

Defining genetic disease

Catherine Dekeuwer

► **To cite this version:**

Catherine Dekeuwer. Defining genetic disease. Philippe Huneman; Gérard Lambert; Marc Silberstein. Classification, disease and evidence : new essays in the philosophy of medicine, 7, Springer, pp.147-164, 2015, History, philosophy and theory of the life sciences, 978-94-017-8886-1. 10.1007/978-94-017-8887-8_7. hal-01791752

HAL Id: hal-01791752

<https://hal-univ-lyon3.archives-ouvertes.fr/hal-01791752>

Submitted on 15 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Catherine Dekeuwer

« Defining Genetic disease », in *Classification, Disease and Evidence. New essays in the philosophy of medicine* (P. Huneman, G. Lambert, M. Silberstein M eds.), Springer, 2015 : 147-164

Abstract The concept of genetic disease refers to the idea that one or more genes are the cause of disease. Under this definition, problems arise when it comes to the use of the term “cause”. Moreover, genes alone cannot explain the development of a disease; environmental causes are also at play. So when is giving primary importance to genetic causation justified? In his paper “The Concept of Genetic Disease” (2004), David Magnus differentiates three competing concepts of genetic disease related to three approaches of causality. He concludes that none is really acceptable. Finding insufficiencies in each approach, he ultimately adopts a definition of genetic disease that arises from medical uses. In this contribution, I will specify the three conceptions of genetic causality and defend the idea that they function together in the definition of a disease as a genetic disease.

I wish to contribute to the philosophical reflection on the definition of genetic disease (Hull 1979 ; Hesslow 1984 ; Sterelny and Kitcher 1988 ; Gifford 1990 ; Gannett 1999 ; Sober 2001 ; Kitcher 2003 ; Magnus 2004) by showing that three criteria are necessary, together, in order to consider a disease to be genetic. Before entering into the subject, however, I would like to point out two things. First, the concept of disease itself is difficult to define. Does this difficulty prevent, however, any attempt at defining certain diseases as genetic? On the contrary, I think that the analyses presented here could contribute to the general discussion of what disease is. Second, there is the problem of figuring out where to begin when it comes to defining genetic diseases. The concept of genetic disease is based on the idea that one or several genes cause it; what is at issue is how we conceive these “causes”. Furthermore, the development of a pathology cannot be attributed to genes alone. The problem then is to know why, for some diseases, researchers assign greater importance to genetic causes than to environmental ones. Does finding a solution to this problem require a critical re-evaluation of the ways the term “genetic disease” is currently used? Or does it require instead a conceptual analysis of genetic causality that would clearly delineate the line between genetic and non-genetic diseases? And how does the history of medicine and biology factor into this research? It seems to me that the uses of the expression “genetic disease” evoke both a series of problems that doctors seek to resolve, as well as the concepts they use to achieve such solutions. Understanding the notion of genetic disease therefore demands a twofold understanding of the term’s concept as well as its history. One of David Magnus’s articles entitled “The Concept of Genetic Disease” (Magnus 2004) serves as the basic frame of reference for my presentation of the three approaches to genetic causality. Magnus distinguishes three competing concepts of genetic disease, which correspond to three approaches to causality. According to Magnus, none of them is truly satisfactory, which leads him to consider the uses of the concept in order to understand why genetic causality is often privileged in studies of certain diseases. However, whereas Magnus critiques the insufficiencies of each approach and reduces the meaning of genetic disease to its medical applications, I am claiming here that the idea of these three conceptions of causality, when rendered more precisely, function together within the definition of certain diseases as genetic disease.

Problems Involving Genetic Causality

Magnus distinguishes three approaches to genetic causality, each with its own set of issues: it is thus impossible, according to Magnus, to settle on one definition of genetic disease in the absence of any firm epistemological foundation. In the first part of this chapter, I would like to review these three approaches and discuss their critiques. First, a disease is genetic if it results from the direct causal action of one or more genes. In order to explain the concept of direct causal action, Magnus uses an article by Fred Gifford (1990, p. 329) in which he argues that in order to be considered genetic, “the trait must be the specific effect of some genetic cause, that the trait must be described or individuated in such a way that it is properly matched to what the gene causes specifically”. Magnus does not even take the time to refine his critique of this concept of causality: for him, the development of a disease is so complex that it is useless to try to find its direct genetic causes. However, on this point he confuses direct causality with specific causality. Gifford does not mention direct causality; he uses the concept of specific effect defined as a correspondence established by researchers between, on one hand, what the gene specifically causes, and on the other, the phenotype as it is described and individuated. The effect is specific if the modification of a gene has some effect on the considered trait but not on other traits. The phenotype must be individuated (i.e. correspond to one precise unit of description), neither too broad nor too narrow. Gifford first takes the example of language acquisition. If genes do have an effect on aptitude for learning a language, the ability to learn French is not genetic: the trait is too specifically individuated. Another example: hypercholesterolemia cannot be considered a genetic disease. It is necessary to distinguish familial and sporadic forms of this disease, by individuating the reported phenotype from a genetic cause. This perspective has the advantage of directing our attention to the need to clearly identify and delineate phenotypes for which genetic causes may be sought. The absence of a clear description and individuation for schizophrenia, for instance, has made research on its genetic determinants quite difficult (Maziade et al. 2003). The concept of specific effect raises questions related to molecular pathology. Gifford explains that it would be better to give less importance to the question of knowing if a trait is genetic in order to focus instead on “what are the steps involved in the biosynthetic or developmental pathway?” (Gifford 1990, p. 328). And yet, “Molecular pathology seeks to explain why a given genetic change should result in a particular clinical phenotype (f). Molecular pathology requires us to work out the effect of a mutation on the quantity or function of the gene product, and to explain why the change is or is not pathogenic for any particular cell, tissue or stage of development” (Strachan and Read 2004, p. 418). Molecular pathology aims to answer this type of question: “Why should loss function of the FMR1 protein, involved in transporting RNA from nucleus to cytoplasm, cause mental retardation and macro-orchidism (Fragile-X syndrome)?” (Ibid., p. 416.) In this perspective, a gene would be a cause of a disease in the sense that the identification of a DNA sequence would allow different steps in the process that leads to an individual becoming ill to be specifically explained. This approach to causality (once it is properly understood) relies then on genes’ specific influence on the ways in which organisms develop. It falls under the category of an explicative approach as it relates to the biological individual where a causal history of specific effects produced by one or several genes is retraced. The second relevant approach to causality when it comes to the definition of genetic disease relates to populations rather than individuals. A disease is genetic if “in that population, the covariance of the trait with some genetic factor(s) is greater than the covariance of the trait with other (non-genetic) factors” (Magnus 2004, p. 235). Magnus again uses one of Gifford’s criteria: “trait is genetic (with respect to population P) if it is genetic factors which ‘make the difference’ between those individuals with the trait and the rest of

population P" (Gifford 1990 , p. 333). This concept of cause is statistical and only has meaning relative to populations. Analysis of variance is effectively used to measure the causal contribution of genetic and environmental factors in a given population. According to Magnus, this approach is not convincing either, precisely because it is population-relative: as a function of the studied population, a factor can be considered genetic or not. Identification of a genetic factor of a disease thus depends on the population selected for the study, which makes the concept of genetic disease too relative. Magnus makes a close study of one proposed by Hesslow (1984) and takes the example of a village well contaminated by a pathogen. Only half the villagers fall ill and the researchers thus assume that their genes confer resistance to the pathogen. On the scale of the village, the covariance of genetic factors of the disease is close to 1, whereas the covariance of the disease and the pathogen is close to 0.5. The disease is, relative to the village's population, genetic. On the planetary scale however, the disease would probably be environmental. For Magnus, this type of statistical approach to causality is not a good one for defining certain diseases as genetic. I would like to add here that evidence of a risk factor is precisely not the identification of a cause. It is helpful to delve deeper into this critique and distinguish several statistical approaches to genetic causes. Magnus refers to a disease to which the villagers would be more or less susceptible. These are association studies that allow for the identification of the genetic terrain's determinants, but not those of genetic diseases. The notion of the terrain is relevant when the goal is to find out the genetic determinants that explain certain individuals' resistance to diseases with an environmental etiology, such as a virus. Association studies are thus adaptive: they aim to correlate a difference in allelic frequency for the same locus among non-related subjects who are or are not affected by the pathology. An allele is "associated" with a disease if it is present more frequently among the ill than among the unaffected. It confers an increased risk of disease compared with the general population, expressed by the relative risk. The choice of reference populations is thus quite crucial, but this is not, in my mind, the main problem with these association studies. Finding out why certain individuals are, for example, more susceptible to viral diseases or cancers than others does indicate some susceptibility factors, but it would be quite difficult to justify the claim that these diseases are genetic. It would be difficult to defend the claim that AIDS is a genetic disease even if there are reported cases of resistance to HIV and even if we could identify genetic factors associated with resistance to HIV. This is where the confusion lies in the field of genetics and genetic disease. Association studies are especially relevant for the study of multifactorial diseases with complex heredity such as asthma, diabetes, or certain forms of Alzheimer's¹. For these diseases, a familial aggregation is often observed, but the mode of transmission often remains unknown. Their study requires a method for researching these diseases' genetic determinants without knowing their mode of transmission, and the association study responds to this requirement. In 1993, such association studies demonstrated a strong statistical correlation between a polymorphism of the APOE gene (apolipoprotein E), named e4, and the most frequent forms of Alzheimer's disease, which are non-Mendelian forms. This correlation between the e4 allele of the APOE gene and the sporadic form of Alzheimer's disease has since been confirmed by independent studies; it at least holds true for populations that descend from common European ancestors. It has been calculated that Alzheimer's disease is two to three times more frequent in the group of heterozygote individuals that carry an e4 allele than it is in the general population. The disease is 9–15 times more frequent for homozygotes, which carry this allele in two copies. The APOE gene is thus referred to as a gene of susceptibility for sporadic forms of Alzheimer's: more than half of people with the disease carry the e4 allele, but many who have this allele will never develop the disease. Moreover, the absence of this allele in a person does not mean that he or she will not develop Alzheimer's. Yet

¹ These multifactorial diseases are often also multigenic diseases: many genes work together in the development of the pathology. Campion (2001), Strachan and Read (2004), Feingold (2005).

these associations are often weak, the studies difficult to reproduce, and researchers struggle to define the physiopathological meaning of an association between a marker allele and an increased disease frequency. The risk factor often lacks any biological meaning, which makes it difficult or even impossible to interpret it as a genetic cause. For example, researchers today are trying to associate markers called SNP (Single Nucleotide Polymorphism) with complex pathologies, but they know already that these markers have no biological function. If the genetic markers do not intervene in the appearance of pathologies, it is difficult to place multifactorial diseases in the genetic disease category. Genetic linkage studies involve diseases of Mendelian heredity for which alterations of one gene have a major and specific effect on the phenotype (Huntington's disease, cystic fibrosis and hereditary breast and ovarian cancer are all examples). These studies rely on the co-segregation of a phenotype and certain alleles in families. For a monogenic disease like Huntington's, for example, some genetic markers have been identified via linkage studies. They have made it possible to predict which individual in a family is at risk; genetic tests were first performed in 1986. At that time, however, the protein coded by the gene near the marker and the molecular mechanisms of the disease were not identified. Linkage analysis does not therefore identify the "gene that causes the disease"; only the marker is identified. In 1993 a gene was identified and research on the Huntington gene began. To summarize: linkage analyses localize genes involved in the appearance of monogenic diseases whose mode of transmission is known. Statistical analysis, in this case, relies on knowing the disease's mode of transmission, which gives meaning to the notion of "genetic" disease. However, as in the case of association studies, as long as the gene's function is unknown, it is hard to extrapolate a causal relationship from a link. All of these examples serve to show that the statistical approach alone does not give us a way to understand pathological processes and only sometimes allows us to identify risk factors. It does lead though to tools which allow us to identify people who are predisposed to certain diseases. And finally, it underscores the importance of the hereditary aspect of diseases considered genetic. According to Magnus, the example of the contamination of a well shows that we intuitively consider the pathogenic agent to be the disease's cause because we believe we are capable of acting on this cause (by cleaning the well water for instance). This third approach to causality is no longer a matter of scientific arguments but rather of interests involving medical treatment that explains why certain diseases are considered genetic. From this perspective, a disease is genetic when its genetic determinants appear easier to manipulate than its environmental ones in order to treat or prevent the disease. Magnus is careful to state, however, that there is no such thing as genetic therapy; the therapeutic justification of this instrumental approach is therefore largely compromised. Moreover, certain diseases that are said to be genetic are medically treated with environmental modifications. A special diet, for example, is enough to prevent the appearance of symptoms characteristic of phenylketonuria. Yet should this instrumental approach be reduced to its preventive dimension? Diseases would then be considered as genetic when a genetic test could be put to market that allows people to predict who is and is not at risk of the disease. This solution clearly shows the current limits of medical care for the great majority of diseases we refer to as genetic: most often, a genetic test prevents the disease by allowing for the selection of births. Magnus considers each of these concepts of causality in turn before rejecting them all. But why not consider these three approaches to causality as all contributing together to the definition of a disease as genetic? These concepts do not seem to be competing: they refer to three aspects of medicine. In medical practice, genetic research effectively leads to specific care regimes for individuals and families. Research toward a better understanding of diseases involves localizing genetic determinants as well as explaining their physio-pathological role. In order to better understand why it is so difficult to define one concept of genetic disease, an important distinction must be made. Lenny Moss (2004) distinguishes two concepts of the gene: the gene-P (for prediction) and the gene-D (for development). The first concept appears in the context of linkage studies that identify markers or "disease genes". In this case, the link between the gene and

the disease is strong enough to make predictions for it, even if nothing is known about the gene's specific mode of action. This concept thus has an essentially instrumental value: the gene is defined by its relationship to a phenotype, but knowing the DNA sequence or the way the gene behaves is not necessarily to make a prediction as to who will be ill in a family. On the other hand, prediction allows for prevention. The D-gene is defined as a resource for development. This resource is in itself indeterminate relative to the phenotype. In this last case, the DNA sequence is known and simply considered an element of developmental processes without reference to a particular phenotype. The same gene can be considered as either a P-gene or a D-gene. For example, certain versions of the BRCA1 gene are correlated to a heightened risk of breast and ovarian cancers. In this case, BRCA1 is considered to be a P-gene. But when the BRCA1 gene is not considered as the "breast cancer gene", but rather as a model for the protein synthesis present in several cells and tissues, then it is a D-gene. These proteins can be studied in each cell and tissue context without taking the phenotype (breast cancer) into account. If the P-gene is a prediction tool used in order to obtain a medical or economic benefit, then the D-gene is of explanatory value. With these conceptual distinctions, it is now important to develop some examples that lay out the argument for a common functioning of these three approaches to causality in order to come up with a definition of genetic disease.

The Concept of Molecular Disease

The concept of genetic disease must be understood within the framework of research on molecular pathology. Here I would like to explain this concept in more detail in addition to pointing out how it is related to medical research on the hereditary transmission of diseases. Magnus cites Linus Pauling's work and recalls the importance of the concept of "molecular disease" that appeared in his famous 1949 article (Pauling et al. 1949 , cf. Feldman and Tauber 1997). But Magnus does not pay enough attention to the specificity of Pauling's approach and its connection to the geneticist James Neel's work (1949). To better understand the concept of genetic disease at play here, it is useful to return to a simple question. Why is it that, although certain individuals only suffer from sickle-cell disease at high altitudes, this disease is classified as genetic rather than environmental? The answer lies partly in a molecular analysis of sickle-cell disease carried out in the late 1940s and 1950s' (Pauling et al. 1949 ; Ingram 1957 ; Ingram and Stretton 1959). At the beginning of the 1949 article, Pauling's team points out that sickle-cell is characterized by severe anemia, which results from cellular abnormalities. The anemic trait is a less-severe form of the disease, which is not felt by most individuals who have it. This disease is defined, at the cellular level, by two characteristics: the special "sickle" form which red blood cells take, and their rigidity. Conversely, red blood cells of individuals who are not sick are flexible, concave on both sides, and disk-shaped. Two hypotheses can explain the process by which the red blood cells take the sickle shape. The first is the cellular hypothesis: red blood cells' rigidity is explained by the cell membrane's properties. The second is molecular: this process is linked to chemical and physical properties of hemoglobin, a protein present in these cells. Pauling's team selected the molecular hypothesis and studied these chemical properties by comparing hemoglobin in anemic individuals, individuals with an anemic trait, and those lacking anomalous red blood cells. The notion of molecular disease thus comes partly from the distinction of three levels of analysis: clinical, cellular, and molecular. It also rests on the choice of explicative hypothesis of the molecular level. DNA is also a molecule, and it is easy enough to complete the schema by referring to an even more elementary level: genes that code proteins. Yet to understand what signifies a molecular cause, we must be aware of the fact that researchers try to find the best explanation of sickle-cell disease's characteristic phenomenon. The best one is molecular: it is possible to distinguish two types of hemoglobin that differ by their electric charge, which also explains their difference in shape. Let's pause for a moment

to go through this causal explanation. Hemoglobin, whose function is to carry oxygen in the organism, is described in biochemical terms, as a succession of amino acids (some of which are electrically charged). Because of the protein's electrical charge, each molecule of abnormal hemoglobin carries a region complementary to another hemoglobin molecule's region: these molecules can thus associate. The association of hemoglobin molecules leads to their alignment in the cell: the latter then takes the form of a sickle. This molecular explanation is both specific and complete.

It is but a short step from the concept of molecular disease to that of genetic disease. In 1957, Ingram analyzed amino acids comprising the two types of proteins and found the amino acid modified. He explains: "the sequence of base-pairs along the chain of nucleic acid provides the information which determines the sequence of amino-acids in the polypeptide chain for which the particular gene, or length of nucleic acid, is responsible. A substitution in the nucleic acid leads to a substitution in the polypeptide" (Ingram 1957, p. 328). The terms "lead to" and "are responsible for" express a causal relationship, which is why researchers consider the DNA sequence to be the primary cause of the disease. In the case where an organism is the subject of study, it is effectively impossible to go any deeper than the DNA sequence to explain the abnormal protein's presence. The example of the discovery of the "cystic fibrosis gene" also explains how disease comes to be seen as genetic. In 1985, the main symptoms of this disease were well known: chronic lung and respiratory infections and, in certain cases, insufficient pancreatic enzymes. Measuring chlorine levels in sweat allowed for diagnosis of the disease. Beginning in 1983, researchers knew that tissue of those with the disease was not permeable enough for chlorine; the disease could have perhaps been caused by a problem with chlorine ion transport. But they did not understand the "cause" of these symptoms and wanted to find a "basic defect" that caused the disease: despite intensive research effort, the basic defect in CF remains unidentified. Therefore, the metabolic abnormalities on which the biochemical studies are based are probably secondary or tertiary consequences of the primary defect (Tsui et al. 1985, p. 1054). The direct strategy of studying the abnormal protein that Pauling used is impossible in the case of cystic fibrosis: the pathology's primary explanatory defect is unknown. The inverse genetic method allowed researchers to identify the cause of cystic fibrosis. Philip Kitcher describes this research strategy: "Even though biomedical researchers may initially be entirely ignorant about the physiological processes that go awry in a particular disease, knowing how that disease is transmitted in a sufficiently large sample of families, they can sometimes isolate the locus that is responsible. The strategy is to find genetic markers (...) associated with the transmission of the disease, confine the locus to a particular chromosomal region (...) pick out candidate genes, and ultimately, clone and sequence the desired gene. Knowledge of the gene may then yield enough understanding of the protein to provide insight into the causal basis of the disease" (Kitcher 1994, p. 522). In the case of cystic fibrosis, the gene responsible for the pathology was identified and sequenced in 1989: it codes the CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) protein, which is involved in ion transport. The concept of a candidate gene is important: starting with a linkage study, a region of DNA is localized. Sequencing provides researchers with a long list of nitrogenous bases in which they can locate several groups that correspond to genes. The problem is thus one of knowing which, among all these possible genes, could be "the cystic fibrosis gene". Starting with the DNA sequence makes it possible to determine the protein's amino acid sequences and, in some cases, its function.

Researchers can therefore identify an order of priority as a function of structures and functions of the proteins that are coded by candidate genes. For some of these genes, it is plausible that their modification causes symptoms that are characteristic of the disease. The order of priority is thus defined based on physiological and pathological, rather than genetic, data. The gene coding of the CFTR protein is selected in the following way (Riordan et al. 1989): we know that the conductivity of the chlorine ion across cell membranes diminishes in ill individuals; a good candidate gene would thus

be a DNA sequence that codes a protein that contributes to the formation of an ionic channel or a DNA sequence that codes for a protein involved in the regulation of ionic channels. Second, we identify the RNA that corresponds to the DNA of CFTR in lung, colon, and sudoriparous tissues; this localization matches up with symptoms characteristic of the pathology. Finally, CFTR is analyzed: two groups of amino acids form a field capable of crossing a cell's plasma membrane; this data matches up with the hypothesis that the disease results from a dysfunctional ionic channel. In addition, certain amino acids of this protein are susceptible to form linking fields for molecules that intervene in regulation processes. And finally, the order of amino acids in this protein is compared to that of proteins in other species whose function is known. This comparison allows for the conclusion that the protein coded by the candidate gene is likely involved in ion transport back and forth in cells' plasma membrane. It is the coherence of this physiopathological data that gives meaning to the concept of genetic disease. Let's not forget one final observation: the reverse genetic study follows a genetic linkage study and is thus only possible if the method of disease transmission is known. What, then, is a genetic disease? It is a hereditary disease whose fundamental (molecular) defect has been identified at the level of DNA. This defect explains the disease's characteristics on higher levels (cellular, tissue, physiological). The importance of a disease's hereditary dimension to its definition as a genetic disease can also be highlighted in the example of sickle cell disease. Ingram claims: "The latter [hemoglobin] is an abnormal protein which is inherited in a strictly Mendelian manner; it is now possible to show, for the first time, the effect of a single gene mutation as a change in one amino-acid of the hemoglobin polypeptide chain for the manufacture of which that gene is responsible" (Ingram 1957 , p. 326). It is precisely this identity between the gene as hereditary factor and the gene as DNA sequence that explains the importance of this research in elaborating the concept of genetic disease. Up until 1949, sickle-cell disease was considered to be a disease of dominant transmission and variable expressivity (variable expressivity referring to the fact that the disk-shaped red blood cells only took the sickle shape when oxygen pressure dropped). Geneticists assumed at the time that a single copy of a mutated gene was enough to explain the frequency of individuals in a family who were more or less anemic; the same allele of the gene was expressing itself in different ways. Neel, in a 1949 (Neel 1949) article, showed that sickle cell was transmitted recessively. His predecessors were mistaken because the clinical entity had not been well defined: the anemic trait and sickle-cell anemia were not sufficiently distinct from one another. Neel distinguished phenotypes (the anemic trait and sickle-cell anemia), drew up genealogical trees and observed the frequency of individuals who had and did not have the anemic trait and sickle-cell anemia. His calculations showed that the hypothesis of a recessive transmission had to be the better one. Pauling explained (Pauling et al. 1949) that he arrived at this same conclusion before Neel's article was published. The experiment with molecular biology in 1949 that indeed led to thinking that the mode of the disease's transmission was recessive; the cells of individuals with an anemic trait contained two types of hemoglobin, whereas healthy individuals' cells and anemic individuals' cells had only one type. Each type of protein could thus be translated from different alleles of the same gene. Research in classical genetics and molecular biology, even when carried out separately, confirmed this. One objection does arise for the defense of genetic disease as a concept that relies on this double dimension of hereditary transmission and explanation at the molecular level of certain pathologies. What about complex genetic diseases whose heredity is not well known and that are only associated with specific risk factors?

Statistics Are Blind

Magnus's analysis also failed to account for the connection between statistical studies on genetic risk factors and molecular research that, as we have seen, are linked to physiopathology. It is clear why: statistical methods identify P-genes whose specific activities remain unknown at the molecular level. In the case of diseases that are transmitted following Mendel's laws, the "gene that causes the disease" thus had first and foremost an instrumental predictive value. Today, the Genome Wide Association Study (GWAS) is an association study. The goal is to show a difference in allelic frequency at the level of one single locus in individuals who are or are not affected by common diseases. As is the case for the P-gene, nothing is known about this allele's specific action on the pathology's development. Is Magnus right then to separate the explicative and statistical approaches? Before I answer this question, it is important to point out a main connection between statistical studies and hereditary studies (which, incidentally, echoes the connection between hereditary and molecular dimensions to genetic diseases). To understand the genetic component of multifactorial diseases with complex heredity whose appearance depends on the interaction of several genes and of environmental factors, it is necessary to demonstrate their hereditary dimension. In their reference book, *Human Molecular Genetics*, Tom Strachan and Andrew Read (2004) separate the study of genetic diseases into two groups. One chapter, "Identifying human disease gene", is dedicated to methods that localize genes of diseases that are monogenic and Mendelian. Another chapter, "Mapping and identifying genes conferring susceptibility to complex diseases", deals with methods that identify genes for susceptibility and predisposition. Here, the way the authors present the difference between Mendelian and complex diseases is quite instructive. In effect, they explain that nobody would argue with the idea that a disease is genetic if it clearly follows a Mendelian mode of transmission, which assumes that a single gene is transmitted in a family. On the contrary, for complex traits, it is necessary to prove that genetic factors are involved in the disease's development. And for that to happen, research on the disease's transmission in genealogies is necessary. This type of research has, for example, led to the distinction between two different forms of Alzheimer's disease: a familial form (FAD or Familial Alzheimer's Disease) which is transmitted in a dominant autosomic manner, as well as a sporadic one, linked to APOE e4. Three predisposition genes (PSEN1, PSEN2 and APP) confer a very high risk of developing early-onset FAD. FAD is rare (5 %); the sporadic form is more frequent. There are also several ways of working with multifactorial diseases to calculate their genetic component (for a review, cf. Feingold 2005 and Campion 2001), but three steps are always required: "To show that the disease is familial, to show that this familial tendency is due to genetic factors, and, finally, to identify the genes involved" (Feingold 2005, p. 927). Research on genetic determinants thus rests on studies of the disease's mode of transmission in genealogies. Finally, the most effective group of statistical methods is most meaningful when the research takes heredity into account. Familial aggregation studies attempt to find out the prevalence of a disease within relatives compared to their prevalence within the general population. Twin studies compare the similarity of monozygotic and dizygotic twins. Segregation studies collect genealogical trees and model the number of genes involved. For the association studies that do not assume any study of relatives but simply that of two populations (affected and not), a study of the affected subject and its two parents completes the analysis (Feingold 2005, p. 930). In a recent article in *Nature* (Manolio et al. 2009), the authors explain that GWAS have led to the identification of hundreds of genetic variations associated with diseases or complex human traits. However, these variants explain only a small proportion of heritability: it is thus necessary to find research strategies that go beyond GWAS to explain the rest. The second connection Magnus misses concerns the links between statistical and explicative approaches to diseases. Association studies are often accompanied by research on candidate genes. Their selection is based on the coherence between their function and the characteristics of the pathology being studied. This makes sense: the difficulties that arise in the study of complex disease are linked to studies' statistical properties, for example problems with the threshold of significance or the population choice. They are also linked to the weakness of the relative

risk conferred by each susceptibility allele. For each different pathology being studied, it is thus necessary to try to look elsewhere for what these alleles might modify in characteristic biological pathways. Two examples serve to better explain this strategy. The first is Jean Dausset's article on histocompatibility systems and cancer risk (Dausset 1968). After Dausset showed in 1958 that the histocompatibility complex is hereditary, he wanted to explain why certain cancers, such as chronic lymphoid leukemias, are very frequent in some populations and nearly absent in others. When, in 1968, he discovered an association between antigenic differences characteristic of mouse populations and their resistance (or not) to a leukemia virus, Dausset immediately raised the following question: how to explain this association? In this study, the presence of specific alleles of the H-2 histocompatibility system was linked to a resistance to cancer; however, a link is not an explanation. "Resistance genes" (Dausset 1968 , p. 1397) at locus H-2 or near it, might have explained this resistance, but no experiment "directly demonstrates the involvement of the H-2 locus itself" (ibid., p. 1398). Finally, according to the immunological explanation of the resistance phenomenon, the concept of cause in this context would designate a molecule that would prevent the virus from penetrating the cell. Demonstrating the role of the products of gene resistance in recognition phenomena of the antigen or of virus penetration in the cell was, at the time, beyond researchers' reach. This experiment only shows a correlation between certain alleles and a diminished frequency of virus contamination; it did not demonstrate causality. On the contrary, it showed that statistical research becomes meaningful in the context of attempts to explain genes' effects on pathological processes. The second example is more contemporary. The difficulties encountered in statistical studies led researchers to progress toward showing evidence of more adequate relationships between more clearly described and individuated phenotypes and candidate genetic factors that may be causes. Dominique Campion reminds us that linkage analyses, which rely on knowing a disease's transmission mode, are most often inadapted for the study of complex diseases. Association studies, for their part "are most interesting when it comes to candidate genes" (Campion 2001 , p. 1139). A gene always intervenes in a biological pathway to whose modification it contributes. According to Campion, "for a long time, it was more or less implicitly understood that Mendelian and multigenic diseases involved two radically different biological characteristics" (ibid., p. 1144). In reality, in both cases genes intervene to modify biological pathways. In the case of monogenic disease, a single gene's mutation causes a major disruption in the pathway, which has important function consequences. In the case of multigenic diseases, the modifications are not as severe, but the effects of several involved risk factors on a single biological pathway can accumulate to the point where they cross a threshold that causes the pathology to appear. What then is the strategy for identifying good candidate genes? It is to first identify the biological pathways involved in a pathology's development. This analysis clarifies why Magnus's "re-examination" of Huntington's disease is not as astonishing as he claims. This disease is linked to a characteristic repetition of CAG codons in a portion of chromosome 4. We know that it is difficult to predict whether individuals with between 30 and 40 repetitions will or will not be sick. This is a problem if we assume that the "Huntington's disease gene" works in an "all or nothing" way. But if the disease results from a disturbance in a biological pathway and if the effects of gene mutations are quantifiable, then it is easier to understand why the disease functions as a series of thresholds being reached and passed. The variable expressivity of monogenic diseases, such as symptoms' severity or the timing of their appearance, is also better explained by the threshold definition that corresponds to protein interactions with other variables involved in the biological pathway. Ultimately, it is not surprising that different genetic variants or even different genes have similar effects on biological pathways. For cystic fibrosis, there is an inventory of several different variants whose effects are more or less deleterious. Thalassemia is linked to mutations in several different genes. In order to find the biological pathways that are disturbed in the context of genetic diseases, it is often necessary to identify Mendelian sub-entities and intermediary phenotypes are also often necessary. This is the case with diseases like

schizophrenia, where difficulties in reproducing linkage and association studies have led researchers to identify intermediate phenotypes associated with a more readable genetic determinism. Campion (2001), for instance, points out difficulties with slow eye tracking and sensorial filtering, which are two endophenotypes associated with schizophrenia. Maziade's team (Maziade et al. 2003) refers to "dimensional" phenotypes (aggregates of symptoms) or neurocognitive phenotypes associated with the disease that are genetically less complex. Finally, there are genetic studies, which, in their difficulty, point researchers toward different nosographies: "schizophrenia" increasingly tends to be reduced to different etiologies while its borders with bipolar disorders become more blurred. So, are we justified in considering multifactorial diseases as "genetic" diseases? It does not seem to be the case that researchers working on the genetics of complex diseases are trying to show that asthma or bipolar disorders are genetic diseases. Rather, they are using researching strategies that allow them to better understand the genetic factors involved in these diseases, since a better grasp of a disease's physiopathological development can lead to the identification of treatment options based on an improved understanding of interactions between genes and the environment. Genetic studies can even improve diagnoses: the HLA-B27 variant now helps diagnose ankylosing spondylitis.

The Choice of Genetics

I would like to return one last time to the issue of knowing why, among all the possible causes of a disease, the genetic ones are often privileged. Magnus ends his article with the following idea: epistemological answers to this question are insufficient. All that is left is to examine the use of the concept of genetic disease. He explains: "Labeling a disease as 'genetic' is to make an implicit claim that, for that disease, understanding and therapy will best come about through research at the genetic level. In other words smuggled into the very conceptual classification is a set of commitments about the best way to allocate resources and the best way to do good science and medicine" (Magnus 2004 , p. 240). Yet for Magnus, we are cruelly lacking in empiric arguments justifying the financing of genetic research. On the other hand, this remark demonstrates why it is necessary to identify values beyond the scientific ones that would lead to this type of research being privileged over others. Looking at Kitcher's (2000) analysis of reasons that lead to genetic determinism's dominance in biology and medicine, despite our recognition of the environment's role in human development and behavior, serves to explain Magnus's argument in more depth. Kitcher's article is a response to Richard Lewontin, who "diagnoses errors that have seduced influential scholars and their readers into believing vulgar slogans about genes and destiny" (ibid., p. 283). Lewontin (1992) criticizes the popular conception of genetic determinism by pointing out that it ignores the interaction between genetic and non-genetic factors. This is why extending the category of genetic disease is socially dangerous: it leads us to think that what is "genetic" is "inevitable" and veers quite close to the belief that the only solution to "genetic" problems is the selection of individuals based on genetic criteria. For Lewontin, this popular conception belies a deeper concern: biology should, at least in part, be conceived of differently. Following Lewontin's reasoning Susan Oyama defends the argument that the oppositions biologists work on, for instance that between innate and acquired, or gene and environment, are not relevant (Oyama 1985). Focusing attention on the gene as a causal factor is simply an abstraction of complex causal situations that unfairly prioritizes certain determinants of the phenotype. Yet Kitcher aims to show that the interactionist concept, which separates genes from the environment, should not be rejected either. He argues: "Genetic determinism persists not because of some subtle error in conventional ideas about the general character of biological causation but because biologists who are studying complicated traits in complex organisms are prone to misapply correct general views" (Kitcher

2000 , p. 284). He is thus defending a “democratic” concept of environmental and genetic causes and explains why, in fact, scientists privilege genetic causes. To demonstrate his argument Kitcher analyzes the reaction norm, a commonly used concept in biology. The underlying strategy in genetic determinist research on biological traits “begins by isolating certain properties of organisms for exploration of their causal impact, regarding the phenotype as the product of contributions from particular kinds of DNA sequences, on one hand, and from everything else, on the other. It goes on to inquire how the phenotype varies as the DNA sequences are held constant and as other factors (the cytoplasmic constitution of the zygote, the molecules passed across cell membranes, etc.) change” (ibid., p. 285). The genotype’s reaction norm is the graphic representation of this strategy: a phenotype’s genetic determinism is defined as its relative invariance, given a single genotype in all environments. For opponents of the interactionist argument, the assumptions that are required to build the reaction norm, for instance the genotype phenotype distinction, must be reconsidered. Kitcher, meanwhile, is trying to point out in his article that this tool is scientifically valid. Isolating certain causal factors by holding them constant in order to see how the effect varies when other factors are modified is legitimate. But if this tool is scientifically valid, scientists often use it incorrectly. The interactionist effectively recognizes that there are several causes involved in development. Kitcher thus defends the principle of “causal democracy”, where it is also scientifically justifiable to propose a causal analysis of a particular environmental factor by observing what happens when a genotype is modified: “The democracy principle accords no special privilege to the representations that foreground the role of genes” (Kitcher 2000 , p. 290). Why, then, do biologists insist on the primacy of genetic causes?

The first reason Kitcher gives is that it is pragmatic: researchers believe that new technologies that come out of molecular biology can improve their understanding of certain diseases and certain behaviors. Kitcher explains that behavioral genetics promises much, because it must be possible to use DNA sequencing techniques to identify alleles shared among different individuals in a population. In this case, if the causally relevant environmental factors were identified (which poses the larger challenge to this type of study) it would be possible to study phenotypes’ variation as a function of a genotype held constant when environmental factors change. In the medical field, he uses the example of research on alcoholism and addiction, and explains why scientists are attempting to identify genetic causes: “They begin with genetic causes not because they are convinced that these are the most important (that the norms of reaction for certain ‘addictive’ genotypes are virtually flat) but because they want to unravel the neurochemistry, and they see the investigation of genotypes as a thread that will lead them into the tangle” (ibid., p. 295). And, in a sense, this is what is happening, since this type of strategy identifies multiple nosological entities for a pathology that was previously considered singular. Modes of transmission for these traits in families and eventual genetic interactions are better understood. Our understanding of pathologies improves. In the case Kitcher discusses, researchers hope that the inverse genetic method and knowledge of the DNA sequence will lead to an understanding of molecular changes in the brain. These scientists believe that they understand the molecular details of interactions carried out between the organism and the environment, which differ between addicted and non-addicted people. Here, then, it is unnecessary to assume a social deformation of science that would rely on socially valued norms to justify the importance granted to research on genetic determinants of diseases. Yet according to Kitcher, we would assume that a study of the reaction norms of “violence alleles” would be socially encouraged for reasons that are not scientific or medical. Under the principle of causal democracy, it would be possible to carry out a study in which the environment would be held constant and the variation of phenotypes would be observed as a function of this causal factor. In this case, according to Kitcher, we could see that for a single genotype, the phenotype varies based on environment and we could thus conclude that the environment is a causal factor in violent behavior. Kitcher, however, writes: “In a society that

consistently and callously turns its back on programs that might aid the unfortunate and sees taxation as a form of robbery rather than a necessary means to social cooperation, the investigation I have outlined has no obvious point" (ibid., p. 296). On the contrary, studies that hope to show that keeping a genotype constant and varying environments leads to an invariable phenotype (the reaction norm is flat or nearly flat) have a greater chance of finding funding. In this case, detecting this genotype in individuals would identify a predisposition to violent behavior. This type of study would thus reinforce the idea that social solutions are hopeless: if an individual is predisposed to violence, there is little chance that school, for instance, would help him or her escape this fate. Genetics are considered inevitable, despite the fact that all scientists agree that genes alone determine no trait.

This way of thinking has been subject to numerous critiques since the 1990s. The concept of "geneticization" arose as a criticism of the extension of a genetic vision of man accompanied by scientific discoveries. Abby Lippman (1991) and Henk Ten Have (2001) view it as a process that affects medicine as well as the broader society. Geneticization redefines individuals according to their genes and creates a new language to describe, interpret and understand human life: it explains human differences by their genetic differences. This trend is deterministic and reductive, and it bears serious consequences for the definition of genetic disease. To begin with, in medicine, extending the category of genetic disease empties the term of its specificity; it would cease to carry any nosographic meaning. Thus, if the notion of genetic terrain is to have meaning, it must not lead to every disease being "genetic" under the pretext that some alleles are associated with a greater risk for a given disease. Second, geneticization has consequences for the instrumental approach to cause. Stating that X is genetic is, according to the critique of geneticization, to abandon any public policy that tries to improve environmental conditions, since "genetic" here is the equivalent of "inevitable". At the same time, considering a disease genetic is to also believe that the easiest course of treatment lies at the genetic level. In fact, the most common type of management of genetic diseases is fetal screening with genetic tests. The possible consequences of such processes are indeed worrisome. The extension of the concept of genetic disease effectively leads to practices that raise the specter of eugenics. In current debates, the slippery slope argument is frequently raised to underscore the danger of extending prenatal or preimplantation diagnostic practices in the context of genetic predisposition. For example, authorizing such diagnostic testing for diseases that could manifest themselves later on and whose likelihood of appearing is not equal to 1, such as hereditary ovarian and breast cancers, could open the door for birth selections based on much more trivial criteria. From this perspective, there is much at stake in the definition of genetic disease.

Conclusion

Three criteria thus function together to define a disease as genetic. The first concerns specific causality that allows for the understanding of the pathology's development. A gene is in this context an explicatory unit for a specific modification of an important biological pathway within a pathology's framework. The second criterion is the link between these explanations and the hereditary dimension of diseases, approached using the framework of statistical studies. The last criterion is the perspective of preventive or curative medical care that is compatible with ethical imperatives. The articulation of these criteria could be extended to many more timely reflections on the concept of disease. For example, homosexuality was removed from the DSM (Diagnostic and Statistical Manual of Mental Disorders) in 1973. Consider this: if homosexual practices had been correlated to a group of genetic risk factors at the time, would they still have been pulled from the list of mental disorders?

References

- Campion D (2001) Dissection génétique des maladies à hérédité complexe. *Médecine/sciences* 17:1139–1148
- Dausset J (1968) Les systèmes d'histocompatibilité et la susceptibilité au cancer. *Presse Med* 76(28):1397–1400
- Feingold J (2005) Maladies multifactorielles un cauchemar pour le généticien. *Médecine/sciences* 11(21):927–933
- Feldman D, Tauber A (1997) Sickle cell anemia: reexamining the first molecular disease. *Bull Hist Med* 71(4):623–650
- Gannett L (1999) What's in a cause? The pragmatic dimensions of genetic explanations. *Biol Philos* 14:349–374
- Gifford F (1990) Genetic traits. *Biol Philos* 5(3):327–347
- Hesslow G (1984) What is a genetic disease? On the relative importance of causes. In: Nordenfelt L, Lindahl BIB (eds) *Health, disease and causal explanation in medicine*. Reidel, Dordrecht, pp 183–193
- Hull R (1979) Why genetic disease? In: Capron AM et al (eds) *Genetic counseling facts, values and norms*. Alan R. Liss, New York, pp 57–69
- Ingram VM (1957) Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* 180:326–328
- Ingram VM, Stretton AO (1959) Genetic basis of the thalassemia diseases. *Nature* 184:1903–1909
- Kitcher P (1994) Who's afraid of the human genome project? In: Hull DL, Ruse M (eds) *The philosophy of biology*. Oxford University Press, Oxford, 1998, pp 522–535
- Kitcher P (2000) Battling the undead. How (and how not) to resist genetic determinism. In: Kitcher P (ed) *In Mendel's mirror. Philosophical reflections on biology*. Oxford University Press, Oxford, 2003, pp 283–300
- Kitcher P (2003) *In Mendel's mirror: philosophical reflections on biology*. Oxford University Press, Oxford
- Lewontin R (1992) *Biology as ideology. The doctrine of DNA*. Richard Harper Perennial, New York
- Lippman A (1991) Prenatal genetic testing and screening: constructing needs and reinforcing inequities. *Am J Law Med* 17:15–50
- Magnus D (2004) The concept of genetic disease. In: Caplan A, McCartney J, Sisti D (eds) *Health, disease, and illness*. Georgetown University Press, Washington, DC, pp 233–242
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Maziade M, Merette M, Chagnon Y-C, Roy M-C (2003) Génétique de la schizophrénie et de la maladie bipolaire. *Médecine/sciences* 19(10):960–966

- Moss L (2004) *What genes can't do*. MIT Press, Cambridge, MA
- Neel JV (1949) The inheritance of sickle cell anemia. *Science* 110:64–66
- Oyama S (1985) *The ontogeny of information. Developmental system and evolution*. Cambridge University Press, Cambridge
- Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. *Science* 110:543–548
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelcák Z (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245:1066–1073
- Sober E (2001) The meaning of genetic causation. In: Buchanan A, Brock D, Daniels N, Wikler D (eds) *From chance to choice. Genetics and justice*. Cambridge University Press, Cambridge, pp 347–370
- Sterelny K, Kitcher P (1988) The return of the gene. *J Philos* 85:339–361
- Strachan T, Read AP (2004) *Human molecular genetics 3*. Garland Science, New York
- Ten Have HAMJ (2001) Genetics and culture: the geneticization thesis. *Med Health Care Philos* 4:295–304
- Tsui L-C, Buchwald M, Barker D, Braman JC, Knowlton R (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230:1054–1057