Telephone (312) 236-1600 *Fax (312) 236-0801*

William Blair & Company
Limited Liability Company

222 West Adams Street Chicago, Illinois 60606

PE CORPORATION – CELERA GENOMICS GROUP

(CRA)

December 9, 1999 (99-146)
Basic Report

Winton Gibbons (312) 364-8371 wgg@wmblair.com
Adam Chazan (312) 364-8418 amc@wmblair.com

Price: $72 1/8 ($14 3/16–$79)
Fiscal Year Ends: June

Fiscal Year	Earnings Per Share	Revenue ($ millions)	Revenue Growth
1999	$(1.58)	$12.5	198%
2000E	$(3.71)	$38.0	203%
2001E	$(4.50)	$65.2	72%
2002E	$(4.04)	$123.8	90%
Calendar 1999E	$(2.74)	$23.9	200%
Calendar 2000E	$(4.35)	$47.0	97%
Calendar 2001E	$(4.39)	$88.7	88%

Earnings Per Share Growth		Return on Average Equity	
1997-1999	NM	1999	NM
Long-term	33%	2001E	NM

Book Value Per Share (June 1999):	$11.42	Dividend:	None
Insider Ownership:	5%	Common Shares:	25.1 million
Sales (Fiscal 1999):	$12 million	Market Value:	$1.9 billion

Investment Opinion: Speculative Strong Buy

We view PE Corporation's Celera Genomics Group as the future leading provider of genomics knowledge. Genomics is the sequencing and deciphering of genetic information. In our opinion, this information will be critical for human drug discovery and development, as well as for agriculture, basic life science research, and forensics. Currently, we estimate this market to be $715 million, growing 24% compounded annually. The ultimate market potential is likely to be vast, more than $5 billion to $10 billion, with the market size, growth, and characteristics difficult to predict even five years from now. We believe that the company has fundamentally changed the gene sequencing and discovery process by moving to a bottom-up, whole-genome shotgun sequencing approach from the top-down, directed sequencing approach used by public institutions and competitors. It has accomplished this by creating an unparalleled gene sequencing "factory," coupled with the most powerful private supercomputer in the world and many of the world's experts in genomics and bioinformatics. This leapfrog approach into a revolutionary market has many inherent risks, but we believe this is the best approach to capture the potentially substantial rewards. Consequently, we recommend that investors who can accept the risks buy the stock of this exciting company.

Celera—Speculative Strong Buy

In our opinion, Celera will become the leading supplier of knowledge regarding genomics and molecular biology, a revolutionary market that should experience exponential growth. We believe that the company has changed the rules in this nascent market by 1) building a gene sequencing factory with more monthly throughput than the cumulative effort of all government-funded projects; 2) constructing the world's largest, nongovernment parallel processing supercomputer to capture and manipulate the data generated; and 3) employing an unprecedented group of world-class and prolific scientists and software engineers.

Summary Income Statement
($ in millions)

Fiscal years ends June 30	1999	% Revenue	2000E	% Revenue	2001E	% Revenue	2002E	% Revenue
Revenues	$12.5	100%	**$38.0**	**100%**	**$65.2**	**100%**	**$123.8**	**100%**
R&D	48.4	386%	**151.7**	**400%**	**199.3**	**306%**	**242.2**	**196%**
SG&A	27.2	217%	**44.1**	**116%**	**54.6**	**84%**	**57.4**	**46%**
Total Operating Expense	$75.6	603%	**$195.8**	**516%**	**$253.9**	**389%**	**$299.6**	**242%**
Operating Income	(63.1)	-503%	**(157.8)**	**-416%**	**(188.7)**	**-289%**	**(175.8)**	**-142%**
Other Income (expense)	1.2	10%	**7.5**	**20%**	**2.6**	**4%**	**5.4**	**4%**
Earnings Before Taxes	($61.9)	-493%	**($150.3)**	**-396%**	**($186.1)**	**-285%**	**($170.4)**	**-138%**
Benefit from Income Taxes (% of EBT)	22.3	-36%	**54.1**	**-36%**	**67.014**	**-36%**	**61.43**	**-36%**
Net Income	($39.6)	-315%	**($96.2)**	**-253%**	**($119.1)**	**-183%**	**($109.0)**	**-88%**
EBITDA	($59.4)		**($137.4)**		**($157.0)**		**($137.0)**	
EPS	($1.58)		**($3.71)**		**($4.50)**		**($4.04)**	
Shares Outstanding	25,100		**25,927**		**26,449**		**26,982**	

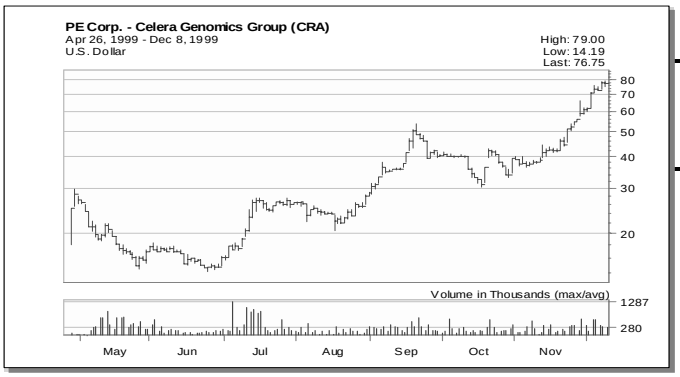
Year-over-year Growth	**1999**	**2000E**	**2001E**	**2002E**
Revenue	366.3%	**202.7%**	**71.8%**	**89.8%**

Summary of Balance Sheet
($ in millions)

	1999	2000E	2001E	2002E
Cash and Equivalents	$71.5	**$135.7**	**$22.2**	**$119.3**
Working Capital	$299.4	**$203.1**	**$85.9**	**$171.8**
Shareholders Equity	$293.9	**$197.7**	**$78.5**	**$1,118.8**

Summary of Cash Flows
($ in millions)

	1998	1999E	2000E	2001E
Net Cash Provided by Operations	($23)	**($56)**	**($80)**	**($62)**
Net Cash Used in Investing	($96)	**($29)**	**($34)**	**($38)**
Net Cash Provided by Financing	$190	**$150**	**($50)**	**$197**
Net Cash Increase (decrease)	$71	**$64**	**($163)**	**$97**



Quarterly EPS

	1999	2000E	2001E	2002E
1Q	($0.15)	($0.75)	($1.14)	
2Q	($0.32)	($0.88)	($1.14)	
3Q	($0.47)	($0.97)	($1.13)	
4Q	($0.63)	($1.10)	($1.10)	
Year	($1.58)	($3.71)	($4.50)	($4.04)
Calendar	($2.74)	($4.35)	($4.39)	

Source: Company financials; FactSet; William Blair & Company, L.L.C. estimates

Contents

Summary of Investment Recommendation: Speculative Strong Buy

We believe that Celera Genomics is well positioned to become the leading supplier of genomic information to the pharmaceutical, biotechnology, and agribusiness communities, serving a market that we currently estimate to be $715 million and is conservatively expected to grow more than 24% compounded annually for the foreseeable future. Genomics is a scientific discipline marked by the sequencing and deciphering of genetic information, with the purpose of applying it to understand basic biology, elucidate the mechanisms of disease, derive more-effective therapeutics less expensively and more quickly, and improve the quality and yield of agricultural products. Celera has garnered a great deal of attention as it aims to sequence completely the human genome at least two years ahead of the international government/academic consortium effort. This massive and controversial undertaking is just the *beginning* of the company's efforts to improve understanding of our biological surroundings.

To achieve these ends Celera has built an unparalleled sequencing facility along with the world's most-powerful private computing facility; codified the intellectual capital of some of the world's leading genomic research facilities, namely The Institute for Genomic Research (TIGR); and has leveraged its relationship with its sister company, PE Corporation's PE Biosystems Group. We believe that Celera is faced with remarkable growth opportunities to which it may apply its immense infrastructure and present and yet-to-be-acquired knowledge, leading to revenue growth of 100% compounded annually.

Celera arose as a tracking stock out of a reorganization and recapitalization of Perkin-Elmer Corporation in April 1999, which we discuss in more detail in appendix A. This process essentially divided the corporation into three pieces: the PE Corporation, with its two distinct tracking stocks, Celera Genomics (CRA) and PE Biosystems (PEB), a life science instrumentation firm, and its declining analytical instrument business of the former Perkin-Elmer. PE Corporation sold this last business to EG&G, which officially renamed itself PerkinElmer in October 1999, with the ticker symbol PKI. PE Corporation itself now maintains no listed stock, only the tracking stocks of CRA and PEB.

Our investment recommendation for Celera is based on the following five key factors.

1) We believe that the revolutionary market for molecular biological knowledge databases and analysis is particularly attractive.

The current state of the world is placing severe demands on a number of the earth's resources. As the population bulges and people come to expect increasing improvements in the quality of life globally, stresses are being placed on the pharmaceutical industry to deliver more-effective therapeutics at rates and costs that satisfy shareholders; on agribusiness to improve products and processes to feed the world; and on law enforcement to ensure safety from and dissuade crime. The science of genomics offers solutions to these problems, yet before the promise of this technology is fully realized, a basic foundation must be laid to enable understanding and facilitate targeted research. The mass sequencing, annotation, and distribution of genomic information provides this foundation and represents an attractive business opportunity. Currently, we estimate that the market for genomic information is $715 million globally, growing more than 24% compounded annually. Customers for this information, which may be provided on a fee-for-service or subscription basis, include multinational pharmaceutical companies, biotechnology companies, and agribusinesses.

2) In our opinion, Celera has established unparalleled capabilities to create, analyze, and distribute this knowledge.

Although the goals of Celera may seem daunting, we believe the company has assembled the critical pieces that will ensure success. The company has built an industrial scale gene-sequencing factory around the successful technology of PE Biosystems' ABI Prism 3700. The capacity of this facility dwarfs that of any competitor, private or public, with monthly output that roughly equals that of all the genetic information generated and compiled in the public record (GenBank) since 1982. Celera has teamed with Compaq Computer to build the world's most-powerful private supercomputer to facilitate the compilation, manipulation, and analysis of the vast amounts of data that will be generated. Lastly, the company has assembled an incredible team of talent, including Nobel Prize winners, who meld molecular biology, computer science, and information technology to accomplish the goal of developing *the world's leading* resource for genomic information.

3) As the company continues to develop its data and applications, it should create remarkable growth opportunities.

In addition to Celera's base business of offering subscription access to its growing offering of value-added, annotated genomic sequence databases, the company may be faced with other paths for growth. The development of industrial-scale genomics may be applied to the next discipline in the chain of biology—proteomics, the study of proteins, which are the molecular agents of genes. The information technology and computational infrastructure might lead to opportunities in experimental planning and data collection in the model of an application service provider (ASP). There also exists the opportunity to capitalize on acquired intellectual property through licenses on both the genetic information and any information technology innovations that might arise. Lastly, Celera's Web-based distribution scheme lends itself to some form of business-to-business e-commerce, which in our opinion is the least lucrative opportunity that may be presented.

4) We expect the company to achieve five-year compounded annual revenue growth of 100%.

This growth should be driven by an increasing number of customers and wider acceptance of this information in existing and new applications. We expect the company's revenue to grow 145% compounded annually over the next few years, increasing to $124 million in fiscal 2002, from $12.5 million in fiscal 1999. This would represent a 9% share of the anticipated $1.4 billion market for these products in 2002. We expect operating expense as a percentage of revenue to decline steadily as revenues rapidly increase and cost-growth slows; research and development (R&D) likely will continue to increase as these expenditures translate directly into the company's information and service products. SG&A expense growth should rapidly decline as the sales and support groups reach critical mass.

5) Furthermore, we believe that a premium valuation is justified on the basis of this unprecedented opportunity.

The market that Celera hopes to address is potentially enormous, well into the billions of dollars. However, the company's near-term valuation entails some consideration. Genomics is a research and development based industry. Consequently, we believe that metrics related to the relative value of technology should help investors assess valuations for the cohort of genomics companies. For example, Celera is trading at a technology value (market capitalization minus cash) of about $1.9 billion, almost 12 times its projected R&D spend for fiscal 2000. This is more than the median ratio of 8 for other pure-play knowledge based firms performing gene sequencing or gene expression analysis. Nonetheless, it is about half that for knowledge-based firms that have a pharmaceutical component. In addition to its knowledge database and interface, Celera intends to provide additional *genomic services* to assist its customers in applying the knowledge, which we believe should elevate its valuation from that of the pure-play, knowledge-based cohort closer to that of the other cohort. The genomics field has the potential to revolutionize drug development, agriculture and other endeavors such as forensics. We believe that this is the vision for Celera, and consequently, the company should have a premium valuation.

Risks

As Celera was created essentially *de novo*—with substantial and costly capacity—to leapfrog existing sequencing and genomic knowledge efforts, it possesses a number of notable risks. *However, we believe that to capture the potentially enormous opportunities that this field represents requires this boldness and magnitude of action.*

Developmental-stage company. Celera has a limited number of customers and subsequent revenue. In addition, it is spending a considerable amount of money to build its genomics factory and supercomputing facility, as well as to populate and annotate its database with the actual sequencing data. The company is doing this in the hope that the market demand will grow as the knowledge-based products and applications are developed. We strongly believe that more-than-sufficient underlying need and demand exists for this type of information.

Ahead of its time. Similar to the previous issue, a question arises not over need, but over utility. *Can pharmaceutical firms effectively use genomic information that is growing exponentially?* If not, Celera may be building a product for which the market is not yet ready. However, this is a double-edged sword, as it also could lead to new services and applications by the company to address the potential deficiencies of its customers. Also, the potential issue is both one of how to glean useful information, as well as how to apply it practically in the drug development sense. Bottlenecks may arise in other technologies such as high-throughput-screening, other service providers such as contract research organizations, or within regulatory agencies such as the FDA who are unprepared for the potential flood of new drugs that may be developed.

Already lost the race. If one listens to the public pronouncements of Human Genome Sciences or Incyte, among others, the race to discover the human genome is over, and all valuable genes have been patented. While this position is open for debate for both whether most genes have been discovered, as well as the quality of the patent submissions to date (and is discussed further in the section entitled Exponential Growth Opportunities), we believe that gene-based intellectual property per se is not critical to Celera's business model and success. *More important to the company's knowledge-based business model is having a comprehensive, high-quality, and compatible database, in addition to the algorithms, processing capacity, and human expertise to turn the data into usable information and knowledge.* Out of this resource and capacity, Celera should be able to develop and sell knowledge-based applications independent of gene-based intellectual property. Nonetheless, on the basis of Celera's novel gene discoveries to date, in both Drosophila and human genomes, intellectual property could add a valuable income source for the company.

Behind the times. A corollary to the previous risk is that we already are in what some scientists have called the post-genomic world. The belief is that we know enough about genes and that the real action is with gene expression and proteins (see appendix B: Molecular Biology). Again, while there may be some truth to this possibility, gene expression and protein databases would benefit substantially from a comprehensive genome database; thus the company's current efforts would not be wasted. Furthermore, if this post-genomic view turns out to be correct (or we should say *when*), *we strongly believe that PE Corporation would establish either an expression or proteomics factory or establish links to a sufficient number of smaller sites to again change the rules of competition*. This new capacity and capability would benefit from the three existing components and subsequent knowledge of Celera.

Efficacy of technology platforms: shotgun sequencing, Compaq Alphas, Oracle databases, and Bioinformatics tools. Celera's approach is to change the rules of competition in this field. Consequently, it is pushing technology and techniques to the limits. For example, the whole-genome shotgun sequencing approach developed by Craig Venter, President of Celera and Chairman of TIGR, has proved itself on simple organisms such as

H. Influenzae; however, will it work on more-complex creatures, such as humans? To test this approach, the company first sequenced Drosophila, the fruit fly. While its genome is about one-twentieth the size of a human's, the fruit fly genome contains sections of DNA that are more difficult to interpret than human DNA, in that they contain many repeated and similar sequences. As it now appears that the company has successfully sequenced Drosophila, we are confident that it will be successful on the larger scale with the human genome. In addition to the sequencing technique, the company also is pushing the limits of information technology. The company will be generating more than 10 terabytes of finished data per year that must be collected into a database system, annotated, and subsequently processed to address customer queries. Much of the process will require processor-intensive pair-wise comparisons of sequence data. To address these extreme requirements, Celera has allied with Compaq and Oracle. Lastly, the company needs to be able to annotate and visualize the data in useful formats for its customers. The company has hired a large number of skilled software engineers and algorithm experts to do this.

Intellectual property: genes and databases. We foresee Celera generating intellectual property (IP) and trade secrets along at least three dimensions: gene and gene products (e.g., proteins); processes, techniques, and systems to identify these gene-related elements; and information technology-based algorithms, databases, and interfaces.

1) There is a great deal of controversy regarding the patenting of genes themselves, in addition to the gene products. In addition, many competitors claim to have already established priority on most genes, which we believe is both an empirical as well as patent law question. First, only completing the entire genome will show for what proportion of genes patents have been filed. Second, patent law and interpretation is not fully clear on priority: Can gene fragments be patented or does one need the full-length gene, or is a computational assignment of gene function sufficient or does one have to prove it experimentally? In any case, as previously discussed (and we also expound on this in the section entitled "Exponential Opportunity"), *we believe that the Celera model is predominantly knowledge-based with regard to the genes and gene products, rather than IP-based,* although this component appears as if it may add more value than the company originally believed when founded. Not only does this reduce the company's exposure to patent uncertainty, but the company could become a white knight for life science research by putting most (but not all) of its gene and gene product claims into the public domain.

2) The company also could generate considerable expertise, trade secrets, or intellectual property regarding the actual process by which it generates sequence data. Some of this may benefit sister company PE Biosystems as well. Legal protections for these type of inventions are more well defined, and we believe no more risky than those associated with any production or medical technology enterprise.

3) Lastly, the company intends to develop IT-based algorithms, databases, and interfaces that need to be protected. This also is an area where, for the most part, we believe that the company faces no more risk than other IT-based firms. However, the protection of the proprietary database is critical, and to some extent is ambiguous, especially in the United States. In Europe, there is *the Database Copyright Protection Act.* In the United States, there are legislative discussions to pass similar laws. In the meantime, query and analysis algorithms, computing capacity, and controlled access to the database should provide sufficient safeguards.

Business model and skills. We believe that the company has substantial scientific and software engineering talent. However, we believe that this embryonic industry and evolving business model require exceptional business talent. Furthermore, the appropriate business-oriented individuals need to be able to bridge the scientific concepts with the business model. We believe PE Corporation so far has been able to identify sufficient talent, including Craig Venter, the president.

Genomics backlash. There currently are discussions in Europe and Japan to label, and in some cases ban, GM (genetically modified) food or GMOs (genetically modified organisms). In addition, there are serious ethical considerations regarding the use of human genomics information, especially during assisted reproduction or for health insurance purposes. While the rational consideration of these issues should be of great importance to all of us, we believe that the possibility of a genomics backlash will have limited long-lasting effect for three reasons. First, addressing these issues requires substantial information, which likely will be provided by tools such as those provided by Celera. Furthermore, much of the genomics information used in agriculture is used for crossbreeding purposes, traditional crop management techniques, or research, rather than for the more controversial genetic engineer. Second, many environmentalists should become allies to genomics once they realize they would make tradeoffs—between destruction of rain forests for farmland, devastation of coral reefs due to overuse of pesticides and fertilizers, and the filling up of landfills with plastics that are not biodegradable—versus avoiding these catastrophes by leveraging genomics. Third, there are billions of people outside the highly developed countries that desperately need the food that this technology could provide. For example, the World Health Organization (WHO) is trying to establish a program to develop *golden rice*—rice that can be grown easily and that contains most or all of the basic nutrients needed. Regardless, research needs to be broadly and scientifically conducted with regard to both unintended consequences, as well as ethical considerations, and PE Corporation intends to spend millions of dollars through a foundation to promote awareness, discussion, and understanding.

Tracking stock. As discussed in appendix A, Celera shareholders are shareholders of PE Corporation with assets allocated to reflect its distinct business and performance. *Celera is not a separate legal entity*, and therefore its shareholders would be responsible for liabilities incurred by PE Biosystems (PEB), the other tracking stock of PE Corporation. *There is no separate stock for PE Corporation*. The tracking stock limits the rights of Celera shareholders, as both entities share a common board of directors and senior officers and Celera shareholders have voting rights proportional to its stock price relative to that of PE Biosystems. In addition, Celera stock may be converted to PE Biosystems stock at any time at the discretion of the board of directors, or vice versa. If this occurs, the proportional exchange rate will include a 10% premium of the average ratio for the preceding 20 days, if the event is not taxable, or no premium if taxable. Financing activities occur on the combined corporate level, which could benefit or hurt Celera stockholders, although currently we believe that it would lower Celera's cost of debt. *In addition, PE Biosystems is expected to generate sufficient net income so that Celera should receive tax benefits from all the losses it generates,* although the corporate reorganization originally limited this benefit to a total of $75 million. Despite these risks, we believe that the tracking-stock approach significantly benefits both sets of shareholders by aligning performance metrics and incentives with each entity's unique business model.

Prologue to the Revolution

A great deal of concern has been raised about the impending Y2K problem, although the potentially more farther-reaching "Y6B" passed on October 12, 1999, without much notice. On that day that the six billionth person was born, a milestone of tremendous population growth. Current estimates anticipate the Earth's population to reach 7.5 billion by 2020, and 8.9 billion in 2050, as shown in figure 1 and geographically in table 1. At this rate, the population is increasing roughly 80 million per year. To put this number in perspective, it would be like adding the populations of France, Greece, and Sweden each year, or the population of Philadelphia each week. Coupled with the population growth has been the remarkable expansion of life expectancy. World life expectancy has increased more than 20 years since 1950, to reach 65, and WHO conservatively expects this to increase to 75 by 2050. Combined, this growth and expansion of life expectancy will lead to exponential global aging. For instance, the number of people over 80 years old should increase from

approximately 66 million in 1999, to at least 370 million in 2050. Lastly, the earth's population also should become even more urbanized as the population dwelling in cities grows nearly 2 times, to more than 2 billion in the next 25 years.

Table 1
PE Corporation - Celera Genomics Group
Population Distribution of the World

Population (in billions)

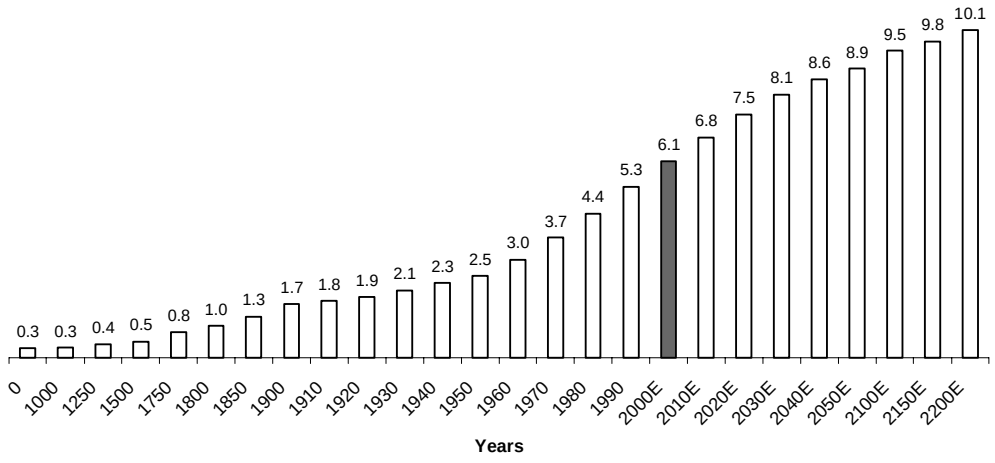
	Years							
Major area	1750	1800	1850	1900	1950	1999	2050E	2150E
Africa	0.11	0.11	0.11	0.13	0.22	0.77	1.77	2.31
Asia	0.50	0.64	0.81	0.95	1.40	3.63	5.27	5.56
Europe	0.16	0.20	0.28	0.41	0.55	0.73	0.63	0.52
Latin America	0.02	0.02	0.04	0.07	0.17	0.51	0.81	0.91
North America	0.00	0.01	0.03	0.08	0.17	0.31	0.39	0.40
Oceania	0.00	0.00	0.00	0.01	0.01	0.03	0.05	0.05
World	0.79	0.98	1.26	1.65	2.52	5.98	8.91	9.75

Percentage Distribution

	Years							
Major area	1750	1800	1850	1900	1950	1999	2050E	2150E
Africa	13.4%	10.9%	8.8%	8.1%	8.8%	12.8%	19.8%	23.7%
Asia	63.5%	64.9%	64.1%	57.4%	55.6%	60.8%	59.1%	57.1%
Europe	20.6%	20.8%	21.9%	24.7%	21.7%	12.2%	7.0%	5.3%
Latin America	2.0%	2.5%	3.0%	4.5%	6.6%	8.5%	9.1%	9.4%
North America	0.3%	0.7%	2.1%	5.0%	6.8%	5.1%	4.4%	4.1%
Oceania	0.3%	0.2%	0.2%	0.4%	0.5%	0.5%	0.5%	0.5%
World	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Source: United Nations; William Blair & Company, L.L.C. estimates

Figure 1
PE Corporation - Celera Genomics Group
Estimated World Population Growth
(in billions)



Source: United Nations Population Division; William Blair & Company, L.L.C. estimates

These substantial growth numbers lead us to raise some basic questions in three areas: pharmaceuticals, farming/agribusiness, and forensics.

Pharmaceuticals

Tremendous advances in medicine in recent history facilitated the expansion of life expectancy; yet our past successes are double edged. Overuse of antibiotics has led to the rise of microbes resistant to medicine's once effective and abundant tools. The incidence of penicillin-resistant Streptococcus pneumonia, which causes meningitis, pneumonia, and

ear infections, has increased in the United States from 5% in 1992, to 25% currently, as reported by the Centers for Disease Control. As the population ages, the diseases and problems that medical science are asked to solve become more complex (for example, cancer or Alzheimer's disease). In addition, as wealth accumulates people likely will demand higher standards of health care, placing even greater demands on the current system. Excess wealth could be diverted to cure previously tolerated ills. The success of Viagra is just one example. Detailed understanding of the etiology or progression of a disease is required before therapies and diagnostics should be contemplated. At what level must discoveries be made to offer the depth of understanding that will allow medicine and the pharmaceutical industry to progress further?

Farming/Agribusiness

Can current levels and methods of agriculture support the world's growing nutritional needs? Although there appears to be plenty of food in the developed countries, an estimated 840 million people worldwide were undernourished in 1998. The food gap—the difference in the amount of food produced and the amount demanded—is expected to double from 94 million tons in 1998, to 228 million tons in 2025. The people in greatest need often are not able to grow the food locally, nor can they afford to have it imported. Furthermore, overuse of fertilizers, pesticides, and herbicides can have a deleterious effect on the rest of the environment. What can be done in the agriculture industry to solve this problem?

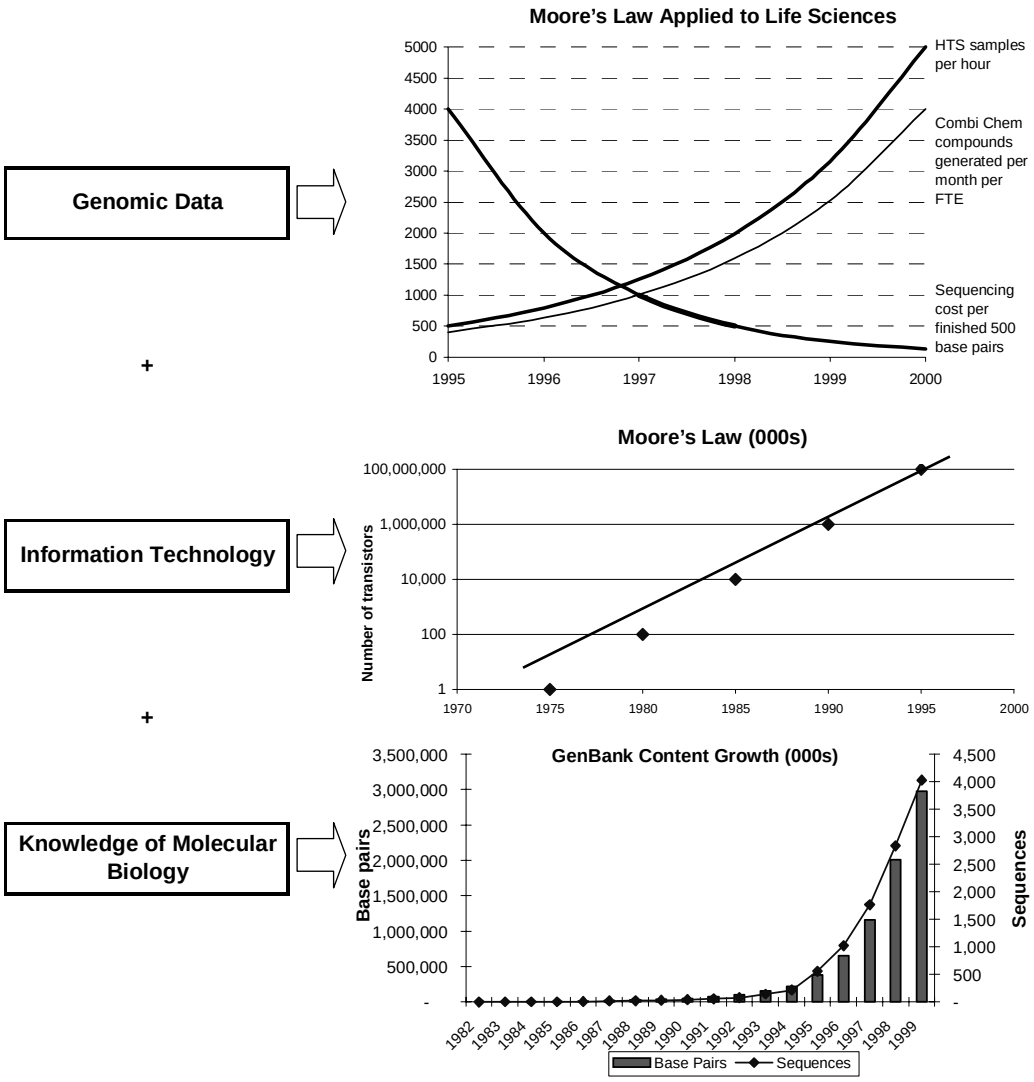
Forensics

This long population boom will result in continued urbanization of the earth. Cities will become more crowded. Studies have correlated the size of a city with increasing levels of crime. In the face of the world's swelling cities, what kinds of tools can be offered to law enforcement that will be effective in solving crime, while dissuading potential criminals by removing the city's veil of anonymity?

Understanding and applying genetic information, a discipline called genomics, as well as other aspects of molecular biology provides the foundation to solve these and other problems. To say the answer lies within our genes, while clichéd, may be one of the grossest understatements of our time. However, why is this revolution only occurring now? The foundation of genetics dates as far back as the work done by Mendel with pea plants in the late 1800s, and scientists have been applying principles of this discipline for many years. As figure 2 shows, advances in three areas have enabled the genomics revolution: life science instrumentation, information technology, and networking. Life science research efforts were long hindered by expensive, labor-intensive processes. Incremental discoveries, let alone major breakthroughs, could take years because of these hurdles. The advancement on the analytical instrument technology and the application of automation are allowing scientists to undertake projects that were previously thought unimaginable due to cost and complexity. Inherent in the complexity of these large-scale projects was the sheer amount of data generated. Life scientists were, and still are, faced with the limitations of information and computing technology. As computing has become less expensive and more accessible, researchers have sought to use these tools to advance their work. As an example, the National Institutes of Health's working group on biomedical computing has reported that biological researchers now spend 95% of their time on average in front of a computer. The National Science Foundation reports that 12% of investigators using the nation's supercomputing centers are biologists, accounting for 25% of all computing cycles, representing an increase of 54% from 1997 to 1998. The NSF also reported that two-thirds of requests to use these facilities by biologists were turned down due to lack of sufficient resources. Networking is the last advance that is enabling the genomic revolution we are experiencing. Communication and resource sharing has long been the hallmark of scientific research, yet this has previously taken the form of stodgy journals and associations, which are notoriously slow, unresponsive and not interactive. The rise of the Internet has

facilitated the transfer of information and tools at a pace and ease never seen before. Two researchers in narrowly defined disciplines located half a world away can share data in real time. Databases such as GenBank provide other examples of what is now possible.

Figure 2
PE Corporation - Celera Genomics Group
Driving Forces



Source: GenBank; Intel; Company financials; Roche Biochemicals; Northwestern University; NIH; DOE; William Blair & Company, L.L.C. estimates

Genomics can advance the efforts of those involved in medical/pharmaceutical research, farming/agribusiness, and forensics.

- ***Pharmaceuticals.*** Understanding disease, either inherited or acquired, at its genetic roots should allow for significantly more accelerated discovery of targeted, cost-effective therapies with fewer side effects. These processes should replace the plodding, expensive methods currently employed.

- ***Farming.*** The application of genomics to agribusiness should enable another green revolution that promises to increase yields and traits of crops and livestock, while displacing the use of pesticides and unsound farming practices.

- ***Forensics.*** The distribution of tools and cataloging of genetic information should enable law enforcement to identify criminals quickly, while simultaneously discouraging criminal behavior.

This genomics revolution, although just beginning, should be one of the most important and far-reaching ever witnessed. Participation in this area as an investor should yield sufficient rewards if one makes well-thought and patient investments.

Attractive Market

We believe that there is a revolution occurring in molecular biology and medicine. This revolution has arisen from a convergence of *needs and funding* in *medicine and agriculture* with a *critical mass of knowledge* in the *life sciences*; *advances in technology such as microfluidics, robotics,* and *molecular biology reagents; and availability of powerful* and *inexpensive computing power.* We consider the market for enabling information technology and related databases for molecular biology to be attractive. We estimate that the worldwide genomics and related markets will total roughly $1.1 billion in 1999, growing 23% annually, as shown in table 2. With its wide-ranging applications, including health care, agriculture, and forensic sciences, we believe that genomics will lead a life science revolution into the next millennium.

Table 2
PE Corporation - Celera Genomics Group
Market Sizes
($ in millions)

Worldwide	1995	1996	1997	1998	1999E	2000E	2001E	2002E	2003E	2004E	CAGR (95-98)	CAGR (99-04)
Bioinformatics	200	255	330	350	420	500	600	720	865	1,040	20%	20%
Molecular Biology Database and Analysis	60	215	325	475	715	960	1,170	1,425	1,740	2,120	99%	24%
Total	$260	$470	$655	$825	$1,135	$1,460	$1,770	$2,145	$2,605	$3,160	47%	23%
Segment Mix	**1995**	**1996**	**1997**	**1998**	**1999E**	**2000E**	**2001E**	**2002E**	**2003E**	**2004E**		
Bioinformatics	77%	54%	50%	42%	37%	34%	34%	34%	33%	33%		
Molecular Biology Database and Analysis	23%	46%	50%	58%	63%	66%	66%	66%	67%	67%		

Sources: Company financials; The Scientist; Frost and Sullivan; Theta Corporation; Business Communications Company; Life Tech; ABRF; Genetic Engineering News; Phrma; Instrument Business Outlook; Phortech; William Blair & Company, L.L.C. estimates

By melding the fields of molecular and cellular biology with classical genetics and computational science, genomics seeks to decipher the information contained within our genetic code, DNA. Deoxyribonucleic acid, or DNA, is found in every living organism and provides all the directions, called genes, necessary to create and sustain life. The sum of all the information or genes in a type of organism is call its genome. The goal of genomics is to understand these complicated directions and essentially produce blueprints that detail the location and function of genes within an organism. Implicit in this process is the understanding of how these DNA blueprints are turned into RNA or ribonucleic acid, an intermediate messenger molecule, and lastly, functional proteins whose job is to carry out the blueprint's designs. This process is described in detail in appendix B, and involves the unwinding and separation of the DNA double helix into individual strands, similar to unzipping a zipper. This allows one of the strands to be copied into mRNA (messenger RNA), a molecule that contains one more oxygen than DNA. The final steps involve the translation of mRNA into long strings of amino acids that subsequently are folded and processed into a functional protein. The ultimate goal is to establish what role each protein plays, and how alterations in the genetic code and these proteins may positively or negatively affect the organism, whether human, animal, microbial, or plant.

Information provided by genomic research should spur exciting revolutions in drug discovery and development, disease management, diagnostics, agriculture, and forensics. For genomics to deliver on these promises, researchers require enabling technologies that can generate, analyze, and store vast amounts of genetic and molecular biology information. This establishes a rapidly growing market with huge potential for both facilitating medical technology and the information it creates.

Current Methods of Drug Discovery Are Inadequate and Inefficient

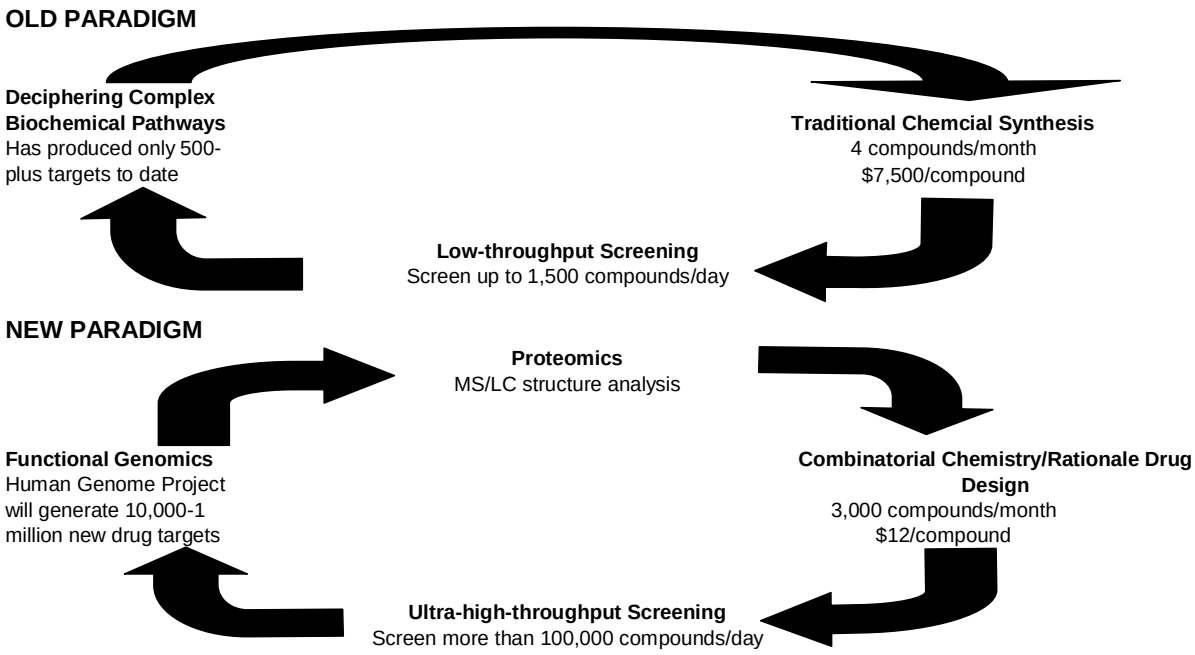
The traditional drug discovery process constrains pharmaceutical and biotechnology companies to the point where they are forced to recycle old targets for drug interaction instead of searching for new, more-powerful ones. This is a direct consequence of searching for targets and potential therapeutic compounds (leads) leads without understanding the underlying biology and causes of disease.

Disease has a genetic component. In recent years, researchers have uncovered numerous links between genes and disease. Disease can occur because of 1) the hereditary passage of gene mutations; 2) the accumulation of mutations in essential genes such as those controlling cell division and cell death (e.g., leading to cancer when cell growth is uncontrolled); and 3) the presence of an infectious agent (e.g., pathogenic bacteria). To date, roughly 5,000 diseases have been identified that are linked to the inheritance of a number of mutated genes. These disorders include Muscular Dystrophy and Cystic Fibrosis. In addition to such inheritable diseases, there also is a link between genetic code and predisposition to diseases that manifest themselves later in life. Such disease etiologies are far more complicated and most likely involve alterations in coupled genes at various points in time. The most obvious example is cancer, which typically is the result of accumulated mutations in crucial regulatory regions of the genome. Lastly, there are countless cases in the world each year of infectious disease, such as HIV, that result from the presence of pathogenic organism. These pathogens contain genetic material that directs the infection process and can serve as sites for therapeutic intervention. With all this in mind, it is likely that the best way to treat disease is to systematically and comprehensively understand the underlying causes and to develop therapies that target them directly, rather than previous approaches that relied more on trial and error based on an incomplete understanding of disease.

The traditional drug discovery paradigm has a bottleneck. To date, most drugs have been identified through both a serendipitous and tedious process of screening. This hit-and-miss or bind-and-grind approach involves screening a small number of organic molecules in disease model organisms and systems, as shown in figure 3, on the following page. The success of this traditional approach is a function of both the number of molecules tested and the number of suitable targets to be screened. Because of increasing health care demands for new therapeutics, much effort has been directed at maximizing both these variables. Researchers have scoured the rainforests for unique compounds, developed new approaches to organic synthesis of molecules (combinatorial chemistry), and automated the screening process to increase throughput. Despite these efforts, there has not been any significant increase in the number of lead molecules available for screening. In addition, attempts to increase the volume of therapeutic targets historically have been unsuccessful. Therefore, the traditional drug discovery paradigm does not offer a mechanism to significantly expand the number of drug targets or lead compounds significantly. *Science* magazine estimates that only 500 molecular targets have yielded successful therapeutic products out of the roughly 100,000 potential points of intervention. In addition, there also is a lack of drug target variety—of the approximately 3,000 human metabolic drugs on the market, 60% are directed at only one (G-protein coupled receptors) of the 4,500 or so classes of proteins. In the case of infectious disease, there are about 70 known target molecules in pathogens and parasites out of the thousands that are believed to exist. As a result of this limited anti-pathogenic arsenal, the CDC and FDA estimate that infectious disease last year alone cost roughly $60 billion in health care

treatment, lost wages, and production waste. Most infected patients are treated with a regimen of antibiotics; however, nonjudicious usage of these drugs has led to an exponential increase in drug-resistant bacteria. This overuse of antibiotics has created resistant "supermicrobes" that threaten to reach epidemic levels if new targets and drugs are not discovered quickly. For example, a rare strain of the pathogenic bacterium s*taphyloccocus aureus* is resistant to all forms of antibiotic treatment. If infected with such "supermicrobes," a person is faced with no treatment possibilities.

Figure 3
PE Corporation - Celera Genomics Group
Drug Screening Approaches



Source: Various company reports; Interviews; *Med Pro Month*; *In Vivo*; McKinsey; Boston Consulting Group; Wilkerson Group; A.T. Kearney; William Blair & Company, L.L.C. estimates

Current drug discovery methods can neither support the growing need for therapeutics nor sustain earnings growth of pharmaceutical and biotechnology companies. The Boston Consulting Group (BCG) estimates that for the pharmaceutical industry a 5.3% gap exists between the earnings growth expected by investors and what is possible using historical operating methods. This is due to an unprecedented level of patent expirations, as well as a late-stage drug development pipeline that is insufficient to satisfy growth expectations. In addition, there are other potential risks looming, such as ongoing cost-containment efforts by insurers and providers. To bridge this gap, pharmaceutical and biotechnology companies likely will require a new drug discovery and development paradigm to quickly fill development pipelines, and stratify and extend markets, for example a paradigm that provides detailed information regarding molecular biology and genetics. Our genetic composition (genotype) plays a significant role in determining our physical attributes (phenotype) and health status. Therefore, the study of genomic information would likely be vitally important to the discovery and development of new treatments. BCG estimates that 0.8% of the earnings gap discussed could be closed by reducing development time by 16 months. Another 0.7% could be closed through a reduction in approval times by 12 months, and at least an additional 0.8% closed if more, high-quality drugs were available for in-licensing. Furthermore, McKinsey & Company estimates that the cost of developing a new drug could be reduced by as much as one-third by employing genomic technology.

Genomics Is Vitally Important

The inability of traditional drug discovery methods to supply adequate levels of targets and leads is the result of not understanding the underlying causes of disease. To elucidate these disease mechanisms, researchers must acquire information on the most basic, structural-level genes. The exciting field of genomics provides a logical new paradigm to drug discovery. Using revolutionary new genetic analysis equipment, genomics, combined with HTS and million-molecule combinatorial chemistry libraries, should contribute a vast array of effective new drugs faster and less expensively.

Genomics will provide a fundamental understanding of disease. Unlike the traditional drug discovery process, genomics provides the most fundamental understanding of disease. In essence, genomics furnishes an understanding of disease pathology and the various alterations and interactions that occur within and between proteins, cells, and drugs. To uncover such information, researchers first need to locate and sequence genes and then characterize gene function through techniques such as functional and comparative computational genomics.

Assigning function to a particular gene or set of genes is one of the primary goals of genomics. These genes may be human or they could be from an infectious organism. The process, functional genomics, utilizes technologies such as expression arrays and proteomics to observe the amounts and/or types of proteins produced from various cell types. The premise behind this form of expression analysis is that in a given cell type, normal genes tend to express functional proteins at a consistent level, whereas genes related to disease may over- or underproduce these proteins, or produce defective ones. Thus, scientists can compare genes from diseased and nondiseased individuals and isolate the ones that exhibit a different "expression pattern." This provides scientists with a hint of what gene and protein types are causing disease and potentially can serve as a site for drug action. Additionally, genes or proteins that are necessary for the function of an infectious organism can be targeted to combat that organism. Expression analysis is highly dependent on fast, highly automated screening systems that only have become available in recent years. With the advancement of new technology, researchers can screen more than 100,000 compounds or cells per day.

Comparative computational genomics relies on extensive databases comprised of genes of known function from a wide range of organisms. Because of evolutionary conservation, many organisms share genes with similar, if not identical, function, explaining one role of model organisms in research. Consequently, it is feasible to compare the less complicated genome from *Drosophila melanogaster* (fruity fly), for example, to one much more complex, such as humans. An example is identification of a mutation within the Pax-6 gene that resulted in an eyeless fruit fly. It also causes mice to be blind, and the same mutation in the human Pax-6 gene causes Aniridia (no iris phenotype), a form of human blindness. Consequently, many companies such as Celera are racing to assemble the largest and most diverse reference databanks composed of sequenced and annotated genomes. This information can then be used to cross-reference unknown genes sequences from other species and determine their function through homology and identify unique targets within pathogens and disease cells.

Genomics should provide a tremendous range of new drug targets. Both the aforementioned techniques provide scientists with a keen understanding of disease mechanisms and, more importantly, produce targets for drug action. By providing insight into individual genotypes, genomics offers a much wider range of targets over current methods. In theory, one could design drugs that act on any one of the steps outlined in appendix B, from the gene itself to mRNA and ultimately proteins. PhRMA, a U.S. pharmaceutical trade organization, estimated that there currently are roughly 1,000 drug targets that have been identified to date, with only about 40% (400-500) yielding suitable sites for drug action. However, it estimates that the Human Genome Project alone will produce at least 10,000 new drug targets.

Genomics should allow for more effective screening and treatment of disease. In recent years, researchers have uncovered genes that appear to play a role in determining whether an individual is at an elevated risk of developing disease. One such example is the discovery of HER-2/neu gene associated with cancer cells. Normal cells carry two copies of HER-2/neu gene and maintain a relatively small number of protein receptors encoded by this gene on their surface. The low number of expressed receptors is used to limit the amount of cell division and thus prohibits uncontrolled cell growth and cancer. However, researchers have found that some forms of breast cancer cells carry extra copies of the gene and a resulting overabundance of the cell surface receptors. Research indicates that this abnormality is not inherited; rather, these mutations are acquired during the course of a woman's life. Researchers estimate that 20% to 30% of women with breast cancer have extra copies of the HER-2/neu genes in their tumors, which amounts to as many as 60,000 cases a year in the United States alone. Armed with this information, scientists set out to create a molecule that could block the action of these "extra" receptors and prevent unchecked growth. The result is Herceptin sold by Genentech, an antibody that attaches to the receptor and prevents the binding of growth factors. A diagnostic test is also available that can identify who is likely to respond to this treatment. This example illustrates the enormous potential of genomics. By starting with a discovery about the basic genetics of cancer, scientists were able to design a specific drug to counteract the abnormality.

Pharmacogenomics Should Have Numerous Beneficial Applications

Pharmacogenomics can be thought of as a marriage of functional genomics and molecular diagnostics. It appears that most diseases are not the result of single gene mutations, but rather the failure of a network of interacting genes, thereby making drug development much more difficult. Further compounding this difficulty is that while disease symptoms might appear to be uniform, individual-to-individual variations in these polygenic networks may make one person respond well to a drug while producing toxic side effects in someone with a different genotype. Pharmacogenomics attempts to correlate these polygenic variations with differential responses to the same drug leads. This should accelerate drug discovery and development by defining specific populations that will benefit most from a drug. Potential applications for pharmacogenomics include drug metabolism, targeted clinical trial recruiting, and disease management.

The future appears bright for SNPs. At the heart of pharmacogenomics are SNPs (single nucleotide polymorphisms, pronounced "snips"), single base pair alterations in a segment of DNA that occur in at least 1% of the population. In essence, when comparing many different individuals' DNA, it is typical to find differences of a single base pair at a rate of one out of every 1,000 base pairs. SNPs are exciting, as they have the potential to be excellent markers for disease because mutations often lead to alterations in protein expression and function. To date, they have led to the genetic elucidation of diseases such as sickle-cell anemia, which is the results of a change in a single letter of DNA. While these polymorphisms have been studied for years in academic labs, only recently, with the development of high-speed sequencing machinery from companies such as PE Biosystems, has their commercial potential been fully appreciated. Pharmaceutical and genomics companies are rushing to catalogue the estimated 30 million human SNPs, distributed across the genome, in hopes of producing new drug targets and corresponding therapies. In addition, many believe that old drugs, previously deemed too toxic to certain patients, might be revived or successfully prescribed to appropriate patient populations. Diagnostic tests would indicate patients that might exhibit adverse reactions or not benefit from a specific drug therapy.

A potential application of SNPs and pharmacogenomics involves drug metabolism. The CDC reports that each year 100,000 people die in the United States from unwanted pharmaceutical side-effects. With the aid of gene sequencing and mapping, researchers have uncovered SNPs within key metabolic genes that have been correlated with increased drug efficacy and also toxicity. An example of the power of pharmacogenomics is the discovery of the Cytochrome p450 oxidase SNPs in the liver. This particular enzyme

appears to be responsible for the metabolism of many classes of drugs. Some polymorphisms result in expedited drug metabolism, thereby preventing any drug action. Other polymorphisms result in metabolic enzymes with decreased or no activity and can lead to toxic side effects and death. An example of this is Hoechst Marion Roussel's Seldane allergy drug, which recently was pulled off the market because of toxic side-effects observed in a small population of patients. It was discovered that individuals possessing a particular Cytochrome p450 variant were unable to metabolize the drug properly when taken in conjunction with erythromycin, a commonly prescribed antibiotic. If Hoechst had been able to identify these potential subgroups in Phase I clinical trials, it could have alerted physicians about the possibility of adverse reactions or scrapped the project and saved millions of dollars in development costs.

Pharmacogenomics and clinical trials. One also can imagine the value of genetic information in the context of clinical trials. Selection of a test population for a drug trial is one of the most important and time-intensive steps in drug development and can greatly affect the fate of a drug. Using SNPs, one can select subpopulations that may experience untoward effects and proceed with those individuals that would most benefit. By selecting patients with the highest drug responses, pharmaceutical companies and CROs could reduce the number of people in each phase of the trial. According to *In Vivo,* if patients are preselected on the basis of drug response, the number of patients in Phase II trials could be reduced by as much as half while the statistical power of the test remained the same. In addition, pharmaceutical companies can save large sums of money if screening eliminates nonpromising compounds earlier in the clinical trial process. This increases the number of drugs on the market and at the same time decreases the high costs of clinical trials. Currently, only 10% of NCEs (new chemical entities) make it through clinical trials at a cost of roughly $500 million per compound over a 10- to 15-year period from discovery to development.

Genomics Should Have a Tremendous Effect on Agriculture

Much of our current understanding of genetics can be traced back to the work of Gregor Mendel, who uncovered the mysteries of heredity using peas nearly a century before Watson and Crick established the structure of DNA in 1952. The study of plant physiology, metabolism, and biochemistry has provided a keen insight into highly conserved biological processes that are observed in humans, animals, and microbes. For thousands of years, agriculturists have used crossbreeding techniques to produce healthier, higher-yielding crops and livestock. Whether they did it knowingly or not, farmers introduced desirable traits such as increasing yields, pest and disease resistance, and the ability to adapt to adverse climates. Therefore, use of genomics in agriculture has had a long history and likely will play an even more vital role in the future.

More mouths to feed. As previously mentioned, the global population will reach more than 6 billion in 1999. An increasing population will require productivity-enhancing technologies to maintain an adequate food supply. The National Center for Genetic Information currently reports that one U.S. farmer provides food for 128 people—94 in the United States and 34 in the remainder of the world. While it is true that an increasing population will require more food, there are additional drivers for agricultural improvement. The National Corn Growers Association notes that changes in the global economic and social structure are leading to demands for different types of high-quality food and crop-derived products such as corn oil. While traditional agriculture techniques such as hybrid corn breeding have increased production efficiency, they are time-consuming and highly labor intensive. Consequently, a new technology, such as genomics, likely will be employed.

With an understanding of the biochemical and cellular mechanisms in plants, scientists should be able to engineer herbicide and pest-resistant plants with minimal effect on the environment. We already have begun to witness the benefits with products such as Monsanto's Bollgard cottonseed. Last year, Monsanto cottonseed was planted on roughly 2.5 million of the entire 14 million acres of cotton in the United States. At a seed price of

about $33 per acre, this amounts to approximately $83 million. However, it is estimated that farmers spend roughly $400 million-$500 million on insecticides annually, or an average of about $70 per acre of non-Bollgard seed. Thus, not only were farmers saving money, but they also were able to provide a more productive harvest while sparing the environment from excessive amounts of pesticides. Nearly half of the United States' soybean fields already are being planted with genetically enhanced seeds

Genomics Should Offer Exciting Opportunities in Disease Management and Nontraditional Health Care Markets Such as Forensics

In addition to aiding in the development of therapeutics, genomics and its associated technologies can be applied to the large disease management market. Physicians or clinical reference labs potentially could employ genomics information and technologies to detect and monitor disease progression. PE Biosystems has produced an HIV genotyping application that can help pharmaceutical companies understand the relationship between HIV sequences and resistance to specific drugs. PE Biosystems currently is involved in a clinical trial to determine if this test can be used to help tailor the most effective drug regimen against the virus. Because HIV has a high mutation rate, it has the ability to develop resistance to many of the commonly used HIV cocktail drugs such as AZT and protease inhibitors. By measuring viral load (the amount of viral particles in the system) and identifying sequences of the mutation-prone regions within the protease and reverse transcriptase genes, physicians can choose an HIV cocktail that should be most effective in combating the virus. In addition, one can imagine the use of real-time PCR and DNA detection systems, which have the ability to amplify and quantify small traces of DNA in clinical settings to provide faster, more-sensitive screening of patients. This same type of instrumentation also can be used outside health care and in fields such as forensics. Law enforcement agencies currently are setting up initiatives to compile DNA samples from criminals in hopes of constructing an extensive database. Such a database would be used to identify criminals through biological specimens left at crime scenes or elsewhere. In 1998, the FBI opened a national DNA database that contains samples from sex offenders in each of the 50 states. In Europe, similar programs have been implemented, although with a slightly different set of genetic markers, as shown in table 3. Other potential forensic uses are listed in table 4. Applying high-speed, automated genomics equipment to these areas will increase the efficacy of results and decrease the investments in time-intensive labor.

Table 3
PE Corporation - Celera Genomics Group
National/International Genetic Loci Standards

Locus	CODIS	ENFSI
D16S539	X	
D7S820	X	
D13S317	X	
D5S818	X	
CSF1PO	X	
TPOX	X	X
TH01	X	X
Vwa	X	X
FGA	X	X
D21S11	X	X
D8S1179	X	X
D18S51	X	X
D3S1358	X	X
Amelogenin		X

CODIS- Federal Bureau of Investigation's Combined DNA Index System
ENFSI- European Network of Forensic Science Institutes DNA Working Group
Source: Promega Corporation; FBI; ENFSI; William Blair & Company, L.L.C. estimates

Table 4
PE Corporation - Celera Genomics Group
Uses of Genetic Forensics for Idenification

1. Identify potential suspects whose DNA matches evidence from a crime scene
2. Exonerate people wrongly accused
3. Establish familial relationships, such as paternity
4. Identify victims of catastrophes, such as plane crashes
5. Match organ donors
6. Detect microorganisms, for example those that may infect food
7. Determine animal or plant pedigrees
8. Identify endangered species
9. Authenticate consumer products, such as wine

Source: United States Human Genome Project; William Blair & Company, L.L.C. analysis

Consequently, We Believe that Genomics and Molecular Biology Technology Is an Enormous Market With Strong Growth Potential

Big pharmaceutical and biotechnology companies are jumping on the genomics bandwagon, in addition to academic research. This is illustrated by the large financial and staffing commitments from major pharmaceutical and biotechnology companies to genomic endeavors. We estimate that pharmaceutical companies will spend almost $54 billion on R&D in 1999. Of that, we believe that approximately $17 billion was allocated toward the discovery and development of new drugs. Furthermore, we estimate that approximately 15% of these budgets, or almost $2.5 billion, will be devoted to genomics projects.

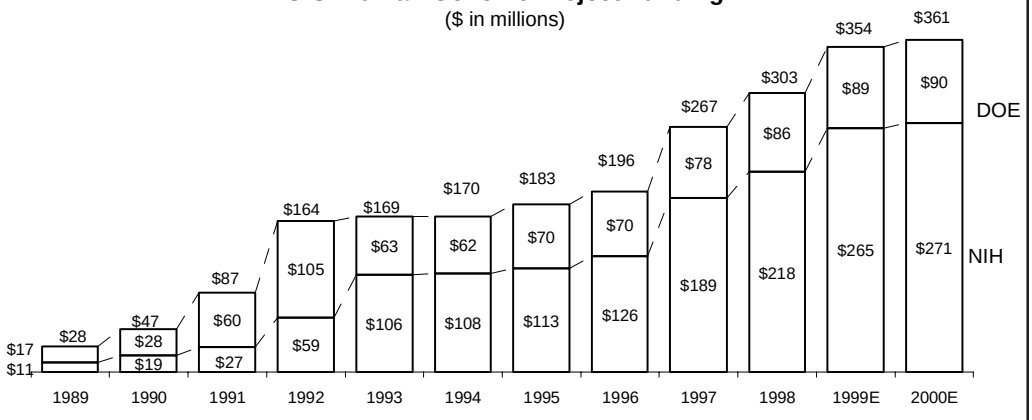
Because of this major commercial push, the U.S. government progressively is increasing funding for genomics projects such as the Human Genome Project. Started in 1990, the Human Genome Project is a $3 billion, 15-year program to determine the more than 3 billion nucleotides and locate the estimated 100,000 genes in humans, as shown in table 5. Spending has been considerable to date, as shown in figures 4 and 5, on the following page, and scientists have made progress, as shown in table 6, on page 21.

Table 5
PE Corporation - Celera Genomics Group
Human Genome Project Timeline

1984	U.S. Office of Technology Assessment describes value of human genome reference sequence
1985	U.S. Office of Health and Environmental Research commission Sante Fe Conference to assess feasabliilty of Human Genome Project (HGP)
1986	HGP with $5.3 million pilot budget.
1987	DOE sets 15-year plan and designates genome centers. NIH begins funding.
1988	Human Genome Organization (HUGO) founded to organize global efforts.
1990	DOE and NIH present joint 5-year HGP plan to Congress, commencing 15-year project.
1991	Data repository established for human chromosome mapping.
1992	Low-resolution genetic linkage map of entire genome published.
1993	DOE and NIH revise five-year goals. Genetic and Insurance Information recommendations released.
1994	Genetic mapping five-year goal achieved one year early. Genetic privacy act proposed.
1995	High- and moderate-resolution physical maps released for chromosomes 16,19,3,12, and 22.
1996	Wellcome Trust sponsors strategy meeting for international coordination.
1997	High resolution maps for chromosomes X and Y completed. Joint Genome Institute form to implement high-throughput activities.
1998	Perkin Elmer forms Celera to sequence entire human genome by 2001. DOE and NIH revise five-year plan to complete HGP by 2003.
1999	DOE and NIH announce that draft, but incomplete sequence to be finished by 2000.

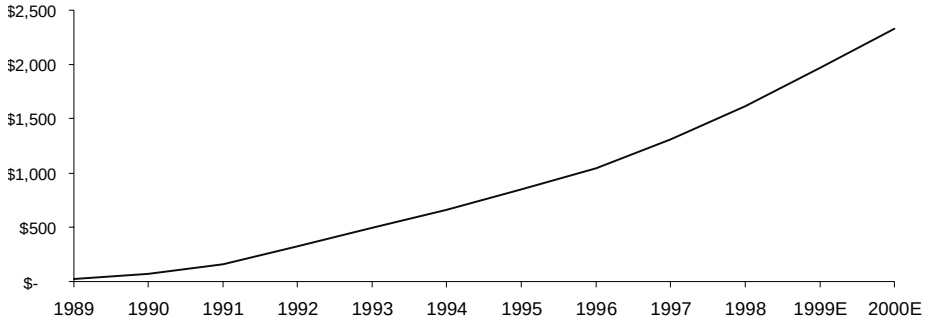
Source: DOE; NIH; Willam Blair & Company, L.L.C. analysis

Figure 4
PE Corporation - Celera Genomics Group
U.S. Human Genome Project Funding
($ in millions)



Source: NIH; DOE; NHGRI; William Blair & Company, L.L.C. estimates

Figure 5
PE Corporation - Celera Genomics Group
Cumulative U.S. Human Genome Project Funding
($ in millions)



Source: NIH; DOE; NHGRI; William Blair & Company, L.L.C. estimates

Genomics: a universe of companies. As a result of the extraordinary promise offered by genomics, there exists a large and growing market for genomic data and the instruments and services that are utilized to expedite the process of gathering, analyzing, and applying genomic information. At present, the genomics universe comprises approximately five types of companies. The first type provides tangible products such as enabling hardware, software, and reagents. These companies produce the DNA sequencers, PCR machines, HTS machinery, and reagents that are driving genomics research in both the academic and industrial arenas. The second type of company provides contract research services, such as target discovery, high-throughput screening, and combinatorial chemistry, for big pharmaceutical companies. Many of the large pharmaceutical companies do not possess the necessary equipment or expertise to screen the thousands of compounds that are created as possible leads. Consequently, a significant portion of this work is contracted out to companies that specialize in high-throughput screening or expression arrays. A third class of companies can be described as gene hunters. These companies use the hardware and software provided by the enabling medical technology companies to track and identify potential drug targets. Typically, a company of this type would form collaborative agreements with major pharmaceutical companies in different disease areas. Such companies derive revenue from milestone payments and equity investments from their pharma partners. A fourth class is companies that generate genomic information and provide outside access via subscription databases or collaborative agreements. These companies focus on creating comparative databases composed of a wide array of organisms and allowing outside

parties to access this wealth of information via annual subscriptions. In essence, they provide genomic information with which pharmaceutical and biotechnology companies can create new drugs or diagnostic tests. Lastly, there are the product suppliers: pharmaceutical, biotech, and diagnostic companies.

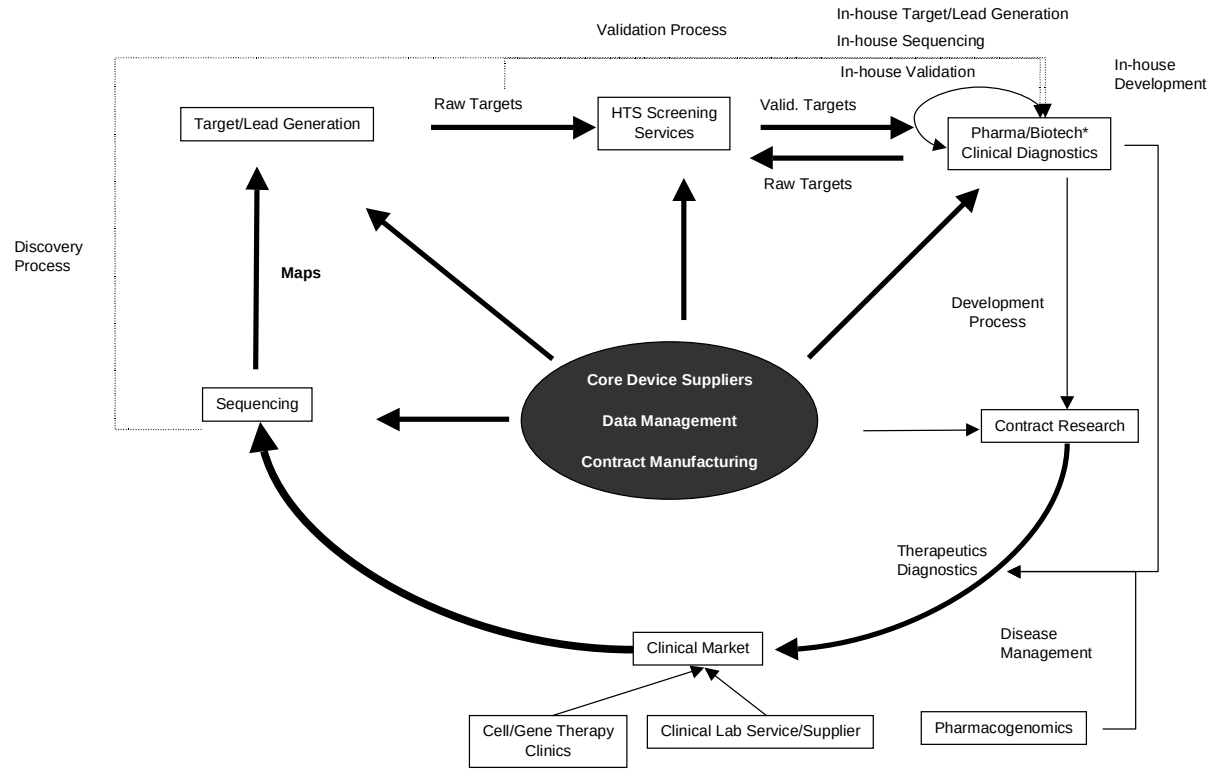
Table 6
PE Corporation - Celera Genomics Group
Human Genome Project Sequencing Progress*

Chromosome	Effective Size** (kb)	Sequence Completed (kb)	Percentage Finished
1	263,000	22,067	8.4%
2	255,000	17,445	6.8%
3	214,000	8,915	4.2%
4	203,000	11,220	5.5%
5	194,000	13,479	6.9%
6	183,000	43,038	23.5%
7	171,000	77,006	45.0%
8	155,000	8,633	5.6%
9	145,000	5,192	3.6%
10	144,000	3,945	2.7%
11	144,000	8,420	5.8%
12	143,000	21,319	14.9%
13	98,000	2,000	2.0%
14	93,000	15,734	16.9%
15	89,000	2,118	2.4%
16	98,000	16,632	17.0%
17	92,000	28,643	31.1%
18	85,000	3,366	4.0%
19	67,000	15,017	22.4%
20	72,000	20,406	28.3%
21	39,000	18,678	47.9%
22	43,000	33,013	76.8%
X	164,000	62,259	38.0%
Y	59,000	3,050	5.2%
Total	3,213,000	461,595	14.4%

* As of 11/17/99 ** excludes repetitive DNA
Source: NIH; GenBank; William Blair & Company, L.L.C. estimates

Presently, many companies exist in the genomics universe with a wide range of technologies and strategies, as shown in figure 6, on the following page. As with biotechnology, it is difficult to determine which one of these companies ultimately will produce successful therapeutic or diagnostic products. However, we can say that regardless of which company ultimately prevails, they all will require enabling technologies and services to complete their endeavors. We strongly believe that genomics will be vital to the future of health care, agriculture, and forensics. Consequently, we believe that the most high-quality investments are in companies that provide the enabling medical technology for genomic and molecular biology research.

Figure 6
PE Corporation - Celera Genomics Group
The Genomics Landscape



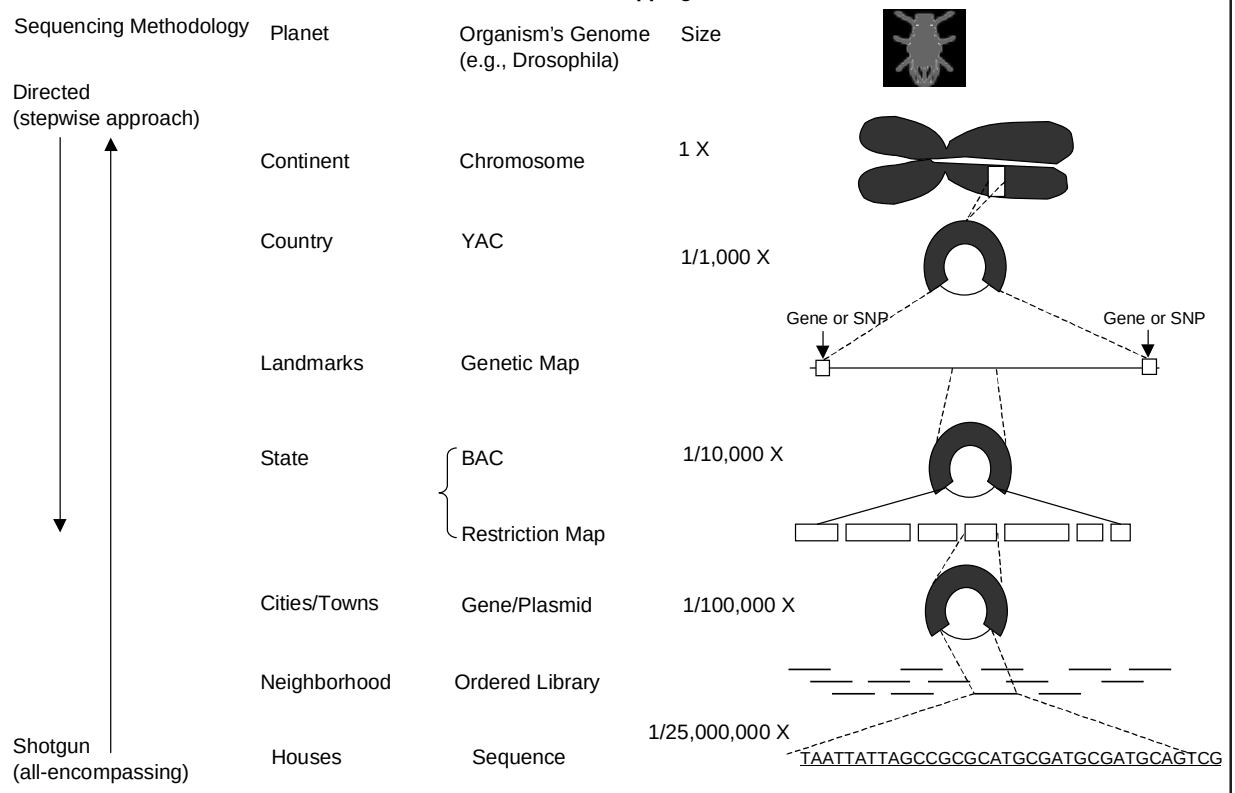
*Includes: cell/gene therapy development companies
Source: William Blair & Company, L.L.C. analysis

Unparalleled Capabilities

PE Corporation created Celera to change completely the rules in gene discovery and more broadly in the development and use of molecular biology knowledge. Prior to the company's whole-genome approach, both public and private efforts collectively used a top-down strategy, directed sequencing, as shown in figure 7. We liken this to a hierarchy and collection of maps. The other organizations first identified the countries, then the states, then the cities, and so on, drilling down to specific addresses only if they appeared interesting. In contrast, Celera intends to operate like an interactive map, much like one would find for some Internet map programs such as MapQuest. By building the map from the bottom up, one can access all the specific addresses—base pairs, genes, chromosomes, organism, or various hierarchies in between. One also avoids critical data comparability issues and the potential to lose data in the process.

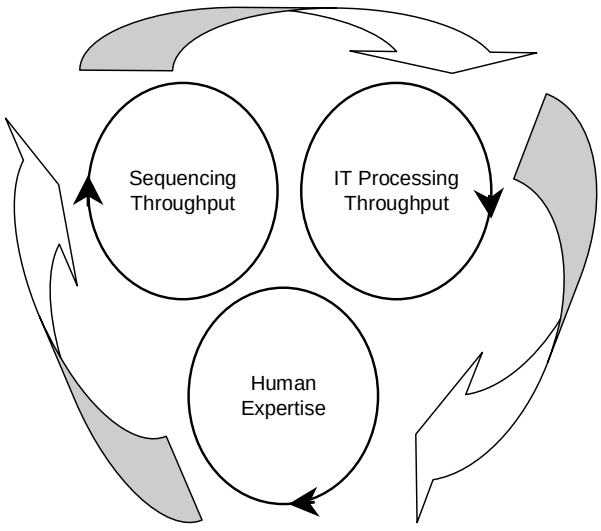
To accomplish this, the company brought together three reinforcing components (as shown in figure 8) that alone were to be *the best in the world* and together should establish a self-reinforcing system that should build large and unique competitive barriers. The first component is the *unprecedented gene sequencing factory* of 300 ABI Prism 3700 DNA analyzers that can sequence in one month more DNA than has been added to the GenBank database ever. The second is the world's largest private parallel processing supercomputer facility composed of 1200 Compaq Alpha processors. The third is the hiring and networking of vast human expertise in molecular biology and software engineering, such as algorithm development. In our opinion, the output of the subsequent system (that is, Celera) already has proved itself in principle and is generating *high-quality* data and information at unprecedented rates.

Figure 7
PE Corporation - Celera Genomics Group
Genetic Mapping



Source: DOE; William Blair & Company, L.L.C. analysis

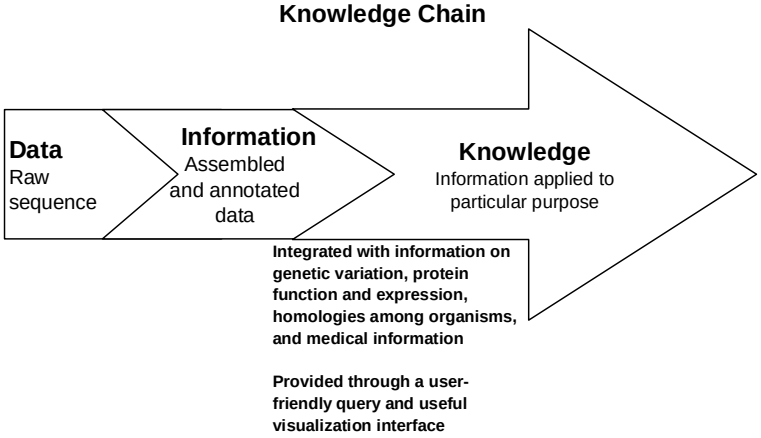
Figure 8
PE Corporation - Celera Genomics Group
Core Resources and Capabilities



Source: Industry interviews; William Blair & Company, L.L.C. analysis

We believe that at its foundation, the company is a vehicle for transforming genomic data into knowledge, as shown in figure 9. While the company has been likened to *Bloomberg* for biologists—both comprehensive and ubiquitously used—we strongly believe that this oversimplifies and underrepresents what the company can do to build knowledge. At the heart of the company is high-quality data generation, not just reporting. This data is intended to meet the "Bermuda Standard," as shown in table 7, set by the international humane genome consortium at the Bermuda meetings sponsored by the Wellcome Trust. It is this quality, in addition to the quantity of data generated, that should help to establish a proprietary position for the company. A competitive company not generating all the data from scratch, such as Celera intends to do, would face at least two barriers: the enormous data input task, and, as important, *data quality*. As table 8 shows, the public databases have many sources of errors. This high-quality data must be fed into *proprietary algorithms* and a *powerful enough computing system* that can parse it, match it with other known sequences, and annotate it to create information. We believe that this task requires extremely sophisticated and scarce genomic and bioinformatic knowledge. The annotated genome data must finally be integrated with information on genetic variation, protein function and expression, homologies among organisms, and medical information. It also should be provided through a user-friendly interface that allows for easy and appropriate queries, and provides useful representation or visualization of the information, to complete its transformation into knowledge. We believe that this is the vision for Celera—*not* just to be the *Bloomberg* of genetic data.

Figure 9
PE Corporation - Celera Genomics Group
Knowledge Chain



Source: William Blair & Company, L.L.C. analysis

Table 7
PE Corporation - Celera Genomics Group
Bermuda Accuracy Standard for Gene Sequencing

I.	**Less than 1 error in every 10,000 base pairs of DNA**
II.	**No ambiguity in order or location**
III.	**No gaps**

Source: 11th Annual Genome Sequencing Annual Conference

Table 8
PE Corporation - Celera Genomics Group
Sources of Quality Problems Within Public Databases

Type of error or difficulty	Comments
Inclusion of DNA from unrelated organism	2% of GenBank may include
Missing base pairs	
Incorrect sequence	
Gene mapped to the wrong chromosome	
Partial genes labeled complete	
Slight variations of same gene	
Complex gene with various splicing	1/3 of genes may be alternately spliced
Function misclassified	15% of GenBank annotation unverified or out of date
Organism misclassified	
Alternate spellings of gene names Different names for same gene	Database search, including GenBank, failed to find 26% mouse genes known to match human genes and 17% of the reverse.

Source: *Science*; *Bioinformatics*; William Blair & Company, L.L.C. estimates

Unprecedented Gene Sequencing Factory

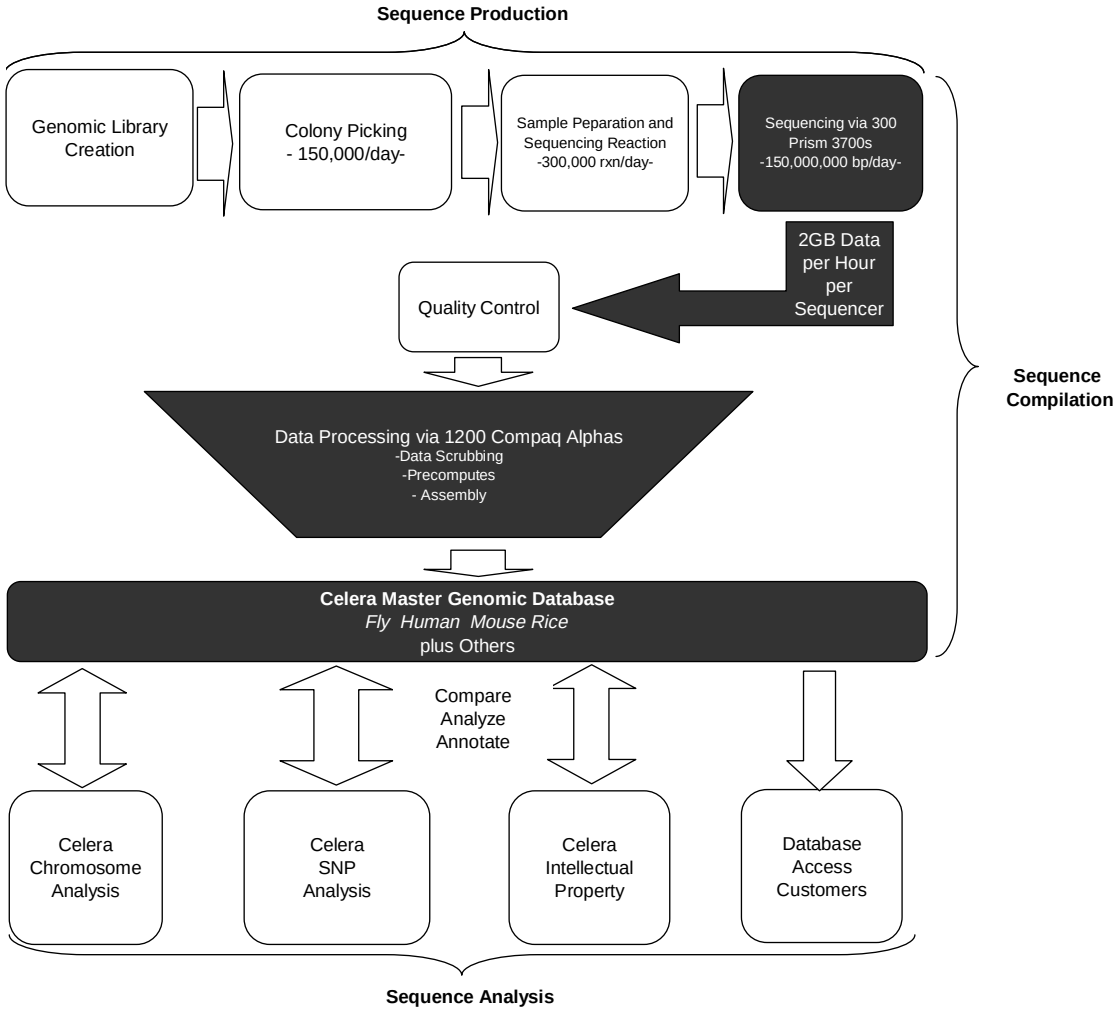
Gene sequencing. The company built the highest-capacity DNA sequencing process and facility, as shown in figure 10, on the following page. This unique factory is designed around 300 of PE Biosystems' latest sequencers, the Prism 3700, a 96-column *high-speed*, capillary electrophoresis system. This capacity rivals the entire high-throughput capacity of the public, human genome project international consortium, as shown in table 9, on the following page. The Prism 3700 sequencer is *fully automated* and can be *run for 24 hours unattended.* Each sequencer can generate about 100 million finished base pairs (bp) per year (about 150 million raw bp), leading to an effective capacity of *30 billion finished bp per year* for the entire factory (roughly 45 billion raw). To put this in context, GenBank, the public repository for genetic sequences, had 2.2 billion *total* bp in its database prior to the formation of Celera. To support these instruments, the company uses robots to *pick* about 150,000 bacterial colonies per day that have human DNA spliced into them (bacteria are a convenient way to manage and maintain the DNA). The whole set of processes is set up as a high-capacity, routine production facility, minimizing waste and maximizing output.

Link with PE Biosystems provides incremental value. Celera's close ties to PE Biosystems enhances its sequencing capability, while at the same time providing PE Biosystems with priceless feedback normally not available from customers. For example, on the basis of Celera's feedback the ABI Prism 3700 cabinet was redesigned early in production to improve fluid handling and ease of use. We believe that the two organizations are working on multiplexing dye systems that would allow the analysis of more than one strand of DNA per capillary, effectively multiplying the capacity and throughput of this system and other genetic analysis systems from PE Biosystems.

World's Most Powerful, Nongovernment Supercomputer

Celera is building the world's largest nongovernment supercomputer—as the world's largest DNA sequencing factory will generate enormous amounts of data, which also must be processed, stored, and subsequently analyzed and distributed. Matching comparisons such as those used in genomics research grow exponentially as data is generated. Each sequencer generates two gigabytes of raw data per hour, 600 gigabytes in total for the entire facility. The data must be processed and cleaned to generate high-quality, finished base pairs daily. These must then be compared with other known sequences to identify relationships and structures (e.g., what part of a gene or what kind of gene), as well as be added to the assembly process to finish the entire genome (as discussed in the appendix on Shotgun Sequencing). The sequences produced each day will be compared with every publicly available gene, gene fragment, protein, and EST, a process that is expected to take 18 hours. Also, the company each day expects to add 15 to 20 gigabytes of finished data to the database.

Figure 10
PE Corporation - Celera Genomics Group
Celera's Genome Sequencing Process



Source: Company interviews; Compaq; William Blair & Company, L.L.C. estimates

Table 9
PE Corporation - Celera Genomics Group
Human Genome Project International Consortium Capacity

Manufacturer	Sequencer	Number
PE Biosystems	ABI 377	~250
PE Biosystems	ABI 3700	~230
Amersham	MegaBACE	~110
Licor	IR2	~70
Total		~660
People	~1,100	
Raw bp/day	~76,000,000	

Source: NIH; National Human Genome Research Institute; William Blair & Company, L.L.C. estimates

The supercomputer itself consists of 1,200 interconnected Compaq Alphas, and we estimate that the entire computing facility will have cost approximately $70 million upon completion. The company has a strategic alliance with Compaq to maintain and build both companies' leadership in bioinformatics. This supercomputer compares with the world's most powerful supercomputer at Sandia Labs designed to simulate nuclear weapons and consisting of 9,152 Intel CPUs. In contrast, Incyte and the not-for-profit Sanger Center—two large competing organizations—have approximately 220 and 250 Compaq Alpha CPUs, respectively. As parallel processing supercomputing power grows exponentially as more CPUs are added, the computing power gap among competitors is wider than the numbers appear to indicate at first glance. Each Alpha processor can perform more than 250 billion sequence comparisons per hour, or 1.3 trillion floating point operations per second. In practice, the Celera system will be able to perform a "BLAST search" to match DNA sequences about 1,000 times faster than public databases and will offer its own public version 10 times faster than those. In addition the company will likely use pre-computes for certain valuable, yet time consuming analysis to further enhance performance. As shown in table 10, biologists already rely substantially on bioinformatics computing, and the trend appears to be increasing still, leading to ever-increasing needs for speed.

Table 10
PE Corporation - Celera Genomics Group
Bioinformatics Usage

- Biomedical researchers spend 90% of time in front of a computer
- Usage of biological databases increasing 10%-15% per month
- NCBI searches (e.g., GenBank)
  - –1991 195/day
  - –1994 5,000/day
  - –1998 600,000/day
  - –2002E 25,000,000/day
- Total determined DNA sequences
  - –1991 71 million base pairs (bp)
  - –1994 217 million bp
  - –1998 2,008 million bp (*Celera built-out capacity 3 billion bp/month*)
- At the NSF super computing center
  - –Two-thirds of requested biomedical cycles still turned down despite doubling of cycles available
  - –12% of users are biomedical and account for 25% of cycles--a 54% increase in one year

Source: 1999 NIH survey; William Blair & Company, L.L.C. estimates

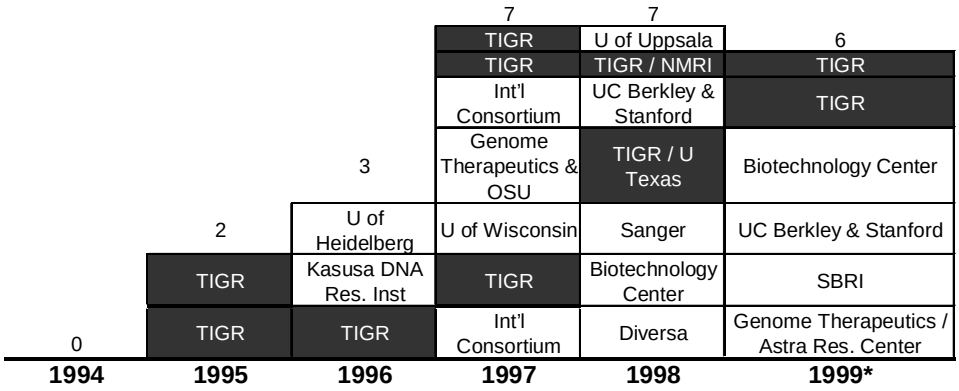
World-class Expertise: Productive Scientists and Software Engineers

To capture the value of the unprecedented gene factory and largest, nongovernment supercomputer, Celera added experts in genomics and bioinformatics, with about one-third biological scientists, one-third software engineers, and one-third hardware engineers. More than 60% are in Maryland at the gene sequencing factory and supercomputer facility, with most of the remainder located in California at "Celera West," the "wet" biology group arising from the combined former Genscope and AgGen businesses of PE Corporation. The scientists and engineers not only are chosen for their intellectual expertise, but also for their productivity and discipline. We believe that the company is focused on building a team of experts who have succeeded with their goals, and therefore do not know how to lose a race. At the head of Celera is J. Craig Venter, Ph.D., founder, and still chairman of The Institute for Genomics Research, probably the leading genome sequencing organization before Celera. At TIGR, he pioneered the whole-genome shotgun sequencing technique now used at Celera, which led to the first three whole genomes ever sequenced. Prior to TIGR, Dr. Venter was at the NIH, where he co-invented the EST (expressed sequence tag) approach to gene sequencing that formed the foundation for Human Genome Sciences, a firm focused on creating genomic-derived pharmaceuticals. The front end of the sequencing process—the creation of gene libraries—is overseen by Nobel Laureate Hamilton Smith, M.D., an expert in biochemistry and microbial genetics who, along with Dr. Venter at TIGR, sequenced the first bacterial genome, *H. influenzae*. The company's chief medical officer is Samuel Broder, M.D., who headed the U.S. National Cancer Institute from 1989 to 1995

before joining IVAX as head of R&D. His role is to integrate the knowledge generated into medical applications. Mark Adams, Ph.D., is the head of Genome programs and was co-inventor with Dr. Venter of the EST approach. He also comes from TIGR, where he headed DNA sequencing. For algorithm development, the company hired Eugene Meyers, Ph.D., who co-developed the ubiquitous BLAST search tool for DNA and protein sequences. Dr. Meyers was part of the original group to propose shotgun sequencing for the whole human genome in 1996; he followed his assertions with a published scientific paper supporting his contentions in 1997. Robert Millman is the patent director at Celera. He was previously with Millennium Pharmaceuticals, and he also worked at two law firms involved in intellectual property for genomics, Morrison and Foerster, and Sterne, Kessler, Goldstein and Fox. In addition, before his law career he taught graduate-level laboratory classes in molecular biology and immunological techniques. Overall, we believe that the company has been able to and should continue to be able to attract the highest-caliber scientific and bioinformatic talent in large part due to its exciting and ground-breaking mission, and the infrastructure it has built to execute this mission.

The company maintains strong relationships with TIGR at all levels to sustain its access to other leading-edge personnel and technologies. TIGR also was granted 5% of Celera's stock through options to help align it with the company's achievements. TIGR, a not-for-profit organization, was founded in 1992 by Dr. Venter in Maryland near Celera's current location. It has achieved considerable success and a strong reputation since. For example, in one year it identified about one-third of expressed human genes using the EST method invented by Drs. Venter and Adams. In addition, scientists at TIGR *sequenced the first three microbial genomes*, *H. influenzae*, *M. genitalium*, and *M. jannaschii*, and they have continued to contribute substantially to whole sequencing for microbial genomes, as shown in figure 11.

Figure 11
PE Corporation - Celera Genomics Group
Number of Whole Sequences Completed for Microbial Genomes
(organization that sequenced the organism)



* partial year
Source: Various scientific journals; William Blair & Company, L.L.C. estimates

Results to Date: Proof of Principle

Now that the company has built much of the unparalleled infrastructure envisioned, what are the results to date? Celera started the Drosophila sequencing in May 1999, and by the end of August it ran the 6X coverage data through the assembler program (see appendix B for more information on coverage, the assembler program, and their implications). Redundancy in sequencing helps to avoid breaks or gaps in a sequence, as sampling done when a CD player reads a Compact Disc helps to avoid skips when playing music. At that point, the assembler correctly designated 97% of the continuous stretches of DNA correctly. By September 17, the company had completed the 10X coverage needed to fully assemble

the Drosophila genome and publish the results before year end 1999. Along the way, the company filed approximately 1,500 novel gene patents and released much of the sequence data to the public database, GenBank. Novel gene patents from Drosophila could be quite useful in a number of ways. Perhaps most obvious, they could be licensed to a firm developing insecticides. Potentially more important applications may exist in studying or correcting human disease. As table 11 shows, there are a number of similarities between human genes that cause disease and those of a fruit fly. As the proteins which arise from genes are often *conserved* or remain quite similar among many organisms, such as the fruit fly or mouse, these *model organisms* can often be critical to our understanding of the human disease, as well as development and testing of treatments to alleviate or cure the diseases.

Table 11
PE Corporation - Celera Genomics Group
Human Diseases With Greater Than 50% *Homology* With a Fruit Fly Gene

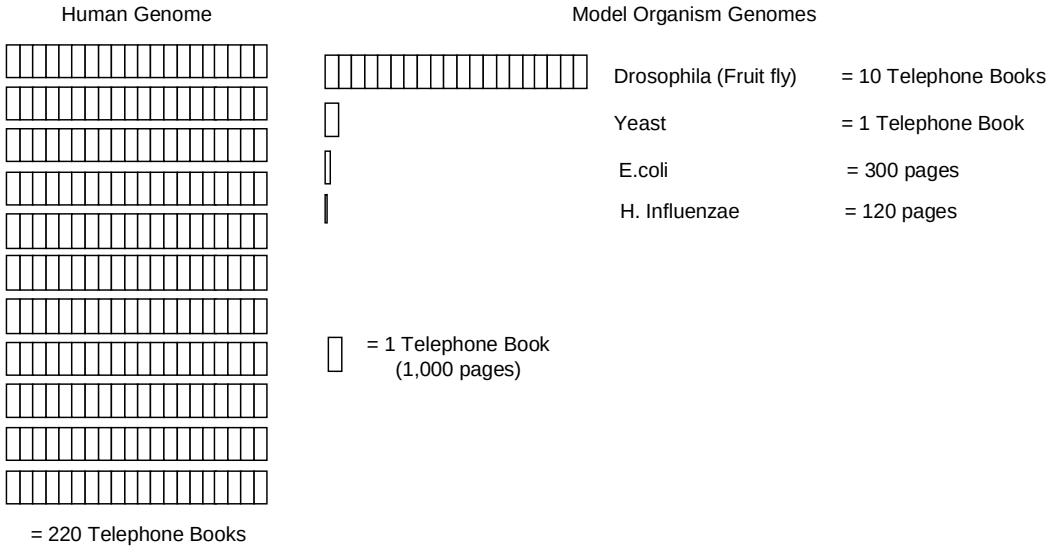
Adenomatous Polyposis Coli	Fragile-X Syndrome	Myotubular Myopathy 1 X-linked
Adrenoleukodystrophy X-linked	Glycerol Kinase Deficiency	Neurofibromatosis Type 1
Agammaglobulinemia X-linked	Gonadal Dysgenesis	Neurofibromatosis Type 2
Alzheimers' Disease (Chromosome 1)	Hereditary Multiple Exostoses Type 1	Niemann-Pick Disease Type C
Alzheimers' Disease (Chromosome 14)	Hereditary Multiple Exostoses Type 2	Oculopharyngeal Muscular Dystrophy
Amyotrophic Lateral Sclerosis	Hereditary Non-Polyposis Colon Cancer	Peutz-Jeghers Syndrome
Aniridia	Holt-Oram Syndrome	Rieger Syndrome Type 1
Ataxia Telangiectasia	Hyperexplexia	Situs Inversus
Best Macular Dystrophy	Long QT Syndrome	Thomsen Disease
Cystinosis	Miller-Dieker Lissencephaly	Waardenburg Syndrome
Duchenne Muscular Dystrophy	Multiple Endocrine Neoplasia Type 1	Wilms Tumor
Dystonia	Myotonic Dystrophy	

Source: Berkley Drosophila Genome Project; William Blair & Company, L.L.C. analysis

Celera has now essentially completed *and* annotated the Drosophila genome. In November, the company conducted a two week workshop with 80 scientists (47 external and 33 internal) to complete the annotation. Out of the process emerged 13,000 genes, in contrast to the 3,000 genes previously known. As an example of the power of the newly completed database, one scientist who had been looking for *three years* to find a p53 homolog in Drosophila found one in *30 seconds* using the Celera database. The p53 gene found in humans appears quite important in the development of *cancer.* Identifying the same gene in *Drosophila* could provide a useful model for research. Additionally, empirical tests matching Drosophila (fruit fly) and Arabidopsis (a plant) genes to ESTs (expressed sequence tags), which are used by competing sequencing firms, appears to indicate that the approach by Celera yields least 2 times as many actual genes as competitors.

Drosophila has allowed the company to prove and refine its shotgun approach on a model organism much larger than those sequenced before, as well as fine tune its process for filing patents. To put this in context, *H. influenzae*, the first microbial genome sequenced, is only 1.8 million base pairs long, compared with 140 million for Drosophila and 3.5 billion for humans. This difference also is shown in figure 12, on the following page. The original assembler program for *H. influenzae* took *one day* to generate the complete sequence at TIGR. In contrast, the assembler used by Celera for Drosophila in the supercomputer takes 36 hours for 10X coverage, but only takes *5 minutes* for *H. influenzae.*

Figure 12
PE Corporation - Celera Genomics Group
Genome Size Comparison



Source: DOE; NIH; NCBI; William Blair & Company, L.L.C. analysis

Most important, we believe that the company is well on track to complete the full human genome by the end of calendar 2000, one year earlier than originally planned. We base this assessment on the Drosophila results, and the recent human sequencing of more than 3.1 billion base pairs—with 1X coverage (3.5 billion base pairs) expected by the end of calendar year 1999. Also by the end of 1999, we expect that the company should have identified roughly 1 million SNPs, considerably more than the 150,000 SNPs sought by the SNP consortium—consisting of 10 pharmaceutical firms—by April 2001. We now expect the company will only need to sequence human DNA through June 2000 to complete the genome. Along the way, we anticipate that the company also will file patent applications, as discussed in the following section. To assemble the full human genome, the company estimates it will take 90 days of total computing time for the assembler, although the assembler can be run concurrently with data gathering. Thus the complete human genome should be available well before year end 2000. The company has signed three early access partners—Amgen, Novartis, and Pharmacia & Upjohn—accessing its Human Gene Index, Drosophila Genome database, and human genome database. It also has begun gene discovery programs with Rhone-Poulenc Rohrer for humans and RhoBio for maize. Most recently, the company has signed an agreement with Pfizer, that not only includes access to all Celera databases, but also to *genomic services* as well. These genomic services are to help Pfizer identify novel genes useful for developing new drugs. Drug targets will be licensed to Pfizer nonexclusively, but therapeutic proteins identified as part of the agreement would be licensed exclusively, with Celera's receiving both milestone payments and royalties. As table 12 shows, there are many large customers interested in genomic and molecular biology data, and they appear interested in continuing to add new sources and new types of information.

Table 12
PE Corporation - Celera Genomics Group
Customer Matrix

	Gene Discovery				Genomic-derived Therapeutics				Gene Expression		
Customer	Celera	Incyte	Genset	Genome Therapeutics	Human Genome Sciences	Millenium	Myriad	Hyseq	CuraGen	Gene Logic	Number
Novartis	X	X		X	X		X				5
Smith Kline Beecham		X	X		X			X		X	5
Bayer		X		X		X	X				4
Hoffman La Roche		X			X	X			X		4
Pharmacia & Upjohn	X	X	X		X						4
Schering-Plough		X		X	X		X				4
AHP			X			X				X	3
Eli Lilly		X				X	X				3
Genentech		X			X				X		3
Hoechst		X		X						X	3
Monsanto		X				X	X				3
Pfizer	X	X				X					3
Abbott		X	X								2
Astra				X		X					2
BMS				X		X					2
Glaxo		X							X		2
J&J		X	X								2
Merck				X	X						2
Organon		X								X	2
Pioneer Hi-bred					X				X		2
RPR	X	X									2
Schering AG		X					X				2
Synthlabo			X		X						2
Zeneca		X								X	2
Amgen	X										1
Ariad		X									1
BASF		X									1
Becton Dickinson						X					1
Biogen									X		1
BioMerieux				X							1
Chiron								X			1
Dow Chemicals				X							1
DuPont									X		1
Japan Tobacco										X	1
Millenium		X									1
Novo Nordisk		X									1
Pasteur Merieux Connaught					X						1
Procter & Gamble										X	1
Rhobio	X										1
Takeda					X						1
Wyeth						X					1
Number of agreements	6	22	6	9	11	10	6	2	6	7	85
Number by Segment			43			29			13		85

Source: Companies financials; William Blair & Company, L.L.C. estimates

In table 13, on the following page, we contrast Celera's focused strategy with key competitors and related companies. We have segmented those into organizations involved predominantly in gene discovery, genomic-derived therapeutics and gene expression. The competitor Incyte deserves discussion. It is an older organization that began as a company to identify specific disease-related genes for partners, similar to Millenium and Human Genome Sciences, although these later two organizations intend to transform themselves into pharmaceutical firms. Incyte has moved beyond that original focused vision into supplying contract sequencing services, microarray technology and "shrink-wrapped" bioinformatic software. Of particular note is Incyte's entry into the genome database arena. It now provides a number of partially complete genome products that we believe are based, at least in part, on an earlier technical approach. This approach uses cDNA (complementary DNA) that is made from the (mRNA) template used to transcribe a protein. As such it lacks a number of DNA structures found in place in a chromosome that may be involved in DNA regulation, transcription or other function. In addition, it may be difficult to locate the position of these types of sequences appropriately in the genome structure. In other words, valuable information is likely to be lost for both these reasons. In contrast, Celera is using the newer, whole-genome shotgun sequencing approach discussed in Appendix B. We believe that this approach captures all the information needed in a very high-quality manner and should provide for the first time the necessary, complete reference upon which other analysis can be performed.

Table 13
PE Corporation - Celera Genomics Group
Competitive Strategies

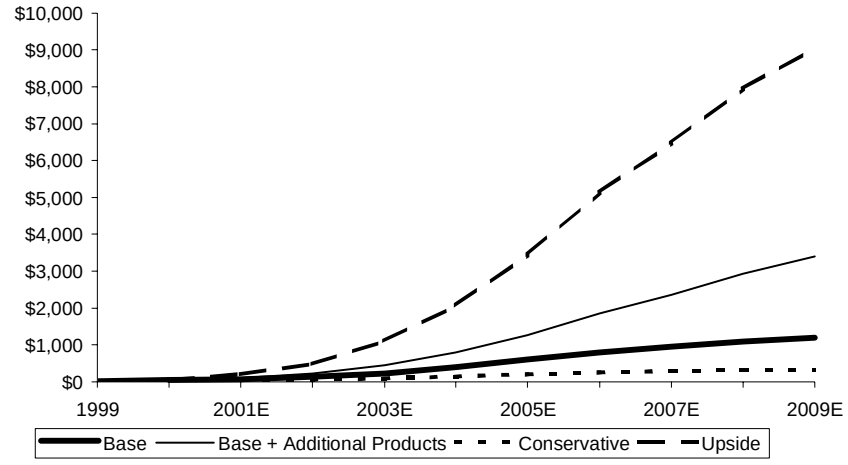
	Gene Discovery			Genomic-derived Therapeutics		Gene Expression	
	PE Celera	**Incyte**	**Genset**	**Millenium**	**Human Genome Sciences**	**Gene Logic**	**Curgagen**
Aim	To become the definitive resource of genomic information through: compilation of whole genome databases and development of analytical tools	To provide genomic databases, contract sequencing, bioinformatics, and microarray tools to researchers	To develop SNP maps for use in gene discovery and pharmecogenomics	To develop therapeutic candidates using genomic information and methodology for customer and own pipeline	To develop therapeutic candidates using genomic information and methodology for customer and own pipeline	To provide gene expression and discovery services utilizing proprietary methods, as well as bioinformatics solutions	To provide gene expression and discovery services utilizing proprietary methods
Database compilation product	Yes	Yes	Yes	No	No	Yes	Yes
Gene discovery services	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IP importance	Low	Moderate	Moderate	High	High	Moderate	Moderate
Mkt. Cap. ($ in millions)	$1,855	$960	$270	$3,845	$2,840	$290	$635
Revenue ($ in millions)	$12.5	$147.6	$29.2	$186.5	$26.1	$12.6	$16.6
# of Collaborators	6	22	6	10	11	7	6

Source: Interviews; various company financials and literature; William Blair & Company, L.L.C. estimates

Remarkable Growth Opportunities

By building a definitive resource for genomic information, Celera should be able to provide various knowledge-based products to a variety of customers at prices that create substantial revenue. The base business case used in our financial model would create expected revenue of $400 million in 5 years or 19% market share and $1.2 billion in 10 years, as shown in figure 13, and table 14 on the following page. Adding foreseeable products, projects, and value from intellectual property could lead to revenue of $3.4 billion over the same time period. If the various products are more successful than we envision, the annual revenue could be as high as $9 billion in 10 years. Our most conservative estimate would yield revenue of $140 million in five years, or 7% market share, and $320 million in ten years. As the business model we envision is one of selling knowledge, there is considerable operating leverage that could generate even larger profits. Furthermore, if the more conservative case occurs, we believe that the company could make significant reductions in expenses, especially R&D, to preserve profitability.

Figure 13
PE Corporation - Celera Genomics Group
Revenue Scenarios



Source: William Blair & Company, L.L.C. estimates

There also are other factors that support the basic model shown in figure 14, on the following page. However, the factors also could produce quantitative upside not included in our model or sensitivity analysis. Aside from establishing the basic foundation and interfaces for genomic products, we believe that Celera will build and even create significant knowledge of many aspects of molecular biology. In addition, it should be able to create new products and competitive barriers through thoughtful use of information technology. Lastly, the company should continue to build an intellectual property estate that, depending on how patent laws evolve, could create even more value. We have chosen to be conservative regarding patents in all our modeling, as this is the most difficult factor to predict. There probably are other unforeseen applications as well.

Table 14
PE Corporation - Celera Genomics Group
Revenue Models

Base Case	1999	2000E	2001E	2002E	2003E	2004E	2005E	2006E	2007E	2008E	2009E
Human Genome	$3	$23	$42	$80	$130	$186	$226	$265	$292	$321	$353
SNP	0	0	4	18	49	109	179	237	289	318	350
Comparative	0	0	1	5	20	54	119	196	259	316	348
Ag.	5	5	5	7	16	35	57	76	92	102	112
Other*	5	10	13	13	15	16	18	19	21	23	26
Total Base	$13	$38	$65	$124	$230	$399	$599	$793	$954	$1,081	$1,189

	1999	2000E	2001E	2002E	2003E	2004E	2005E	2006E	2007E	2008E	2009E
Base	$13	$38	$65	$124	$230	$399	$599	$793	$954	$1,081	$1,189
Incremental Projects	0	6	17	49	108	213	359	544	749	957	1,150
Other Applications	0	0	0	0	0	21	82	220	426	671	909
IP Total	0	0	22	42	105	157	232	295	225	215	154
Total	$13	$44	$105	$215	$442	$791	$1,273	$1,852	$2,354	$2,924	$3,402

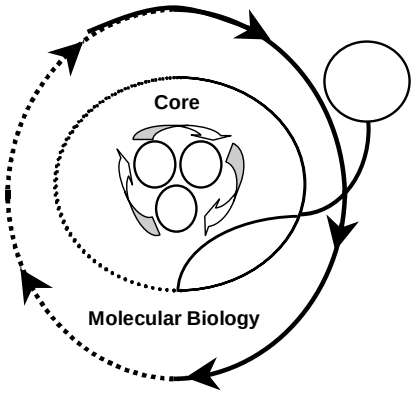
Revenue	1999	2000E	2001E	2002E	2003E	2004E	2005E	2006E	2007E	2008E	2009E
Base	$13	$38	$65	$124	$230	$399	$599	$793	$954	$1,081	$1,189
Base + Additional Products	$13	$30	$105	$215	$442	$791	$1,273	$1,852	$2,354	$2,924	$3,402
Conservative	$13	$25	$32	$53	$87	$142	$201	$252	$287	$308	$320
Upside	$13	$43	$210	$483	$1,104	$2,070	$3,448	$5,135	$6,476	$7,947	$9,047

Year-over-year	1999	2000E	2001E	2002E	2003E	2004E	2005E	2006E	2007E	2008E	2009E
Base		203%	72%	90%	86%	74%	50%	32%	20%	13%	10%
Base + Additional Products		143%	243%	106%	105%	79%	61%	45%	27%	24%	16%
Conservative		98%	27%	68%	64%	63%	42%	25%	14%	7%	4%
Upside		246%	385%	130%	129%	87%	67%	49%	26%	23%	14%

* Other includes IP licenses, royalties, e-commerce, and other opportunities

Source: Various company financials; Interviews; PhRMA; BCG; McKinsey; Frost & Sullivan; William Blair & Company, L.L.C. estimates

Figure 14
PE Corporation - Celera Genomics Group
Celera Knowledge Creation Cycle



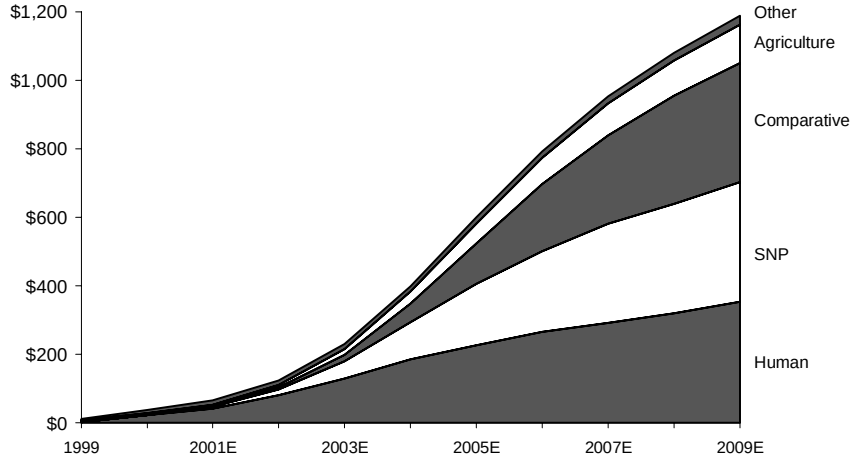
Source: Industry interviews; William Blair & Company, L.L.C. analysis

Base Business Case

Our base business case, which leads to $1.2 billion in revenue in 2009, is built around five major product lines to five target customer groups, as shown in figure 15. Of the five product lines, we envision four will be database products—human genome, human SNPs, comparative organism genomes, and agricultural genomes—and the fifth will be services such as specific gene discovery or pharmacogenomic programs. We foresee the completed human genome database as comprehensively annotated to identify all genes, their locations, their known function (from literature or Celera's own "wet" research), or their computationally derived function. We see the SNP database as having two components, one with evenly spaced SNPs that could be used for linkage studies as the competitor

Genset is doing, and another with SNPs that affect gene expression or protein function, again derived from literature, Celera's "wet" lab, or computationally. This information could be used in pharmacogenomics, for example. The last database for comparative genomics across many organisms may be the most exciting. In addition to the company's own sequencing activity for humans, Drosophila, the mouse, and rice, as well as public genomes, we believe that the company likely could become the distribution and analysis vehicle for genomes developed by other third parties. We conclude this as Celera should have the most comprehensive data on the key model organisms of humans, mice, and fruit flies; unmatched computing power, algorithms, and data storage; and high-powered interfaces with the target customers. Lastly, we believe that the company will continue to provide some services through its Celera "West" organization—a combination of PE Corporation's GenScope and AgGen businesses—that has capabilities to work with actual model organisms in its "wet" lab, and combine these with the company's own databases for gene discovery.

Figure 15
PE Corporation - Celera Genomics Group
Base Revenue Model



Source: William Blair & Company, L.L.C. estimates

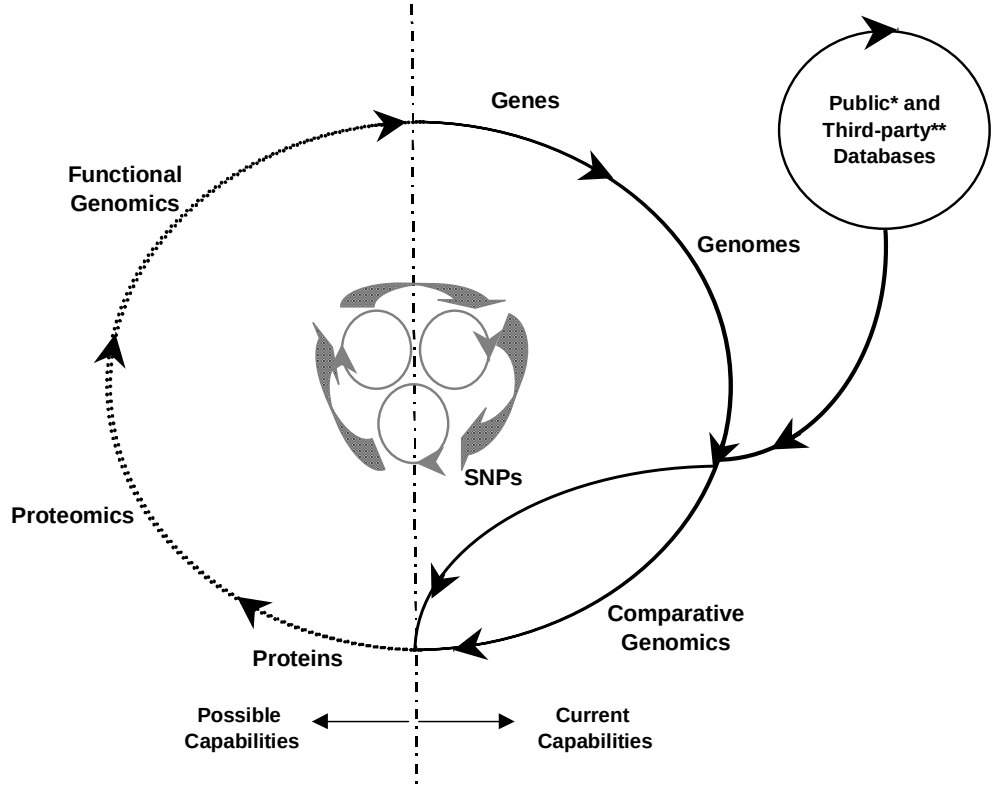
We also see five customer groups: the top 20 major pharmaceutical companies (or the top 3 similar agriculture firms); minor pharmaceutical firms (small chemical entities) such as those that specialize in specific indications or delivery; biotechnology firms, both human and agricultural based; academic and government life science researchers; and clinical diagnostic firms. Pricing likely will vary across the customer groups. We estimate that for the major pharmaceutical and agricultural firms, each database subscription will cost somewhat more than the $5 million per year paid by the early access partners, rising 10% annually for *new* subscriptions to reflect increasing value. The price compares with the hundreds of thousands to millions of dollars of profit that are lost for *every day* it takes a new drug to come to market, not including the likely benefits of improved drugs. For diagnostic, smaller pharmaceutical or biotech firms we could envision more limited or focused databases costing $500,000 to $2.5 million annually, again with 10% annual price increases for *new* subscribers. For life science researchers, we anticipate more of a per-individual subscription approach—like scientific journals or software, costing at least $300 per scientist, although low enough to inhibit cheating.

To provide the databases and services to this range of customers, Celera is building significant information technology capabilities, in our opinion. These leverageable capabilities included genome assembly programs, algorithms for computational comparisons across genomes, and query and visualization interfaces.

Building Knowledge of Molecular Biology

In addition to the products and capabilities that we foresee in the base case, we anticipate that the company will build even more knowledge in molecular biology, as shown in figures 14, 16, and 17 that could translate into revenue upside. As previously mentioned, we believe that the company could become the focal point for genomes developed by third parties. Furthermore, we believe that the company will add to its genomic capabilities. For example, as figure 16 shows, we could see logical database products based on gene expression, proteomics, or protein-oriented databases, and functional genomics or systems biology approaches. We believe that each of these databases could be sold to the same targeted customer groups at the same types of subscription rates and could lead to an additional $900 million annually by 2009. Also, by building these similar wet biology capabilities at Celera West, we estimate that the company could add another $1.2 billion in annual service revenue by 2009.

Figure 16
PE Corporation - Celera Genomics Group
Molecular Biology Knowledge

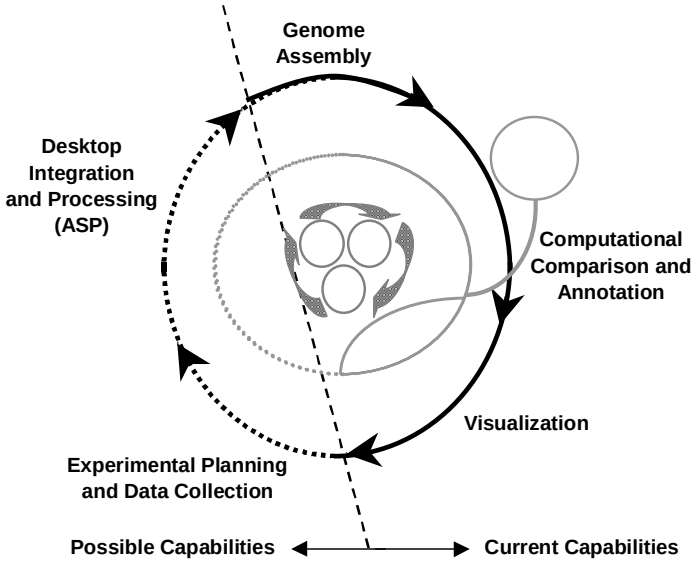


* Added to Celera system ** Provided through Celera
Source: Industry interviews; William Blair & Company, L.L.C. analysis

Information Technology Platforms, Algorithms, and Applications

To achieve the database objectives and provide the information in formats useful for customers, we believe that the company will continue to build leading-edge information technology capabilities. As shown in figure 17, these skills could be further applied to better lock in customers and provide more services. For example, Celera could add experimental planning and customer-proprietary data collection to its own tools. In addition, the company could create an integrated bioinformatics desktop, with analytical tools available on its supercomputer, in an ASP (application service provider) format—such as TurboTax by Intuit, which one can use completely through the Internet. At this point we have not modeled any incremental revenue for these IT additions in our sensitivity analysis, choosing instead to view them as a means to capture and retain customers for the database products.

Figure 17
PE Corporation - Celera Genomics Group
Information Technology Platforms/Algorithms



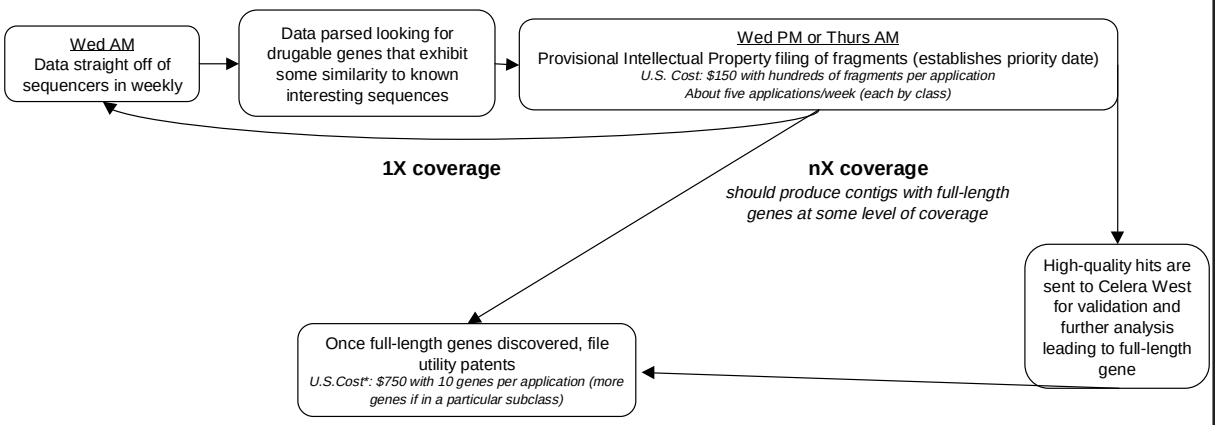
Source: Industry interviews; William Blair & Company, L.L.C. analysis

Intellectual Property Possibilities

We envision that Celera will establish intellectual property in three fields. The first, genetic sequences and use, is the most obvious, but probably the least straight forward. The second is in the area of information technology, where we believe the firm needs to innovate with algorithms and interfaces. The third relates to the factory process and equipment to perform the sequencing itself.

Genetic sequences and use. As shown in figure 18, on the following page, the company is actively filing for genetic sequence and use patents. This is done weekly for provisional patents and followed up as necessary with full-length-gene final patent applications when a discovery meets the company's priority list, as shown in table 15, on the following page. The company already has filed 1,500 novel full-gene patents for Drosophila and anticipates filing only *150 to 300 high-quality full-gene patents for humans*. To establish priority for the full-length final patents, the company files provisional patents on potentially thousands of gene fragments. Promising gene fragments are either completed by Celera West using its biological wet lab or through additional coverage from the normal sequencing process. Because of the various issues regarding genetic patents, *we have included essentially no IP revenue in our base business model*. However, we estimate that this type of IP could increase revenue by a peak of about $300 million annually by 2006 and contributes to our additional product scenario—which also includes the incremental database products and services mentioned previously—leading to potential revenue in 2009 of $3.4 billion, as shown in table 14 and figures 19 (on the following page) and 13. The potential genetic IP revenue is based on license fees of $1 million to $1.5 million per patent licensed, with potential milestone payments of up to $40 million on an estimated low percentage patents that are deemed "drugable." *Over the 10-year scenario horizon, we do not anticipate any royalties for actual drug sales even in the upside scenario, as no drug would likely be approved yet, although a royalty payment for sales of an approved drug could readily be part of an agreement.*

Figure 18
PE Corporation - Celera Genomics Group
Gene Patenting Process



* European cost: $2,500 per application, but $1million-$3 million to issue

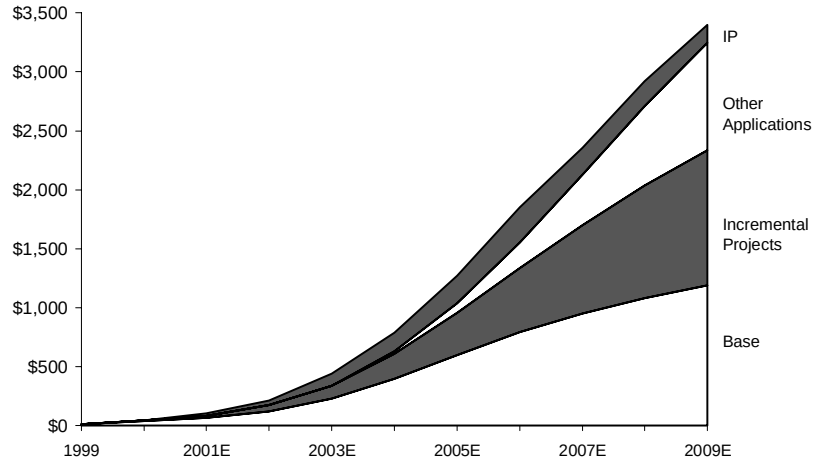
Source: Interviews; William Blair & Company, L.L.C. analysis

Table 15
PE Corporation - Celera Genomics Group
Genomic IP Priority

Value	Description
High	Use of comparative genomics to find previously unknown gene type Real long-term value Celera's strength
Middle	Orphan, open reading frame Discovered using public models and tools Elucidation of function can rapidly follw Patent status less clear
Low	Structurally obvious and easily identified Discovery validates earlier identification (e.g., by EST) Patent status clear

Source: Interviews; William Blair & Company, L.L.C. estimates

Figure 19
PE Corporation - Celera Genomics Group
Base Revenue Plus Additional Products



Source: William Blair & Company, L.L.C. estimates

As we briefly mentioned in the section on risks, there are a number of issues regarding patents and patentability of genetic information and use. Foremost is the availability of novel gene patents. Two of Celera's major competitors, Incyte and Human Genome Sciences, both claim to have filed provisional patents for virtually all the human genes. While *we do not believe this should have any effect on Celera's knowledge-based model*, the range of patents and priority of those already filed are still to be determined. The question is empirical for the specific genes: have they all been found? The question is legal for the priority: is an EST fragment found in the cell sufficient for patenting (as has been tried in many cases by competitors) or does one need a full-length gene, its function, and location? Furthermore, differences exist among regional laws. In the United States, it appears that one can file a sequence with the likely function computationally determined, and fine tune the sequence after filing. In Europe, it appears that one may need to file an *error-free* sequence with function determined biologically. In either case, Celera is set up to achieve accurate sequences not requiring fine-tuning and can perform biological work at Celera West.

Other IP: information technology and gene sequencing process. Given the need for Celera to maintain state-of-the-art capabilities in gene sequencing and IT, we believe that IP opportunities will arise in these areas. However, we have not modeled any upside even into our sensitivity analysis. We believe that gene-sequencing IP easily could be licensed to PE Biosystems, and if substantive, PE Biosystems would be able to capitalize well on these inventions and provide a royalty stream back to Celera. Leveraging information technology IP would be more problematic, but at least should help build barriers for the company's own offerings and may be of interest to partners such as Compaq and Oracle.

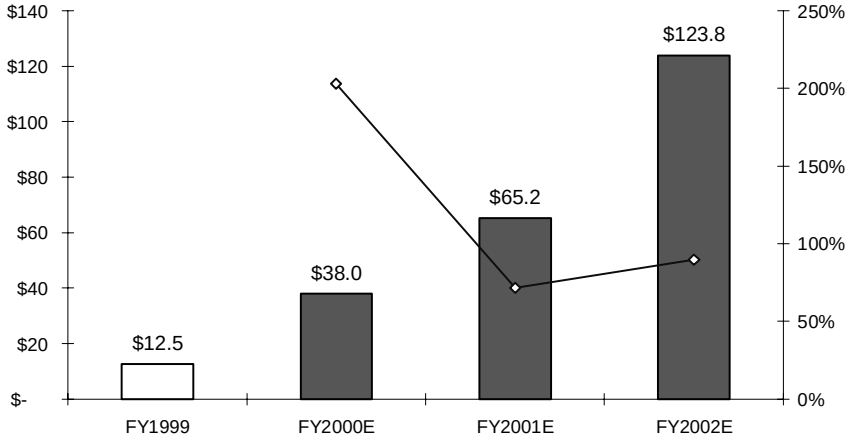
100% Compounded Annual Revenue Growth Expected

Celera should experience more than 100% compounded annual revenue growth over the next few years. This growth should be driven by several factors, including exponential revenue growth and plateauing R&D and SG&A expense costs. In addition, this growth could be accelerated by a rapid and wide acceptance of Celera's product offering, as well as the potential windfall in royalties and IP licenses that may result from genomics-based research.

Revenue Should Grow 145% Compounded Annually Through 2002

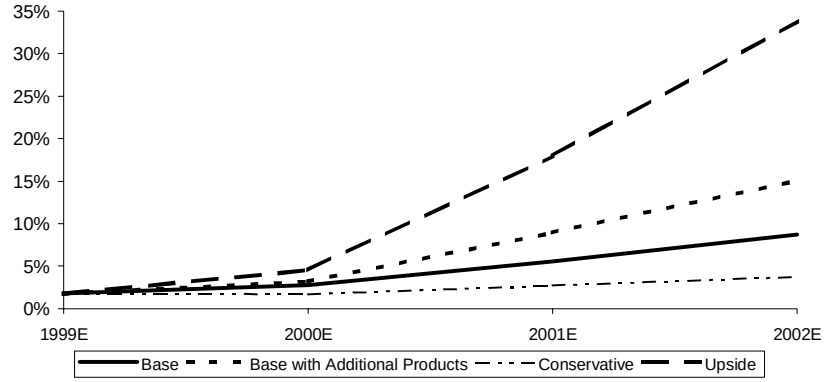
Additional customers and new product revenues should drive Celera's revenue growth. The market for molecular biology databases and related analysis is estimated to be $715 million globally in 1999, forecast to grow at least 24% annually for the foreseeable future. This growth should be driven by an increasing number of customers and wider acceptance of this information in existing and new applications. We expect the company's revenue to grow 145% compounded annually over the next few years, as shown in figure 20, on the following page, increasing to $124 million in fiscal 2002, from $12.5 million in fiscal 1999. This would represent 9% market share of the $1.4 billion market for these products in 2002, as shown in figure 21, on the following page.

Figure 20
PE Corporation - Celera Genomics Group
Annual Revenue and Growth
($ in millions)



Source: Company financials; William Blair and Company, L.L.C. estimates

Figure 21
PE Corporation - Celera Genomics Group
Market Share Growth



Source: Various company financials; The Scientist; Frost and Sullivan; Theta Corporation; Business Communications Company; Life Tech; ABRF; Genetic Engineering News; Phrma; Instrument Business Outlook; Phortech; William Blair & Company, L.L.C. estimates

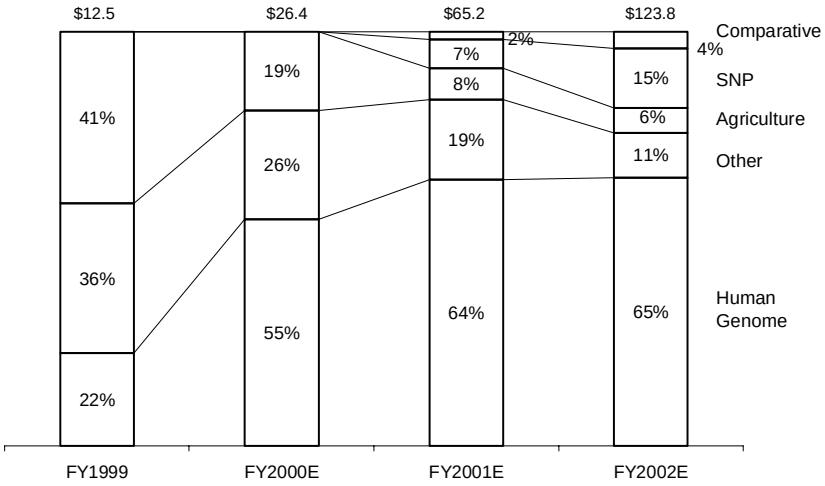
The company derived the bulk of its revenue in 1999 from several research collaborations, including the $18 million three-year joint effort with RhoBio S.A. to discover genes in maize using expression analysis. Additional alliances such as with Rhone-Poulenc Rorer for the discovery of “drugable” targets for asthma, cancer, and cardiovascular disorders should contribute an expected $13.3 million by 2002. Combined, agriculture and other products contributed 77% of revenue in 1999, yet are expected to decline to 17% in 2002 as revenue from the database products increases, as shown in table 16 and figure 22.

Table 16
PE Corporation - Celera Genomics Group
Product Line Sales
($ in millions)

Revenue	FY1999	**FY2000E**	**FY2001E**	**FY2002E**
Human Genome Total	$2.8	**$22.8**	**$41.9**	**$80.2**
SNP Total	$0.0	**$0.0**	**$4.5**	**$18.0**
Universal Total	$0.0	**$0.0**	**$1.2**	**$4.9**
Ag.	$5.2	**$5.0**	**$5.0**	**$7.4**
Other	$4.5	**$10.2**	**$12.6**	**$13.3**
Total Base	$12.5	**$38.0**	**$65.2**	**$123.8**
Year-over-year growth	FY1999	**FY2000E**	**FY2001E**	**FY2002E**
Human Genome Total		**715%**	**84%**	**92%**
SNP Total				**300%**
Universal Total				**300%**
Ag.		**-4%**	**0%**	**48%**
Other		**124%**	**24%**	**5%**
Total Base		**203%**	**72%**	**90%**
100% of Revenue	FY1999	**FY2000E**	**FY2001E**	**FY2002E**
Human Genome Total	22%	**60%**	**64%**	**65%**
SNP Total			**7%**	**15%**
Universal Total			**2%**	**4%**
Ag.	41%	**13%**	**8%**	**6%**
Other	36%	**27%**	**19%**	**11%**
Total Base	100%	**100%**	**100%**	**100%**

Source: Company financials; William Blair & Company, L.L.C. estimates

Figure 22
PE Corporation - Celera Genomics Group
Annual Revenue Mix
($ in millions)



Source: Company financials; William Blair & Company, L.L.C. estimates

Revenue from the keystone Human Genome database is expected to grow more than 200% compounded annually over the next few years, increasing to more than $80 million in 2002, from only $2.8 million in 1999, as shown in table 16. Initial revenue for this product will come from Celera's three early access partners, Amgen, Novartis, and Pharmacia UpJohn, each of which has agreed to five-year, $5 million annual access subscriptions that will run through 2003. These customers are receiving discounted pricing as they help Celera define the product offering and work through the issues encountered early in the sequencing process. These customers' agreements may also allow them access to other databases, which include the Celera Human Gene Index and the Celera Drosophila database, yet their subscriptions do not include access to future offerings such as the SNP and comparative databases. Celera has also recently signed an agreement with Pfizer for access to its full-range of current database products, as well genomic services. These genomic services, for which the company should receive milestone payments and royalties, are intended to help Pfizer find novel genes and proteins for use as drug targets. In addition, some of the proteins may be useful as drugs themselves. The success of the Celera Human Genome database will hinge on additional subscribers from the pharmaceutical and biotechnology industry, as well as enlisting academic communities. Prices for later subscribers will increase beyond the $5 million per year offered to early access partners, while a unique pricing scheme is being developed for academic users to foster its acceptance as an adjunct to the public resources currently available. We anticipate that the high-quality databases and the powerful analysis tools being developed will increase the appeal of this product for both groups of users.

Beyond the Human Genome, we expect Celera to create additional database products such as a SNP database, as well as additional genomic databases of various other organisms that will be indispensable to extract the complete value of the Human Genome and find applications in areas such as basic research and agriculture. Work on these projects already has begun, and revenue is expected in new subscriptions or add-ons in 2001, contributing roughly $23 million in 2002.

Operating Expenses Are Expected to Level Off by 2002

We expect operating expense as a percentage of revenue to decline steadily as revenues rapidly increase and cost growth slows, as shown in table 17. R&D will continue to increase as these expenditures translate directly into the company's information and service products. Growth should slow as headcount reaches desired levels of higher than 400, and the company continues to achieve efficiencies from its factory approach. SG&A expense growth should rapidly decline as the sales and support groups reach critical mass and the company becomes fully self sufficient from PE Corporation in these areas, which explains the large increase in 1999 as a result of the company's formal founding and recapitalization.

Tax benefits should remain stable at 36% for the foreseeable future, as losses from Celera offset profit from PE Biosystems at the PE Corporation level. We expect PE Biosystems to maintain profits sufficient to be able to utilize the Celera losses for the foreseeable future.

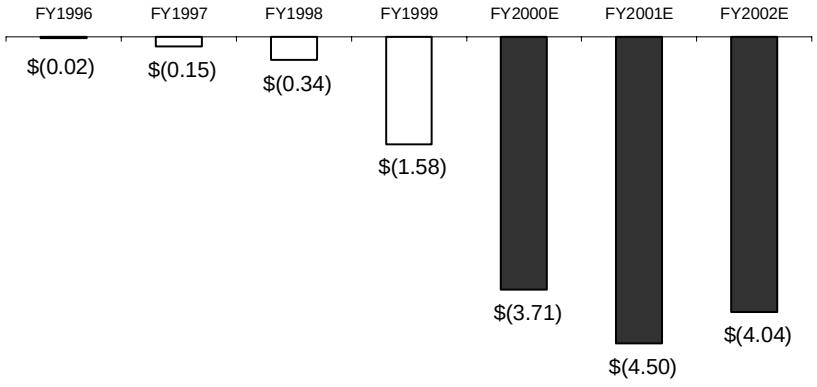
Net Losses and Losses per Share to Increase as Investment in Future Growth Continues

The continued investment in R&D to support development of the full range of products will lead Celera to experience increasing net losses at first, peaking in 2001 at $119 million, and then expected to decline to $109 million in 2002. As a result, earnings per share should decline until 2001 and then improve, as shown in figure 23.

Table 17
PE Corporation - Celera Genomics Group
Annual Income Statement
($ in millions)

Fiscal year ends June 30	1998	1999	**2000E**	**2001E**	**2002E**
Net Revenues	$4,211	$12,541	**$37,956**	**$65,227**	**$123,795**
Costs and Expenses					
R&D	10,279	48,448	**$151,653**	**$199,298**	**$242,249**
SG&A	6,725	27,194	**$44,148**	**$54,644**	**$57,377**
Total Costs and Expenses	17,004	75,642	**195,801**	**253,943**	**299,625**
Operating Loss	(12,793)	(63,101)	**(157,846)**	**(188,715)**	**(175,830)**
Interest Income		1,245	**7,521**	**2,567**	**5,441**
Loss Before Income Taxes	(12,793)	(61,856)	**(150,324)**	**(186,149)**	**(170,390)**
Benefit for Income Taxes	4,478	22,268	**54,109**	**67,014**	**61,340**
Net Loss	($8,315)	($39,588)	**(96,216)**	**(119,135)**	**(109,049)**
EPS	($0.34)	($1.58)	**($3.71)**	**($4.50)**	**($4.04)**
Shares Outstanding	24,280	25,100	**25,927**	**26,449**	**26,982**
Year-over-year Growth					
Revenue	366.3%	197.8%	**202.7%**	**71.9%**	**89.8%**
Operating Expenses	174.1%	344.8%	**140.8%**	**29.7%**	**18.0%**
Operating Income	141.3%	437.6%	**129.5%**	**19.6%**	**-6.8%**
Net Loss	241.3%	539.9%	**214.3%**	**123.8%**	**91.5%**
EPS	136.1%	422.3%	**107.5%**	**21.4%**	**-10.3%**
100% of Revenue	1998	1999	**2000E**	**2001E**	**2002E**
Net Revenues	100.0%	100.0%	**100.0%**	**100.0%**	**100.0%**
Costs and Expenses					
R&D	244.1%	386.3%	**399.6%**	**305.5%**	**195.7%**
SG&A	159.7%	216.8%	**116.3%**	**83.8%**	**46.3%**
Total Costs and Expenses	403.8%	603.2%	**515.9%**	**389.3%**	**242.0%**
Operating Loss	-303.8%	-503.2%	**-415.9%**	**-289.3%**	**-142.0%**
Interest Income	0.0%	9.9%	**19.8%**	**3.9%**	**4.4%**
Loss Before Income Taxes	-303.8%	-493.2%	**-396.1%**	**-285.4%**	**-137.6%**
Benefit for Income Taxes	-35.0%	177.6%	**-36.0%**	**-36.0%**	**-36.0%**
Net Loss	-197.5%	-315.7%	**-253.5%**	**-182.6%**	**-88.1%**

Figure 23
PE Corporation - Celera Genomics Group
Annual EPS



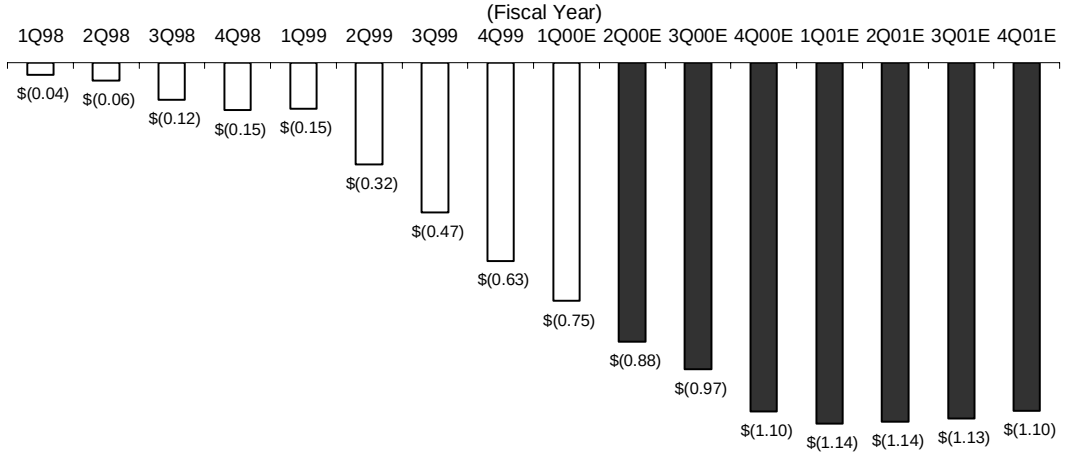
Source: Company financials; William Blair & Company, L.L.C. estimates

Our quarterly estimates are shown in table 18 and figure 24. Revenues and expenses do not yet exhibit any signs of seasonality, and are expected to grow steadily as customers are added and operations continue.

Table 18
PE Corporation - Celera Genomics Group
Quarterly Income Statement
($ in millions)

Fiscal year ends June 30	Q199	Q299	Q399	Q499	Q100	Q200E	Q300E	Q400E	Q101E	Q201E	Q301E	Q401E
Net Revenues	$3,916	$1,715	$1,798	$5,112	$8,300	**$8,715**	**$10,215**	**$10,726**	**$11,415**	**$14,676**	**$17,938**	**$21,199**
Costs and Expenses												
R&D	4,677	8,260	13,300	22,211	32,200	**35,742**	**39,674**	**44,038**	**46,240**	**48,552**	**50,979**	**53,528**
SG&A	4,792	5,814	6,988	9,600	8,400	**10,800**	**11,880**	**13,068**	**13,098**	**13,737**	**14,088**	**13,721**
Total Costs and Expenses	$9,469	$14,074	$20,288	$31,811	$40,600	**$46,542**	**$51,554**	**$57,106**	**$59,337**	**$62,289**	**$65,067**	**$67,250**
Operating Loss	(5,553)	(12,359)	(18,490)	(26,699)	(32,300)	**(37,827)**	**(41,339)**	**(46,380)**	**(47,923)**	**(47,612)**	**(47,130)**	**(46,051)**
Interest Income			107	1,138	2,000	**2,194**	**1,976**	**1,352**	**1,066**	**784**	**500**	**216**
Loss Before Income Taxes	(5,553)	(12,359)	(18,383)	(25,561)	(30,300)	**(35,633)**	**(39,363)**	**(45,028)**	**(46,856)**	**(46,829)**	**(46,630)**	**(45,834)**
Benefit for Income Taxes	1,944	4,325	6,435	9,564	10,900	**12,828**	**14,171**	**16,210**	**16,868**	**16,858**	**16,787**	**16,500**
Net Loss	($3,609)	($8,034)	($11,948)	($15,997)	($19,400)	**($22,805)**	**($25,192)**	**($28,818)**	**($29,988)**	**($29,970)**	**($29,843)**	**($29,334)**
EPS	($0.15)	($0.32)	($0.47)	($0.63)	($0.75)	**($0.88)**	**($0.97)**	**($1.10)**	**($1.14)**	**($1.14)**	**($1.13)**	**($1.10)**
Shares Outstanding	24,689	24,942	25,254	25,507	25,733	**25,862**	**25,991**	**26,121**	**26,252**	**26,383**	**26,515**	**26,647**
Year-over-year Growth												
Revenue	247.2%	-56.2%	4.8%	184.3%	62.4%	**5.0%**	**17.2%**	**5.0%**	**6.4%**	**28.6%**	**22.2%**	**18.2%**
Operating Expenses	38.8%	48.6%	44.2%	56.8%	27.6%	**14.6%**	**10.8%**	**10.8%**	**3.9%**	**5.0%**	**4.5%**	**3.4%**
Operating Income	-2.5%	122.6%	49.6%	44.4%	21.0%	**17.1%**	**9.3%**	**12.2%**	**3.3%**	**-0.6%**	**-1.0%**	**-2.3%**
Net Loss	97.5%	222.6%	148.7%	133.9%	121.3%	**117.6%**	**110.5%**	**114.4%**	**104.1%**	**99.9%**	**99.6%**	**98.3%**
EPS	-2.9%	120.4%	46.9%	32.6%	20.2%	**17.0%**	**9.9%**	**13.8%**	**3.5%**	**-0.6%**	**-0.9%**	**-2.2%**
100% of Revenue	**Q1 99**	**Q2 99**	**Q3 99**	**Q4 99**	**Q1 00**	**Q2 00E**	**Q3 00E**	**Q4 00E**	**Q1 01E**	**Q2 01E**	**Q3 01E**	**Q4 01E**
Net Revenues	100.0%	100.0%	100.0%	100.0%	100.0%	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**
Costs and Expenses												
R&D	119.4%	481.6%	739.7%	434.5%	388.0%	**410.1%**	**388.4%**	**410.6%**	**405.1%**	**330.8%**	**284.2%**	**252.5%**
SG&A	122.4%	339.0%	388.7%	187.8%	101.2%	**123.9%**	**116.3%**	**121.8%**	**114.7%**	**93.6%**	**78.5%**	**64.7%**
Total Costs and Expenses	241.8%	820.6%	1128.4%	622.3%	489.2%	**534.0%**	**504.7%**	**532.4%**	**519.8%**	**424.4%**	**362.7%**	**317.2%**
Operating Loss	-141.8%	-720.6%	-1028.4%	-522.3%	-389.2%	**-434.0%**	**-404.7%**	**-432.4%**	**-419.8%**	**-324.4%**	**-262.7%**	**-217.2%**
Interest Income	0.0%	0.0%	6.0%	22.3%	24.1%	**25.2%**	**19.3%**	**12.6%**	**9.3%**	**5.3%**	**2.8%**	**1.0%**
Loss Before Income Taxes	-141.8%	-720.6%	-1022.4%	-500.0%	-365.1%	**-408.9%**	**-385.3%**	**-419.8%**	**-410.5%**	**-319.1%**	**-260.0%**	**-216.2%**
Benefit for Income Taxes	-35.0%	-35.0%	-35.0%	-37.4%	-36.0%	**-36.0%**	**-36.0%**	**-36.0%**	**-36.0%**	**-36.0%**	**-36.0%**	**-36.0%**
Net Loss	-92.2%	-468.5%	-664.5%	-312.9%	-233.7%	**-261.7%**	**-246.6%**	**-268.7%**	**-262.7%**	**-204.2%**	**-166.4%**	**-138.4%**

Figure 24
PE Corporation - Celera Genomics Group
Quarterly EPS
(Fiscal Year)



Source: Company financials; William Blair & Company, L.L.C. estimates

Celera's Balance Sheet and Cash Flow Should Support Operations

Celera maintains a healthy balance sheet and a strong cash position, as shown in tables 19 and 20. The company will receive the remainder of the originally allocated $300 million that remains from the recapitalization in the form of a $150 note from PE Biosystems in fiscal 2000. The completion of the facility build-out in 1999 should curtail capital expenditures at levels between $25 million to $40 million annually. The company's continued spending in support of operations

should result in a significant cash burn, leading the company to access the capital markets by 2002. We have modeled a $200 million debt offering for 2002, as this appears to be the preferred method to avoid dilution and increase PE Corporation's overall tax shield.

Table 19
PE Corporation - Celera Genomics Group
Balance Sheet
($ in millions)

Fiscal year ends June 30	1999	2000E	2001E	2002E
Assets				
Current assets				
Cash	$71,491	**$135,710**	**$22,166**	**$119,324**
Accounts receivable	3,276	**2,725**	**5,424**	**10,329**
Note receivable from the PE Biosystems group	150,000			
Tax benefit receivable from the PE Biosystems group	9,935	**-**	**-**	**-**
Prepaid expenses and other current assets	3,454	**2,464**	**2,986**	**3,614**
Total Current Assets	$238,156	**$140,900**	**$30,576**	**$133,268**
Property and equipment	$109,700	**$138,494**	**$171,569**	**$208,544**
Accumulated depreciation and amortization	5,508	**25,984**	**57,707**	**96,530**
Property and equipment, net	104,192	**112,509**	**113,862**	**112,014**
Other long-term assets	2,372	**2,883**	**3,505**	**4,260**
Total Assets	$344,720	**$256,292**	**$147,943**	**$249,541**
Liabilities and Group Equity				
Current liabilities				
Accounts payable	$19,861	**$31,275**	**$37,361**	**$44,694**
Accrued salaries and wages	4,179	**7,138**	**8,406**	**9,934**
Deferred Revenues	12,032	**1,073**	**2,120**	**4,023**
Other accrued expenses	9,281	**13,705**	**16,140**	**19,073**
Total current liabilities	$45,353	**$53,191**	**$64,027**	**$77,724**
Other long-term liabilities	5,500	**5,450**	**5,400**	**202,350**
Total Liabilities	$50,853	**$58,641**	**$69,427**	**$280,074**
Commitments and contingencies				
Group equity (deficit)	$293,867	**$197,651**	**$78,516**	**($30,533)**
Total Liabilities and Group Equity	$344,720	**$256,292**	**$147,943**	**$249,541**

Table 20
PE Corporation - Celera Genomics Group
Statement of Cash Flows
($ in millions)

Fiscal year ends June 30	1999	2000E	2001E	2002E
Operating Activities				
Net loss	($44,894)	**($96,216)**	**($119,135)**	**($109,049)**
Adjustments to net cash used by operating activities				
Depreciation and amortization	3,757	**20,476**	**31,723**	**38,823**
Long-term compensation programs	2,802			
Changes in operating assets and liabilities				
(Increase) decrease in accounts receivable	(2,520)	**551**	**(2,699)**	**(4,905)**
Increase in tax benefit receivable from PEB	(9,935)	**9,935**	**0**	**0**
(Increase) decrease in prepaid expenses and other assets	(3,458)	**990**	**(521)**	**(628)**
Increase (decrease) in accounts payable and other liabilities	31,496	**7,838**	**10,836**	**13,697**
Net Cash Used in Operating Activities	($22,752)	**($56,426)**	**($79,797)**	**($62,062)**
Investing Activities				
Additions to property and equipment	($94,541)	**($28,794)**	**($33,076)**	**($36,974)**
Acquisitions and investments, net	(1,236)	**(511)**	**(621)**	**(755)**
Net Cash Used By Investing Activities	($95,777)	**($29,305)**	**($33,697)**	**($37,730)**
Financing Activities				
Net cash allocated from the PEB group	$188,535	**$150,000**	**$0**	**$0**
Increases in long-term liabilities		**(50)**	**(50)**	**196,950**
Proceeds from stock issued for group stock plans	1,485			
Net Cash Provided By Financing Activities	$190,020	**$149,950**	**($50)**	**$196,950**
Net Change in Cash and Cash Equivalents	$71,491	**$64,219**	**($113,544)**	**$97,158**
Cash and Cash Equivalents Beginning of Period	$0	**$71,491**	**$135,710**	**$22,166**
Cash and Cash Equivalents End of Period	$71,491	**$135,710**	**$22,166**	**$119,324**

Premium Valuation Justified for Unprecedented Opportunity

The market that the company hopes to address is potentially enormous, well into the billions of dollars, and the impact of its operations and products could be far-reaching. However, Celera's near-term valuation demands some consideration. Genomics is a technology- and R&D-intensive industry. Consequently, we believe that metrics related to the relative value of technology should help investors assess valuations within the cohort of genomics companies. For example, Celera is trading at a technology value (market capitalization minus cash) of about $1.9 billion, almost 12 times its projected R&D spend for fiscal 2000, as shown in table 21. This is more than the median ratio of 8 for other pure-play knowledge based firms performing gene sequencing or gene expression analysis. Nonetheless, it is about half that for knowledge-based firms that have a pharmaceutical component. In addition to its knowledge database and interface, Celera intends to provide additional *genomic services* to assist its customers in applying the knowledge, which we believe should elevate its valuation from that of the pure-play, knowledge-based cohort closer to the other cohort.

Table 21
PE Corporation - Celera Genomics Group
Comparative Company Valuation Table

Company	Symbol	Price 12/6/99	Mkt. Cap ($ in mil.)	Revenue	LTM R&D Spend*	Cash ($ in mil.)	Technology Value (TV)**	R&D/ Rev	TV / Rev	TV / R&D	EPS*** 1998	1999	2000E	2001E
Knowledge Firms														
Sequencing														
PE Celera	CRA	$72.75	$1,855.4	$12.5	$152.0	$80.0	$1,775.4	1216.0%	142.0	11.7	($0.34)	($1.58)	($3.71)	($4.50)
Incyte	INCY	$33.69	$958.6	$147.6	$131.9	$82.1	$876.5	89.3%	5.9	6.6	$0.75	($0.96)	($0.67)	($0.01)
Genset	GENXY	$12.00	$268.5	$29.2	$44.9	$34.2	$234.3	154.0%	8.0	5.2	($0.76)	($1.01)	($0.64)	$0.41
Expression														
Curagen	CRGN	$47.06	$632.8	$12.6	$24.3	$43.3	$589.5	193.1%	46.9	24.3	($1.47)	($1.64)	($1.48)	($0.85)
Gene Logic	GLGC	$14.63	$290.4	$16.6	$26.9	$20.5	$269.9	161.9%	16.3	10.0	($0.62)	($0.89)	($0.35)	$0.15
Genome Therapeutics	GENE	$6.63	$122.3	$24.0	$26.7	$32.9	$89.3	111.2%	3.7	3.3	($0.87)	($0.34)	($0.22)	($0.19)
Pharma / knowledge firms														
Millenium Pharmaceuticals	MLNM	$104.88	$3,843.0	$186.5	$147.0	$225.2	$3,617.9	78.8%	19.4	24.6	$0.33	($0.02)	$0.03	$0.74
Human Genome Sciences	HGSI	$123.63	$2,840.8	$26.1	$55.9	$288.6	$2,552.2	214.5%	98.0	45.7	($1.03)	($1.74)	($2.16)	($1.89)
Myriad	MYGN	$36.50	$344.2	$25.3	$23.4	$11.3	$332.8	92.5%	13.1	14.2	($1.05)	($1.06)	($0.90)	($0.32)
Hyseq	HYSQ	$8.44	$110.0	$8.0	$19.4	$34.8	$75.2	241.4%	9.4	3.9	($1.27)	($1.44)	($1.03)	
Total	Mean		$1,126.6	$48.8	$65.2	$85.3	$1,041.3	255.3%	36.3	15.0	$ (0.63)	$ (1.07)	$ (1.11)	$ (0.72)
	Median		$488.5	$24.7	$35.9	$39.1	$461.1	157.9%	14.7	10.9	$ (0.82)	$ (1.04)	$ (0.79)	$ (0.19)
	Total / Weighted		$11,265.8	$488.4	$652.3	$852.9	$10,412.9	133.6%	21.3	16.0	-	-	-	-
Knowledge Firms	Mean		$688.0	$40.4	$67.8	$48.8	$639.1	320.9%	37.2	10.2	$ (0.55)	$ (1.07)	$ (1.18)	$ (0.83)
	Median		$461.6	$20.3	$35.9	$38.8	$429.7	157.9%	12.1	8.3	$ (0.69)	$ (0.99)	$ (0.66)	$ (0.10)
	Total / Weighted		$4,127.9	$242.4	$406.6	$293.0	$3,834.9	167.7%	15.8	9.4	-	-	-	-
Pharma / knowledge firms	Mean		$1,784.5	$61.5	$61.4	$140.0	$1,644.5	156.8%	35.0	22.1	$ (0.76)	$ (1.07)	$ (1.02)	$ (0.49)
	Median		$1,592.5	$25.7	$39.7	$130.0	$1,442.5	153.5%	16.3	19.4	$ (1.04)	$ (1.25)	$ (0.97)	$ (0.32)
	Total / Weighted		$7,138.0	$245.9	$245.7	$559.9	$6,578.0	99.9%	26.7	26.8	-	-	-	-

* CRA estimated FY2000 R&D expenditure due to exponential ramp up in spending ** Technology value =Mkt. Cap - Cash ***Fiscal-year estimates

Source: First Call; FactSet; William Blair & Company, L.L.C. estimates for CRA

We believe that at its foundation, the company is a vehicle for transforming genomic data into useful knowledge. While the company has been likened to *Bloomberg* for biologists—both comprehensive and ubiquitously used—we strongly believe that this oversimplifies and under represents what the company can do to build knowledge. At the heart of the company is high-quality data generation, not just reporting. It is this quality, in addition to the quantity and utility of data generated, that should help to establish a proprietary position for the company. Competitive companies not generating all the data from scratch, such as Celera intends to do, face at least two barriers: the enormous data input task, and, as importantly, *data quality*. This high-quality data must be fed into *proprietary algorithms* and a *powerful enough computing system* that can parse it, match it with other known sequences, and annotate it to create information. The annotated genome data must finally be integrated with information on genetic variation, protein function and expression, homologies among organisms, and medical information. It also should be provided through a user-friendly interface that allows for easy and appropriate queries, and

provides useful representation or visualization of the information, to complete its transformation into knowledge. Celera also intends to help its customers apply this knowledge through its *genomic services.* This knowledge and application of biology has the potential to revolutionize drug development, agriculture and other endeavors such as forensics. We believe that this is the vision and competitive edge for Celera—*not* just to be the *Bloomberg* of genetic data, and consequently, the company should have a premium valuation.

Additional information in available upon request.

DJIA:	11134.79
S&P 500:	1408.11
NASDAQ:	3594.21

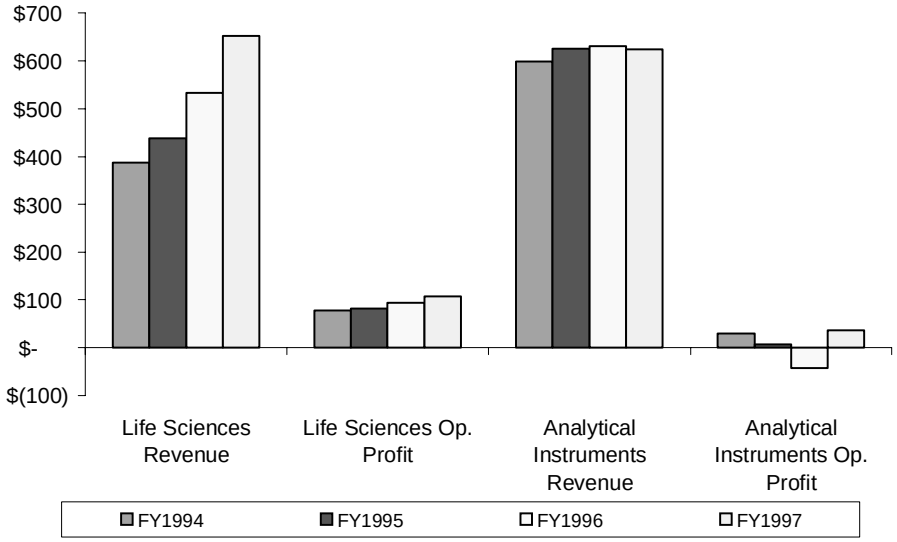
The prices of the common stock of other public companies mentioned in this report follow:

Compaq	$24 3/4
Genentech	$97
Oracle	$78 9/16
PE Biosystems	$95
PerkinElmer	$43 5/8

Appendix A: PE Corporation Vision and Recapitalization

PE Biosystems arose out of a reorganization and recapitalization of Perkin-Elmer begun at the end of 1995. Tony White, Chairman, President and CEO of PE Corporation, joined Perkin-Elmer in September 1995 from Baxter, where he was an executive vice president. He had a vision of *transforming Perkin Elmer into a life sciences company*. At the time, Perkin Elmer comprised two businesses, Analytical Instruments, supplier of analytical chemistry instruments and Applied Biosystems, maker of life science instruments. Applied Biosystems was growing and had about $440 million in revenue with almost 20% operating profit. In contrast, the Analytical Instruments group was stagnant with about $630 million in sales and little operating profit, as illustrated in figure 25.

Figure 25
PE Corporation - Celera Genomics Group
Perkin Elmer Business Performance*
($ in millions)



* prior to acquisition of Perseptive Biosystems

Source: Company financial statements; William Blair & Company, L.L.C. estimates

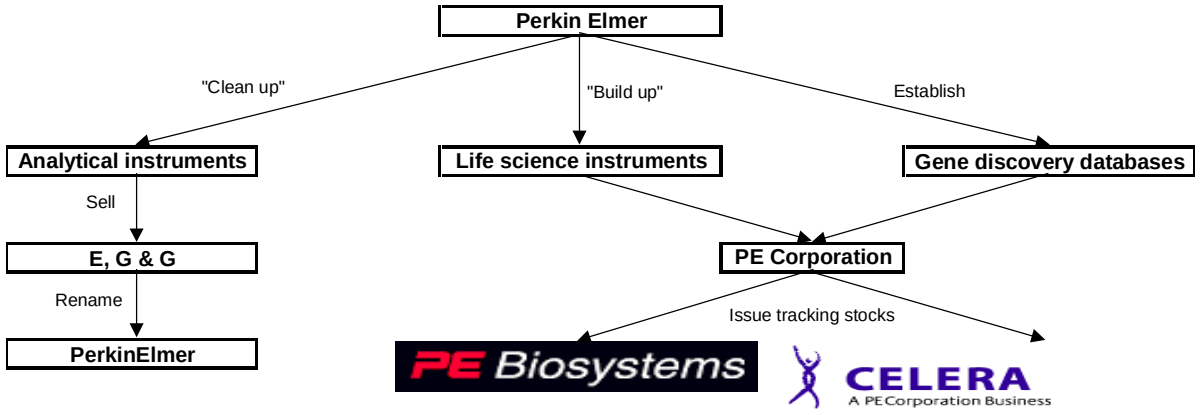
Table 22
PE Corporation - Celera Genomics Group
Business Development Timeline

DATE	EVENT
11/9/99	Collaboration with **Illumina,** developer of fiber optic array platform
6/9/99	Disposed of invested interest in **Tecan**
4/28/99	**PEB** and **CRA** tracking stocks begin trading
3/8/99	Disposed of **Analytical Instruments** to EG&G
5/19/98	Collaboration with **Aclara BioSciences**, a leader in microfluidic technology
5/9/98	Formation of **Celera Genomics**
2/18/98	Strategic partnership with **Hitachi** for electrophoresis based genetic analysis
1/22/98	Acquires **PerSeptive Biosystems**, leading mfg of mass spectrometry and accessories
12/18/97	Investment in **Tecan** to facilitate development of HTS systems for molecular medicine
11/24/97	Acquires **Molecular Informatics**
8/14/97	Collaboration with **Biometric Imaging** to develop HTS system
6/19/97	Strategic partnership with **Hyseq** for chip technology development
2/19/97	Acquire **GenScope**, a genomics company focused on gene expression in living cells
7/1/96	Investment in **Parcel**, a bioinformatics company
4/19/96	Acquire **Tropix,** leader in chemillumenescent detection technology
1993	Acquired **Applied Biosystems**
1989	JV with Canadian Sciex, to form **PE Sciex**

Source: Annual Reports; William Blair & Company, L.L.C. research

Over the next four years, the company proceeded to acquire and build its life sciences portfolio, as highlighted in table 22, with the PerSeptive Biosystems acquisition in 1998 the largest. Additionally, in March 1998 the company formed Celera Genomics. Lastly, the company prepared the Analytical Instrument group for sale. In March 1999, the company sold this group to EG&G, which subsequently renamed itself PerkinElmer, ticker symbol PKI, in October 1999. This process of dividing and reorganizing the original Perkin-Elmer into its current form as PE Corporation is depicted in figure 26.

Figure 26
PE Corporation - Celera Genomics Group
Reorganization and Recapitalization of Perkin Elmer



Source: Company financials; William Blair & Company, L.L.C. analysis

`

Recapitalization

As table 22 shows, tracking stocks for PE Biosystems (PEB) and Celera Genomics (CRA), began trading on April 28, 1999. At that time, approximately 50 million shares of PEB and 25 million shares of CRA were distributed to holders of the original Perkin-Elmer shares, ticker symbol PKI. Subsequently, on July 26 PEB's stock split 2 for 1.

Below are some specific procedural aspects of the tracking stock structure set up by PE Corporation.

Voting. Each share of PEB stock has one vote on matters concerning PE Corporation. The voting rights of CRA share holders are determined as a ratio of the average market value of CRA stock, over a 20-trading-day ending on the 10th trading day prior to the day of record for a vote, divided by the value of a share of PEB stock over the same time period. For example, if shares of CRA are trading at $60 and shares of PEB are trading at $80, then CRA shares have 0.75 votes. It should be noted that there are approximately four times as many shares outstanding for PEB; thus, in the previous example, the shareholders of PEB, as a whole, have about 5 time the number of votes as the shareholders of CRA, as a whole, in votes concerning PE Corporation.

Conversion of either stock to the other. The board of directors may at any time chose to convert either stock, PEB or CRA, into the other at a conversion ration of 110% of the average market value for a 20-trading-day period. If this event is deemed taxable, then the ratio would be only 100%. We view this possibility as more of a *reinvestment risk* for shareholders, rather than a fundamental stock price risk.

Dividends. PE Corporation chose to maintain its quarterly dividend of $0.085 per share, paying it to PEB shareholders. CRA is not expected to pay a dividend for the foreseeable future.

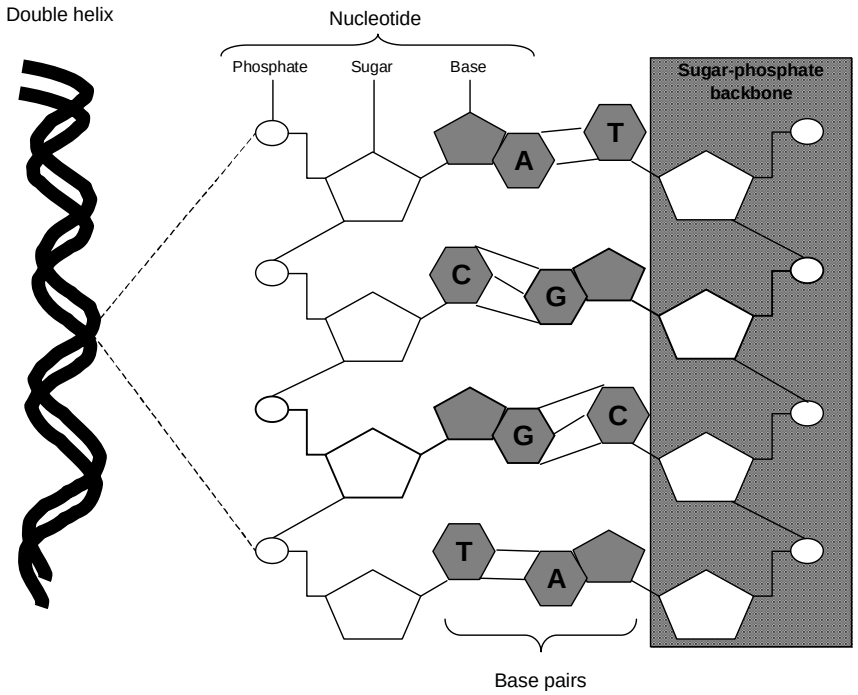
Appendix B: Molecular Biology

Living things are made up of cells, numbering from one for simple organisms such as bacteria, to trillions such as those that make up a human being. Each one of these cells contains a copy of that organisms' genome. A genome is the complete set of genetic information that determines what an organism is and what it does in response to its environment. The genome is like an uncompiled computer program, which when processed by the cell's machinery (enzymes) gives rise to the cell. At its most fundamental level the genome is made up of deoxyribonucleic acid, better known as DNA. DNA of specific, discrete sequences form packets of information called genes. These genes are strung together on a scaffold known as a chromosome, and the group of chromosomes in total makes up the genome. Molecular biology is the discipline that deals with the study of genetic information, with the hopes of deciphering and using it for various applications such as pharmaceutical, farming and forensics.

DNA

DNA (and its cousin RNA) forms the basis of all genetic information. Its now familiar double helix structure, shown in figure 27, was determined in 1953 by Watson and Crick. The DNA molecule's basic shape is that of a tightly wound double helix, which upon closer inspection resembles a ladder. The ladder is made up of basic units or letters called nucleotides. A nucleotide consists of sugar and phosphate molecules attached to a nitrogenous base, of which there are four types: adenine (A), cytosine (C), guanine (G), and thymine (T). The sugars and phosphates are strung together to form the molecule's backbone, while the bases span the space between the backbones to form the rungs of the ladder. The bases pair by specific, weak interactions; *A pairs with T* and, *G pairs with C*, with each strand forming the complement of the other. The human genome is made up of more than 3 billion of these base pairs. Discrete sequences of these bases encode the genetic information known as genes that give rise to proteins such as insulin and collagen.

Figure 27
PE Corporation - Celera Genomics Group
Basic DNA Structure



Source:William Blair & Company, L.L.C. analysis

Each time a cell duplicates, its genome must be copied as well to provide sufficient instructions to each daughter cell. DNA's ladder like structure and redundancy through its complementary strands ensures the fidelity of the information. The molecule replicates by "unzipping" down its axis, with each strand providing the template which an enzyme called DNA polymerase copies the complementary strand. DNA "unzips" again to allow other enzymes to copy the information into RNA, a process called transcription, which provides the program for proteins.

Genes

The long strands of DNA make up discrete sequences which encode for proteins which perform roles in the cell, either structurally in the case of collagen, a component of skin and bones, or functionally, in the case of insulin or hemoglobin in blood cells. The human genome is estimated to have over 100 thousand genes, which encode for proteins that are expressed at varying levels throughout the life of an organism. These genes in aggregate constitute less than 10% of the genome. The remaining genetic material represent regulatory sequences that coordinate the transcription or replication of the DNA, as well as non-essential DNA that serves little purpose in the life of the organism, yet may yield practical utility, such as the case with "microsatellite" repeats used as forensic markers. The high fidelity of the replication process ensures that a gene's functionality is preserved with each duplication, yet minor changes over time coupled with natural selection have yielded the great diversity we now observe, as well as certain diseases such as cancer. Thus a gene that performs a particular function in a bacteria or mouse, often has a close cousin (in sequence) in a human being. Comparisons such as these allow for the use of model organisms to study disease, as well as the opportunity to conduct comparative, functional genomics studies to decipher the function of genes and proteins.

Genes' discrete nature allows them to be shuffled from one section of the genome to another, as well as across species. This allows us to moves genes from one organism to another to suit our needs, giving rise to gene therapy and genetically modified foods. However nature also exploits this property as part of the evolutionary process, where by genetic material is passed between species. This has resulted in multi-drug resistant strains of bacteria that now threaten the health of those in hospitals.

Chromosomes

Genes and other sequences are arranged linearly on long DNA strands. The sheer volume of this material would prove unruly and lead to errors unless contained in some manner. The DNA molecules are combines with proteins to form a scaffold upon which the DNA is packed into tight bundles called chromosomes. Chromosomes not only provide convenient packaging, but also ensure the proper distribution of genetic material to daughter cells at replication.

The number of chromosome varies as shown in table 23, on the following page. A human carries 23 pairs of chromosomes while the fruit fly only 4. As mentioned earlier, one function of chromosomes is to ensure the proper number is passed on to each daughter upon replication. Errors do occur. Down's syndrome is the result of trisomy 23, or three copies of the twenty-third chromosome.

Table 23
PE Corporation - Celera Genomics Group
Genome Size and Complexity

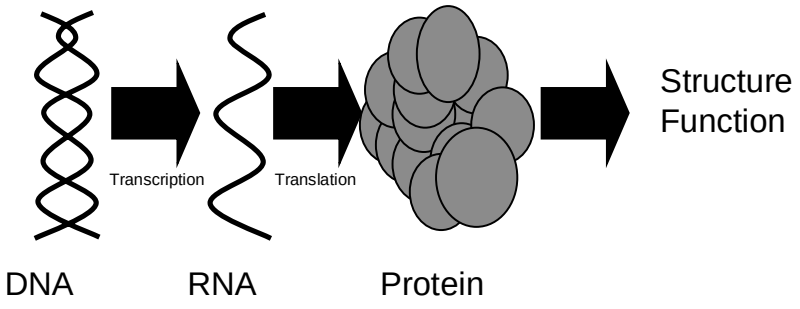
Organism	Genome Size (base pairs)	Relative Size (E.coli = 1)	Length of DNA (mm)	Number of Chromosomes
Virus				
SV70	5,000	0.00125	0.0017	1
T2	200,000	0.05	0.68	1
Bacteria				
Mycoplasma	300,000	0.075	0.1	1
E.coli	4 million	1	1.36	1
Fungi				
Yeast	20 million	5	68	16
Animals				
Fruit Fly	200 million	50	70	8
Chicken	2 billion	500	680	78
Human	3 billion	1500	1700	46
Plants				
Peas	9 billion	4500	3100	14

Source: The World of the Cell; William Blair & Company, L.L.C. estimates

Proteins

Proteins can be thought of as the physical agents of genes, performing some role in the organism either structural or functional. Proteins are made up of long strands of molecules called amino acids, of which there are 20. Genes give rise to proteins, which are said to be expressed in a two-step process, shown in figure 28. The first step, transcription, involves the processing of DNA into an intermediary message called mRNA. To facilitate this the DNA unwinds, and enzymes use the gene's DNA as a template for the single stranded, complementary mRNA. The mRNA travels through the cell where it encounters another set of molecules known as ribosomes, which use the mRNA template as instructions to build proteins out of the amino acid building blocks. Ribosomes read the mRNA in series of triplets called codons. Each codon corresponds to a particular amino acid, as shown in table 24. It should be noted that there is redundancy to help ensure that errors caused by the transcribing enzymes are not propagated in the proteins encoded for by the mRNA. These proteins then carry out their role inside or outside of the cell. The human genome is likely to encode over 100 thousand to 1 million distinct proteins, which make up the functioning human body.

Figure 28
PE Corporation - Celera Genomics Group
Central Dogma



Source: William Blair & Company, L.L.C. estimates

Table 24
PE Corporation - Celera Genomics Group
Codon Table

First position	Second Position of Codon												Third position
	T			C			A			G			
T	TTT	Phe	[F]	TCT	Ser	[S]	TAT	Tyr	[Y]	TGT	Cys	[C]	T
	TTC	Phe	[F]	TCC	Ser	[S]	TAC	Tyr	[Y]	TGC	Cys	[C]	C
	TTA	Leu	[L]	TCA	Ser	[S]	TAA	*Ter*	[stop]	TGA	*Ter*	[stop]	A
	TTG	Leu	[L]	TCG	Ser	[S]	TAG	*Ter*	[stop]	TGG	Trp	[W]	G
C	CTT	Leu	[L]	CCT	Pro	[P]	CAT	His	[H]	CGT	Arg	[R]	T
	CTC	Leu	[L]	CCC	Pro	[P]	CAC	His	[H]	CGC	Arg	[R]	C
	CTA	Leu	[L]	CCA	Pro	[P]	CAA	Gln	[Q]	CGA	Arg	[R]	A
	CTG	Leu	[L]	CCG	Pro	[P]	CAG	Gln	[Q]	CGG	Arg	[R]	G
A	ATT	Ile	[I]	ACT	Thr	[T]	AAT	Asn	[N]	AGT	Ser	[S]	T
	ATC	Ile	[I]	ACC	Thr	[T]	AAC	Asn	[N]	AGC	Ser	[S]	C
	ATA	Ile	[I]	ACA	Thr	[T]	AAA	Lys	[K]	AGA	Arg	[R]	A
	ATG	Met	[M]	ACG	Thr	[T]	AAG	Lys	[K]	AGG	Arg	[R]	G
G	GTT	Val	[V]	GCT	Ala	[A]	GAT	Asp	[D]	GGT	Gly	[G]	T
	GTC	Val	[V]	GCC	Ala	[A]	GAC	Asp	[D]	GGC	Gly	[G]	C
	GTA	Val	[V]	GCA	Ala	[A]	GAA	Glu	[E]	GGA	Gly	[G]	A
	GTG	Val	[V]	GCG	Ala	[A]	GAG	Glu	[E]	GGG	Gly	[G]	G

Coded Protein Name

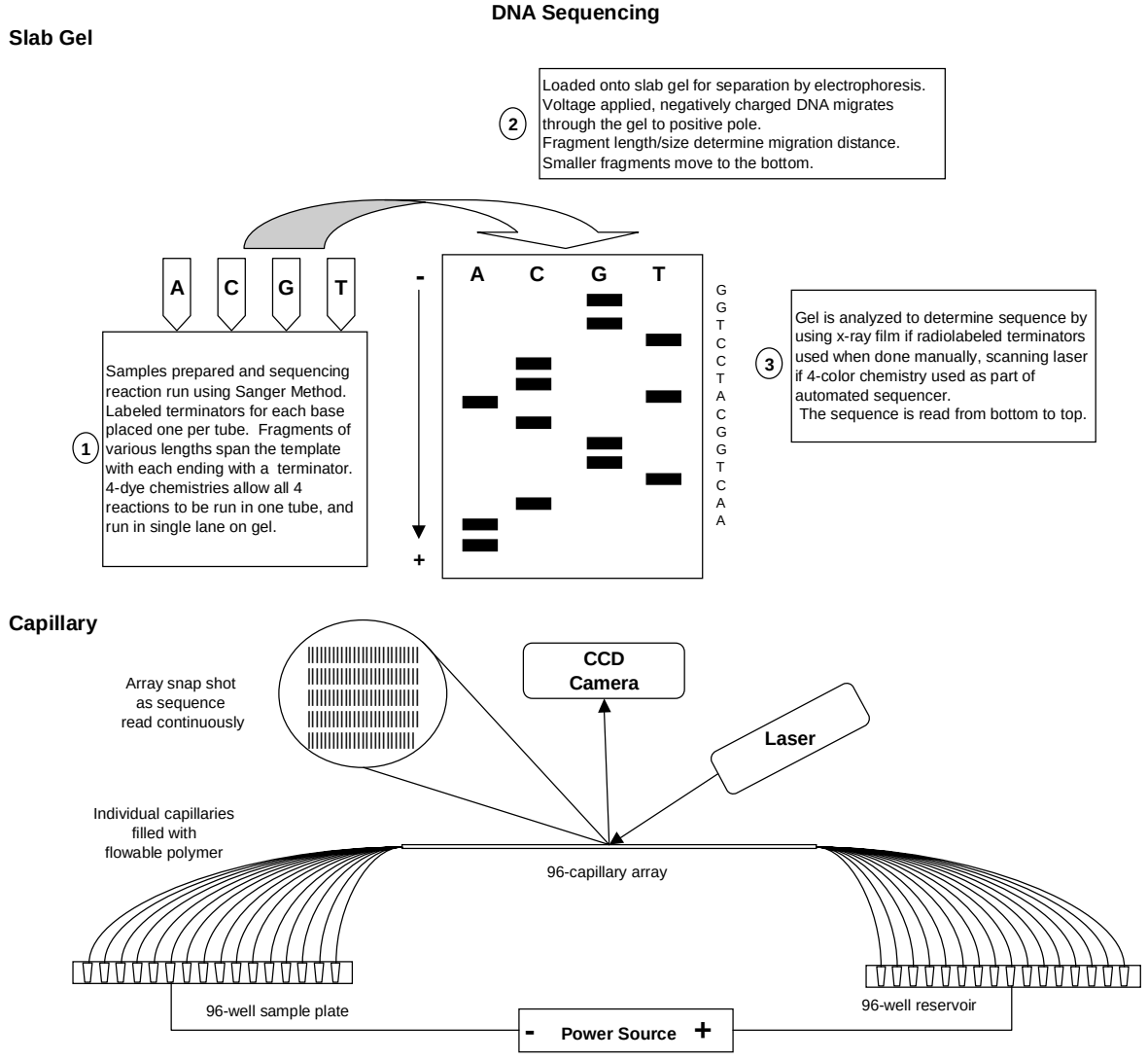
Phenylalanine Leucine	Serine Proline	Tyrosine Histidine	Cysteine Tryptophan
Isoleucine Methionine Valine	Threonine Alanine	Glutamine Asparagine Lysine	Arginine Serine Glycine
		Aspartate Glutamate	

Source: NIH; Principles of Biochemistry; William Blair & Company, L.L.C. analysis

DNA Sequencing

The core of understanding biology of an organism at the molecular level involves the decoding of its genetic material. Getting at basic sequences allows for the construction of maps of various resolutions, and ultimately complete genomes that will be offered by groups such as Celera and the NIH. In principle, the current tools used to tease apart this information are based on the Sanger method of DNA sequencing, which can be thought of in three discrete steps: sample preparation, sequencing, and analysis, as shown in figure 29, on the following page. Sample preparation involves isolating and purifying the DNA from the specimen from which it was collected, this may be from bodily fluids or tissue samples. The DNA to be sequenced, called the template, is parceled out into four test tubes, and then combined with a mixture of the four nucleotides, random stretches of DNA called primers, and a modified DNA polymerase enzyme which makes copies of DNA off of the template. Also added to each tube are one of four (A, T, G, C) nucleotide analogue terminator molecules, that is labeled with either a radioactive molecule or a colored dye. The sequencing reaction starts when the random primers attach to the template DNA at the sequences that are complementary along the length of the template strand. The DNA polymerase recognizes this paired structure, and begins synthesizing a copy of the DNA using the nucleotides in the chemical mixture. Occasionally a terminator is incorporated into the growing strand of newly synthesized DNA, causing the DNA polymerase to fall off of the template. The mixture of chemicals and the primers are random enough to ensure that one terminator molecule has been incorporated at the end of sufficient copies of various lengths of the template, corresponding to each base in the sequence. The reaction is then stopped and sample is prepared for the separation phase.

Figure 29
PE Corporation - Celera Genomics Group
DNA Sequencing



Source: DOE, NIH, William Blair & Company, L.L.C. analysis

Sequencing relies on the separation of these mixed DNA fragments generated by the previously described reaction. Current methods rely on electrophoresis, which is the movement of charged molecules through a liquid under the influence of an electrical field. The samples are placed on top of a liquid/gel matrix that in most cases resembles Jell-O. An electrical charge is applied to the gel, driving the negatively charged DNA molecules through the gel towards the positive pole. The gel matrix acts as a sieve, slowing the migration of larger fragments through the gel, as a result large fragments are found at the top of the gel while short one are at the bottom. Sequencing is carried out on one of two platforms, either on slab gels, or in capillaries. Slab gels are large sheets of gel, the Jell-O, sandwiched between two glass plates. Samples are placed at the top of the gel, electric current is applied, the fragments migrate, and the gel is scanned at the end of the run with a laser, if dyes are used, and the image is captured for analysis. The ABI Prism 377 is an example of this sequencing format. Capillary electrophoresis is carried out in thin glass capillaries that are filled with the sieving polymer. Samples are placed at one end of the capillary, current is applied, and the capillary is continuously scanned to provide information in real-time. The ABI Prism 3700 is an example of this platform. Capillary machines provide several advantages over slabs, making them useful for large scale sequencing efforts. Slab gels

are cumbersome since the gel is toxic and must be prepared just prior to running a sequencing experiment. The gels have low voltage loads, which result in longer run times, reducing the throughput. Finally the platforms are labor intensive, especially in the sample loading process. Capillary machines such as the 3700 may be fully automated, utilizing flowable polymers that remove the gel pouring step, and allow for higher voltage applications that yield faster run times and higher throughput.

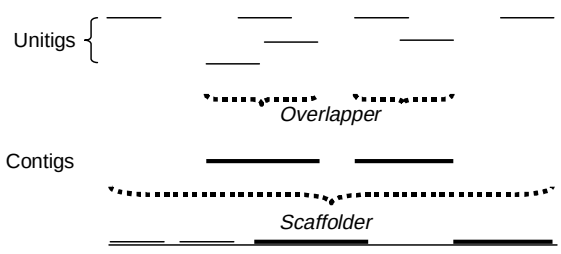
Sequencing Strategies

Large-scale sequencing projects are tremendous undertakings requiring careful planning. A crucial process that must be decided is the sequencing strategy to be used, either directed sequencing, or whole genome shotgun sequencing. The publicly funded human genome project employs the directed sequencing strategy, while Celera and TIGR employ whole-genome shotgun sequencing. Directed sequencing, as previously illustrated in figure 7, involves the creation of clone libraries that represent small pieces of a given genome pasted into a surrogate chromosome from either yeast or bacteria. A particular clone in a library is selected for sequencing and primers—designed to recognize the junction of the artificial chromosome and the genomic DNA—are used to begin the sequencing. A stretch of this genomic DNA is sequenced, and then the end of this read sequence provides the information to create the next sequencing primer, to run the next reaction for that particular clone. This process is repeated over and over, allowing scientists to "walk" along the DNA with one long read. As scientists make their way through the library, the genome is knitted together to create maps of various resolutions, eventually yielding the complete sequence. This strategy is very labor intensive and time consuming since it is contingent on the previous sequence and the creation of overlapping primers before processing, yet *yields contiguous data continuously* as the project progresses.

Whole-genome Shotgun Sequencing

The whole genome shotgun sequencing method was developed by Craig Venter, President of Celera, at The Institute for Genome Research (TIGR), in the mid 1990s as an attempt to more efficiently decode genomes for use as research tools. Whole-genome shotgun sequencing relies on sophisticated mathematical algorithms and high-powered computers to generate complete, detailed maps of the genome under study "all at once," bypassing the intermediate maps. For whole-genome shotgun sequencing, the genome under study is randomly shattered into many small pieces, and incorporated into artificial chromosomes to form a library much like the directed approach. This process may be performed up to ten times, providing up to 10X coverage, with each coverage randomly broken up in different ways. The pieces come in three sizes, 2 thousand, 10 thousand and 150 thousand base pairs, or letters, long. These various pieces are all sequenced from each end, yielding *short* sequences of about 600 base pairs in length or unitigs. The sequences are passed into a powerful computer running an assembler program, which consists of algorithms that tile the fragments, looking for overlapping sequences to piece together, as if completing a puzzle, as illustrated in figure 30.

Figure 30
PE Corporation - Celera Genomics Group
Shotgun Sequencing and Assembly



Source: GSAC; industry interviews; William Blair & Company, L.L.C. analysis

Assembler program. Celera had four basic design criteria for the assembler program. First, the program was written to make all the sure moves first, getting progressively more aggressive if necessary. Second, the program was to identify sequences in all the highly repetitive areas and leave these for last. Third, the program needed the capacity to incorporate all available data. This data may include information in the public domain. Lastly, the program had to provide a complete audit trail.

We believe that Celera's assembler program is robust and impressive. The Celera assembler program was able to determine the genome for *H. influenza* in *5 minutes* versus the *24 hours* the same task took originally at TIGR. For Drosophila, the assembler took 36 hours, and the team estimates that it will take a total of 90 days to complete the human genome. At least part of this 90 day total can be concurrent with the sequencing.

The Celera assembler program has 6 components: 1) the screener; 2) the overlapper; 3) the unitiger; 4) the scaffolder; 5) resolution of repeat sections—Repeat Rez I, II, and III; and 6) the consensus builder. The screener masks the heterochomatin and ribosome DNA. The overlapper identifies sequences with greater than 40 base pairs, or letters, overlapping—allowing for up to 6% mismatch. The unitiger assembles unitigs, or the 600 base pair sequences mentioned above, into longer contigs and identifies unique remaining unitigs. There is about a 130 fold reduction in the number of pieces at this point. The scaffolder maps the contigs and unitgs relative to each other to identify the positions within the entire genome. The Repeat Rez routines fill in the *gaps* caused by highly repetitive DNA. Finally, a consensus genome is determined, with SNPs, single nucleotide polymorphisms, resulting as each individual carries DNA from both its mother and father.

ESTs

In addition to the complete consensus genome with whole genes, other tools and approaches are used to look for gene fragments. One can identify genes that have been expressed or transcribed into mRNA. By using an enzyme, reverse transcriptase, that can make DNA from RNA, one can create cDNA or complementary DNA. cDNA made this way is termed an EST or expressed sequence tag, as it represents an expressed gene. It has also been processed by the cell to remove non-coding gene regions from coding gene regions and is thus not a complete gene. Current empirical estimates, based on the Drosophila and Arabidopsis (a plant) genomes, are that ESTs found will only represent 33 to 50% of the complete genome, missing rarely expressed genes for example.

Appendix C: Glossary

Antisense. Stretches of complementary DNA that pair with specific sequences of expressed mRNA. Antisense therapeutics stops disease at the genetic level by blocking the production of disease-related proteins.

BAC (Bacterial Artificial Chromosome). Common vector used for mapping and sequencing. Larger than a cosmid, but smaller than a YAC.

BDGP. Berkley Drosophila Gene Project.

Bioinformatics. The development of new tools for the analysis of genomic and molecular biological data, using the combined power of biology, mathematics, and computers.

cDNA (complementary DNA). Single-stranded DNA that is complementary to messenger RNA, representing the expression of genes in a given cell.

Combichem (combinatorial chemistry). A method of generating diverse compound libraries used to develop the starting blocks of pharmaceutical compounds, as well as in the material sciences.

Contig. Groups of clones representing overlapping regions of a genome.

CRO (contract research organization). Outsourcing service provider to the pharmaceutical industry that set up, run, and monitor clinical trials for the FDA approval process.

DNA. (deoxyribonucleic acid). The molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

Expression. The detectable effect of a gene. The appearance of an inherited trait.

Forensics. Pertaining or applicable to personal injury, murder, and other legal proceedings.

High Throughput Screening (HTS). The process of running assays to elucidate the function of protein, or molecule, used on the development of pharmaceuticals.

Homolog. A member of a pair of identical chromosome parts with respect to their construction and genetic content.

Microsatellite. Highly polymorphic DNA marker comprised of mononucleotides, dinucleotides, trinucleotides or tetra-nucleotides that are repeated in tandem arrays and distributed throughout the genome. They are used for genetic mapping.

Mitochondria. Principal energy source of the cell. Small, membrane-bound cellular structure responsible for converting nutrients into the energy-yielding compound adenosine triphosphate (ATP) to fuel the cell's activities. Mitochondria are found in eukaryotic cells (cells with a nucleus contained within a membrane) and carry a different genetic code than a cell's nucleus.

Model Organism. Organisms used in life science research that have sufficient evolutionary relationships to humans, such as fruit flies and mice.

Ortholog. Genes from same organism with sequence homology, i.e. interleukin 1 and interleukin 2.

Paralog. Genes from different species with similar functionality, i.e. human and mouse interleukin 1.

Pharmacogenomics. The study of genetically determined variations in responses to drugs in humans or in laboratory organisms.

Proteomics. Study of proteins.

Ribosome. Small cellular components composed of specialized ribosomal RNA and protein; site of protein synthesis.

RNA (ribonucleic acid). A chemical found in the nucleus and cytoplasm of cells. It plays an important role in protein synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, each serving a different purpose.

SNP. (Single Nucleotide Polymorphism). Difference in DNA sequence differing in a single base pair.

YAC (Yeast Artificial Chromosome). A vector used to clone DNA fragments. It is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells.