



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Analytical Biochemistry 320 (2003) 55–65

ANALYTICAL
BIOCHEMISTRY

www.elsevier.com/locate/yabio

Fluorescent in situ sequencing on polymerase colonies

Robi D. Mitra,^{a,1} Jay Shendure,^a Jerzy Olejnik,^b
Edyta-Krzyszowska-Olejnik,^b and George M. Church^{a,*}

^a *Lipper Center for Computational Genetics and Department of Genetics, Harvard Medical School,
200 Longwood Ave., Boston, MA 02115, USA*

^b *AmberGen, Inc., 1106 Commonwealth Ave., Boston, MA 02215, USA*

Received 7 January 2003

Abstract

Integration of DNA isolation, amplification, and sequencing can be achieved by the use of polymerase colonies (colonies) and cycles of fluorescent dNTP incorporation. In this paper, we present four advances that bring us closer to sequencing genomes cost-effectively using the polony technology. First, a polymerase trapping technique enables efficient nucleotide extension by DNA polymerase in a polyacrylamide matrix and eliminates loss of enzyme during sequencing cycles. Next, we present two novel types of reversibly dye-labeled nucleotide analogues, show that DNA polymerase can incorporate these analogues, and demonstrate that the dyes can be removed by thiol reduction or light exposure. Using these nucleotides, we have sequenced multiple colonies in parallel. In addition, we have found that a high density of colonies can be achieved with minimal overlap between adjacent colonies by limiting the concentration of free primer in the polony amplification reactions. Finally, we have developed software for automated image alignment and sequence calling.

© 2003 Elsevier Science (USA). All rights reserved.

Kurzweil's version of Moore's law observes that the number of calculations per second per dollar has been doubling every 2.3 years since 1900 [1]. Analogous exponential improvements have occurred in macromolecular sequencing since the 1940s [2]. Can this trend continue? Although over 100 genomes have been sequenced, we have only scratched the surface of the valuable range of sequence space. The ability to sequence routinely and cheaply cell-lineage-specific transcriptomes, the diversity of antibody responses, or microbial biomes would provide a wealth of information for biomedical research and diagnostics. Most exciting is the prospect that the cost of sequencing could be reduced to the point where the full determination of one's genome would become part of routine health care. Of course, for this to be feasible, the cost of DNA sequencing must be reduced by three to five orders of magnitude, as resequencing a single human genome is currently estimated

at more than ten million dollars [3]. Achieving this goal would have an almost immediate impact on our understanding of the genetics of common and rare diseases, comparative genomics, mRNA expression analysis, population genetics, and functional genomics.

Several groups are working to develop disruptive third-generation sequencing technologies to enable cost-effective, whole-genome resequencing. Many of these approaches involve sequencing a single DNA molecule, either by repeated cycles of single base extension [4], real-time monitoring of the incorporation of dye-labeled nucleotides by DNA polymerase [5–7], or monitoring of conductivity changes in a nanopore as a single-stranded DNA molecule is passed through it [8,9]. Other approaches first amplify the DNA template and then analyze the product by either pyrosequencing [10,11] or hybridizing DNA to a high-density oligonucleotide microarray [12,13]. These approaches all have the potential to greatly speed the rate of DNA sequencing; however, most have yet to document details of cost and accuracy.

We are adapting polymerase colony technology to enable the cost-effective sequencing of genomes and transcriptomes [14,15]. This technology enables the

* Corresponding author. Fax: 1-617-432-7266.

E-mail address: church@arep.med.harvard.edu (G.M. Church).

¹ Current address: Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110.

amplification of millions of individual DNA molecules via the polymerase chain reaction (PCR) in an acrylamide gel attached to the surface of a glass microscope slide. Because acrylamide restricts the diffusion of the PCR amplification products, each single molecule included in the reaction produces a spherical colony of DNA, which we have termed polony (short for polymerase colony). A similar technology, the molecular colony technique, grows nucleic acid colonies in acrylamide using Q beta replicase or PCR [16,17]. Here we describe our progress in developing a protocol for sequencing polonies in parallel by performing repeated cycles of primer extension with fluorescent deoxynucleotides. To enable this protocol, we developed and characterized two types of reversibly dye-labeled deoxynucleotide analogues. These analogues utilize linkers that can be cleaved by reducing agent or near-UV light. We found that DNA polymerase correctly incorporated these nucleotides and that the fluorophore could be quickly removed. We also describe a technique, polymerase trapping, which greatly increases the efficiency of extension by DNA polymerase in acrylamide gels. Finally, we demonstrate short (8-bp) sequencing reads on multiple polonies amplified from single DNA molecules and discuss the remaining hurdles to fully enabling this technology. Because tens of millions of femtoliter-scale polonies [14] can be amplified on a single microscope slide, this technology could enable extremely low-cost DNA sequencing.

Materials and methods

Oligonucleotide primers and templates

The sequences for all primers and templates used in this study can be found in the supplementary information (http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm).

Polony amplification

Polony amplification was performed as described previously [14,18], except the acrylamide gels were 40 μm thick and BSA² was added to the gel mix at a final concentration of 0.2%. To pour these gels, we used slides partially covered with a Teflon coating (Erie Scientific) which served as a spacer between the glass surface of the slide and a coverslip (Fisher Scientific). These slides were covered with mineral oil to prevent evaporation and thermally cycled (MJ Research).

Polony denaturation

After polony amplification, we removed the unattached DNA strand by incubating in 70 °C denaturing buffer (70% formamide, 1 \times SSC) and performing electrophoreses in 0.5 \times tetrabromoethane with 42% urea for 1 h at 5–10 V/cm. We washed the slides 2 \times 4 min in wash buffer 1 (10 mM Tris–HCl, pH 7.5, 50 mM KCl, 2 mM EDTA, 0.01% Triton X-100).

Determining the efficiency of nucleotide extension

For the experiments performed without using polymerase trapping, oligonucleotide templates OT4 and OT2 were annealed to primer Seq 1 and spotted on glass microscope slides. Then 1 μl of 10 μM hybridized primer:template was added to 4 μl of dH₂O and 5 μl of 2 \times acrylamide mix (25% glycerol, 1.6% TEMED, 6 ng/ml riboflavin, 7.96% acrylamide, 0.4% bis-acrylamide). Then 0.2 μl of this mixture was spotted in quadruplicate onto four glass microscope slides (Acrylate-treated glass slide, CEL associates) and the droplets were photopolymerized by exposure to white light. An extension mixture containing polymerase and unlabeled dTTP (1 \times sequenase buffer, 4 mM DTT, 100 $\mu\text{g}/\text{ml}$ BSA, 380 ng/ μl *Escherichia coli* single-stranded binding protein (United States Biochemical), 0.01% Triton X-100, and 0.5 μM dTTP) was added to the slide for 0, 90, or 360 s. Each slide was then washed with wash buffer 2 (10 mM Tris–HCl, pH 7.5, 50 mM KCl, 0.01% Triton X-100) and Cy5-dUTP was added to the slide (1 \times sequenase buffer, 4 mM DTT, 100 $\mu\text{g}/\text{ml}$ BSA, 380 ng/ μl *E. coli* single-stranded binding protein (United States Biochemical), 0.01% Triton X-100, 0.5 μM dTTP, and 0.169 U/ μl sequenase v2.0 (USB)). The slides were washed in wash buffer 1 and scanned with a GSI Lumonics confocal scanner.

For the experiments employing polymerase trapping, the primer Seq1 was annealed as described above. Then 1 μl of each annealed oligonucleotide was added to 17 μl of acrylamide gel mixture (40 mM Tris–HCl, pH 7.3, 25% glycerol, 1 mM DTT, 6% acrylamide (5% C), 17.4 U sequenase v2.0 (USB), 15 $\mu\text{g}/\text{ml}$ *E. coli* single-stranded binding protein (USB), 0.1 mg/ml BSA). Then, 1 μl of 1.66% TEMED and 1 μl of 1.66% APS were added and 0.2 μl of each mixture was pipetted onto two acrylate-treated microscope slides (CEL Associates). Extensions were performed as described above.

Extensions of polonies with unlabeled nucleotides

Template LC-1 was used in a polony amplification reaction and the slides were denatured as described above. The acrylamide gel was covered with a frame seal chamber (MJ Research) and annealing mix (0.25 μM PR1-R, 6 \times SSPE, 0.01% Triton X-100) was

² Abbreviations used: BSA, bovine serum albumin; DTT, dithiothreitol; FISSEQ, fluorescent in situ sequencing; FITC, fluorescein isothiocyanate.

added over the gel. The slides were heated at 94 °C for 2 min and then at 60 °C for 15 min. Then 15 µl of polymerase trapping mix (BST DNA polymerase 6000 U/µl, *E. coli* SSB (45 ng/µl), 14.25% acrylamide, 0.75% bis-acrylamide, 25% glycerol, 0.01% Triton X-100, 1 mM DTT, 0.1 mg/ml acetylated BSA, 0.083% TEMED, 0.083% APS) was pipetted over the gel; the slide was covered with a No. 2 18 × 18-mm coverslip and allowed to polymerize. Unlabeled nucleotide additions were performed by submerging the slides in 55 °C unlabeled extension mix (20 mM Tris-HCl, pH 8.8, 10 mM MgCl₂, 50 mM KCl, 0.5 mg/ml BSA, 0.01% Triton X-100, 2 µM appropriate deoxynucleotide triphosphate) for 3 min. After each nucleotide addition, slides were washed in wash buffer 1 for two 5-min washes and then in wash buffer 2 for 1 min. Labeled nucleotide additions were performed by attaching a frame seal chamber to the slide, pipetting 65 µl labeled extension mix (20 mM Tris-HCl, pH 8.8, 10 mM MgCl₂, 50 mM KCl, 0.5 mg/ml BSA, 0.01% Triton X-100, 2 µM FITC-dCTP, 2 µM Cy3-dUTP), and washing 2 × 5 min in wash buffer 1 and 1 × 1 min in wash buffer 2. The slides were scanned, and the fluorescence of each polony was quantified using ImageQuantNT (Molecular Dynamics).

FISSEQ

Templates T1–T5 were used to grow polonies, and the slides were denatured as described above. A frame seal chamber was attached to the slide and annealing mix (0.25 µM PR1-R, 6 × SSPE, 0.01% Triton X-100) was poured over the gel. The slides were heated at 94 °C for 2 min and then at 60 °C for 15 min. Unannealed primer was removed by washing the slides 2 × 4 min in 2 × SSPE. Then 15 µl of polymerase trapping mix (Klenow DNA polymerase 2200 U/µl, *E. coli* SSB (45 ng/µl), 14.25% acrylamide, 0.75% bis-acrylamide, 25% glycerol, 0.01% Triton X-100, 1 mM DTT, 0.1 mg/ml acetylated BSA, 0.083% TEMED, 0.083% APS) was pipetted over the gel; the slide was covered with a No. 2 18 × 30-mm coverslip and allowed to polymerize. Slides were washed in wash buffer 1 and equilibrated with Klenow extension buffer (50 mM Tris-HCl, pH 7.3, 5 mM MgCl₂, 0.01% Triton X-100). Extension reactions were performed by adding 1 µM Cy5-SS-dNTP in extension buffer to the slides for 1 min, and then washing in wash buffer 2 for 7 min. After washing, slides were scanned, and the fluorescence was removed by washing slides in bleaching buffer (500 mM β-mercaptoethanol, 100 mM Tris-HCl, 50 mM KCl, pH 9.4) for 5 min and then rinsing 7 × 30 s in wash buffer 2. The free sulfhydryls were alkylated by adding alkylation buffer (8 µg/ml iodoacetamide, 300 mM Tris-HCl, 50 mM KCl) for 5 min. After washing, the next cycle was performed.

Results

Overview of polony sequencing

Polony sequencing is performed in three steps.

Step 1: make library of linear DNA molecules with universal priming sites (Fig. 1A). Each molecule in the library will contain a variable region flanked by two constant regions. The constant regions contain primer-binding sites to allow universal amplification by PCR. This type of library was first created to perform SELEX experiments [19].

Step 2: amplify polonies in a polyacrylamide gel (Fig. 1A). A thin polyacrylamide gel is poured onto a glass microscope slide. The library made in step 1 is amplified in this gel by performing PCR. At the end of the reaction, each single template molecule gives rise to a polony. As many as 15 million polonies can be amplified on a single slide. An acrydite modification [20,21] is included at the 5' end of one of the primers, so that one strand of the amplified DNA is covalently attached to the polyacrylamide matrix. This allows further enzymatic manipulations and washing to be performed on the attached DNA. The details of polony amplification have been previously described [14].

Step 3: sequence polonies by fluorescent in situ sequential quantitation (Fig. 1B). The immobilized DNA is denatured, the unattached strand is washed away, and a universal primer is hybridized to the template. A primer extension reaction is performed [22,23]. The gel is then scanned using a scanning fluorescence microscope. If a polony has incorporated the added base, it will fluoresce, revealing the identity of the template base immediately 3' of the annealed primer. The fluorescence is then removed by cleaving the linker between the fluorophore and the nucleotide and washing away the fluorophore. The cycle is repeated by adding a different dye-labeled base, washing away unincorporated nucleotide, and scanning the gel. In this way, the sequence of every polony on the gel can be determined in parallel.

The efficiency of nucleotide extension

Because FISSEQ requires a number of sequential base extensions, the incorporation of the correct base must be highly efficient to enable accurate sequencing. For example, if only 85% of the primer:template molecules in a polony are extended each time a correct base is added, after six extensions, only (0.85)⁶, or 38%, of the primer:template molecules will have correctly incorporated the added nucleotides. The remaining 62% of the molecules will be “out of phase” because they did not incorporate the correct base at an earlier cycle. To estimate the efficiency of nucleotide incorporation by DNA polymerase, polyacrylamide spots containing immobilized oligonucleotide templates were polymerized

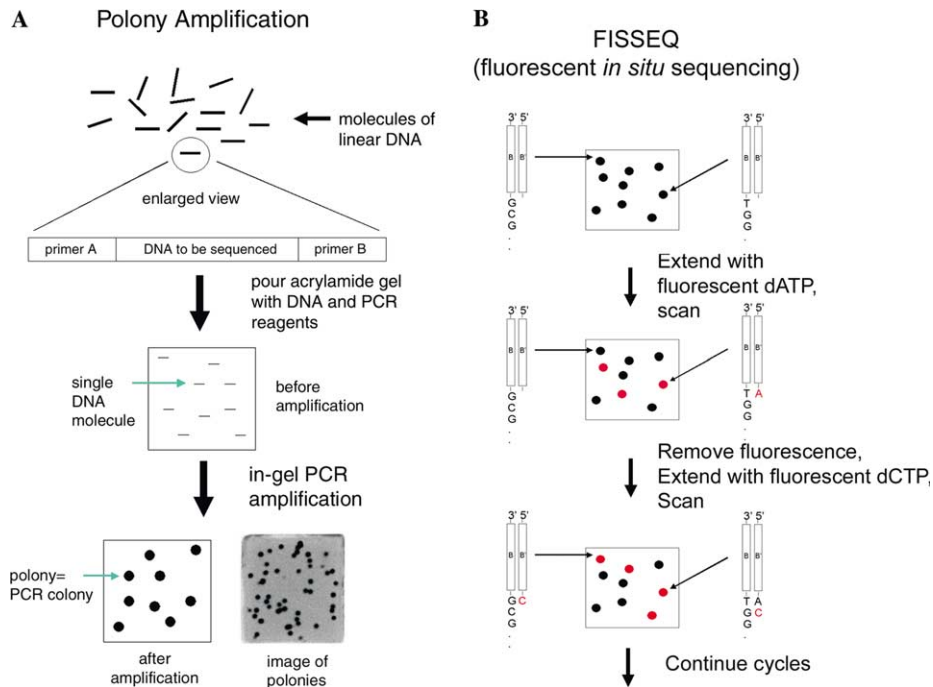


Fig. 1. (A) Polony amplification. A library of linear DNA molecules with universal priming sites is PCR amplified in a polyacrylamide gel. A single template molecule gives rise to a polymerase colony or polony. (B) Fluorescent in situ sequencing. Polonies are denatured, and a sequencing primer is annealed. Polonies are sequenced by serial additions of a single fluorescent nucleotide.

on three glass slides, and a sequencing primer was annealed (Fig. 2A). One template required the addition of dTTP (or its analogue dUTP) for correct incorporation. The other, a negative control, required the addition of dCTP for correct incorporation. Slides were incubated with unlabeled dTTP and DNA polymerase for varying amounts of time, ranging from 15 s to 6 min. A control slide was not extended with dTTP. Cy5-labeled dUTP and DNA polymerase were then added to all slides, and the amount of fluorescent signal on each slide was compared to estimate the efficiency of unlabeled dTTP incorporation (Fig. 2A). If the extensions went to completion, we would expect to see no fluorescent signal on slides incubated with unlabeled dTTP before the fluorescent extension. However, even after 6 min of incubation, there was still significant fluorescent signal incorporated (Figs. 2A and B), indicating that the extension with dTTP did not go to completion. By quantifying the signal, we estimated the efficiency of the 6-min dTTP extension to be approximately 85%. As discussed above, a value this low would not permit the accurate sequencing of more than a few bases. By plotting the measured signal versus the time of dTTP incubation (Fig. 2B), we observed that the fraction of primer:template molecules extended can be approximated by the function $F = 1 - (0.66e^{-t/12.8s} + 0.34e^{-t/500s})$. From this equation, we concluded that even significantly longer extension times would not significantly increase extension efficiency. We repeated the experiment using thermostable polymerase and

extending at higher temperatures, but this did not increase extension efficiency (data not shown).

To solve this problem, we developed a technique called “polymerase trapping,” in which DNA polymerase is allowed to bind to the primer:template duplex, acrylamide monomer is added to the reaction, and polymerization is initiated, trapping the polymerase in a complex with the primer:template duplex. The polyacrylamide matrix prevents the polymerase from diffusing away from the primer:template, so that every primer that is extended during the first cycle of extension will continue to be extended at later cycles. We repeated the experiment described above using polymerase trapping and estimated that in this case, extension efficiency reaches 99.8% (Fig. 2C). Trapping the polymerase within a polyacrylamide matrix also has the advantage that the enzyme is not replenished with each nucleotide addition, thus reducing the cost of the reaction. We determined that trapped polymerase molecules will remain bound to the primer:template complex for over 72 h without diffusing away (supplementary information, http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm).

As a second method for estimating the extension efficiency when polymerase trapping is employed, we amplified an 87-bp synthetic oligonucleotide in a polony reaction. The polonies were denatured, a sequencing primer was annealed, and DNA polymerase was trapped on the immobilized DNA. Serial, single-nucleotide additions were performed with unlabeled nucleotides for a

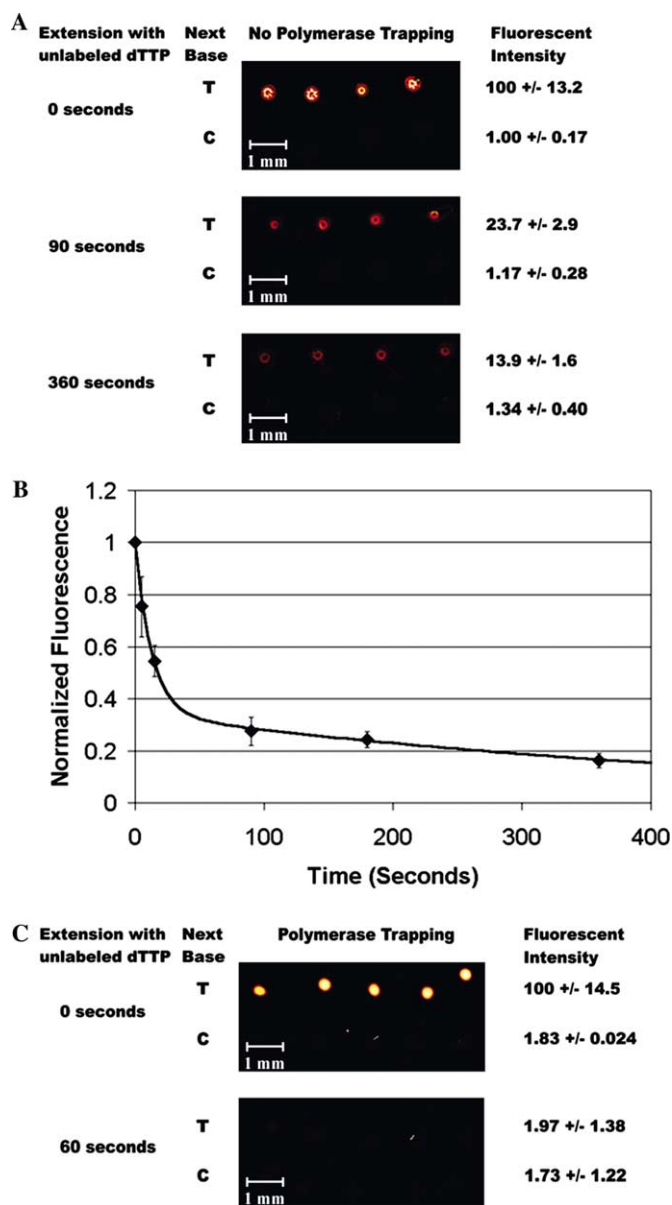


Fig. 2. Estimating the efficiency of the extension reaction. (A) Two templates were immobilized in acrylamide spots, the first requiring a T to extend the annealed primer and the second requiring a C. Unlabeled dTTP and DNA polymerase were added to the slide for the indicated amount of time, followed by a 6-min incubation with fluorescently labeled dUTP and DNA polymerase. In all cases there was incorporation of the fluorescent nucleotide, indicating that the primer:template complexes were not efficiently extended when the unlabeled nucleotide was added. (B) The experiment was repeated using different incubation times with unlabeled dTTP, and the incorporated fluorescence was plotted versus time. (C) The experiment was repeated using the polymerase trapping technique. There was no fluorescent signal incorporated if the primer:template molecules were first incubate with unlabeled dTTP for 60 s.

number of cycles until the next base required to extend the growing strand was either a C or a T. A mixture of FITC-labeled dCTP (green in Fig. 3) and Cy3-labeled dUTP (red in Fig. 3) was then added to the slide to determine whether the correct base was incorporated. If the primer:template molecules in each polony had become dephased during the course of the unlabeled extensions, one would expect to see high levels of the incorrect base incorporated and therefore signal in both the red and the green channels; however, even after 26

single-base additions, which caused the DNA polymerase to extend the sequencing primer by 34 bases, only the correct channel displayed significant signal (slide 4 in Fig. 3).

Characterization of fluorescent nucleotides with a disulfide linker

We synthesized four fluorescent deoxynucleotide analogues containing a disulfide linker between the

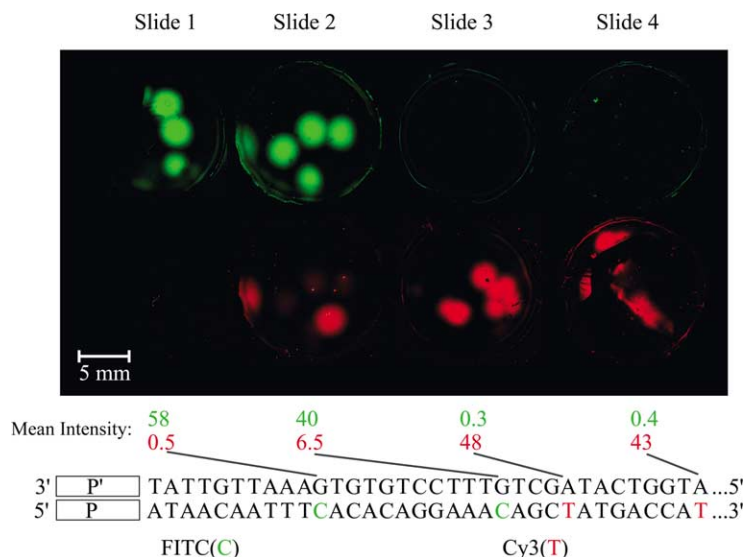


Fig. 3. Multiple rounds of nucleotide addition on polonies. Polonies were amplified using a known DNA template (sequence shown below slide image). Polonies were subjected to multiple rounds of extension with unlabeled nucleotides (the added nucleotides are shown above each image) followed by the addition of a mixture of FITC-dCTP and Cy3-dUTP. For slide 1, 7 rounds of nucleotide addition were performed, extending the primer 11 bases. For slide 2, 15 rounds of nucleotide addition were performed, extending the primer 22 bases. The fluorescent region in the lower right portion of the red channel image is due to a primer-dimer artifact. For slide 3, 19 rounds of nucleotide addition were performed, extending the primer 26 bases. For slide 4, 26 rounds of nucleotide addition were performed, extending the primer 34 bases. For this slide, the gel was torn during the course of the extensions, but polonies can be seen and quantified in the upper left corner of the slide. (This experiment was repeated several times, but in all experiments the polyacrylamide gel became torn after 25–34 extensions.) The mean incorporated fluorescent intensity for each channel is shown below the images of the slides. In all cases, only the expected fluorescent base was incorporated, indicating that the primer:template complexes remained in phase through 34 extension reactions.

fluorophore and the nucleotide (Fig. 4A and http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm). The design of these analogues was inspired by the published structure of the chemically cleavable biotinylated nucleotide analog, Bio-12-SS-dUTP [24]. To determine whether DNA polymerase would recognize these fluorescent analogues, we performed a nucleotide extension reaction on amplified polonies using one of them, Cy5-SS-dCTP. We used five different templates for generating polonies, with only one template requiring a C to correctly extend the sequencing primer. The result of this experiment is shown in Fig. 4B. In this image, the green (532-nm) and red (635-nm) channels are merged. The sequencing primer was labeled at its 5' end with a Cy3 dye, so that all polonies on the slide are visible in the green channel (Fig. 4C). A fraction of the polonies display fluorescent signal in the red channel (yellow polonies), indicating a successful extension (Fig. 4B). We verified that only the correct polonies incorporated the fluorescent base by performing additional single base extensions with dye-labeled nucleotides and were able to match every polony on the gel with one of the five template sequences. This experiment demonstrated that DNA polymerase recognizes the Cy5-SS-dCTP nucleotide analogue and correctly incorporates the nucleotide during a single-base extension reaction.

To establish that fluorophore incorporated by polymerase can be removed, we incubated the slide in a

buffer containing β -mercaptoethanol and monitored the fluorescence decay (Figs. 4C and D). The fluorescent signal decayed exponentially, with a time constant of 7 s. To determine the incorporation efficiencies of these fluorescent analogues relative to that of the corresponding natural deoxynucleotide, we performed extension reactions using varying ratios of each and quantified the incorporated fluorescence (supplementary information, http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm). In addition to the Cy5-SS-dNTP molecules just described, we also synthesized dye-labeled nucleotide analogues with a photocleavable linker [25], Cy5-PC-dUTP, and BO-DIPY-FL-PC-dUTP. DNA polymerase also incorporated the Cy5-PC-dUTP analogue correctly during single-base extension and the dye could be removed by exposure to near-UV light. The design, synthesis, and characterization of this molecule can be found in the supplementary materials (http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm).

Sequencing polymerase colonies

As a proof of principle for polony sequencing, we used five synthetic oligonucleotides (templates T1–T5, see http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm) as templates for polony amplification on two slides. Approximately 40 polonies

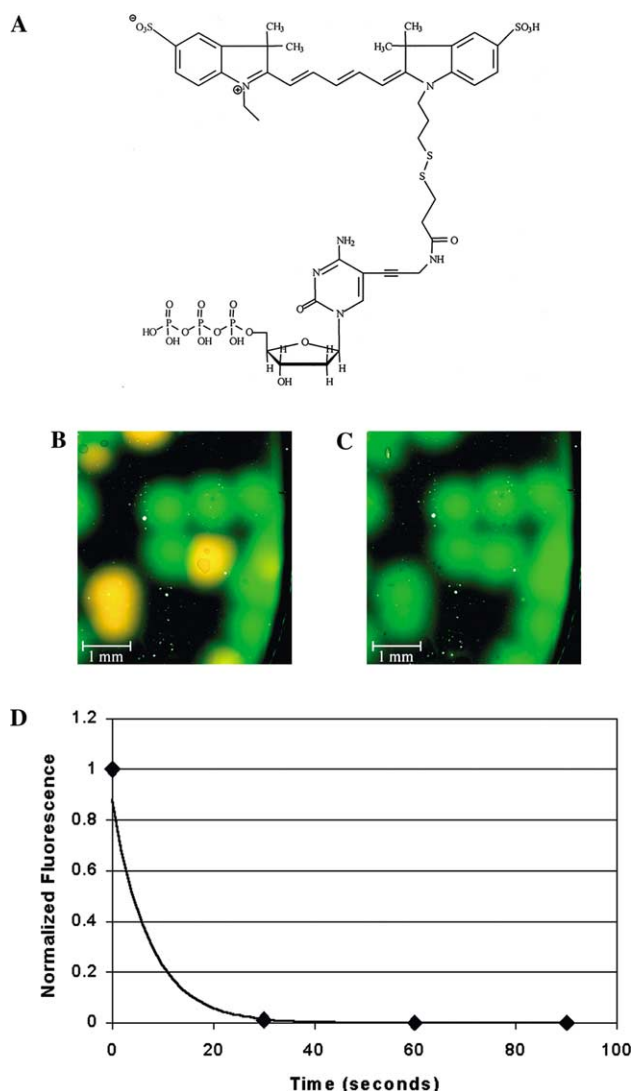


Fig. 4. (A) Structure of Cy5-SS-dCTP. This molecule has a disulfide bond between the fluorophore and the deoxynucleotide so that the label can be removed after incorporation. (B) Extension of an immobilized primer:template with Cy5-SS-dCTP. The yellow polonies indicate that the DNA polymerase recognizes the nucleotide analogue. (C) After a wash in buffer containing β -mercaptoethanol, the fluorescence can be removed from the slide. (D) The signal decays exponentially with a time constant of 7 s.

were amplified on each slide. We denatured the polonies, hybridized to a Cy3-labeled sequencing primer, and trapped the polymerase onto the immobilized primer:template duplex. We then performed serial base additions with the Cy5-SS-dNTPs according to the protocol outlined in Fig. 1B (no unlabeled nucleotides were used in this reaction). Slide images from the first four cycles of base addition are shown in Fig. 5A. Here, the green (Cy3-labeled sequencing primer) and red (Cy5-labeled base) channels are merged. Polonies that have not incorporated the added base are green, and polonies that have incorporated the added base are yellow. We performed 10 extension cycles on slide 1 and 14 exten-

sion cycles on slide 2 before the polyacrylamide gel detached from the surface of the slide.

To computationally identify polonies and call a sequence for each polony we wrote a software package, PolCall (see supplementary information, http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENT_ARY_INFO.htm). This package aligns the images collected from each cycle and determines a sequence for every pixel. Pixels are then combined into objects based on their signature sequence, and a series of filters are applied to remove objects due to background noise and objects that represent the overlap of two polonies. This software was used to analyze the images from the polony sequencing cycles. Only a portion of each slide was analyzed, as some regions of the gel were torn during the sequencing process. This analysis was performed for slide 1 (Fig. 5B, Table 1) and slide 2 (Fig. 5C, Table 2).

In total, 24 polonies were analyzed, and five unique sequences were identified. The called sequences are labeled S1–S5 in Table 1. Sequences S1–S3 exactly match the template sequences T1–T3, indicating that these molecules were accurately amplified and sequenced. Sequence S4 has the same base composition as the template sequence T4; however, we were unable to correctly identify repeated base pairs. This was expected since 100% fluorescent base was used and fluorophore quenching affects signal linearity. S5 is similar to the expected sequence T5, but there are two bases that were called present in S5 that are not present in the template sequence (Table 2), a G in the second position and a T in seventh position. These errors were systematic, as there were a total of 5 polonies that all called the sequence S5. The longest accurate read achieved in this experiment was 8 bp (sequences S2 and S3 on slide 2).

Polony exclusion at high densities

In the polony sequencing experiment described above, pixels corresponding to regions of the slide where two polonies overlapped were automatically removed from the analysis by the PolCall software. This step is necessary to maximize accuracy. However, we would also like to maximize the number of useful pixels per scan because we project the cost of scanning to be the most costly component of this technology. Thus, we hypothesized that by decreasing the concentration of free primer in the polony amplification reaction, we could reduce the amount of overlap between adjacent polonies. This would be due to the exhaustion of the free primer in areas of high polony density, which would inhibit further polony growth. To test this hypothesis, we amplified two polony templates using a concentration of free primer reduced by threefold. Primer extension was performed using Cy3- and Cy5-labeled nucleotides to distinguish the two species. Fig. 6 demonstrates that polony overlap can be reduced using this

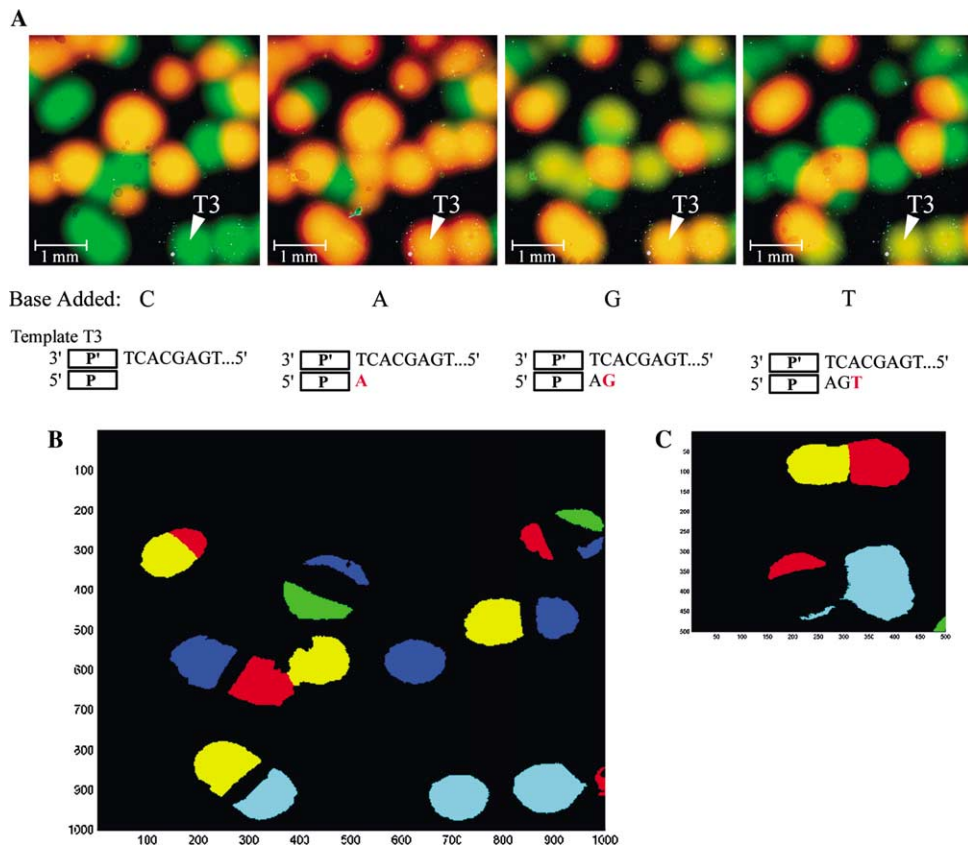


Fig. 5. FISSEQ on polonies. (A) First four cycles of fluorescent base additions on slide 1. The sequencing primer is Cy3 (green) labeled, and the fluorescent base is Cy5 (red) labeled. Polonies that have incorporated the added base appear yellow on the image; green polonies have not incorporated the base. The sequence of one of the included templates, T3, is shown below the images and the corresponding polony is labeled in each of the images. The order of base addition was C,A,G,T. The yellow polonies display different Cy5 intensities due to sequence context effects on dye fluorescence. (B) Automated polony base calling for slide 1, which was subjected to 10 FISSEQ cycles. This is the same portion of the slide that is displayed in (A). Each called sequence was assigned a color. The called sequences, their color assignments, and the corresponding template sequence can be found in Table 1. (C) Automated polony base calling for slide 2, which was subjected to 14 FISSEQ cycles. The called sequences, their color assignments, and the corresponding template sequence can be found in Table 2.

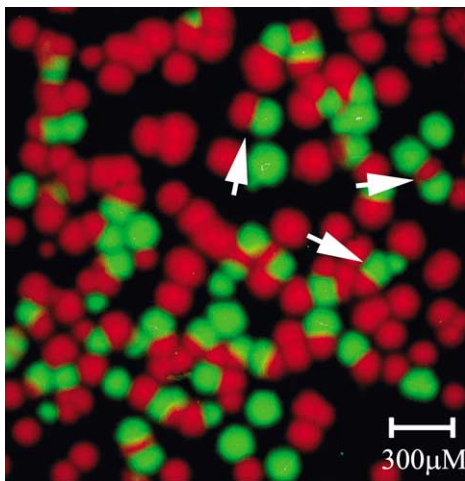


Fig. 6. Polony exclusion at high densities. When free primer is made limiting, adjacent polonies do not overlap due to inhibited growth by neighboring polonies.

strategy. Also, numerous adjacent polonies display a deformed shape (arrows in Fig. 6) indicating growth inhibition due to the presence of a neighbor.

Discussion

There are three biochemical sources of error that will likely determine the maximum read length attainable using the FISSEQ approach: mispriming, misincorporation, and incomplete extension. Mispriming occurs when a sequencing primer anneals nonspecifically to a template molecule and is extended upon addition of nucleotide. Mispriming can also occur when the 3' end of a single-stranded DNA template forms a hairpin, which is then extended upon base addition. Misincorporation occurs when a nucleotide is incorporated opposite a noncomplementary template base. Once this

Table 1
Called sequences from slide 1 (10 cycles)

Object color	Sequence ID	Called sequence	Expected sequence	Template sequence	Number of polonies	Mismatched bases
Green	S1	CACACA	CACACA	T1: CACACACACACACTC	2	0
Red	S2	GTGT	GTGT	T2: GTGTGTGTGTGTGTGTC	4	0
Lt blue	S3	AGTGC	AGTGC	T3: AGTGCTCACACACGTGATC	3	0
Dk blue	S4	CAGCGA	CAGCGA	T4: CAGCCGAACGACCGATC	5	0
Yellow	S5	AGTGT	ATGT	T5: ATGTGAGAGCTGTCGTC	4	1

Table 2
Called sequences from slide 2 (14 cycles)

Object color	Sequence ID	Called sequence	Expected sequence	Template sequence	Number of polonies	Mismatched bases
Green	S1	CACACACA	CACACACA	T1: CACACACACACACTC	1	0
Red	S2	GTGTGT	GTGTGT	T2: GTGTGTGTGTGTGTGTC	2	0
Lt blue	S3	AGTGCTCA	AGTGCTCA	T3: AGTGCTCACACACGTGATC	2	0
Yellow	S5	AGTGTGTA	ATGTGA	T5: ATGTGAGAGCTGTCGTC	1	2

happens, subsequent incorporation will be less efficient. Incomplete extension occurs when the incorporation reaction does not go to completion so that only a fraction of the primer:template molecules in a polony have correctly incorporated the added nucleotide. It is important to minimize misincorporation and incomplete extension as these types of error accumulate exponentially with the number of extensions.

By comparing the fluorescent signal of the two templates extended in Fig. 2A, we can estimate an upper limit for the error due to mispriming and misincorporation. When dUTP was added to the samples, the mismatched templates displayed only 1% of the fluorescent signal of the perfect match template. To estimate how much of this erroneous signal was due to misincorporation, as opposed to mispriming, we repeated the single-base extension experiment on these templates using several sequencing primers that were not complementary to any region of the template. We found that the fluorescence signal was consistently equal to or slightly greater than the previously measured signal, suggesting that most, if not all of the erroneous signal in Fig. 2 is due to mispriming (data not shown). Unlike misincorporation and incomplete extension, error due to mispriming does not accumulate exponentially with the number of base additions, so that the mispriming that occurs in Fig. 2 should not substantially limit the read length of FISSEQ. Mispriming can be reduced by performing the single base extensions at higher temperatures. For this reason, we used BST polymerase and performed the single base extensions at 55°C for the FISSEQ experiment shown in Fig. 5. It should be possible to further reduce mispriming by including a sequence at the 3' end of the DNA template during the initial amplification step that forms a stem loop struc-

ture when the template is made single stranded. The 3' end of the template molecule can then act as a primer for DNA sequencing [26]. This hairpin reaction should be highly specific, reducing the chance of mispriming. This approach also eliminates the need for the additional primer required for the extension reaction.

The results in Fig. 2B demonstrate that the extension reaction does not proceed to completion if polymerase trapping is not employed. We hypothesize that the reason that the extension reaction did not go to completion is that we were unable to saturate the primer:template molecules with DNA polymerase. Consequently, adding more polymerase at each cycle resulted in the dephasing of the extended primers. To ensure efficient extension, we developed a protocol in which polymerase is added only at the beginning of the sequencing reaction, is "trapped" on the primer:template duplex, and remains trapped for all subsequent extensions. When this protocol is applied, the extension efficiency increases substantially and reaches over 99.8% (Fig. 2C). Primer:template molecules that do not have a trapped polymerase molecule during the first cycle of extension are not extended during any cycle, so they do not contribute to the dephasing of the fluorescent signal. The results in Fig. 3 demonstrate that as many as 34 extensions can be performed without significant dephasing of the templates or reduction in fluorescent signal. Further extensions have not yet been attempted, but by estimating the extension efficiency from Figs. 2C and 3, we predict that the primer can be extended 150 bases without significant dephasing. It is possible that the maximum length of an extension product may be limited because the polymerase is trapped in an acrylamide pore. However, a 34-bp extension corresponds to a 11.9-nm translocation of the DNA polymerase on the DNA

template, a length greater than the predicted pore size (2 nm) of a 15% acrylamide gel with 5% cross-linker [27]. Therefore, it seems that acrylamide pore size does not limit FISSEQ read length in the useful range in which we are working, possibly because of polymer flexibility.

The FISSEQ reactions shown in Fig. 5 and summarized in Tables 1 and 2 produced the correct sequence for three of the five templates. In this experiment, the lengths of the sequencing reads were limited by the fact that the gels became detached from the glass slides after 10 (slide 1) or 14 (slide 2) cycles. This experiment was repeated several times, but, in all cases, gel detachment occurred after 10–14 cycles. We are exploring two potential solutions to this problem. First, we are working to reduce the thickness of the acrylamide gels from 40 μm to between 1 and 5 μm . Thinner gels are subject to lower shear forces during washing and due to gel swelling, and our initial experience indicates that decreasing the thickness can substantially improve gel stability. We are also experimenting with changing the chemistry by which the acrylamide is attached to the slide. The experiments described here used a glass microscope slide coated with bind-silane, and silane bonds are prone to hydrolysis [28]. If a clear polymer slide is used instead of glass [29,30], a variety of different attachment chemistries can be employed.

In its current embodiment, the FISSEQ protocol has trouble calling repeated bases. For example, the FISSEQ reaction summarized in Table 2 did not correctly call the sequence for template T4. Sequence S4 is the sequence expected for template T4, except that the repeated bases in T4 are called as a single base in S4 (e.g., CAGCC is called CAGC). A challenge to resolving repeated bases is the quenching between the adjacent fluorophores, so that the fluorescent signal does not increase linearly with the number of incorporated bases, but instead decreases (data not shown). To resolve this issue, we have attempted using a mixture of fluorescent nucleotide and nonfluorescent nucleotide, hoping to see a linear increase in signal proportional to the number of bases added (as is seen in pyrosequencing [11,31]). However, our initial results indicate that the relative incorporation of fluorescent nucleotide to natural nucleotide changes in different sequence contexts, again making it difficult to correctly call multiple base repeats (data not shown). Once a large amount of sequence is obtained from FISSEQ reactions, it may be possible to correlate sequence context with incorporated fluorescence. We could then use this knowledge to improve the accuracy of calling repeated bases. A second solution to the problem of calling repeated bases is to use nucleotide analogues with a reversibly blocked 3' OH group [32,33]. This dye terminator strategy has the advantage that only one base is added at each cycle, so that precise quanti-

tation of incorporated fluorescence is not necessary. However, it is critical that the reaction used to unblock the 3' OH group goes to completion, otherwise the DNA molecules in a polony will rapidly dephase. Also, the dye terminator strategy will require more FISSEQ cycles for a given read length and likely will have higher polymerase misincorporation rates. We are currently investigating both approaches.

In addition to miscalled sequences caused by multiple incorporation of the same nucleotide, we observed another type of error. The sequence of template T5 was not correctly determined by the FISSEQ protocol. By examining Table 2 we see that the G added in the 2nd cycle of base addition and the T added in the 12th cycle were erroneously incorporated. We are uncertain why there is fluorescent signal incorporated in the T5 polonies at these cycles, but the incorporations are repeatable over multiple polonies and experiments (Table 1 and data not shown). It is unlikely to be due to misincorporation at one of the extension cycles, as this would terminate the extension reaction [34]. One possibility is that these incorporations are due to misprimed extensions at the 3' end of the immobilized DNA. Further work is needed to understand the source of this error.

The innovations presented here bring us closer toward our goal of sequencing genomes using polymerase colonies. We have demonstrated short sequence reads on multiple polonies and plan to extend these results by attempting longer read lengths, growing polonies at higher densities, and improving the accuracy of our reads. Although this technology is still in its early stages of development, we believe that it has the potential to greatly reduce the cost of DNA sequencing. If the hurdles described above can be overcome, we estimate the resequencing of a diploid human genome using polony sequencing would cost less than \$6000, at least three orders of magnitude less than current costs. The details of this estimate can be found in the supplementary information (http://arep.med.harvard.edu/polony/FISSEQ_SUPPLEMENTARY_INFO.htm). We have also developed other applications for polymerase colonies including digital genotyping, long-range haplotyping, and quantifying alternate splicing.

Acknowledgments

We thank Benjamin Williams, Vincent Butty, Jun Zhu, Martin Steffen, Xiaohua Huang, and Greg Porreca for discussions and ideas with regard to polony protocol development. We thank Ting Wu, Aimee Dudley, Barak Cohen, Martin Steffen, Vasu Badarinarayana, and members of the Church lab for helpful discussions and critical readings of the manuscript. This work was supported by the US Department of Energy (DE-FG02-87-ER60565) and DARPA (BIO-COMP).

Note added in proof

We recently became aware of a relevant publication [35] that describes the synthesis and characterization of a photocleavable fluorescent nucleotide similar to the one described in our supplementary material.

References

- [1] R. Kurzweil, <http://www.kurzweilai.net/articles/art0134.html>, (2002).
- [2] F. Sanger, Sequences, sequences and sequences, *Annu. Rev. Biochem.* 57 (1988) 1–28.
- [3] E. Pennisi, The Institute for Genomic Research meeting. Gene researchers hunt bargains, fixer-uppers, *Science* 298 (2002) 735–736.
- [4] M.A. Osborne, W.S. Furey, D. Klenerman, S. Balasubramanian, Single-molecule analysis of DNA immobilized on microspheres, *Anal. Chem.* 72 (2000) 3678–3681.
- [5] S.-C. Tu, J.M. Briggs, X. Gao, S.H. Hardin, R. Willson, Real-Time Sequence Determination, World Patent Organization #WO0204680.
- [6] E. Chan, Methods and products for analyzing polymers, US Patent #6,355,420.
- [7] J. Korlach, M. Levene, S.W. Turner, D.R. Larson, M. Foquet, H.G. Craighead, W.W. Webb, A new strategy for sequencing individual molecules of DNA, *Biophys. J.* 80 (2001) 65501.
- [8] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, D. Branton, Rapid nanopore discrimination between single polynucleotide molecules, *Proc. Natl. Acad. Sci. USA* 97 (2000) 1079–1084.
- [9] S. Winters-Hilt, W. Vercoutere, V.S. DeGuzman, D. Deamer, M. Akeson, D. Haussler, Highly accurate classification of Watson-Crick basepairs on termini of single DNA molecules, *Biophys. J.* 84 (2003) 967–976.
- [10] J.M. Rothberg, J.S. Bader, Method of sequencing a nucleic acid, US Patent #6,274,320.
- [11] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, P. Nyren, Real-time DNA sequencing using detection of pyrophosphate release, *Anal. Biochem.* 242 (1996) 84–89.
- [12] N. Patil, A.J. Berno, D.A. Hinds, W.A. Barrett, J.M. Doshi, C.R. Hacker, C.R. Kautzer, D.H. Lee, C. Marjoribanks, D.P. McDonough, B.T. Nguyen, M.C. Norris, J.B. Sheehan, N. Shen, D. Stern, R.P. Stokowski, D.J. Thomas, M.O. Trulson, K.R. Vyas, K.A. Frazer, S.P. Fodor, D.R. Cox, Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21, *Science* 294 (2001) 1719–1723.
- [13] K.A. Frazer, J.B. Sheehan, R.P. Stokowski, X. Chen, R. Hosseini, J.F. Cheng, S.P. Fodor, D.R. Cox, N. Patil, Evolutionarily conserved sequences on human chromosome 21, *Genome Res.* 11 (2001) 1651–1659.
- [14] R.D. Mitra, G.M. Church, In situ localized amplification and contact replication of many individual DNA molecules, *Nucleic Acids Res.* 27 (1999) e34.
- [15] R.D. Mitra, Polony sequencing: a DNA sequencing technology, Ph.D. Thesis, Massachusetts Institute of Technology, 2000.
- [16] H.V. Chetverina, A.B. Chetverin, Cloning of RNA molecules in vitro, *Nucleic Acids Res.* 21 (1993) 2349–2353.
- [17] H.V. Chetverina, T.R. Samatov, V.I. Ugarov, A.B. Chetverin, Molecular colony diagnostics: detection and quantitation of viral nucleic acids by in-gel PCR, *Biotechniques* 33 (2002) 150–152, 154, 156.
- [18] R.D. Mitra, V.L. Butty, J. Shendure, B.R. Williams, D.E. Housman, G.M. Church, Digital genotyping and haplotyping with polymerase colonies, *Proc. Natl. Acad. Sci.* 100 (2003) 5926–5931.
- [19] B.S. Singer, T. Shtatland, D. Brown, L. Gold, Libraries for genomic SELEX [published erratum appears in *Nucleic Acids Res.* 1997 Nov 1;25(21):4430], *Nucleic Acids Res.* 25 (1997) 781–786.
- [20] F.N. Rehman, M. Audeh, E.S. Abrams, P.W. Hammond, M. Kenney, T.C. Boles, Immobilization of acrylamide-modified oligonucleotides by co-polymerization, *Nucleic Acids Res.* 27 (1999) 649–655.
- [21] A.V. Vasiliskov, E.N. Timofeev, S.A. Surzhikov, A.L. Drobyshev, V.V. Shick, A.D. Mirzabekov, Fabrication of microarray of gel-immobilized compounds on a chip by copolymerization, *Biotechniques* 27 (1999) 592–594, 596–598, 600 passim.
- [22] A. Kolchinsky, A. Mirzabekov, Analysis of SNPs and other genomic variations using gel-based chips, *Hum. Mutat.* 19 (2002) 343–360.
- [23] M.H. Shaper, K.K. Leuther, A. Nguyen, M. Scott, K.W. Jones, SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing, *Genome Res.* 11 (2001) 1926–1934.
- [24] M.L. Shimkus, P. Guaglianone, T.M. Herman, Synthesis and characterization of biotin-labeled nucleotide analogs, *DNA* 5 (1986) 247–255.
- [25] J. Olejnik, E. Krzymanska-Olejnik, K.J. Rothschild, Photocleavable affinity tags for isolation and detection of biomolecules, *Methods Enzymol.* 291 (1998) 135–154.
- [26] M. Ronaghi, B. Pettersson, M. Uhlen, P. Nyren, PCR-introduced loop structure as primer in DNA sequencing, *Biotechniques* 25 (1998) 876–878, 880–882, 884.
- [27] P.G. Righetti, S. Caglio, M. Saracchi, S. Quaroni, ‘Laterally aggregated’ polyacrylamide gels for electrophoresis, *Electrophoresis* 13 (1992) 587–595.
- [28] N. Nishiyama, T. Ishizaki, K. Horie, M. Tomari, M. Someya, Novel polyfunctional silanes for improved hydrolytic stability at the polymer-silica interface, *J. Biomed. Mater. Res.* 25 (1991) 213–221.
- [29] J.P. McNamee, J.R. McLean, C.L. Ferrarotto, P.V. Bellier, Comet assay: rapid processing of multiple samples, *Mutat. Res.* 466 (2000) 63–69.
- [30] P. Reinhardt-Poulin, J.R. McLean, Y. Deslauriers, W. Gorman, S. Cabat, M. Rouabhia, The use of silver-stained comets to visualize DNA damage and repair in normal and Xeroderma pigmentosum fibroblasts after exposure to simulated solar radiation, *Photochem. Photobiol.* 71 (2000) 422–425.
- [31] M. Ronaghi, M. Uhlen, P. Nyren, A sequencing method based on real-time pyrophosphate, *Science* 281 (1998) 363–365.
- [32] M.B. Welch, K. Burgess, Synthesis of fluorescent, photolabile 3'-O-protected nucleoside triphosphates for the base addition sequencing scheme, *Nucleosides Nucleotides* 18 (1999) 197–201.
- [33] M.L. Metzker, R. Raghavachari, K. Burgess, R.A. Gibbs, Elimination of residual natural nucleotides from 3'-O-modified-dNTP syntheses by enzymatic mop-up, *Biotechniques* 25 (1998) 814–817.
- [34] M.J. Donlin, S.S. Patel, K.A. Johnson, Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction, *Biochemistry* 30 (1991) 538–546.
- [35] Z. Li, X. Bai, H. Ruparel, S. Kim, N.J. Turro, J. Ju, A photocleavable fluorescent nucleotide for DNA sequencing and analysis, *Proc. Natl. Acad. Sci. USA* 100 (2003) 414–419.