Re-Examining the Gene in Personalized Genomics

JORDAN BARTOL

phjnb@leeds.ac.uk

Centre for History and Philosophy of Science, University of Leeds

Abstract Personalized Genomics companies (PG; also called 'Direct-to-Consumer Genetics') are businesses marketing genetic testing to consumers over the Internet. While much has been written about these new businesses, little attention has been given to their roles in science communication. This paper provides an analysis of the gene concept presented to customers and the relation between the information given and the science behind PG. Two quite different gene concepts are present in company rhetoric, but only one features in the science. To explain this, we must appreciate the delicate tension between PG, academic science, public expectation, and market forces.

1.0 Introduction

The concept of the gene has a long and multifarious history. It has been pronounced 'dead' (Gray 1992) only to be later revived (Neumann-Held 1999), heralded as the hallmark scientific concept of the 20th century (Fox Keller 2000), then declared obsolete in light of more precise biological taxonomy (Fogel 2000), all the while being redefined countless times for countless purposes. In spite of this controversy, the gene has not disappeared from view. Spurring and capitalizing on the public fascination with genetics, a new type of commercial enterprise has recently arisen offering genetic testing directly to consumers. Several such Internet-based companies are now thriving. These Personalized Genomics companies (sometimes called 'Direct-to-Consumer Genetics' or 'Retail Genetics') have aroused a great deal of attention from governments rushing to erect regulations, from journalists sensationalizing scientific advancements, and, perhaps most importantly, from consumers eager to partake in the gains of modern science. Amid all of this excitement, it is important to pause for a moment and examine the concept that is central to Personalized Genomics (PG). I believe it is prudent to determine what, precisely, the 'gene' in Personalized Genomics really is and what it can purportedly do.

Scientific concepts in commercialized science are an important topic. These concepts are powerful and their effects are wide-reaching. For many, popular culture and the media have long been the only sources of education about molecular biology and genetics (Bates 2005). Yet information from these sources is rarely in-depth, and has been hotly criticized (Kua et al 2004). PG is a new source of information about genetics, one which customers may view as more reliable than other sources (compare Bates 2005 and Kaufmann 2012). It is therefore prudent to determine what sorts of knowledge about genes and genetics the public is liable to receive from PG. Studies have already shown that most customers view PG as a source of knowledge about disease (McGuire et al. 2009) and that many intend to use knowledge gained to inform health decisions (Kaufmann 2012). Though customers report a high-degree of confidence in their understanding of PG science, that confidence may be misplaced. Leighton and colleagues (2012) report that customers' frequently misinterpret PG results in spite of self-assessments to the contrary. This is worrying when we consider that over 1/3 of prospective customers expect results to be equivalent to medical diagnoses (McGuire et al. 2009).

Customers' poor comprehension of PG science has been attributed to both the extremely high literacy demands of PG websites (Lachance et al. 2010) and the public's generally poor understanding of statistics (Leighton et al. 2012).   I suggest that examining the concept of the gene will shed new light on the nature of this problem. Further, misapprehension the science, via misunderstanding of the gene, is a problem

that extends beyond the scope of PG tests themselves. Knowledge of genetics will go on to inform future decisions about and assessments of genetic and biological claims. Lay conceptions of genetic causation have implications for topics from racism and prejudice (Condit 2011; Bates 2005), to more everyday topics like fairness in sport (eg. see exchange in Cohen 2008). Beyond these practical implications, epistemic consequences are a virtual certainty as false beliefs about genetics form the basis for yet further misunderstanding. Evaluating the quality of information on offer is therefore important for both customers' understanding of PG tests themselves, and for their understanding of science, more generally.

2.0 Personalized Genomics

Although there are a number of types of business to which 'Personalized Genomics' might refer, I will focus only on one such business model here. These are 'complete' genome scans, so-called because they assess large samples of DNA from across the genome and forward predictions about a wide range of traits. I have chosen to focus on three companies as representative of this class: Navigenics, 23andMe, and deCODEme.

2.1 THE SCIENCE BEHIND PERSONALIZED GENOMICS

Personalized Genomics companies operate almost exclusively online. Although some companies have attempted to sell their products in drug store chains, these moves have been unsuccessful (see Lakhman 2010). Testing kits, which range in price from $200 to over $2000, generally contain a saliva tube or cheek swab. Once the DNA is mailed and processed, the consumer accesses test results via the company website. In some instances, an ongoing fee is charged for continued access and genetic counselling.

From the customer's saliva sample, PG companies are able to extract DNA. The DNA is scanned for a pre-determined set of data points, which are interpreted in light of massive DNA-trait correlation databanks. These databanks have been established thanks to a new type of study known as the Genome Wide Association Study (GWAS). Older and more labour intensive genome mapping techniques required hypothesis-driven research. This forced researchers to go into gene-trait correlation studies with a target genetic marker in mind, searching for just that marker (called the 'candidate gene approach'). It is now possible to sample an extremely wide array of markers from across the genome, approaching GWAS without *a priori* hypotheses about which markers will be important (Hunter & Chanock 2010). In this spirit, PG tests and GWAS proceed with up to two million data points per subject. This process begins by focusing in on the areas in which human DNA typically varies; these are locations where a single nucleotide has been substituted for another, the results of copying errors at some point in the ancestry. There are an estimated 11million such Single Nucleotide

Polymorphisms (SNPs) in the human genome (Kruglyak 2008), though not all of these are known. It is thus only a subset of all extant SNPs that are included in tests, and a much smaller subset of those SNPs are actually known to confer any information about traits.[1]

SNPs known to have meaningful correlations (negative or positive) with traits are known as 'risk alleles', and the degree of association is known as the 'risk association'. These data furnish the basis of PG assessments. With knowledge of which risk alleles a customer possesses PG companies search for specific risk alleles, determining relative risk for a number of different traits.

For each trait, the PG customer receives three figures:

(1) the population frequency of the trait;

(2) the subjective risk assessment;

(3) the adjusted risk.[2]

Population frequency (1) is the percentage of the population that will have the trait in question. Usually, this population is not the general population, but that population narrowed by gender and/or ethnicity, where possible.[3] The average risk of Ulcerative Colitis, for European white males, for instance, is 0.77%. This means that 7.7 in 1000 members of this population will develop Ulcerative Colitis in their lifetime. The subjective risk assessment (2) is the sum total of risk associations from the risk alleles present in the customer. If, for instance, the customer has a risk allele with a risk association of 1.3, then he is at a 30% increased risk relative to the population mean (1). Multiplying the population risk (1) by the subjective risk (2) gives us the total adjusted risk (also called the 'absolute risk') (3). To apply this to colitis, this customer would have an adjusted risk of 1.001% [0.77(1.3) = 1.001].

(1) 0.77%

(2) 1.3

(3) 1.001%

That means that ~10 of 1000 males of European descent with this risk allele will develop Colitis.

---

[1] It is interesting that PG companies use the number of SNPs included in their tests as marketing tools. Some boast of using 2 million where others used only 1.5. Since only a few thousand of those SNPs confer any information, the gross amount of SNPs tested is little indication of test quality.

[2] Some companies omit (2), requiring you to infer the subjective risk, or omit (3), requiring you to calculate adjusted risk.

[3] There are some interesting and important questions about the degree to which this is accomplished. Many ethnic groups are not represented in the GWAS studies on which PG results are based. See (Mountain et al., 2007).

Customers are provided with lists of various traits and conditions along with information about any risk alleles where applicable. Descriptions of traits are also provided. Reporting results in this fashion ought to provide the customer with an idea of her risk, relative to the risk of members of the population of which she is a part.

2.2 PROBLEMS FOR THE PG APPROACH

When research first began into SNP-trait correlations, the hope was that they would uncover the genes responsible for nearly all heritable illnesses. Although genetic variants implicated in many traits have been uncovered, the magnitude of these results have been disappointing. As one researcher comments, "common [genetic] variation is packing much less of a phenotypic punch than expected" (Goldstein 2009, p.1696). Often, traits with estimated high heritability have been only partially accounted for in GWAS results.[4] The risk associations yielded by GWAS have often been so small that when used by PG they confer marginal increases in estimated risk and are thus often too low to be of clinical value.

It is tough to pin down the precise average risk association used by PG, but most sources cite figures ranging from 1.1 to 1.5 (Hudson 2008; Hunter et al. 2008; McGuire & Burke 2008). A risk association of this magnitude would increase one's risk to between 10% and 50% above that of the regular population. To put this in perspective, for a disease with a population frequency of 10%, a risk-assessment of 1.5 (on the high end of the average) would raise the person's personal risk to 15%. That is not much of an increase, hardly enough to invoke a sense of urgency. What type of intervention might be available, desirable, or necessary for a 15% risk that is neither available nor appropriate for a 10% risk?

Many of the behavioural changes recommended on the basis of small increases in risk are behaviours that are recommended of the general public anyhow. Referring to this obstacle, one group of researchers warned, "until the genome can be put to useful work, [one would be] better off spending their money on a gym membership or a personal trainer" (Hunter, Khoury et al. 2008). Though it may be interesting to know that you carry a marginal increase in risk for some trait or condition, the practical value of that information is often quite small. John P. Ioannidis, one of the most widely published commentators on GWAS, notes that many people may be unaware of the low clinical utility of these predictions,

> [T]he utility of the genetic tests rests on the brittle assumption that there is a very specific and noncontestable risk threshold that leads to very different action plans. (Iaonnidis 2009)

---

[4] Perhaps the most discussed of these is height, which is estimated to be 80-90% heritable, but for which SNPs account for only 5%. The issue of missing heritability is a complex one. The poor risk associations discussed here are but a symptom of this greater problem. On the problem of heritability estimates, see (Sesardic 2005). On the problem of missing heritability, see (Maher, 2008).

Marginal increases in risk do not permit any change in recommended preventive measures. Unless the risk associations utilized in PG can drastically increase, it is unclear that the results will translate into meaningful changes to clinical practice.

The poor results offered by PG tests may go unnoticed. Indeed, customers demonstrate a propensity to overestimate the significance of the results they receive (Leighton et al. 2012). A defender of PG might wonder whether some utility might come, if not from the results themselves, then from increased knowledge and awareness of genetics and molecular biology. In the next section, I analyze the concept of the gene presented in PG and ask whether this tuition compensates for the poor clinical utility, and, perhaps equally importantly, whether it may contribute to or counteract customers' demonstrated ignorance of the nature of genetic risk.

3.0 What and Where are the Genes in Personalized Genomics?

PG websites, press releases, and advertising feature the gene prominently. From video tutorials to terminology sheets, the gene crops up with considerable frequency. Note that my account of PG reports presented above, however, the term 'gene' does not feature once. Indeed, risk is based on SNPs. One can present the material, as I have, in such a way as to render the relationship between SNPs and genes an unnecessary part of understanding of risk reports. Yet PG companies make the gene necessary for understanding the science by featuring the concept so prominently and by connecting DNA to risk via genes. Customers will surely wonder just what genes are and what they can do. Whatever the gene concept at work in PG, we should find evidence of this presented on their websites, written in the information provided to customers and physicians, and embedded in their scientific practices. Uncovering information about the gene concept on which PG relies will require first analysing the relationship between SNPs and genes, then analyzing how SNPs and genes are taken to affect traits, and the role accorded to variables other than DNA.

There are at least two points of contact for the gene in PG. The first of these is in the science behind PG, the second is in the customer perception of genetics. Examining the former will involve examining the technology and techniques on which the industry relies[5]; the latter requires an examination of company websites, publications, and

---

[5] To be clear, we cannot access what the PG scientists *really* think genes are. To access scientists' inner thoughts isn't possible. The best we can do is to analyze the material they present to their customers and examine their scientific practices. How they actually conceive of and use the concept is beyond the scope of this sort of investigation. For more on the epistemic and psychological barriers to reconstructing scientific concepts, see (Waters, 2004).

customer reports. Whether the gene presented to customers aligns with the gene concept used in PG science is an open question.

3.1 SNPS AND GENES

To determine the gene concept at work in the collection of data, it is best to begin by determining how genes are conceptualized in the context of GWAS, the science that features most prominently in PG practice.

PG companies are somewhat indiscriminate with regard to the selection of SNPs. Roughly 57% of the trait-associated SNPs identified by GWAS fall within protein-coding DNA (Hindorff et al. 2009). By including some variations occurring outside of these regions, GWAS are also capable of capturing risk allele correlations for traits affected by variations in other regions – sequences such as regulatory regions, for instance. When SNPs falling in non-coding regions are identified, the standard approach is to determine the variant in a protein-coding region for which that SNP is an indicator. A common suggestion is that we treat SNPs as 'surrogates' for variants in protein-coding regions (Hunter & Chanock 2010).[6] The authors of a recent review of GWAS results suggest that we should not be too quick to assume that risk association in a non-coding SNP is the 'true' cause, and not merely an indicator of the 'true' cause of the trait in a protein-coding region elsewhere (Hindorff et al. 2009). Similarly, an early review forecasting progress for GWAS reminds,

> [M]ost of the [correlating DNA sites] are more likely tagging SNPs than the causal sites, so it is not clear what their relationships to … the true disease-promoting variants are. (Gibson & Goldstein 2007, p. 931)

Practice is thus to treat SNPs as a resource for probabilistic claims about traits while assuming that they are indicative of a variation in a protein-coding region that serves as the cause. The SNP approach is grounded in a very typical notion of genes as protein-coding DNA. This view, typical among molecular biologists, is often called a 'molecular gene' (Waters 1994; Griffiths & Stotz 2007; cf. 'nominal gene' in Burian 2004). Although this is a common view, it is worth noting that the molecular gene is not the only category of gene concept in use. A more general view of the gene is in use among branches of population genetics, biometrics, and plant breeding. This concept requires no physical referent, molecular or otherwise. It acts as an instrumental variable, useful for tracking

---

[6] This likely owes, at least in part, to early Linkage Disequilibrium (DL) mapping, a technique for mapping polymorphisms, which evolved into contemporary GWAS. Early DL approaches, relying on restriction fragment length polymorphisms and familial transmission data, were predicated on the fact that polymorphisms shared by individuals related ancestrally are often surrounded by shared alleles at nearby loci. The polymorphism is thus treated as a *marker* for the 'true' source of the trait(s) in question. In contemporary studies, however, samples are not restricted to family lines, and thus researches ought not assume so readily that a polymorphism is indicative of shared alleles. Yet this caveat may have been neglected, and surrogate assumption adopted erroneously, in contemporary GWAS (see Kruglyak, 2008).

phenotypic transmission. A number of philosophers have highlighted the importance of recognizing this gene in addition to a molecular gene (see Lenny Moss' 'Gene-P', 2003 and Raphael Falk's 'instrumental gene' 1984, 1986).

The next question is how PG companies present genes and SNPs to the public. I began my search on the three PG companies' websites, the amount of information contained on which is staggering. One can find slide shows and animations dedicated to explaining the basics of DNA and SNPs, as well as information for physicians on how to incorporate PG testing into their practice. It is also possible to locate the complex sets of algorithms by which risk is calculated, links to GWAS papers, lab and counsellor accreditations, SNP databases, and more. Yet one thing you will not find is a clear, straightforward definition of a gene and its relationship to SNPs. Most companies do provide adequate definitions of SNPs – as variations in single nucleotides – but fail to provide much clarity on what SNPs do, their relationship to genes, and their locations. The best one can do is tease apart the gene-SNP relationship from statements contained in distant parts of the websites.

In the physician information FAQ section of Navigenics' website there is a clue into the gene concept on offer. Under the FAQ 'Is a SNP the same as a gene mutation?' the response is,

> No, not necessarily. A SNP may represent a point mutation (meaning a known disease-causing variant in the genetic code). Individual disease-causing mutations are rare; they occur in less than one percent of the population. However, SNPs used to assess common diseases … are common variants found in more than one percent of the population. SNPs are most often thought of as 'predisposition' genetic markers, not as disease causing genetic markers.[7] (Navigenics 2010)

This is the closest one gets to an explanation of the SNP-gene relationship within the information provided to physicians. Though it is not clear precisely what they take the relationship between SNPs and genes to be, this resembles the SNP-as-surrogate view common in the field. On an information page about 'Genetic Markers' we find a clearer definition:

> SNPs may not actually cause the condition, but we know they are either on or close to genes that increase the risk of that condition and are therefore used as markers of increased predisposition to that condition.[8] (Navigenics 2010)

A short definition on a terminology sheet identifies 'gene' with a molecular gene account.

> Genes are segments of DNA dotted throughout the genome that contain blueprints for various proteins.[9] (Navigenics 2010)

---

[7] http://www.navigenics.com/visitor/for_physicians/physician_faqs/
[8] http://www.navigenics.com/visitor/about_us/our_science/genetic_markers/
[9] http://www.navigenics.com/visitor/genetics_and_health/terminology/

Taken together, this appears to be much the same as the SNP-gene relationship assumed in GWAS. Given the difficulty in obtaining this explanation, however, it is doubtful that the customer would decipher this.[10]

deCODEme is perhaps the least forthcoming about the relationship between SNPs and genes. Though their 'Genetics Explained' guide implicitly identifies genes with both chromosomes and the whole genome, and uses 'gene' as an umbrella term for multiple possible SNPs at a given locus, they also offer an explicit definition of the gene.

> [Genes] are relatively small segments of chromosomes, where the sequence of DNA nucleotides encodes a recipe for making a protein.[11] (deCODEme, 2010)

They also explain what is done when a SNP is identified as a risk-allele.

> Usually, further research is needed to find out whether this SNP or another nearby is the true genetic cause of the characteristic in question.[12] (deCODEme 2010)

Again, this explanation seems to fit with the SNP-gene relationship and molecular gene conception used in PG science.

3.2 SNPS, GENES, AND TRAITS

Given this presentation of the SNP-gene relationship, how is the relationship between SNPs, protein-coding genes, and human traits presented? For the answer to this question, I think we can turn first to 23andMe.

23andMe provides a "Keywords for Genetics" information sheet, which, like the other companies, gives a typical molecular gene account.

> Each kind of protein tool has its own blueprint, or gene, located in the cell's nucleus. Genes can be turned on or off in different cells at different times. (23andMe 2010b)

Further down the sheet, we find the following explanation of SNPs:

> A SNP is a site in the genome where a single DNA 'letter' often differs from person to person. Some (but not all) SNPs appear to be associated with variation in different people's phenotypes. (23andMe 2010b)

On their own these definitions tell us little about the mechanism linking SNPs to traits; but when we examine the diagram by which they are accompanied, it becomes clear how SNPs are meant to fit. The diagram depicts a gene as a section of DNA, a molecular gene. Within that gene falls a SNP, where a Guanine nucleotide has been substituted for a Cytosine. That SNP results in the translation of a different protein from the gene,

---

[10] The difficulty in piecing together information about the science will only be compounded by the extremely high reading level required of PG users. Lachance et al (2010) found that the reading level of PG websites was 15, at least 6 grades above the average reading comprehension of US citizens.

[11] http://www.deCODEme.com/genetic-code

[12] http://www.deCODEme.com/genes-traits-diseases

which resides on the bitter taste region of the tongue. Proteins coded without the SNP are amenable to broccoli, and those coded with the SNP are not. Relying heavily on imagery of DNA and proteins as structure-specific shapes, they depict the protein as either broccoli shaped or not. The result of the SNP is a structurally different protein, which, according to the diagram, is not broccoli shaped so the bitter taste receptors on the tongue are protected from the broccoli. The result is broccoli enjoyment. Explaining this phenomenon, the account continues:

> In this example, a SNP in the gene encoding the protein that responds to bitter flavors can have C [cytosine] or G [guanine] variants – leading to a big difference in phenotype! (23andMe 2010b)

For want of a single guanine, enjoyment of broccoli is lost.

The broccoli example makes clear that 'gene' refers to molecular gene, and that SNPs are taken to reside within those genes. This latter claim is false in many cases since SNPs may occur anywhere on the DNA. It is likely that they chose the broccoli example because it is one of the rare cases where a single SNP can be linked directly to a change in a protein and a change in a trait. SNPs affect genes either directly, by changing the coding sequence, or indirectly, perhaps by affecting a regulatory region. But this latter possibility is not communicated to the customers. Two things are clear: (1) 'Genes' are protein-coding genes. (2) SNPs are important because of their relationship to mutations in those genes; those mutations affect protein synthesis and this in turn is where differences in traits emerge. Simplified information about the location of SNPs is not on its own problematic, but insofar is it is used to communicate inaccurate information about genetic causation, it should be a source of concern. The use of simplified examples is not uncommon. In science communication, where the information to be communicated is often immensely complex, simplifications are a necessity. Yet we must be careful that the simplification does not come at too great a cost. In this instance, it may.

Assertion (2) is not entirely incorrect. There do exist rare cases in which a single nucleotide variation in a molecular gene directly determines phenotype via a change in protein structure. Yet this is not the only mechanism behind trait determination. This depiction represents the etiology of a minority of traits; but the broccoli example sheet does not comment on the question of how much trait variation is caused in this way. Does the direct action of SNPs account for all, most, some, or little variation? Is there any way a SNP would not lead to a phenotypic change? There is nothing in the depiction that would provide an answer to these important questions, let alone prompt a customer to ask them.

Similar characterizations of the SNP-gene-trait relationship can be found on the deCODEme site.

> Small differences in the sequence of DNA nucleotides of a particular gene can lead to differences in the structure and behavior of the proteins they encode. It is these

differences, in turn, that account for the variable characteristics of the people around you.[13] (deCODEme 2010)

Elsewhere on the site, a similar sentiment is expressed more explicitly:

[T]he fact that our DNA can change, given enough time, explains why we are all different in size, shape, color and many other characteristics. Such differences are the result of the many SNPs that have arisen in the DNA of our species and its predecessors.[14] (deCODEme 2010)

Unlike the 23andMe example, these two statements do *not* leave open the question of how much variation is caused in this way. Rather, they seem to imply quite strongly that *all* of your traits are the direct, unmediated result of SNPs. A different gloss appears elsewhere:

Most of our features, internal and external are determined or influenced *in some way* by such variations in our DNA, i.e. by SNPs. Each of us is different because we carry a unique combination of genetic variants, including SNPs.[15] (deCODEme 2010; *emphasis added*)

This final statement is importantly different from the first two and the lack of consistency is noteworthy. This claim leaves the door open for other influences acting in concert with genetic variants. Again, this implies a causal connection from SNP→protein→trait and fails to make clear to what extent this mechanism accounts for trait formation.

With the discussion of these two companies in mind, we should be concerned about three things. First, it is disconcerting that both of these examples imply that traits are the direct result of changes to nucleotide sequence. While this is the case in some instances, many SNPs for which these companies test fall outside of molecular genes. It is also well-established that changes to regulatory and other non-protein-coding regions can have phenotypic effects of similar magnitude to their protein-coding gene counterparts (see eg. Mastron et al 2004). Traits in these instances may not be a result of a change in polypeptide sequence, but in other changes in transcription patterns. That PG may be unable to communicate this to customers who lack an understanding of DNA architecture is perhaps understandable, but worth noting. On the other hand, it is important to note that SNPs are just one of a number of different types of mutations relevant to diseases. Although SNPs are very common, there are larger mutations that may carry larger phenotypic effects. These larger mutations are not covered in PG tests, and their very existence is not discussed. Second, we should be concerned that these characterizations of gene action leave out any mention of other influences. Other factors independent of mutation play large roles in the direction of many traits. In some

---

[13] http://www.deCODEme.com/genetic-code
[14] http://www.deCODEme.com/genetic-variation
[15] http://www.deCODEme.com/genetic-variation

instances the mechanisms within which protein-coding genes operate are determined by external influences, not just the presence of a given nucleotide sequence in the DNA. In the sections of their web pages explaining the SNP-gene-trait relationship, these companies make no mention of non-DNA factors and their role in trait production. In fact, the broccoli example from 23andMe seems to have been selected specifically to avoid discussion of other factors. Finally, some of these selections, particularly the broccoli example and the first deCODEme quote, seem to suggest an inevitability about gene-trait causal links.

This first concern is ultimately one about the extent to which DNA architecture and organization is taught to customers. It may be easier to couch everything in terms of coding regions rather than explain the role of other DNA regions, too. This does not seem to translate into a problem for the scientific practice of PG companies, given that they do test for SNPs falling in non-coding gene regions, even if this is not communicated to customers. The failure to mention the importance of mutation other than SNPs is worth considering. To the extent that customers hope to gain an understanding of biology, epidemiology, and genetic causation, this presents a problem, one to which I will return in the final section of this paper. This problem is compounded by the second and third concerns, which have more radical implications for the way in which genetic causation is understood by customers. In the next section, I examine a different message contained in PG information, albeit one originating from very different parts of the webpages. In these sections, factors other than DNA are discussed in a manner quite contradictory to the message just examined.

3.3 SNPS, GENES, AND TRAITS … IS THAT ALL?

From the investigations reported in the previous section, one might expect that PG companies make no mention of non-DNA factors at all. When explaining the science on which they rely, they seem to imply quite strongly that SNPs in protein-coding DNA are the only important causal factor behind traits. Yet the reports customers receive from PG companies relaying test results are strikingly different. Consider the following:

> Environmental factors can also affect our visible traits, either negatively or positively (deCODEme 2010)

A similar quote can be found on every risk assessment page of a 23andMe report:

> Environmental contributions to a trait may be small (earwax type) or large (type 2 diabetes). Environmental factors range from things that are more or less under one's control, such as diet and obesity, to things outside one's control, such as childhood influences or geographic location. (23andMe 2010)

Statements like these can be found as footnotes to risk assessments, or under FAQ sections about why genes cannot be used for diagnosis or whether genes are one's destiny. In most cases they are not easy to come across, often requiring linking off main

11

pages or finding hyperlinks embedded within risk assessments. That these statements are difficult to find while appearing at odds with the information about SNPs, genes, and traits reviewed above should not go without notice. A customer who happened to find one group of information but not the other might be misled, and customers who find both will likely be confused. Given this juxtaposition, combined with the difficulty in finding statements in which the role of environment is emphasized, it is perhaps unsurprising that when we dig a little deeper we find that this emphasis on other factors is rather hollow.

When viewing the risk-assessments from several PG companies, one is now able to review a list of known environmental risk factors affecting a given trait. Often these are quite general, nearly always including smoking, age, ethnicity, gender, and weight. Little indication is given as to the relative importance of any of these factors, or how a customer might go about determining the extent to which one of these factors affects her. Customers cannot turn to original research for this information, as sources are rarely provided. 23andMe does have a link labelled 'sources' at the end of the list, but these are actually monozygotic twin studies used to calculate heritability figures, not sources for the risk factors themselves. One could certainly forgive a customer for failing to notice that these studies were not, in fact, source research for the environmental risk factors. deCODEme provides sources, but only a minority of the time. In most cases the customer is left with little information about the nature of any additional risks conferred.

This trend is repeated throughout PG websites and reports. The contribution of environmental influences is acknowledged explicitly, but is not utilized in risk-assessments and is not presented in detail sufficient for the customer to understand the environmental impact herself.

23andMe's reports each contain a heritability report, presented as a graph representing the genetic versus environmental components of a trait; or, "how much genetics contributes to a trait or the risk of a condition" (23andMe 2010). The chart for heart attack, for instance, reads '38-57% attributable to genetics'. These figures, the result of monozygotic twin studies, are problematic for a number of reasons. First, these studies assess the traits of identical twins to determine how frequently traits are attributable to their shared genetics, and how often to their (presumably) different environments. This methodology has been subject to a great deal of scrutiny and should be viewed with a touch of scepticism (see Hawkes 1997). Second, it is unclear that these studies provide an estimate of the 'genetic versus environmental' components of a trait. It is more likely that the 'genetic' components measured actually indicate heritability, which may be reflective of modes of transmission other than that captured in nucleotide sequence (Sesardic 2005; Fox Keller 2010). Finally, the customer might think that these 'genetic v. environmental' figures had in some way been taken into account during the

calculation of her risk assessment. Indeed, there is no clear indication to the contrary. Yet environmental risk factors are not factored into the risk assessment.

One further case deserves attention. Each of the three PG companies emphasizes the role of family history in determining susceptibility to diseases. Navigenics has a large section of their webpage explaining that family history provides a great tool to which PG scans can greatly contribute.[16] 23andMe and deCODEme offer ancestry analyses, the predictive virtues of which each company extols. On Thanksgiving (U.S.), which also happens to be 'National Family History Day' (U.S.), Navigenics issued a press release, addressed as an open letter to the surgeon general. The letter celebrated the value of collecting family history health information.

> At a time when families are gathered, we are reminded that our family's health history offers a window into our own health, providing insight on some of the heritable factors that may predispose us to particular health conditions. Family history – particularly gathered through the online platform – is an easy, accessible means by which we all can take greater control over our health. A seemingly small step, family history could provide us with a foundation on which we can collectively build a new era of disease prevention. (Navigenics 2010, November 23)

This information is all correct. The internet *is* an excellent means by which to collect family history information, and family history *is* a powerful tool for disease prediction. The latter is precisely why standard clinical practice for assessing many traits and conditions begins with a simple family history; for generations this was the only such metric available. Today, in the midst of genetic testing, family history remains a powerful tool for predicting risk for many diseases from diabetes to various cancers. There are well-established models of familial risk associations based on heritability patterns that can be used by clinicians and patients.

The utility of family history is no accident. Some of the reasons behind the predictive power of family history are now understood. Of course one reason is that disease-causing nucleotide variants are heritable. But there are other reasons, too. Many epigenetic disease risk factors, such as methylation patterns, are inherited in predictable familial patterns, as are behaviours, traditions, diet, and habitat. All of these may confer risks for different diseases and traits. Importantly, much of the risk information conferred by familial assessments is not captured in a SNP-based genetic test. That these three companies so publically endorse the value of knowing one's family history is laudable. What is surprising, however, is that none of these PG companies incorporates family history data into the calculation of risk assessments. Nor are any mechanisms provided for customers to make these assessments themselves.[17]

---

[16] http://www.navigenics.com/visitor/genetics_and_health/family_history/

[17] Some recent work has shown just how fruitful (though difficult) it can be to include clinical metrics, like family history, into risk assessments relying on DNA. See (Ashley et al., 2010).

PG repeatedly acknowledges the importance of variables for trait prediction, yet does nothing to incorporate those variables into risk-assessments. An interesting question is whether customers, steeped in PG literature about family history and environment, are aware that no such information impacts the predictions they receive.

3.4 WHAT MESSAGE DO CUSTOMERS GET?

It is not clear what the take-home message for PG customers is intended to be. There are two distinct versions of genetic causation presented. In each case, 'genes' are quite clearly molecular genes. Yet the etiological structure within which those genes operate receives two very different presentations. The discrepancy lies in the level of autonomy given to the gene in genetic causation. In the first version, coming from information about SNPs, genes, and traits, genes are autonomous determiners of traits. This view, which appears more frequently on web pages and in informational literature, videos, and tutorials, emphasises the role of SNPs in determining traits by generating variation in proteins. The second version, arising from clauses in risk-assessments about the role of environment and from public praise of family history, features a less-autonomous gene. On this view, genes are important, but other variables are important, too. This take on genetic causation closely resembles what philosopher Paul Griffiths (2006) calls the 'informational gene', according to which molecular genes contain information about a trait, which is either permitted or denied by latent features of the environment. These two views are related insofar as the informational gene is a weak form of determinism. Though certainly not without problems (see Griffiths' 2006 discussion), the informational gene is far better represented in scientists' thinking about genes. The deterministic approach, on the other hand, has very few, if any, proponents within the academy. While it is unsurprising that the deterministic gene is absent in PG science, it should come as some surprise that it is present in PG rhetoric.

The difficulty inherent in reconciling these two presentations of genetic causation is striking. Are protein-coding DNA autonomous causal agents, or are they working with the environment? The answer cannot be 'both'. It is impossible to fit environmental influences into the picture of SNP-driven genetic causation outlined in the broccoli diagram. Perhaps the motive behind these two very different characterizations is that each appeals to a different view of genetic causation extant in the population. While many people view genes as discrete sections of DNA acting autonomously to effect phenotype, there are likely others who understand that the environment plays a role as well, opting for a more holistic view of genes as the system of DNA, environment, and development. Persons of either persuasion could find something to latch onto within the pages of PG sites.

It is certainly worth comment that the locations of these two views suggest a possible rhetorical strategy. A customer's first encounter with PG, perhaps while

browsing the sites, reading the educational literature, trying to decide whether or not the $300 price tag is worth it, she will encounter the very deterministic presentation. It is not difficult to see how PG companies would benefit from this view. The more causal power attributed to DNA, the better the predictions forwarded using DNA. If the broccoli diagram is taken at face value and traits are thought to be the mechanistic result of SNP-induced changes in protein structure, it is hard to see how a customer would fit chance or under-determination into the picture. The customer's perception of the predictive power of PG tests would increase drastically. In fact, believing herself to be in possession of knowledge of the precise mechanism behind genetic causation, she may approach her decision to purchase a test with a great deal more confidence. It is only after she has purchased a test, of course, that she begins to receive the second story about genetic causation. With results in hand and decisions to be made, she is now cautioned that genes are not her destiny; the environment plays a role, too!

This is difficult to explain. Why present this latter view, which after all is more representative of the stance of the scientific community, if it is not reflected in the science? Literature on lay attitudes toward determinism may be informative, here. As Celeste Condit (2009) reports, members of the public have been found hold multiple beliefs about genetic causation, some deterministic, others more holistic. These different sets of beliefs, or different 'mental models', are triggered differentially by different contexts and goals. It is known, for instance, that members of the public will be highly deterministic when asked about the origins of disease, but far less deterministic when asked about the origins of behaviour (Condit 2011). Similarly, individuals with the goal of maintain health optimism may appeal to environment, while those maintaining fatalistic beliefs appeal to genetic determinism. Very similar results were found by a group of researchers studying patient interpretations of PG results (O'Neill et al. 2010). What is puzzling is that people maintain and use these two models of genetic causation entirely separately. That is, they hold *incompatible* views of determinism and gene-environment interactionism, with no regard to the manifest inconsistency. In any single context, beliefs appear consistent because only one mental model is used. It is only on a protracted view that the inconsistency comes out. Individuals appear to have taken in messages about the role of environment in genetic causation, but ignore that knowledge when it is inconvenient.

These two models of genetic causation roughly accord with the two views of the gene seen here. With two views of the gene on offer, subjects can pick whichever one suits the context and their goals. Providing support for both views of causation allows customers to cherry-pick the account they require. The way in which PG facilitates this psychological phenomenon is unfortunate. PG provides customers with putative justification for the deterministic model. The broccoli diagram and similar materials provide what appears to be a molecular mechanism for genetic determinism, allowing customers to continue to rely on the determinist model, believing themselves to be in

15

possession of scientific knowledge in its support. Yet the determinist model is scientifically untenable. It is therefore likely to lead to poor decisions and further false beliefs. We should hope for a public that does not posses beliefs about determinism as a legitimate way of thinking about genetics. The explanations provided by PG are a step in the wrong direction.

4.0 Personalized Genomics and Academic Science

It should come as no surprise that the risk associations utilized in PG are unacceptably low. As we have seen, the calculation of risk is grounded in a narrow conception of genetic causation that leaves no room for interactions between DNA, environment, and development. When predictions are being provided for traits like diabetes and various cancers, for which the environment plays so crucial a role, surely the recognition of systems of interacting causal factors is a more appropriate approach than one grounded solely in the DNA. Indeed, PG companies themselves extol the virtues of taking such an approach!

If the emphasis on DNA and neglect of additional variables is a potential cause of PG's principal scientific shortfall, then why is the importance of environmental information frequently acknowledged, but never utilized? One possibility is that it is simply too difficult to obtain data about customers' environment; but I think there are good reasons to partially discharge this excuse.

Some environmental information is easily accessible to PG companies, yet is not obtained. There is a wealth of environmental information, such as smoking behaviour, level of exercise, ethnicity, family history, and geographic location that we know is relevant to the prediction of many traits. All of this could be obtained with a simple questionnaire. In fact, customers have to fill out questionnaires about sex, name, birth date, etc. when they register for their PG results. It would not be difficult to add a few more questions. It might even be possible to get customers to input more direct indices of health such as geographic location, blood pressure, cholesterol, insulin levels, average heart rate, etc. These are indices to which many health-conscious persons have access.

Perhaps the critic might respond that customers would never take the time to provide such data. Thanks to an unlikely source, there is now strong evidence that this objection is incorrect. A recent initiative by 23andMe demonstrates that customers are more than willing to answer *a lot* of additional questions about their health, habits, traits, and wellbeing, if they think it may improve the accuracy or quality of their tests (Eriksson et al. 2010). Unfortunately, the questions 23andMe has chosen to ask are largely trivial (eg. Are you naturally a night person or a morning person? Does raw broccoli taste bitter to you?). Collecting data via survey is an easy yet neglected way to obtain useful environmental variables.

There is also some important information stored in the DNA sample itself, which needs only to be extracted. The saliva samples and buccal swabs used to obtain customers' DNA contain not only the DNA, but also information about the health and nutrition of the individual.

Environmental data is certainly available to be harvested; whether it can be easily utilized is another question. A second possible explanation for the failure to include environmental data is thus that, though available, it is difficult to use. Interpreting environmental data requires established trait-variable correlations. Correlations between SNPs and traits are known because of GWAS, but no endeavour of comparable scope or magnitude exists for establishing correlations between environmental variables and traits.[18] This issue is beyond the scope of PG, it concerns the scientific community itself. We simply have not seen variable-trait correlation studies on the magnitude that GWAS have taken place. Why not? It may be that funding for this type of research is limited. It may be that fewer researchers are interested in this type of research. It may be because environmental variables are less tractable than SNPs. It may be that this research is viewed as secondary to SNP-trait correlations. Whatever the answer, this points to a curious gap in mainstream biological sciences and, by extension, an unexpected way in which commercialized science may rectify that problem.

The fact that trait-variable correlations have not been established does not mean that it is not possible. In fact, PG companies are in an excellent position to do just that. They have access to very large customer populations and those customers have a demonstrated willingness to provide large quantities of personal information (Eriksson et al. 2010). This overcomes the main obstacle to any large variable-trait correlation study, which is the recruitment of willing subjects. PG companies are also in the unique position of being able to cross-reference environmental data with genetic data. The volume of variable-gene-trait correlations that could be established is great. Unfortunately, no such work has yet been initiated.[19]

If it is not the case that environmental information is difficult to obtain, and if PG companies have the power to create the infrastructure to utilize that information, then why doesn't this happen? This is a difficult question to answer. I suggest two possible explanations. On the one hand, it may be that PG companies truly believe in the predictive power of DNA. Perhaps PG, like many researchers (see Hunter, Altshuler, et al. 2008), are holding out for larger pools of GWAS data, or better mechanisms for

---

[18] In the wake of criticism of GWAS, some EWAS (Environmental-wide association studies) have been suggested. Yet these have not been funded or pursued with anything like the enthusiasm of GWAS. See (Pate, 2010; Turkheimer 2012)

[19] This may be set to change, a new company, Personalis, promises to integrate whole-genome scans with clinical risk information (family history, behaviour, etc.). It will be interesting to watch whether, if successful, Personalis prompts other PG companies to follow suit, as I predict.

translating GWAS research into clinical practice, which will yield stronger SNP-trait correlations. Yet it is difficult to reconcile such a belief with their public emphasis on environmental variables, and it is hard to imagine that they see no benefit in including such data.

I think a more plausible explanation, one that highlights the importance of examining commercialized science as distinct from mainstream science, is that these companies have no prerogative or precedent to take environmental data into account. Increases in the predictive power of their assessments are not being demanded by the market and it would run contrary to PG's interests qua business to implement any such changes. The cost of this, balanced against prospective returns, is simply too high at present for a for-profit enterprise to bear. PG companies are still attracting flocks of customers, they are still receiving attention in the press, and they are still raising impressive amounts of capital from investors (see 23andMe 2011, Jan. 7; Navigenics 2010, Feb. 2). PG companies are for-profit enterprises, not academic researchers. Their goal is not necessarily to determine the best metric by which to predict traits; it is not to provide the best results to customers; it is not to operate in a way that is consistent with the concepts, laws, and maxims of academic biology. Their goal it is to earn profit and expand the market. If profit and predictions overlap, all the better, but the former must come first.

This is not to say that the incorporation of environmental factors is not on the distant horizon. Now that more and more people are receiving scans, PG needs to find ways to keep those customers coming back while at the same time attracting new clients. If it became apparent that one company, by including environmental data in their screens, forwarded predictions with markedly stronger risk-associations, and perhaps with a better track record of identifying genuine health concerns (though this latter possibility would be difficult to demonstrate) surely that company would see an increase in sales. Profit and better predictions would overlap. Other companies would not be far behind. Yet switching business models in this way is not yet mandated by the market. Consumers do not seem to care that the risk-assessments they receive are low. They may not even notice. Until the market demands it, I do not think we should expect a radical upswing in the amount of environmental data utilized.

PG appears to benefit from the gene concepts presented. The longer customers are able to retain deterministic ideas, the longer they will seek these tests in the face of evidence of their poor utility. Until potential customers become aware that the predictions they receive could be improved by the addition of more variables, there is no incentive for PG companies to do so. The interesting flip-side of this, however, is that if this *does* happen, then PG companies could help effect a conceptual change within the public perception of genes, genetics, and traits.

5.0 Conclusion

Commercialized personalized genetic testing is in a unique position at the intersection of rhetoric, commerce, scientific practice, and scientific progress. These dynamics shape the conceptual and scientific landscape. Risk associations are weak because they fail to take into account a sufficient breadth of information. The information required for a better approach is available, but neglected. It is neglected because there exist no mechanisms to interpret and utilize it. There exist no such mechanisms in part because the manifest attitude toward genetic and epidemiological research focuses principally on DNA and ignores other factors. Commercialized genetics is in a unique situation to rectify this problem. Though capable, PG companies do not utilize these mechanisms because there exists no market demand for the inclusion of other variables.

A brief look through the highly-active online community of, eg., 23andMe, reveals that PG has likely succeeded in raising customers' interest in their health. This analysis suggests that the next step is to ensure that the information they provide is maximally informative. When it comes to informing customers about genes and faithfully representing their scientific practices, PG companies come up short. It should hardly take a detailed reconstruction of the information they provide to extract the gene concepts presented. Customers hoping to learn about genes, traits, and molecular biology will likely leave their interactions with PG either confused or misinformed. This confusion may well explain customers' misunderstanding of PG results. This leaves a lingering and important question about the influence of PG on the public understanding of science: what do customers think genes are, and what do they think PG companies are using to generate their test results? Perhaps therein lies important empirical work.

BIBLIOGRAPHY

23andMe.: 2010, http://www.23andme.com Accessed December 9, 2010.

23andMe.: 2010b, *Keywords for Genetics*. Retrieved from:
     https://www.23andme.com/gen101/graphics/genetics/ [March 1, 2011]

23andMe.: 2011, January 7, *23andMe Announces Additional Investment in Series C Financing* [press release]. Retrieved from: https://www.23andme.com/about/press/20110107/

Ashley E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., Dudley, J. T., Ormond, K. E., Pavlovic, A., Morgan, A. A., Pushkarev, D., Neff, N. F., Hudgins, L., Gong, L., Hodges, L. M., Berlin, D. S., Thorn, C. F., Katrin, S., Hebert, J. M., Woon, M., Sagreiya, H., Whaley, R., Knowles, J. W., Chou, M. F., Thakuria, J. V., Rosenbaum, A. M., Zaranek, A. W., Church, G. M., Greely, H. T., Quake, S. R., Altman, R. B.: 2010, 'Clinical Assessment Incorporating a Personal Genome', *The Lancet* **375**(9725), 1525 - 1535, doi:10.1016/S0140-6736(10)60452-7.

Bates, B. R.: 2005. 'Public culture and public understanding of genetics: A focus group study' *Public Understanding of Science* **14**(47), 47–65

Burian, R.M.: 2004, 'Molecular Epigenesis, Molecular Pleiotropy, and Molecular Gene Definitions', *History and Philosophy of the Life Sciences* **26**, 59-80.

Cohen.: 2008, 'Volleyball and the Public Perception of Genetics', http://scienceblogs.com/worldsfair/2008/08/beach_volleyball_and_the_publi.php

Condit, C. M., Gronnvoll, M., Landau, J., Shen, L., Wright, L., & Harris, T. M.: 2009, 'Believing in Both Genetic Determinism and Behavioral Action: A Materialist Framework and Implications', *Public Understanding of Science* **18**(6), 730-746.

Condit, C. M.: 2011, 'When do People Deploy Genetic Determinism? A Review Pointing to the Need for Multi-Factorial Theories of Public Utilization of Scientific Discourses', *Sociology Compass* **5**(7), 618-635.

deCODEme.: 2010, http://www.decodeme.com/ [accessed Dec. 9, 2010)]

Eriksson N., Macpherson J. M., Tung J. Y., Hon L. S., Naughton B., et al.: 2010, 'Web-based, Participant-driven Studies Yield Novel Genetic Associations for Common Traits', *PLoS Genetics* **6**(6): e1000993.

Fogel, T.: 2000, 'The Dissolution of Protein Coding Genes in Molecular Biology'. In Beurton, P., Falk, R., & Rheinberger, H. J. (eds.) *The Concept of the Gene in Development and Evolution*, Cambridge University Press, Cambridge UK, pp.3-25.

Fox Keller, E.: 2000, *The Century of the Gene*, Harvard University Press, Cambridge MA.

Fox Keller, E.: 2010, *The Mirage of a Space Between Nature and Nurture*, Duke University Press, NC.

Gibson, G., Goldstein, D. B.: 2007, 'Human Genetics: The Hidden Text Behind Genome-wide Associations', *Current Biology* **17**(21), 929-932.

Grey, R. D.: 1992, 'The Gene is Dead'. In. Griffiths, P. E. (ed.) *Trees of Life*. Kluwer, Dordrecht, pp.165-210.

Griffiths, P. E.: 2006, 'The Fearless Vampire Conservator: Philip Kitcher, Genetic Determinism and the Informational Gene'. In Rehmann-Sutter, C., & Neumann-Held, E. M. (eds), *Genes in development: Rethinking the molecular paradigm.* Duke University Press, NC, pp.175-198.

Griffiths, P. E. & Stotz, K.: 2007, 'Gene'. In Michael Ruse and David Hull (eds.), Cambridge Companion to Philosophy of Biology, Cambridge University Press, pp.85-102.

Hawkes, C. H.: 1997, 'Twin Studies in Medicine – What Do They Tell Us?' *Quarterly Journal of Medicine* **90**(5), 311-321.

Hindorff, L. A., Praveen, S., Heather, A. J., Erin, M. R., Jayashri, P. M., Francis, S. C., Manolio, T. A.: 2009, 'Potential Etiologic and Functional Implications of Genome-wide Association Loci for Human Diseases and Traits', *Proceedings of the National Academy of Sciences of the United States* **106**(23), 9362-9367.

Hudson, K.: 2008, 'The Health Benefits of Genomics: Out with the Old, in with the New', *Health Affairs* **27**(6), 1612-15.

Hunter, D. J., Altshuler, D., & Rader, D. J.: 2008, 'From Darwin's Finches to Canaries in the Coal Mine – Mining the Genome for New Biology', *New England Journal of Medicine* **358**(26), 2760-2763.

Hunter, D. J., Khoury, M., & Drazen, J. M.: 2008, 'Letting the Genome Out of the Bottle – Will we Get our Wish?' *The New England Journal of Medicine* **358**(2), 105-107.

Hunter, D. J. & Chanock, S. J.: 2010, 'Genome Wide Association Studies and 'The art of the Soluble'', *Journal of the National Cancer Institute* **102**(12), 1-2.

Ioannidis, J. P. A.: 2009, 'Personalized Genetic Prediction: Too Limited, too Expensive, or too Soon?', *Annals of Internal Medicine* **150**(2), 139-141.

Kaufman, D.J., Bollinger J.M., Dvoskin, R. L., & Scott, J. A. (2012). Risky business: Risk perception and the use of medical services among customers of DTC personal genetic testing. Journal of Genetic Counseling, 21(3), 413–422.

Kua, E., Reder, M., & Grossel, M. J.: 2004, 'Science in the News: A Study of Reporting Genomics', *Public Understanding of Science* **13**, 309-322.

Kruglyak, L.: 2008, 'The Road to Genome-wide Association Studies', *Nature Genetics* **9**, 314-318.

Lakhman, K,: 2010, May 25, 'First Walgreens, Now House Calls: The Increasingly Bizarre Predicament of DTC Genetic Testing', *The Sample*. Retrieved from: http://www.genomeweb.com/blog/first-walgreens-now-house-calls-increasingly-bizarre-predicament-dtc-genetic-tes

Lachance, C. R., Erby, L. A. H., Ford, B. M., Allen, V. C., & Kaphingst, K. A.: 2010, 'Informational Content, Literacy Demands, and Usability of Websites Offering Health-related Genetic Tests Directly to Consumers', *Genetics Medicine* **12**(5), 304-312.

Leighton, J. W., Valverde, K., Bernhardt, B. A.: 2012, 'The general public's understanding and perception' *Public Health Genomics* **15**

Maher, B.: 2008, 'The Case of the Missing Heritability', *Nature* **456**, 18-21. doi:10.1038/456018a

Maston, G. A., Evans, S. K., & Green, M. R.: 2006, 'Transcriptional Regulatory Elements in the Human Genome', *Annual Review of Genomics and Human Genetics* **7**, 29-59.

McGuire, A. L., & Burke, W.: 2008, 'An Unwelcome Side Effect of Direct-to-consumer Personal Genome Testing: Raiding the Medical Commons', *The Journal of the American Medical Association* **300**(22), 2669–2671.

McGuire, A. L. Diaz, C., M., Wang, T., & Hilsenbeck, S. G.: 2009, 'Social Networkers Attitudes' Toward Direct-to-consumer Personal Genome Testing', *American Journal of Bioethics* **9**(6-7), 3-10.

Mountain, J., Hsu, A., Macpherson, M., Naughton, B.: 2007, '*White paper 23-02: Estimating genotype-specific incidence in the context of ethnic variation*', [White paper] Retrieved from: https://23andme.https.internapcdn.net/res/6295/pdf/23-02_Incidence_And_Ethnicity.pdf

Navigenics.: 2010, http://www.navigenics.com/

Navigenics.: 2010, November 23, '*An open letter to the surgeon general*', [press release] Retrieved from: http://www.navigenics.com/visitor/about_us/press/releases/surgeon_general_1 12310/

Navigenics.: 2010, February 2, '*Navigenics raises $18 million in series C financing',* [press release] Retrieved from: http://www.navigenics.com/visitor/about_us/press/releases/financing_020310/

Neumann-Held, E.: 1999, 'The Gene is Dead – Long Live the Gene! Conceptualizing Genes the Constructionist Way'. In Koslowski, P. (ed.) *Sociobiology and Bioeconomics: The Theory of Evolution in Biological and Economic Theory*. Springer, NY, pp.105-137.

O'Neill, S. C., McBride, C. M., Alford, S. H., Kaphingst, K. A.: 2010, 'Preferences for Genetic and Behavioral Health Information: The Impact of Risk Factors and Disease Attributions', *Annals of Behavioral Medicine* **40**(2), 127-37.

Patel C. J., Bhattacharya J., Butte A. J.: 2010, 'An Environment-wide Association Study (EWAS) on Type 2 Diabetes Mellitus', *PLoS ONE* **5**(5): e10746. doi:10.1371/journal.pone.0010746

Sesardic, N.: 2005, *Making Sense of Heritability*. Cambridge: Cambridge University Press.

Turkheimer, E.: 2012, 'Genome Wide Association Studies of Behavior are Social Science'. In K. S. Plaisance & T.A.C. Reydon (Eds.) *Philosophy of Behavioral Biology*, Springer, NY, pp.43-64.

Waters, K.: 1994, 'Genes Made Molecular', *Philosophy of Science* **61**, 163-185.

Waters, K.: 2004, 'What Concept Analysis in Philosophy of Science Should be (and why competing philosophical analyses cannot be tested by polling scientists)', *History and Philosophy of the Life Sciences* **26**(1), 29-58.