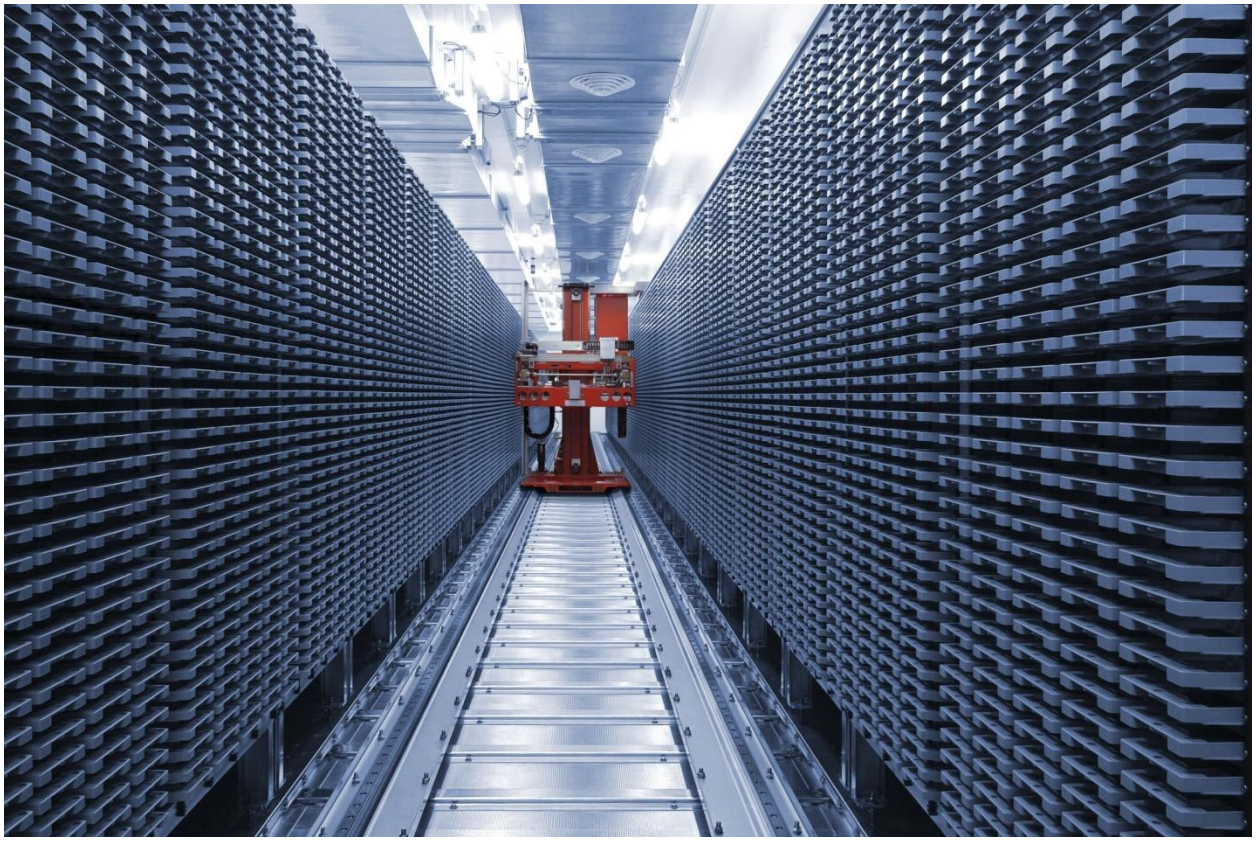
Council for Responsible Genetics

Do You Know Where Your DNA Is?

Genetic Privacy and Non-Forensic Biobanks



Aditi Shah

August 2014

Council for Responsible Genetics
5 Upland Road, Suite 3
Cambridge, MA 02140
Email: crg@gene-watch.org
www.councilforresponsiblegenetics.org

Contents

Introduction

What are DNA Databases?

A DNA database, or biobank, is a collection of people's DNA samples/data that often derive from one's blood, tissues, or saliva. There are several different types of DNA databases, including forensic (criminal), military, medical, research, and commercial. While forensic and military DNA databases target specific members of society, medical, research, and commercial biobanks are open to everyone. Medical DNA databases often include DNA samples along with medical information and are commonly found in hospitals and health care facilities which store the DNA samples so that researchers can use them for various causes. Research DNA biobanks are specifically used to enable certain research organizations to study particular genetic diseases. Commercial DNA databases are direct-to-consumer genetic testing services that store and utilize people's DNA samples to help them learn more about their ancestry, health, and more. Medical, research, and commercial biobanks consist of genetic data from people who voluntarily submit their DNA, voluntarily meaning that they are not being compelled, not necessarily that they are appropriately informed. Participation in medical and research DNA databases is free and these databases serve noble causes and have highly positive intentions, as they aim to use the DNA samples to find possible treatments and ways to prevent destructive diseases such as cancer and heart disease. Consumers must pay a fee to participate in commercial DNA biobanks, which also have beneficial intentions, as they provide people with the opportunity to learn about their family tree and receive health reports without a physician. Today, medical, research, and commercial DNA biobanks have expanded and grown immensely in their reach and popularity throughout the United States. While the growth of these beneficial non-forensic DNA databases is positive, genetic privacy concerns have developed due to the lack of appropriate regulations to protect genetic privacy in an age where the potential for one's DNA to be in a database is greater than ever before. A close evaluation of findings on medical, research, and commercial DNA biobanks and the policies for each demonstrates the need for an adequate regulatory framework that allows these databases to prosper while protecting people's genetic privacy at the same time.

What is DNA and Why is it Important?

DNA, deoxyribonucleic acid, is a molecule made out of nucleic acids that can be found in every cell in our body and forms the genetic information of each living organism. Consequently, DNA is often noted as the "blueprint of biological life", as it gives instructions for an organism's functioning and development. A single DNA molecule is double stranded and has sequences of four bases: adenine (A), thymine (T), cytosine (C), and guanine (G). Humans share 99.9% DNA in common, but 0.1% of the DNA is what differentiates the genetic makeup of one human being from another. This 0.1% difference might result from a single-nucleotide polymorphism (SNP), when one letter in the DNA sequence of a person is different from that of another person, or due

to the DNA segments having different sized fragments.[1] While a difference of 0.1% might seem trivial, this variation between human DNA sequences actually accounts for what makes people genetically unique.

DNA is important not only because it makes everyone biologically different from one another, but also because it is the unique identifier that people are born with, and cannot change. Unlike other personal items which can be used to identify people, such as passports, cell phones, social security number, and credit cards, DNA cannot be replaced or changed so that no one can identify a person based on their DNA sample. DNA is inescapable and irreplaceable, and for that reason, it is perhaps the most private and personal possession people hold. DNA databases were first established in the 1980s and were initially used as forensic databases to be able to identify criminals and as military databases to help recognize deceased military members based on their remains.[2] In the following years, hospitals began to establish medical databases to make DNA samples available for research purposes and private organizations started to establish research databases to study specific diseases and conditions. Likewise, commercial biobanks began to develop as they used scientific and technological advancements to attract those curious about their ancestry and health.

Medical DNA Databases

What are Medical DNA Databases?

Medical DNA databases began to form in the late twentieth century when hospitals and other health care organizations started collecting and storing DNA samples. The DNA samples found in medical biobanks are mostly used to provide various medical research institutions with DNA samples to work with for their individual studies. Private hospitals and medical centers as well as state governments are commonly known to keep medical DNA databases and use them for a variety of purposes. The state government medical DNA databases include newborn DNA biobanks, which consist of DNA samples from babies usually taken immediately after they are born during the screening process.[3] The medical DNA biobanks which hospitals and health care facilities have comprise of DNA samples from people who voluntarily donate their DNA to the institution, who are often promised that their samples will be stored anonymously. Also, medical centers and hospitals often store relevant medical and family history along with the DNA sample in the database to help researchers looking for specific samples.[4] These medical databases have extremely positive and helpful motives, as they intend to serve as banks for scientists researching various diseases in hopes of finding a treatment or cure. The following section specifically looks at newborn biobanking, while the later section focuses on other types of medical DNA databases.

Newborn DNA Banking: What exactly happens?

Newborn DNA biobanking, or the storage of newborn DNA samples in state government databases, began in the U.S. in July 1997. The newborn DNA samples are gained from newborn screening, which began in the U.S. in the 1960s with scientist Robert Guthrie who found out how to test for phenylketonuria (PKU), a genetic disorder, in babies.[5] Today, however, newborns are screened for at least thirty different conditions and their blood samples are also being stored in state government laboratories. This process involves taking a small sample of blood by pricking the baby's heel and placing the blood on a card that is then tested for different genetic disorders and conditions. After the test results are obtained, some of the residual blood sample that remains on the card is stored in the state government newborn DNA databases.[6] Newborn screening has become so popular that 98% of the 4.3 million babies born each year in the U.S. are genetically screened.[7] Newborn DNA biobanks are often advertized and applauded as precious resources for medical research. However, the problem is that many states are storing newborn blood samples without the permission or informed consent of the parents. Often, parents are provided a consent form to sign without proper explanation soon after their child is born, a form which new parents are unable to properly understand given the anxious environment and lack of actual knowledge that their child's blood sample will be stored after screening.[8]

State and Federal Laws for Newborn DNA Banking

The state and federal laws regulating the retention of newborn blood samples emerged in the past three decades in response to the growing interest of states to store the samples. A federal law that regulates the use of humans in scientific research is called the Common Rule (45 CFR 46), or the Federal Policy for Protection of Human Subjects, which lists stipulations for getting parental consent and requirements for informed consent. According to this policy, for research involving "the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens" and data which is "recorded by the investigator in such a manner that subjects cannot be identified", the Common Rule provisions for informed consent do not apply.[9] Since states claim that the stored newborn DNA samples are "de-identified", the Common Rule essentially enables newborn blood samples to be stored and used for research without informed parental consent. Another federal law which regulates newborn screening and has influenced the state policies regarding storage of newborn blood samples is the Newborn Screening Saves Lives Act, which President Bush signed in 2007. This act provides grants to hospitals and institutions which administer newborn screening programs and includes no provisions regarding having the informed consent of parents prior to the screening and storage.[10] Because of this act, seven states have extended their retention period for the newborn blood samples since 2007 and it gives states the power to decide policies on obtaining parental consent.[11] Many states do not have policies or laws which require parental consent for newborn screening and storage of the blood samples and in four states, California, Maine, Utah, and Washington, collected newborn residual bloodspots become state property.[12] Since states regulate newborn biobanks, the amount of time they decide

to store the blood samples varies from six weeks to indefinitely and seven states say that they do store newborn blood samples indefinitely.[13]

Privacy Concerns with Newborn DNA Biobanks

Lack of Informed Consent from the Parents

Privacy concerns with newborn DNA banking stem from the lack of informed consent from the parents. Informed consent involves the parents giving permission to the hospitals and the state government to store their child's DNA with complete awareness of the potential risks, benefits, the people who will be able to access the sample, and how it will be used. Most parents in the U.S. do not give permission for their child's residual blood sample to be stored in the state government's biobank and thus are often unaware that their child's DNA is in a state database.[14] According to the Council for Responsible Genetics, many states do not have specific regulations about getting consent from the parents for the storage of their child's blood sample.[15] Informed consent is an important issue with newborn blood sample storage due to the difference between the importance of the biobanks that state health departments promote and the actual benefits they lead to. For instance, while the newborn DNA biobanks are praised for being used to help study and treat diseases, most of them are actually used for quality control of screening tests.[16] This is relevant to the issue of informed consent because the lack of actual substantial benefits from the storage of newborn blood samples shows that states should not use "protecting public health" as an excuse for not obtaining consent from parents. Most states have adopted an "opt-out" model which allows parents to choose to refuse screening, but do not involve an "opt-in" process with parents having to approve or give permission to screening.[17] The problem with opting-out is that some parents may want their child to be screened but not want their residual blood sample to be stored in the biobank. Instead, various advocacy groups including the Newborn Screening Task Force and the President's Council on Bioethics support "opt-in" policies which would involve parents giving permission for screening and for storage, separating the two procedures to allow parents to choose that which they are comfortable with.[18] Since the Common Rule allows states to keep newborn blood samples without parental approval and forty eight states practice the opt-out policy, there is an apparent lack of informed parental consent for newborn biobanking.

Genetic Privacy: "Anonymous" Samples

State Departments of Health claim that their newborn biobanks contain samples that have been "de-identified" or are "anonymous", using both words interchangeably. "Anonymous" can be defined as the personally identifiable information of the person whose DNA sample is in the biobank being permanently destroyed, while "de-identified" data only means that the personally identifiable information is detached from the sample, but can be re-accessed in the future.[19] This

issue of anonymity is important for newborn biobanking because several parents are concerned about whether their child's genetic data is actually stored anonymously in the databases. Even though states often claim that the newborn biobanks contain de-identified data, various findings show that in reality, it is not actually possible to store DNA samples in a way so that no one can be identified from the database. For instance, there was a study that showed that even if a small amount of personal information was present, researchers were able to identify people from the database.[20] Further, scientists have argued that the Newborn Screening Translational Research Network, which allows researchers access to the residual blood spots to advance their newborn screening research, is secured to prevent intermixing of databases.[21] However, there is a lack of evidence to show that this is true, as there is no guarantee that unauthorized personnel cannot use it for their own means, especially since the data is not actually "de-identified". Given this, the anonymity of newborn biobanks raises further privacy concerns for parents because the issue of how secure these biobanks are becomes a problem. For example, in New York, there are paper cards with identifying information for each newborn blood sample kept in the storage facilities.[22] Considering this, it is imperative that appropriate safeguards and security are present to prevent the information from becoming too accessible.

Lack of Transparency

The lack of transparency of newborn biobanks also contributes to the privacy concerns that parents have. There is little information on exactly how the state newborn biobanks operate, where these biobanks are located, who has access to the biobanks, and how secure they actually are. One of the issues with newborn biobanking is that the warehouses where the states keep the residual blood samples are unknown. For example, in Indianapolis, officials did not allow news investigators to visit or know the location of the warehouse where they store the newborn blood samples.[23] With limited knowledge on where the samples are stored, it is difficult to know how they are kept and whether any personal information is stored with them. Likewise, it is unclear exactly who the researchers that can use the blood samples are and for what purposes they may use them for. This is concerning for parents since without knowing who might be viewing their child's blood samples and for what, it is reasonable for them to worry that anyone might be able to access the data and use it for purposes they would not give consent for. For instance, findings suggest that the blood samples are sometimes given to researchers along with the baby's name, and in Minnesota, more than twenty scientific papers were published using newborn samples.[24] Actually, many state newborn biobanks contain samples that are decades old and are no longer useful for scientific research, yet they remain in the warehouses without any apparent purpose. Also, since little is known about how the biobanks operate, parents do not know how properly secured these warehouses with the physical samples and databases with the genetic profiles are, leading to parents worrying about people being able to break in and possibly misuse their child's DNA. The lack of information and complete knowledge regarding how newborn biobanks work leave numerous unanswered questions and concerns for parents who might feel much safer with their child's DNA in these databases if they knew more about them. In fact, if newborn biobanks

became more transparent, their operation and their efficiency may improve as State Departments of Health would become aware that the public is keeping an eye on their biobanks.

Problems with Current Regulations

One of the reasons why such genetic privacy concerns exist with newborn biobanks is that there is an apparent lack of proper regulations to address these concerns. For example, the fact that the federal Common Rule does not necessarily apply to newborn biobanks is extremely concerning because it essentially allows for states to collect and store newborn residual blood samples without informed parental consent. Instead, obtaining parental consent for storing their child's DNA in the state biobank needs to be enforced by the Common Rule, giving parents the ability to decide whether or not they want their child's blood sample to be stored. The dangerous outcomes of not obtaining parental approval were demonstrated in the cases of *Bearder v. State of Minnesota* (MN, 2011) and *Beleno v. Texas Department of State Health Services*, which did not actually go to trial. Both cases involved parents suing the states for violating the privacy of their children by storing their genetic information without their consent. The Beleno case did not go to trial since the Department of State Health Services and the hospital settled and agreed to destroy all samples obtained without parental consent before May 2009, resulting in 5.3 million blood spots being removed.[25] In the Bearder case, nine families sued Minnesota for storing the blood spots without parental consent and in this case, the Supreme Court of Minnesota held that the state needed to destroy samples after seventy-one days of retention and the state since then has removed samples and changed its laws to require informed parental consent.[26]

These cases demonstrate that federal regulations are needed to ensure that the parents are completely aware of the actual potential risks and benefits of newborn biobanking rather than the exaggerated positive aspects which states and hospitals emphasize. Additionally, it is critical that there are regulations to enable the protection of the residual blood samples and extracted genetic data so that false guarantees of "anonymous" and "de-identified" data are not made and instead, federal oversight is provided to ensure that the databases have adequate safeguards and security. Furthermore, even though federal laws such as the Genetic Information Nondiscrimination Act of 2008 (GINA) exist to prohibit specific types of genetic discrimination, federal regulations are needed to ensure that unauthorized access to the residual newborn blood samples does not occur at all, rather than acts to prevent misuse of that data. While additional regulations are needed to properly educate parents about the actual benefits and risks of newborn biobanking and to ensure that it is done only with their informed consent, a proposed bill known as the Newborn Screening Saves Lives Reauthorization Act of 2013 has passed in the House of Representatives and is now waiting on the Senate.[27] While this act aims to expand the newborn screening programs of states, it is important to consider the impact this might lead to on the storage of newborn residual blood samples and the privacy implications this might have, as it is crucial that federal acts and policies begin treating newborn biobanking as a separate issue from newborn screening.

Other Types of Medical DNA Databases

Other types of medical DNA databases include those which involve people voluntarily donating their DNA sample to be collected and stored in databases found in hospitals and health care facilities. As mentioned earlier, the DNA samples and medical information stored in these databases are provided to researchers studying a variety of diseases and conditions. Unlike the newborn biobanks, these types of medical DNA databases often involve volunteers specifically looking to participate, rather than having their blood or saliva sample be automatically stored in the database. These medical DNA databases provide a rich source of genetic and medical data for scientists hoping to use this data to find treatments for different disorders, and are particularly appealing because of their noble motive of furthering medical research to help people suffering from various conditions and disorders. The following three examples demonstrate medical DNA databases that are not newborn biobanks.

Example #1: Mayo Clinic Biobank

Mayo Clinic is a medical practice and health care organization that was founded in 1889 and is based in Minnesota with hospitals in Florida and Arizona.[28] Mayo Clinic has grown to be one of the largest medical practice groups in the world and it provides medical care for patients, education facilities including graduate school and medical school, and it conducts research. The Mayo Clinic Biobank, established in 2009, is part of Mayo Clinic's endeavor to help researchers and pharmaceutical companies to study the role of DNA in health. Their biobank accepts blood samples and health information from both current and former Mayo Clinic patients.[29] In order to participate, patients need to complete the enrollment package, sign the consent form, and fill out a questionnaire regarding their health, lifestyle, and family history. To reach their goal to enroll 50,000 patients in the database by 2015, Mayo Clinic also offers every participant items worth a total of $20 as compensation.[30] Under their privacy policy, they say that the samples they receive are not stored with the person's name, address, birth date, social security or mayo clinic number and note that if a person is identified through their sample, the federal act GINA exists to protect them.[31]

Example #2: "BioMe" Mount Sinai Biobank

Another example of a voluntary medical biobank is the Mount Sinai Medical Center's DNA database. The Mount Sinai Medical Center is located in New York and has one of the world's largest biobanks which contains blood and/or saliva samples from more than 25,000 patients.[32] Launched in 2007, their biobank is similar to the Mayo Clinic biobank, as it gives various researchers genetic and phenotypic information to work with and advance their own medical studies. In order to participate in BioMe, the Mount Sinai Hospital patients must meet with their doctor at Mount Sinai, who would then have them speak with a BioMe recruiter who would explain the procedures. Once someone decides to participate, they would have to sign a

consent form, a health questionnaire, allow access to their medical records, and provide a blood sample. Under their privacy policy, they claim that the samples and health information will only be identified using a code and that no personal information will be shared without permission.[33]

Privacy Issues with Voluntary Medical Biobanks

Both the Mayo Clinic Biobank and the Mount Sinai Biobank demonstrate how voluntary medical DNA databases function and what their purpose is. While both examples contain privacy policies and claim that the genetic privacy of their participants will be protected, the unfortunate reality is that there is not enough evidence to support their claims. In fact, there are shortcomings and flaws with their policies and federal laws to control these databases that create a gap between what volunteers expect and what actually happens. The specific issues that lead to these privacy concerns are: who has access to the medical records and health information, lack of actual "de-identified" DNA samples, and the ability to identify volunteers from their DNA samples even if no names are attached. The privacy concerns with voluntary medical biobanks are discussed in detail in the section following the research DNA databases, as the same privacy concerns apply to both types of biobanks.

Research DNA Databases

Voluntary Research DNA Databases

Voluntary research DNA databases are similar to medical DNA databases in that they both involve the use of DNA samples which participants willingly submit. However, research databases, unlike medical ones, are used by scientists and research institutions to study specific diseases and conditions. Voluntary research DNA databases also emerged in the late twentieth century with advancements in biotechnology and increasing interest in examining DNA to be able to understand the causes and possible treatments of certain disorders. Research biobanks were largely created with the intention of finding ways to prevent or treat these diseases and to help large populations of people suffering from devastating diseases such as breast cancer and diabetes. Specific examples of types of voluntary research DNA databases are discussed below.

Example#1: Kaiser Permanente: RPGEH

Kaiser Permanente is an integrated health care organization formed in 1945 and based in California. It has a division of research that launched an initiative entitled the Research Program on Genes, Environment, and Health (RPGEH). This research program has its own biobank with DNA samples from more than 500,000 California Kaiser Permanente members and its aim is to understand which genes and environmental factors influence certain diseases.[34] Participants must be members of Kaiser Permanente in Northern or Southern California and sign the consent form

prior to submitting a blood or saliva sample. Currently, one of their research projects is to study the genetic factors that might influence bipolar disorder, funded through grants from the National Institute of Mental Health (NIMH). Another research project they are working on is studying the genetic factors which influence prostate cancer in African-American men, funded through grants from the National Cancer Institute.[35] In their privacy policy, they inform participants that there is an institutional review board (IRB) that considers the privacy regulations that exist and that each participant's medical information will be de-identified. They also state that all of the samples and information are stored as a database in a computer system that is located in a secure setting.[36]

Example #2: Alzheimer's disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) formed in 2004 and is a global project that involves collecting and storing DNA samples to research possible ways to avoid and treat Alzheimer's disease.[37] Alzheimer's disease is the most common type of dementia that often results in memory loss and behavioral and thinking problems, as the disease hurts and kills brain cells.[38] Scientists believe that a combination of genetic and environmental factors can cause the disease, and the aim of ADNI is to learn more about what influences the disease to find ways to treat or prevent it. ADNI 2 is a new project launched in 2013 as the third phase of ADNI, aiming to further develop the findings of the ADNI project and more specifically detect the earliest signs of Alzheimer's disease. Researchers plan to carry out the study until 2017 and they are looking for 550 volunteers between 55 and 90 years old to participate in clinical experiments.[39] Using the information collected through these studies, the ADNI has compiled an extensive DNA database and between 2012 and 2013, ADNI sequenced DNA from 818 people and then released the data to researchers connected with ADNI.[40] ADNI continues to build its biobank as it acquires further participants around the world.

Example #3: Personal Genome Project

In 2005, Harvard Medical School launched the Personal Genome Project, a program that aims to collect DNA samples from participants to create and use an extensive DNA database for medical research. The Personal Genome Project is different from other research DNA databases in that it intends to create a public database and informs participants that the genetic information that they supply can and will be publically displayed and published. The specific scientific goal of this project is to understand the connection between people's genetic information (DNA and genes) and their traits (medical information and physical traits) and to make their study public.[41] To participate, volunteers must sign the consent form which states that "data will not be kept or made available by the PGP (Personal Genome Project) in a confidential or anonymous fashion", as they warn prospective participants that there is no guarantee that the information they provide could not be used to identify them.[42] To make sure that each participant who signed the consent form understands the terms and conditions listed in the form, all of the participants must take an exam testing their comprehension of the consent form and must achieve a perfect score on it to enroll in the study. The PGP also distributes safety questionnaires to the participants every three

months asking them to share their experiences and positive/negative effects of their participation on their lives.[43]

Privacy Concerns with Voluntary Medical and Research DNA Databases

Lack of Informed Consent: People's Expectations vs. Reality

One of the problems with voluntary medical and research DNA databases is the lack of actual informed consent, leading to an apparent contrast between what participants who donate their DNA sample expect and what really occurs. Informed consent involves participants having a complete and accurate understanding of all of the details regarding the project or program they will participate in, including knowledge of any of the risks and negative consequences. Signing a consent form is supposed to show that the person agrees to participate in the project/study while knowing exactly what their participation involves, what it might lead to, and any potential risks. The truth, however, is that most people who sign consent forms and participate in medical and research DNA databases do not do so with informed consent. This lack of informed consent is reflective of the misunderstanding between what volunteers anticipate from their participation and the reality, as there are certain actualities of medical and research DNA databases that are not clearly conveyed to participants before they submit their DNA sample. For example, these biobanks often have a lack of transparency which disables participants from having complete knowledge about how these databases operate, how and where their genetic data is stored, who has access to the databases, and how long their data will be kept there. Likewise, there is a lack of federal policies to make these databases more transparent or to regulate them so that people are more adequately informed of what will happen to their DNA. The lack of informed consent and federal regulations to appropriately address this issue are explored further in the following sections as they discuss specific points of concern which reflect the disparity between people's expectations and the reality.

Identifying People from these Databases

Various scientists and researchers have tested the promise of genetic privacy that many medical and research DNA databases claim through their usage of "de-identified" genetic data. Through their experiments, it is evident that it is actually relatively easy to identify people based on their DNA in medical and research biobanks. For example, in 2008, Daniel Craig, a geneticist at a research institute in Phoenix, Arizona called TGen, conducted a study to see if he would be able to identify people based on their DNA sample from their massive biobank. He discovered a method to use the four million differences in the DNA sequences to identify people, even if their DNA was only 0.1% of the database.[44] Erich Schadt at the Mount Sinai School of Medicine had

conducted a similar study and found that the RNA expression data that is part of the genetic data stored in the database could be used not only to identify people but also to reveal a person's age, weight, and whether or not they are diabetic or had viruses like HIV.[45] In January 2013, Yaniv Erlich, a geneticist at the Whitehead Institute in Massachusetts, was able to identify five people from their DNA through a random selection from 1,000 people.[46] Further research conducted by the Chief Technology Officer Latanya Sweeney on re-identification shows that it can occur by linking people's recorded medical and genetic information with their recorded personal data and using what is in common to identify the person.[47] To understand how to prevent re-identification, Sweeney and other researchers have developed certain methods such as "k-anonymity" to avoid being able to link information to find people from databases.[48] However, while more efficient privacy protection methods are being developed, it is crucial that participants donate their DNA with the understanding that current privacy settings do not enable the complete anonymization of their data, since not attaching peoples' names to their DNA sample does not prevent others from finding out who they are. Thus, these findings express the need for better methods to disable re-identification and the need for hospitals and research institutions to better educate volunteers on the current realities of data anonymization.

Security: How Secure are these Databases?

Knowing that it is possible to identify people from medical and research DNA databases makes it even more important and necessary for these databases to be secure and have adequate safeguards. Having a strong security system may not completely prevent people from being able to identify participants from a database, but it may prevent unauthorized access and misuse of the genetic profiles. The only security provision that is mandated by law is for medical and research DNA databases to have de-identified genetic data so that one's name, birth date, address, social security number and other such personal information are detached from the samples. However, other security measures are not specifically mentioned by the databases and are not enforced by federal or state regulations. For example, Kaiser Permanente informs participants that their data will be stored in a secure computer system that is locked in a secure facility and uses electronic security measures and a firewall.[49] The problem with their description of their security system is that it is rather vague and does not sufficiently answer some security questions such as who has access to this computer system and how long is the information stored there. Even though many of these research organizations, including Kaiser Permanente, state that only a small number of authorized scientists and staff have access to these genetic profiles, there still remains concerns about whether they are dealing with personally identifiable information, since participants often expect them not to. Likewise, how long their genetic data will be kept there is another security concern, as medical and research DNA databases store the DNA samples and the genetic data indefinitely unless otherwise stated. Further, while the specific security systems biobanks have remain ambiguous, it is also difficult to assess how secure these biobanks are because of the lack of transparency. After participants submit their DNA sample to medical and research databases, they are not aware of exactly who is using their DNA and how, leading to more privacy as well

as security concerns. To remedy this, the Interscience Molecular Oncology Laboratory (IMOL) suggested a user-centricity security system which involves the person who donated their DNA to know where their DNA is being stored and to be notified every time their genetic data is used.[50] They also recommend security measures such as authentication demands any time the scientists and staff try to access or use the databases.[51] These measures and further reforms with security allow for increased transparency of medical and research biobanks, which will enable volunteers to have some degree of control over their genetic data and will establish powerful safeguards to prevent misuse of the data.

Bleeding with Other Types of DNA Databases

Another privacy concern with medical and research DNA databases is the possibility of the genetic information in those biobanks bleeding with other kinds of DNA databases, including forensic databases. Since it is possible to identify people from their DNA sample in a database, it is reasonable to believe that the people trying to identify volunteers and looking at DNA samples can be anyone, including police officers and law enforcement authorities. For example, a man in Scotland who voluntarily donated his DNA sample to a research DNA database, thinking that his identity will be anonymous, was convicted of knowingly infecting a woman with HIV using the DNA sample he submitted as evidence.[52] This case reveals the difference between what people expect when donating their DNA sample to these medical and research biobanks and the reality of what occurs. The man in Scotland, just as the millions of Americans who donate their DNA sample for medical research, participate in these projects/studies with the expectation that their identity will be kept anonymous so that their genetic privacy is protected. The benefits of such interaction between databases include that there is a broader database to find criminals and help crime investigation. However, using medical and research biobanks for forensic needs can create a sense of uneasiness and discomfort for people who participate under the expectation that their DNA will only be seen by a select number of scientists who would not know if it is their sample. The identification of people from databases, the lack of adequate security, and the potential for bleeding with other types of DNA databases question the regulations that might exist to prevent such breaches of genetic privacy, an issue that is discussed in a later section.

What Should the Volunteers Expect?

Given the gap between what people expect from their participation in these medical and research DNA databases and what actually occurs, it is necessary for prospective volunteers to understand what they should and should not expect. Volunteers should not expect their genetic information to be de-identified or anonymous, as the various studies previously mentioned show that it is almost impossible to 'de-identify' DNA or genetic data. Participants should also expect third parties, police officers, and people other than the scientists associated with the particular medical or research biobank to be able to access and use their genetic information for anything, including identifying who they are. Also, particularly for research DNA databases such as PGP which publically display people's genetic profiles and certain personal and medical information

online, participants should be aware that they may be vulnerable to discrimination because of the unrestricted access and potential for misuse of their DNA. For medical DNA databases such as the Mayo Clinic Biobank, participants should be aware that the geneticists looking at their DNA may be able to easily identify them using their medical information and that there is no guarantee that their information will not be used in ways they might not be comfortable with. These points demonstrate the need for prospective volunteers interested in donating their DNA to medical and research biobanks to be properly educated about the realities of having their DNA in the database so that the disparity between what volunteers expect and the reality decreases significantly.

Lack of Proper Regulations for Medical and Research DNA Biobanks

Current Regulations in Place

The current regulations in place to protect the genetic privacy of people include GINA, the Genetic Information Nondiscrimination Act of 2008, which is designed to prohibit health insurers and employers from discriminating against people based on their genetic information. To protect people from genetic discrimination, GINA prohibits health insurers and employers from requesting or requiring people to provide any kind of genetic information, which includes family medical history.[53] Another federal law that aims to protect people's genetic and medical privacy is HIPAA, the Health Insurance Portability and Accountability Act of 1996, which was established to allow people to access their own medical records, knowledge of how their medical information may be used, and prohibits genetic discrimination in group health coverage.[54] Also, HIPAA requires that medical information used in genetic research must be de-identified so that eighteen personal identifiers including social security number, name, as well as unique traits, are removed from the medical information.[55] The other federal law which regulates human research and has set the guidelines for informed consent is the Common Rule, which states that medical and research institutes seeking participants must include the purpose of their research, the right to confidentiality, all potential risks, possible benefits, and more.[56] The Common Rule applies to scientific research organizations that are funded or supported by the federal government. For the research institutions that are not federally supported, the Common Rule requires an Institutional Review Board (IRB) such as the one that Kaiser Permanente has. An example of a state law that intends to protect people's genetic privacy is California's Confidentiality of Medical Information Act (CMIA), which prohibits giving or using medical information for reasons other than research to allow the person's identity to remain private.[57]

Why these Regulations are Insufficient

While there are a few federal and state laws attempting to protect people from genetic discrimination, none of them specifically address many of the most important privacy concerns with medical and research DNA databases. The federal regulations mentioned above also have shortcomings which imply that not all types of genetic discrimination are outlawed. For example, despite being a significant protection against genetic discrimination, GINA does not address life insurance, disability insurance, and long-term care insurance.[58] Likewise, it can be very difficult to show that genetic discrimination occurred, as the Personal Genome Project's study guidelines state "you might never know whether your employer found your PGP data and read about your genetic findings".[59] Perhaps even more concerning than the possibilities for discrimination as a result of health insurers and employers identifying people through medical and research biobanks is the fact that there are no federal or state laws specifically regulating these databases to protect people's genetic privacy. For instance, there is no federal law explicitly prohibiting those outside of private medical and research organizations, such as law enforcement officials, from searching their databases and possibly using them to identify people. There are thus no laws placing limits on who can access these databases, enabling the misuse of genetic information. There is also a lack of federal laws for how medical and research biobanks store the genetic data anonymously or any administrative bodies to oversee and control how secure these databases are. There are also no regulations for research DNA databases that are public, like the PGP, to protect people's genetic information from being misused, as although the PGP informs the participants that their genetic privacy will not be protected, they do not explain how the participants will be protected if their genetic information is used against them. There is thus a concerning lack of federal laws that address the privacy concerns specifically correlated with medical and research biobanks.

Limitations of the Common Rule

The previously mentioned federal Common Rule is an important policy for research on humans, which includes the medical and research DNA databases. The Common Rule sets out guidelines for informed consent, including requiring that each participant knows the purpose of the research, how their participation may benefit them or others, and all of the foreseeable risks to them by participating.[60] However, this policy has many limitations that disable it from being an appropriate and sufficient regulatory framework for informed consent. For example, under its requirements for informed consent, the Common Rule allows the IRB to not include or to change some of the provisions for consent if doing so does not harm the rights of participants.[61] Given that the IRB generally consists of scientists, lawyers, and doctors who are often interested in the studies that the research company or organization is doing, it is difficult to believe that this part of the Common Rule does not provide a signal for the IRBs to alter consent forms in ways that would benefit their groups. For instance, the IRB of a group may decide to remove the need to explain the circumstances in which they would destroy participants' samples and genetic data, which the Common Rule itself only vaguely describes how to include in the consent form. Also,

even if a research institution has a consent form as the Common Rule requires, there is no federal agency to oversee whether these research facilities actually carry out what they tell volunteers in the consent form. For example, in 2010, the Havasupai Indians living in the Grand Canyon sued Arizona State University after finding out that instead of using their blood samples only to study the rate of diabetes in their tribe, they used them to study mental illness and the tribe's origins.[62] Even though the Havasupai Indians won the case, this incident reflects significant shortcomings with current legislation and that substantial regulations still need to be created to establish actual informed consent so that research organizations cannot use people's DNA for anything without their permission.

Commercial DNA Databases/Biobanks

What are they?

Commercial DNA databases developed in the United States during the late twentieth and early twenty-first century and have expanded over the past few years due to their growing appeal and popularity. DNA databases are often types of direct-to-consumer genetic testing companies, which sell DNA kits and analyze the samples for various purposes, including informing people of any potential medical risks, ancestry, and even relationship compatibility. Using this method, direct-to-consumer genetic testing companies offer customers the opportunity to perform genetic testing without a doctor or third party. Many of these companies have also become genetic social networking websites, as they provide customers with the chance to share their genetic data with others online to potentially find people they might share genetic traits with and be related to. An example of a commercial DNA database that tests for relationship compatibility is called Instant Chemistry, a Canadian based group that involves single individuals and couples sending in their DNA sample to be tested for their compatibility with others in the database or with each other for couples.[63] Even more popular are genealogical databases that allow people to trace their ancestry and health-related ones that enable people to find out about potential medical risks. Commercial DNA databases also have positive motives, as they enable people to learn more about their DNA, fill their family tree, find potential partners, and help adoptees find their biological parents while not needing a physician. The following sections provide and discuss examples of health related and genealogical DNA databases.

Health Related Commercial DNA Databases: Personal Genome Services

Example#1: 23andMe

An example of direct-to-consumer genetic testing for medical risks and health reporting is the company 23andMe. Created in 2006 and based in California, 23andMe was established to give health reports through their Personal Genome Service (PGS) and ancestry information to the customers who send in their DNA samples. Until November 22, 2013, 23andMe's PGS offered customers heath reports showing people if they had any diseases or conditions, their chances for developing certain disorders, their genetic predisposition to traits such as baldness, and how they may respond to certain drugs.[64] The U.S. Food and Drug Administration (FDA) had discontinued 23andMe's PGS by sending a warning letter to the CEO of the company on November 22, 2013 after finding concerning flaws with the accuracy of the health reports. As a result, 23andMe has stopped sending health reports and it now only provides ancestry information and un-interpreted raw genetic data. Under their core values, the company states that it aims to allow people to have control over their genetic information and the means to share their data with the other 23andMe customers and to engage in genetic research.[65] Today, 23andMe is one of the largest commercial DNA databases in the world and has acquired more than 650,000 customers.

The privacy concerns with 23andMe's services generate from the disparity between what customers believe 23andMe does with their genetic information and what it is actually aiming to accomplish. After purchasing the DNA kit to submit their DNA in the form of saliva, customers need to register their kit, which involves entering the barcode, agreeing to their terms of service, filling out their research consent form, and choosing their account.[66] This research consent form seeks permission from customers for their participation in 23andMe's research project known as 23andMe Research, which aims to support and publish scientific discoveries.[67] Customers have the option to decline participation in the research program but still receive their ancestry results and raw genetic data, but their privacy policy states that their genetic and personal data may still be used for other purposes such as quality control and improving their services.[68] Also, the terms of service and privacy policy do not explicitly explain how the DNA samples will be used, which has led to differences between what customers expect and what 23andMe does. For example, the Genetic Privacy Network states that while customers give their DNA to 23andMe expecting it to be used to allow them to learn about their ancestry and genetic data, 23andMe has actually been trying to establish an expansive biobank using their customers' genetic data.[69] While 23andMe has mentioned this in the news, its goal to create a biobank is not clearly conveyed to consumers reading their terms of service. This is important because it changes the way customers perceive their participation and the privacy risks they might be considering, especially while knowing so little about how the data is stored anonymously, who may be able to access it, and the security settings. The lack of transparency and customers not knowing exactly what they are purchasing

demonstrates the need for 23andMe and such commercial genetic testing companies to be more appropriately upfront about their intentions and actions so that consumers are better informed.

Example #2: Interleukin Genetics Inc.

Interleukin Genetics Inc., another direct-to-consumer genetic testing company, formed in 2006 and based in Waltham, Massachusetts, provides customers genetic testing for disease risk as well as weight management, heart health, nutritional needs, and bone health.[70] The company provides people with personalized health reports for $169, and involves customers mailing their DNA samples (cheek swabs) to be analyzed. Interleukin then provides the customers with their results, which they can access online or through the mail and their purchase includes the choice of meeting with a genetic counselor to help them interpret their results. Customers are also able to purchase genetic tests for others who will have their own account and their company does not need to be FDA approved because their tests are laboratory developed tests and are only required to follow the Clinical Laboratory Improvement Amendment (CLIA) standards.[71] Their company policy also holds that they will store the DNA data results in their database for seven years as the law requires, even though the DNA is destroyed within ten days of the report being completed.[72] It continues to operate today, offering genetic tests to help guide people about their health.

Shortcomings with Interleukin's policies and the lack of strong governmental regulations demonstrate why their services may generate certain genetic privacy concerns. For example, the CLIA standards only establish rules for quality control of the laboratories and do not include any specific regulations to ensure the privacy or anonymity of the data.[73] Likewise, since the FDA's approval is not necessary, there is no regulatory board to ensure that Interleukin actually removes the DNA samples when it says it does. While there is a lack of proper governmental regulations, there are also points of the privacy policy that are concerning as well. For instance, it holds that their privacy terms do not apply to the privacy standards of third parties that Interleukin might be partners with. Their privacy policy also notes that third party affiliates and employees will have access to certain personally identifiable information, which may include one's name and address, yet it does not specifically state which rules and regulations are in place to prevent affiliates and employees from misusing this information.[74] Claiming that their system is highly secure and that the employees and affiliates would need the password to access the data is also not comforting or enough to prove that the privacy of consumers is protected. Another point of concern is that they collect and track how people use their website and say that their data will not be connected back to each individual, but there is no guarantee that individual users cannot be identified using this. These shortcomings with Interleukin's privacy policies and the lack of federal oversight express why Interleukin and other personal genomic services need stringent governmental regulations to respect and protect the genetic privacy of customers.

Ancestry Commercial DNA Databases and Genetic Social Networking

Example #1: AncestryDNA

One of the largest genealogical commercial DNA databases belongs to AncestryDNA, a branch of the Ancestry.com family that uses DNA samples from its customers to help trace their ancestry and family tree. AncestryDNA was established in 2011 and since then, it has acquired more than 200,000 DNA samples from people all over the world.[75] It also uses autosomal DNA testing rather than y-chromosome or mitochondrial DNA testing to search for genetic markers at more than 700,000 locations in a person's genome to find potential matches and to examine their ancestry.[76] To participate, customers pay $99 to send in their DNA sample by using the DNA kit and to receive an ancestry report showing them their possible ethnic makeup, geographic origins of their DNA, and allow them to connect with other customers who may be related to them based on partially matching genetic data. AncestryDNA has acquired previous genetic testing ancestry databases such as the Sorenson Molecular Genealogy Foundation to grow its database and allow more potential DNA matches and ancestry results. Their privacy policy states that DNA samples are stored in their database and can be utilized along with the personal information for research purposes. Also, they state that the DNA samples are analyzed and processed in a third party lab in the U.S. which would only have access to the genetic information (DNA samples) submitted.[77]

Example #2: Genebase

Genebase is another genealogical commercial DNA database that offers ancestry services similar to AncestryDNA, but specific components set it apart and allow it to be noted as a prime example of a genetic social networking website. Genebase has 1,674,492 active members and it includes a special service called "DNA Reunion" which makes it unique amongst ancestry DNA databases.[78] DNA Reunion, created in 1998, is a free service which enables people to access the open DNA database which people who have tested at any laboratory can use to upload and share their genetic data to find potential matches.[79] DNA Reunion, unlike other genealogical databases, does not require that people have their DNA tested through Genebase and it does not enforce any limits on who can search the database or how many times.[80] People can also use the database to search for potential matches themselves, unlike AncestryDNA which only allows people to see the profiles of their DNA matches. DNA Reunion essentially allows people to upload their DNA data online into a public database which then leads to genetic social networking. People looking to fill their family tree and parents and children searching for one another to reconnect might be interested in Genebase's services because it provides a service that is more focused on finding relatives and connecting with others based on shared genetic data.

Example #3: National Geographic: The Genographic Project

The National Geographic's "The Genographic Project" is an initiative slightly different from AncestryDNA and Genebase that involves the same element of storing people's DNA to help understand human ancestry. Unlike the two previously mentioned companies, the National Geographic's Genographic Project is a research project which began in 2005 and aims to trace human ancestry collectively as well as individually. For $159, consumers can purchase a DNA kit and send in their DNA sample to be analyzed to receive results that show them where their ancestors may have lived and how they may have migrated.[81] While people are able to see their individual results online using the unique access code assigned to each DNA kit, the purpose of storing the samples is for the team of scientists with the National Geographic Society to be able to use to study human origins and migration patterns. Part of their policy is that all of the results can and will be published in public domain and may be presented on television, the Internet, the radio, and books.[82] Furthermore, in their privacy policy, the National Geographic Society states that they may share information with other organizations it is affiliated with unless participants opt-out of this. They also claim that third party online advertizing agencies may have access to personal information such as age and gender to offer other services that might be of interest.[83] The Genographic Project thus aims to help individuals learn more about their ancestry and to understand human origins by collecting, building, and studying genetic data from their biobank.

Privacy Concerns with Commercial DNA Databases

Surreptitious DNA Collection

Surreptitious DNA collection is when people take or collect a sample of someone else's DNA without their knowledge or permission. An example of surreptitious DNA collection is the police collecting leftover saliva or hair from a crime scene which they would then use to analyze the DNA and find potential suspects. In the new age of personal genomics, surreptitious DNA collection is becoming increasingly popular outside of criminal investigation purposes. Today, the advanced genetic testing technologies and amount of commercial genetic testing companies help enable people to covertly collect DNA samples for their own means, and there is a lack of federal regulations to prohibit this.[84] The New York Times reported an example of surreptitious DNA collection in 2007, when a woman waited outside of a McDonalds for a man who she felt she could be related with to throw out his coffee cup so that she could use it to collect his DNA and confirm her suspicion.[85] The article reported similar cases of people wanting to get a DNA sample from their potential relatives and have it tested to see if they are in fact related, which is possible by sending that DNA sample to certain commercial genetic testing companies that do

not have policies or safeguards to prevent people from doing so. Some of the commercial genetic testing companies have controls in place that aim to prevent people from submitting DNA from people without their consent. For example, 23andMe requires a certain quantity of saliva to be submitted that takes about five to twenty minutes to collect as a step to prevent customers from surreptitiously submitting DNA.[86] However, other direct-to-consumer genetic testing services do not have adequate safeguards to prevent customers from sending in other people's DNA samples. For example, even though AncestryDNA and Genebase state that people must submit their own DNA or that of those they are the legal guardian of, their terms do not include specific measures to prevent people from covertly submitting other people's DNA.[87] Likewise, there is an apparent lack of federal and state policies to prohibit surreptitious DNA testing even though doing so may lead to breaches of genetic privacy and possible lawsuits. For example, a California law prohibits surreptitious DNA testing for health purposes but does not prohibit it for other purposes such as ancestry and paternity.[88] In order to respect the genetic privacy of people who do not want their DNA to be collected and sued, there is a necessity for both the commercial DNA companies and the federal government to adopt strict regulations and appropriate safeguards that would prohibit surreptitious DNA collection.

Genetic Privacy: What about the Third Party Viewers?

Another issue that represents the gap between what customers expect from commercial genetic testing companies and what actually occurs is that of third party viewers and of people other than those that customers expect being able to view their genetic profiles. AncestryDNA, Genebase, and the Genographic Project all note in their privacy policies that DNA samples that customers submit can be viewed and processed by their third party groups. For example, with AncestryDNA, the lab in which they process the DNA samples would also be considered a third party, since the lab is not owned by AncestryDNA. Similarly, most commercial DNA databases involve online payment transactions which might occur through specific companies that are also considered third parties. Also, for Genebase's "DNA Reunion" and the National Geographic's Genographic Project, the entire public can be seen as a third party because these services allow public access to their collected data. Many commercial DNA companies are also highly vague in their explanations of the roles of third parties in their privacy policies, while AncestryDNA and the Genographic Project explicitly note that they will not be responsible for the privacy terms of their third party affiliates and advise customers to go over their privacy policies as well.[89]

While such claims are troubling, a May 2014 lawsuit demonstrates how an ancestry DNA commercial company called Family Tree DNA, which is involved with the Genographic Project, violated one man's genetic privacy. The lawsuit was filed by a man from Alaska named Michael Cole who purchased a Family Tree DNA kit and was outraged after finding out that the results of his DNA test, his DNA kit number, and his personal information including his email address and full name were made publically available online.[90] Even after consulting Family Tree DNA with this, Cole found that by just searching his name online, people could see his genetic information and realized that his information had been shared with RootsWeb, part of Ancestry.com.[91] While

this case continues, it is clear that these ancestry DNA commercial companies cannot guarantee to protect the personal and genetic privacy of customers. Given the possibilities for thirty party members to view their genetic data, customers may ask, what security measures and safeguards do these companies have to prevent unauthorized access? Although many of these commercial genetic testing companies state that they use secure storage systems for the DNA samples and de-identify the samples sent for testing at the laboratories, their exact security procedures need transparency. The lack of actual knowledge on exactly who is able to access these databases and the impact of third party groups and affiliates on genetic privacy demonstrate the need for clear limitations on who can view what information from these biobanks.

Genetic Privacy: Bleeding Between Databases

Another concern about the genetic privacy of customers who submit their DNA to these databases is the potential for bleeding between different types of databases. Many companies say in their privacy policy that if required by law enforcement authorities, they will disclose personal information without the approval of the individual. This represents the potential for bleeding, or interaction, between commercial and forensic DNA databases. If law enforcement officials are able to access commercial DNA databases to assist them in their criminal investigations, what is the difference between the people who willingly submitted their DNA to learn more about their ancestry and the people who were compelled to submit their DNA upon arrest? According to the policies of commercial DNA companies and federal law, there is no actual difference since they allow police to use their databases if required. Along with creating a sense of uneasiness for the customers, this type of interaction shows that the genetic privacy of the customers is almost non-existent, as their DNA samples might become subject to forensic search if a court order or legal proceeding says it is necessary. While some may argue that this helps investigators solve crimes, this does not detract from the larger concern that doing so invades the genetic privacy of millions of people, including the customers and their relatives who share their DNA. The potential for the sharing of information between databases also shows the disparity between people's expectations and the reality, as customers usually do not expect police officers to be able to make use of these commercial databases to find suspects. Also, as mentioned earlier, DNA cannot be canceled as a credit card can be if it is caught in the wrong hands. With privacy policies which admit there is a potential for bleeding between databases and a lack of federal regulations to deny this, there is an evident need for reformative policies and better education for consumers so that they know this.

Lack of Regulations for Commercial DNA Databases

What are the Current Regulations (for Commercial DNA Databases)?

The current regulations that relate to commercial DNA databases include the previously mentioned Genetic Information Privacy Act (GINA) of 2008, which prohibits health insurers and employers from discriminating against people based on their genetic information. Although this act does not directly regulate commercial DNA databases, it serves to help protect people from being fired or denied health insurance because of their genetic data so that if health insurers and employers gain access to people's DNA samples from DNA Reunion, they could not use that to discriminate against people. Unlike for medical and research DNA databases, HIPAA does not apply to commercial DNA databases since HIPAA only regulates covered entities, and direct-to-consumer genetic testing companies are not considered covered entities. Likewise, the federal Common Rule does not apply to commercial DNA databases because the Common Rule only requires informed consent for companies with federal funding, and commercial databases are privately funded.[92] GINA is thus the only current federal regulation which indirectly addresses commercial genetic testing companies.

Why these Regulations are Insufficient

The current regulations for commercial DNA databases are substantially insufficient to protect the genetic privacy of individuals submitting their DNA to these companies. Although GINA protects people from genetic discrimination to a certain extent, limitations within GINA itself disable it from protecting people from all forms of genetic discrimination. Moreover, there is a significant lack of actual federal laws to protect people's genetic privacy by regulating these commercial biobanks. For example, since the Common Rule does not apply, commercial genetic testing companies do not need to seek explicit informed consent from consumers, allowing them to not necessarily include all details in the consent forms. Likewise, since HIPAA also does not apply to commercial DNA databases, there is no guarantee that people will have access to their own genetic data, know exactly how it is being used, or that the data will be de-identified. Even though many commercial genetic testing companies have their own rules and regulations, federal oversight is necessary to make sure that these databases do not jeopardize the genetic privacy of the consumers. Also, rather than just federal laws prohibiting genetic discrimination, policies are needed to prevent unauthorized people from gaining access to these databases, as the invasion of genetic privacy depends on people being able to access genetic data without people's permission. A federal regulatory framework and governmental oversight for these various commercial DNA databases would help consumers and the companies by protecting genetic privacy and allowing the companies to become more transparent and reliable, allowing all both parties to benefit.

Conclusion: The Need for Reform

Mitigating the Privacy Concerns with Non-Forensic Biobanks

Although there are numerous privacy concerns raised with regards to medical, research, and commercial DNA databases, the positive motives and benefits these biobanks have suggest that there is a need to reform the entire system so that these databases continue without so many privacy concerns. The current system is missing substantial federal and state policies to address and resolve people's privacy concerns and proper education for people about the laws that are there to protect their genetic privacy. Given the significant beneficial intentions of non-forensic databases, including furthering medical research and providing people with novel opportunities, getting rid of them is not a solution. Instead, because of the various risks that are associated with non-forensic biobanks, it is necessary to have stringent regulations and federal oversight so that scientists may be able to use donated genetic data to research diseases without compromising the genetic privacy of donors as well as their close relatives. The regulatory reform that all three of these types of non-forensic biobanks require is the pressing need for informed consent. Medical, research, and commercial DNA databases all generate privacy concerns largely because of the misunderstanding between what people anticipate from giving their genetic information to these biobanks and what actually occurs. A strong federal regulatory framework which is lacking can remove this misunderstanding if there is legal enforcement of the federal Common Rule in a way that makes the current gap between what people expect and the realities of biobanking diminish. Further, federal oversight is needed to ensure that these databases are properly secure to prohibit unauthorized personnel from being able to access the samples and genetic data and to make the security systems more transparent so that people are aware of where their DNA is and how it is being used. People's genetic privacy can be protected if they receive proper education about non-forensic databases, actual knowledge about the existing laws and policies, more transparency so that they are aware of how these biobanks operate, and a strong federal regulatory framework to control them.

References

[1] Mei-Hsin Cheng, "Understanding Genetics", *Genetic Testing*, The Tech Museum of Innovation, last accessed 06 Aug 2014, http://genetics.thetech.org/ask-a-geneticist/telling-2-dna-samples-apart.
[2] "Governor's Office of Crime Control and Prevention in Maryland", *History of DNA Databases*, last accessed 06 Aug 2014, http://www.goccp.maryland.gov/dna/database.php.
[3] Twila Brase, "Baby DNA Warehousing", *Gene Watch Page*, Council for Responsible Genetics, Jan-Apr 2014, http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=526.
[4] Ibid.
[5] Vani Kilakkathi, "Newborn Screening In America: Problems and Policies", Council for Responsible Genetics, 2012, http://www.councilforresponsiblegenetics.org/pageDocuments/WNMAKEPP1P.pdf, 5.
[6] Ibid.
[7] Ibid., 4.
[8] Ibid., 13.
[9] "Code of Federal Regulations", *United States Department of Health and Human Services*, 15 Jan 2010, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
[10] Ibid.
[11] Kilakkathi, 15.
[12] Ibid.
[13] "State by State Government Newborn Blood & Baby DNA Retention Practices", *Citizens' Council on Health Care*, 2009, http://www.cchfreedom.org/pr/50_States-Newborn_Blood_Retention_Policies_FINAL.pdf.
[14] "Tag Archives: Newborn Screening", *Genetic Privacy Network*, 25 July 2014, http://geneticprivacynetwork.org/tag/newborn-screening/.
[15] Kilakkathi, 4.
[16] Jeremy Gruber, "CRG Testifies Against Rollback of Newborn Privacy Protections", *Council for Responsible Genetics*, 17 March 2014, http://www.councilforresponsiblegenetics.org/pageDocuments/KKUAMS7VBV.pdf.
[17] Kilakkathi, 21.
[18] Kilakkathi, 24.
[19] Kilakkathi, 16.
[20] Ibid.
[21] "Newborn Screening Translational Research Network (NBSTRN)", *Newborn Screening Translational Research Network (NBSTRN)*, 30 Nov 2012, https://www.nichd.nih.gov/research/supported/Pages/nbstrn.aspx.
[22] Kilakkathi, 16.
[23] "Indiana storing blood & DNA of 2 million children without parents' consent", *Genetic Privacy Network*, 14 July 2014, http://geneticprivacynetwork.org/tag/newborn-screening/.
[24] Elizabeth Cohen, John Bonifield, and Jennifer Bixler, "The Government Has Your Baby's DNA", *CNN*, Cable News Network, 04 Feb 2010, http://www.cnn.com/2010/HEALTH/02/04/baby.dna.government/.
[25] Kilakkathi, 20.
[26] Kilakkathi, 23.
[27] "Newborn Screening Saves Lives Reauthorization Act of 2013 (H.R. 1281)", *GovTrack.us*, last accessed 06 Aug 2014, https://www.govtrack.us/congress/bills/113/hr1281.
[28] "Overview Mayo Clinic Biobank", *Mayo Clinic*, last accessed 06 Aug 2014, http://www.mayo.edu/research/centers-programs/mayo-clinic-biobank/overview.
[29] "For Participants Mayo Clinic Biobank", *Mayo Clinic*, last accessed 06 Aug 2014, http://www.mayo.edu/research/centers-programs/mayo-clinic-biobank/participant-information.

[30] Ibid.
[31] Ibid.
[32] "BioMe Facts - Icahn School of Medicine at Mount Sinai", *Icahn School of Medicine at Mount Sinai*, last accessed 06 Aug 2014, http://icahn.mssm.edu/research/institutes/institute-for-personalized-medicine/innovation-and-technology/biome-platform/biome-facts.
[33] "Information for Participants - Icahn School of Medicine at Mount Sinai", *Icahn School of Medicine at Mount Sinai*, last accessed 06 Aug 2014, http://icahn.mssm.edu/research/institutes/institute-for-personalized-medicine/innovation-and-technology/biome-platform/information-for-participants.
[34] "Kaiser Permanente Division of Research", *The Research Program on Genes, Environment, and Health*, last accessed 06 Aug 2014, http://www.dor.kaiser.org/external/DORExternal/rpgeh/index.aspx.
[35] "Kaiser Permanente Division of Research", *Building a World-Class Genetics Research Program*, last accessed 06 Aug 2014, http://www.dor.kaiser.org/external/DORExternal/research_report/research_rpgeh.aspx.
[36] "Kaiser Permanente Division of Research", *Privacy and Confidentiality*, last accessed 06 Aug 2014, http://www.dor.kaiser.org/external/DORExternal/rpgeh/privacy.aspx?ekmensel=194f64c3_47_51_btn link.
[37] Michael W. Weiner, "ADNI Home", *Alzheimer's Disease Neuroimaging Initiative ADNI,* last accessed 07 Aug 2014, http://www.adni-info.org/Home.aspx.
[38] "Alzheimer's Disease Fact Sheet", *National Institute on Aging*, National Institutes of Health NIH, Sept. 2012, last updated 16 July 2014, http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-fact-sheet.
[39] "World Wide Alzheimer's Disease Neuroimaging Initiative", *Alzheimer's Association*, Alzheimer's Association, last accessed 07 Aug 2014, http://www.alz.org/research/funding/partnerships/WW-ADNI_north-america.asp.
[40] "ADNI Full Genetic Sequences Now Available for Download | ALZFORUM", *ALZFORUM Networking for a Cure*, Biomedical Research Forum, LLC., 4 Oct 2013, http://alzforum.org/news/research-news/andi-full-genetic-sequences-now-available-download.
[41] "Personal Genome Project: Harvard Medical School", *Personal Genome Project: Harvard Medical School*, last accessed 07 Aug 2014, http://www.personalgenomes.org/harvard/about-pgp.
[42] Church M. George, "Patient Consent Form", *Personal Genome Project*, Harvard Medical School, 18 Feb 2014, http://www.personalgenomes.org/static/docs/harvard/PGP_Consent_2014-02-18_online.pdf.
[43] "How it Works", *Personal Genome Project: Harvard Medical School*, last accessed 07 Aug 2014, http://www.personalgenomes.org/harvard/howitworks.
[44] "Researchers Raise Privacy Concerns Over DNA Databases", *iHealthBeat*, iHealthBeat-Reporting Technology's Impact on Health Care, 17 June 2013, http://www.ihealthbeat.org/articles/2013/6/17/researchers-raise-privacy-concerns-over-dna-databases?view.
[45] Ibid.
[46] Gina Kolata, "Poking Holes in Genetic Privacy", *The New York Times*, The New York Times, 17 June 2013, http://www.nytimes.com/2013/06/18/science/poking-holes-in-the-privacy-of-dna.html?pagewanted=all&_r=0.
[47] Latanya Sweeney, "k-Anonymity: A Model for Protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10, no. 5 (2002), 557-570, http://arbor.ee.ntu.edu.tw/archive/ppdm/Anonymity/SweeneyKA02.pdf.
[48] Ibid., 559.
[49] "Kaiser Permanente Division of Research", *Privacy and Confidentiality*, last accessed 06 Aug 2014, http://www.dor.kaiser.org/external/DORExternal/rpgeh/privacy.aspx?ekmensel=194f64c3_47_51_btn link.
[50] Yannis Stamatiou, "Security Aspects of DNA Databases", Interscience Molecular Oncology Laboratory (IMOL), University of Ioannina Cancer Biobank Center (UICBC), 2010, http://www.imol.edu.gr/home/Publications-Presentations_files/secure_dna_databank_new.pdf.

[51] Ibid.
[52] Christine Rosen, “"Liberty, Privacy, and DNA Databases”, *The New Atlantis*, no.1 (Spring 2003), http://www.thenewatlantis.com/publications/liberty-privacy-and-dna-databases, 37-52.
[53] Jeremy Gruber, “The New Genetic Information Nondiscrimination Act”, *Gene Watch Page*, Council for Responsible Genetics, May 2009, http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=526>.
[54] “HIPAA and Privacy”, Genetic Privacy Network, last accessed 07 Aug 2014, http://geneticprivacynetwork.org/hippa-privacy/.
[55] “HIPAA PHI: List of 18 Identifiers and Definition of PHI”, *UC Berkeley Committee for Protection of Human Subjects*, last accessed 07 Aug 2014, http://cphs.berkeley.edu/hipaa/hipaa18.html.
[56] “Code of Federal Regulations”, *United States Department of Health and Human Services*, 15 Jan 2009, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
[57] “CMIA”, Genetic Privacy Network, last accessed 07 Aug 2014, http://geneticprivacynetwork.org/genetic-research/.
[58] Jeremy Gruber, “The New Genetic Information Nondiscrimination Act”, *Gene Watch Page*, Council for Responsible Genetics, May 2009, http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=526>.
[59] Madeline Ball, “Personal Genome Project Study Material”, Personal Genome Project: Harvard Medical School, last accessed 07 Aug 2014, http://www.personalgenomes.org/static/docs/harvard/v20120430-study-guide.pdf, 20.
[60] “Code of Federal Regulations”, *United States Department of Health and Human Services*, 15 Jan 2009, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
[61] Ibid.
[62] Amy Harmon, “Indian Tribe Wins Fight to Limit Research of Its DNA”, *The New York Times*, The New York Times, 21 Apr 2010, http://www.nytimes.com/2010/04/22/us/22dna.html?pagewanted=all.
[63] “Frequently Asked Questions (FAQ)”, Instant Chemistry Inc., last accessed 07 Aug 2014, http://instantchemistry.com/faqs/.
[64] “Status of our health-related genetic reports”, 23andMe Inc., last accessed 07 Aug 2014, https://www.23andme.com/health/.
[65] “Core Values”, 23andMe Inc., last accessed 07 Aug 2014, https://www.23andme.com/about/values/.
[66] “How it works”, 23andMe Inc., last accessed 07 Aug 2014, https://www.23andme.com/howitworks/.
[67] “Research Consent Document”, 23andMe Inc., last accessed 07 Aug 2014, https://www.23andme.com/about/consent/.
[68] “Full Privacy Statement”, 23andMe Inc., last accessed 07 Aug 2014, https://www.23andme.com/about/privacy/#Full.
[69] “Genetics for the People?”, Genetic Privacy Network, 07 Feb 2014, http://geneticprivacynetwork.org/tag/council-for-responsible-genetics/page/5/.
[70] “About Interleukin Genetics”, Interleukin Genetics Inc., last accessed 07 Aug 2014, http://www.ilgenetics.com/content/about-interleukin.
[71] “What is Genetic Testing, How it Works and How to Get it | Inherent Health FAQ”, Inherent Health Insight into Wellness, Interleukin Genetics Inc., 2009, http://www.inherenthealth.com/faq.aspx.
[72] Ibid.
[73] “Clinical Laboratory Improvement Amendments (CLIA)”, CMS.gov Centers for Medicare & Medicaid Services, 27 Feb 2014, http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html?redirect=/clia/.
[74] “Privacy Policy”, Inherent Health Insight into Wellness, Interleukin Genetics Inc., 15 June 2009, http://www.inherenthealth.com/privacy-policy.aspx.
[75] “Ancestry.com | Press Releases.”, *Ancestry.com | Press Releases*, last accessed 07 Aug 2014, http://corporate.ancestry.com/press/press-releases/2013/10/AncestryDNA-Now-a-More-Comprehensive-DNA-Test-for-Exploring-Ethnic-Origins/.

[76] "AncestryDNA-Frequently Asked Questions", AncestryDNA, last accessed 07 Aug 2014, http://dna.ancestry.com/atFAQ.
[77] "Ancestry DNA Privacy Statement", AncestryDNA, 20 March 2013, http://dna.ancestry.com/legal/privacyStatement.
[78] "Genebase :: Home", *Genebase :: Home*, last accessed 07 Aug. 2014, http://www.genebase.com/.
[79] "Frequently Asked Questions", DNA Reunion, Genebase, last accessed 07 Aug 2014, http://www.dnareunion.org/article.php?type=start&page=4&l=n&__atoken=__NONE__.
[80] Ibid.
[81] "The Genographic Project - Frequently Asked Questions - About the Project", *Genographic Project*, National Geographic, last accessed 07 Aug. 2014. https://genographic.nationalgeographic.com/faq/about-project/.
[82] Ibid.
[83] Ibid.
[84] "Covert DNA Collection", Genetic Privacy Network, last accessed 07 Aug 2014, http://geneticprivacynetwork.org/covert-dna-collection/.
[85] Amy Harmon, "Stalking Strangers' DNA to Fill In the Family Tree", *The New York Times*, The New York Times, 01 Apr 2007, http://www.nytimes.com/2007/04/02/us/02dna.html?_r=1&oref=slogin.
[86] CRG Staff, "In their words: 23andMe", *Gene Watch Page,* Council for Responsible Genetics, last accessed 07 Aug 2014, http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=283.
[87] "US Terms and Conditions", AncestryDNA, 20 March 2013, http://dna.ancestry.com/legal/termsAndConditions#4;"Terms and Conditions", Genebase, last accessed 07 Aug 2014, http://www.genebase.com/term.php.
[88] "Covert DNA Collection", Genetic Privacy Network, last accessed 07 Aug 2014, http://geneticprivacynetwork.org/covert-dna-collection/.
[89] "AncestryDNA-Frequently Asked Questions", AncestryDNA, last accessed 07 Aug 2014, http://dna.ancestry.com/atFAQ.
[90] Kyla Asbury, "Alaska class action lawsuit says Family Tree DNA posted info on public websites", Legal Newsline, Legal Newsline Legal Journal, 16 May 2014, http://legalnewsline.com/issues/class-action/249202-alaska-class-action-lawsuit-says-family-tree-dna-posted-info-on-public-websites.
[91] CRG Staff, "Update: DNA Ancestry Company Accused of Sharing Personal Data", *Gene Watch Page*, May-July 2014, http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=535.
[92] Sara Fendrick, "The Role of Privacy Law in Genetic Research", A Journal of Law and Policy for the Information Society 4, no. 3, 2008, http://moritzlaw.osu.edu/students/groups/is/files/2012/02/Fendrick_Formatted_Final.pdf.