Ashley Tehranchi
BIOC258 6/8/12

An Analysis Of Personal Genomics

INTRODUCTION

Currently we find ourselves in the age of genomics and with one foot in the door of personal genomics. As the price of whole genome sequencing becomes cheaper every year, more of the general public is becoming concerned with personal genomics as it relates to their health. This paper addresses the current state of personal genomics, potential directions for improvement and the future of the field. In particular, I will address the benefits and risks of companies offering personal genomics services, the uses and implications of personal genomics, the future directions, and how we may predict diseases from personal genomics datasets.

BENEFITS/RISKS OF PERSONAL GENOMICS COMPANIES

Currently, there exist two direct-to-consumer personal genomics companies, 23andMe and Navigenics. 23andMe is available to the public for a fee of $300, while Navigenics is available to public only through a doctor's prescription and the price varies[1,2]. These companies aim to provide the same service but have a different philosophy fueling their companies. 23andMe allows the public to directly access their genome either because they believe everyone has the right to know their genome, or

because they believe this approach will be more profitable. Navigenics, on the other hand, only allows someone to access his or her genome with doctor supervision. This is either because they believe that people are not qualified to look at their genome and make their own decisions, or because they believe it will be more profitable. There is no doubt, however, that 23andMe is winning the profits race, likely because of its accessibility to the public.

The major difference between 23andMe and Navigenics is the services offered by each. 23andMe uses an Illumina OmniExpress Plus to analyze about 1,000,000 SNPs that cover the entire genome[1]. Navigenics on the hand uses an Affymetrix 6.0 chip and tests for only 28 health conditions, and 12 medication sensitivities[2]. Navigenics, however, does have a more clinical focus because they require doctor supervision and offer genetic counseling for patients, which 23andMe does not.

Benefits

One of the biggest advantages of personal genomics is that it can tell you with nearly 100% accuracy whether you are a carrier for genetic diseases such as cystic fibrosis and breast cancer. This information can be extremely useful to know if, for example, you and your partner are both carriers for cystic

fibrosis, so that you will be aware that any children you have will have a 25% chance of having the disease.

Additionally, personal genomics can offer insight into your current calculated risk of diseases like Alzheimer's Disease as well as your sensitivity to certain drugs such as Warfarin. While these calculated risks are not the complete picture, they do give a sense of awareness and can be a note of caution to users.

Another potential benefit of sequencing all of these individuals is that companies may be able to use all the genomic data that they collect to conduct their own GWASs to improve their prediction abilities. This would benefit the companies because the individual would cover the cost. This also benefits the individual because they can receive better risk predictions. Additionally, this information could be shared with the scientific community to further research that would also benefit these companies.

Disadvantages/Risks

While there are many advantages to personal genomics, it seems that currently, there may be more disadvantages for the general public. Most of the SNPs currently tested by companies address disease risk, disease carrier status, and drug responses but

fail to do so for many diseases/drugs with significant accuracy[3]. This is because the scientific community has yet to understand all of the genetics underlying any given disease and can only predict a small percentage of risk. The level of predictable risk however, is becoming better all the time as progress is made towards understanding how the genome can predict disease risk. Additionally, it should be remembered that because a SNP is associated with a disease, does not directly translate to causation of a disease. As our current disease detection ability stands, we sometimes call a disease association using tag SNPs (SNPs that rely on linkage disequilibrium (LD) to detect tag SNPs that are proxies for the causal variants)[4]. This can easily be misleading and cause someone to believe a strong association based on a SNP that is not causal.

Another potential disadvantage of personal genomics is that most of the studies have only been performed in Europeans[5]. Although this is beneficial for Europeans it neglects other populations, which can vary substantially in causative disease SNPs. While companies like 23andMe do indicate European only SNPs, it is not extremely clear to visualize on their website. Also, this does not rule out the possibility that your risk is the same as the European risk or how it may vary. This indicates that some of the information presented by companies pertaining to causative disease SNPs may in fact be false for an individual of an

ethnicity other than European. If this were the case for an individual, it may follow that the individual would wrongly infer their disease risk or lack of risk according to the information presented by the company. It should be noted that the disadvantage of having mostly European studies to draw from is of no fault of the companies that offer the services of personal genomics but instead of the funding granted for such research[6].

I think it is important for the market to have both types of companies/services available so that people who would like help interpreting their genome can do so and others who just want the data and can interpret it themselves can do so. I do also believe that everyone who has their genome sequenced should be used for data collection and have a complete history annotated for their genome and have their identity removed. This data should also be collected in a consistent manner between all current and future companies and be publicly available. As more and more people sequence their genomes, we can simultaneously build up a database larger than any that has been built and hopefully, in the near future, these companies will be able to offer sequencing of the entire genome, which will aid researchers tremendously in figuring out disease genetics.

THE USEFULNESS AND IMPICATIONS OF PERSONAL GENOMICS

One question to consider before sequencing your genome is, how useful would it be and what can I learn from it? This question cannot truly be answered until after the sequencing, as you cannot know what you may learn until you have the facts in front of you. For example, the drug Warfarin is a blood thinner used to treat heart attacks. Having the right dose of the drug is critical for survival: too little and the patient's blood will clot, too much and the patient will bleed excessively. Warfarin sensitivity has several associated SNPs that can predict if you may require more, less, or the average dose of the drug. If you were one of the individuals that required more or less Warfarin, having this information and sharing this with your doctor could save your life in the event of a heart attack. Additionally, in the case of cystic fibrosis, if a couple considering having children found they were both carriers for the variant, they may reconsider having children given their carrier status. In cases such as these, sequencing your genome would be of tremendous value.

One shortcoming of personal genomics is that it cannot answer questions like, will I get Parkinson's Disease? This is perhaps the type of question most people want an answer to but one we cannot begin to answer. This is because we cannot explain all of the heritability of a disease and have not even begun looking at other factors influencing disease, such as environmental

factors. What information companies can offer regarding disease risk such as Parkinson's or Alzheimer's Disease is minimal because our current understanding of the genetics is minimal. As we learn more, our predictive abilities will increase. A potential drawback to giving the layperson information such as, "you have a 2-fold increased risk for Parkinson's Disease" is that this information may be misconstrued. An individual may believe that a 2-fold increased risk means that they will get Parkinson's. Of course this is not true, but that may not be how the information is perceived. This could cause the individual to be depressed or cause them to act out of character or respond drastically. Because of this possibility, companies offering these services should be explicit about actual disease risk, compared to an elevated risk over average, as well as remembering to take into account ethnicity as an additional factor.

When addressing the question of usefulness, we must also consider the question of actionable treatment. It remains that for all cases of diseases reported, there is not any medically translatable patient cure, although there is usually some form of treatment. For diseases such as Huntington's, companies do not report results because there is no form of treatment. While companies do not report this information for ethical reasons, it may good to offer it as an option so that people can choose for

themselves if they want the information, as they may decide to live their lives differently given that information. For other diseases such as Parkinson's or Alzheimer's there are forms of treatment but no cure. This information may lead to difficult decisions about whether or not to take drugs to impede progression of the disease and at what age to take them. Also, one should consider the side effects of a given drug as they may outweigh the potential benefits. Additionally, at what threshold of elevated risk should drug therapy be prescribed? All of these are important considerations when discussing treatment and also require approval of your doctor who may not know what to do with your genomic data.

FUTURE DIRECTIONS FOR PERSONAL GENOMICS

One question society always contemplates is, what does the future hold? While we cannot know for sure, we can make some educated guesses. It is foreseeable that in the not too distant future we could see genetic counselors becoming an integral part of every doctor's office. A genetic counselor could give us a prescription for tests to run of diseases we may be at risk for and that prescription would be given to our doctor who could run the tests and decide how to proceed with treatment.

Currently, it is not feasible to have a place for personal genomics in the doctor's office, mainly for two reasons. First,

we must consider that doctors vary in age from new medical students to those who have been practicing medicine for more than half a century. Even new doctors, fresh from medical school, are not being taught how to interpret a genome. Additionally, it is not practical for all the doctors in the country to retroactively learn about personal genomics and how to apply it. Secondly, we do not see personal genomics in the doctor's office today because some doctors who do not understand genomics, are also not accepting of incorporating genetics into their treatment plans via a genetic counelsor. This idea is not completely unfounded as the numbers of GWASs that have translated into significant treatments have been minimal and so doctors may not see the long-term benefits of genomics. As genomics begins to translate into medicine, we will see more and more doctors incorporating genomics into their practices. While some people believe that doctors should be both the genetic counselor and the doctor, I think there is too much knowledge required for an individual to be an expert in this many fields. A genetic counselor should be a geneticist and be up to date on all relevant literature and be able to filter new studies as to whether they are scientifically accurate or not, while a doctor should be an expert in the physical and anatomical nature of the human body. While these two areas are

very related, they are not equal and require ones full attention to detail as the life of an individual depends on it.

HOW TO PREDICT DISEASE FROM GENETICS

One major goal of genomics has been to attain the power to predict diseases with the intention of learning how to cure them. The key to achieving this idea was thought to be that simply sequencing the human genome would reveal all the answers but in fact, in comparison to what scientists expected, this turned out to be a quite lofty goal. It has been shown that the amount of heritability that can be explained today is at best around 20-30% for well-studied cases and even less for all others[7]. This low percentage of explained heritability has incurred more questions than answers, namely, why can we explain so little if we have unlocked the genetic code?

There are many potential answers to this question including some that we cannot fathom, however the simplest answer is that we are not looking in the right places. Currently there are two ideas about why we cannot account for all of the heritability; (1) All the variants affecting heritability have yet to be discovered and/or (2) The way which we calculate heritability is faulty and inadvertently creates "phantom heritability". The only way to address undiscovered heritability is to continue to do more GWASs with excellent controls and parameters and more

statistical power to continue to unveil new SNPs. However, the method by which we calculate heritability is an aspect that can be investigated further. The current method by which we calculate heritability is to take the ratio of the proportion of the phenotypic variance explained by the additive effects of known variants and divide that by the proportion of the phenotypic variance attributable to the additive effects of all variants that include undiscovered variants[8]. We can calculate all variants but must infer those variants that have not been discovered and in doing such, may overestimate this number and thereby underestimate the heritability of a set of variants[7]. This overestimation could, in part, be due to epistatic interactions where they are counted for each epistatic interaction when there is only one variant. Therefore, it may be possible that we are inflating the amount of missing heritability because of a failure to account for epistasis. Another current issue concerning how we screen for diseases via genetics is, who are we screening? It is known that the vast majority of GWAS performed have used populations of European ancestry. While there has been some effort recently to use cohorts of varying populations from all over the world, these attempts have been overshadowed by European only studies. The significance of performing studies in mostly one ethnicity has a couple of major implications. (1) All other non-European

ethnicities have SNPs that vary from Europeans and may also vary in how that SNP is manifested. Additionally, because it remains unknown the effects of a given SNP, these individuals may not be aware they have a disease or may falsely believe they have a disease and may miss out on treatment options or receive unnecessary treatments, respectively. (2) In todays world it is becoming increasingly difficult to categorize an individual as a particular race. Many people, especially Americans, have mixed backgrounds and ethnicities and for these individuals they do not belong to just one race. Even among many GWASs that claim that they only investigated Europeans for example, this is determined by the patient's self-reported race that may not reflect their correct ancestry. This could lead to confounding results in data analysis because these individuals may falsely deflate a real SNP association for a particular disease. The answer to solving this problem is to continue to sequence more people from all over the world until we can sequence a large number of people from every inhabited place or, ideally, every inhabitant. If we can do this, we can eliminate the concept of race and can build a predictive framework for diseases based on knowledge we gain from the entire world.

Such a type of predictive model could be used to incorporate not only genetics, but also other factors that also contribute to disease. One major contributing factor to disease is the

underexplored environmental impact. We know that the environment can play an influential role in diseases; for example, air pollution can cause an increased risk for cancer. It is then logical to assume that we cannot begin to accurately predict diseases if we do not begin to incorporate environmental conditions into the equation. How to solve this problem, again begins with collecting more genomes but also simultaneously collecting a family history as well as an environmental history. An environmental history meaning a history of where an individual was conceived and born, maternal habits while pregnant, parental child rearing habits, places of residences and so forth. By collecting all this data we can look for connections and patterns in environmental conditions and build on the genetics to make a more accurate predictive model.

CONCLUSION

Over a decade ago scientists brought us the human genome, and since then we have begun exploring the genome and looking for insight into the implications of our genetic code. We are in the beginning of the personal genomics era, which has opened many doors for researchers as well as the average person. Personal genomics is the future of medicine and will likely continue to be a fruitful field for many years to come.

REFERENCES

1. www.23andme.com
2. www.navigenics.com
3. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature 2009;461:724-726 [PubMed 19812653]
4. Stram, D. O. Tag SNP selection for association studies. Genet. Epidemiol., 2004:27: 365-374. [PubMed 15372618]
5. Haga SB. Impact of limited population diversity of genome-wide association studies. *Genet Med* 2010;**12**:81-84. [PubMed 20057316]
6. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. Trends Genet. 2009:25,489-494. [PubMed 19836853]
7. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci USA. 2012;109:1193-1198. [PubMed 22223662].
8. Sang Hong Lee, Naomi R. Wray, Michael E. Goddard, Peter M. Visscher, Estimating Missing Heritability for Disease from Genome-wide Association Studies, The American Journal of Human Genetics. 2011:88,3,294-305. [PubMed 21376301]