Journal of Autoimmunity 96 (2019) 123–133

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

A stochastic epigenetic Mendelian oligogenic disease model for type 1 diabetes

Chester A. Alper[a,b,*], Charles E. Larsen[a,b], Michael R. Trautwein[a], Dennis R. Alford[a,1]

[a] Program in Cellular and Molecular Medicine, Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA
[b] Department of Pediatrics, Harvard Medical School, 25 Shattuck Street, Boston, MA, 02115, USA

ARTICLE INFO

Keywords:
Disease association
Epigenetics
Genetics
HLA
Monozygotic twins
Type 1 diabetes

ABSTRACT

The incidence of type 1 diabetes (T1D) and some other complex diseases is increasing. The cause has been attributed to an undefined changing environment. We examine the role of the environment (or any changing non-genetic mechanism) in causing the rising incidence, and find much evidence against it: 1) Dizygotic twin T1D concordance is the same as siblings of patients in general; 2) If the environment is responsible for both the discordance among identical twins of patients with T1D and its rising incidence, the twin concordance rate should be rising, but it is not; 3) Migrants from high-to low-incidence countries continue to have high-incidence children; 4) TID incidence among the offspring of two T1D parents is identical to the monozygotic twin rate. On the other hand, genetic association studies of T1D have revealed strong susceptibility in the major histocompatibility complex and many optional additive genes of small effect throughout the human genome increasing T1D risk. We have, from an analysis of previously published family studies, developed a stochastic epigenetic Mendelian oligogenic (SEMO) model consistent with published observations. The model posits a few required recessive causal genes with incomplete penetrance explaining virtually all of the puzzling features of T1D, including its rising incidence and the specific low T1D incidence rates among first-degree relatives of patients. Since historic selection against any causal gene could prevent T1D, we postulate that the rising incidence is because of increasing population mixing of parents from some previously isolated populations that had selected against different causal genes.

1. Introduction

Defining complex genetic disease etiology is currently mostly a task of determining the optimal mathematical equations that quantify the relationship between population phenotypic and genotypic variance and then explaining the relevance of this relationship to causality. As pointed out recently [1], this search is conducted within a paradigm assuming "genetic variance in a population [is] due to a large number of Mendelian factors, each making a small additive contribution to a particular phenotype, the so-called 'infinitesimal model'" [2]. This "genes of small effect" paradigm is now usually assumed for all complex traits, including both quantitative and binary traits [1,3]. The dominant viewpoint further assumes environmental inputs modify (or "contribute to") the structural relationships and effects ("risk") of many (dozens to hundreds) of these genes of small effect. This paradigm assumes there is no single causal pathway for any complex genetic trait and, indeed, presumes there might be multiple pathways to complex genetic disease onset. For simplicity, we call this current paradigm the environmentally-triggered risk-based polygenic (ERP) model.

Type 1 diabetes (T1D) is a complex genetic disease of unclear etiology despite decades of research. Many features of T1D natural history are incompatible with or are completely ignored by the ERP model. From our perspective, which is focused on pedigree-based genomic structural and genetic linkage and association analyses, we conclude that ERP is an impediment to understanding T1D etiology and probably that of some other complex genetic diseases. We support calls for developing a new approach to unraveling complex genetic disease [1,4]. We propose an alternative simpler model that fits all observed data of T1D as a stochastic epigenetically-triggered Mendelian oligogenic (SEMO) disease (see Box 1; based on, but modified from, a summary in reference (ref.) 5). We cite reports that both contradict ERP and support the SEMO model. While many models are capable of approximating observed reality, we show that ERP assumptions lead both to inappropriate methods and to incorrect conclusions regarding T1D

* Corresponding author. Program in Cellular and Molecular Medicine, Boston Children's Hospital, 3 Blackfan Circle, CLS_03, Boston, MA, 02115, USA.
E-mail address: chester.alper@childrens.harvard.edu (C.A. Alper).
[1] Deceased.

https://doi.org/10.1016/j.jaut.2018.09.006
Received 3 August 2018; Accepted 12 September 2018
Available online 08 October 2018
0896-8411/

Box 1
SEMO Model Hypotheses, Equations and Conclusions (adapted from Ref. [5])

I. Most models of complex genetic disease assume:

Phenotype = Genotype + Environment

and calculate "heritability" (h^2) as the ratio of genetic (σ_G^2) to phenotypic (σ_P^2) variances.

II. The SEMO model proposes that, at least for some complex genetic diseases:

$$\text{Phenotype Frequency (P)} = \text{Disease Susceptible Freq. (D)} \cdot \text{Non—Genetic Effect Freq. (N)} \quad (1)$$

and,

$$N = fn\ [\text{Environmental Effect(s) (E)} + \text{Stochastic Effect(s) (S)}]** \quad (2)$$

And, for a disease in which there are a limited number (e.g., 3) of required disease susceptibility genes (X, Y and Z), and, e.g., susceptibility at **loci X** and **Y** is **each** Mendelian **recessive**, and locus **Z** is Mendelian **dominant***, where the frequencies (p) of aggregate susceptibility alleles at locus X are pX, pY at locus Y and pZ at locus Z, and where alleles sort independently (are unlinked):

$$D = (pX)^2 \cdot (pY)^2 \cdot \{(pZ)^2 + [2(pZ) \cdot (1 - pZ)]\} \quad (3)$$

III. **Monozygotic twin (MZT) pairs are essentially genetically identical.** Therefore, all MZTs of patients with the disease **are fully genetically susceptible** to the disease, and **D = 1**. Therefore, disease frequency (eqn. (1)) among MZTs of probands ascertained by disease status reduces to:

$$P = N \quad (4)$$

Therefore:

$$N = \text{MZT concordance} \quad (5)$$

Likewise, the total population comprising D (eqn. (3)**) is**, *for the purposes of disease susceptibility*, genetically **identical to MZT pairs with a disease proband.** Therefore, eqn. (5) **applies to all individuals who are genetically fully susceptible to the disease**.

IV. **E (**eqn. (2)**) *might be* a component of N**, but dizygotic twins (DZTs) of T1D probands have the same disease concordance as all siblings. Therefore, the more "common" environment shared by twins has no effect on T1D frequency; the only *variable* "E" possible among twins is E_u, that unique to each individual. For 50 years, within experimental error, T1D MZT concordance has neither changed nor differed worldwide. Furthermore, the geoethnic and temporal variability of MZT concordance is unrelated to T1D frequency. Therefore, **for T1D, N is constant** and **geoethnic and temporal *differences* in incidence are explained by differences in genetic susceptibility (D) alone**.

V. T1D models assuming $N = E_u$ (i.e., stochastic effects do not exist; the non-genetic effect is entirely environmental), must assume E_u (for T1D) is constant temporally, geographically and in different populations. T1D pathogenesis models assuming E_u is only part of N *and* hypothesize E_u varies must explain, based on eqn. (2), how the opposite S has changed such that N is constant. **The SEMO model posits (for simplicity) that, for T1D, E = 0 and** eqn. (2) **reduces to:**

$$N = fn\ S \quad (6)$$

*For T1D, we hypothesize, based on available data, all susceptibility genes are recessive.
**The model posits epigenetic mechanisms could underlie E and S *separately*.

causality. Table 1 summarizes the conceptual and explanatory differences between these two models. We make specific predictions and we suggest different approaches to genetic analyses of T1D pathogenesis.

We begin by discussing the many puzzles of T1D etiology, including both those for its genetic and non-genetic components. Next, we examine evidence for and against environmental causes of T1D, especially because a changing environment is often blamed for the rising incidence of T1D. After concluding that the environment is unlikely to play any role in T1D pathogenesis, we present evidence that the non-genetic component of T1D, as defined by concordance for T1D in monozygotic twin (MZT) pairs, neither differs worldwide nor has recently changed. Then, because the major histocompatibility complex (MHC) is the region most strongly correlated with T1D susceptibility, we develop the model's genetic component using the MHC relationship to T1D. We also consider the findings in families followed for nearly 60 years in which both parents have T1D, and we present the SEMO model's non-genetic component. Finally, we evaluate the fit of the SEMO model's consequences to observed facts, and we predict particular features of T1D genetics.

2. Puzzling features of T1D

T1D has many puzzling features. First, T1D onset, after a variable autoimmune prodrome, is usually not until mid-childhood, adolescence or, rarely, even adulthood. Second, since complex genetic diseases like T1D are far more common than monogenic diseases, the MHC haplotypes that mark T1D susceptibility alleles are not rare but are common variants [6]. We posit there may be several susceptibility alleles of any causal gene. A third puzzling feature of T1D is the rarity of any family history – nearly 90% of T1D patients have no affected relative [7]. A distressing enigma of T1D is its rising incidence of 2%–3% per year [8–10]. Along with this rise is the changing genetic composition of T1D patient populations, at least in some countries, with a decrease in the frequencies of the high-risk T1D susceptibility MHC markers HLA-DR3 and HLA-DR4 (old nomenclature) [10]. The basis for the high T1D "risk" of HLA-DR3/4 heterozygotes [11] compared with HLA-DR3/3 or -DR4/4 homozygotes is also puzzling. A final puzzling T1D feature is that the roughly 50% T1D incidence among the offspring of two parents who have T1D [12] is the same as the concordance rate of MZTs of patients with the disease [13].

Table 1
Comparing and contrasting ERP and SEMO methods and explanations for T1D[a].

Component/Question	SEMO Method	ERP Method
Genomic architecture	Direct pedigree-based haplotype determination	Genotype or imputation-based haplotype estimation
Genetic association to T1D	Patient *vs* FC (not in any patient) haplotype comparison	Patient *vs* Control genotype, allele or imputed haplotype comparison
Identifying T1D genes	Recessive alleles at a limited number of chromosome 6 loci with FC frequencies near population frequencies	Statistically-significant different (and odds ratios > 1) frequencies in patients *vs* controls; then, nearby "immune-related" or "likely." Dozens identified.
Phenomenon	**SEMO Explanation**	**ERP Explanation**
Rising incidence	Parental subpopulation admixture and selective selection	Unspecified environmental change
50% of MZTs of patients have T1D	Age-related random epigenetics changes susceptibility to disease	Unspecified environmental factor(s) trigger(s) disease
DZT rate = sib rate	T1D susceptibility purely genetic	Not considered (incompatible)
50% of offspring of two parents with T1D have T1D (= MZT rate)	All disease alleles are full susceptibility	Not considered (incompatible)

Shown are the methods used to study the genetics of T1D and the explanations that the two complex genetic disease models provide for specific known phenomena associated with T1D.
[a] Abbreviations: T1D = type 1 diabetes; FC = family control; DZT = dizygotic twin; MZT = monozygotic twin.

No causal gene or molecular mechanism for this disease associated with autoimmune loss of pancreatic beta cells has been definitively established. Instead, methods for identifying genes "associated" with T1D have focused on statistical calculations of risk (e.g., odds ratios). Such methods do not require the context of a biological model nor any known (or fixed) genetic architecture. Most recent genetic association studies do not even use a family-based analysis comparing phased patient alleles or haplotypes with family controls [14]. Instead, results are deemed significant if case-control studies are replicated (or *p* values lowered) in larger or different populations.

The mode of inheritance of a gene or genetic region for a genetic trait is a key element necessary to understand the trait's genetic etiology. For T1D, the mode of inheritance of MHC-related disease susceptibility alleles or haplotypes was originally determined to be clearly Mendelian recessive. Later, "complex," "mixed" or disease "heterogeneity" models became fashionable. Since 1990, the question of the MHC-associated mode of inheritance has largely been abandoned. For non-MHC genes, the question has rarely been raised. A strong argument has recently been made [1] for re-examining the many variable genes of small effect assumptions as originally put forth by Sir Ronald Fisher a century ago [15]. In this context, it is essential to distinguish causal genes (with clear modes of inheritance) from subpopulation markers that only correlate with increased T1D risk [16,17] but are likely irrelevant to T1D pathogenesis.

3. Twin studies in T1D

Approximately 40–50% of MZTs of T1D patients also have T1D [7]. Since these twins carry virtually the same gene variants, the less than 100% concordance has been termed "incomplete penetrance" of genetic susceptibility. It has been *assumed* for decades that incomplete penetrance reflects some (combination of) triggering causal environmental factor(s) or event(s). For T1D, the difference in concordance between MZTs (40–50%) and dizygotic twins (DZTs) (6–10%) has been ascribed to multiple additive risk-elevating genes that contribute to "heritability" [5]. When applied to T1D [18], this analysis ascribes up to 66%–72% of the risk for T1D to heritability, leaving, presumably, 28%–34% to the environment. However, penetrance is actually the combination of the genetic contribution and *all non-genetic* contributions (whether of environmental origin or not) that convert full genetic susceptibility (e.g., in the MZT of a patient) to disease. The extent to which the non-genetic component contributes to disease pathogenesis must be determined for every complex genetic disease, and one should never *assume* the "environment" is synonymous with the non-genetic component. Indeed, stochastic epigenetic changes to DNA may be more common than environment-induced changes [19].

3.1. Is there a role for the environment in T1D pathogenesis/penetrance?

The conviction that exposure to (an) environmental factor(s) is partly responsible for the disease is so strong that it is invoked to explain the observed rapidly changing MHC allele and haplotype composition of T1D patients [20]. Proponents of this view claim [21] this observed rapid genetic change could not have a genetic basis because a purely genetic mechanism could not operate so rapidly. Therefore, in this view, the observed rapid genetic change must represent the changing genetic response to a postulated but undefined rapidly changing environment [20]. In contrast, we provided a purely genetic explanation for the rapid genetic change [22,23]. Although there has been a recent resurgence in suspicion that the microbiome [24] or, more frequently, viral infections [25–28] could play such an environmental role, "there is still little evidence" [28]. Thus, the ERP model search for (an) essential environmental trigger(s) of T1D has been unsuccessful, despite almost a century of intensive investigation. While a full critique of this common (and usually uncited) assumption would require far more extensive space than we have here, we propose that the findings presented below sufficiently argue against *any* environmental factor as a causal basis for penetrance of genetic susceptibility to T1D acting *in utero* or in childhood within a family or in general. At the very least, the environment should not be presumed to play any role in this disease.

The seasonal variation in the onset of clinical signs of T1D has been used to implicate the environment, particularly viral infection, in T1D pathogenesis. The peak incidence of the initial clinical signs of T1D occurs in the autumn-winter months in the northern [29] and southern [30] hemispheres and is coincident with or follows closely epidemics of mumps, rubella and coxsackie B4 virus [29]. Several observations make it unlikely that infection with any of these viruses is involved in penetrance in T1D. From studies of the occurrence of autoantibodies and glucose metabolism in genetically susceptible relatives of patients we now know that pancreatic β-cell destruction begins many months or years before symptoms of T1D appear [31–33]. Overt disease does not occur until 90% or more of β-cell mass is destroyed. More importantly, the presence of antibodies to the implicated viruses does not correlate with T1D in discordant MZTs [34] nor in newly diagnosed T1D patients in general [35], indicating that infection with those viruses is unrelated to the presence of T1D. Finally, thanks to vaccines, infections with at least two of these viruses (mumps and rubella) plummeted during the last third of the 20th and into the 21st centuries, a period during which

T1D incidence increased significantly.

If exposure to an environmental factor during intrauterine life or early childhood causes penetrance (i.e., T1D pathogenesis in genetically fully susceptible individuals), DZTs of patients should have higher concordance than siblings of patients in general, despite similar 50% sharing of variant genes, since DZTs usually share pre- and early postnatal environments more completely than do sibs in general. The reported sib and DZT rates have varied widely over the years, but the best T1D twin study to date suggests the DZT rate is quite similar to that found in all sibs. In a very large study of twins in Finland [36], concordance for T1D was the same in DZTs (7.4%) as in sibs in general (6.3% up to age 34 and 9.6% up to 60 years of age [37]). Moreover, antibodies against insulin and other pancreatic β-cell antigens, which are signs of impending T1D, were found at the same rate in DZTs as in sibs in general [38]. These two facts alone are sufficient to challenge the assumption that the environment plays a significant role in T1D pathogenesis. Furthermore, if the environment were required to cause T1D, T1D should appear more or less simultaneously in each member of any pair of affected MZTs (or DZTs or sibs). This is not the case. Several years commonly separate the onset in affected family members [7].

If the environment is a cause of the high incidence of T1D in certain countries, families who move from a high incidence to a low incidence country should, after the move, have children with a lower incidence in the new environment. On the contrary, children of Sardinian (high incidence) families who moved to mainland Italy (low incidence) or Germany (also low) continued to have high incidence [39,40]. A contrasting argument is made based on several studies of southern Asian (low incidence) families who moved to high incidence countries such as the United Kingdom [41,42] or Sweden [43]. These studies showed a higher T1D incidence among the children of such immigrants than that reported for southern Asia. These studies, however, may have two problems. First, they were registry studies, not interviews, so some ethnically-mixed parents may have been involved. Second, rigorous complete studies of T1D incidence in southern Asia, including considering internal regional and genetic variation, are lacking. Moreover, other studies (with, admittedly, some of the same analysis issues) of southern Asian migrants to Sweden [44,45] revealed retained low incidence among their children.

3.2. The non-genetic component is constant and cannot explain variable T1D incidence

While the above observations provide strong circumstantial evidence against a role for (an) environmental factor(s) in T1D pathogenesis, the question remains: Does any *variable* non-genetic effect correlate with the variable incidence of T1D? The hypothesis that some non-genetic element influences geographic, demographic or temporal differences in T1D incidence has rarely been tested directly. Below, we provide predictions based on a role in T1D pathogenesis for the non-genetic component (whether or not it includes an environmental subcomponent) and report the findings that belie the predictions.

If the non-genetic component (as represented by the failure of about half the MZTs of T1D patients to have T1D) plays a role in the rising incidence of T1D in the general population, then concordance for T1D in MZTs should also be rising. In fact, there does not appear to be any increase in MZT (or DZT) concordance over the past five decades [13,36,46–55], as is evident in Fig. 1, despite wide variations in results. Fig. 1 displays all of the world's published data from non-overlapping studies on twin T1D concordance rates using reports cited in the figure legend. We do not include a theoretical line proposing how the results would differ if the non-genetic component were variable because no one has proposed a quantitative formula for such change(s) nor how it/they might vary worldwide.

If the non-genetic component were responsible for a significant percentage of the variable incidence of T1D, countries with a low incidence might be expected to have a lower MZT concordance than those with a high incidence. As can be seen in Fig. 1, MZT concordance in Japan [49] with a T1D incidence of 1.5–2.5/100,000 is similar to that in Finland with an incidence over 25 times higher at 42–64/100,000. In fact, MZT incomplete penetrance, the supposed consequence of a putative environmental cause, is the same (within experimental error) in all countries (each, presumably, with a unique environment). Moreover, there was no increase in the very low incidence in Japan over the prior decade [49], despite the similar MZT concordance as in Finland. Taken together, these observations support a purely genetic basis for T1D susceptibility and its rising incidence, with no evidence for quantitative variability in the non-genetic component.

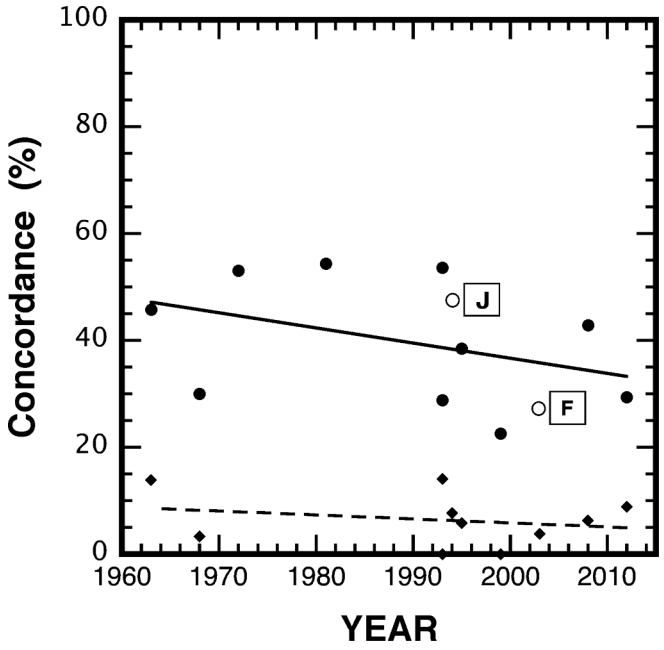

Fig. 1. MZT and DZT concordance for T1D over the last several decades and in various parts of the world. T1D MZT (•) and DZT (♦) pairwise concordance (calculated as in reference (ref.) 13) and the trends of each as a function of time (—, MZT; - -, DZT) are shown. Data are from ref. 36 and publications cited in refs. 13, 36 and 46, except that: a) ref. 36 is a more recent study of the Finnish T1D twins than the one reported in ref. 46 (and therefore the earlier one is *not* included here); and, b) ref. 49 is the corrected reference (although the same data) as that cited in ref. 46. The Finnish result from ref. 36 is labeled "F" and the Japanese MZT concordance result from ref. 49 is labeled "J" (both results displayed as **o**).

4. The MHC and T1D

The human MHC, a highly polymorphic genomic region of over 3 Mb on human chromosome 6p21, shows by far the strongest genetic linkage to T1D of any genomic region [56,57]. We conclude, therefore, at least one required (i.e., causal) T1D gene must exist within or near the MHC. Although the strongest statistical correlation to T1D is the HLA-DR, DQ block of the class II region [58], the identity of the MHC T1D susceptibility gene (or genes) remains unknown. This is partly because all of the MHC genetic markers of susceptibility to or protection from T1D are parts of conserved extended haplotypes (CEHs) [14,59,60], and no single locus or combination of loci shows a complete correlation with T1D. Two other complications have prevented identification of the MHC susceptibility gene(s): a) at least one other required T1D gene must exist outside of the MHC region; and, b) T1D exhibits incomplete penetrance in subjects fully genetically susceptible to the disease.

The methodology by which the MHC region has been studied has had profound consequences both for understanding MHC genomic structure and for localizing the T1D susceptibility gene(s).

Technological advances over the past 50 years, from antibody-determined specificities and soluble serum and DNA microsatellite variants to single nucleotide "markers" to full locus and haplotype DNA sequencing, have played a role. More importantly, although the MHC is the most intensively studied region of the human genome for genetic polymorphisms, reports over the past 40 years with respect to T1D have largely focused on statistical correlation of such variants to identify both "imputed" haplotypes and genetic "risk." However, until long-range sequencing capability (of individual chromosomes–paternal and maternal variants separately) is optimized, family-deduced haplotype-based studies are essential for identifying *both* genomic architecture at the individual and population levels as well as complex disease susceptibility genes [14,23,59–67].

For example, MHC CEHs are identified by direct segregation analysis in families and are defined by their *HLA-C*, *-B*, complement gene (*CFB*, *C2*, *C4A*, *C4B*) and *HLA-DRB1*, *-DQB1* markers [59]. These very large (1.4–4.5 Mb) stretches of fixed MHC DNA are essentially identical in many apparently unrelated individuals [60–67]. In European Caucasian populations, about 30 CEHs have a frequency ≥0.5% and account for nearly half of all MHC haplotypes [66]. Defining individual haplotype structure is the critical first step. Imputation and statistical correlation methods have largely been employed to avoid the difficult, time-consuming and expensive method of collecting and analyzing pedigrees, but the latter is currently the only method to determine genomic architecture accurately and precisely.

A second critical factor is the variable subpopulation distribution of such well-defined MHC variants. Evidence that the MHC marks susceptibility to T1D was first obtained by association studies in which positivity rates of MHC alleles in patients were compared with those in an "ethnically-matched" control population [68,69]. Such evidence is, however, not definitive. Ethnic matching helps to reduce confusing a purely subpopulation genetic marker with a genetic marker for a causal gene (which could also be a subpopulation marker, since T1D incidence differs considerably among various European subpopulations). But the best way to minimize population stratification in an association study is family segregation analysis [14] to determine the frequencies of alleles and haplotypes. This method yields sets of disease haplotypes (occurring in patients) and family control (FC) haplotypes for comparison. FC haplotypes should be defined as those not found in any patient in a family [14].

Finally, the mode of inheritance of MHC-related disease susceptibility alleles or haplotypes must be known to identify the T1D MHC susceptibility locus or loci. For T1D, affected sib pair (ASP) sharing of MHC haplotypes is consistent with recessive inheritance [6,70,71]; dominant inheritance is excluded [6]. In addition, among T1D patients, the genotype distributions of complement factor B (*CFB*) [72,73] and HLA-DR [74] as well as of HLA-DR, DQ haplotypes [75] fit the Hardy-Weinberg distribution in some populations consistent with recessive inheritance. However, there is an excess of HLA-DR3/4 heterozygotes in some Caucasian patient populations [76,77], suggesting a higher

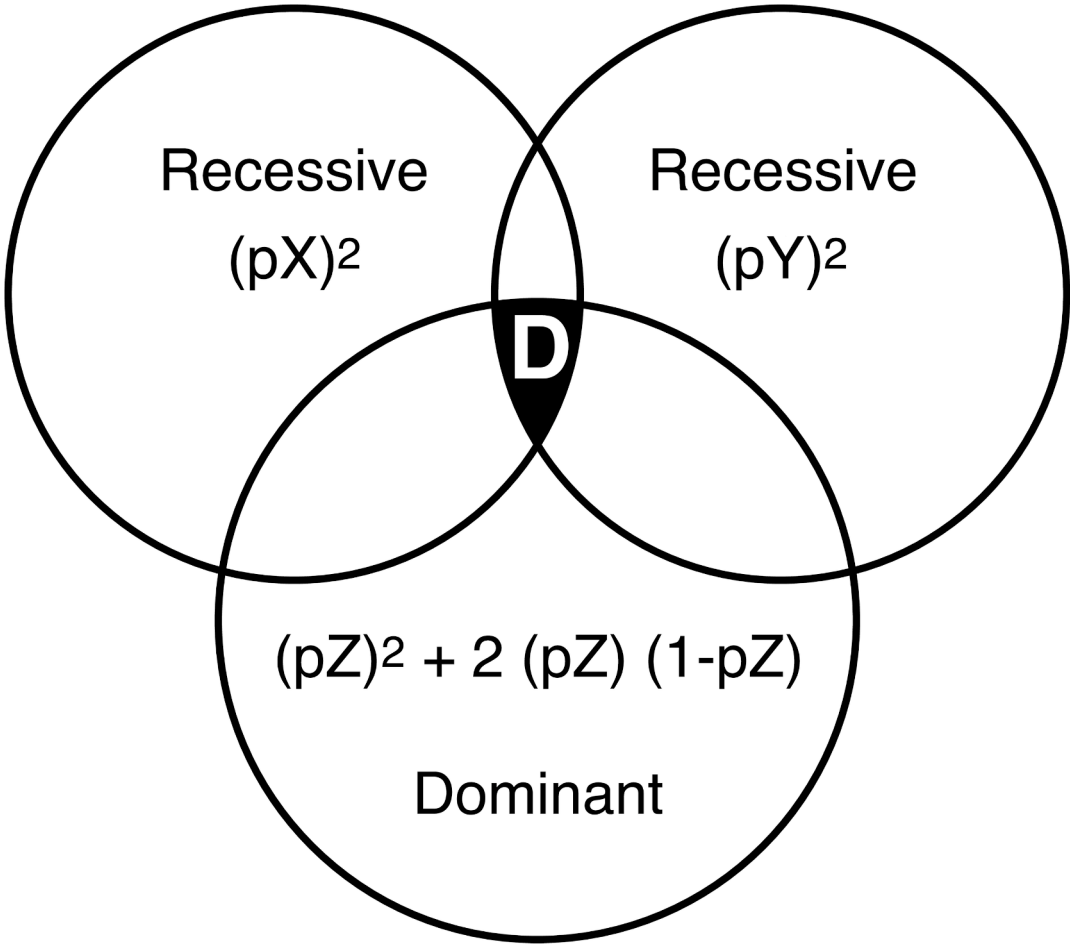

Fig. 2. Schematic presentation of the genetic component of the SEMO model for T1D involving three required susceptibility loci, of which two are recessive and one is dominant. For T1D, all three genes may be recessive. The population frequency (p) of aggregate susceptibility alleles at each of three hypothetical genetic loci (X, Y and Z) is used to calculate population susceptible frequencies (circles with equations) for each locus. Where the three susceptibility circles overlap (D), individuals are genetically susceptible to the disease. Thus, the frequency of population susceptibles (D) is the product of the frequencies of potential susceptibility frequencies at each gene. (Modified version with permission from ref. 22.)

T1D risk for all such individuals (although the vast majority of such high-risk individuals in the general population do not have T1D). Since there are populations without such an excess [74,75], the question should be why the distortion occurs only in some populations, not whether MHC recessive inheritance is ruled out, as has been falsely claimed [78].

5. The genetic component of the SEMO model for T1D

We propose a purely genetic component for *susceptibility* to T1D [6,22,23] in which an age-related stochastic (randomly distributed but invariant) intrinsic process (*not* a variable environmental process) causes penetrance of that susceptibility [61,79,80]. Both the difference in T1D frequency between patients' sibs in general (around 6–10%) and MHC-identical sibs (about 16–17%) and the MHC recessive mode of inheritance strongly suggest the MHC-marked gene(s) is (are) causal, i.e., *required for disease susceptibility*. Given the frequency gap of T1D concordance among MHC-identical sibs (16–17%) and MZTs (40–50%), and just as the MHC region T1D susceptibility gene is Mendelian recessive, we posit a small number of *required* Mendelian recessive or (less likely) dominant non-MHC genes. The term "oligogenic" in the SEMO model refers to these few (two to four or some other very limited number of) required genes. Our viewpoint is that, if monogenic traits and polygenic traits exist, it is highly plausible (particularly for T1D) that oligogenic traits also exist.

The SEMO model [6,22,23] stipulates a multiplicative (epistatic) interaction [71,81,82] of the required genes such that the general population prevalence of T1D-susceptible individuals is equal to the multiplication product of the frequencies of susceptible persons at each locus. That frequency multiplied by 40–50% (MZT concordance) gives T1D prevalence. Fig. 2 is a schematic of the genetic component of the SEMO model [22]. Although three genes are shown, the model could involve any small number of genes (at least two and probably three). The non-MHC gene(s) may be recessive, dominant or a mixture of dominant and recessive. The evidence we present below strongly suggests only recessive inheritance for each of the required genes.

Fig. 3 shows a hypothetical family based on the SEMO model in which a single member has a polygenic disease with three required recessive susceptibility genes, *1*, *2* and *3* [6]. Disease susceptibility alleles are designated by *D* (*1D*, *2D* and *3D*) and protective or normal or non-susceptibility alleles by *N* (1N, 2N, and 3N). Shown here is one of 27 possible mating pairs with the potential to generate offspring homozygous for *D* alleles at all three susceptibility loci. The healthy non-susceptible father I-1 is homozygous for susceptibility alleles at locus *2* and the healthy non-susceptible mother I-2 is homozygous for susceptibility alleles at locus *1*. Of their eight children, only II-2 and II-7 are genetically susceptible; they have only susceptibility alleles at all three susceptibility loci. Only II-7 has T1D, however. Although Fig. 3 shows two of eight children with susceptibility for T1D, given the parental genotypes in this family, only one child in 16 would be expected, on average, to be susceptible (i.e., homozygous for susceptibility alleles at all three required susceptibility loci). Because of incomplete penetrance (50%), only one in 32 on average would have T1D.

Fig. 4 shows the relationship between allele/haplotype sharing by ASPs and aggregate susceptibility allele frequency at that locus for recessive and dominant MHC inheritance in the SEMO model [6]. It is clear that, for T1D, the MHC is recessive and dominant inheritance is excluded. For multiple sclerosis (MS), a complex MHC-associated autoimmune disease, either mode of inheritance fits the sharing and no conclusion can be drawn (Fig. 4).

We previously estimated the aggregate frequency of MHC-linked susceptibility alleles in the general population [6] using the average MHC allele/haplotype sharing by ASPs as reported in many studies [71] of 55% sharing 2, 38% sharing 1 and 7% sharing 0. We used the possible parental genotypes that could produce susceptible offspring,

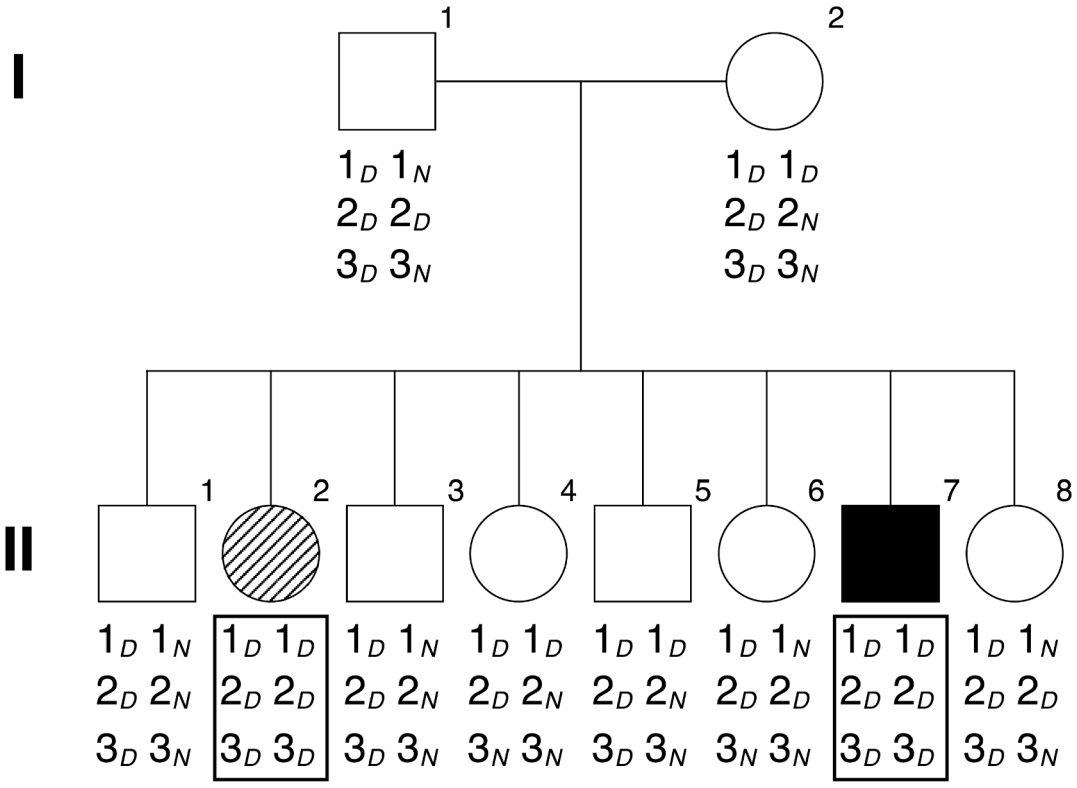

Fig. 3. Postulated disease allele distribution for a recessive 3-locus model in a family with one child with T1D (II-7) (dark square). Susceptibility alleles are "*D*" and protective alleles are "*N*." We have assumed three required recessive loci for T1D susceptibility, 1, 2 and 3, that interact multiplicatively with each other [6]. Note that individual II-2 (cross-hatched circle) is also genetically susceptible to T1D but disease susceptibility is not penetrant.

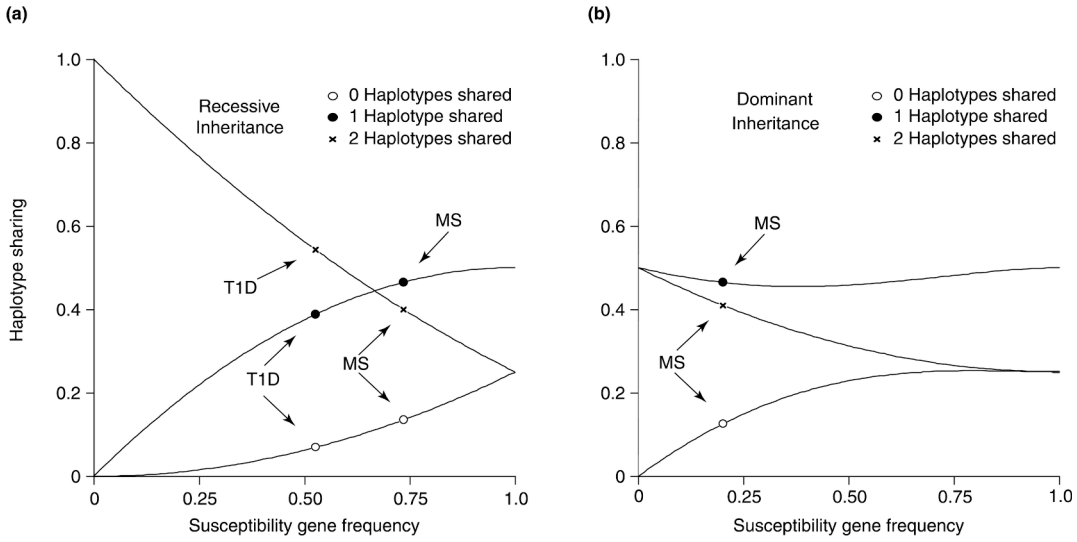

Fig. 4. Predicted frequencies of 2, 1 and 0 alleles/haplotypes shared for recessive (a) and dominant (b) inheritance of an MHC susceptibility gene at various aggregate allele frequencies of D. Observed allele/haplotype distributions for MS and T1D are marked. Note that for MS, the distribution fits either form of inheritance. There is no dominant solution for TID. (Reproduced with permission from refs. 6 and 61.)

assuming Mendelian recessive inheritance and the Hardy-Weinberg equilibrium, to yield the frequency of MHC T1D-susceptible individuals in a European population. This frequency is the fraction of ASPs who share no MHC haplotypes (0.07) multiplied by 4 or 0.28 [6]. In families where there is no ASP sharing of MHC alleles/haplotypes, all MHC alleles/haplotypes mark T1D susceptibility.

Since the MHC-linked gene for T1D is recessive, the square root of 0.28 (0.525) is the aggregate T1D susceptibility allele frequency. From prevalence estimates of T1D (four per 1000) and an estimate of penetrance (for convenience, we used 50%), the total fraction of Caucasians genetically susceptible to T1D is 0.008. This divided by the frequency of MHC susceptibles among European Caucasians (0.28) gives the frequency of susceptible individuals at all non-MHC susceptibility loci (0.029). If there are two unlinked recessive non-MHC susceptibility loci of equal frequency, each would have an aggregate susceptibility allele frequency of 0.41 (since $(0.41)^2$ x $(0.41)^2 = 0.029$). If three such genes exist, their aggregate susceptibility allele frequency would be 0.55 (since $(0.55)^2$ x $(0.55)^2$ x $(0.55)^2 = 0.029$).

A consequence of the SEMO model for T1D is that selection against *any* one of the required loci (selective selection) [22] would result in selection against this previously fatal disease. If today's two parents are the descendants of different previously isolated European subpopulations in which negative selection against *different* susceptibility genes occurred, their children will have a higher risk of T1D because they are more likely than either parent to carry a complete set of all of the *required* susceptibility alleles (Fig. 5a). Thus, specific subpopulation admixture could result in gene complementation in which a higher frequency of the offspring than either parental population have full genetic susceptibility at all *required* genes. While such admixture of parental populations is likely, given increasing globalization and population migration, we stress that not *all* admixture would result in increasing incidence. Some subpopulation admixture would not result in increased incidence overall for two groups (Fig. 5b), but aggregate admixture of many subpopulations could explain the rising incidence of T1D.

Four lines of evidence indicate that HLA-DR3 and -DR4 mark two such previously isolated European subpopulations [23]. Below, we summarize the results from that publication. First, self-reported admixture of European subpopulations among parents in control and T1D families in the Boston area is increasing by 25%–30% per generation. The parents of T1D patients had twice the self-reported European subpopulation ethnicity disparity (54%) than that of parents in control families (27%) ($p < 0.001$). Second, parents who transmitted HLA-DR3 to their T1D children had significantly lower frequencies of HLA-A2 in their own *non-transmitted* MHC haplotypes than those who transmitted HLA-DR4 ($p < 0.001$). Thus, HLA-DR3-transmitting parents had a different distribution of *HLA-A* alleles on their FC (*non-disease*) haplotypes than HLA-DR4-transmitting parents, indicating the two parental populations likely had different subpopulation ancestries. Third, parents who transmitted HLA-DR3 to their children with T1D also had significantly different insulin gene (*INS)* allele frequencies as compared with parents who transmitted HLA-DR4 to their children with T1D ($p < 0.05$). Fourth, parents of T1D patients were significantly *deficient in* HLA-DR3/4 heterozygotes ($p < 0.01$), reflecting their distinct origins. With the current increase in ethnic mixing, including from populations with non-DR3 and non-DR4 high-risk markers, we expect to see dilution of these classical HLA-DR3 and HLA-DR4 high-risk T1D susceptibility markers among European Caucasian patients and this has been reported [10]. Although underlying differences exist in autoantibody marker progression between DR3 and DR4 carriers [83,84], there are a number of possible explanations for this that do not require one to hypothesize differential genetic subtypes of T1D. In any case, DR3 and DR4 variants (and their highly linked HLA-DQ covariants) are not different genes but different sets of alleles on the same genes. These may simply be marking linked allele variants at the causal MHC susceptibility gene(s).

Although Finland has the highest T1D incidence in the world, there are marked differences in incidence within the country [85]. The highest incidence in Finland occurs where two genetically distinct (southwestern and northern, based on Y chromosome markers) populations [86] meet. Consistent with this observation is the difference in *HLA-DQB1* allele frequencies in patients and newborns in general between these regions [87]. Similarly, in Great Britain the highest T1D incidence occurs near Plymouth [88] where Celts (with a high frequency of HLA-DR3) and Anglo-Saxons (with a high frequency of HLA-DR4) meet [89]. In Hawaii, the incidence of T1D is higher among people of mixed European-Native Hawaiian background than among people of either ethnic group alone [90]. Although neither NZB nor NZW mice have lupus, their offspring have a high incidence [91], suggesting an animal model for selective selection and parental admixture. We stress that not *all* population admixture results in increased incidence. As shown in Fig. 5b, if two separate populations selected against the *same* required T1D susceptibility gene, their admixed T1D incidence would be the average of the two populations' prior

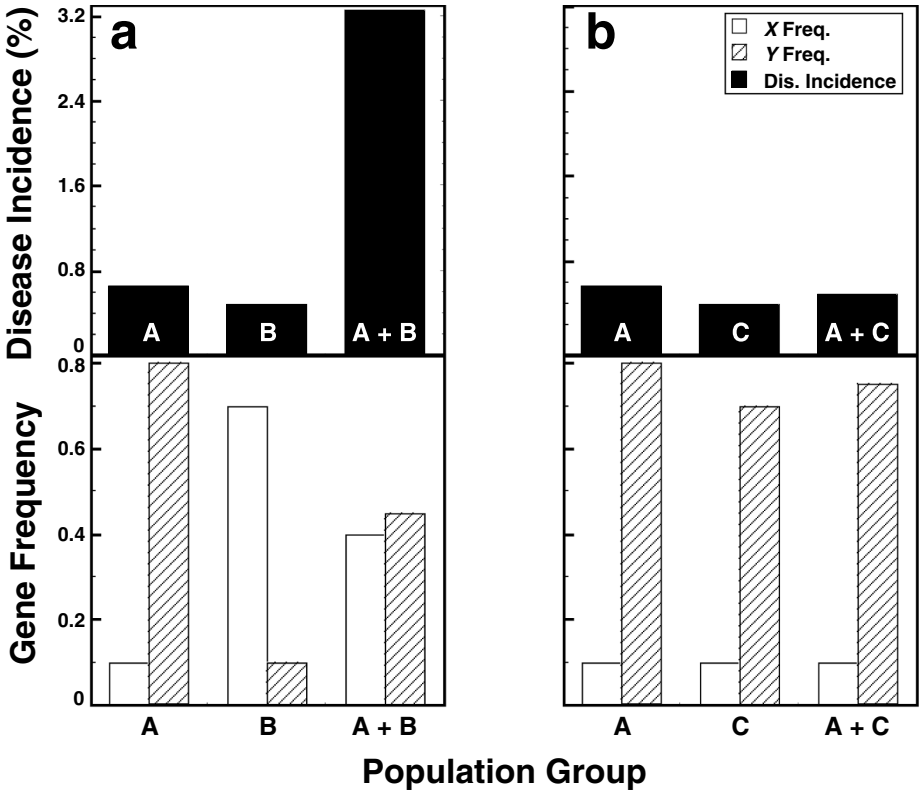

Fig. 5. Hypothetical susceptibility gene frequencies and disease susceptibility in previously separate and currently admixed populations. Before admixture, populations A, B and C have similar low susceptibility to disease (A = 0.64%, B and C = 0.49%). Populations A and B have selected against different susceptibility genes and thus have inversely proportional frequencies of X and Y (Fig. 5a), whereas populations A and C have selected against the same susceptibility genes and have similar frequencies of X and Y (Fig. 5b). After genetic mixing in equal parts, the disease susceptibility in the offspring of A + B is five times or more than that of A + C or that of the parental populations, A, B or C. (Modified version with permission from Ref. [22].)

incidences. This is in contrast to the high incidence where different loci were selected against (Fig. 5a).

5.1. The fit of the genetic component with observed facts

The genetic component of the SEMO model provides estimates of the fractions of affected sibs, children and parents of T1D patients [6] that agree closely with observed fractions [7]. The model explains why susceptibility alleles are common in the general population and yet only a small percentage of people with high-risk HLA-DR3/4 genotypes have T1D. It also explains why so few T1D patients have affected relatives. The aggregate frequency of T1D MHC susceptibility alleles of 0.525 fits well with the observation that the sum (0.30) of HLA-DR3 (around 0.14) and HLA-DR4, DQ8 (around 0.16) frequencies in a typical American European population constitutes nearly 70% of T1D susceptibility alleles. The model also explains why only a few (2%–3%) of MHC T1D-susceptible persons, who are over 25% of the Caucasian population, are susceptible to T1D: they lack *required* non-MHC gene susceptibility.

The SEMO model, combining historic selective negative selection against different susceptibility loci and current parental admixture of previously isolated subpopulations, explains: (a) the rising incidence of T1D (and, perhaps, other polygenic diseases such as lupus, MS and Crohn's disease); (b) the “high risk” and excess frequency of HLA-DR3/4 heterozygotes beyond the Hardy-Weinberg expectation among some T1D patient populations; (c) the changing MHC haplotype and allele composition of T1D patient populations; and (d) the declining frequencies of traditional HLA-DR3 and -DR4 susceptibility markers in some T1D patient populations. The model predicts that the rising incidence of T1D will subside as the mixed populations approach homogeneity as may already be starting in Sweden [92,93].

The SEMO model also predicts the offspring of two T1D-affected parents would all be fully T1D genetically susceptible if all causal genes are recessive. The ERP model would not make this prediction. If the SEMO model is correct, the offspring incidence would be identical to MZT concordance and all offspring would have susceptibility alleles at all T1D causal genetic loci. The SEMO model also predicts in these families that pairs of offspring, whether affected or not, would have random sharing of all loci on chromosome 6 of two, one or no haplotypes (25%, 50%, 25%). Families with two T1D parents have been followed since 1955 in Karlsburg, Germany [12]. The overall incidence among 58 offspring in 46 families was 43% in 2001, remarkably identical to MZT concordance and therefore consistent with the SEMO model. Moreover, the incidences during the first three decades of study were 54%, 42% and 50%; like MZT concordance, they did not rise. Since the early environments of the children were obviously more varied than those among MZTs of T1D patients, and yet the incidence of T1D was identical to that of MZTs of T1D patients, a randomly-distributed yet invariant (i.e., stochastic) epigenetic explanation [19] applicable to all susceptible individuals is far more likely than some particularly pathogenic environment among these particular German families.

The fact that the proportion of T1D offspring of two T1D parents is identical to MZT concordance for T1D suggests (but does not prove) that the non-MHC causal gene(s) is (are) also recessive. If any causal gene were dominant and both parents were heterozygotes, a quarter of their offspring would not be susceptible to T1D. This would result in a lower incidence of T1D in the children than MZT concordance. Overall, the findings in the families with both parents affected by T1D are inconsistent with the ERP model but are completely consistent with the SEMO model. It is time to abandon the ERP view that T1D is the result of a strong effect of (a) MHC gene(s), the environment and a variable number of potential non-required genes of small effect that interact additively.

6. The non-genetic component of the SEMO model for T1D

The evidence against a variable external trigger (e.g., the “environment”) converting a fully susceptible person to a T1D patient leads us to conclude that a randomly distributed (stochastic) event involving

one of the causal genes is an intrinsic trigger of this process [61,79]. If an hypothesized external trigger were to be the underlying basis of the stochastic intrinsic process, it would, presumably, have to be a common and invariable molecule, pathogen or process worldwide. Although there may be different biological mechanisms or events that exhibit a random distribution (e.g., B- or T-cell antigen receptor generation), our hypothesis is that the stochastic event is most likely epigenetic [19]. This interpretation is supported by the observed divergence of expression of different individual genes in healthy pairs of MZTs [94] with increasing age. Therefore, epigenetics could explain the later childhood onset of T1D.

To understand the non-genetic component fully, one must know the molecules involved, the cellular/tissue site(s) of expression of the relevant genes, and the event timing. There is no information about any of those elements. Several concepts help to identify elements of the non-genetic component of T1D. First, we propose T1D concordance among MZTs of T1D patients not only defines penetrance in the MZTs but also in all unrelated genetically fully-susceptible persons in the population. Second, we have postulated [61,79] an age-related stochastic epigenetic change that alters expression of at least one (or perhaps even all) of the causal genes for T1D as the underlying mechanism for MZT discordance. However, it is also possible that such an epigenetic change is triggered at a non-susceptibility gene (i.e., regardless of that gene's allele status) *by and not upon* one of the causal T1D genes. Finally, any future hypotheses regarding the non-genetic mechanism should be consistent with the evidence that the non-genetic component frequency or rate (i.e., MZT concordance (Fig. 1) and the incidence among offspring of two parents with T1D) appears to be both constant over time and similar in different ethnic groups.

7. Future directions and testable predictions of the SEMO model

The SEMO model is consistent with what is currently known regarding T1D genetics and onset. We propose several required steps toward unraveling T1D pathogenesis. Several questions must be addressed to test the model further. The first and chief task is to identify the required causal T1D susceptibility genes. Based on linkage results, we believe this must focus on family-based full sequence searches within the MHC and surrounding regions of chromosome 6. Since essentially all genetic linkage to T1D is on chromosome 6, we predict the non-MHC causal genes must also be on chromosome 6. This includes the entire short arm and the centromeric third of the long arm of chromosome 6.

Elements of the putative epigenetic effect must be identified next. If specific causal genes were known, changes in their expression states could be tested for correlation with T1D in unrelated healthy persons, T1D patients, concordant and discordant MZTs and autoantibody positive and negative offspring of parents with T1D. It is critical that the appropriate tissue and/or cell type(s) be identified and analyzed. It will be necessary to study expression changes before, during development and after disease occurs. The SEMO model suggests that gene expression changes relevant to T1D pathogenesis should be found (at some critical time point) in all (future) patients with T1D as well as in 40–50% of unaffected individuals (who are not fully genetically susceptible at *all* required loci). However, no healthy non-autoantibody positive MZT of a patient with T1D should show these epigenetic changes nor should *any* long-term MZT non-progressor (whether autoantibody positive or not). Below, we list several predictions based on the SEMO model.

1. All causal T1D susceptibility genes are on chromosome 6. Family-based haplotype analyses [14] of non-MHC genes will contradict prior reports of association of non-chromosome 6 putative susceptibility genes.
2. The HLA-DR3/DR4, DQ8 excess among T1D patients is not general but involves only some specific CEHs and cohorts, reflecting specific subpopulation stratification and susceptibility gene complementation (like that shown in Fig. 5).
3. MHC allele/haplotype sharing (and sharing throughout the entirety of chromosome 6) by ASPs among the children of two T1D parents will be random, i.e., 25% sharing 2, 50% sharing 1 and 25% sharing none, since all alleles/haplotypes at susceptibility loci are susceptibility.
4. The rising frequency in T1D incidence will level off as subpopulation mixing approaches equilibrium.
5. Expression of one of the causal genes (or of (a) gene(s) under the expression control of one of the causal genes) is different at one time in one cell type critical for T1D initiation in discordant MZTs and among the T1D-affected offspring as compared to the unaffected offspring of two T1D parents.

8. Conclusion

No causal gene or genetic model for T1D has been definitively established, although the MHC and most of human chromosome 6p are strongly genetically linked to the disease. The dominant analytical method of genetic epidemiology, risk analysis, has led to the ERP model, based on speculation that many non-MHC genes of small effect and the environment participate, to varying extents, in T1D pathogenesis. However, non-chromosome 6 genes show little to no genetic linkage to T1D, and we provide strong evidence that T1D incidence is not influenced by the environment. We propose a Mendelian-based SEMO causal model in which a limited number of required recessive chromosome 6 genes participate in T1D pathogenesis along with a stochastic (i.e., randomly distributed) but invariable (likely epigenetic) process that converts a susceptible person into a patient. At the very least, we hope our perspective and proposed methodologies offer new avenues to understand hidden trait complexity [95]. The SEMO model is for T1D specifically. Elements of the model might apply to other complex genetic diseases, but we are not attempting to explain the basis for autoimmunity in general–just one specific autoimmune disease: T1D.

Acknowledgments

This work was supported by Juvenile Diabetes Research Foundation grants 5-2007-788 and 1-2008-472, National Institutes of Health grant HL-29583 and institutional funds from the Immune Disease Institute/Program in Cellular and Molecular Medicine, Boston Children's Hospital. We thank Drs. Richard J. Grand, Åke Lernmark, Joseph A. Majzoub, Arjun K. Manrai, David G. Nathan, Eileen Remold-O'Donnell, Robert S. Schwartz, Claes B. Wollheim, Frank A. Wollheim and Ms. Zareen Vadva for critical and helpful comments on the manuscript.

Abbreviations

ASP	affected sib pair
CEH	conserved extended haplotype
DZT	dizygotic twin
ERP	environmentally-triggered risk-based polygenic
FC	family control
GSE	gluten-sensitive enteropathy
MHC	major histocompatibility complex
MS	multiple sclerosis
MZT	monozygotic twin
ref	reference
SEMO	stochastic epigenetically-triggered Mendelian oligogenic
T1D	type 1 diabetes

Author contributions

CAA and CEL wrote the manuscript. DRA and MT contributed ideas and criticism, and all authors contributed to editing the manuscript.

Conflicts of interest

The authors have no relevant conflict of interest to disclose.

References

[1] R.M. Nelson, M.E. Pettersson, Ö. Carlborg, A century after Fisher: time for a new paradigm in quantitative genetics, Trends Genet. 29 (2013) 669–676.
[2] R.A. Fisher, The Genetical Theory of Natural Selection, Clarendon Press, Oxford, UK, 1930.
[3] J.P.A. Ioannidis, Making optimal use of and extending beyond polygenic additive liability models, Hum. Hered. 80 (2015) 158–161.
[4] E. Génin, F. Clerget-Darpoux, Revisiting the polygenic additive liability model through the example of diabetes mellitus, Hum. Hered. 80 (2015) 171–177.
[5] P.M. Visscher, W.G. Hill, N.R. Wray, Heritability in the genomics era–concepts and misconceptions, Nat. Rev. Genet. 9 (2008) 255–266.
[6] C.A. Alper, D.P. Dubey, E.J. Yunis, Z. Awdeh, A simple estimate of the general population frequency of the MHC susceptibility gene for autoimmune polygenic disease, Exp. Clin. Immunogenet. 17 (2000) 138–147.
[7] L.J. Raffel, M.O. Goodarzi, Diabetes mellitus, in: D.L. Rimoin, J.M. Connor, R.E. Pyeritz, B.R. Korf (Eds.), Emery and Rimoin's Principles and Practice of Medical Genetics, sixth ed., Churchill-Livingstone, New York, 2013, pp. 1–58 (Chapter 86).
[8] J.O. Craig, J.A. Ford, A.C. McCuish, Rising incidence of childhood diabetes, Br. Med. J. 1 (6061) (1977) 639.
[9] T.H. Lipman, L.E. Levitt Katz, S.J. Ratcliffe, K.M. Murphy, A. Aguilar, I. Rezvani, C.J. Howe, S. Fadia, E. Suarez, Increasing incidence of type 1 diabetes in youth: twenty years of the Philadelphia pediatric dabetes registry, Diabetes Care 36 (2013) 1597–1603.
[10] K.M. Gillespie, S.C. Bain, A.H. Barnett, The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes, Lancet 364 (2004) 1699–1700.
[11] T.A. Aly, A. Ide, M.M. Jahromi, J.M. Barker, M.S. Fernando, S.R. Babu, L. Yu, D. Miao, H.A. Erlich, P.R. Fain, K.J. Barriga, J.M. Norris, M.J. Rewers, G.S. Eisenbarth, Extreme genetic risk for type 1A diabetes, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 14074–14079.
[12] I. Rjasanowski, I. Kőlting, W. Kerner, Frequency of diabetes transmission from two type 1 diabetic parents to their children, Diabetes Care 26 (2003) 2219–2220.
[13] L. Nisticò, D. Iafusco, A. Galderisi, C. Fagnani, R. Cotichini, V. Toccaceli, M.A. Stazi, The Study Group on Diabetes of the Italian Society of Pediatric Endocrinology and Diabetology, Emerging effects of early environmental factors over genetic background for type 1 diabetes susceptibility: evidence from a nationwide Italian twin study, J. Clin. Endocrinol. Metab. 97 (8) (2012) E1483–E1491.
[14] D. Raum, Z. Awdeh, E.J. Yunis, C.A. Alper, K.H. Gabbay, Extended major histocompatibility complex haplotypes in type I diabetes mellitus, J. Clin. Invest. 74 (1984) 449–454.
[15] R.A. Fisher, The correlation between relatives on the supposition of Mendelian inheritance, Trans. R. Soc. Edinb. 52 (1918) 399–433.
[16] S.E. Hodge, What association analysis can and cannot tell us about the genetics of complex disease, Am. J. Med. Genet. 54 (1994) 318–323.
[17] A.M. Madsen, S.E. Hodge, R. Ottman, Causal models for investigating complex disease: I. A primer, Hum. Hered. 72 (2011) 54–62.
[18] J.N. Hirschhorn, Genetic epidemiology of type 1 diabetes, Pediatr. Diabetes 4 (2003) 87–100.
[19] A. Petronis, Epigenetics as a unifying principle in the aetiology of complex traits and diseases, Nature 465 (2010) 721–727.
[20] M. Knip, R. Veijola, S.M. Virtanen, H. Hyöty, O. Vaarala, H.K. Åkerblom, Environmental triggers and determinants of type 1 diabetes, Diabetes 54 (Suppl 2) (2005) S125–S136.
[21] J. Pitkäniemi, P. Onkamo, J. Tuomilehto, E. Arjas, Increasing incidence of Type I diabetes – role for genes? BMC Genet. 5 (2004) 5.
[22] Z.L. Awdeh, C.A. Alper, Mendelian inheritance of polygenic diseases: a hypothetical basis for increasing incidence, Med. Hypotheses 64 (2005) 495–498.
[23] Z.L. Awdeh, E.J. Yunis, M.J. Audeh, D. Fici, A. Pugliese, C.E. Larsen, C.A. Alper, A genetic explanation for the rising incidence of type 1 diabetes, a polygenic disease, J. Autoimmun. 27 (2006) 174–181.
[24] A. Paun, C. Yau, J.S. Danska, The influence of the microbiome on type 1 diabetes, J. Immunol. 198 (2017) 590–595.
[25] M.E. Craig, S. Nair, H. Stein, W.D. Rawlinson, Viruses and type 1 diabetes: a new look at an old story, Pediatr. Diabetes 14 (2013) 149–158.
[26] S.J. Richardson, N.G. Morgan, A.K. Foulis, Pancreatic pathology in type 1 diabetes mellitus, Endocr. Pathol. 25 (2014) 80–92.
[27] A. Op de Beeck, D.L. Eizirik, Viral infections in type 1 diabetes mellitus–why the ß cells? Nat. Rev. Endocrinol. 12 (2016) 263–273.
[28] T. Rodriguez-Calvo, S. Sabouri, F. Anquetil, M.G. von Herrath, The viral paradigm in type 1 diabetes: who are the main suspects? Autoimmun. Rev. 15 (2016) 964–969.
[29] J.E. Craighead, Viral diabetes in man and experimental animals, Am. J. Med. 70 (1981) 127–134.
[30] E.V. Moltchanova, N. Schreier, N. Lammi, M. Karvonen, Seasonal variation of diagnosis of type 1 diabetes mellitus in children worldwide, Diabet. Med. 26 (2009) 673–678.
[31] A.G. Ziegler, M. Rewers, O. Simell, T. Simell, J. Lempainen, A. Steck, C. Winkler, J. Ilonen, R. Veijola, M. Knip, E. Bonifacio, G.S. Eisenbarth, Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children, JAMA 309 (2013) 2473–2479.
[32] R. Moulder, R. Lahesmaa, Early signs of disease in type 1 diabetes, Pediatr. Diabetes 17 (Suppl 22) (2016) 43–48.
[33] J.P. Krischer, K.F. Lynch, Å. Lernmark, W.A. Hagopian, M.J. Rewers, J.-X. She, J. Toppari, A.-G. Ziegler, B. AkolkarTEDDY Study Group, Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study, Diabetes Care 40 (2017) 1194–1202.
[34] P.G. Nelson, D.A. Pyke, D.R. Gamble, Viruses and the aetiology of diabetes: a study in identical twins, Br. Med. J. 4 (5991) (1975) 249–251.
[35] T. Tuvemo, G. Dahlquist, G. Frisk, L. Blom, G. Friman, M. Landin-Olsson, H. Diderholm, The Swedish childhood diabetes study III: IgM against coxsackie B viruses in newly diagnosed type 1 (insulin-dependent) diabetic children–no evidence of increased antibody frequency, Diabetologia 32 (1989) 745–747.
[36] V. Hyttinen, J. Kaprio, L. Kinnunen, L. Koskenvuo, J. Tuomilehto, Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs, Diabetes 52 (2003) 1052–1055.
[37] T. Lorenzen, F. Pociot, P. Hougaard, J. Nerup, Long-term risk of IDDM in first-degree relatives of patients with IDDM, Diabetologia 37 (1994) 321–327.
[38] M.J. Redondo, P.R. Fain, J.P. Krischer, L. Yu, D. Cuthbertson, W.E. Winter, G.S. Eisenbarth, DPT-1 Study Group, Expression of beta-cell autoimmunity does not differ between potential dizygotic twins and siblings of patients with type 1 diabetes, J. Autoimmun. 23 (2004) 275–279.
[39] G. Bruno, G. Pagano, F. Faggiano, A. De Salvia, F. Merletti, Effect of Sardinian heritage on risk and age of onset of type 1 diabetes: a demographic case-control study of Sardinian migrants, Int. J. Epidemiol. 29 (2000) 532–535.
[40] S. Ehehalt, P. Popovic, S. Muntoni, S. Muntoni, A. Willasch, R. Hub, M.B. Ranke, A. Neu, DIARY Group Baden-Wuerttemberg, Incidence of diabetes mellitus among children of Italian migrants substantiates the role of genetic factors in the pathogenesis of type 1 diabetes, Eur. J. Pediatr. 168 (2009) 613–617.
[41] N.T. Raymond, J.R. Jones, P.G.F. Swift, M.J. Davies, I.G. Lawrence, P.G. McNally, M.L. Burden, R. Gregory, A.C. Burden, J.L. Botha, Comparative incidence of type I diabetes in children aged under 15 years from south Asian and white or other ethnic backgrounds in Leicestershire, UK, 1989 to 1998, Diabetologia 44 (Suppl 3) (2001) B32–B36.
[42] R.G. Feltbower, H.J. Bodansky, P.A. McKinney, J. Houghton, C.R. Stephenson, D. Haigh, Trends in the incidence of childhood diabetes in south Asians and other children in Bradford, UK, Diabet. Med. 19 (2002) 162–166.
[43] U. Söderström, J. Åman, A. Hjern, Being born in Sweden increases the risk for type 1 diabetes - a study of migration of children to Sweden as a natural experiment, Acta Paediatr. 101 (2012) 73–77.
[44] A. Hjern, U. Söderström, Parental country of birth is a major determinant of childhood type 1 diabetes in Sweden, Pediatr. Diabetes 9 (2008) 35–39.
[45] J. Ji, K. Hemminki, J. Sundquist, K. Sundquist, Ethnic differences in incidence of type 1 diabetes among second-generation immigrants and adoptees from abroad, J. Clin. Endocrinol. Metab. 95 (2010) 847–850.
[46] A. Huber, F. Menconi, S. Corathers, E.M. Jacobson, Y. Tomer, Joint genetic susceptibility to type 1 diabetes and autoimmune thyroiditis: from epidemiology to mechanisms, Endocr. Rev. 29 (2008) 697–725.
[47] M.S. Gottlieb, H.F. Root, Diabetes mellitus in twins, Diabetes 17 (1968) 693–704.
[48] B. Harvald, M. Hauge, Selection in diabetes in modern society, Acta Med. Scand. 173 (1963) 459–465.
[49] A. Matsuda, T. Kuzuya, Diabetic twins in Japan, Diabetes Res. Clin. Pract. 24 (Suppl) (1994) S63–S67.
[50] R.D.G. Leslie, Diabetic twin studies, in: R.D.G. Leslie (Ed.), Causes of Diabetes: Genetic and Environmental Factors, John Wiley and Sons, Hoboken, NJ, 1993, pp. 61–82.
[51] D. Kumar, N.S. Gemayel, D. Deapen, D. Kapadia, P.H. Yamashita, M. Lee, J.H. Dwyer, P. Roy-Burman, G.A. Bray, T.M. Mack, North-American twins with IDDM. Genetic, etiological, and clinical significance of disease concordance according to age, zygosity, and the interval after diagnosis in first twin, Diabetes 42 (1993) 1351–1363.
[52] A.H. Barnett, C. Eff, R.D.G. Leslie, D.A. Pyke, Diabetes in identical twins. A study of 200 pairs, Diabetologia 20 (1981) 87–93.
[53] N. Tajima, A. Morimoto, Epidemiology of childhood diabetes mellitus in Japan, Pediatr. Endocrinol. Rev. 10 (Suppl 1) (2012) 44–50.
[54] V. Harjutsalo, L. Sjöberg, J. Tuomilehto, Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study, Lancet 371 (2008) 1777–1782.
[55] S.E. Neville, K.S. Boye, W.S. Montgomery, K. Iwamoto, M. Okamura, R.P. Hayes, Diabetes in Japan: a review of disease burden and approaches to treatment, Diabetes Metab. Res. Rev. 25 (2009) 705–716.
[56] P. Concannon, W.-M. Chen, C. Julier, G. Morahan, B. Akolkar, H.A. Erlich, J.E. Hilner, J. Nerup, C. Nierras, F. Pociot, J.A. Todd, S.S. Rich, The Type 1 Diabetes Genetics Consortium, Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the type 1 diabetes genetics consortium, Diabetes 58 (2009) 1018–1022.
[57] G. Morahan, M. Mehta, I. James, W.-M. Chen, B. Akolkar, H.A. Erlich, J.E. Hilner, C. Julier, J. Nerup, C. Nierras, F. Pociot, J.A. Todd, S.S. Rich, Tests for genetic interactions in type 1 diabetes. Linkage and stratification analyses of 4,422 affected sib-pairs, Diabetes 60 (2011) 1030–1040.

[58] X. Hu, A.J. Deutsch, T.L. Lenz, S. Onengut-Gumuscu, B. Han, W.-M. Chen, J.M.M. Howson, J.A. Todd, P.I.W. de Bakker, S.S. Rich, S. Raychaudhuri, Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk, Nat. Genet. 47 (2015) 898–905.
[59] Z.L. Awdeh, D. Raum, E.J. Yunis, C.A. Alper, Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man, Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 259–263.
[60] M.A. Degli-Esposti, A.L. Leaver, F.T. Christiansen, C.S. Witt, L.J. Abraham, R.L. Dawkins, Ancestral haplotypes: conserved population MHC haplotypes, Hum. Immunol. 34 (1992) 242–252.
[61] C.E. Larsen, C.A. Alper, The genetics of HLA-associated disease, Curr. Opin. Immunol. 16 (2004) 660–667.
[62] C.A. Alper, C.E. Larsen, D.P. Dubey, Z.L. Awdeh, D.A. Fici, E.J. Yunis, The haplotype structure of the human major histocompatibility complex, Hum. Immunol. 67 (2006) 73–84.
[63] J.R. Bilbao, B. Calvo, A.M. Aransay, A. Martin-Pagola, G. Perez de Nanclares, T.A. Aly, I. Rica, J.C. Vitoria, S. Gaztambide, J. Noble, P.R. Fain, Z.L. Awdeh, C.A. Alper, L. Castaño, Conserved extended haplotypes discriminate HLA-DR3-homozygous Basque patients with type 1 diabetes mellitus and celiac disease, Gene Immun. 7 (2006) 550–554.
[64] T.A. Aly, E. Eller, A. Ide, K. Gowan, S.R. Babu, H.A. Erlich, M.J. Rewers, G.S. Eisenbarth, P.R. Fain, Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype, Diabetes 55 (2006) 1265–1269.
[65] W.P. Smith, Q. Vu, S.S. Li, J.A. Hansen, L.P. Zhao, D.E. Geraghty, Toward understanding MHC associations: partial resequencing of 46 distinct HLA haplotypes, Genomics 87 (2006) 561–571.
[66] A. Szilágyi, Z. Bánlaki, E. Pozsonyi, E.J. Yunis, Z.L. Awdeh, A. Hossó, K. Rajczy, C.E. Larsen, D.A. Fici, C.A. Alper, G. Füst, Frequent occurrence of conserved extended haplotypes (CEHs) in two Caucasian populations, Mol. Immunol. 47 (2010) 1899–1904.
[67] C.E. Larsen, D.R. Alford, M.R. Trautwein, Y.K. Jalloh, J.L. Tarnacki, S.K. Kunnenkeri, D.A. Fici, E.J. Yunis, Z.L. Awdeh, C.A. Alper, Dominant sequences of human major histocompatibility complex conserved extended haplotypes from *HLA-DQA2* to *DAXX*, PLoS Genet. 10 (2014) e1004637.
[68] L.P. Ryder, M. Christy, J. Nerup, P. Platz, M. Thomsen, HLA studies in diabetics, Adv. Exp. Med. Biol. 119 (1979) 41–48.
[69] A. Svejgaard, P. Platz, L.P. Ryder, Insulin dependent diabetes mellitus, in: P.I. Terasaki (Ed.), Histocompatibility Testing 1980, UCLA Tissue Typing Laboratory, Los Angeles, 1981, pp. 638–656.
[70] P. Rubinstein, N. Suciu-Foca, J.F. Nicholson, Genetics of juvenile diabetes mellitus. A recessive gene closely linked to HLA-D and with 50 per cent penetrance, N. Engl. J. Med. 297 (1976) 1036–1040.
[71] G. Thomson, A two locus model for juvenile diabetes, Ann. Hum. Genet. 43 (1980) 383–388.
[72] D. Raum, C.A. Alper, R. Stein, K.H. Gabbay, Genetic marker for insulin-dependent diabetes mellitus, Lancet 313 (1979) 1208–1210.
[73] D. Raum, Z. Awdeh, C.A. Alper, BF types and the mode of inheritance of insulin-dependent diabetes mellitus (IDDM), Immunogenetics 12 (1981) 59–74.
[74] P. Rubinstein, M. Walker, N. Mollen, C. Carpenter, S. Beckerman, N. Suciu-Foca, R. McEvoy, F. Ginsberg-Fellner, No excess of DR*3/4 in Ashkenazi Jewish or Hispanic IDDM patients, Diabetes 39 (1990) 1138–1143.
[75] F. Cucca, F. Muntoni, R. Lampis, F. Frau, L. Argiolas, M. Silvetti, E. Angius, A. Cao, S. De Virgiliis, M. Congia, Combinations of specific DRB1, DQA1, DQB1 haplotypes are associated with insulin-dependent diabetes mellitus in Sardinia, Hum. Immunol. 37 (1993) 85–94.
[76] P. Rubinstein, M. Walker, F. Ginsberg-Fellner, Excess of DR3/4 in type 1 diabetes. What does it portend? Diabetes 35 (1986) 985–989.
[77] A.G. Cudworth, E. Wolf, The genetics of type 1 (insulin-dependent) diabetes, Curr. Probl. Clin. Biochem. 12 (1983) 45–64.
[78] J.I. Rotter, C.E. Anderson, R. Rubin, J.E. Congleton, P.I. Terasaki, D.L. Rimoin, HLA genotypic study of insulin-dependent diabetes the excess of DR3/DR4 heterozygotes allows rejection of the recessive hypothesis, Diabetes 32 (1983) 169–174.
[79] C.A. Alper, Z. Awdeh, Incomplete penetrance of MHC susceptibility genes: prospective analysis of polygenic MHC-determined traits, Tissue Antigens 56 (2000) 199–206.
[80] C.A. Alper, Z. Husain, C.E. Larsen, D.P. Dubey, R. Stein, C. Day, A. Baker, H. Beyan, M. Hawa, T.O. Ola, R.D. Leslie, Incomplete penetrance of susceptibility genes for MHC-determined immunoglobulin deficiencies in monozygotic twins discordant for type 1 diabetes, J. Autoimmun. 27 (2006) 89–95.
[81] S.E. Hodge, Some epistatic two-locus models of disease. I. Relative risks and identity by descent distributions in affected sib pairs, Am. J. Hum. Genet. 33 (1981) 381–395.
[82] N. Risch, Linkage strategies for genetically complex traits. I. Multilocus models, Am. J. Hum. Genet. 46 (1990) 222–228.
[83] J.P. Krischer, K.F. Lynch, D.A. Schatz, J. Ilonen, Å. Lernmark, W.A. Hagopian, M.J. Rewers, J.-X. She, O.G. Simell, J. Toppari, A.-G. Ziegler, B. Akolkar, E. BonifacioTEDDY Study Group, The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study, Diabetologia 58 (2015) 980–987.
[84] A. Katsarou, S. Gudbjörnsdottir, A. Rawshani, D. Dabelea, E. Bonifacio, B.J. Anderson, L.M. Jacobsen, D.A. Schatz, Å. Lernmark, Type 1 diabetes mellitus, Nat. Rev. Dis. Primers 3 (2017) 17016.
[85] M. Karvonen, J. Rusanen, M. Sundberg, E. Virtala, A. Colpaert, A. Naukkarinen, J. Tuomilehto, Regional differences in the incidence of insulin-dependent diabetes mellitus among children in Finland from 1987 to 1991. Childhood Diabetes in Finland (DiMe) Study Group, Ann. Med. 29 (1997) 297–304.
[86] R.A. Kittles, M. Perola, L. Peltonen, A.W. Bergen, R.A. Aragon, M. Virkkunen, M. Linolla, D. Goldman, J.C. Long, Dual origins of Finns revealed by Y chromosome haplotype variation, Am. J. Hum. Genet. 62 (1998) 1171–1179.
[87] J. Ilonen, H. Reijonen, A. Green, A. Reunanen, M. Knip, O. Simell, H.K. Åkerblom, Geographical differences within Finland in the frequency of HLA-DQ genotypes associated with type 1 diabetes susceptibility. The Childhood Diabetes in Finland Study Group, Eur. J. Immunogenet. 27 (2000) 225–230.
[88] H.X. Zhao, R.A. Moyeed, E.A. Stenhouse, A.G. Demaine, B.A. Millward, Space-time clustering of childhood type 1 diabetes in Devon and Cornwall, England, Diabet. Med. 19 (2002) 667–672.
[89] S. Leslie, B. Winney, G. Hellenthal, D. Davison, A. Boumertit, T. Day, K. Hutnik, E.C. Royrvik, B. Cunliffe, Wellcome Trust Case Control Consortium, International Multiple Sclerosis Genetics Consortium, D.J. Lawson, D. Falush, C. Freeman, M. Pirinen, S. Myers, M. Robinson, P. Donnelly, W. Bodmer, The fine-scale genetic structure of the British population, Nature 519 (2015) 309–314.
[90] S.L. Patrick, J.K. Kadohiro, S.H. Waxman, J.D. Curb, T.J. Orchard, J.S. Dorman, L.H. Kuller, R.E. LaPorte, IDDM incidence in a multiracial population. The Hawaii IDDM Registry, 1980-1990, Diabetes Care 20 (1997) 983–987.
[91] J.B. Howie, B.J. Helyer, The immunology and pathology of NZB mice, Adv. Immunol. 9 (1968) 215–266.
[92] Y. Berhan, I. Waernbaum, L. Torbjörn, A. Möllsten, G. Dahlquist, Swedish Childhood Diabetes Study Group, Thirty years of prospective nationwide incidence of childhood type 1 diabetes. The accelerating increase by time tends to level off in Sweden, Diabetes 60 (2011) 577–580.
[93] G.G. Dahlquist, L. Nyström, C. Patterson, Incidence of type 1 diabetes in Sweden among individuals aged 0-34 years, 1983-2007, Diabetes Care 34 (2011) 1754–1759.
[94] M.F. Fraga, E. Ballestar, M.F. Paz, S. Ropero, F. Setien, M.L. Ballestar, D. Heine-Suñer, J.C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T.D. Spector, Y.Z. Wu, C. Plass, M. Esteller, Epigenetic differences arise during the lifetime of monozygotic twins, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 10604–10609.
[95] T. Fournier, J. Schacherer, Genetic backgrounds and hidden trait complexity in natural populations, Curr. Opin. Genet. Dev. 47 (2017) 48–53.