

Review

The genetics of non-insulin-dependent diabetes mellitus: tools and aims

M.I. McCarthy¹, P. Froguel², G.A. Hitman¹

¹ Medical Unit, London Hospital Medical College, London, UK

² Centre d'Etude du Polymorphisme Humain, Paris, France

Non-insulin-dependent diabetes mellitus (NIDDM) represents a major health problem in both developed and underdeveloped countries and efforts to understand the pathogenesis of glucose intolerance provide a significant scientific challenge for the coming decade. Few facts concerning the development of NIDDM are beyond dispute or controversy: in particular, the relative importance of genetic and environmental contributions to the pathogenesis of NIDDM has engendered fevered debate between advocates of nature and of nurture. It is not within the scope of this review to rehearse these arguments but it seems overwhelmingly likely that, in the majority of populations, both genetic and environmental influences interact to determine individual risk of NIDDM [1].

The focus of most recent genetic research in NIDDM has been the identification of specific genetic determinants which influence individual predisposition to glucose intolerance: the resulting progress has been sluggish [1]. The hope is that, ultimately, such knowledge will clarify the sequence of pathophysiological events which produces glucose intolerance and lead to more rational treatments and preventive approaches. It might be feasible, if the number of genes involved is limited, to use genetic markers to predict individual risk of future diabetes, permitting specific targeting of such measures.

Key words Non-insulin-dependent diabetes mellitus, genetics, linkage analysis, population association studies, complex inheritance.

Corresponding author: Dr. G. Hitman, Medical Unit, Royal London Hospital, London E1 1BB, UK

Abbreviations: MODY, Maturity onset diabetes of the young; ARP, affected relative pair; ADA, adenosine deaminase; SSCP, single stranded conformational polymorphism, MIDD, maternally inherited diabetes with deafness.

What are the problems in the genetic study of NIDDM?

Genetic heterogeneity

Studies into the genetics of NIDDM are hindered by a number of obstacles (Table 1): the greatest of these is that NIDDM is, without doubt, not a single homogeneous disease. On the contrary, heterogeneity is evident in virtually all aspects of NIDDM (Table 2) and is likely to be the result of both genetic heterogeneity and environmental diversity. Genetic heterogeneity may be allelic (phenotypic diversity is explained by different alleles at the same locus as in Duchenne/Becker muscular dystrophy) and/or non-allelic (explained by the action of a number of different loci). In the latter case, a number of genes (polygenes) might interact to determine individual glucose tolerance status: different genes (or combinations of genes) might be important in different diabetic individuals and the situation might be further complicated by ethnic differences in the most significant diabetogenes.

Confused pathophysiology and diagnosis

A second difficulty lies in ignorance regarding the 'primary' defect in NIDDM. Although this may simply reflect the fact that there are diverse pathophysiological routes to glucose intolerance, dispute continues between those who hold that the beta cell is the site of the fundamental defect and those who feel that it is insulin action which is defective [2]. Certainly once hyperglycaemia is present, defects of both insulin action and secretion are evident: effort has therefore been directed towards the study of individuals who, though normoglycaemic, are at an increased risk of future diabetes (e.g. offspring of diabetic parents) in the hope that the earliest (poten-

Table 1. Principal problems with genetic studies in NIDDM

Problem	Features
Heterogeneity	Genetic heterogeneity – allelic – non allelic Environmental diversity Gene-environment interactions
Complex pathophysiology	Complex network of intermediate traits Feedback effects of disease on traits and gene expression Critical pathophysiological events not known
Family collection	Late onset of disease Premature mortality
Diagnostic difficulties	Tests of and criteria for glucose tolerance Subtypes of diabetes

tially primary) defect will be identified. However, the concept of a primary defect may itself be misleading [2]. Consider the hypothesis that NIDDM is due to a defect of betacell number, perhaps as a result of mutations in a gene determining the regenerative capacity of the beta cell following insult. Owing to a redundancy in beta-cell provision, individuals inheriting such a diabetogene might maintain normoglycaemia whilst they remain insulin-sensitive: however, development of insulin resistance (e.g. due to obesity) in later life would expose the beta-cell insufficiency with consequent hyperglycaemia. To the geneticist, the beta-cell defect is 'primary'. To the physiologist using standard techniques (e.g. clamps) to study the patient longitudinally, the earliest defect detected may be insulin resistance, since methodologies to detect subtle beta-cell dysfunction are less well-developed or have limited applicability in large studies (e.g. measures of insulin pulsatility).

Absence of a clear pathophysiological framework for NIDDM means that the easy route (i.e. from defective protein to defective gene, as followed, for example, in thalassaemia) to the identification of diabe-

togenes has not been available; instead, the search has been conducted at the genetic level, seeking mutations which associate with or are linked to diabetes. Identification of the gene(s) harbouring disease-predisposing mutations should ultimately permit reconstruction of the pathophysiological sequence of events.

Further confusion surrounds the diagnosis of diabetes in general and of NIDDM in particular. Glucose levels are continuous within populations and current diagnostic criteria are based on future risk of specific diabetic complications [3]. It is not inevitable that such diagnostic thresholds have any particular significance from the point of view of genetic or environmental parameters determining the development of hyperglycaemia: these are likely to influence glucose levels throughout the range of glucose tolerance. Any attempt to dichotomise glucose tolerance into 'diabetic' and 'normal' is arbitrary and squanders information. Besides, many of the diagnostic tools are poorly reproducible and inaccurate [4].

If NIDDM represents a heterogeneous group of conditions it would be desirable to dissect out distinct disease-entities: careful characterisation might reduce heterogeneity and allow closer correlation of genotype and phenotype. This has proved a useful strategy in the case of maturity-onset diabetes of the young (MODY) [5,6] and maternally-inherited diabetes and deafness (MIDD) [7] which represent relatively discrete subtypes of NIDDM, but further extension of this approach within typical NIDDM presents difficulties. Even the 'classic' division between the two major forms of diabetes (IDDM and NIDDM) is increasingly suspect [8, 9]: a significant proportion of subjects with clinically-diagnosed NIDDM seem to have a form of slowly-evolving IDDM [10]. Although identification of such an 'autoimmune' subgroup might permit clarification of the NIDDM phenotype, the boundaries are indistinct and many individuals will resist categorisation.

Table 2. Evidence for heterogeneity within NIDDM

Aspect of study	Findings	Examples
Phenotypic	NIDDM patients, though obese as a group, show wide anthropometric variability	Obese vs thin, early onset vs late onset
Physiological	Studies of intermediary metabolism reveal no single biochemical marker of pre-diabetes or diabetes	Predominant insulin resistance vs insulin secretory failure
Ethnic	NIDDM is present in high prevalence in a wide range of unrelated ethnic groups? Convergent evolution	Pimas, South Indians, Nauruans, Australian Aboriginals
Inherited diseases	NIDDM features in a number of inherited diseases of widely varied pathogenesis	Prader Willi, mitochondrial myopathies, insulin resistance syndromes
Genes in MODY	At least three genes contribute to the genetic basis of MODY	Glucokinase, ADA-linked, unlinked to either
Genes in NIDDM	Although a number of minor genes have been identified, no major gene has been uncovered	Insulin, insulin receptor, glucokinase, mitochondrial, HLA

The association between NIDDM and obesity adds further complexity since obesity is itself a multifactorial disease with a strong genetic component [11]: the search for diabetogenes (at least in obese NIDDM subjects) is therefore complicated by the contribution of obesity-related genes. The hypothesis that, in some ethnic groups at least, NIDDM is but one component of a larger syndrome (syndrome X) [12] suggests that hyperinsulinaemia rather than hyperglycaemia might be the appropriate phenotype to study.

One solution to these diagnostic difficulties (which share the effect of weakening correspondence between genotype and phenotype) would be to follow the course charted by researchers seeking genetic determinants of vascular disease [13]. Recognising that atheroma is a complex process resulting from the interaction of multiple genetic and environmental variables, research has focused on intermediate traits which lie between the genetic and the disease level e.g. cholesterol, fibrinogen levels. Correspondence between genotype and traits will inevitably be closer than between genotype and clinical disease. Equivalent intermediate traits in NIDDM might be insulin secretory capacity, beta-cell mass, insulin sensitivity, etc.; these have the disadvantage of being less easy to measure with accuracy and reproducibility. A further drawback is that in diabetes the hierarchy of genes – traits – disease is complicated by feedback effects whereby hyperglycaemia modifies the intermediate traits through physiological mechanisms (glucotoxicity) or through medical intervention (e.g. insulin, weight loss). Consequently intermediate traits measured in glucose-intolerant subjects reflect both ‘bottom-up’ and ‘top-down’ influences and are difficult to interpret: if intermediate traits relevant to NIDDM are to be studied, paradoxically it is individuals with normal glucose tolerance who need to be characterised.

Difficulties accumulating family resources for study

Linkage analysis, the standard methodology used in the detection of disease susceptibility genes generally relies on the collection of families segregating for disease. NIDDM is a disease of late onset associated with premature mortality: both these factors militate against the collection of multigenerational families optimal for linkage analysis.

Genetic models in NIDDM and the limitations of linkage analysis

The loci responsible for a variety of single-gene disorders e.g. cystic fibrosis, familial polyposis coli have been identified in recent years by linkage techni-

ques. Linkage analysis seeks to provide evidence for the cosegregation within families of a pair of loci (in this context, one is a postulated disease-predisposing locus, the other a marker locus at known genomic location): cosegregation implies that the two loci are only rarely separated by meiotic recombination and that they lie close together in the genome. Diseases successfully mapped in this way have features which make them well suited to linkage analysis: early disease-onset, high penetrance, clear definition, homogeneity and classic Mendelian inheritance (i.e. they have ‘simple’ genetic models). For these same reasons, linkage analysis has been the method of choice for the identification of diabetes susceptibility genes in MODY [5, 6, 14] which has (by definition) early onset and autosomal dominant inheritance.

Simple genetic models may also adequately describe the segregation of NIDDM in populations at extreme risk of NIDDM (e.g. Pima Indians and Nauruan islanders): in both of these populations, segregation analysis has been supportive of a major gene determining diabetes susceptibility [15, 16]. The bimodality of glucose tolerance observed in several high-risk populations has also been regarded as confirming the segregation of a single gene [1]: however, an alternative explanation is that intermediate levels of glucose tolerance are ‘unstable’ and that bimodality is an artefact of rapid transition from normal to abnormal glucose tolerance.

By extension from the success of linkage analysis in single gene disorders, analogous ‘simple’ genetic models have regularly been employed in linkage analyses involving typical NIDDM, despite evidence that in most populations at less extreme risk of NIDDM, inheritance is inconsistent with segregation at a single locus [17,18]. NIDDM in these populations displays complex inheritance patterns: it does not clearly segregate within families like the simple ‘Mendelian’ diseases, rather it aggregates. Individuals who carry a particular diabetogene may not express disease (due to incomplete penetrance) and affected individuals will not always possess the diabetogene under study (due to phenocopies and genetic heterogeneity). Additionally, observations supporting excess maternal transmission of NIDDM [19, 20] suggest further mechanisms (such as imprinting and mitochondrial mutations) which may need to be incorporated in any realistic genetic model of NIDDM.

Such effects as those described can generally be accommodated within linkage analyses but the price to be paid is usually a substantial reduction in power: the number of pedigrees (or affected relative pairs (ARPs)) required to detect or exclude linkage escalates rapidly as the proportion of the disease due to the locus under study declines [21, 22]. For example, assuming a locus linked to disease in 50% of pedigrees with NIDDM (which many would regard as optimistic) well over 100 nuclear families (comprising

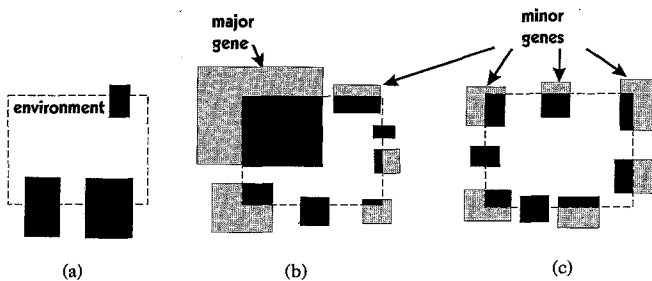


Fig. 1a-c. Representation of possible genetic models for NIDDM. In each of the three models shown, the dotted line encloses an area denoting environmental influences predisposing to NIDDM. Each small square represents a single theoretical diabetogene, which may be expressed as diabetes (black areas, generally but not exclusively in the presence of permissive environment) or as normal glucose tolerance (grey areas). Model (a) represents the situation in MODY and possibly in populations at extreme risk of NIDDM. The other models represent alternative models for typical NIDDM with (b), and without (c) a major gene locus for NIDDM

both parents and four offspring, two affected and two unaffected) are required to detect linkage to that locus under reasonable assumptions of marker information content and gene penetrance [22]. Such numbers may not appear excessive but the late onset and early mortality of NIDDM make the collection of even modest numbers of pedigrees suitable for linkage analysis surprisingly troublesome [23].

Parametric (or classic) linkage analyses require that disease-locus parameters such as the penetrance function and gene frequency are specified before analysis is performed. Unfortunately, these values are unlikely to be known with any certainty in complex diseases such as NIDDM: the researcher may be obliged to make an educated guess or to test multiple disease-locus models (differing perhaps in mode of inheritance, number of phenocopies, levels of penetrance, diagnostic categories) for linkage to the candidate gene or marker locus of interest. Two-point linkage analysis (that is, tests of linkage between a disease locus and a single marker locus) is relatively robust to misspecification of the genetic model, so this is a reasonable approach provided modifications are made to avoid increasing the type 1 error rate [24]. This is not equally true of multipoint mapping (where an attempt is made to place a disease locus in relation to a number of linked marker loci): failure to allow adequately for incomplete penetrance at a disease locus may force the analysis to deduce that localisation of the disease-locus between a pair of flanking markers would only be possible in the unlikely event of double recombination: linkage may in consequence be falsely rejected [24].

Less widely appreciated are limitations in interpreting analyses which claim to exclude linkage or of results in the intermediate range which neither definitively support nor exclude linkage. A locus which

contributes to disease in only a minority of families or one which merely increases the risk of disease (without being either necessary or sufficient for disease expression) [25] may be difficult or impossible to detect by linkage methods when the majority of evidence points away from close linkage. As an extreme illustration, if inappropriate (high penetrance, low phenocopy) genetic models were used for linkage analysis of a complex, heterogeneous (but entirely genetically determined) disease, it is conceivable that disease loci could, paradoxically, be excluded from the entire genome [26]. Exclusion mapping as employed in simple Mendelian diseases is not possible in complex diseases since exclusion of a candidate locus under a single-locus model for NIDDM, can never exclude a minor genetic role for that locus.

It remains unclear at this stage which of these two statements concerning NIDDM is nearer the true state of affairs (Fig. 1.): a) the genetic predisposition to typical NIDDM is determined by a large number of different loci (a combination of polygenicity within the individual, and of multigenicity and ethnic heterogeneity within the population) each individually responsible for a minor portion of the predisposition: if so, then standard linkage techniques will be of limited value unless the complexity and heterogeneity within NIDDM can be in some way simplified (as in the case with MODY and MIDD), b) there are a limited number of major genes (perhaps acting on a polygenic background) individually accounting for a sizeable fraction of the genetic contribution. If so, linkage analysis techniques should permit their identification, provided sufficient families (or ARPs) are gathered and typed for sufficient polymorphic markers. This optimistic view opines that the situation is less complex at the genomic level than it appears phenotypically, citing the examples of cystic fibrosis where allelic variation at a single locus explains observed phenotypic diversity and of Hirschsprung's disease, previously thought to be a polygenic disease, which has been mapped to a single locus.

In the meantime, the challenge is to improve the linkage methodology to deal more effectively with the problems posed by complex diseases. One approach is to use segregation analysis techniques to determine relevant parameters of the disease-locus as a preliminary to linkage analysis. Using the computer program COMDS in South Indian pedigrees segregating for NIDDM, support was obtained for a major genetic locus accounting for around 40% of the variance in the liability to affection and acting on a background of polygenes, shared family environment and random environmental influences [27]. Such a disease locus may be satisfactorily represented within linkage analyses by ascribing reduced penetrance and high phenocopy rates to the disease locus (albeit with attendant loss of power). Alternatively,

it may be possible to combine segregation and linkage analysis in a single procedure, an approach which maximises the information extracted from segregating pedigrees [28].

Secondly, linkage analysis techniques better suited to the study of complex disease are becoming available [29, 30]; these provide a significant boost in the power to detect multiple loci contributing to disease.

Non-parametric methods of linkage analysis (affected sibling-pair, affected relative pair (ARP) methods) [31] offer a realistic alternative to classic methods given concerns about inadequate specification of the genetic model and have proved successful in complex diseases such as hypertension [32] and insulin-dependent diabetes [33]. These techniques rely on the detection of regions of increased identity-by-descent in ARPs, and require no a priori assumptions regarding dominance, allele frequencies or penetrance functions (hence, the term 'non parametric'). ARPs are easier to collect than entire pedigrees and, since affected individuals are more informative for linkage than their unaffected relatives (who may yet develop disease) such an approach is more economical of laboratory effort. Methodological uncertainties remain concerning the appropriate corrections to be made when identity-by-state (as opposed to identity-by-descent) relationships are employed and in the statistical treatment of multiple sibships. The total number of ARPs required for such an analysis in a multifactorial disease depends on the genetic architecture of the disease and on the density and information content of the marker map available, but 200–300 ARPs should suffice to detect a locus which doubles the risk of disease [21, 34].

Total genome search vs candidate gene approach

Overall strategies for the identification of NIDDM-susceptibility genes are represented by two extremes: the total genomic search and the candidate gene approach [35]. The total genomic search option relies (at present) on the application to NIDDM of techniques which have been successful in simple Mendelian disorders. Families segregating for disease are typed at a large number of highly polymorphic markers scattered throughout the genome and regions linked to disease are identified; given the reliance on linkage methodology, the limitations in the analysis of 'typical' NIDDM are clear. However, such limitations need not apply in models of NIDDM such as MODY and may be less relevant in high-risk groups such as the Pima Indians. Indeed, linkage between glucose intolerance and the adenosine deaminase (ADA) locus on chromosome 20q has been conclusively demonstrated using just these techniques in the RW pedigree, a massive family segregating for MODY [14]. It should be noted

that identification of the chromosomal region linked to disease is only a first step: homing in on the specific defective gene within such a region (which may be several megabases long) remains a formidable task.

A complementary approach is to apply methods of total genomic searching to animal models of NIDDM; human chromosomal regions syntenic (homologous) with regions implicated in the causation of diabetes in rodent models of NIDDM would then be candidates for study. This approach has proved successful in hypertension, a major gene for which maps to chromosome 10 in the stroke-prone spontaneously hypertensive rat [36]. The homologous human region lies on chromosome 17q and includes the angiotensin converting enzyme gene.

Positive findings will provide invaluable insights into pathophysiological routes capable of leading to diabetes but there is no assurance that genes implicated in these 'model' disorders will also be major players in typical NIDDM. The loci so far implicated in around 50% of families with MODY (glucokinase and ADA-linked) are responsible, at most, for 5% of the genetic contribution to NIDDM [37]. New approaches to total genomic searching have been advocated but remain unproven: these include genomic mismatch scanning [38] and linkage disequilibrium mapping [39] (see below).

Most studies of NIDDM in populations not at extreme risk of the disease have followed a candidate gene approach. Candidates have generally been promoted because of their known (or presumed) biochemical function (e.g. insulin, insulin receptor, glucose transporters); however, the choice of potential candidates for a role in the development of NIDDM is a wide one (Table 3).

Non linkage-based approaches

Association studies

Population association (case-control) studies have been extensively used in the study of NIDDM [1]: the image of such studies has become tarnished owing to the perception that they generate inconsistent results (some of this inconsistency might reflect the heterogeneity of diabetes e.g. between ethnic groups). In these studies, cases (diabetic subjects) and controls (e.g. normoglycaemic individuals with no family history of diabetes) are compared by typing at a polymorphic marker closely linked to a candidate gene of interest. Any significant differences in allele and/or genotype frequencies between the two groups are generally considered to indicate that the marker lies in linkage disequilibrium with a pathogenic mutation in the adjacent candidate gene (or conceivably another gene in the vicinity).

Table 3. Potential mechanisms by which candidate genes for NIDDM may be identified

Type of candidate	Reason for candidacy	Examples
Biochemical	Known or presumed function	Insulin receptor GLUT 2
Models of NIDDM	Locus implicated in MODY or region syntenic to locus in animal model	ADA Glucokinase
Inherited diabetes	Locus implicated in inherited disease which includes diabetes	Mitochondrial
Associated condition	Loci involved in obesity, syndrome X or appetite	Neuropeptide Y Dopamine receptor D2 Lipoprotein lipase B3 adrenergic receptor Glucocorticoid receptor
Differential expression	Product differentially expressed in diabetic and normal tissues	Rad

This is, however, not the only explanation of a positive association [24]: spurious positive results may result from chance (type 1 error) or from inadequate matching of case and control populations (particularly for ethnicity). Conversely, false negative results may arise in such a study if the number of subjects studied is inadequate or if the control group is diluted by individuals destined to develop diabetes in the future. More importantly, a major genetic effect may be obscured when marker alleles and candidate gene mutations are in linkage equilibrium and/or when there are multiple mutations in the candidate gene with no single mutation-marker association paramount. Indeed, the power of association studies is critically dependent on the linkage relationships between alleles at the marker and the candidate gene [40]. Under the most auspicious circumstances e.g. mutation in disequilibrium with a rare marker allele, association studies have the capacity to detect subtle genetic contributions to disease which would almost certainly remain undetected by linkage analysis [25].

It can be argued that the disappointing history of association studies in NIDDM limits their future utility. This is probably an excessively pessimistic view: when the correct candidate gene is selected (e.g. apolipoprotein E and Alzheimer's disease, HLA and IDDM, angiotensin converting enzyme and cardiovascular disease) association studies have proven worth [41, 42]. However, strenuous efforts are essential to reduce sources of error, through judicious choice of control subjects (matched for ethnicity, body mass index etc.) or, where the data are available, by treating unaffected family members as control subjects [43]. Cross-sectional studies of (normal) populations may allow individual intermediate traits to be analysed at suitable candidate loci e.g. individuals in the upper and lower deciles for a measure of insulin resistance might be compared. Subphenotyping within NIDDM should help to reduce complexity: for example comparison of diabetic and non-diabetic obese subjects may 'amplify' the effects of genes determining insulin secretion. Fi-

nally, the development of more detailed genomic maps offers the opportunity to exploit more effectively the linkage disequilibrium relationships on which association studies depend. The use of association studies to fine map disease loci in isolated populations (linkage disequilibrium mapping) has already been demonstrated [39]. In theory, similar techniques could be applied to search chromosomal regions of interest (or even the entire genome) to identify stretches which associate with NIDDM: this approach would be most likely to bear fruit in isolated populations likely to show a strong founder effect for a limited number of diabetogenic mutations.

Molecular scanning and sequencing

The development of novel techniques and increasing automation have made practical more direct study of candidate genes and a variety of methodologies are available for the detection of mutations [44]. The most widely-applied approach in NIDDM has been to screen candidate loci for mutations using single stranded conformational polymorphism (SSCP) followed by sequencing to identify the molecular basis of any variants identified. One success of this direct approach has been the identification of mutations in the mitochondrial genome in pedigrees with MIDD [7].

One disadvantage all these methods share is that *cis*-acting mutations (e.g. mutations upstream of the gene which might influence transcription) will not be identified unless specifically sought: the other major obstacle lies in the evaluation of any mutations identified. It can be surprisingly difficult to establish whether or not a mutation identified in a candidate gene is contributing to disease [45]. The fact that a mutation is present at higher frequency in cases than control subjects is not itself proof since the mutation/polymorphism may simply be in linkage disequilibrium with a mutation elsewhere in the gene. Co-segregation of the mutation with disease may be im-

Glossary

Anticipation	a tendency apparent in certain diseases for earlier disease onset and/or increased disease severity in successive generations
<i>Cis</i> -acting mutation	mutations lying up or downstream of the coding portion of the gene e.g. in the promoter region which might influence gene function
Differential display	used to detect differences in mRNA expression between cell types: a subset of the mRNA from the cells is amplified by the polymerase chain reaction using a series of arbitrary primers, and comparison made between the products displayed on a polyacrylamide gel
Exclusion mapping	the use of negative linkage analysis results to build up a map of areas from which a disease gene has been excluded
GMS	genomic mismatch scanning: a methodology under development which aims to rapidly identify those genomic regions shared by pairs of affected relatives: such regions are likely to be enriched for disease-susceptibility genes
Identity-by-descent	when a pair of relatives shares a particular allele at a given locus, that allele is considered as shared 'identity-by-descent'(IBD) if it represents the 'same' allele inherited from a single ancestor
Identity-by-state	when a pair of relatives shares a particular allele at a given locus, that allele is considered as shared 'identity-by-state'(IBS) if it is not possible to determine whether the allele represents the 'same' allele inherited from a single ancestor or 'different' alleles derived from two different ancestors
Illegitimate transcription	Inappropriate expression of a gene: e.g. cells from tissues which do not usually express glucokinase, may do so when established as cell lines
Linkage disequilibrium mapping	a technique for disease gene mapping based on the detection of linkage disequilibrium effects in the genomic region neighbouring an ancestral disease-predisposing mutation
Linkage equilibrium	when alleles at two linked loci show random association, they are in linkage equilibrium: if certain allele combinations (haplotypes) occur more or less often than expected, the loci are in linkage disequilibrium and alleles at one locus will act as 'markers' for alleles at the other
Multipoint mapping	linkage analysis applied simultaneously to multiple loci: often used to place a disease gene within a previously ordered map of marker loci
Non-parametric linkage analysis	linkage analysis techniques which require no assumptions to be made concerning the mode of inheritance of the disease locus (e.g. sib-pair methods)
Parametric linkage analysis	linkage analysis methodologies which require the parameters of the disease locus to be explicitly stated before analysis (as in lod score methods)
Penetrance	the probability of observing a particular phenotype, given the corresponding genotype e.g. a fully penetrant dominant gene would imply that presence of one or more mutant alleles always leads to the disease phenotype
Phenocopies	affected individuals in whom disease status is due to environmental rather than genetic effects (sporadic cases)
Phenotype	the observable characteristics of an individual
Segregation analysis	techniques which aim to determine the mode of inheritance of a condition from the study of the patterns of segregation within families
SSCP	single-stranded conformational polymorphism detection: allows a candidate gene to be screened for the presence of mutations, by making use of the fact that migration of single-stranded DNA through a non-denaturing polyacrylamide gel depends on sequence as well as size
Subtractive hybridisation	used to detect differences in mRNA expression between cell types: single stranded DNA is prepared from RNA and hybridised in asymmetric quantities to produce a 'subtraction' library enriched for those mRNA species relatively overexpressed in one of the cells
Synteny	refers to genes located on the same chromosome: also used to denote chromosomal regions which are homologous across species
Trait	any phenotypic characteristic determined by gene expression

possible to demonstrate if the gene is only a minor player. Assessment of the consequences of the mutation on secondary and tertiary protein structure may provide supportive evidence but ultimately expression studies may be necessary to support a pathogenic role.

Prospects for the future

Rapid progress in the development of laboratory and analytical techniques offers the expectation that significant advances will be made in the study of the genetic basis of NIDDM. Exemplars such as MODY

and animal models represent experiments of nature with simpler genetic architecture than typical NIDDM which are tractable to present methodologies in linkage analysis (including total genomic search). Just as the identification of glucokinase in MODY pedigrees has provided vital clues to processes which may fail in NIDDM, characterisation of the disease-locus linked to ADA (and of further loci linked to MODY) will provide additional, perhaps unexpected, insights. Even if such genes play only a small role themselves in the genetic predisposition to NIDDM, they may highlight more rewarding candidate genes. Similarly, the recognition of mitochondrial defects in MIDD raises interesting questions about the role of mitochondrial defects (encoded in either the mitochondrial or the nuclear genome) in determining insulin secretion and action.

The total genomic search option in typical NIDDM remains problematic with current technology principally due to uncertainties concerning the appropriate genetic model to apply. Means by which to dissect NIDDM into more homogeneous subtypes may be revealed through physiological studies, although the basis for any such categorisation must remain simple and acceptable to patients if it is to be applied to characterise entire families. As denser, more complete genetic maps are developed and as automation progresses, the mechanics of typing will cease to be the rate-limiting step: genetic heterogeneity within NIDDM and the consequent need for large numbers of pedigrees (or ARPs) is more likely to prejudice success. Collection of suitable, accurately-characterised family material remains a significant obstacle to progress and international collaborative input to establish such resources (as cell lines) is essential. In this regard, the collection of ARPs may be more practical: techniques currently under development offer the promise of increasing the power of genomic searching by ARPs (e.g. genomic mismatch scanning [38]). Linkage disequilibrium mapping in isolated populations remains an exciting but unproven alternative strategy in complex disease.

In the meantime, the candidate gene approach will remain the mainstay of most NIDDM research. Information from MODY and from animal models will provide new candidate genes for study, as will further elucidation of the cellular mechanisms of insulin action and secretion. Techniques such as subtractive hybridization [46] or differential display [47] allow patterns of gene expression in relevant tissues or cell lines to be analysed according to genetic background (e.g. two strains of rodent, diabetic vs non-diabetic humans) or environmental conditions (e.g. ambient glucose concentration) to identify genes of potential relevance to the diabetic state. Systematic scanning for mutations in the c(complementary)DNA of potential candidate genes obtained from readily accessible tissues (muscle, adipose) offers promise: alterna-

tively when there is no easy access to tissues expressing relevant candidates (e.g. beta cells) advantage can be taken of illegitimate transcription to amplify (using the polymerase chain reaction) the cDNA from lymphoblastoid cell lines [48].

Any postulated candidate gene may ultimately prove to be a major diabetogene, a minor modifying gene or (in most cases) to have no role at all in the development of NIDDM: given the relative strengths and weaknesses of association, scanning and linkage analyses in these alternative circumstances it is clear that no single methodology can claim universal discrimination and that the contribution of any candidate gene can only be reliably assessed after a number of studies conducted using different methodologies in different ethnic groups (as illustrated in recent studies of the insulin gene in IDDM [33, 49]).

Finally, we must not close our eyes to the possibility of novel genetic mechanisms leading to disease. The excess family history of NIDDM inherited through the maternal line [19, 20] is consistent with mutations within the mitochondrial genome [7] or with genetic imprinting. Recently a number of neurological diseases have been attributed to unstable trinucleotide repeats which provide a molecular basis for the clinical observation of anticipation [50].

The elucidation of the genetic basis of NIDDM remains an important and achievable goal. To realise that goal, international collaborations between multidisciplinary groups will be essential together with an aggressive approach to funding which embraces academic bodies, governments and industry.

Acknowledgements. M.McC is a Medical Research Council Training Fellow. We thank Dr. R. Turner (Oxford), Dr. S. O'Rahilly (Cambridge) and Professor N. Morton (Southampton) for stimulating and helpful discussions over the years; and Dr. A. Hattersley (Birmingham) for comments on the manuscript.

References

1. McCarthy M, Hitman GA (1993) The genetic aspects of non-insulin-dependent diabetes mellitus. In: Leslie RDG (ed) *The causes of diabetes*. Wiley, London, pp 157-183
2. Turner R, O'Rahilly S, Levy J, Rudenski A, Clark A (1989) Does type II diabetes arise from a major gene defect producing insulin resistance or β -cell dysfunction? In: Nerup J, Mandrup-Poulsen T, Hökfelt B (eds) *Genes and gene products in the development of diabetes mellitus*. Elsevier, Amsterdam, pp 171-183
3. World Health Organisation Study Group (1985) *Diabetes mellitus*. WHO Tech Rep Ser, no 727
4. Home P (1988) The OGTT: Gold that does not shine. *Diabet Med* 5: 313-314
5. Froguel Ph, Vaxillaire M, Sun F et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* 356: 162-164

6. Hattersley AT, Turner RC, Permutt MA et al. (1992) Linkage of type 2 diabetes to the glucokinase gene. *Lancet* 339: 1307–1310
7. van den Ouweland JMW, Lemkes HHPJ, Ruitenbeek W et al. (1992) Mutation in mitochondrial tRNA^{Leu(UUR)} gene in a large pedigree with maternally-transmitted type II diabetes mellitus and deafness. *Nature Genetics* 1: 368–371
8. Tuomilehto-Wolf E, Tuomilehto J, Hitman GA et al. (1993) Genetic susceptibility to non-insulin dependent diabetes mellitus and glucose intolerance are located in the HLA region. *BMJ* 307: 155–159
9. Freinkel N, Metzger BE, Phelps RL et al. (1985) Gestational diabetes mellitus: heterogeneity of maternal age, weight, insulin secretion, HLA antigens and islet cell antibodies and the impact of maternal metabolism on pancreatic beta-cell and somatic development on the offspring. *Diabetes* 34 [Suppl 2]: 1–7
10. Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR (1993) Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. *Diabetes* 42: 359–362
11. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE (1990) The body-mass index of twins who have been reared apart. *New Engl J Med* 322: 1483–1487
12. Reaven G (1988) Role of insulin resistance in human disease. *Diabetes* 37: 1595–1607
13. Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL (1992) Biological complexity and strategies for finding DNA variations responsible for inter-individual variation in risk of a common chronic disease, coronary artery disease. *Ann Med* 24: 539–548
14. Bell GI, Xiang K-S, Newman MV et al. (1991) Gene for non-insulin-dependent diabetes mellitus (maturity onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc Natl Acad Sci USA* 88: 1484–1488
15. Yamashita T, Mackay W, Rushforth N, Bennett P, Houser H (1984) Pedigree analyses of non-insulin dependent diabetes mellitus (NIDDM) in the Pima Indians suggest dominant mode of inheritance. *Am J Hum Genet* 36 [Suppl]: 183S (Abstract)
16. Serjeantson SW, Zimmet P (1991) Genetics of non-insulin dependent diabetes mellitus in 1990. *Bailliere's Clin Endocrinol Metab* 5: 477–493
17. Rich SS (1990) Mapping genes in diabetes: genetic epidemiological perspective. *Diabetes* 39: 1315–1319
18. Cook JTE, Hattersley AT, Levy JC et al. (1993) Distribution of type II diabetes in nuclear families. *Diabetes* 42: 106–113
19. Thomas F, Balkau B, Vauzelle-Kervroedan F, Papoz L and the CODIAB-INSERM-Zeneca Study Group (1994) Maternal effect and familial aggregation in NIDDM. The CODIAB study. *Diabetes* 43: 63–67
20. Alcolado JC, Alcolado R (1991) Importance of maternal history of non-insulin dependent diabetic patients. *BMJ* 302: 1178–1180
21. Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46: 229–241
22. Lathrop GM (1989) The power of linkage studies of Mendelian and multifactorial genetic disease in man. In: Nerup J, Mandrup-Poulsen T, Hökfelt B (eds) *Genes and gene products in the development of diabetes mellitus*. Elsevier, Amsterdam, pp 373–379
23. Cook JTE, Page RCL, O'Rahilly S et al. (1993) Availability of type II diabetic families for detection of diabetes susceptibility genes. *Diabetes* 42: 1536–1543
24. Ott J (1991) *Analysis of human genetic linkage*. Revised edn. Johns Hopkins University Press, Baltimore
25. Greenberg DA (1993) Linkage analysis of “necessary” disease loci versus “susceptibility” loci. *Am J Hum Genet* 52: 135–143
26. Clerget-Darpoux F, Bonaïti-Pellié C (1993) An exclusion map covering the whole genome: a new challenge for genetic epidemiologists? *Am J Hum Genet* 52: 442–443 (Letter)
27. McCarthy M, Hitman GA, Shields DC, Morton NE, Snehalatha C, Mohan V, Ramachandran A, Viswanathan M (1994) Family studies of non-insulin-dependent diabetes mellitus in South Indians. *Diabetologia* (in press)
28. Risch N (1984) Segregation analysis incorporating linkage markers. I. single-locus models with an application to type 1 diabetes. *Am J Hum Genet* 36: 363–386
29. Schork NJ, Boehnke M, Terwilliger JD, Ott J (1993) Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. *Am J Hum Genet* 53: 1127–1136
30. Lander ES, Botstein D (1986) Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. *Cold Spring Harbor Symposia on Quantitative Biology*, Volume LI: 49–62
31. Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. *Am J Hum Genet* 42: 315–326
32. Jeunemaitre X, Soubrier F, Kotelevtsev YV et al. (1992) Molecular basis of human hypertension: role of angiotensinogen. *Cell* 71: 169–180
33. Julier C, Hyer RN, Davies J et al. (1991) Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* 354: 155–159
34. Feingold E, Brown PO, Siegmund D (1993) Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am J Hum Genet* 53: 234–251
35. Clerget-Darpoux F, Bonaïti-Pellié C (1992) Strategies based on marker information for the study of human diseases. *Ann Hum Genet* 56: 145–153
36. Jacob HJ, Lindpaintner K, Lincoln SE et al. (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* 67: 213–224
37. Stoffel M, Bell KL, Blackburn CL et al. (1993) Identification of glucokinase mutations in subjects with gestational diabetes mellitus. *Diabetes* 42: 937–940
38. Nelson SF, McCusker JH, Sander MA, Kee Y, Modrich P, Brown PO (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nature Genetics* 4: 11–18
39. Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1993) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics* 2: 204–211
40. Cox NJ, Bell GI (1989) Disease associations: chance, artifact or susceptibility genes? *Diabetes* 38: 947–950
41. Scott J (1993) Apolipoprotein E and Alzheimer's disease. *Lancet* 342: 696
42. Cambien F, Pourier O, Lecerf L et al. (1992) Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature* 359: 641–644
43. Knapp M, Seuchter SA, Baur MP (1993) The haplotype-relative-risk (HRR) method of analysis of association in nuclear families. *Am J Hum Genet* 52: 1085–1093
44. Grompe M (1993) The rapid detection of unknown mutations in nucleic acids. *Nature Genetics* 5: 111–117

45. Almind K, Bjørbaek C, Vestergaard H, Hansen T, Echwald S, Pedersen O (1993) Aminoacid polymorphisms of insulin receptor substrate-1 in non-insulin dependent diabetes mellitus. *Lancet* 342: 828-832
46. Reynet C, Kahn CR (1993) Rad: a member of the ras family overexpressed in muscle of type II diabetic humans. *Science* 262: 1441-1444
47. Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257: 967-971
48. Sun F, Knebelmann B, Pueyo ME et al. (1993) Deletion of the donor splice site of intron 4 in the glucokinase gene causes maturity-onset diabetes of the young. *J Clin Invest* 92: 1174-1180
49. Lucassen AM, Julier C, Beressi J-P et al. (1993) Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nature Genetics* 4: 305-310
50. Sutherland GR, Richards RI (1993) Dynamic mutations on the move. *J Med Genet* 30: 978-981