COMPLETE GENOMICS INC
Form 10-K
March 30, 2011
Table of Contents

UNITED STATES

SECURITIES AND EXCHANGE COMMISSION

Washington, D.C. 20549

Form 10-K

(Mark One)

x **ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the fiscal year ended December 31, 2010

OR

¨ **TRANSITION REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934**

For the transition period from_____ to _____

Commission file number: 001-34939

Complete Genomics, Inc.

(Exact Name of Registrant as Specified in its Charter)

Delaware	20-3226545
(State or Other Jurisdiction of Incorporation or Organization)	(I.R.S. Employer Identification No.)

2071 Stierlin Court

Mountain View, California	94043
(Address of Principal Executive Offices)	(Zip Code)

(650) 943-2800

(Registrant s Telephone Number, Including Area Code)

Securities registered pursuant to Section 12(b) of the Act:

Title of Each Class	Name of Each Exchange on Which Registered
Common Stock, $0.001 par value	The NASDAQ Stock Market LLC

Securities registered pursuant to Section 12(g) of the Act:

None

Indicate by check mark if the registrant is a well-known seasoned issuer, as defined in Rule 405 of the Securities Act. Yes ¨ No x

Indicate by check mark if the registrant is not required to file reports pursuant to Section 13 or Section 15(d) of the Act. Yes ¨ No x

Indicate by check mark whether the registrant: (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months (or for such shorter period that the registrant was required to file such reports), and (2) has been subject to such filing requirements for the past 90 days. Yes x No ¨

Indicate by check mark whether the registrant has submitted electronically and posted on its corporate Web site, if any, every Interactive Data File required to be submitted and posted pursuant to Rule 405 of Regulation S-T during the preceding 12 months (or for such shorter period that the registrant was required to submit and post such files). Yes ¨ No ¨

Indicate by check mark if disclosure of delinquent filers pursuant to Item 405 of Regulation S-K is not contained herein, and will not be contained, to the best of registrant s knowledge, in definitive proxy or information statements incorporated by reference in Part III of this Form 10-K or any amendment to this Form 10-K. ¨

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, or a smaller reporting company. See the definitions of large accelerated filer, accelerated filer and smaller reporting company in Rule 12b-2 of the Exchange Act. (Check one):

Large accelerated filer	¨	Accelerated filer	¨
Non-accelerated filer	x (Do not check if a smaller reporting company)	Smaller reporting company	¨

Indicate by check mark whether the registrant is a shell company (as defined in Rule 12b-2 of the Exchange Act). Yes ¨ No x

As of February 28, 2011, the number of outstanding shares of the registrant s common stock, par value $0.001 per share, was 25,976,693.

DOCUMENTS INCORPORATED BY REFERENCE

Certain sections of the registrant s definitive Proxy Statement for the 2011 Annual Meeting of Stockholders to be filed with the Securities and Exchange Commission pursuant to Rule 14A not later than 120 days after end of this fiscal year covered by this Form 10-K are incorporated by reference into Part III of this Form 10-K.

Table of Contents

COMPLETE GENOMICS, INC.

FORM 10-K FOR THE FISCAL YEAR ENDED DECEMBER 31, 2010

TABLE OF CONTENTS

Table of Contents

PART I

All statements in this Annual Report on Form 10-K that are not historical are forward-looking statements within the meaning of the federal securities laws. These include statements regarding our anticipates, beliefs, estimates, intentions, strategies or the like. Such statements are based on our current expectations and are subject to a number of factors and uncertainties that could cause actual results to differ materially from those described in the forward-looking statements. We cannot assure you that actual results or business conditions will not differ materially from those projected or suggested in such forward-looking statements as a result of various factors, including, but not limited to, those discussed in Risk Factors contained in Item 1A of this Annual Report on Form 10-K. Unless required by law, we do not undertake to release publicly any updates or revisions to any forward-looking statements contained herein to reflect any change in our expectations with regard thereto or any change in events, conditions, or circumstances on which any such statements are based.

ITEM 1. BUSINESS

Initial Public Offering

On November 16, 2010, we closed our initial public offering (the IPO) of 6,000,000 shares of common stock at an offering price of $9.00 per share, resulting in net proceeds of approximately $47.2 million, after deducting underwriting discounts, commissions and offering expenses.

Overview

We are a life sciences company that has developed and commercialized an innovative DNA sequencing platform, and our goal is to become the preferred solution for complete human genome sequencing and analysis. Our Complete Genomics Analysis Platform, or CGA Platform, combines our proprietary human genome sequencing technology with our advanced informatics and data management software and our innovative, end-to-end, outsourced service model to provide our customers with data that is immediately ready to be used for genome-based research. We believe that our solution can provide academic and biopharmaceutical researchers with complete human genomic data and analysis at an unprecedented combination of quality, cost and scale without requiring them to invest in in-house sequencing instruments, high-performance computing resources and specialized personnel. By removing these constraints and broadly enabling researchers to conduct large-scale complete human genome studies, we believe that our solution has the potential to significantly advance medical research and expand understanding of the basis, treatment and prevention of complex diseases.

We believe that our complete human genome sequencing technology, which is based on our proprietary DNA arrays and ligation-based read technology, is superior to existing commercially available complete human genome sequencing methods in terms of quality, cost and scale. In the DNA sequencing industry, complete human genome sequencing is generally deemed to be coverage of at least 90% of the nucleotides in the genome. Because we have optimized our technology platform and our operations for the unique requirements of high-throughput complete human genome sequencing, we are able to achieve accuracy levels of 99.999% at a total cost that is significantly less than the total cost of purchasing and using commercially available DNA sequencing instruments and information process technology and then performing all the required sequence data assembly and analysis. We believe that we will be able to further improve our accuracy levels and reduce the total cost of sequencing and analysis, enabling us to maintain significant competitive advantages over the next several years. Because our technology resides only in our centralized facilities, we can quickly and easily implement enhancements and provide their benefits to our entire customer base. Our goal is to be the first company to sequence and analyze high-quality complete human genomes, at scale, for a total cost of under $1,000 per genome.

From the earliest days of the field of genomic sequencing to the present, companies and organizations that have achieved sequencing milestones in quality, cost and scale have immediately announced and/or published

Table of Contents

these sequencing results. We regularly and actively monitor publications and have compared the parameters of our sequencing process and the sequencing results of competitive commercially available technologies announced in these various publications. We are currently unaware of any scientific publications by competitors publicly announcing superior sequencing results. Based on the above, we believe that our complete human genome sequencing technology provides a superior combination of quality, cost and scale when compared to existing commercially available complete genome sequencing methods, when taking into consideration the total cost of purchasing, operating and maintaining the instruments and information systems necessary for complete human genome sequencing.

While our competitors primarily sell DNA sequencing instruments and reagents that produce raw sequenced data, requiring their customers to invest significant additional resources to process that raw data into a form usable for research, we offer our customers an end-to-end, outsourced solution that delivers research-ready genomic data. As the cost of complete human genome sequencing declines, we believe the basis of competition in our industry will shift from the cost of sequencing to the value of the entire sequencing solution. We believe that our integrated advanced informatics and data management services will emerge as a key competitive advantage as this shift occurs.

Our genome sequencing center, which began commercial operations in May 2010, combines a high-throughput sample preparation facility, a collection of our proprietary high-throughput sequencing instruments and a large-scale data center. Our customers ship us their samples via common carrier services such as Federal Express and United Parcel Service. We then sequence and analyze these samples and provide our customers with finished, research-ready data, enabling them to focus exclusively on their single highest priority, discovery.

In 2010, we sequenced over 800 complete human genomes, including more than 300 in the fourth quarter of 2010, and have an order backlog at December 31, 2010 of over 1,000 genomes. Our customers include some of the leading global academic and government research centers and biopharmaceutical companies. At present, our facility has the capacity to sequence and analyze over 400 complete human genomes per month. We expect this capacity to increase between two- and three-fold by year-end 2011 as we deploy additional sequencers and increase the throughput of our sequencing process through software refinements and component upgrades. In future years, we plan to construct additional genome centers in the United States and other international strategic markets to accommodate an expected growing global demand for high-quality, low-cost complete human genome sequencing on a large scale.

In December 2010, we announced that for the last 500 complete human genomes that we had sequenced, an average of over 98% of the genome was read at 10-fold or greater coverage. In addition, our software made high confidence calls of an average of over 95% of the genome and over 95% of the exome.

Market Overview

Background

Every organism has a genome that contains the full set of biological instructions required to build and maintain a living example of that organism. The information contained in a genome is stored, or encoded, in deoxyribonucleic acid, or DNA, a nucleic acid that is found in each cell of the organism. DNA is divided into discrete units called genes, which carry specific information necessary to perform a particular biological function, such as instructions for making proteins. The chemical building blocks that make up each gene are the molecules adenine, cytosine, guanine and thymine, labeled as A, C, G and T, respectively, which are known as nucleotide bases. Human DNA has approximately three billion nucleotide bases, and their precise order is commonly known as the DNA or genetic sequence.

Studying how genes and proteins differ between species and among individuals within a species, or genetic variations, helps scientists to determine their functions and roles in health and disease and, we expect, will continue to drive advancements in medical research and diagnostics.

Table of Contents

Genetic Analysis Market

Genetic analysis products comprise instruments and consumables, as well as associated hardware, software and services directly involved in the study of DNA and RNA. The medical research market consists of laboratories generally associated with universities, medical research centers and government institutions, as well as biotechnology and pharmaceutical companies. In the longer term, we believe genetic analysis tools will likely play a critical role in molecular diagnostics. By detecting small, individual genetic differences, we believe molecular diagnostic tests could be used to identify predisposition to or the presence of a disease, to select appropriate medication and dosage, and to monitor disease progression and response to treatment.

The primary genetic analysis methods traditionally used by genetic researchers fall into three categories: DNA sequencing, genotyping and gene expression analysis. DNA sequencing is the process of determining the exact order, or sequence, of the individual nucleotides in a DNA strand so that this information can be correlated to the genetic activity influenced by that segment of DNA. Genotyping is the process of examining certain known mutations or variations in the DNA sequence of genes to determine whether the particular variant can be associated with a specific disease susceptibility or drug response. Gene expression analysis is the process of examining the molecules that are produced when a gene is activated, or expressed, to determine whether a particular gene is expressed in a specific biological tissue.

The Importance of Complete Human Genome Sequencing

One of the most difficult challenges facing the genetic research and analysis industry is improving our understanding of how genes contribute to diseases that have a complex pattern of inheritance. For many diseases, multiple genes each make a subtle contribution to a person s predisposition or susceptibility to a disease or response to a drug treatment protocol. Accordingly, we believe that unraveling this complex network will be critical to understanding human health and disease. We believe that sequencing complete human genomes is the most comprehensive and accurate method by which to achieve these objectives and improve our understanding of human disease. However, the cost and complexity associated with complete human genome sequencing have been prohibitively high for researchers and have slowed our progress in understanding the genetic underpinnings of disease.

Due to these limitations, many researchers have been using an alternative approach in which a small portion of the genome, referred to as the exome, is targeted, enriched and sequenced, which requires less than 5% of the sequencing compared to sequencing required for a complete genome. However, important areas of the genome lie outside of the exome, such as the promoter regions that control gene expression and other conserved regions of the genome that are believed to perform regulatory functions. Moreover, current exome selection technologies are inefficient, typically sequencing a lower percentage of the exome than can be sequenced by complete human genome sequencing. Within the next several years, we believe the combination of the low cost of sequencing complete human genomes and the advancements in complete human genome sequencing will drive the adoption of complete human genomes sequencing and away from targeted exome sequencing.

Complete Genomics Solution

Although other sequencing technologies have led to dramatic reductions in cost and improvements in quality and throughput for complete human genome sequencing, they were designed as general-purpose instruments for sequencing the DNA or RNA of plants, animals, bacteria and viruses. More specifically, these technologies were not designed solely for sequencing large numbers of complete human genomes.

We have developed a novel approach focused on complete human sequencing. We combine our proprietary human genome sequencing technology, which achieves accuracy levels of 99.999%, with our advanced informatics and data management software and our innovative, end-to-end service model, to deliver research-ready genomic data at a total cost that is significantly less than the total cost of purchasing and using commercially available DNA sequencing instruments and the required information management hardware and

Table of Contents

software. We believe this novel outsourced solution overcomes the key limitations of other sequencing technologies and addresses the unmet needs of the complete human genome research market.

Proprietary Sequencing Technology

There are two primary components of our proprietary human genome sequencing technology: DNA nanoball, or DNB, arrays and combinatorial probe-anchor ligation, or cPAL, reads. Our patterned DNB arrays, due to their small size and biochemical characteristics, enable us to pack DNA very efficiently on a silicon chip. We have developed a proprietary process that causes the DNA to adhere to desired spots on the chip, while conversely preventing the DNA from adhering to the area between these spots. This enables us to affix individual particles of DNA to over 90% of these spots, leading to increased efficiency in nanoarray assembly. In addition, we have developed a highly accurate cPAL read technology, which enables us to read the DNA fragments efficiently using small concentrations of low-cost reagents while retaining extremely high single-read accuracy.

We believe this unique combination of our proprietary DNB and cPAL technologies is superior in both quality and cost to other commercially available approaches and provides us with significant competitive advantages. As reported in the January 2010 edition of *Science*, we sequenced a complete human genome at a consumables cost of approximately $1,800 and with a consensus error rate of approximately 1 error in 100,000 nucleotides. Our read accuracy was further validated by one of our customers, the Institute for Systems Biology, or ISB, as published in *Science Express* in March 2010. To our knowledge, based on our review of scientific publications in the genome sequencing field announcing sequencing results, there are no commercially available technologies that have achieved quality and cost comparable to our sequencing results. We have identified and are developing additional performance enhancements to our core technologies that we believe will enable us to maintain significant competitive advantages over the next several years. As we implement these technological enhancements, our goal is to be the first company to sequence and analyze high-quality complete human genomes, at scale, for a total cost of under $1,000 per genome.

Advanced Informatics and Data Management Software

Sequencing complete human genomes generates substantial amounts of data that must be managed, stored and analyzed. While many users of instrument-based sequencing systems have historically conducted their own in-house data analysis on a limited number of genomes, many of these users lack the computing, storage and network bandwidth necessary to manage the massive data sets generated by larger scale complete human genome studies. In response to this need by our customers, we have built a genomic data processing facility with computing infrastructure for managing both small- and large-scale genomic sequencing projects.

There are two major components of our complete data management solution: assembly software and analysis software. Assembly is the process of using computers to organize all of the overlapping 70-base nucleotide sequences to reconstruct the complete human genome. Our proprietary assembly software uses advanced data analysis algorithms and statistical modeling techniques to accurately reconstruct over 90% of the complete human genome from approximately two billion 70-base reads. After assembling the genomic data, we use our analysis software to identify and annotate key differences, or variants, in each genome.

By using our analytical tools and data management software, our customers can significantly reduce their investments in computing infrastructure. Our customers are provided with reliable access to assembled and annotated sequence data in multiple formats to ease data sharing and comparative analyses. In addition, our data storage options provide flexibility and allow customers to customize their data management strategy based on their particular business and scientific requirements. We have also developed a suite of open source analytical tools, called CGA™ Tools, designed to enable our customers to rapidly analyze the data we generate from their samples. As the reagent cost of sequencing declines, we believe that the cost and complexity of data analysis and management will emerge as the primary limiting factor for conducting complete human genome analysis.

Table of Contents

Innovative, End-to-End, Outsourced Solution

While our competitors primarily sell DNA sequencing instruments and reagents that produce raw sequenced data, requiring their customers to invest significant additional resources to process that raw data into a form usable for research, we offer our customers an end-to-end, outsourced solution that delivers research-ready genomic data. Our genome sequencing center combines a high-throughput sample preparation facility, a collection of our proprietary high-throughput sequencing instruments and a large-scale data center. Our customers ship us their samples via common carrier services such as Federal Express and United Parcel Service. We then sequence and analyze these samples and provide our customers with finished, research-ready genomic data, enabling them to focus exclusively on their single highest priority, discovery.

Our customers are not required to purchase expensive sequencing instruments and high-performance computing resources to sequence and analyze large sets of complete human genomes. Our outsourced service model enables our customers to offload to us the complex processes of sample preparation, sequencing, computing and data storage and management. We believe our services will expand the potential addressable market by enabling a broad base of researchers who may lack sufficient capital and the specialized personnel necessary to build and operate a sequencing laboratory, or who have historically been constrained by the high total cost of sequencing, to conduct large-scale complete human genome studies.

Customer Benefits

We believe our end-to-end solution provides the following advantages to our customers:

High-Quality Data. Our technology delivers what we believe is the industry s highest quality complete human genome data.

Cost-Savings. Our customers are not required to purchase expensive sequencing instruments and high-performance computing resources or hire the necessary specialized personnel to sequence and analyze large sets of complete human genome data.

Speed at Scale. Our customers can often complete their large-scale projects more quickly by using our services than by using commercially available sequencing instruments.

Ease of Use. We believe our customers can avoid the difficulty and time-consuming process of purchasing and operating their own sequencing instruments and can outsource the entire process to us, from sample preparation to delivery of research-ready data.

Operational Flexibility. By outsourcing their large-scale complete human genome sequencing projects to us, our customers can free up the capacity of in-house instruments to run smaller or more targeted sequencing projects and applications.

Technological Flexibility. As DNA sequencing technology improves, our customers have available to them the latest technology that we have developed, and they avoid the risk of their expensive instruments becoming technologically obsolete.

Enables Customers to Focus on Discovery. Outsourcing offloads the operational burdens of managing large-scale genome sequencing projects and enables our customers to focus their resources on research, which can reduce the time to discovery.

Table of Contents

Customers and Applications

Customers

We have more than 40 past and current customers, including the following:

Academic Medical Center University of Amsterdam

Brigham & Women s Hospital

Broad Institute of MIT and Harvard

Children s Hospital of Philadelphia

Eli Lilly and Company

Erasmus Medical Centre in Rotterdam, the Netherlands
Flanders Institute for Biotechnology (VIB)

Genentech, Inc.

Hudson Alpha Institute for Biotechnology

Institute of Cancer Research United Kingdom

Institute of Molecular Medicine at the University of Texas Health Science Center at Houston
Institute for Systems Biology

Ontario Institute for Cancer Research

Pfizer Inc.

SAIC-Frederick, Inc., National Cancer Institute

University of North Carolina

University of Texas Southwestern Medical Center

Selected Customer Examples

SAIC-Frederick, Inc., National Cancer Institute Pediatric Cancer Study

Our project with SAIC-Frederick, Inc., the prime contractor for the National Cancer Institute s research and development facility in Frederick, Maryland, involves sequencing and analyzing 50 tumor-normal pairs, comprising 100 complete human genomes, over a six-month period, to identify patterns relating to the genesis of cancerous tumors in children. This study may potentially lead to improved diagnosis and treatment of pediatric cancers. This project forms part of the National Cancer Institute s Therapeutically Applicable Research to Generate Effective Treatments, or TARGET, Initiative. TARGET seeks to use genomic technologies to rapidly identify valid therapeutic targets in childhood cancers so that new, more effective treatments can be developed. It is currently focusing on five childhood cancers: acute lymphoblastic leukemia, acute myeloid leukemia, neuroblastoma, osteosarcoma and Wilms tumor. Our contract with SAIC-Frederick contains an option for SAIC-Frederick to engage us to sequence 564 additional cancer cases, comprising an additional 1,128 complete human genomes, over an additional 18-month period.

Institute for Systems Biology Miller Syndrome Study

Our project with Dr. Leroy Hood of the ISB involved sequencing the complete genomes of a four-member nuclear family, including two healthy parents and their two children who suffer from two genetic disorders: Miller Syndrome and primary ciliary dyskinesia. The data we provided allowed ISB researchers to pinpoint the causal gene and subsequently confirm that gene s role in Miller Syndrome, a disease in which the genetic basis had previously evaded detection. The results were published in *Science Express* in March 2010 and have led to two follow-on projects with the ISB to sequence an additional 122 and 615 genomes respectively.

Genentech, Inc. Non-Small Cell Lung Cancer Study

Our project with Genentech, Inc. (a member of the Roche Group) compared the complete human genome sequences of a primary non-small cell lung tumor with nearby non-tumorous tissue taken from the lung of a long-term smoker. This project was the first complete human genome sequence of a primary non-small cell lung tumor

Table of Contents

and matched normal tissue. Comparison of these sequences revealed both known (KRAS G12C) and novel mutations in numerous oncogenes and led to the discovery of numerous somatic mutations. The data we delivered allowed Genentech to measure the rate of smoking-induced mutations accumulated over time and resulted in a publication in *Nature* in May 2010.

University of Texas Southwestern Medical Center Hypercholesterolemia Study

Our project with Dr. Jonathan Cohen of University of Texas Southwestern involved sequencing the complete genome of an 11 month old breast-fed girl with cholesterol-laden deposits and very high blood cholesterol levels. Doctors ruled out a diagnosis of a disease called sitosterolemia, based on certain blood test results. However, sequencing of the girl s genome revealed a mutation in a relevant gene. This finding indicated that the infant definitely has sitosterolemia, despite the prior contradictory test results. The major finding of this study is that complete genome sequencing identified the culprit mutations and provided a definitive diagnosis in a patient who did not have the classical hallmark features of the disease in question. This finding demonstrates that complete genome sequencing can be a valuable aid to diagnose and treat genetic diseases, even in individual patients. These results were published by the *Oxford University Press* in August 2010.

Applications

Potential applications for our complete human genome sequencing service include:

Cancer Research. Researchers are sequencing cancer genomes and comparing them to normal genomes, which are referred to as tumor-normal pairs, to identify the mutations in cancer genomes. We believe understanding these mutations will guide development of new cancer therapeutics and diagnostics and ultimately enable doctors to select the best course of therapy based on the specific mutations found in a tumor.

Mendelian Disease Research. There are thousands of Mendelian diseases, or diseases that have been found to run in families, and are accordingly likely to have a significant genetic component. However, the genetic cause of most of these diseases is currently unknown. By sequencing the complete genomes of the affected families, we believe the genetic causes of these Mendelian diseases can be discovered, which could lead to the development of novel diagnostics and therapeutics.

Rare Variant Disease Research. Diseases such as central nervous system disorders, cardiac disease and certain metabolic disorders that appear broadly in the population are thought to be caused by rare variants. Large-scale studies of affected individuals may help to identify the disrupted pathways and lead to the development of novel diagnostics and therapeutics.

Clinical Trial Optimization. We believe that selecting or stratifying patients on the basis of their genetic profiles could enable the preferential admission of high responders into a clinical trial. This stratification could enable the trial to reach its conclusion with fewer patients and lower costs and result in faster clinical trials and drug commercialization.

In addition to these research studies, we expect future clinical applications to include:

Companion Diagnostics. We believe that therapeutics that are not first-line treatments for the general population may be elevated to first-line treatments or used in combination therapies for subsets of the population that share a common genetic profile. Complete human genome studies may unlock new market opportunities for these therapies or combination therapies.

Cancer Pathology. We believe that analyzing complex cancer genomes that involve large and unpredictable structural changes will be most reliably and economically implemented using complete human genome sequencing. According to the National Cancer Institute SEER Cancer Statistics, there are approximately 1.5 million new cases of cancer diagnosed each year in the United States.

9

Table of Contents

Universal Diagnostics. As medical records technology and public health policy advance, we believe that large numbers of people will have their complete human genomes sequenced and stored in their electronic medical records for use by their physicians in managing their health care decisions.

Our Strategy

We intend to become the leading complete human genome sequencing and analysis company and the preferred platform for human genome discovery by:

Continuing to Deliver the Highest Quality Genomic Data and Analysis at a Low Total Cost. By continuing to deliver the highest quality research-ready data and by enabling our customers to avoid the cost, complexity and risks associated with purchasing and operating the instruments and computing resources required to undertake complete human genome sequencing, our goal is to become the preferred solution for our customers.

Maintaining and Strengthening our Technology. We plan to continue to conduct research and product development activities to further improve quality, reduce costs, increase throughput and reduce our turnaround time. We plan to further develop the biochemistry, informatics, instrumentation and software that we believe together make up the industry s most robust solution. We will also seek to continually improve our operational processes and analysis software.

Capitalizing on our Scalable Model. Due to the highly scalable nature of our service model, we believe we are well positioned to serve customers looking to sequence a small number of genomes as well as customers who are looking to rapidly sequence a very large number of genomes.

Establishing Ourselves as the Leader in Outsourced Complete Human Genome Sequencing. We intend to continue to focus exclusively on complete human genome sequencing. We believe that this focus will put us in a strong position to become the preferred platform for complete human genome sequencing.

Expanding Globally to Increase Capacity and Reach New Markets. We expect to enter into partnership agreements with domestic and international organizations to build additional genome sequencing centers around the world. These genome sequencing centers will increase our sequencing capacity, provide us with improved access to global markets and expand our revenue opportunities.

Exploring Strategic Partnerships and Collaborations. We expect to explore opportunities for strategic partnerships and collaborations with commercial and research organizations to leverage our genome sequencing technology with the strengths of these organizations to further develop and expand the applications for our sequencing technology.

Expanding Applications for the Use of our Technology. While our current focus is on providing complete human genome solutions primarily to academic and biopharmaceutical researchers, we believe that as we sequence and deliver more complete human genomes to our customers, our growing understanding of the genetic basis of human disease may lead to future applications in areas such as cancer pathology.

Our Human Genome Sequencing Platform Technology

Our proprietary human genome sequencing platform consists of three major technologies: our proprietary human genome sequencing technology, our high-throughput process automation technology and our complete data management solution.

Proprietary Sequencing Technology

There are two primary components of our proprietary human genome sequencing technology: DNB arrays and cPAL reads.

Table of Contents

DNB Arrays

We have developed a novel approach to preparing fragmented DNA for reading on our sequencing instruments. Using a biochemical process for copying DNA, we reproduce each DNA fragment in a manner that connects all of the copies together in a head-to-tail configuration, forming a long single molecule of connected nucleotides. We have developed proprietary techniques for causing each long single molecule to consolidate, or ball up, into a small particle of DNA that we call a DNB. The DNBs are approximately 200-300 nanometers in average diameter. Each DNB contains hundreds of copies of the 70 bases of DNA we are seeking to read in each fragment.

The small size and biochemical characteristics of our DNBs enable us to pack them together very tightly on a silicon chip. We use established photolithography processes developed in the semiconductor industry to create a silicon chip that has a grid pattern of small spots. The small spots are approximately 300 nanometers in diameter, and the center of each spot is separated by approximately 700 nanometers from neighboring spots. Each silicon chip has approximately 2.8 billion spots in an area 25 millimeters wide and 75 millimeters long. We have developed a proprietary process that causes the DNA to adhere to these spots, which we refer to as sticky spots, while conversely preventing the DNA from adhering to the area between the sticky spots. When a solution of DNBs is spread across the chip, the DNBs adhere to the sticky spots, with one DNB per spot. We have also developed proprietary techniques to fill over 90% of the sticky spots with exactly one DNB. We refer to the silicon chip filled with DNA as a DNA nanoball array. Each finished DNA nanoball array contains up to 180 billion bases of genomic DNA prepared for imaging.

cPAL Read

To read the sequence of nucleotides in each DNB, we have developed a highly accurate proprietary ligase-based DNA reading technology called cPAL. Our cPAL technology uses the naturally occurring ligase enzyme, which accurately distinguishes between the A, C, T and G nucleotides, to attach fluorescent molecules that light up with a different color for each of the four nucleotides. By imaging the color lights of a DNB array and decoding the color images, we can determine the sequence of nucleotides in each DNB. A key characteristic of our cPAL technology is its high accuracy of reading very short five-base sequences of DNA. We have developed a proprietary technique for preparing the DNA fragments so that we can read seven five-base segments from each of the two ends of the DNA fragment for a total of 70 bases from each fragment. We have also developed proprietary software that generally reconstructs over 90% of the complete human genomes from these 70 base reads from each fragment.

Advantages of our DNB arrays and our cPAL technology over other commercially available DNA sequencing technologies include:

High Accuracy. Our cPAL technology has very high single-read accuracy due to the intrinsic nature (high accuracy) of the ligase enzyme. By reading each nucleotide multiple times, we achieve a consensus error rate equal to approximately 1 error in 100,000 nucleotides.

No Accumulation of Errors. Many other DNA sequencing methods employ sequential processes that cause errors to accumulate as each successive nucleotide is read, which results in a higher potential error rate for each successive nucleotide. Our cPAL technology reads each nucleotide independently, and as a result there is no accumulation of errors, which enables us to read successive bases without increasing our error rate.

Low Reagent Cost. Our cPAL technology uses low concentrations of low-cost commodity reagents. Our DNB arrays achieve a very high density of DNA on each array, which reduces the quantity, or volume, of reagents we use compared to other DNA array approaches. The combination of low concentration of low-cost reagents and smaller quantities results in lower reagent costs compared to other commercially available DNA sequencing methods.

Table of Contents

High Throughput Process Automation

There are five major components of our high-throughput process automation technology: high-throughput sample preparation, high-throughput sequencing instruments, high-performance computing infrastructure, workflow automation software and service delivery technology.

High-Throughput Sample Preparation

Our high-throughput sample preparation technology consists of step-by-step protocols for preparing DNA for sequencing and pipetting robots that automatically execute these protocols. We prepare genome samples in batches of 88 and load the samples into a 96-well plate (the other eight wells in the plate contain known, or reference, DNA that we use to monitor the quality of the sample preparation process). A sample preparation run processes four 88-sample plates for a total of 352 genomes per run. We are building the instruments and staffing capacity to perform two runs in parallel for a total of 704 genomes prepared for sequencing. The result of a sample preparation run is up to 352 genomes loaded onto flow slides, ready to be loaded on sequencing instruments. Our sample preparation capacity can be scaled by adding additional sample preparation instruments and staff as needed.

High-Throughput Sequencing Instruments

Our sequencing instruments consist of a fluidics robot that pipettes multiple types of chemical reagents (including fluorescent molecules) onto the flow slides and an imaging system that records images of the fluorescent molecules attached to the DNA. Each sequencing instrument processes 18 flow slides at a time. The 18 flow slides are robotically moved back-and-forth from the fluidics robot to the imaging system. While one flow slide is being imaged, the other 17 flow slides are prepared with reagents or waiting for the imager to become available. A sequencing run takes approximately 11 days. Currently, our sequencing instruments can generate between 70 and 100 gigabases of usable data from each flow slide in an 11-day run. To sequence a complete human genome at an average redundancy of 40 times requires 120 gigabases of usable data. We expect to make continued enhancements in our technology to further increase the amount of usable data we get from each flow slide.

High-Performance Computing Infrastructure

We have built a genomic data processing facility that consists of approximately 7,500 core processors and 2,250 terabytes (a terabyte is one thousand gigabytes) of high-speed disk storage. Our sequencing instruments are connected to our data center by a network connection that transfers data at a rate of 30 gigabits per second. Our data center has the capacity to perform all of the required computation for several hundred genomes per month. We plan to expand our data center as needed, and we expect to make continued enhancements to our software to further increase the efficiency of our data center.

Workflow Automation Software

Our workflow automation software tracks each sample from arrival at our facility to delivery of research-ready data to the customer. Sample tracking is accomplished through bar codes. Each 96-well plate of samples has a bar code, and each flow slide has a bar code. The instruments that process plates and flow slides have bar code readers attached to them. User interfaces to our workflow automation software allow us to track the progress of each sample throughout sample preparation, sequencing and computing. We are also developing a web-based customer portal to enable customers to track their projects real-time throughout the sequencing process.

Service Delivery Technology

Our cloud-based data delivery system is based on our vendor relationship with Amazon Web Services, or AWS. We upload our customers finished genomic data to AWS, which AWS then copies to hard disks and ships the hard disks to our customers. Our customers also can pay AWS to store their data on an ongoing basis.

Table of Contents

Complete Data Management Solution

There are two major components of our complete data management solution: assembly software and analysis software.

Assembly Software

Assembly is the process of using computing methods to organize the overlapping 70-base nucleotide sequences to reconstruct the complete genome. We have developed a proprietary approach to assembly that uses a combination of advanced data analysis algorithms and statistical modeling techniques to reconstruct over 90% of the complete human genome from approximately two billion 70-base reads. We have designed our assembly software to run in parallel across our large network of Linux computers.

As reported in *Science* published in January 2010, we generated high-quality base calls in as much as 95% of the genome, identifying between 3.2 million and 4.5 million sequence variants per genome processed. Detailed validation of one genome dataset demonstrated a consensus error rate of approximately 1 error in 100,000 nucleotides.

Analysis Software

After assembling the genomic data, we use our analysis software to identify key variants in each genome and automatically annotate the genomic data. We have developed a suite of open source analytical tools, called CGA™ Tools, designed to enable our customers to rapidly analyze the data we generate from their samples. For example, we offer a tool facilitating the comparison of two genomes, enabling the quick determination of where the genomes differ. We are also developing additional analytical tools, such as a tumor-normal comparison tool designed to allow cancer researchers to compare a cancer genome to the normal genome from which it was derived, a family analysis tool designed to enable researchers to compare parental genomes with the genomes of their children and a large-scale genome browser designed to allow researchers to compare the hundreds of genomes sequenced in a large-scale study.

Technology Strategy

We plan to continue to advance our complete human genome sequencing and analysis technology in four major areas:

Array Density. Our unique grid patterned arrays currently consist of a 700 nanometer grid. We may reduce the grid size to 250 nanometers and correspondingly reduce the diameter of the sticky spots and DNA nanoballs. If successful, this improvement will increase the density of the DNA on an array by a factor of eight, which will decrease the reagent cost of sequencing a given amount of DNA by a factor of eight.

Instrument Speed. Our unique grid patterned arrays enable us to align the grid pattern of the DNA on the array with the grid pattern of the pixels in the detector, allowing us to image our arrays with very short exposure times. We may increase the speed of our instruments by acquiring and deploying new cameras that take images and transfer data at approximately six times the speed of our existing cameras. We may also increase the number of cameras per sequencing instrument from two to four.

Process Automation. As our instruments get faster, we intend to improve our sample preparation, process automation and data management technologies to process and deliver an increasing number of genomes to our customers with reduced turnaround time.

Analytic Software. We continue to improve and extend our analytic capabilities through the development of software designed to decrease genome assembly time and address specific application requirements of our customers. For instance, in December 2010, we started providing copy number variation and structural variation results to our customers.

Table of Contents

As we implement a combination of these technology enhancements, our goal is to be the first company to sequence and analyze high-quality complete human genomes, at scale, for a total cost of under $1,000 per genome.

Sales, Marketing and Customer Support

We sell our complete human genome sequencing service through our direct field sales and support organizations. Our sales process with each new customer typically involves undertaking a small project, or pilot program, which enables the customer to become familiar with our outsourced solution and research-ready data. We then work with our customers to expand the relationship to larger projects.

Sales

We have assembled a highly experienced and technically qualified field sales team, many of whom hold a Ph.D. or other advanced degree in a relevant scientific field. Each of these sales managers brings a network of extensive contacts in our targeted customer segments. The sales group develops business opportunities and obtains orders for our complete human genome sequencing service by proactively identifying, qualifying and visiting well-funded prospects at major companies, institutions and universities.

Marketing

Our marketing group has developed and maintains the Complete Genomics brand, increases market awareness and generates demand for our solution through a variety of methods. First, we have created and continue to maintain a clear media presence via our website, press releases, interviews and articles that reinforce our market presence and scientific credibility. Second, we generate demand by promoting the company via marketing programs and by attending and exhibiting at relevant tradeshows and conferences. Third, we continue to evolve our marketing strategy by tracking market trends, understanding customer needs and developing appropriate products and programs. The marketing group also fulfills traditional product management requirements, such as defining our service and application strategy and roadmap, including partnering strategies, and developing sales tools, training materials and competitive analyses for the sales group.

Customer Support

We are committed to supporting our customers through a network of scientific applications staff based in both Mountain View and locations near our most concentrated customer bases. This team currently consists mostly of Ph.D.-level scientists with extensive bioinformatics experience. Our scientific applications team works with customers to address technical questions related to our service offering and provide detailed training and support.

Most of the training and support efforts are focused on helping customers understand and use the large amounts of data that are delivered as part of multi-human genome sequencing projects. We supplement these efforts with a team of Mountain View-based bioinformatics support specialists.

Research and Development

Our research and development team brings together a variety of technical disciplines required for the development of a high-throughput sequencing system for commercial human genome sequencing services and includes DNA engineers, biochemists, molecular biologists, chemists, mathematicians, statisticians and electrical, mechanical, optical and software engineers. As of December 31, 2010, we had approximately 71 employees engaged in research and development, many with Ph.D.s. These professionals apply their skills in disciplines including:

- biochemistry (sample preparation, DNA array preparation, DNA sequencing assay);
- hardware (optics, fluidics, mechanical design, flow slides);

Table of Contents

software (algorithms, instrument software, genome sequencing software, bioinformatics);

information technology (high-performance data center management);

semiconductors (mask design, surface chemistry); and

process automation.

Our research and development teams are engaged in developing new applications for our technologies, including:

Cancer Genomes. We are developing new methods for sequencing the complex structural variations found in cancer genomes, such as duplications, deletions and translocations in tumor genomes. We believe these methods will enable the research community to better understand the genetic basis for cancer.

Diploid Sequencing. We have invented and are developing a method for independently sequencing the maternal and paternal chromosomes. We believe this independent chromosome sequencing will be required for many molecular diagnostics, because multiple variants within a gene may or may not affect both copies of the gene.

Human Methylomes. We are researching possible methods of sequencing the human methylome, which we believe will be important in understanding cancer genomes.

Human Transcriptomes. We are researching possible methods of sequencing the human transcriptome, which, when combined with complete human genomes, we believe will shed light on the genetic basis of human disease and drug response.

In the years ended December 31, 2010, 2009 and 2008, we spent $21.7 million, $22.4 million and $23.6 million, respectively, on company-sponsored research and development activities.

Intellectual Property

Our success depends in part upon our ability to obtain and maintain intellectual property rights with respect to our products, technology and know-how, to prevent others from infringing these intellectual property rights and to operate without infringing the proprietary rights of others. Our policy is to seek to protect our proprietary position by, among other methods, filing patent applications related to our proprietary technology, inventions and improvements that are important to the development and conduct of our business. We also rely on trade secrets and know-how to develop and maintain our proprietary position.

Our patent strategy is to seek broad patent protection on new developments in genome sequencing technology, and also file patent applications covering new implementations of our technology. Additionally, we file new patent applications directed at equipment and software that are used in conjunction with our genome sequencing technology.

Our core genome sequencing technology originated at Callida Genomics, Inc., or Callida, in the laboratory of Radoje (Rade) Drmanac, Ph.D., our Chief Scientific Officer and one of our co-founders. Dr. Drmanac played an important role in high-throughput sequencing of whole genomes using ligation-based sequencing reactions performed on microarrays. In March 2006, we entered into a license agreement with Callida pursuant to which we exclusively licensed from Callida the relevant patent filings relating to the use of the technology in random arrays, or arrays of genomic DNA fragments wherein the position of any specific fragment on the array is not predetermined, and probe anchor ligation, which we utilize in our commercial sequencing technology. Under this license agreement, we also obtained a nonexclusive license under additional patent filings owned by Callida that permits us to use the random array technology without infringing such additional patents. In exchange for the licenses, we issued to Callida 13,333 shares of our common stock, paid $1.0 million in cash for repayment of certain promissory notes held by Callida and agreed to pay $250,000 each year until the earlier of (a) March 28,

15

Table of Contents

2012 and (b) such time as our common stock is traded on a national exchange and the shares issued to Callida are