Personalized Prognosis and Diagnosis of Type 2 Diabetes – Vision or Fiction?

Günter Müller

Department of Biology I, Genetics, Biocenter, Ludwig Maximilian University, Munich, Germany

Abstract

Typical civilization diseases, such as type 2 diabetes, share several features: their worldwide frequency, the complexity of the underlying pathogenic mechanisms, heterogeneity in the phenotypes and their multifactorial nature due to a wide variety of possible combinations of disease susceptibility or protective genes in different tissues and negative or positive environmental factors. This is in sharp contrast to classical inherited diseases, such as Huntington's chorea, which are often caused by complete loss- or gain-of-function mutations in a single gene. The causative polymorphisms of susceptibility genes, however, are characterized by relatively subtle alterations in the function of the corresponding gene products, i.e. low penetrance and effect size, which do not support the pathogenesis per se, and by their high frequency; these two characteristics result in high expenditures for their identification and a rather low predictive value. In the future, the reliable and early diagnosis of common diseases will thus depend on the determination of all (or as many as possible) polymorphisms of each susceptibility gene together with the corresponding gene products and the me-

tabolites emerging thereof for each individual. Great hopes are currently associated with systems biology to cover these demands in time (i.e. along the pathogenesis) and space (i.e. in all relevant tissues). Copyright © 2010 S. Karger AG, Basel

Introduction

Personalized medicine – or the use of information about a person's genes, proteins, metabolites and environment to prevent, diagnose and treat disease – has been much talked about in recent years. And some observers are wondering what the excitement is all about, cumulating in the following statement: 'Personalized healthcare is nothing new. Doctors have always tried to fit the therapy to the patient if possible. But what's happened more recently is that we have begun to go a level deeper. We are now exploring the biology of disease and treatment at the molecular level.' Molecular medicine does not per se define personalized medicine, but molecular tools are important as they should enable greater relevance of the information provided by diagnostic tests. As 'personalized' means different things to different people, additional complementary ways of characterizing diagnostics may further help distinguish different shades of gray in the

Dr. Günter Müller
Sanofi-Aventis Pharma Germany, TD Metabolism, Industrial Park Hoechst, Bldg. H821
DE–65926 Frankfurt am Main (Germany)
Tel. +49 69 305 4271, Fax +49 69 305 81901
E-Mail Guenter.Mueller@sanofi-aventis.com

spectrum of personalized medicine. In the strictest sense, personalized-medicine diagnostics may consist exclusively of companion diagnostics, which are by definition geared towards supporting a therapy decision for a particular drug, patient by patient. At a more permissive end of the spectrum, personalized-medicine tests may include: (1) early diagnostics, which encompasses diagnostic products permitting the detection of a disease at very early stages of its development and thus gives more treatment options; (2) prognostics, which means diagnostics that predict or estimate the risk of developing a particular condition based on phenotypic (e.g. transcriptomic, proteomic or metabolomic) parameters or genomic (e.g. hereditary or gene-based) characteristics, and (3) all other types of diagnostics. One may indeed argue that if a diagnostic were not designed to inform treatment decisions for individual patients – one way of defining a personalized-medicine diagnostic – it would not make much sense.

Drug approval agencies, including the Federal Drug Administration (FDA) and the European Medicines Agency (EMEA), are encouraging greater use of biomarkers and diagnostics in drug development and prescribing decisions, thus promoting the concept of companion diagnostics for drugs. The FDA has recently started reporting a list of genomic biomarkers that it considers valid to guide the appropriate clinical use of approved drugs. The list is being updated on a quarterly basis and counted 32 valid genomic biomarkers in mid-September 2009. Most drug labels in the list provide pharmacogenomic information with no immediate recommendation for genetic testing. However, testing is 'recommended' or 'required' in a few cases. On March 20, 2009, four biomarkers were 'required' to be tested – three for cancer, and one for infectious disease indications.
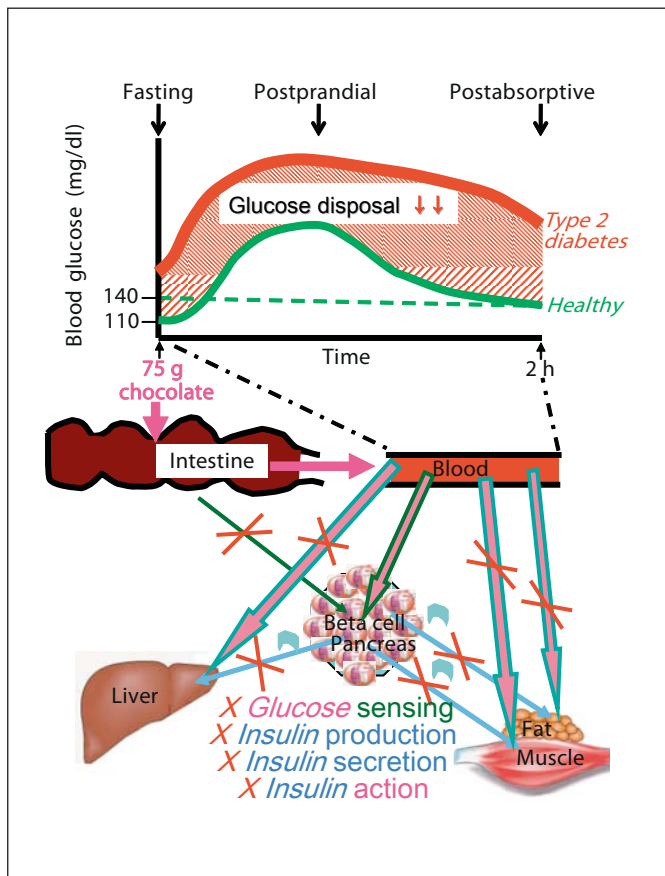
We are still at the start of the process if we consider that only four biomarkers are 'required' to be tested. However, the FDA was prompted to publish its list following a marked increase in the number of approved drug labels containing pharmacogenomic information over the last decade. The FDA estimates that 10% of approved drug labels now contain pharmacogenomic information and this is expected to continue increasing. The EMEA's communication on the requirement for biomarker testing is less transparent than the FDA's, but its initiatives should not be overlooked. For example, the EMEA played a key role in requiring biomarker testing for Amgen's Vectibix (active principle panitumumab, human monoclonal antibodies against epidermal growth factor receptor), following the FDA's accelerated approval without specific testing requirements. The EMEA also requires biomarker testing for a larger number of drugs: in mid-2009 accounting for at least 11 such drugs. Greater harmonization between regulatory agencies has to be developed over time through greater consultation but also following pressure from clinician communities as stakeholders in one country push to implement practices already included in drug labels in other countries.
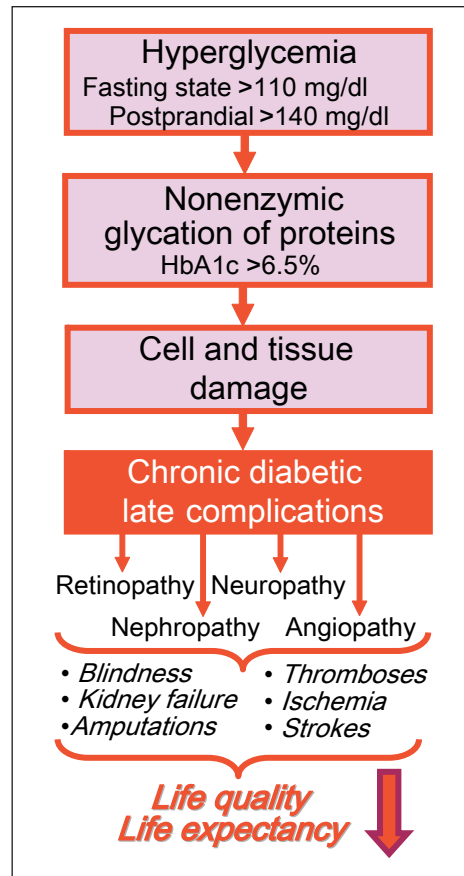
Type 2 Diabetes – The Pathogenesis of a Common Disease

Meanwhile, 5–10% of the individuals over 40 years of age in the Western world will suffer from diabetes during their lifetime, with increasing tendency. Most of them (about 95%) suffer from type 2 diabetes with late-onset clinical manifestations in most cases. A minor portion is affected by type I diabetes, often fully developed already between the 15th and 25th year of life and most likely completely different from that of type 2 diabetes [1, 2]. Due to the serious and long-lasting consequences for the individual's fate and welfare, the reliable and early diagnosis as well as the efficacious and safe therapy of type 2 diabetes is of tremendous importance for the patients as well as for the national health care systems.

Frank type 2 diabetes is characterized and caused by excessively high blood glucose levels already in the fasting state and even higher levels in the postprandial state (fig. 1). After consumption of carbohydrate-rich food (e.g. 75 g chocolate), the glucose absorbed by intestinal cells is released into the circulation, resulting in the transient increase in blood glucose levels for about 1 h. Thereafter, the glucose is removed from the blood into the peripheral tissues, predominantly muscle, fat and liver (fig. 1), and stored after conversion to glycogen and lipid, resulting in a decline in blood glucose levels to almost fasting levels after 2 h. In the course of type 2 diabetes, utilization of the blood glucose (glucose disposal) by the peripheral tissues is significantly impaired and, consequently, glucose accumulates in the circulation. Initially, elevated blood glucose is not accompanied by clinical insults, one could argue – unfortunately – given the missing chance of early diagnosis, before the manifestation of irreversible pathogenic mechanisms. However, elevated blood glucose leads to damage to certain proteins from the very beginning. During exposure to high glucose concentrations, a variety of extra- and intracellular proteins – in particular long-lived serum, matrix and cytoskeletal proteins – become covalently coupled to glucose

Fig. 1. Regulation of blood sugar and major organs and physiological mechanisms involved.

Fig. 2. Pathogenesis of type 2 diabetes with its late complications.

and some of its derivatives (fig. 2, hyperglycemia). This spontaneous nonenzymic glycation of proteins, which is diagnostically monitored with the biomarker hemoglobin A1c (HbA1c), impairs their function irreversibly. Over the years (5–25 years), relevant cells and tissues will become increasingly damaged and trigger diabetic late complications. These predominantly affect the eyes (blindness), kidneys (renal failure), peripheral nervous system (pain), brain (mental problems), small vessels (atherosclerosis, ischemia, stroke) and large vessels (thromboses, amputations) (fig. 2). Despite intensive efforts in diagnosis and therapy, type 2 diabetes is currently still associated with considerable reductions in life quality and life expectancy.

What are the reasons for the reduced glucose disposal by the peripheral tissues, which ultimately results in diabetic late complications caused by elevated blood glucose levels? For proper glucose disposal, the peripheral tissues depend on endocrine signals, in particular insulin, the blood glucose-regulating hormone (fig. 1, arrows). The pancreatic beta cells function as sensors for blood glucose and respond with the production and secretion of insulin. In addition, the beta cells receive direct hormonal information (e.g. glucagon-like polypeptide-1) on the amount of glucose just taken up with the food from special intestinal cells, which also operate as glucose sensors (fig. 1, arrow). The insulin released by the beta cells into the circulation tells the muscle, adipose and liver cells to take up the glucose and convert it to glycogen and/or lipids. This metabolic insulin action requires a multitude of components for intracellular signaling and the terminal transport steps and enzymic pathways [1–3]. Defects in those components of glucose sensing, insulin production and secretion, or insulin action in the glucose-sensing or utilizing cells will inevitably result in inadequate and delayed glucose disposal from the circulation and concomitantly impaired glucose tolerance. Thus impaired glucose disposal may be caused by defects (fig. 1, crosses) in
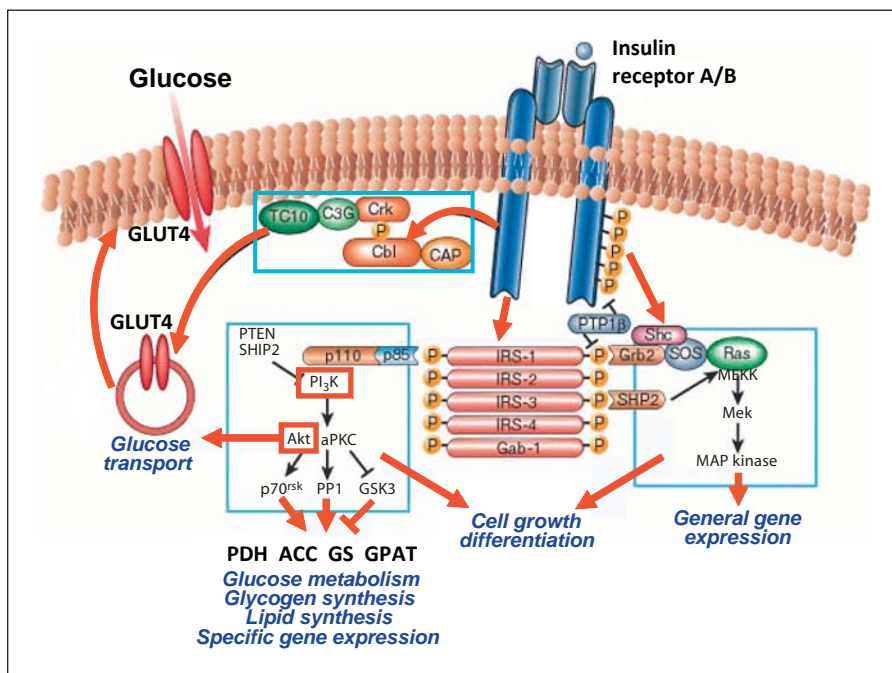
Fig. 3. Molecular mechanisms for the stimulation of glucose utilization by insulin in muscle, liver and adipose cells. The binding of insulin to the insulin receptor triggers autophosphorylation and activation of the intrinsic tyrosine kinase which in turn phosphorylates the adaptor proteins IRS1–4 and Gab-1. Their interactions with signaling proteins activate at least three signal transduction cascades, such as the Cbl/CAP, PI_3K/Akt and Ras/MAP kinase pathways, which control the vesicular trafficking of glucose transporter protein-4 (GLUT4) from intracellular vesicles to the plasma membrane, the activity of rate-limiting enzymes of glucose and lipid metabolism, such as pyruvate dehydrogenase (PDH), acetyl-CoA carboxylase (ACC), glycogen synthase (GS) and glycerol-3-phosphate acyltransferase (GPAT), and the specific and general gene expression. Together, this leads to the uptake of glucose and its metabolism to glycogen and lipid in response to insulin in relevant glucose-utilizing tissues upon their adequate growth and differentiation, which is also regulated by insulin-dependent gene expression. Only some of the major signaling and terminal effector proteins are shown.

the beta cells or intestinal cells affecting glucose sensing, in the beta cells affecting insulin production and secretion and in the muscle, adipose and liver cells affecting insulin action. Each of these processes depends on a multitude of genes and gene products including signaling proteins and terminal effector proteins, such as the glucose transporter, glucokinase, ATP-dependent potassium channel (K_{ATP} SUR1/Kir6.2) for glucose sensing by the beta cells and insulin itself for insulin production by the beta cells as well as the insulin receptor with its isoforms A and B, insulin receptor substrate proteins 1–4 and Gab-1), glucose transporter protein-4, pyruvate dehydrogenase, acetyl-CoA carboxylase, glycogen synthase and glycerol-3-phosphate acyltransferase for insulin action in the muscle, adipose and liver cells (fig. 3). It is currently estimated that 5–10% of the human genes, i.e. 1,500–3,000, are directly or indirectly involved in glucose disposal and regulation [4]. In principle, each of these

genes is a candidate for a so-called disease gene and/or susceptibility gene for type 2 diabetes, i.e. functional alterations may lead to impaired glucose disposal.

Disease Genes

Any variant (i.e. polymorphism or mutation) of a (putative disease or susceptibility) gene coding for a protein may affect its promoter, exon, intron or terminator areas with varying functional consequences involving different molecular mechanisms (single nucleotide exchange, duplication, insertion, translocation, inversion, deletion) (fig. 4). If they are located in exons or at exon-intron boundaries, they will directly affect the structure of the protein in most cases and thereby lead to drastically reduced or, in rare cases, increased activity. If they are located in introns or promoter regions, they will cause
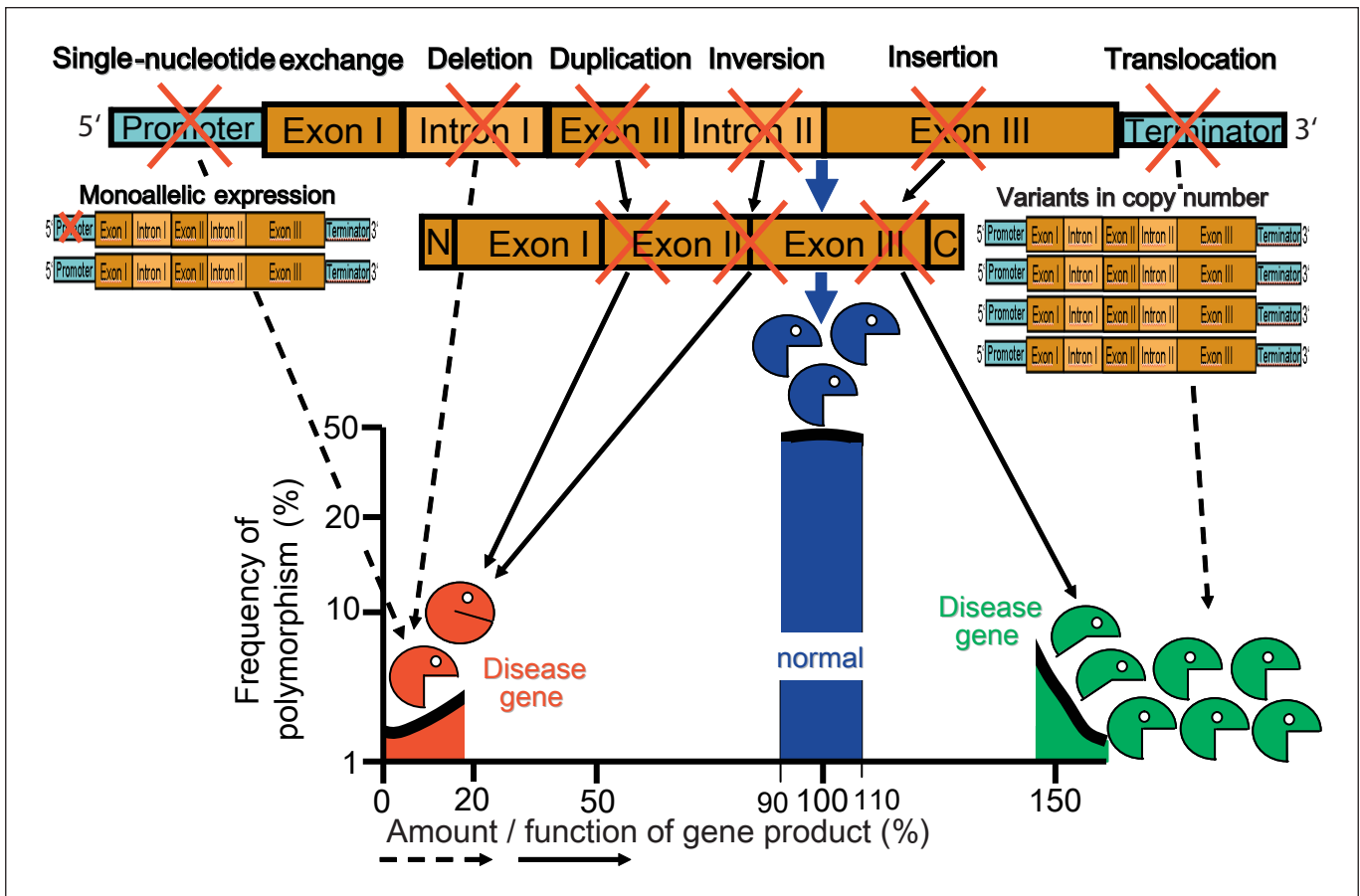
Fig. 4. Molecular mechanisms, frequency and functional consequences of disease genes.

drastically reduced or, in rare cases, increased protein synthesis. Moreover, monoallelic expression and variants in the copy number of genes may contribute to an enormous diversification of the human genome with corresponding phenotypic consequences (fig. 4). With the complete description of all integration sites of the most frequent copy number variants and the introduction of novel methods for their identification and characterization in the total genome, it should be possible to evaluate the physiological relevance of copy number variants of disease genes (and susceptibility genes, see below) to the pathogenesis of type 2 diabetes. All these polymorphisms and mutations may already be manifested in the germ line cells and may thus be inherited or occur in a subset of somatic cells during life.

On the basis of our knowledge of the molecular mechanisms of blood glucose regulation, it is tempting to speculate on polymorphisms or mutations in single but essential genes for glucose disposal that represent the primary cause of type 2 diabetes. In fact, it has been known for more than 100 years, since Sir Archibald Garrod's *Inborn Errors of Metabolism,* that the complete defect of a single gene product, for instance phenylalanine hydroxylase, can lead to severe metabolic diseases, such as phenylketonuria. Subsequently, typical mendelian inheritance, i.e. monogenic inheritance, has been delineated for a number of human diseases from family studies [5, 6]. There has been rapid progress in defining the etiological genes for monogenic diabetes, which reflects the relative simplicity of gene discovery in single-gene disorders. The candidate gene approach has been remarkably successful in defining monogenic diseases. This reflects that key rate-limiting steps in insulin secretion and action are known, and severe mutations affecting these processes and the underlying proteins will result in beta cell dysfunction or insulin resistance. Examples of this approach include the

genes encoding insulin [7], glucokinase [8, 9], the two subunits of K_{ATP} Kir6.2 [10] and SUR1 [11, 12], the peroxisome proliferator-activated receptor PPARγ [13], and the insulin receptor with its isoforms A and B [14]. Finding human subjects with mutations in these candidate genes has allowed confirmation of a critical role in humans of the encoded protein, helped define the structure and function of the protein and allowed confirmation of the associated pathophysiology, such as abnormal glucose sensing in glucokinase mutations [15]. In fact, single mutations in the genes coding for insulin receptor A/B or glucokinase have subsequently been identified in patients with severe phenotypes, so-called leprechaunism and maturity-onset diabetes of the young type 2 (MODY2) [3, 4]. In addition to the missing regulation of the blood glucose, the affected patients suffer from severe growth and developmental retardation. The presence of one copy of the polymorphic/mutant gene coding for the insulin receptor or glucokinase suffices for both diseases to manifest themselves, i.e. they are dominantly inherited. However, the candidate gene approach has not led to the identification of novel pathways in glucose homeostasis.

Completely unexpected pathways critical for insulin secretion and action were identified by pedigree analysis of the association of a phenotype with a genetic marker within affected families or populations, i.e. positional cloning of novel monogenic type 2 diabetes genes. The most striking example was the identification of the transcription factor, hepatic nuclear factor 1A *(HNF1A)*, as the gene linked to the phenotype of MODY and mapped to chromosome 12q [16]. Before this finding, it was not known that *HNF1A* was expressed in the beta cell, and diabetes had not been noticed in the *hnf1α*-knockout mouse [17], although it was noticed subsequently [18]. This result rapidly led to the detection of mutations in other hepatic transcription factor genes, *HNF4A* [19] and *HNF1B* [20], also shown to cause MODY. These findings have led to a totally new area of beta cell biology seeking to explain why haploinsufficiency of these genes resulted in progressive beta cell dysfunction [21, 22]. Other monogenic defects in glucose utilization, such as leprechaunism, are recessively inherited, i.e. they only become manifest in the presence of both maternal and paternal copies of the disease gene, such as mutations in the insulin receptor. Mutations in the gene encoding the lipolytic enzyme carboxyl-ester lipase, which is responsible for the hydrolysis of cholesterol esters, were also unexpectedly found to cause distinct MODY phenotypes through positional cloning [23]. Carboxyl-ester lipase is only expressed in the pancreatic acinar cells, so beta cell dysfunction was unexpected. Further studies of the mechanism will clarify the close relationship between the exocrine and endocrine pancreas. Following mapping to chromosome 1q21, familial partial lipodystrophy was shown, to arise from mutations in the *LMNA* gene, encoding lamin A/C [24]. Mutations in *LMNA* can also result in myopathy, dilated cardiomyopathy, or atypical progeria [25]; the biology of how these mutations alter fat distribution is still incompletely understood. In addition, for many genetic syndromes, such as Wolcott-Rallison and immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndromes, a discrete cluster of clinical features including diabetes was initially recognized as a clinical syndrome and subsequently the responsible gene was identified [26, 27]. In these cases, discovery of the gene gave new biological insights but only limited insights into the phenotype. Nevertheless, taken together positional cloning has led to exciting novel pathways of glucose homeostasis.

A key result has been that the vast majority of genes single mutations of which cause early-onset diabetes exert reduced beta cell function rather than increased insulin resistance. Heterozygous haploinsufficiency results in dominant early-onset diabetes for many beta cell genes, including *GCK, HNF1A, HNF4A,* and *HNF1B,* but this is not seen in insulin resistance genes. This shows that even when faced with severe insulin resistance, a healthy beta cell is usually able to compensate, but no compensation is possible in the presence of marked insulin deficiency. There are many mechanisms of beta cell dysfunction in monogenic diabetes, including reduced beta cell development, failure of glucose sensing, and increased destruction of beta cells. Importantly, individuals with type 2 diabetes typically display concomitant defects in both insulin secretion and action. While it is axiomatic that hyperglycemia implies some degree of relative or absolute failure of beta cell function, there has been a long-standing debate about the relative importance (even 'primacy') of the two processes in the pathogenesis of type 2 diabetes. Notwithstanding the efforts of epidemiologists and physiologists, this may be one debate where genetics in general and positional cloning in particular (precisely because of the focus on inherited rather than acquired phenomena) may provide the answers. The relative prevalence of mutations causing monogenic forms of diabetes suggests that mutations in beta cell-related processes are a more frequent cause of severe early-onset diabetes than mutations influencing insulin action. Studies of the relative heritability of indexes of beta cell function and insulin action in the gen-
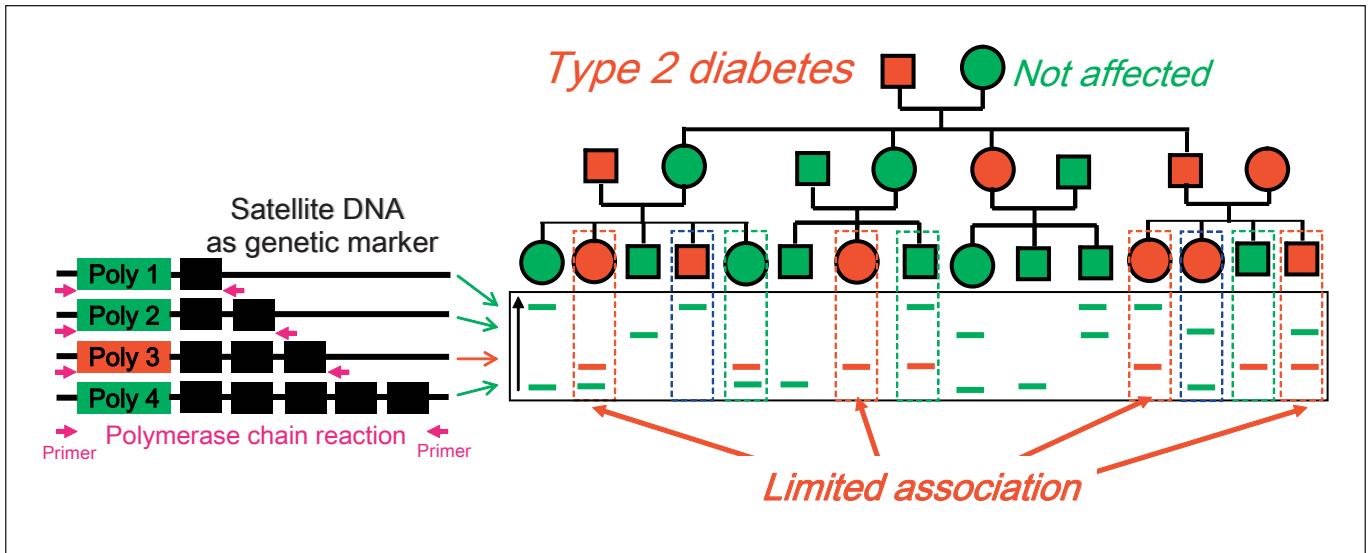
Fig. 5. Identification of disease and susceptibility genes by positional cloning.

eral population also suggest a preponderance of beta cell effects [28].

However, the drastic and relatively homogeneous phenotypes with rather early-onset and complete or at least partial independence of the environment, as in leprechaunism and MODY, is characteristic for disease genes, such as mutations in the *IRA/B* and *HNF4A* genes. The underlying (almost) complete absence in the function or synthesis of the relevant gene product (fig. 4) is, however, very rare and apparently not causative of typical type 2 diabetes, with its frequent and late manifestations and its rather moderate, nevertheless harmful phenotype of considerable heterogeneity and dependence on the environment. Monogenic and syndromic forms account for only a small though highly informative proportion of cases of nonautoimmune diabetes. The challenge for medical science lies in bringing equivalent mechanistic insights and translational benefits to the hundreds of millions of people already affected by type 2 diabetes or at risk. There is abundant evidence that individual susceptibility is influenced by both the combination of genetic variation at multiple sites and a series of environmental exposures [28]. Tracking down the specific genetic variants involved has been more difficult than for monogenic forms of the disease since the correlations between genotype and phenotype are far weaker.

Susceptibility Genes

In fact, most patients with type 2 diabetes harbor fully functional insulin receptors, glucokinases or HNF transcription factors and transmission from generation to generation apparently does not obey mendelian laws. Pedigree analysis typically reveals (fig. 5) that healthy parents have affected offspring and vice versa. Nevertheless, an increased frequency of type 2 diabetes can often be derived from most pedigrees of affected twins, families and populations. The concordance rate for type 2 diabetes is about 70% for monozygotic and 20–30% for dizygotic twins in comparison to 10% as commonly observed for Western populations. The individual risk for the development of type 2 diabetes in a lifetime accounts for about 40% in case of individuals with either an affected father or an affected mother, with a trend to a higher risk in case of an affected mother; the risk approaches 70% in case of both parents suffering from the disease. Moreover, first-degree relatives of type 2 diabetics are faced with a three-fold elevated disease risk compared to nonrelatives. The pronounced ethnic differences in the frequency of type 2 diabetes, which have been known for decades, may also be indicative of a genetic component. Studies of large series have shown that the known type 2 diabetes susceptibility variants (see below) influence clinically relevant phenotypes, such as disease progression [29], risk of complications and therapeutic response [30]. However, it does not follow that those differences

suffice to provide clinically information relevant to individual patients. Indeed, the modest effect sizes of the variants identified to date mean that their individual impact is likely to be limited. This is best illustrated by considering variants in *TCF7L2* [30, 31]. Genome-wide association (GWA) studies have demonstrated that variants of this gene have the strongest effect on diabetes risk currently known (see below). Assuming an average lifetime risk of type 2 diabetes of about 10%, this percentage would decrease to about 7.5% in an individual with no copies of the risk allele, whereas the lifetime risk of an individual with two copies would increase to 14.5%. Taken together, there is no doubt about a strong genetic background for type 2 diabetes, i.e. about the operation of susceptibility genes in the affected twins, families and populations during its pathogenesis. This holds true for the majority of the typical complex and multifactorial 'civilization' diseases, among them cardiovascular and (neuro)degenerative diseases as well as the most common types of cancer. During the past 50 years of the worldwide epidemic increase in the number of type 2 diabetic patients, our environmental conditions have changed to the typical Western lifestyle of permanent food availability and physical inactivity. Our genes, however, have not changed to any significant degree during this period. This does not exclude an important role of the genes involved in glucose utilization and of their polymorphisms in the current, rapidly increasing prevalence of type 2 diabetes since they ultimately determine the response of our organism to the altered lifestyle.

Identification of Disease and Susceptibility Genes

The traditional method for the identification of disease and susceptibility genes relevant to the rare monogenic and common polygenic diseases is linkage analysis. However, the chromosomal regions identified by linkage analysis of several thousand families, which segregate the relevant gene locus, typically exceed 30 cM in length and may contain several hundred genes. In the past, those extended chromosomal regions were associated with the pathogenesis of certain multifactorial diseases, such as cancer and asthma, in the course of linkage analyses using pronounced structural chromosomal aberrations, such as deletions, inversions, duplications or translocations. Subsequently, positional cloning was used for fine mapping. For this, as the first step, the gene apparently associated with the disease is attributed to a limited region within a chromosome (preferably less than 1 cM =

1% recombination frequency) during sequential rounds of linkage analysis of a genetic marker and a phenotype within affected family pedigrees. The genetic marker may be located in the immediate neighborhood of the disease gene rather than within it. If appropriate, the copies of highly repetitive satellite DNA with often unknown structural or regulatory function are used as genetic markers for linkage analysis. In the next step, each positional candidate assigned to a given chromosomal interval will be tested for its functional relevance to the pathogenesis of the disease in biochemical and cell biological experiments and for the existence of the relevant polymorphisms/mutations in affected families and populations in epidemiological studies.

For monogenic diseases, linkage analysis and positional cloning often lead to the unambiguous delineation of the underlying disease gene. However, so far the use of those strategies for the identification of susceptibility genes for complex multifactorial common diseases has been of only limited success [32, 33]. This is certainly due to the typically rather limited causal relationship between the mutations or polymorphisms of a single gene and the resulting phenotype. In the example of figure 5, satellite DNA is closely associated with the susceptibility gene and therefore used as genetic marker for the prevalent polymorphisms '1–4'. Among them, the predisposing polymorphism '3' is associated with three copies of the satellite DNA, the other nonpredisposing polymorphisms '1', '2' and '4' are also linked to this satellite DNA, but in different copy numbers each. The copy numbers are determined via amplification of the complete satellite DNA by polymerase chain reactions using appropriate primers and subsequent analysis of its migration behavior (and thus its length) by gel electrophoresis (fig. 5). Accordingly, 4 type 2 diabetic patients harbor polymorphism '3' (2 homozygous, 2 heterozygous), 2 patients have polymorphism '1' or '2' and '4' and 2 healthy probands have polymorphism '3' (homozygous, heterozygous). This analysis demonstrates that the association of polymorphism '3' with type 2 diabetes is higher compared to a random distribution. However, there are carriers of polymorphism '3' who apparently are not affected by the disease and vice versa. In consequence, the disease risk for carriers of polymorphism '3' is considerably higher than for noncarriers; however, far away from the 100% as holds true for monogenic diseases. Apparently, one or several additional susceptibility genes are implicated in the pathogenesis but are not associated with the genetic marker used (here satellite DNA). This limited association of a single polymorphism with a phenotype is typical for susceptibility

genes. Carriers of the polymorphism may or may not become affected by the disease; however, their relative risk for developing the disease is higher than for noncarriers, albeit the difference (i.e. effect size) may be very small (see below).

This complexity inherent in the identification of disease and susceptibility genes by linkage analysis and positional cloning caused many researchers to favor the alternative approach of directly studying the biological function of the candidate gene ('candidate gene approach'). However, this strategy turned out to be rather disappointing too, mainly due to the often insufficient knowledge of the molecular mechanisms underlying the pathogenesis of multifactorial diseases. As a consequence, it appears that in many cases the candidate genes have been chosen in a rather arbitrary and biased fashion.

In contrast, the identification of susceptibility genes by analysis of linkage disequilibrium is based on the determination of the allelic distribution in distinct populations delineating the recognizable consequences of not recognizable recombination events along the historic development of the population [34]. In general, the identification of susceptibility genes by linkage disequilibrium is more precise and reliable compared to that by linkage analysis provided that the genetic marker used is located in the immediate neighborhood. The extent of linkage disequilibrium within populations typically encompasses 10,000–50,000 base pairs, which is 1,000-fold less than observed within families, enabling a much more precise and reliable localization of susceptibility genes. Consequently, a successful strategy for the mapping and identification of susceptibility genes for complex multifactorial diseases may rely on the initial rough linkage analysis for determination of chromosomal regions associated with a certain probability threshold. Subsequently, after successful replication of the findings on the basis of additional human populations and independent data sets, the chromosomal regions with the strongest linkage are used for a more precise localization via analysis of the linkage disequilibrium. If the gene loci studied do not show pronounced heterogeneity and if a set of genetic markers of sufficient density is used, preferably single-nucleotide exchanges, the signal of association of a linked chromosomal region will be of sufficient strength and can easily be detected. This approach has already led to the identification of a number of susceptibility genes. But its broad application for common diseases and major health problems, in general, is hampered by the enormous expenditure for the adequate analysis of linkage disequilibrium covering large chromosomal regions.

In addition, two technical problems seriously reduce the value of many association studies performed on the basis of linkage analysis or linkage disequilibrium [32, 35]. These studies have to be performed on populations which are representative for future prediction based on analysis of their genetic profile. Study protocols relying on the comparison of affected and nonaffected probands, such as those designed for case-control studies, should not be used to calculate the predictive value of certain combinations of susceptibility genes. In particular, as a consequence of selection, the affected probands may be overrepresented versus the healthy population. This is often the case in studies designed for the detection of disease or susceptibility genes by association. Genetic profiles (as well as traditional risk factors, see below) should preferably be examined in large cohort and longitudinal study populations in which disease prediction is intended to be performed in the future. Furthermore, the predictive value is generally overestimated if the identification of a single polymorphism or a combination of polymorphisms for one or several susceptibility genes has been derived from the same study population. It is therefore of major importance that the results of GWA studies be reproduced with at least one additional independent population prior to publication.

The successful elucidation of the association through linkage analysis or linkage disequilibrium does not necessarily imply a causal relationship between the susceptibility gene and the disease. Linkage or linkage disequilibrium relies on a close neighborhood of gene loci, which reduces the probability of segregation by homologous recombination. However, it may persist over generations by chance rather than as a consequence of tight mechanistic coupling between susceptibility gene(s) and the pathogenesis. It is therefore of tremendous importance to demonstrate causality in subsequent biochemical, cell-biological and animal studies concerning the molecular mechanism that translates genetic information into a disease phenotype [35]. For this, manipulation of the expression of disease or susceptibility genes using overexpression, silencing, transgenic and knockout approaches are of particular value.

Despite these difficulties, 18 susceptibility genes for type 2 diabetes have been identified during the last decade with the help of the methods described [1–4, 6, 34]. However, they account for only about 1% of the theoretical number (see above). Additional susceptibility genes will certainly be detected in the near future. Some of the genes reported so far are shown in figure 6. They are distributed over the complete genome, i.e. 23 chromosomes and the mitochondrial DNA, with polymorphisms pre-
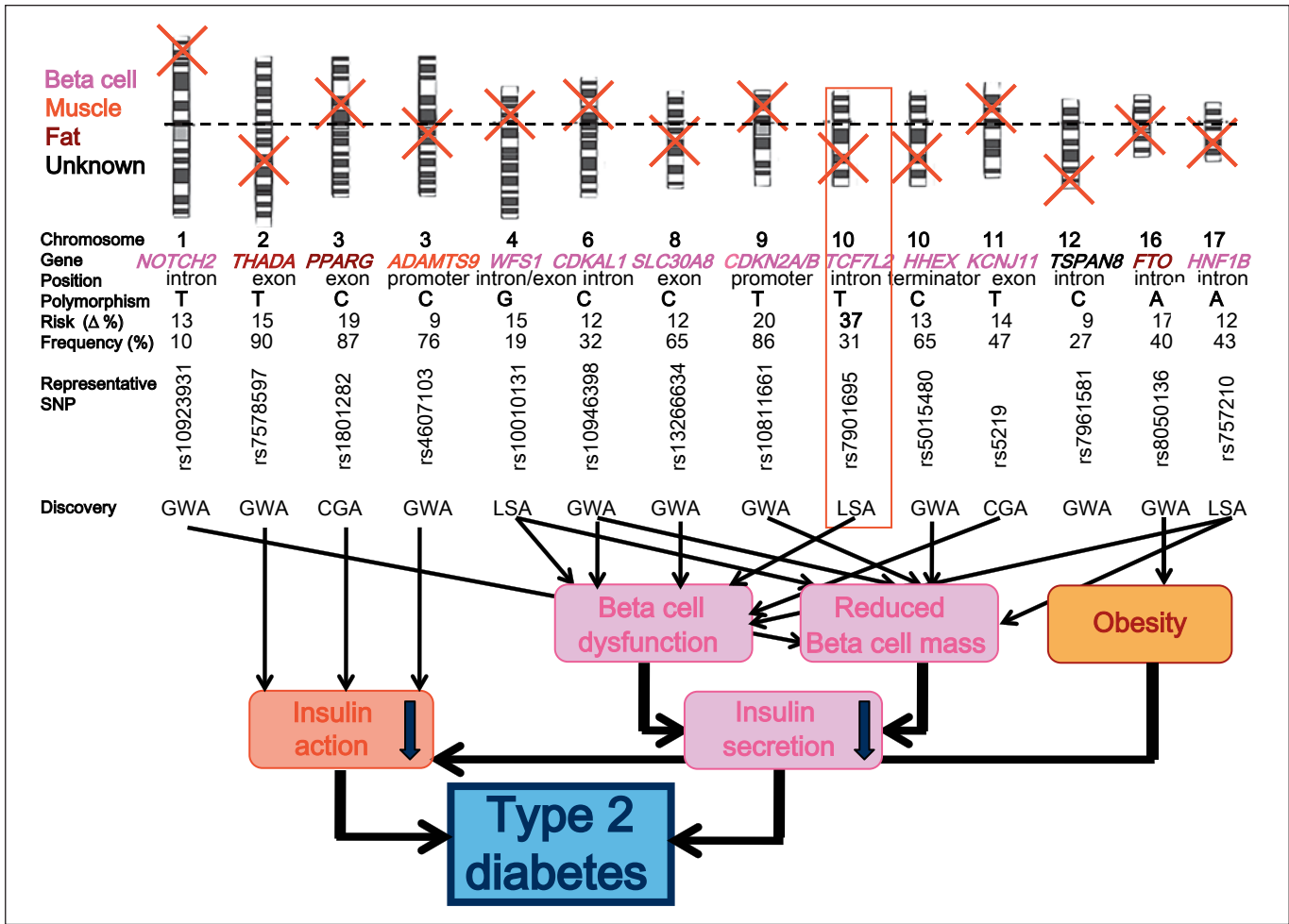
Beta cell / Muscle / Fat / Unknown

Chromosome	1	2	3	3	4	6	8	9	10	10	11	12	16	17
Gene	NOTCH2	THADA	PPARG	ADAMTS9	WFS1	CDKAL1	SLC30A8	CDKN2A/B	TCF7L2	HHEX	KCNJ11	TSPAN8	FTO	HNF1B
Position	intron	exon	exon	promoter	intron/exon	intron	exon	promoter	intron	terminator	exon	intron	intron	intron
Polymorphism	T	T	C	C	G	C	C	T	T	C	T	C	A	A
Risk (Δ %)	13	15	19	9	15	12	12	20	37	13	14	9	17	12
Frequency (%)	10	90	87	76	19	32	65	86	31	65	47	27	40	43
Representative SNP	rs10923931	rs7578597	rs1801282	rs4607103	rs10010131	rs10946398	rs13266634	rs10811661	rs7901695	rs5015480	rs5219	rs7961581	rs8050136	rs757210
Discovery	GWA	GWA	CGA	GWA	LSA	GWA	GWA	GWA	LSA	GWA	CGA	GWA	GWA	LSA

Beta cell dysfunction — Reduced Beta cell mass — Obesity

Insulin action ↓ — Insulin secretion ↓

Type 2 diabetes

Fig. 6. Major susceptibility genes and pathogenic mechanisms for type 2 diabetes identified so far (2008). SNP = Single-nucleotide polymorphism; CGA = candidate gene approach; GWA = genome-wide association; LSA = large-scale association.

dominantly encompassing single-nucleotide exchanges (but also other molecular mechanisms) in exons, introns, promoters and terminators (fig. 4). Interestingly, the majority of the susceptibility genes are specifically expressed in the beta cell, a few of them in muscle, adipose and liver tissues, which resembles the distribution of disease genes for monogenic type 2 diabetes (see above). Most importantly, the relative contribution of the *TCF7L2* gene – for which the most significant disease association so far has been reported in several independent studies – to the development of type 2 diabetes has been estimated at only 34%. As the absolute risk of developing type 2 diabetes is about 10% in Western populations, the individual risk for carriers of the *TCF7L2* polymorphism increases to about 13%. Moreover, *TCF7L2* and many of the other polymorphisms are present in the total population at a relatively high frequency. Both characteristics, i.e. a limited contribution to the individual disease risk (effect size) and pronounced frequency, are typical for the risk alleles of susceptibility genes. In contrast, carriers of the rare risk alleles of disease genes are faced with a theoretical disease risk of 100%.

Of course, individual small effect sizes can amount to more when considered collectively, and it is true that genetic testing (for the known 18 genes) can identify subsets of individuals who have inherited particularly high or low numbers of risk alleles and therefore markedly differ in individual risk [36]. However, the numbers of individuals in these 'extreme' high- and low-risk groups are comparatively small, and in many, their risk will already

be obvious from conventional factors (e.g. family history, BMI, previous gestational diabetes, see below). When the information from the known type 2 diabetes susceptibility variants is examined using approaches such as receiver-operating curve analysis, which are better suited for evaluating the performance of diagnostic tests at the population level, the results look far less spectacular.

As mentioned above, success in the identification of disease and susceptibility genes impacting on individual risk of type 2 diabetes has come from two distinct approaches of gene discovery. The first, linkage mapping within monogenic and syndromic families, has identified rare but highly penetrant causal variants. The second, large-scale association mapping, is now yielding growing numbers of common variants. These have modest effect sizes and low penetrance, at best. Several genes have features that were unraveled by both approaches. For example, mutations in *KCNJ11, PPARG, WFS1,* and *HNF1B* are causal for syndromic variants and/or monogenic forms of diabetes, while common variants in these same genes influence predisposition to typical type 2 diabetes. While common variants in glucokinase do not influence type 2 diabetes risk per se, they have a clear impact on fasting glucose levels within the population. In any case, the genotype-phenotype relationships revealed by these gene discovery efforts will highlight the pathways involved as prime candidates for beneficial therapeutic or preventative manipulation, a view reinforced by the fact that at least two of the genes involved in both monogenic and multifactorial forms of diabetes *(PPARG, KCNJ11)* encode the targets of the two major classes of the currently available antidiabetic drugs, the thiazolidinediones and sulfonylureas, with the most widely used representatives rosiglitazone and glimepiride [37]. The corresponding gene products, transcription factor PPARγ in adipose cells and the SUR1 subunit of the K_{ATP} in pancreatic beta cells, play critical roles in the insulin-stimulated whole-body glucose utilization and the glucose-stimulated insulin secretion, respectively. However, it is remarkable that the associations of the corresponding polymorphisms *(PPARG P12A, KCNJ11 E23K)* with type 2 diabetes are relatively weak.

It is obvious that the two 'flavors' of polymorphism – rare and highly penetrant on the one hand and frequent and low penetrant on the other – are not the only options when it comes to the variants that might influence disease susceptibility. It seems probable, even highly likely, that between these extremes there is a class of medium frequency, medium penetrance variants that have escaped the gaze of the gene mappers until now. The penetrance of such variants would be too low to generate mendelian patterns of segregation and the frequencies would be too low to be covered by current GWA approaches. Despite this, such variants have particularly attractive translational properties. For example, a variant the risk allele of which has a frequency of 1% and produces in a per-allele OR of about 3 would provide greater predictive power than the known variants in *TCF7L2.* Polymorphisms with such characteristics are increasingly being reported in other disease states, such as breast cancer and hyperlipidemia [38, 39], and have even been reported in type 2 diabetes [40]. In principle, just 30 such variants across the genome could explain the observed familial aggregation of type 2 diabetes much better than the current set of common, low-penetrance polymorphisms. Such a pool of polymorphisms would also provide an excellent tool for individual diabetes risk prediction, generating a discriminative accuracy on receiver-operating characteristic analysis close to 80%. The advent of new high-throughput sequencing technologies, along with large-scale association analysis, brings variants in this class within reach of genetic discovery and should allow researchers to evaluate the contribution to disease susceptibility attributable to polymorphisms that lie between the extremes, where attention has previously been focused. Many other challenges are to be faced and opportunities to be realized in the years ahead. The first of these consists in extending the range of polymorphisms that are accessible to scrutiny, beyond the low-frequency variants referred to above, to a systematic evaluation of structural polymorphisms (insertions, deletions, translocations and duplications) and variants that influence the methylation status [41]. Another lies in characterizing the association signals that have been found. Large-scale resequencing and fine-mapping strategies will be required to recover the full allelic spectrum of causal polymorphisms and thereby obtain the most precise quantification of the genetic effects attributable to each locus. The part played by nonadditive interactions between different genetic loci and between susceptibility variants and environmental exposures (see below) needs to be charted, and discovery and replication studies need to be extended beyond the European populations that have been the focus of much of the current research.

Heterogeneity

In general, a single polymorphism or mutation in a disease gene may lead to a more or less pronounced downregulation (or in rare cases upregulation) of the function
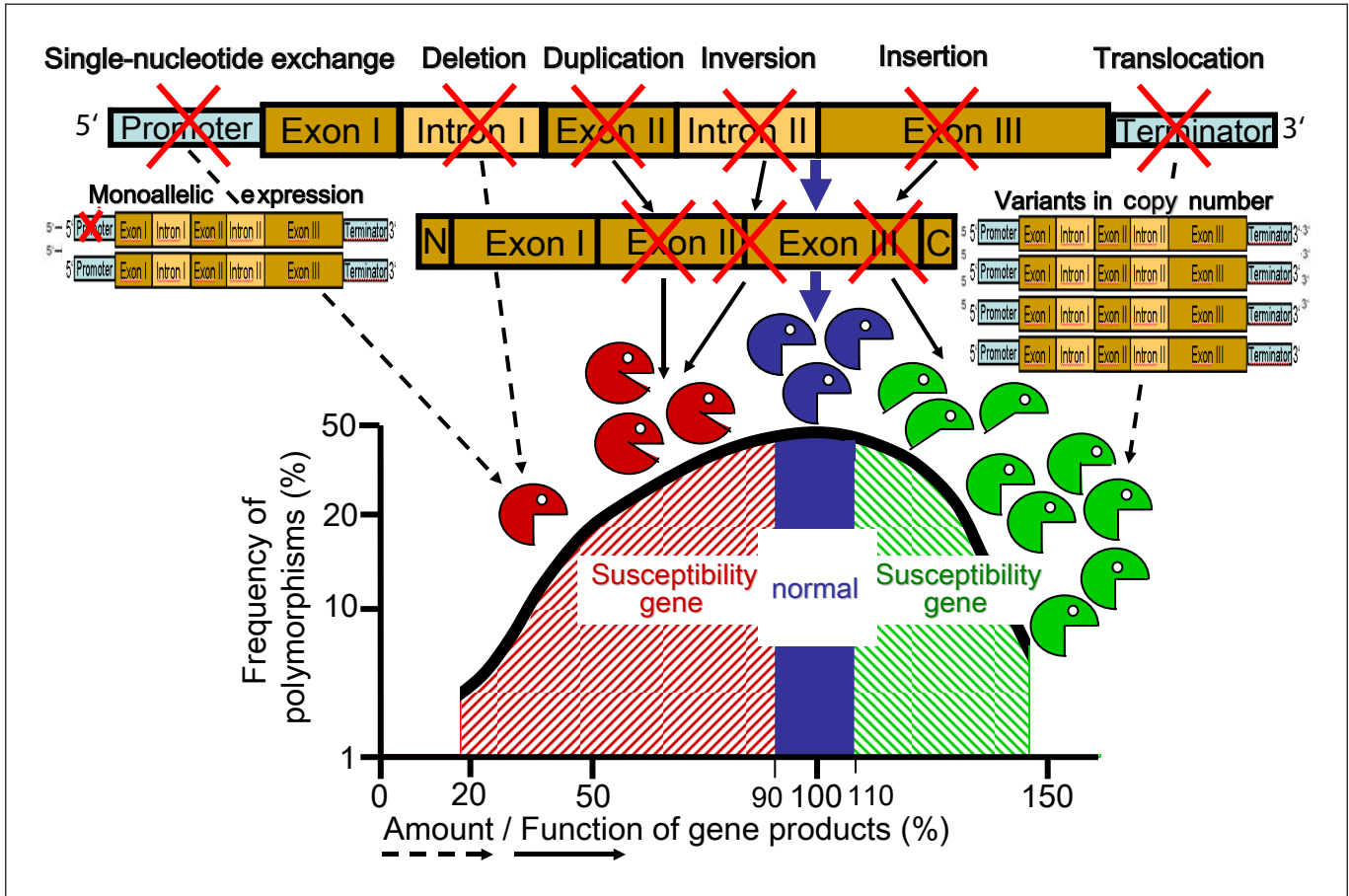
Fig. 7. Molecular mechanisms, frequency and functional consequences of susceptibility genes (hatched areas).

or amount of the corresponding gene product, which per se is sufficient for induction of the corresponding pathogenic pathway (fig. 4). Additional polymorphisms or mutations in other genes or environmental factors are not required. This explains the relatively homogeneous phenotype of monogenic diseases, such as leprechaunism and MODY. In contrast, a single polymorphism in a susceptibility gene, which typically is accompanied by a rather moderate reduction (or in rare cases elevation) of the function or amount of the corresponding gene product, per se is not adequate for triggering the disease phenotype (fig. 7). The participation of additional polymorphisms in other genes and/or specific environmental factors is required. In combination, this results in the complex polygenic inheritance of common diseases, which does not follow the mendelian laws. Thus susceptibility genes apparently form the 'gray zone' (fig. 7, hatched areas) within the 'gaussian' distribution of the

amount or function of the polymorphic gene product between the 'healthy' wild-type genes coding for (almost) normal gene products in the center of the curve, i.e. at the highest frequency, and the disease genes coding for more or less functionally impaired or overactive gene products on the flank of the curve, i.e. at lower frequency (fig. 7). Moreover, as many of the causal polymorphisms are situated in noncoding regions of susceptibility genes, often at some distance from the coding sequence, they will often have subtle, spatially and/or temporally restricted effects. In such circumstances, gathering experimental evidence of their functional impact will be very difficult.

As a consequence, a polymorphism in a single susceptibility gene involved in glucose utilization and expressed in relevant tissues, such as the pancreas, muscle, fat, liver and intestinum will not be sufficient per se for induction of type 2 diabetes. In the example of figure 7, 5, 7 or 15% less functionality of beta cells, muscle or adipose cells
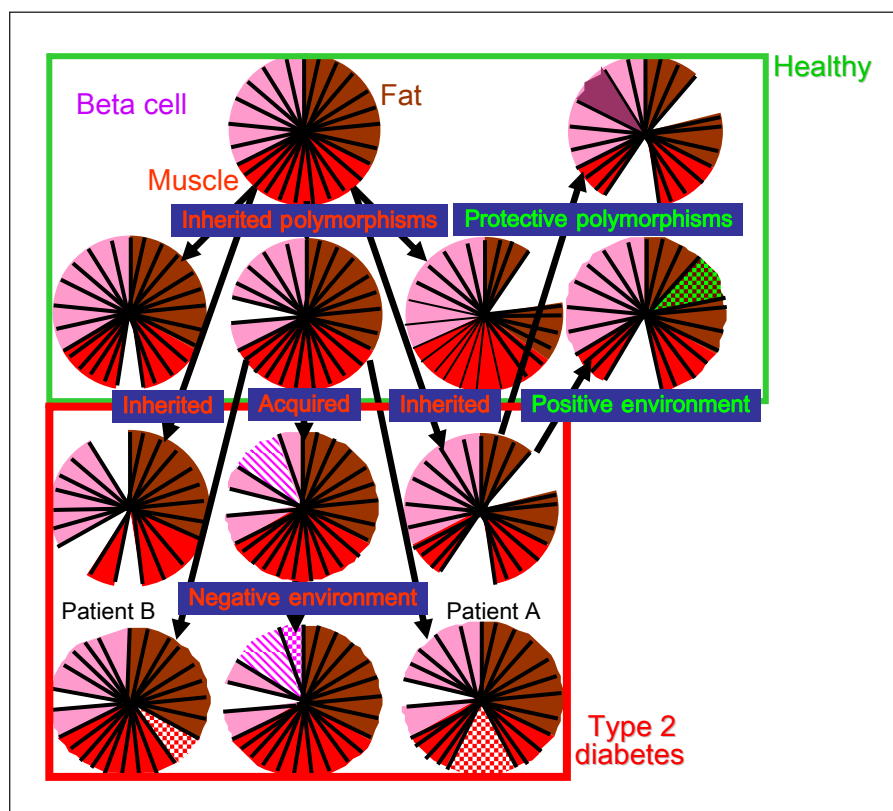
Fig. 8. Concerted actions of inherited and acquired polymorphisms, as well as positive and negative environmental factors on the pathogenesis of type 2 diabetes.

may still enable some glucose utilization, compatible with almost normal blood glucose homeostasis. Additional polymorphisms in the same or other tissues inherited from the parents or acquired in somatic cells during a lifetime are required to trigger the pathogenesis. Thus the combination of 5% (inherited) and 19% (acquired) functional impairment in the beta cells, 7% in muscle cells and 11 and 12% in beta cells or 15% in adipose cells and 19% in muscle cells may lead to diminished glucose utilization which over years fuels the development of type 2 diabetes. The situation is even more complex since polymorphisms or mutations in protective genes of tissues and organs involved in glucose utilization may compensate for functional weaknesses in the same or distinct tissues. Thus 15 and 19% functional impairment of adipose and muscle cells may be overcome by 12% higher activity of 'empowered' beta cells (fig. 8). In fact, this compensatory mechanism is operating in the majority of obese humans who would suffer from significantly impaired glucose utilization by their muscle and liver cells in response to normal physiological insulin levels, i.e. from apparent peripheral insulin resistance. However, the elevated supraphysiological insulin secretion which can be maintained during

the whole lifetime by the 'empowered' beta cells drives the adequate flux of glucose into peripheral tissues and thereby prevents the pathogenesis of type 2 diabetes. Only those obese subjects who lack 'empowered' beta cells will develop frank type 2 diabetes. Strikingly, different polymorphisms of the same gene may play either a predisposing or a protective role, e.g. in the glucose-induced apoptosis or in the glucose-induced proliferation/differentiation of the beta cells.

In conclusion, the manifestation of a huge number of combinations of the various susceptibility genes and their multiple polymorphisms in disease-relevant tissues is apparent (fig. 8). This explains the extraordinary genetic heterogeneity of common diseases, in general, and type 2 diabetes, in particular. With the exception of monozygotic twins, the probability for the identity of the genetic profile, i.e. specific combination of all the polymorphisms in the disease-relevant susceptibility genes, between 2 diabetic patients is almost zero. It seems to be very low even in 1st-degree relatives as exemplified by the following calculation. Under the assumptions that a combination of 12 homozygous polymorphisms in recessive susceptibility genes constitutes a complete causal pathogenic mecha-
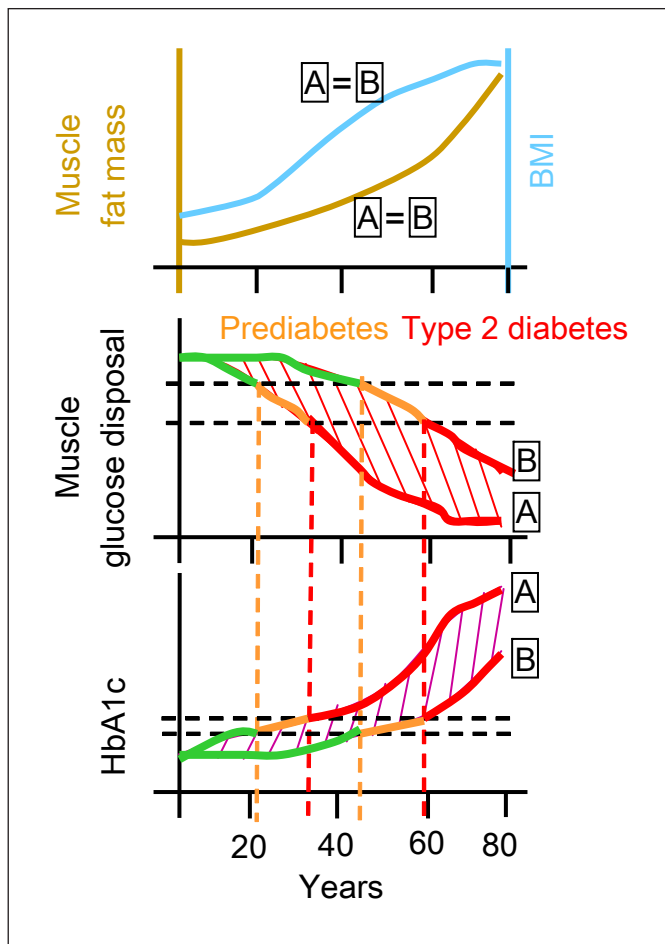
Fig. 9. Individual variability (between patients A and B) during the pathogenesis of type 2 diabetes.

and 19% reduced activities in adipose and muscle cells, respectively, may not be sufficient to induce the pathogenesis (fig. 8). However, a 'negative' environment further impairs the functionality of the same or different relevant tissues, by 8, 5 or 16% in adipose, beta cells or muscle cells, respectively, which, together with the marginally reduced beta cell activity (by 5%) caused by the polymorphism, will induce the pathogenesis (fig. 8). For type 2 diabetes, the impact of the environment is most dramatic with regard to the individual lifestyle, in particular, nutrition and physical exercise. With increasing body weight as is reflected in the increase in the BMI, the deposition and accumulation of fat in the adipose tissue, and not only there, but also and even more so in other tissues, such as muscle, liver and beta cells, is accelerated (fig. 9). In order to prevent excessive storage of neutral lipids, those cells respond with downregulation of the insulin-stimulated glucose uptake and metabolism, which provide essential building blocks for lipid synthesis, by molecular mechanisms, which are not yet fully understood. The resulting 'physiological' insulin resistance with impaired glucose disposal predominantly in muscle and liver tissues leads to impaired glucose tolerance, i.e. steep increase in blood sugar after food consumption (fig. 1), which via a stage of prediabetes (with HbA1c levels below a critical threshold) proceeds to full-blown type 2 diabetes (fig. 9). Importantly, the disease-relevant cells and tissues of different individuals (patients A and B in fig. 9) may respond to the same negative environmental factors (development of BMI) with the same (lipid deposition in muscle) and/or different (muscle glucose disposal) sensitivity, which is also determined by susceptibility genes. This explains the well-known observations that despite identical lifestyle some patients (here A) will pass the state of prediabetes and develop frank type 2 diabetes at a considerably earlier time point than others (here patient B). This reflects the heterogeneity of the inherited and acquired polymorphisms in susceptibility genes which control the sensitivity of tissues relevant to glucose utilization (here muscle) leading to the deposition of neutral lipids.

The impact of environmental factors on the glucose-utilizing organs may be positive as well. In the example of figure 7, the functionality of adipose cells, which would be reduced by 16% due to a polymorphism in a susceptibility gene, becomes fully restored by appropriate nutrition and thereby counteracts the development of type 2 diabetes, although the functional impairment (by 15%) of the muscle cells persists. Thus, susceptibility genes and environmental factors in close cooperation contribute to

nism (see below) and that only the mother or the father is a carrier of this combination and suffers from the disease, the probability for inheritance of exactly this combination is $0.5^{12} = 0.024$ %, or 0% in case of the other parent being a heterozygous carrier of each of the 12 polymorphisms or a noncarrier of even only one (or several) of these polymorphisms.

Environment

Glucose disposal and its control by the relevant tissues are not only determined by the individual genetic profile. Negative and positive factors contributed by the environment modulate the function of the glucose-utilizing organs as well. In a 'positive' environment, even numerous but functionally moderate polymorphisms, such as 15

the considerable heterogeneity in the individual pathogenic developments of type 2 diabetes, in particular, and common diseases, in general.

Evolution

It is evident that, as a result of the intimate relationship between biological and cultural evolution, natural selection favored the spread of polymorphisms and mutations with moderate impact on the amount and/or function of the corresponding gene products that are associated with significant advantages for survival and vitality during certain periods of mankind. A plausible explanation for the interaction between genes and environment during the pathogenesis of type 2 diabetes is provided by the 'thrifty-gene hypothesis'. According to Neel [42], the probability of survival of individuals living in an environment with unsteady nutrient supply, i.e. of our hunting and gathering ancestors during periods of famine, was higher with the greater capacity for energy storage during periods of nutrient abundance. Energy storage is much more efficient as deposition of neutral lipids, preferably in the abdominal rather than subcutaneous adipose tissue compared to glycogen storage in muscle and liver tissues. The observed early accumulation of abdominal adipose tissue in 1st-degree relatives of type 2 diabetic patients may thus be interpreted as the result of natural selection for carriers of genes that guarantee efficient energy storage and, as a prerequisite, utilization of the glucose taken up by food. Relevant genes may trigger excessive insulin secretion by beta cells in response to food uptake in parallel to those provoking transient 'physiological' insulin resistance of muscle and liver cells, but leaving unaltered the insulin sensitivity of adipose cells. Thereby, the food calories are efficiently taken up and predominantly stored in the adipose tissue in the long term without the danger of hypoglycemia in the course of immediate disposal of the blood glucose by muscles and the liver. Importantly, the continuous changes between nutrient abundance and deprivation as were typical for hunters and gatherers prevented the development of a permanent 'pathological' insulin resistance and the excessive accumulation of lipids and, as a consequence, the development of obesity. In the course of the neolithic revolution, with the invention of agriculture and farming and subsequent industrialization in the Western world, this advantage of the 'physiological' insulin resistance in combination with hyperinsulinemia, i.e. supraphysiological plasma insulin levels, was blunted. The permanent

and excessive availability of food with high calorie content reversed the former advantage for survival into a disadvantage. In conclusion, according to the thrifty-gene hypothesis, complex combinations of polymorphisms and mutations in genes controlling metabolic pathways which facilitated the survival of our ancestors increase the susceptibility of the present population to the development of type 2 diabetes.

The 'thrifty-phenotype hypothesis' provides an alternative explanation for the intimate interplay between genes and the environment. According to Hales and Barker [43], intrauterine malnutrition leads to delayed and impaired development of the molecular mechanisms of glucose utilization and energy storage, i.e. of an efficient action of insulin in muscle, liver and adipose tissues, insulin production and secretion by the beta cells as well as glucose sensing by beta cells and intestinal cells. During later life, calories taken up in excess will not be handled properly by concerted actions of these cells and will not be deposited in the adipose tissue, exclusively, but will be stored as lipid in muscle and liver, as well. This bypass further drives impaired glucose utilization and development of type 2 diabetes into a vicious circle. For the malnourished embryo, the diminished insulin-dependent glucose utilization is of advantage since it allows the rapid access to calories contained in the limited amount of food available to meet the short-term energy requirements. In conclusion, the thrifty-phenotype hypothesis attributes the susceptibility genes a special role in determining the sensitivity of beta cells, muscle, liver, adipose and intestinal cells for rapid energy mobilization during intrauterine malnutrition.

Predictive Value

The complex interplay between polymorphisms in a multitude of susceptibility genes (each with a relatively moderate functional impact, i.e. effect size) and a broad array of environmental factors results in a huge number of theoretically possible combinations which ultimately reflect the causal chains of events, i.e. pathogenic mechanisms leading to common multifactorial diseases, such as type 2 diabetes. According to Rothman and Greenland [35], the understanding of a complete causal pathogenic mechanism provides a combination of the conditions, events or (single) components (at the physiological and/or molecular level) sufficient and necessary for the development of the disease. Thus, this combination will inevitably trigger the pathogenesis. Consequently, if all
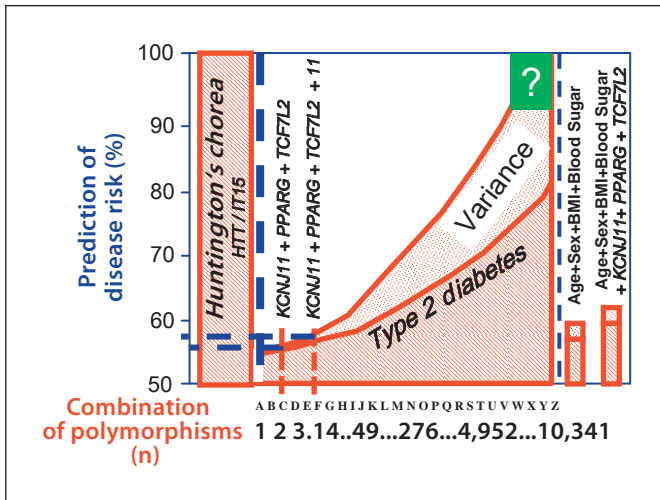
Fig. 10. Prediction of the risk for monogenic (Huntington's chorea) and polygenic diseases (type 2 diabetes) on basis of the combination (A to Z) and number of polymorphisms (1–10,341) in the relevant disease (huntingtin, *HTT/IT15*) or susceptibility *(KCNJ11, PARG, TCF7L2)* genes, respectively, as well as traditional risk factors and biomarkers. BMI = Body mass index.

the components constituting a pathogenic mechanism are present, the disease risk is 100%, and if one or several components are missing, the risk is 0%. Chance or indetermination does not play a role in this model of the complete causal pathogenesis. However, most complex common diseases rely on several or many distinct complete causal pathogenic mechanisms which often interact in complex fashion. The resulting 'nonlinearity' of the pathogenesis in time and space (i.e. tissue) makes it often very difficult to decipher the causal chain of events.

The elucidation of a complete causal pathogenic mechanism for a common disease necessitates the selection of a specific combination of single components from the multitude of theoretically possible combinations which inevitably lead to the pathogenesis. This is a tremendous scientific challenge. For the reasons mentioned above, the number of possible combinations is extremely high and typically by far exceeds the number of probands even of large association studies. In case of the simultaneous analysis of 12 susceptibility genes with 3 polymorphisms each, the number of possible combinations accounts for $3^{12} = 531,441$. Consequently, in a large study encompassing 30,000 probands, each possible combination including that of healthy controls will be detected only once at best, even if certain pathogenic

mechanisms, i.e. combinations, occur at a higher frequency than others. The consequences for the prognosis and diagnosis as well as the therapy of common diseases are severe [32]. It will remain difficult to finally prove by association (i.e. linkage analysis or linkage disequilibrium, see above) that a certain combination of polymorphisms in relevant susceptibility genes that have been detected in affected probands only, constitutes a complete causal pathogenic mechanism. In fact, the probability of finding a specific combination once identified in more than one proband is extremely small. Furthermore, the value of the identification of a unique combination of polymorphisms in susceptibility genes for the prediction of a common disease is rather low. This holds true even under the prerequisite of the successful demonstration that it defines a complete causal pathogenic mechanism. Only very few (if any) humans share the identical or a very similar genetic profile with the affected individual from whom the specific combination has been originally derived.

For these reasons, it seems unlikely that large or very large numbers of identified susceptibility genes would significantly increase the predictive power for type 2 diabetes compared to the traditional risk factors, such as age, sex, BMI, fasting blood glucose. Thus, the 100% predictive value typical for monogenic diseases (e.g. Huntington's chorea), will probably never be reached for common diseases (fig. 10). This is exemplified best by the observation that certain combinations of 3 and 14 out of the hitherto demonstrated susceptibility genes for type 2 diabetes (fig. 6) increase the disease risk from 50% (probability of throwing a coin) only to 55 and 58%, respectively (fig. 10). Strikingly, on the basis of traditional risk factors solely, type 2 diabetes is currently being predicted with about 59% probability; upon combination with the 3 susceptibility genes, this probability only increases to 61% (fig. 10). Unfortunately, a number of polymorphisms in certain susceptibility genes for type 2 diabetes are likely to induce or modulate the development of these traditional risk factors or to affect their consequences for the disease phenotype. However, only those polymorphisms of susceptibility genes will improve the predictive value compared to traditional risk factors, which operate via hitherto unknown pathogenic mechanisms and/or determine the expression of risk factors that have escaped detection and measurement so far. The likelihood for the operation of pathogenic mechanisms that have not been recognized so far will vary with certain subsets of common diseases rather than being uniform. This nourishes the hope that the predictive value of the genetic profile will increase
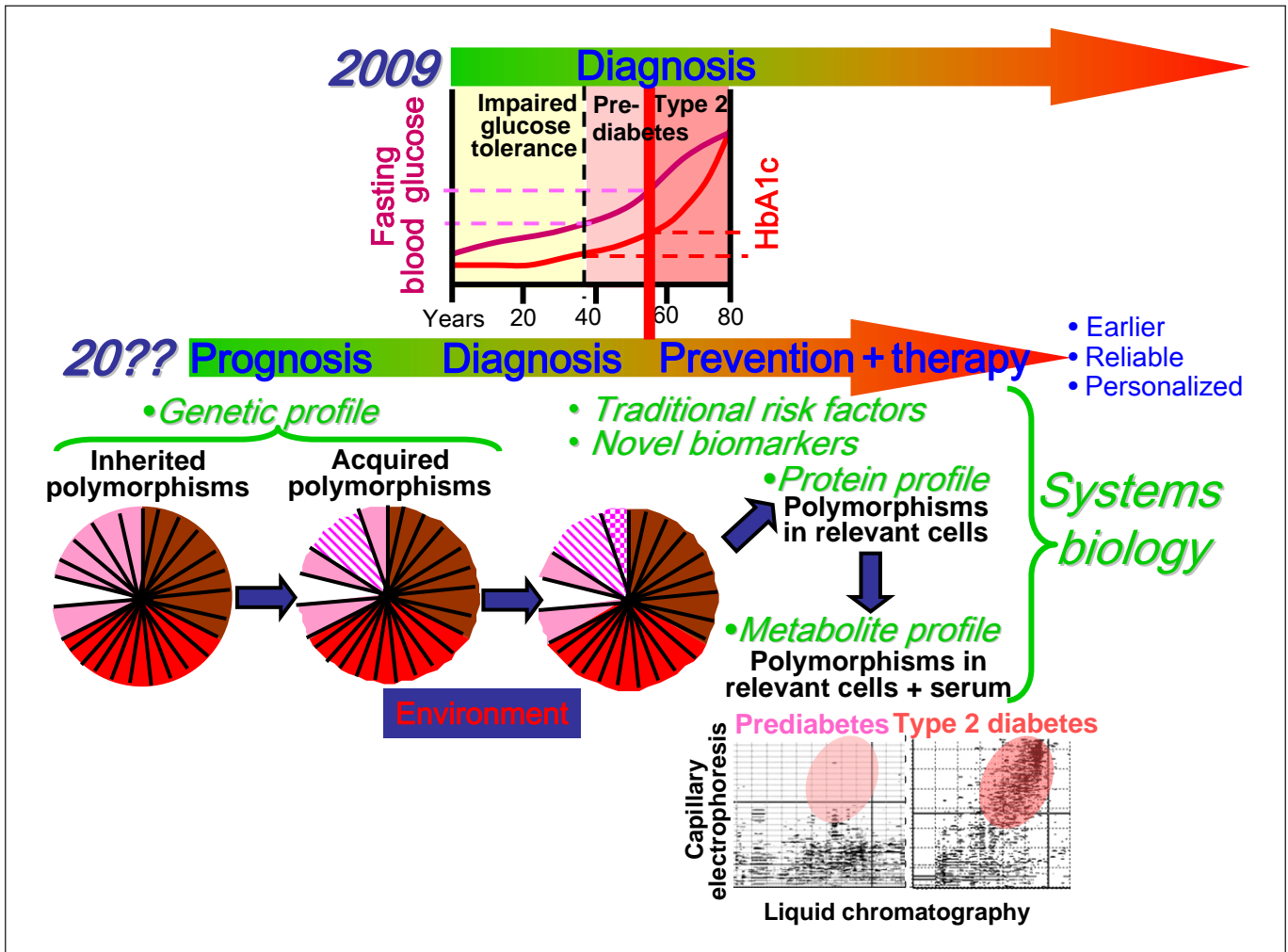
Fig. 11. Replacement of 'nonpersonalized' diagnosis of type 2 diabetes prevalent in 2009 on the basis of impaired glucose tolerance, fasting blood sugar and HbA1c levels by personalized diagnosis in the future on the basis of the gene, protein and metabolite profiles as well as traditional risk factors and novel biomarkers.

upon the inclusion of an increasing number of 'novel' susceptibility genes involved in hitherto unknown pathogenic mechanisms and approach the 100% value more closely than on the basis of the previously identified polymorphisms (fig. 10).

Importantly, proteins and metabolites, such as lipids, amino acids and metabolites, are 'mechanistically closer' to the pathogenic mechanism than the genes directly or indirectly controlling their synthesis. Therefore, analysis of the polymorphic patterns of proteins and metabolites reflecting the gene product variants – which directly or indirectly emerge from the expression of the corresponding susceptibility genes – in the relevant tissues and serum of probands may further enhance the prognostic

and diagnostic predictive power. This is true in particular for the determination of the proteins and metabolites at a very early time point in the pathogenesis (see below). Nevertheless, the predictive power typical for monogenic diseases will never be reached for common diseases, irrespective of the higher variance generally observed for common compared to monogenic diseases. In any case, the identification of novel pathogenic mechanisms and associated risk factors or biomarkers, such as proteins and metabolites, will represent an additional independent and valuable outcome emerging from future GWA studies for susceptibility genes with improved predictive power.

Personalized Prognosis and Diagnosis

A number of susceptibility genes and polymorphisms associated with type 2 diabetes to a certain degree have been identified so far (fig. 6). Nevertheless, the current diagnosis is exclusively relying on traditional risk factors, i.e. the serum parameters, fasting blood glucose and HbA1c, and is often restricted to routine medical examinations in elderly probands (fig. 11). However, at this late time point of diagnosis, most type 2 diabetic patients are already suffering from impaired glucose tolerance and even slightly to moderately elevated fasting blood glucose levels, which per se are not associated with acute symptoms and pain, but have already set into motion the vicious pathogenic circle ultimately leading to late diabetic complications. Unfortunately, early symptoms, such as slight nerve pain, excessive thirst or deterioration of visual power, which often accompany the prediabetic stage, and moderately elevated HbA1c levels, are often being overlooked during routine medical examinations unless the patient is aware of familial clustering of type 2 diabetes. Admittedly, the timely recognition of elevated levels of fasting blood glucose, plasma insulin and HbA1c as well as of peripheral insulin resistance and impaired glucose tolerance is not very demanding from a technical point of view or from the time frame available due to the long-lasting pathogenesis, which typically requires years to decades from the states of impaired glucose tolerance and pre-diabetes to frank type 2 diabetes. However, these parameters do not provide any information about the complex etiology of the disease, which, as outlined above, differs from patient to patient. Some forms may have their origin in defective glucose sensing by the beta cells or intestinal cells, other forms in defective insulin production and secretion, and certain forms in defective insulin action in muscle, liver or adipose cells, and others may originate from combinations thereof. However, this information is a prerequisite for personalized therapy addressing the pathways of glucose sensing, insulin production, insulin secretion or insulin action in the relevant target cells with its crucial advantages in efficacy and safety if the primary cause of the disease is aimed at rather than (a multitude of) secondary consequences. Therefore, determination of the complete gene, protein and metabolite profiles covering all inherited and acquired polymorphisms in the relevant susceptibility genes and gene products could be helpful for an earlier and more reliable prognosis and diagnosis of type 2 diabetes with critical advantages for its prevention and therapy (fig. 11).

This is exemplified in the pattern of small-molecule metabolites in glucose-utilizing tissues, such as muscle, as revealed by two-dimensional separation by capillary electrophoresis and liquid chromatography [44, 45]. The patterns considerably differ between probands analyzed at the stage of prediabetes and frank type 2 diabetes with quantitative differences ranging from 'black to white' to 'dark gray to light gray' (fig. 11). The elucidation of the corresponding protein patterns, which is facilitated by modern protein chip technology [46], would reveal comparable differences. This challenging, seemingly 'global' approach, so-called systems biology, encompasses the analysis of all genes (genomics), proteins (proteomics) and metabolites (metabolomics) with the aim of identifying their quantitative changes in expression and their mutual relationships along time and space [45, 46]. Early prognosis and diagnosis based on systems biology, which combines genomics, proteomics, metabolomics, traditional risk factors and novel biomarkers, may ensure the maximally possible predictive power (fig. 11).

However, it is not yet clear whether personalized information of this kind (particularly when other pertinent factors such as age, ethnicity, family history and BMI are not explicitly taken into account) will lead individuals toward beneficial changes in health-related behaviors [47] or alterations in their clinical management. Indeed, if such information were to be poorly presented, there is a danger that overestimation of the deterministic qualities of genetic information could motivate individuals toward counterproductive changes in their lifestyle (through unwarranted fatalism or feelings of personal immunity). The final challenge lies in placing systems biology into a translational context. The clinical utility and validity of gene, protein and metabolite diagnostics have already been established in monogenic diseases, such as MODY, where such testing can influence clinical practice and treatment. It is a major challenge to establish how to use analogous knowledge from the identification of predisposing polymorphisms of genes, proteins and metabolites for common diseases such as type 2 diabetes, to improve patient care. In the future, systems biology will be the prerequisite for an earlier and more reliable diagnosis, prognosis and therapy, tailored to the individual pathogenic mechanism and thereby justify the hopes commonly associated with personalized medicine in the public [37, 48–53]. No doubt, this necessitates a broad understanding of the putative personal consequences and advantages of systems biology-based diagnosis in the public, which will critically depend on corresponding individualized therapeutic options. It is clear from the daily

medical doctor's experience and the increasing incidence of type 2 diabetes in all regions of the world with unlimited access to calorie-rich diets that the mere offering of dietetic and exercise measures, albeit their benefits hold true for almost any common disease and individual, does not justify intensified efforts in personalized prognosis and diagnosis.

Conclusions

Common or civilization diseases, such as type 2 diabetes, coronary heart diseases and degenerative diseases, are caused by a complex multifactorial interplay of genes and environmental factors. Recent developments in genomic sciences have led to the discovery of numerous disease and susceptibility genes. This nourished great hope in personalized medicine with its promise of reliable prognosis, early diagnosis as well as efficient and safe therapy of common diseases in the near future based on the determination of individual genetic profiles. However, association and clinical studies have also revealed the rather limited predictive value of the current analysis of gene polymorphisms for common diseases, even if it is combined with that of traditional risk factors. However, there are good reasons for assuming that the detection of additional susceptibility genes in combination with the analysis of proteins, metabolites as well as traditional and novel biomarkers will improve the predictive value to the degree necessary and sufficient for a personalized prognosis and diagnosis. Nevertheless, the considerable expenditure for a personalized prognosis and diagnosis will be justified only in combination with the availability of a personalized therapy.

References

1 Prokopenko I, McCarthy MJ, Lindgren CM: Type 2 diabetes: new genes, new understanding. Trends Genet 2008;24:613–621.
2 Ridderstrale M, Groop L: Genetic dissection of type 2 diabetes. Mol Cell Endocrinol 2009; 297:10–17.
3 Lindgren CM: Mechanisms of disease: genetic insights into the etiology of type 2 diabetes and obesity. Nature Clin Pract 2007;4: 156–163.
4 Florez JC: The genetics of type 2 diabetes: a realistic appraisal in 2008. J Clin Endocrinol Metab 2008;93:4633–4642.
5 Pearson H: One gene, twenty years. Nature 2009;460:165–169.
6 McCarthy MI, Hatterlsley AT: Learning from molecular genetics. Novel insights arising from the definition of genes for monogenic and type 2 diabetes. Diabetes 2008;57: 2889–2898.
7 Stoy J, Edghill EL, Flanagan SE, Ye H, Paz VP, Pluzhnikov A, Below JE, Hayes MG, Cox NJ, Lipkind GM, Lipton RB, Greeley SA, Patch AM, Ellard S, Steiner DF, Hattersley AT, Philipson LH, Bell GI: Insulin gene mutations as a cause of permanent neonatal diabetes. Proc Natl Acad Sci USA 2007;104: 15040–15044.
8 Hattersley AT, Turner RC, Permutt MA, Patel P, Tanizawa Y, Chiu KC, O'Rahilly S, Watkins PJ, Wainscoat JS: Linkage of type 2 diabetes to the glucokinase gene. Lancet 1992; 339:1307–1310.
9 Froguel P, Vaxillaire M, Sun F, Velho G, Zouali H, Butel MO, Lesage S, Vionnet N, Clement K, Fougerousse F, Tanizawa Y, Weissenbach J, Beckmann JS, Lathrop GM, Passa P, Permutt MA, Cohen D: Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature 1992;356:162–164.
10 Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, Howard N, Srinivasan S, Silva JM, Molnes J, Edghill EL, Frayling TM, Temple IK, Mackay D, Shield JP, Sumnik Z, van Rhijn A, Wales JK, Clark P, Gorman S, Aisenberg J, Ellard S, Njolstad PR, Ashcroft FM, Hattersley AT: Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350;2004:1838–1849.
11 Proks P, Arnold AL, Bruining J, Girard C, Flanagan SE, Larkin B, Colclough K, Hattersley AT, Ashcroft FM, Ellard S: A heterozygous activating mutation in the sulphonylurea receptor SUR1 (ABCC8) causes neonatal diabetes. Hum Mol Genet 2006;15: 1793–1800.
12 Babenko AP, Polak M, Cave H, Busiah K, Czernichow P, Scharfmann R, Bryan J, Aguilar-Bryan L, Vaxillaire M, Froguel P: Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. N Engl J Med 2006;355:456–466.
13 Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, Chatterjee VK, O'Rahilly S: Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 1999;402:880–883.
14 Kadowaki T, Bevins CL, Cama A, Ojamaa K, Marcus-Samuels B, Kadowaki H, Beitz L, McKeon C, Taylor SI: Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. Science 1988; 240:787–790.
15 Byrne MM, Sturis J, Clement K, Vionnet N, Pueyo ME, Stoffel M, Takeda J, Passa P, Cohen D, Bell GI, et al: Insulin secretory abnormalities in subjects with hyperglycemia due to glucokinase mutations. J Clin Invest 1994; 93:1120–1130.
16 Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, Southam L, Cox RD, Lathrop GM, Boriraj VV, Chen X, Cox NJ, Oda Y, Yano H, Le Beau MM, Yamada S, Nishigori H, Takeda J, Fajans SS, Hattersley AT, Iwasaki N, Pedersen O, Polonsky KS, Turner RC, Velho G, Chevre JC, Froguel P, Bell GI: Mutations in the hepatic nuclear factor 1α gene in maturity-onset diabetes of the young (MODY3). Nature 1996;384:455–458.
17 Pontoglio M, Barra J, Hadchouel M, Doyen A, Kress C, Poggi Bach J, Babinet C, Yaniv M: Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome. Cell 1999; 84:575–585.

18 Dukes ID, Sreenan S, Roe MW, Levisetti M, Zhou YP, Ostrega D, Bell GI, Pontoglio M, Yaniv M, Philipson L, Polonsky KS: Defective pancreatic beta-cell glycolytic signaling in hepatocyte nuclear factor-1α-deficient mice. J Biol Chem 1998;273:24457–24464.

19 Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, Fajans SS, Signorini S, Stoffel M, Bell GI: Mutations in the hepatocyte nuclear factor 4α gene in maturity-onset diabetes of the young (MODY1). Nature 1996;384:458–460.

20 Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn B, Lindner T, Yamagata K, Ogata M, Tomonaga O, Kuroki H, Kasahar T, Iwamoto Y, Bell GI: Mutation in hepatocyte nuclear factor-1β gene (TCF2) associated with MODY. Nat Genet 1997;17:384–385.

21 Servitja JM, Ferrer J: Transcriptional networks controlling pancreatic development and beta cell function. Diabetologia 2004;47:597–613.

22 Maestro MA, Cardalda C, Boj SF, Luco RF, Servitja JM, Ferrer J: Distinct roles of HNF1β, HNF1α, and HNF4α in regulating pancreas development, beta-cell function and growth. Endocr Dev 2007;12:33–45.

23 Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, Nermoen I, Eide SA, Grevle L, Bjorkhaug L, Sagen JV, Aksnes L, Sovik O, Lombardo D, Molven A, Njolstad PR: Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. Nat Genet 2006;38:54–62.

24 Hegele RA, Anderson CM, Wang J, Jones DC, Cao H: Association between nuclear lamin A/C R482Q mutation and partial lipodystrophy with hyperinsulinemia, dyslipidemia, hypertension, and diabetes. Genome Res 2000;10:652–658.

25 Capell BC, Collins FS: Human laminopathies: nuclei gone genetically awry. Nat Rev Genet 2006;7:940–952.

26 Delepine M, Nicolino M, Barrett T, Golamaully M, Lathrop GM, Julier C: EIF2AK3, encoding translation initiation factor 2-α kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nat Genet 2000;25:406–409.

27 Wildin RS, Ramsdell F, Peake J, Faravelli F, Casanova JL, Buist N, Levy-Lahad E, Mazzella M, Goulet O, Perroni L, Bricarelli FD, Byrne G, McEuen M, Proll S, Appleby M, Brunkow ME: X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet 2001;27:18–20.

28 Stumvoll M, Goldstein BJ, van Haeften TW: Type 2 diabetes: principles of pathogenesis and therapy. Lancet 2005;365:1333–1346.

29 Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C, et al: Genetic prediction of future type 2 diabetes. PLoS Med 2005;2:1299–1307.

30 Pearson ER, Donnelly LA, Kimber C, Whitley A, Doney ASF, McCarthy MI, et al: Variation in TCF7L2 influences therapeutic response to sulfonylureas: a GoDARTs study. Diabetes 2007;56:2178–2182.

31 Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38:320–323.

32 Janssens ACJW, van Duijn CM: Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet 2008;17:R166–R173.

33 McCarthy MI, Smedley D, Hide W: New methods for finding disease-susceptibility genes: impact and potential. Genome Biol 2003;4:119.

34 Florez JC, Hirschhorn JN, Altshuler D: The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annu Rev Genomics Hum Genet 2003;4:257–291.

35 Rothman KJ, Greenland S: Causation and causal interference in epidemiology. Am J Public Health 2005;95;S1:S144–S150.

36 Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al: Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3:e374.

37 Müller G: The mode of action of glimepiride – beyond insulin secretion. Curr Med Chem 2005;5:499–518.

38 Stratton MR, Rahman N: The emerging landscape of breast cancer susceptibility. Nat Genet 2008;40:17–22.

39 Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC: Population-based resequencing of ANGPLT4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet 2007;39:513–516.

40 Hegel RA, Cao H, Harris SB, Hanley AJG, Zinman B: The hepatic nuclear factor-1α G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. J Clin Endocrinol Metab 1999;84:1077–1082.

41 McCarroll SA; Altshuler DM: Copy-number variation and association studies of human disease. Nat Genet 2007;39:S37–S42.

42 Neel V: Diabetes mellitus: a 'thrifty' genotype rendered detrimental by progress? Am J Hum Genet 1962;14:352–362.

43 Hales C, Barker D: Type 2 diabetes mellitus: the thrifty phenotype hypothesis. Diabetologia 1992;35:595–601.

44 Figeys D, Ducret A, Yates JR 3rd, Aebersold R: Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray tandem mass spectroscopy. Nat Biotechnol 1996;14:1579–1583.

45 Xiayan L, Legido-Quigley C: Advances in separation science applied to metabonomics. Electrophoresis 2008;29:3724–3736.

46 Figeys D, Pinto D: Proteomics on a chip: promising developments. Electrophoresis 2001;22:208–216.

47 Janssens AC, Gwinn M, Subramonia-Iyer S, Khoury MJ: Does genetic testing really improve the prediction of future type 2 diabetes? PLoS Med 2006;3:e114.

48 Müller G: Concepts and options for current insulin research and future anti-diabetic therapy. Recent Res Develop Endocrinol 2002;3:199–218.

49 Edwards KL, Hutter CM, Wan JY, Kim H, Monks SA: Genome-wide linkage scan for the metabolic syndrome: The GENNID study. Obesity 2008;16:1596–1601.

50 Risch N, Merikangas K: The future of genetic studies of complex human diseases. Science 1996;273:1516–1517.

51 McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. Curr Diab Rep 2003;3:159–167.

52 Altshuler D, Daly M: Guilt beyond a reasonable doubt. Nat Genet 2007;39:813–815.

53 Janssens ACJW, Pardo MC, Steyerberg EW, van Dujin CM: Revisiting the clinical validity of multiplex genetic testing in complex diseases. Am J Hum Genet 2004;74:585–588.