Research Article

The Missing Heritability in T1D and Potential New Targets for Prevention

Brian G. Pierce,[1] Ryan Eberwine,[2] Janelle A. Noble,[3] Michael Habib,[4] Hennady P. Shulha,[1] Zhiping Weng,[1] Elizabeth P. Blankenhorn,[2] and John P. Mordes[4]

[1] *Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA*

[2] *Department of Microbiology and Immunology, Center for Immunogenetics and Inflammatory Diseases, Drexel University College of Medicine, Philadelphia, PA 19129, USA*

[3] *Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA*

[4] *Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA*

Correspondence should be addressed to John P. Mordes; john.mordes@umassmed.edu

Type 1 diabetes (T1D) is a T cell-mediated disease. It is strongly associated with susceptibility haplotypes within the major histocompatibility complex, but this association accounts for an estimated 50% of susceptibility. Other studies have identified as many as 50 additional susceptibility loci, but the effect of most is very modest (odds ratio (OR) <1.5). What accounts for the "missing heritability" is unknown and is often attributed to environmental factors. Here we review new data on the cognate ligand of MHC molecules, the T cell receptor (TCR). In rats, we found that one allele of a TCR variable gene, Vβ13A, is strongly associated with T1D (OR >5) and that deletion of Vβ13+ T cells prevents diabetes. A role for the TCR is also suspected in NOD mice, but TCR regions have not been associated with human T1D. To investigate this disparity, we tested the hypothesis *in silico* that previous studies of human T1D genetics were underpowered to detect MHC-contingent TCR susceptibility. We show that stratifying by MHC markedly increases statistical power to detect potential TCR susceptibility alleles. We suggest that the TCR regions are viable candidates for T1D susceptibility genes, could account for "missing heritability," and could be targets for prevention.

1. Introduction

Type 1 diabetes (T1D) is a T cell-mediated autoimmune disease that afflicts a million persons in the USA [1, 2]. It is a polygenic disorder resulting from the interaction of multiple gene variants [3] and environmental factors [4]. No approved methods are currently available for its prevention or reversal [5]. Most interventions targeted at curing human T1D have focused on either "secondary" or "tertiary" prevention, that is, treating individuals who either have the disease or are at risk based on family history and autoantibody titers [5]. To date, no intervention has achieved the degree of success required for clinical adaptation [6].

New strategies for primary prevention in susceptible individuals would be advantageous, attacking the problem before it starts or at its earliest stages [7]. Primary prevention,

however, requires accurate predictive genetic tools. Treatment of individuals who would have remained diabetes-free poses serious pragmatic and ethical issues.

The major genetic loci for diabetes susceptibility are within the human leukocyte antigen (HLA) region, specifically those encoding HLA-DR and DQ antigens, with a less significant independent contribution from HLA class I genes [8–10]. Several high-risk HLA class II haplotypes account for ~40% of the predisposition to T1D, with an odds ratio of ~6.8, but accounting for the remaining 60% is an unresolved problem [3]. The insulin genes VNTR, *PTPN22*, and *CD25* are associated with odds ratios >1.5, and rare alleles of *IFIH1* have an odds ratio (OR) near 0.5 [11]. More than 40 non-MHC genes/regions, most involved in immune responses, have statistically significant associations, but with OR <1.5 [11, 12].

Unfortunately, although low-resolution HLA-genotyping will identify most individuals at risk for T1D, only 1/15 or ~7% of individuals with one of the highest risk HLA genotypes (known as "DR3/DR4") will actually become diabetic [13]. Additional genetic knowledge has not yet significantly improved prediction. An early estimate of sibling relative risk (λs) for T1D was estimated to be quite high at 15 [14]. Predictions of T1D, incorporating both HLA and all currently known loci, generate a λs of only 5 [3, 15], although a recently reported strategy based on combining multiple risk alleles appears to hold promise [16].

2. The TCR and "Missing Heritability"

A possible explanation for our inability to predict T1D accurately based on genotyping is the effect of environmental perturbants [21]. There is clear seasonal and geographic variation in the onset of T1D [22], and miniepidemics of the disease have been documented [23]. Viral infection is thought to be the most likely perturbant, and it remains a topic of intense investigation [24]. Although there is no good evidence for direct infection of pancreatic beta cells, the immune response to infection might easily provoke disease onset in genetically predisposed individuals. Most of the genes and loci identified by genome-wide association study (GWAS) analyses of T1D are involved in immune responses [11], and the interaction of random infection with such genes ("environmental genetics") is a plausible way to account for the "missing heritability."

We would like to suggest, however, that there may be an overlooked genetic element that has not been detected for technical reasons, specifically the genome-encoded parts of the T cell receptor (TCR). The TCR is the cognate partner of major histocompatibility complex (MHC) molecules in the peptide-MHC (pMHC) unit (Figure 1), and T1D is clearly a T cell-mediated disease. Nonetheless, there is very little evidence that germline TCR haplotype is important in susceptibility to T1D. There are, however, linkages to TCR in human autoimmune diseases other than T1D. TCR genotype has been implicated in multiple sclerosis (MS) [25]. There are also well-documented associations of TCR genotype with other forms of autoimmunity including Sjogren's syndrome [26, 27] and narcolepsy [28–30], which has a *TCRA* bias.

Of course, because approximately 10^{15} V-(D)-J recombined TCRs are possible, it is not surprising that a role for germline-encoded TCR usage, with far less diversity than the recombined genes, has met with skepticism. New data, however, suggest that the genome-encoded TCR is likely to play a critical, previously unrecognized role in the pathogenesis of T1D. Here we review our data that point to a role for genome-encoded TCR susceptibility to T1D and then present new quantitative analyses that attempt to account for the failure of previous studies to detect such an effect.

3. Evidence from the Rat

3.1. Gene Mapping. Type 1-like autoimmune diabetes, both spontaneous and inducible, is relatively common among
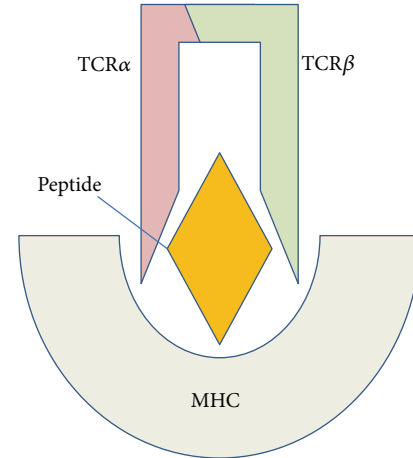

FIGURE 1: The trimolecular TCR-pMHC complex that is fundamental in T1D susceptibility.

TABLE 1: Type 1 diabetes frequency in rats as a function of MHC and TCR genotypes.

TCR	MHC	Diabetes	Strains
Vβ13a present	*RT1B/Du*	High susceptibility to diabetes	BBDP and BBDR LEW.1WR1 and LEW.1AR1-iddm KDP PVG.*RT1u*
Vβ13a absent	*RT1B/Du*	Low penetrance of T1D	WF
Vβ13a present	Non-*RT1B/Du*	No T1D	Many

Coassociation of Class II MHC haplotype and TCR usage in T1D in rats [17–19].

inbred rat strains that, like humans, express a high-risk class II MHC haplotype. In rats, this is designated *RT1B/Du* [19, 31, 32]. We have previously reported that *Iddm14* (formerly designated *Iddm4*) is a dominant non-MHC susceptibility locus important for both spontaneous and induced autoimmune diabetes in multiple rat strains [18, 33–38].

Studies of *Iddm14* in eight *RT1B/Du* rat strains led to the identification of a susceptibility haplotype in the *Tcrb-V* locus [18]. Sequencing and single nucleotide polymorphism (SNP) haplotype mapping revealed that 6 rat strains susceptible to diabetes (KDP, BBDR, BBDP, LEW.1WR1, LEW.1AR1-iddm, and PVG-RT1u) all share one allele of the beta chain variable region gene *Tcrb-V13* (designated *Tcrb-V13S1A1*) [20]. Three rat strains that are resistant to, or confer resistance to, diabetes in genetic studies, all express different alleles, either *Tcrb-V13S1A2* in the case of BN and WF rats, or *Tcrb-V13S1A3P* in the F344 rat [20]. These polymorphisms are of interest because the *Tcrb-V13S1A1* gene product, designated Vβ13a, is used more by CD4+ than CD8+ cells [20]. Taken in the context of additional data available from studies of rat T1D, our findings suggest that it is the combination of MHC and TCR that in large measure determines susceptibility to T1D in the rat. As summarized in Table 1, only those rats that express

both RT1B/Du and Vβ13a are highly susceptible to T1D. In the absence of Vβ13a, rats with a high-risk MHC are relatively resistant to T1D, and in the absence of RT1B/Du essentially no rats develop autoimmune diabetes.

These genetic observations, consistent with a critical role for germline TCR usage in T1D in the rat, led us to hypothesize that allele-specific TCR targeting could substantially prevent disease. This hypothesis was confirmed in multiple model systems, described below.

3.2. Depletion of Vβ13+ T Cells Prevents Poly I:C-Triggered T1D. LEW.1WR1 rats have a normal immunophenotype and develop T1D spontaneously at a rate of 2.5% and after treatment with polyinosinic: polycytidylic acid (poly I:C, a TLR3 and IFIH1 ligand) at a rate of 90–100% [39]. After documenting that anti-Vβ13 monoclonal antibody (mAb) reduces the number of Vβ13+ T cells *in vivo* by about 60%, we compared diabetes frequency in rats treated with either anti-Vβ13 mAb or control mouse anti-human OKT8. A second trial compared diabetes frequency in rats treated with anti-Vβ13 mAb, depleting anti-Vβ16 mAb, or vehicle. Diabetes frequency in rats treated with poly I:C and anti-Vβ13 mAb was 10% (2/20). In contrast, diabetes frequency in controls averaged 85% (34/40, $P < 0.001$) [17]. Histologic study showed significantly less insulitis and nearly complete preservation of beta cell insulin in animals treated with depleting anti-Vβ13 [17].

3.3. Depletion of Vβ13+ T Cells Prevents Virus-Triggered T1D. We also tested a model of triggered diabetes induced by viral infection. Rats were given a small priming dose of poly I:C followed by infection with Kilham rat virus (KRV). The priming dose of poly I:C is by itself nondiabetogenic but increases the penetrance of virus-triggered diabetes from ~40% to ~100% [40]. Diabetes frequency in anti-Vβ13-treated rats was 30% (3/10) as compared with 80% (8/10, $P = 0.03$) in both anti-Vβ16 mAb treated animals and untreated controls.

3.4. Depletion of Vβ13+ T Cells Prevents Spontaneous T1D. BBDP rats develop spontaneous T1D at a rate of 60–90% [31]. We treated cohorts of BBDP rats with vehicle, anti-Vβ13 mAb, or anti-Vβ16 mAb. Treatment with anti-Vβ13 mAb through 100 days of age completely prevented diabetes, whereas diabetes occurred in vehicle injected and anti-Vβ16 mAb treated rats at rates of 40% and 70%, respectively ($P < 0.01$). Among rats still nondiabetic at the end of the experiment, there was substantial "simmering" insulitis in rats treated with anti-Vβ16 mAb or vehicle [40].

These prevention studies were supplemented by additional immunological data showing a critical role for Vβ13+ T cells early in T1D pathogenesis.

3.5. CD4+Vβ13+ T Cells Are Abundant in Islets Early in the Disease Process. The animal models of "triggered" T1D that we use have well-defined kinetics and relatively rapid onset. This allows us to harvest islets from animals very early during disease onset and study the infiltrating inflammatory cells. By day 5 CD4+Vβ13+ T cells are remarkably abundant in the prediabetic islet [40], reaching a peak on day 10, when overt diabetes is first detectable.

3.6. Vβ13/Jβ mRNA Transcripts in Prediabetic Islets Are Skewed. Upon cloning the Vβ13+ transcripts from early prediabetic islets, we observed significant skewing of the TCRβ repertoire, with pauciclonal expansion of Vβ13-CDR3 sequences from islet T-cells compared to a high diversity of Vβ13-CDR3s in spleen. These data suggested that antigen-specific expansion of Vβ13+ T cells occurs in the islets of prediabetic rats. We also observed skewed TCR-Jβ usage in islet-infiltrating Vβ13+ T cells, with overrepresentation of Jβ1.3 and underrepresentation of Jβ2.1 relative to peripheral T cells. Spleen Vβ13+ T cells from poly I:C treated and untreated rats display skewing of individual Jβ segments. In addition, the representation of different Jβ segments in Vβ16+ T cell transcripts was not skewed in the islets or in the periphery. These results strongly support a role for Vβ13+ T cells in the early recognition of antigen in islets. In addition, we showed that the TCR-Vα5 repertoire is skewed among islet homing, sorted Vβ13+ T cells. This is exciting because TCR-Vα5D-4 is frequently used in the mouse T-cell response to islet antigen and recognizes insulin B:9-23 [41, 42]. Collectively, these data indicate that an oligoclonal Vβ13 response to pancreatic beta cells exists early in progression to autoimmune diabetes.

4. Evidence from the NOD Mouse

In the NOD mouse, Abiru et al. have observed a dramatic TCR α chain restriction (predominantly Vα5) in the recognition of insulin autoantigen [44]. Retrogenic NOD strains expressing Vα5D-4 α chains with many different CDR3 sequences show that even those derived from TCRs recognizing islet-irrelevant molecules develop anti-insulin autoimmunity [41]. The germline encoded Vα5D-4 T cell receptor targets a primary insulin peptide in NOD mice [41]. In addition, induction of insulin autoantibody production by helper T cells bearing Vα5D-4 α chains can be abrogated by the mutation of two amino acid residues in CDR1 and CDR2 sequences of *TRAV5D-4*. *TRAV13-1*, the human ortholog of murine *TRAV5D-4*, was also capable of inducing *in vivo* anti-insulin autoimmunity in the NOD mouse [41].

Nevertheless, the *Vα* locus has never been detected in mouse linkage studies to discover T1D genes. TCR-Vα5D-4 is polymorphic in mice, with 5 identified alleles differing by only 2–4 amino acids, but it also belongs to a family of paralogous genes with >92% homology (International ImMunoGeneTics information system or IMGT, http://www.imgt.org/). There is a paralog in B6 that has the same CDR1 and CDR2 as the important NOD paralog, *VA5D4*04. There are two non-CDR amino acid differences in this B6 paralog, which by their position are unlikely to contribute to the trimolecular complex [45]. This means that both NOD and B6 strains could transmit an effective diabetogenic allele of a TCR-Vα5D-4, and thus there would be no apparent linkage to the *Vα* locus (resulting in the *Idd* designation) in a backcross or F2 [46].

		CDR1		CDR2	

T1D-S LSETGVTQSPRYAIIQESQTVFFSCDPVSGHQSLYWYQQHRGQGPRLLVYFQNKVDIDKSQLPKDRFFAVRTEGVNSTLKIQSAKQSDSATYLCASS
T1D-R LSETGVTQSPRYAIIQESQTVSFWCDPVSGHERLYWYQQHRAQGPRLLVYFQNEVDIDKSQLPKDRFFAVRTEGVNSTLKIQSAKQSDSATYLCASS

Figure 2: Nonsynonymous amino acid sequence alignments of beta chain CDR1 and CDR2 regions of Vβ13a, found in T1D susceptible (T1D-S) rats, and Vβ13b, which is found in T1D resistant (T1D-R) rats [17, 20]. Both the CDR1 and CDR2 regions (red) exhibit differences (all indicated in yellow). Vβ13a is encoded by *Tcrb-V13S1A1* and has been found in the BBDR, BBDP, LEW.1WR1, LEW.1AR1-*iddm*, KDP, and PVGRT1u strains, all of which are T1D susceptible; Vβ13b is encoded by *Tcrb-V13S1A2* and is found in T1D resistant WF and BN rats [17]. Another allele, *Tcrb-V13S1A3P*, is a pseudogene found in the resistant F344 rat [17, 20].

A separate issue is whether a Vα (and TCR-Vα5D-4 in particular) selectively recognizes a permissive diabetogenic MHC molecule. A hint that this might occur is found in the mouse, where paralogs/alleles of TCR-Vα5D-4 likely influence differential selection on polymorphic MHC haplotypes. This is supported by the finding that the expressed repertoire of TCR-Vα5 paralogs in mice has a different frequency distribution in the periphery of mice with different MHC-II haplotypes [47]. In contrast, *Iddm14* (TCR-Vβ13-A1) was discovered in rats in a linkage study because this TCR allelic polymorphism exists among multiple strains bearing the same high-risk MHC-II haplotypes.

5. Data That Point toward a Mechanism Explaining the Role of TCR Genotype

5.1. Sequence Data in the Rat. The gene products of *Tcrb-V13S1A1* and *Tcrb-V13S1A2* encode different amino acid sequences for both the CDR1 and CDR2 regions of the beta chain [20]. This polymorphism distinguishes WF and other T1D-resistant strains from BBDR, BBDP, KDP, and LEW.1WR1 T1D-susceptible strains, all of which share the same class II MHC [18]. CDR1 and CDR2 sequences are encoded within each *Tcrb-V* allele and are not altered by the combinatorial processes that create the CDR3 regions of the TCR (Figure 2).

The Vβ13 sequences shown in Figure 2, which differentiate our susceptible and resistant alleles of *Tcrb-V13,* differ in both CDR1 and CDR2 and are consistent with emerging data on structural elements of the TCR-pMHC synapse that affect not only peptide recognition, but also binding affinity and peptide registration [48].

5.2. Structural Analyses. Crystal structures of the TCR-pMHC reveal the importance of the CDR1 and CDR2 regions in the human immunological synapse (recently reviewed in [48]). It is well accepted that CDR1 and CDR2 are critical for T cell-MHC restriction [49], and new data reveal how they interact with MHC helices to produce unanticipated and potentially important effects. In one study of the TCR "energetic landscape" it was noted that CDR1 and CDR2 loops act in a major way to stabilize the ligated CDR3 loops [50].

Another recent crystallographic study was designed specifically to address the question of whether shared germline contacts within the TCR-pMHC would persist despite distinct CDR3-peptide contacts in the model system, and they do [51]. The authors concluded that, "…a TCR utilizing entirely distinct chemistries to recognize different peptides exhibits highly persistent germline-mediated contacts."

Studies by Sethi et al. have reported the crystal structure of a TCR from a patient with multiple sclerosis that engages its pMHC ligand in an unusual manner [52]. The TCR is bound in a highly tilted orientation that prevents interaction of the TCR-α chain with the MHC class II β chain helix. In this structure, only a single germline-encoded (i.e., CDR1 or CDR2) TCRVβ loop engages the MHC protein. Furthermore, the reduced interaction surface with the peptide may facilitate TCR cross-reactivity.

Finally, a very recent study shows biased TCR usage against HLA DQ8-restricted gliadin peptides in persons with celiac disease [53]. These new data show that TCR usage biased to $TRBV9^*01$ underpins the recognition of HLA-DQ8-α-1-gliadin. More importantly for our hypothesis, they show that "*all* CDRβ loops (not just CDRβ3) interact with the gliadin peptide." They proved that "…Leu37β from the CDR1β loop, and Tyr57β from the CDR2β loop are the "hot spot" residues underpinning the SP3.4 TCR-DQ8-glia-α1 interaction providing a basis for the $TRBV9^*01$ bias." This is precisely what our rat data predict to be true in T1D (which is often comorbid with celiac disease). Interestingly, rat Vβ13 is polymorphic in the analogous CDR1β position 37 "hot spot" described for celiac disease, strengthening the notion that allelic polymorphism in Vβ13 may influence pMHC interaction.

6. TCR Allelism in the Human Genome

A number of factors could affect investigations of the role of genome-encoded TCR sequences in human T1D. One is the issue of paralogs, which we discuss using the NOD Vα5D-4 as an example. Like the NOD, humans could also have multiple paralogs that are capable of binding insulin, and thus, some linkage studies would not find a T1D gene in the Vα region if the parents each had one or more suitable Vα paralogs. However, GWAS evaluates many more individuals than a linkage study does, and would have more power to detect those individuals who do not possess a suitable allele of a Vα chain paralog that detects insulin autoantigen. Deep sequencing has not been performed on the Vα and Vβ regions in large numbers of people, so it is premature to suggest that such individuals do not exist. To date, however, the human *TCRα* locus does seem significantly less complex than in rodents due to rat- and mouse-specific gene duplication events and/or human specific gene convergence

```
HG00328        MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
gi|33525       MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
gi|161728643   MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
gi|642998      MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
gi|44887460    MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
gi|312225      MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
161728659      MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS

TRBV11-2*01    MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
1231           MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
1201           MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
00275          MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
00040          MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGnKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLEDSAVYLCASS
TRBV11-2*02    MGTRLLCWAALCLLGAELTEAGVAQSPRYKIIEKRQSVAFWCNPISGHATLYWYQQILGQGPKLLIQFQNNGVVDDSQLPKDRFSAERLKGVDSTLKIQPAKLENSAVYLCASS
```

FIGURE 3: Genomic TRBV11-2 DNA sequences from Caucasian samples in the 1000 genomes database (including HG00328 (and HG00361, HG00320, HG00111, HG00310, HG00247, HG00256, HG00231, HG00127, HG00103, HG00117, HG0032, not shown)), five expressed sequences from the dbEST (indicated with gi numbers), two alleles of TRBV11-2 from IMGT, and four of the sequences we obtained at Drexel (1231, 1201, 00275, 00040) were translated and compared. Only TRBV11-2*02 shows any nonsynonymous change (D105N, underlined, yellow highlight), and it is substantiated by one transcript (M33235). The D105N change (which is residue 98 using IMGT numbering) forms a salt bridge with a positive residue (Arg75) in the Vβ domain but is away from the pMHC (see Supplementary Figure 1 available online at http://dx.doi.org/10.1155/2013/737485). Accordingly there could be some functional/structural consequence of that SNP, but given its location (not in any CDR region) and conservative nature (Asp to Asn) it is unlikely that this is a functionally significant change. There is, however, no proof that this is the case. Nucleotide substitutions reflecting known SNPs (rs183490568, rs149749379, rs148941368, rs139187012, rs76976752, rs34112565, rs17163285, rs7777952, rs17281, rs17163283, rs17280, rs11505614, rs57147993, rs10375465, and rs17279) were commonly observed among these sequences. Using these SNPs, four haplotypes (two homozygous and two heterozygous) were observed among the four Drexel sequences.

(IMGT), so that human paralogs might in fact not complicate the genetic identification of a T1D gene in the Vα region.

Another factor that could affect investigations of the role of genome-encoded TCR sequences in human T1D is the possibility of extensive undocumented allelism. To investigate this possibility, we performed a preliminary study and analyzed one TCR isotype, TRBV11-2, in detail as proof of principle; it was selected because it is the human homolog of rat Tcrb-V13. From the 1000 Genomes database [57], we retrieved several TRBV11-2 sequences in which clear null alleles were present (some with open reading frames but no possibility of use in a TCR, as they did not have proper predicted cysteines or tryptophans at key residues ([57], (http://www.imgt.org/)). Among the remaining 1000 Genomes Caucasian sequences were multiple examples bearing the most common SNP variants, which supports the expected average minor allele frequency (MAF) of several of the SNPs in this region (0.21–0.43 among Caucasians, 1000 Genomes database). We sequenced a small number of samples at Drexel, to determine the level of polymorphism in this TCR element. The predominant TRBV11-2 allele among our sequences is supported by numerous ESTs and cDNAs (Figure 3). Interestingly, the UCSC database (hg19) [58] has a reference sequence for TRBV11-2 with a stop codon in the leader sequence, but all four of these TRBV11-2 genes have an open reading frame throughout the entire gene. The low read depth of sequencing, and the likely heterozygosity in the 1000 Genomes database, can contribute to ambiguous calls, leading us to question all but the most substantiated alleles of 11-2. Therefore, we searched the NCBI EST/cDNA databases for known transcripts with homology to TRBV11-2 (using BLAST), and these confirm the major alleles we found by sequencing. Our results compare with only two amino acid substitution alleles recognized previously in IMGT.

To generate a comprehensive analysis of genomic TCR effects in T1D, it will be necessary to collect alleles from each chromosome of heterozygous individuals. This information is missing from any available database, as most report a single sequence from each "person" in the cohort studied. As it is likely that people are polymorphic for TCR Vβ regions, we expect that many will be heterozygous. Thus, data are needed to fill the gap in our knowledge about variation in TRAV and TRBV as well as to contribute information on linkage disequilibrium (LD) between the many different isotypes within the TCR genomic complexes.

7. Molecular and Analytical Basis for Identification of Human TCRs Involved in T1D

7.1. Identifying and Mapping Known Human TCR Polymorphisms. To examine the possibility of germ-line human TCR variants having a functional impact on the immune repertoire and a potential impact on autoimmunity and T1D in particular, we extended the analysis of polymorphism in TRBV and compiled a comprehensive set of nonsynonymous polymorphisms in the human TRAV and TRBV genes in regions known to contact the pMHC (Table 2) [56]. In addition to the CDR1 and CDR2 loops, we included the TCR N-termini and HV4 loops, which often have 1-2 residues contacting pMHC [59, 60]. Our results (Table 2) indicate that there is a notable degree of polymorphism in both the TCRα and TCRβ chains, in regions with direct impact on pMHC binding. This was also noted in a previous study of human TRAV gene diversity where the authors found that nucleotide diversity was "substantially higher in the CDRs versus the FRs (framework regions)" [56]. This diversity is observed in

TABLE 2: Previously identified polymorphisms in *TRAV/TRBV* genes near the pMHC interface.

TCR Location[1]	Polymorphism(s)[2]	Genes[2]
N-term	N2D	*TRAV9-2*
CDR1α	V27M, G29V, **G29R**, **N30S**, P30E, P30Q, N31D, **Y32S**	*TRAV36, TRAV12-2, TRAV8-4, TRAV14-1, TRAV38-1,* **TRAV20**
CDR2α	F55S, **Q56E**, A57G, **V57M**, S58T, **T58I**, A59G, **K59E**, Q61E	*TRAV12-2, TRAV1-1, TRAV8-4, TRAV14-1,* **TRAV25, TRAV8-7, TRAV26-2, TRAV38-1**
CDR1β	A30V, N30E	*TRBV7-7, TRBV6-6*
CDR2β	Q55H, Q57H, V57I, D58N, G60D, S60C, Q60H, L61I	*TRBV9, TRBV19, TRBV30, TRBV15, TRBV20-1, TRBV10-1, TRBV3-1*
HV4β	G84E	*TRBV7-2*

[1]Region of the TCR variable domain tertiary structure. CDRs are as defined by Lefranc et al. [54], with CDR2 extended by one residue at the N-terminus to account for pMHC contacts with this position. [2]From IMGT [55], as well as additional data from Mackelprang et al. [56] (in bold italics). IMGT TCR residue numbering used.

a total of 19 positions and 22 genes, including the *TRBV9* gene, which, as noted above, is implicated in celiac disease [53].

It is not certain whether the particular polymorphism in TRBV9 (Q55H) plays a key role in recognition of the gliadin-HLA-DQ8 celiac disease antigen (see above), though a structural and biophysical study of the TK3 TCR, which is encoded by the *TRBV9* gene, found that this particular variant affected the electrostatic makeup and structure of pMHC recognition for HLA-B*35:01 and an EBV peptide [61].

To provide a structural context for these TCR sequence polymorphisms, we mapped their positions onto the structure of a complex of a TCR with a Class II MHC and peptide (Figure 4). While not all TCRs exhibit the same docking geometry, this particular complex [62] has a typical pMHC docking angle (49°) and is, thus, approximately representative of a variety of known complexes [63]. These polymorphisms in germ-line encoded TCR genes have a clear potential to impact the pMHC recognition, and, as the figure indicates, primarily via interaction with MHC helices. Additionally, a subset of these variants (including TCRβ position 55 in the case of the TK3 TCR [61]) have the potential to directly interact with the peptide as well. It is worth noting that additional positions outside of those analyzed here can potentially impact pMHC binding, given that TCRs are known to exhibit long-range energetic effects [64, 65] and long-range dynamic coupling between distal TCR sites [66]. It should also be noted that Table 2 is very likely an incomplete representation of TCR polymorphisms, and next generation sequencing of T cells from diseased populations, in conjunction with larger-scale studies of exome sequencing (including the 1000 Genomes project [67]), should yield valuable data on TCR polymorphisms and insights into their impact on autoimmune diseases like T1D and control of infection.

7.2. Improving Detection via Targeted Genotyping and Next Generation Sequencing.

The question remains, then: why have such putative TCR alleles not been detected in GWAS studies? To address this issue, we performed a quantitative
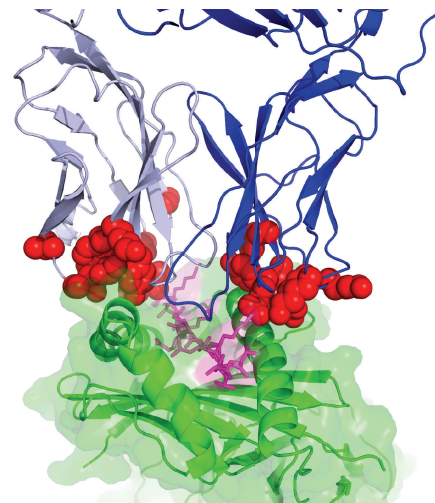

FIGURE 4: TRAV and TRBV polymorphic positions shown on a TCR/pMHC complex structure. Polymorphic positions are red, TCRα is light blue, TCRβ is blue, peptide is magenta, and MHC is green. Structure shown is HA1.7 TCR/HA peptide/HLA-DR4 (Protein Data Bank [43] ID 1J8H). The figure was generated using PyMOL (http://www.pymol.org/).

analysis. Assume that some *TCRV* allele, call it "*Vx*," is important for T1D in individuals with a particular HLA-DRB1 allele (e.g., HLA-DRB1*03:01, which will be referred to here as DR3). We base this assumption on our data showing that Vβ13a is important for T1D in MHC Class II *RT1Bu* rats. Our remaining analysis is as follows.

7.2.1. Stratification by DR Allele.

The ability to detect a causative "*Vx*" TCR allele may be dramatically improved by stratifying disease population by HLA-DR allele. Here we consider stratification by HLA diplotype (rather than presence of one haplotype alone), based on the impact of trans-encoded HLA proteins on T1D susceptibility and immune function [8, 68], as observed for the high-risk trans-encoded HLA-DQA1*05:01/DQB1*03:02. Additionally, this will control for the potential confounding effects of another
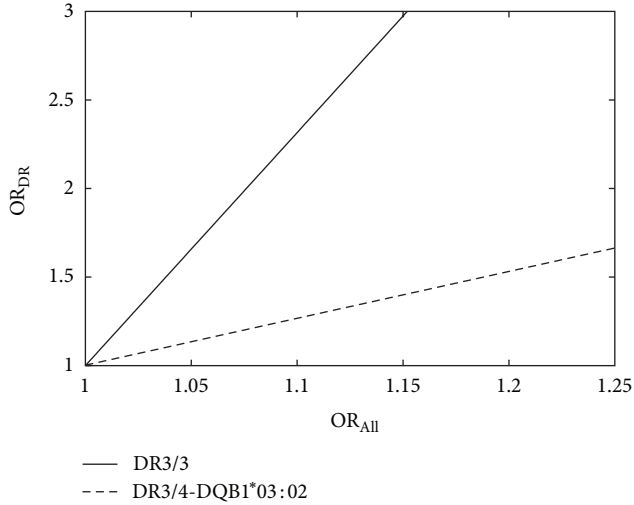
FIGURE 5: Calculated improvement in odds ratio (OR) when stratifying by HLA-DR-DQ genotype (OR_{DR}) versus analyzing all T1D patients (OR_{All}) for detection of a putative risk TCR allele. Shown are data for homozygous HLA-DR3/3 and the high-risk DR3/4-$DQB1^*03:02$ diplotype.

HLA allele whose interaction with a different TCR could confer either additional susceptibility or resistance. However, such analysis (and stratification) can be applied for single alleles (i.e., the percentage of the T1D population with at least one copy of an allele), though haplotype frequencies (as given in Table 2 of Erlich et al. [8]) would not suffice. This is because they provide overall frequency of haplotypes within all diploid genomes (not distinguishing between homozygous and heterozygous).

As an example, we will focus on one of the high-risk HLA alleles, DR3, but the argument can be applied to any allele of interest. The frequency of DR3/3 homozygotes in T1D patients is 7.6% [8]. Assuming that the "Vx" allele is relevant in the context of HLA-DR3/3 and not the remainder of the HLA-DR allele combinations, the odds ratio seen for all T1D subjects (OR_{All}) is scaled accordingly from the odds ratio of "Vx" within the DR3/3 T1D population ($OR_{DR3/3}$) by the DR3/3 frequency:

$$OR_{All} = 0.076 * OR_{DR3/3} + 0.924. \tag{1}$$

A comparison of stratified versus unstratified ORs is shown in Figure 5, for both DR3/3 and another diplotype (HLA-DR3/4-$DQB1^*03:02$), which is more frequent in T1D (38% of the T1D population with European ancestry [8]) and contains the high-risk trans-encoded HLA-$DQA^*05:01$/HLA-$DQB^*03:02$ noted above. Even for a relatively modest value of OR_{All} (1.1, which is below the level of most identified genes in GWAS studies [11]), the OR for DR3/3 becomes appreciably higher (2.3) and somewhat higher (1.27) for DR3/4-$DQB1^*03:02$. Given these odds ratios, for a study with 2000 patients, 2000 controls and background allele frequency of 10%, power to detect the risk TCR allele increases from 16% to 75% for DR3/4-$DQB1^*03:02$, and to 100% for DR3/3 (G^*Power version 3.1, Fisher's two-tailed

exact test, $\alpha = 0.05$). This implies that for previous studies that did not include stratification by HLA alleles, detection of a "Vx" allele may have been confounded by dilution of the odds ratio.

This analysis helps explain why a TCR α chain locus associated with autoimmune narcolepsy was identified using GWAS [29]. That study was stratified by default because a single HLA haplotype ($DRB1^*15:01$-$DQB1^*06:02$) is seen in 90% of cases [69]. That study did not identify the specific allele of a TCR that confers narcolepsy, though one might be identified using next generation sequencing (NGS).

7.2.2. NGS versus GWAS Array Coverage. Traditionally, analysis of polymorphisms in T1D has involved genome-wide association studies (GWASs), which rely on arrays with defined sets of probes, to infer the genotype in different regions of the locus. An analysis of linkage disequilibrium (LD) at the human *TRAV* locus found that it is highly variable, leading the authors to conclude that, "even with relatively dense coverage, it is unlikely that a genotyping strategy (as opposed to a resequencing strategy) will provide adequately dense coverage extending into the *V* genes" [56]. We examined this contention further using a dataset of *TRAV* and *TRBV* SNPs downloaded from the UCSC genome browser [58] (dbSNP [70] build 130), selecting only SNPs found within TRAV and *TRBV* exons from IMGT reference sequences [55]. This resulted in a total of 322 SNPs (158 in *TRAV* exons and 164 in TRBV exons). We used the SNAP web server [71] to calculate the degree of correlation (r^2) of these SNPs with three arrays used in previous studies of GWAS in T1D (Affymetrix GeneChip 500K, Illumina 550K Infinium and Affymetrix SNP Array 6.0) [72] and found average r^2 of approximately 0.3, and approximately 25% of SNPs with r^2 of 0.8 or higher. This coverage is somewhat lower than observed for common genome-wide SNPs from European ancestry for the Affymetrix GeneChip 500 (commensurate with values for the less common SNPs with MAFs of 1–5%) [73], but since these SNPs were not classified by frequency it is not possible to conclude that *TRAV/TRBV* coverage is lower than the rest of the genome, or determine the coverage of SNPs above a certain frequency threshold. Regardless, in light of the previous findings of variable LD at these loci noted above, these data support the notion that average r^2 for *TRAV* and *TRBV* SNPs is markedly lower than 1. As correlation is directly proportional to power of GWAS studies [74], it follows that there is likely at least some reduction in power to detect putative TCR alleles, reinforcing the strategy of NGS to genotype TCR polymorphisms.

8. Conclusion

The striking observations that have recently been made in the NOD mouse and in multiple strains of rats that are used to model T1D strongly suggest that elements of the TCR, encoded at the level of the genome and not subject to V-(D)-J recombination, may play a critical role in T1D susceptibility in an MHC-dependent fashion. Assuming these animal model systems are reliable, the likelihood is

high that the germline TCR regions are also important for susceptibility to human T1D. This is not a new idea, but no supporting data have been reported to date. However, our quantitative analyses show that cogent statistical flaws in previous approaches could account for the negative findings. If an HLA-dependent TCR susceptibility paradigm can be identified for human T1D in the future, it could open the path to new ways of preventing the disease, either by targeting specific alleles of the TCR with deletion approaches or by using small molecules to interfere with the specific TCR-pMHC synapses that lead to the disease.

Acknowledgments

References

[1] T. L. Van Belle, K. T. Coppieters, and M. G. Von Herrath, "Type 1 diabetes: etiology, immunology, and therapeutic strategies," *Physiological Reviews*, vol. 91, no. 1, pp. 79–118, 2011.

[2] J. E. Tooley, F. Waldron-Lynch, and K. C. Herold, "New and future immunomodulatory therapy in type 1 diabetes," *Trends in Molecular Medicine*, vol. 18, no. 3, pp. 173–181, 2012.

[3] C. Polychronakos and Q. Li, "Understanding type 1 diabetes through genetics: advances and prospects," *Nature Reviews Genetics*, vol. 12, no. 11, pp. 781–792, 2011.

[4] M. Trucco, "Gene-environment interaction in type 1 diabetes mellitus," *Endocrinología y Nutrición*, vol. 56, no. 4, pp. 56–59, 2009.

[5] D. K. Wherrett and D. Daneman, "Prevention of type 1 diabetes," *Endocrinology and Metabolism Clinics of North America*, vol. 38, no. 4, pp. 777–790, 2009.

[6] A. W. Michels and G. S. Eisenbarth, "Immune intervention in type 1 diabetes," *Seminars in Immunology*, vol. 23, no. 3, pp. 214–219, 2011.

[7] L. Zhang and G. S. Eisenbarth, "Prediction and prevention of Type 1 diabetes mellitus," *Journal of Diabetes*, vol. 3, no. 1, pp. 48–57, 2011.

[8] H. Erlich, A. M. Valdes, J. Noble et al., "HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk analysis of the type 1 diabetes genetics consortium families," *Diabetes*, vol. 57, no. 4, pp. 1084–1092, 2008.

[9] A. Ide, S. R. Babu, D. T. Robles et al., ""Extended" A1, B8, DR3 haplotype shows remarkable linkage disequilibrium but is similar to nonextended haplotypes in terms of diabetes risk," *Diabetes*, vol. 54, no. 6, pp. 1879–1883, 2005.

[10] J. A. Noble, A. M. Valdes, M. D. Varney et al., "HLA class I and genetic susceptibility to type 1 diabetes: results from the type 1 diabetes genetics consortium," *Diabetes*, vol. 59, no. 11, pp. 2972–2979, 2010.

[11] P. Concannon, S. S. Rich, and G. T. Nepom, "Genetics of type 1A diabetes," *New England Journal of Medicine*, vol. 360, no. 16, pp. 1646–1654, 2009.

[12] D. J. Smyth, V. Plagnol, N. M. Walker et al., "Shared and distinct genetic variants in type 1 diabetes and celiac disease," *New England Journal of Medicine*, vol. 359, no. 26, pp. 2767–2777, 2008.

[13] M. Rewers, T. L. Bugawan, J. M. Norris et al., "Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY)," *Diabetologia*, vol. 39, no. 7, pp. 807–812, 1996.

[14] J. L. Davies, Y. Kawaguchi, S. T. Bennett et al., "A genome-wide search for human type 1 diabetes susceptibility genes," *Nature*, vol. 371, no. 6493, pp. 130–136, 1994.

[15] D. G. Clayton, "Prediction and interaction in complex disease genetics: experience in type 1 diabetes," *PLoS Genetics*, vol. 5, no. 7, Article ID e1000540, 2009.

[16] C. Winkler, J. Krumsiek, J. Lempainen et al., "A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes," *Genes & Immunity*, vol. 13, no. 7, pp. 549–555, 2012.

[17] Z. Liu, L. Cort, R. Eberwine et al., "Prevention of type 1 diabetes in the rat with an allele-specific anti-T cell receptor antibody: Vβ13 as a therapeutic target and biomarker," *Diabetes*, vol. 61, no. 5, pp. 1160–1168, 2012.

[18] J. P. Mordes, L. Cort, E. Norowski et al., "Analysis of the rat Iddm14 diabetes susceptibility locus in multiple rat strains: identification of a susceptibility haplotype in the Tcrb-V locus," *Mammalian Genome*, vol. 20, no. 3, pp. 162–169, 2009.

[19] K. E. Ellerman and A. A. Like, "Susceptibility to diabetes is widely distributed in normal class II(u) haplotype rats," *Diabetologia*, vol. 43, no. 7, pp. 890–898, 2000.

[20] M. Stienekemeier, K. Hofmann, R. Gold, and T. Herrmann, "A polymorphism of the rat T-cell receptor β-chain variable gene 13 (BV13S1) correlates with the frequency of BV13S1-positive CD4 cells," *Immunogenetics*, vol. 51, no. 4-5, pp. 296–305, 2000.

[21] Z. Laron, "Interplay between heredity and environment in the recent explosion of type 1 childhood diabetes mellitus," *American Journal of Medical Genetics*, vol. 115, no. 1, pp. 4–7, 2002.

[22] A. Green and C. C. Patterson, "Trends in the incidence of childhood-onset diabetes in Europe 1989–1998," *Diabetologia*, vol. 44, no. 3, pp. B3–B8, 2001.

[23] T. H. Lipman, Y. Chang, and K. M. Murphy, "The epidemiology of type 1 diabetes in children in Philadelphia 1990–1994: evidence of an epidemic," *Diabetes Care*, vol. 25, no. 11, pp. 1969–1975, 2002.

[24] N. van der Werf, F. G. M. Kroese, J. Rozing, and J. L. Hillebrands, "Viral infections as potential triggers of type 1 diabetes," *Diabetes/Metabolism Research and Reviews*, vol. 23, no. 3, pp. 169–183, 2007.

[25] C. T. Watson, A. E. Para, M. R. Lincoln et al., "Revisiting the T-cell receptor alpha/delta locus and possible associations with multiple sclerosis," *Genes and Immunity*, vol. 12, no. 2, pp. 59–66, 2011.

[26] S. J. Manavalan, J. R. Valiando, W. H. Reeves et al., "Genomic Absence of the Gene Encoding T Cell Receptor Vβ7.2 Is Linked to the Presence of Autoantibodies in Sjögren's Syndrome," *Arthritis and Rheumatism*, vol. 50, no. 1, pp. 187–198, 2004.

[27] R. A. Kay, C. J. Hutchings, and W. E. R. Ollier, "A subset of Sjogren's syndrome associates with the TCRBV13S2 locus but not the TCRBV2S1 locus," *Human Immunology*, vol. 42, no. 4, pp. 328–330, 1995.

[28] J. L. Black, "Narcolepsy: a review of evidence for autoimmune diathesis," *International Review of Psychiatry*, vol. 17, no. 6, pp. 461–469, 2005.

[29] J. Hallmayer, J. Faraco, L. Lin et al., "Narcolepsy is strongly associated with the T-cell receptor alpha locus," *Nature Genetics*, vol. 41, no. 6, pp. 708–711, 2009.

[30] T. Miyagawa, M. Honda, M. Kawashima et al., "Polymorphism located in TCRA locus confers susceptibility to essential hypersomnia with HLA-DRB1*1501-DQB1*0602 haplotype," *Journal of Human Genetics*, vol. 55, no. 1, pp. 63–65, 2010.

[31] J. P. Mordes, P. Poussier, A. A. Rossini, E. P. Blankenhorn, and D. L. Greiner, "Rat models of type 1 diabetes: Genetics, environment, and autoimmunity," in *Animal Models of Diabetes: Frontiers in Research*, E. Shafrir, Ed., pp. 1–39, 2007.

[32] T. Awata, D. L. Guberski, and A. A. Like, "Genetics of the BB rat: association of autoimmune disorders (diabetes, insulitis, and thyroiditis) with lymphopenia and major histocompatibility complex class II," *Endocrinology*, vol. 136, no. 12, pp. 5731–5735, 1995.

[33] A. M. Martin, M. N. Maxson, J. Leif, J. P. Mordes, D. L. Greiner, and E. P. Blankenhorn, "Diabetes-prone and diabetes-resistant BB rats share a common major diabetes susceptibility locus, *iddm4*: additional evidence for a "universal autoimmunity locus" on rat chromosome 4," *Diabetes*, vol. 48, no. 11, pp. 2138–2144, 1999.

[34] E. P. Blankenhorn, L. Rodemich, C. Martin-Fernandez, J. Leif, D. L. Greiner, and J. P. Mordes, "The rat diabetes susceptibility locus *iddm4* and at least one additional gene are required for autoimmune diabetes induced by viral infection," *Diabetes*, vol. 54, no. 4, pp. 1233–1237, 2005.

[35] J. P. Mordes, J. Leif, S. Novak, C. Descipio, D. L. Greiner, and E. P. Blankenhorn, "The *iddm4* locus segregates with diabetes susceptibility in congenic WF.*iddm4* Rats," *Diabetes*, vol. 51, no. 11, pp. 3254–3262, 2002.

[36] L. Hornum, C. DeScipio, H. Markholst et al., "Comparative mapping of rat *iddm4* to segments on HSA7 and MMU6," *Mammalian Genome*, vol. 15, no. 1, pp. 53–61, 2004.

[37] E. P. Blankenhorn, C. Descipio, L. Rodemich et al., "Refinement of the *iddm4* diabetes susceptibility locus reveals TCRVβ4 as a candidate gene," *Annals of the New York Academy of Sciences*, vol. 1103, pp. 128–131, 2007.

[38] J. M. Fuller, M. Bogdani, T. D. Tupling et al., "Genetic dissection reveals diabetes loci proximal to the *gimap5* lymphopenia gene," *Physiological Genomics*, vol. 38, no. 1, pp. 89–97, 2009.

[39] J. P. Mordes, D. L. Guberski, J. H. Leif et al., "LEW.1WR1 rats develop autoimmune diabetes spontaneously and in response to environmental perturbation," *Diabetes*, vol. 54, no. 9, pp. 2727–2733, 2005.

[40] R. S. Tirabassi, D. L. Guberski, E. P. Blankenhorn et al., "Infection with viruses from several families triggers autoimmune diabetes in LEW.1WR1 rats: prevention of diabetes by maternal immunization," *Diabetes*, vol. 59, no. 1, pp. 110–118, 2010.

[41] M. Nakayama, T. Castoe, T. Sosinowski et al., "Germline TRAV5D-4 T cell receptor sequence targets a primary insulin peptide of NOD mice," *Diabetes*, vol. 61, no. 4, pp. 857–865, 2012.

[42] L. Zhang, J. M. Jasinski, M. Kobayashi et al., "Analysis of T cell receptor beta chains that combine with dominant conserved TRAV5D-4*04 anti-insulin B:9-23 alpha chains," *Journal of Autoimmunity*, vol. 33, no. 1, pp. 42–49, 2009.

[43] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[44] N. Abiru, D. Wegmann, E. Kawasaki, P. Gottlieb, E. Simone, and G. S. Eisenbarth, "Dual overlapping peptides recognized by insulin peptide B:9–23 T cell receptor AV13S3 T cell clones of the NOD mouse," *Journal of Autoimmunity*, vol. 14, no. 3, pp. 231–237, 2000.

[45] K. W. Wucherpfennig, E. Gagnon, M. J. Call, E. S. Huseby, and M. E. Call, "Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling," *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 4, Article ID a005140, 2010.

[46] S. Ghosh, S. M. Palmer, N. R. Rodrigues et al., "Polygenic control of autoimmune diabetes in nonobese diabetic mice," *Nature Genetics*, vol. 4, no. 4, pp. 404–409, 1993.

[47] E. A. Simone, L. Yu, D. R. Wegmann, and G. S. Eisenbarth, "T cell receptor gene polymorphisms associated with anti-insulin, autoimmune T cells in diabetes-prone NOD mice," *Journal of Autoimmunity*, vol. 10, no. 3, pp. 317–321, 1997.

[48] K. W. Wucherpfennig, "The first structures of T cell receptors bound to peptide-MHC," *Journal of Immunology*, vol. 185, no. 11, pp. 6391–6393, 2010.

[49] B. C. Sim, L. Zerva, M. I. Greene, and N. R. J. Gascoigne, "Control of MHC restriction by TCR V(α) CDR1 and CDR2," *Science*, vol. 273, no. 5277, pp. 963–966, 1996.

[50] N. A. Borg, L. K. Ely, T. Beddoe et al., "The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition," *Nature Immunology*, vol. 6, no. 2, pp. 171–180, 2005.

[51] J. J. Adams, S. Narayanan, B. Liu et al., "T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex," *Immunity*, vol. 35, no. 5, pp. 681–693, 2011.

[52] D. K. Sethi, D. A. Schubert, A. K. Anders et al., "A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC," *Journal of Experimental Medicine*, vol. 208, no. 1, pp. 91–102, 2011.

[53] S. E. Broughton, J. Petersen, A. Theodossis et al., "Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease," *Immunity*, vol. 37, no. 4, pp. 611–621, 2012.

[54] M. P. Lefranc, C. Pommié, M. Ruiz et al., "IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains," *Developmental and Comparative Immunology*, vol. 27, no. 1, pp. 55–77, 2003.

[55] M. P. Lefranc, V. Giudicelli, C. Ginestoux et al., "IMGT, the international ImMunoGeneTics information system," *Nucleic Acids Research*, vol. 37, no. 1, pp. D1006–D1012, 2009.

[56] R. Mackelprang, R. J. Livingston, M. A. Eberle et al., "Sequence diversity, natural selection and linkage disequilibrium in the human T cell receptor alpha/delta locus," *Human Genetics*, vol. 119, no. 3, pp. 255–266, 2006.

[57] D. L. Altshuler, R. M. Durbin, G. R. Abecasis et al., "A map of human genome variation from population-scale sequencing," *Nature*, vol. 467, no. 7319, pp. 1061–1073, 2010.

[58] L. R. Meyer, A. S. Zweig, A. S. Hinrichs et al., "The UCSC genome browser database: extensions and updates 2013," *Nucleic Acids Research*, vol. 41, no. 1, pp. D64–D69, 2012.

[59] Q. Kaas, E. Duprat, G. Tourneur et al., "IMGT standardization for molecular characterization of the T-cell receptor/peptide/MHC complexes," *Immunoinformatics*, vol. 1, pp. 19–49, 2008.

[60] B. G. Pierce and Z. Weng, "A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes," *Protein Science*, vol. 22, no. 1, pp. 35–46, 2012.

[61] S. Gras, Z. Chen, J. J. Miles et al., "Allelic polymorphism in the T cell receptor and its impact on immune responses," *Journal of Experimental Medicine*, vol. 207, no. 7, pp. 1555–1567, 2010.

[62] J. Hennecke and D. C. Wiley, "Structure of a complex of the human α/β T cell receptor (TCR) HA1.7, Influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRBI*0401): insight into TCR cross-restriction and alloreactivity," *Journal of Experimental Medicine*, vol. 195, no. 5, pp. 571–581, 2002.

[63] M. G. Rudolph, R. L. Stanfield, and I. A. Wilson, "How TCRs bind MHCs, peptides, and coreceptors," *Annual Review of Immunology*, vol. 24, pp. 419–466, 2006.

[64] B. Moza, R. A. Buonpane, P. Zhu et al., "Long-range cooperative binding effects in a T cell receptor variable domain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 26, pp. 9867–9872, 2006.

[65] B. G. Pierce, J. N. Haidar, Y. Yu, and Z. Weng, "Combinations of affinity-enhancing mutations in a T cell receptor reveal highly nonadditive effects within and between complementarity determining regions and chains," *Biochemistry*, vol. 49, no. 33, pp. 7050–7059, 2010.

[66] W. F. Hawse, M. M. Champion, M. V. Joyce et al., "Cutting edge: evidence for a dynamically driven T cell signaling mechanism," *Journal of Immunology*, vol. 188, no. 12, pp. 5819–5823, 2012.

[67] G. R. Abecasis, A. Auton, L. D. Brooks et al., "An integrated map of genetic variation from 1,092 human genomes," *Nature*, vol. 491, no. 7422, pp. 56–65, 2012.

[68] A. W. Michels and M. Nakayama, "The anti-insulin trimolecular complex in type 1 diabetes," *Current Opinion in Endocrinology, Diabetes and Obesity*, vol. 17, no. 4, pp. 329–334, 2010.

[69] J. Faraco and E. Mignot, "Immunological and genetic aspects of narcolepsy," *Sleep Medicine Research*, vol. 2, pp. 1–9, 2011.

[70] S. T. Sherry, M. H. Ward, M. Kholodov et al., "DbSNP: the NCBI database of genetic variation," *Nucleic Acids Research*, vol. 29, no. 1, pp. 308–311, 2001.

[71] A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'Donnell, and P. I. W. De Bakker, "SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap," *Bioinformatics*, vol. 24, no. 24, pp. 2938–2939, 2008.

[72] J. C. Barrett, D. G. Clayton, P. Concannon et al., "Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes," *Nature Genetics*, vol. 41, no. 6, pp. 703–707, 2009.

[73] I. Pe'er, P. I. W. De Bakker, J. Maller, R. Yelensky, D. Altshuler, and M. J. Daly, "Evaluating and improving power in whole-genome association studies using fixed marker sets," *Nature Genetics*, vol. 38, no. 6, pp. 663–667, 2006.

[74] C. C. A. Spencer, Z. Su, P. Donnelly, and J. Marchini, "Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip," *PLoS Genetics*, vol. 5, no. 5, Article ID e1000477, 2009.