SECURITY OF OUR PERSONAL GENOME

By Gregory H. Smith

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Master of Science

in the Department Informatics of,

Indiana University

August 2003

Table of Contents

Introduction

Our personal genome, which is the map of our DNA, is our ultimate source of identity, which should be given our highest concern for security. The primary approach used for securing any highly sensitive health care data such as our genome would be to guard against any personal identity information being associated with the data. The belief that nameless data records eliminates risk and would be a benefit to research is the common pretense for how we manage our health data systems. However, the incredible advances that we are seeing with computational power and more affordable and sophisticated DNA sequencing software may be creating a problem greater then the benefit that it is providing. Now we must be concerned about all data in the health care systems that could provide a link to accessible identity free data. Old data records or samples that provide possibilities of DNA sequence matching to existing identity free genomic data presents a whole new problem. How might this change the face of health care? Will further advances in technology make it impossible for us to secure our personal health information? Solutions could lead to restricting our ability to improve health care or it could force us to rely more heavily on ethical judgment to protect the rights of patients.

The unprecedented rate of recent advances in information technologies along with improved speed, economy and accuracy of mapping the human genome has created serious concerns about the usage and security of this new highly sensitive genetic data. Our knowledge of DNA has come along way in the 50 years since James Watson and Francis Crick first presented their discovery of the double helix. The discovery timeline

has been crowded in recent years starting with the U.S, Department of Energy's Human Genome Initiative in 1986 and culminating in completion of the Human Genome Project in 2003. The exponential growth of genomic scientific accomplishment now forces us to assume new milestones will arrive sooner then later.

> **BETHESDA, Md., April 14, 2003 –** The International Human Genome Sequencing Consortium, led in the United States by the National Human Genome Research Institute (NHGRI) and the Department of Energy (DOE), today announced the successful completion of the Human Genome Project more than two years ahead of schedule.

The completion of the original public funded project is a landmark; however, the anticipated opportunities from this project spawned numerous privately funded projects in recent years. Incredible progress has been made to the point where scientists are taking Craig Venter's prediction of a $1000 personal genome by 2012 very seriously.[1] We are already accumulating vast amounts of genetic data on patients that are proving valuable in diagnosing and treating a select number of diseases. The problem that we now face is that a lot of this data may be more sensitive with respect to privacy issues then we could have imagined just a few years ago. When the Human Genome Project, HGP, was started the guidelines for the protection of the privacy of donors actually took into account the fact that the anonymity of the donors was only secure because the technology required to determine identity was not readily available outside of the HGP. Our computational abilities are still following Moore's Law which states that we will "Double raw performance every 18 months". This computational power is extremely

valuable for solving our scientific challenges, however, it is also providing the capability to access and correlate our nation's immense amount of medical data. This creates the opportunity for conversion to sensitive identifiable data that can be misused.

In the movie "Gattaca" we saw futuristic testing that instantly determined one's genetic makeup for the purpose of human genetic engineering and population discrimination. We would hope that this would never occur in our culture, where our personal freedoms were restricted by the use of our genomic information but we must start to be more aware of the genetic trail that we leave throughout our lives. Our genomic information can be found in many places, hopefully we still have an opportunity to secure it, but we must also envision and prepare for the extreme possibility that is presented in "Gattaca" where our important genetic sequences could be instantly analyzed from samples of our dead skin or hair.

This report will focus on the specific aspects of securing our personal genome mostly under the context of the protection of highly sensitive medical health data. However, the issues that relate to genetic data are far broader and must be reviewed in order to give proper background for this report. The initial background section for this report provides some insight into the categories that relate to the use or protection of genetic data.

Background

An individual's personal genome is the total complement of their genetic information. The genome consists of structures called chromosomes that are composed of very long double strands of DNA. Each human cell contains 23 pairs of chromosomes. One-half of each pair is inherited from an individual's mother and the other half of the pair is inherited from an individual's father. The 23rd pair is composed of the X and Y sex chromosomes. Each chromosome contains many genes which are the basic subunits of life, however, these genes account for less then 2% of the genome. Chromosomes are located in the part of the cell called the nucleus. The long, double strand of DNA contained in each chromosome are made up of nucleotides which are composed of phosphates, a sugar, and four different nitrogen-containing bases. These bases in DNA: (A)adenine, (G)guanine, (T)thymine, and (C)cytosine make up the side-by-side arrangement of bases along the strand (e.g., ATTGCCT). It is the difference in the arrangement of these bases on each strand of DNA that leads to the uniqueness of each person's genetic makeup. There are now believed to be less then 25,000 genes in a human genome, and expression of these genes leads to the production of a large number of proteins that perform most of our life functions and make up the majority of our cellular structures. These proteins are large complex molecules made up from 20 subunits called amino acids. The constellation of all the proteins in a cell is known as the proteome. And a person's genotype is their genetic identity, the specific combination of genes that they have in their cells.

Relevant Genetic Research Issues

- **Privacy – protection of identity**
    - **Sensitivity of Genetic Data**
- **Data Security – protection of personal genome**
    - **Authentication**
    - **Encryption**
- **Ethics**
    - **Cloning (Therapeutic and Stem Cell)**
    - **Discrimination**
        - **Job**
        - **Insurance Coverage (both ways)**
- **Legislation**
    - **HIPAA**
    - **State Laws for public disclosure**
    - **Disability Insurance Regulation**
- **Genetic Testing**
    - **Physician Controlled**
    - **Personal choice**
- **Medical Record Systems**
- **Forensics**
- **Research**

Privacy

The issue of privacy in relation to genetic data is all about the protection of the identity attached to the data. Why is this important? Well unfortunately it did not carry as much importance in the early days of DNA sequencing. In fact, we knew so little about most of the DNA data that we considered it a scientific wasteland. However based on our knowledge today this information takes on new importance and sensitivity. Much of this early genetic data in research databases and patient medical records exists without the level of security precautions that we might apply to the same data today. The general security strategy is that of anonymity, meaning medical data is safe as long as the owner's identity is protected. This strategy was valid before the advent of the powerful data search engines and sophisticated DNA search algorithms. But today we must consider all personal health data as vulnerable to providing a link to identity which would then render the data valuable with concern for misuse.

In this age of rapid advances in technology, new medical records systems and data transfer techniques are implemented in the interest of improving care before the security can be developed sufficiently to protect the information. The challenge is to maintain the rights of patient confidentiality without limiting the quality of the health care or the potential for valuable research. Today there is a legal motivation to comply with goals of maintaining a patient's privacy. Liability and civil lawsuits have always influenced privacy policy, however, with the heightened concern over highly sensitive medical data the government has stepped in with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). HIPAA which took effect on April 14, 2003 was originally

authored to deal with protecting the security and privacy of personal health care records, and it still maintains that primary focus along with adding various administrative regulations. But it is important to realize that HIPAA has and will hence forth change how we deal with the privacy of health care records.[2]

There are many approaches being taken to hopefully guarantee our privacy, however, this is where technology may be our enemy. Studies done at Carnegie Mellon University by Malin and Sweeney have shown that our identity is just a data search away from being discovered even with just minimal nameless data such as zip code, date of birth, gender. They show that with this data 87% of the U.S. population can be identified.[3] Unfortunately this type of data has never been considered worthy of protection and is accessible for many health records. Research shows that by combining any specific health visit information such as a doctor's appointment with these typical nameless data items a person can be identified 98% of the time.[3] The lesson to be learned here is that if identity is this easy to ascertain with typical access to public health records how can we ever guarantee privacy for our personal health information which in turn could provide the identity link to our genomic information.

Sensitivity of Genetic Data

The availability of genetic data and the possible knowledge that it represents has created the most difficult question for justifying the risk for the rewards. The obvious concern is the high sensitivity of this data due to its impact on personal privacy. But of a greater concern may be the discoveries of the future that could make this data even more sensitive. The technology is advancing rapidly to provide a higher quality of genetic data

in less time. Combining the ease of identifying a person from general health records with the ease of matching DNA sequences and one realizes a situation where a segment of DNA data may be enough to provide a positive match to an assumed anonymous DNA record. An additional concern comes from the new capabilities which exist to match DNA with protein information, which is evolving out of the study of proteomics. If these advances in technology are combined with a plummeting cost for analyzing our personal genome, as with the possibility of the $1000 genome within 10 years, then the overwhelming challenge that face the guardians of our medical records starts to become evident.

Data Security

Security for the protection of our personal digital health information is a huge topic that has many parallels with the procedures used for the protection of financial data. We start with a base strategy of physical security or restriction of physical access to the computer and network. Next, the concept of controlling access to data through authentication and then the protection of the data through encryption provides a security structure for most computer applications. Medical record systems utilize this structure and it is taken to the most advanced levels for the protection of the personal identity of sensitive health care data. And for the most part computer security works when it is adhered to properly. However, compromises of any security system can occur resulting in the loss of valuable information. The question then becomes what risk are we willing to accept when it comes to the protection of our personal genetic data.

Authentication:

Can there be a perfect authentication system? Most systems rely on something known such as a password. The passwords are encrypted with guidelines to ensure security, but theft or human error is far too common to be able to rely on this type of protection. An additional layer of security may rely on something possessed such as a key or smart card. Many believe that biometrics is the answer, but the weakness generally lies in the theft of the biometric signature or in inaccuracy of the technology. The concern over the use of biometrics for the protection of medical data by the health care providers can be summed up by a quote from Tal Moise, VP of Business Innovations at Clarian Health Services.

> "Currently, we have decided not to rely upon biometrics to provide access control to highly sensitive medical data. No industry standards have been established which indicate the permissible value for false positive authentications. Without that standard, Clarian would potentially create a liability for itself by way of having selected to low of a thresh-hold. No device has been shown to be 100% accurate; however, the more expensive devices on the market tend to have a lower false positive authentication. It is not reasonable for the organization to spend $3000 per device for a .2% false positive when there are $30 devices with a .3% false positive read. The reality is that the .1% difference could affect access to your data that is when the liability is created. The industry needs to have an approved standard to control this risk, given that by selecting the technology standard ourselves, when a data compromise occurs then it could be perceived as

> our fault. Yet, if a compromise occurs with a password based system then it is typically viewed as the user's fault for having distributed the password."

Authentication is a major challenge that if it is implemented correctly allows medical records systems to provide selective data access which means that it controls who can access what information. This selectivity is quite powerful, but it also creates a situation where the effectiveness of the information security comes into question. How does one decide who has access to what data? We might assume that our personal physician would have access to our genetic data, but what if we end up in an emergency medical situation where we could benefit from other health care providers having access to this information. These scenarios start to build in options to our medical record systems that only complicate the overall integrity of the process.

Encryption of data

Encryption of data is now a common option for securing data communication and for the storage of specific data. The use of a Private Key Infrastructure, PKI, with properly managed digital signatures will provide the security needed for sensitive data. But as with most security systems, improper use and lack of protection of private keys or passwords will result in the compromise of the system. The added challenge for PKI is also the convenience and acceptance of use. Diligence to properly administer the PKI and the commitment to use digital signatures can offer us a way to secure our genetic data. It probably comes down to our acceptance that genetic data must be secured so at a minimum we will utilize encryption for access and storage of the data.[4,5] Encryption could also be taken to the extreme with regard to the actual protection at the DNA level.

Research is being done to mask the identifiable genetic data by mixing in DNA base pair masks into the actual DNA samples.[6] This may seem to be an extreme precaution, but it shows what level of protection would actually be required to secure stored sample of DNA. It probably justifies some research into using coded masks for scrambling the base pairs of stored DNA sequences.

Ethics

Ethical considerations have followed the science of genetics from the very beginning. Exploring or manipulating the building blocks of life seems to conjure up serious ethical debate about what humans should and should not be delving into. In many situations, decisions made with knowledge gained from genetic data have no overseeing authority other then the ethical climate of the scientific community. But the genetic knowledge does exist and it has great value if used properly. So the ethical debate is very important if not critical to the future of genetic research and development. In fact it may turn out that ethical management of our personal genome is what we will need to rely upon if it turns out that we will not be able to secure or control this information with technology. The major genomic based topics for ethical discussion center on cloning or discrimination based on misuse of genomic information.

Cloning

Cloning presents a different twist on the use of genetic information. Most people assume human cloning means producing an exact duplicate of a human being. There has yet to be a confirmed existence of a successful human clone. However, the concept is starting

to play out in areas such as the creation of designer babies. Cloning does represent the ultimate controversial use of one's personal genome. The ethical debate over the evolution of cloning may also be the most visible. Again this issue exists because of the scientific and technological advances made possible from the utilization of our genetic knowledge. It is generally accepted that the cloning of human beings is a dangerous path to explore. However, the potential from therapeutic cloning and the use of stem cell research has such great potential that we can be guaranteed that where there is funding, which today is primarily private, there will be research and development in cloning.

Stem Cell

We will assume that our world will properly address human cloning and that therapeutic cloning based on stem cell research will probably evolve with various forms of regulation. The issues that come up with respect to therapeutic cloning and genetic security have to do with management of the embryonic cell line. Without going into a complete description of stem cell research, suffice it to mention that an embryonic cell minus its nucleus receives a nuclear genetic material from a cell of interest such as skin or cardiac for cloning. Cell multiplication occurs that provides a source for transplant or repair for the specific cell type. So the genetic security concerns for this process seems to be in selection and protection of the chosen embryonic cells. The real fears are not so much with identity and protection of specific genetic data, but with the unknowns that could exist with this type of research. A good lesson here is that we need to be patient. Possibly a lesson should be reviewed from the scientists at the 1975 Asilomar Conference who agreed to suspend research involving recombinant DNA technology until potential

risks could be evaluated. The competition to come up with breakthrough scientific discoveries which translate into fame or fortune is hard to resist.

Discrimination

Discrimination can be an outcome of the misuse of personal genetic information. Hence there is a need for legislative or ethical guidelines to insure the proper use for information derived from our personal genome. There are many forms of discrimination but the most commonly highlighted areas with regard to genetic information are with employers and health insurers. The Americans with Disabilities Act (ADA) which is discussed in the Legislative section below deals with discrimination based on a disability. A genetic disorder can be classified as a disability which allows for protection by the ADA..

In recent testimony before Congress, Dr. Francis Collins, Director of the National Human Genome Research Institute at the National Institutes of Health, noted:

> while genetic information and genetic technology hold great promise for improving human health, they can also be used in ways that are fundamentally unjust. Genetic information can be used as the basis for insidious discrimination. . . . The misuse of genetic information has the potential to be a very serious problem, both in terms of people's access to employment and health insurance and the continued ability to undertake important genetic research.[7]

A landmark case relating to discrimination by an employer involved Burlington Northern Santa Fe Corp. illegally testing workers to determine if they were genetically predisposed to Carpal Tunnel Syndrome. This was the government's first case against workplace

DNA discrimination and the company was fined 2.2 million dollars.[8] This case not only set conditions for the type of testing that could be done but also for the consent that must be obtained prior to testing or use of genetic information.

There is great fear that insurance companies will take advantage of access to medical records to decline coverage to individuals who they know will generate costly medical expenses in the future. This is of great concern with the new capabilities for determining future health and disease conditions from personal genetic data. We have never had this type of crystal ball potential for diagnostics before. Now the alternate side of this discrimination can occur when individuals procure large health or life insurance coverage when they know that they will become the victim of a future medical condition.

Legislation

Government intervention is becoming more and more a part of the securing genomic information. Laws that set guidelines for control of genetic data tend to be represented by general health data privacy legislation. The umbrella legislation that has created quite a stir in the medical community are the HIPAA regulations that are primarily intended to manage privacy protections required for health care providers. But legislation is also provided at the state level for the control of health data disclosures which are important for regional health care monitoring. The health insurance industry is regulated at the state and national level which just adds to the complicated regulatory web which exists to protect us but not at the expense of benefiting us.

The national government has influenced state governments with HIPAA, but states have also been venturing more deeply into genetics based legislation. A majority of states have enacted laws that strictly prohibit the use of genetic information for risk selection and risk classification in health insurance. Six states prohibit genetic discrimination in life and disability insurance without actuarial justification. Seventeen states restrict insurer use of genetic information in life, disability or long-term care insurance in some manner. Nine states have laws pertaining to human cloning. Laws in 16 states require informed consent for a third party to either perform or require a genetic test or to obtain genetic information. Twenty-three states require informed consent to disclose genetic information.[9]

All states assume medical information to be confidential, but they also recognize the increasing capabilities for storing and transferring this data escalates the challenge of protecting privacy. Laws in all states restrict access to medical records. An important question is whether genetic information should be protected generally, as another component of health data, or by special genetic data privacy laws.

Legislation to deal with discrimination due to genetic defects is essentially covered by the Americans with Disabilities Act (ADA). This law protects individuals who have an impairment that substantially limits them in a major life activity, who have a record of such impairment, or who are regarded as having such impairment. Although loosely written, genetic impairment is essentially covered by the ADA.[10]

A new trend in post 9/11 legislation can be seen with the recently legislated Terrorist Identification Database Act of 2003 which is buried deep within the Justice department's

secretly drafted Domestic Security Enhancement Act of 2003. Also known as the Patriot Act II, it would empower the Attorney General to collect DNA samples for the purpose of "detecting, investigating, prosecuting, preventing or responding to terrorist activities." This option for data disclosure has been provided for in the HIPAA regulations. This type of legislation has benefit beyond homeland security since it would enable government health authorities to respond to health alerts such as with Anthrax or smallpox scrares. The value of the genetic information supposedly would be in the ability to respond more effectively to a threat to specific population area based on genetic demographics.

Genetic Testing

Public and private testing for genetic disease or predisposition for a disease, genetic carrier status and paternity testing are becoming more common. The screening of newborns for genetic disorders by state public health programs is estimated at 4 million infants annually. This is justified by the discovery of abnormalities that can result in severe problems, mental retardation or even death.[9] Testing for genetic mutations that can predict with varying degrees of probability the likelihood an individual will develop cancer is now commercially available. However, information about the usefulness of cancer susceptibility genetic testing in screening, prevention and treatment remains incomplete.[11] There are experts that believe these tests should only be performed in the context of carefully designed clinical research, but testing is now available both with and without physician referral.

Genetic testing has medical, psychological, ethical, social and financial implications that will need to be dealt with sooner then later. Of course the fact that we are now able to perform genetic testing is a major reason why we need guidelines for dealing with these implications. But most of all we just need to be aware that the issues do exist.. One common concern such as the question about the counseling that may be required by clinical workers for the proper disclosure of test results to patients is just the beginning of the psychological and social issues. The financial implications are far reaching with respect to the economics of the health insurance industry and government funded medical coverage. If the tests are beneficial then the wealthy will pursue them which will open up many new ethical dilemmas.

Physician Controlled

Genetic testing has mainly been offered within the context of medical diagnosis and treatment as a specialist discipline. The options available to physicians today for requesting specific genetic tests allow for the finest health care that we have ever known. Now our medical professionals must add this expertise to their resumes and this should be a positive thing. However, the exposure of an incorrect diagnosis due to lack of testing or improper interpretation may open up additional liabilities that will strain an already fragile malpractice insurance situation. This concern has the potential to increase the cost of health care even more.

Personal Choice

It is now possible to personally request genetic tests for paternity verification, forensic evidence and various medical diagnoses. This is due to a simplification of the tests which in turn has made them affordable. The fact that genetic tests are being requested is a result of a growing understanding by the public of medical information and a corresponding desire to make health decisions based on this information. This shift to personal responsibility and partnership in health decisions has been encouraged by governments in countries with more socialized medicine.[12] It is being encouraged in the U.S. by the commercial testing companies. The concern that private testing creates is the personal detriment that could be caused by the lack of counseling that should accompany sensitive medical diagnostic results. This has been seen with the private HIV testing services. This is not to say that the testing organizations are not providing proper counseling, but it does call for a need for policies and guidelines.

Medical Record Systems

It is important to discuss the role that medical record systems play in the control of sensitive health data such as genetic information. There has been debate for and against the centralization of our health care data that may be referred to as a National Health Database, NHD. Support of a NHD primarily comes from the potential for improved health care and research benefits. The heavy criticism comes from the potential "big brother" growth of government involvement. However, what needs to be understood is that we are much closer to having a NHD than we may realize. The growth and

interconnectivity of all of the individual medical record systems running health care in this country forces us to pay more attention to the gatekeepers of our sensitive health data. The recent report "Connecting for Health, a Public-Private Collaborative", released by the Markle Foundation's Data Standards Working Group points out the need for standardization of our medical data and communication strategies to facilitate more effective sharing of data for the overall improvement of health care.[13]

A National Health Database may never have the security to justify the storage of genetic data, at least in standard database fields. However, the possibility could exist for the inclusion of genetic data into a NHD if the data itself was encrypted possibly with an unbreakable double key. This relates back to authentication and encryption issues, but when considering the increased sharing of medical information that is inevitable, the NHD will need to have a failsafe security options. There will be many groups of information users that need controlled access to various portions of our medical information. This is guaranteed because the driving force behind medical record systems is to increase productivity and to justify the overall cost many user communities such as insurers and employers along with health providers will need to contribute. Hence these communities would need specialized access control.[14]

Forensics

The use of genetic testing for forensic science is an area that is generating a substantial amount of genetic data with an entirely different purpose. That purpose is now heavily based on homeland security but it has traditionally been about criminal identification. The issues of misuse of data in forensics are parallel to that which is of concern to the

medical community. The primary goal of the test analysis is to determine identity from a DNA sample by keying on a few unique DNA locations. This does not tend to increase the problem of excess DNA sequence information being misused, but the issue of identity misuse does cause concern. This identification causes concern that our government may know too much about us and that it could lead to other types of discrimination.

The forensic issue that is probably most concerning to those monitoring the ethical use of personal genetic information is the growing databank of DNA results from the testing of criminals and suspected criminals. This is now spreading to new requests to get DNA profiles of foreigners from countries with known terrorist connections. The FBI's national DNA database contains 1.4 million profiles of convicted offenders. However, this would be greatly increased if Attorney General John Ashcroft receives his request for a billion dollars over the next 5 years to reduce the backlog in DNA testing for criminal cases.[15] The UK, France, Germany, the Netherlands, Austria, Switzerland, Canada and Australia all operate forensic DNA databases of similar construction to the U.S. and a number of European bodies are concerned with co-operation and standardization.[16]

The primary accepted use for genetic data has been for judicial evidence and prosecution. The value of genetic identification is invaluable in placing a person to a crime scene or proving paternity rights for real parents. But what if genetic information was used to prohibit parental rights based on a genetic predisposition for a disease?[17]

Research

The utilization of genetic information for research purposes is the driving force behind many new medical and pharmaceutical discoveries. The public sector has justified genomic research for all of the potential that genetic data could yield for medical diagnosis and treatment. But it is the pharmaceutical industry that drives research because of the blockbuster potential that a genetically engineered drug represents. Early use of genetic databanks yielded promising results that encouraged the pharmaceutical industry to invest heavily in the bioinformatics research required for extraction of information from the genetic data. However, the gold mines have not been found specifically for genetic based drugs or cures. But success has come from the use of genetic information which can be used to influence prescription and medication usage, the area of pharmacogenetics. The concern of our researchers is from a fear that privacy concerns will make it more difficult to gain access to the wealth of medical related genetic data.

Most genetic research organizations that work with human biological samples have guidelines for proper and ethical management of the samples and data resulting from the research. One foundation for these guidelines comes from the National Bioethics Advisory Commission (NBAC) that outlines that ethical researchers must pursue their scientific aims without compromising the rights and welfare of human subjects.[18] Recent advances with the technologies of genetic sequencing and the management of the data has created new concerns about protecting the privacy of the individuals who willingly submitted samples for research prior to these capabilities.

Policies are in place today that require informed consent with respect to the genetic uses for human biological materials that are needed for support of the biomedical research community. However, it is crucial that the more than 282 million specimens already in storage remain accessible under appropriate conditions and with appropriate protections for the individuals who supplied this material. The new genetic databanks that have emerged in recent years have been created more as privately funded versus publicly funded. The most famous genetic databank which is managed by the Icelandic Company deCODE is leveraging the advantages of a homogenous population and excellent historical genealogical population records. The policies for personal privacy are extensive since regulation is provided by the Iceland government.[18] However, other genetic databanks are being driven entirely by commercial ventures such as the brokering of information by the First Genetic Trust Company out of Deerfield, IL. What we have seen from the commercial ventures are major efforts for providing ultimate security for the personal genetic data. Proof of security is the foundation of their product; it is what they must sell to entice participation in their research studies.[2]

Risk vs. Benefit

A review of the risks and benefits from the areas of genetic data use.

Genetic Issue	Risk	Benefit
Privacy - Identity Protection	Loss of freedom and opportunity plus risk of discrimination	Genetic knowledge for improved health care
Genetic Data Sensitivity	DNA is our ultimate personal identification, it must be protected	Medical benefits from known information with potential for future knowledge
Data Security - Authentication	Key to your most personal information, may restrict access by those who could help you	Front line of protection of your genomic privacy, allows you to control your own genetic security
Data Security - Encryption	Utilization of genetic data is more difficult	Second line of security for your genomic privacy, could allow you to secure most sensitive data
Ethics - Cloning	Mistakes could be catastrophic	Therapeutic cloning could revolutionize healing
Ethics - Discrimination	Genetic data if misused could cost you employment, health coverage and personal freedom	Is it a personal choice? Genetic data could prove beneficial as proof of health and hereditary strength
Legislation	Government control could reduce the benefits from our genomic knowledge	Recourse for protecting our genomic privacy as well as national security
Genetic Testing Physician Controlled	More complex diagnosis with potential for malpractice and higher medical costs	Potential for improved diagnosis of a level never known to medical science before
Genetic Testing Personal Choice	Psychological damage from the health information	Personal control of health decisions
Medical Records Systems	Genetic data may be too risky to place in general medical records systems	Improved health care from access to genetic data from medical records systems
Forensic Genetic Data	Loss of personal freedoms without proper control of criminal DNA databases	Benefits of genetic evidence for criminal identification and terrorism prevention
Genetic Research	Security of research data is a changing entity that could jeopardize subject privacy	Potential of medical and pharmaceutical discovery

Methods

This thesis is not a typical scientific study, however, it is a scientific investigational study and there are methods involved. The methods are heavily dependant upon the use of extensive literature review to uncover all of the latest developments in the area of medical health data security with the specific emphasis on genetic data. But this report is also based on investigative evidence from a medical and academic environment that represents all aspects of today's genomic landscape. The methods are further refined by the information technology, IT, perspective presented by the author, an IT professional of 25 years, who has maintained a unique relationship with scientific computing and who writes this thesis in fulfillment for a Master of Science in Bioinformatics.

The working environment for this thesis is the Indianapolis campus of Indiana University and Purdue University known as IUPUI. The author is the Director of Information Technology for the Purdue School of Engineering and Technology. The Indiana University Medical School located at the same campus provides numerous opportunities for direct investigation and confirmation for the various medical and research data handling issues. This medical community offers a unique benefit from the relationship the Medical School has with the hospital operating organization, Clarian Health. Representing University, Riley and Methodist hospitals, Clarian is listed as one of the best hospitals in America by *U.S. News and World Report* in 2002. Another added benefit comes from the current association Clarian has with Cerner Corporation for the

implementation of their Millennium clinical information system. All of these organizations provided valuable input for this study.

Securing our Genome

It is true that there is an immense amount of sensitive medical health data that exists in various data repositories around the world. Most of this data is considered to be protected from unauthorized access. But this data has been produced over a period of time where we have seen adjustment of security policy in order to keep up with technological advances that convert old anonymous data into sensitive personal information. This access to old data is not a major privacy epidemic that threatens to expose a large portion of our population. But the lessons are applicable to the larger amounts of more recent data that we feel is secure. The increasing power and sophistication of technology will continue to break down the walls of privacy that we feel are impenetrable today.

We will first take a historical view of various types of genetic data and the events that have created such great potential use for this genomic information. We will then focus on the current usage of our personal genomic information which will expose our concerns for securing our personal genetic data. This will all lay the foundation for discussion about what security is required with respect to the creation, utilization and protection of our personal genome by the various data providers and managers of our personal health data.

Historical

Personal medical records have historically existed as folders in physical filing systems maintained by the government or a health care provider. Even today the folder based filing system is the norm with the electronic medical records, EMR, beginning to provide duplication primarily for productivity. These physical records generally do not contain advanced medical data that would be equated to the genomic era, however, they do contain a wealth of information that could now be used to confirm a genetic condition as well as supply genealogical links. Data such as actual disease occurrence, blood and biochemical data could now be used to build a very complete genetic profile. And these records are relatively available to the medically qualified public. Supporting data such as the U.S. census even with the 70 year moratorium for access to personal data can be valuable. So we need to be concerned about these old records, even if their antiquity may be their greatest security. The effort required to access and incorporate this old data into current technology does not justify the expense, meaning the transcription to digital format. However, there are situations that justify the effort.

Genomic Era

The concept of genetics may have started with Gregor Mendel's theory of heredity and then became tangible with Watson and Crick's discovery of DNA 50 years ago. Sequencing of DNA and genetic testing has only been around for about 25 years and human DNA sequencing with any significant data has only been available for a handful of years. Some of the first personal genetic data could be tied to genetic disease testing

such as for Cystic Fibrosis or Sickle Cell Anemia which tested for a symptom or trait that confirmed the disease. The ability to sequence and manipulate DNA in the late 70's offered beneficial opportunities for DNA based drug development which launched hundreds of new biotechnology companies. The first public discussion about mapping the human genome occurred in 1984 and with Polymerase Chain Reaction, PCR, and automation available in the late 80's the Human Genome Project was launched.[19]

Human Genome Project

The Human Genome Project, HGP, was an international effort officially launched in October of 1990. The project was planned to last 15 years for an estimated cost of 3 billion dollars, but rapid technological advances accelerated completion date of the working draft in June of 2000 with the completion of the project on April 14, 2003. The finished sequence produced by the HGP covers about 99 percent of the human genome's gene-containing regions, and it has been sequenced to an accuracy of 99.99 percent. The project goals were to determine the complete sequence of the 3 billion DNA base pairs, identify all human genes, and make them accessible for further biological study.[20] The HGP is also known for the public vs. private efforts to complete the sequencing and the great debate it has created in the scientific community. This was highlighted by the simultaneous publication of the historic Genome Issues of Nature supporting the public effort of the Human Genome Project and Science publishing a version by Celera Genomics.

Human Genetic Milestones of the Genome ERA

Here are a few events outside of the Human Genome Project since 1990 that helped to shape the genome era that now presents so many options for human application.

The U.S. Department of Energy (DOE) and the National Institutes of Health (NIH) establish the Ethical, Legal and Social Implications (ELSI) program in recognition of the issues that would arise from the HGP.

1990: The first use of gene therapy to treat severe combined immunodeficiency (SCID) or ADA Deficiency, a very rare disease known better from the movie "The Boy in the Bubble". This was the first occurrence of an altered gene being placed in a human patient to treat a disease. This event was important because it presented incredible hope and motivation for the potential that should be realized from genetic research. However, the ADA example of gene therapy has not proliferated into other similar success stories.[21]

The development of the ABI Prism 3700 which automated DNA sequencing in 1995 changed the face of the DNA mapping procedure. This automation of sequencing was the breakthrough that was needed to allow research to think beyond the HGP.

1996: Development of the GeneChip®: The Department of Biochemistry at Stanford and Affymetrix introduce a technological breakthrough in gene expression and DNA sequencing technology with the introduction of DNA chips, small glass or silica microchips that contain thousands of individual genes that can be analyzed

simultaneously. Since then, DNA Chip technology has become a growth industry as new tools for making, probing, imaging, and analyzing arrays are introduced almost daily.[21]

The first cloning of a mammal (Dolly the sheep born February 1997) was performed by Ian Wilmut and colleagues, from the Roslin institute in Scotland. This cloning was important to the human genome era because of the fear that it generated for potential misuse of the genetic research taking place.

In December of 1998, the parliament of Iceland passed a bill that allowed for the creation of a centralized database of all the Icelandic peoples' genealogical, genetic, and personal medical information. The parliament then granted an exclusive contract to deCODE genetics, a biomedical company, giving deCODE access to the national health records for purposes of genetic research and drug development.[22]

In 1998 two research teams, led by James Thompson (UW Madison) John Gearhart (Johns Hopkins) succeed in growing human Embryonic Stem (hES) cells, pleuripotent, self renewing cells with the demonstrated ability to differentiate in vitro into all three embryonic germ layers. Stem cell research has been at the center of ethical genetic controversy ever since.[21]

Human Cloning Prohibition Act of 2001, HR 2505, a ban on all human cloning, either for reproduction or for therapeutic cloning to derive immunologically compatible embryonic stem cells was passed by the U.S House of Representatives. The bill, that did not become law, would have made cloning a crime punishable by up to 10 years in prison. Some

states have passed such a law and many countries such as the UK and Japan have been struggling with this type of legislation.

Medical Records

It has been an amazingly short span of time that has brought us to a point in history where we have to make major decisions about how we hope to control biotechnology so that our personal genomic freedoms are not lost forever. Medical records are the major source of concern for the protection of personal privacy. The issues that relate to the securing of our personal genome are similar to the issues of managing our medical records. The transition to electronic records that are accessible by many and solely dependent upon basic computer security sets the stage for the more complicated issues of dealing with genetic data being generated by ever sophisticated biotechnology. But this is all a part of a changing landscape of the medical record management systems that not only have the goals of efficient management of data but goals of more efficiency and productive health care systems.

Electronic Medical Records

The primary repository for our personal health records will be our medical records stored by our primary health care provider. These records are generally stored as hard copy but the concern for our personal privacy centers upon the Electronic Medical Records, EMR. New medical records management software is spreading rapidly through the U.S. health care providers. Old systems are being updated, databases are being linked together, and

web access portals are becoming common. All of this is suppose to be secure and HIPAA will help control any weaknesses that could jeopardize our privacy. Overall this computerization of our health care system is good. So if there are security risks then the value of the improved health care probably justifies it.

How does this digitization of health care connect to our concerns about protecting our personal genome? Direct connection today is not a concern. Our personal genomes do not exist so they are not stored in our HMO's computer system. However, an incredible amount of personal data is ending up in these data repositories. The survival of the health care industry is dependent upon this automation which is driven by the processing of insurance claims and compliance with government regulations. Consolidation of health care providers has been the trend for the last decade. With this corporate consolidation comes the information systems consolidation. Patients are tracked by identification numbers that are cross referenced across many filing systems or databases. The access boundaries for these information systems are merged and redesigned. The gatekeepers to the information are laid off and then outsourced. So is there a problem? That is a question that must always be avoided since the trust in the integrity of our health care system must be protected. There is an assumption that computer administration staff is highly skilled and capable of handling all of these procedural and operational issues. Unfortunately the Information Technology workforce is only human comprised of a lot of talent and capable of making a lot of errors.

Behind all of this activity to build better EMR systems we are seeing the emergence of what will become a "National Health Database" of interconnected health care

management systems that are evolving out of the need for national and global efficiencies similar to what has taken place with our interconnected financial systems.

National Health Database?

The lack of support that that resulted in the downfall of the Clinton administration's national health plan leads us to believe that the American public would never agree to the creation of a national health database. However in a survey conducted by the Foundation for Accountability (FACCT) to gain a better understanding of the internet-using public's interest in using computers to manage their healthcare, it was found that seventy percent of consumers would be interested in using some or all the aspects of an electronic personal health record.[13]

Interconnected Health Information Network

The arguments for improved quality of care, reduced errors meaning saved lives and better health management have fueled spirited legislative debate but have not resulted in legislation for the creation of an interconnected health network that would be resemble a National Health Database. However, the exact model for a national health database is actually coming together as an interconnected network of medical records systems based on existing industry standards. It is referred to as the "operable" standards set that contains the HL7 v2.x data interchange standard, the HL7 Reference Information Model, the DICOM standard for imaging, the NCPDP SCRIPT prescription drug information standard, the LOINC vocabulary for laboratory tests, the IEEE/CEN/ISO 1073 medical device communication standard, the ASC X12 administrative transaction standard, HL7

Data Types, Clinical Document Architecture (CDA), and the HL7 Clinical Context Management Specification (CCOW).[13] The health care industry that needs the benefits of electronically linking hospitals with federal and state health agencies, will probably come together through private collaboration with the public sector. Bringing this to fruition is a pioneer facilitator for promoting emerging information and communication technologies, the Markle Foundation. Helping to equip the effort is the initiative, Connecting For Health, that leveraged the recent concerns for more accurate secure medical records, coupled with concerns for combating bioterrorism and global health threats such as SARS. What we see emerging is an extremely expedient effort that has united over 100 organizations to bring the American healthcare system into the information age.

Connecting for Health is a public-private collaborative of the Markle Foundation that will advance an interconnected, electronic national health information infrastructure by focusing on adopting national clinical data standards for interoperability, ensuring secure and private transmission of medical information and working to understand consumers needs and expectations from an interconnected health information system.[13]

Over the next year, a group of participating hospitals, technology companies and government agencies will use Markle's Healthcare Collaborative Network to share de-identified or anonymous information on patients' clinical procedures, lab results, prescriptions and diagnostic summaries. A key part of the network, the Markle group and its supporters envision, is a "personal health record", PHR, that would gather patient's medical data from doctors, hospitals, pharmacies and insurers together in one Internet-

based record.[23] There are a number of existing health data interchanges that are laying the groundwork for an interconnected health care network. The New England Healthcare EDI Network, NEHEN, is a prime example of collaboration between healthcare providers in a regional network. Made up of over a dozen major health organizations in New England, NEHEN has enabled them to securely transmit millions of HIPAA compliant administrative transactions since the network's inception in 1998. What this all means is that the digital structure is soon to exist that will not only fulfill the goals for a nationally interconnected health information network, but will lend itself to valuable data mining efforts that will be justified under the banner of medical research.

Standardized Medical Terminology

This interconnected health information network is not the only piece of the puzzle. The other major challenge to overcome is to standardize the medical terminology used for all of the medical data that is stored. This problem has recent support from the U.S. Department of Health and Human Services for the development of a standardized electronic medical records system based on a common medical terminology database.

> Tommy G. Thompson, Secretary of Health and Human Services, announced on July 1, 2003, an agreement with the College of American Pathologists (CAP) that will make SNOMED Clinical Terms (SNOMED CT®) available to U.S. users at no cost through the National Library of Medicine's Unified Medical Language System® (UMLS).[24]

SNOMED CT is a major driving force for the standardization of medical terminology. The Cerner Corporation was one of the first major health care information technology companies to fully integrate SNOMED CT into their Electronic Medical Record structure. Mark Hoffman, Cerner's Genomic Solutions Manager can see the value of the SNOMED decision "We definitely see the national license as a positive development and worked with the government to encourage that decision."

Do these developments translate into a modern digital health care system for the U.S. in the coming years? Not at all, they just lay the groundwork and help break down some public and private political barriers. The real challenge is the money needed to implement the new systems. But the major health care providers such as Kaiser Permanente are moving forward with major investments and the trend appears to be that America's health care providers have accepted the need to invest more in their Information technology.

HIPAA

The Health Insurance Portability and Accountability Act of 1996 ("HIPAA"), established the Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule"), which for the first time, creates a set of national standards for the protection of certain health information. The Privacy Rule standards address the use and disclosure of individuals' health information by organizations subject to the Privacy Rule, as well as standards for individuals' privacy rights to understand and control how their health information is used. A major goal of the Privacy Rule is to assure that individuals' health

information is properly protected while allowing the flow of health information needed to provide and promote high quality health care and to protect the public's health and well being. The Rule strikes a balance that permits important uses of information, while protecting the privacy of people who seek care and healing. Given that the health care market place is diverse, the Rule is designed to be flexible and comprehensive to cover the variety of uses and disclosures that need to be addressed.[2]

The Privacy Rule applies to health plans, health care clearing houses, and to any health care provider who transmits health information in electronic form in connection with transactions by the covered entities of HIPAA. The Privacy Rule protects all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. This protected information referred to as “Individually identifiable health information” is information, including demographic data, which relates to:

- the individual’s past, present or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual,

Individually identifiable health information includes many common identifiers (e.g., name, address, birth date, Social Security Number). There are no restrictions on the use or disclosure of de-identified health information. De-identified or anonymous health information neither identifies nor provides a reasonable basis to identify an individual.[2]

However, here lies the question for what data is really de-identified especially when dealing with the ultimate identifier of our genetic data.

HIPAA is legislation that has officially identified the best practice ethical operating procedures that had already evolved in the health care industry. The fact that the HIPAA legislation was created validates that control and security of health care data was in question. HIPAA is good for the fact that official guidelines were needed and initially HIPAA is dramatically changing the face of health record management because of the reengineering of the health care data systems. There are the specific compliances but HIPAA also presents a general rule of minimal disclosure of an individual's health information. A covered entity must make reasonable efforts to use, disclose, and request only the minimum amount of protected health information needed to accomplish the intended purpose of the use, disclosure, or request. This is forcing all of the covered entities to review and generally re-engineer their data handling policies. However, it has also helped to define the rules of disclosure that do allow access to data by many public and governmental agencies.

Allowable Disclosures

The key to HIPAA rests with the individual providing authorization for use of their personal information. This has created a documented authorized flow of data in the health industry which had always been in place with implied consent. Since the HIPAA covered entities now have authorization to handle the patient's personal information they

are also allowed to disclose this information to 12 identified national priority purposes which are:

- That which is required by Law.
- Public Health Activities. These activities would include control or prevention of disease or other threats to public health along with identification of individuals suspected of exposure to a health concern. The regulation of FDA reporting and disclosures of information for the compliance to agencies such as Occupational Safety and Health Administration (OHSA), the Mine Safety and Health Administration (MHSA), or similar state law.
- Victims of Abuse, Neglect or Domestic Violence
- Health Oversight Activities. Which would include audits and investigations necessary for oversight of the health care system and government benefit programs?
- Judicial and Administrative Proceedings
- Law Enforcement Purposes. There are a number of purposes relating to assisting in the location and capture of a fugitive, prevention of crime or for the prosecution of a crime.
- Decedents. This would be primarily for identification of deceased persons or determination of the cause of death.
- Cadaveric Organ, Eye, or Tissue Donation.

- Research. This authorization of information for use by research is primarily to help insure that health information is protected and that any use is properly monitored by governing authorities. But it does open the door for the use of information by research for the benefit to health that it can provide to the general public. There is also provision for the creation of a limited data research set that can be created if it ensures anonymity.

- Serious Threat to Health or Safety. This priority has been greatly influenced by the post 9/11 anti-terrorist activities.

- Essential Government Functions. This would be the additional government catch all for functions such as: assuring proper execution of a military mission, conducting intelligence and national security activities that are authorized by law, providing protective services to the President, making medical suitability determinations for U.S. State Department employees, protecting the health and safety of inmates or employees in a correctional institution, and determining eligibility for or conducting enrollment in certain government benefit programs.

- Workers' Compensation. In order to comply with, workers' compensation laws and other similar programs providing benefits for work-related injuries or illnesses.

HIPAA gives us a framework that helps us deal with the explosion of health care data being generated in the U.S. It does not speak directly to the unique privacy needs that are being created by genetic data, but HIPAA gives us a foundation from which to build a baseline for handling sensitive health care data. From this start, we could hope to better

identify the additional privacy requirements that will surface around the handling of our personal genomic data.

Data Becomes Information

When our personal genome is readily available there will be sophisticated medical information systems available to store it. But before that happens there will already be plenty of personal medical data that can be correlated into information as informative as our personal genome. The key to this valuable information is not just the specific pieces of data, but the information that can be derived from the data. The data search and computational algorithms that are available today will devour data in a national health information network to produce incredibly valuable information for medical and pharmaceutical research.

The computerization of our health care records is justified on many fronts for the overall improvement of our health care. Health care providers, medical insurance providers and employers all will benefit from access to EMR's. Safe guards will be in place to insure that accessed data will not be misused. Actually the marketing tact of the Connecting for Health Markle Project is that all records will be Personal Health Records, PHR's and the individual will have total control of them. The individual will authorize access to their own records. That should be a fairly easy process for a population that is computer savvy. Actually this will be a monumental problem if real security and flexibility is desired. There will probably be some basic options for a patient such as to allow their health care provider access which would include associated doctors. This would require

patients to make decisions such as; if they would like quality health care then would they be willing to allow access to their records, not a difficult choice to make. All of the access rights decisions will be driven by the patient's need for service, such as allowing access to their insurance company if they want coverage. The question about access by research would probably be influenced by their desire to help find cures for their own health problems. And would all of fears be put to rest because access will follow rules of anonymity or secured privacy laid out by complying with regulations such as have been provided by HIPAA?

Genetic Data

The definition and clarification of what makes up genetic data is very loosely defined. In this section the specific types of gene based medical information will be discussed along with how it is being utilized, stored and protected.

Genetic Medical Records

The occurrence of genetic data in our current medical records is very limited if we defined genetic data as specific to base pair sequences. However, an increasing number of genetic tests are becoming available as a result of recent and rapid advances in biomedical research. This test data is now becoming a part of our current medical records. There are a number of different reasons genetic tests are performed. These include the following:[25]

- Clinical genetic testing (diagnosing current or future disease)

- Pharmacogenomics (assessment of therapeutic drug treatment)
- Identity testing for criminal investigations or forensics studies
- Parentage or paternity testing
- Tissue typing for transplantation
- Cytogenetics (chromosome analysis)
- Infectious disease testing.

There are hundreds of genetic tests that can be run, most which either identify the presence of variant genes that cause disease or confirm a diagnosis suggestive of a genetic disorder. Some diseases that have a genetic component include: Alzheimer's, Bone Marrow Disorders, Breast, Ovarian and Colon Cancer, Cystic Fibrosis, Down Syndrome, Leukemia, Lupus, Osteoarthritis, Sickle Cell Anemia and Thalassemia. The tests may be specific to DNA abnormalities or they may test or measure the presence of genetic end products such as proteins or metabolites. These tests generate a lot of valuable medical data that when viewed alone may not convey that much information, but when analyzed against all other data could uncover medical information that we might not authorize.

Let us take for example typical test results for a particular mutation for BRCA1 or BRCA2 such as a data entry such as 185delAG which indicates a 2 base pair deletion of AG at nucleotide 185. This specific test result alone just tells us about a specific mutation that happens to be related to breast cancer and does not confirm anything. However, what if this data is placed in context with a patient's genealogy, for example they happen to be Ashkenazi Jewish, and you have a different set of statistics due to a

study performed at the Baylor College of Medicine.[26] Again this is an issue today because of the ability to cross reference so much electronic health care information and the ease of identification of a patient's identity. We must keep in mind that a BRCA test result of 185delAG is a fairly typical piece of medical test data that is going to be handled with typical security attributed to an EMR.

A very common source of genetic test data occurs at birth with standard genetic screening test performed shortly after birth. This is now the beginning of a data trail that will remain with an individual for life. An initial challenge is the identification of a newborn that does not yet have official identification such as a name let alone a universal identification code. The American Academy of Pediatrics has issued policy statements that address the issues of EMRs for pediatrics. This is a more recent concern since most EMR policy has been developed for managing adult medical records. An additional challenge for pediatric records is the compliance with specific privacy laws of the individual states. It is interesting to note that the policy statement specifically references embryo donor situations now. "EMR systems must provide protection of information on a patient's genetic information, including newborn metabolic screening results. This protection must extend to those who are genetically unrelated to their parents (eg, those born after donor embryo procedures)."[27]

The concern for properly managing this new genetic test data rests with the control of the intelligent data analysis that an individual may or may not want to have their records exposed to. So as this genetic test data finds its way to the new PHR's, additional security and privacy burdens are placed on the caretakers of the medical records. But

will the security sensitivity of these records reflect the face value of the data or the potential value of an intelligent analysis?

Genetic Databanks

Genetic Data Repositories have been springing up to feed the appetites of the pharmaceutical companies hoping to capitalize from the potential of gene based drug development. These large databanks have been created from genetic records taken from stable populations with good genealogical records or they have been created from commercial ventures for the purpose of supporting research along with offering advantageous genetic information to the participants.

Iceland's Genetic Databank

The most impressive example of a population based genetic databank is Iceland's decision to include genetic data in their national health database and then grant exclusive 12 year contract rights to deCODE genetics (http://www.decode.com), an Icelandic biomedical company. These actions allow deCODE to combine genetic information with the genealogical and health records of each Icelander in order to create a comprehensive database. From the beginning deCODE formed financial alliances with pharmaceutical companies for rights to specific genetic disease research. Most notably was the original partnership with Roche that included a large amount of seed capital. deCODE'S researchers have already mapped disease genes linked to more than 25 common diseases and identified 14 specific disease genes. Today the partnerships read like a who's who of pharmaceutical research and the fortunes of deCODE appear to be proving that the

commercialization was a good decision for all involved. With so much personal information available to a private enterprise, scientists and policy makers are watching the endeavor closely to see how the ethical, legal, and business aspects are resolved. There are also many questions still being asked about the apparent collaboration between deCODE and the Iceland government. [18,22]

Other countries and health providers are following Iceland's lead in combining health and genetic data on large populations. The hopes and promises point to the potential for "personalized" medicine with plenty of promises for security and privacy. Estonia hopes to follow the Iceland model for capitalizing on its homogenetic population without the controversies over the implied consent and business relationships. They have launched a 3-year $2.5 million pilot project, targeting 10,000 donors that will rely on a health questionnaire rather then medical records and offer participants access to their genetic profile unlike the Iceland plan. Another high visibility project is the BioBank UK that plans to enroll 500,000 donors with an initial budget of $66 million. This database would be the largest population database based on interview surveys for subjects 45 to 69 years of age with a plan to track them for 10 years with no offer of personal benefits from the study. BioBank UK would also draw from national health care records and along with a opportunity to factor in lifestyle information.[28]

Looking at activity in the U.S. we find discussions by federal health representatives that point toward larger versions of the BioBank UK project. A new program that is very interesting is the Marshfield Personalized Medicine project starting up in Wisconsin. Lead by the non-profit Marshfield Medical Research Foundation with a focus on the rural

population surrounding the small town of Marshfield, WI, the project hopes to enroll 40,000 donors for the database with no offer of profile reports to be given in return. This project hopes to produce a powerful new database that combines DNA scans and extensive electronic medical history. Researchers hope to identify the more challenging multi-gene based diseases with tricky epidemiological questions such as factoring in effects of exposure to sunlight or alcohol consumption. This project has significant support from state funding justified by the potential for economic development. These projects along with the other databanks mentioned have committed to free access to all population data for university research teams. And as would be expected all of these public based databanks have guaranteed that privacy of the participants will be maintained and security is the number one priority.[28]

Private Biobanks

As these various publicly funded databases continue to emerge many private companies mostly in the United States are amassing large "biobanks" of DNA and tissue samples. These companies are in the business of supplying pharmaceutical or academic researchers with the materials needed for disease and drug discovery research. Donors are recruited with offers of specialized diagnoses and ultimate assurances of privacy. But ethicists are throwing much caution to the wind on these private ventures since there are no regulations that these companies must abide by.

One of the largest commercial biobanks is Genomics Collaborative, Inc. of Cambridge, Mass. They claim to have samples from 120,000 people from all over the world. A

common strategy for acquiring samples is to contract with physicians and medical centers to solicit for donors. Another company, DNA Sciences, Inc., in Fremont, CA, has collected 3,000 of its 18,000 samples via Website recruitment. These companies are also utilizing the Internet for selling access to their data such as with First Genetic Trust, (http://www.firstgenetic.net) enTRUST™ genetic bank. These commercial biobanks have to be viewed with the same concerns that would apply to any business venture. Guarantees and contracts don't carry much weight when companies fail or merge. Some states have laws specifying ethical standards for handling DNA samples, but federal regulations are weak at best. This may be the time for federal legislation to move on a law to license biobanks similar to one enacted by Iceland in 2000.[28]

Research Data

There are many scientists and ethicists who question whether there is really a need for large population databases. There is no proof that similar results cannot be realized from existing smaller research studies without exposure to the larger security risks inherent with a population database. Additionally there are existing biobanks that have evolved typically from health care providers that have great untapped potential. Health Databases from the Mayo Clinic, EPIC and various cancer societies highlight some of the research data that is also available. There is also a wealth of data available in smaller research studies typically at university medical research centers.

Security Concerns

Up to now this report has mostly provided genetic background information and the state of genetic data management. It is evident that there is an explosion of genetic information sweeping through our health care services, research institutions and pharmaceutical companies. There have not been any major publicized security breaches and so discussions of security concerns are based on the potential and likelihood of an incident. However, privacy concerns may not be tied to security. Many groups have access to our genetic records and for the most part they are governed by their industry's business code of ethics. But these lines are very gray with respect to controlling access to the data, meaning many groups have access to sensitive data that they should not, but the complexity of the data systems cannot account for every situation.

The foremost responsibility for health care data management is to keep the process running so people can be cared for. When control of medical information is under the control of a patient's physician there is the potential for privacy to be maintained. However, in today's medical facilities it is rare that a single physician can provide effective care without granting access of patients records to many health care workers. So in many situations we rely on integrity to manage our data access rules. Unfortunately in this post Enron world we no longer can rely entirely on integrity. Another area of concern comes from the value that genetic information carries and the opportunities it presents for groups to retrieve and analyze in order to produce information of value to certain service industries. The following sections will expand upon the inherent

weaknesses of data security and the hidden possibilities that exist from the computational power that we do not fully understand.

Privacy

One of the most important promises of the Physician's Hippocratic Oath is that of the confidentiality and the trust of the doctor-patient relationship. This promise is a cornerstone of medical history and may be one of the most trusted. However, the challenge to maintain patient privacy has increased as medicine has progressed beyond the traditional family doctor. A 1982 article by Dr. Mark Siegler illustrated the difficulty of keeping medical records confidential in a typical hospital where up to 75 medical professionals needed access to his patient's records.[29] The world of the family doctor may be gone and the world of a single health care provider may no longer exist and unfortunately the world where our medical records are truly confidential may also be an impossibility.

The concern for confidentiality of our medical records is now magnified by the need to protect our genetic information which is essentially treated as general medical information and is stored in a variety of electronic data repositories. The foundation of privacy is to insure that any information about an individual is never linked to that individual. In health care it is critical that information be shared for the sake of delivering quality care. HIPAA has laid some groundwork for the proper management of medical information to help control that privacy. The basics of HIPAA's approach are to circulate medical records without identity. Identity in our society today typically refers to

your name, address and your identifying codes most notably your social security number. If a medical record can avoid any link to such information then anonymity could be guaranteed. That was before the internet, huge powerful databases and the concept of electronic data mining. Today the process for maintaining one's anonymity probably needs to be reevaluated.

The work by Malin and Sweeney mentioned in the Overview of Privacy section of this report, presents the shocking reality that privacy in this electronic environment may be impossible to guarantee. Their research shows the ease at which a medical record can be linked to an individual with little more then a zip code, date of birth and gender. This should not be that surprising to us if we consider what has been going on for a number of years with the sophistication of electronic market survey analysis. Most of us would be shocked at the information that is known about all aspects of our lives driven by the value this information carries for steering targeted marketing strategies. The internet has provided the mechanism for this to occur, and we have to accept the fact that anyone with any buying potential at all has probably had their identity compromised. And it is probably safe to assume based on economic tendency, that anyone who has had genetic information generated about them would fall into this internet demographic.

Are we really naive enough to believe that our social security numbers, SSN, are confidential? The attention that is given to maintaining the importance of protecting our SSN is only about keeping the population's fears under control. As the need to protect sensitive medical records gains importance, it is unfortunately too late, the expertise required to compromise the security has already been developed. It has been developed

in response to a business need. Data is important as shown by the value of marketing research data, and personal health data is also valuable. Proper evaluation of an individual's medical information may mean economic success for a company dependent upon the costs associated with the medical health of that individual. Hence the financial motivation exists to take advantage of the opportunity, and federal or state regulations or a code of ethics is probably not going to stand in the way.

If privacy is impossible to maintain then is there still a chance that further advances in technology can provide security that can protect our privacy? Unfortunately it is the advance of genetic technology that is out pacing the advances in information security. Security is required administrative overhead that allows us to move forward in good conscience. Even if one realizes the hopelessness of maintaining privacy there is still the belief that data security will disallow any of our genetic information from falling into these processes of sophisticated data search and retrieval systems.

Genetic Data Security

The goal of all health data management computer systems is to create the ultimate secure working environment that accomplishes the goals for managing and utilizing health care information for the delivery of the highest quality of health care. What gets in the way of these goals is the economic reality of what it costs to provide such a system. The technology exists to totally protect any data that is considered to be sensitive to the privacy rights of the patient. However, the determination of the sensitivity associated with the various types of medical data is first needed. One of the main issues debated

with respect to genetic data is how it should be treated with respect to sensitivity. Is it just sensitive medical data or does it need to be classified in an entirely higher level of security access? The ramifications of this debate have a lot to do with the possible security. If genetic data is just sensitive medical data then it falls under the typical high security access requirements of a typical EMR. That just means that it has restricted access typically to health care providers and standard good practice security measures will be applied. But what if genetic data does need to be classified at a higher security level? Does a higher security level exist that fulfills the requirements for genetic records in today's EMR systems?

The answers to questions about genetic security are difficult and time consuming and unfortunately do not carry a high enough priority with the current management strategies for genetic data. Genetic data is allowing the health care provider to offer superior service to patients which is the driving force for the survival in this highly competitive and expensive health care industry. Genetic data is being handled with the security tools available to the industry today. These tools handle authentication, secure transfers of data on the network and the encryption of messaging or email. There is nothing unique about the EMR security. The tools exist that will accomplish the job, but the challenge is to define the job.

Authentication

The weaknesses of password authentication are primarily based on human error. The authentication systems today with rules about the number of characters and their content

are virtually uncrackable. But authentication is mostly for internal security, controlling those who have access to the medical records system. Vulnerabilities with authentication contribute to compromises of sensitive data, but are really insignificant with respect to the majority of incidents where data is compromised. The real vulnerability is with the enforcement of the authentication. The inconveniences associated with requiring strict authentication are generally the reasons why the process fails. Doctors just do not want to replace the clipboard with a workstation login for every patient encounter, let alone be told to change their password on a frequent basis. And the economics of the EMR systems don't work when we incorporate the costs of data transfer from paper to computer so where do you think the process fails. It fails with relaxed security that is a result of compromises by administration and users under the disguise of efficiency. Help is on the way with new proximity based biometric smartcard authentication options, but again this is not the critical link, this is just the most publicized.

Encryption

The use of encryption is the key to physical security. The basic concept of a public and private key that is used to encrypt and de-encrypt sensitive health care data solves the problem. If all the details and procedures could be worked out there would be no security problems. But unfortunately a sophisticated encryption system requires the management of a complex key that presents many of the same challenges from securing authentication in the traditional password system. And it also goes back to the old problem of inconvenience. Not only an inconvenience for the user to be on top of the process, but also for the computer administration to implement the system. We would hope that our

computer administration would always have a handle on this management. But the information technology support in health care does not tend to keep pace with the needs. This is a very real problem based on the fact that the health care industry does not invest heavily in Information Technology, IT. The trend has been for about a 2 percent budget for IT, however, predictions now point to a health care industry that is probably going to shift IT spending to about 10 percent of the budget.

The standard for applying encryption to genetic data has been set by the Iceland deCODE approach which utilizes a third party encryption system. The third party which is the Iceland government's Data Protection Commission initially handled the security of the encryption key with a small number of government employees. However, this process which was slow and vulnerable to human weaknesses has since been replaced by an automated Identity Protection System developed by deCODE and monitored by the Data Protection Commission.[30] The driving force to develop what may be the most secure genetic database encryption system is the need to maintain identity of the samples. The common security approach is to go down the path of anonymity; however, deCODE contends that maintaining identity is critical since the future cannot be predicted, and that future benefits may be lost if identity links are not maintained.

The importance of this approach is the fact that an independent third party supposedly controls the access to the sensitive data, thereby offering the individual security assurance and an option for the possible removal of their personal genetic information. The Markle Connecting for Health project talks about the creation of a personal health record controlled by the patient. If this could actually be designed then a lot of security and

privacy concerns could be eliminated. But there are valid reasons why an individual may be better off without having this control. Greg Smith's initial thesis research focused on biometrics and how one's biometric template information could be used to encrypt their highly sensitive personal health data. In fact options such as multi-level encryption keys were researched for the design of a National Health Database. This did not lead to valid options since biometric data would still need to be reduced to an encryption key that would be as vulnerable as a randomly generated code.[14] But this discussion warrants further study as an option for providing direct personal control for securing personal genomic information.

The use of encryption for securing sensitive health care data is not in question. The question is what and when to encrypt the data. Encryption should occur at the database field level, and with newer EMR systems this can be accomplished without sacrificing to much in the way of performance or usability. But the decisions required for what data warrants encryption level security and then the application of the security structure to carry out the decisions do come with a high cost. The complexity of an EMR system that provides complete security controlled by the patient/physician and allows authorized access to various health care workers and related employment and financial organizations is quite high. The economics of providing health care in this country tell us that this system will settle for a security solution far short of that.

Email and Health Data Web Portals

The need for communication of health care data with patients, health care providers, government agencies and insurers forces the use of email and other internet based services. Incorporating encryption measures into these communications is becoming the norm; however, again we have a system that leaves itself open for misuse. The potential for a compromise of a data transfer is generally not the concern. The concern is for the growing number of data systems that have to support these transfers. Email and web database systems which have to manage the vast amounts of data that is transferred present the risk. A prime example would be the Microsoft email and web products. The products Exchange and SQL Server have large market shares, but they are also plagued by numerous security vulnerabilities inherent to the recent versions of the Windows operating system. Now consider how much sensitive health care data resides on Microsoft Servers. This is a know vulnerability that should be given more attention. And this is not just a Microsoft problem; we have a computer software industry that has been driven to satisfy usability at the expense of maintainability. Again this is just an economic reality of the world we live in. So the serious vulnerability facing the protection of our health care data may not rest with users or EMR systems, but rather with the repositories that facilitate the exchange of the data.

These repositories which are often sophisticated database systems are generally quite secure. However in many cases data must be exported to presentation applications, the most common being the health or patient data web sites. The competition to provide patient access to personal health records via the Web is a new strategic initiative of most

of the larger health care providers. This information delivery presents another challenge for defining what data needs to be reported and what level of security is adequate.

The need for sharing health care data across the networks is a requirement of our interlinked health care system. Genetic testing is performed over a diverse geographic map which has samples transported and results returned. Tests that are performed must be communicated to insurers and government agencies. Results need to be returned to patients for timely informed decisions to be made. Health provider websites are in great competition to offer personal health data to the patient. All of this need to provide data is putting an incredible strain on the IT infrastructure which is just trying to keep up with the process. It is no wonder that there is no time or motivation to investigate more fully the security needs specific to genetic data. The overall system can only hope to keep up with the massive job of managing health care data in general.

Genetic Data Mining

Data Mining that builds Data Warehouses is a common practice for companies that require knowledge to be extracted from the vast amounts of computer data. The same concept applies to genetic information. A similar data mining technique has been under development for a number of years by the major pharmaceutical companies in order to consolidate the vast amounts of DNA research data that exists in the many public databases such PubMed and Entrez. An example of such a tool is the Biological and Chemical Information Integration System, (BACIIS), which has been developed within the School of Engineering and Technology at IUPUI.

> BACIIS uses a semantic data model to integrate dynamic, heterogeneous, autonomous, geographically distributed, and semi-structured web-databases with total data source transparency. It allows the execution of global multi-database queries that extend beyond the boundaries of individual databases.[31]

These types of data mining tools are already picking up personal genetic information without specifically searching for it.

Interpretation of Genetic Data

The surface value of most medical data is not considered genetically sensitive. However, technology has allowed us to run powerful correlations of the data to make interpretations that lead to genetic information. An example may start with a sequence of medical records where specific tests are run which would typically signal a genetic disease concern. Then link pharmacological data that might identify specific levels of proteins that would signal the possibility of an irregular gene production. If an identity can be applied to these records which we know can occur, you then have a candidate for further investigation. This could occur with a single nucleotide polymorphism, SNP, data that could be used to match the identity to what was thought to be a totally anonymous record in a public research database. Now you have far more information about an individual then any group should be in possession of. Unfortunately the nature of genetics and heredity also allows for the linkage of information to everyone in their family tree for a few generations back to identify many more individuals with a possible genetic concern.

Genetic Records are Not Anonymous

The ultimate identity of an individual is their genome. Hence it cannot be said that a genetic record is anonymous. Everyone's genome is a unique map which would provide an identity if sequencing was available for everyone. But even with incomplete sequence records, unique identity can be confirmed through matching unique DNA polymorphisms. This is why we now have a problem with the large genetic databanks that exist under the security justification that the records would be anonymous. We know how easy it is to place identity with existing medical records. Combine the ease of identifying a medical record with the valuable information that could then be tied to that identity such as specific gene mutation information and you have a situation where today's powerful search engines and bioinformatics tools could find ways to identify genetic records that exist in these databanks.

Genetic Spyware

The category of spyware software relates primarily to background processes or daemons that end up on computers for the purpose of acquiring information about computing activity and data that moves through that computer. This technique is extremely valuable process for building targeted marketing information for individuals. When you evaluate spyware at the client level it appears to be very simple and relatively harmless. However, when you evaluate the huge data repositories that spyware feeds then you realize the incredible power that spyware enables. Applying this spyware strategy to the acquisition of genetic data is not that far fetched. If there is the potential for profit then the

application will follow. The form of these not so legitimate genetic databases would be to first consolidate all publicly available genetic records. Then the acquisition of genetic information through unscrupulous spyware technology could be compared and added to these data repositories. These communications should be protected, but one can question how tight security will be on say the frequent data checks of the new anti-terrorism DNA databases or the transfer of "anonymous" test information between labs. Then assume that some of the most advanced hacking and decoding techniques could be put into the process and we could then envision how a genetic profile could be created for all people who are involved with genetic medical situations.

So what if there was a market for personal genetic information that could be accumulated via these spyware tactics. This is not to say that legitimate companies who might benefit from information such as this would ever participate in the actual acquisition of unauthorized personal genetic data. But it could be conceived that these companies would be willing to run a genetic credit check on someone through these spyware built databases. The economic benefits probably justify the risk. Spotting a small percentage of high health risk individuals in order to decline employment of insurance coverage could be the difference between fortune or failure.

Ethical Concerns

The Genomic Era has brought about incredible advances in genetic technology but it has also spawned a commitment to consider the ethical implications that accompany the genomic journey. The early discovery of the process of recombinant DNA in 1973 by

Boyer and Cohen whereby DNA from one organism was combined with DNA of another brought about humanity's greatest fears. Fears of human engineering, cloning and run-away mutations caused the scientific community to call for an unprecedented moratorium on scientific experiments with DNA. Scientists were actually shocked by the possiblility that their research was uncovering. They understood that the economic implications of genetic research on a global stage required an increased attention to be paid to the social implications and responsibilities. Even though the moratorium was voluntary it was honored by almost all of the scientific community. In February 1975, 150 scientists from 13 countries, along with attorneys, government officials and members of the press, met at the Asilomar Conference Center to discuss recombinant DNA work, consider whether to lift the voluntary moratorium and, if so, establish strict conditions under which the research could proceed safely. The moratorium was replaced with rules and cautions for genetic research. At no other time has the international scientific community voluntarily ceased the pursuit of knowledge before any problems occurred, imposed regulations on itself and been so open with the public.[32]

Ethical, Legal and Social Issues, ELSI

We can probably point to the Asilomar Conference as the birth of a research commitment to the ethical, legal and social issues associated with genetic research. One of the main goals of the Human Genome Project was to address the Ethical, Legal, and Social Implications (ELSIs) that would:

- Analyze and address implications of identifying DNA sequence information for individuals, families, and communities.
- Facilitate safe and effective integration of genetic technologies.
- Facilitate education about genomics in nonclinical and research settings.[21]

Proof of the impact that ELSIs are having on biotechnology is evident by numerous Bioethics Research Centers that have been created in academia and health care organizations in the last decade. But just like the fear that originated from the rapid discoveries surrounding the manipulation of DNA in the 70's. Are we now experiencing new rapid developments in genetic research that may caution the scientific community to again consider a moratorium to make sure that we have the ELSIs covered?

Regulatory Organizations

An outcome of the 1973 moratorium on genetic research was the establishment of the Recombinant DNA Advisory Committee (known as the RAC) by the National Institutes of Health, NIH, in October of 1974. The RAC developed a set of Guidelines that were first published in 1976 and have been revised periodically since then. These Guidelines include a comprehensive description of facilities and practices intended to prevent unintended release or inadvertent exposure to either genetically modified organisms or recombinant DNA. Compliance with the Guidelines is mandatory for investigators at institutions receiving NIH funds for research involving recombinant DNA.[33]

Since then many agencies have been setup or roles of existing groups have changed to better deal with the genetic issues confronting the scientific, academic, commercial, government and public communities of our world. The big players in the United States will continue to be the NIH, Department of Health and Human Services, HHS, the Presidents Council on Bioethics, The Food and Drug Administration, FDA, the Office for Human Research Protections, OHRP, the National Science Foundation, NSF, and the Department of Energy, DOE. These large organizations have chartered various regulatory or advisory committees to mostly act as advisors or watchdogs over the industry. Actual legislation for addressing genomic concerns has not met with much success. What we tend to see is that the United States is more concerned about the regulation of the research, scientific and pharmaceutical consequences of human genome activities.

The United Kingdom has played a major role in this genomic era with milestone scientific contributions such as the cloning of the lamb, Dolly, and the controversies of the development of a National Biobank. In 1999 the British government formed the Human Genome Commission, HGC, to address regulation of genetic products and processes. The HGC's overriding priorities in biotechnology and genetic modification are to protect the health of the public and to protect the environment. They divided their responsibilities into three areas: the Advisory Committee on Genetic Testing, the Advisory Group on Scientific Advances in Genetics and the Human Genetics Advisory Commission. More direct government involvement has come about in the UK due to the

existence of socialized medicine. This is also the case in other countries such as Canada that have already established national health plans and databases.

Conclusion

The expected conclusion for this topic would be that we must be more concerned about the security of genetic information. But it is far more complicated then that. Everything should be more secure and software systems should be failsafe and IT professionals should be more highly trained, etc. Possibly our conclusions should be to limit the proliferation of genetic research that is driven by the financial greed under the charter of improved health for mankind. This is about dealing with the incredible advances in technology with a mixing of pride from scientific accomplishment that is signaling us to be more concerned about protecting our genetic privacy. The conclusions are also about accepting the reality that this issue exists and because of the great good that can come from exploring our genome we have to deal with the fact that our genomic information will become a more active part of our existence.

Unfortunately if the answer truly required that we slow down the genomic research engine, the power and momentum of the pharmaceutical industry probably cannot be curtailed. Big pharma is so heavily committed to the development of its next block buster drug that it cannot afford slow downs. The high cost of research and development require timely successes for the companies to survive. This is not to say that the pharmaceutical industry is not doing all that it can to protect our genetic privacy. What it is saying is that this industry creates great demand for the many forms of genetic data. The demand that is being fulfilled by public and private research ventures that are

establishing collaborations across many entities dilutes any centralized authority's ability to guarantee that all is being carried out safely.

Concern for the rights and privacy of patients can get lost in the larger population privacy issues, but the root of securing our personal genome begins with securing every piece of personal health data ever generated about us. However, securing our medical records only controls the access list to that information which unfortunately for the sake of a quality health care system has become quite large. We now must drill down to greater detail concerning our medical records. We must identify the current as well as potential sensitivity of all health care data. In some cases this may require that we go back and reclassify data and possibly apply additional security to existing records. Our medical record systems and biobanks need to be pushed to greater sophistication. We need to reconsider the sensitivity of data with respect to information it might offer through linkage to other sources of data. This may mean that we must redesign network access rights. But it may mean that we need to investigate how ethical standards could be applied for the use of this genetic information.

Unfortunately the problem of securing our personal genome is a result of the advanced technology that provides access to more data and interprets that data with more powerful knowledge building algorithms. Are we ready to accept the consequences from the compromise of our personal genetic information? All indications show that we are. Human objection is strong against the controversial issues such as cloning, but not so strong as to totally ban it. There is an underlying belief by individuals that mankind will do what is right and that their liability will be covered by the system. The intrigue of

great medical breakthroughs from genetic discovery far outweighs our fears of genetic discrimination. If that is how we proceed then we place great ethical responsibility on the health care industry and all entities that try to regulate or utilize the personal genetic information that they will produce.

Discussion

There are a number of key areas of discussion about securing our personal genome. We know that the technology exists and that it is becoming faster and more affordable. Access to our personal genome is a reality today for those who could justify the cost. But what does that knowledge do for us? The ability to look into the crystal ball of our future health is a 2 sided sword of good and bad news. However, if we can fix the bad news, than there is probably no cost too great. Now consider how this gets complicated as the costs of acquiring our personal genome come within the grasp of a typical health insured individual. The questions really start to surface.

Health Care Economics

How does the health care system determine what genetic testing should be done or even if an entire genome map should be provided per patient? Will it all be paid for under the justification of preventative medicine? The economics of the system will dictate what medical costs can be absorbed by the patients or insurance providers. We can hope that biotechnology will create incredible and affordable cures for genetic disease and repairs for genetic mutations. Unfortunately the trend of what it takes for pharmaceutical companies to profit from research and development does not equate to a low priced cures or treatments. Looking at pharmaceutical research into genetic disease today we see a shotgun approach that is very expensive. The use of improved technology in conjunction with knowledge gained from the evolving Biobanks around the world could change this

to a more targeted production of genetic based drugs or therapies. Many health problems could be taken care which would improve our quality of life. A simple question to ask at that point might be the value of the quality of an extended life. What new health problems will now be able to surface after we eliminate such a broad array of genetic based problems? Will it all be worth it? How will we justify not providing these services to the less fortunate.

The concerns raised over the justification of providing new cures and therapies for genetic disease are a good problem to have. This could all play out if genetic discovery goes well and treatments are easy to replicate. But results so far indicate that our genetic make up is so unique, that dealing with treatment for genetic mutations is extremely complicated and does not seem to present standard therapies for specific diseases. This can be understood by the complicated set of biochemical relationships that tend to be the consequences of genetic problems. We have come a long way in getting a map of the human genome, but we may still a long way from understanding the biochemical relationships of our gene products.

If we find that the pursuit of gene based cures and therapies does not move more easily and costs continue to stay high then we begin to place extreme pressure on our health care system to manage the services. The United States already has the most expensive per capita health care costs in the world by a fairly large margin. If we see the global economic competition heat up more, the U.S. will have great difficulty in justifying such a large component of the gross national product being devoted to health care. At that point we will still have genetic based medical care, but it will be developed only to cater

to the rich. And the social pressures that will mount from this will be unhealthier to society then the diseases that could be cured. Would we really be any worse off if we had never discovered how to work at the DNA level?

Genetic Discrimination

The information in this report and the general pulse of society indicate that genetic medicine and the potential that it holds is good and will be cultivated further with an acceptance of the risks to privacy and biological catastrophe. It is a foregone conclusion that our genetic privacy does not exist and may never be regained, but this is not a major concern as of yet because the consequences have not developed to directly affect the people. What we do not know does not tend to hurt us. Have we not always been discriminated against by insurers and employers? Have we not come to accept the in discrepancies that exist between the ends of the socio-economic scale? What will be so different about people having to deal with the denial of health insurance due to a propensity for heart disease or refusal of employment because of knowledge of future health problems? The difference might be that there may be no protection for the privileged portion of the population. Genetic discrimination may actually target the wealthy since they have a longer trail of quality health data.

It would be an interesting reaction for society's privileged to have to deal with the risks of discrimination weighed against fountain of youth value from advanced genetic diagnostics and medicine. Would we then see a concerted effort to improve security of the computer systems housing this personal health data? It could create some interesting

splits of the health care system. We could possibly see more of a country club exclusive model for health care providers taking care of the privileged with socialized medicine absorbing the remainder of the population. At least a structure such as that would allow us to know where we stood. What remains the scariest scenario is that we never really know if our personal health data is being used to discriminate against us. That we never know that it is possible to analyze our personal health data to arrive at discriminatory information. That may be the accomplishment of this thesis report to present an alert that the possibilities for misuse of our health and specifically our genomic information does exist.

References

1. Venter, Craig, The Institute for Genome Research's (TIGR)'s 14th International Genome Sequencing and Analysis Conference, October, 2002

2. "Summary of the HIPAA Privacy Rule", Office for Civil Rights, U.S. Department of Health and Human Services, April 11, 2003, (http://www.hhs.gov/ocr/hipaa)

3. Malin B., Sweeney L., "Compromising Privacy in distributed Population-Based Databases with Trail Matching: A DNA Example", CMU-CS-02-189. December 2002

4. Pangolos G., Mavridis C., Ilioudis C., Georgiadis C. "Developing a Public Key Infrastructure for a Secure Reginal e-Health Environment". Methods Inf Med 2002; 41 pp 414-8

5. Brandner R., Van der Haak M., Hartman M., Haux R., Schmucker P. "Electronic Signature for Medical Documents – Integration and Evaluation of a Public Key Infrastructure in Hospitals". Methods Inf Med 2002; 41 p p321-30

6. Kawazoe Y., Toshikazu S., Yamamoto M., Ohuchi A. "A Security System for the Personal Genome Information at DNA level". IEEE Computer Society Bioinformatics Conference (CSB'02)

7. "Genetic Information: Legal Issues Relating to Discrimination and Privacy", Congressional Research Service - Report No. RL30006 (July 19, 2001)

8. Szekely, Peter, "Railroad to Pay $2.2 Million in DNA Test Case Illegally Testing Workers for Genetic Defects", Reuters News Service, May 8, 2002. (http://www.mindfully.org/GE/GE4/Railroad-Workers-Genetic-Defects8may02.htm)

9. "Genetics Laws and Legislative Activity", National Conference of State Legislatures, 2003, (http://www.ncsl.org/programs/health/genetics/charts.htm)

10. "Position Paper on Genetic Discrimination Legislation", National Council on Disability, March 4, 2002, (http://www.ncd.gov/newsroom/publications/geneticdiscrimination_positionpaper.html)

11. "Cancer Susceptibility Genes:Testing For", Section 3 Clinical & Administrative Policy & Procedure Manual, Lahey Clinic, Burlington, Mass, (http://www.lahey.org/Pdf/Ethics/PDF_Policy/5119.pdf)

12. "The supply of genetic tests direct to the public, A consultation document", Report by the United Kingdom's Human Genetics Commission, July 2002, (http://www.hgc.gov.uk/testingconsultation/testingconsultation.pdf)

13. "Connecting for Health, A Public-Private Collaborative", Markle Foundation, All Working Group Reports, Surveys and Recommendations, June 5, 2003, (http://www.connectingforhealth.org)

14. Smith, Greg H., Website presenting research for securing highly sensitive health care data by use of Biometric IDs, IUPUI, (http://www.biometricid.org)

15. Anderson, Curt, "Ashcroft Seeks $1B for DNA Testing", Associated Press, March 11, 2003

16. "Whose hands on your genes?", Report by the United Kingdom's Human Genetics Commission, November 8, 2000, (http://www.hgc.gov.uk/business_consultations2maintext.pdf)

17. Weisfeld, NE, "Mapping public policy on genetics", Gene Therapy, 2002, 9, pp 662-666

18. "The deCode Approach", Decode Genetics. Iceland, Website (2003), (http://www.decode.com)

19. Collins, Francis S., Green, Eric D., Guttmacher, Alan E., Guyer, Mark S., "A Vision for the Future of Genomic Research". Nature April 24, 2003, 422 pp 835-847

20. "Genome Programs of the U.S. Department of Energy Office of Science", Founder of the Human Genome Project, (http://DOEgenomes.org)

21. Marrs, Kathleen A., "Biology 540 - Topics in Biotechnology", (http://www.biology.iupui.edu/biocourses/Biol540)

22. Hlodan, Oksana, "For Sale: Iceland's Genetic History", An Action Bioscience.org Publication, June 2000 (http://www.actionbioscience.org/genomic/hlodan.html)

23. Landro, Laura, "Wired Patients", The Wall Street Journal Online Commentary, July 1, 2003, (http://webreprints.djreprints.com/780271040687.html)

24. "SNOMED Clinical Terms® To Be Added To UMLS® Metathesaurus®", July 1, 2003, (http://www.nlm.nih.gov/research/umls/Snomed/snomed_announcement.html)

25. "The Universe of Genetic Testing", Lab Tests Online, (http://www.labtestsonline.org/understanding/features/genetics.html)

26. Roa, Benjamin B., Boyd, Alfred A., Volcik, Kelly, Richards, C. Sue, "Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2", Nature Vol 14, No 2, October 1996, p 185

27. "Special Requirements for Electronic Medical Record Systems in Pediatrics", Pediatrics, Vol 108, No 2, August 2001, pp 513-515

28. Kaiser, Jocelyn, "BIOBANKS: Population Databases Boom, From Iceland to the U.S., Science Nov. 8, 2002 298: pp 1158-1161

29. Siegler, Mark, "Confidentiality in Medicine – A Decrepit Concept", New England Journal of Medicine, Vol 307, No 24, December 9, 1982, pp 518-521

30. Overby, Stephanie, "Iceland's Dilemma: Privacy vs. Progress", IT and Health Data, CIO Magazine, July 15, 2001

31. Ben-Miled, Zina, "Biological and Chemical Information Integration System", (BACIIS), http://baciis.engr.iupui.edu

32. "Ethics", Biotechnology Industry Organization, BIO, http://www.bio.org/er/ethics.asp

33. Recombinant DNA Advisory Committee, Office of Biotechnology Activities of the NIH, http://www4.od.nih.gov/oba