Exploring the Key Informational, Ethical and Legal Concerns to the Development of Population Genomic Databases for Pharmacogenomic Research

Julia Gladstone
Bryant University, Smithfield, RI, USA
jgladsto@bryant.edu

Abstract

The completion of a high quality comprehensive sequence of the human genome has lead to the discovery of genetic links to complex diseases and the development of target drugs. Population genetic databases (PGDs) are a powerful resource to the systematic study of the genetic component of disease; in the quest to understand the impact of genetic factors on drug response data from laboratory experiments, computational methods and clinical studies must be integrated. The establishment of a pharmacogenomics knowledge base entails complex information management balanced with the appropriate legal and ethical standards. This article reviews some of the ontology development challenges and examines the need to establish a legal structure which creates the proper intellectual property incentives to develop PGDs and the ethical standards that must be upheld so that the terms of informed consent are legitimate.

Keywords: database, intellectual property, ethics, biotechnology, genetics, privacy

Introduction

As time passes man becomes more sophisticated at unraveling the mysteries and miracles of life. The completion of a high quality comprehensive sequence of the human genome has lead to the discovery of genetic links to complex diseases and the development of target drugs. This field of biotechnology which applies bioinformatics and genomics to the development of pharmaceutical products is known as “pharmacogenomics,” literally pharmaco- meaning drug and genomics meaning the study of genes and their function (Binzak, 2003). The digitization of information has moved scientific research from the wet lab to the computer lab. Large scale population based DNA collections, population genetic databases (PGDs), are proving to be particularly useful in sorting through genetic variants and environmental factors in order to discern which permutations are significant contributors to disease. As the scientific and the pharmaceutical communities move forward towards providing “individualized medicines” many legal and ethical issues need to be addressed.

The literature that discusses the ethical, legal and social implications (ELSI) of the human genome project (HGP), pharmacogenomics and population genetic databases is vast, ranging from common law rules of ownership and “reach through” agreements to privacy rights and profit sharing. This
Exploring the Key Informational, Ethical and Legal Concerns

article focuses on the need to establish a legal structure which creates the proper intellectual property incentives to promote the development of population genetic databases and the ethical standards that must be upheld in creating these databases. This article will describe the database protection laws that impact the development of PGDs together with ethical policies around issues of informed consent can influence the growth of this research.

The first section, *the Development of the Science of Pharmacogenomics*, of this article explains the background of the (HGP), the field of bioinformatics and the main trends in the study of population genetic databases. *An Analysis of Intellectual Property in Genomics* sets out the intellectual property rules that have been used to create incentives in the development of databases in general and in genomics more particularly. The ethical considerations that are involved in obtaining informed consent are reviewed in the third section, *What is Informed Consent: An Ethical Duty* and the Conclusion provides an overview of the how policy decisions regarding intellectual property rights and human beneficence can encourage growth in the field of pharmacogenomics.

The Development of the Science of Pharmacogenomics

The deciphering of the chemical structure of deoxyribonucleic acid (DNA) began with the work of James Watson and Frances Crick in 1953. The study of genomes has been an international project aimed at obtaining a detailed map and complete sequence of a variety of organisms since the late 1980s. Prior to the sequencing of the human genome 20 species’ genomes were available for viewing and use on the Internet. The first organisms to be sequenced were viruses. Genetic analysis of sequences of organisms other than Homo Sapiens established a common language of genetics and the similarity of genes meant genetics was being advanced for all species.

The HGP is the term used to describe a collective group of projects from around the world organized in 1980 to create an ordered set of DNA sequences from known chromosomal locations and to develop new computational methods for analyzing genetic maps and DNA sequences (Binzak, 2003). In 1988 the Polymerase Chain Reaction (PCR) was devised which greatly increased the rapidity of DNA manipulation. Microarray technology or “lab on a chip” techniques are the more recent biomedical informatics techniques have greatly expanded genome-wide discovery. The data capacity afforded by microarray technology extends the molecular understanding of disease in order to find pathways to modulate it.

Although several countries participate in the HGP the United States initially made the largest financial contribution through the Department of Energy (DOE) (Macer, 2000). The DOE has had a long term interest in the effects of radiation on human health, and knowledge about DNA would inform this investigation. Regardless of the political posturing which led to the massive funding of the HGP, politicians soon became interested and involved. The National Institute of Health was selected as a cosponsor because of the need for biologists and it now spearheads the HGP. Through the late 1980s and 1990s the number of human genes that were sequenced grew exponentially and additional countries and private enterprise joined in the effort. In June 2000 the human genome was sequenced to a level of sufficient quality that it was considered by scientists to be complete.

It is fortunate that the original researchers saw the genome as more than merely a bundle of genetic information and chose to sequence the entire genome rather than only the “coding DNA;” the organization in the context of the surrounding information is important (Bentley, 2004). Only 3-5% of the genome contains genes which code for protein, yet the remaining “junk” DNA is proving to have an enormous amount of functionally important information, especially as the science of pharmacogenomics evolves (Macer, 2000). The DNA that has been sequenced for the HGP is actually a composite of different human tissue cell lines from several people. Less than
1% of the nucleotides in our DNA vary between different people so for the initial genome sequencing it seems not to matter whose genome is sequenced.

DNA is a molecule with four chemical components called nucleotide bases that are linked together into long sequences, two of these sequences are paired together creating a chemical structure called a double-stranded helix. There are only four distinct nucleotides that make up the double helix; these are adenine (a), cytosine (c), guanine (g) and thymine (t). It is the combination of these four bases, their sequence in a strand of DNA, that give human life its character. A genome, therefore, is the continuous thread of DNA sequences on all 46 chromosomes in the human cell (Gardner, 2003). The human genome sequence is estimated to comprise 3 billion pairs of these letters. A gene is a length of DNA that contains sufficient information to code for or make a protein. The human genome contains approximately 31,000 genes.

A mutation occurs where a nucleotide base is different from what is normally found at that location on the chromosome. With the current level of scientific knowledge anyone can have their entire genome sequenced and compared to the prototype homo sapiens genome to find discrepancies. Certain mutations may be harmful and cause disease while others may have no impact on how the protein functions. When a mutation occurs in more than 1% of the population, thereby indicating a commonality to the variation it is called a single nucleotide polymorphism (SNP). SNPs have great significance for biomedical research and several private and public groups have organized the SNP Consortium based upon principles of open access of information similar to the HGP, to encourage further discoveries in this area.

The success of the HGP in sequencing the human genome would not have much value beyond that of its raw data without the tools and expertise of researchers in bioinformatics. The greatest use of sequence information is found in comparing it to genomic structures the function of which is known. Bioinformatics is an “in silico” biology which conceptualizes biology in terms of molecules and applies informatics techniques derived from math, computer science and statistics to understand and organize the information associated with these molecules on a large scale (Greenbaum, 2003). Bioinformatics offers a multi-pronged approach to mining, analyzing and deciphering genomic data. Databases that are composed of complex architectures designed to integrate different forms of data such as DNA sequences, metabolic pathway data, qualitative protein-protein interactions or three dimensional atomic coordinates are extracted, collected and coalesced into more definitive resources.

Bioinformatics provides the methods and algorithms for extracting this data. As the study of monogenic and complex multigenic diseases become more prevalent it has become important to integrate information not only from the traditional bioinformatics field an example of which follows, but also medical informatics and clinical and epidemiological research. It is the elucidation of the human genome which has driven the synergy between bioinformatics and medical informatics; biomedical informatics is a new field of knowledge creation which presents significant management and organizational challenges.

In the effort to understand the impact of genetic factors on drug response data from laboratory experiments (medical informatics), computational methods (bioinformatics) and clinical studies must be integrated. The sheer volume of this data together with the differences in how the information is described and concepts expressed presents challenges for structuring and managing pharmacogenetic data and knowledge. There are several efforts ongoing to develop common common ontologies for information processing and retrieval which is the foundation for building a useful database. (Klein, 2004)

Much of the basic genetic discovery is based upon sequence homology between known and unknown proteins. Having the ability to search through a database of similar sequences and output a list of sequences which is statistically similar to the one you have uncovered is a meaningful first
Exploring the Key Informational, Ethical and Legal Concerns

step in the bioinformatics research process. Basic Local Alignment Search Tool (BLAST) which can be accessed on the Internet offers this service and is used by a variety of researchers to start their research process.

Bioinformatics is the science responsible for revealing the DNA sequences which comprise the human genome and the development of genomic databases. Population genetics deals with the characteristics of genes within a population; it evolves from epidemiological research studies (Hsieh, 2004). Isolated populations are ideal groups for the study of genetic diseases because evolutionary and environmental factors are held nearly constant when comparing variations between individuals within the population. Historically, population geneticists have focused on small isolated groups of people to make a correlation between phenotypic characteristics and genetic matches. The results have lead to the identification of single mutation diseases such as Down’s syndrome, Huntington’s disease and sickle cell anemia (Racine, 2003).

The usefulness of population genetic databases increased significantly in terms of scale, scope and format with the information from the HGP. As genetic relationships can be analyzed at the DNA level the underlying causes of complex diseases can be understood. Drugs can be designed to target the sources of the disease at the level where protein coding takes place. deCODE genetics (deCODE) which is the biotechnology company that created the population genetic database in Iceland has eight major common diseases under investigation. In the research of heart attacks, deCODE scientists have discovered variations in the gene encoding the 5-lipoxygenase activating protein that double the risk of the disease by increasing the production of leukotrienes. deCODE has a license for the DG031 compound that inhibits leukotriene synthesis, and Phase II clinical trials are now underway.

The use of populations to study genetic variations can involve many approaches and many goals. Target drug discovery which is exemplified in the deCODE case is only one of several possible purposes or aims (Hsieh, 2004). There are more than one hundred population genetic database studies ongoing at the present time with varying goals. The International HapMap Project is an international research consortium including China, Canada, U.S. Japan and the U.K. which collects DNA samples from volunteers in Nigeria, Japan and the U.S. The goal of the HapMap Project is to determine common patterns of sequence variance in the human genome and to make it freely available to the public. Since 1999 the United Kingdom has been developing a biobank consisting of DNA blood samples and lifestyle information on 500,000 volunteers age 40 to 70 in order to study how genetic and environmental factors affect the risk of developing common diseases. Similarly, in Quebec, Canada a population genetics project called CARTaGENE involving 50,000 individuals aimed at producing a map of genomic variation in health determinants is underway (Racine, 2003).

An Analysis of Intellectual Property in Genomics

The goals of the researchers are numerous and varied yet improving the well being of human life is a generally acceptable goal for all projects. PGDs can lead to discoveries about human disease and we posit that we should encourage the responsible development of these databases. This requires creating the proper legal and economic incentives to spur innovation and contribution to these projects.

Intellectual property rights provide a financial incentive for researchers to devote time and energy into developing socially useful inventions which are expensive to produce yet inexpensive to replicate (Rai, 2001). Genomic databases that are derived from human populations are an example of such property. Data compilations have been protected under U.S. Copyright Law as long as the database consists of original creative works. In 1991 the U.S. Supreme Court ruled that owners of a database could not claim copyright protection for a database solely because of the “sweat of the

The problem for many scientific fields and the databases that they service is that they are considered merely compilations of facts and thus will not qualify for U.S. copyright protection. This lack of legal protection has not lead to an open access to genomic databases, however, because protections are found through the trade secret model, licensing and technological self help (Greenbaum, 2003). The Digital Millennium Copyright Act (DMCA) prohibits the circumvention of technological locks that owners of proprietary information such as databases can use to prevent access. In some cases the DMCA may even prevent fair use. The benefits of the trade secret model have been employed at different stages of discovery by participants in the pharmacogenomics industry. It is not uncommon either, for database providers to license out information for hundreds of thousands of dollars.

Outside the United States, in the EU in particular, the law provides protection for databases. The EU Database Privacy Directive, Directive 96/9 EC which was fully implemented by 2001 affords a sui generis protection for databases produced in the EU. The Directive grants the database owner some copyright protection for the aesthetics of the database and a more general protection against unfair extraction of significant portions of the data. This sui generis protection lasts for 15 years from each time the database is changed which is in essence grants a perpetual protection (Greenbaum, 2003).

The tension that is created by the various forces, namely federal law versus common law and technology, competing to master the nature of the intellectual property interests in databases reflects the much broader debate about the proper scope of intellectual property rights in “basic research;” pharmacogenomics is a quintessential example. Can we encourage innovation without enclosing access to knowledge? Molecular biology was initially conceived as a highly theoretical abstract inquiry of applying the analytical methods of physics and chemistry to the study of life as a collection of biological macromolecules. At the time of Watson and Crick’s determination that the double helix structure served DNA’s functional purposes by encoding cell protein, most of this research was being conducted in academic institutions and it was funded either by the government or nonprofit foundations.

The basic science norms which guided this research is know popularly today as “open science.” These traditional scientific norms promoted a public domain, freely available scientific information, independent choice of research topics and respect for uninhibited scientific invention. The term “communalism” was coined to express the view that scientific knowledge constituted a common heritage (Rai, 1999). The recognition by scientists of their dependence upon a cumulative effort is best reflected by Newton’s famous remark, “If I have seen farther it is by standing on the shoulders of giants,” (Rai, 1999). The strongest driver under this norm is invention itself. Individualism, innovation and contribution to the common stock of knowledge leads to the highest levels of recognition and prestige. Hence, the norm of invention leads scientists to compete vigorously.

In the mid 1970s there was an economic and legal shift in the United States in favor of stronger intellectual property laws which affected the traditional norms in science. Congress found that putting results in the public domain was not efficient to the goal of developing marketable commercial products. The aim of simply promoting science seems to have taken a back seat. Establishing monopoly rights in inventions through patent law was believed to be the necessary incentive to translate innovations into marketable products; patents were not necessarily being used as an incentive to create (Rai, 1999). Thus several technology transfer statutes were enacted and in 1982 Congress created The U.S. Court of Appeals for the Federal Circuit by the merging of the
Exploring the Key Informational, Ethical and Legal Concerns

U.S. Court of Claims and the U.S. Court of Customs and Patent Appeals. the special Federal Circuit Court of Appeal to hear patent cases.

In 1980 the Bayh-Dole Act was passed which gave universities the ability to seek patent rights on their federally sponsored research. The tenor of the Bayh Dole Act was not to encourage invention but rather, “to use the patent system to promote the utilization of inventions arising from federally supported research or development.” While universities did not have the resources to commercialize research discoveries, the grant of patent rights allowed them to attract licensees with the resources to undertake such commercialization (Rai, 1999).

Developments in case law surrounding the definition of patentable subject matter also closed off the open landscape of basic science research. In 1980 the Supreme Court held in the case of Diamond v Chakrabarty, 447 U.S. 303, (1980) that a genetically engineered strain of bacteria with, “markedly different characteristics from any found in nature,” was patentable. It was also in this decision that the Court declared that patentable subject matter, “included anything under the sun made by man.” The 1998 U.S. Court of Appeals for the Federal Circuit (Federal Circuit) decision in State Street Bank and Trust v Signature Financial Group, 149 F.3d 1368, (Fed Cir 1998) held that an abstract idea such as a mathematical algorithm is patentable if it has a practical utility, catalyzed genomics information development.

The Federal Circuit has retained the position that “products of nature” are not per se patentable yet it routinely upholds patents on purified or isolated forms of molecules which occur in nature as long as the physiological function has been defined. The granting of patents on gene sequences where the protein for which they code has been identified is now taken for granted. Efforts to patent smaller pieces of DNA sequences that are significantly further away from practical commercial application and without a known function which are known as express sequence tags (ESTs) have been denied, however (Brudno, 1999).

As mentioned earlier in this article SNPs which represent those areas of the human genome which differ by one genome pair in at least one percent of the population. SNPs are very useful in understanding the disease process. In 1999 the Wellcome Trust and ten pharmaceutical companies joined together to create a high quality map of genetic markers/SNPs which is available for public use. Scientists believe that these common variations or SNPs can help to pinpoint subtle genetic differences that predispose some, but not others to such diseases as arthritis, Alzheimer’s, cancer, diabetes and depression. The SNP Consortium was guided by open science norms as they forced disclosure of SNPs discoveries into the public domain for the betterment of mankind.

It is clear that there is much heated debate and controversy over determining what approach to intellectual property is best suited to promote genomic and pharmacogenomics research. The dollar value of this industry is estimated to be several billion dollars and finding cures for common killer diseases is invaluable. It might be reasonable to conclude based upon the economic underpinnings of intellectual property law that broadening patentable subject matter to include or protect a large portions of scientific discoveries which contribute to pharmacogenomics research would create financial incentives to spur innovation. This level of analysis may not be appropriate for evaluating judicial decisions however. Yet unsurprisingly, a U.S. Commerce Department Study was conducted in the Spring of 2004 to “enhance the vitality and overall operation of the patent system.”

What is Informed Consent: An Ethical Duty

As the discussion in previous section illustrates the legal norms surrounding the biotechnology industry influence the direction of development of PGDs. The ways in which the ethical issues raised by this research are worked out is also critical to the maturation of the pharmacogenomics field. Genetics research involves access to human biological data which raises concerns about
patient privacy, confidentiality and consent (Hsieh, 2004). Genetic population studies affect the interests of more than just the sampled subjects. It impacts families and communities and can raise challenges for the protection of rights of large and small social groups with regard to stigmatization and discrimination. An example of this can be seen by the discontinued efforts of the Human Genome Diversity Project (HGDP) which was one of the first PGDs created to study human genetic diversity. The HGDP which was an extension of the HGP was established to collect DNA samples from various isolated populations to learn about human history and biological relationships in order to create a resource for the benefit of all humanity and for the scientific community worldwide (Hsieh, 2004). There was much criticism and fear that the HGDP would exploit indigenous populations and create global health discrimination by stigmatizing the populations with certain disease characteristics, and consequently the funding was curtailed and the program was discontinued (Hsieh, 2004).

The ethical duties that are presented in a PGD project can be grouped into two general and not mutually exclusive categories; the distinction is helpful and this article examines the ethical responsibilities primarily from the latter perspective. There is the ethical duty of a member of a community to donate a sample to the database versus the duties of those who are conducting the research. Duties to contribute to PGD have historically based justifications and are well founded in principles of beneficence. Under this ethical perspective, the benefits to the population become the critical justifications for the study itself; sharing benefits with participants also prevents exploitation.

The ethical concerns which the researcher must adhere to have a more immediate impact on the studies’ subjects and will likely have a more direct impact on the near future development of PGDs and pharmacogenomics. As an international consensus begins to form most policymakers and legislators will refer to the report of the UNESCO International Bioethics Committee on Bioethics and Human Population Genetics Research (1995) which states:

The ethics of population genetics should be formulated with reference to the minimal agreed human values as expressed in international human rights law. These agreed human values are based upon the recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family.

The research approach and methodologies used by the various PGDs with regard to obtaining consent, maintaining confidentiality of information and providing access and disclosure are not uniform (Hsieh, 2004). Currently, the literature suggests that the information which is given to individuals prior to participation in these projects does not sufficiently describe all of the components of the program and therefore informed consent is not being given by the donor. The deepest moral concerns stem from this doctrine of informed consent. Requiring a subject’s consent before performing medical research is the universal standard in scientific research and is codified in professional codes such as the World Medical Association’s Declaration of Helsinki, Ethical Principles For Medical Research Involving Human Subjects. Yet, determining whether informed consent has been given is a fact specific inquiry which will vary depending upon which information is given, who is asking for permission and whether consent is being given in conjunction with diagnostic treatment (Caufield [include last name of other authors here], 2003).

Typically a consent protocol will require the researcher to inform the donor of the purpose and nature of the research, the risks and benefits of the research, the method of study and the potential consequences. Written consent will be required for permission to follow up with the donor, to access medical and other records, to use the donor’s samples for various analysis and testing and for permission to talk with the donor’s physician. Informed consent procedures emphasize that donors should not be coerced into consent when they believe that their health could be compromised (Hsieh, 2004).
The biological information which is received can be used for multiple purposes, however, some of which may not be known at the time the sample is taken. It is quite common for biobanks to ask subjects for “open ended permission” to conduct future research projects on the samples obtained (Hsieh, 2004). This is requiring the donor to make a decision about risks and benefits of an unknown and unspecified future protocol. It is literally impossible to categorize this permission as “informed,” to call it such seems meaningless and ethically troublesome. Consent forms may also be misleading to the donor by suggesting that the banked tissue or blood has little or no market value and would otherwise be thrown away. In addition, the actual forms used may include a hospital or clinic letterhead where the study is actually being conducted by a commercial enterprise. Those PGDs that operate on a presumed or general consent model argue for its justification based upon efficiency where the costs of acquiring informed consent are prohibitive or practically impossible and individual privacy is protected. “Presumed consent” sometimes mistaken as “informed consent” is often allowed for existing samples that are sufficiently anonymous or unlinked to personally identifying information. “Informed consent” requires the researcher to describe the purpose of the research itself, yet genetic population databases may be used for multiple purposes inside and outside the specific research arena. Ethicists in the field have commented that the presumed consent model should be allowed only where there are additional safeguards in place; a key safeguard is Institutional Review Board (IRB) oversight. The misuse of “informed consent” not only harms the safety and privacy of the research subject, it violates respect for international values of individual autonomy and human dignity.

Conclusion

Much of the knowledge which is generated in finding the genetic causes of diseases and in introducing genomics based drugs is protected by patent law. The compilations of information used in the process, namely the families of databases employed in bioinformatics and genomics, are currently not protected under U.S. law. The tension between innovation and access in terms of legal decision making in regards to the development of PGDs which causes confusion and stymies growth. During the 1990s and into the early 2000s the U.S. patent Office appeared to be granting patents on “basic research” which was cause for concern in many academic and policy venues. In Spring 2004 U.S. Department of Commerce completed a study and produced a report with several recommendations to tighten the patent approval process which ensures greater access to pharmacogenomics research. These recommendations include reinvigorating the “nonobviousness standard” and creating an “open review procedure” for third parties to challenge recently issued patents before an administrative patent judge. These changes to the patent system will align the legal protections for the data underlying the genomic databases more closely with the more loose legal protections on genomic databases. Leaving databases that are merely factual compilations unprotected by copyright law and tightening the requirements for patent protection are both approaches which foster the open science model of innovation.

The “open ended consent form” that is commonly used in collecting DNA samples from large scale populations and which grants the researcher permission to use the sample for future, yet undetermined research, makes the donor’s consent senseless. This lack of informed consent not only violates standards of ethical practice, it may cause participants to withdraw or refuse to participate in such projects, and most importantly, these practices threaten society’s human autonomy.

The sequencing of the human genome has lead to the establishment of large scale genetic data banks which allow linking of DNA information with individual’s clinical and non medical information to explore the genetic basis of disease and to develop target drugs. The genomics revolution will change the role of prescription drugs in society (Rai 2001). Pharmacogenomics will reduce the cost and increase the access to new improved drugs worldwide which will potentially
improve the lives of millions. As discussed throughout this article pharmacogenomics is a new science and industry which is developing technical protocols to maximize its efficiency. The protections granted by the intellectual property laws that evolve and the maintenance of high ethical standards in developing PGDs will effect the growth of the pharcogenomics industry. There are no stark choices to be made either, rather we seek to find the level of human rights and beneficence which will encourage sufficient innovation to keep science progressing.

References


Biography

Julia Alpert Gladstone is an Associate Professor of Law at Bryant University where she teaches law courses with a multidisciplinary approach. Her research focus is examining the dynamic between law and technology which is revealed so brilliantly in the area of intellectual property law. She has published in all areas of intellectual property law most recently focusing on the database phenomenon as it has emerged in the field of genomics