Fluorogenic DNA sequencing in PDMS microreactors

Peter A Sims[1,2], William J Greenleaf[1,2], Haifeng Duan[1] & X Sunney Xie[1]

We developed a multiplex sequencing-by-synthesis method combining terminal phosphate–labeled fluorogenic nucleotides (TPLFNs) and resealable polydimethylsiloxane (PDMS) microreactors. In the presence of phosphatase, primer extension by DNA polymerase using nonfluorescent TPLFNs generates fluorophores, which are confined in the microreactors and detected. We immobilized primed DNA templates in the microreactors, then sequentially introduced one of the four identically labeled TPLFNs, sealed the microreactors and recorded a fluorescence image after template-directed primer extension. With cycle times of <10 min, we demonstrate 30 base reads with ~99% raw accuracy. Our 'fluorogenic pyrosequencing' offers benefits of pyrosequencing, such as rapid turnaround, one-color detection and generation of native DNA, along with high detection sensitivity and simplicity of parallelization because simultaneous real-time monitoring of all microreactors is not required.

High-throughput sequencing methods are revolutionizing biology and promise to substantially impact the future of medicine. However, reductions in cost and improvements in turnaround time and sample preparation will be necessary for sequencing to realize its scientific and diagnostic promise. Commercially available clonal sequencing-by-synthesis technologies can be divided into two major classes. First, in pyrosequencing schemes, typified by the 454 Genome Sequencer (Roche)[1], native nucleotide species are serially introduced to sequencing reactions, and a transient bioluminescence signal is detected upon nucleotide incorporation[1,2]. The semiconductor sequencing scheme implemented in the Ion Torrent Personal Genome Machine (Life Technologies) also uses serial introduction of native nucleotides but avoids the use of an enzymatic cascade by detecting released hydrogen ions electrochemically[3]. Second, in the fluorescence detection sequencing scheme, used by Illumina's HiSeq 2000, DNA polymerase incorporates a fluorescently labeled, reversible terminator nucleotide, resulting in the conjugation of a different fluorophore to DNA for each nucleotide species[4]. The identity of the base is indicated by the color of the resulting fluorescence, and subsequent base incorporations require chemical removal of the fluorophore and terminating moiety. Similar chemistry is used in the HeliScope (Helicos) for single-molecule sequencing[5]. In competing fluorescence-based sequencing-by-ligation technologies, such as SOLiD (Life Technologies)[6], Polonator (Danaher)[7] and cPAL (Complete Genomics)[8], short, fluorescently labeled oligonucleotide probes are hybridized, ligated to anchor strands and imaged to identify bases in the DNA template. The workflow for sequencing by ligation also involves a distinct step to remove labeled components and the introduction of many different labeled and unlabeled oligonucleotides.

Each class of sequencing methodology has distinct advantages and disadvantages. The native nucleotides used in pyrosequencing and semiconductor sequencing allow the synthesis of native DNA and require no chemical removal step after incorporation to continue the sequencing process. Therefore, sequencing can progress as fast as the nucleotides can be introduced, incorporated, detected and replaced, leading to relatively rapid turnaround times on the order of hours[1]. In addition, longer read lengths have been achieved with pyrosequencing[1]. But the detection of a transient luminescence or electrochemical signal requires constant monitoring of every clonal population, which has thus far severely limited throughput. Also, luminescence and electrochemical detection schemes used in pyrosequencing and semiconductor sequencing, respectively, are less sensitive when compared to fluorescence detection.

Fluorescence-based schemes achieve massive throughput per run because the fluorescence signal of an incorporated base-labeled nucleotide can be probed at any time, allowing for a scanning-based readout of an area not limited by detector size[4–8]. The scalability of these methods allows the generation of $>10^{11}$ bases per run[8], whereas pyrodetection sequencers are currently limited to ~10^9 bases per run[1]. Finally, the inherent sensitivity of fluorescence allows fluorescence-based sequencing methods to use $>10^3$ times fewer DNA templates than pyrodetection methods[1,3], reducing reagent volumes and costs[4–8]. However, the multiple chemical steps required in each sequencing cycle result in a more complex workflow, limiting sequencing speed and read lengths.

The single-molecule, real-time approach developed by Pacific Biosciences uses four terminal phosphate–labeled nucleotides, each of which is conjugated to a different colored fluorophore[9]. This method allows for a fluorescence-based readout along with generation of native DNA and extremely rapid turnaround times[9]. However, the transient fluorescence signal requires constant monitoring with high sampling rates and single-molecule sensitivity. The single-molecule nature of the readout, as well as the fact that binding rather

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA. [2]These authors contributed equally to this work. Correspondence should be addressed to X.S.X. (xie@chemistry.harvard.edu).

RECEIVED 24 FEBRUARY; ACCEPTED 20 MAY; PUBLISHED ONLINE 12 JUNE 2011; DOI:10.1038/NMETH.1629

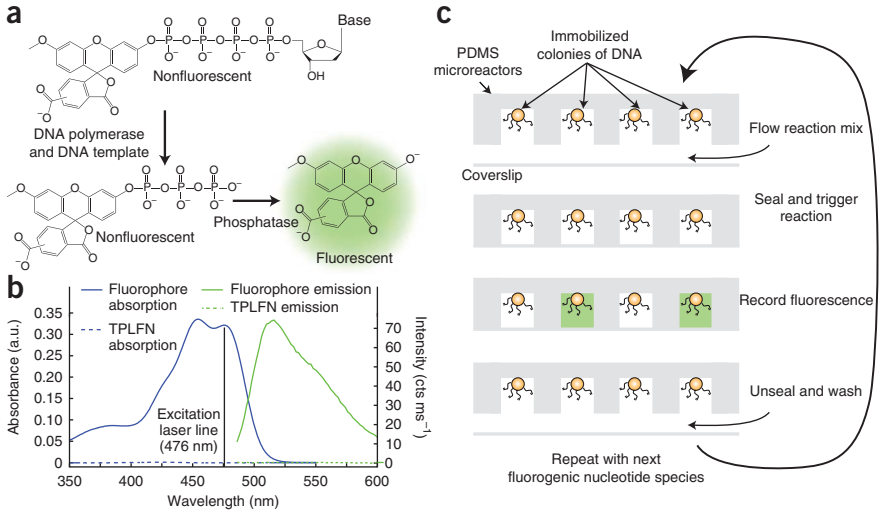

Figure 1 | Fluorogenic pyrosequencing chemistry and workflow. (**a**) A nonfluorescent TPLFN is incorporated by DNA polymerase into a DNA primer-template, releasing a labeled, nonfluorescent polyphosphate, which is digested by phosphatase to generate a fluorescent product. (**b**) Absorption and fluorescence emission spectra of the TPLFN and fluorescent product. (**c**) DNA is immobilized on beads in PDMS microreactors, which are loaded with a reaction mixture containing DNA polymerase, phosphatase and one of four TPLFNs at low temperature, sealed and heated to trigger primer extension. Fluorophores are generated in microreactors that contain DNA templates in which the base adjacent to the primer is complementary to the introduced TPLFN. The array is imaged with a fluorescence microscope, unsealed and washed before the cycle is repeated.

than incorporation is observed, leads to higher single-read error rates compared to clonal methods[9].

Here we describe fluorogenic pyrosequencing, an approach that uses terminal phosphate–labeled fluorogenic nucleotides (TPLFNs)[10–12] to allow for a pyrosequencing workflow with a non-transient, fluorescence-based signal, combining benefits of both classes of clonal sequencers. Fluorogenic pyrosequencing combines scalability with fast cycle times and, unlike commercially available fluorescence-based sequencers, uses a simple, one-color wide-field imaging system. The microreactor flow-cell platform is ideal for low reagent consumption and can be readily integrated into conventional microfluidic devices for sample preparation.

RESULTS

Fluorogenic pyrosequencing workflow

In fluorogenic pyrosequencing, each nucleotide species is labeled at the terminal phosphate position with an identical dye moiety (**Fig. 1a**), forming a nonfluorescent substrate (**Fig. 1b**). Upon incorporation of a TPLFN into an immobilized DNA template by DNA polymerase, native DNA is generated and a labeled polyphosphate molecule is released. This polyphosphate chain is rapidly digested by phosphatase, yielding a fluorophore (**Fig. 1a**,**b**) that is trapped in resealable PDMS microreactors and detected by a charge-coupled device (CCD) camera. This process is repeated for all four TPLFNs, each of which has the same label.

Upon incorporation of a TPLFN, the detectable species released by DNA polymerase diffuses away, destroying the spatial correspondence between base incorporation and the presence of fluorophores. To solve this problem, DNA templates are spatially arranged in individual, sealable microreactors fabricated using soft lithography in PDMS (**Fig. 1c**). These reactors can be sealed reversibly against a coverglass, isolating individual sequencing reactions and trapping the generated fluorophores[13]. The microreactors, each of which contains ~5,000 copies of a primed DNA template, are cooled to a temperature where DNA polymerase is relatively inactive and loaded with DNA polymerase, phosphatase and one of the four species of TPLFNs. TPLFNs are excellent substrates for DNA polymerase[14], and because TPLFNs are nonfluorescent (**Fig. 1b**), they can be present in the sample at relatively high concentrations without concern for fluorescence background. The microreactors are then sealed, and the temperature of the sample is raised to allow nucleotide incorporation. If the introduced TPLFN is complementary to the DNA template base adjacent to the primer, fluorophores will be generated and trapped in the microreactor (**Fig. 1c**). In homopolymeric regions with multiple, identical bases in the DNA template adjacent to the primer, primer extension will continue until the identity of the template base changes. Hence, the number of fluorophores generated should be proportional to homopolymer length. After the nucleotide incorporation reaction has gone to completion, the microreactor array is imaged with a one-color fluorescence microscope, unsealed and washed. This cycle can be automated and repeated in a large array to sequence multiple DNA templates in parallel. Each sequencing cycle results in the generation of physically trapped fluorophores in the microreactors, which can be detected with high sensitivity at any time, obviating the need for real-time monitoring.

Fluorogenic pyrosequencing chemistry

Terminal phosphate–labeled nucleotides synthesis[15–17] is simple compared to that of many other nucleotide analogs, and the same conjugation strategy can be applied to all four nucleotides[10,11,14]. Although TPLFNs, in which the label is attached to the γ phosphate, are rather poor substrates for DNA polymerase[10,11], extension of the phosphate chain to increase the distance between the labeling moiety and the polymerase active site and recover the native charge of a nucleotide triphosphate has led to considerable improvements[10,11]. To test fluorogenic pyrosequencing, we synthesized four δ-labeled nucleotide tetraphosphates (**Fig. 1a**) that are incorporated by *Bst* DNA polymerase. We conjugated all four nucleotides to 3′-*O*-methyl-5(6)-carboxyfluorescein, which is a fluorogenic dye excitable at 476 nm with extremely high fluorogenic contrast (**Fig. 1b**) that does not diffuse appreciably into PDMS under the conditions of our experiment. It is easily synthesized from the commercially available fluorophore 5(6)-carboxyfluorescein and can be phosphorylated with standard phosphoramidite chemistry[18], allowing conjugation to all four dNTPs (**Supplementary Note 1** and **Supplementary Figs. 1** and **2**) to generate TPLFNs (**Fig. 1a**).

We used *Bst* DNA polymerase (large fragment) because it incorporates TPLFNs efficiently, has good strand-displacement activity,

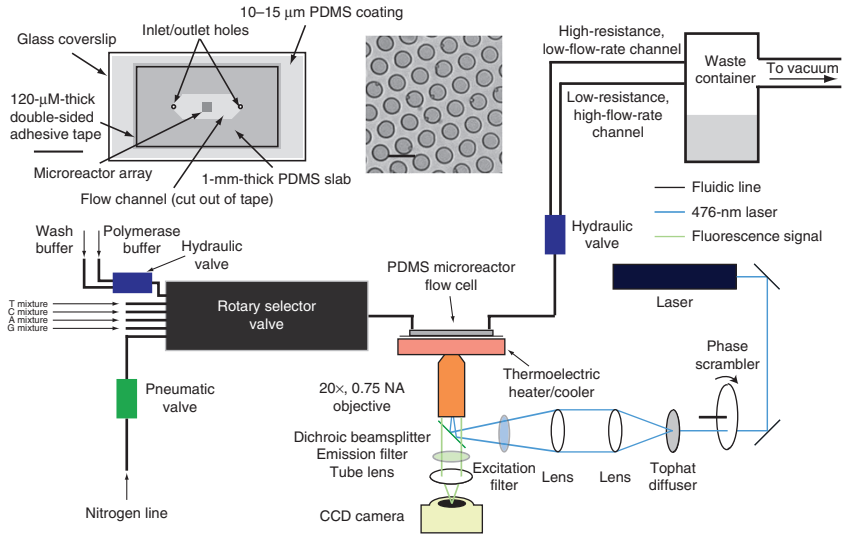

Figure 2 | Schematic of the fluorogenic pyrosequencing system. The sequencer comprises three modules: imaging, fluidics and temperature control. The imaging module is a one-color epifluorescence microscope composed of a laser, halogen lamp, CCD camera, microscope frame, filter set and a 20× air objective with some additional optics (phase scrambler and tophat diffuser) to flatten the field of view. The lamp provides bright-field transmission images of the microreactor array that are used for maintaining the proper focal plane. A rotary selector valve, hydraulic and pneumatic valves, a nitrogen line and a vacuum line comprise the fluidics module, which controls sealing, reagent flow and washing. The vacuum line is used to both deliver reagents to the sample and also seal the PDMS microreactor array against the lower surface of the flow cell. The thermoelectric temperature controller cools the microreactor array while reagents are loaded into the reactors to avoid substantial nucleotide incorporation before sealing. This module then heats the array after sealing to allow rapid primer extension. A schematic of a PDMS microreactor flow cell and a bright-field image of the PDMS microreactors is shown at top middle. The microreactors are ~5 μm in diameter, 3.8 μm in height with center-to-center distance of 7.5 μm. Some of the microreactors contain 1-μm polystyrene beads. Scale bars, 1 cm (left) and 7.5 μm (right).

a very temperature-sensitive incorporation rate (high above 50 °C but low below 5 °C) and no exonuclease activity. We introduced calf intestinal alkaphosphatase along with the polymerase to rapidly digest the released 3′-*O*-methyl-5(6)-carboxyfluorescein triphosphate, generating fluorescent product. After the microreactors were sealed, we triggered primer extension by raising the temperature to 55 °C, driving TPLFN incorporation and release of fluorescent 3′-*O*-methyl-5(6)-carboxyfluorescein to completion within 2 min.

PDMS microreactor flow cell

Fluorogenic pyrosequencing occurs in a flow cell that facilitates rapid fluid exchange while minimizing sample volume[1,3–8]. The instrumentation required for fluorogenic pyrosequencing consists of three modules: imaging, fluidics and temperature control (**Fig. 2** and Online Methods). We constructed flow cells from PDMS[19] (Online Methods). We generated arrays of micrometer-sized pillars on a silicon wafer using conventional photolithography and used this silicon master to generate PDMS microreactor arrays with soft lithography[13,19]. The regular array format has the potential for greatly simplified data acquisition and analysis while providing a straightforward means of increasing sample density and throughput[8].

PDMS is largely transparent to visible light and has low autofluorescence relative to other polymers used in microfluidics[20], making it compatible with fluorescence microscopy. Although PDMS is largely inert, it can be treated with oxygen plasma to form reactive groups for functionalization[19]. This feature allows biotinylation[21] of PDMS microreactors, facilitating capture of streptavidin-coated beads to which DNA templates are bound (Online Methods). PDMS is also a flexible elastomer that can seal reversibly to flat surfaces[13]. Similar platforms involving silica microreactors sealed to a silicone gasket have also been applied to single-molecule enzymology[22]. We constructed PDMS microreactor flow cells consisting of a microreactor array–containing slab, a spacer and a glass coverslip coated with a thin layer of PDMS (**Fig. 2**). Sealing was accomplished by applying vacuum to the outlet of the flow cell, causing the elastomeric PDMS slab that forms the top of the flow cell to deform and seal to the PDMS coating on the glass surface (Online Methods and **Supplementary Fig. 3**). This process is repeated in every sequencing cycle to facilitate microreactor loading, fluorescent product trapping and washing.

Fluorogenic pyrosequencing performance

A subset of the fluorescence images we obtained during a 60-cycle fluorogenic pyrosequencing run are shown in **Figure 3a**; each

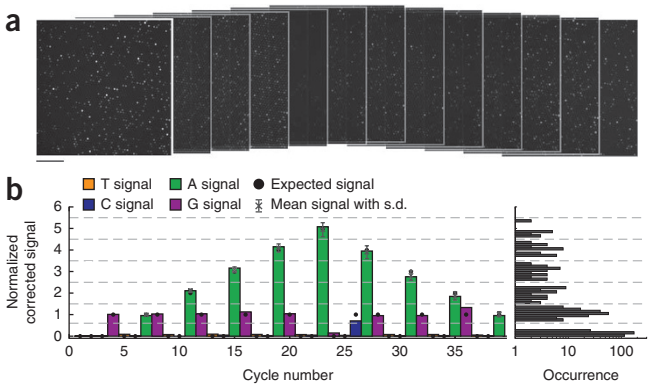

Figure 3 | Fluorogenic pyrosequencing images and homopolymer analysis. (**a**) Sequential fluorescence images of a PDMS microreactor array obtained during the first 13 cycles of a 60-cycle fluorogenic pyrosequencing run. Signal heterogeneity between reactors is due to either the presence of multiple beads or a homopolymeric region of the DNA template, which leads to the generation of more than one fluorophore per DNA copy. Scale bar, 74 μm. (**b**) Representative sequence of homopolymeric DNA. The colored bars represent a corrected (Online Methods) sequencing trace from a single microreactor containing the HL sequence. Dashed lines mark the intensity ranges used to determine homopolymer lengths. Signal intensities expected from this template sequence (black dots) show that this sequencing trace is error-free. The average and s.d. of the intensities for all analyzed traces (n = 15) fell within the expected intensity range. Intensity levels for all sequencing cycles from the analyzed HL traces were plotted as a histogram (right).

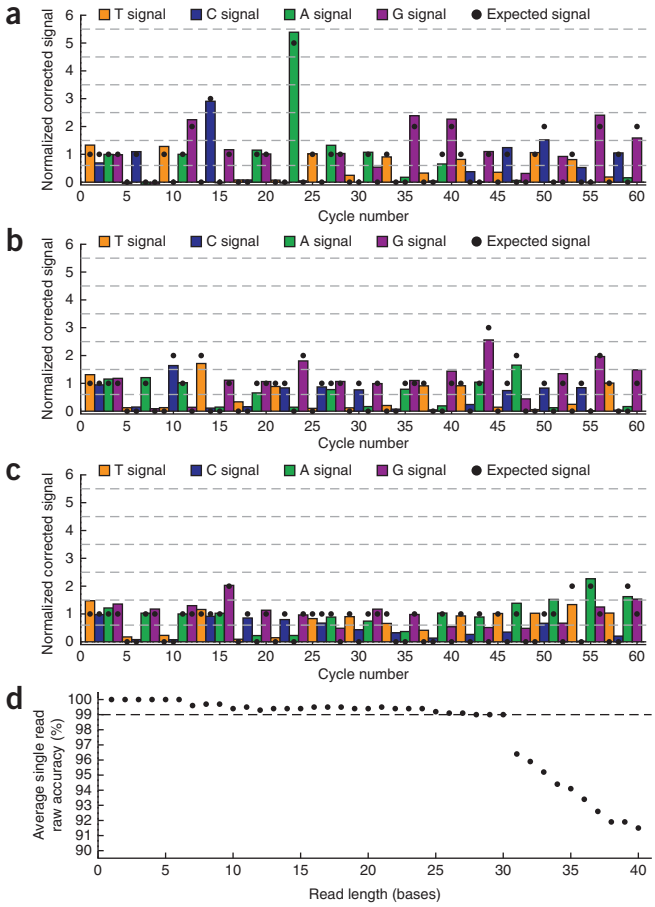

Figure 4 | Sequencing traces of quasi-random sequences and analysis of global errors. (**a**–**c**) Corrected sequencing traces from individual microreactors containing the R1 (**a**), R2 (**b**) and R3 (**c**) DNA templates. Dashed lines mark the intensity ranges used to determine homopolymer lengths, and expected signals from the template sequence are shown as black dots. (**d**) Global analysis of accuracy as a function of read length for 36 analyzed R1, R2 and R3 traces (n = 36).

image corresponds to the measured fluorescence generated after one TPLFN reaction mixture was introduced to the sample. We loaded 1-μm polystyrene beads into some of the microreactors (not visible in the fluorescence images) each with ~5,000 copies of immobilized DNA template. Each of the beads was decorated with one of four different DNA oligo test sequences (**Supplementary Table 1**). The first test sequence was a 'homopolymer ladder' (HL) oligo (**Fig. 3b**), and the other three were random polymers with very different sequences (R1, R2 and R3) (**Fig. 4** and **Supplementary Table 1**). Data presented in **Figures 3** and **4** are from the same run.

The HL oligo comprises polythymine (poly(T)) homopolymers (up to 5 bases long) interrupted by single guanine (G) or cytosine (C) bases. Unlike for reversible terminator chemistry, in which primer extension cannot proceed beyond the first base in a homopolymer within a single cycle, primer extension continues through the whole length of a homopolymer in pyrosequencing. The fidelity of fluorogenic pyrosequencing depends critically on the ability to quantitatively differentiate the amounts of signal generated at homopolymers of different lengths[1–3]. Ideally, the signal is directly proportional to homopolymer length. Homopolymer sequencing accuracy is one of the major challenges of methods with a pyrosequencing-type workflow such as fluorogenic pyrosequencing. The system achieved a single-read raw accuracy of >99%, and 80% of these HL-oligo sequencing traces were error-free. The origins of this accuracy can be visualized from the average and s.d. of the signal generated in each cycle for all HL oligo–containing microreactors analyzed from this run (**Fig. 3b**). A histogram of the signal distribution for all analyzed HL oligo–containing microreactors across all sequencing cycles shows that signals corresponding to zero, one, two, three, four and five bases are well separated (**Fig. 3b**, right).

To test this sequencing methodology on arbitrary sequences, we designed three quasi-random DNA templates with very different sequence characteristics and generated individual sequencing traces from the three oligos (**Fig. 4a**–**c**). The displayed sequencing trace from the R1 sequence was error-free to 36 bases with one error in the 42-base read. The sequencing trace from the R2 sequence was error-free for the entire 41-base read. The sequencing trace from the R3 sequence was error-free to 30 bases with six errors in the entire 40-base read. A global analysis of all analyzed traces for all three random oligos sequenced in this experiment (36 traces, 12 traces for each sequence; **Fig. 4d**) revealed that our current system maintained a mean single-read raw accuracy of at least 99% (an error rate of 1% or less) to a read length of 30 bases.

The reduction in accuracy beyond ~30 bases (**Fig. 4d**) was due to signal decay and dephasing, which occur in many clonal sequencing systems[1,4,7]. Data presented in **Figures 3** and **4** were background-subtracted and corrected for signal decay (Online Methods and **Supplementary Fig. 4**) but not corrected for dephasing.

DISCUSSION

Fluorogenic pyrosequencing has the potential for dramatically improved performance over this proof-of-principle demonstration. The current cycle times of <10 min can be reduced by increasing the speed of nucleotide incorporation and the rate of fluidic exchange (for example, by using positive pressure and optimizing flow-cell geometries). Terminal phosphate–labeled nucleotides with longer phosphate chains have been shown to impart better catalytic efficiency compared to the δ-labeled nucleotides used here, making incorporation more rapid[9–11,14]. The time required for washing the system can be reduced with enzymes that thoroughly digest excess nucleotides between cycles[1,2]. Finally, more advanced data-processing algorithms to correct for dephasing will increase the tolerance of our system to faster cycle times[1,23,24].

The observed signal decay can occur for several reasons, including DNA loss resulting from misincorporation, primer-template melting, digestion by enzymatic impurities or even DNA dissociation. Dephasing occurs as a result of either incomplete extension, in which primer extension for a subpopulation of DNA templates falls behind, or carry-forward, in which incomplete washing causes a subpopulation of DNA templates to advance ahead of the main population[1,4,23,24]. Commercial sequencing platforms use computer algorithms to model and correct for dephasing in combination with sophisticated image processing, leading to increases in read length[23,24]. We expect these improvements to our sequencing chemistry, wash-cycle efficiency and data analysis to yield increases in read length and accuracy.

Because real-time monitoring is unnecessary, DNA samples in two flow cells may be sequenced simultaneously with one sample undergoing wash and TPLFN incorporation cycles while the other is imaged. Although imaging a very large array will inevitably lead to increased cycle times, additional increases in throughput can be realized by fabrication of smaller microreactors and the implementation of high-speed fluorescence imaging schemes such as time-delay-integration line scanning. We performed fluorogenic pyrosequencing in arrays of ~2.5-μm-diameter microreactors (**Supplementary Fig. 5**), each of which contain only ~2,000 copies of DNA template. These microreactors are ~8× smaller in volume than those shown in **Figures 2** and **3**, and the arrays are also >2.3-fold denser. Higher-density arrays may be combined with more efficient sample loading methods that are not limited by Poisson statistics to increase the fraction of DNA-containing microreactors[1,8] (**Supplementary Fig. 6**). Increases in array density and loading efficiency will lead to decreases in reagent consumption per base (**Supplementary Note 2**). The off-the-shelf polymerase and phosphatase enzymes currently dominate the reagent cost. Fabrication and material costs of the PDMS microreactor flow cells are extremely low and highly scalable. Despite having similar chemistry, fluorogenic pyrosequencing has the potential for much lower reagent costs than conventional pyrosequencing because of the relatively small DNA copy number (thousands versus ~10 million for pyrosequencing)[1] and feature size required in fluorescence-based approaches. At the same time, TPLFN chemistry is considerably simpler than that used in commercial fluorescence-based platforms.

Rapid reduction in the reagent costs per sequenced base in commercial sequencers has made up-front instrumentation cost increasingly important. For sequencing systems that use an optical readout, the illumination and detection equipment, including lasers, optics and cameras, contribute substantially to instrument cost. However, fluorogenic pyrosequencing requires only one-color fluorescence detection. Furthermore, DNA colonies in fluorogenic pyrosequencing are arranged in a regular array[8] allowing more efficient use of the detector than in most commercial platforms, in which colonies are immobilized randomly[4–7]. The realization of high-speed, scanning-based fluorescence detection is substantially simplified if spectral separation is unnecessary and the sample can be spatially registered with the detector.

Sample preparation costs are also becoming more critical, particularly as diagnostic applications are contemplated. Fluorogenic pyrosequencing is, in principle, compatible with all commercially available methods for library preparation and clonal amplification. We have demonstrated the sequencing of DNA oligos immobilized on polymer beads, making clonal amplification by emulsion PCR a natural extension of our current technique[25]. However, because the number of template DNAs required for sequencing is small ($<10^4$), less efficient but simpler on-chip amplification methods such as bridge PCR[4] or solid-state rolling-circle amplification[26] are also applicable. Additionally, our PDMS microreactor platform can be readily integrated into PDMS microfluidic devices for processing complex samples, single-cell manipulation, extraction of genetic material, whole-genome amplification and targeted, on-chip amplification[27–29] (**Supplementary Note 3**). This potential for seamless microfluidic integration opens the door to sample-to-sequence analysis of raw biomaterials on a unitary microfluidic device.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by US National Institutes of Health National Human Genome Research Institute (HG005097-01) to X.S.X. and US National Institutes of Health National Human Genome Research Institute Recovery Act Grand Opportunities grant (1RC2HG005613-01) to X.S.X. W.J.G. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2000 09). We acknowledge J. Foley, S. Song, L. Song, G. Holtom and K. Fiala for valuable discussions and technical assistance. This work was performed in part at the Center for Nanoscale Systems, which is a part of Harvard University and a member of the National Nanotechnology Infrastructure Network, which is supported by the National Science Foundation award ECS-0335765.

AUTHOR CONTRIBUTIONS

P.A.S., W.J.G. and X.S.X. conceived the fluorogenic pyrosequencing concept. P.A.S. and W.J.G. constructed the sequencing apparatus. H.D. synthesized the TPLFNs. P.A.S. collected and analyzed the data. W.J.G. carried out the microfabrication. P.A.S., W.J.G., H.D. and X.S.X. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
2. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84–89 (1996).
3. Rothberg, J.M. *et al.* Methods and apparatus for measuring analytes using large-scale FET arrays. International patent WO 2010/008480 A2 (2010).
4. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
5. Harris, T.D. *et al.* Single molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
6. McKernan, K.J. *et al.* Sequence and structure variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).
7. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
8. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
9. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
10. Kumar, S. *et al.* Terminal phosphate-labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* **24**, 401–408 (2005).
11. Sood, A. *et al.* Terminal phosphate-labeled nucleotides with improved substrate properties for homogeneous nucleic acid assays. *J. Am. Chem. Soc.* **127**, 2394–2395 (2005).
12. Kozlov, M., Bergendahl, V., Burgess, R., Goldfarb, A. & Mustaev, A. Homogeneous fluorescent assay for RNA polymerase. *Anal. Biochem.* **342**, 206–213 (2005).
13. Rondelez, Y. *et al.* Microfabricated arrays of femtoliter chambers allow single molecule enzymology. *Nat. Biotechnol.* **23**, 361–365 (2005).
14. Korlach, J. *et al.* Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* **27**, 1072–1083 (2008).
15. Ankilova, V.N., Knorre, D.G., Kravchenko, V.V., Lavrik, O.I. & Nevinsky, G.A. Investigation of the phenylalanyl-tRNA synthetase modification with gamma-(p-azidoanilide-)-ATP. *FEBS Lett.* **60**, 172–175 (1975).
16. Yarbrough, L.R. Synthesis and properties of a new fluorescent analog of ATP: adenosine-5′-triphosphoro-γ-1-(5-sulfonic acid) napthylamidate. *Biochem. Biophys. Res. Commun.* **81**, 35–41 (1978).

17. Grachev, M.A. & Zaychikov, E.F. ATP gamma-anilidate: a substrate of DNA-dependent RNA-polymerase of *Escherichia coli*. *FEBS Lett.* **49**, 163–166 (1974).
18. Takakusa, H., Kikuchi, K., Urano, Y., Kojima, H. & Negano, T. A novel design method of ratiometric fluorescent probes based on fluorescence resonance energy transfer switching by spectral overlap. *Chem. Eur. J.* **9**, 1479–1485 (2003).
19. McDonald, J.C. & Whitesides, G.M. Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Acc. Chem. Res.* **35**, 491–499 (2002).
20. Piruska, A. *et al.* The autofluorescence of plastic materials and chips measured under laser irradiation. *Lab Chip* **5**, 1348–1354 (2005).
21. Liu, D., Perdue, R.K., Sun, L. & Crooks, R.M. Immobilization of DNA onto poly(dimethylsiloxane) surfaces and application to a microelectrochemical enzyme-amplified DNA hybridization assay. *Langmuir* **20**, 5905–5910 (2004).
22. Gorris, H.H., Rissin, D.M. & Walt, D.R. Stochastic inhibitor release and binding from single enzyme molecules. *Proc. Natl. Acad. Sci. USA* **104**, 17680–17685 (2007).
23. Leamon, J.H. & Rothberg, J.M. Cramming more sequencing reactions onto microreactor chips. *Chem. Rev.* **107**, 3367–3376 (2007).
24. Erlich, Y. *et al.* Alta-cyclic: a self-optimizing base caller for next-generation sequencing. *Nat. Methods* **5**, 679–682 (2008).
25. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **100**, 8817–8822 (2003).
26. Hatch, A., Sano, T., Misasi, J. & Smith, C.L. Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection. *Genet. Anal.* **15**, 35–40 (1999).
27. Marcus, J.S., Anderson, W.F. & Quake, S.R. Microfluidic single-cell mRNA isolation and analysis. *Anal. Chem.* **78**, 3084–3089 (2006).
28. Fan, H.C., Wang, J., Potanina, A. & Quake, S.R. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* **29**, 51–57 (2011).
29. Tewhey, R. *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.* **27**, 1025–1031 (2009).

ONLINE METHODS

Fabrication of PDMS microreactor arrays. Masters for soft lithography were generated from 3-inch silicon test wafers (University Wafer) coated with SU-8 2005 (MicroChem) resist to a depth of ~3.8 µm. This layer was pre-baked for 1 min at 65 °C and 2 min at 95 °C. This resist-coated wafer was then exposed through a chrome on glass mask patterned with a 2 × 2 mm array of 5 µm diameter close-packed circular transparent features spaced by 7.5 µm (center to center). Exposure was carried out in hard-contact mode (5 s pre-exposure) with an exposure energy of ~130 mJ cm^{-2} (i-line). The wafer was post-baked for 1 min at 65 °C and 1 min at 95 °C, then developed under agitation for 1 min. Features were checked with a reflected light microscope and profilometer. Then the wafer was hard-baked for 5 min at 150 °C. The wafer was then exposed to 1*H*,1*H*,2*H*,2*H*-perfluorooctyltrichlorosilane (Alfa Aesar) vapor under vacuum for ~2 h to facilitate lift-off of the PDMS. PDMS (Sylgard 184, Dow Corning) was thoroughly mixed 10:1 (base:curing agent) and degassed under house vacuum. About 9 g of degassed PDMS was poured onto the 3-inch silicon wafer master and allowed to cure for more than 2 h at ~75 °C. This slab was then gently peeled from the silicon master and used to construct PDMS microreactor flow cells. A similar protocol was followed for fabrication of the higher-density array described in **Supplementary Figure 5**.

Surface chemistry and flow cell assembly. To stably immobilize streptavidin-coated polystyrene beads conjugated to DNA templates in the PDMS microreactors, the inner walls of the PDMS microreactors were covalently functionalized with biotinylated polyethylene glycol (PEG) using the procedure described below.

A slab of PDMS (2 cm long, 1 cm wide and ~1 mm thick) containing a 2 mm × 2 mm microreactor array was cut from a PDMS chip using a razor blade. To maximize the amount of functionalization that occurs on the inner walls of the microreactors relative to the interstitial walls (that is, the part of the PDMS microreactor array that contacts the surface to which the array seals), conventional plasma oxidation of the PDMS surface was applied to a sealed device, so that air plasma will mainly contact the inner walls. The slab was sealed against a clean, glass microscope slide such that the microreactor array was facing down with air trapped in the microreactors. The PDMS slab, still sealed to the glass slide, was placed in a plasma sterilizer (Harrick) and exposed to air plasma for 1 min, generating hydroxyl groups inside the microreactors[30]. The PDMS slab was then immediately removed from the glass slide, exposed to hydrogen chloride gas for 10 s and placed face-up on a 40 °C heat block in a glass desiccator. A glass vial containing 0.2 ml of 3-mercaptopropyltrimethoxysilane (Gelest) was opened and placed near the PDMS slab on the 40 °C heat block. The contents of the desiccator were then put under vacuum for 10 min using a diaphragm pump (U-500, Buchi) to allow vapor deposition of the thiol-functionalized silane, which can react with the silanol groups generated on the microreactor walls by plasma oxidation. About 50 µl of maleimide-PEG5000-biotin (Laysan Bio) at 1 mg ml^{-1} in phosphate-buffered saline (PBS; pH 7.4; Dulbecco) was deposited onto the microreactor array, which was then placed under vacuum for about 1 min to load the aqueous solution. The maleimide group on the functionalized PEG will react with the thiols on the microreactor walls, resulting in biotinylated microreactors. After a 30-min incubation at room temperature (24 °C), the PDMS slab was rinsed with MilliQ water and dried with nitrogen gas.

A sheet of double-sided adhesive tape (0.12-mm-thick, Grace Bio-Labs) was cut using a VersaLaser into multiple 2 cm × 1 cm pieces each with a hexagonal-shaped channel cut out of its center (**Fig. 2**). One such piece was attached to the microreactor side of the PDMS slab such that the microreactor array was centered in the hexagonal channel. Two holes were punched on either end of the channel to accommodate inlet and outlet tubing.

Robust sealing is easier to achieve if the interstitial walls of the PDMS microreactor array and the flat surface against which they seal are both hydrophobic. However, the majority of the flow channel should be hydrophilic to minimize nonspecific binding of protein, enable facile loading of aqueous solutions and avoid trapping air bubbles. During plasma treatment, which is used to make the flow channel surfaces hydrophilic, the two sealing surfaces and surface chemistry was protected from oxidation. About 4 µl of 60% glycerol in water was deposited onto the microreactor array, and the PDMS slab was placed under vacuum for 1 min to load the mixture into the microreactors. Additionally, a small, circular piece of PDMS (3–4 mm diameter, 0.5–1 mm thick) was sealed to the center of a PDMS-coated glass coverslip. The PDMS slab and PDMS-coated coverslip were then treated with air plasma in a plasma sterilizer (Harrick) for 20 s. Both the PDMS microreactor array and the region of the PDMS-coated coverslip under the sealed circular piece of PDMS were protected from oxidation, but all other surfaces were oxidized and became very hydrophilic. At this point, the remaining side of the double-sided adhesive tape was used to attach the microreactor-containing PDMS slab to the PDMS-coated coverslip after removal of the circular PDMS blocking piece from the coverslip. Attaching these two components formed a PDMS flow cell with a hexagonal channel (**Fig. 2**), which was immediately filled with water.

Automated fluorogenic pyrosequencer. The automated fluorogenic pyrosequencing instrument consists of three modules: imaging, temperature control and fluidics. All three modules were controlled by a single computer using software that was home-written in LabVIEW (National Instruments) and C/C++ (compiled with LabWindows, National Instruments). A central LabVIEW VI controlled the entire workflow and displayed images and system status.

The imaging module was constructed from an inverted epi-fluorescence microscope frame (Nikon TE2000). Excitation light from an argon ion laser tuned to 476 nm (Innova 300 FRED, Coherent) was delivered to the sample through a diffuser (ThorLabs) to flatten the field of view and a 20×, 0.75 NA air objective (UPLANSAPO20x, Olympus). The objective was also used to collect the resulting fluorescence and image it onto an electron-multiplying (EM)-CCD camera (Cascade 512B, Roper Scientific). Before fluorescence image acquisition, the bright-field transmission image of the microreactor array was optimized using a correlation-based autofocus routine. The software-based autofocus program actuated a rotary stepper motor (McMaster-Carr) to control the microscope focus knob. After autofocusing, a 0.5 mm diameter region of the sample was exposed to 476-nm laser light for 0.5–5.0 s at 0.7 µW $µm^{-2}$, and a 365 µm × 365 µm field of view was imaged onto the EM-CCD camera (no EM gain, 1-MHz digitization). Laser and lamp exposures were controlled

by two shutters (Uniblitz and MAC-6000, LEP). Image acquisition was controlled by the central LabVIEW VI, which calls C/C++ functions that facilitate communication with a PCI card (PVCAM, Roper Scientific). Shutter and stepper motor control was accomplished by sending digital signals with LabVIEW through a NI-DAQ counter/timer card (PCI-6602, National Instruments).

The temperature control module was constructed from commercially available components. Four Peltier heater-cooler units (TE-31-1.4-1.15, TE Technology) were connected in series and mounted in a diamond formation on a heavy aluminum block. A thin aluminum plate with a hole in the center to allow imaging was attached to the exposed side of the Peltier units, and a small thermistor (MP-2444, TE Technology) was attached to the plate with good thermal contact. The PDMS microreactor flow cell was mounted on top of this plate, and the plate temperature could be cycled between −5 °C and 100 °C using a bipolar controller (TC-36-25-RS232, TE Technology) and power supply (PS-12-8.4, TE Technology).

The fluidics module controlled reagent flow, rinsing and sealing of the microreactor array. It consisted of a low-dead-volume rotary selector valve (MHP7970-500-4, Rheodyne), a vacuum line, a nitrogen line, and several hydraulic and pneumatic solenoid valves. Fluids were introduced to the flow cell using negative pressure. On the inlet side of the flow system, each of the four nucleotide reaction mixtures was assigned to a port on the rotary selector valve controlled by the central LabVIEW VI and digital signals from the counter/timer card. A fifth selector valve port was split into two channels using a hydraulic Y-valve (Lee Company). One channel was used to introduce wash buffer and the other to introduce the polymerase buffer. The sixth selector valve port was connected to a nitrogen line and controlled by a pneumatic solenoid valve (Action Automation and Controls). A large waste container under vacuum was connected to this outlet. The waste container input port was split into two channels by a hydraulic valve, one with high resistance for slow flow and one with low resistance for fast flow. The hydraulic and pneumatic valves were controlled by the central LabVIEW VI through a USB (universal serial bus) controller built according to the instructions on the Stanford Microfluidics Foundry website (http://www.stanford.edu/group/foundry/Building%20Your%20Own%20Valve%20Controller.html).

Microreactor array sealing was accomplished by the fluidics module in two steps that occur in rapid succession (**Supplementary Fig. 3**). Once the appropriate reaction mixture has been loaded in the microreactors, the fluidics module injects nitrogen into the flow cell at very low pressure (<1 pound per square inch), removing almost all liquid from the flow cell. The liquid that has been loaded into the microreactors remains in the microreactors, but the interstitial walls of the flow cell and the lower surface of the flow cell (the two sealing surfaces) become dry. In the second step, the nitrogen line is closed and the vacuum line is activated, causing the PDMS slab component of the flow cell to deform and bend downward toward the lower surface. This allows the two hydrophobic sealing surfaces to come in contact with each other, sealing the individual microreactors. Sealing can be reversed by simply releasing the vacuum and opening the flow cell to atmospheric pressure.

Experimental details. Four main buffers were used in fluorogenic pyrosequencing, which we define here. Polymerase incubation buffer consisted of 20 mM Tris-hydrochloride (pH 8.7), 20 mM sodium chloride, 10 mM ammonium chloride and 0.1% Tween-20. High-salt buffer consisted of 50 mM Tris-hydrochloride (pH 8.0), 1 M sodium chloride, 1 mM EDTA and 0.1% Tween-20. Wash buffer consisted of 20 mM Tris-hydrochloride (pH 8.7), 20 mM sodium chloride, 10 mM ammonium chloride, 0.1 mM EDTA and 0.1% Tween-20. Reaction buffer consisted of 20 mM Tris-hydrochloride (pH 8.7), 20 mM sodium chloride, 10 mM ammonium chloride, 1 mM manganous chloride and 0.1% Tween-20.

One-micrometer streptavidin-coated polystyrene beads (Bang's Labs) were washed three times in polymerase incubation buffer by centrifugation and resuspended to a final concentration of 30 pM in high-salt buffer. Dual-biotinylated primer was added to the beads at a final concentration of 150 nM (~5,000 primers per bead), and the mixture was incubated at room temperature for 1 h on a roller. A dual biotin-streptavidin linkage was more stable during multiple thermocycles than a single biotin-streptavidin linkage. The primerbeads were then washed twice by centrifugation in high-salt buffer and resuspended in high-salt buffer to a final concentration of 30 pM. The primerbeads were distributed into different tubes, one tube for each DNA template to be sequenced. Single-stranded DNA templates, each of which is complementary to the dual-biotinylated primer, were added to the primerbeads at a final concentration of 1 μM and incubated with the primer for at least 2 h at 4 °C on a roller. The primer-template beads were then washed four times by centrifugation in polymerase incubation buffer and resuspended in high-salt buffer to a final concentration of 15 pM.

A PDMS microreactor flow cell was constructed as described above and filled with high-salt buffer. After incubating the flow cell with high-salt buffer for 10 min at room temperature, a mixture of primer-template beads at a final concentration of 15 pM was introduced to the flow cell, which was incubated upside-down for 1 min at room temperature so that the beads would sink into the biotinylated microreactors and bind. Excess beads were then washed away with polymerase incubation buffer, and the flow cell was mounted on the sequencing apparatus by taping the coverslip to the aluminum plate component of the temperature controller.

The fluidics module was then loaded with six different solutions including the four nucleotide reaction mixtures, wash buffer and polymerase buffer. Polymerase buffer consists of polymerase incubation buffer with 200 units ml^{-1} *Bst* large fragment DNA polymerase (New England Biolabs). The four nucleotide reaction mixtures were composed of reaction buffer, 200 units ml^{-1} *Bst* large fragment DNA polymerase, 2.4 units ml^{-1} calf intestinal alkaline phosphatase (New England Biolabs), and either 3.0 μM dT4P-δ-3′-*O*-methylcarboxyfluorescein or 1.8 μM dC4P-δ-3′-*O*-methylcarboxyfluorescein or 2.5 μM dA4P-δ-3′-*O*-methylcarboxyfluorescein or 2.5 μM dG4P-δ-3′-*O*-methylcarboxyfluorescein. Calf intestinal alkaline phosphatase is active at a broad range of temperatures, does not appreciably digest TPLFNs and digests unlabeled nucleotide impurities, which might compete with TPLFNs for incorporation. Because these enzymes are available from multiple commercial sources, fluorogenic pyrosequencing requires only inexpensive, native, off-the-shelf enzymes.

The device was then connected to the fluidics module, and the imaging, temperature control and fluidics modules were tested to ensure proper autofocusing, thermocycling and sealing, and

doi:10.1038/nmeth.1629

to verify the flow rates of the different solutions. Once this initialization process was complete, the sequencing run was started. During each cycle, the flow cell was first washed for 5 min with wash buffer to remove the previous reaction mixture. The flow cell was cooled to ~1 °C during the last 30 s of the wash cycle. The flow cell was then incubated for 30 s with polymerase buffer. Next, the appropriate nucleotide reaction mixture was loaded into the microreactors, and the array was sealed (<10 s). After sealing the microreactor array, the sample temperature was increased to ~55 °C (~15 s), and the reaction mixture was incubated for 120 s. After incubation, the microreactor array was cooled to ~30 °C (<15 s) and a bright-field autofocus routine was used to focus the microscope (5–10 s). The cycle concludes with the acquisition of a fluorescence image, data transfer and unsealing of the microreactor array (~5 s). During each cycle, the flow cell was first washed for 5 min with wash buffer to remove the previous reaction mixture, cooled to ~1 °C and incubated for 30 s with polymerase buffer. Next, the appropriate nucleotide reaction mixture was loaded into the microreactors, and the array was sealed and incubated at ~55 °C for 2 min before autofocusing and imaging. This cycle was then repeated.

Data processing and base-calling analysis. Fluorogenic pyrosequencing data were processed using in-house–written programs that accept a list of coordinates of initial positions for each microreactor of interest. The program identifies the pixels associated with each microreactor in the list in each image frame, computes their integrated intensities and removes the background from each of the resulting intensity traces by subtracting the integrated intensity trace of a nearby microreactor that does not contain a bead. This results in a background-subtracted intensity (BSI) trace for each microreactor of interest.

BSI traces were processed using a very simple automated algorithm. All four test oligos have a C at the fourth template position. Microreactors for which the BSI exceeds a certain threshold at this position, indicating the presence of multiple DNA-coated beads, were eliminated from the analysis. Additionally, the R1, R2 and R3 test oligos each contain an AGTC 'key sequence' in the first four template positions. This key sequence was used to assess the performance of the fluidics module and sample quality at the beginning of a run (a similar measure is taken in the 454 pyrosequencing platform)[1]. A step-by-step sample application of the data processing algorithm is provided in **Supplementary Figure 4**. First, all BSIs were normalized by the BSI obtained in the fourth cycle of each trace because the identity of this fourth base is the same for every DNA template sequenced. The normalized traces were then thresholded to identify cycles in which the incorporation of single bases occurred (normalized intensities between ~0.5 and ~1.5 were considered single-base cycles during this first pass). Typically, the normalized BSI trace exhibited a signal decay of <1% per cycle. This signal decay was first treated by fitting the single-base calls determined from this initial normalization to a single exponential (**Supplementary Fig. 4a**). An initial correction was applied by multiplying the entire normalized BSI by the inverse of the computed exponential decay. This first corrected trace (**Supplementary Fig. 4b**) was used to reconsider which cycles actually correspond to single-base incorporations. The original, normalized BSI trace was then re-fit with a single exponential using this new set of cycles considered to be single-base incorporations (**Supplementary Fig. 4c**). A second corrected trace (**Supplementary Fig. 4d**) was then generated by multiplying the original, normalized BSI by the inverse of the second calculated exponential decay. This second corrected trace was used to determine the final, normalized intensity trace from which the final base calling was based. A set of normalized intensity thresholds was applied to the second corrected trace, and based on these preliminary base calls, an average single-base intensity was computed for the whole trace. The trace was then normalized by this value which results in the final normalized intensity trace (**Supplementary Fig. 4e**). Intensity thresholds (~0.5–1.5 for a single base, 1.5–2.5 for a two-base homopolymer, 2.5–3.5 for a three-base homopolymer and so on) were then applied to the final normalized intensity trace to generate the final base-calling trace in which all single-base calls are assigned an intensity of 1, all two-base homopolymer calls were assigned an intensity of 2 and so on. (**Supplementary Fig. 4f**). The DNA sequence that is complementary to the original template sequence can be read from this final base-calling trace. All input parameters and thresholds for data analysis were applied globally to all analyzed traces in an automated fashion.

30. Frimat, J.P. *et al.* Plasma stenciling methods for cell patterning. *Anal. Bioanal. Chem.* **395**, 601–609 (2009).