Can Medicine Be Predictive?

José M. Mato

There is a nautical chart attributed to Christopher Columbus, possibly drawn up in 1492 following the conquest of the Muslim city of Granada by the Catholic Monarchs, and obviously made before he set sail on the voyage that would lead to the discovery of America. The map stretches from the south of Scandinavia to the mouth of the river Congo, showing all the Mediterranean ports of Europe and Africa in detail. The enormous space that Columbus dedicated to the Atlantic Ocean is conspicuously lacking in detail. In all probability, this huge blank space served not only to mark the frontier of the known world and therefore the potential expansion of world knowledge — knowledge that would indeed double over the course of the following century. It also opened up a route for the imagination and the adventure of sailing through it, a route travelled by numerous sixteenth and seventeenth-century explorers who in most cases were destined to remain anonymous.

Columbus's portolan chart shows us where any geographical investigation would be carried out while there were still unexplored lands on the earth. In the same way, progress at the frontier of knowledge — be it in chemistry, physics or mathematics, or in more applied fields such as biology, geology, medicine, social science, or nature conservation — has been produced, since the beginning of time, by opening new routes for the imagination and setting out on adventures into the unknown in search of solutions based on hard facts, just like a geographical expedition to fill in an incomplete map. The scientist, as opposed to the philosopher or the ideologue, does not possess that quest for totality nor does he try to explain the world through a system.

Blanca Muñoz, *Espacio arrugado* (detail), 2000

The scientist progresses via modifications and additions to theories that already exist by delving into the details. Darwin did not invalidate the biological and geological knowledge of his day; he incorporated it to give it validity within the scope of the evolution of species through natural selection.

At the end of the twentieth century, however, a phenomenon appeared which might be called "genocentric." It consists in considering genes as the sole elements responsible for what we are and how we look. And we often forget that genes give few clues, not only to the personality and appearance of a person, but also to their state of health. We should be clear that in biology, when speaking of things such as genes, genotypes and genomes, we are not speaking of an objective physical reality — i.e., a phenotype — but of information. The genomes of two unrelated people are approximately a 99.9-percent match: they differ by just one part in a thousand. How is it possible that people so genetically alike can be so physically different? How can they have such diverse phenotypes, so that even genetically identical twins are not the same, nor have the same personality, nor suffer from the same diseases? Only a few decades ago, geneticists had to defend the idea that the development of any species was influenced not just by the environment but also by genes. The current situation has changed so radically that we now have to defend the idea that the environment and chance play a fundamental role in the risk of suffering common diseases, in the formation of personality, or in ageing.

Molecular biology, via the direct study of genes, has developed a deterministic concept of biology that rests on the principle of specificity. According to this concept, which has been crucial in the understanding of the function of genes and the proteins they encode, each genome has its corresponding unique three-dimensional structure — a unique phenotype. This way of thinking has given excessive value to the biology of systems, to a way of carrying out research in biosciences which consists in measuring everything (genes, proteins, metabolites) in a biological system in the hope that upon analysing this huge amount of information, new properties of the system will emerge which allow an integral understanding. This holistic focus, diametrically opposed to the reductionist method that has dominated and driven biomedical research for over 150 years, is simply a mathematical model — a way of understanding biology and trying to explain it through a theoretical construct.

The key question is: How can phenotype diversity be produced from a single genotype? The issue is that specificity is a qualitative, abstract term. In biology, the interactions between molecules (DNA, RNA, proteins and metabolites) are characterized not by exclusivity, but by the multiplicity of possible interactions between some molecules and others. The problem is that in biology we cannot carry out reverse engineering: it is not possible to discover how a cell works based on a mathematical system which incorporates thousands of measurements of its internal components, simply because there is no single solution, no predefined design. From this perspective, the human being (or any other organism) is no longer the fulfilment of a linear

Ryoichi Kurokawa, *Mol*, 2012

genetic program, but the result of an open process in which individuality springs from certain genetic information interacting between other information existing at that moment. This vision of biology allows an understanding of how multiple phenotypes can be formed from a single genome, and how environment and chance select, at each moment, one from among all the possible phenotypes.

The rise of systems biology is a temporary fad. Future scientific research in biology and biomedicine will continue to be dominated by the spontaneous response of the researcher towards a question, working directly on the cell components, solving direct rather than reverse problems, opening new routes for the imagination, and seeking a solution based on hard facts. From this vision of biology that seeks to solve problems, predictive medicine — currently so fashionable — is not realistic, because it is one thing to know that a certain gene mutation increases the risk of suffering some type of cancer in the general population (e.g. to 50 percent) and another thing to be able to tell a person whether they belong to the 50 percent who are going to develop the tumour or the 50 percent who are not. The fact that not all those with an identical mutation in the same gene will go on to develop a certain type of cancer is the greatest proof that biological reality is extremely complex, that within cells there are multiple

compensatory and adaptive mechanisms, that we need to distinguish clearly between a person's genome (the genotype) and their characteristics (the phenotype), and that biology is not deterministic, nor does it rest on the principle of specificity. It thus follows that illness is individual and medicine cannot be based on statistics.

Even in those diseases caused by mutations in a single gene, as in the case of phenylketonuria, the correlation between genotype and phenotype is not always close. Phenylketonuria is a rare hereditary disease caused by the presence in the body of toxic levels of an amino acid called phenylalanine, due to mutations in an enzyme called phenylalanine hydroxylase. Phenylalanine is an amino acid obtained from diet; it is needed to synthesize

In biology, the interactions between molecules are characterized not by exclusivity, but by the multiplicity of possible interactions between some molecules and others. The problem is that we cannot carry out reverse engineering: it is not possible to discover how a cell works based on a mathematical system simply because there is no single solution, no predefined design

proteins in an organism. If not treated in time, the levels of phenylalanine increase in the bodies of those with phenylketonuria, in most cases causing severe and irreversible intellectual disability. Although the genotype predicts the biochemical phenotype (the concentrations of phenylalanine in the blood), it does not predict the clinical phenotype (the appearance of intellectual disability). Thus, some people with phenylketonuria and high plasma concentrations of phenylalanine display normal intelligence, and some siblings sharing the same genotype have different metabolic and clinical phenotypes. The mechanisms that cause these differences in cerebral pathogenesis, despite the presence of comparable concentrations of phenylalanine, are unknown. So even in the case of a monogenic disease such as phenylketonuria, where the effect of the genetic mutation is the accumulation of a specific amino acid, the pathogenic mechanism is complex at the metabolic and cognitive level, and it is not possible to predict intellectual disability from the genotype.

The differences between the genomes of two people are mainly variations in one gene, consisting of the substitution of one base for another in the DNA sequence. These polymorphisms of a single nucleotide in a gene receive the name of SNP (single nucleotide polymorphism). Around 12 million SNPs have been identified in the human genome. Each person's genome contains a unique SNP profile, a specific genotype, which is a result of genetic variation between individuals. During the last decade numerous studies have been carried out in which hundreds or even thousands of people have been genotyped using microarrays of up to 5 million SNPs, covering the whole human genotype, with the aim of identifying specific genetic

variants implicated in common conditions such as obesity, diabetes or hypertension, which are all associated with increased morbidity and mortality.

In practical terms, this research has provided little in the way of results. Let us analyse, for example, the case of obesity. Obesity is currently one of the most serious global public health problems because it is associated with some of the principal causes of death in the world such as diabetes, cardiovascular disease, and some types of cancer, and in recent decades it has reached epidemic proportions. According to the World Health Organisation,[1] around 1.3 billion people in the world suffer from obesity or are overweight — that is 18 percent of the global population. In the United States, this proportion is around 38 percent. Mexico has also surpassed 30 percent. Australia and the United Kingdom are at around 25 percent; and Spain, Germany and Finland have surpassed 15 percent. The countries with the slimmest populations are India, Indonesia and Japan.

Obesity is a result of an energy imbalance produced when people ingest more calories than they burn. How do genes influence obesity? Between 1969 and 1979, the percentage of adults in the United States who were not overweight remained stable at around 74 percent. Ten years later, in 1989, this percentage had gone down to 63 percent, and in 1999 only 39 percent of the adult population were slim, a percentage that according to the most recent data has now fallen to 34 percent. Clearly, most of the blame for this situation does not lie with the genes. A lack of physical activity and an excessive intake of calorie-rich foods are the main causal factors in the global obesity epidemic. However, not all people exposed to the same environment become obese, nor do they all suffer from the same obesity-related health problems, suggesting that some part of the weight difference between adults is linked to genetic factors. It must be made clear, however, that only on very rare occasions is obesity produced by the mutation of a single gene (monogenic obesity), and in the majority of cases obesity — as with other common conditions — is the result of multiple and complex interactions between genes, proteins and metabolites with diet and environmental factors, which are not well understood.

Through studies comparing thousands of obese and non-obese people, variants related with the development of obesity have been identified in around fifteen genes. In practical terms, what use is it for a person to know whether his or her genome contains one or more of these genetic variants associated with obesity? It is no use at all. If the person is slim, he or she must obviously carry on with the same lifestyle and pay no attention to the information. If the person is obese, two things must be made clear: 1) that it is impossible to know if the gene or genes in his or her genome have had any influence on their being overweight because, as we have said, biological events do not happen in a linear manner, and therefore we cannot

1. WHO, http://apps.who.int/bmi/index.jsp

conclude that, for example, 5 kilos of excess weight in a person correspond to such and such a gene, and another 3 kilos correspond to another; and 2) that in any case, an increase in physical activity and a reduction in daily calorie intake will reduce weight.

The rapid advances in DNA sequencing technology have made sequencing of the whole genome of a person a technically and economically feasible objective. The X-Prize Foundation recently offered a 10-million-dollar prize to the first research team capable of sequencing the complete genome of 100 centenarians in 30 days at a cost of up to 1 000 dollars per genome, with an error of less than one base per million base pairs sequenced. What benefits could this technology currently offer for the health of the individual? To answer this question, it is important to emphasize that clinical reasoning is basically Bayesian. In other words, the predictive value associated with a diagnostic test varies when it is applied to populations with indices of prevalence very different to those of the studied condition. For example, in a person diagnosed with excessive levels of iron, the gene mutation known as HFF is a highly reliable predictor of a diagnosis of hereditary hemochromatosis. However, in a population that has not been preselected for high levels of iron, the presence of the same mutation confers only a slight risk of developing clinical symptoms. Equally, in a person diagnosed with hypercholesterolemia, the mutation of certain genes implicated in cholesterol metabolism is a highly reliable predictor of the diagnosis of familial hypercholesterolemia. However, in a population that has not been preselected for high levels of cholesterol, the presence of these same mutations has little predictive value regarding the risk of developing hypercholesterolemia. These results speak for themselves of the importance of interpreting the results of studies of genetic variations within an adequate medical context. Carrying out the routine sequencing of the genome in healthy people means separating the diagnostic test from the medical question, which could lead to the diagnosis of a clinical condition when in reality it does not exist (false positives). One example would be the erroneous diagnosis of hemochromatosis in a person with a low risk of developing the condition, simply because he or she shows a mutation of the HFF gene. Routine genome sequencing could likewise lead to the identification of false negatives — diagnosing people as healthy when they have non-genetic forms of a condition, or forms produced by mutations of other genes, the complex reorganization of chromosomes, etc. In practical terms, it is hard to see how the routine sequencing of the genome in healthy people could have a positive effect on the health of the individual.

Medicine, like insurance companies, works with general statistics, and experts base their conclusions and recommendations regarding public health policies on studies that conclude, for example, that obesity is a serious global public health problem and states must follow active policies that promote exercise and a good diet. But these studies can say nothing about the individual risk of being obese in one country or another. To draw an analogy, drivers involved in

Ryoichi Kurokawa, *Mol*, 2012

few traffic accidents are deemed to have a positive future and therefore enjoy a bonus on their insurance premiums. Conversely, drivers with a history of traffic accidents have a poor outlook so they pay a penalty regardless of where they live. In the same way, when the weight, blood pressure, glucose, triglycerides and cholesterol of a person are normal, his or her future is deemed good — and that is the received bonus; but if these parameters are not normal then his or her future is bad, and that is his or her penalty. Because it is not written in a person's genes that he or she cannot modify their lifestyles and reduce their weight, blood pressure, glucose, triglycerides and cholesterol, and thus improve their future. And neither is it written in a person's genes that they cannot be slim in the United States and obese in India. It is the phenotype (weight, blood pressure, glucose, triglycerides, cholesterol, etc.) and not the genotype that provides greater information regarding a person's future.

That is why medicine is individual, and it must adapt to the person and work with specific cases rather than with statistics. Let us take the case of breast cancer. Over 1.3 million women worldwide are diagnosed with breast cancer each year. Although in developed countries survival rates at five years following diagnosis increased from 40 percent in the 1950s to 86 percent in 2004, not all breast cancers are the same and some women fare better than others. When a

patient is diagnosed with breast cancer, the tumour(s) are classified into four different phenotypes according to how oestrogen receptors (ER positive), the progesterone receptor (PR positive), or any receptor from the epidermal growth factor receptor family (HER2 positive) are expressed on the cell surface. If none of the previous three are manifest, it is classed as triple-negative breast cancer. The study of a great number of cases has demonstrated that when tumours are governed by oestrogens, progesterone or HER2 there are effective treatments with a survival rate at five years of 93 percent, while for triple-negatives, conventional chemotherapy together with surgery and radiotherapy is the only available treatment, and survival is reduced to 77 percent. Clearly, knowing whether the tumour displays certain receptors (ER, PR or HER2) or not — i.e. knowing the tumour phenotype — is crucial in choosing the best type of treatment for each patient. It is the phenotype, the fact that the surface of the tumour cells expresses certain receptors (ER, PR or HER2), together with the clinical history of the patient, which determines the best treatment in each case.

Another example which clearly illustrates why medicine must be flexible and work with specific cases rather than statistics is that of lung cancer, a condition for which there is still no treatment. Every year, over 1.3 million people in the world die from lung cancer, making this the most fatal type of cancer ahead of stomach cancer (approximately 740 000 deaths a year) and liver cancer (approximately 700 000 deaths a year). We have discovered that around 5 percent of lung cancer cases are governed by a type of mutation that reorganizes a single gene that encodes a protein called ALK. Tests have shown that approximately 60 percent of ALK-positive patients respond well to treatment with a molecule which inhibits the activity of ALK and which is currently undergoing clinical research. The reason why 40 percent of the cases of ALK positive lung cancer do not respond to this treatment is unknown. This example teaches us that even in the most favourable circumstances in which the guiding principle has its origins in the mutation of a single gene, the lack of linearity in the chemical processes that govern human physiology and physiopathology produces a diversity of phenotypes indicating, once again, the need for medicine to be individual and to adapt to the person.

Research into the molecular basis of common chronic conditions such as obesity, hypertension, cardiovascular disease, cancer, and diabetes has become the main focus of genetic epidemiology. Consequently, there is renewed interest in studying the relationship between disease and racial descent. Numerous studies have documented differences in the frequencies of diverse common conditions — such as cardiovascular disease — in correlation with racial descent. Is this conclusion justified? Because if it is correct, this is where we should be looking for the genotypic bases of common diseases. Firstly, bear in mind that the current concept of race is based on the experience of naming and organizing populations found during the rapid expansion of European countries during the seventeenth century. It follows that we are dealing with a social construct with economic, legal, socio-political, and also biological

ingredients in various proportions that, due to this malleability, have been used to justify diametrically opposed objectives ranging from committing genocide to improving public health. Secondly, although it is possible to group people according to their geographical continental origin by analysing mutations on their DNA microsatellites, the relevance for health of these results has not been demonstrated. These microsatellites are DNA sequences with a high level of mutation, generally in non-coding DNA areas, where a small fragment is repeated consecutively. Microsatellites are used as genetic markers in paternity tests and in population studies, but it is not clear that these differences are relevant when it comes to health.

The globalization of common chronic conditions suggests, however, that variations in the state of health between populations are due more to differences in exposure to environmental causes than to genetic variations

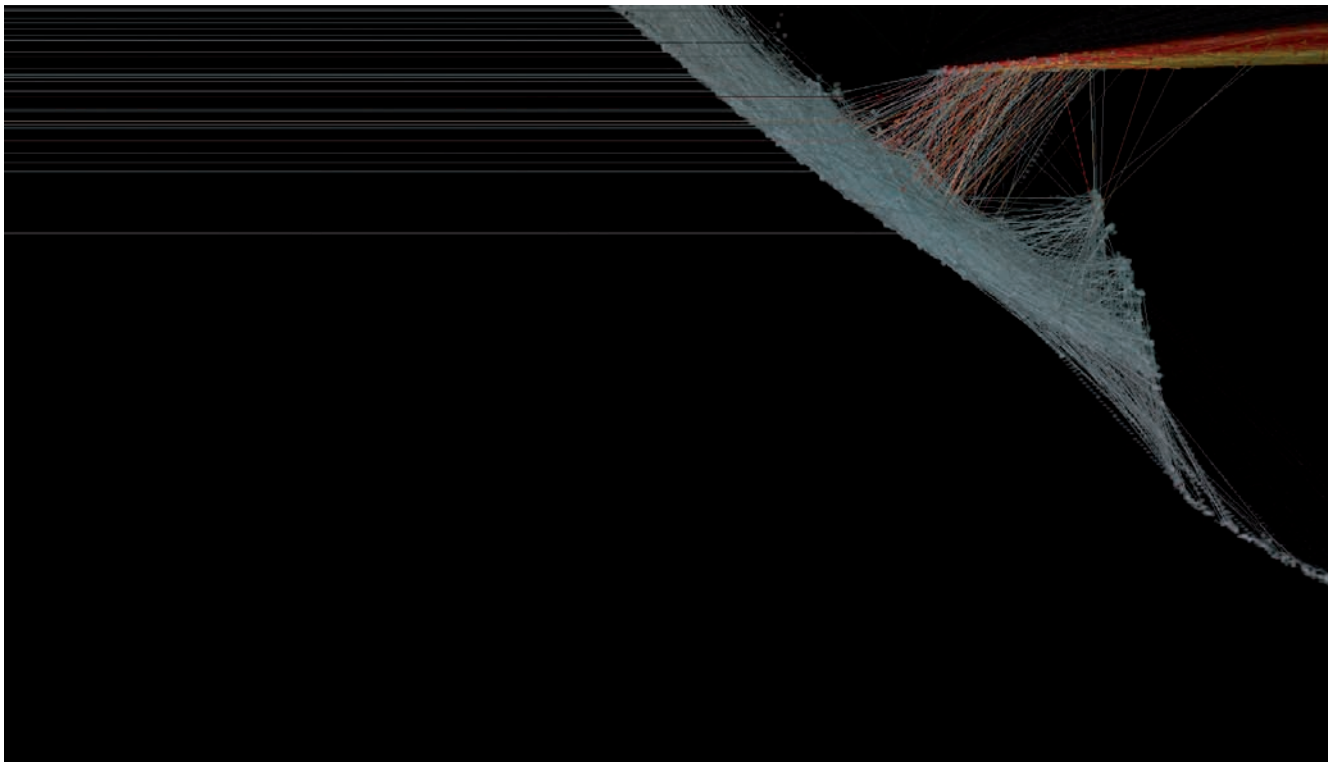
Although it is true that many genetic diseases vary in frequency between populations, these diseases are generally rare and their frequency is unconnected to geographical continental origin, i.e. to racial descent. Let us consider Tay-Sachs disease, cystic fibrosis and thalassemia, three examples of rare diseases that occur with a greater frequency in some populations than in others. Tay-Sachs disease is a rare hereditary disease caused by the mutation of a gene that encodes a protein called Hex-A. Without this protein, neurons accumulate toxic levels of special types of lipid called gangliosides. This build-up affects the central nervous system, usually producing severe physical and mental disability that causes death after just a few years of life. Tay-Sachs disease is common in people of central European Jewish descent, but not among other Europeans nor in people descendent from Sephardic Jews: it is independent of racial descent. Now take cystic fibrosis, a rare hereditary disease produced by the mutation of a gene that encodes a protein called CFTR, which regulates the movement of chlorine and sodium ions through the epithelium membrane. The mutation causes thick secretions that obstruct and infect numerous organs, chiefly the lungs. Its frequency is also independent of geographical continental origin. And lastly thalassemia — a type of hereditary anaemia where the synthesis of one or more of the protein chains that make up haemoglobin is altered — is frequent in various populations from Italy to Thailand. It is not associated with a particular continent or race either.

Although we know very little about the genetic variants that predispose people to common chronic conditions, it is frequently assumed that minority groups — in general Blacks and Latin-Americans — are genetically more predisposed to suffer from any common chronic condition.

The globalization of common chronic conditions suggests, however, that variations in the state of health between populations are due more to differences in exposure to environmental causes than to genetic variations. Without the context provided by variables such as levels of education, socio-economic status, occupation, diet and place of residence, race is not useful in making predictions regarding health. In diverse European countries, the United States, Australia and Canada, there is a socio-economic gradient related with diet whereby people with a higher socio-economic status tend to have a healthier diet, characterized by a greater consumption of fruit, vegetables, skimmed milk, and a lower consumption of fats. Consequently, among women in richer countries, where a slim body is socially valued, there is a negative gradient between weight and socio-economic and educational level. However, for women in countries with other socio-cultural values and/or a medium or low economic level, it is more common to find a positive relationship between economic status and weight. Thus while in the United States or in Spain obesity is divided more or less equally among women and men, in Saudi Arabia, Algeria, Egypt or Russia the number of obese women is approximately double that of men.

There is an equally complex relationship between socio-economic status and other risk factors — such as tobacco and alcohol consumption and a lack of physical activity — associated with some of the diseases which result in higher mortality, including cancer, cardiovascular disease or diseases of the respiratory or digestive system. These data indicate that, taken together, the net effect of racial descent on the health of a person is small, and in no case is it greater than that imposed by economic or socio-cultural factors.

It is metabolism — the set of biochemical reactions taking place within an organism in order to keep it alive — and not genes that has the greatest influence on phenotype. The human genome contains around 2800 genes that encode the enzymes catalysing the biochemical reactions taking place in the human body. Thus, approximately 10 percent of a person's genes are dedicated to metabolism. These biochemical reactions not only allow a living being to grow, be differentiated, maintain its structure and reproduce. They also enable it to respond to environmental changes such as temperature, oxygen concentration, food, or the presence of toxic substances. An organism's metabolism is regulated on multiple levels in order to maintain stable conditions inside its cells regardless of environmental changes: a process known as homeostasis. Homeostatic imbalance is the basis of many diseases, not only in the most common ones such as diabetes or obesity, but also in others such as cancer, immunodeficiency, hepatic conditions and genetic disorders. In the human body, we have identified nearly 5 000 different metabolites, over 75 percent of which are lipids. The largest class of lipids is the phospholipids, which encompass more than 2 000 different molecular species. Following these in complexity are the triglycerides, with over 1 000 different species; then come the cholesteryl esters and the fatty acids, each with over 100 different species.



Ryoichi Kurokawa, *Syn*, 2012

The classical function of lipids is to form the lipid bilayer of membranes and provide energy: the oxidation of a gram of fat produces 9 kilocalories, while carbohydrates and proteins produce just 4 kilocalories per gram. Yet lipids also carry out other essential functions, such as regulating cell proliferation, differentiation and death, as well as oxidative stress and inflammation. The lipids carrying out these functions are known as lipid mediators. Key among the lipid mediators are the eicosanoids, a group of lipids with pro-inflammatory effects derived from arachidonic acid. The importance of the eicosanoids is clear if we consider that the anti-inflammatory drug aspirin works by preventing the production of eicosanoids. Moreover, prostacyclin, another member of the eicosanoid family, is used as a medicine in the treatment of portal hypertension, and as a vasodilator. Other lipid mediators derived from arachidonic acid are the lipoxins and the endocannabinoids. Lipoxins have anti-inflammatory effects, and endocannabinoids are lipids that attach to the same cannabinoid receptors upon which marijuana acts. In addition to having psychoactive effects, the endocannabinoids contribute to the regulation of the immune system and the development of portal hypertension, the chief complication in hepatic cirrhosis. Other lipid mediators are the resolvins and protectins, which derive from linolenic acid, also known as omega-3 fatty acid, and they receive the generic name

of docosanoids. To these we must add the lipid mediators derived from phospholipids, especially the platelet-activating factor or PAF and lysophosphatidic acid, plus those derived from the sphingolipids (ceramide, ceramide 1-phosphate, sphingosine, and sphingosine 1-phosphate) which regulate platelet aggregation, inflammation and anaphylaxis, in addition to the proliferation, differentiation and death of numerous types of cells. Finally there are hydroxycholesterol and oxycholesterol which, as their names suggest, are lipids derived from cholesterol. They also regulate important cell functions.

If we compare the small number of lipid mediators that have been identified (around fifty) with the 4 000 or so lipids present in the human metabolome and about which little or nothing is known of their biological function, it is reasonable to assume that among them are probably tens or hundreds of lipids whose biological properties go far beyond forming cell membrane bilayers or serving as fuel to generate energy. Moreover, the majority of lipid mediators that have been identified are derived from omega-3 and omega-6 fatty acids, and it is therefore likely that there are other unknown lipid families performing important biological functions. The coming decades are likely to see an acceleration in the identification of new lipid mediators, providing the basis for new treatments for a wide range of conditions from obesity, diabetes and cardiovascular disease to cancer and neurodegenerative diseases. In fact, it has recently been discovered that adminisering an unusual phospholipid called dilauroylphosphatidylcholine, or DLPC, has an anti-diabetic effect and improves the condition of fatty liver disease in mice.

Deciphering metabolic changes such as those produced during the growth of a tumour cell, and comparing them with those occurring during normal cell division, could help us develop new therapeutic procedures as yet unexplored. Over recent decades, cancer research has been chiefly dedicated to the study of cell division mechanisms, i.e. to identifying genes and proteins implicated in this process; only recently has more attention been paid to tumour metabolism. Unlike normal cells which generate the energy they need mainly through mitochondrial respiration (a process called oxidative phosphorylation), cancerous cells use aerobic glycolysis to obtain energy, a phenomenon named the “Warburg effect” after its discoverer, the great German biochemist Otto Warburg. Aerobic glycolysis is a less efficient way of generating energy from glucose than oxidative phosphorylation: while aerobic glycolysis generates two ATP molecules per glucose molecule, oxidative phosphorylation generates up to 36 molecules. How is this metabolic reprogramming in tumour cells produced? What is the advantage for tumour cells of having a less efficient metabolism, at least in terms of ATP production, than normal cells? The simplest explanation is that if tumour cells extracted maximum energy from glucose, they would not be able to use the carbon atoms from these molecules to produce biomass — the lipids, proteins and nucleic acids needed for growth. And don’t forget, growth is the sole objective of a tumour cell.

In support of this hypothesis, numerous cases of metabolic pathway reprogramming have been discovered in cancerous cells, with this reprogramming resulting in greater biomass synthesis and reduced energy production. For example, an increase in serine synthesis has been seen in breast cancer cells. Serine is a non-essential amino acid synthesized from 3-phosphoglycerate, a metabolite of glucose. Serine is necessary not only for protein synthesis, but also for the synthesis of another amino acid called glycine. Glycine is in turn needed for protein synthesis, glutathione (one of the chief antioxidant molecules) and purines (essential in DNA and RNA synthesis): in other words, for biomass. Likewise in numerous tumour cells, including breast cancer cells, an increased synthesis and consumption of glycine has been observed. Indeed, we have seen that the flow from 3-phosphoglycerate to glycine is accelerated in various tumour cells. The synthesis of glycine requires the presence of tetrahydrofolate, or THF, a derivative of folic acid or vitamin B_9. It is no coincidence that since the 1950s, antifolates — a group of compounds that inhibit the synthesis of THF, the most well-known of which is methotrexate — have been used to treat various types of cancer, including breast cancer, as well as other conditions such as rheumatoid arthritis and Crohn's disease. Experiments have shown that blocking serine synthesis inhibits the proliferation of breast cancer cells, which seems to indicate opportunities for new "metabolic" therapies aimed at inhibiting serine and glycine synthesis in tumour cells.

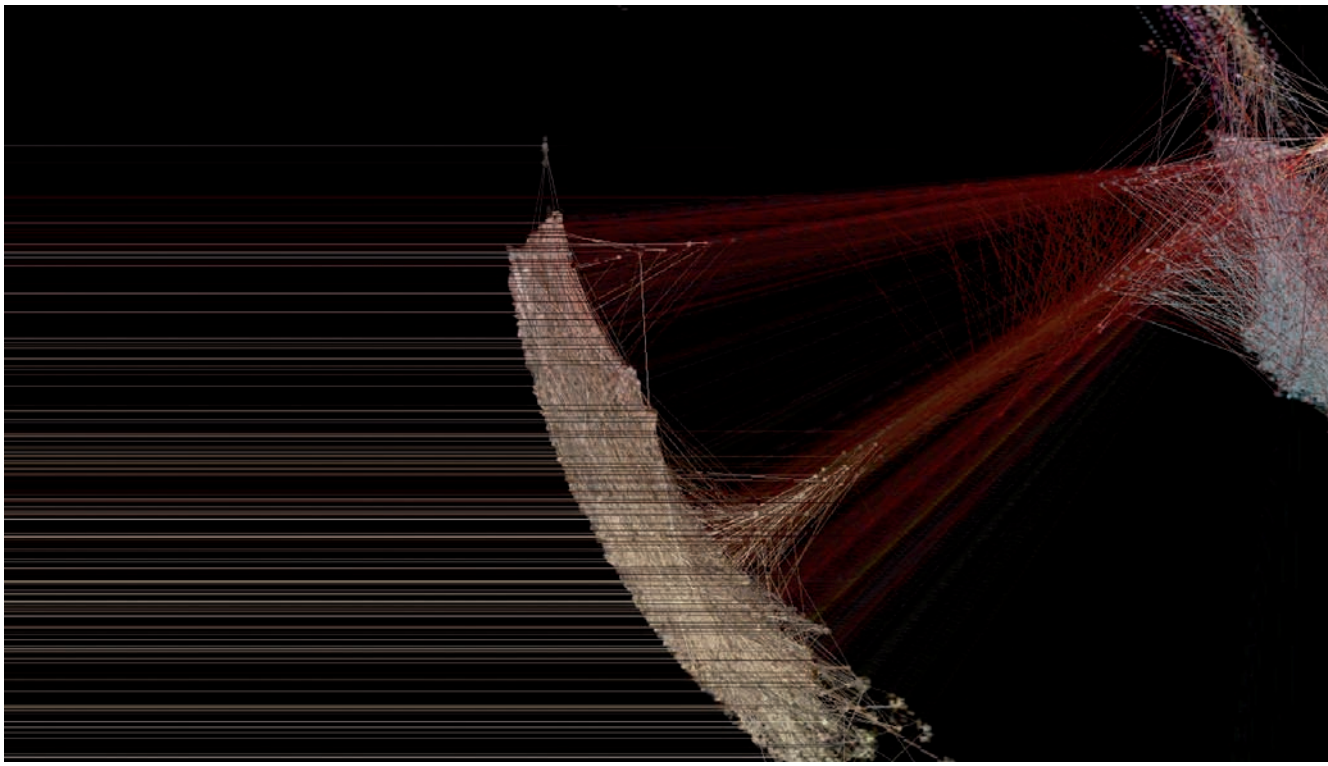
The coming decades are likely to see an acceleration in the identification of new lipid mediators, providing the basis for new treatments for a wide range of conditions from obesity, diabetes and cardiovascular disease to cancer and neurodegenerative diseases

Another interesting case of metabolic reprogramming is that which takes place in liver cancer. The tumour cells synthesize and consume less S-adenosylmethionine (SAMe) than normal hepatic cells. SAMe is a derivative of methionine, an essential amino acid (i.e. we cannot synthesize it ourselves) which is needed for protein synthesis. SAMe carries out various vital functions in the cell. One of them is to donate a carbon atom originating from the methionine (called the methyl group) in a great number of biochemical reactions known collectively as transmethylation reactions. Another function of SAMe consists in channelling the flow of the rest of the carbon atoms from the methionine to the mitochondria in order to be catabolized through an intricate series of biochemical reactions (the transulphuration pathway), thereby generating energy. SAMe synthesis is not the only reaction related to the metabolism of methionine that is reprogrammed in liver cancer cells. Liver cancer also sees a decrease in the

biosynthesis of sarcosine, a reaction that uses glycine and SAMe, and in the decarboxylation of SAMe, the first step in the synthesis of a group of compounds called polyamines. The reduction in SAMe biosynthesis and its role in synthesizing sarcosine and polyamines enables the hepatic tumour cells to redirect the metabolism of methionine towards the synthesis of proteins needed for growth. Experiments have shown the importance of this metabolic reprogramming, demonstrating that the genetic manipulations that block the metabolic axis linking methionine to sarcosine, via SAMe, are sufficient to induce the spontaneous appearance of liver cancer. These results point to opportunities for new therapies that aim to restore the metabolism of methanine in liver cancer. Tests have shown that treatment with SAMe reduces growth in hepatic tumour cells and induces their death.

When not accompanied by malnutrition, calorie restriction — diets that reduce the number of ingested calories — improves people's general health (including blood pressure, glucose, triglycerides and cholesterol) and increases longevity in many animal species including rodents. Although the effect of calorie restriction on longevity in primates, including humans, is open to debate, these studies are important because they can provide information regarding diet strategies that improve general health without the side effects produced by excessive weight loss. For example, when *Drosophila* flies are submitted to calorie restriction they live longer, even though their fertility diminishes (one of the side effects of excessive weight loss). But if methionine is added to the diet this side effect disappears. In mice, the restriction of methionine in the diet also increases longevity, while a severe deficiency of this amino acid produces malnutrition, fatty liver, inflammation, fibrosis, and liver cancer. These results indicate that methionine acts as a metabolic rheostat in mice, and both excessive and insufficient levels of this amino acid can affect health. This work also demonstrates that not all calories are the same and that what matters is the type of feeding. An understanding of how a diet restricted in certain nutrients can improve general health or increase longevity could help us identify therapeutic targets that offer the beneficial effects of calorie restriction with minimal side effects.

When we speak of metabolism, we immediately think of the thousands of biochemical reactions taking place within a person's body, but we usually forget the metabolic reactions performed by the micro-organisms living in the intestine. The intestinal microbiome is made up of 10^{13} to 10^{14} micro-organisms, whose collective genome contains at least 100 times more genes than our own genome. This microbiome contains thousands of genes implicated in the synthesis and metabolism of polysaccharides, carbohydrates, amino acids and lipids, as well as other compounds such as vitamins, isoprenoids and methane. Many of these molecules pass through the intestinal wall and circulate freely in the blood, together with those metabolites synthesized endogenously. In other words, human metabolism is a fusion of its endogenous metabolism with that of the micro-organisms living in its intestine. The case of trimethylamine



Ryoichi Kurokawa, *Syn*, 2012

(TMA) is a good example of this metabolic fusion. Intestinal flora — not endogenous human metabolism — can convert choline from phosphatidylcholine (the most common phospholipid in our diet) into TMA. In turn, a group of hepatic enzymes known collectively as FMOs oxidise TMA to TMA oxide, or TMAO. And TMAO activates the synthesis of cholesterol in macrophages which, in mice, promote the appearance of plaques in the arteries and the development of atherosclerosis. Although the impact of the microbiome on human health is subject to debate, changes in microbiome composition have been associated with obesity and certain inflammatory diseases. Therefore, the study of the microbiome and its metabolome provides enormous potential in the search for new diagnostic biomarkers and treatments.

The rapid advances in mass spectrometry and nuclear magnetic resonance technologies have made it possible to quickly and simultaneously determine and quantify around 1 000 different metabolites from a serum sample of a few microliters. In all probability, it will soon be possible to determine the complete metabolome — some 5 000 distinct compounds. How can studying the metabolome improve the health of the individual? The human metabolome is an ocean of biomarkers. In the future, studying the metabolome in biological fluids — serum, urine, faeces, sweat, tears — and tissue samples will provide biomarkers to

assist not only in the early diagnosis and prognosis of numerous complex diseases, but also in monitoring each individual's response to treatment. Of course, as we have established, there is no point in determining the concentration of thousands of phospholipids, triglycerides, cholesteryl esters and fatty acids in the serum or other biological sample from a person if the data are not then interpreted within an adequate medical context. The experience of diagnosing non-alcoholic fatty liver disease (NAFLD) teaches us that considering aspects that at first glance may appear of little relevance, such as the body mass index,[2] can be decisive when using metabolomics as a diagnostic tool. NAFLD is a progressive disease that ranges from the simple accumulation of fat in the liver (steatosis) to steatosis with inflammation, necrosis and fibrosis, a condition called non-alcoholic steatohepatitis, or NASH. Patients with NASH have a greater risk of developing cirrhosis and cancer of the liver: between 10 and 20 percent of people with NASH develop cirrhosis within 10 years. NAFLD affects 1 in every 4 adults in developed countries, and approximately 30 percent of those with NAFLD have NASH. Although the majority of people with NASH are obese, this disease is also found in slim people, but in a lesser proportion. Few people with NASH are aware they have the condition, as it presents few or no symptoms. Imaging techniques such as ultrasound or magnetic resonance can reveal the presence of fat in the liver, but they cannot differentiate NASH from simple steatosis. The hepatic biopsy, while considered the gold standard, is an invasive, subjective and costly procedure that is not exempt from complications (it carries a risk of death of 0.01 percent) and is prone to sampling errors. Due to these limitations and an increase in the prevalence of NAFLD (it is the most common liver condition in developed countries), the identification of biomarkers that allow the diagnosis of this disease is urgent and necessary. The study of lipidomics has identified a small group of lipids that allow us to distinguish between NASH and steatosis using a serum sample. Surprisingly, tests have shown that this lipid profile — the footprint that serves to diagnose NASH — is dependent on BMI: the lipids that serve to differentiate between NASH and steatosis are different in slim, obese or morbidly obese people. Therefore, carrying out routine metabolome testing to diagnose NASH in people who have no symptoms and who are not preselected using BMI or within an adequate clinical context could lead to false diagnoses.

The development of complex molecular tests based on DNA or RNA profiles, proteins or metabolites, carries a series of problems inherent to high-performance techniques where huge quantities of data are analysed. When thousands of molecules are examined using a relatively small number of patient samples, it is easy to find false correlations with the diagnosis of a certain condition or with the effectiveness of a treatment. Some years ago a group of

2. BMI, obtained by dividing the weight in kilos by the square of the height in metres.

researchers from the University of Duke in North Carolina published various works that concluded that certain genetic expression profiles — genome fingerprints — could predict the response to chemotherapy in a diverse range of cancers. As a consequence of this research, the university set up a clinical trial in which these genome fingerprints were used to guide the clinician in the treatment to be followed. Years later, various statistics experts found errors in the treatment of the published data, which invalidated the conclusions drawn. In the end the clinical trial was suspended, and some of the patients who had taken part in the trials, and families of patients who had died, sued the University of Duke and the researchers involved in the work. This case demonstrates how it is possible to be too hasty in taking tests based on genome research to the clinic and to the market. Inevitably, if we search among thousands of variables, we will find a group that correlates significantly with these thousands of variables plus one, such as responding or not to a certain treatment. In 2010, a genome study carried out in centenarians concluded that certain genetic variants were associated with exceptional longevity. This work had widespread and immediate repercussions in the media, but at the same time it was received with scepticism by geneticists. It was certainly an unexpected discovery to find that particular genetic variants had such a marked effect on as complex a characteristic as longevity. As was seen later, the results of this study were incorrect because its authors had not demonstrated sufficient rigour in obtaining the data and extracting their statistics. One year later, the authors retracted the conclusions of the work. These and other cases remind us of the importance of designing experiments correctly, avoiding technical errors, validating genome fingerprints using new blind samples (so that the researcher does not know in advance which samples belong to the patients who respond to the treatment) from different hospitals, and using adequate statistical tools. Genomics is not the only branch of life sciences that has quality-control problems, and works are all too frequently published in which massive quantities of data containing avoidable errors are handled.

In the future, studying the metabolome in biological fluids and tissue samples will provide biomarkers to assist not only in the early diagnosis and prognosis of numerous complex diseases, but also in monitoring each individual's response to treatment

Scientific progress is underpinned by two firm beliefs: 1) that the checking by equals, or peer review, is the touchstone for determining the quality, credibility, and scientific rigour of a work or research project; and 2) that research corrects itself — scientific errors are always identified and corrected. What we usually forget is that self-correction in research does not just depend

on competition between researchers, but also on the connection between research and applications. In other words, if during the process of knowledge generation and translation — both in the academic world and in business — the maximum standards of quality, rigour and ethics which science demands are not met, then the results of this badly done research will never lead to new products and applications. There is evidence that the bias in scientific research in recent decades is not random: data show that the two pillars that support scientific progress — peer review and self-correction — are not working as well as they should. There is more tolerance of badly done science and, while there are many factors that contribute to explaining the high failure rate in translational research, this behaviour makes a negative contribution to the attainment of new diagnostic tests and medicines.

All high-performance technologies generate errors and systematic biases which to the unwary, untrained eye, or to one unfamiliar with the technology, could appear to be very interesting results, thereby leading to erroneous conclusions. Minimizing the frequency of these technical errors in the design of experimental protocols and in data analysis is a mission for all researchers, but mainly for the directors of doctoral thesis research, chief researchers and project heads, in the academic world and in business. High-performance techniques or "omics" have the potential to revolutionize medical practice in the coming decades, but it is increasingly clear that these problems will have to be addressed beforehand.

The fact that not all people with an identical mutation develop a certain disease shows that biological reality is complex; there are compensatory mechanisms and it is necessary to distinguish between a person's genome (genotype) and his or her traits (phenotype). When speaking of genes and genomes, we are not speaking about an objective physical reality — that is to say, a phenotype — but about information. Moreover, this fact reveals that biology is not deterministic nor does it rest on the principle of specificity. This vision allows an understanding of how multiple phenotypes can be formed from a single genome, and how environment and chance select, at each moment, one from among all those possible phenotypes. From this perspective, it is evident that disease is individual and that medicine cannot be predictive, but must adapt to the person, working with specific cases and not statistics.

BIOGRAPHY

José M. Mato

CIC bioGUNE

José M. Mato (Madrid, 1949) studied Chemistry at the University of Madrid and obtained his doctorate from the University of Leiden. He has developed his professional activities in the Jiménez Díaz Foundation, the Biozentrum at the University of Basilea, the University of Chapel Hill in North Carolina, the National Institutes for Health (NIH) in Bethesda, the University of Pennsylvania, the Thomas Jefferson University in Philadelphia, the Higher Council of Scientific Research (*Consejo Superior de Investigaciones Científicas*, CSIC) and the University of Navarra. He has been President of CSIC (1992–96) and is currently Director General of bioGUNE (Bilbao) and biomaGUNE (San Sebastián).

In 2004 he was awarded the National Prize for Medical Research for his work on the interrelation between metabolism and methionine, fatty liver and hepatocellular carcinoma (HCC). His research work led him to identify a group of genes implicated in the synthesis and catabolism of S-adenosylmethionine, and to outline its function during the accumulation of fat in the liver, cirrhosis and the development of HCC, as well as in other normal processes such as hepatic regeneration and differentiation.

José M. Mato is the author of over 250 publications, and the holder of various patents. His work has been recognized with many distinctions, including the CJ Kok Prize for excellence in doctoral work from the University of Leiden (1979); the GB Morgagni Medal for research in the area of diabetes and metabolic diseases (1989); the Lennox K Black International Prize for excellence in biomedical research, from Thomas Jefferson University (1994); and the Jiménez Díaz Commemorative Lecture (2011). He has participated in the creation of various companies, particularly OWL, a biotechnology company based on metabolomics.