Hist. Phil. Life Sci., 33 (2011), 497-514

Genetic Susceptibility to a Complex Disease: The Key Role of Functional Redundancy

Gaëlle Debret[1], Camille Jung[2], Jean-Pierre Hugot[2], Leigh Pascoe[3], Jean-Marc Victor[1] and Annick Lesne[1,4,*]

[1]*Laboratoire de Physique Théorique de la Matière Condensée (LPTMC) UMR 7600 CNRS & Université Pierre et Marie Curie-Paris 6, France*

[2]*Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France & Université Paris-Diderot, Paris, France & Assistance Publique Hôpitaux de Paris, Hôpital Robert Debré, Paris, France*

[3]*Fondation Jean Dausset (CEPH), Paris, France*

[4]*Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France*

** Corresponding author: LPTMC case 121, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris Cedex 05, France*

ABSTRACT - Complex diseases involve both a genetic component and a response to environmental factors or lifestyle changes. Recently, genome-wide association studies (GWAS) have succeeded in identifying hundreds of polymorphisms that are statistically associated with complex diseases. However, the association is usually weak and none of the associated allelic forms is either necessary or sufficient for the disease occurrence. We argue that this promotes a network view, centred on functional redundancy. We adapted reliability theory to the concerned sub-network, modelled as a parallel array of functional modules. In our model, as long as one module remains active, the function correlated with the respective disease is ensured and disease does not occur. Genetic factors reduce the initial number of available modules while environment, contingent surroundings, personal history, epigenetics, and some intrinsic stochasticity influence their persistence time. This model reproduces age-specific incidence curves and explains the influence of environmental changes. It offers a new paradigm, according to which disease occurs due to a lack of functional elements, depending on many idiosyncratic factors. Genetic risk assessed from GWAS is only a statistical notion with no direct interpretation at the individual level. However, genomic profiling could be useful at population level in devising models to guide decisions in health care policy.

(Acronyms: CD Crohn's disease; GWAS genome-wide association studies; SNP single nucleotide polymorphism)

KEYWORDS - Genetic susceptibility, genome-wide association studies, gene-environment interaction, biological networks, redundancy, reliability theory

Introduction

A disease can be seen as the loss of one or a few biological functions in specific cell types. A first class is composed of Mendelian disorders where a mutation in one or both copies of a single gene is both necessary and sufficient to explain the disease. In this case occurrence is fully explained by a genetic factor. A second class is composed of multifactorial diseases, also called complex diseases, in which both genetic susceptibility and environmental factors are implicated in the aetiology. The involvement of a genetic component is inferred from an increased disease risk among relatives of affected individuals. A global and quantitative identification of the genetic determinants is now possible with genome-wide association studies (GWAS). Overall they show that genetic factors, namely specific allelic forms, although significantly associated with the disease, are neither sufficient nor necessary. This fact can be summarized by saying that each disease-associated allelic form has an incomplete penetrance. We use here the term "allelic form" for each of the possible molecular sequences of a gene or genome stretch, rather than "allele," which also designates one of the two copies of a gene. "Penetrance" is an acknowledged though only statistical notion describing the phenotypic impact of an allelic form. More precisely, penetrance is the proportion of individuals carrying a particular allelic variation that also express an associated trait. In medical genetics, the penetrance of a mutation is the proportion of individuals carrying the mutation who exhibit clinical symptoms. Equivalently, penetrance is the probability of developing the disease or trait given the presence of the genotype in question. Incomplete penetrance means that no one-to-one relationship exists between the presence of the allelic form and the observation of the trait. It is generally related to multifactorial disease (although monogenic diseases also can have incomplete penetrance). Explaining the incomplete penetrance of disease-associated allelic forms evidenced by GWAS is one of the challenges of our modeling study.

On the other hand, a modest concordance rate among monozygous twins (that is, the probability that two individuals sharing identical genotypes also share the same clinical syndromes) and observed changes of incidence (that is, the fraction of new cases of disease in the population over one year) over a period of a few years demonstrate that epigenetic and/or environmental factors are also important. However, what is called gene-environment interaction, namely the interplay between

environmental and genetic factors in complex disease aetiology, is far from being understood nor even described in a quantitative way (Caspi and Moffitt 2006; Mitchell 2009). Accordingly, the notions of genetic susceptibility, causality, and even the very notion of disease, require a re-assessment for complex multifactorial diseases.

In this study, we discuss the necessary change of paradigm and offer an operational interpretation of GWAS results, together with a critical discussion of their aim of quantitatively assessing genetic risk. In the context of public health, a proper evaluation of genetic risk is essential at two levels: at the individual level, with the hope of developing personalized medicine based on genomic profiling; at the population level, to make reliable predictions about the future incidence of a disease and decide on efficient preventive actions and policies. The results of GWAS, showing no necessary or sufficient genetic determinants, lead us to question the relevance of genomic profiling as providing a basis for preventive treatments and personalized medicine. We claim that prediction at the individual level requires a systemic understanding of the mechanisms involved in the onset of complex diseases (see also the discussion by Bertolaso [2011] and by Gross [2011]). The aim of our modeling approach is to propose a first step in this direction.

Genetic epidemiology and GWAS

An emerging viewpoint is that genome sequencing makes classical epidemiology obsolete and promotes genetic epidemiology (Weeds 2006). An emblematic approach is provided by genome-wide association studies, comparing the presence of hundreds of thousands of single nucleotide polymorphisms (SNPs) in cohorts of patients to their distribution in samples of control individuals. GWAS have been recently conducted for many complex diseases (Wray et al. 2008) in large cohorts of several thousands of patients displaying the same degree of disease according to a standard classification (Kutschenko 2011). Their careful statistical analysis identified many loci (up to 70 for Crohn's disease [Franke et al. 2010]) having particular SNPs significantly associated with the disease, however with only modest odds ratios, mostly between 1.1 and 2 (the largest is around 4 for the main Crohn's disease susceptibility loci; see Hugot et al. 2001). These loci will be henceforth termed "risk alleles" whatever their form, moreover extending the notion of allele to any kind of genomic sequence, whether coding for a gene or not. The "odds ratio" is defined for a pair of allelic forms as the ratio of the odds of each form, where the "odds" is the ratio of the conditional probability of being ill

(given this form) to the conditional probability of being healthy (see also Pascoe et al. 2007 for more details). In non-technical terms, the odds ratio is the relative ratio in a case-control study. An allelic form is statistically correlated to the disease occurrence when the odds ratio is above some threshold. This threshold is fixed as a function of the size of the cohort by the requirement of having a sufficient statistical significance.

However, these studies show no clear causal relation. Having a risk allele in the form statistically correlated with the disease (what is currently termed with some prejudice the "variant form," taking as a norm the form correlated with the absence of this specific disease) or even a set of risk alleles in a variant form, is neither necessary nor sufficient for the appearance of the disease. Some control subjects even have more risk alleles in their variant form than some cases (Weersma et al. 2009a; 2009b). On the other hand, it can be shown from the histograms of the number of variant risk alleles in cases and controls respectively, that the probability of being ill increases according to an exponential law with respect to the number of variant risk alleles, clearly showing the importance of the genetic components discovered by GWAS. This can be used to support the notion that the disease is not due to a rare variant in each individual but rather correlates with the accumulation of common variants (Debret et al., in preparation).

GWAS, thus, strongly challenge the notion of risk alleles in case of complex diseases. Certainly odds ratios provide a statistically well-founded link between having a given allelic form and displaying the disease. However such a link can by no means be considered as causal. It is only a statistical correlation at the population level. The notion of genetic susceptibility of an individual, and presumably the very notion of causation for such diseases, have to be reevaluated (see also the discussion on causal claims in medicine by Russo and Williamson [2011]). One has to understand how the presence of specific allelic forms favors the disease occurrence and to clarify in what respect the presence of the same allelic forms in the genome of two individuals can lead to the disease or not, as shown in monozygous twin studies (Baranzini et al. 2010; Katsnelson 2010). Certainly, environmental factors play an essential role, as do possibly epigenetics, life history, individual contingencies, or some intrinsic stochasticity. By "intrinsic stochasticity" we mean the randomness occurring in, e.g. intracellular processes, in particular due to the numerous and highly sensitive degrees of freedom involved in these processes. An acknowledged example is the stochasticity of gene expression, contributing to the variance observed in the phenotype. Intrinsic stochasticity is opposed to extrinsic stochasticity, where the probabilistic nature of the outcome is assumed to be due to random external influences.

At this point, a legitimate question is whether it makes sense to look

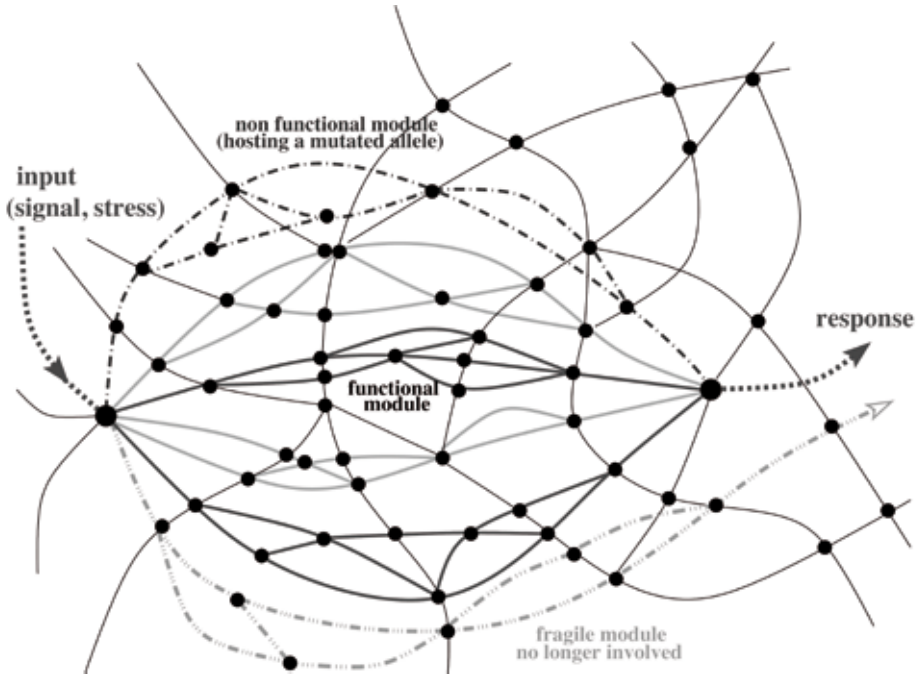

Fig. 1 - Idiomatic outline of the individual functional sub-network whose dysfunction results in a disease, and associated molecular phenotype.
Thin edges represent the whole biological network composed of all reactions and interactions potentially at work at the molecular level in the organism. Our model is based on functional redundancy: a given function can be achieved by independent functional modules (also called functional pathways in the literature). These modules are identified with non-overlapping sets of edges, delineated in bold, each able to produce the proper functional activity or response. In each module, the alleles (the nodes on the figure) having variant forms with different activities are identified with the disease-associated alleles detected in genome-wide associations studies. Among the modules, some (dashed-dotted) are intrinsically non-functional (hosting a non-functional mutated allele, e.g. nonsense or missense allele). Some others (dashed-triple-dotted) are no longer involved in the function. Typically some of their elements are used in a competing function, which depends on the genotype, individual life history, environment, and some intrinsic stochasticity. The actually functional modules (continuous bold lines), and the alleles they involve, define the molecular phenotype of the given individual relative to the considered function. Quantitative modeling of the disease occurrence is based on reliability theory applied to these redundant functional modules, considering that each module is characterized by a persistence time (see Fig. 2).

for a causal allelic form and what kind of biological mechanisms such a term should cover and account for. We here argue that it may be more productive to take a network view of the biological function whose impairment results in the disease. Such a network view gives, in particular, a central role to the ensuing functional redundancy (see Fig. 1).

Network-based redundancy

Most of the biological components of the cell are organized in complex networks of interactions (Barabasi et al. 2011; Carter 2005). In such network structures, a given function may be achieved by any of several interacting elements. Typically, the network structure accommodates a certain degree of redundancy (Tautz 1992) so that if one element is altered, others could still ensure the function. Accordingly, this redundancy contributes to the functional robustness of the network with respect to the impairment of some elements, e.g. due to mutations or changes in life style or external conditions (Lesne 2008). Network approaches of biological functions and systems biology promoted the notion of "functional module" (Barabasi et al. 2011; Sieberts and Schadt 2007; Wagner et al. 2007) namely, a connected sub-network able to achieve a given function. In some contexts, e.g. metabolic networks, functional modules are also termed "functional pathways." They should not be confused with other (though possibly related) notions of "modules," e.g. structural, developmental, or evolutionary modules (see the critical analysis in Mitchell 2006). We propose to consider redundancy at a more integrated level, that of functional modules. It has been recently proposed that disease-causing elements should be sought at this level (Chen et al. 2010; Rossin et al. 2011; Wang et al. 2010). We moreover suggest that causation can be understood only in giving the central role to functional redundancy.

Specifically, a signal can act through several paths between the initial signaling event and the activation of the element actually responsible for the functional response (Fig. 1). In our case, this defines a functional sub-network associated with the function whose breakdown is associated with the disease. Functional redundancy refers to the existence of several redundant functional modules for achieving the same function. In this view, risk alleles correspond to genomic sequences that are involved with one of these functional modules in one way or other. For instance, these alleles can code for an involved protein or a non-coding RNA, or bind an involved regulatory factor. Risk alleles in their disease-associated form typically prevent the modules to which they belong from being functional for the function of interest (see Fig. 1, dash-dotted modules) or induce a fast destabilization of the module by being rerouted into another competing function (see Fig. 1, dashed-triple-dotted modules).

The fact that the presence of risk alleles in their disease-associated form is neither necessary nor sufficient for the occurrence of the disease leads us to consider an additional level of variability. Namely, we propose that not all functional modules are involved in the function in

a given individual: "what can occur" is not "what does occur" (Carter 2005). Those actually involved are not the same from one individual to another. The origin of this individual variability is diverse: possibly, stochasticity in the transcriptional regulatory network, contingent life history (e.g. intra-uterine development or breeding; see Waterland and Rached 2006), epigenetics, or on-going plasticity in relation to the microbiome (Dupré 2011). The functional sub-network has to be seen as a dynamic system conditioned by the genotype but also influenced by the environment and individual history, being different from one individual to the other and even at different periods of life in the same individual. Accordingly, the actually involved modules are also age-dependent.

Our proposal: an idiosyncratic age-specific molecular functional phenotype

We propose the notion of "molecular functional phenotype" to describe the intermediate level between the genotype and the observed phenotype, according to which an individual implements a function and its regulation in a given environment. It is specific to the considered function. It describes how the set of functional modules is partitioned for the given individual at a given age into (i) the constitutively impaired modules, typically because they involve an allele carrying a deleterious mutation, (ii) the modules no more used in the considered function, typically because one of their elements has been rerouted into another competing function, and (iii) the modules actually recruited for the achievement of the function and its regulation (Fig. 1). This molecular phenotype can also be dissected and presented at the level of alleles involved (via the product of their expression, or as regulatory sequences) in the corresponding functional modules. In this alternative view, alleles are either (i) in a non-functional form, (ii) in a functional form but not involved in the function or (iii) functional and actually involved.

With this notion of molecular functional phenotype defining the idiosyncratic, age-specific, and partly stochastic implementation of the sub-network relevant to the considered function, it follows that two individuals in the same environment with the same risk alleles, e.g. monozygous twins, may display different responses to their environment and have different disease status. Another more indirect support for this model is the fact that odds ratios of risk alleles may vary significantly from one population to another one (Ioannidis et al. 2009). What matters for the disease occurrence is the set of functional modules that each individual has developed and recruited to achieve the function of interest. This

consideration contributes to understanding the observed incomplete penetrance of the risk alleles with regard to disease susceptibility. It is also consistent with the fact that the genotype does not fully determine whether someone will develop a disease and it integrates life history and various environmental factors.

The presence of multiple risk alleles in their disease-associated form favors the disease occurrence insofar as it reduces functional redundancy of the network and shrinks the set of possible ways to achieve the biological function. In our model, a residual individual variability remains among people sharing the same genotype (e.g. monozygous twins or, in the context of a given disease, people carrying the same disease-associated allelic forms). It lies in the different sets of functional modules (or their components at a more molecular level) that are being used to perform the considered function. This notion reflects a change of paradigm from deleterious mutations to functionally missing modules, which depend not only on the whole set of alleles involved in the function, but also on all the internal or external factors controlling their expression and usage. This view is reminiscent of the "endophenotype" (Gottesman and Gould 2003) introduced in psychiatry, namely an intermediary level between the genotype and the pathological phenotype at which some features associated with the disease are delineated and more easily related to the genotype variability. Our notion of functional phenotype also substantiates the notion of "functional substrate" depending on the genotype and affected by the environment (as proposed by Caspi and Moffitt 2006) as the locus of the gene-environment interplay in complex disease aetiology.

Network reliability theory

In order to explain why someone becomes ill during their lifetime, we investigate the conditions for the failure of their individual functional sub-network (associated with the function whose breakdown results in the disease, Fig. 1). For this, we adapted to complex diseases the general theory of systems failure. Also known as reliability theory, it describes the probability of a system completing its expected function during an interval of time or, conversely, its probability of failure (Barlow et al. 1965). In Gavrilov and Gavrilova (2001), a central role is given to the redundant architecture of the system, which is ideally adapted to the network structure and functional redundancy of biological systems. Failure of the system appears as an essentially systemic property arising from a network built from non-aging elements (here the alleles indeed do not

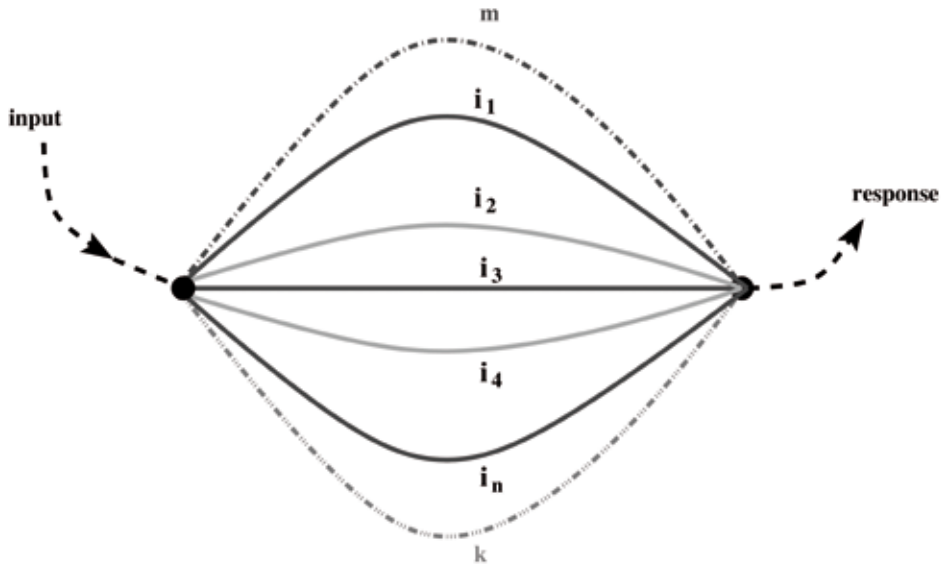

Fig. 2 - Simplified functional sub-network used in reliability study, extracted from the full biological network sketched in Figure 1.
Only redundant functional modules are represented. They are non-overlapping by construction. Here each is reduced to a single entity (a path on the sketch, with the same graphical convention as in Fig. 1: dashed-dotted for intrinsically non-functional modules *m*, dashed-triple-dotted for non-involved modules *k*, continuous bold line for actually functional modules *i*). Each of the actually functional modules is characterized by a persistence time depending on the genotype, individual life history, environment, and some intrinsic stochasticity. This characteristic time corresponds to the average duration before the module is destabilized and its elements become involved in another competing function. The disease occurs when all modules are no longer functional. This approach leads to an excellent agreement with the age-specific incidence curve of Crohn's disease.

experience any somatic mutations and no protein degradation processes are of relevance). In this model, disease results from the progressive inactivation with time of the functional modules initially used to ensure the function (at birth or at the initiation of the biological function). Each module can dynamically become destabilized due to perturbations experienced by the organism and no longer fulfill its function. The failure may be induced by a targeted bifurcation, where some element of the module is rerouted towards a more demanding function after some specific change in the environment, for example. It can also lie at the level of gene expression and its regulation, meaning that a gene is no longer expressed at sufficiently high level to operate properly. Disease

development is thus a dynamic process, corresponding to the escape of redundant modules from their original functional regime (e.g. kinetics of the metabolic reactions) under the accumulated influences of various internal and external factors (modules *i* becoming *k* in Fig. 2; see also Gross 2011). Disease arises when all functional modules have failed. In this regard, disease-associated allelic forms are not severally effective for disease: a single allele has no specific action. Only jointly do they have a decisive causal impact in preventing the functional modules to achieve their function 2011 properlyduring the lifetime of the individual.

At the individual level, the main ingredient for the quantitative implementation of our reliability model is the average duration τ, henceforth termed "persistence time," during which a given functional module *i* (Fig. 2) contributes to the function of interest in the considered individual. In technical terms, the failure of each functional module is assumed to be an exponential process with hazard rate $1/\tau$. Allocating one and the same persistence time τ to every module leads to a strong disagreement between the predicted age-specific incidence curves and the data obtained for several complex diseases, such as Crohn's disease (Molinié et al. 2008), schizophrenia (Lia et al. 2007), or multiple sclerosis (Phadke and Downie 1987). We, thus, introduce a specific persistence time τ (i; G, E, E') for each module *i*, which moreover depends on the genotype G of the individual, on the environment E (nutriments, bioclimatic conditions, pollutants, which are most often shared between individuals close to each other), and also on the individual history, epigenetics, contingent surroundings, some intrinsic stochasticity and, in fact, all the other functions and processes taking place inside the organism. This latter set of influencing factors is summarized as a strictly idiosyncratic internal environment E' that is independent of the genotype (e.g. differing even in monozygous twins). Modules hosting a mutated allele, where the mutation affects the ability of the module to fulfill the function (modules *m* in Fig. 2), are initially non-functional or fail very rapidly, i.e. τ (i; G, E, E') is close to 0. The persistence time heterogeneity of the different modules matches the heterogeneity of their functional efficiencies, which is in strong agreement with the aetiology of complex diseases. People with no or only a few disease-associated risk alleles are likely to have at least one robust functional module; that is, at least one module with a long persistence time (longer than their life span). Accordingly, they have a very low probability to fall ill during their lifetime. For a rare disease, such as Crohn's disease or multiple sclerosis, these people form the huge majority of the general population. A noticeable point of our model is that the persistence time of a module is determined not only by the variant forms of the alleles involved in this module. Persistence time is also

an individual feature, as indicated by the dependence on the internal environment E'. Even for individuals sharing the same disease-associated allelic forms, the set of active functional modules at a given age (see Figs. 1 and 2) can be quite different. Accordingly, their disease susceptibility (that is, the probability to fall ill during their lifetime) in a given environment or faced with the same perturbations can differ markedly.

The effect of fragile modules, i.e. modules having a short persistence time (shorter than the individual life span), play a role only in those people having no robust modules in their molecular phenotype. This situation occurs when the alleles involved in the modules (through their expression product or as genomic regulatory binding sites) are either in an ill-adapted variant form, not expressed, or recruited in another function. Having a short persistence time, these fragile modules do not dramatically influence the total prevalence (that is, the fraction of cases of the disease in the population at a given time). Indeed they fail almost with certainty within the lifespan of the individual. They are the main determinants of the age of the onset of the disease.

We emphasize that disease-associated allelic forms play a role only in being ill-adapted, so that the module in which they are implicated has a short persistence time. They are not actively deleterious and rather correspond to a shortfall, responsible for a lack of robustness and efficiency. Variants with large odds ratios correspond to mutations spoiling the functionality of the most adapted modules. Of course, the larger the number of disease-associated allelic forms in a given individual, the more fragile are his/her functional modules and, hence, the larger the risk that they all fail during a lifetime so that disease eventually sets on.

Epidemiological data

In order to confront our scenario with epidemiological data (age-specific incidence curve; that is, the curve representing the rate of falling ill for a given age, plotted as a function of age) we have to derive the prediction of our model at the population level. The description becomes statistical. Grouping, for simplicity, persistence times into two classes, namely short and long times, a functional module is then characterized by the probability that its persistence time is long. In this setting, we were able to compute a parameterized expression of the probability that at least one module remains functional (non-occurrence of the disease) and, conversely, the distribution of the time at which all modules have failed and disease sets on (Debret et al. in preparation). An excellent agreement with the incidence curve of Crohn's disease (see data in Molinié et

al. 2008) allowed us to fit the parameters of our model (for instance, the number of redundant functional modules associated with the underlying function, which is about 12 in the case of Crohn's disease). Importantly, the expression is very discriminating and other fits are not able to match the data, even qualitatively. Our model is also able to fit epidemiological data for other complex diseases like schizophrenia (see data in Lia et al. 2007) or multiple sclerosis (see data in Phadke and Downie 1987). All the incidence curves for these diseases, including Crohn's disease, have an initial exponential-like increase, reaching a peak, and then decreasing rather slowly and eventually saturating on a plateau at the greatest ages.

As explained above, the quantitative analysis of our model and its predictions are based on statistical reliability theory. Communication network engineers have introduced several different sub-notions of reliability, namely availability, performability, survivability, or sustainability (Medhi 1999). For instance, in our case, network availability fails when a component gets preferentially involved in another function. Performability characterizes the efficiency of the paths beyond their mere connectivity. In our case, we assume that the regulatory sub-network is fault-tolerant until all modules have failed, reminiscent of the reliable network made of unreliable elements (Moore and Shannon 1956). This engineering viewpoint is likely to offer a fruitful analogy and further concepts could certainly be transposed to complex diseases.

Gene-environment interplay

Several complex diseases have a very recent history of rising prevalence, e.g. increasing by a factor of ten up to a value of 1/600 for Crohn's disease (Kappelman et al. 2007) and 1/500 for multiple sclerosis (Koutsouraki et al. 2010), which clearly indicates a response to environmental or lifestyle changes. Our network reliability model offers a new way of viewing the underlying interaction (using the current terminology) between environmental and genetic factors, beyond their observed statistical relationship, if any (Mitchell 2009). In our model, the involvement of an allele (through its expression product or as a regulatory sequence) into a functional module is partly controlled by the environment. The same goes for the possible recruitment of the module to achieve the function. An environmental change or a new environmental factor, can destabilize some functional modules that were previously robust, so that more individuals develop the disease, who would not have developed it previously. Quantitatively, this can be incorporated in our model at the individual level through a modification of the persistence times of the

functional modules. At the population level, this corresponds for each module to a change of its probability to have a long persistence time, possibly leading to a dramatic change in the incidence curve (Debret et al., in preparation).

Discussion

We have defined a disease as the breakdown of some biological function. Analysis of the network of protein interactions identified by double hybrid studies in yeast (Feldman et al. 2008) shows that the connectivity of genes of Mendelian disorders is low whereas it is intermediate for genes involved in complex disorders and high for lethal disorders. This observation suggests that for Mendelian disorders, the key function has a low redundancy, hence there is no place for an influence of the environment or internal stochasticity: the defective allele is causally effective whatever the context, i.e. Mendelian disorders display a (near) complete genetic causality. In our model, Mendelian diseases correspond to the impairment of some essential and non-sustainable functional pathway as a result of a genetic mutation. In other words, the underlying function can be achieved by only one functional module, with no redundancy. This leads to the prediction that any allele involved in this single module (when existing in different variant forms with different functional efficiencies) should be detected as a risk allele, which is actually the case in some Mendelian diseases, e.g. 13 risk alleles in Fanconi anemia (Rossin et al. 2011).

We postulate that complex diseases correspond to the progressive inactivation of alternative functional modules (lipid metabolism regulation in obesity, immune response in Crohn's disease, multiple sclerosis and type II diabetes, etc). Typically, this happens because some of the elements of functional modules are rerouted into a competing function. Disease appears as a systemic feature and an emergent property of the molecular network associated with the underlying function. Accordingly, genetic variations are not defective in themselves, but relative to a context, an environment, a set of perturbations, and all the more to an individual and his or her history, e.g. environmental conditions during early development. We suggest that it might well be that a disease-associated allelic form in a given environment appears to be a protective variant in another environment. In our view, the difference in causality between Mendelian and complex diseases (the two extremes of a continuous spectrum) is a shift from a genetic cause to a genetic risk, which we are able to quantify and of which we are able to explain the dependence upon environmental factors.

Our explanatory model is based on the supposition of a network architecture of the biological functions. However an explicit knowledge of the network nodes and edges is not required: the only network feature invoked in our analysis is the redundancy of functional modules (see Fig. 2). The function is impaired only when all the modules that could perform the function have failed. This failure can be due either to a dynamic destabilization of the module, an expression switch-off of some involved elements, or their re-routing into another function. It is presumably all the more rapid when the elements involved in the module are ill-adapted and the module is used as an auxiliary or rescue pathway (short persistence time). In contrast, modules involving well-adapted elements (corresponding through expression or as regulatory sites to alleles in their non-disease-associated form) have a high probability to remain functional during the whole lifespan of an individual (long persistence time). Our model allows that several diseases with close aetiologies can share risk alleles (in agreement with the experimental analysis in Goh et al. 2008) since these alleles only contribute in their disease-associated form through the absence of the functional modules which rely on them, compelling the individual to have recourse to other modules to implement the required function.

In our view, it does not make sense to discriminate between the proportion of risk explained by risk alleles and that due to the environment, as is currently done (in a moreover linear way) at the level of risk assessment. Each individual has a susceptibility profile to a given disease, depending on his or her genotype G, its environment E, and even its internal environment E' (meaning the internal context in which the considered function is satisfied which itself depends on the way the other functions are implemented). Accordingly, we introduced a novel notion, termed the age-specific molecular functional phenotype of the individual relative to a given function, describing the idiosyncratic usage of the genotype and implementation of the sub-network underlying the function (Figs. 1 & 2). Individual susceptibility to disease depends on this molecular functional phenotype. It arises in our model through the fragility, i.e. the short persistence time, of the alternative modules involved in the considered function (Figs. 1 & 2). The fewer modules are available and the more fragile they are, the greater the risk for the individual to develop the disease. Genetic factors (risk alleles in the disease-associated form) play a role in reducing the initial number of available modules. Jointly with environmental factors, they also influence the persistence times of the functional modules.

Recently, a study reported no evidence for genetic, epigenetic, or transcriptome differences that could explain disease discordance between

monozygous twins (where only one develops the disease) in multiple sclerosis (Baranzini et al. 2010; Katsnelson 2010). The authors performed a comparison of the fully sequenced genomes (on a single pair of twins), of one million SNPs (on three pairs of twins), i.e. a measurement analogous to that of GWAS, of epigenetic marks and of expression levels of key alleles (transcriptome data), without finding any significant difference. Several aspects of our model are consistent with this study: (i) SNPs do not reveal any necessary nor sufficient condition for disease development; (ii) the absence of differences of expression in key genes suggests that the determining factor is either the expression level of auxiliary genes with a mild and redundant regulatory role, a difference in the usage of the key gene product (rerouting of the products, bifurcation of the pathways and reactions), or both since usage and expression are linked by feedback loops; and most importantly, (iii) the role of environmental factors (expected by physicians in multiple sclerosis) needs to be understood (if possible mechanistically and not only assessed on statistical grounds) and quantified.

Conclusion and perspectives

Overall, our scenario introduces a far-reaching change of paradigm in complex disease susceptibility. Disease is postulated to originate causally in a lack (non-use or inactivation) of functional modules rather than in the presence of detrimental alleles. We thus explain why risk alleles in their disease-associated form determined by GWAS are neither sufficient nor necessary for the occurrence of disease. They are not noxious allelic forms and their only deleterious effect is to prevent the modules in which they are involved to be functional, reducing the number of possible ways of achieving the pertinent function. In other words, complex diseases follow from the lack of positive causes, rather than from the presence of a negative cause. The risk of developing the disease and the age at which disease sets on, are controlled by the robustness or fragility of the functional modules as quantified by their persistence times. The actual difference maker (Waters 2007) is thus the set of functional modules and their persistence times. The aetiology of the disease relies on these functional modules: robust ones (characterized by long persistence times) determine whether the disease will occur, fragile ones (characterized by short persistence times) determine when the disease will occur. The therapeutic target would be to protect those initially functional modules. Mutated alleles contribute only in being non-functional and ruling out the modules relying on them.

GWAS are useful at the population level, for epidemiology, and to devise models for predictions and simulations in order to guide public health decisions. The necessity of a careful distinction between explanations at the individual and population levels cannot be overemphasized. It seems delicate (and presently out of reach) to found personalized medicine on the sole basis of GWAS and genome profiling, because these methods give population results. Prediction at the individual level would require an understanding of the detailed mechanisms involved in the onset of the disease (presented here in an as yet abstract fashion) and the detection of their malfunction. We suggest that a relevant basis for personalized medicine would be the difficult reconstruction (or at least partial knowledge) of molecular functional phenotypes, for instance by crossing qualitative information on previous or other diseases (either for the individual or its relatives) and physiological specificities.

Acknowledgements

This paper greatly benefited from the talks and discussions during the First European Advanced Seminar in the Philosophy of Life Sciences, held at the Brocher Foundation, Geneva, Switzerland, September 6-11, 2010, and especially from the precise and immensely relevant comment made by Mila Petrova (Exeter) on the presentation given there of the above work and ideas by Annick Lesne. The work of the research team is supported by a grant from the French Agence Nationale de la Recherche, program GENOPAT 2009 (project CROHN MODEL).

References

Barabasi A.L., Gulbahce N. and Loscalzo J., 2011, "Network medicine: a network-based approach to human disease", *Nature Reviews Genetics*, 12: 56-68.

Baranzini S.E., Mudge J., van Velkinburgh J.C. et al., 2010, "Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis", *Nature*, 464: 1351-1356.

Barlow R.E., Proschan F. and Hunter L.C., 1965, *Mathematical theory of reliability*, New York: Wiley.

Barrett J.C., Hansoul S., Nicolae D.L. et al., 2008, "Genome-wide association defines more than 30 susceptibility loci for Crohn's disease", *Nature Genetics*, 40: 955-962.

Bertolaso M., 2011, "Hierarchies and causal relationships in the interpretative models of the neoplastic process", *History and Philosophy of the Life Sciences*, 33: 515-536.

Carter G.W., 2005, "Inferring network interactions within a cell", *Briefings in Bioinformatics*, 6: 380-389.

Caspi A. and Moffitt T.E., 2006, "Gene-environment interactions in psychiatry: joining forces with neuroscience", *Nature Reviews Neuroscience*, 7: 583-590.
Chen X., Wang L., Guo M., Barnard J. and Zhu X., 2010, "Pathway-based analysis for genome-wide association studies using supervised principal components", *Genetic Epidemiology*, 34: 716-724.
Debret G. et al., 2011, "A reliability network model applied to the genetic susceptibility of Crohn's disease", in preparation.
Dupré J., 2011, "Emerging sciences and new conceptions of disease. Or, beyond the monogenomic differentiated cell lineage", *European Journal of Philosophy of Science*, 1: 119-132.
Franke A., McGovern D.P.B., Barrett J.C. et al., 2010, "Meta-analysis increases to 71 the tally of confirmed Crohn's disease susceptibility loci", *Nature Genetics*, 42: 1118-1125.
Feldman I., Rzhetsky A. and Vitkup D., 2008, "Network properties of genes harboring inherited disease mutations", *Proceedings of the National Academy of Sciences of the USA*, 105: 4323-4328.
Gavrilov L.A. and Gavrilova N.S., 2001, "The reliability theory of aging and longevity", *Journal of Theoretical Biology*, 213: 527-545.
Goh K.I. Cusick M.E., Valle D. et al., 2007, "The human disease network", *Proceedings of the National Academy of Sciences of the USA*, 104: 8685-8690.
Gottesman I.I. and Gould T.D., 2003, "The endophenotype concept in psychiatry: etymology and strategic intentions", *American Journal of Psychiatry*, 160: 636-645.
Gross F., 2011, "What can systems biology tell us about diseases", *History and Philosophy of the Life Sciences*, 33: 477-496.
Hugot J.P., Chamaillard M., Zouali H. et al., 2001, "Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease", *Nature,* 411: 599–603.
Ioannidis J.P.A., Thomas G. and Daly M.J., 2009, "Validating, augmenting and refining genome-wide association signals", *Nature Reviews Genetics*, 10: 318-329.
Katsnelson A., 2010, "Twin study surveys genome for cause of multiple sclerosis", *Nature*, 464: 1259. (Commentary of Baranzini et al. 2010).
Kappelman M.D., Rifas-Shiman S.L., Kleinman K. et al., 2007, "The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States", *Clinical Gastroenterology and Hepathology*, 5:1424-1429.
Koutsouraki E., Costa V., and Baloyannis S., 2010, "Epidemiology of multiple sclerosis in Europe: A review", *International Review of Psychiatry,* 22: 2-13.
Kutschenko L.K., 2011, "Why classify diseases? Revisiting the epistemological function of medical classification systems", *History and Philosophy of the Life Sciences*, 33: 583-602.
Lesne A., 2008, "Robustness: confronting lessons from physics and biology", *Biological Reviews*, 83: 509-532.
Lia X., Sundquista J. and Sundquista K., 2007, "Age-specific familial risks of psychotic disorders and schizophrenia: A nation-wide epidemiological study from Sweden", *Schizophrenia Research,* 97: 43-50.
Medhi D., 1999, "Network reliability and fault tolerance", *Encyclopedia of Electri-*

cal and Electronics Engineering, Vol. 14, edited by J.G. Webster, New York: Wiley, 213-218.
Mitchell S.D., 2006, "Modularity - More than a buzzword?", *Biological Theory*, 1: 98-101.
Mitchell S.D., 2009, *Unsimple truths: Science, Complexity, and Policy,* Chicago: The University of Chicago Press.
Moore E. and Shannon C., 1956, "Reliable circuits using less reliable elements", *Journal of The Franklin Institute,* 262: 191-208, 281-297.
Pascoe L., Zouali H., Sahbatou Z. and Hugot J.P., 2007, "Estimating the odds ratios of Crohn disease for the main CARD15/NOD2 mutations using a conditional maximum likelihood method in pedigrees collected via affected family members", *European Journal of Human Genetics* 15: 864-871.
Phadke J.G. and Downie A.W., 1987, "Epidemiology of multiple sclerosis in the north-east (Grampian region) of Scotland-an update", *Journal of Epidemiology and Community Health*, 41: 5-13.
Rossin E.J., Lage K., Raychaudhuri S. et al., 2011, "Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology", *PLoS Genetics*, 7: e1001273.
Russo F. and Williamson J., 2011, "Epistemic causality and evidence-based medicine", *History and Philosophy of the Life Sciences*, 33: 563-582.
Sieberts S.K. and Schadt E.E., 2007, "Moving towards a system genetics view of disease", *Mammalian Genome*, 18: 389-401.
Tautz D., 1992, "Redundancies, development and the flow of information", *Bioessays*, 14: 263-266.
Wagner G.P., Pavlicev M. and Cheverud J.M., 2007, "The road to modularity", *Nature Reviews Genetics*, 8: 921-931.
Wang K., Li M. and Hakonarson H., 2010, "Analysing biological pathways in genome-wide association studies", *Nature Reviews Genetics*, 11: 843-854.
Waterland R.A. and Rached M.T., 2006, "Developmental establishment of epigenotype: a role of dietary fatty acids?", *Scandinavian Journal of Food and & Nutrition*, 50: 21-26.
Waters K.C., 2007, "Causes that make a difference"*, The Journal of Philosophy*, 104**:** 551-579.
Weeds D.L., 2006, "Rethinking epidemiology", *International Journal of Epidemiology*, 35: 583-586.
Weersma R.K., Stokkers P.C.F., Cleynen I. et al., 2009a, "Confirmation of multiple Crohn's disease susceptibility loci in a large Dutch-Belgian cohort", *American Journal of Gastroenterology*, 104: 630-638.
Weersma R.K., Stokkers P.C.F., van Bodegarden A.A. et al., 2009b, "Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort", *Gut*, 58: 388-395.
Wray N.R., Goddard M.E. and Visscher P.M., 2008, "Prediction of individual genetic risk of complex disease", *Current Opinion in Genetics and Development*, 18: 257-263.