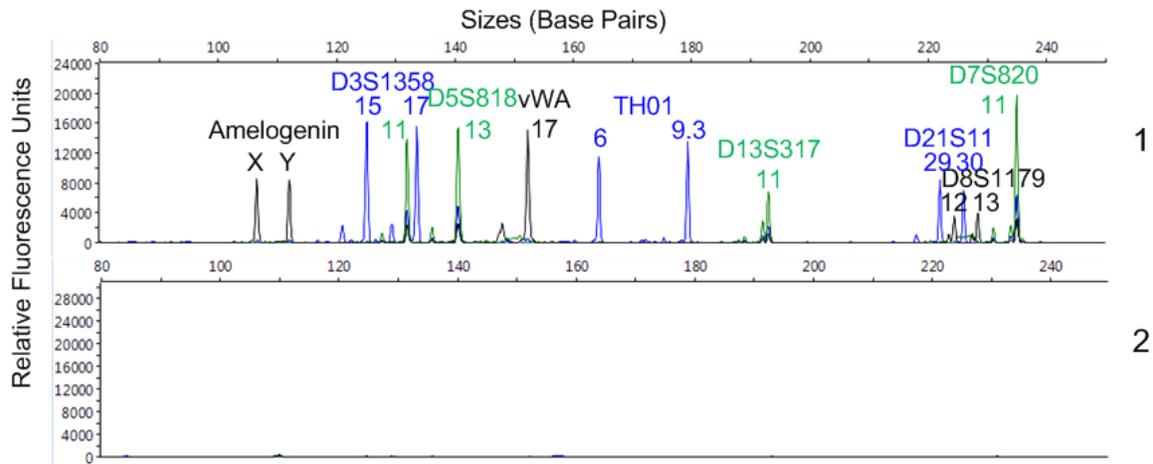

30

Figure 8. Single-cell STR profiles resulted from microdroplets containing 0.15 GM09948 human (male) lymphoid cells and 0.9 beads per droplet on average. 30 cycles of emulsion PCR is performed. Under the statistically dilute conditions it is expected that approximately 14% beads are bound with all 9 STR products. Totally 20 samples are tested during secondary PCR (25 cycles) based on 0.9 beads per reaction (corresponding to 0.126 positive beads per reaction) on average. There is 1 analysis that is positive (shown in 1) and 19 analyses that are null results, which is consistent with the theoretical value of 12% predicted by Poisson distribution. Panel 2 shows a typical null result.

The mixtures of GM09947 (female) and GM09948 (male) human lymphoblast cells with different ratios were then tested to verity the selectivity and sensitivity of single-cell STR typing. To reduce the possibility of cell aggregation in agarose feed solution, the cells were further diluted to achieve approximate 0.1 cells per droplet while the bead concentration remained unchanged (0.9 microbeads/droplet), predicting that 9.5% of the beads should be positive. The number of the secondary PCR reactions was increased to 40, and the bead concentration was increased to more than 1 (1.5 and 2) beads/reaction resulted in 0.1425 and 0.19 positive beads/reaction. No peaks appeared in the negative samples, which confirmed that negative beads did not affect the PCR amplification starting from positive beads. Therefore, the high bead concentration in secondary PCR ensured that more useful positive data could be obtained from fewer PCR reactions, thus greatly reduced the consumption of PCR reagents and the cost of the methods. As shown in **Figure 9**, when the cell ratio was 1:1, 5 samples were positive based on 1.5 bead/reaction (0.1425 positive bead/reaction), which closely agrees with the theoretically predicted Poisson distribution of 13%. Among the 5 positive samples, we could detect the complete STR profiles from a single GM09948 cell and a GM09947 cell. A mixed STR profile containing all the peaks from the two cell types was also detected, which was likely caused by cell sticking. Additional low-level peaks in panel 2 were likely stutter peaks. Allelic drop-out of a TH01 peak was observed in the profile from a GM09947 cell, possibly due to preferential amplification in the individual cell or locus heterogeneity in this cell line. Similarly, PCR failure of the TH01 allele was also observed in a mixed profile, though the heights of the peaks specific for GM09948 cells were much lower. These partial profiles still provide useful information to exclude potential contributors. As shown in **Figure 10**, when the GM09947 (female) to GM09948 (male) cell ratio increased to 2:1, 8 samples were positive based on 2 bead/reaction (0.19 positive bead/reaction), which generally consistent with the theoretical prediction of 17%. Among the 8 positive samples, 3 single GM09947 cells, 1 single GM09948 cell and 4 mixed

31

cells were detected. The heights of the peaks specific for GM09948 cells were much lower in 1 mixed profile. Although acceptable profiles were obtained, more cells should be further tested to validate the method at different dilution ratios.
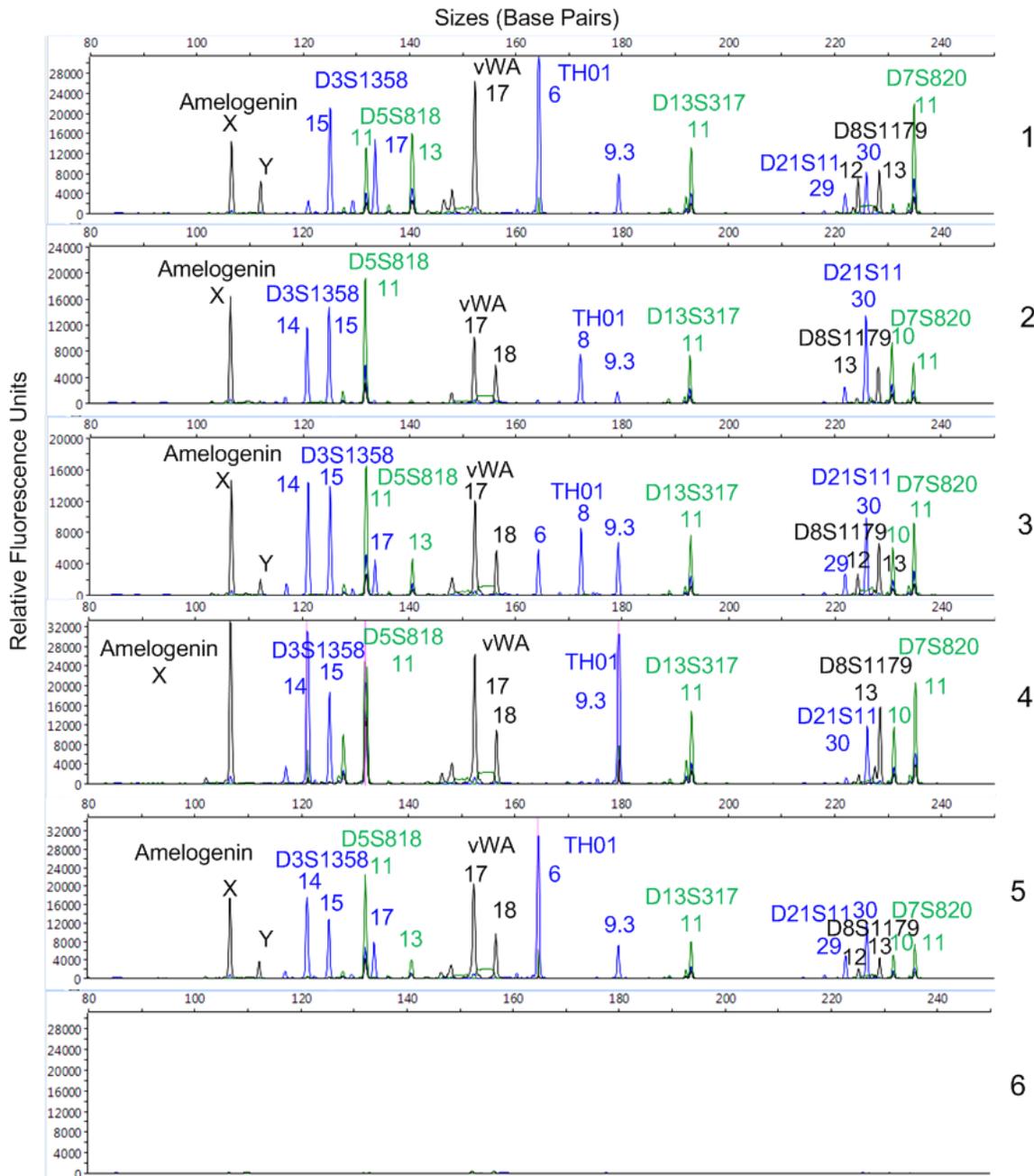

Figure 9. Single-cell STR profiles resulted from microdroplets containing 0.1 GM09947 (female)/GM09948(male) human lymphoid cell mixture (1:1) and 0.9 beads per droplet on average. 30 cycles of emulsion PCR are performed. Under the statistically dilute conditions it is expected that approximately 9.5% beads are bound with all 9 STR products. Totally 40 samples are tested during secondary PCR (25 cycles) based on 1.5 beads per reaction

32

(corresponding to 0.1425 positive beads per reaction) on average. There are 5 analyses that are positive (shown in 1 to 5) and 35 analyses that are null results, which is consistent with the theoretical value of 13% predicted by Poisson distribution. Panel 1 demonstrates the STR profile of a GM09948 cell. Panel 2 and 4 demonstrate the STR profile of GM09947 cells. Panel 3 and 5 contain all the peaks of both cell types. Panel 6 shows a typical null result.
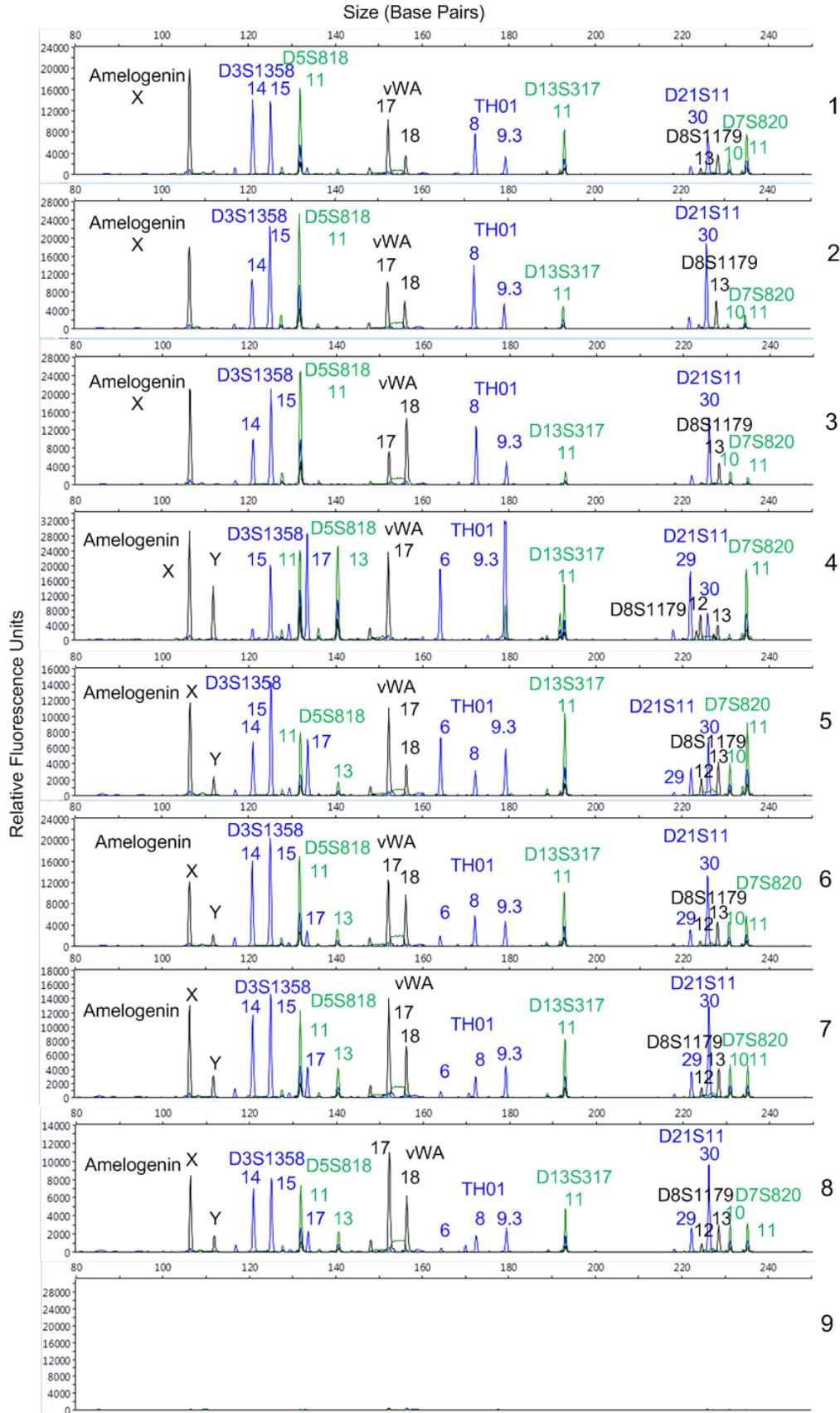
Figure 10. Single-cell STR profiles resulted from microdroplets containing 0.1 GM09947 (female)/GM09948(male) human lymphoid cell mixture (2:1) and 0.9 beads per droplet on average. 30 cycles of emulsion PCR are performed. Under the statistically dilute conditions it is expected that approximately 9.5% beads are bound with all 9 STR products. Totally 40 samples are tested during secondary PCR (25 cycles) based on 2 beads per reaction (corresponding to 0.19 positive beads per reaction) on average. There are 8 analyses that are positive (shown in 1 to 8) and 32 analyses that are null results, which is consistent with the theoretical value of 17% predicted by Poisson distribution. Panels1, 2 and 3 demonstrate the STR profile of GM09947 cells. Panel 4 demonstrates the STR profile of a GM09948 cell. Panel 5, 6, 7 and 8 contain all the peaks of both cell types. Panel 9 shows a typical null result.

4. Conclusions

We have demonstrated a droplet microfluidics-based method to type single cells with high throughput and high sensitivity. The expected profiles of nine STR loci could be successfully detected from pure and mixed single cells from standard GM09947 (female) and GM09948 (male) human lymphoid cells with high single-genome integrity. Taking advantage of the uniform nanoliter agarose droplets, the sensitivity, resolution, reliability, robustness and speed was significantly enhanced for STR analysis of single cells. A large number of single-cell PCR reactions can be simultaneously performed in a PCR tube, greatly increasing the throughput of the assay. This high throughput is necessary to produce a statistically valid result when performing single cell studies. Compared with conventional differential extraction method, our method is not limited to the separation of female epithelial cells and male sperms involved in sexual assault samples, and the efficiency of cell separation is substantially enhanced by dramatically diluting the cell suspension. Although the actual analysis time is not significantly reduced, the result interpretation becomes much easier due to the cleaner separation of mixtures. These improvements will lead to more accurate and faster results at crime laboratories in cases of evidence samples containing low amounts of cells or mixed cells such as small samples left on surfaces and samples collected after sexual assault crimes.

To further improve the accuracy and reliability of our method, lower cell concentrations will be tested to avoid mixed STR profiles generated from cell aggregates, and more realistic cell types will be studied to explore the ability to extend our methods to real world samples. More STR loci will then be included in PCR to meet the requirements of current forensic labs, though the primer concentrations will need to be further optimized to achieve balanced results. Ultimately samples

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

obtained from forensic casework such as samples from oral swab will be processed and analyzed. We will also explore the extent to which lysed non-cellular DNA in the input sample might contaminate our single cell typing results. The PCR protocol will be further optimized because PCR inhibitors and contaminants (e.g. environmental elements or natural contaminants) exist in the samples usually reduce the amplification efficiency. However, this problem is dramatically reduced in impact with our approach because the statistical dilution of the target cells also dilutes the contaminants and random DNA templates so they are less likely to interfere with each individual microdroplet PCR reaction. In a bulk reaction any contaminant effects the entire PCR reaction while with microdroplets the contaminant only impacts the microdroplet that it ends up in. This is a big advantage for droplet based single cell methods. We believe that the single cell typing methods presented here will have a major impact on the ability of labs to type low concentration, mixed and touch evidence in the future.

5. References

1. Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T. & Chakraborty, R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* 55, 175-189 (1994).

2. Andersen, J.F.*, et al.* Further validation of a multiplex STR system for use in routine forensic identity testing. *Forensic Sci. Int.* 78, 47-64 (1996).

3. Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y. & Budowle, B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 20, 1682-1696 (1999).

4. Butler, J.M. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J. Forensic Sci.* 51, 253-265 (2006).

5. Jobling, M.A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* 5, 739-751 (2004).

6. Vallone, P.M., Hill, C.R. & Butler, J.M. Demonstration of rapid multiplex PCR amplification involving 16 genetic loci. *Forensic Sci. Int. Genet.* 3, 42-45 (2008).

7. Gill, P., Jeffreys, A.J. & Werrett, D.J. Forensic application of DNA 'fingerprints'. *Nature* 318, 577-579 (1985).

8. Wiegand, P., Schurenkamp, M. & Schutte, U. DNA extraction from mixtures of body fluid using mild preferential lysis. *Int. J. Legal Med.* 104, 359-360 (1992).

9. Yoshida, K.*, et al.* The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. *Forensic Sci. Int.* 72, 25-33 (1995).

10. Chen, J., Kobilinsky, L., Wolosin, D., Shaler, R. & Baum, H. A physical method for separating spermatozoa from epithelial cells in sexual assault evidence. *J. Forensic Sci.* 43, 114-118 (1998).

11. Schoell, W.M., Klintschar, M., Mirhashemi, R. & Pertl, B. Separation of sperm and vaginal cells with flow cytometry for DNA typing after sexual assault. *Obstet. Gynecol.* 94, 623-627 (1999).

12. Horsman, K.M.*, et al.* Separation of sperm and epithelial cells in a microfabricated device: potential application to forensic analysis of sexual assault evidence. *Anal. Chem.* 77, 742-749 (2005).

13. Bienvenue, J.M., Duncalf, N., Marchiarullo, D., Ferrance, J.P. & Landers, J.P. Microchip-based cell lysis and DNA extraction from sperm cells for application to forensic analysis. *J. Forensic Sci.* 51, 266-273 (2006).

14. Norris, J.V.*, et al.* Acoustic differential extraction for forensic analysis of sexual assault evidence. *Anal. Chem.* 81, 6089-6095 (2009).

15. Sanders, C.T., Sanchez, N., Ballantyne, J. & Peterson, D.A. Laser microdissection separation of pure spermatozoa from epithelial cells for short tandem repeat analysis. *J. Forensic Sci.* 51, 748-757 (2006).

16. Li, C.X., et al. DNA profiling of spermatozoa by laser capture microdissection and low volume-PCR. *PLoS ONE* 6, e22316 (2011).

17. Bruck, S., et al. Single cells for forensic DNA analysis--from evidence material to test tube. *J. Forensic Sci.* 56, 176-180 (2011).

18. Pereira, J., Neves, R., Forat, S., Huckenbeck, W. & Olek, K. MtDNA typing of single-sperm cells isolated by micromanipulation. *Forensic Sci. Int. Genet.* 6, 228-235 (2012).

19. Song, H., Chen, D.L. & Ismagilov, R.F. Reactions in droplets in microfluidic channels. *Angew. Chem. Int. Ed.* 45, 7336-7356 (2006).

20. Vyawahare, S., Griffiths, A.D. & Merten, C.A. Miniaturization and parallelization of biological and chemical assays in microfluidic devices. *Chem. Biol.* 17, 1052-1065 (2010).

21. Theberge, A.B., et al. Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology. *Angew. Chem. Int. Ed.* 49, 5846-5868 (2010).

22. Kintses, B., van Vliet, L.D., Devenish, S.R. & Hollfelder, F. Microfluidic droplets: new integrated workflows for biological experiments. *Curr. Opin. Chem. Biol.* 14, 548-555 (2010).

23. Kumaresan, P., Yang, C.J., Cronier, S.A., Blazej, R.G. & Mathies, R.A. High-throughput single copy DNA amplification and cell analysis in engineered nanoliter droplets. *Anal. Chem.* 80, 3522-3529 (2008).

24. Zeng, Y., Novak, R., Shuga, J., Smith, M.T. & Mathies, R.A. High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Anal. Chem.* 82, 3183-3190 (2010).

25. Novak, R., et al. Single-cell multiplex gene detection and sequencing with microfluidically generated agarose emulsions. *Angew. Chem. Int. Ed.* 50, 390-395 (2011).

26. Geng, T., Novak, R. & Mathies, R.A. Single-cell forensic short tandem repeat typing within microfluidic droplets. *Anal. Chem.* 86, 703-712 (2014).

27. Baroud, C.N., Gallaire, F. & Dangla, R. Dynamics of microfluidic droplets. *Lab Chip* 10, 2032-2045 (2010).

28. Margulies, M., et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380 (2005).

6. Dissemination of Research Findings

We have published a manuscript entitled "Single-cell forensic short tandem repeat typing in microfluidic droplets" in *Analytical Chemistry* based on the work (reference 26). The results were also presented in the Frontiers of Single Cell Analysis Conference held at Stanford, CA, on September 5-7, 2013, and the AAFS 66th Annual Scientific Meeting held at Seattle, WA, on February 17-22, 2014. We are also in discussions about transferring this technology to our collaborators including the Virginia Department of Forensics, the Palm Beach Sheriff's Office, and Richard Satcher at RTI International, NC.