The genetic basis for type 1 diabetes

Kay L. Mehers and Kathleen M. Gillespie*

Diabetes and Metabolism, Department of Clinical Science at North Bristol, University of Bristol, UK

Background: Type 1 diabetes (T1D) is characterized by autoimmune destruction of insulin-producing β-cells in the pancreas resulting from the action of environmental factors on genetically predisposed individuals. The increasing incidence over recent decades remains unexplained, but the capacity of identifying infants at highest genetic risk has become an increasing requirement for potential therapeutic intervention trials.

Sources of data: Literature searches on T1D and genes were carried out, and key papers since the 1970s were highlighted for inclusion in this review.

Areas of agreement: Early genetic studies identified the most important region for genetic susceptibility to T1D—the human leukocyte antigen genes on chromosome 6; later shown to contribute approximately half of the genetic determination of T1D. The other half is made up of multiple genes, each having a limited individual impact on genetic susceptibility.

Areas of controversy: Historically, there have been many controversial genetic associations with T1D, mostly caused by underpowered case–control studies but these are now decreasing in frequency.

Areas of growth: The functional effect of each gene associated with T1D must be investigated to determine its usefulness both in risk assessment and as a potential therapeutic target.

Areas timely for developing research: Recently identified copy number variants in DNA and epigenetic modifications (heritable changes not associated with changes in the DNA sequence) are also likely to play a role in genetic susceptibility to T1D.

Keywords: type 1 diabetes/genes/HLA class II/islet autoantibodies

Introduction

Accepted: November 6, 2008
*Correspondence to: Dr K. M. Gillespie, Medical School Unit, Southmead Hospital, Bristol BS10 5NB, UK. E-mail: k.m.gillespie@bristol.ac.uk

The immune system protects by being able to specifically differentiate between host cells and infectious agents. In autoimmunity, however, this system breaks down: for instance, in type 1 diabetes (T1D), insulin-producing β-cells are subjects to specific attack by the host immune system. T1D is often considered a condition of the young but is also frequently diagnosed in adults, and markers of autoimmune
diabetes have been identified in the slowly progressive form of autoimmune diabetes, latent autoimmune diabetes in adults. Autoimmune diabetes therefore affects an age spectrum from the very young onwards.

T1D is an evolving disease, changing in incidence over the second half of the twentieth century, with a particularly rapid relative increase in those diagnosed under the age of 5.1 Overall, the incidence is predicted to be 40% higher by 2010, compared with 1998.2 Epidemiological studies highlight the geographical differences in the incidence of T1D, with the highest occurring in Caucasoid populations, mainly in Northern Europe (40/100 000/year in Finland), and the lowest in Asia and South America (0.1/100 000/year), resulting in an incidence pattern with a North and South hemisphere divide.2 Given the range of complications associated with T1D such as nephropathy, retinopathy, neuropathy, coronary heart disease and peripheral vascular disease, improved understanding of the pathogenesis and identification of preventative therapies are the major aims of both physicians and politicians. In order to intervene very early in the autoimmune process, accurate prediction of those at risk of future T1D is required.

Islet autoimmunity can already be predicted accurately by detecting multiple islet autoantibodies and their characteristics.3,4 Combinations of antibodies to insulin (IAA),5–7 glutamic acid decarboxylase (GADA)8 and the tyrosine phosphatase IA-29,10 are found in individuals at risk or who have recently developed T1D. In the last year, the presence of antibodies in the zinc transporter (ZnT8) has been added to this list.11 Islet autoantibodies are however indicative of ongoing autoimmune β-cell destruction. It is now clear that the autoimmune process begins early in life: islet autoantibodies are detected at the age of 5 years in most future T1D cases;12 in many by the age of 2 years13 and antibodies to INS (generally the first to appear) have been detected as early as 6–12 months.14 Studies of neonatal diabetes suggest that most cases of diabetes diagnosed before 6 months are unlikely to be autoimmune, but those diagnosed after the age of 6 months have the genetic characteristics of T1D.15 Interventions aimed at preventing islet autoimmunity would therefore require identification of infants at genetic risk of T1D.

Genetic susceptibility is important in the development of T1D. The lifetime risk for a member of the general population is often quoted as 0.4%. This increases to >1% if the mother has diabetes and intriguingly to >3% if the father has T1D.16 The sibling risk is 6% (15 times greater than in a member of the general population).17 The classic indicator of the role of genetics in any disease is found by comparing concordance rates in monozygotic (MZ) vs. dizygotic (DZ) twins. In Finland, with the highest incidence of T1D, concordance rates for T1D were compared in 44 MZ and 183 DZ twin pairs from a
population-based twin cohort of 22,650 pairs. Concordance rates of 27% and 3.8%, respectively, were reported. This study showed that early age of onset in one twin increased the risk in the co-twin. Calculations to estimate the role of genetic vs. environmental factors in this study showed that 88% of the phenotypic variance was due to genetic factors, and the remaining variance is due to unshared environmental factors. A prospective analysis of 134 pairs of monozygotic twins from the UK and 53 pairs from the USA, discordant for diabetes, also emphasized the importance of age at onset in one twin to risk of future diabetes for the co-twin. Twins of individuals diagnosed at age 24 or younger had a 38% probability of diabetes by 30 years of discordance. This drops to only 6% in those diagnosed after the age of 24 years.

These concordance rates emphasize the importance of genetics in T1D, but also clearly demonstrate that having certain combinations of genes is not sufficient to cause T1D. Environmental triggers therefore modulate the onset of T1D in genetically susceptible individuals. Environmental factors have been implicated in the recent rapid increase in T1D incidence, because the gene pool cannot change quickly enough to account for the rapid rate of increase of T1D. Three independent studies, comparing genetic susceptibility in individuals developing T1D currently compared with those developing diabetes at least a generation ago, show that more people with less genetic susceptibility are now developing autoimmune diabetes. It might therefore be argued that genetic susceptibility is now less important than in previous generations, but the susceptibility genes have not altered; merely, the number that an individual requires to precipitate diabetes in a ‘permissive’ environment. Identification of these environmental determinants is proving more elusive than the susceptibility genes. Several studies are ongoing to try to identify the environmental determinants of T1D with particular emphasis on viruses and diet. An international consortium, the Environmental Determinants of Diabetes in the Young (TEDDY) (http://teddy.epi.usf.edu/), has been established to study genetic–environmental interactions, including gestational events, childhood infections or other environmental factors after birth in relation to the development of pre-diabetic autoimmunity and T1D.

Calculating genetic risk for a complex disease is challenging. Over the last three decades, the study of T1D has led the field in the identification of genes underlying complex multifactorial diseases. Unlike single gene disorders, which are inherited in distinct predictable Mendelian patterns, in multifactorial diseases such as T1D, identification of the combination of underlying causative genes is still a work in progress. Several different strategies have been used in efforts to identify T1D susceptibility genes. Among the most successful were early studies of gene frequencies in individuals with T1D, compared with controls.
**Human leukocyte antigen**

Genetic variants (or alleles) in the highly polymorphic human leukocyte antigen (HLA) on chromosome 6p21.3 (Fig. 1) can lead to functional differences in how fragments of protein are presented to the immune system. In the early 1970s, several groups investigated and found associations between T1D and HLA class I. Later, lymphocyte-defined HLA-D antigens, HLA class II DR4 and DR3 were shown to be more closely associated with T1D, and the combination of two susceptible alleles together, DR3/DR4, produced a higher risk genetic combination. The HLA contains lots of genes close together that are transferred from the parent to the child in adjacent ‘DNA chunks’ or haplotypes that are said to be in ‘linkage disequilibrium’. It can therefore be difficult to establish exactly which gene is having an observed effect. It is now generally considered that the principal susceptibility markers for T1D are HLA class II DQB1*0302 on the DR4 haplotype and DQB1*0201 on the DR3 haplotype (Fig. 1).

We now know that 90–95% of the young children with T1D carry either or both susceptibility haplotypes, whereas the protective DR2-DQB1*0602 is present in <0.1%. HLA-mediated susceptibility represents ~50% of the genetic susceptibility to T1D. HLA class II haplotypes have been ranked in a risk hierarchy. Those in the general population with the highest risk genotype DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302 have a 5% absolute risk of getting diabetes by the age of 15 years. A Belgian study

![Diagram of chromosome 6 with MHC genes and markers for T1D susceptibility](image)

**Fig. 1** MHC genes on chromosome 6 confer almost 50% of genetic susceptibility to T1D.
suggests that genetic strategies can be used to identify 10% of the general population that harbours most future cases of T1D. However, the majority of this genetically at-risk population will never develop islet autoimmunity. It is possible that screening the general population for high-risk HLA genes with islet autoantibody follow-up could represent a strategy to identify most future cases of T1D. This strategy has been used by the DIPP (Diabetes Prediction and Prevention Trial) study in Finland and by Diabetes Autoimmunity Study in the Young (DAISY) study in the USA. In DIPP, all newborn infants in a defined region carrying HLA DQB1 genotypes conferring susceptibility to T1D were observed from birth for the appearance of diabetes-associated autoantibodies. As described in a recent report, the DIPP study screened 116 720 consecutively born infants for high-risk HLA DQB1 genes, and 17 397 (6.7%) infants had increased genetic risk and 11 225 consented to follow-up for islet autoantibody screening. Two different islet autoantibodies were detected in at least two consecutive samples in 328 (2.9% of those consented). In the prospective DAISY cohort study, 27 000 newborns were screened for high-risk HLA genotypes and 1135 were identified for follow-up. These and other natural history of T1D studies have demonstrated that HLA genes are associated with the autoantibody profile. IA2-A are associated with specific DR alleles, IAA are associated with DR4, GADA are associated with DR3 and the recently discovered ZnT8 antibodies are associated with a single base pair change in the SLC30A8 gene.

HLA-DPB1, another HLA class II gene, has been shown to modulate HLA class II-mediated risk independent of HLA DRB1 and DQB1, and independent effects of HLA class I genes that have recently been reported in detail. Also, studies in siblings have identified ‘extreme’ high-risk haplotypes probably due to functional variants within or adjacent to the HLA. The remaining 50% of the genetic susceptibility results from contributions from an increasingly large selection of T1D susceptibility genes.

**INS gene**

In 1983, another case–control study identified a second locus linked with susceptibility to T1D: a variable number tandem repeat region in the promoter of the INS gene, which is important for regulating how much INS is produced. Alleles in this region are divided into three classes distinguished by the number of DNA base pair repeats. Class I alleles have a mean of 570 base pairs, class II alleles 1200 base pairs and class III alleles have 2200 base pairs. There is a higher frequency of
class I alleles in individuals with T1D, whereas protection from T1D is associated with the class III allele.\textsuperscript{46} In studies of INS gene expression, it has been shown that class I alleles are associated with higher INS expression in the pancreas when compared with class III alleles, but the opposite is true in the thymus where class I alleles are expressed at 2–3-fold lower levels. This is likely to alter the selection of T cells in the thymus and may therefore influence tolerance to INS.\textsuperscript{47,48}

**Whole genome screens**

In the 1990s, advances in genetic technology allowed detailed maps of polymorphisms (or genetic variants), in particular, microsatellites and single nucleotide polymorphisms (SNPs), throughout the genome to be developed. With the addition of semi-automated genotyping technology, it became possible to analyse genetic markers all over the human genome for associations with T1D. The initial screens suggested possible linkage to an array of markers.\textsuperscript{49–51}

Areas of the genome linked to T1D were assigned IDDM numbers. For instance, the HLA (IDDM1) and the INS susceptibility locus (IDDM2) were confirmed in all studies. IDDM 1–18 have been allocated to various regions of the genome, but, in 1998, simultaneous whole genome screens of 187 families (with follow-up in 429 additional families)\textsuperscript{52} and 356 affected sib pairs\textsuperscript{50} yielded conflicting results. A follow-up study of 767 families by Cox et al.,\textsuperscript{53} including a new collection of 225 families with T1D combined with data from the previous screens,\textsuperscript{50–52} confirmed some of the loci, but diminished support for others. The success of this combined analysis highlighted the need for worldwide genome screening collaborations to increase the numbers of patients and controls that could be analysed, thereby increasing the statistical power of these experimental strategies.

**CTLA-4**

In 1996, the cytotoxic T-lymphocyte antigen (CTLA-4) gene encoded on chromosome 2q33 was recognized as a further T1D susceptibility gene.\textsuperscript{54} CTLA-4 is a surface molecule found on activated T cells that produces a negative signal for T cell activation. It is thought that inherited changes in the CTLA-4 gene expression can increase T cell self-reactivity and therefore play an important role in autoimmune diseases such as T1D. In 2003, Ueda et al.\textsuperscript{55} unequivocally proved the role of this molecule in susceptibility to T1D.
**PTPN22**

More recently in 2004, *PTPN22*, a gene found on chromosome 1p13 that encodes lymphoid protein tyrosine phosphatase, was found to be associated with susceptibility to T1D in another case–control study. An SNP in the *PTPN22* gene potentially contributes to susceptibility to T1D because of increased negative regulation of T cell activation. It has been shown that 30.6% of the people with TID compared with 21.3% healthy controls are heterozygous for the 1858C/T polymorphism. This highlights the complexity of genetic susceptibility to T1D, as the T1D-associated allele is relatively common in the general population. The identification of PTPN22 as a T1D susceptibility gene emphasizes the importance of T cell activation, as demonstrated in Figure 2.

**IL2RA/CD25**

In 2005, Vella et al. reported that the interleukin 2 receptor alpha (IL2RA) region on chromosome 10p15 was associated with T1D. IL2RA encodes the α-chain of the IL-2 receptor complex (also referred to as CD25), which is responsible for binding IL-2, a key player in the proliferation of regulatory T cells. IL2R has also been associated with T1D in the well-characterized model of T1D, the non-obese diabetic (NOD) mouse. The identification of IL2RA/CD25 as a T1D susceptibility gene underscores the importance of T cell activation, as demonstrated in Figure 2.
In early 2007, Qu et al. confirmed the association of IL2RA with T1D and identified two SNPs associated with the increased risk of T1D. Later that year, Lowe et al. analysed 288 SNPs and identified the SNP ss52580101 to be the most closely associated with T1D. They showed that the SNPs with the weakest association with T1D were those involved in increased soluble IL2RA expression. Thus, T1D-susceptible alleles are also associated with decreased concentrations of IL2RA, suggesting a possible biological mechanism for autoimmunity through reduced binding of IL-2 (an essential component for the survival and proliferation of regulatory T cells—suppressors of autoreactivity).

**IFIH1**

In 2006, Smyth et al. identified IFIH1 (interferon induced with helicase C domain 1) on chromosome 2q24.3 as the sixth gene to be strongly associated with T1D, where the frequency of IFIH1 polymorphisms was analysed in 4253 individuals with T1D, compared with 5842 healthy controls. In 2008, Qu et al. carried out a follow-up study on IFIH1, in which they genotyped five SNPs in 1767 individuals with T1D. The strongest association was found with rs1990760 (as reported previously by Smyth et al.). IFIH1 is thought to contribute to innate immune responses by releasing the cytokine interferon-gamma and inducing apoptosis of virally infected cells. This molecule may therefore provide molecular insights into some epidemiological studies that have suggested that viruses and enteroviruses, in particular, may contribute to the initiation of T1D, and evidence for Coxsackie B4 infection was recently demonstrated in the pancreas of individuals with T1D.

**Vitamin D receptor (VDR)**

Vitamin D has important immunomodulatory properties, and the active form vitamin D3 (1,25 dihydroxyvitamin D3) has been shown to inhibit T cell proliferation. In 1992, Mathieu et al. demonstrated that the administration of vitamin D3 in the NOD mouse prevented the onset of T1D, and in 1999, the EURODIAB study demonstrated that a protective effect to T1D in infancy was conferred by vitamin D supplementation.

Genetic studies of vitamin D-associated genes and T1D have been complicated. A comprehensive analysis by Nejentsev et al. in 2004 identified no association between T1D and the VDR gene after
examining 97 SNPS in up to 3763 T1D families from the UK, USA, Finland and Romania. More recently, in 2007, Bailey et al.\textsuperscript{70} investigated the association of polymorphisms in the CYP27B1 and CYP24A1 genes (involved in the activation and inactivation, respectively, of the vitamin D precursor enzyme 1α-hydroxylase) in T1D. No association was identified for the CYP24A1 gene; however, an association with T1D was identified for the CYP27B1 gene on 12q13.1–q13.3, in which the C allele of rs10877012 was significantly associated with increased risk of T1D. A potential genetic explanation for the observed effects of vitamin D in T1D has been identified, and further studies of vitamin D in T1D are ongoing.

Genome wide association studies

Recent genome wide association (GWA) studies operate at a much higher level of statistical power compared with previous relatively underpowered whole genome screens. For instance, the results of analysis of 2000 individuals for seven major diseases and a shared set of approximately 3000 controls were reported in 2007.\textsuperscript{71} In T1D, significant linkages were shown for 12q24, 12q13, 16p13, 18p11, 12p13 and 4q27. Interestingly, these loci do not overlap with the previously identified IDDM loci. Further analysis of 4000 individuals with T1D, 2997 T1D family trios and 5000 controls confirmed 12q24, 12q13, 16p13 and 18p11.\textsuperscript{71}

The apparent success of GWA studies comes not only from increased power due to the increased number of patients and controls available for analysis, but is also attributable to the availability of high-throughput genotyping arrays and the HapMap project that defined areas of linkage disequilibrium throughout the genome to allow maximum genetic information to be obtained from analysing a minimum number of genetic polymorphisms.

Susceptibility genes in 12q24 and 12q13 have yet to be identified, but KIAA0350 (CLEC 16A) and PTPN2 represent the most likely candidates in 16p13 and 18p11, respectively.

KIAA0350

KIAA0350, also known as CLEC 16A, is a gene of unknown function on chromosome 16p13.2, identified by cDNA library sequencing. Its structure, predicted from the coding sequence, suggests that it has a C-type lectin-binding domain, indicating that it may be a cell surface receptor. It has been reported to be expressed in immune cells,\textsuperscript{72}
particularly B lymphocytes, dendritic cells and NK T cells. Over the next year, we can predict that a lot more detail about the role of KIAA0350 will emerge.

**PTPN2**

Phosphotyrosine protein phosphatase, non-receptor 2 (PTPN2) is involved in the activation of STAT-1 (signal transducer and activator of transcription), which regulates many immune molecules.

More than 10 genetic regions that underlie susceptibility to T1D have now been identified (Figure 3). Many already have candidate genes, and further SNP analysis and re-sequencing will be required to identify the chromosome 12 T1D genes, although ERBB3e has been suggested as a possible candidate for the chromosome 12q13 locus. It is reasonable to suggest that more associations will be reported and replicated in the next year or two. It will be interesting to observe whether any more of the original ‘IDDM’ loci will be identified by GWA methods.

The techniques described so far in this review only test for the association of genetic variants with disease. It is increasingly clear that other molecular mechanisms may also influence the genetic contribution to disease. Sequencing analysis has shown that some loci are deleted or present in multiple copies. If such a region contains transcription factor binding sites or enhancers, then there could be a functional or

![Fig. 3 Odds ratio for susceptibility alleles for T1D-associated genes or regions (adapted from Todd et al.71).](http://bmb.oxfordjournals.org/)

---

K. L. Mehers and K. M. Gillespie

British Medical Bulletin 2008;88
phenotypic effect. These copy number variants (CNV) are not necessarily detected by current high-throughput genotyping methods. Such variants have already been associated with cytogenetic syndromes\textsuperscript{73} as well as protection from HIV,\textsuperscript{74} but their role in complex diseases such as T1D remains to be elucidated.

The other unknown is potentially the role of imprinting and epigenetics. Genomic imprinting is the phenomenon by which some genes are preferentially expressed if they are maternally inherited, whereas others are expressed if they are paternally inherited. Imprinting has already been implicated in a number of human diseases including transient neonatal diabetes.\textsuperscript{75} Several studies have suggested differences in paternal and maternal inheritance patterns in T1D, making imprinting an important possible mechanism, and the importance of imprinting in T1D is also likely to be unraveled over the next few years but genomic imprinting is not the only means by which transcription is altered. In particular, microRNAs (miRNAs) that have been identified in viruses, plants and animals have the capacity to regulate gene expression in a sequence-specific manner. A greater understanding of miRNAs in T1D could provide unique insights into how environmental agents can alter the risk of disease.

In conclusion, as shown in Figure 4, it has taken over 30 years, but the scientific community is increasingly able to define the importance of individual genes in susceptibility to T1D. It is, as yet, unclear what percentage of absolute genetic risk can now be measured by combining

Fig. 4 Progress in the identification of T1D susceptibility alleles: the last 35 years.
all the known risk alleles, but it is unlikely to be of help in diagnosing T1D or in predicting the onset of diabetes in an individual case. The short-term effects are most likely to be observed in identifying high-risk subsets of the populations for islet autoantibody follow-up. In the long-term, understanding of the roles of these genes and their molecular pathways may potentially lead to targeted therapies for children known to be at risk of T1D in the future.

**Acknowledgements**

We thank Mimi Chen for reading the manuscript and her helpful insights.

**Funding**

KH is funded by a University of Bristol PhD scholarship.

**References**

British Medical Bulletin 2008;88


15 Edghill EL, Dix RJ, Flanagan SE et al. (2006) HLA genotyping supports a nonautoimmune etiology in patients diagnosed with diabetes under the age of 6 months. Diabetes, 55, 1895–1898.


41 Cruz TD, Valdes AM, Santiago A et al. (2004) DPB1 alleles are associated with type 1 diabetes susceptibility in multiple ethnic groups. Diabetes, 53, 2158–2163.