Title
Integrated Microfluidic Systems for Genetic Analysis

Author
Cronier, Samantha Anne

Publication Date
2011

Peer reviewed|Thesis/dissertation

Integrated Microfluidic Systems for Genetic Analysis

By

Samantha Anne Cronier

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with the University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Richard A. Mathies, Chair

Professor Tejal Desai

Professor George Sensabaugh

Fall 2011

Copyright Page

Abstract

Integrated Microfluidic Systems for Genetic Analysis

by

Samantha Anne Cronier

Joint Doctor of Philosophy in Bioengineering

University of California, Berkeley

and

University of California, San Francisco

Professor Richard A. Mathies, Chair

Recent work in the areas of microfluidic technology and innovative new chemistries have made possible tremendous improvements in our ability to determine the nucleotide sequence of DNA. These advances are already changing the way doctors diagnose and treat human disease and enabling scientists to undertake genetic studies never before possible. In this dissertation I provide a brief history of DNA sequencing methods as well as a general description of microfluidic technologies, in particular those used for genetic analysis. Next, I describe my dissertation research on the development of a highly efficient fully integrated microfluidic platform for Sanger DNA sequencing, including automated thermal cycling, purification, concentration and in-line injection of the extension fragments for microchip capillary electrophoresis separation. The two-layer glass device that I developed features two independently operated valve-free systems, comprised of a 200 nL thermal cycling reactor with resistive temperature detector, a 1.2 nL *in situ* photopolymerized oligonucleotide affinity capture gel for post reaction clean-up and inline injection, and an 18-cm long capillary electrophoresis channel for separation. Integration of the efficient photopolymerized affinity gel capture allows sequencing from only 350 - 500 attomoles of starting DNA template. Using this device, I was able to sequence 507 ± 31 bases at 99% accuracy. In addition, I show that this method is compatible with single cell genetic analysis techniques (SCGA) by sequencing from the small amounts (~100 attomoles) of amplified DNA bound to agarose microbeads that can be produced from single cells.

This dissertation concludes with a discussion of the future of DNA sequencing and the feasibility of performing single cell DNA sequencing using the Microbead Integrated DNA Sequencing (MINDS) method.

Table of Contents

List of Figures

Dedication

This dissertation is dedicated to my family:

my adventurous, curious seafaring parents, Ron and Sandy
my delightful talented sister, Brigitte
and finally,
my brilliant, compass-headed patient husband Mike.

Thank you for your steadfast love and support.

Chapter 1. DNA Sequencing and Microfluidics

Introduction

The past thirty years have seen dramatic advances in DNA sequencing technology, driven by funding from the National Institutes of Health and the Department of Energy. For the first time, large-scale studies examining the genomes of multiple individuals – at great depth and coverage – allow us to examine individual genomes,[1] to identify and classify new species,[2] to characterize populations,[3] and to determine all mRNAs being expressed in sample at an instant in time.[4] Possibly the most important application for the future will be the sequencing and comparison of many human genomes, which has the potential to reveal much about human genetic variation and genome-related diseases, such as cancer. DNA sequencing is critical for basic biological research, yet it is still an expensive and time consuming endeavor to determine the complete sequence of a complex genome, such as the three billion base pair human genome (or six billion base pair *diploid* genome) or the 2.5 billion base pair *Zea mayes* genome. Emerging technologies promise to deliver extremely high throughput sequencing runs, but lack the quality required for *de novo* sequencing, in particular for resolving structural variations (SVs) or regions of the genome rich in repeating nucleotides, such regions as are major sources of genetic variation and disease origins. This chapter covers the history of DNA sequencing, including an overview of competing technologies, and introduces the goal for my dissertation work: the development of a fully integrated microfluidic single-molecule Sanger sequencing device.

History of Sanger Sequencing

Fred Sanger and Alan R. Coulson reported their method for accurately reading the sequence of DNA in 1977, providing a conceptually simple and practical method for reading the genetic code.[5] The technique, illustrated in Figure 1, and described in detail below, involved the use of a DNA polymerase to generate multiple "extension fragments" from an original template and then identifying the sequence by separating the individual bands using gel electrophoresis. The so-called Sanger sequencing method proved to be a robust and automatable approach to DNA sequencing, gaining popularity over the Maxam-Gilbert approach due the use of fewer toxic chemicals and less radioactive material. Over the next 30 years, there were significant advances in the read length, throughput, accuracy and automation of the pipeline due to significant work by many contributors.

The fundamental chemistry of the Sanger technique has changed little since its invention in 1977. The key is the use of an unnatural dideoxynucleotide triphosphate (ddNTP), which lacks the –OH group required for DNA chain extension. A typical reaction, illustrated in Figure 1, contains a mixture of natural deoxynucleotide triphosphates (dNTPs), ddNTPs, template DNA, primer DNA and polymerase enzymes. The reactants are thermally cycled between a primer-annealing step (50-55 ˚C), an extension step (72 ˚C), and a denaturing step (95 ˚C), in which the copied DNA dehybridizes from the template. The result is the generation of many truncated copies of the original DNA template molecule. All the "extension fragments" begin with the same sequence of DNA, but are randomly terminated at differing locations based on the stochastic incorporation of ddNTPs by the polymerase enzyme. Thus, for each base in the sequence there is an extension fragment with

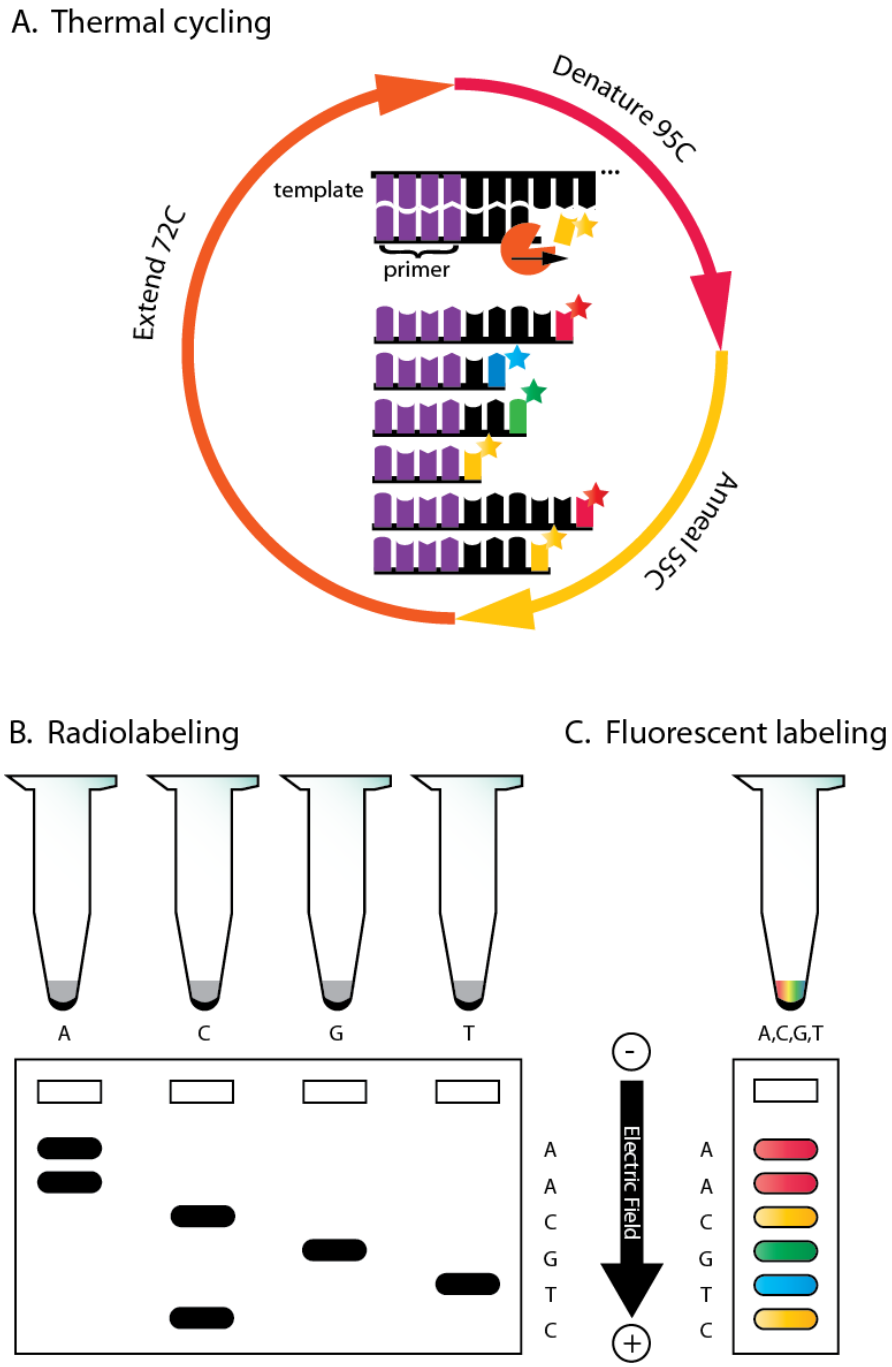

Figure 1. Sanger sequencing reaction. (A) The key feature of the Sanger sequencing reaction is the use of labeled ddNTPs, known as terminators. During thermal cycling, a polymerase generates multiple truncated copies of an original template DNA molecule. (B) Initially, DNA bands were visualized by using radioactively labeled nucleotides and the different nucleotides were distinguished by running the reactions in four separate tubes and four separate lanes in a slab gel. (C) By labeling the ddNTPs with four spectrally unique fluorescent dyes, it was possible to perform the reaction in a single tube and the separation in a single lane. This was a major advancement toward the goal of automation.

corresponding size. The sequence can then be "read" by separating the DNA fragments based on their sizes using gel electrophoresis.

Initially, such reactions were performed in four separate vessels, one for each of the four ddNTPs (A, C, G and T), and then size-separated in four separate electrophoresis lanes of a polyacrylamide slab gel, as illustrated in Figure 1B. The use of radioactively labeled nucleotides meant the bands could be visualized by radiography, and the sequence decoded. Fluorescently labeled primers,[6,7] with a specific fluorophore for each of the A, C, G and T reactions, allowed all four bases to be run in a single gel lane and later, fluorescently labeled terminators[8] allowed the reactions to take place in the same tube. Dye terminator chemistry has additional advantages. Because any primer site can be used, it is a more flexible technique, not requiring synthesis of four dye-labeled primers. Second, occasional termination events resulting from the incorporation of a dNTP do not result in a labeled fragment, whereas in dye primer chemistry, this results in a spurious peak. Still, fluorescent dyes created other problems, such as differences in electrophoretic mobility imparted by the different fluorophores as well as unequal peak heights due to differences in molar absorption at the excitation wavelength employed. These problems were overcome by the use of energy transfer dyes with matched electrophoretic mobilities.[9]

Other advances in automating Sanger sequencing came from improved electrophoretic separation methods. The original cross-linked polyacrylamide slab gels were replaced by gel-filled capillaries with automatically replaceable non-cross-linked matrices that significantly improved speed and efficiency.[10–12] Covalent[13] and non-covalent[14] surface coatings were developed to combat electroosmotic flow (EOF) which occurs with non-crosslinked gels in capillaries. In practice, it is not possible to sequence a whole chromosome directly because the length of a read is limited by the electrophoresis step. Above a certain molecular weight, DNA migration through a gel matrix is governed by reptation, in which the DNA molecules preferentially elongate along the direction of the electric field as they move through the gel. In this regime, the resolving power (ability to separate a fragment of N bases from N+1 bases) diminishes.[15] The quality of the sequence produced is described by the read-length and associated degree of certainty on the call of each base. Read lengths of as long as 1300 bases have been reported,[16] but typical values for commercial automated sequencers are in the range of 700 bases in 2 hrs with "raw" per-base accuracies of 99.999%.

The advances described above were driven by the need for faster, cheaper and more automated sequencing for the Human Genome Project. This 13-year mega project costing $2.7 billion, relied entirely on Sanger sequencing. Researchers were able, for the first time, to decode an entire human genome, 3 billion basepairs (bp), with the draft published in 2001.[17] Over the course of the project, the state of the art was advanced so significantly that the project was completed ahead of schedule. Astonishingly, at the outset of the HGP, the aim was to sequence 500 Mb/year at <$0.25/base, but the throughput achieved by the end of the project, in 2002, was >1400 Mb/year at only $0.09/base. The result was the sequencing of 99% of the gene-containing portion of the human sequence, finished to 99.99% accuracy. In fact, this is the only "finished-grade" sequence to date.[18] The data

generated by the HGP is available to the public and stored, along with complete or partial genomes of 100,000 species, in Genbank, the NIH database of annotated DNA sequences. Figure 2 plots the amount of sequence data stored in Genbank, which has doubled every 18 months since its inception in the 1980s, an exponential growth rate matching Moore's Law for integrated circuits.[19,20] As of April 2011, Genbank contained an astounding 124 billion bases.

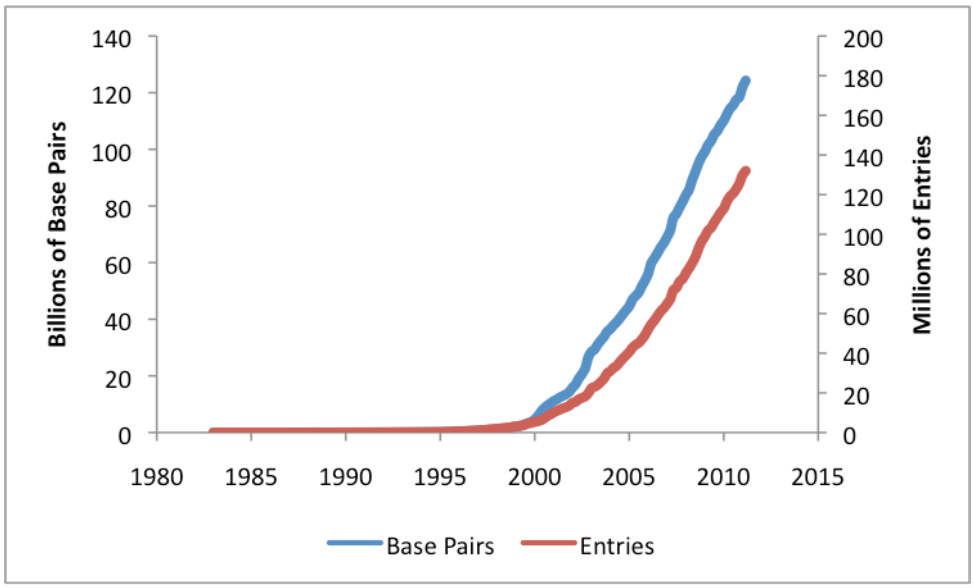

Figure 2. The number of base pairs stored in Genbank has approximately doubled every 18 months since 1982. ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

Following the completion of the HGP, incentive for the development of new sequencing methods continued to build. First, despite the remarkable success of the HGP, sequencing a human-sized genome remained a monumental effort. Second, with the complete sequence of the *Homo sapiens* genome available, the ability to make use of short-read technologies became more feasible, as the existing sequence could be used as a reference with which to align the shorter reads. Additionally, various molecular methods under development used sequencing as a means to study a broad range of biological phenomena, such as gene expression, novel single nucleotide polymorphism (SNP) discovery, chromosomal conformation, epigenetic modifications, and protein-DNA interaction. Finally, the most easily accomplished and high-payoff advances to the Sanger sequencing pipeline had largely been exploited by 2003, but advances in other fields, such as microscopy, nucleotide biochemistry, surface chemistry, polymerase engineering, data storage and computation opened many avenues for innovation.

Recently Developed Sequencing Technologies

The following section provides an overview of more recent technologies, including those which have been developed to the point of commercialization, and those which are still under development.

Table 1. Comparison of next generation sequencing technologies. Data for this table comes mainly from a review by Metzker[18], with additional information from Pushkarev[33] and the Pacific Biosciences website.

Platform	Library/ Template Preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost	Price of Reagents	Pros	Cons	Best Applications	Refs
Roche/454's GS FLX Titanium	Fragment, mate-pair, emPCR	Pyro-sequencing	400	0.35	0.45	$500,000	~$4000	Relatively fast, longer reads allow better alignment in repetitive regions	High reagent cost; high error rates in homopolymer regions	de novo assembly of small genomes (insect, bacteria); medium scale exome capture; 16S in metagenomics	Metzker, 2010
Illumina/ Solexa's Genome Analyzer	Fragment, mate-pair, solid-phase	Reversible Terminators	75 to 100	4 or 9	18 or 35	$540,000	~$4000	Most widely used NGS platform	Low multiplexing capability	Variant discovery by whole-genome resequencing or whole-genome exome capture; gene discovery in metagenomics	Metzker, 2010
Life/APG's SOLiD 3	Fragment, mate-pair, emPCR	Cleavable Probe Sequencing by Ligation	50	7 or 14	30 or 50	$595,000	~$4000	Two-base encoding allows for error correction	Long run times	Variant discovery by whole-genome resequencing or whole-genome exome capture; gene discovery in metagenomics	Metzker, 2010
Polonator 0.007	Mate pair only; emPCR	Non-cleavable Probe Sequencing by Ligation	26	5	12	$170,000		Least expensive platform; open source to adapt alternative NGS chemistries	Users must do quality control themselves; shortest of NGS read lengths	Bacterial genome resequencing for variant discovery	Metzker, 2010
Helicos BioSciences HeliScope	Fragment, mate-pair, single molecule	Reversible Terminators	32	8	37	$999,000	~$18000	Unbiased representation of fragments for genome and sequencing based applications	High error rates compared with other reversible terminator platforms	Seq-based methods	Pushkarev, 2009
Pacific Biosciences	Fragment only, single molecule	Real-time	964	1	unknown	unknown		Best read length	Highest error rates compared to other NGS chemistries	Full-length transcriptome sequencing; capable of discovery large structural variants and haplotype blocks	Metzker, 2010, pacificbiosciences.com

Recently developed technologies that aim to rapidly sequence vast amounts of DNA at low cost are referred to as next-generation sequencing (NGS) technologies, summarized in Table 1. Among these, most aim to improve on the Sanger methodology by simplifying and parallelizing the template preparation steps and performing the sequence determination in a highly parallelized fashion. These methods are limited to significantly fewer bases (20-400 per read) than Sanger sequencing (500-1000 bases) , which complicates the data assembly required to reconstruct an entire genome and limits the ability to sequence regions high in repeats or with structural variations.

Although each platform has a unique methodology, the general template preparation and sequence determination workflows are similar. First, genomic DNA is randomly fragmented and amplified using a PCR-based method such as polonies, emulsion PCR or bridge PCR, the result of which is a library of amplified templates which are spatially separated on a flat substrate. The sequence determination is accomplished by alternating enzymatic reactions (ligase or polymerase) and high resolution imaging of the substrate.

The nuances, advantages and disadvantages of commercially available platforms are discussed below.The first next generation method to be commercialized was the Roche/454 pyrosequencing system illustrated in Figure 3.[21] The library of template DNA can be generated by any method that produces a population of short (300 to 800 bp) adapter-flanked fragments (i.e., DNA fragments of unknown sequence with short adapters of known sequence on either end). Clonal amplification is accomplished by statistically dilute emulsion PCR, the products of which are captured on 28 μm beads. The un-tethered strands of DNA are dehybridized and the ssDNA-containing beads are selectively enriched by hybridization (beads containing no DNA wash away). Universal primers are annealed and beads are distributed into individual wells of a size such that only one bead can be accommodated in each. Smaller beads containing other required enzymes are also packed in the wells. One side of the well-containing plate is used as a flow cell to deliver reagents and the other side contains fiber optic connections to a CCD detector. Sequencing is accomplished by sequentially flushing unlabeled dNTPs over the wells, one base at a time. The incorporation events are detected by a chain of enzymatic reactions whereby pyrophosphate (PPi) is converted to light.

The biggest limitation of the 454 method is the difficulty in correctly calling homopolymer stretches, such as AAA or CCC. Because there is no capping between incorporations of each nucleotide, the difference between one or more of the same base must be measured based on the intensity of the signal, which is more difficult to distinguish than the presence or absence of an incorporation event. Because of this, the most common forms of errors are insertions or deletions, rather than substitutions. The most significant advantage of the 454 method is the relatively long read-length (330 bases/read) as compared to other next generation sequencing strategies, which out-weighs the relatively higher cost per base for some applications. The throughput is about 400 Mb/run.

Currently, the most widely used NGS platform is the Illumina/Solexa Genome Analyzer. This four-color cyclic reversible termination-based method (Figure 4) was based on the work of Turcatti and colleagues and the merger of four companies: Solexa, Lynx Therapeutics, Mantela Predictive Medicine and Illumina. Template preparation can be done by any method that produces fragments up to several hundred bases long with universal adapters on both ends. Amplification is accomplished by bridge PCR on a solid substrate, which generates ~1000 molecules per cluster and several million clusters in each of eight independent lanes. Sequences are determined by flowing a mixture of the four fluorescently labeled reversible terminators and imaging with total internal reflection fluorescence (TIRF) in four colors (using two lasers). After imaging, the clusters are regenerated for the next cycle by chemically cleaving the terminating moiety. The read-length, of about 75 bases, (100 with mate-pairs) is limited by signal decay and dephasing. The most common errors are substitutions, and studies have indicated that there is a bias against AT-rich and GC-rich regions as well as an increase in errors when the previous base is a G. [22–24] The throughput is either 18 or 35 Gb/run.

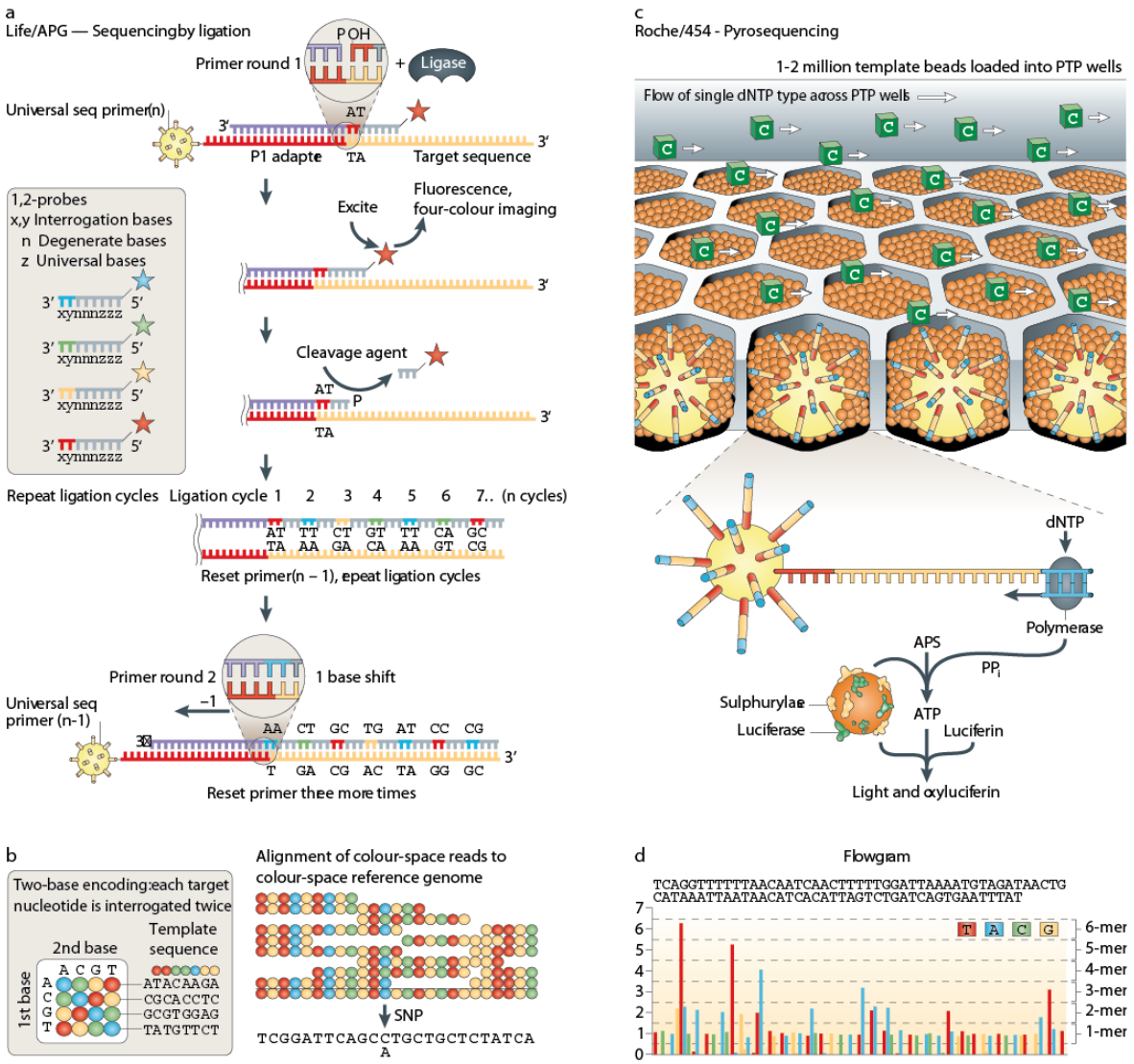

Figure 3. (A) A four-color sequencing by ligation method using Life Technologies' support oligonucleotide ligation detection (SOLiD) platform is shown. After annealing of a universal primer, a library of 1,2-probes is added and imaged in four colors. The ligated 1,2-probes are chemically cleaved with silver ions to generate a 5′-PO4 group. This cycle is repeated nine more times. The extended primer is then stripped and four more ligation rounds are performed, each with ten ligation cycles. The 1,2-probes are designed to interrogate the first (x) and second (y) positions adjacent to the hybridized primer, such that the 16 dinucleotides are encoded by four dyes (colored stars). (B) The dinucleotide labeling scheme. Each template base is interrogated twice and compiled into a string of color-space data bits. The color-space reads are aligned to a color-space reference sequence to decode the DNA sequence. (C) Pyrosequencing using Roche/454's Titanium platform. DNA-amplified beads are loaded into individual PicoTiterPlate (PTP) wells along with additional beads, coupled with sulphurylase and luciferase. The fiber-optic slide is mounted in a flow chamber and the underneath of the fibre-optic slide is directly attached to a high-resolution charge-coupled device (CCD) camera. (D) The light generated by the enzymatic cascade is recorded as a series of peaks called a flowgram. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews | Genetics (reference citation), copyright 2009. http://www.nature.com/nrg/journal/v11/n1/full/nrg2626.html

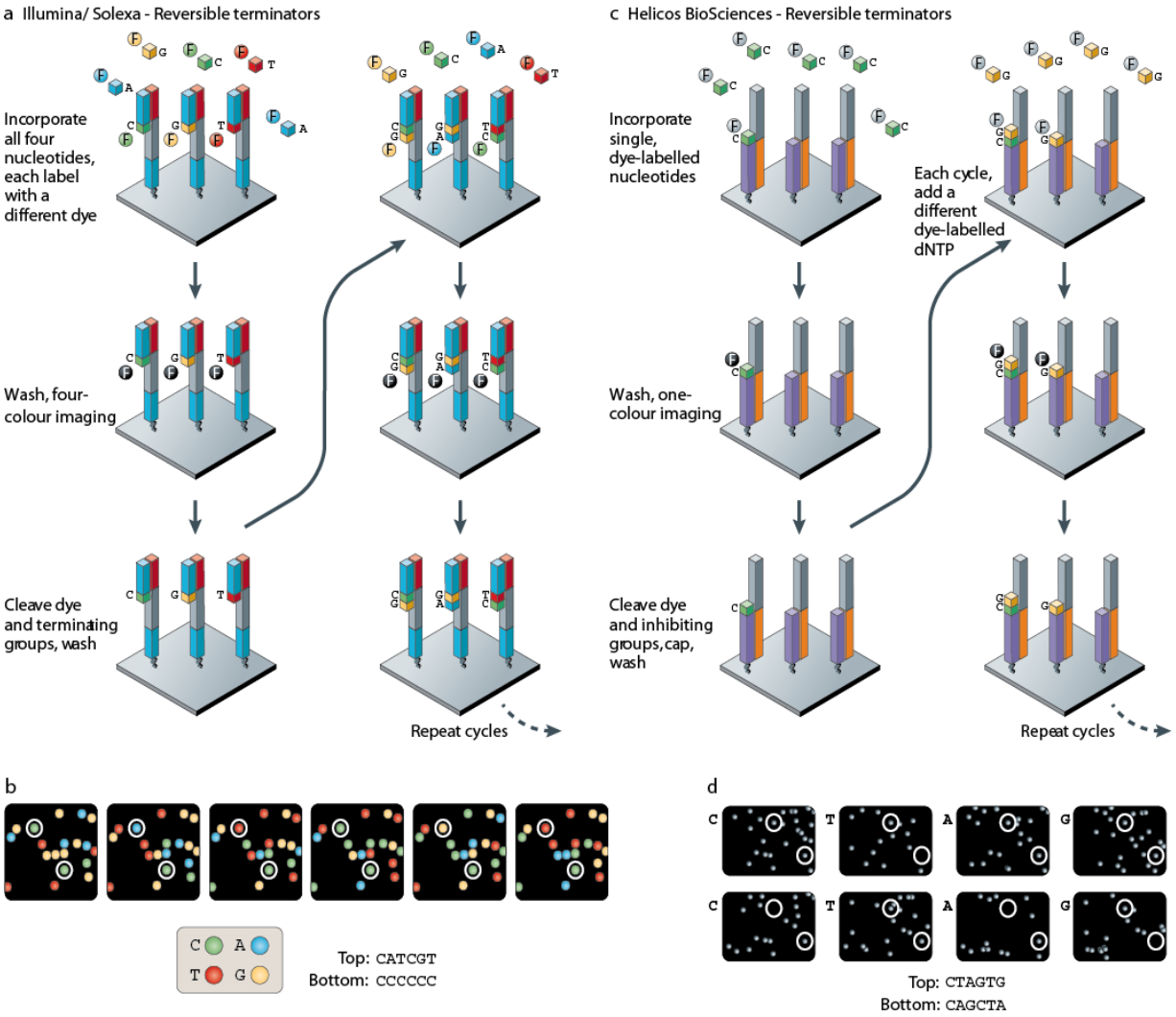

Figure 4. Illumina and Helicos methodologies. a) The four-color cyclic reversible termination (CRT) method uses Illumina/Solexa's 3′-O-azidomethyl reversible terminator chemistry using solid-phase-amplified template clusters. Following imaging, a cleavage step removes the fluorescent dyes and regenerates the 3′-OH group using the reducing agent tris(2-carboxyethyl)phosphine (TCEP). b) The four-color images highlight the sequencing data from two clonally amplified templates. c) Unlike Illumina's terminators, the Helicos Virtual Terminators are labeled with the same dye and dispensed individually in a predetermined order, analogous to a single-nucleotide addition method. Following total internal reflection fluorescence imaging, a cleavage step removes the fluorescent dye and inhibitory groups using TCEP to permit the addition of the next Cy5-2′-deoxyribonucleoside triphosphate (dNTP) analogue. The free sulphhydryl groups are then capped with iodoacetamide before the next nucleotide addition (step not shown). d) The one-color images highlight the sequencing data from two single-molecule templates. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews | Genetics (reference citation), copyright 2009. http://www.nature.com/nrg/journal/v11/n1/full/nrg2626.html

The SOLiD system of sequencing, by Life/APG is based on the work of Shendure and colleagues[25] and McKernan *et al*.[26] This ligation-based method, described in Figure 3, begins with a library generation process similar to those mentioned above. Fragments are amplified using emPCR and captured on 1 μm paramagnetic beads. The beads containing bound DNA are selectively captured and immobilized on a solid substrate in a disordered array. After annealing a universal primer, a library of 1,2-probes is added. Under appropriate conditions complementary oligos will be selectively hybridized and ligated to the universal primer. The eight base long oligos (octamers) are designed to interrogate only the first two bases adjacent to the primer and all sixteen combinations of two bases are encoded using a two-base encoding scheme. In this somewhat complicated scheme, only four dyes are used to label the sixteen two-base combinations, so the labeling is redundant, for example, AA and CC are both labeled with the blue dye in Figure 3b. Following imaging, the probes are cleaved with silver ions between positions 5 and 6 to remove the label and regenerate the phosphate group. This process is repeated, generating a color call every fifth base (positions 5, 10, 15, …) and then the primer is denatured and the cycle repeats with a primer one nucleotide shorter to query positions 4, 9, 14, …. The color calls are then arranged sequentially and converted from "color space" to nucleotide sequence by aligning to a reference sequence. The seven-day runs produce read-lengths of about 50 bases and 30 or 50 Gb/run.

Advantages of the SOLiD system include the inherent error correction of the two-base encoding, which essentially interrogates each base twice. Substitutions are the most common error types, and as in the Illumina system, there is an underrepresentation of AT- and GC-rich regions.

In contrast to the previous systems, the HeliScope platform, in Figure 4, (by Helicos BioSciences) is a single molecule method, requiring no amplification step. Libraries are generated by random fragmentation followed by poly-A tailing. The fragments are randomly hybridized to a poly-T-coated substrate. Fluorescently labeled nucleotides are cyclically added and incorporated in a sequence-dependent manner by a polymerase. Chemical cleavage of the dye follows highly sensitive imaging tiling the entire surface after each base addition. Several hundred cycles yield asynchronous 25-base reads for each template. At the 2009 Advances in Genome Biology and Technology meeting Helicos reported the sequencing of the *C. elegans* genome using just 7 of the instrument's 50 lanes in one 8-day run, a throughput of about 37 Gb/run. The data comprised 2.8 Gb of high-quality data, from 25 base reads of 0, 1 or 2 errors, which covered 99% of the genome.

Helicos Biosciences was unable to secure long-term financing from outside funding sources and has entered into bridge loan agreements according to their 2010 annual report; nevertheless, the HeliScope system is notable for a few reasons. First, although homopolymers could be an issue, just as for 454, the fact that single-molecules are used for templates means that dephasing is not an issue, thus multiple incorporations can be decreased by limiting the number of nucleotides available. In addition, a characteristic quenching function can be used to determine how many bases were added (G, GG or GGG, for example). Second, the process of sequencing generates a copy of the initial template,

which is tethered to the surface. The original template can be dehybridized, and the copy-strand can be sequenced, in what is called two-pass sequencing, generating *phred*-like quality scores approaching 30 (99.9% accuracy). The dominant error type (at 2-7% for one-pass and 0.2-1% for two-pass) is deletion, due to occasional un-labeled and non-fluorescing bases. In contrast, substitution errors are very low at 0.01-1% for one-pass and nearly 0.001% for two-pass sequencing.

Another single molecule sequencing system is being commercialized by Pacific BioSciences. In this system, single molecules are sequenced by single polymerases anchored to the bottom of zero-mode wave guide (ZMW) wells. The four fluorescently-labeled phospho-linked nucleotides are all present in solution, without terminating moieties, meaning the sequence can be read in "real-time". The continuous sequence of extensions are determined by the fluorescence pulse recorded when a nucleotide is held in the range of the ZMW during incorporation. The ZMW wells solve the dilemma that polymerases perform optimally with fluorescently labeled nucleotides in the micromolar concentration range, but single molecule fluorescent detection works best in the pico- to nanomolar range. Fluorescently labeled nucleotides may diffuse in and out of the detection volume, but when a nucleotide is incorporated into the chain it is held in place on the order of milliseconds, resulting in a pulse of fluorescence over the background signal. Errors are

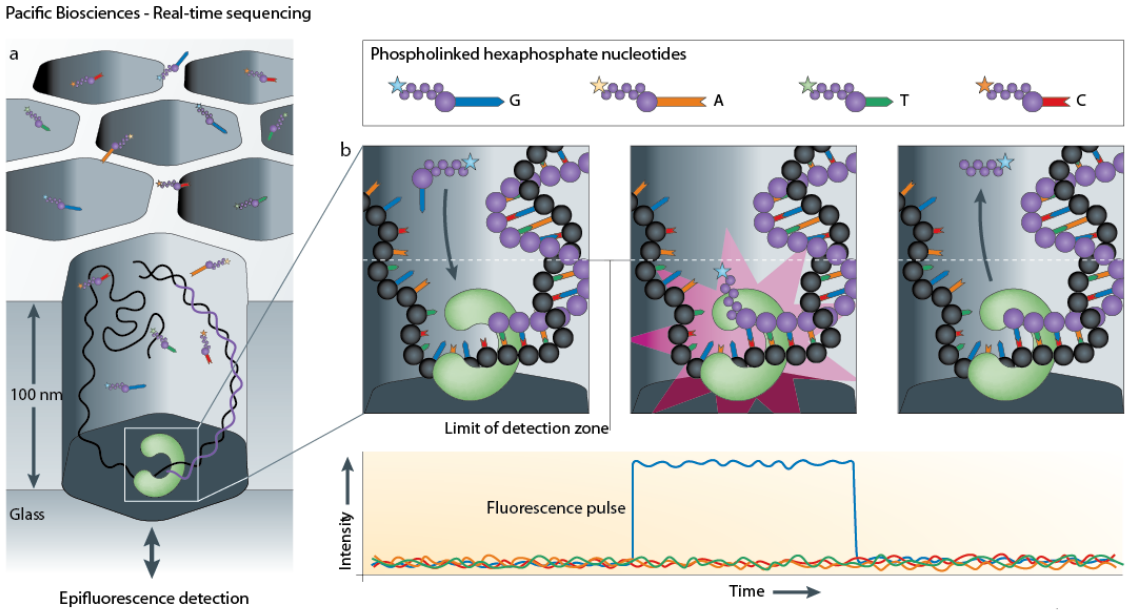

Figure 5. Pacific Biosciences' four-color real-time sequencing method. (A) The zero-mode waveguide (ZMW) design reduces the observation volume, therefore reducing the number of stray fluorescently labeled molecules that enter the detection layer for a given period. (B) The residence time of phospholinked nucleotides in the active site is governed by the rate of catalysis and is usually on the millisecond scale. This corresponds to a recorded fluorescence pulse, because only the bound, dye-labeled nucleotide occupies the ZMW detection zone on this timescale. The released, dye-labeled pentaphosphate by-product quickly diffuses away, dropping the fluorescence signal to background levels. Translocation of the template marks the interphase period before binding and incorporation of the next incoming phospholinked nucleotide. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews | Genetics (reference citation), copyright 2009. http://www.nature.com/nrg/journal/v11/n1/full/nrg2626.html

generally due to stochastic events, thus can be dramatically reduced by resequencing the templates. This is accomplished by using the φ29 polymerase, a strand-displacing enzyme capable of continuously resequencing a circular template. Using this method, the 38-fold base coverage of the E. coli genome was reported at the 2009 AGBT meeting with 99.3% genome coverage and 99.999% consensus accuracy from 964-base average reads.[18]

Ion Torrent (recently acquired by Life Technologies) has developed a semi-conductor-based sequencing technology that uses the precise fabrication technology of the semiconductor industry to create a high density array of wells that carry out sequencing by synthesis and detect the release of hydrogen ions. Because this technique does not rely on light detection, the process is simplified and the cost is reduced by eliminating the imaging system required for most next-generation sequencers. Still, the process requires PCR amplification of target DNA and utilizes a wash-and-scan process that ultimately limits the read length (about 200 bases) and throughput (about 100 Mb in a two hour run) to be similar to other wash-and-scan sytems.[27]

Many as yet not commercialized avenues are under exploration including nanopore, and direct imaging using scanning tunneling or transmission electron microscopy.[27,28] Nanopore sequencing consists of a biological or synthetic pore in a membrane and the determination of bases is accomplished by measuring the change in electrical current or optical signal as a single molecule of DNA is threaded through the pore. This technique has the potential to process extremely small amount of DNA extremely quickly. Because it is a direct sequencing approach, there is no need for PCR or labeling methods that can lead to biases in the population templates in the genomic library. In addition, there is no fundamental limit to the length of DNA molecules that can be sequenced. Work to use single molecule thick layers of graphene or nanotubes as a pore-support is also under development.[29,30]

De novo assembly, resequencing

Because of the finite read lengths output by any sequencing method, a large DNA molecule cannot be sequenced directly in once continuous read - it must be broken into smaller pieces, sequenced, and then reassembled using computers. This approach, known as whole genome shotgun sequencing, is illustrated in Figure 6. First, genomic DNA is isolated and sheared into appropriate insert sizes (usually 2 or 10 kb). Next, the inserts are ligated into lasmids containing known priming sites and transformed into *Escherichia coli*. Using a reporter gene, colonies containing template DNA can be robotically picked and used to amplify DNA for sequencing. That DNA is then used as the template for Sanger sequencing and then purified before CAE.

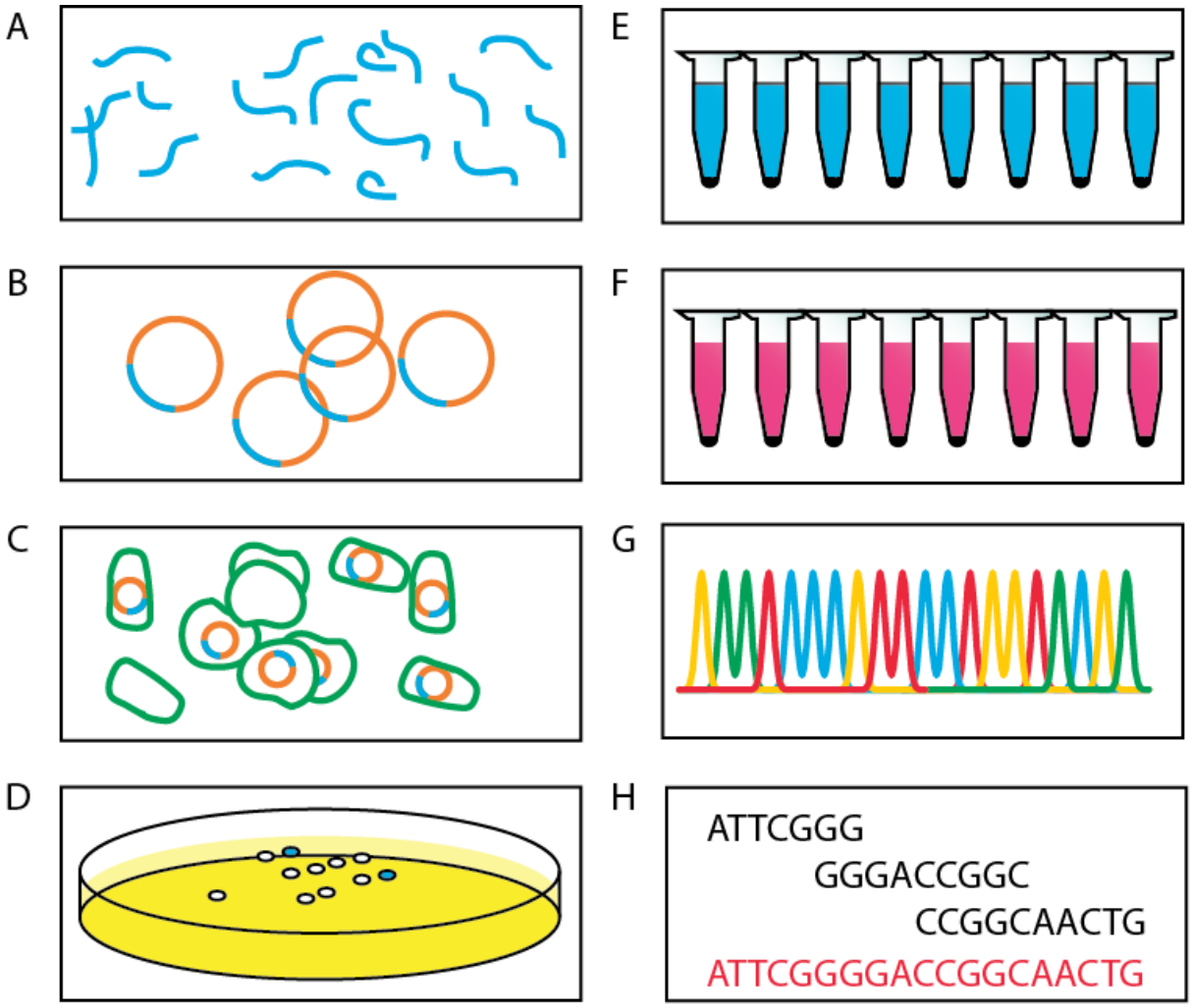

Figure 6. Whole genome shotgun sequencing pipeline. (A) Genomic DNA is sheared into random fragments and size selected. (B) Fragments are cloned into an appropriate plasmid. (C) Plasmids are cloned into E. coli. (D) E. coli are plated and successful colonies are robotically picked using reporter genes. (E) Successful colonies are picked, grown, and the plasmid is isolated. (F) Sanger sequencing reaction and purification are performed using universal primers. (G) Electrophoretic separation and fluorescence detection yields traces, which are assigned base calls by computer. (H) Contigs of sequences are assembled by computer based on overlapping regions.

The *de novo* assembly of human genomes is still not economical, but because of the existence of the high quality (finished) reference sequence produced by the HGP, it is possible to "resequence" genomes for less than $100,000. Reads can be aligned to the reference to give a comprehensive profile of the mutations and polymorphisms in an individual genome. Using NGS technologies is both cheaper and faster than using automated Sanger sequencing in its traditional form, and single nucleotide variants (SNVs) can be detected; however, there are several limitations. First, the short read lengths characteristic of NGS technologies are difficult to place in repetitive regions, and it is possible that the regions may not even exist in the reference due to gaps or SVs which can be as large as several megabases.[31] This can be remedied to some extent, by the use of paired-end reads, which help to localize a short read by limiting it to within a known distance of another read. Systematic variability in local sequence coverage can also be an issue, and may vary between platforms.

The importance of SVs cannot be overstated. SVs, consisting of inversions and copy number variations (CNVs) like insertions, deletions and duplications, affect more bases in the human genome than SNPs and arise more often. In fact, CNVs have been shown to play a larger role in human evolution, disease susceptibility and genetic diversity than SNPs.[32] Assembly problems for short read technologies may be mitigated by concomitant use of various platforms, but in general, it is also not clear how much information is lost by the use of short-read technologies. For example, 900,000 SVs and 3.2M SNVs were found when the Sanger-sequenced Venter genome was compared to the reference genome (the human genome, an assembly from a number of donors). The SVs accounted for more of the variant bases. In contrast, the 454-sequenced Watson genome found 3.3M SNVs, but far fewer SVs, which could indicate that a substantial portion of the variability was missed because of the short reads. Single molecule sequencing techniques also provide a unique form of information about CNVs since there is no amplification or cloning bias. After aligning the reads to the refernce genome, regions of CNV are over-represented in frequency. A locus of heterozygous deletion, for example, would have only half the read density of the surrounding genomic region.[33]

Microfluidic devices and DNA Sequencing

Microfluidic devices manipulate fluids on the μL to nL volume scale, a more appropriate scale for single molecule analysis. The field grew out of microelectromechanical systems (MEMS) in the 1980s and is now a respected area with multitudinous applications,[34] including biological assays, such as PCR, DNA sequencing, RT-PCR, forensics, protein crystallization, high throughput screening for drug development, pH sensing, and cytological studies. Microfluidic systems enjoy many of the advantages that their semi-conductor forefathers do, such as precise fabrication of microscale components, economies of scale in producing multiple devices and reduced energy and reagent consumption. Microfabricated devices for capillary electrophoresis were first introduced in the 1990s by Manz et al.,[35] and have proven highly effective for fast, accurate separations with high throughput.[36–39]

Components of Microfluidic Devices for Capillary Electrophoresis

Device Substrate

Various materials, including plastics and silicon, have been used in the fabrication of microfluidic devices and the choice of material depends largely on the application in question. For example, fluorescence detection requires that the material be transparent to light at the wavelengths employed and performs best when the material has low native background fluorescence. Silicon fabrication is well characterized, but its opacity is a challenge for microfluidic device development, which often relies on optical detection and it is inappropriate for integrating electrophoretic separations because of its conductivity. Plastics offer ease of fabrication and low manufacturing costs, thus providing an attractive option for disposable devices. Glass, however, has much lower native fluorescence, and is more stable when subjected high temperatures and harsh chemicals, thus it can be cleaned and reused. In addition, the surface chemistry of glass has been thoroughly characterized and there are several known methods for modifying or suppressing electroosmotic flow (EOF) in glass channels, whereas this is a more of a challenge in plastic devices. For these reasons, glass is usually the best choice for CE with fluorescent detection.

Fluorescence detection

Optical detection of fluorescently labeled molecules can be accomplished by several methods including photomultiplier tube (PMT) or CCD camera. Laser induced fluorescence (LIF) coupled with confocal detection can produce excellent sensitivity, down to just a few million labeled molecules per band. In the Mathies group, this is accomplished with the versatile Berkeley confocal radial scanner diagramed in Figure 7 and described in detail elsewhere.[40]

On-chip fluid handling

Since most biological molecules require the presence of water for analysis and manipulation, sample routing on chip requires fluidic manipulation. This is accomplished using microfabricated valves which, when serially arranged in groups of three or more, can function as peristaltic pumps. Four broad categories of valves have been demonstrated in the literature, including (i) active mechanical valves either pneumatically[41] or piezoelectrically[42] controlled, (ii) active non-mechanical valves (phase change)[43], (iii) passive mechanical (check valves)[44], and (iv) passive non-mechanical (hydrophobic) valves.[45]

The PDMS-glass hybrid valves developed and extensively characterized in the Mathies group by Grover et al.,[46] are normally-closed pneumatically-controlled valves that are easily incorporated for sample routing on glass microfluidic devices. The principles of their operation are shown in Figure 8. By applying a vacuum to the displacement chamber located above the valve the flexible PDMS membrane is pulled up allowing fluid to flow freely through the channel. Under atmospheric pressure or applied pressure the valves remain closed. Networks of valves can be combined and multiplexed to form complex architectures for diverse applications.

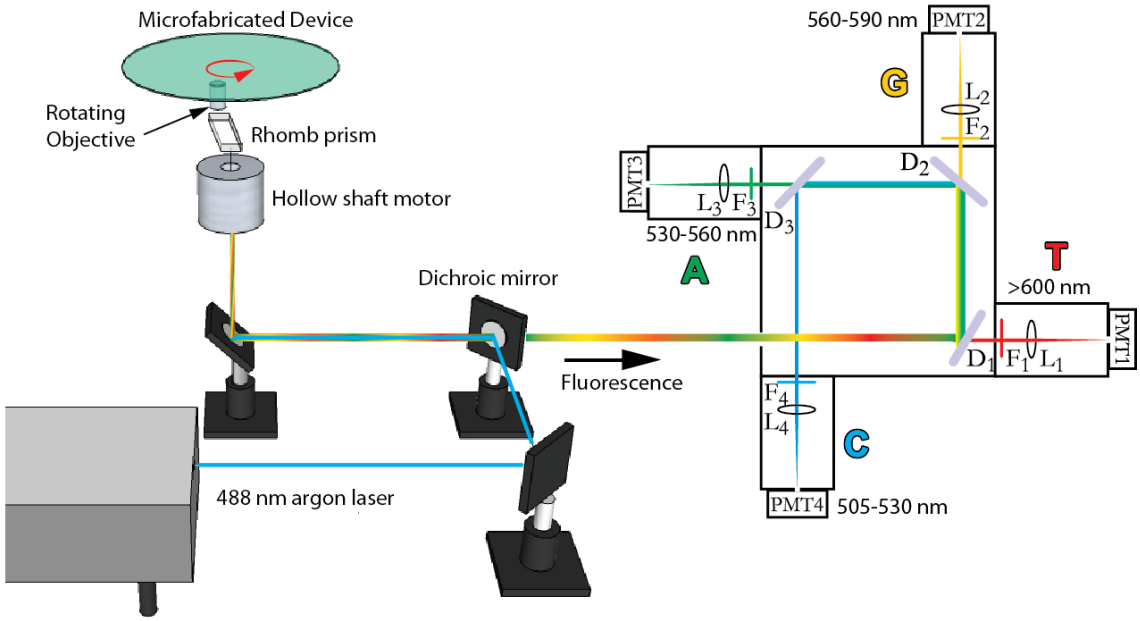

Figure 7. The Berkeley radial confocal scanner apparatus designed by Jim Scherer in the Mathies lab. Laser light at 488 nm is reflected by a dichroic beam splitter before passing through the hollow center of the stepper motor. Then, the light is laterally offset by 1cm using a rhomb prism before going through an objective lens (20X NA 0.4) focused on the array of channels in microfabricated device. The objective simultaneously focuses the excitation light on the channel and collects the emission light from each channel as it scans through the array at 5 Hz. Emitted light passes through the dichroic mirror to the array of PMTs. Four color detection is accomplished using three dichroic mirrors (D1, D2, D3) reflecting light <600 nm, <560 nm, and < 530 nm, respectively. Band pass filters F1, F2, F3 and F4 (20 nm gap centered at 605, 580, 550, and 520 nm) precede each spatially filtering pinhole. Adapted from Shi *et al.*[39]

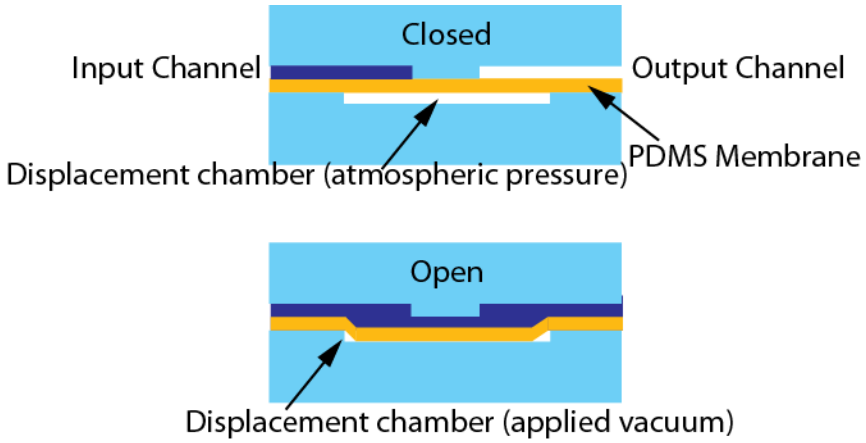

Figure 8. A PDMS valve is formed by sandwiching a flexible featureless PDMS membrane (here, in yellow) between two photopatterned glass wafers. Discontinuous channels are closed (top) until a vacuum is applied to the displacement chamber via a control line, deforming the PDMS membrane so that fluid can flow freely between the two portions of the fluidic channel (bottom).

Heating and temperature control

Many biological assays such as PCR, RT-PCR and Sanger sequencing require accurate temperature control and rapid cycling between temperatures. To accomplish this, either external or internal heaters and temperature sensors must be integrated with the device. Contact heaters in the form of Peltier devices and thin film heaters in polysilicon[47], titanium/platinum[48] and indium tin oxide,[49] have been integrated with microdevices. In addition non-contact heating in the form IR radiation[50] has also been demonstrated. The choice of heater depends on the ease of fabrication and integration as well as specific design requirements for a particular application. In general, thin film heaters use less power than Peltier heaters, and their fabrication on the surface of a monolithic glass substrate is straightforward. Feedback control using photolithographically pattered resistive temperature detectors[51] enables the accurate temperature control required for thermal cycling.

Microbead Integrated DNA Sequencing (MInDS)

Motivation

Recent studies have revealed the prevalence and significance of large SVs (>1kbp) in the human genome. SVs are thought to encompass more polymorphic basepairs than SNPs[52,53] and a more significant impact on phenotypic variation[54]. Copy number variations are associated with diseases such as glomerulonephritis[55], HIV[56], Crohn's disease[57] and psoriasis[58]. It is anticipated that future studies will reveal that SVs play an even bigger role in human genetics; however, next generation sequencing technologies, such as those

described previously, have difficulty mapping to repeat-rich regions and duplications, making them poorly suited to the *de novo* detection of structural variations.[59] Sanger sequencing chemistry is valuable because of its unique long read lengths and microfluidic devices, composed of elements described above, provide a platform that is ideally suited for advancing Sanger sequencing to its ultimate molecular limits.

Major improvements to Sanger sequencing, such as labeling with energy-transfer dyes and the use of capillary array electrophoresis were implemented in the 1990s, but significant inefficiencies still exist.[60–62] Paegel et al.'s analysis determined that the injection of DNA into a sequencing capillary is merely 0.1% to 1% efficient due to non-ideal injection geometry and injection competition of DNA with buffer ions. The use of this non-ideal injection method necessitates the production of huge amounts of DNA, that can only be generated using large-scale cloning setups that require space-demanding, complex robotic handling systems. This cloning system has not changed significantly since the beginning of the HGP; however, current sensitive detection methods based on LIF require only 100 attomoles of starting template to generate significant extension fragments for sequencing. This is an amount of DNA that can be generated using the emulsion PCR-based SCGA microbead system developed by Kumaresan *et al*.[63] Thus, replacing the cumbersome cloning step with the microbead amplification system would provide enough DNA for sequencing – provided that the extension fragments can be purified and injected with a very high efficiency (approaching 100%).

Significant work has been done to eliminate inefficient steps and develop a device capable of performing the nearly 100% injection required to sequence from SCGA microbeads. Standard sample handling steps are both a source of inefficiencies and require volumes appropriate for human or robotic handling (μL scale). Blazej et al. overcame this by developing an integrated Sanger sequencing bioprocessor based on the integrated PCR reactors developed by Lagally et al.[64,65] This device, consisting of a 200 nL reaction chamber and an affinity capture gel purification system followed by an inline injection structure, was capable of producing 500 bases of high quality sequence from 1 fmol DNA template. The injection structure in this chip limited the sensitivity of the device because only 15% of the purified sample was injected. An improved inline injection method was subsequently developed by Blazej et al.[66] and used to purify and concentrate dye primer sequencing extension fragments that were generated off-chip. This proof of principle experiment indicated that replacing the cross injection with an inline injection would increase the efficiency of injection such that fully integrated, on-chip sequencing from 100 attomoles of template DNA would be possible. Before this could be accomplished, some hurdles remained, including making the inline injection compatible with dye-terminator chemistry (simpler and easier to automate than dye primer chemistry) and increasing the read length, which had been limited to 350 bases at 99% accuracy.

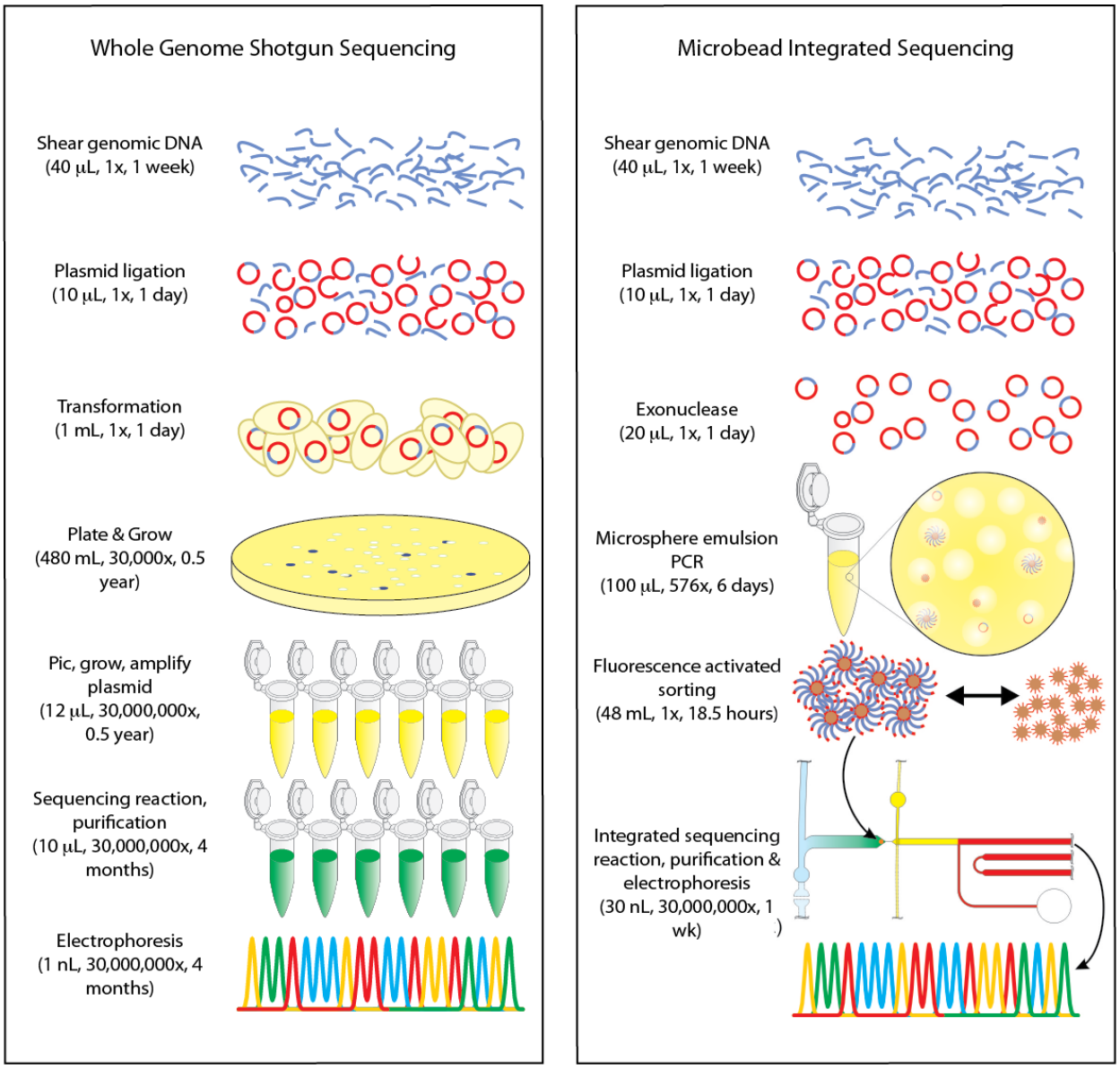

Figure 9. A comparison of traditional whole genome shotgun sequencing with the MINDS process. Plating and growing colonies to amplify DNA is a time consuming manual process. Emulsion PCR can amplify DNA in discrete droplets in a one-step reaction.

Detailed Explanation of Proposed MINDS Process

The following is a description of a proposed high throughput long read sequencing method based on the efficient sequencing bioprocessor described above with the microdroplet generator described below. This system is referred to as microbead integrated DNA sequencing or MINDS.

Step 1. Micro-droplet Generator (μDG) for Genomic Library Generation

Emulsion PCR (emPCR) has been shown to be a valuable tool for sequencing and mutation detection,[21,67,68] which allows for massively parallel single molecule PCRs by statistically diluting templates in the aqueous phase. This compartmentalization of millions of simultaneous nL scale reactions in a single PCR tube increases the relative concentration in each reaction and reduces the effects of contaminants. However, the conventional bulk production of emulsion droplets using agitation has drawbacks due to high shear that can damage DNA, and the broad size distribution of the droplets formed. The use of microfluidic droplet generation platforms, that produce uniform sized droplets with less shearing, has shown increased performance in bioassays.[69,70]

Figure 10 shows a schematic representation of a single nozzle in the microdroplet generator. Stable, uniform emulsions of aqueous droplets in oil are generated at the intersection of the aqueous and oil input channels. By statistical dilution of template DNA and primer-labeled microbeads it is possible to trap one bead and one template (on average) in a droplet of PCR reagents, enabling single-molecule amplification. The droplets are generated at high throughput and routed into a 0.6 mL tube for thermal cycling. After PCR, the emulsion is broken and the beads containing bound amplicons are recovered and sorted (because of the statistical dilution, not all beads will have PCR products) using fluorescently activated cell sorting (FACS).

Recently, Zeng *et al*. [71] developed a microfluidic emulsion generator array, building on the work of Kumaresan *et al*..[63] This array of 96 droplet-generating nozzles, illustrated in Figure 11 makes use of on-chip pumping which allows for tunable control of droplet size (~1-10 nL) and extremely uniform size distributions (relative standard deviation ~3%). This device was used to amplify a 624 bp product from single molecules which yielded 150 amol/bead of amplicon from a single template molecule at a rate of 5 Hz, or 2.4 million droplets per hour!

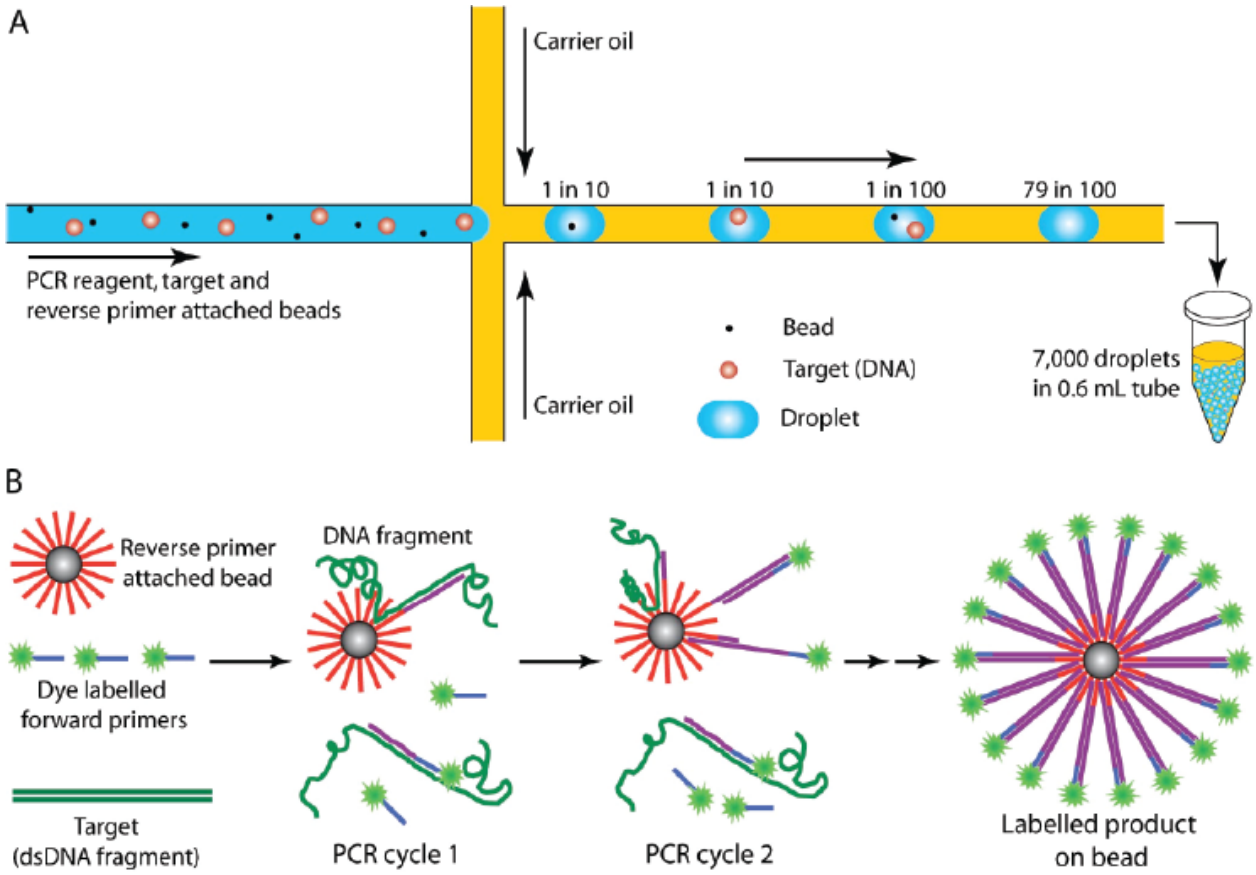

Figure 10. (A) Target DNA and primer-beads are mixed with the PCR reagents (blue) at very dilute statistical concentrations and pumped through the microdroplet generator. Monodisperse nanoliter volume droplets of PCR reagents are formed at the intersection of the oil (yellow) and aqueous flows and routed into a tube for thermal cycling in a conventional bench-top thermal cycler. The number of template DNA molecules and beads in each droplet is controlled by their concentration in the bulk fluid and the volume of the droplets. (B) Only droplets containing both a target DNA molecule and a primer-bead will be viable for PCR. The use of fluorescently labeled forward primer ensures the product on the beads with bound amplified material will be distinguishable in flow cytometry.

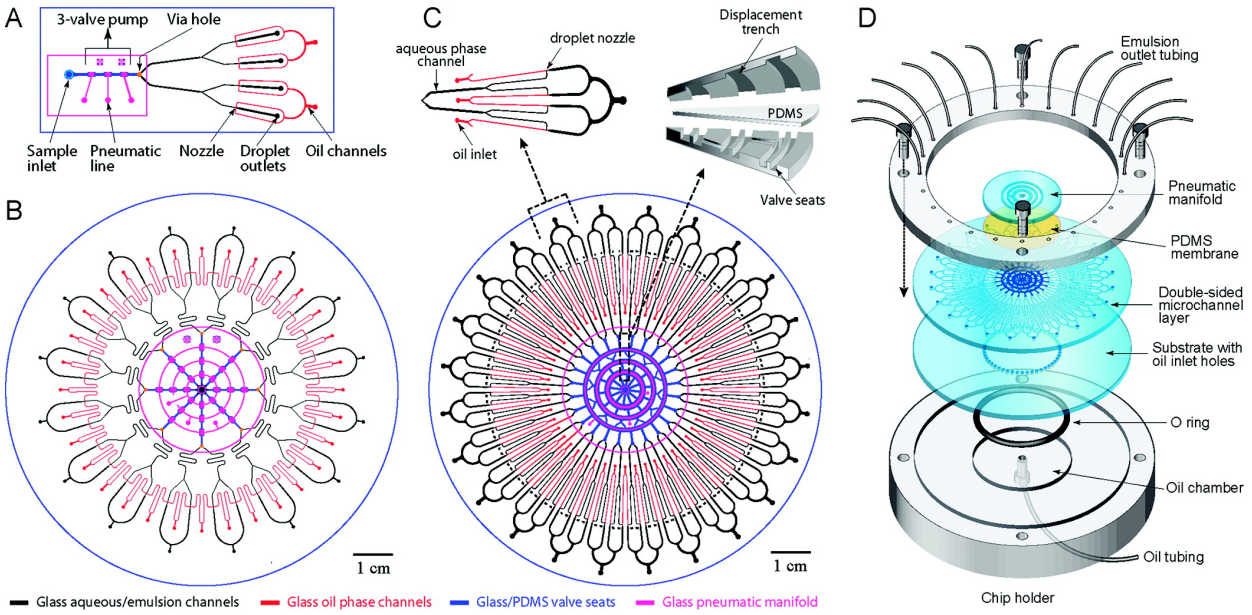

Figure 11. Microfluidic emulsion generator array (MEGA) devices. a) Layout of a glass/PDMS/glass hybrid four-channel MEGA device with a pneumatically controlled three-valve micropump integrated to drive four nozzles for droplet generation. b) Design of a 32-channel MEGA device using an array of eight identical micropumps to operate 32 nozzles simultaneously. Two adjacent emulsion channels are combined to increase device density. c) Layout of 96-channel MEGA on a 4 in. wafer composed of a single ring pump and 96 droplet generators. Inset: close-ups of a single repeating unit composed of four T-shaped nozzles (left) and the pump structure schematically showing three pairs of coaxial ring-shaped valves and displacement trenches (right). d) Exploded view of the complete four-layer 96-channel MEGA device and the plexiglass assembly module used to infuse oil and to collect the generated emulsion. (add permission to reproduce image once thesis title is fixed.)

While typical shotgun whole genome sequencing requires time consuming and expensive library generation, the MINDS process massively parallelizes and simplifies this step. Genomic DNA is sheared into 2 or 10 kb fragments as in the traditional method, but instead of transforming the insert-containing plasmids into *Escherichia coli*, the plasmids are amplified using the microdroplet generator. Following emPCR, the droplets are broken and the beads are ecovered for sorting (presence or absence of amplified DNA) using FACS. The total library of DNA fragments for a human genome can be sorted in 150 hours, or on an as-needed basis. Beads containing more than one template per reaction (expected frequency of <5% based on Poisson distribution) should be distinguishable based on their expected two-fold signal increase. If multi-template beads are not noticed during sorting, their sequencing reads will contain multiple superimposed sequences and can be discarded after basecalling. Fluorescent labeling for FACS can be accomplished with fluorescently labeled DNA probes to known sequence or by thiazole orange (TO) intercalating dye-labeling of the dsDNA.

Step 2. Integrated Sanger Sequencing Bioprocessor
In order to sequence from the template-beads, it is necessary to spatially confine single beads in isolated reactors for nL-scale Sanger sequencing reactions. An automated 96-lane microfluidic bioprocessor containing the elements described above would be ideal for sample routing, on-chip reactions, purification and CE separation. Initially, a prototype single-channel device is developed as the subject of my dissertation research.

A schematic of the integrated inline bioprocessor is shown in Figure 12. FACS sorted beads containing clonally amplified template DNA are routed to the on-chip thermal cycling reactor with ET dye terminator sequencing reagents. Following rapid thermal cycling, the extension fragments are electrophoretically driven to the 200 μm x 200 μm x 30 μm photopolymerized capture gel, where they are selectively purified from residual reagents (primers, salts, and especially ET dye-labeled ddNTPs) and concentrated by approximately 150X.

Photopolymerized capture gels provide several advantages over the inline-injection described in Blazej *et al*..[66] First, the use of photopolymerization allows for precise patterning of small capture gels (below 150 μm x 200 μm) in an easily automated fashion. This is an improvement over the previous method, in which capture gel was loaded by hand using a syringe and could not produce gels smaller than ~1 cm. Second, use of cross-linked polyacrylamide improves the stability of the gel. Even with the application of high separation electric fields (150 V/cm) the photopolymerized capture gel remains stable and

anchored in place, improving on the previous method, in which a linear polymer gel was used. In the former system, the covalently bound capture oligo would migrate down the channel during CE, dragging capture gel with it, and leading to band broadening and decreased resolution.

Scope of the Dissertation

This dissertation presents my work in three areas. First, the inline injection method is improved for compatibility with dye-terminator sequencing chemistry and the resolution is improved by increasing the stability of the capture gel. Second, the improved inline method is integrated with on-chip thermal cycling of free solution template DNA. Finally, a device is presented proving the feasibility of the MINDS concept: an integrated bioprocessor for sequencing from bead-bound targets. In the final chapter, a discussion is presented analyzing the prospects for developing a high-throughput device based on this single channel prototype and other related applications are explored.

Chapter 2. Integrated Inline Bioprocessor for DNA Sequencing

Abstract

A highly efficient fully integrated microfluidic platform for Sanger DNA sequencing, including automated thermal cycling, purification, concentration and in-line injection of the extension fragments for microchip capillary electrophoresis separation. The two-layer glass device that I developed features two independently operated valve-free systems, comprised of a 200 nL thermal cycling reactor with resistive temperature detector, a 1.2 nL *in situ* photopolymerized oligonucleotide affinity capture gel for post reaction clean-up and inline injection, and an 18-cm long capillary electrophoresis channel for separation. Integration of the efficient photopolymerized affinity gel capture allows sequencing from only 350 - 500 attomoles of starting DNA template. Using this device, I was able to sequence 507 ± 31 bases at 99% accuracy. In addition, I show that this method is compatible with single cell genetic analysis techniques (SCGA) by sequencing from the small amounts (~100 attomoles) of amplified DNA bound to agarose microbeads that can be produced from single cells.

Introduction

The draft and finished version of the sequence of the human genome is now available due to the successful completion of the Human Genome Project. [17,72,73] This information is valuable for identifying diseases,[74,75] studying disease susceptibility,[76] elucidating mechanisms of human evolution,[77] developing pharmaceuticals[78] and understanding human genetics. Sanger sequencing chemistry was the fundamental tool employed for the Human Genome project, and despite the recent development of massively high throughput sequencing platforms, it is likely to remain the gold standard sequencing method for many applications.[21,79–82] Much work has been done toward the aim of streamlining and modernizing the Sanger sequencing pipeline, particularly in the area of microfabricated devices. [83–85] Advantages of lab-on-a-chip technologies for sequencing applications include low sample and reagent consumption, rapid sensitive CE separations, efficient integration of various chemical reactions in one planar device and amenability to automation without the need for large robotic systems.[86–88]

Traditional Sanger sequencing coupled to CE separation has impressive read lengths, even up to 1300 bases in some cases,[16] but lacks the sensitivity needed for single cell sequencing. We are working to develop more highly sensitive methods for genetic analysis. The first step toward this goal was a fully integrated device, capable of performing thermal cycling, post reaction purification and separation on-chip.[65] This chip has demonstrated the production of up to 556 bases at 99% accuracy, but employed an inefficient cross-injector resulting in a minimum starting template requirement of 1 fmol of DNA template. To address this inefficiency, we developed a new injection method based on an inline affinity capture gel. When tested in a prototype chip (capture and separation only), the results indicated that the amount of template DNA needed for sequencing could be reduced from 1 fmol to 100 attomoles; however, because the affinity capture gel was made without photopatterning and used a linear polyacrylamide matrix, the device suffered from limited read lengths and tedious manual gel loading.[66]

In addition, we have recently reported a novel high performance single cell genetic analysis technique based on microdroplet generator technology.[63,71,89] This powerful technique combines high-throughput microfluidic droplet generation with single cell multiplex polymerase chain reaction (PCR). We demonstrated multilocus, single cell sequencing using the DNA from single beads. Following PCR, the beads were recovered, diluted to the statistical level and then loaded in a 96-well plate for re-amplification. The products were then run in agarose gels, and where a band was present, it was cut out and the recovered DNA was sequenced. Using this method it was possible to sequence from multiple single cells, although the two rounds of amplification and gel-band excision was a time-consuming process.

Here we present a novel fully integrated microfluidic sequencer that uses a photopolymerized inline injection method capable of interfacing with microbeads carrying amplified target DNA. This system combines the high sensitivity of inline injection with the stability and reproducibility of in-situ photopolymerized affinity gels to generate high quality sequencing reads from only 350 to 500 attomoles of starting DNA template. Because of the stable photopolymerized capture gel, there is no need for off-chip pressure-driven flow or valving, making this device simple to operate and easily automatable. We demonstrate sequencing from bead-bound PCR amplicons targeted to detect the t(14;18) translocation, a mutation associated with 85-90% of follicular lymphoma cases, as a model sequencing target.[90,91] This represents a first step toward automated devices for characterizing the co-occurrence of somatic cell mutations associated carcinogenesis on a single cell level.

Materials and Methods

Fabrication and device control. The sequencing bioprocessor illustrated in Figure 12 is fabricated using two 500 μm thick, 100-mm-diameter borofloat glass wafers (Precision Glass and Optics, Santa Ana, CA). Two independent analysis systems are patterned on one wafer, each consisting of three functional modules: a 200 nL thermal cycling reactor (Fig. 1B), a 1.2 nL cross linked polyacrylamide capture gel (Fig. 1A) and an 18 cm (effective length) serpentine capillary electrophoresis (CE) channel. The bottom wafer is patterned with electrical features and the top wafer contains the microfluidic components. For the top wafer, the mask pattern is photolithographically transferred (KS Aligner, Suss MicroTec, Santa Clara, CA) to a 2000 Å amorphous silicon hard mask and microfluidic features are etched with concentrated hydrofluoric acid (49% HF) to a depth of either 30 μm (for the channels) or 60 μm depth (for the reactors). All channels are 200 μm wide except in the hyperturn regions. Four-point RTDs are fabricated on the top side of the bottom wafer by a 2 min aqua regia etch of a 200 Å Ti/2000 Å Pt sputtered layer at 90 ˚C. Access holes for the channels (1.1 mm diameter) and RTD contacts (1.5 mm) are drilled in the top wafer and then the top and bottom wafers are aligned and bonded in a vacuum furnace at 650 ˚C for 5 hours.

Thermal cycling is accomplished using a 12-mm square surface heater (7.8 Ohms; Minco, Minneapolis) mounted on the bottom of the device directly under the reactor. RTD

connections are made by spring loaded pins (Everett Charles Technologies, Pomona, CA). A 1 mA current is passed through the outer leads and the temperature-dependent voltage drop across the inner leads is measured using a signal conditioner (5B31-01; Analog Devices) and a data acquisition board (National Instruments). The temperature of the reactor is controlled by a proportional integral derivative (PID) controller that delivers 0-8 V for powering the heater and applies active cooling with pressurized air.

Polymer synthesis. Separation gel (4% LPA with 6 M urea in 1X TTE) and backing gel (3% LPA in 1X TTE) were prepared using previously published methods. (Paegel, 2001, 96 lane) The cross linked affinity capture matrix was prepared by sparging a solution of 3% T 12% C acrylamide containing 20 μM acrydite-modified capture oligonucleotide, (5'-acrydite-cctcccgtatcgtagttatcta, IDT) in 1X TTE buffer with N_2 gas for 10 min before adding 0.2% (w/v) VA-086 photoinitiator (Wako Chemicals, Richmond, VA).

Bioprocessor preparation. Before operation, all fluidic channels are cleaned with piranha (3:1 H_2SO_4: H_2O_2) at 65 ˚C for 10 min, rinsed with DI water, and coated with a dynamic coating solution (1:1 methanol:DEH-100, The Gel Company, San Francisco, CA) for 5 min. In a dark room, the separation channel is filled with capture gel monomer solution and 20 μL of 5% linear polyacrylamide (LPA) is placed on all wells to inhibit any flow of the solution during polymerization. The capture gel is patterned using a chrome mask with a 200 by 400 μm window for UV illumination (365 nm, 10 mW/cm^2, 8 min) from a mercury arc lamp on a Nikon inverted microscope (TE2000U). Unpolymerized monomer solution is removed from the chip by vacuum and replaced with 1X TTE in the CE channel and 3% LPA backing gel in 1X TTE in the sample arm (light blue in Fig. 1A).

DNA sequencing. Sequencing reagents containing 1X DYEnamic ET dye terminator sequencing premix (GE Healthcare, Fairfield, CT), 701 bp pUC18 PCR product template (500 attomol), 50 fmol of sequencing primer (5'–tccatagttgcctgactccc-3') are loaded into the thermal cycling reactor via the sample inlet port and sealed in place with 20 μL of 5% LPA on the sample inlet and outlet ports. The reaction is rapidly thermal cycled for 35 cycles of 95 ˚C (12 s), 50 ˚C (10 s), and 60 ˚C (60 s) and then electrophoretically driven to the capture gel by applying an electric field of 20 V/cm between the sample inlet port and the waste port for 40 min at 35 ˚C. Negatively charged species, including extension fragments, salts, unincorporated nucleotides and primers, are driven through the capture gel. After capture is complete, an electrophoretic wash step of 5 V/cm for 5 min from waste to cathode and 50 V/cm for 3 min from cathode to sample outlet removes any residual reaction components. In-house polymerized 4% LPA containing 6 M urea in 1X TTE is loaded from the anode at 300 psi. Extension fragments are released from the photopolymerized capture gel at 67 ˚C and immediately inline injected at 150 V/cm with 100 s pull-back voltages of 150 V at sample outlet and waste ports. Bands are detected 1 cm from the anode by laser-induced fluorescence on the Berkeley confocal rotary scanner[92] and processed with the Cimarron base caller 3.12 (NNIM, Sandy, UT).

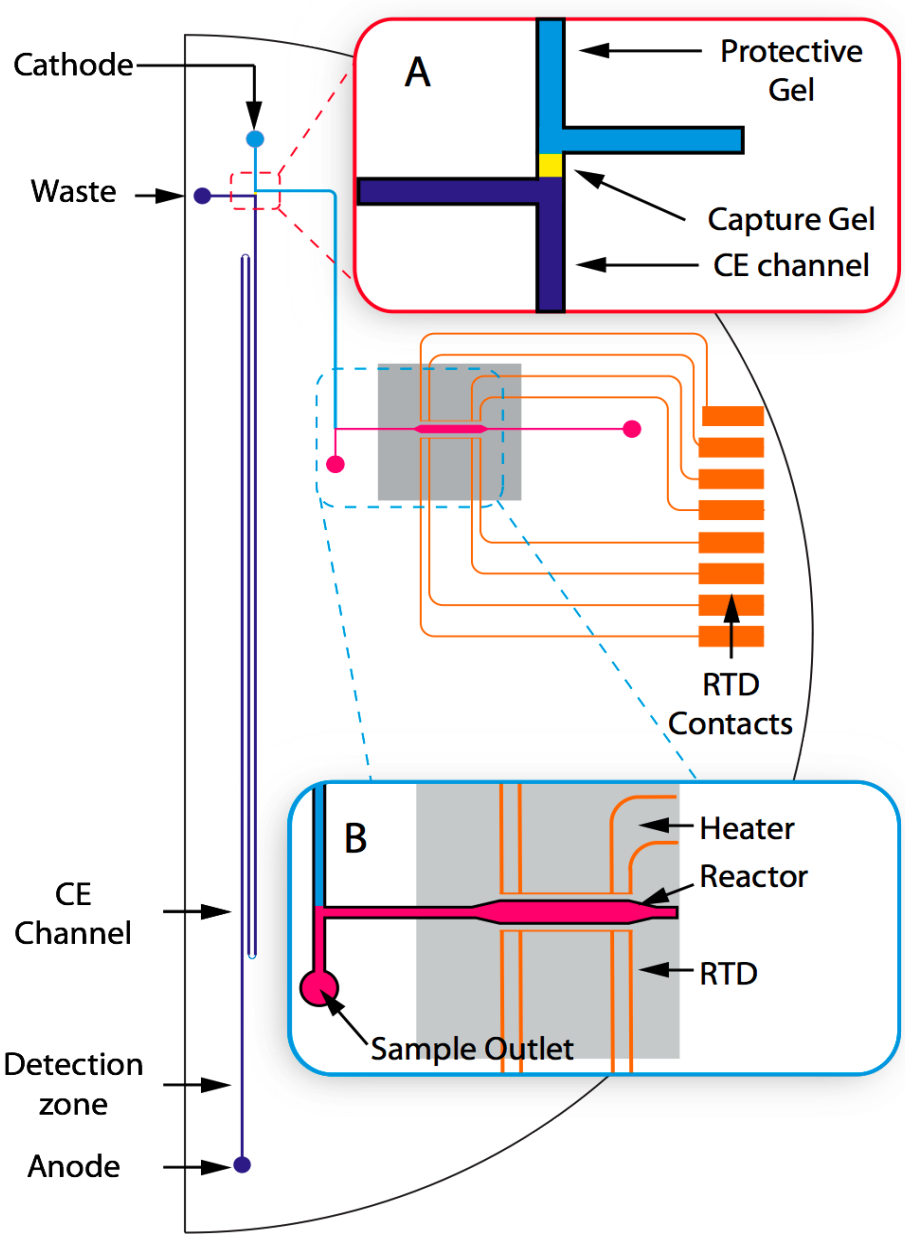

Figure 12. Integrated Sanger sequencing microfluidic processor fabricated on one half of a 100 mm diameter borofloat glass wafer. The device has three functional modules, which include a Sanger sequencing reactor, a capture gel for purification and separation, and a CE separation channel. (Box A) Expanded view of the 1.2 nL oligo capture gel (3% acrylamide / 12% bis) photopolymerized in the 200 mm offset between the waste arm and the 18 cm separation channel (dark blue). The capture gel is isolated from the reactor by a 3% LPA protective gel in the sample arm. (Box B) Expanded view of the 200 nL reactor. A microfabricated RTD and a heater are used for thermal cycling.

Affinity-capture optimization. The sequencing bioprocessor was placed on the Berkeley confocal scanner[92] with the detection point 3 cm after the capture gel, and equilibrated at 25, 30, 35, 40, 45 or 55 ˚C. 0.1 nM FAM-labeled oligonucleotide complementary to the capture oligonucleotide was loaded in the reactor and captured by applying a 20 V/cm electric field. A fluorescent signal was detected when any uncaptured DNA passed by the detector. Finally, the voltage was stopped and the bioprocessor was heated to 70 ˚C and equilibrated to release the captured DNA before injecting at 20 V/cm. Log-normal peak fits were obtained using PeakFit v4 (Systat Software, Inc., San Jose, CA).

Capture efficiency was found to be dependent on capture temperature as expected for a hybridization-based capture. At temperatures above 35 ˚C, the amount of perfectly complementary DNA passing through the gel without hybridizing begins to increase, as shown in Figure 13. Thus, a capture temperature of 35 ˚C was used for all experiments.

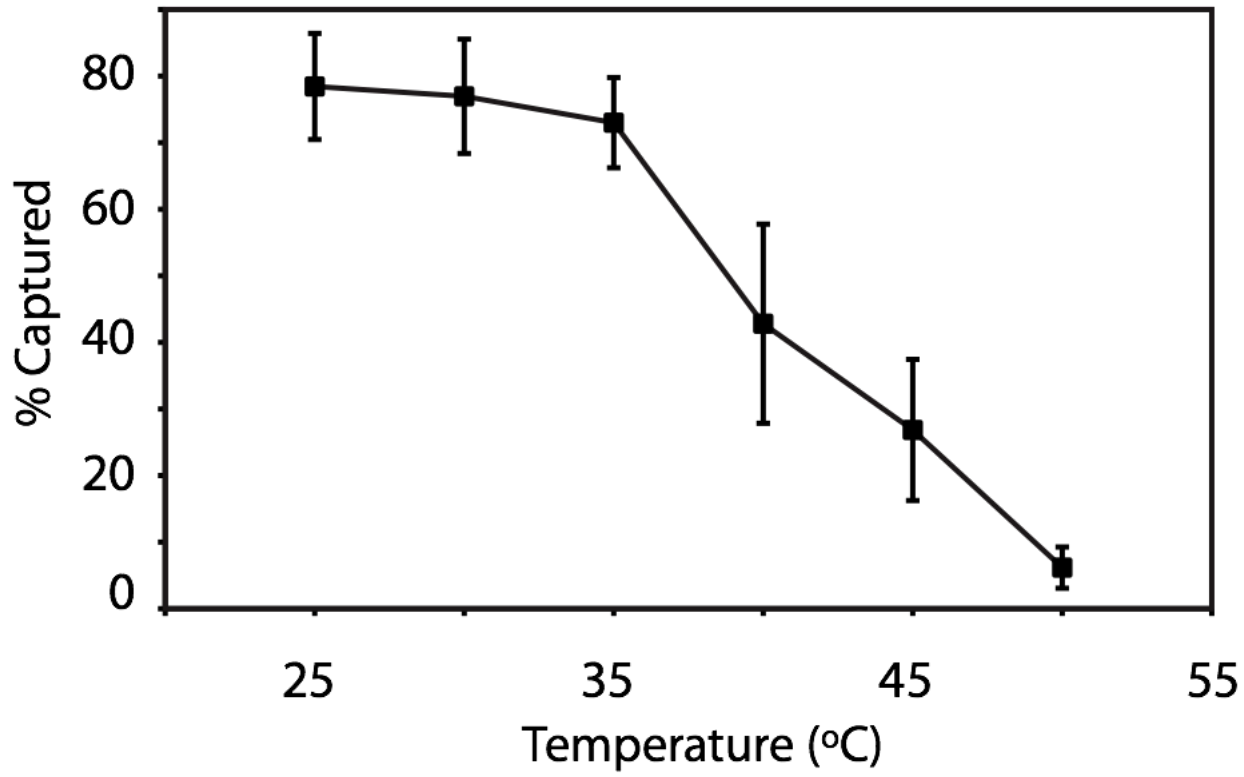

Figure 13. Capture efficiency of complementary oligonucleotide as a function of capture temperature at 20 V/cm. Fluorescence detection was performed 3 cm downstream from the capture gel. An electric field of 20 V/cm was applied between the sample and anode ports so that the amount of fluorescently labeled DNA passing through the gel without being captured could be detected. After all the DNA from the reactor had either captured or passed through the gel, the entire device was raised to 70 oC and captured material was driven through the detection region by a 20 V/cm electric field. The percent captured material is calculated as peak area at 70 oC divided by the total area (captured peak area plus uncaptured peak area) for 25, 30, 35, 40, 45 and 50 oC (n=3).

Results and Discussion

To demonstrate the functionality of the integrated microfluidic sequencing processor, 500 attomol of pUC18 PCR amplicon (701 bp) was sequenced by on-chip reaction, purification and CE separation. DYEnamic ET dye terminator cycle sequencing reagents combined with primer and target DNA were thermally cycled and the contents of the reactor were then electrophoresed through the capture gel, as shown in Figure 14, by applying an electric field of 20 V/cm between the sample inlet and waste ports for 40 min. Higher capture voltages were explored, but above 20 V/cm the plug was not stable during capture. With the chip held at 35 ˚C, the *in-situ* polymerized cross-linked polyacrylamide capture gel binds successfully extended sequencing fragments while allowing the passage of residual reagents. The accumulation of bound sequencing fragments can be seen in Figure 14B-14D. In particular, there is complete removal of the dye-labeled ddNTPs, whose high concentrations can confound sequence readout even at low levels of residuals.[93]

The backing gel separating the capture gel from the sequencing reactor serves two purposes in the device. First, it provides a passive barrier that blocks the flow of the sequencing reagents during thermal cycling. Second, it provides an important barrier to diffusion during injection. After the purified DNA is thermally released from the capture gel, its diffusion back toward the reactor is limited by the backing gel. Peak widths were found to increase significantly when the device was operated without backing gel.

The photopolymerized gel measures ~200 x 200 x 30 μm, thus constraining the length of the DNA plug that is ultimately injected to 200 μm, while containing sufficient capture oligonucleotide (1.2 nL x 20 uM = 24 fmol) to capture the theoretical maximum number of extension fragments generated from 500 attomoles of starting template DNA (500 attomol x 35 cycles = 17.5 fmol). The mechanical strength and pore size of polyacrylamide gels can be tuned by adjusting the %T and %C (T = total acrylamide, C = cross linker).[94,95] In order to leverage the differential mobility between the separation matrix and capture matrix to effect stacking, a gel for which DNA has a relatively high mobility was desired for the capture gel. At relatively low %T, increasing %C increases the electrophoretic mobility.[95] A 3% T 12% C gel yielded the highest mobility (estimated to be 2.5 E-4 cm^2/Vs for a 200 bs fragment) while still maintaining the mechanical properties of a solid. Since the mobility of a similarly-sized fragment in the separation gel was $\mu_{200bs} \approx 1.5$ E -4 cm^2/Vs, the use of a 3% T 12% C capture gel caused stacking of the injected band during injection that would not occur with, for example, a 5% T 5% C capture gel, for which $\mu_{200bs} \approx 0.65$ E -4 cm^2/Vs.[96,97]

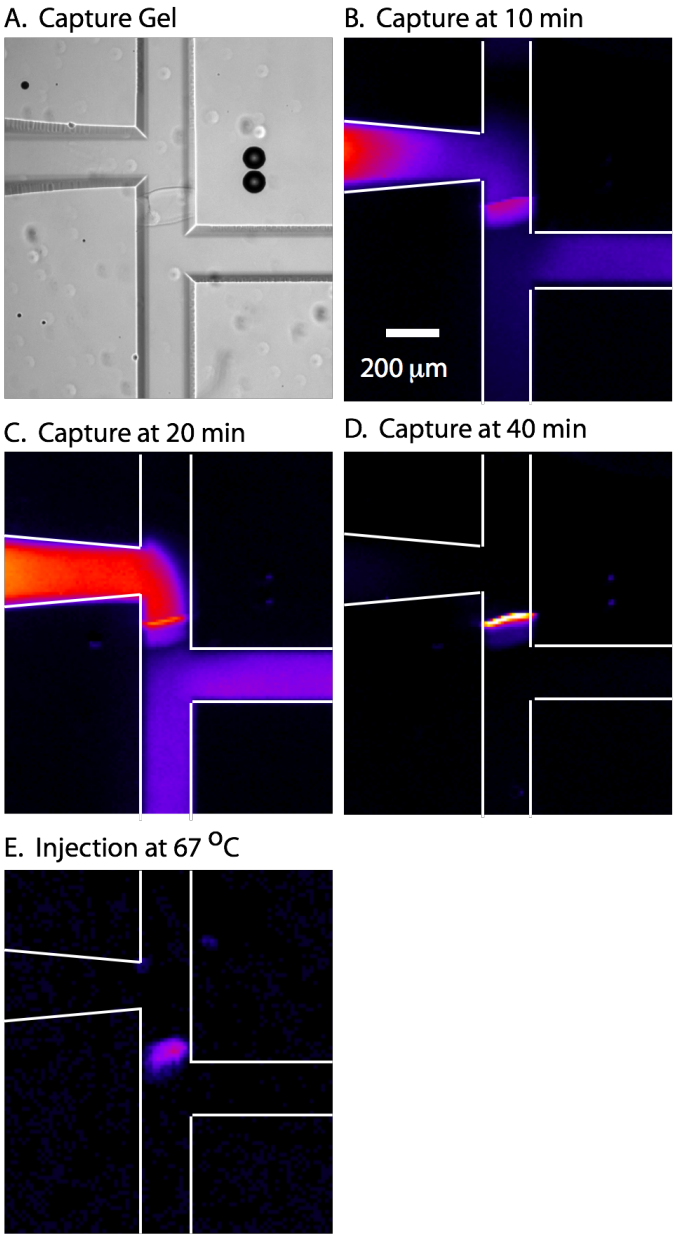

Figure 14. False color fluorescence images of post-reaction sample concentration and cleanup. (A) The 1.2 nL capture gel before capture. (B) 35 V/cm electric field applied from the reactor to the waste drives dye terminator sequencing fragments through the 1.2 nL oligo capture gel (left to right) held at 35 oC. (C) Sucessfully extended DNA fragments are captured via hybridization to the oligonucleotides in the capture gel. ET dye labeled-ddNTPs pass through the gel. (D) Final band of purifed and concentrated sequencing fragments. (E) After heating to 67 oC to release the purified fragments, a 150 V/cm electric field is applied from the cathode to the anode to effect separation.

After an electrophoretic washing step to remove any residual ddNTPs, the entire device is ramped to 67 ˚C, well above the T_m of the capture oligo (52.5 ˚C) to dehybridize the sequencing fragments from the capture gel. A separation electric field of 150 V/cm is then applied from cathode to anode. The injected plug of DNA can be seen entering the CE channel in Figure 14E.

Representative sequencing data produced with the microfluidic sequencer is presented in Figure 15. Alignment of three runs to the known pUC18 sequence using a 99% accuracy cutoff yielded an average read length of 507 ± 31 bases. The complete reaction on-chip yields similar results to capture and separation of sequencing fragments produced with an off-chip reaction (average read length of 524 ± 48 bases). Figure 16 shows phred[98] quality scores for each base call with the predicted read accuracy at each fragment length.

To demonstrate the feasibility of sequencing from amplified DNA on our SCGA microbeads, we performed on-chip sequencing using the DNA on 7 microbeads. These beads contained a mean of 50 attomoles PCR product amplified from genomic DNA of RL cells. Figure 17 shows the partial sequence of the t(14;18) breakpoint region, including the unique insert sequence between chromosomes 14 and 18, which matches the expected insert sequence for this cell line.[91]

The current device improves upon and combines features from two of our previously published devices: the integrated bioprocessor[65] and the inline injection device.[66] The former was also a fully integrated system (reaction, purification and separation), but required at least 1 fmol starting template DNA because of the inefficient cross-injector system employed. The latter device demonstrated a novel inline injection technique, which improved injection efficiency, but did not integrate on-chip thermal cycling and required delicate loading of two linear gels relying on fluidic resistance to hold them in place. The instability of this linear gel system made it difficult to incorporation in integrated devices.

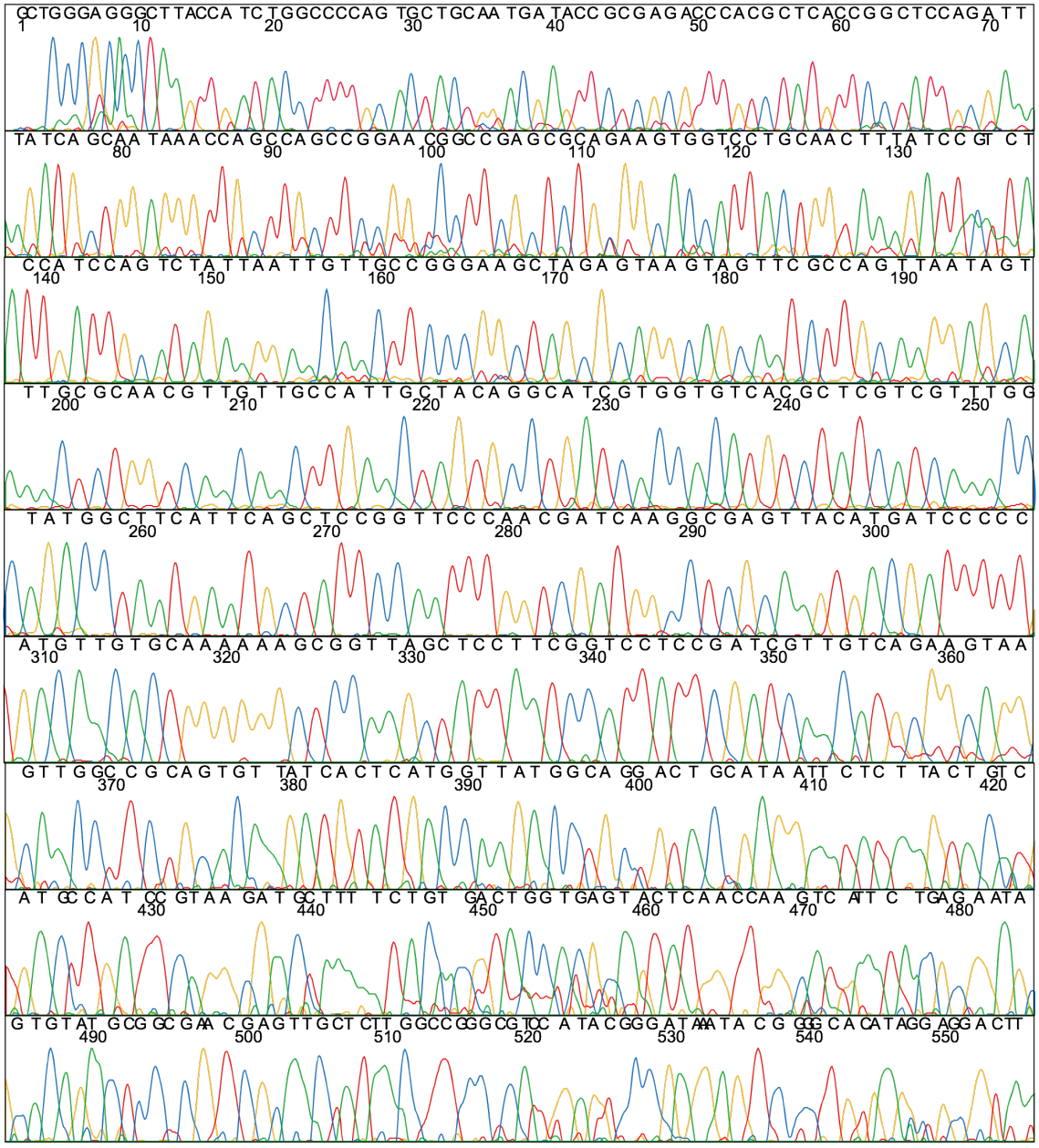

Figure 15. Sequence generated from on-chip reaction, purification and concentration with inline photopolymerized injection. Sanger sequencing reaction generated from 500 amol of 701 bp pUC18 PCR amplicon are separated in 40 min at 67 ˚C and 150 V/cm using a 4% linear polyacrylamide with 6 M urea in 1x TTE buffer. The on-chip reaction consisted of 30 cycles of 95 ˚C for 15 s, 50 ˚C for 15 s and 60 ˚C for 55 s. Automatic basecalls by the Cimarron base caller and base numbers are shown above the trace.

The system presented here advances the state of the art in four ways. (1) The photopolymerized gel utilized here shows significant increases in read length over our previous inline device, which only produced 365 bases 99% accuracy.[66] This increase is likely due to the remarkable stability of the photopolymerized gel. In the previous system, the capture was performed using a linear gel (5% LPA) polymerized off-chip and then manually loaded in the offset between the sample and waste arms. This gel was held in place simply by its fluidic resistance and could be seen moving ~400 um during the 15 min electrophoretic wash at 12 V/cm. It is assumed that the gel must have moved substantially during the 30 min separation under 100 V/cm leading to band broadening. Thus the improved sequence quality in the system here is likely due to the improved physical stability of the photopolymerized crosslinked capture gel. (2) An additional advance over our previous inline system is the compatibility with dye terminator sequencing chemistry, which allows for four-color sequencing reactions in a single reactor (rather than four separate reactors, as in dye primer sequencing). In dye primer sequencing the four versions of the primer must be synthesized, with four distinct dyes. The Sanger sequencing reaction is performed in four separate vessels, one for each dideoxynucleotide base. Dye terminator sequencing chemistry is more amenable to automation because of the reduced complexity. (3) Compared to our integrated device, we see a 2-fold reduction in required template for high quality on-chip sequencing. This reduction is due to the high-efficiency of the inline injection, about 80% compared to the cross injection, which is only 15% efficient.[65] (4) Because the photopolymerized gel is highly mechanically stable, it effectively partitions the device into three compartments, the contents of which cannot move. This obviates the need for microvalves and off-chip syringes for fluidic control because the movement of charged species can be simply, automatically and directly controlled by application of electric field.

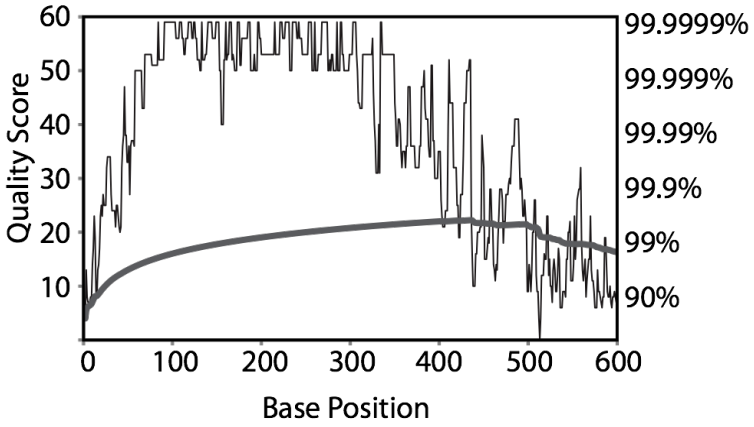

Figure 16. Basecall accuracies as predicted by PHRED. The thin line plots the PHRED quality score at for each individual basecall and the thick line plots the predicted read accuracy at each length.

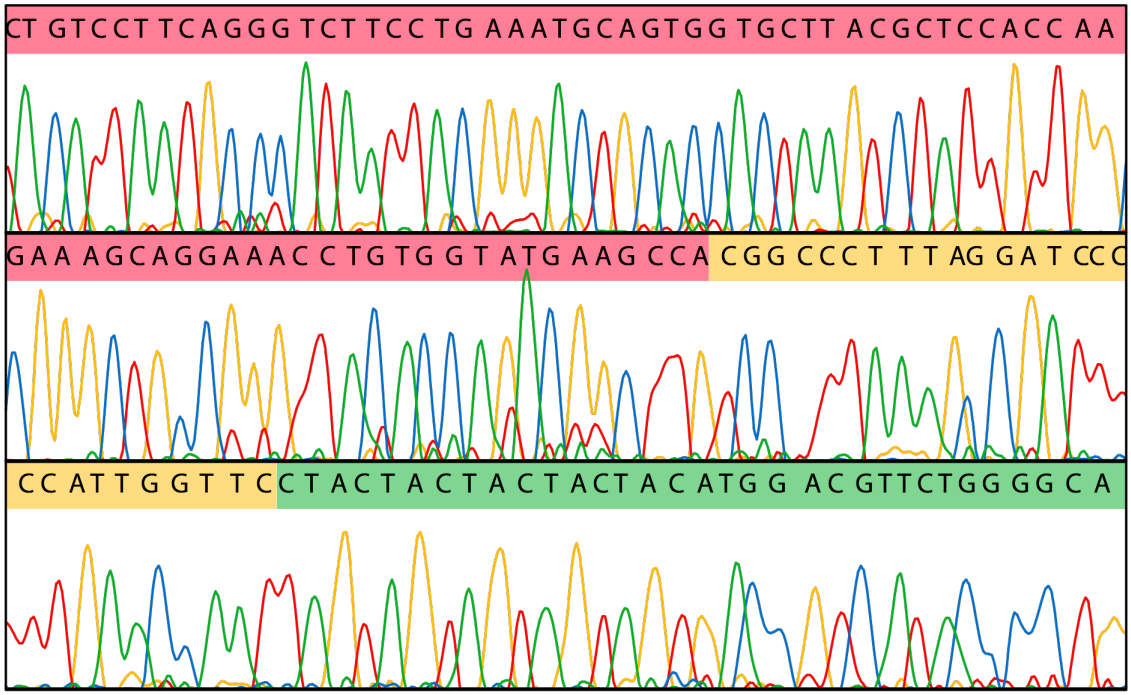

Figure 17. Partial sequence of t(14;18) breakpoint region. This trace was generated on-chip using amplified DNA from RL cells on 7 beads as a template for Sanger sequencing. The unique insert sequence between chromosome 18 (red) and chromosome 14 (green) is highlighted in yellow and matches the expected insert sequence. The amount of amplicon on 7 beads corresponds to about 350 attomoles.

Conclusions

We have demonstrated a fully integrated inline injection Sanger sequencing microfluidic device capable of thermal cycling, purification and separation of 507 bases at 99% accuracy. By utilizing an *in situ* photopolymerized capture gel, the injection process is greatly simplified and improved compared to our previous sequencing devices. This method is compatible with dye terminator sequencing chemistry and enables sequencing from a few attomoles of template DNA. This device can also be applied to DNA sequencing from microbeads carrying amplified single molecule targets, such as those generated by our microdroplet generator.[71]

Chapter 3. Prospects

The future of sequencing

In future years, DNA sequencing will likely play a larger role in both medicine and science as advances in technology allow for the development of economical sequencing instruments for diverse applications. If the goals set forth by the NHGRI to reduce the cost of sequencing a human genome to $1000 are met, then we will likely enter the era of genomic sequencing as a diagnostic tool used by clinicians to diagnose and prevent disease, in addition to ongoing scientific pursuits of *de novo* whole genome sequencing, complete genome resequencing, targeted genomic resequencing, paired end sequencing, metagenomic sequencing of populations, transcriptome sequencing, small RNA sequencing, and others. This brings up many issues including the issue of how to interpret such a flood of information. Recent advances in sequencing and diagnostic tests have outpaced our understanding of disease in many areas and may lead to a problem of over diagnosis and over treating (e.g , prostate specific antigen test for prostate cancer). Significant investment should be made toward developing guidelines to serve as a framework for interpreting this information.[99]

In terms of which specific sequencing chemistry or instrument is likely to dominate future sequencing efforts, it is likely that many technologies will co-exist due to the disparate nature of various sequencing applications. Although the next generation sequencing platforms were developed to be high throughput and low cost, the actual cost of running any of the instruments is very expensive. For example, the GS-FLX, Illumina and SOLiD instruments cost about $7000, $6300 and $7700/per run and take up to 10 days to complete a run. Because of the high cost associated with every press of the "start" button, it is common practice to check the quality of libraries using Sanger sequencing before doing so. In addition, researchers wanting to make the most of finite resources will likely want to limit the region of interest to a certain area of the genome known to be associated with a disease phenotype, or with a selection of genes associated with a disease pathway. These considerations imply that a medium throughput device capable of highly accurate and fast sequencing with the ability to target specific genomic areas would be of great interest to the scientific and medical communities. Sanger sequencing combined with electrophoretic separation remains by far the best option for sequencing in terms of read length and quality, and remains the workhorse of genome sequencing centers, but we still lack a way of parallelizing it to meet the demands of speed and price.[100] Microchip based methods are uniquely poised to solve this problem.

A recent review by Fredlake *et al.* compared Sanger sequencing plus electrophoresis, as currently employed, with the three most commonly employed next generation sequencing instruments. When comparing the cost of generating *de novo* sequence of a human genome with 30x coverage, the 454, Illumina and SOLiD systems required $6,100,000, $80,000 and $130,000, respectively. To generate the same quality sequence using Sanger sequencing, which requires only 5x coverage, the costs are $5,000,000. Thus, as currently employed, the next generation methods are better suited for low cost massively high throughput sequencing. However, for applications such as targeted sequencing of, for example, human leukocyte antigen (HLA) compatibility testing, in which there is a targeted region of interest, or in human STR typing, where highly accurate base calls are required in as short a

time as possible, the next generation sequencers are too cumbersome and ill suited for these applications.

The MINDS process described in Chapter 1 has many advantages, including reduced sample and reagent volumes, shortened analysis times due to shorter separation channels, reduced time needed for sample preparation, integration of the individual steps of sequencing and the ability to parallelize the sequencing separations by using a multi-lane device. To make this device a reality, there are two major milestones that must be reached. (1) The amount of product amplified from single molecules must be sufficient to enable robust sequencing using the integrated microchip sequencer. (2) A multi-lane device must be designed capable of sorting single beads into individual Sanger sequencing reactors and simultaneously processing multiple CE runs.

Feasibility of single cell sequencing with MINDS system

The work presented in Chapter 2 shows sequencing from as few as 7 beads. This raises the question of the feasibility of single cell sequencing. The limit of detection for the radial scanner used in these experiments for fluorescein in 1x TAE buffer in a microchannel is 10 pM. Shi et al. calculated the number of fluorescent molecules needed to detect a peak to be about 1 million, or about 1.6 attomoles.[40] To sequence a 500 base molecule we would then need 500 million Sanger extension fragments. Assuming perfect capture efficiency, a polymerase efficiency of 70% and 35 thermal cycles, we would then require

$$500 \times 10^{6} \div 35 \ cycles \div 0.7 \div 6.02 \times 10^{23} = 34 \ attomoles$$

of amplified DNA per bead for sequencing. In practice, the injection efficiency for inline injection has been measured to be ~80% (see Chapter 2) and the amount of DNA per peak needed to generate high quality sequence is ~ 4 million.[65] Thus, to *reliably* generate enough fragments for high quality sequencing, 170 attomoles of amplified DNA are needed per bead.

Kumeresan et al. reported that their best primer coupling reactions yielded 4.4 fmol primer per bead.[63] Using these beads, they were able to generate beads with ~50 attomoles of amplified 624 bp pUC18 DNA from single molecule PCR. In theory, each of these beads contains enough material for a sequencing reaction on chip (>34 attomoles), but if we consider the practical requirements (170 attomoles), they fall short by about 3 times. Increasing the sensitivity of the three times could be accomplished by improving the efficiency at either the emPCR step, the Sanger extension step, the injection step, or the detection step.

The amount of amplicon generated in the PCR is determined by the number of primer molecules coupled to the bead as well as the PCR conditions. There are several ways to increase the amount of amplicon per bead. First, is to use larger beads. The number of amplicons bound to the bead is obviously limited by the number of primers coupled to the bead; however, the amount of primers bound to the bead (0.15 to 4.4 fmol/bd) exceeds the

amount of amplicon/bd, so it is unlikely that increasing the degree of primer labeling will increase the degree of amplicon labeling. The mean diameter for the beads used in these experiments is 34 μm, but the size distribution is such that about 10% of the beads have diameters over 40 μm. These larger beads contain proportionally more primer and amplicon. Filtering the beads to obtain only those over 40 um in diameter could increase the mean amplicon attachment from ~50 to ~90 attomoles.

Increasing the efficiency of the Sanger sequencing reaction and capture would likely yield small but significant increases in signal strength, and with the addition of new scanning and detection technologies the target amount of 170 attomoles of amplicon per beads required for single cell Sanger sequencing seems within reach.

Single cell forensics

Building on the work presented above and previous work done on the Mathies lab, a promising area for further exploration is in single cell forensics. Single tandem repeat (STR) typing is a robust method widely used in the forensics community for establishing the identity of individuals involved in crimes. There is huge demand for reliable methods of typing in low-DNA, mixed DNA or degraded DNA samples which are difficult or impossible to interpret. As the amount of DNA decreases below 1 ng, the chances of obtaining a partial profile increases, reducing the certainty of the identification. One alternative to this problem is to perform mitochondrial sequencing, since there are multiple copies of mtDNA for each genomic DNA copy. However, this does not help in situations with mixed DNA samples. Existing methods that attempt single cell forensic analysis require the use of a micromanipulator or laser under a microscope to isolate single cells.[101] At the time of publication, there have been no single cell STR traces generated with this method, but this technique would have limited usefulness because it is rather tedious.

Typical Process Flow for Forensic STR Typing	Process Flow for Proposed μDG Forensic STR Typing
Cheek swab	Cheek swab
Incubate swab in chelating solution to remove Mg (Reduces activity of DNAses)	Incubate swab in chelating solution to remove Mg (Reduces activity of DNAses)
Lyse cells and extract DNA	Single cell STR PCR in microdroplets
RT-PCR to quantify DNA in sample	FACS to sort successfully amplified target beads
STR PCR	Electrophoresis
Electrophoresis	

Table 2. Comparison of existing bulk STR typing procedure and proposed high-throughput single cell STR typing procedure.

Building on our emPCR single cell genetic analysis methods,[89] illustrated in Figure 18, it should be possible to develop a single cell forensic typing system. Rather than performing STR on samples containing DNA from a mixture of cells, our µDG can be used to generate millions of single cell STR reactions, each containing the full 16 plex STR reaction. This method, outlined in Table 2, will allow easy interpretation of mixed DNA samples since each cell will produce a single STR trace. One advantage of this method is that there will be inherent peak balancing. A major challenge for multiplex PCRs is to identify conditions such that all the amplicons are amplified with approximately equal representation. This challenge will be overcome using our µDG STR since, despite imbalances in the PCR, the 16-primer beads will limit the number of any specific amplicon that can be bound, the electropherogram will have a maximum number of amplicons equal to about 100 attomoles/16 = 6.25 attomoles, which is within our STR detection limits, as shown in Figure 19. Another advantage to switching to µDG-based STR typing is that each bead will carry highly purified and *concentrated* amplified STR fragments, because all amplified material from each single cell reaction will be bound to a 34 µm bead.

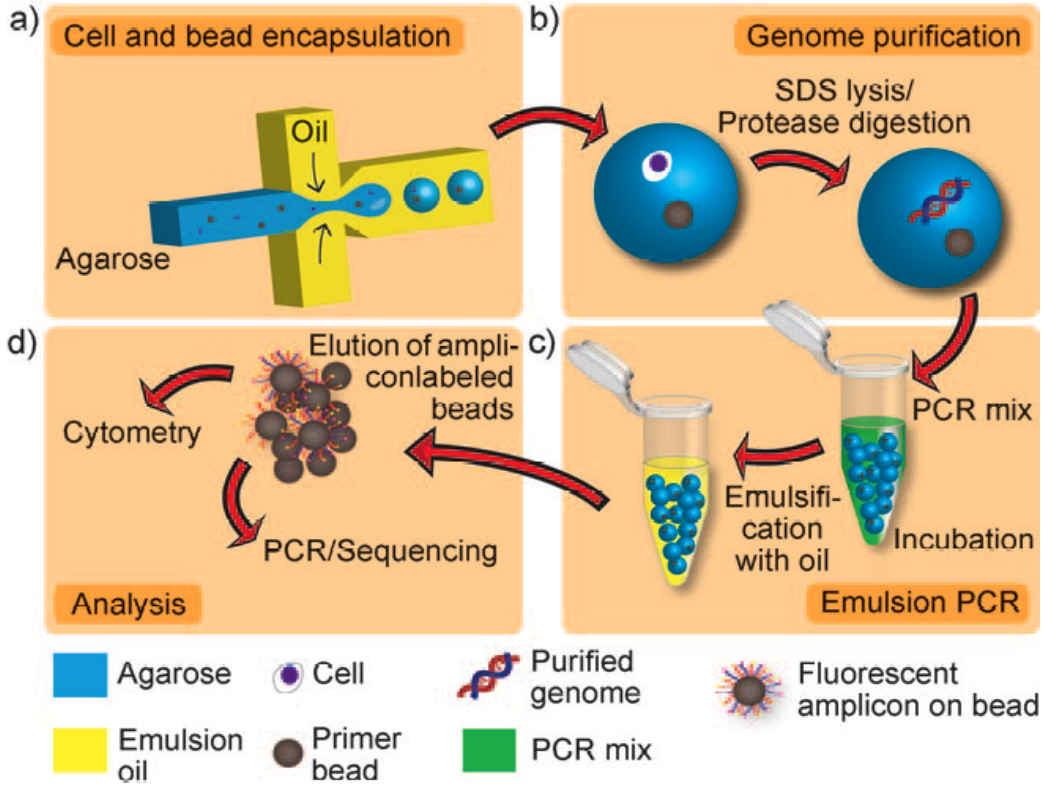

Figure 18. Workflow diagram showing the use of agarose-emulsion droplets for the genetic analysis and multilocus genetic analysis of single mammalian cells. a) Single cells are microfluidically encapsulated together with primer-functionalized beads in agarose-gel droplets. b) The genomes of single cells are released in the gel droplets upon SDS lysis and digestion with proteinase K according to a standard protocol. c) The agarose droplets are equilibrated in PCR buffer containing fluorescent forward primers, emulsified with oil by mechanical agitation, and thermally cycled. d) Following multiplex PCR amplification, beads carrying the PCR products are released by breaking the emulsion and melting the agarose. The fluorescent beads are then rapidly quantified and/or sorted by flow cytometry, and directed to CE devices for genetic analysis. SDS=sodium dodecyl sulfate.

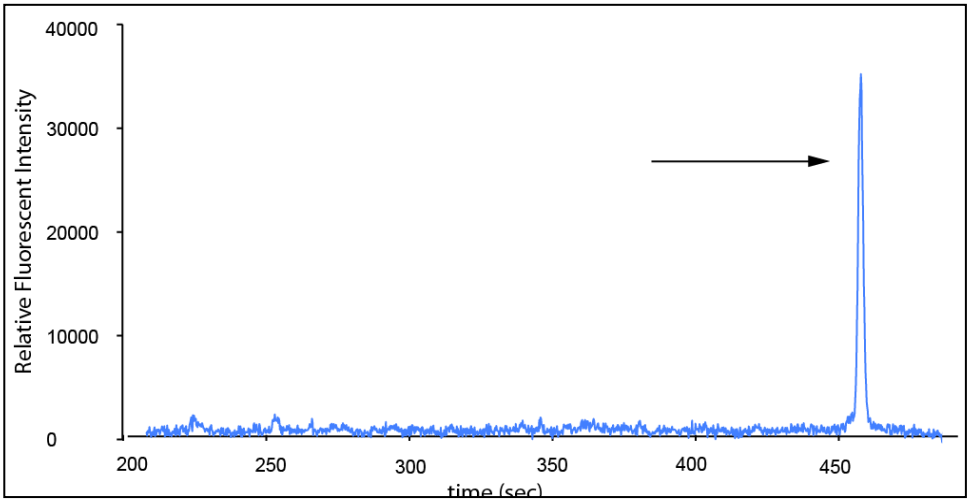

Figure 19. Fluorescent detection of 2 attomoles of FAM labeled DNA oligo (25 bases) using our integrated photopolymerized capture gel system (shown in Figure 20), for purification and concentration. 200 nL of 10 pM oligo was loaded in the reactor and captured in the photopolymerized gel by oligo hybridization. Detection levels of the device are sufficient to enable µDG-based 9-plex STR typing.

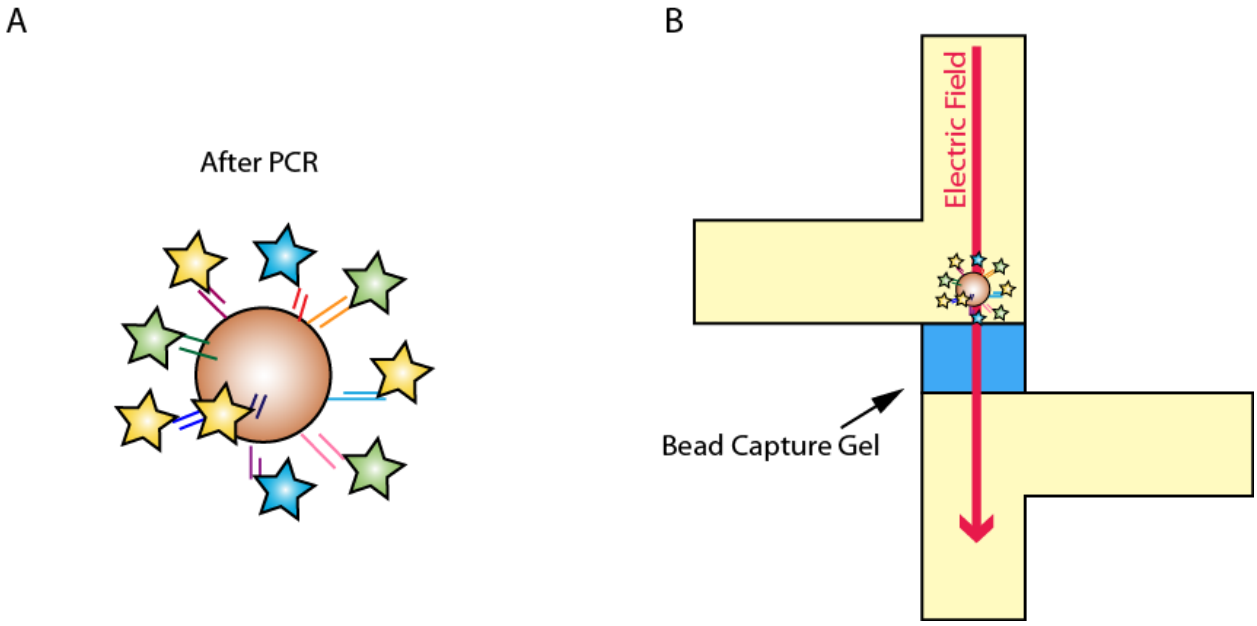

Figure 20. Illustration of the µDG based single cell STR typing method. A) After the 9-plex PCR, beads carrying multiple labeled PCR products are recovered by breaking the emulsion, washing away excess reagents. FACS is used to isolate beads containing amplified DNA. B) Individual beads are captured in photopolymerized capture gels by electrophoresing toward and into the capture gel. Next, a 5% linear polyacrylamide gel is loaded above the bead-capture gel to trap the bead in the offset region and flush out any extra beads. As in our previous designs, the labeled strands of DNA are dehybridized by heating the chip to 65 °C before applying a separation electric field of 150 to 200 V/cm.

Post-PCR bulk washing steps to break the emulsion and remove excess reagents will eliminate the need for on-chip purification and concentration, further streamlining our previous approach,[102] and eliminating losses due to DNA extraction. The photopolymerized capture gels can be used as bead-capture sites, as diagrammed in
Figure 20. Single beads will be driven against the top edge of the gel by an electric field and the presence of exactly one bead will be confirmed by visual inspection under a microscope. If necessary, a 5% LPA gel will be loaded above the capture gel to fix the location of the captured bead and flush out any extra beads. The labeled strands will be released from the bead by heating the whole chip to 67 ˚C and then applying an electric field for CE separation, as in our previous designs. Scaling up to 48 or 96 lanes would be reasonable with this relatively simple glass-glass chip.

In the case of sperm or degraded DNA, each μDG reaction will only represent a subset of the alleles, so multiple cells will have to be typed to reconstruct a complete profile, but this is easily accomplished due to the high throughput of the method, and still benefits from the dramatically increased concentration (in the droplets that contain cells) compared to a bulk reaction.

Conclusions

The technology needed to genetically interrogate single cells has advanced rapidly in the last ten years. Single cell experiments are most meaningful when a significant proportion of the population can be sampled, thus it is critical to develop high throughput automated systems. Microfabricated structures are ideally suited for single cell-scale reactions, automated high throughput single cell manipulation, and analysis because their features can be tuned to the desired length and volume scales. The work presented in this thesis toward the goal of single cell sequencing represents an advance, that points the way to targeted single cell sequencing for analysis of somatic cell mutations, single cell STR typing or even whole genome *de novo* sequencing.

References

(1) Levy, S.; Sutton, G.; Ng, P. C.; Feuk, L.; Halpern, A. L.; Walenz, B. P.; Axelrod, N.; Huang, J.; Kirkness, E. F.; Denisov, G.; Lin, Y.; MacDonald, J. R.; Pang, A. W. C.; Shago, M.; Stockwell, T. B.; Tsiamouri, A.; Bafna, V.; Bansal, V.; Kravitz, S. A.; Busam, D. A.; Beeson, K. Y.; McIntosh, T. C.; Remington, K. A.; Abril, J. F.; Gill, J.; Borman, J.; Rogers, Y.-H.; Frazier, M. E.; Scherer, S. W.; Strausberg, R. L.; Venter, J. C. PLoS Biol 2007, 5, e254.

(2) Venter, J. C.; Remington, K.; Heidelberg, J. F.; Halpern, A. L.; Rusch, D.; Eisen, J. A.; Wu, D.; Paulsen, I.; Nelson, K. E.; Nelson, W.; Fouts, D. E.; Levy, S.; Knap, A. H.; Lomas, M. W.; Nealson, K.; White, O.; Peterson, J.; Hoffman, J.; Parsons, R.; Baden-Tillson, H.; Pfannkoch, C.; Rogers, Y.-H.; Smith, H. O. *Science* **2004**, *304*, 66 -74.

(3) Jung, J. Y.; Lee, S. H.; Kim, J. M.; Park, M. S.; Bae, J.-W.; Hahn, Y.; Madsen, E. L.; Jeon, C. O. *Appl. Environ. Microbiol.* **2011**, *77*, 2264-2274.

(4) Wang, Z.; Gerstein, M.; Snyder, M. *Nat Rev Genet* **2009**, *10*, 57-63.

(5) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463-5467.

(6) Smith, L. M.; Fung, S.; Hunkapiller, M. W.; Hunkapiller, T. J.; Hood, L. E. *Nucleic Acids Res* **1985**, *13*, 2399-2412.

(7) Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. *Nature* **1986**, *321*, 674-679.

(8) Prober, J.; Trainor, G.; Dam, R.; Hobbs, F.; Robertson, C.; Zagursky, R.; Cocuzza, A.; Jensen, M.; Baumeister, K. *Science* **1987**, *238*, 336 -341.

(9) Ju, J.; Ruan, C.; Fuller, C. W.; Glazer, A. N.; Mathies, R. A. *Proceedings of the National Academy of Sciences* **1995**, *92*, 4347 -4351.

(10) Cohen, A. S.; Najarian, D.; Smith, J. A.; Karger, B. L. *Journal of Chromatography A* **1988**, *458*, 323-333.

(11) Ruiz-Martinez, M. C.; Berka, J.; Belenkii, A.; Foret, F.; Miller, A. W.; Karger, B. L. *Analytical Chemistry* **1993**, *65*, 2851-2858.

(12) Karger, B. L.; Foret, F.; Berka, J. In *High Resolution Separation and Analysis of Biological Macromolecules Part B: Applications*; Academic Press, 1996; Vol. 271, pp. 293-319.

(13) Hjertén, S. *Journal of Chromatography A* **1985**, *347*, 191-198.

(14) Mandabhushi, R. S. *Electrophoresis* **1998**, *19*, 224-230.

(15) Lumpkin, O. J.; Déjardin, P.; Zimm, B. H. *Biopolymers* **1985**, *24*, 1573-1593.

(16) Zhou, H.; Miller, A. W.; Sosic, Z.; Buchholz, B.; Barron, A. E.; Kotler, L.; Karger, B. L. *Anal. Chem* **2000**, *72*, 1045-1052.

(17) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C. *Nature* **2001**, *409*, 860-921.

(18) Metzker, M. L. *Nat Rev Genet* **2010**, *11*, 31-46.

(19) Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Wheeler, D. L. *Nucleic Acids Res* **2008**, *36*, D25-30.

(20) Moore, G. E. *Electronics Magazine* **1965**, *38*.

(21) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y.-J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Goodwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jirage, K. B.; Kim, J.-B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, *437*, 376-380.
(22) Dohm, J. C.; Lottaz, C.; Borodina, T.; Himmelbauer, H. *Nucleic Acids Research* **2008**, *36*, e105.
(23) Hillier, L. W.; Marth, G. T.; Quinlan, A. R.; Dooling, D.; Fewell, G.; Barnett, D.; Fox, P.; Glasscock, J. I.; Hickenbotham, M.; Huang, W.; Magrini, V. J.; Richt, R. J.; Sander, S. N.; Stewart, D. A.; Stromberg, M.; Tsung, E. F.; Wylie, T.; Schedl, T.; Wilson, R. K.; Mardis, E. R. *Nat Meth* **2008**, *5*, 183-188.
(24) Harismendy, O.; Ng, P. C.; Strausberg, R. L.; Wang, X.; Stockwell, T. B.; Beeson, K. Y.; Schork, N. J.; Murray, S. S.; Topol, E. J.; Levy, S.; Frazer, K. A. *Genome Biol* **2009**, *10*, R32.
(25) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005**, *309*, 1728 -1732.
(26) McKernan, K.; Blanchard, A.; Kotler, L.; Costa, G. Reagents, methods, and libraries for bead-based sequencing. US Patent 20080003571A1, Jan. 3, **2008**.
(27) Schadt, E. E.; Turner, S.; Kasarskis, A. *Human Molecular Genetics* **2011**, *20*, 853.
(28) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. *Nat Biotech* **2008**, *26*, 1146-1153.
(29) Bayley, H. *Nature* **2010**, *467*, 164-165.
(30) Liu, H.; He, J.; Tang, J.; Liu, H.; Pang, P.; Cao, D.; Krstic, P.; Joseph, S.; Lindsay, S.; Nuckolls, C. *Science* **2010**, *327*, 64-67.
(31) Korbel, J. O.; Urban, A. E.; Affourtit, J. P.; Godwin, B.; Grubert, F.; Simons, J. F.; Kim, P. M.; Palejev, D.; Carriero, N. J.; Du, L.; Taillon, B. E.; Chen, Z.; Tanzer, A.; Saunders, A. C. E.; Chi, J.; Yang, F.; Carter, N. P.; Hurles, M. E.; Weissman, S. M.; Harkins, T. T.; Gerstein, M. B.; Egholm, M.; Snyder, M. *Science* **2007**, *318*, 420 -426.
(32) Stankiewicz, P.; Lupski, J. R. *Annu. Rev. Med.* **2010**, *61*, 437-455.
(33) Pushkarev, D.; Neff, N. F.; Quake, S. R. *Nat Biotech* **2009**, *27*, 847-850.
(34) Gravesen, P.; Branebjerg, J.; Jensen, O. S. *Journal of Micromechanics and Microengineering* **1993**, *3*, 168-182.
(35) Manz, A.; Graber, N.; Widmer, H. M. *Sensors and Actuators B: Chemical* **1990**, *1*, 244-248.
(36) Woolley, A. T.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11348 -11352.
(37) Fredlake, C. P.; Hert, D. G.; Kan, C.-W.; Chiesl, T. N.; Root, B. E.; Forster, R. E.; Barron, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 476 -481.
(38) Paegel, B. M.; Emrich, C. A.; Wedemayer, G. J.; Scherer, J. R.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 574 -579.

(39) Emrich, C. A.; Tian, H.; Medintz, I. L.; Mathies, R. A. *Anal. Chem* **2002**, *74*, 5076-5083.
(40) Shi, Y.; Simpson, P. C.; Scherer, J. R.; Wexler, D.; Skibola, C.; Smith, M. T.; Mathies, R. A. *Anal. Chem* **1999**, *71*, 5354–61.
(41) Unger, M. A.; Chou, H.-P.; Thorsen, T.; Scherer, A.; Quake, S. R. *Science* **2000**, *288*, 113 - 116.
(42) Li, H. Q.; Roberts, D. C.; Steyn, J. L.; Turner, K. T.; Yaglioglu, O.; Hagood, N. W.; Spearing, S. M.; Schmidt, M. A. *Sensors and Actuators A: Physical* **2004**, *111*, 51-56.
(43) Pal, R.; Yang, M.; Johnson, B. N.; Burke, D. T.; Burns, M. A. *Anal. Chem* **2004**, *76*, 3740-3748.
(44) Li, B.; Chen, Q.; Lee, D.-G.; Woolman, J.; Carman, G. P. *Sensors and Actuators A: Physical* **2005**, *117*, 325-330.
(45) Gong, H.; Ramalingam, N.; Chen, L.; Che, J.; Wang, Q.; Wang, Y.; Yang, X.; Yap, P. H. E.; Neo, C. H. *Biomed Microdevices* **2006**, *8*, 167-176.
(46) Grover, W. H.; Skelley, A. M.; Liu, C. N.; Lagally, E. T.; Mathies, R. A. *Sensors and Actuators B: Chemical* **2003**, *89*, 315-323.
(47) Erill, I.; Campoy, S.; Rus, J.; Fonseca, L.; Ivorra, A.; Navarro, Z.; Plaza, J. A.; Aguiló, J.; Barbé, J. *J. Micromech. Microeng.* **2004**, *14*, 1558-1568.
(48) Liu, C. N.; Toriello, N. M.; Mathies, R. A. *Anal. Chem* **2006**, *78*, 5474–5479.
(49) Fiorini, G.; Chiu, D. *Biotechniques* **2005**, *38*, 429-446.
(50) Hühmer, A. F. R.; Landers, J. P. *Analytical Chemistry* **2000**, *72*, 5507-5512.
(51) Lagally, E. T.; Emrich, C. A.; Mathies, R. A. *Lab on a Chip* **2001**, *1*, 102-107.
(52) Redon, R.; Ishikawa, S.; Fitch, K. R.; Feuk, L.; Perry, G. H.; Andrews, T. D.; Fiegler, H.; Shapero, M. H.; Carson, A. R.; Chen, W.; Cho, E. K.; Dallaire, S.; Freeman, J. L.; Gonzalez, J. R.; Gratacos, M.; Huang, J.; Kalaitzopoulos, D.; Komura, D.; MacDonald, J. R.; Marshall, C. R.; Mei, R.; Montgomery, L.; Nishimura, K.; Okamura, K.; Shen, F.; Somerville, M. J.; Tchinda, J.; Valsesia, A.; Woodwark, C.; Yang, F.; Zhang, J.; Zerjal, T.; Zhang, J.; Armengol, L.; Conrad, D. F.; Estivill, X.; Tyler-Smith, C.; Carter, N. P.; Aburatani, H.; Lee, C.; Jones, K. W.; Scherer, S. W.; Hurles, M. E. *Nature* **2006**, *444*, 444-454.
(53) Wong, K. K.; deLeeuw, R. J.; Dosanjh, N. S.; Kimm, L. R.; Cheng, Z.; Horsman, D. E.; MacAulay, C.; Ng, R. T.; Brown, C. J.; Eichler, E. E.; Lam, W. L. *The American Journal of Human Genetics* **2007**, *80*, 91-104.
(54) Levy, S.; Sutton, G.; Ng, P.; Feuk, L.; Halpern, A. *PLoS Biol* **2007**.
(55) Aitman, T.; Dong, R.; Vyse, T.; Norsworthy, P. *Nature* **2006**.
(56) Gonzalez, E.; Kulkarni, H.; Bolivar, H.; Mangano, A. *Science* **2005**.
(57) Fellermann, K.; Stange, D. E.; Schaeffeler, E.; Schmalzl, H.; Wehkamp, J.; Bevins, C. L.; Reinisch, W.; Teml, A.; Schwab, M.; Lichter, P.; Radlwimmer, B.; Stange, E. F. *The American Journal of Human Genetics* **2006**, *79*, 439-448.
(58) Hollox, E.; Huffmeier, U.; Zeeuwen, P.; Palla, R. *Nature genetics* **2007**.
(59) Kidd, J.; Cooper, G.; Donahue, W.; Hayden, H. *Nature* **2008**.
(60) Ju, J.; Ruan, C.; Fuller, C.; Glazer, A.; Mathies, R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**.
(61) Heiger, D. N.; Cohen, A. S.; Karger, B. L. *J. Chromatogr., A* **1990**, *516*, 33-48.
(62) Huang, X.; Quesada, M.; Mathies, R. *Anal. Chem* **1992**.
(63) Kumaresan, P.; Yang, C. J.; Cronier, S. A.; Blazej, R. G.; Mathies, R. A. *Anal. Chem* **2008**, *80*, 3522-3529.
(64) Lagally, E.; Emrich, C.; Mathies, R. *Lab on a Chip* **2001**.

(65) Blazej, R. G.; Kumaresan, P.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7240 -7245.
(66) Blazej, R. G.; Kumaresan, P.; Cronier, S. A.; Mathies, R. A. *Anal. Chem* **2007**, *79*, 4499-4506.
(67) Diehl, F.; Li, M.; He, Y.; Kinzler, K. W.; Vogelstein, B.; Dressman, D. *Nat Meth* **2006**, *3*, 551-559.
(68) Li, M.; Diehl, F.; Dressman, D.; Vogelstein, B.; Kinzler, K. W. *Nat Meth* **2006**, *3*, 95-97.
(69) Teh, S.-Y.; Lin, R.; Hung, L.-H.; Lee, A. P. *Lab Chip* **2008**, *8*, 198.
(70) Beer, N. R.; Hindson, B. J.; Wheeler, E. K.; Hall, S. B.; Rose, K. A.; Kennedy, I. M.; Colston, B. W. *Anal. Chem* **2007**, *79*, 8471-8475.
(71) Zeng, Y.; Novak, R.; Shuga, J.; Smith, M. T.; Mathies, R. A. *Anal. Chem* **2010**, *82*, 3183-3190.
(72) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigó, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.;

Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. *Science* **2001**, *291*, 1304-1351.

(73) International Human Genome Sequencing Consortium *Nature* **2004**, *431*, 931-945.

(74) Mayer, A. N.; Dimmock, D. P.; Arca, M. J.; Bick, D. P.; Verbsky, J. W.; Worthey, E. A.; Jacob, H. J.; Margolis, D. A. *Genetics in Medicine* **2011**, *13*, 195-196.

(75) King, M.-C. *Science* **2011**, *331*, 1026.

(76) Estivill, X.; Armengol, L. *PLoS Genet* **2007**, *3*, e190.

(77) Uddin, M.; Goodman, M.; Erez, O.; Romero, R.; Liu, G.; Islam, M.; Opazo, J. C.; Sherwood, C. C.; Grossman, L. I.; Wildman, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3215 - 3220.

(78) Chin, K.-V.; Selvanayagam, Z. E.; Vittal, R.; Kita, T.; Kudoh, K.; Yang, C. S.; Wong, Y. F.; Cheung, T. H.; Yeo, W.; Chung, T. K. H.; Lin, Y.; Liao, J.; Shih, J. W.; Yap, S. F.; Lin, A. W. *Drug Dev. Res.* **2004**, *62*, 124-133.

(79) Bennett, S. *Pharmacogenomics* **2004**, *5*, 433-438.

(80) Bennett, S. T.; Barnes, C.; Cox, A.; Davies, L.; Brown, C. *Pharmacogenomics* **2005**, *6*, 373-382.

(81) Kidd, J. M.; Cooper, G. M.; Donahue, W. F.; Hayden, H. S.; Sampas, N.; Graves, T.; Hansen, N.; Teague, B.; Alkan, C.; Antonacci, F.; Haugen, E.; Zerr, T.; Yamada, N. A.; Tsang, P.; Newman, T. L.; Tuzun, E.; Cheng, Z.; Ebling, H. M.; Tusneem, N.; David, R.; Gillett, W.; Phelps, K. A.; Weaver, M.; Saranga, D.; Brand, A.; Tao, W.; Gustafson, E.; McKernan, K.; Chen, L.; Malig, M.; Smith, J. D.; Korn, J. M.; McCarroll, S. A.; Altshuler, D. A.; Peiffer, D. A.; Dorschner, M.; Stamatoyannopoulos, J.; Schwartz, D.; Nickerson, D. A.; Mullikin, J. C.; Wilson, R. K.; Bruhn, L.; Olson, M. V.; Kaul, R.; Smith, D. R.; Eichler, E. E. *Nature* **2008**, *453*, 56-64.

(82) Shendure, J.; Porreca, G.; Reppas, N.; Lin, X. *Science* **2005**.

(83) Woolley, A. T.; Mathies, R. A. *Anal. Chem* **1995**, *67*, 3676-3680.

(84) Paegel, B. M.; Emrich, C. A.; Wedemayer, G. J.; Scherer, J. R.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 574 -579.

(85) Aborn, J. H.; El-Difrawy, S. A.; Novotny, M.; Gismondi, E. A.; Lam, R.; Matsudaira, P.; Mckenna, B. K.; O'Neil, T.; Streechon, P.; Ehrlich, D. J. *Lab Chip* **2005**, *5*, 669.

(86) Dittrich, P. S.; Manz, A. *Anal Bioanal Chem* **2005**, *382*, 1771-1782.

(87) Mauk, M. G.; Ziober, B. L.; Chen, Z.; Thompson, J. A.; Bau, H. H. *Ann. N. Y. Acad. Sci* **2007**, *1098*, 467-475.

(88) West, J.; Becker, M.; Tombrink, S.; Manz, A. *Anal. Chem* **2008**, *80*, 4403-4419.

(89) Novak, R.; Zeng, Y.; Shuga, J.; Venugopalan, G.; Fletcher, D. A.; Smith, M. T.; Mathies, R. A. *Angew. Chem. Int. Ed.* **2011**, *50*, 390-395.

(90) d' Amore, F.; Chan, E.; Iqbal, J.; Geng, H.; Young, K.; Xiao, L.; Hess, M. M.; Sanger, W. G.; Smith, L.; Wiuf, C.; Hagberg, O.; Fu, K.; Chan, W. C.; Dave, B. J. *Clinical Cancer Research* **2008**, *14*, 7180 -7187.

(91) McHale, C. M.; Lan, Q.; Corso, C.; Li, G.; Zhang, L.; Vermeulen, R.; Curry, J. D.; Shen, M.; Turakulov, R.; Higuchi, R.; Germer, S.; Yin, S.; Rothman, N.; Smith, M. T. *JNCI Monographs* **2008**, *2008*, 74 -77.

(92) Scherer, J. R.; Kheterpal, I.; Radhakrishnan, A.; Ja, W. W.; Mathies, R. A. *Electrophoresis* **1999**, *20*, 1508-1517.
(93) He, Y.; Pang, H.-M.; Yeung, E. S. *J. Chromatogr., A* **2000**, *894*, 179-190.
(94) Anseth, K. S.; Bowman, C. N.; Brannon-Peppas, L. *Biomaterials* **1996**, *17*, 1647-1657.
(95) Holmes, D. L.; Stellwagen, N. C. *Electrophoresis* **1991**, *12*, 612-619.
(96) Harke, H. R.; Bay, S.; Zhong Zhang, J.; Rocheleau, M. J.; Dovichi, N. J. *J. Chromatogr., A* **1992**, *608*, 143-150.
(97) Brahmasandra, S. N.; Burke, D. T.; Mastrangelo, C. H.; Burns, M. A. *Electrophoresis* **2001**, *22*, 1046-1062.
(98) Ewing, B.; Hillier, L.; Wendl, M. C.; Green, P. *Genome Res* **1998**, *8*, 175 -185.
(99) Hudson, T. *Science* **2011**, *331*, 547.
(100) Shendure, J.; Ji, H. *Nat Biotech* **2008**, *26*, 1135-1145.
(101) Brück, S.; Evers, H.; Heidorn, F.; Müller, U.; Kilper, R.; Verhoff, M. A. *Journal of Forensic Sciences* **2011**, *56*, 176-180.
(102) Liu, P.; Li, X.; Greenspoon, S. A.; Scherer, J. R.; Mathies, R. A. *Lab Chip* **2011**, *11*, 1041.