An Analysis of the Next Generation DNA Sequencing Technology Market

by
Joseph Graham

Submitted to the System Design & Management Program
In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2007]
January 2007

© 2007 Joseph Graham
All Rights Reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or part.

Signature of Author

Joseph Graham
System Design and Management Program
January 2007

Certified By

Fiona Murray
Thesis Supervisor
Assistant Professor, Management of Technology Innovation & Entrepreneurship

Certified By

Noubar Afeyan
Thesis Supervisor
Senior Lecturer, MIT Entrepreneurship Center

Certified By

Patrick Hale
Director
System Design and Management Program

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
FEB 0 5 2008
LIBRARIES

BARKER

An Analysis of the Next Generation DNA Sequencing Technology Market

by
Joseph Graham

Submitted to the System Design & Management Program
on January 19, 2007 in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Engineering and Management

Abstract

While there is no shortage of successful and failed biotechnology ventures, it is still very difficult to gage, a priori, how a new company will fare in this industry. In many cases new biotechnology ventures are driven by rapidly evolving technology and emergent customer needs, both unpredictable by nature. Also, the Biotech Industry faces increased public and federal scrutiny as companies attempt to navigate murky ethical and legal waters. This thesis will explore the ongoing development of the next generation DNA sequencing market in an effort to predict exactly which factors will play a role in determining who will ultimately succeed. This will be accomplished through an analysis incorporating a combination of historical precedents in this industry and traditional market theories. The goal is to produce a set of dimensions along which to judge the current and future participants in this market in order to determine which are most likely to succeed.

Thesis Supervisor: Fiona Murray
Assistant Professor, Management of Technology Innovation & Entrepreneurship

Thesis Supervisor: Noubar Afeyan
Senior Lecturer, MIT Entrepreneurship Center

Table of Contents

Table of Figures 7
Introduction 9
Chapter 1: Background on DNA Sequencing 11
1.1 Human Genome Project 12
1.2 History of the Technology 12
1.3 Market Leader 15
1.4 Technology Assessment 17
1.4.1 Quality 18
1.4.2 Complexity of Sample Preparation 19
1.4.3 Read Length 20
1.4.4 Throughput 22
1.4.5 The Nature of the Data 24
1.5 New Entrant Opportunity 24
1.6 New Applications 25
1.6.1 Gene Expression 25
1.6.2 Genomic Variation 26
1.7 Examples of Next Generation Sequencing Technology (NGST) Companies 28
1.7.1 454 Life Sciences 28
1.7.2 Solexa Inc. 29
1.7.3 Agencourt Personal Genomics 29
1.7.4 Helicos Biosciences Corp. 29
Chapter 2: Conceptual Perspectives on Market Evolution 33
2.1 Technology Push vs. Market Pull 33
2.2 Level of Disruption 33
2.3 Dominant Design 35
2.4 Intellectual Property 36
2.5 Synthesis 37
Chapter 3: Survey 41
3.1 Survey Methodology 41
3.2 Survey Design 41
3.3 Data Analysis 43
3.3.1 Market Overview 44
3.3.2 Technology 47
3.3.3 Intellectual Property 48
Chapter 4: Conclusions 53
References 57
Appendix A: Four Color Sanger Sequencing 61
Appendix B: NIH Next Generation Sequencing Grant Awardees 65
Overview 66
Round 1 – Awarded October 14, 2004 67
Round 2 – Awarded August 8, 2005 70
Round 3 – Awarded October 4, 2006 72
Appendix C: Next Generation Sequencing Company Overview 75
454 Life Sciences 76

Solexa Inc. 78
Sequencing by Ligation – Church Group and Agencourt Personal Genomics 83
Helicos Biosciences 85
Appendix D: Next Generation Sequencing Survey 91
Part 1 – Background 92
Part 2 – Market Overview 93
Part 3 – Technology 94
Part 4 – Intellectual Property 95
Part 5 – Funding 96
Appendix E: Next Generation Sequencing Survey Data 97

Table of Figures

Figure 1 - DNA Overview ... 11
Figure 2 - NCBI Trace Submissions March 2001 to November 2006 ... 14
Figure 3 - Cost to Sequence 1000 Base pairs at the Broad Institute vs. Moore's Law ... 15
Figure 4 - ABI's Revenue from Sequencing Products 2001-2006 ... 16
Figure 5 - ABI Revenue from All DNA Products 2001-2006 ... 17
Figure 6 - Average Error Rates of Thermostable DNA Polymerase During PCR ... 19
Figure 7 - Unique Sequence for Varying Read Length in the Human Genome ... 21
Figure 8 - Unique Sequence for Varying Read Length in λ-phage and E.Coli K12 ... 21
Figure 9 - Typical Sequencing Process ... 23
Figure 10 - Illumina and Affymetrix Revenue 2001-2005 ... 26
Figure 11 - Methods for Detecting Structural Variants in the Human Genome ... 27
Figure 12 - Next Generation Sequencing Company Timeline ... 31
Figure 13 - Christensen's View on Disruptive Technology ... 34
Figure 14 - Utterback's Phase Diagram for New Companies ... 35
Figure 15 - New Market Evolution ... 39
Figure 16 - Ranking of the Factors Motivating Entry into the Sequencing Technology Market ... 44
Figure 17 - Ranking of the Attractiveness of Potential Markets for DNA Sequencing Technology ... 45
Figure 18 - Expected Customer Base for New Sequencing Technologies ... 46
Figure 19 - Expected Relationship to Existing Technologies ... 47
Figure 20 - Ranking of Effort Allocation to Develop the Different Technology Components ... 48
Figure 21 - Perceived IP Benefit ... 49
Figure 22 - Influence of IP Constraints on Technology Design ... 50
Figure 23 - Potential IP Assimilation ... 50
Figure 24 - Comparison of New Technologies to 3730 Performance ... 53

Introduction

The promise of the $1000 genome is a future of personalized medicine where genetically-linked diseases are targeted at their source and eradicated. Progress towards this goal has been slow going and the basic science behind the leading technology has remained unchanged for over ten years. Recently, however, the quest for rapid, inexpensive DNA sequencing technology has been reinvigorated with the emergence of a handful of legitimate new technology companies as well as a host of academic efforts.

The last major round of innovation in DNA sequencing technology occurred in the late 90's and resulted in the emergence of a sole technology provider (Applied Biosystems) which has continued to dominate the market. This begs the question: will history repeat itself; will the latest crop of technology developers be filtered down to one dominant sequencing technology provider that will dominate this market for the next ten years? This question is a complicated one that infers the ability to accurately predict how companies respond in a new market setting with only a short and rapidly evolving history to rely on. Being able to predict the response of well established industries to the introduction of new technology is difficult enough, for relatively new ones it becomes even more so.

This thesis is motivated by a desire to present a novel assessment of this market. In an attempt to sort out this puzzle it is necessary to first identify how this market is evolving including what the true market needs are today and what emergent market needs may surface. Once these needs are identified, the ability of the new technologies to address them can be quantified to some extent based on their claimed performance specifications, strength of intellectual property, and financial security. The weight of these factors will be determined based on the real world experiences of the companies operating in this space.

An introduction to DNA and sequencing will be provided in Chapter 1 in order to familiarize the reader with the history of this ground breaking science. Chapter 2

presents some widely held theories on how to characterize emergent technologies and their evolving market. Whether or not the next generation sequencing technologies conform to these widely held notions on market evolution will be explored. To support the correlation and divergence of this market from traditional models, a survey was circulated to some of the leaders in this field. The results are presented in Chapter 3. Finally, Chapter 4 discusses some conclusions and predictions that can be drawn regarding DNA sequencing and its future, while also suggesting some areas for future research.

Chapter 1: Background on DNA Sequencing

Even before Watson and Crick's historic unraveling of the structure of Deoxyribonucleic Acid (DNA) in 1953[1], scientists realized the importance of this molecule and its implications. DNA serves as the storehouse for inherited traits by encoding the information necessary for every cell making up an organism to behave and interact in a manner that sustains life. Consisting of the nucleic acids, adenine, thymine, guanine, and cytosine, referred to as DNA bases, which are linked together in long chains to form chromosomes, DNA serves as an intricate template which the cellular machinery uses to produce proteins and monitor cellular processes. This is done through a cascade of events in which genes, or specific encoding sequences scattered throughout an organism's DNA, are read and converted.

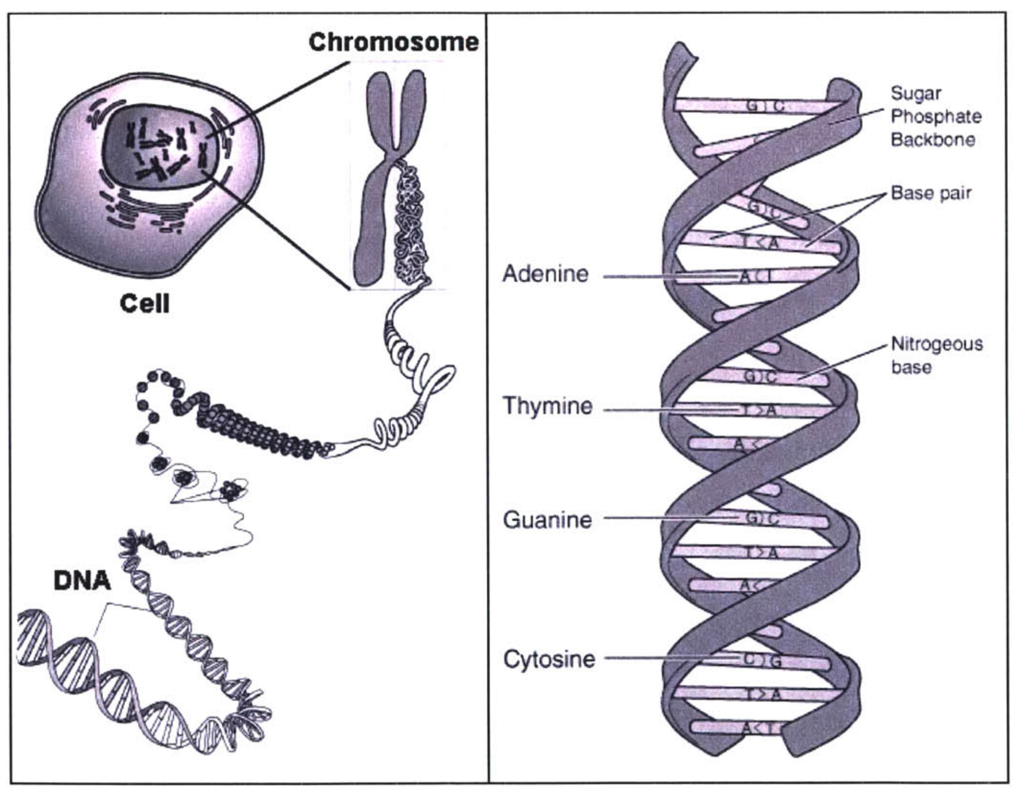


Figure 1 - DNA Overview[2]

---

[1] Watson, J., Crick, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737, 1953.
[2] From http://www.koshland-science-museum.org/ and http://biology.quickseek.com/

1.1 Human Genome Project

A Genome refers to the entire complement of DNA that defines an organism. For humans the genome consists of ~3 billion bases of DNA arranged into 13 pairs of chromosomes. Each of the trillions of cells in the human body, except white blood cells contains an entire copy of that person's genome. While each person is believed to be ~99.9% the same genetically, the 0.1% difference correlates to about 3 million bases. In 1990 the NIH and DOE announced an ambitious project to sequence the entire genome of a representative group of humans that would serve as a template for further research.[3] Having a reference genome, it was surmised, would enable researchers to more easily elucidate genetically linked diseases, identify potential drug targets, and gain a deeper understanding of cell function and the cell-environment interaction.

Originally slated to take about 15 years, the sequencing effort finished years ahead of schedule thanks to rapid improvement in DNA sequencing technologies driven by competition. The publication of 'The Initial Sequencing and Analysis of the Human Genome' in Nature magazine in 2001[4], represented a remarkable step forward for scientific research and ushered in the post-genomic era where a reference genome is the first step toward the ability to rapidly and reliably sequence and compare DNA sequence.

1.2 History of the Technology

The ability to systematically sequence DNA was introduced in 1977 when Fred Sanger published his groundbreaking paper 'DNA Sequencing with Chain-Terminating Inhibitors'[5] which ultimately earned him a Nobel Prize. It took over another 20 years before the technology had advanced enough in regards to throughput and cost to even suggest tackling something as large as the human genome.

---

[3] U.S. Department of Health and Human Services and Department of Energy, Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years, April 1990.

[4] International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. *Nature*, Vol. 409: 860-921, 2001.

[5] Sanger, F., Nicklen, S., Coulson, A. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Science*,74(12):5463-5467, December 1977.

The problem with DNA sequencing is one of scale and throughput. A strand of DNA measures just 22 to 24 angstroms across and the average unit (a base pair) is only 3.3 angstroms long resulting in the need for some means to translate the information encoded by the sequence of bases into a detectable signal. Sanger sequencing relies on DNA polymerase combined with fluorescently labeled bases to create many copies of an unknown strand of DNA incorporating a detectable fluorescent signal. This method then relies on a sieving matrix to sort out the labeled DNA by size enabling the determination of both the identity and order of the sequence. The basic process is outlined in Appendix A.

Since its invention Sanger sequencing has been improved through the incorporation of thermally-stable polymerase, utilization of four different fluorescent dyes, invention of thermal cycling instruments, and introduction of capillary electrophoresis. Each of these improvements led to an increase in the amount of DNA that could be sequenced in a given time period and a reduction in the complexity of the operation. While many attribute the accelerated completion of the Human Genome Project to a competition that emerged with a private company (Celera Genomics), which was also attempting to sequence the Human Genome ahead of the public effort, it is certain that the timeline for both efforts benefited from the introduction of capillary sequencers.

Figure 2 reflects the rapid increase in trace submissions (a trace consisted of a continuous sequence of DNA typically around 700 base pairs long) to the National Center for biotechnology Information (NCBI, the database where all of the NIH funded sequence is made publicly available) between 2001 and 2006. This curve directly correlates with the introductions of new capillary sequencers by Applied Biosystems, the ABI 3700 in the late 90's and the ABI 3730 in 2002. While there were other technology providers attempting to meet the demand for low cost, high-throughput sequencing, these instruments dominated. This was reflected in the fact that they were the technology choice of three of the largest sequencing centers (The Whitehead Institute, The Sanger Center, and Celera Genomics).

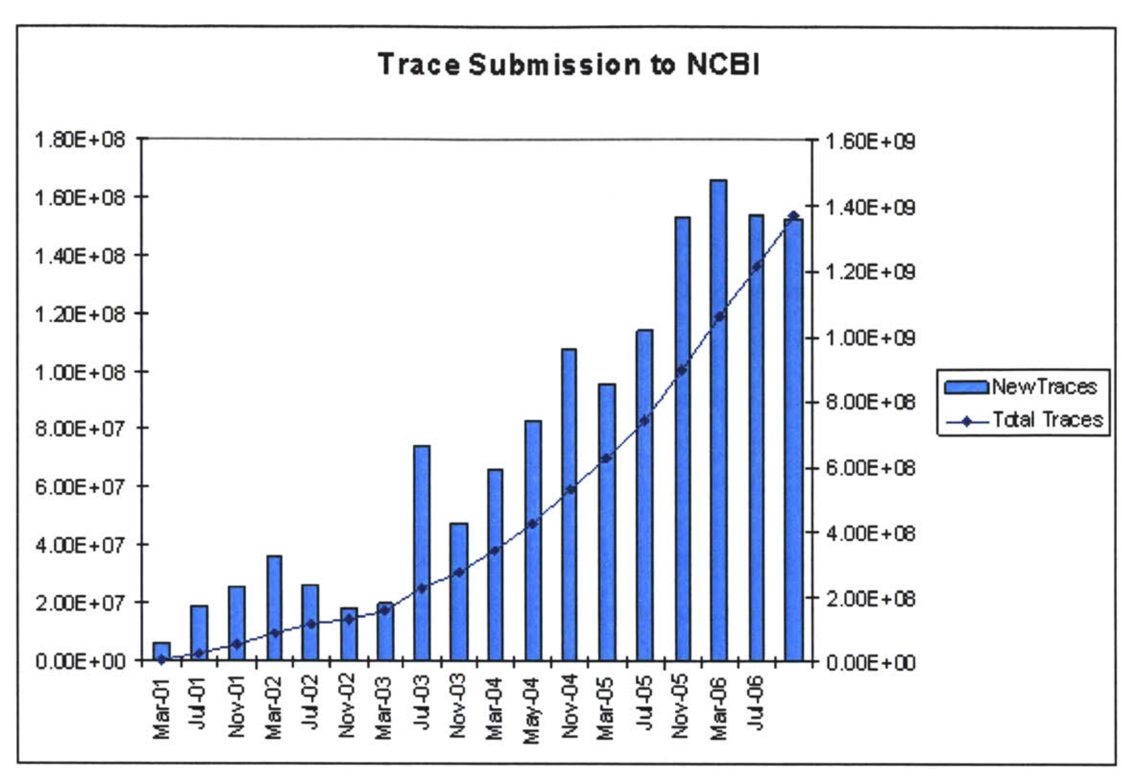


Figure 2 - NCBI Trace Submissions March 2001 to November 2006[6]

Advances in technology were accompanied by the emergence and scale-up of large genome sequencing centers which were able to take advantage of economies of scale and innovative new production methods to drive down the cost of sequencing down. This advantage was achieved through the introduction of automation and the ability to negotiate reduced pricing from vendors with larger volumes. Figure 3 shows the declining cost of sequencing 1000 bases of DNA experienced at the Broad Institute of MIT and Harvard (originally the Genome Sequencing Center of the Whitehead Institute) mapped against Moore's Law. Moore's Law reflects the fact that the cost per transistor has been declining inversely as the achievable density of transistors on a chip. This annual 2 fold decrease in cost per year has held up over the last 40 years and has often been compared to the declining cost of DNA sequencing. This comparison is interesting when considered in terms of what the scaling of chip technology has enabled and what the scaling of sequencing technology hopes to enable.

[6] From NCBI website, http://www.ncbi.nlm.nih.gov/

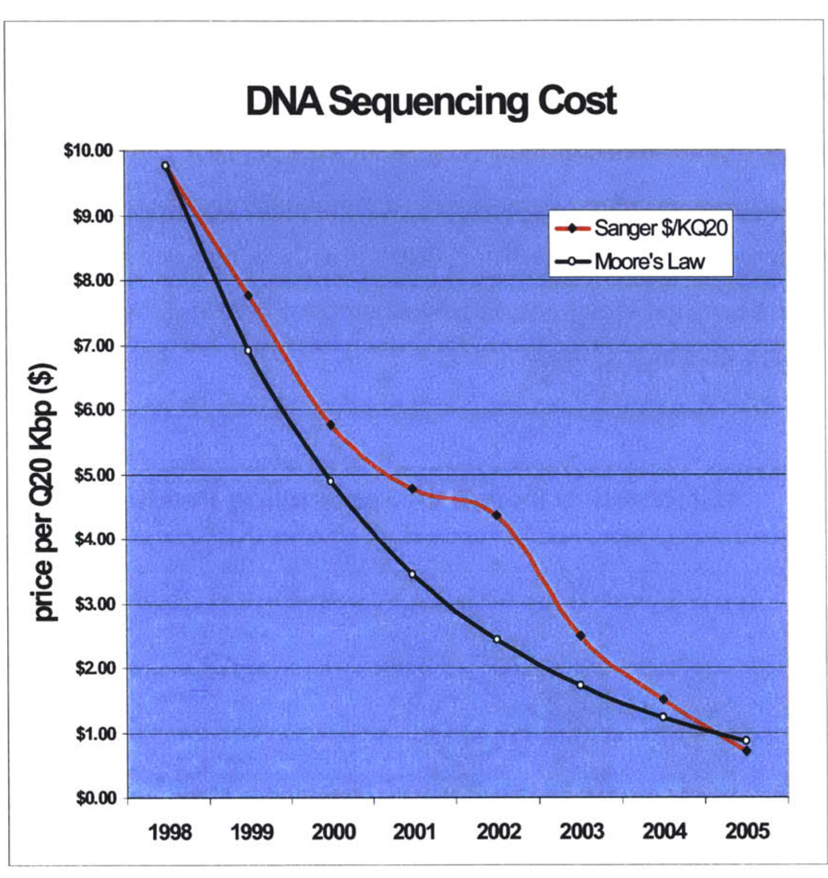


Figure 3 - Cost to Sequence 1000 Base pairs at the Broad Institute vs. Moore's Law[7]

1.3 Market Leader

As mentioned earlier the Human Genome Project helped to enhance Applied Biosystems' position as the market leader in DNA sequencing technology. Strategic alliances with the larger labs involved in the HGP enabled ABI's 3700 instrument to quickly outsell its nearest competitor, Amersham Biosciences. By the time the 3730 (an upgrade from the 3700's 16 capillaries to 96 along with the addition of automated sample loading) was introduced in May of 2002, ABI was already entrenched in the market, enjoying annual sales of over $700 million by 2001.

Looking at ABI's declining profits and revenue from DNA sequencing products from 2001 to 2006 (Figure 4), it is not obvious what makes this market so attractive to new

[7] Nicol, R. Production Sequencing at the Broad Institute, 2006.

entrants. There are at least three explanations for this decline. First, although the NIH continues to fund the sequencing of large organisms, there has been no additional capacity scale-up at the large centers resulting in a net decrease in instrumentation sales. At the same time, the centers continue to drive the cost per base down as they refine and optimize their internal sequencing processes (thereby reducing the quantity of consumables they need to purchase from ABI). Finally, with a reference genome in hand, many researchers have set their sites on new avenues of investigation, utilizing technologies other than sequencing.

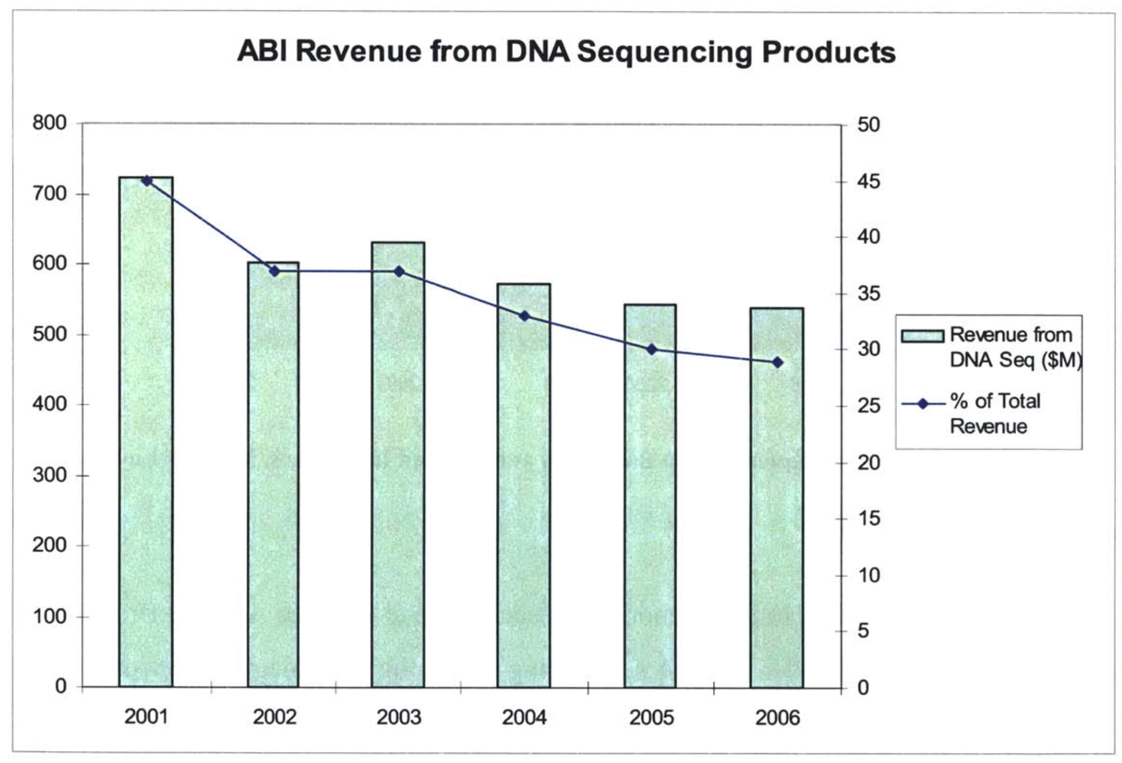


Figure 4 - ABI's Revenue from Sequencing Products 2001-2006[8]

One of the new areas of investigation is gene expression analysis, which will be discussed in greater detail later in the new applications section. ABI also has products that support this type of research which are reflected in Figure 5. Given the noticeable upward trend in revenues from these additional products, it is not difficult to see how a

[8] From ABI Annual Reports 2002-2007

technology that both encompassed and extended the capabilities of the existing technologies generating this revenue would be considered a reasonable investment.

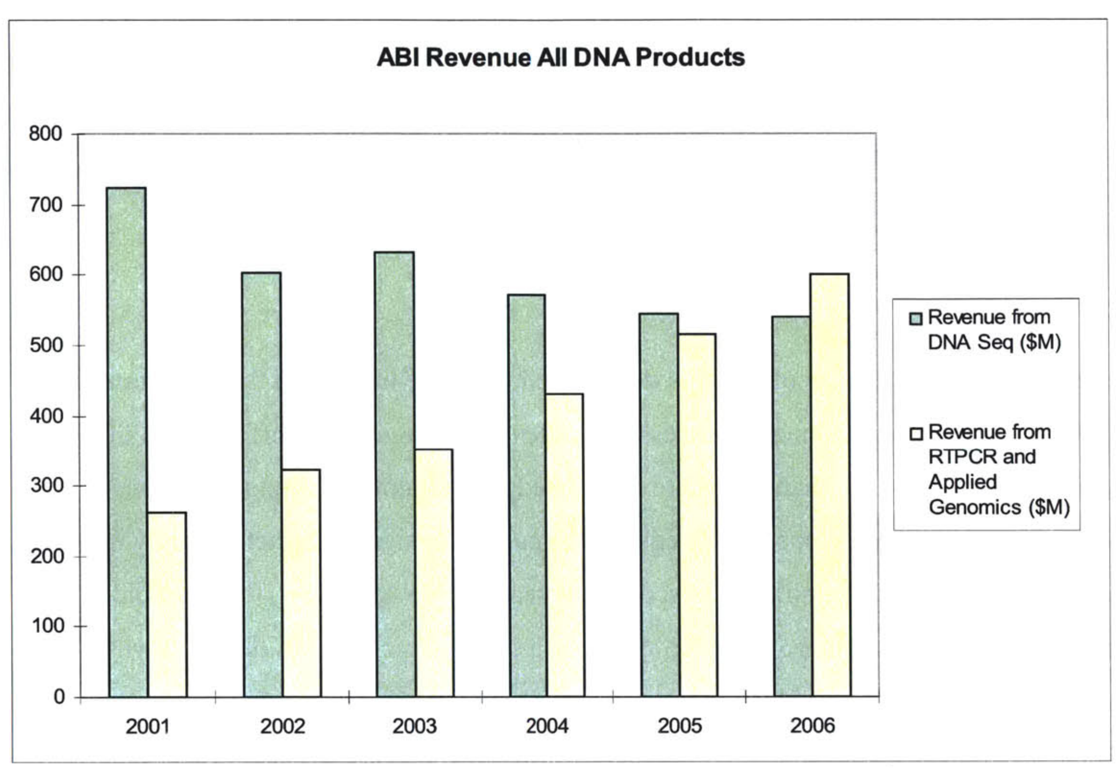


Figure 5 - ABI Revenue from All DNA Products 2001-2006[9]

1.4 Technology Assessment

The performance of DNA sequencing technology is assessed along a few different dimensions that vary in importance depending on the application being addressed. The primary dimensions are: Quality; Complexity of Sample Preparation; Read Length; Throughput (Bases/Hour, $/Base); and the Nature of the Final Data. For optimal analysis, DNA sequencing technology would be able to read off an entire chromosome, end to end, without interruption (this would be >200 million bases for human chromosome 1) and perhaps be able sequence an entire human genome in the same amount of time it takes a cell to create a copy of its genome when dividing (approximately 10 hours during the synthesis phase of the cell cycle). Unfortunately, the greatest continuous length of DNA commonly sequenced by the market leading technology is five orders of magnitude less

[9] From ABI Annual Reports 2002-2007

than the typical chromosome (~800 base pairs) and takes more than an hour to generate. A decreased signal to noise ratio is the primary obstacle to extending Sanger sequencing far beyond this threshold. This is why the current approach to sequencing relies massively redundant sequencing feeding complex algorithms that take data from many short reads and reassemble them into larger continuous segments based on overlaps. These short reads are characterized by the parameters listed above which this section will explore further.

1.4.1 Quality

The quality of a sequence of DNA is determined to be the likelihood that an errant base was reported in the sequencing process. In order to determine the performance of a sequencing instrument, many runs are conducted with samples of known sequence. Correct and errant sequence is then equated to particular signal characteristics which are referenced in later runs and used to score new reads. The ABI 3730, for example, relies on laser excitation of fluorophores to generate a signal which is then detected with a CCD camera. Characteristics of the received signal such as intensity, duration, and conformation to an expected wavelength are then used to score the individual base calls.

For the HGP a standard was created known as Phred20 which represented the minimum quality needed for a base to be submitted into NCBI. Phred20 equates to one error in every hundred bases reported. In order to improve on this error rate, sequencing projects rely on levels of coverage. In the HGP the typical base received 10X coverage, meaning the same base location was observed in at least 10 different reads.

A human's genome is a diploid meaning there are two sets of chromosomes (one inherited from each parent) resulting in two different but correct sets of sequence for each person. In addition, genetic mutations, such as those seen in some cancers, can introduce additional sequence variation. As a result, it is necessary for sequencing technology to have the capacity to differentiate between an errant base and true genetic variation which has proven very difficult with current methods.

1.4.2 Complexity of Sample Preparation

The Phred20 standard is meant to report on errors arising strictly from the detection instrumentation itself, another source of error in DNA sequencing, however, is the sample preparation process. Prior to actually being loaded onto a sequencing instrument, the sample of interest undergoes any number of rigorous preparation steps, each of which may introduce errors into the sequence that may be reported as high quality bases later on. Processes such as clonal amplification, polymorphous chain reaction, rolling circle amplification, and the Sanger sequencing reaction, generate many copies of the target sequence and each have error rates associated with them. While these error rates are typically very low (see Figure 6), the prep may utilize multiple amplification steps that are exponential in nature and involve hundreds of thousands of molecules, meaning any error can propagate and amplify, possibly rendering it indistinguishable from actual genetic variation.

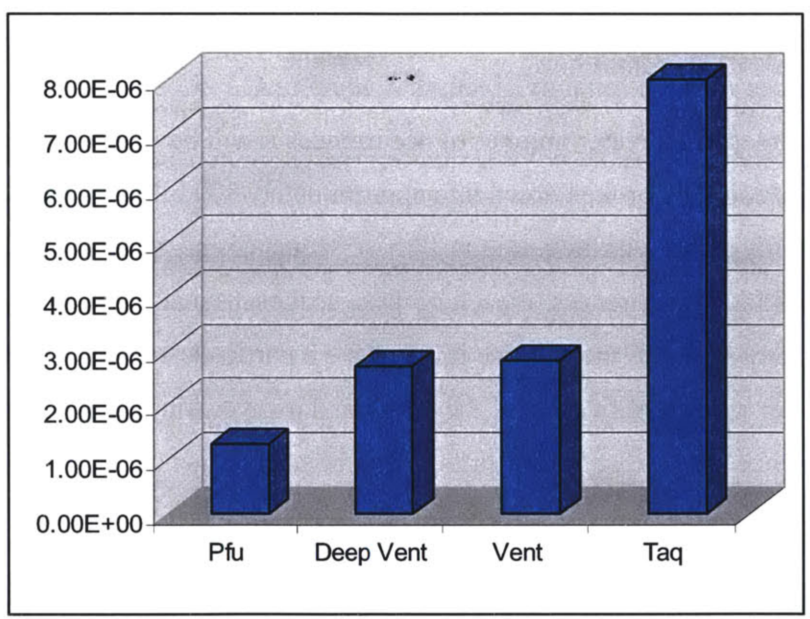


Figure 6 - Average Error Rates of Thermostable DNA Polymerase During PCR
The error rates in this chart were calculated using the equation ER = mf (bpxd), where mf is the mutation frequency, bp is the number of detectable sites, and d is the number of template doublings[10]

[10] Cline, J., et al. PCR Fidelity of Pfu DNA Polymerase and other Thermostable DNA Polymerases. *Nucleic Acids Research*, Vol. 24(18): 3547, 1996

There are other less quantifiable errors that can be attributed to the level of complexity inherent in sample prep. Many of these can be reduced through automation and refinement, but the most effective way to eliminate errors is through simplification of the process. In addition to limiting errors, a streamlined process speeds up the learning curve for potential customers making it a much easier sell. The total amount of amplification steps along with the total time to prep one sample will be used in this technology assessment as a proxy for a complexity measurement. The standard Sanger sequencing process consists of 3 amplification steps that occur over a 33 hour time span.

1.4.3 Read Length

As mentioned earlier, the most advanced sequencing methods to date rely on read lengths that are on the order of 800 bases that are assembled together based on overlapping sequence to form larger segments or contigs. Read length is a very important characteristic of sequencing data because it translates directly into the ability to accurately assemble a read. DNA is a basically a quaternary system with the number of unique sequences possible for a given length X equal to 4^X. Consequently, the shorter the length of a read, the greater number of occurrences it will have in a genome. The human genome could be broken down into approximately 300 billion 10 base sequences and there are only about 1 million unique 10 base sequences possible. Thus, one would expect to find ~3000 occurrences of each of these sequences making it difficult to determine exactly where in the genome it may have a particular read may have originated from. This is an oversimplification of the problem, however, since it is well known that there are sequence biases that further limit the number of sequence variations for any read length that actually occur.

In their 2005 Nuclei Acids Research paper 'An Analysis of the Feasibility of Short Read Resequencing', Whiteford et al. investigated the utility of short read lengths based on their ability to accurately reassemble them onto a reference sequence. Figure 7(a) shows the percentage of the Human Genome that can be uniquely characterized by reads of a given read length. The authors argue that an approach that relied on read lengths of at least 43 base pairs could only hope to uniquely characterize ~90% of the genome. Figure

8 (a) and (b) shows the same analysis for the much smaller genomes of λ–phage and E.Coli K12. In both of these cases, read lengths of 15 base pairs is adequate to account for over 95% of the genome.

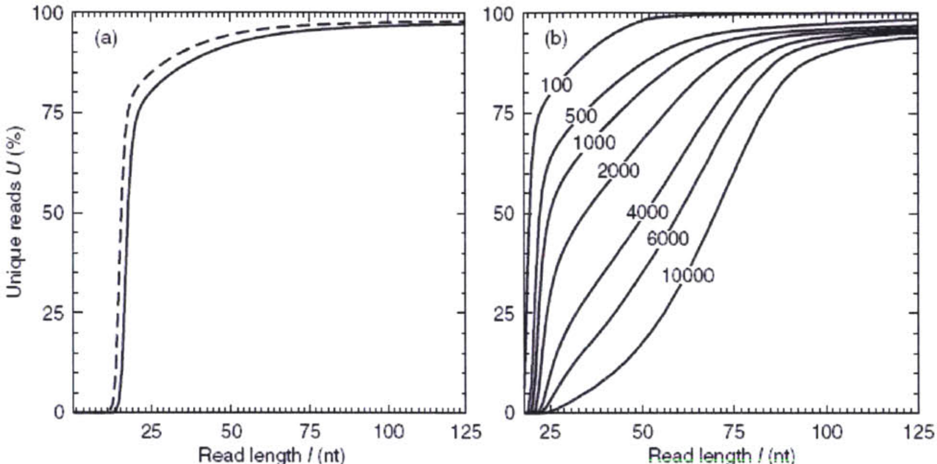


Figure 7 - Unique Sequence for Varying Read Length in the Human Genome
On the right is the percentage of unique sequence in the Human Genome for given read lengths and on the right is the percentage of Chromosome 1 covered by contigs greater than a threshold length as a function of read length[11]

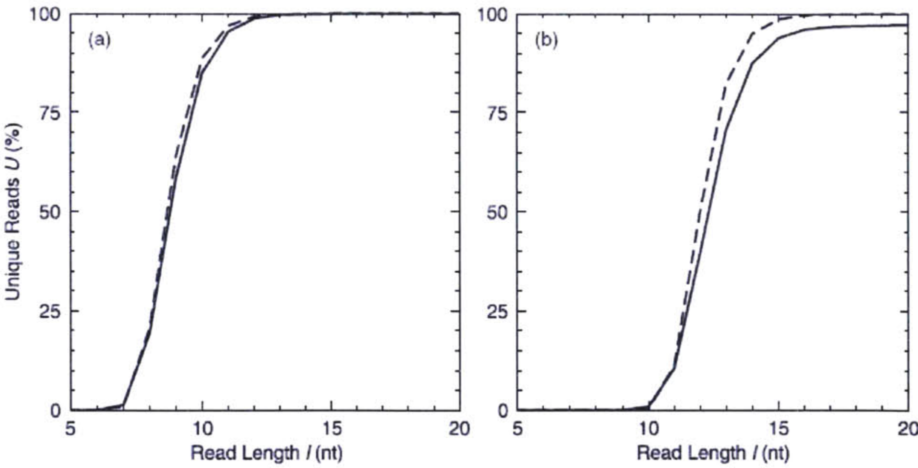


Figure 8 - Unique Sequence for Varying Read Length in λ–phage and E.Coli K12
The percentage of unique sequence in the λ–phage(a) and E.Coli K12 (b) genomes for given read lengths[12]

[11] Whiteford, N., Haslam, N., Weber, G., Prugel-Bennett, A., Essex, J., Roach, P., Bradley, M., Nelyon, C. An Analysis of the Feasibility of Short Read Resequencing. *Nucleic Acides Research*, 33(19) p.3, 2005.
[12] Whiteford, N., Haslam, N., Weber, G., Prugel-Bennett, A., Essex, J., Roach, P., Bradley, M., Nelyon, C. An Analysis of the Feasibility of Short Read Resequencing. *Nucleic Acides Research*, 33(19) p.4, 2005.

This study puts the importance of read length into perspective. With the available human genome template, the authors argue that 'whole genome resequencing will be limited with current technology' but that 'partitioning the problem, by focusing on single chromosomes or by neglecting the more difficult and repetitive parts of the genome, makes the problem more tractable'[13].

Much effort has been devoted in the development of new sequencing technologies to extend the read length as far as possible. The greatest limitation for many of the next generation efforts is the fact that the detection method is serial and noise is introduced with each successive base measurement making each additional base in the read more difficult than the one before it and severely limiting the opportunity for substantial improvement. One way some technologies have sought to extend their read length is through the use of paired ends. Paired Ends refers to the ability to sequence both ends of a continuous stretch of DNA of known length. If you knew the approximate length of a strand of DNA, two 25 base pair reads from its ends, is essentially one 50 base pair read for the assembly algorithm.

It is obvious that read length will play a major role when determining the utility of new sequencing efforts, but it is not yet obvious what the optimal read length will be for the different applications. With more reference sequence to rely on every day and more complex analysis algorithms being employed on more powerful computing hardware, it is probable that the theoretical required read lengths will continue to shrink. The simple fact remains, however, that longer read lengths are better.

1.4.4 Throughput

The throughput of a sequencing technology is often what is advertised, although it can be difficult to determine exactly what is being referenced. For the purposes of this analysis, throughput will consist of both the time and cost components of an associated technology.

---

[13] Whiteford, N., Haslam, N., Weber, G., Prugel-Bennett, A., Essex, J., Roach, P., Bradley, M., Nelyon, C. An Analysis of the Feasibility of Short Read Resequencing. *Nucleic Acides Research*, 33(19) p.5, 2005.

In order to establish a metrics for comparison, some simplification will be required. Consider a representative sequencing process as characterized by Figure 9.

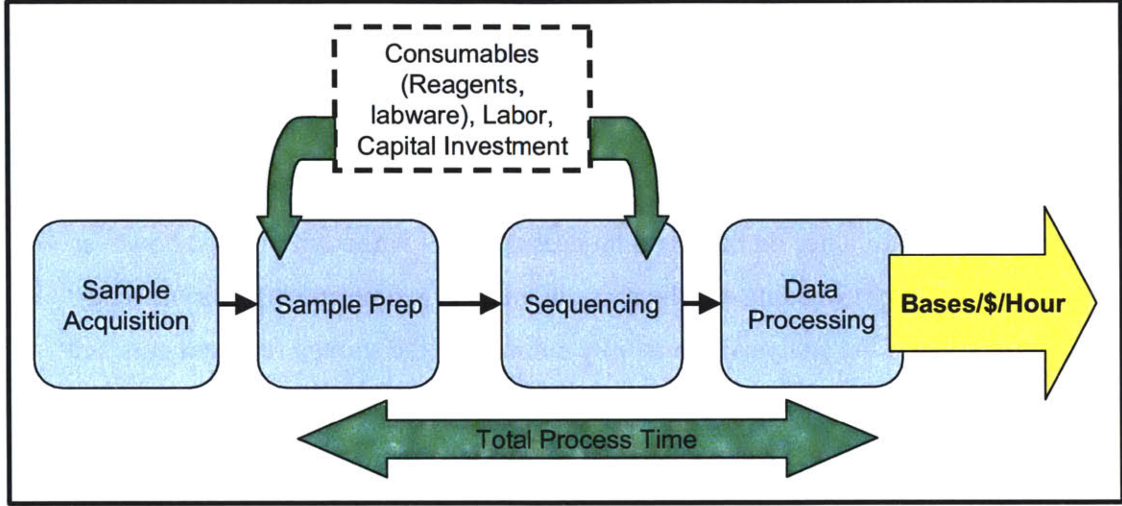


Figure 9 - Typical Sequencing Process

The inputs to this process, aside from the actual sample, are the consumables, such as proprietary reagents and labware (sample plates, pipettes, etc.), labor, and an upfront capital investment in the sequencing detection instrumentation. Assuming that the cost of the instrumentation will amortize over the course of many sequencing runs, this component can be ignored, leaving consumables and labor as the primary determinants of cost. The output of a sequencing process is bases of DNA.

For the ABI 3730, it costs approximately $0.001 per base of sequence in labor and consumable costs. As mentioned earlier, it takes approximately 33 hours to prepare a sample and another 1.1 hours to run a sample on the detector with each run yielding around 80,000 bases of sequence. Since all sample prep can be highly parallelized, the time associated with this portion of the process can be ignored and throughput can be calculated based on a constant feed to the detector; this translates into a throughput of 73,000 bases/hour at a cost of $73. This means you could sequence the human genome once over in 41,000 hours (~5 years) for a cost of $3 million with a single ABI instrument.

1.4.5 The Nature of the Data

Fast, inexpensive computational power has enabled researchers to extract meaningful information more efficiently from ever growing data sets. This trend has contributed significantly to the DNA sequencing efforts. Complex signal processing combined with even more complex statistical analysis, has opened the doorway to high throughput detection technologies that can afford to measure significant amounts of spurious data without impacting their ability to generate an even greater amount of meaningful information. This reliance on back end processing is not without it pitfalls, however. Time and effort spent weeding out the noise in the data is time and effort sacrificed elsewhere. Even if the analysis step is fully automated, the storage requirements and transmission rates for the data generated can pose a formidable challenge to a typical research lab. Also, while complex statistical algorithms may help extend the usefulness of short read lengths (as discussed previously) they can only go so far before they begin to impact the reported quality of the reads. It is for these reasons that it is important to keep in mind what type of data is being generated and how much post processing is involved with a particular sequencing technology when assessing its utility.

1.5 New Entrant Opportunity

Although dramatic improvements have occurred in the rate and quality of DNA sequencing over the last decade, large projects are still primarily relegated to the large genome centers by virtue of their infrastructure and specialized knowledge. Following the completion of the HGP, many new specialized sequencing applications emerged for which traditional Sanger sequencing was clearly not ideal. The ability to quickly and inexpensively generate smaller sets of sequence information for many individuals promised an instant market for whoever could invent it first. In an effort to further spur on the development of new technologies, the NHGRI awarded a number of sequencing grants:

> *NHGRI's near-term goal is to lower the cost of sequencing a mammalian-sized genome to $100,000, which would enable researchers to sequence the genomes of hundreds or even thousands of people as part of studies to identify genes that contribute to cancer, diabetes and other common diseases.*

Ultimately, NHGRI's vision is to cut the cost of whole-genome sequencing to $1,000 or less, which would enable the sequencing of individual genomes as part of medical care. The ability to sequence each person's genome cost-effectively could give rise to more individualized strategies for diagnosing, treating and preventing disease. Such information could enable doctors to tailor therapies to each person's unique genetic profile.[14]

The NHGRI awarded three rounds of grants in 2004, 2005, and 2006 to 30 different researchers and companies totaling over $83 million (a list of awardees can be found in Appendix B). This was followed in 2006 by calls for applicants for the XPrize award for genomics that would award $10 million to the effort which could:

develop radically new technology that will dramatically reduce the time and cost of sequencing genomes, and accelerate a new era of predictive and personalized medicine. The X PRIZE Foundation aims to enable the development of low-cost diagnostic sequencing of human genomes.[15]

With the promise guaranteed market and the fanfare created by such publicized funding opportunities, it is easy to understand the allure of this market for both technology developers and potential investors.

1.6 New Applications

Inherent in the definition of the Next Generation Sequencing Technologies is the ability to extend the utility of sequence information through new applications. Gene expression analysis, mutation detection, copy number variation, genotyping, and targeted resequencing are just some of the new areas that being explored.

1.6.1 Gene Expression

Gene expression profiling has been an important tool in biological research for some time. Just knowing a person's genetic makeup does not tell the whole story. The specific genes and the rate at which they are translated depend on a number of factors that are difficult to measure. One method that has been employed successfully to determine quantitatively, which genes are active, has been to rely on the levels of mRNA (an intermediary in the

[14] NHGRI press release: NHGRI Seeks Next Generation of Sequencing Technologies. October 14, 2004.
[15] From Xprize website, http://genomics.xprize.org/

gene expression cascade) in a cell. A technology that could improve upon current performance of this measurement would be an instant success.

Currently companies such as Illumina and Affymetrix market proprietary technology for conducting gene expression analysis. As Figure 10 shows, both of these companies experienced revenue growth between 2001 and 2005 that would be attractive to any new entrant.

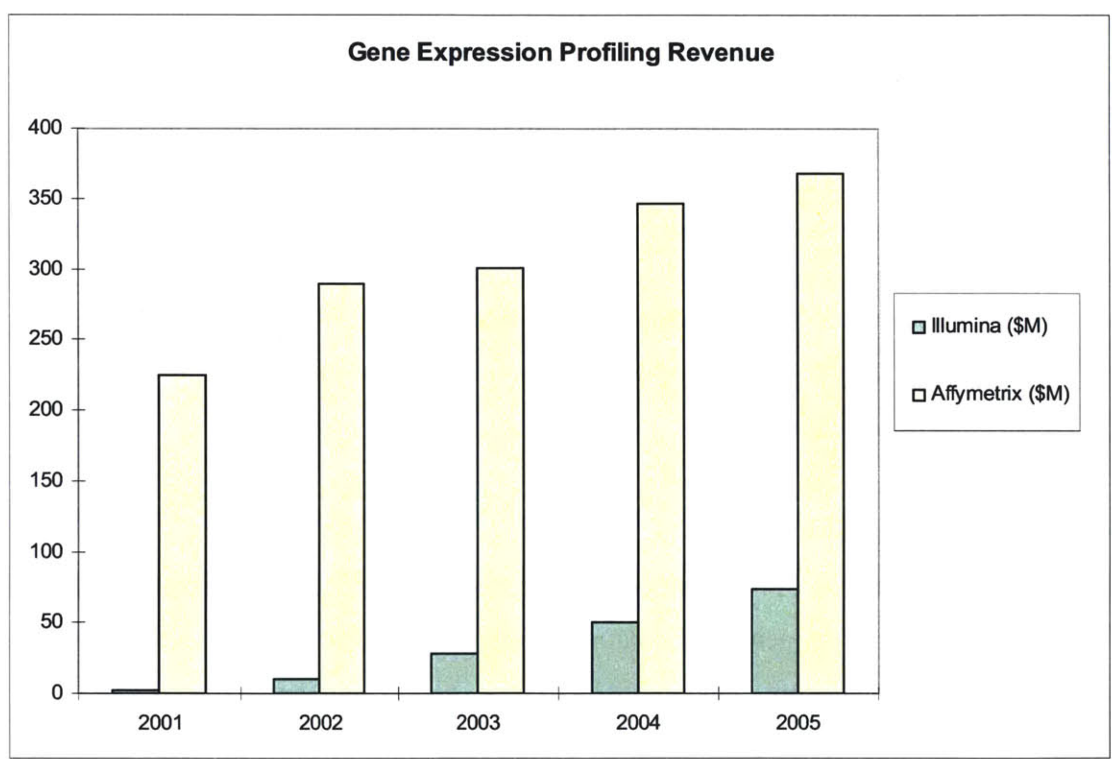


Figure 10 - Illumina and Affymetrix Revenue 2001-2005[16]

1.6.2 Genomic Variation

Initially Single Nucleotide Polymorphisms (SNPs) were thought to comprise the majority of interspecies genomic variation. This led to the formation of the International Hap Map Consortium, which sought to decipher a roadmap of SNPs that could be used as a shortcut to determine underlying structural variation in the genome which could highlight evolutionary change as well as identify possible disease markers. In 2006, however, a

[16] From Illumina, Affymetrix Annual Reports 2002-2006

paper by Feuk et al. reviewed recent research resulting from genome-scanning technologies which revealed that deletions, duplications and large scale copy number variants (CNVs), insertions, inversions, and translocations are all present in the genome at levels that far exceed the number of SNPs. Later in 2006, Redon et al. estimated that the level of CNVs within the genome was somewhere on the order of 12%, showing that the initial estimates of 99.9% person to person identity was a gross overestimate[17].

The implication of the high level of genomic variation is that simple SNP measurement assays will most likely not provide an adequate detection method. In addition, most of the genome scanning technologies being utilized cannot provide a complete assessment of genomic variation. As seen in Figure 11, only the highly inefficient 'Southern Blotting' and 'Sequence-assembly comparison' methods can hope to capture information on all types of variation.

Method	Translocation	Inversion	LCV (>50 kb)	CNV indel (1–50 kb)	Small sequence variants (<1 kb)
Genome-wide scans					
Karyotyping	Yes (>3 Mb)	Yes (>3 Mb)	Yes (>3 Mb)	No	No
Clone-based array-CGH	No	No	Yes (>50 kb)	No	No
Oligonucleotide-based array-CGH	No	No	Yes (>35 kb)	Yes (>35 kb)	No
SNP array	No	No	Yes	Yes	Yes (SNPs)
Sequence-assembly comparison	Yes	Yes	Yes	Yes	Yes
Clone paired-end sequencing (fosmid)	Yes	Yes (breakpoints)	Yes (>8 kb of deletions)	Yes (>8 kb of deletions; <40 kb of insertions)	No
Targeted scans					
Microsatellite genotyping	No	No	Yes (deletions)	Yes (deletions)	Yes
MAPH	No	No	Yes	Yes	Yes
MLPA	No	No	Yes	Yes	Yes
QMPSF	No	No	Yes	Yes	Yes
Real-time qPCR	No	No	Yes	Yes	Yes
FISH	Yes	Yes	Yes	Yes	No
Southern blotting	Yes	Yes	Yes	Yes	Yes

Detection limits are shown in parentheses where applicable. The emphasis is on those approaches that are used for the detection of submicroscopic variants, although karyotyping is also shown. For comparison, each technology's ability to detect smaller sequence variants (<1 kb) is also shown. CGH, comparative genome hybridization; CNV, copy-number variant; FISH, fluorescence in situ hybridization (including metaphase, interphase and fibre FISH); indel, insertion or deletion; LCV, large-scale CNV; MAPH, multiplex amplifiable probe hybridization; MLPA, multiplex ligation-dependent probe amplification; QMPSF, quantitative multiplex PCR of short fluorescent fragments; qPCR, quantitative PCR.

Figure 11 - Methods for Detecting Structural Variants in the Human Genome[18]

[17] Redon, R., et al Global Variation in Copy Number in the Human Genome. *Nature*, 444 p.444, Nov. 2006.
[18] Feuk, L., Carson, A., Scherer, S. Structural Variation in the Human Genome. *Nature Reviews Genetics*, 7 p.88, Feb. 2006.

Once again it is obvious that a sequencing method that could provide the same information as the existing technologies listed in Figure 11, in a more cost and time efficient manner, would have no problem conquering the market. Whether or not the Next Generation Sequencing Technologies can meet the required specifications will be an important determinant in their possible success.

1.7 Examples of Next Generation Sequencing Technology (NGST) Companies

While there are many researchers pursuing new sequencing technologies, a handful of companies have taken an early lead in establishing themselves as the next big advancement. This section will briefly highlight four of these efforts: 454 Life Sciences; Solexa Inc.; Agencourt Personal Genomics; and Helicos Biosciences Corp. A timeline of some of the major events in the evolution of these companies is shown in Figure 12. Combined, these companies offer tremendous insight into the sequencing market. The experiences of these companies and the dimensions along which they have chosen to innovate and compete, is the basis for the analysis in Chapter 2 and the survey detailed in Chapter 3. In other words, the best way to determine the important aspects of this market is through the observation of those who are attempting to enter it. It is important to note both the similarities and differences in the expected specifications for each of these technologies.

It is interesting that in addition to the companies explored here, there are a set of efforts looking at even more radical sequencing solutions. While those closest to market seek to make the $100,000 genome reality, the $1000 is being actively pursued by those employing completely different technologies such as Nanopores and Force Spectroscopy (see Appendix B).

1.7.1 454 Life Sciences

Started in June of 2000, 454 Life Sciences was the first company to bring a product to market, installing their first unit in April of 2005. Relying on a miniaturized Pyrosequencing platform combined with emulsion PCR (as outlined in Appendix C) 454 published results in 2005 that showed their technology was capable of generating about

500,000 100 base pair reads in 4 hours with a Phred20 quality[19]. While this represented a considerable increase in raw throughput over the 3730, the technology was not well suited for some applications due to a difficulty sequencing long sequence repeats. 454 is currently a wholly owned subsidiary of Curagen Corporation and has an exclusive distribution agreement with Roche

1.7.2 Solexa Inc.

Solexa Inc. is the oldest of the companies that will be discussed and has undergone considerable change in its lifetime. Following its acquisition of Lynx Therapeutics Inc. in early 2005, Solexa emerged as one of the front runners in this market. Solexa shipped its first instruments to customers in July of 2006 with published specifications of 1 billion bases per run broken into 25 base lengths with a pass rate of 90% delivered in a 3 day run. This process relies on a bridge PCR step followed by extension sequencing with reversible terminators (see Appendix C for more details). Solexa was acquired by Illumina Inc. in November of 2006.

1.7.3 Agencourt Personal Genomics

Agencourt Personal Genomics was a spin-off of Agencourt Bioscience following its acquisition by Beckman Coultier. APG was then acquired by ABI in May of 2005. APG technology is a variation on sequencing by ligation which was initially licensed from George Church's lab at Harvard. Appendix C explains the Polony technology employed at the Church lab. Like the Polony approach APG's technology relies on emulsion PCR and sequencing by ligation, and hopes to achieve ~25 base pair reads on a chip yielding 1 billion bases in a 3 day run.

1.7.4 Helicos Biosciences Corp.

Helicos Biosciences Corp. is the only one in this group that claims to be true 'single molecule sequencing'. This is due to the fact that there is no initial PCR step prior to sequencing, which is accomplished via extension sequencing with reversible terminators

[19] Margulies, M., et al. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature*, Vol.437:378, September 2005.

(see Appendix C for more details). Helicos' instrument, due for a 2007 release, promises 1 billion bases of high quality sequence per day.

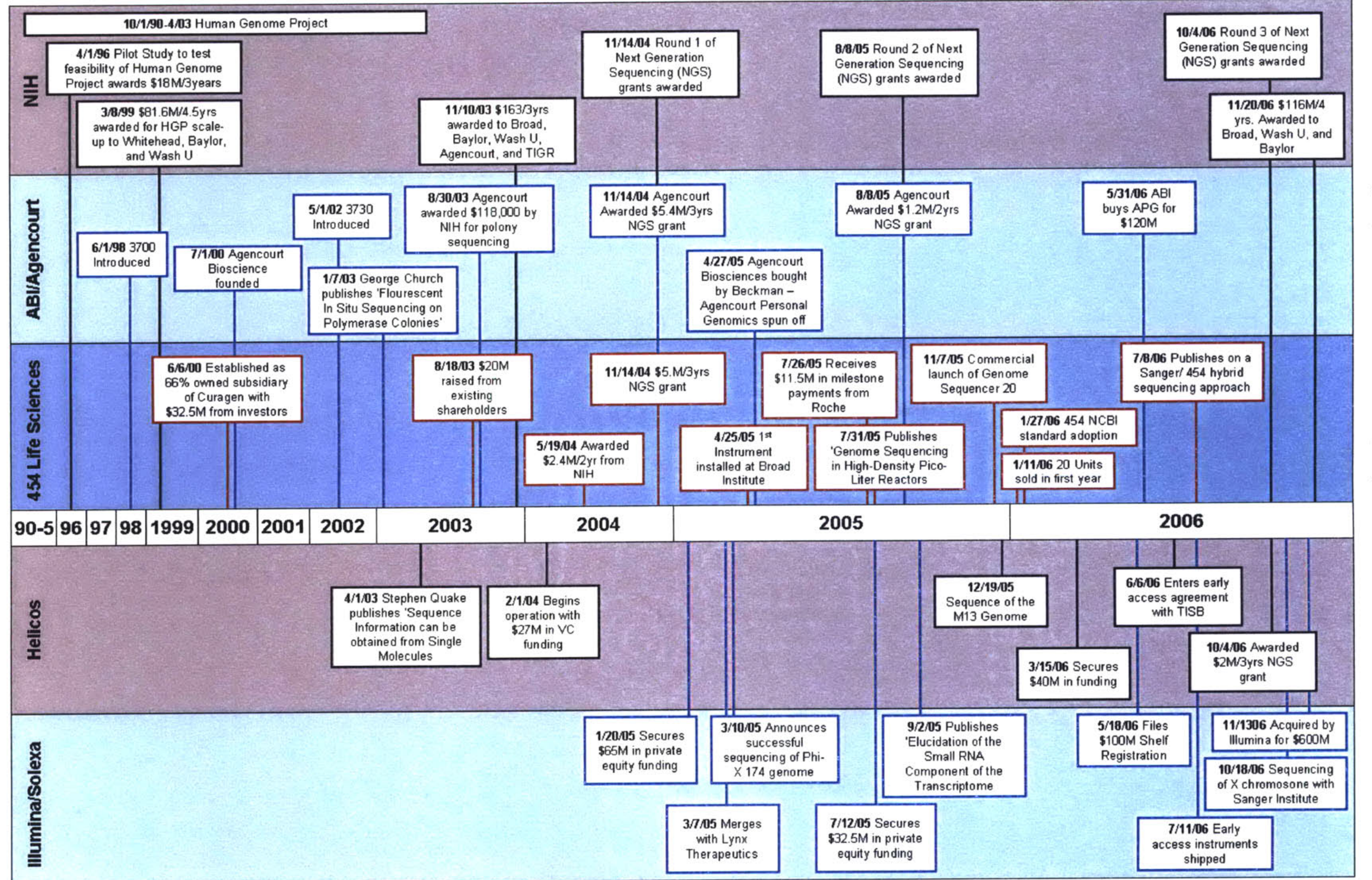


Figure 12 - Next Generation Sequencing Company Timeline

Chapter 2: Conceptual Perspectives on Market Evolution

The development and maturation of new technology markets has been examined from many different theoretical perspectives, all of which offer some insight into the underlying mechanisms dictating this evolution. The following sections will explore a handful of these theories that provide an insightful description of the Next Generation Sequencing Market.

2.1 Technology Push vs. Market Pull

There have been many studies arguing the prevalence of either 'Technology Push' or 'Market Pull' as the primary determinant of industry evolution. 'Technology Push' is the concept that industry is driven by new technological invention and discovery[20]. 'Market Pull', on the other hand argues that underlying unmet market needs determine how industries change and which endeavors are ultimately successful[21]. The fact that there is support for both sides reinforces the view that it is most likely a combination of these forces and their interplay driving any new market. The degree of this interplay is determined by many factors, internal and external, including maturity of the industry, level of competition, and intensity of demand. Knowing which force is the primary driver of a market (aware that this can and does change over time) is an important step in being able to assess the achievements of the participants.

2.2 Level of Disruption

The concept of disruptive technology is an apt description of most emerging technology markets. As defined by Christensen in *The Innovator's Dilemma*, disruptive technologies are those which:

> *...bring to market a very different value proposition than had been available previously. Generally disruptive technologies underperform established products in mainstream markets. But they have other features that a few fringe (and generally*

[20] Chidamber, S. R., Kon, H. B. A Research Retrospective of Innovation Inception and Success: The Technology-Push Demand-Pull Question. *International Journal of Technology Management*, 9(1), 1994.
[21] Chidamber, S. R., Kon, H. B. A Research Retrospective of Innovation Inception and Success: The Technology-Push Demand-Pull Question. *International Journal of Technology Management*, 9(1), 1994.

new) customers value. Products that are based on disruptive technologies are typically cheaper, simpler, smaller, and, frequently, more convenient to use[22].

Christensen also argues that 'markets that do not exist cannot be analyzed' referring to the inability of established companies to predict and prepare for the impact of a disruptive technology.

Figure 13 outlines the evolution of disruptive technologies with reference to the established product. As the chart indicates, not all 'invading technologies' are successful in their attempts to overtake the current market leader. The ability to accurately assign such a development curve to a new technology company would be an invaluable tool in predicting that company's ultimate level of success in the market.

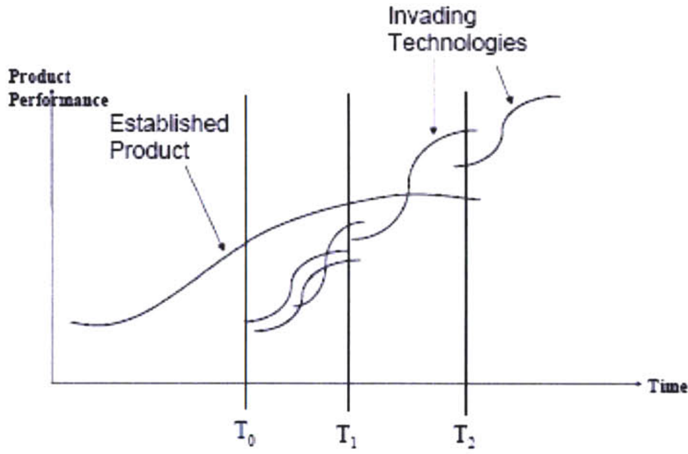


Figure 13 - Christensen's View on Disruptive Technology

One way to define a company's development curve is suggested by Utterback in *Mastering the Dynamics of Innovation*[23]. Figure 14 illustrates the two types of innovation proposed for a new technology company. Initially there is a high level of product innovation as the technology is altered and refined in an attempt to address a range of applications. This is followed by a period of process innovation, where

[22] Christensen, C. *The Innovator's Dilemma.* HarperCollins, New York, NY 2003. p.xviii
[23] Utterback, J. Mastering the Dynamics of Innovation. Harvard Business School Press. Boston, MA 1996. p. 91

production costs are reduced in response to a reduction in the need to support a high degree of product variation. The combined area under the two of these curves can be interpreted in order to generate the product performance S curves seen in Figure 13. Therefore, determining the ongoing level of product innovation can lead to a prediction of the shape of a company's S curve and its advancement along it. This can be reflected in forward looking statements by the company concerning planned technology improvements and the degree to which a company is focused on particular markets as mentioned earlier.

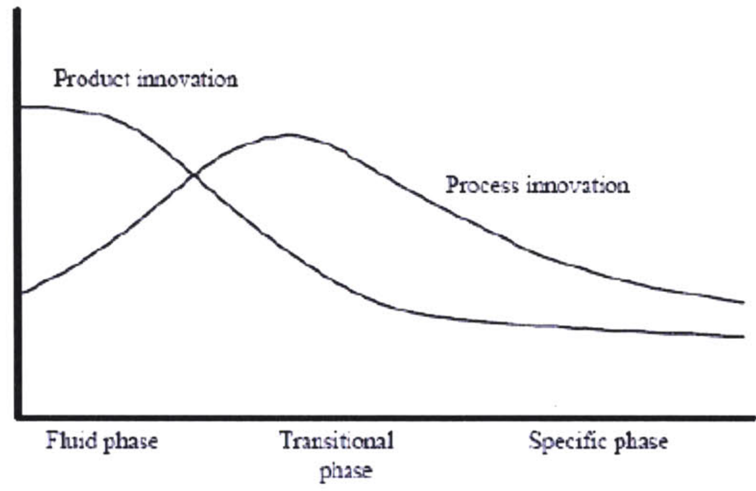


Figure 14 - Utterback's Phase Diagram for New Companies[24]

2.3 Dominant Design

As a market evolves the natural progression is towards what is known as a dominant design environment. As identified by Tushman and Anderson, a dominant design results when 'a technological breakthrough, or discontinuity, initiates an era of intense technical variation and selection'[25]. A dominant design typically represents the optimal solution available to meet the identified market needs. Inherent in this definition is the fact the dominant design is not necessarily the most technologically advanced solution, but rather

[24] Utterback, J. Mastering the Dynamics of Innovation. Harvard Business School Press. Boston, MA 1996. p. 91

[25] Tushman, M. L., Anderson, P. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Administrative Science Quaterly*, Vol. 35, 1990.

the one that addresses the financial and accessibility demands of the consumers while delivering the necessary technological performance. As defined by James Utterback in Mastering the Dynamics of Innovation, a dominant design:

> *...is the one that wins the allegiance of the marketplace, the one that competitors and innovators must adhere to if they hope to command significant market following. The dominant design usually takes the form of a new product (or set of features) synthesized from individual technological innovations introduced independently in prior product variants.*[26]

Once established, a dominant design is subject to competition from other copycat technologies that seek to capitalize on the market standardization. The degree to which the different competitive technologies in a given market vary is a good indication of how far along the path to a dominant design it is.

2.4 Intellectual Property

In all technology industries, intellectual property plays an integral role in enabling a company to succeed. A company needs to innovate and to support its innovation with a foundation of solid IP in addition to providing the resources to defend it. At the same time, a company must have extensive knowledge of the existing IP in its market space in order to avoid wasting time developing technology it can never own and possibly falling victim to costly lawsuits brought by its competitors. In his work 'Competition, Cooperation, and Innovation', Teece addresses the problems of both too little and overaggressive IP strategies. As he explains, one major market failure involves the inability of firms to prevent other firms from incorporating their technology resulting in a 'free rider' environment. Even when direct imitation is prevented, he adds, 'there is likely to be a technological neighborhood illuminated by the innovation that is not foreclosed by the patent'[27]. How to measure the strength of IP is a different matter. Lacking litigation as a reference for the strength of a company's IP, the size of the portfolio and the importance allotted it are good indicators.

---

[26] Utterback, J. Mastering the Dynamics of Innovation. Harvard Business School Press. Boston, MA 1996. p. 24

[27] Teece, D. J. Competition, Cooperation, and Innovation: Organizational Arrangements for Regimes of Rapid Technological Progress. *Journal of Economic Behavior and Organization*, 18(1):1-25, June 1992.

2.5 Synthesis

The topics discussed in the previous section were meant to provide an evaluation framework in order to help provide a clearer picture of how this new technology market may evolve. First it is necessary to discuss how each of these topics relates specifically to the NGST market.

While it is difficult to determine the degree to which a company is being driven by the force of market pull, it is possible to gage the amount of application focus a company exhibits. This, in turn, can be viewed as a reflection of market influence. For the Next Generation Sequencing companies, this is a function of the potential applications that were highlighted in section 1.6. All of the applications discussed are potentially very lucrative, but they also require a significant amount of work to develop. The implied trade off here is to focus on a particular application in order to secure that customer base in favor of pursuing a larger customer base by addressing many applications with less effectiveness.

When the NGST market is evaluated with respect to the traditional de novo sequencing application, the technology certainly seems to fit the definition of a 'disruptive technology'. The shorter lengths and reduced quality of the reads generated by the NGSTs, would have rendered them highly inefficient if they were employed exclusively in the Human Genome Project. Also, while they seem to enjoy advantages in other parameters (200 fold increase in throughput and a 100 fold reduction in cost) described in Section 1.4, as of yet, none of the NGST have managed to corner the market. As far as the inability to analyze a disruptive market, the fact that the next big applications in Genomic sequencing have only generally been identified is merely an indication of the relative maturity of this market. For this market Figure 13 can be used to map the progress of these new technologies to the established ABI technology.

Regarding the concept of dominant design, all of the companies described earlier share a high degree of similarity. All rely on a massively parallel approach to sequencing with

reduced read length. This fact seems to suggest that the market is more or less converging on a technology standard for the next generation in DNA sequencing. But as mentioned earlier, there are many efforts underway that rely on an entirely different approach to sequencing which represent an entirely different tier of development. Determining which dominant design appears to be relevant to a given technology will play a role in determining how to differentiate these different tiers of development.

The NGST market is currently full of competitors all vying for the same customers. The fact that they are all relying on a scientifically similar approach (see Appendix C) in order to address the customers' needs is certain to result in some overlapping technology. It is likely that once the handful of competitors mentioned earlier all have instruments on the market, some IP infringement suits will follow. This conclusion is based on the history of this market (ABI has been involved in numerous IP infringement lawsuits) and recent occurrences (ABI has recently sued Solexa over patents relating to SBL which threaten their technology acquired with APG). When this happens it will be the company with the deepest pockets and largest IP portfolio that will win out, assuming there is no dramatic advantage to any of the different technologies.

Based on the above observations, it is possible to postulate some measurements of success for this market. The four dimensions that can be extracted are: Degree of Market Focus; Phase of Development; Degree of Differentiation from Competition; and Strength of IP. Figure 15 is meant to show an overview of how these dimensions interact.

Illustrated in the figure is how, as a technology development effort evolves, its application pool shrinks, or to put it another way, its Degree of Market Focus is enhanced as the developers identify and focus on the most promising applications. Also represented, though not explicitly, in this figure is the Development Phase which is the relationship between the forces of market pull and technology push. As the technology migrates toward a dominant design, the strength of Market Pull increases representing Utterback's transition from the 'Fluid' to the 'Transitional' and finally the 'Specific Phase'.

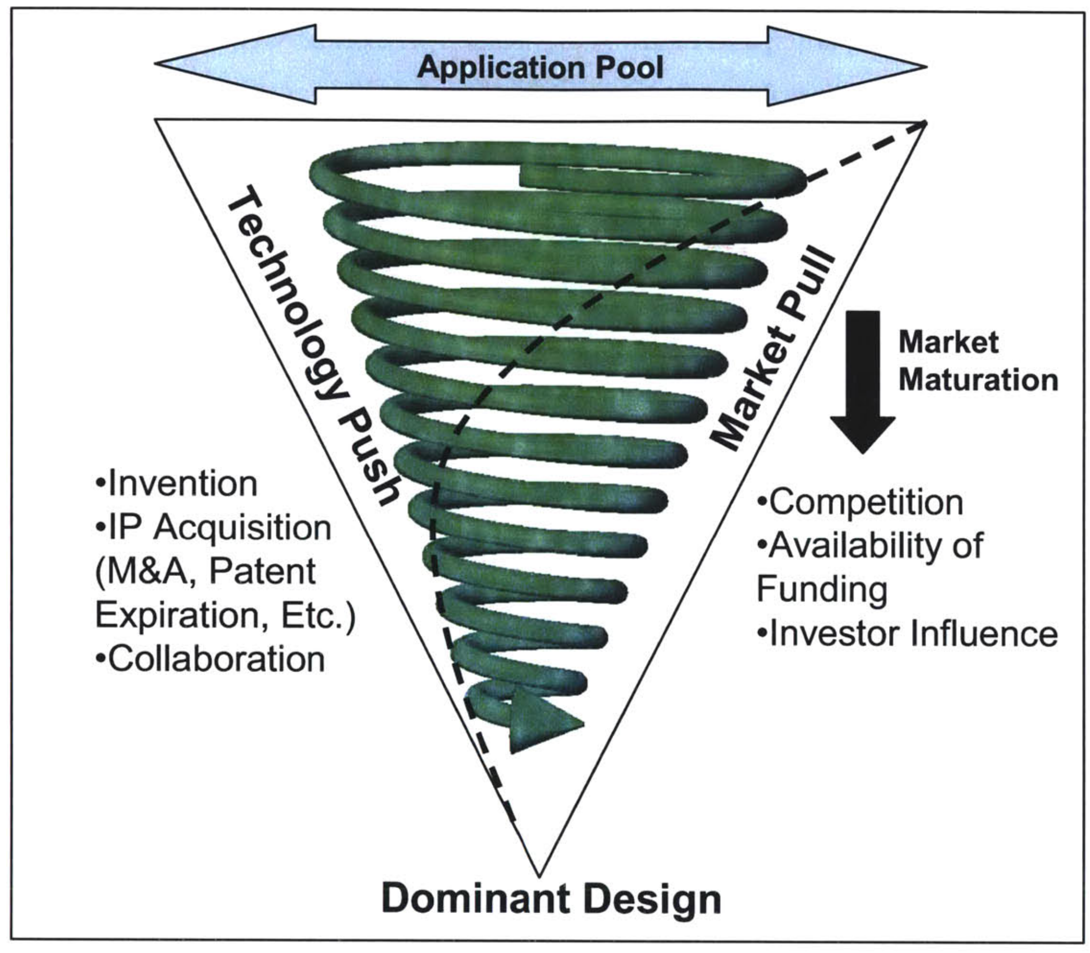


Figure 15 - New Market Evolution

The influence of the 'Strength of IP' dimension can be seen in Figure 15 through the diminishing influence of Technology Push. This represents the fact that as the market moves toward the dominant design, there is much less room to innovate due to the impediment of competitive IP. Finally the 'Degree of Differentiation' dimension is represented when Figure 15 is interpreted not as representative of a particular effort, but as an overview of the entire market. As the market as a whole migrates towards the Dominant Design, the Degree of Differentiation for all of the competitors is reduced (implied here is the fact that, for each application, a different dominant design may surface).

With this view of the market in mind, a survey was constructed in order to gather information from the participants in the NGST market. The primary goal of the survey is to test whether or not these dimensions can be used to provide an accurate gage of the markets relative maturity and future prospects. If this is the case, these dimensions can then be used in order to track the progress of competitors in this market and as a template for the analysis of future new biotechnology markets.

Chapter 3: Survey

3.1 Survey Methodology

The purpose of this survey is to gather information from a range of participants in the NGST market in order to gain a deeper understanding of their motivation and what they view as their primary challenges. The survey was distributed to individuals associated with both private and academic technologies, as well as to institutional financiers. By surveying participants with a range of backgrounds, the intent is to shed light on the technology drivers at different points in the evolution of the efforts in question. A list of survey candidates was generated from a combination of a search of leading companies in the commercial sequencing market and the awardees list of the NIH 'Next Generation Sequencing Technology' grants. The mechanism for identifying the leading companies was a search of press releases and recent conference participation. The major criteria necessary for a firm or lab to be included was that the investigator or company in question had to claim a goal of sequencing all or part of a human genome at a significantly reduced cost from today's benchmark of ~$10M.

3.2 Survey Design

The survey is included in Appendix D. It is divided in four sections: Background; Market Overview; Technology; Intellectual Property; and Funding. The 'Background' section serves to identify the responder's role, affiliation, and historical involvement. This was important for grouping responses in order to determine how the different participant categories view this market. The 'Market Overview' section explores what the respondents consider to be the primary market opportunities for their technology and what effect their technology will have on the established market. It is comprised of five questions including the following:

- What effect did the following factors have on the decision to develop your technology?
    - Faith in the superiority of the technology
    - Market opportunity
    - Level of competition

- Availability of Funding
- IP Availability

- How do you rank the following potential markets for your technology?
    - De Novo sequencing
    - Resequencing
    - Mutation detection
    - Copy number identification
    - Genotyping
    - Expression analysis
- How do you rank the following potential customers?
    - Pharma (Biopharma) companies
    - Diagnostic companies
    - Academic labs
    - Government labs
- How would you describe your relationship to these existing technologies?
    - ABI 3730
    - Affymetrix Genechip
    - Illumina Bead Array
    - Luminex xMap

The 'Technology' section is meant to gather information on what the novel aspects of the respondents' technologies are and what their targeted performance specifications look like. This section has four questions total with the most critical being:

- To the best of your knowledge, please rank the amount of effort devoted to each of the listed components in order to develop your technology.
    - Core biology (custom enzymes, dyes, etc.)
    - Assay design (component scaling, cycling optimization, etc.)
    - Mechanical integration (fluidics, optics, etc.)
    - Image analysis software
    - Read quality analysis software

The questions under 'Intellectual Property' relate to how the respondents value their intellectual property and how IP in general has dictated their development effort. These issues are examined with six questions including:

- What do you consider the primary benefits of your IP?
  - Freedom to operate (no blocking IP)
  - Protection from competition/imitation
  - Leverage for bargaining with potential partners/buyers
  - Cross licensing opportunities
  - Credibility for funding
- Have you had to alter your design at all in order to avoid potential IP conflicts?
- If you had access to all of the IP available in this market (including that of your competitors) do you believe you could use this to improve your product?
  - Yes, absolutely
  - Maybe, I am not familiar enough with the available technologies
  - Not likely

Finally the 'Funding' section serves to gather information on the degree of public funding each respondent is receiving.

The survey was sent to 45 candidates comprised of both companies and researchers. The initial response of 4 was increased to 7 with a follow-up round.

3.3 Data Analysis

While the response to the survey was not overwhelming the data gathered did provide a interesting insight regarding the Market. With a sampling of four company and three academic responses the data does include responses from both of the primary participant groups in this market. Due to the sensitive nature of some of the questions on the survey, not all respondents answered the survey in its entirety. As a result, the following analysis

will concentrate only those questions which were answered by a majority. The response data, in its entirety appears in Appendix E.

3.3.1 Market Overview

The answers for question 8 in the survey 'What effect did the following factors have on the decision to develop your technology?', are represented in Figure16 weighted by the relative importance assigned by each of the respondents. The largest discrepancies between the two viewpoints can be seen in 'Market Opportunity', 'Availability of Funding', and 'IP Availability'.

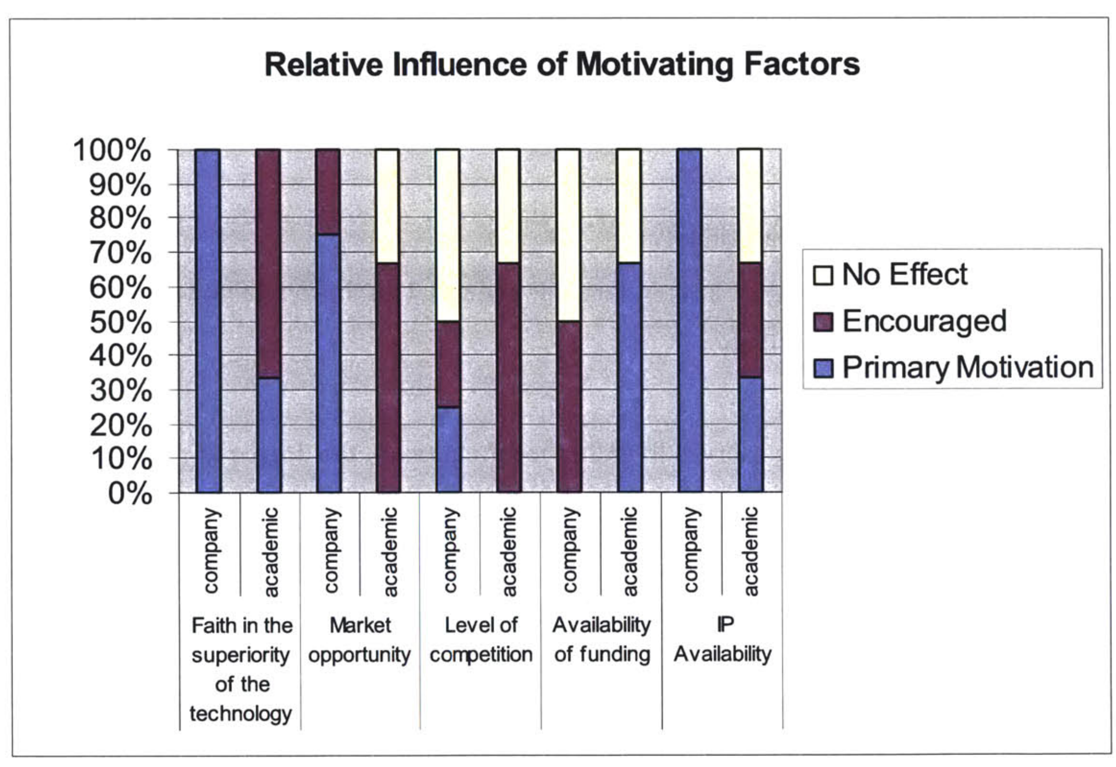


Figure 16 - Ranking of the Factors Motivating Entry into the Sequencing Technology Market

These differences should be expected since Academic endeavors are theoretically less focused on the market implications of their technology and are more reliant on external funding to feed their efforts. Private companies, on the other hand, are very focused on the market and are reluctant to take in to much external funding for risk of diluting their stake in the company. The difference of opinion over the importance over IP is most

likely due to the fact that academic efforts may consider themselves somewhat sheltered from infringement lawsuits due to the fact that they are not actively pursuing a market and pose no competitive threat. This environment is likely to change, however, and academic endeavors, particularly in the life sciences should be aware of the IP landscape.

Question 9, 'How do you rank the following potential markets for your technology?', provides a snapshot of what the 7 respondents consider the most promising markets for their technologies. The distribution for the (shown in Figure 17) is highly predictable with the applications having the highest current demand ranking as the most attractive. It is also interesting to note that the applications that already have alternative technologies available such as genotyping and copy number identification, are viewed as less attractive. Whether this is due to the competitive threat from the established technologies or the difficulty in making headway into an established market is not obvious.

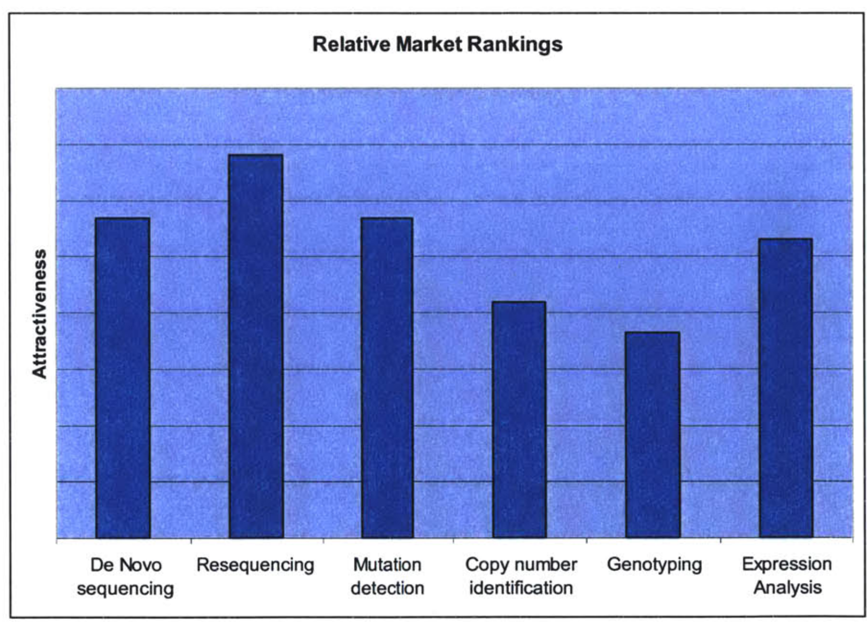


Figure 17 - Ranking of the Attractiveness of Potential Markets for DNA Sequencing Technology

Question 11 asks 'How do you rank the following potential customers?'. As Figure 18 shows, Pharmaceutical and Biopharmaceutical companies are viewed as the greatest potential customers. To date these drug developing companies have utilized DNA

sequencing technology with the hope it could play a major role in identifying potential drug targets and speeding the development of new drugs by providing a rich set of diagnostic data reflecting patient response. Unfortunately, there are very few cases of successful utilization of the existing technology. The NGSTs could represent a fundamental shift in the way drug development is accomplished if they are able to accomplish these goals more effectively.

Government and academic labs remain targeted customers for this technology, which makes sense since it has been research at these institutions that has not only provided momentum to genomic science but has also resulted in a plethora of new applications. Finally, the fact that there is a spread of responses regarding the potential of these technologies to serve as diagnostic tools, shows that there is a concentration by these efforts to delivering an exportable technology to end users rather than developing specific applications for the technology that could be sold as a service.

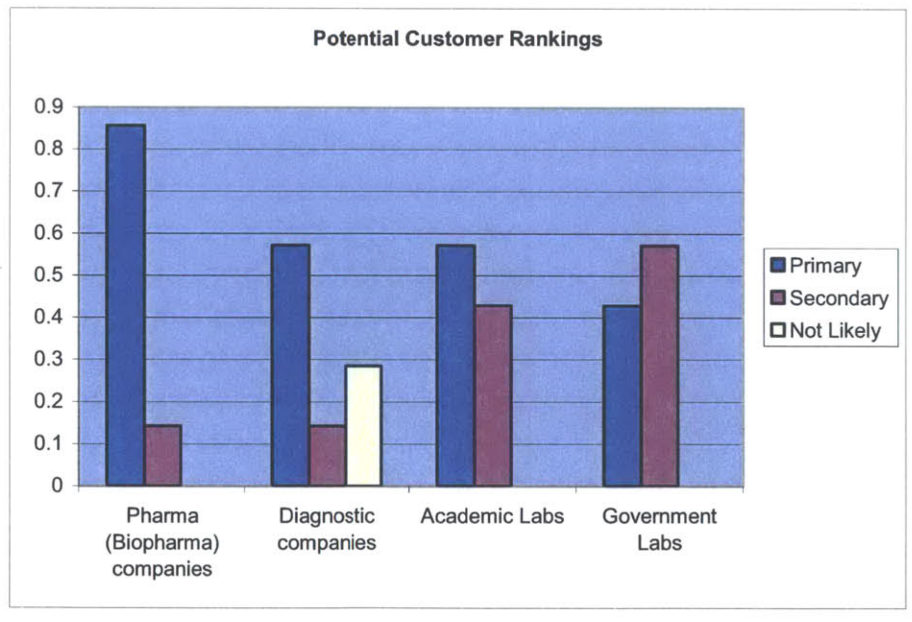


Figure 18 - Expected Customer Base for New Sequencing Technologies

While question 11 may have shown a relative consensus of how the respondents viewed potential customers, question 12 ‘How would you describe your relationship to these existing technologies?’ reflects a much more varied view of the competitive landscape

concerning existing technologies. The companies listed in Figure 19 each have products in the potential markets reflected in question 9, but are each viewed very differently by the respondents. The take away message from this question is that no one is quite sure what these relationships will ultimately look like. Up until the publication of 'A Sanger/Pyrosequencing Hybrid Approach for the Generation of High-Quality Draft Assemblies of Marine Microbial Genomes'[28] in 2006 by Venter et al., for instance, 454 was actively marketing itself as a replacement for traditional Sanger sequencing.

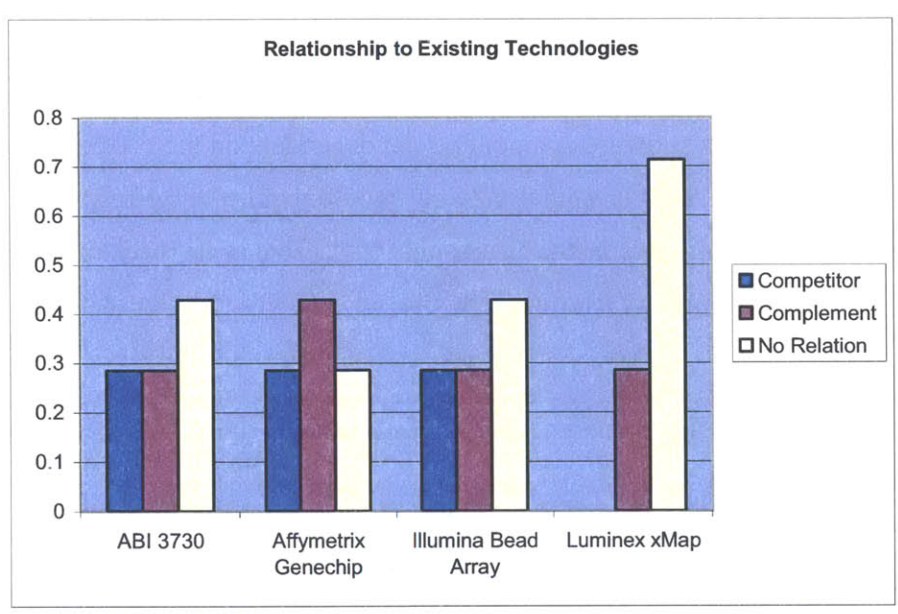


Figure 19 - Expected Relationship to Existing Technologies

3.3.2 Technology

Details regarding technological specifications are well guarded before IP has been secured and an instrument has been released. Consequently it is difficult not only to gain insight into the underlying science behind these technologies but also to extract meaningful specifications that can be used to compare the different efforts.

Question 14 attempts to shed some light on how the technologies vary by asking the respondents 'To the best of your knowledge, please rank the amount of effort devoted to

[28] Venter, JC, et al. A Sanger/Pyrosequencing Hybrid Approach for the Generation of High-Quality Draft Assemblies of Marine Microbial Genomes. PNAS, 103(30), p.4

each of the listed components in order to develop your technology'. Based on what is known about these technologies, the categories that were rated were: Core Biology; Assay Design; Mechanical Integration; Image Analysis Software; and Read Quality Analysis.

Figure 20 shows a fairly uniform distribution for the 7 respondents which makes sense when put in context. While the problems associated with the core biology and assay design are typically solved as a precursor to pursuing the development of a technology, the mechanical integration, which is shown in the graph to demand the most effort, is what is needed to scale the science and bring a product to market. The image analysis and read quality analysis software components are dependent on the effectiveness of the mechanical components, so it is natural that they would require less effort. This result highlights the fact that ultimate instrument is only as useable as its mechanics allow it to be, making this a determining factor in predicting any NGSTs ultimate success.

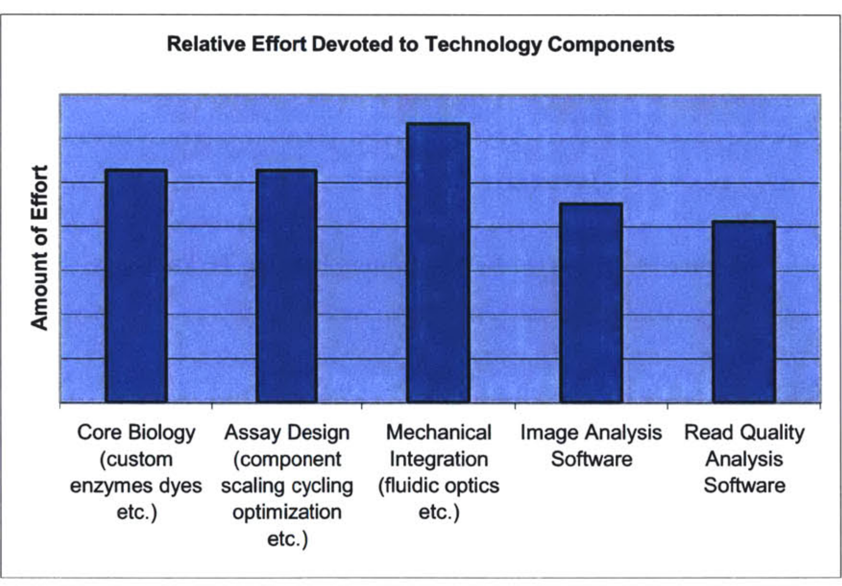


Figure 20 - Ranking of Effort Allocation to Develop the Different Technology Components

3.3.3 Intellectual Property

Like the underlying technology, intellectual property details are well guarded by new technology developers. The reasoning for this is to protect any unpatented secrets from

getting into the public realm and also to prevent competitors from attempting to gain blocking IP.

When asked 'What do you consider the primary benefits of your IP?' in question 18, the responses shown in Figure 21 reflect that the NGST companies (the majority of the academic efforts failed to answer this question and are excluded from the results) were most concerned with their ability to develop their technology without having to worry about competitors. Each of the benefits were ranked on a five point scale (from 'Very Important' to 'Not Important') by the four respondents and these rankings were then averaged for the group to get the relative importance of each to the rest of the group. While adding to the companies overall value was also considered a primary benefit of IP, there was much less interest in cross licensing opportunities.

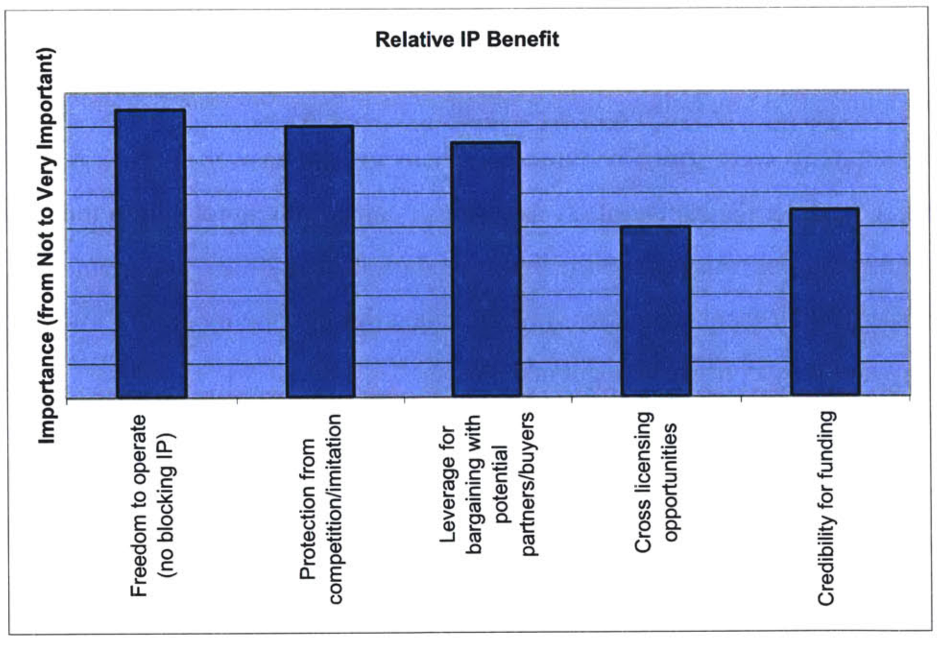


Figure 21 - Perceived IP Benefit

The results for Question 20 'Have you had to alter your design at all in order to avoid IP conflicts?' are shown in Figure 22. The fact that so many respondents have had to alter their technology shows how little differentiation there is among the many NGSTs. This

also supports the earlier observation in section 2.4 that there is bound to be litigation over IP infringement at some point down the road in this technology space.

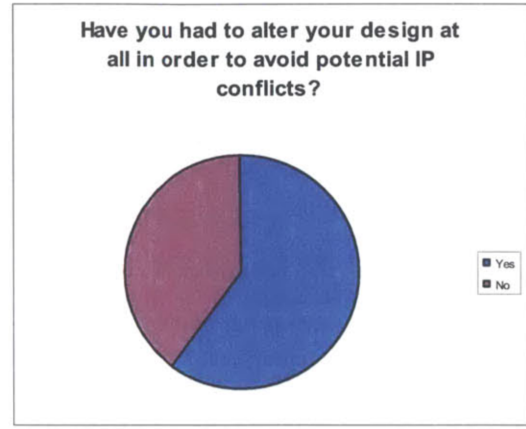


Figure 22 - Influence of IP Constraints on Technology Design

When asked 'If you had access to all of the IP available in this market (including that of your competitors) do you believe you could use this to improve you product?' in question 21, the respondents were split. The inference here with an answer of 'No', is that the respondents consider their technology superior or completely unrelated to the other technologies. A 'Yes' answer, on the other hand implies that the respondents feel that the performance of their technology could be extended through the incorporation of existing technology that they don't have legal access to.

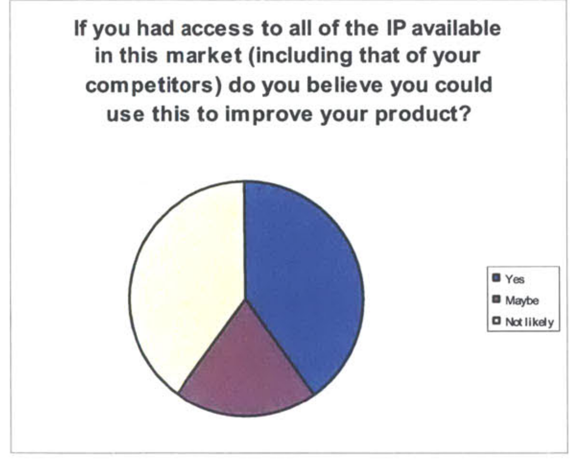


Figure 23 - Potential IP Assimilation

The most interesting conclusion that can be drawn from the answers to the intellectual property questions is the fact that IP appears to have an inhibiting influence in this market. Based on the respondents input, it would seem that a superior technological platform would be attainable if the IP in this space was universally available. This does not bode well for participants, since, if a superior solution is available, it is only a matter of time before someone determines how to exploit it.

Chapter 4: Conclusions

There development of Next Generation Sequencing Technology has many characteristics that make it an interesting study in how new markets evolve. With no revenue and a customer base that has yet to be defined, the companies involved in this market defy analysis by straightforward quantitative methods. This work has sought to enable such analysis through an approach consisting of identifying the meaningful performance specifications, potential applications, and relevant theoretical market trajectories on which it might be mapped. The most interesting conclusions resulting from this concern the role of intellectual property and the fact that the market appears to be a classic example of the introduction of a disruptive technology.

The definition of this market as disruptive is reflected in the fact that the defined performance parameters for the contending technologies do not surpass existing technology in all facets. Also supporting the disruptive definition is that the developers of this technology seek to extend the market with new applications as opposed to merely hijacking the customer base for established applications. The fact that so many efforts are being launched with only a vague understanding of the true potential revenue, while not unheard of, is another part of what makes this market so interesting.

	Read Length	Quality	Cost/Base	Bases/Hour	Sample Prep Complexity	Data Post Processing
454	-10X	-5%	--	100X	↓	↑
Solexa (Illumina)	-20X	?	-100X	150X	↓	↑↑
Helicos	-20X	?	-100X	?	↓↓	↑↑
APG (ABI)	-20X	?	-100X	?	↓	↑↑

Figure 24 - Comparison of New Technologies to 3730 Performance
The green arrows represent a reduction in the complexity of sample prep due to the elimination of amplification steps. The red arrows represent an increase in the amount of data processing.

As discussed earlier, the primary measurable metrics for DNA sequencing technologies are quality, read length, cost per base, and bases per hour. The more qualitative measurements of this technology include the complexity of the sample prep and the amount of post processing necessary for the data generated. Figure 24 provides a breakdown of how some of the leading new technologies compare to the performance of Sanger sequencing on the ABI 3730. As the chart reflects, there are trade-offs with these new technologies which are only acceptable if the market continues to evolve towards the predicted applications and customers are willing to adapt to the new demands of these technologies, such as infrastructure to support the large data sets and intensive post processing.

Provided that the customers can be won over, the factors that will determine their selection were identified in Chapter 2 to be: Degree of Market Focus; Phase of Development; Degree of Differentiation from Competition; and Strength of IP. Customers are seeking a well developed technology that has been proven to address their chosen application better than the competition. While the strength of IP may not play an active role in customer decisions from the onset, this will become more important as the market matures.

The influence of intellectual property and its restrictions on the development of the NGSTs will most likely be the primary driver of this market. The survey revealed that many of the participants in this market believe they could develop a better product if there was unfettered access to all of the IP available in this space. This will be an important fact as this market matures and companies begin to address any IP infringement issues. This is already reflected in the pending litigation against Solexa by ABI over sequencing by ligation IP. While the SBL technology is not utilized by Solexa's platform, if they do have potential blocking IP to ABI's recently acquired technology, this could be invaluable.

As this market continues to unfold, it will be worth following the development of not only the companies but the applications and customer base as well. This will provide an excellent indication of whether this technology is leading down the road to personalized medicine, or just another placeholder technology that will enable researchers to identify the data that is truly meaningful leading to a new round of innovation.

It will also be interesting to monitor the development of the 2nd generation technologies all ready under development. Will these companies follow the same trajectory as this batch or will they be driven by entirely different applications and customers? Will they go public or become acquisitions like APG and Solexa? Will there be the same low level of differentiation (possible with the high prevalence of nanopore technology already evident) in the next batch? Will funding opportunities like those from the NIH and the xPrize continue to influence the development of this technology or merely serve as distractions? The answers for all of these questions and, more importantly, the reasoning behind them would be excellent candidates for future study

References

[1] Ahmadian, A., Ehn, M., Hober, S. Pyrosequencing: History, Biochemistry, and Future. *Clinica Chimica Acta*, 363: 83-94, 2006.

[2] Almus, M., Nerlinger, E. A. Growth of New Technology-Based Firms: Which Factors Matter?. *Small Business Economics*, pp. 141-154, September 1999

[3] Astebro, T. B., Dahlin, K. B. Opportunity Knocks. *Social Science Research Network*, February 2005.

[4] Balzej, R., Kumaresan, P., Mathies, R. Microfabricated Bioprocessors for Integrated Nanoliter-Scale Sanger DNA Sequencing. *PNAS*, 103(19):7240-7245, May 2006.

[5] Becker, M. C., Lillemark, M. Marketing/R&D Integration in the Pharmaceutical Industry. *Research Policy*, 35: 105-120, 2006.

[6] Bonetta, L. Genome Sequencing in the Fast Lane. *Nature Methods*, 3(2): 141-147.

[7] Brooks, H. The Relationship between Science and Technology. *Reseach Policy*, 23:477-486, 1994.

[8] Chakravorti, B. The New Rules for Brining Innovations to Market. *Harvard Business Review*, pp. 58-67, March 2004.

[9] Chan, E. Advances in Sequencing Technology. *Mutation Research*, 537:13-40, 2005.

[10] Chidamber, S. R., Kon, H. B. A Research Retrospective of Innovation Inception and Success: The Technology-Push Demand-Pull Question. *International Journal of Technology Management,* 9(1), 1994.

[11] Christensen, C. *The Innovator's Dilemma.* HarperCollins, New York, NY 2003.

[12] Christodouliou, N., Amundson, S., Dooley, K. Push and Pull: Sequential, Parallel, or Integrated. Arizona State University

[13] Church, G. M., et al. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, Vol 309:1728-1732, September 2005.

[14] Cline, J., et al. PCR Fidelity of Pfu DNA Polymerase and other Thermostable DNA Polymerases. *Nucleic Acids Research*, Vol. 24(18): 3546-3551, 1996.

[15] Collins, F., et al. A New Five-Year Plan for the U.S. Human Genome Project. *Science*, Vol. 262: 43-46, October 1993.

[16] Collins, F., et al. The Human Genome Project: Lessons from Large-Scale Biology. *Science*, 300: 286-290, April 2003.

[17] Collins, F., et al. New Goals for the U.S. Human Genome Project: 1998-2003. *Science*, Vol. 282: 682-689, October 1998.

[18] Collins, F., et al. The Human Genome Project: Lessons from Large-Scale Biology. *Science*, 300: 286-290, April 2003.

[19] Davies, K. Bio-IT Briefing on: the $1000 Genome. www.bio-itworld.com

[20] Fan, J., Chee, M., Gunderson, K. Highly Parallel Genomic Assays. *Nature Reviews Genetics*, 7: 632-644, August 2006.

[21] Feuk, L., Carson, A., Scherer, S. Structural Variation in the Human Genome. *Nature Reviews Genetics*, 7: 85-97, Feb. 2006.

[22] Fologea, D., et al. Detecting Single Stranded DNA with a Solid State Nanopore. *Nano Letters*, 5(10):1905-1909, 2005.

[23] Garnsey, E. Heffeman, P., Hugo, O., Stam, E. New Firm Growth: Exploring Processes and Paths. January 1004.
[24] Gibbs, R. Deeper into the Genome. *Nature*, 437(27):1233-1234, October 2005.
[25] Gutmann, G. Hybrid Electric Vehicles and Electrochemical Storage Systems – a Technology Push-Pull Couple. *Journal of Power Sources*, 84: 275-279, 1999.
[26] Howells, J. Rethinking the Market-Technology Relationship for Innovation. *Research Policy*, 25:1209-1219, 1997.
[27] International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. *Nature*, Vol. 409: 860-921, 2001.
[28] Jaffe, Adam B. Demand and Supply Influences in R&D Intensity and Productivity Growth. *The Review of Economics and Statistics*, 70(3):431-437, August 1988.
[29] Kamien, M. Schwartz, N. Market Structure and Innovation. *Journal of Economic Literature,* 13(1):1-37, 1975.
[30] Kleinknecht, A., Verspagen, B. Demand and Innovation: Schmookler Re-Examined. *Research Policy*, 19:387-394, 1990.
[31] Kling, J. Ultrafast DNA Sequencing. *Nature Biotechnology*, 21(12): 1425-1427
[32] Kossoff, G. Three-Dimensional Ultrasound – Technology Push or Market Pull? *Ultrasound Obstet. Gynecol.,* 5:217-218, 1995.
[33] Kostoff, R., Schaller, R. Science and Technology Roadmaps. *IEEE Transactions on Engineering Management,* 48(2):132-143.
[34] Kuo, W. et al A Sequence-Oriented Comparison of Gene Expression Measurements Across Different Hybridization-Based Technologies. *Nature Biotechnology*, 24(7):832-840, July 2006.
[35] Leydesdorff, L. The Triple Helix: an Evolutionary Model of Innovations. *Research Policy*, 29:243-255, 2000.
[36] Maine, E., Garnsey, D. Challenges Facing New Firms Commercialising Nanomaterials. Commercialisation of Micro and Nano Systems Conferences.
[37] Margulies, M., et al. Genome Sequencing in Microfabricated High-Density Picolitre Reactors. *Nature*, Vol.437:376-380, September 2005.
[38] Marziali, A., Akeson, M. New DNA Sequencing Methods. *Annual Review of Biomedical Engineering*, 3:195-223, 2001.
[39] McGuire, A., Gibbs, R. No Longer De-Identified. *Science*, 312:370-371, April 2006.
[40] McMillan, G. S., Narin, F., Deeds, D. L. An Analysis of the Critical Role of Public Science in Innovation: the Case of Biotechnology. *Reseach Policy*, 29:1-8, 2000.
[41] Menanteau, P., Lefebvre, H. Competing Technologies and the Diffusion of Innovations: the Emergence of Energy-Efficient Lamps in the Residential Sector. *Research Policy*, 29:375-389, 2000.
[42] Metzker, M. Emerging Technologies in DNA Sequencing. *Genome Research*, 15:1767-1776, 2005.
[43] Meyer, M. Does Science Push Technology? Patents Citing Scientific Literature. *Research Policy*, 29:409-434, 2000.
[44] Mowery, D., Roseberg, N. The influence of market demand upon innovation: a critical review of some Empirical Studies. *Research Policy*, 8:102-153, 1979.

[45] Murmann, J. P., Frenken, K. Toward a Systematic Framework for Research on Dominant Designs, Technological Innovations, and Industrial Change. *Research Policy*, 35:925-952, 2006.

[46] Nelson, R. R. On the Uneven Evolution of Human Know-How. *Research Policy*, 32:909-922, 2003.

[47] PhorTech International, DNA Sequencing.

[48] Redon, R., et al Global Variation in Copy Number in the Human Genome. *Nature*, 444:444-454, Nov. 2006.

[49] Rickne, A., Utterback, J. M. Critical Events in Early Phases of Firm Development.

[50] Roberts, E. B. *Biomedical Innovation*. The MIT Press, 1981.

[51] Robertson, J. The $1000 Genome: Ethical and Legal Issues in Whole Genome Sequencing of Individuals. *The American Journal of Bioethics*, 3(3), 2003

[52] Rogers, Y., Venter, C. Massively Parallel Sequencing. *Nature*, 437:326-327, September 2005.

[53] Rosenberg, N. Science, Invention and Economic Growth. *Economic Journal*, 84:90-108, March 1974.

[54] Rosenberg, N., Nelson, R. R. American Universities and Technical Advance in Industry. *Research Policy*, 23:323-348, 1994.

[55] Sanger, F., Nicklen, S., Coulson, A. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Science*,74(12):5463-5467, December 1977.

[56] Schoonmaker, M., Bernhardt, B., Holtzman, N. Factors Influencing Health Insurers' Decisions to Cover New Genetic Technologies. *International Journal of Technology Assessment in Healthcare*, 16(1):178-189, 2000.

[57] Shendure, J., Mitra, R., Varma, C., Church, G. Advanced Sequencing Technologies: Methods and Goals. *Nature Reviews Genetics*, 5:335-344, May 2004

[58] Smailus, D., Marziali, A., Dextras, P., Marra, M., Holt, R. Simple, Robust Methods for High-Throughput Nanoliter-Scale DNA Sequencing. *Genome Research*, pp. 1447-1450, September 2005.

[59] Souder, W. Improving Productivity through Technology Push. *Research Technology Management*, pp. 19-24, March 1989.

[60] Spann, M. S., Adams, M., Souder, W. E. Measures of Technology Transfer Effectiveness: Key Dimensions and Differences in Their Use by Sponsors, Developers, and Adopters. *IEEE Transactions on Engineering Management*, 42(1):19-29, February 1995.

[61] Teece, D. J. Competition, Cooperation, and Innovation: Organizational Arrangements for Regimes of Rapid Technological Progress. *Journal of Economic Behavior and Organization*, 18(1):1-25, June 1992.

[62] The International HapMap Consortium. A Haplotype map of the Human Genome. *Nature*, 237(27):1299-1320, October 2005.

[63] Tushman, M. L., Anderson, P. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Administrative Science Quaterly*, Vol. 35, 1990.

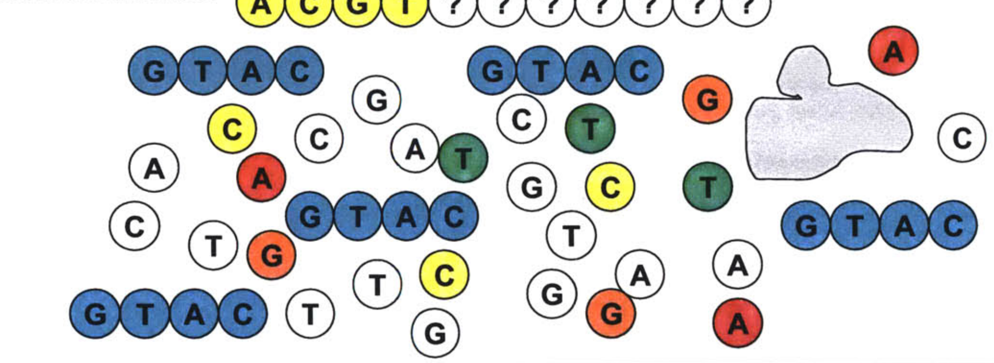
[64] U.S. Department of Health and Human Services and Department of Energy, Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years, April 1990. (http://www.genome.gov/10001477)

[65] Utterback, J. M., Afuah, A. N. The Dynamic "Diamond": A Technological Innovation Perspective. Sloan Working Paper #3872, 35004
[66] Utterback, J. M., Afuah, A. N. Dynamic Competitive Strategies: A Technological Evolution Perspective. Sloan Working Paper #3871, 34790
[67] Venter, JC, et al. A Sanger/Pyrosequencing Hybrid Approach for the Generation of High-Quality Draft Assemblies of Marine Microbial Genomes. PNAS, 103(30):11240-11245. July 2006.
[68] Walsh, S. T., Linton, J. D. Infrastructure for Emergent Industries Based on Discontinuous Innovations. *Engineering Management Journal*,12(2):23-31, June 2000.
[69] Walsh, Steven T., Kirchhoff, Bruce A., Newbert, Scott Differentiating Market Strategies for Disruptive Technologies. *IEEE Transactions on Engineering Management*, 49(4):341-351, November 2002.
[70] Watson, J., Crick, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737, 1953.
[71] Whiteford, N., Haslam, N., Weber, G., Prugel-Bennett, A., Essex, J., Roach, P., Bradley, M., Nelyon, C. An Analysis of the Feasibility of Short Read Resequencing. *Nucleic Acides Research*, 33(19), 2005.
[72] Applera 2002 Annual Report
[73] Applera 2003 Annual Report
[74] Applera 2005 Annual Report
[75] Applera 2006 Q1 Earnings Report
[76] Applera 2006 Q2 Earnings Report
[77] Applera 2006 Q3 Earnings Report
[78] Luminex 2005 Annual Report

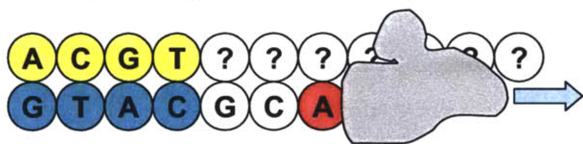
Appendix A: Four Color Sanger Sequencing

Step 1: An unknown strand of DNA is bound to a known strand (yellow)

Step 2: The newly formed DNA is added to a mixture containing short 'Primer' DNA strands (blue) which are complementary to its known portion, DNA Polymerase (grey) which synthesizes complementary DNA strands from existing ones and free bases which may or may not have an attached fluorescent label

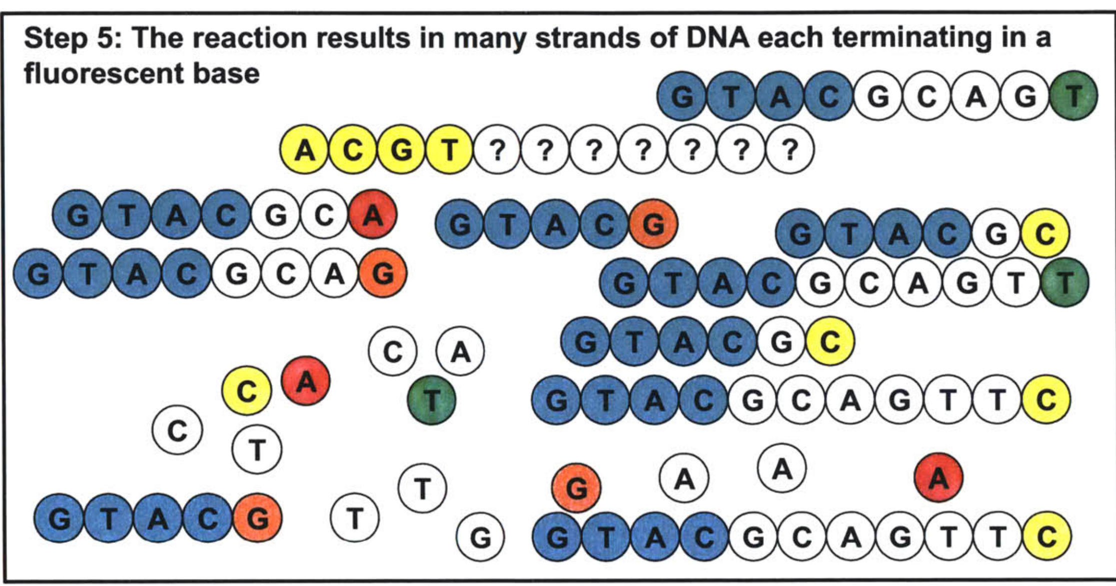


Step 3: The Primer is annealed and the DNA polymerase begins synthesizing a new DNA strand incorporating the free bases until a labeled base is added

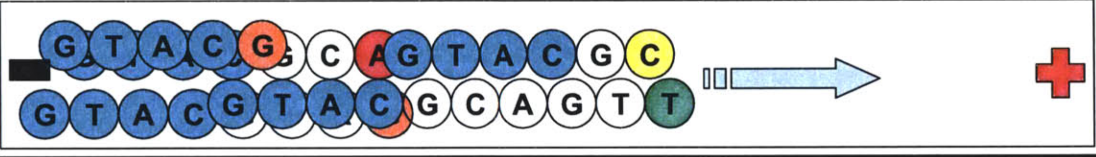


Step 4: The new strand is stripped away and the process repeats many times

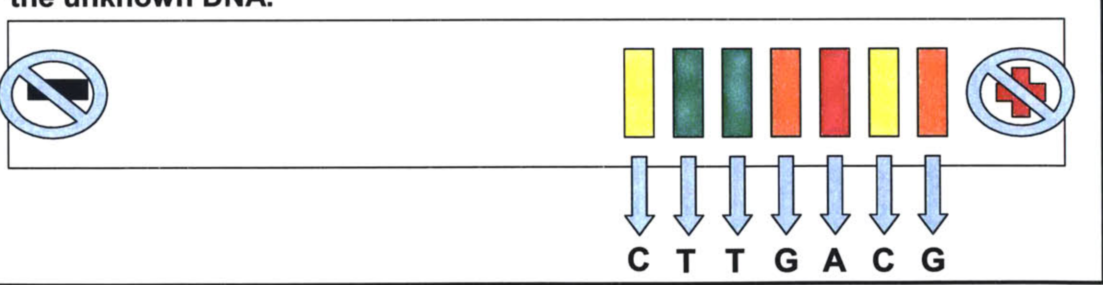
Step 5: The reaction results in many strands of DNA each terminating in a fluorescent base



Step 6: The product is loaded onto a sieving gel and an electric field is applied causing the negatively charged DNA to migrate



Step 7: While the DNA strands of different length would normally migrate at the same rate due to their identical mass to charge ratio, the gel creates a physical barrier in which the shorter strands move faster resulting in a spatial separation of the strands when the electric field is turned off. The fluorescent labels can then be excited and read to determine the sequence of the unknown DNA.



Appendix B: NIH Next Generation Sequencing Grant Awardees

Overview

Round	Category	Awardee	Title	PI	$M	Term
1	$100K	Microchip Biotechnologies Inc.	Microbead Integrated DNA Sequencer (MINDS) System	Stevan Jovanovich	$6.10	3 yr.
1	$100K	Agencourt Biosciences	Bead-based Polony Sequencing	Gina Costa	$5.40	3 yr.
1	$100K	454 Life Sciences	Massively Parallel High Throughput Low Cost Sequencing	Kenton Lohman	$2.00	2 yr.
1	$100K	454 Life Sciences	454 Life SciencesMassively Parallel System Sequencing	Marcel Margulies	$5.00	3 yr.
1	$100K	Li-Cor Inc.	Single-Molecule DNA Seq using Charge-Switch dNTPs	John Williams	$2.50	3 yr.
1	$100K	Baylor College of Medicine	Ultrafast Seq by Synthesis (SBS) for large scale Human Reseq	Micheal Metzker	$2.00	3 yr.
1	$100K	Stanford University	High-Throughput Single Molecules Sequencing	Stephen Quake	$1.80	3 yr.
1	$100K	Stanford Genome Tech. Center	Pyrosequencing Array for DNA Sequencing	Mostafa Ronaghi	$1.80	3 yr.
1	$100K	Columbia University	An Integrated System for DNA Seq by Synthesis	Jingyue Ju	$1.80	3 yr.
1	$100K	Arizona State University	Multiplexed Reactive Sequencing of DNA	Peter Williams	$1.70	3 yr.
1	$100K	University of Florida	Polymerase for Sequencing by Synthesis	Steven Benner	$0.80	3 yr.
1	$100K	Rowland Institute at Harvard	Ultrafast Nanopore Readout Platform for Designed DNA's	Amit Meller	$0.60	2 yr.
1	$1K	University of North Carolina	Nanotechnology for the Structual Interogation of DNA	J. Micheal Ramsey	$2.00	2 yr.
1	$1K	Oak Ridge National Laboratory	Comp. Research & Dev. For Rapid Sequencing Nanotechnology	James Lee	$0.70	3 yr.
1	$1K	Oak Ridge National Laboratory	Exp. Research & Dev. For Rapid Sequencing Nanotechnology	James Lee	$0.75	3 yr.
1	$1K	University of Maine Orono	High-speed Nanopore Gene Sequencing	Scott Collins	$0.85	2 yr.
1	$1K	University of Florida	DNA Sequencing Using Nanopores	Steven Benner	$0.80	3 yr.
1	$1K	University of British Columbia	Nanopores for Trans-membrane Bio-molecule Detection	Andre Marziali	$0.65	3 yr.
1	$1K	Arizona State University	Molecular Reading Head for Single Molecule DNA Sequencing	Stuart Lindsay	$0.55	3 yr.
1	$1K	Stanford University	Single Molecule Nucleic Acid Detection with Nanopipettes	Ronald Davis	$0.45	2 yr.
2	$100K	Agencourt Biosciences	Bead-based Polony Sequencing	Gina Costa	$1.20	2 yr.
2	$100K	State University of New York	Ultra-high throughput DNA Seq Sys-Multi Cap and nl Reaction vol	Vera Gorfinkel	$1.50	2 yr.
2	$100K	Network Biosystems	$100,000 Genome Using Integ. Microfluidic Capillary Electroph.	Greg Kellog	$4.50	3 yr.
2	$1K	Duke University	Droplet-based Digital Microfluidic Genome Sequencing	Richard Fair	$0.51	2 yr.
2	$1K	The Scripps Research Institute	Single-Molecule DNA Sequencing w/ Engineered Nanopores	M. Reza Ghadiri	$4.20	5 yr.
2	$1K	Harvard University	Electronic Sequencing in Nanopores	Jene Golovchenko	$5.20	3 yr.
2	$1K	Visigen Biotechnologies	Real-time DNA Sequencing	Susan Hardin	$4.20	3 yr.
2	$1K	University of California SD	Massively Parallel Cloning and Sequencing of DNA	Xiaohua Huang	$0.75	3 yr.
2	$1K	Columbia University	Modulating Nucleotide Size in DNA for Detection by Nanopore	Jingyue Ju	$0.97	3 yr.
2	$1K	New York University	Haplotype Sequencing Via Single Molecule Hybridization	Bhubaneswar Mishra	$0.59	2 yr.
2	$1K	University of Illinois U-C	Sequencing a DNA Molecule using a Synthetic Nanopore	Gregory Kimp	$2.10	3 yr.
2	$1K	Pacific Biosciences(Nanofluidics)	Real-time Multiplex Single-Molecule DNA Sequencing	Stephen Turner	$6.60	3 yr.
3	$100K	Baylor College of Medicine	Ultrafast Seq by Synthesis (SBS) for large scale Human Reseq	Micheal Metzker	$0.50	1 yr.
3	$100K	Intelligent Bio-Systems	High-Throughput DNA Sequencing by Synthesis Platform	Steven Gordon	$0.43	1 yr.
3	$1K	General Electric Global Research	Closed Complex Single Molecule Sequencing	John Nelson	$0.90	2 yr.
3	$1K	University of North Carolina	Nanoscale Fluidic Techs for Rapidly Seq. Single DNA Molecules	J. Micheal Ramsey	$3.80	4 yr.
3	$1K	University of California SD	Genome Seq. by Ligation Using Nano-Arrays of Single DNA Mol.	Xiaohua Huang	$0.28	1 yr.
3	$1K	Boston University	High-Thr DNA Seq. Using Design Polymers & Nanopore Arrays	Amit Meller	$2.20	3 yr.
3	$1K	Helicos Biosciences	High Accuracy Single Molecule DNA Sequencing by Synthesis	Timothy Harris	$2.00	3 yr.
3	$1K	Lehigh University	Force Spectroscopy Platform for label free Genome Sequencing	Dmitri Vezenov	$0.91	3 yr.
3	$1K	Arizona State University	Fabrication of Univ. DNA Nanoarrays for SBH	Peiming Zhang	$0.90	3 yr.
3	$1K	Case Western University	Large-Scale Nanopore Arrays for DNA Sequencing	Carlos Mastrangelo	$0.82	3 yr.
3	$1K	University of Washington	Engineering MspA for Nanopore Sequencing	Jens Gundlach	$0.61	2 yr.

Round 1 – Awarded October 14, 2004

From

NIH News Release '2004 Release: NHGRI Seeks Next Generation of Sequencing Technologies'

"$100,000" Genome Grants

Stevan B. Jovanovich, Ph.D., Microchip Biotechnologies Inc., Fremont, Calif.
$6.1 million (3 years)
"Microbead INtegrated DNA Sequencer (MINDS) System"
Retaining the advantages of current DNA sequencing methods, including well-developed community infrastructure, commercial availability of reagents and existing analysis software, this group will push Sanger-based sequencing toward its performance limit in a completely automated, bench-top system. The heart of the system will be a microchip-based device that can label and process DNA fragments from individual microbeads in low-volume reactions, followed by ultra-fast separation and analysis on microfabricated capillary electrophoresis channels.

Gina L. Costa, Ph.D., Agencourt Bioscience Corp., Beverly, Mass.
$5.4 million (3 years)
"Bead-based Polony Sequencing"
This group will work to further develop polymerase colony (polony) sequencing technologies. This is a highly parallel sequencing approach that involves synthesizing short regions of identical DNA fragments on magnetic beads, packing millions of them into a chamber and then extending each of those molecules while detecting the addition of fluorescently labeled DNA building blocks or nucleotides.

Kenton Lohman, Ph.D., 454 Life Sciences Corp., Branford, Conn.
$2 million (2 years)
"Massively Parallel High Throughput, Low Cost Sequencing"
and
Marcel Margulies, Ph.D., 454 Life Sciences Corp., Branford, Conn.
$5 million (3 years)
"454 Life Sciences Massively Parallel System DNA Sequencing"
Expanding the capabilities of its sequencing-by-synthesis technology, this group will scale up the system to increase throughput, cut costs and provide the power to sequence genomes of organisms for which no framework of genomic data exists. In order to reduce labor and costs, this method emphasizes the miniaturization of each step, from sample preparation to DNA sequencing. The platform enables one person to fragment, amplify, sequence and assemble an entire genome, regardless of the genome's size.

John Williams, Ph.D., LI-COR Inc., Lincoln, Neb.
$2.5 million (3 years)
"Single-Molecule DNA Sequencing Using Charge-Switch dNTPs"
Sequencing single molecules produces challenges in imaging, but reduces other hurdles to achieving long sequence read length. This group is developing technology to detect the release of reaction products when nucleotides are incorporated into single DNA strands.

Michael L. Metzker, Ph.D., Human Genome Sequencing Center, Baylor College of Medicine, Houston
$2 million (3 years)
"Ultrafast SBS (Sequencing by Synthesis) Method for Large-Scale Human Resequencing"
This team will focus on developing novel fluorescent, photolabile nucleotide terminators for sequencing by synthesis, as well as making improvements to enzymes called DNA polymerases that will support their

accurate incorporation into DNA. This is part of the group's plan to eventually build a full-scale sequencing system.

Stephen R. Quake, Ph.D., Stanford University, Palo Alto, Calif.
$1.8 million (3 years)
"High-Throughput, Single-Molecule DNA Sequencing"
This group will try to improve its sequencing-by-synthesis technology in order to achieve longer reads from very large numbers of single DNA molecules. The key to the technology's single molecule sensitivity is the detection of fluorescence resonance energy transfer on a total internal reflection microscope.

Mostafa Ronaghi, Ph.D., Stanford Genome Technology Center, Palo Alto, Calif.
$1.8 million (3 years)
"Pyrosequencing Array for DNA Sequencing"
The principal investigator of this team is an inventor of pyrosequencing, which uses unmodified nucleotides and polymerases to synthesize DNA and a firefly enzyme to generate a chemiluminescent signal. The group of researchers will work on further developing a highly integrated and parallel format with improved equipment for detection of the chemiluminescent signals resulting in a portable and inexpensive device for low-cost genome sequencing.

Jingyue Ju, Ph.D., Columbia University, New York
$1.8 million (3 years)
"An Integrated System for DNA Sequencing by Synthesis"
One focus of this team's research is novel chemistry that allows a fluorescent molecule attached to a nucleotide to be detected and then removed with a flash of light after its addition to a growing DNA molecule. The researchers will also develop a unique way to attach many thousands of DNA molecules site specifically to a surface to produce a high-throughput device for DNA sequencing by synthesis.

Peter Williams, Ph.D., Arizona State University, Tempe
$1.7 million (3 years)
"Multiplexed Reactive Sequencing of DNA"
This group will use commercially available, fluorescein-labeled nucleotides and off-the-shelf detectors in a practical sequencing-by-synthesis system multiplexed to read more than 10,000 sequences simultaneously. The system will be targeted initially at specific genes and subsequently at whole genomes.

Steven A. Benner, Ph.D., University of Florida, Gainesville
$800,000 (3 years)
"Polymerases for Sequencing by Synthesis"
This group's goal is to engineer a DNA polymerase, which is the enzyme used in cells and in laboratory experiments to synthesize new DNA molecules, that will have optimal characteristics for sequencing by synthesis with chemically-altered nucleotides.

Amit Meller, Ph.D., Rowland Institute at Harvard, Harvard University, Cambridge, Mass.
$600,000 (2 years)
"Ultra-fast Nanopore Readout Platform for Designed DNA's"
A nanometer is one-billionth of a meter, much too small to be seen with a conventional lab microscope. Most groups developing nanopores (holes about 2 nm in diameter) as DNA sequence transducers propose to detect an electrical, or ionic, signal from individual DNA molecules. This group will pursue a novel approach in which a nanopore is used to simultaneously detect electrical and fluorescent signals from many nanopores at one time.

"$1,000 Genome" Grants

J. Michael Ramsey, Ph.D., University of North Carolina, Chapel Hill
$2 million (2 years)

Nanotechnology for the Structural Interrogation of DNA
This group will explore combinations of fabrication technologies to build devices for analysis of single DNA molecules, and use various measurement techniques to extract information from those devices. Experiment is woven together with simulation and modeling to understand the basic physics of molecule-device interaction at this size scale and its implications for device design.

James Weifu Lee, Ph.D., Oak Ridge National Laboratory, Oak Ridge, Tenn.
Two grants: $700,000 (3 years); $750,000 (3 years)
"Computational Research & Development for Rapid Sequencing Nanotechnology" "Experimental Research & Development for Rapid Sequencing Nanotechnology"
This group will develop computational modeling to guide the fabrication of a novel nanotechnology sequencing device, as well as design electronic control and detection experiments. Using this information, the group will build a device in which stretched DNA molecules would be made to pass between sharp electrodes spaced 2 to 5 nanometers apart. It will then test the device to see if it is possible to distinguish between the four types of nucleotides based on differences in a phenomenon called electron tunneling.

Scott D. Collins, Ph.D., University of Maine, Orono
$850,000 (2 years)
"High-speed Nanopore Gene Sequencing"
Skilled in silicon fabrication methods, this group will try to fabricate a nanopore with tiny electrodes and built-in circuits that will be used in experiments that attempt to measure differences in the electron tunneling of individual nucleotides in DNA molecules. Such devices could lay the groundwork for high-speed approaches to sequencing single DNA molecules.

Steven A. Benner, Ph.D., University of Florida, Gainesville
$800,000 (3 years)
"DNA Sequencing Using Nanopores"
This group will produce conical nanopores in a synthetic membrane, coat the pores with gold, modify the pores to control DNA transport and then introduce chemically modified DNA. The goal will be to detect different signals from each of the four types of nucleotides as DNA passes through the pores.

Andre Marziali, Ph.D., University of British Columbia, Vancouver
$650,000 (3 years)
"Nanopores for Trans-Membrane Bio-Molecule Detection"
This group will contribute to understanding how single biological molecules interact with pores inserted into membranes, and how useful information can be derived from those interactions. This study attempts to extend the use of nanopore sensors into living cells.

Stuart Lindsay, Ph.D., Arizona State University, Tempe, Ariz.
$550,000 (3 years)
"Molecular Reading Head for Single-Molecule DNA Sequencing"
Building on the concept of threading DNA through a molecular pore, this group is developing a system in which a chemical ring, acting as a reading head, is used to measure differences in friction as DNA passes through the ring. Theory and experiment are used to understand the observations.

Ronald W. Davis, Ph.D., Stanford University, Stanford, Calif.
$450,000 (2 years)
"Single Molecule Nucleic Acid Detection with Nanopipettes"
This group takes a stepwise approach to single nucleic acid molecule detection, using the nanoscale pore in a pulled glass pipette to measure single DNA molecules with attached nanoparticles. Understanding the electrical signals produced by, and the limits of this technology will contribute to next-generation devices with higher informational content.

Round 2 – Awarded August 8, 2005

From

NIH News Release '**2005 Release: NHGRI Expands Effort to Revolutionize Sequencing**

"$1,000 Genome" Grants

Richard B. Fair, Ph.D., Duke University, Durham, N.C.
$510,000 (2 years)
"Droplet-Based Digital Microfluidic Genome Sequencing"
The near-term goal of this group is to demonstrate how existing droplet-based microfluidic electro-wetting technology can be modified to perform sequencing by synthesis reaction chemistry. This method allows for smaller volumes of materials to be used as well as the decoupling of synthesis and detection steps, resulting in more efficient automation.

M. Reza Ghadiri, Ph.D., The Scripps Research Institute, La Jolla, Calif. and
Hagan P. Bayley, Ph.D., Oxford University, UK.
$4.2 million (5 years)
"Single-Molecule DNA Sequencing with Engineered Nanopores"
This project is a collaborative effort between two laboratories that have experience in nanopore research, protein engineering and molecular recognition. The group will engineer a device with the ability to recognize a nucleotide on the basis of changes in electrical current, as it passes through a membrane with tiny channels known as nanopores.

Jene A. Golovchenko, Ph.D., Harvard University, Cambridge, Mass.
$5.2 million (3 years)
"Electronic Sequencing in Nanopores"
The objective of this project is to develop a general utility instrument to provide inexpensive sequencing that can also be used for projects to recognize genome variation. The group will design novel nanopores articulated with probes to sequentially, and directly, identify nucleotides in very long fragments of genomic DNA based on their unique electronic signals.

Susan H. Hardin, Ph.D., VisiGen Biotechnologies, Inc., Houston.
$4.2 million (3 years)
"Real-Time DNA Sequencing"
This group is developing a sequencing system in which polymerase (an enzyme used to synthesize DNA molecules) and nucleotides act together as direct molecular sensors of DNA base identity. The key to the system is the interaction between a fluorescent polymerase and the nucleotide, which emits a signature detectable in real-time.

Xiaohua Huang, Ph.D., University of California, San Diego, La Jolla.
$750,000 (3 years)
"Massively Parallel Cloning and Sequencing of DNA"
The goal of this project is to develop two innovative technologies: massively parallel, whole-genome amplification and DNA sequencing by denaturation. The resulting system amplifies DNA directly on a microchip, enabling the process of sequencing to be done on a single miniaturized device.

Jingyue Ju, Ph.D., Columbia University, New York.
$970,000 (3 years)
"Modulating Nucleotide Size in DNA for Detection by Nanopore"
This group will design and synthesize modified nucleotides of different sizes, which can be incorporated into DNA. When passed through nanopores, the differences between these modified nucleotides will be easier to detect, producing clean sequencing data.

Bhubaneswar (Bud) Mishra, Ph.D., New York University, New York.
$585,000 (2 years)
"Haplotype Sequencing Via Single Molecule Hybridization"
Investigators from this group will hybridize short DNA probes to genomic DNA fragments to determine sequence information. In addition, they will use optical mapping to create restriction maps to help assemble the genome once it is sequenced. The group will then demonstrate how to combine the sequence and maps into distinct haplotype sequences.

Gregory L. Timp, Ph.D., University of Illinois at Urbana-Champaign.
$2.1 million (3 years)
"Sequencing a DNA Molecule Using a Synthetic Nanopore"
This group will explore the feasibility of sequencing a DNA molecule using a type of silicon integrated circuit. The circuit incorporates a nanopore mechanism with a molecular trap that forces the DNA molecule to oscillate back and forth between electrodes, measuring the electrical signal associated with each specific base.

Stephen W. Turner, Ph.D., Nanofluidics, Menlo Park, Calif.
$6.6 million (3 years)
"Real-Time Multiplex Single-Molecule DNA Sequencing"
This group will leverage their "zero-mode waveguide" technology to detect single nucleotides in real-time, as they are incorporated by a DNA polymerase into a growing DNA molecule. The ultimate goal is to create a real-time, multiplex single-molecule DNA sequencing system that produces sequence reads containing hundreds of thousands of nucleotides.

"$100,000 Genome" Grants

Gina L. Costa, Ph.D., Agencourt Personal Genomics., Beverly, Mass.
$1.2 million (2 years)
"Bead-Based Polony Sequencing"
Supplemental funding is expected to accelerate commercialization of this technology that will use oligonucleotide ligation to read DNA sequence, using bead-based, polymerase colony (polony) sequencing technology.

Vera B. Gorfinkel, Ph.D., The State University of New York (SUNY), Stony Brook, N.Y.
$1.5 million (2 years)
"Ultra High Throughput DNA Sequencing System Based on Two-Dimensional Monolith Multi-Capillary Arrays and Nanoliter Reaction Volume"
This group will develop and implement an efficient method capable of sequencing mammalian size genomes by amplifying single template molecules, and subjecting the product to Sanger sequencing and a highly parallel, capillary electrophoresis separation system.

Greg Kellogg, Ph.D., Network Biosystems, Woburn, Mass.
$4.5 million (3 years)
"$100,000 Genome Using Integrated Microfluidic Capillary Electrophoresis"
This group will work to improve performance of Sanger sequencing and PCR as compared to that attainable using capillary electrophoresis systems. To do so, it will miniaturize and integrate current sequencing technologies, building on its microfluidics platform.

Round 3 – Awarded October 4, 2006

From
NIH News Release **'2006 Release: NHGRI Aims to Make DNA Sequencing Faster, More Cost Effective'**

"$1,000 Genome" Grants

John Nelson, Ph.D., General Electric Global Research, Niskayuna, N.Y.
$900,000 (2 years)
"Closed Complex Single Molecule Sequencing"
This team will use existing enzyme and dye-tagged nucleotide resources, the building block of DNA, in a novel way that will simplify the fundamental, front-end chemistry of massively parallel sequencing-by-synthesis. This method uses the natural catalytic cycle of DNA polymerase to capture just a single DNA base on an immobilized primer/template. A fluorescence scanner will be used to scan and identify hundreds of thousands of molecules at once. Then the cycle will be repeated. This phased award will increase if specific milestones are met in the initial experiments.

J. Michael Ramsey, Ph.D., University of North Carolina, Chapel Hill
$3.8 million (4 years)
"Nanoscale Fluidic Technologies for Rapidly Sequencing Single DNA Molecules"
A nanometer is one-billionth of a meter, much too small to be seen with a conventional lab microscope. Several groups are developing nanopores (holes about two nanometers in diameter) for use as DNA sequence transducers and propose to detect an electrical, or ionic, signal from individual DNA molecules. The goal of this group is to fabricate nanoscale channels in which single molecules of DNA will pass between nano-electrodes that are less than two nanometers apart, to measure an electric current that will identify individual bases.

Xiaohua Huang, Ph.D., University of California, San Diego, La Jolla
$275,000 (1 year)
"Genome Sequencing by Ligation Using Nano-Arrays of Single DNA Molecules"
Using an experimental method for DNA sequencing called "single molecule sequencing by ligation," this project aims to develop a method for fabricating high-density arrays of wells with sub-micrometer dimensions for ordering single nanoparticles and DNA molecules. The investigator will attempt to demonstrate that more than 1 billion individual DNA molecules can be sequenced in massive parallel though a process involving cyclic sequencing by ligation, a process where an enzyme is used to join pieces of DNA together. This phased award will increase if specific milestones are met in the initial experiments.

Amit Meller, Ph.D., Boston University, Boston
$2.2 million (3 years)
"High-Throughput DNA Sequencing Using Design Polymers and Nanopore Arrays"
This group along with their industrial partner LingVitae AS, will continue to implement a novel approach previously funded through this program in which a nanopore is used to simultaneously detect electrical and fluorescent signals from many nanopores at one time. A novel sequencing instrument will be fabricated, along with additional analysis tools, with the aim of producing a viable, low-cost sequencing system.

Timothy D. Harris, Ph.D, Helicos Biosciences Corporation, Cambridge, Mass.
$2 million (3 years)
"High Accuracy Single Molecule DNA Sequencing by Synthesis"
This team of investigators has developed a fully automated instrument capable of sequencing single molecules of DNA on a planar surface. The group is now developing a high-throughput version of this technology for the re-sequencing of whole human genomes. The sequencing strategy involves obtaining

short reads (about 25 DNA bases) from billions of strands of DNA immobilized on a surface inside a reagent flow cell. The research plan aims to advance this strategy to achieve high accuracy, re-sequencing of highly variable genomes and assembly of never-before sequenced genomes.

Dmitri V. Vezenov, Ph.D., Lehigh University, Bethlehem, Penn.
$905,000 (3 years)
"Force Spectroscopy Platform for Label Free Genome Sequencing"
This team will apply force spectroscopy, a technique used to understand the mechanical properties of polymer molecules or chemical bonds, to DNA undergoing arrested polymerization to initially demonstrate one-molecule-at-a-time analysis of changes in molecular mechanics at a resolution of a single base. Using optical, near-field probes, the methods of force spectroscopy can be advanced into techniques having massively parallel format, where millions of single DNA base additions can be followed at the same time. The identification of bases will be done exclusively on the basis of changes experienced by the molecule as a whole. The team aims to fabricate a low cost table-top setup suitable for use in a majority of biological, chemical and hospital laboratories.

Peiming Zhang, Ph.D., Arizona State University, Tempe, Ariz.
$895,000 (3 years)
"Fabrication of Universal DNA Nanoarrays for Sequencing by Hybridization"
Expanding the performance of the sequencing-by-synthesis technology, this group will develop a cost-effective method to fabricate universal DNA nanoarrays using nano-contact printing. The current photolithography technology can cause damage to DNA probes, which the group will strive to avoid by using nano-contact printing. With the nano-sized features, a DNA nanoarray can also improve throughput by offering the ability to accommodate billions of DNA molecules in a small area. Hybridization will be detected by atomic force microscopy.

Carlos H. Mastrangelo, Ph.D., Case Western Reserve University, Cleveland
$815,000 (3 years)
"Large-Scale Nanopore Arrays for DNA Sequencing"
This team will aim to develop highly integrated arrays of nanopores that can be fabricated by lithographic methods, along with on-chip silicon-based electronic circuits and circuit techniques that amplify and isolate their various electrical signals. This group will also design a dipole-sensing methodology, which in principle can distinguish signals from each of the DNA bases. Arrays of nanopores will be constructed on silicon substrates using a self-aligned compositional approach. Quadrature dipole moment detectors will be constructed that yield a signal independent of the rotation of the DNA molecule relative to the electrodes.

Jens Gundlach, Ph.D., University of Washington, Seattle
$605,000 (2 years)
"Engineering MspA for Nanopore Sequencing"
The passage of single-strand DNA through a nanometer-scale pore is driven by an electric field revealing information about the DNA sequence. This method has the potential to become an inexpensive, ultrafast DNA sequencing technique. Most nanopore sequencing approaches involve either the protein pore alpha-hemolysin or artificial pores in inorganic materials. This team will use protein-engineering to tailor an alternative protein pore, Mycobacterium smegmatis porin A (MspA), for nanopore sequencing.

"$100,000 Genome" Grants

Michael L. Metzker, Ph.D., Human Genome Sequencing Center, Baylor College of Medicine, Houston
$500,000 (1 year)
"Ultrafast SBS (Sequencing by Synthesis) Method for Large-Scale Human Resequencing"
This team has developed a novel type of fluorescent nucleotide that is modified for sequencing by synthesis. Their goal is to improve the chemical subunits, called reversible terminators, for use in a system that will ultimately be used to sequence DNA templates in high-density arrays, using a sensitive fluorescence detection system.

Steven Jeffrey Gordon, Ph.D., Intelligent Bio-Systems, Inc., Worcester, Mass.
$425,000 (1 year)
"High-Throughput DNA Sequencing by Synthesis Platform"
The main goal of this project is to develop a high-speed, massively parallel DNA sequencing system using unique base analogues and the sequencing by synthesis approach, in collaboration with a group at Columbia University. This application is focused on the development of the subsystems required to construct high-density sample arrays on glass chips and to run sequencing by synthesis reactions on them in an automated, high-throughput fashion.

Appendix C: Next Generation Sequencing Company Overview

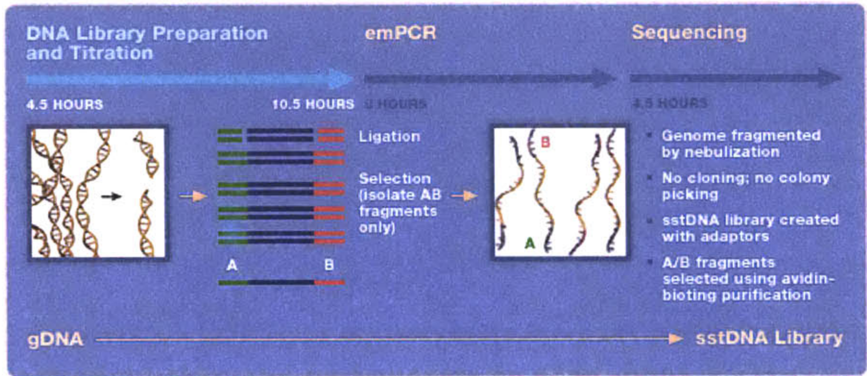
454 Life Sciences

From http://www.454.com/enabling-technology/the-process.asp

DNA Library Preparation

Preparation of the DNA library consists of a few simple steps (Figure 7). Genomic DNA (gDNA) is fractionated into smaller fragments (300-500 base pairs) that are subsequently polished (blunted). Short Adaptors (A and B) are then ligated onto the ends of the fragments. These adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments. Adaptor B contains a 5'-biotin tag that enables immobilization of the library onto streptavidin coated beads. After nick repair, the non-biotinylated strand is released and used as a single-stranded template DNA (sstDNA) library. The sstDNA library is assessed for its quality and the optimal amount (DNA copies per bead) needed for emPCR™ is determined by titration.

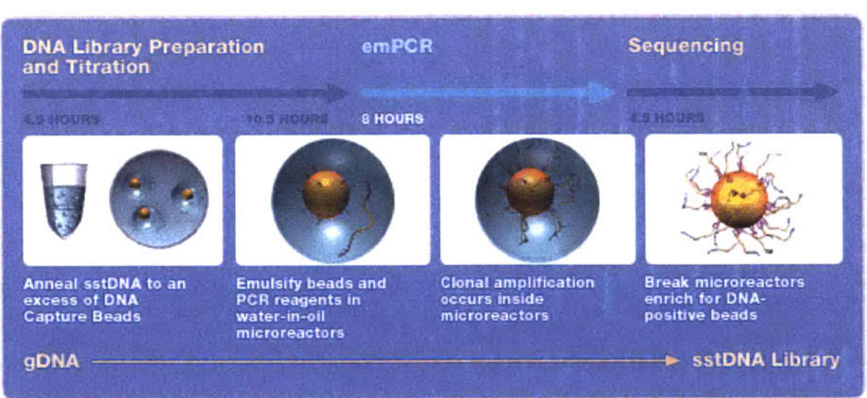
FIGURE 7



emPCR™

The sstDNA library is immobilized onto beads. The beads containing a library fragment carry a single sstDNA molecule. The bead-bound library is emulsified with the amplification reagents in a water-in-oil mixture. Each bead is captured within its own microreactor where PCR amplification occurs. This results in bead-immobilized, clonally amplified DNA fragments.

FIGURE 8

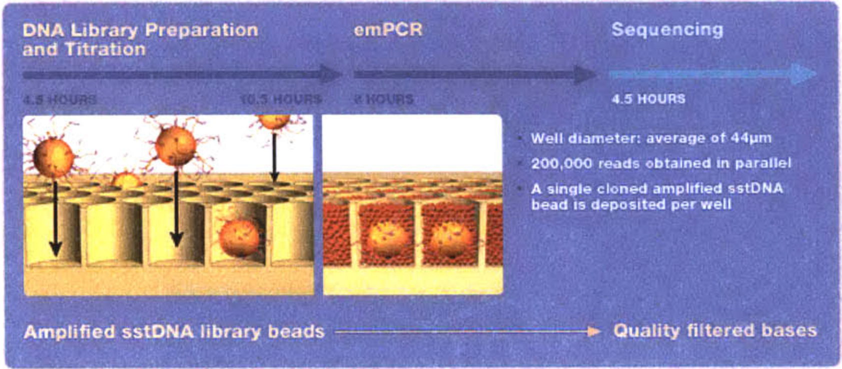


Sequencing

sstDNA library beads are added to the DNA Bead Incubation Mix (containing DNA polymerase) and are layered with Enzyme Beads (containing sulfurylase and luciferase) onto the PicoTiterPlate™ device. The device is centrifuged to deposit the beads into the wells. The layer of Enzyme Beads ensures that the DNA

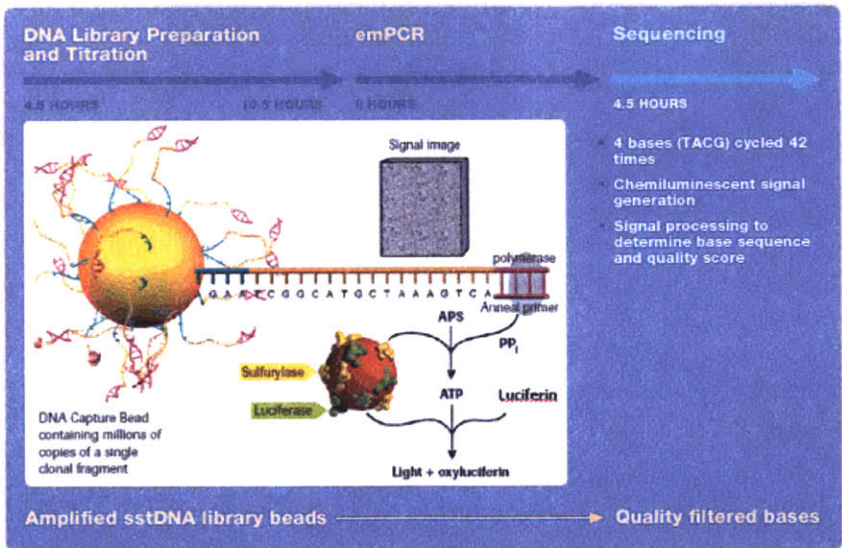
beads remain positioned in the wells during the sequencing reaction. The bead-deposition process maximizes the number of wells that contain a single amplified library bead (avoiding more than one sstDNA library bead per well).

FIGURE 9



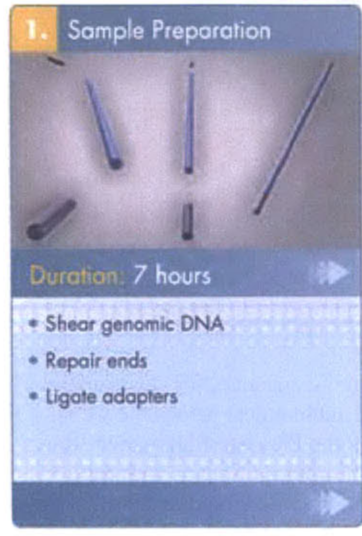
The loaded PicoTiterPlate device is placed into the Genome Sequencer 20™ Instrument. The fluidics subsystem flows sequencing reagents (containing buffers and nucleotides) across the wells of the plate. Nucleotides are flowed sequentially in a fixed order across the PicoTiterPlate device during a sequencing run. During the nucleotide flow, each of the hundreds of thousands of beads with millions of copies of DNA is sequenced in parallel. If a nucleotide complementary to the template strand is flowed into a well, the polymerase extends the existing DNA strand by adding nucleotide(s). Addition of one (or more) nucleotide(s) results in a reaction that generates a light signal that is recorded by the CCD camera in the Instrument. The signal strength is proportional to the number of nucleotides, for example, homopolymer stretches, incorporated in a single nucleotide flow.

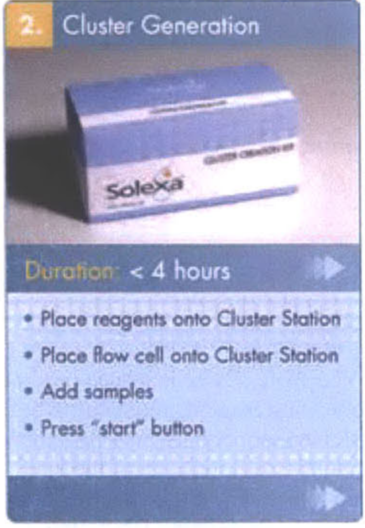
FIGURE 10

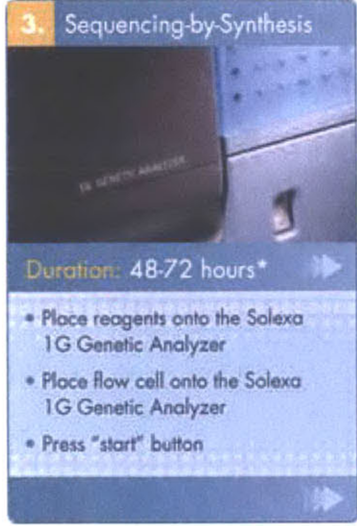


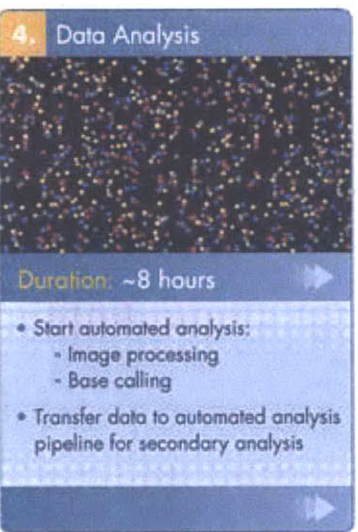
Solexa Inc.

From http://www.solexa.com/technology/sbs.html

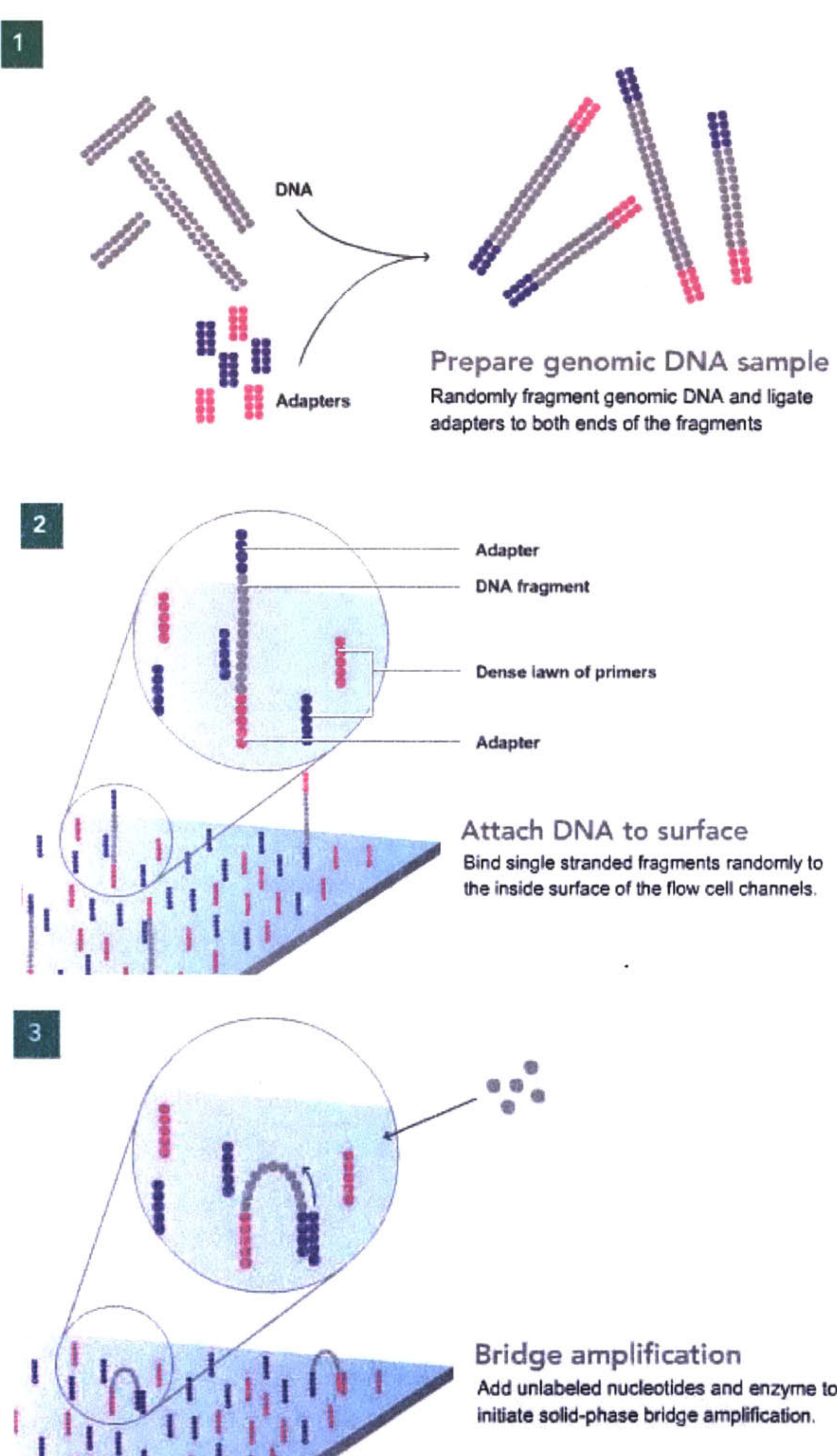


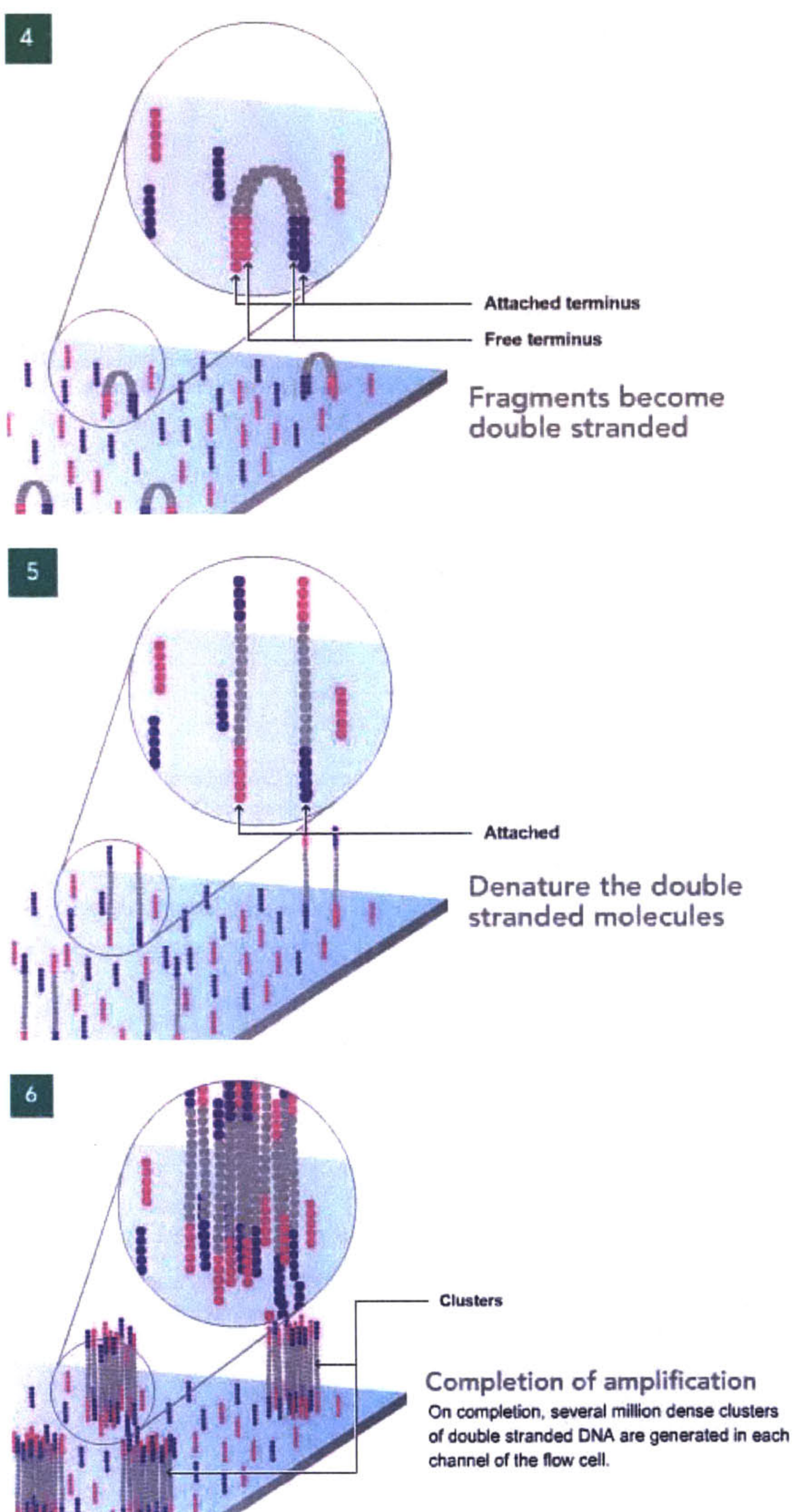




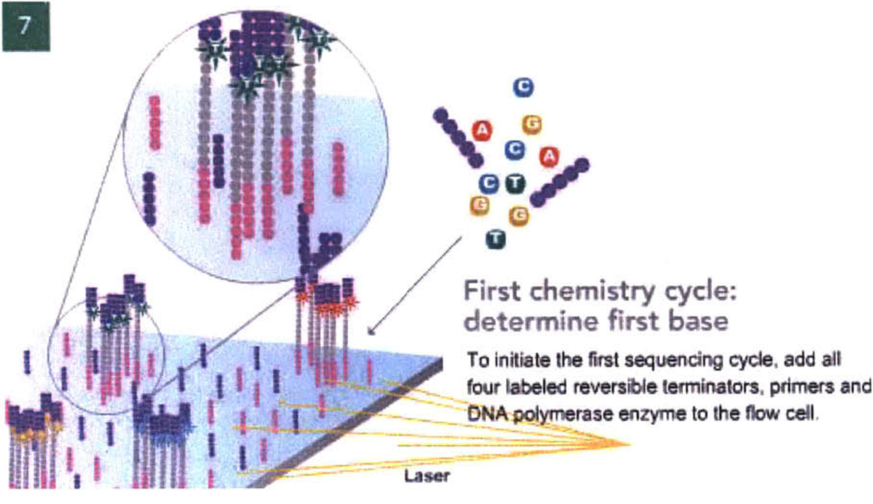


* Duration of the run depends on the desired number of sequencing cycles; complete walk-away automation


1
DNA
Adapters
Prepare genomic DNA sample
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments
2
Adapter
DNA fragment
Dense lawn of primers
Adapter
Attach DNA to surface
Bind single stranded fragments randomly to the inside surface of the flow cell channels.
3
Bridge amplification
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.



4
Attached terminus
Free terminus
Fragments become double stranded
5
Attached
Denature the double stranded molecules
6
Clusters
Completion of amplification
On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.



8

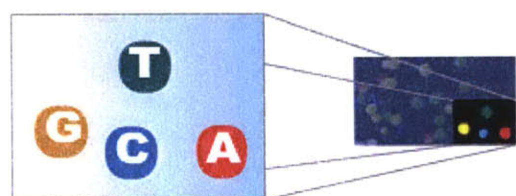
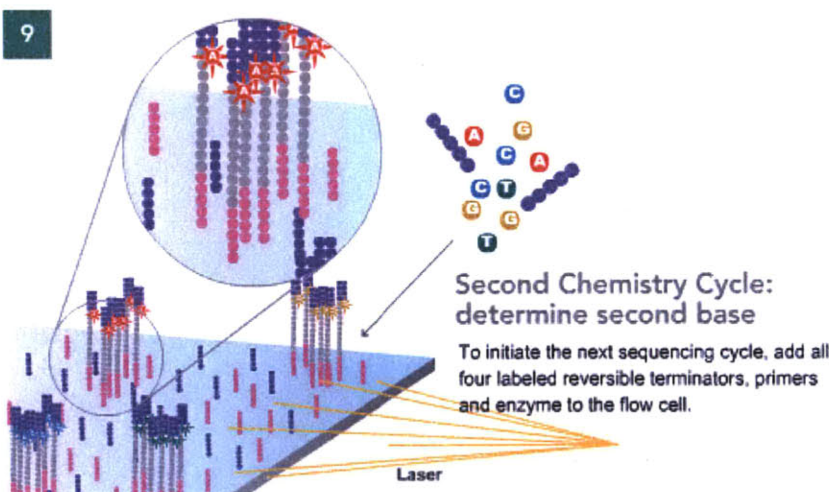


Image of first chemistry cycle

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle

The blocked 3' terminus and the fluorophore from each incorporated base are removed.



10

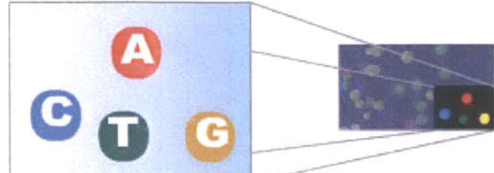
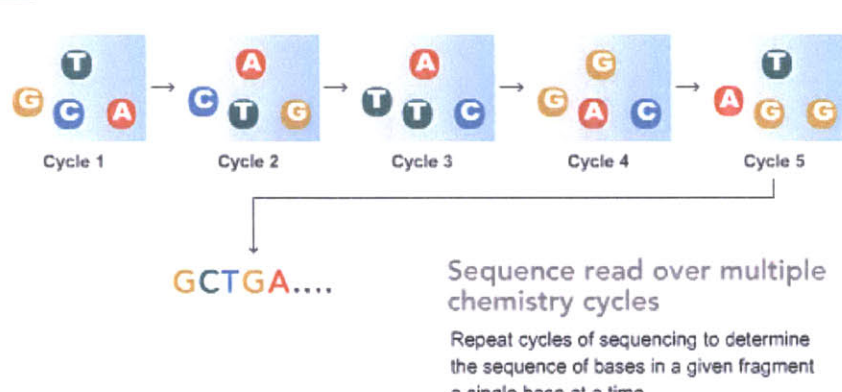


Image of second chemistry cycle is captured by the instrument

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11



Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

Sequencing by Ligation – Church Group and Agencourt Personal Genomics

From Church, G. M., et al. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, Vol 309:1728-1732, September 2005.

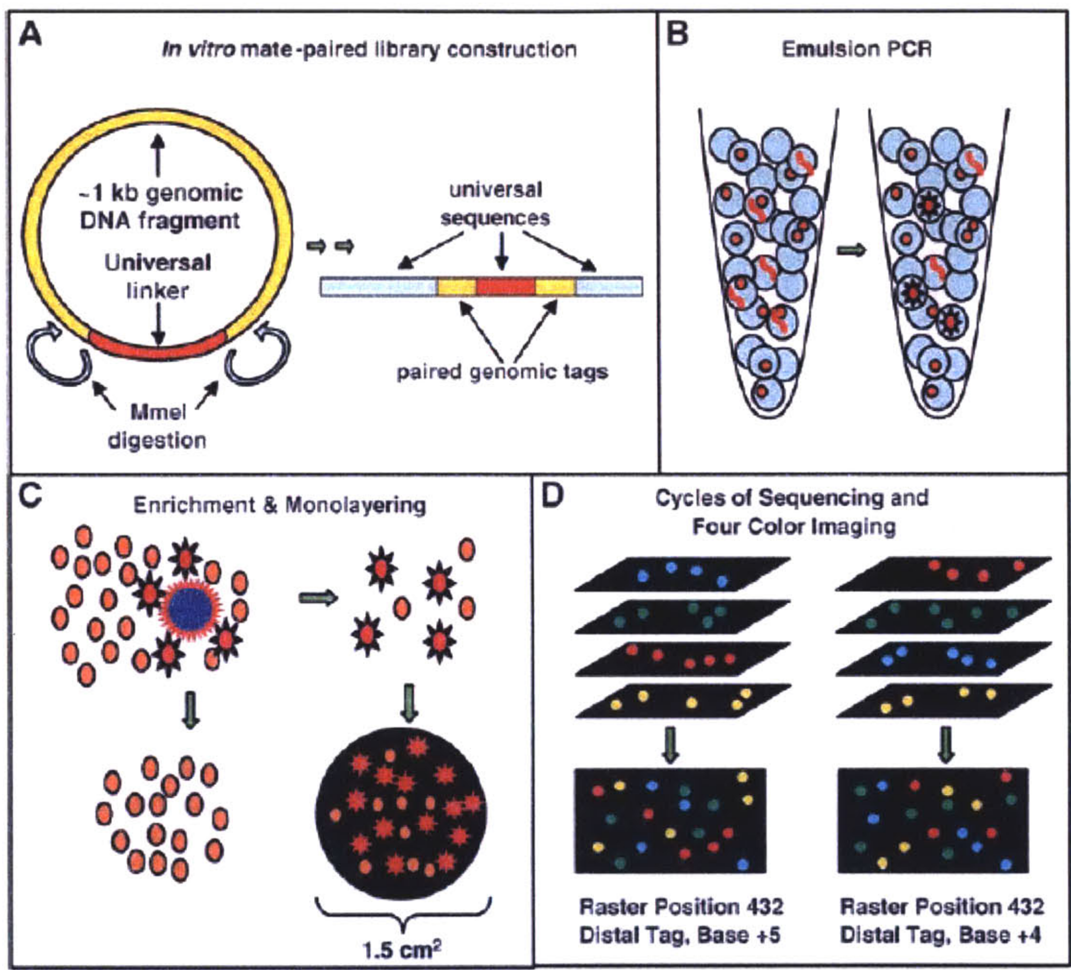
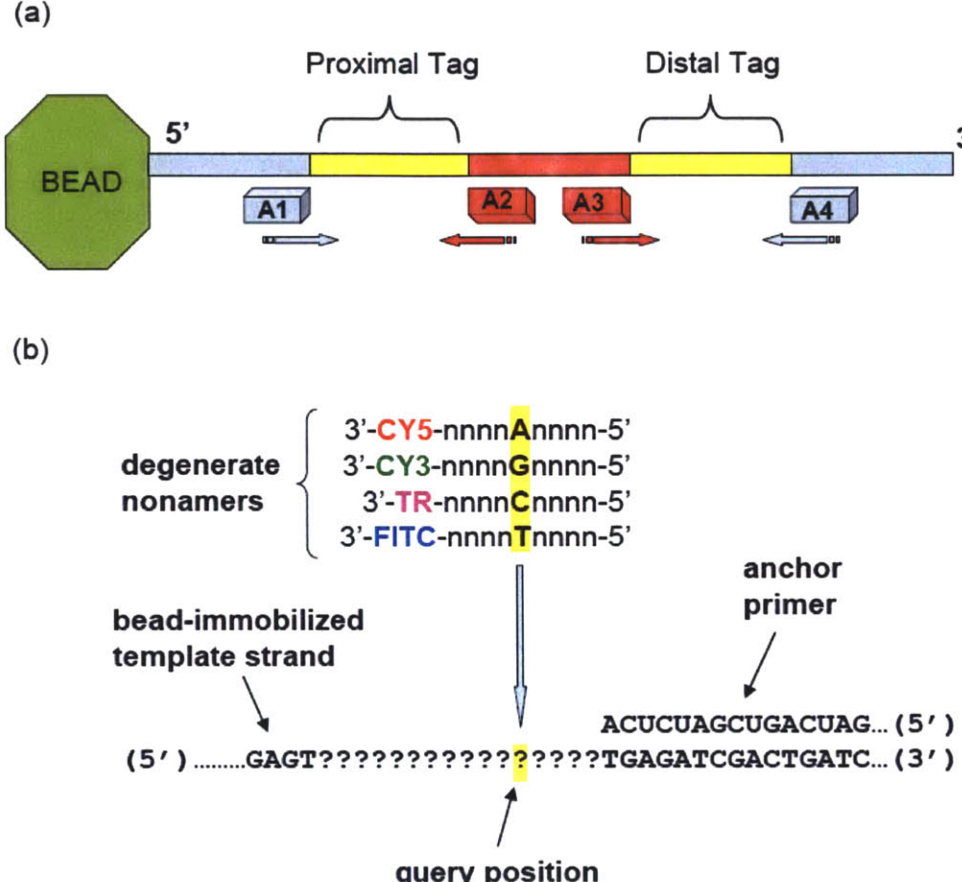


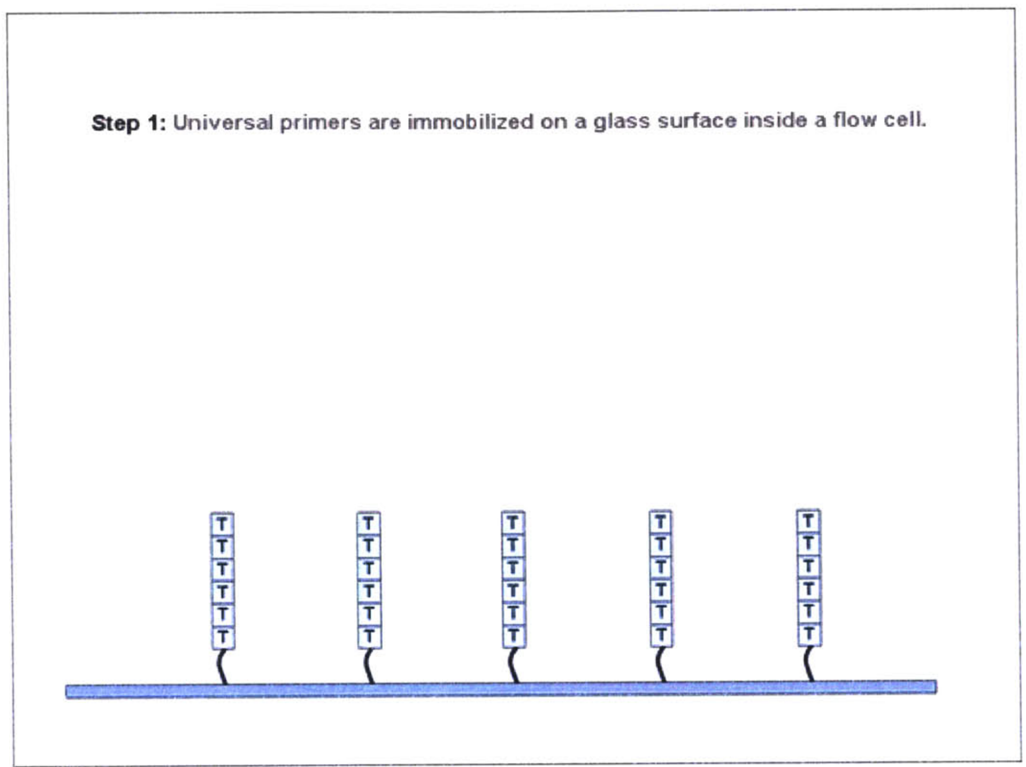
Fig. 1. A multiplex approach to genome sequencing. (A) Sheared, size-selected genomic fragments (yellow) are circularized with a linker (red) bearing Mme I recognition sites (Note S1). Subsequent steps, which include a rolling circle amplification, yield the 134- to 136-bp mate-paired library molecules shown at right. (B) ePCR (*14*) yields clonal template amplification on 1-μm beads (Note S2). (C) Hybridization to nonmagnetic, low-density "capture beads" (dark blue) permits enrichment of the amplified fraction (red) of magnetic ePCR beads by centrifugation (Note S3). Beads are immobilized and mounted in a flowcell for automated sequencing (Note S4). (D) At each sequencing cycle, four-color imaging is performed across several hundred raster positions to determine the sequence of each amplified bead at a specific position in one of the tags. The structure of each sequencing cycle is discussed in the text, Note S6, and fig. S7.

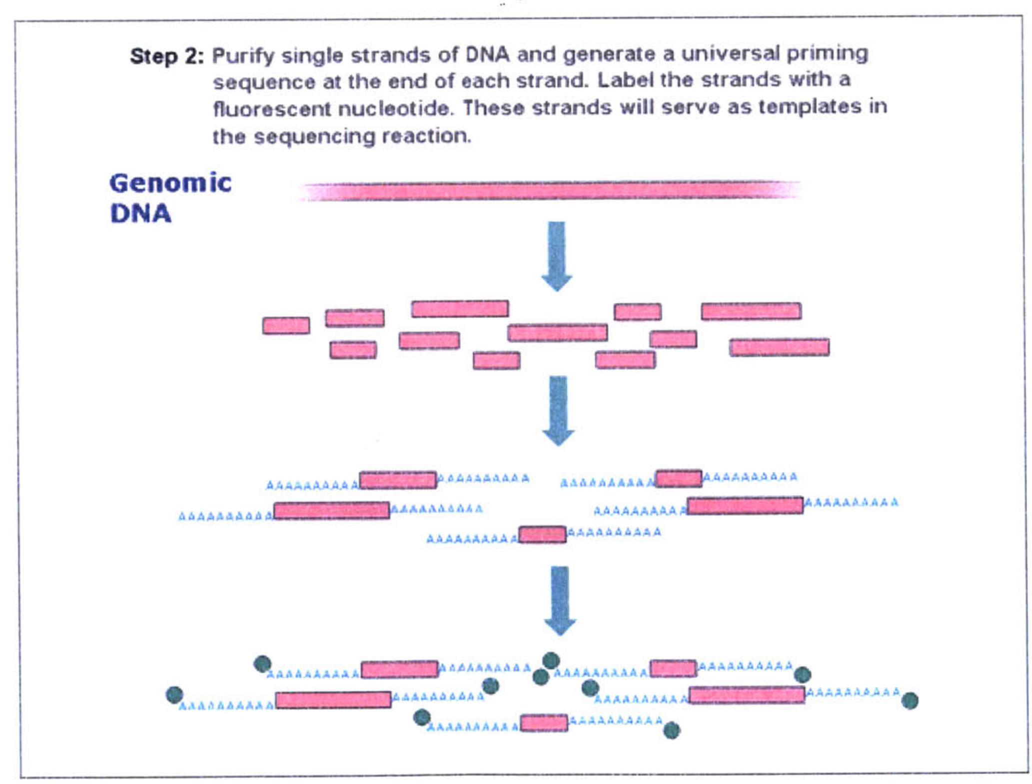


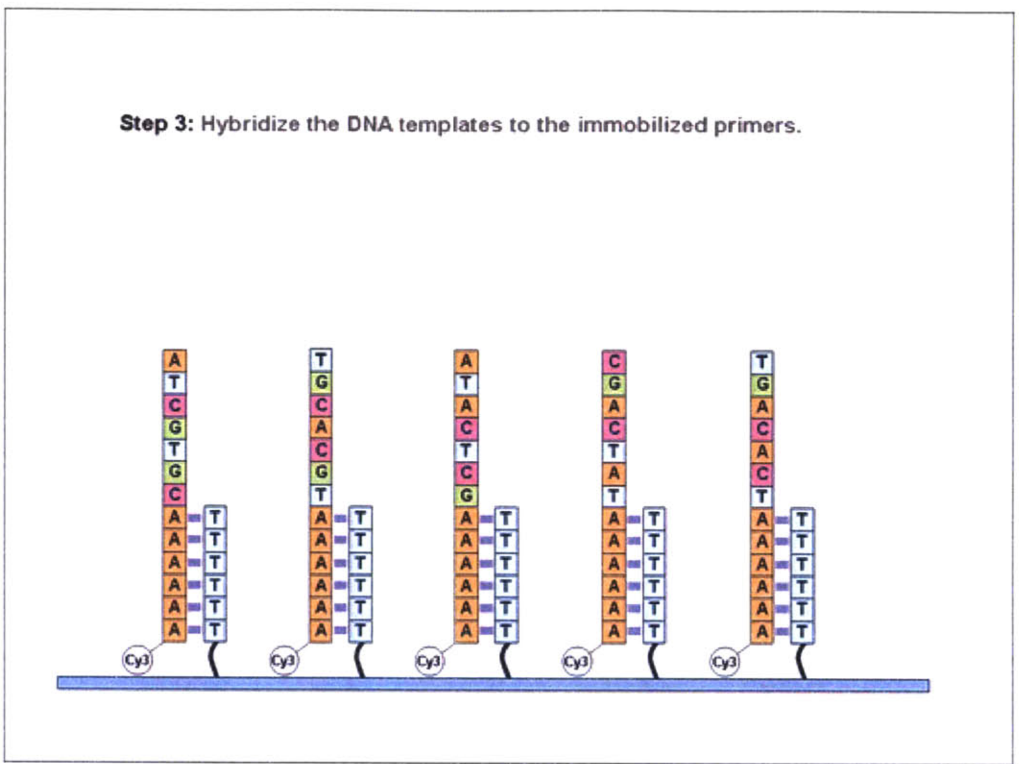
Sequencing By Ligation. The structure of sequencing cycles is described in the text and Supp Note 6. In brief, cycles consist of the following four steps: (a) hybridization of anchor primer, (b) ligation of fluorescent, degenerate nonamers, (c) four color imaging on epifluorescence microscope, (d) stripping of the anchor primer:nonamer complexes prior to beginning the next cycle. The images above are intended to clarify the positioning of anchor primers and the nonamer ligation steps. (a) Hybridization of one of four "anchor primers". The anchor primers are each designed to be complementary to universal sequence immediately 5' or 3' to one of the two tags. A1, A2, A3 and A4, as shown above, indicate the four locations to which anchor primers are targeted relative to the amplicon. Arrows indicate the direction sequenced into the tag from each anchor primer. From anchor primers A1 and A3, we sequence 7 bases into each tag, and from anchor primers A2 and A4, we sequence 6 bases into each tag. Thus, we obtain 13 bp per tag, and 26 bp per amplicon, with 4 to 5 bp gaps within each tag sequence.

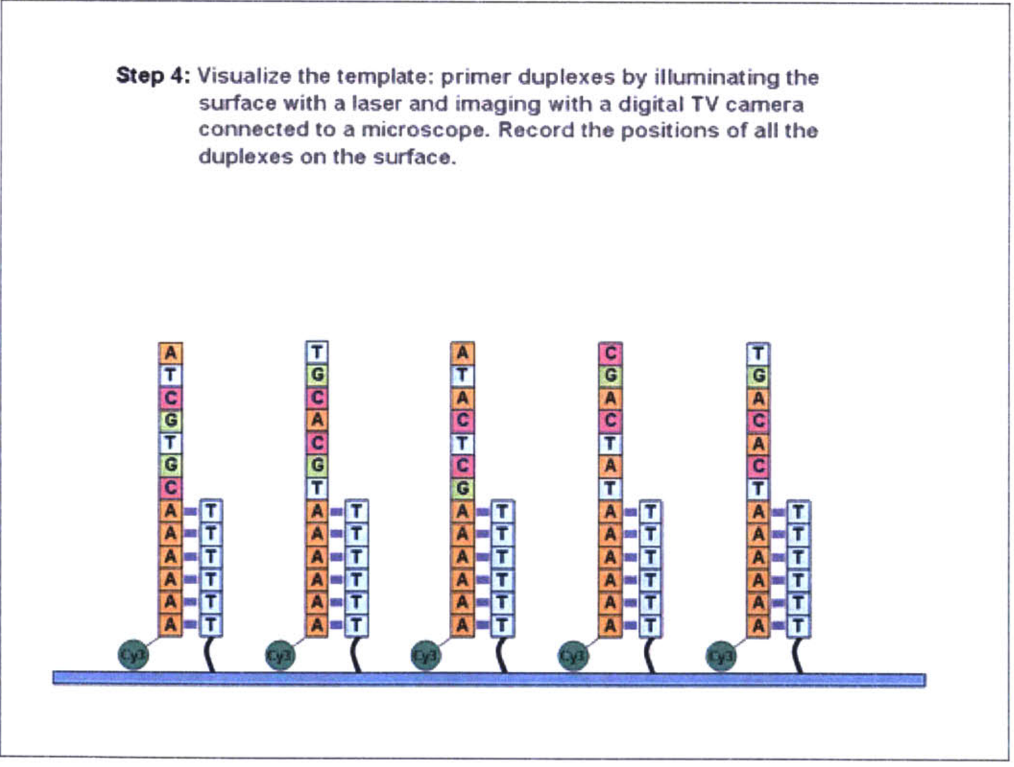
Helicos Biosciences

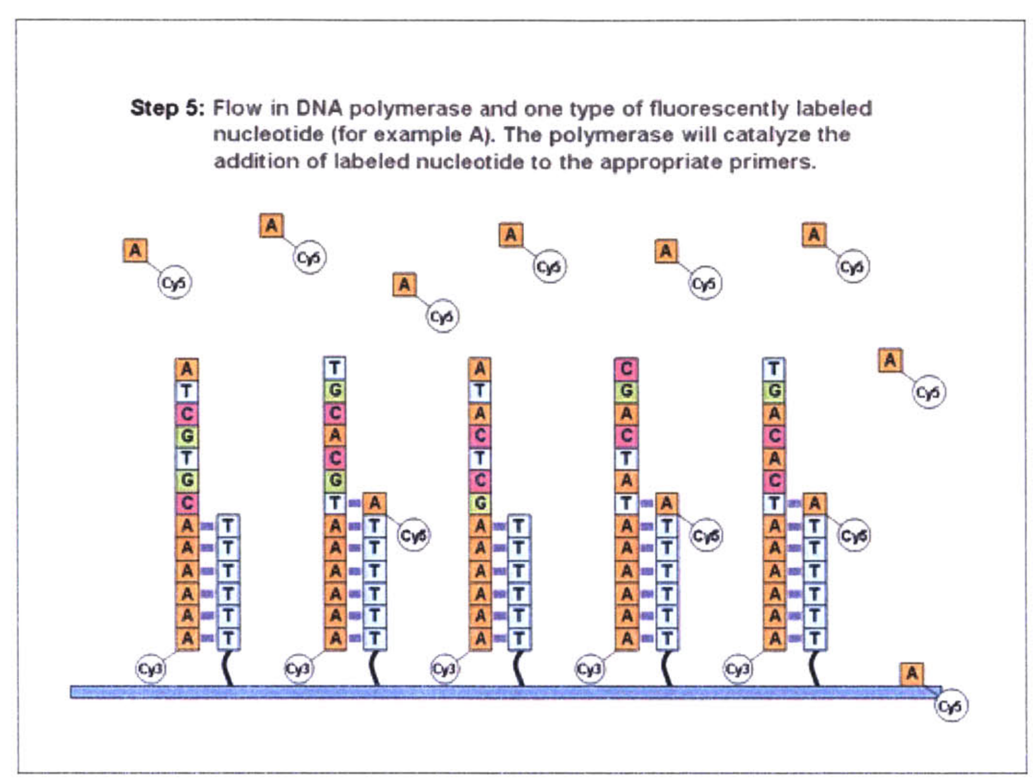
From http://helicosbio.com/default.asp

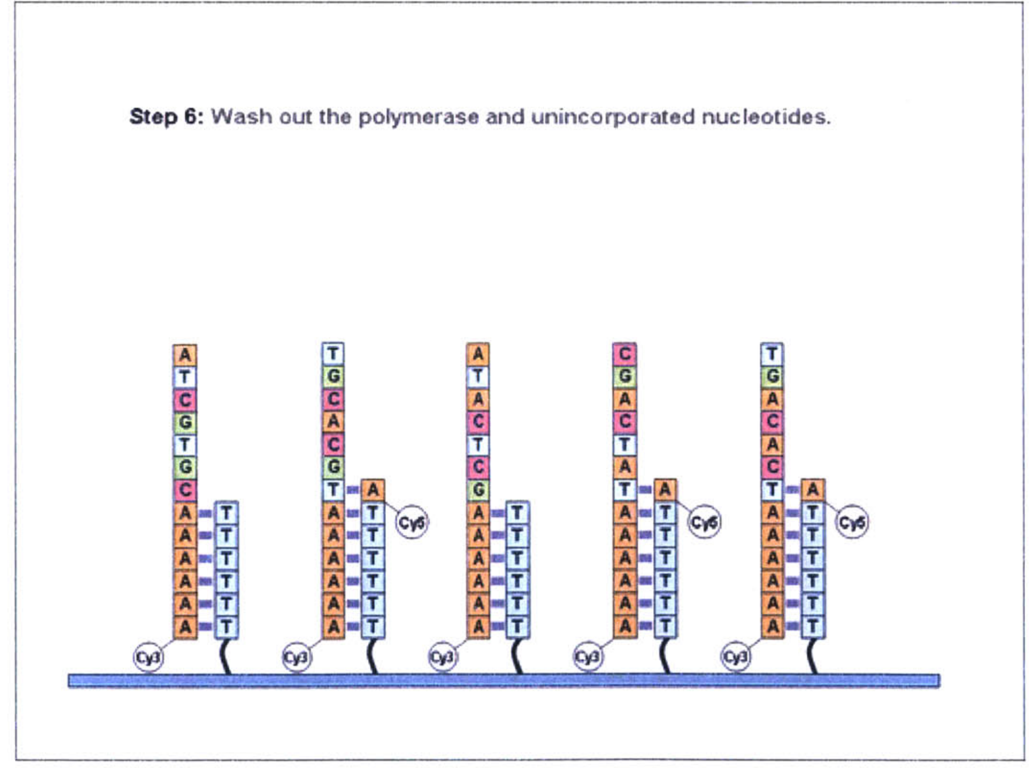


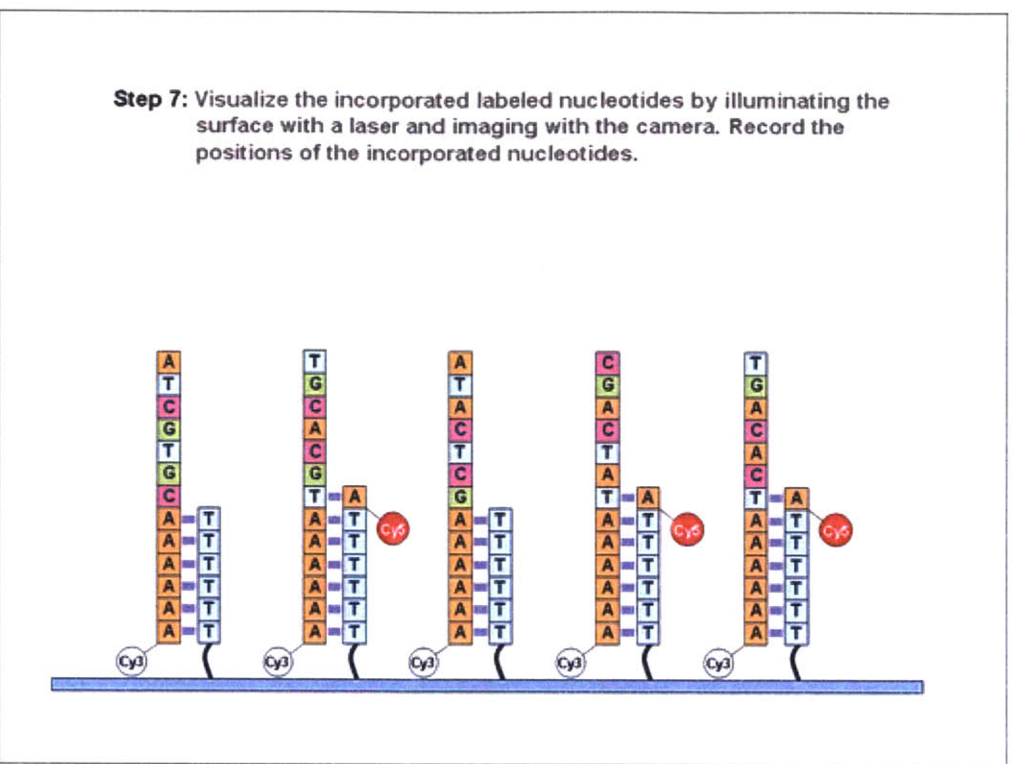


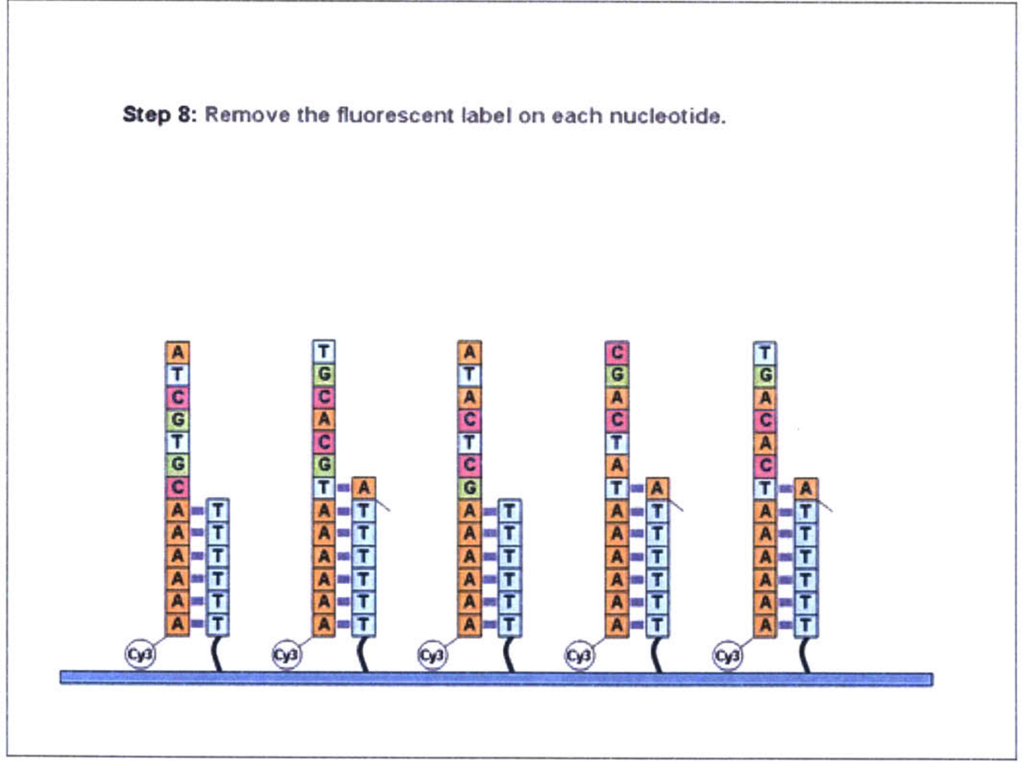

Step 3: Hybridize the DNA templates to the immobilized primers.
Cy3

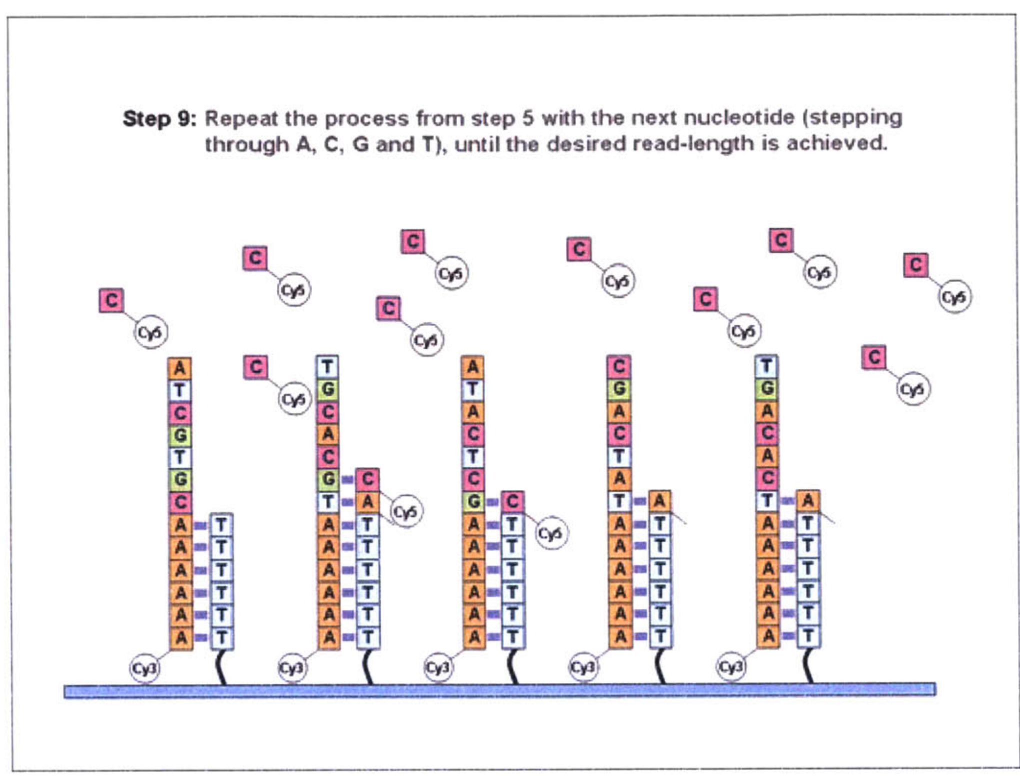

Step 4: Visualize the template: primer duplexes by illuminating the surface with a laser and imaging with a digital TV camera connected to a microscope. Record the positions of all the duplexes on the surface.


Step 5: Flow in DNA polymerase and one type of fluorescently labeled nucleotide (for example A). The polymerase will catalyze the addition of labeled nucleotide to the appropriate primers.


Step 6: Wash out the polymerase and unincorporated nucleotides.


Step 7: Visualize the incorporated labeled nucleotides by illuminating the surface with a laser and imaging with the camera. Record the positions of the incorporated nucleotides.


Step 8: Remove the fluorescent label on each nucleotide.


Step 9: Repeat the process from step 5 with the next nucleotide (stepping through A, C, G and T), until the desired read-length is achieved.

Appendix D: Next Generation Sequencing Survey

Part 1 – Background

Next Generation DNA Sequencing Technology

Exit this survey >>

1. Background

Company/Researcher names will be only be used as reference to group answers from the same efforts but will remain anonymous in the results

1. What is the name of the DNA sequencing company or researcher you are associated with?

2. What type of organization is it?

3. How old is the effort?

4. What is your job title?

5. Are you a company founder?

- Yes
- No

6. How long have you been associated with the technology?

7. Did you have experience developing DNA sequencing technology prior to this effort?

- Yes
- No

Part 2 – Market Overview

Next Generation DNA Sequencing Technology Exit this survey >>

2. Market Overview

8. What effect did the following factors have on the decision to develop your technology?

	Primary Motivation	Encouraged	No Effect	Discouraged	Major Obstacle
Faith in the superiority of the technology					
Market opportunity					
Level of competition					
Availability of funding					
IP Availability					

9. How do you rank the following potential markets for your technology?

	Very Attractive		Attractive		Not Attractive	Not Applicable for my Technology
De Novo sequencing						
Resequencing						
Mutation detection						
Copy number identification						
Genotyping						
Expression Analysis						

10. Based on the applications you listed above, what do you forsee as the market size ($) and customer base for this technology compared to currently employed technologies?

- more customers and larger market value
- more customers and lower market value
- less customers and larger market value
- less customers and lower market value
- Other (please specify)

11. How do you rank the following potential customers?

	Primary	Secondary	Not Likely
Pharma (Biopharma) companies			
Diagnostic companies			
Academic Labs			
Government Labs			

12. How would you describe your relationship to these existing technologies?

	Competitor	Complement	No Relation
ABI 3730			
Affymetrix Genechip			
Illumina Bead Array			
Luminex xMap			

Part 3 – Technology

Next Generation DNA Sequencing Technology Exit this survey >>

3. Technology

13. What do you feel are the major advantages of your technology over those currently being employed for these applications?

- Cost/base
- Accuracy (reduced error rate)
- Throughput (bases/run)
- Other (please specify)

14. To the best of your knowledge, please rank the amount of effort devoted to each of the listed components in order to develop your technology

	Most Effort		Moderate Effort		No Effort
Core Biology (custom enzymes, dyes, etc.)					
Assay Design (component scaling, cycling optimization, etc.)					
Mechanical Integration (fluidic, optics, etc.)					
Image Analysis Software					
Read Quality Analysis Software					

15. If possible, please describe the state of your technology's performance in term's of the following parameters

continuous bases

% quality

total # of bases per read

reads (total to date)

16. When do you expect to have an instrument on the market

- Already have one
- <6 months
- <1 year
- <2 years
- <5 years
- >5 years
- Probably never

Part 4 – Intellectual Property

Next Generation DNA Sequencing Technology — Exit this survey >>

4. Intellectual Property

17. What role do the inventors of your core Intellectual Property (IP) play in your effort

- Primary Investigator
- Founder
- Employee
- Advisor
- Board Member
- No Role
- Other (please specify)

18. What do you consider the primary benefits of your IP?

	Not Important		Important		Very Important
Freedom to operate (no blocking IP)					
Protection from competition/imitation					
Leverage for bargaining with potential partners/buyers					
Cross licensing opportunities					
Credibility for funding					

19. Does your company/lab own or license the following aspects of your technology?

	Owns	Licenses	Both	Don't Know	N/A
Core Biology (custom enzymes, dyes, etc.)					
Assay Design (component scaling, cycling optimization, etc.)					
Mechanical Integration (fluidic, optics, etc.)					
Image Analysis Software					
Read Quality Analysis Software					

20. Have you had to alter your design at all in order to avoid potential IP conflicts

- Yes
- No

21. If you had access to all of the IP available in this market (including that of your competitors) do you believe you could use this to improve your product

- Yes, absolutely
- Maybe, I am not familiar enough with the available technologies
- Not likely

22. If the answer to the previous question was yes, which component of your technology would you improve with the new IP

- Core Biology (custom enzymes, dyes, etc.)
- Assay Design (component scaling, cycling optimization, etc.)
- Mechanical Integration (fluidic, optics, etc.)
- Image Analysis Software
- Read Quality Analysis Software
- Other (please specify)

Part 5 – Funding

Next Generation DNA Sequencing Technology

Exit this survey >>

5. Funding

23. Is your effort a recipient of any of the NIH grants?

- Yes
- No

24. What effect had the NHGRI grants for 'Next Generation Sequencing Technologies' had on your efforts?

- Instrumental
- Encouraging
- Discouraging
- No Effect

25. Why do you feel the grants had this effect?

Appendix E: Next Generation Sequencing Survey Data

What is the name of the DNA sequencing company or researcher you are associated with?

Total Respondents	7
(skipped this question)	0

What type of organization is it?

Academic/Research Lab	3
Company	4
Total Respondents	7
(skipped this question)	0

How old is the effort?

Total Respondents	7
(skipped this question)	0

What is your job title?

Total Respondents	7
(skipped this question)	0

Are you a company founder?

Yes	1
No	4
Total Respondents	5
(skipped this question)	2

How long have you been associated with the technology?

Total Respondents	7
(skipped this question)	0

Did you have experience developing DNA sequencing technology prior to this effort?

Yes	3
No	3
Total Respondents	6
(skipped this question)	1

What effect did the following factors have on the decision to develop your technology?

	Primary Motivation	Encouraged	No Effect	Discouraged	Major Obstacle	Response Average
Faith in the superiority of the technology	5	2	0	0	0	1.29

Market opportunity	3	3	1	0	0	1.71
Level of competition	1	3	3	0	0	2.29
Availability of funding	2	2	2	0	1	2.43
IP Availability	0	5	1	0	0	2.17
Total Respondents	7					
(skipped this question)	0					

How do you rank the following potential markets for your technology?

	Very Attractive		Attractive		Not Attractive	Not Applicable for my Technology
De Novo sequencing	5	0	1	1	0	0
Resequencing	5	1	1	0	0	0
Mutation detection	4	1	2	0	0	0
Copy number identification	3	0	1	2	0	0
Genotyping	2	1	2	0	0	1
Expression Analysis	3	2	0	1	0	0
Total Respondents	7					
(skipped this question)	0					

Based on the applications you listed above what do you forsee as the market size ($) and customer base for this technology compared to currently employed technologies?

	Response Total
more customers and larger market value	4
more customers and lower market value	0
less customers and larger market value	2
less customers and lower market value	0
Other (please specify)	1
Total Respondents	7
(skipped this question)	0

How do you rank the following potential customers?

	Primary	Secondary	Not Likely	Response Average
Pharma (Biopharma) companies	6	1	0	1.14
Diagnostic companies	4	1	2	1.71

Academic Labs	4	3	0	1.43
Government Labs	3	4	0	1.57
Total Respondents	7			
(skipped this question)	0			

How would you describe your relationship to these existing technologies?

	Competitor	Complement	No Relation	Response Average
ABI 3730	2	2	3	2.14
Affymetrix Genechip	2	3	2	2
Illumina Bead Array	2	2	3	2.14
Luminex xMap	0	2	5	2.71
Total Respondents	7			
(skipped this question)	0			

What do you feel are the major advantages of your technology over those currently being employed for these applications?

	Response Total
Cost/base	6
Accuracy (reduced error rate)	5
Throughput (bases/run)	7
Other (please specify)	1
Total Respondents	7
(skipped this question)	0

To the best of your knowledge please rank the amount of effort devoted to each of the listed components in order to develop your technology

	Most Effort		Moderate Effort		No Effort	Response Average
Core Biology (custom enzymes dyes etc.)	3	1	1	0	2	2.57
Assay Design (component scaling cycling optimization etc.)	1	3	2	0	1	2.57
Mechanical Integration (fluidic optics etc.)	1	4	2	0	0	2.14
Image Analysis Software	0	4	1	0	2	3
Read Quality Analysis Software	0	3	1	1	2	3.29
Total Respondents	7					
(skipped this question)	0					

If possible please describe the state of your technology's performance in term's of the following parameters

	Response Total
# continuous bases	3
% quality	3
total # of bases per read	3
# reads (total to date)	3
Total Respondents	3
(skipped this question)	4

When do you expect to have an instrument on the market

	Response Total
Already have one	2
<6 months	1
<1 year	0
<2 years	1
<5 years	1
>5 years	2
Probably never	0
Total Respondents	7
(skipped this question)	0

What role do the inventors of your core Intellectual Property (IP) play in your effort

	Response Total
Primary Investigator	3
Founder	3
Employee	1
Advisor	1
Board Member	0
No Role	0
Other (please specify)	1
Total Respondents	5
(skipped this question)	2

What do you consider the primary benefits of your IP?

	Not Important		Important		Very Important	Response Average
Freedom to operate (no blocking IP)	0	0	1	1	2	4.25
Protection from competition/imitation	0	0	1	2	1	4
Leverage for bargaining with potential partners/buyers	0	0	2	1	1	3.75

Cross licensing opportunities	0	2	2	0	0	2.5
Credibility for funding	1	1	1	0	1	2.75
Total Respondents	4					
(skipped this question)	3					

Does your company/lab own or license the following aspects of your technology?

	Owns	Licenses	Both	Don't Know	N/A	Response Average
Core Biology (custom enzymes dyes etc.)	1	1	1	0	1	2
Assay Design (component scaling cycling optimization etc.)	2	0	0	1	1	2
Mechanical Integration (fluidic optics etc.)	2	0	1	1	0	2.25
Image Analysis Software	2	0	0	2	0	2.5
Read Quality Analysis Software	2	0	0	2	0	2.5
Total Respondents	4					
(skipped this question)	3					

Have you had to alter your design at all in order to avoid potential IP conflicts

	Response Total
Yes	3
No	2
Total Respondents	5
(skipped this question)	2

If you had access to all of the IP available in this market (including that of your competitors) do you believe you could use this to improve your product

	Response Total
Yes	2
Maybe	1
Not likely	2
Total Respondents	5
(skipped this question)	2

If the answer to the previous question was yes which component of your technology would you improve with the new IP

	Response Total
Core Biology	3

Assay Design	1
Mechanical Integration	1
Image Analysis Software	1
Read Quality Analysis Software	1
Other (please specify)	1
Total Respondents	3
(skipped this question)	4

Is your effort a recipient of any of the NIH grants?

	Response Total
Yes	4
No	1
Total Respondents	5
(skipped this question)	2

What effect had the NHGRI grants for 'Next Generation Sequencing Technologies' had on your efforts?

	Response Total
Instrumental	3
Encouraging	3
Discouraging	0
No Effect	0
Total Respondents	5
(skipped this question)	2

Why do you feel the grants had this effect?

Total Respondents	2
(skipped this question)	5