



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc



Pyrosequencing enhancement for better detection limit and sequencing homopolymers

Min Gong^a, Shi-Hui Foo^a, Lifang Lin^a, Edison T. Liu^a, Baback Gharizadeh^b, Sanket Goel^{a,c,*}

^a Genome Institute of Singapore, 60 Biopolis Street, 02-01 Genome, Singapore 138672, Singapore

^b Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

^c Institute of Microelectronics, 11 Science Park Road, Science Park II, Singapore 117685, Singapore

ARTICLE INFO

Article history:

Received 1 September 2010

Available online 15 September 2010

Keywords:

Genome sequencing

Pyrosequencing

Porous beads

Homopolymeric regions

Biotinylated DNA

Streptavidin-coated beads

ABSTRACT

Pyrosequencing is a DNA sequencing technique based on sequencing-by-synthesis enabling rapid and real-time sequence determination. Although ample genomic research has been undertaken using pyrosequencing, the requirement of relatively high amount of DNA template and the difficulty in sequencing the homopolymeric regions limit its key advantages in the applications directing towards clinical research. In this study, we demonstrate that pyrosequencing on homopolymeric regions with 10 identical nucleotides can be successfully performed with optimal amount of DNA (0.3125–5 pmol) immobilized on conventional non-porous Sepharose beads. We also validate that by using porous silica beads, the sequencing signal increased 3.5-folds as compared to that produced from same amount of DNA immobilized on solid Sepharose beads. Our results strongly indicate that with optimized quantity of DNA and suitable solid support, the performance of pyrosequencing on homopolymeric regions and its detection limit has been significantly improved.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The sequencing of entire genomes, their annotation and variations has been the critical enabling factors for all biological research. Following the completion of human genome project, there have been continuous efforts to realize a more accurate and time-and-cost efficient platform for many applications. Primarily, this resulted three high-throughput sequencing platforms: the genome sequencers from Roche/454 [1]; the Genome analyzer from Illumina/Solexa [2] and the SOLiD system from Applied Biosystems [3]. On a positive note, the deliverable throughput from present high-throughput platforms is as high with few hundreds of gigabases per run costing as low as ten thousand dollar. However, owing to their longer sample preparation time (2–7 days), longer sequencing time (2–10 days) and requirement of ultra-high computational and post-processing power, their utilization is limited to only big genomic centers and mainly for *de novo* and resequencing applications. There has been constant work in the detection of genetic variations in a large number of samples representing a broad range of biological material providing an insight into genetic mechanisms of different diseases. In identification of genetic variants such as single nucleotide polymorphism (SNP) genotyping, detection of specific mutations and gene identi-

fication in transcript analysis, in place of a high-throughput platform, a reliable low-throughput platform with a real-time and fast sample-to-answer system, where read-length is not a limiting factor, is a preferred choice. For these applications, a pyrosequencing based platform is becoming very attractive and has been successfully used in many such applications [4–9].

Pyrosequencing is a real-time bioluminescence based sequencing-by-synthesis method catalyzed by four kinetically well-balanced enzymes: DNA polymerase, ATP Sulfurylase, firefly luciferase and apyrase [10]. In the presence of DNA polymerase, complementary nucleotide (A, C, G or T) incorporated into the single-stranded DNA (ssDNA) sample leads to generation of pyrophosphate (PPi) in a quantity equimolar to the number of incorporated nucleotides. The released PPi triggers the ATP Sulfurylase reaction resulting in a quantitative conversion of PPi to ATP. ATP is then consumed by the luciferase for producing bioluminescence which is proportional to the amount of DNA and number of incorporated nucleotides. The unincorporated nucleotides and the generated ATP are degraded by apyrase allowing iterative addition of next nucleotide dispensation [10].

Pyrosequencing can be performed in two different phases – liquid and solid. In liquid-phase pyrosequencing, free DNA template, four enzymes and their substrates are mixed appropriately in a reaction-cell. Solid-phase pyrosequencing utilizes immobilized DNA in a three-enzyme system (DNA polymerase, ATP Sulfurylase and firefly luciferase) where a washing step is performed to remove the excess substrate after each nucleotide addition.

* Corresponding author at: Genome Institute of Singapore, 60 Biopolis Street, 02-01 Genome, Singapore 138672, Singapore. Fax: +65 6808 8305.

E-mail address: goels@gis.a-star.edu.sg (S. Goel).

Template preparation is quite straightforward where solid-phase template preparation is the usual strategy used for generation of a primed DNA template for pyrosequencing [11]. In brief, after generation of biotinylated DNA template by polymerase chain reaction (PCR), double-stranded DNA (dsDNA) samples can be captured onto streptavidin-coated beads. After sedimentation, ssDNA sequencing template can be produced by alkalies denaturation. The non-biotinylated strand in the solution can be used in liquid-phase pyrosequencing whereas immobilized biotinylated strand can be used in both liquid and solid-phase pyrosequencing systems.

Though pyrosequencing has been widely used in various applications [6–9], except for its relative short read-length, pyrosequencing has another two main limitations: the detection sensitivity is significantly limited due to high background signal generated from nonspecific reactivity of luciferase with APS, and difficulty in reading the sequence in homopolymeric regions – stretches of the same base in the DNA sample. In this paper, our work centers on the improvement of pyrosequencing on immobilized DNA sample by targeting the aforementioned two limiting factors – amount of DNA sample and sequencing homopolymeric regions.

2. Materials and methods

2.1. Reagents

Streptavidin-coated Sepharose™ high performance, Silica microspheres and Paramagnetic (Dynabeads™ M280) beads were purchased from GE Healthcare (Uppsala, Sweden), Corpuclar (New York) and Invitrogen (Oslo, Norway) respectively. PyroGold reagents were purchased from Qiagen (Uppsala, Sweden). Plasmid pUC19 and PCR amplification master mix were purchased from Fermentas (Burlington, Canada).

2.2. Oligonucleotides

The oligonucleotides (Table 1) were synthesized and HPLC purified by Sigma (St. Louis, MO). PCR primers pUC19_F and pUC19_R were used for amplifying a 221 bp DNA fragment of pUC19 [12].

2.3. Template DNA preparation

The 221 bp PUC 19 fragment was obtained by PCR reaction with primer pair of pUC19_F and pUC19_R. The amplification was carried out on a Bio-Rad thermo Cycler PCR system (CA, USA) with the following protocol: denaturing at 95 °C for 30 s, followed by 30 thermal reaction cycling (95 °C for 30 s, 60 °C for 30 s and 72 °C for 45 s). After the thermal cycle reaction, the product was incubated at 72 °C for 5 min to ensure the complete extension of the amplified DNA fragment. Amplified 221 bp DNA fragment was purified using QIAquick PCR purification kit (Qiagen).

Biotinylated DNA oligonucleotides were immobilized on three different streptavidin-coated beads, namely Sepharose beads, silica beads and paramagnetic beads. The immobilization was performed by incubating the mixture of DNA and beads in binding buffer (10 mM Tris–HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20, pH 7.6) at room temperature for 30 min. Single-stranded pUC19 DNA fragment was obtained using vacuum preparation protocol according to the manufacturer's instructions. Annealing of sequencing primer (SG_seq) with different biotinylated DNA oligos (SG_122, SG_122m, SG_A, SG_10T and SG_repeated C) or pUC19_seq with single-stranded pUC19 DNA fragment was carried out in the annealing buffer (20 mM Tris–acetate, 2 mM MgAc₂ pH 7.6) at 95 °C for 5 min and then cooled down to room temperature gradually. The resulting immobilized double-stranded DNA strands were washed three times with washing buffer (10 mM Tris–acetate pH 7.6) before sequencing.

2.4. Pyrosequencing

Pyrosequencing was carried out on PyroMark Q96 ID system (Qiagen) using PyroGold reagents according to the manufacturer's protocol. After an initial dispensation of enzyme and substrate mixes, the sequencing procedure was accomplished by stepwise elongation of the primer strand through iterative additions of deoxynucleoside triphosphates and simultaneous detection of resulting bioluminescence emission.

3. Results and discussion

3.1. Optimization of DNA:beads ratio in direct pyrosequencing on immobilized DNA

Pyrosequencing is a relatively straightforward and user-friendly DNA sequencing technique. It has therefore been used in a variety of applications such as SNP genotyping [4,5], DNA methylation analysis and a wide range of other genetic analyses [5]. To perform a good DNA sequencing reaction either in a liquid-phase four-enzyme system or a microfluidic three-enzyme system with fixed amount of substrate and enzymes, it is important to work out the optimized concentration of DNA template in the reaction mixture. Here, we center on the optimization of suitable DNA concentration on beads to favor direct sequencing as immobilized DNA could be sequenced using both liquid and solid-phase systems. Fig. 1 illustrates signals from pyrosequencing on different amounts of immobilized DNA. As shown, an increase in the amount of DNA immobilized on 1.25×10^6 Sepharose beads leads to increasing signal peaks (Fig. 1A–E). However, the pyrograms with expected sequencing peaks are only in the range of DNA concentration (0.3125–15 pmol, Fig. 1A–D). Although, with higher DNA concentration, higher signal peaks were detected for the first few nucleotides but decreasing signal peaks were observed for the following nucleotides (Fig. 1E). This could be explained by the fact that while

Table 1
Oligonucleotide sequences.

Name	Nucleotide sequence (5'→3')
SG_seq	GGACTATAAGATACAGCGTT
SG_122	Biotin-TTCCAAGGAGTCTACGAACGCCTGGTATCTTTATAGTCCATC
SG_122m	Biotin-5'-AATCAGAGTCTACGAACGCCTGGTATCTTTATAGTCCATC
SG_10A	Biotin-Biotin-5'-CATGAGCGTTACGACTTTTTTTTACTGAACGCCTGGTATCTTTATAGTCCA
SG_10T	Biotin-ATCAAGGCCTTCATGAGCGTTACGACAAAAAAGTACAACGCCTGGTATCTTTATAGTCCA
SG_repeated C	Biotin-ATTGGGGGGTTGGGGGATGGGGTCTGGGTATGGATGTAAACG CCTGGTATCTTTATAGTCCA
pUC19_seq	TAAGGGATTTTGGTCATGAG
pUC19_F	Biotin-CTCCCGTATCGTAGTTATCT
pUC19_R	CACGTTAAGGGATTTTGGTC

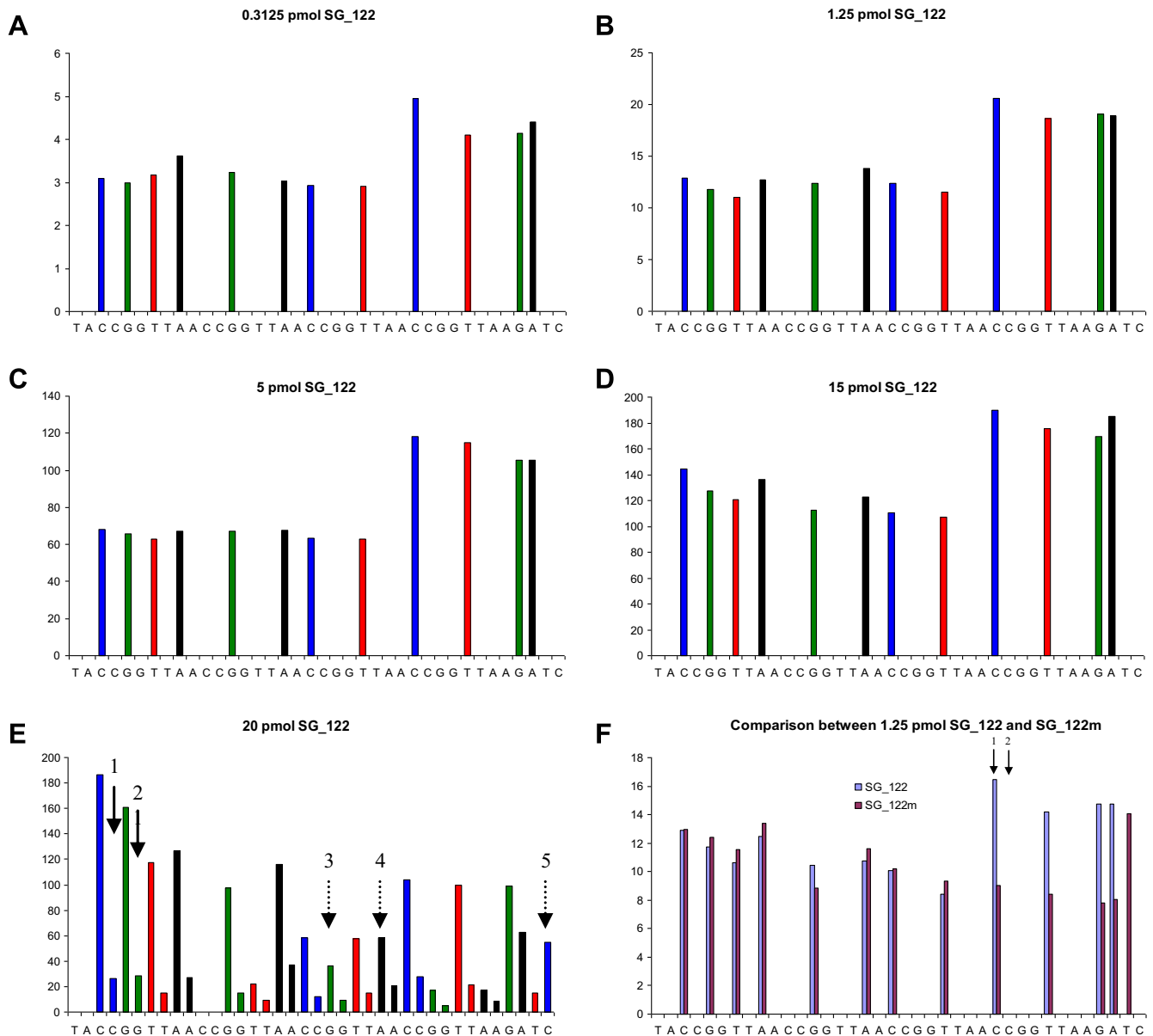


Fig. 1. The effect of DNA:beads ratio on the pyrosequencing data quality. (A–E) Pyrosequencing data for 0.3125–20 pmol of biotinylated SG-122 immobilized on 1.25×10^6 streptavidin-coated Sepharose beads. (F) Comparison of pyrosequencing data for 1.25 pmol SG_122 and SG_122m. The order of nucleotide addition is indicated on the bottom of the traces. The read sequence is indicated above the traces.

sequencing higher amount of DNA template, the fixed amount of APS, present in the sequencing reaction, is consumed quickly. It has been documented that pyrosequencing reactions generate SO_4^{2-} and oxyluciferin which have been recognized to have inhibitory effect on ATP sulfurylase and luciferase, respectively [4]. Hence, higher quantity of DNA accumulates more by-products during pyrosequencing and therefore causes significant decrease in signal intensities after first few correct incorporations. Fig. 1E also shows plus frame shift during pyrosequencing as highlighted with dotted arrows #3–5. This is due to that fixed quantity of apyrase can not completely degrade the extra amount of ATP and dXDP generated during pyrosequencing with higher amount of DNA. In addition, with such high quantity of DNA template, incomplete incorporation of nucleotide could be observed as highlighted with arrows #1 and 2 in Fig. 1E which might be due to lack of DNA polymerase in the reaction. Therefore, the results indicate that pyrosequencing can be carried out directly on 0.3125–15 pmol of immobilized DNA in a liquid-phase pyrosequencing system.

It is also observed that the signal due to double correct incorporations (e.g., 2C, 2G, 2T or 2A) is less than twice the previous single ones. As can be seen in Fig. 1F, the signal intensity due to the first double C incorporation (17, marked by an arrow #1) is less than twice the signal due to previous two correct single C incorporations (12.9 and 10.5, respectively). Interestingly, after dispensing C again (arrow #2), no peak was detected, indicating that previous double C has been incorporated completely. This suggests that while the incorporations take place as expected, there is a continuous loss in signal intensities when sequencing towards increased read-length. For further investigation, pyrosequencing on SG_122m (modified SG_122) was carried out. In SG_122m, the four double nucleotides in SG_122 have been replaced by single nucleotides addition with a double T at the end. As shown in Fig. 1F, the peak heights of double nucleotides in SG_122 were about twice the height of those single nucleotides in SG_122m, even though the overall peak heights drop with increased read-length. The results reinforce that accumulation of by-products during pyrosequencing

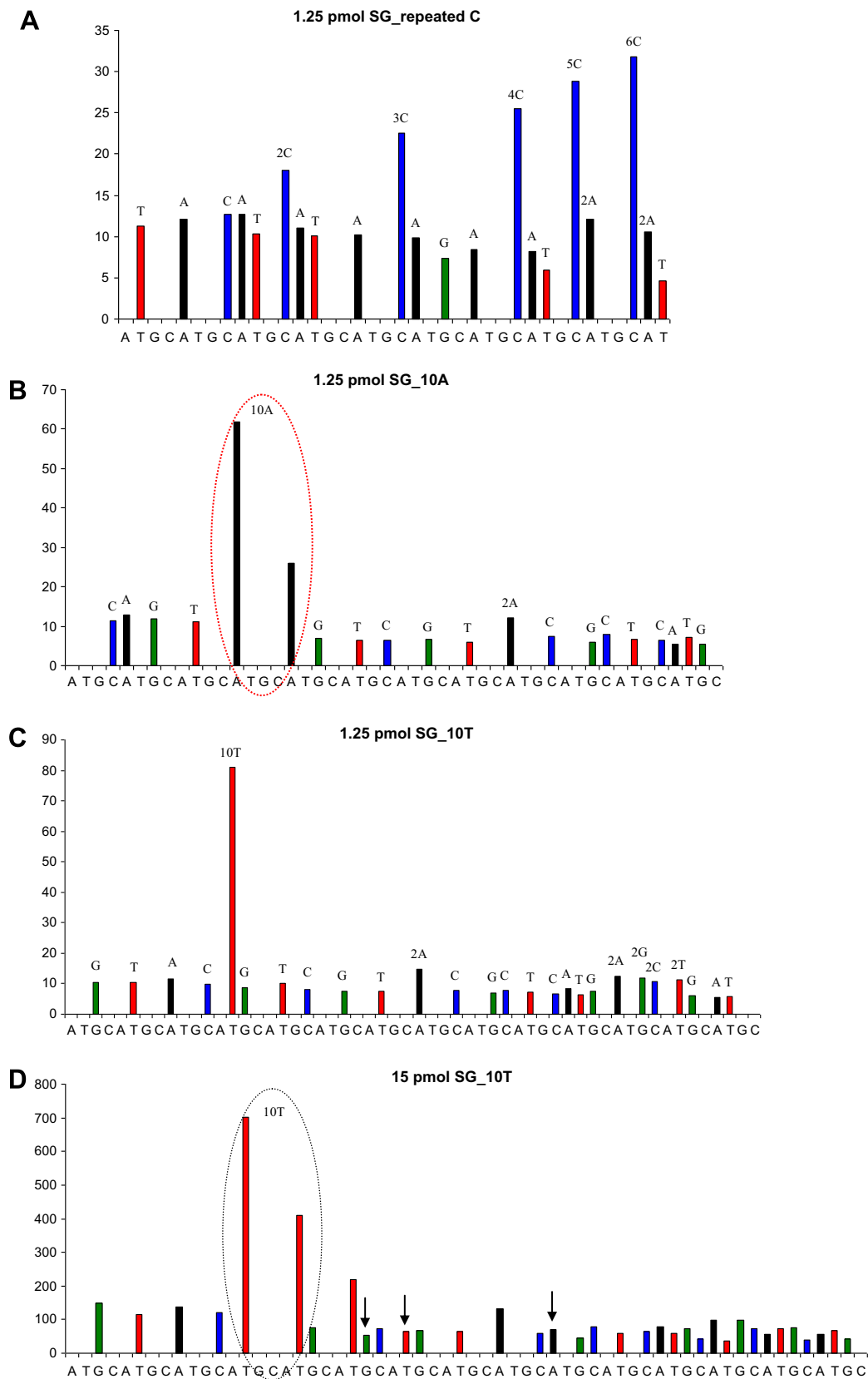
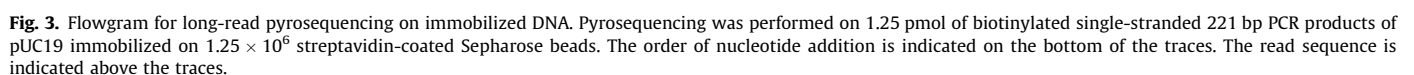


Fig. 2. Pyrosequencing on DNA templates containing various kinds of nucleotides in homopolymeric stretches. (A) Pyrosequencing on SG_repeated c (containing a poly(c) track with 1C, 2C, 3C, 4C, 5C, and 6C). (B) Pyrosequencing on SG_10A (containing a poly(A) track with 10 A). (C) Pyrosequencing on SG_10T (containing a poly(T) track with 10 T). Pyrosequencing was performed on 1.25 \times 10⁶ streptavidin-coated Sepharose beads. (D) Pyrosequencing on 15 pmol immobilized SG_10T. The order of nucleotide addition is indicated on the bottom of the traces. The read sequence is indicated above the traces.

detected after 10T when pyrosequencing was carried out using 15 pmol of DNA (Fig. 2D). Our results suggest that to acquire correct nucleotide incorporation for homopolymeric regions with 10 identical nucleotides, ≤ 5 pmol of DNA should be used for pyrosequencing. In addition, it was successfully demonstrated that pyrosequencing on ≤ 5 pmol immobilized DNA can determine as many as 10 identical adjacent bases incorporated at a time from signal intensities (10A, 10T).

Studies have shown that incorporation of dATP- α -S in T-homo-polymeric regions results in uneven sequence signals and reduced sequence quality directly downstream of such homopolymers [12]. This may be due to the inefficiencies of Klenow polymerase to incorporate completely and apyrase to degrade the analog dATP- α -S completely, resulting in asynchronous and ambiguous sequence peak signals. Interestingly, in the present work (Fig. 2), through the optimization of DNA:beads ratio, the immobilized DNA with homopolymer T (complementary nucleotide A in nucleotide dispensation) has been sequenced successfully. Although the signal heights decreased notably in sequence downstream as compared to that before the 10A region, neither plus nor minus frame shift were detected. In addition, it is also noted that 10 homopolymers of base A could not be fully incorporated instantly with single dispensation of dATP- α -S. This is in agreement with the known fact that the rate of dATP incorporation by Klenow polymerase is very low [17]. Hence, an extra dispensation of dATP- α -S has been introduced to confirm the completion of incorporation of 10A homopolymeric region. In addition, T7 DNA polymerase might be an alternative enzyme to be applied in pyrosequencing as it has shown more adequate for the incorporation reaction with dNTPs than Klenow polymerase especially for the homopolymeric T regions [18].

gions [18].



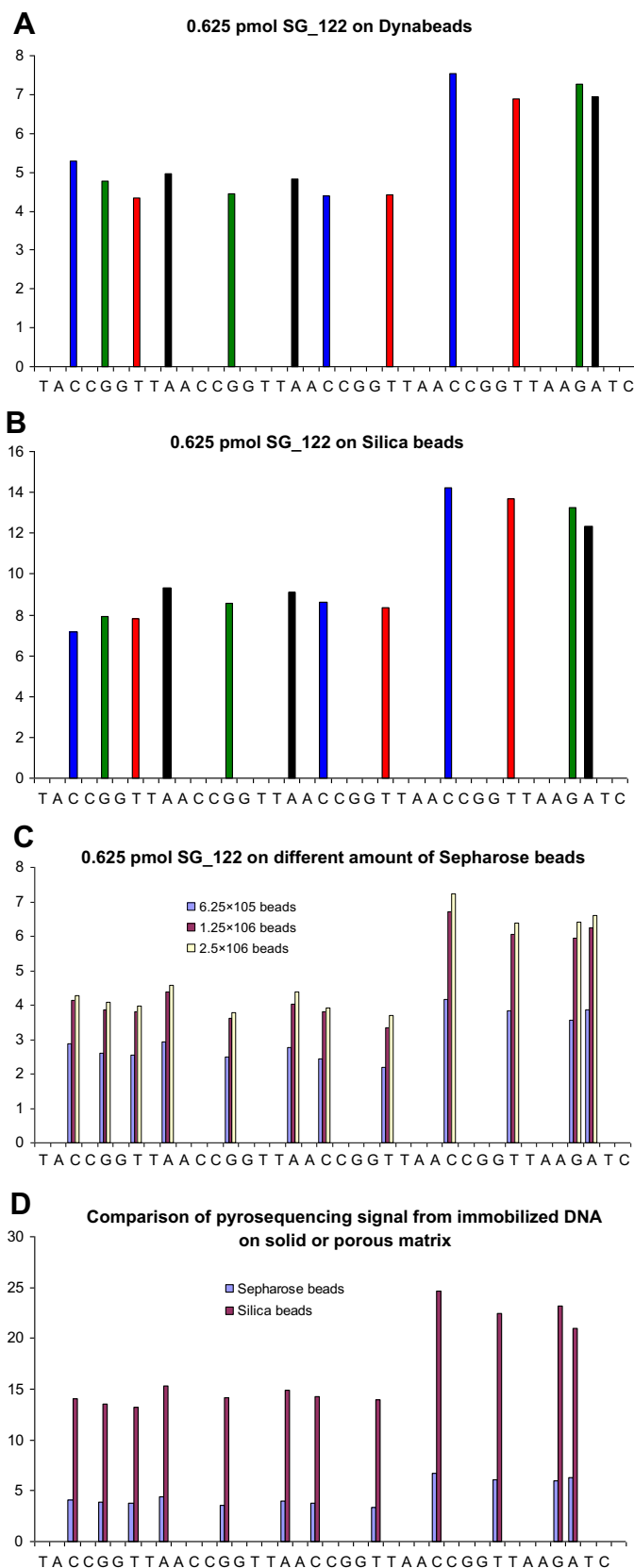
3.3. Long-read DNA sequencing

Long-base reading is desirable in applying pyrosequencing to genome sequencing, microbial typing, and resequencing. The performance of pyrosequencing on optimized amount of immobilized DNA for long-base reading was investigated. As shown in Fig. 3, although the signal intensities to the incorporated base decreased gradually with increasing read-length, more than 80 bases could be successfully sequenced and analyzed in a liquid-phase pyrosequencing system. The decreasing signal intensities in the long-base region might be caused by multiple factors. First, the exonuclease activity of DNA polymerase will cause primer degradation in the long read sequencing which may lead to out of phase sequencing. Second, an accumulation of byproduct during sequencing [4], and third, the dilution produced by the continuous addition of nucleotides. For example, sequencing of 80 bases requires about 100 nucleotide additions, which gives a 40% dilution of the assay volume. The serial dilution due to nucleotides dispensation lowers the enzyme concentrations thereby decreasing their efficiency. With immobilized DNA, by introduction of solid-phase pyrosequencing in a microfluidic system, the dilution effect of nucleotide addition at each reaction cycle and byproduct accumulation in the current plate format could be eliminated.

3.4. Pyrosequencing on DNA immobilized on different matrix

To investigate direct pyrosequencing on DNA sample immobilized on different medias, oligo SG₁₂₂ was immobilized onto 30 micron streptavidin-coated silica beads and 2.8 micron super paramagnetic Dynabeads. As shown in Fig. 4, Pyrosequencing can effectively be carried out on the DNA template immobilized on silica beads or Dynabeads as well. Generally, the streptavidin-coated magnetic beads (Dynabeads) are used for preparing primed ssDNA template for pyrosequencing. The results from this study show that optimal quantity of immobilized biotinylated ssDNA on magnetic beads can be directly annealed with sequencing primer and applied for pyrosequencing. In addition, as paramagnetic beads (2.8 μm) and silica beads (30 μm) has different particle sizes, our results suggest that pyrosequencing on immobilized DNA can be widely carried out with different solid-phase microfluidic systems.

Though pyrosequencing is based on a sensitive bioluminescent assay, it still requires DNA templates at a picomole level. This is due to the detection limit of the enzymatic assay. In many cases, especially in clinical research, a much lower amount of DNA is available. Therefore it is essential to address the DNA detection limit in pyrosequencing. Fig. 4C shows the performance of pyrosequencing when immobilizing same amount of biotinylated DNA on different number of streptavidin-coated Sepharose beads. As can be seen, when same amount of DNA is distributed on more number of beads, or when there is less DNA per bead, the overall signal intensity increases. The system with more beads has two diffusion effects; first the additional surface area provided by higher number of beads results in more uniform distribution of DNA on the beads. The even distribution of DNA creates more efficient diffusion-based mixing of DNA with the substrate and enzymes where efficient mixing is crucial for pyrosequencing [19]. Second, the presence of more beads leads to the occupancy of more volume in the reaction chamber; bringing down the diffusion efficiency in consequence. As shown in Fig. 4C, after first bead-number increment, the rapid increase of the signal indicates that the first diffusion effect is predominantly high. However, after the second bead increment, the signal increase is not very obvious suggesting there is a limit to increase pyrosequencing signal even with utilizing more Sepharose beads to bind same amount of DNA. This limitation can be overcome by using porous beads with more surface area and effective diffusion coefficient. The study using porous



silica beads was carried out to support our hypothesis on the effect of diffusion efficiency on the pyrosequencing signal intensity. Although silica beads with similar size of Sepharose beads will occupy the same volume in the reaction chamber, due to the pores the overall surface area will be higher. In addition, pores all over the beads favor the diffusion-based mixing of substrate and enzymes with well-distributed DNA template. As expected, DNA on silica beads displayed a 3.5-fold higher signal intensity as compared to that produced from DNA on Sepharose beads (Fig. 4D) indicating lower amount of DNA is required to produce distinguishable pyrosequencing signal. Taken together, these findings imply that porous silica beads served as a better matrix for DNA immobilization in pyrosequencing.

In conclusion, we have successfully performed optimization of the amount of DNA for pyrosequencing which allow the sequencing to be carried out for homopolymeric regions containing 10 identical nucleotides. The introduction of porous silica beads for DNA binding significantly increased sequencing detection limit. Hence, running pyrosequencing on optimal quantity of DNA immobilized on silica beads in a microfluidic system could offer several advantages such as improved detection limit, well distinguishable signal for homopolymeric DNA regions and possibly increased read-length.

Acknowledgments

We thank Dr. Heng Hang Tsai and Dr. Yijun Ruan for valuable discussion and providing the facility. We are thankful to Dr. Masafumi Inoue for allowing us to use the Pyromark ID at the Experimental Therapeutics Centre, Agency for Science, Technology and Research, Singapore. We acknowledge funding support by the Joint Council Office (JCO) Grant (CCOG01_002_2008) at the Agency of Science, technology and Research (A STAR), Singapore.

References

- [1] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bembien, J. Berka, M.S. Braverman, Y.J. Chen, et al., Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437 (2005) 376–380.
- [2] S.T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown, Toward the 1000 dollars human genome, *Pharmacogenomics* 6 (2005) 373–382.
- [3] A.A. Antipova, T.D. Sokolsky, C.R. Clouser, E.T. Dimalanta, C.L. Hendrickson, C. Kosnopo, C.C. Lee, L. Zhang, A.P. Blanchard, K.J. McKernan, Polymorphism discovery in high-throughput resequenced microarray-enriched human genomic loci, *J. Biomol. Technol.* 20 (2009) 253–257.
- [4] F. Mashayekhi, M. Ronaghi, Analysis of read length limiting factors in Pyrosequencing chemistry, *Anal. Chem.* 362 (2007) 275–287.
- [5] J. Tost, I.G. Gut, DNA methylation analysis by pyrosequencing, *Nat. Protoc.* 2 (2007) 2265–2275.
- [6] A. Ahmadian, B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyrén, M. Uhlén, J. Lundeberg, Single-nucleotide polymorphism analysis by pyrosequencing, *Anal. Biochem.* 280 (2000) 103–110.
- [7] C. Agaton, P. Unneberg, M. Sievertzon, A. Holmberg, M. Ehn, M. Larsson, J. Odeberg, M. Uhlén, J. Lundeberg, Gene expression analysis by signature pyrosequencing, *Gene* 289 (2002) 31–39.
- [8] J. Doostzadeh, S. Shokralla, F. Absalan, R. Jalili, S. Mohandessi, J.W. Langston, R.W. Davis, M. Ronaghi, B. Gharizadeh, High throughput automated allele frequency estimation by pyrosequencing, *PLoS ONE* 3 (2008) e2693.
- [9] Z. Chaoyong, O. Jacob, H. Anders, Per Eriksson Allele-specific MMP-3 transcription under in vivo conditions, *Biochem. Biophys. Res. Commun.* 348 (2006) 1150–1156.
- [10] M. Ronaghi, M. Uhlén, P. Nyren, A sequencing method based on real-time pyrophosphate, *Science* 281 (1998) 363–365.
- [11] M. Ronaghi, M. Nygren, J. Lundeberg, P. Nyren, Analyses of secondary structures in DNA by pyrosequencing, *Anal. Biochem.* 267 (1999) 65–71.
- [12] B. Gharizadeh, T. Nordström, A. Ahmadian, M. Ronaghi, P. Nyren, Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer, *Anal. Biochem.* 301 (2002) 82–90.
- [13] A. Russom, N. Tooke, H. Andersson, G. Stemme, Pyrosequencing in a microfluidic flow-through device, *Anal. Chem.* 77 (2005) 7505–7511.
- [14] J.D. Parvin, R.J. McCormick, P.A. Sharp, D.E. Fisher, Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor, *Nature* 373 (1992) 724–727.
- [15] R.I. Richards, G.R. Sutherland, Simple repeat DNA is not replicated simply, *Nat. Genet.* 6 (1994) 114–116.
- [16] M. Margulies, M. Egholm, W.E. Altman, et al., Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437 (2005) 376–380.
- [17] G.H. Zhou, N. Gotou, T. Kajiyama, H. Kambara, Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI), *Nucleic Acids Res.* 33 (2005) e133.
- [18] B. Gharizadeh, J. Eriksson, N. Niyruzad, T. Nordstrom, P. Nyren, Improvements in pyrosequencing technology by employing sequenase polymerase, *Anal. Biochem.* 330 (2004) 272–280.
- [19] A. Agah, M. Aghajan, F. Mashayekhi, S. Amini, R.W. Davis, J.D. Plummer, M. Ronaghi, P.B. Griffin, A multi-enzyme model for pyrosequencing, *Nucleic Acids Res.* 32 (2004) e166.