Genes and Immunity (2014) **15,** 218–223

www.nature.com/gene

ORIGINAL ARTICLE

Family studies of type 1 diabetes reveal additive and epistatic effects between *MGAT1* and three other polymorphisms

Z Yu[1], CF Li[2,3], H Mkhikian[4], RW Zhou[4], BL Newton[2] and M Demetriou[2,3,4]

In a recent study on multiple sclerosis (MS), we observed additive effects and epistatic interactions between variants of four genes that converge to induce T-cell hyperactivity by altering Asn-(*N*)-linked protein glycosylation: namely, the Golgi enzyme *MGAT1*, cytotoxic T-lymphocyte antigen 4 (*CTLA-4*), interleukin-2 receptor-α (*IL2RA*) and interleukin-7 receptor-α (*IL7RA*). As the *CTLA-4*, *IL2RA* and *IL7RA* variants are associated with type 1 diabetes (T1D), we examined for joint effects in T1D. Employing a novel conditional logistic regression for family-based data sets, epistatic and additive effects were observed using 1423 multiplex families from the Type 1 Diabetes Genetic Consortium data set. The *IL2RA* and *IL7RA* variants had univariate association in MS and T1D, whereas the *MGAT1* and *CTLA-4* variants associated with only MS or T1D, respectively. However, similar to MS, the *MGAT1* variant haplotype interacted with *CTLA4* ($P = 0.03$), and a combination of *IL2RA* and *IL7RA* ($P = 0.01$). The joint effects of *MGAT1*, *CTLA4*, *IL2RA*, *IL7RA* and the two interactions using a multiple conditional logistic regression were statistically highly significant ($P < 5 \times 10^{-10}$). The *MGAT1–CTLA-4* interaction was replicated ($P = 0.01$) in 179 trio families from the Genetics of Kidneys in Diabetes study. These data are consistent with defective *N*-glycosylation of T cells contributing to T1D pathogenesis.

Genes and Immunity (2014) **15,** 218–223; doi:10.1038/gene.2014.7; published online 27 February 2014

Keywords: type 1 diabetes; *N*-glycosylation; genetic interaction; MGAT1

INTRODUCTION

With the advancement of high-throughput genotyping technologies, hundreds of common genetic variants have been identified for complex human traits, such as type 1 diabetes [T1D, MIM 222100]. However, it has been reported that these genetic variants explain only a small proportion of heritability.[1] Gene–gene interactions are likely a major factor in explaining the mystery of missing heritability,[1] and, thus, characterizing gene–gene interactions is of fundamental importance in unraveling the etiology of complex human diseases. However, successful detection of gene–gene interactions faces many challenges. A major constraint is the issue of multiple hypothesis testing. In a genome-wide search for gene–gene interactions, correcting for the very large number of tests greatly diminishes the power to detect interactions with moderate effects.

Single-gene disorders displaying Mendelian inheritance disrupt molecular pathways at a single step. However, a similar degree of pathway disruption may be obtained through small defects in the multiple genes/environmental inputs that combine to disrupt a single pathway. These interactions may be epistatic or additive and may promote disease only when combined, and are therefore poorly detected by genome-wide association studies. A functional approach that groups candidate variants on the basis of a shared ability to alter a common molecular pathway provides an alternative method to identify interactions. Indeed, we recently reported that multiple environmental factors (vitamin D_3 deficiency and metabolism) and multiple genetic variants (*IL7RA*, *IL2RA*, *MGAT1*, *MGAT5* and *CTLA-4*) converge to dysregulate Golgi *N*-glycosylation and T-cell function in multiple sclerosis (MS).[2–4] Causality of defective *N*-glycosylation in MS is supported by animal data, where genetic- and metabolic-induced alterations control T-cell growth, T_H1/T_H17 differentiation and autoimmunity, including development of a spontaneous MS-like disease in *Mgat5*-deficient PL/J mice.[5–9] In MS, epistatic interactions and additive effects were observed between the four variants and environmental factors resulting in dysregulated *N*-glycosylation. For example, a haplotype of the Golgi *N*-glycosylation enzyme *MGAT1* promotes MS, alters *N*-glycosylation, T-cell activation thresholds and surface expression of anti-autoimmune cytotoxic T-lymphocyte antigen 4 (CTLA-4) in a manner that is sensitive to metabolic conditions, Vitamin D3 signaling, the number of *N*-glycans attached to CTLA-4 (*CTLA-4*, rs231775) and interleukin-7/interleukin-2 signaling modulation by the *IL7RA* (rs6897932) and *IL2RA* (rs2104286) variants. The interaction between the *MGAT1* and *CTLA-4* variants was epistatic, as *CTLA-4* (rs231775) lacks univariate association with MS. In contrast, a non-additive interaction was observed between the *MGAT1* risk variant and a combination of the *IL7RA* and *IL2RA* risk variants, a result consistent with their opposing effects on mRNA levels of the *MGAT1* enzyme. These data suggest that studies examining only univariate association, such as genome-wide association studies, are unlikely to adequately define heritability.

Studies have shown that genetic risk factors and pathways are frequently shared across different autoimmune diseases, albeit not always in the same direction.[10–14] For example, the interleukin-2 receptor-α (*IL2RA*) gene is significantly associated with both MS and T1D;[10,11] however, the direction of the effect may be the same or opposite depending upon the specific variant examined.[11,15] Similarly, *HLA-DR15* is a risk marker for MS but is protective in T1D. These considerations, along with a common molecular target

[1]Department of Statistics, University of California, Irvine, CA, USA; [2]Department of Neurology, University of California, Irvine, CA, USA; [3]Institute for Immunology, University of California, Irvine, CA, USA and [4]Department of Microbiology & Molecular Genetics, University of California, Irvine, CA, USA. Correspondence: Dr M Demetriou, Neurology and Microbiology & Molecular Genetics, University of California, Rm 208, Sprague Hall, 839 Medical Sciences Court, Irvine, CA 92869, USA.
E-mail: mdemetri@uci.edu
Received 13 August 2013; revised 13 January 2014; accepted 17 January 2014; published online 27 February 2014

(that is, *N*-glycosylation), motivated us to hypothesize that the four MS variants we detected[2] may also interact in T1D to determine disease susceptibility. By borrowing the interaction information learned from MS, the burden of multiple testing present in a random genome-wide search is significantly reduced. The most common test for genetic association is the case–control design; however, this can be biased by population stratification. In contrast, a family-based design, such as the Type 1 Diabetes Genetics Consortium (T1DGC), provides inference of association that is robust against population stratification. A common way to analyze family data is with conditional logistic regression (CLR).[16,17] Cordell *et al.*[18] proposed the use of CLR to test genetic interaction between two variants by constructing 15 pseudo controls for each affected child. This approach is difficult to be generalized to examine multiple variants as the number of pseudo controls for each affected child grows exponentially with the number of variants. In addition, analyzing linked variants requires knowledge of recombination rates between variants. One way to avoid these complications is to match each affected child to the pseudo control whose genotype is formed by all of the other non-transmitted alleles by parents. Kotti *et al.*[19] used this matching strategy to test gene–gene interactions. We have recently shown that Kotti's matching strategy is suboptimal for testing gene–gene interactions.[20] Therefore, to test both additive and non-additive genetic effects using the multiplex family data collected by the T1DGC, we utilized an easy-to-implement yet efficient method of constructing pseudo controls. Using this method, we identified additive and non-additive effects of *MGAT1*, *CTLA4*, *IL2RA* and interleukin-7 receptor-α (*IL7RA*) on T1D risk, with an overall *P*-value $< 5 \times 10^{-10}$.

RESULTS

A novel matching strategy

In recent theoretical work[20] we examined and compared CLRs under two matching strategies: 1:1 matching and exhaustive matching. Suppose that we are interested in testing *L* loci. In 1:1 matching, each affected child is matched to its 'anti-self', a pseudo control whose genotype is formed by the non-transmitted alleles. In exhaustive matching, each affected child is matched to 4^L-1 pseudo controls. The two matching strategies at two loci for a case–parent trio are illustrated in Figure 1a. Compared with exhaustive matching, the 1:1 matching strategy is simpler, more straightforward to implement and computationally easier. Furthermore, 1:1 matching does not require knowledge of recombination rates between markers, whereas exhaustive matching does. Intuitively, 1:1 matching utilizes less information from the data. However, we found that 1:1 matching is as efficient as exhaustive matching when the true underlying genetic effects are additive, which requires that there are no intra- or inter-locus interactions.[20] Thus, when the focus is on additive genetic effects, we can safely use 1:1 matching; on the other hand, when the focus is non-additive effects, we should consider exhaustive matching.

On the basis of our prior understanding of *MGAT1* and other genetic variants altering *N*-glycosylation in MS,[2] we expected both additive effects and gene–gene interactions between variants of the following four genes: *MGAT1* (rs7726005 and rs2070924), *CTLA4* (rs231775), *IL2RA* (rs2104286) and *IL7RA* (rs6897932). At the individual gene level, our studies in MS indicate that the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype (rs7726005 and rs2070924) has a dominant effect, whereas single-nucleotide polymorphisms (SNPs) rs231775 (*CTLA4*), rs2104286 (*IL2RA*) and rs6897932 (*IL7RA*) demonstrate additive effects. Between genes, we found that the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype interacts with two sets of SNPs: rs231775 (*CTLA4*), and a combination of rs2104286 (*IL2RA*) and rs6897932 (*IL7RA*). In the following, we show how this prior information can be used to facilitate our construction of pseudo controls for each affected child in the T1DGC study.

The rs2070924 SNP in *MGAT1* is almost in complete linkage disequilibrium with rs7726005. The haplotype formed by these two rare SNPs shows a dominant effect, indicating that exhaustive

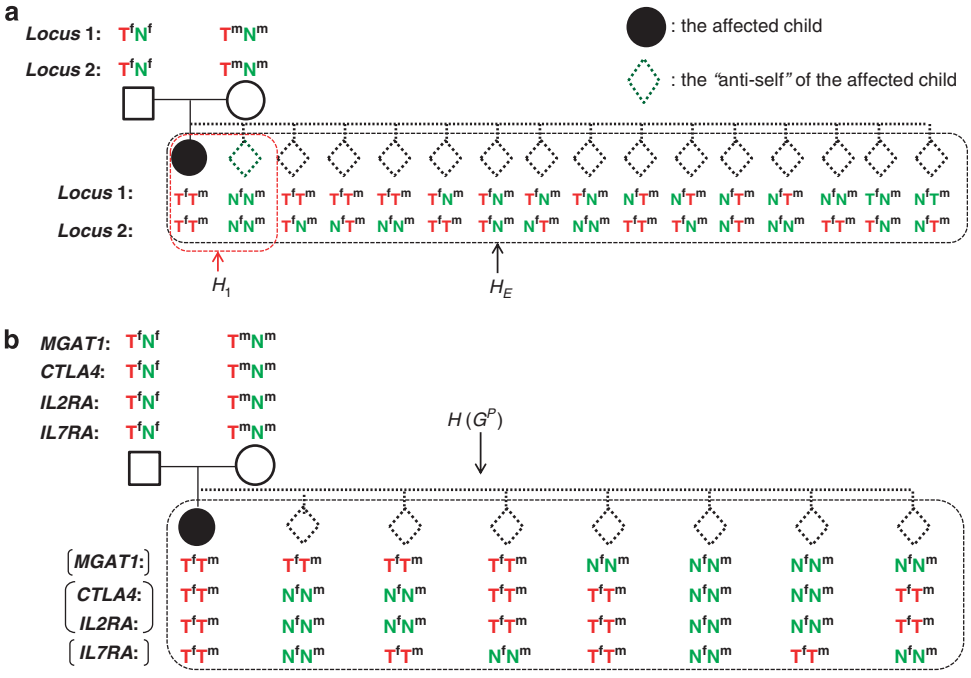

Figure 1. (**a**) The 1:1 and the exhaustive matching strategies for a case–parents trio at two loci. T^f: the allele transmitted to the child by the father. N^f: the allele not transmitted to the child by the father. T^m: the allele transmitted to the child by the mother. N^m: the allele not transmitted to the child by the mother. H_1: all possible offspring genotypes given the couple's genotypes under the 1:1 matching. H_E: all possible offspring genotypes given the couple's genotypes under the exhaustive matching. Under the null hypothesis of no association, the two genotypes in H_1 are equally likely; under the null hypothesis of no association and linkage equilibrium between the two SNPs, the 16 genotypes in H_E are equally likely. (**b**) The matching strategy used to analyze the four SNPs in this study.

Table 1. Individual genetic effects estimated from the T1DGC data

	Alleles	*Freq*	*GRR (95% CI)*	P-*value*
MGAT1 $IV_A/V_{T\text{-}T}$ (rs7726005, rs2070924)	—	0.039	1.06 (0.87–1.29)	0.58
rs231775 (*CTLA-4*)	A, G	G: 0.414	1.17 (1.08–1.26)	5.3×10^{-5}
rs2104286 (*IL2RA*)	A, G	A: 0.775	1.25 (1.14–1.36)	6.9×10^{-7}
rs6897932 (*IL7RA*)	C, T	C: 0.751	1.12 (1.02–1.22)	1.3×10^{-2}

Abbreviations: CI, confidence interval; GRR, genotype relative risk; T1DGC, type 1 diabetes genetics consortium.

matching would be more efficient than 1:1 matching; however, the frequency of the haplotype is rare, and in this case a dominant model is close to an additive model. To reduce the complexity of matching, we used 1:1 matching at *MGAT1*. Both rs2104286 (*IL2RA*) and rs6897932 (*IL7RA*) show additive individual effects in our previous MS study;[2] therefore, we used 1:1 matching at each of the two genes. As there was no evidence of genetic interaction between rs231775 (*CTLA4*) and rs2104286 (*IL2RA*), we assumed that the alleles at rs231775 (*CTLA4*) and rs2104286 (*IL2RA*) are co-transmitted from parents to offspring. In addition, there was no evidence of genetic interaction between rs231775 (*CTLA4*) and rs6897932 (*IL7RA*). Thus, we could also let rs231775 (*CTLA4*) co-transmit with rs6897932 (*IL7RA*) when constructing pseudo controls. Indeed, the two methods presented identical results. Finally, because we wanted to test gene–gene interactions, we considered exhaustive matching among the three groups (*MGAT1*), (*IL2RA*, *CTLA4*) and (*IL7RA*), leading to 1:7 matching. It should be noted that the genes are in linkage equilibrium. Thus, under the null hypothesis, the eight possible offspring genotypes, including that of an affected child and his/her seven matched pseudo controls, are equally likely. We used $H(G^P)$ to denote the eight genotypes given the parental genotype G^P. The final matching strategy to identify both additive and non-additive multi-locus genetic effects of these genes is summarized in Figure 1b.

CLRs

Matched case–control data are often analyzed by CLRs. Let G_i^O be the genotype of the *i*th child among *n* total affected children and G_i^P be the genotype of the parents of the *i*th affected child. Using the matching strategy we described above, the likelihood function of association parameters is

$$L\left(\vec{\beta}\right) = \prod_{i=1}^{n} \frac{\exp\left(\vec{\beta}' G_i^O\right)}{\sum_{\{G_i^* : G_i^* \in H(G_i^P)\}} \exp\left(\vec{\beta}' G_i^*\right)}$$

The form of $\vec{\beta}$ and G_i^O depends upon our model. For example, when testing the association of the dominant *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype in T1D, we assigned G_i^O to 1 if the offspring has at least one copy of the haplotype, and 0 otherwise. Correspondingly, $\vec{\beta}$ is the log of the genotype relative risk (GRR) for carriers of the haplotype to those non-carriers. As another example, when testing the interaction between the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and rs231775 (*CTLA4*), G_i^O is a vector of numerical values with three elements indicating the presence of the *MGAT1* IV/V haplotype, the number of copies of the G allele (that is, the risk allele) of rs231775 (*CTLA4*) and the product of the first two numbers. Correspondingly, $\vec{\beta}$ is a vector of coefficients corresponding to the main effect of the *MGAT1* IV/V haplotype, the main effect of rs231775 (*CTLA4*) and the interaction of the two variants, respectively.

We characterized the significance of additive and non-additive effects utilizing *P*-values. *P*-values of individual terms in a multiple CLR were calculated using the Wald test. When examining the joint effect of multiple terms, we use the likelihood ratio test. CLRs were fitted using the 'clogit' function in the survival package in R (http://cran.r-project.org/package=survival).

Table 2. Stratified analysis of *MGAT1* $IV_A/V_{T\text{-}T}$ using the T1DGC data based on *CTLA4* genotypes

CTLA-4 genotypes	*GRR (95% CI)*	P-*value*
AA	1.53 (1.09–2.15)	1.4×10^{-2}
AG	1.00 (0.75–1.33)	1.0
GG	0.58 (0.36–0.95)	2.8×10^{-2}

Abbreviations: CI, confidence interval; GRR, genotype relative risk; T1DGC, Type 1 Diabetes Genetics Consortium. The *P*-value for interaction is 2.9×10^{-2}.

Gene–gene interactions and joint effects in the T1DGC study

We first examined the individual effects of the four variants. We used the risk alleles (column 3 of Table 1) as the test alleles and the protective alleles as the reference alleles. Many groups have reported association between T1D and *CTLA4*, *IL2RA* and *IL7RA*.[21,22] All variants are significantly associated with T1D except the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype (Table 1). This differs from MS, where all variants were associated, except *CTLA-4*.

Motivated by the interactions between *MGAT1* and *CTLA4*, *IL2RA* and *IL7RA* for MS susceptibility, we tested their genetic interactions for T1D susceptibility. Although the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype does not show univariate association with T1D, it is a protective, neutral and risk allele for AA, AG and GG *CTLA-4* genotypes, respectively (Table 2). For subjects with the AG genotype at rs231775 (*CTLA-4*), the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype shows no association with T1D. For subjects with the AA genotype, that is, the low-risk group based on *CTLA-4*, the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype leads to an increased risk for T1D with a GRR of 1.53 (*P*-value = 0.014). For subjects with the GG genotype, that is, the high-risk group based on *CTLA-4*, the *MGAT1* $V_A/V_{T\text{-}T}$ haplotype has a protective role for T1D with a GRR of 0.58 (*P*-value = 0.028). The different effects of the *MGAT1* $V_A/V_{T\text{-}T}$ haplotype under the three CTLA-4 genotypes suggest a gene–gene interaction between the two variants. The *P*-value for interaction is 0.029. Stratified point estimates of GRRs, 95% confidence intervals and *P*-values can be found in Table 2.

In MS, the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype also interacts with a combination of the *IL2RA* and *IL7RA* risk alleles.[2] Our CLR indicates that the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype also shows differential effects on T1D susceptibility between subjects with four risk alleles of *IL2RA* and *IL7RA* versus other subjects (Table 3). The *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype increases T1D risk for subjects at high risk based on *IL2RA* and *IL7RA*. It has a protective risk in the rest of the population, as can be seen from the point estimates of the GRRs in Table 3. Testing the interaction between them leads to a *P*-value of 0.013.

As the variants we considered here are in linkage equilibrium, the interactions we observed cannot be explained by each other.

Table 3. Stratified analysis of *MGAT1* $IV_A/V_{T\text{-}T}$ using the T1DGC data based on the number of risk alleles in *IL2RA* and *IL7RA*

IL[a]	*GRR (95% CI)*	P-*value*
4	1.30 (0.95–1.77)	0.10
<4	0.90 (0.69–1.16)	0.40

Abbreviations: CI, confidence interval; GRR, genotype relative risk; IL, interleukin; T1DGC, Type 1 Diabetes Genetics Consortium. The *P*-value for interaction is 1.3×10^{-2}. [a]IL is the number of risk alleles in *IL2RA* and *IL7RA*.

Table 4. Multivariate analysis of the four genes, including main effects and two gene–gene interactions, using the T1DGC data

Variable	*GRR (95% CI)*	P-*value*
MGAT1 $IV_A/V_{T\text{-}T}$	1.08 (0.80–1.46)	0.63
CTLA-4	1.19 (1.10–1.28)	2.5×10^{-5}
IL2RA	1.21 (1.11–1.33)	2.1×10^{-5}
IL7RA	1.10 (1.00–1.20)	4.0×10^{-2}
MGAT1 $IV_A/V_{T\text{-}T}$ x *CTLA-4*	1.58 (1.14–2.18)	6.0×10^{-3}
MGAT1 $V_A/V_{T\text{-}T}$ x (*IL* = 4)	0.76 (0.60–0.97)	2.7×10^{-3}

Abbreviation: GRR, genotype relative risk; T1DGC, Type 1 Diabetes Genetics Consortium. The *P*-value (LRT) for the two interaction terms is 2.0×10^{-3}. The *P*-value (LRT) for the joint effect, additive and non-additive effects is 3.9×10^{-10}.

To confirm this and evaluate the overall impact of the variants, we fit a multiple CLR with the four variants and the two interactions. Table 4 indicates that the point estimates of the GRRs and *P*-values for the four variants in the multiple CLR are similar to those from individual CLRs.

The two interaction terms are significant in the multiple CLR, indicating they are still important after accounting for the main effects, consistent with what we observed in Tables 2 and 3. Therefore, we used a likelihood ratio test of two degrees of freedom to examine the joint effects of the two interaction terms. The *P*-value based on the likelihood ratio test is 0.002. Finally, we used a likelihood ratio test of seven degrees of freedom to test the joint effect of both main and interaction effects. The *P*-value is $<5 \times 10^{-10}$.

Replication analysis in the GoKinD study

The *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype is not available in public genome-wide association study data sets. Therefore, to attempt replicating the results above, we genotyped all four variants in 379 trios from the Genetics of Kidneys in Diabetes study (GoKinD). The results show that the only significant individual term is *IL7RA* (*P*-value 0.016). This is not surprising. First, the sample size is much smaller than that of the T1DGC study. Second, the T1D offspring in the GoKinD study have either clear-cut kidney disease or normal renal status despite long-term diabetes.[23] Thus, the combination of the two cohorts does not fully represent the T1D population. It has been estimated that about 30% of T1D patients have kidney disease.[23] We therefore hypothesized that those T1D patients without kidney disease are more similar to the general T1D population and conducted a separate analysis of the two cohorts. Because of the small sample size and the low frequency of the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype, we grouped the genotypes AG and GG at rs231775 (*CTLA-4*) when analyzing the interaction between the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and *CTLA-4*.

Our results reveal that the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype has a significant interaction with *CTLA-4* (*P*-value 0.012) in the cohort free from kidney disease, with the direction of interaction consistent with the one we found in the T1DGC study (Table 5).

Table 5. Stratified analysis of *MGAT1* $IV_A/V_{T\text{-}T}$ using the GoKinD data based on *CTLA4* genotypes

CTLA-4 genotypes	*GRR of MGAT1* $IV_A/V_{T\text{-}T}$			
	Kidney patients[a]		*Kidney controls*	
	GRR (95% CI)	P-*value*	*GRR (95% CI)*	P-*value*
AA	1.75 (0.51–5.98)	0.37	0.83 (0.25–2.73	0.76
AG/GG	0.40 (0.08–2.06)	0.27	0.78 (0.29–2.09)	0.62

Abbreviations: CI, confidence interval; GRR, genotype relative risk; GoKinD, Genetics of Kidneys in Diabetes study. [a]The *P*-value for interaction is 1.2×10^{-2}.

Specifically, in both the T1DGC study and the GoKinD trios free from kidney disease, the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype increases the risk of T1D in patients with the AA genotype at *CTLA-4* and has a protective role in patients with the AG or GG genotype. The interaction is not significant in the GoKinD trios with kidney disease and it is estimated that the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype has a similar role in the AA and AG/GG groups (Table 5).

We also tested the interaction between the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and the *IL2RA* and *IL7RA* polymorphisms. No interaction was identified for either the separate or combined analysis of the GoKinD families. This is not surprising, given the lower effect size of the interaction and reduced power relative to the T1DGC cohort.

DISCUSSION

In this article, we presented two potential gene–gene interactions involved in regulating *N*-glycosylation and possibly T1D susceptibility. Our analysis is knowledge-driven and is motivated by the fact that gene–gene interactions were observed in MS. It is known that both MS and T1D are autoimmune diseases and they share many common pathways. Different from a genome-wide search for gene–gene interactions, our analysis avoids multiple testing and thus potentially improves power. We have observed gene–gene interactions between the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and *CTLA-4* in both the T1DGC study and the GoKinD study. This provides encouraging evidence that the observed interaction is likely to be true. The interaction between the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and *IL2RA* and *IL7RA* is observed in the T1DGC families but not in the GoKinD families. As the T1DGC study is larger than the GoKinD and identifying gene–gene interaction requires large sample sizes, it is likely that the interaction is true but the GoKinD study is not powerful enough to detect it. Independent studies are required to further examine how *MGAT1*, *IL2RA* and *IL7RA* jointly affect T1D susceptibility.

There were several important differences between the results in MS and T1D. Univariate association of the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype was observed in MS but not in T1D, whereas the G allele of *CTLA-4* (rs231775) associated with T1D but not with MS. Yet, in both diseases epistatic interaction was observed between the two variants. The *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype enhanced the risk of MS in combination with the GG and AG genotypes of *CTLA-4* (rs231775), whereas a protective interaction with the GG genotype was present in T1D. These differences are not surprising as many confirmed loci of several autoimmune diseases often show opposite directions.[11,13] Such differences may provide important clues and new tools for understanding the pathogenic process of complex traits. In this case, the observed differences may arise from variances in metabolism between T1D and MS, coupled with the molecular mechanisms by which the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and *CTLA-4* (rs231775) alter *N*-glycosylation and cell

surface expression of the CTLA-4 protein in T cells. The *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype is a gain of function that increases mRNA and protein levels of the Golgi enzyme Mgat1. When metabolism limits substrate availability (that is, UDP-GlcNAc derived from glucose) to the Golgi, the *MGAT1* $IV_A/V_{T\text{-}T}$ gain-of-function haplotype paradoxically lowers *N*-glycan branching by limiting UDP-GlcNAc availability to downstream Golgi enzymes, resulting in reduced cell surface expression of the anti-autoimmune CTLA-4 protein. The G allele of *CTLA-4* (rs231775) decreases the number of *N*-glycans attached to CTLA-4 by 50%, thereby also reducing surface expression of the CTLA-4 protein. Thus, when metabolism limits Golgi substrate (UDP-GlcNAc) availability, the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype and the G allele of *CTLA-4* (rs231775) combine to lower the CTLA-4 cell surface expression,[2] consistent with the genetic interaction observed in MS. In contrast, when metabolism increases Golgi UDP-GlcNAc substrate supply, as occurs with high glucose levels,[7] the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype has the opposite effect on *N*-glycan branching and CTLA-4 surface expression. Thus, under high glucose levels often present in early stages of T1D, the *MGAT1* $IV_A/V_{T\text{-}T}$ haplotype is expected to counteract the G allele of *CTLA-4* (rs231775) to promote CTLA-4 protein expression and inhibit T-cell function, consistent with the protective genetic interaction observed in T1D.

Most existing multi-loci methods for family data only provide an overall significance of multiple loci or specific combinations of alleles. Examples of such methods include haplotype-based methods,[24–30] genotype-based methods,[30–34] family-based multiple dimension reduction[35] and contrasting linkage disequilibrium.[36] In comparison, the method we utilized here has two advantages. First, most of the existing multi-locus methods are based upon 1:1 matching, which is not efficient for testing non-additive effects. Second, these methods are mainly for hypothesis testing. In contrast, our method not only provides the significance level of effects but also provides point estimates of GRRs.

PATIENTS AND METHODS

Data sets

We analyzed Caucasian multiplex or trio families in the T1DGC and the GoKinD.[23] The SNPs were genotyped using a method previously described.[2] We excluded families with missing parents or genotyping errors at any of the four SNPs. For T1DGC, this led to 2858 affected offspring and their parents from 1423 multiplex families. The original goal of the GoKinD study was to identify genetic factors associated with kidney disease within T1D patients. Here, we used these families to examine the four genes. The GoKinD analysis is based upon 379 trio families with both parents genotyped, divided into two cohorts: (1) 200 T1D patients with kidney disease and their parents and (2) 179 T1D patients without kidney disease and their parents. Other information, such as genotyping, can be found in a report from the consortium.[23] To test genetic effects, we used CLR with a novel matching strategy, as described in the following section.

Statistics

To test genetic effects, CLR with a novel matching strategy was used and is described in detail in the results section. The significance of additive and non-additive effects was characterized by *P*-values. *P*-values of individual terms in a multiple CLR are calculated using the Wald test. When examining the joint effect of multiple terms, we used the likelihood ratio test. CLRs were fitted using the 'clogit' function in the survival package in R (http://cran.r-project.org/package=survival).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the National Institute of Diabetes and Digestive and Kidney Diseases for providing access to the T1DGC and GoKinD DNA samples. Research was supported in part by grant R01HG004960 from the National Human Genome Research Institute to ZY and grant R01AI082266 from the National Institute of Allergy and Infectious Diseases to MD.

REFERENCES

1 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
2 Mkhikian H, Grigorian A, Li CF, Chen HL, Newton B, Zhou RW *et al.* Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nat Commun* 2011; **2**: 334.
3 Li CF, Zhou RW, Mkhikian H, Newton BL, Yu Z, Demetriou M. Hypomorphic MGAT5 polymorphisms promote multiple sclerosis cooperatively with MGAT1 and interleukin-2 and 7 receptor variants. *J Neuroimmunol* 2013; **256**: 71–76.
4 Grigorian A, Mkhikian H, Li CF, Newton BL, Zhou RW, Demetriou M. Pathogenesis of multiple sclerosis via environmental and genetic dysregulation of *N*-glycosylation. *Semin Immunopathol* 2012; **34**: 415–424.
5 Demetriou M, Granovsky M, Quaggin S, Dennis JW. Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* 2001; **409**: 733–739.
6 Lee SU, Grigorian A, Pawling J, Chen IJ, Gao G, Mozaffar T *et al.* N-glycan processing deficiency promotes spontaneous inflammatory demyelination and neurodegeneration. *J Biol Chem* 2007; **282**: 33725–33734.
7 Grigorian A, Lee SU, Tian W, Chen IJ, Gao G, Mendelsohn R *et al.* Control of T Cell-mediated autoimmunity by metabolite flux to N-glycan biosynthesis. *J Biol Chem* 2007; **282**: 20027–20035.
8 Grigorian A, Araujo L, Naidu NN, Place D, Choudhury B, Demetriou M. *N*-acetylglucosamine inhibits T-helper 1 (Th1) / T-helper 17 (Th17) responses and treats experimental autoimmune encephalomyelitis. *J Biol Chem* 2011; **286**: 40133–40141.
9 Grigorian A, Mkhikian H, Demetriou M. Interleukin-2, Interleukin-7, T cell-mediated autoimmunity, and N-glycosylation. *Ann NY Acad Sci* 2012; **1253**: 49–57.
10 Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* 2009; **10**: 43–55.
11 Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011; **7**: e1002254.
12 Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JHM *et al.* Shared and distinct genetic variants in type 1 diabetes and celiac disease. *New Engl J Med* 2008; **359**: 2767–2777.
13 Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with snp alleles. *PLoS Genet* 2009; **5**: e1000792.
14 Todd JA. Etiology of type 1 diabetes. *Immunity* 2010; **32**: 457–467.
15 Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, Severson C *et al.* IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 2009; **5**: e1000322.
16 Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/disease association with application to a study of aplastic anemia. *Biometrics* 1991; **47**: 53–61.
17 Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 1993; **53**: 1114–1126.
18 Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002; **70**: 124–141.
19 Kotti S, Bickeboller H, Clerget-Darpoux F. Strategy for detecting susceptibility genes with weak or no marginal effect. *Hum Hered* 2007; **63**: 85–92.
20 Yu ZX, Deng L. Pseudosibship methods in the case-parents design. *Stat Med* 2011; **30**: 3236–3251.
21 Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G *et al.* Genetics of type 1 diabetes: what's next? *Diabetes* 2010; **59**: 1561–1571.
22 Santiago JL, Alizadeh BZ, Martinez A, Espino L, de la Calle H, Fernandez-Arquero M *et al.* Study of the association between the CAPSL-IL7R locus and type 1 diabetes. *Diabetologia* 2008; **51**: 1653–1658.
23 Mueller PW, Rogus JJ, Cleary PA, Zhao Y, Smiles AM, Steffes MW *et al.* Genetics of Kidneys in Diabetes (GoKinD) study: a genetics collection available for identifying genetic susceptibility factors for diabetic nephropathy in type 1 diabetes. *J Am Soc Nephrol* 2006; **17**: 1782–1790.
24 Zhao H, Zhang S, Merikangas KR, Trixler M, Wildenauer DB, Sun F *et al.* Transmission/disequilibrium tests using multiple tightly linked markers. *Am J Hum Genet* 2000; **67**: 936–946.
25 Allen AS, Satten GA. Statistical models for haplotype sharing in case-parent trio data. *Hum Hered* 2007; **64**: 35–44.
26 Zhang S, Sha Q, Chen HS, Dong J, Jiang R. Transmission/disequilibrium test based on haplotype sharing for tightly linked markers. *Am J Hum Genet* 2003; **73**: 566–579.

27 Van der Meulen MA, te Meerman GJ. Haplotype sharing analysis in affected individuals from nuclear families with at least one affected offspring. *Genet Epidemiol* 1997; **14**: 915–920.
28 Dudbridge F, Koeleman BP, Todd JA, Clayton DG. Unbiased application of the transmission/disequilibrium test to multilocus haplotypes. *Am J Hum Genet* 2000; **66**: 2009–2012.
29 Knapp M, Becker T. Family-based association analysis with tightly linked markers. *Hum Hered* 2003; **56**: 2–9.
30 Fan R, Knapp M, Wjst M, Zhao C, Xiong M. High resolution T2 association tests of complex diseases based on family data. *Ann Hum Genet* 2005; **69**: 187–208.
31 Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: A class of tests and the determinants of statistical power. *Hum Hered* 2003; **56**: 18–31.
32 Zhang Z, Zhang S, Sha Q. A multi-marker test based on family data in genome-wide association study. *BMC Genet* 2007; **8**: 65.
33 Shi M, Umbach DM, Weinberg CR. Identification of risk-related haplotypes with the use of multiple SNPs from nuclear families. *Am J Hum Genet* 2007; **81**: 53–66.
34 Lee WC. Testing for candidate gene linkage disequilibrium using a dense array of single nucleotide polymorphisms in case-parents studies. *Epidemiology* 2002; **13**: 545–551.
35 Cattaert T, Urrea V, Naj AC, De Lobel L, De Wit V, Fu M *et al*. FAM-MDR: a flexible family-based multifactor dimensionality reduction technique to detect epistasis using related individuals. *PLoS One* 2010; **5**: e10304.
36 Yu Z, Wang S. Contrasting linkage disequilibrium as a multilocus family-based association test. *Genet Epidemiol* 2011; **35**: 487–498.