Mike J. Mason, Cate Speake, Vivian H. Gersuk, Quynh-Anh Nguyen, Kimberly K. O'Brien, Jared M. Odegard, Jane H. Buckner, Carla J. Greenbaum, Damien Chaussabel, and Gerald T. Nepom

Low HERV-K(C4) Copy Number Is Associated With Type 1 Diabetes

Complement component C4 (C4) is a highly variable complement pathway gene situated ~500 kb from DRB1 and DQB1, the genes most strongly associated with many autoimmune diseases. Variations in C4 copy number (CN), length, and isotype create a highly diverse gene cluster in which insertion of an endogenous retrovirus in the ninth intron of C4, termed HERV-K(C4), is a notable component. We investigated the relationship between C4 variation/CN and type 1 diabetes. We found that individuals with type 1 diabetes have significantly fewer copies of HERV-K(C4) and that this effect is not solely due to linkage with known major histocompatibility complex class II susceptibility alleles. We show that HERV-K(C4) is a novel marker of type 1 diabetes that accounts for the disease association previously attributed to some key HLA-DQB1 alleles, raising the possibility that this retroviral insertion element contributes to functional protection against type 1 diabetes.

Genes encoding complement component C4 (C4) are located on chromosome 6 ~500 kb from the major histocompatibility complex (MHC) class II genes associated with multiple autoimmune diseases, including multiple sclerosis, systemic lupus erythematosus (SLE), celiac sprue, Graves disease, and rheumatoid arthritis (1). Additionally, this region is the most strongly associated with type 1 diabetes (odds ratio [OR] >6.5) (2). The HLA class II alleles most highly associated with type 1 diabetes are at the DQB1 locus, with risk modifiers at the closely linked HLA-DRB1 and class I loci (3,4). Specifically, the alleles with the highest associated risk are DRB1*04/DQB1*03:02 and DRB1*03:01/DQB1*02:01 (OR 11.37 and 3.64, respectively), whereas DRB1*15:01/DQB1*06:02 is a protective allele (OR 0.03) (5). However, extended regions of linkage disequilibrium characterize the MHC region, confounding efforts to attribute susceptibility to single loci or, indeed, to the classical HLA class I and class II alleles themselves. This is particularly notable with DRB1*03:01/DQB1*02:01, which forms an extended haplotype >1 Mb and encompasses the complement gene region (6,7).

C4 varies in three key ways: copy number (CN), isoform, and length. C4 is the second gene in the four-gene RCCX cassette and positioned midway between the MHC class I and MHC class II gene clusters (Fig. 1A). The RCCX cassette comprises STK19 (RP1), C4, CYP21A2, and TNXB (8). CYP21A2 contains a recombination site resulting in mono-, bi-, and trimodular cassettes with one, two, and three functional copies of C4, respectively, while retaining just one functional copy of STK19, CYP21A2, and TNXB (Fig. 1B). Each copy of C4 is either acidic or basic (C4A or C4B) as determined by nucleic acid differences in the 26th exon. C4A deficiency, having zero or one functional copy of C4A, has been linked to both type 1 diabetes and SLE (9–13). Ninety-five percent of C4A and 54% of C4B isoforms contain a 6.5-kb class K human endogenous retrovirus in their ninth intron [HERV-K(C4)] (14,15). HERVs are retroviral elements
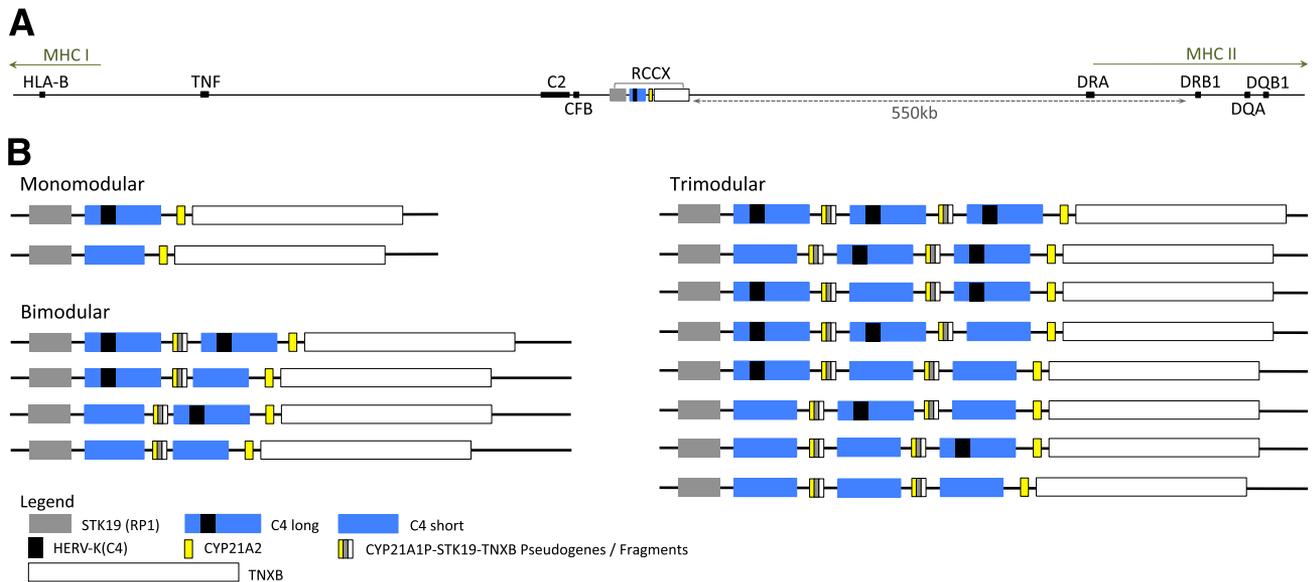
Figure 1—RCCX location and configurations. *A*: RCCX is located approximately midway between the MHC class I and MHC class II regions. This region is gene dense, and most genes have been omitted from this illustration for simplicity. Class II genes have been abbreviated (e.g., HLA-DRB1 is DRB1). *B*: Mono-, bi-, and trimodular RCCX cassette configurations. C2, complement component 2; CFB, complement factor B; TNF, tumor necrosis factor.

that integrated into the human germline during the course of evolution. In addition to long terminal repeats, HERVs contain gag, pro, pol, and env genes and make up ~8–9% of the human genome (16). Recent work has shown that HERVs may affect expression of nearby genes (17). HERV-K(C4) is oriented opposite to C4, resulting in its antisense transcription (18,19). Genes with HERV-K (C4) are termed C4 long (C4L), and those HERV-K(C4) it are termed C4 short (C4S).

Historically, it was believed that individuals had four copies of C4, with a long C4A followed by a long or short C4B on each chromosome and some rare individuals missing either C4A or C4B (9). Earlier studies, however, were limited by methods available at the time and were unable to quantitatively assay C4. Newer methods have revealed that individuals inherit one to three copies of C4 from each parent with many possible isoform and length combinations (8,15,20). We developed a high-throughput, sensitive assay to study the relationship between C4 variability and type 1 diabetes in the context of associated HLA alleles. We show that low HERV-K(C4) CN is a strong marker of type 1 diabetes risk and that it captures the disease associations of the DRB1*03:01/DQB1*02:01 and DRB1*15:01/DQB1*06:02 HLA haplotypes.

RESEARCH DESIGN AND METHODS

Sample Collection and Processing

Samples were obtained from type 1 diabetic patients and healthy donors participating in studies under the JDRF/Diabetes and Control Registry and Repositories. Informed consent was obtained from subjects according to institutional review board–approved protocols at Benaroya Research Institute and Seattle Children's Hospital. DNA was isolated from whole blood drawn using EDTA and from peripheral blood mononuclear cell (PBMC) samples drawn in heparin. RNA samples were collected in Tempus Blood RNA Tubes, extracted using MagMAX for Stabilized Blood Tubes RNA Isolation Kit, and then processed using GLOBINclear (all from Life Technologies). cDNA was prepared with the High Capacity cDNA Reverse Transcription Kit (Life Technologies). Two cohorts were collected from the sample repository: Cohort 1 (n = 107) was randomly sampled from the repository, whereas samples in cohort 2 (n = 110) were selected on the basis of HLA type (Table 1).

C4 CN Estimation

TaqMan quantitative PCR was used to determine C4 variant CN (20). Assays were run on a BioMark HD system (Fluidigm, San Francisco, CA); HLA-DRA was the reference gene. DNA samples were preamplified for 10 cycles with a pool of the five assays, treated with exonuclease I (New England Biolabs), and diluted 1:5 before further use. Cloned amplicons from each target were used to generate a plasmid standard curve. Fourteen samples with known C4 CN from the International Histocompatibility Working Group were used for linear correction (20) (Supplementary Table 1 and Supplementary Fig. 1). Primers and probes for DRA, C4B, and C4S were previously defined (20–22): C4A forward primer 5′CTGTCCCAGCAGCAGGCT3′, reverse primer 5′TCCTCCGACAGGCGCTT3′, and probe 5′TCCAGTGTTAGACAGGAGC3′ and C4L forward

Table 1—Sample information

Cohort	Female/male sex (%)	Race	HLA-DQB1			
			*03:02	*02:01	*06:02	None
Cohort 1 (n = 107)						
C (n = 57)	53/47	96% white, 2% Asian, 2% other (0% Hispanic)	16 (28)	15 (26)	18 (32)	15 (26)
D (n = 50)	56/44	94% white, 2% Asian, 4% other (2% Hispanic)	31 (62)	40 (80)	0 (0)	5 (9)
Cohort 2 (n = 110)						
NR-C (n = 33)	60/40	67% white, 18% Asian, 9% other (12% Hispanic)	0 (0)	0 (0)	0 (0)	33 (100)
NR-D (n = 17)	41/59	76% white, 12% Asian, 12% other (6% Hispanic)	0 (0)	0 (0)	0 (0)	17 (100)
DR4-C (n = 24)	74/25	96% white, 4% Asian (4% Hispanic)	24 (100)	6 (24)	0 (0)	0 (0)
DR4-D (n = 27)	30/70	85% white, 15% other (4% Hispanic)	27 (100)	0 (0)	0 (0)	0 (0)
DQ6-D (n = 9)	62/38	78% white, 22% other (10% Hispanic)	6 (67)	0 (0)	9 (100)	0 (0)

Data are n (%) unless otherwise indicated. NR, no-risk HLAs; C, control subjects; D, type 1 diabetic patients.

primer 5'TCCCAACACAGACAGGAATACG3', reverse primer 5'TTCCCTCCCACAAGACAGTGA3', and probe 5'CAGCCTGTGCCCTGG3'.

HLA Typing

HLA-DRB1 and DQB1 were determined using real-time PCR, as previously described (21,22). We used previously published primers (21,22) for all alleles except DQB1*02:01: forward primer 5'GGGCATGTGCTACTTCACCAA3', reverse primer 5'ACGATCTCTTCTCGGTTATAGATGC3', and MGB probe 5'TGCGTCTTGTGAGCAGA3'.

Statistical Analysis

For simplicity, we used nonparametric tests where possible (Wilcoxon rank sum, Spearman correlation test). Throughout this study, multiple testing corrections were unnecessary given the small number of independent tests. Tests of C4 variants are not independent from each other (total C4 = C4A CN + C4B CN = C4L CN + C4S CN); additionally, C4A and C4B contain HERV-K(C4) at different frequencies (95% and 54%, respectively). Note that because of the slightly better correlation of the C4A CN + C4B CN estimate to known total CN, we use it in relevant significance tests and figures (Supplementary Fig. 1).

RESULTS

Type 1 Diabetes Is Associated With Fewer Copies of HERV-K(C4)

We profiled C4 in a cohort of 50 type 1 diabetic patients and 57 control subjects randomly selected from our sample repository (age at diagnosis: median 11.8 years, mean 18.5, SD 14.7) (Table 1). Patients had significantly fewer copies of total C4, C4A, and HERV-K(C4) (i.e., C4L) (Fig. 2A). The most significant difference was in HERV-K(C4) CN, with most patients having two fewer copies than healthy control subjects. This HERV-K(C4) deficiency is not surprising given previous observations of C4A deficiency in type 1 diabetes and that 95% of C4A genes have the HERV-K(C4) insertion (11–15). However, HERV-K(C4) is the variant with the greatest difference in median CN and greatest significance (2.3 copies, $P = 4.59 \times 10^{7}$), whereas the difference in C4A is much smaller and less significant (1.1 copies, $P = 1.76 \times 10^{5}$).

HERV-K(C4) CN Is Correlated With DRB1*03:01/ DQB1*02:01 and DRB1*15:01/DQB1*06:02

Linkage between specific HLA haplotypes and certain RCCX formations has been previously described (8). We found strong relationships between HERV-K(C4) CN and the DQB1*02:01 and DQB1*06:02 alleles and no relationship with DQB1*03:02 (Fig. 2B). Simple linear regression suggested that each DQB1*02:01 allele an individual possesses is associated with ~1.5 fewer copies of HERV-K(C4), whereas each allele of DQB1*06:02 is associated with ~1.5 additional copies of HERV-K(C4). DQB1*02:01 also has significant associations with total C4, C4A, and C4S CNs, which may explain why both C4A deficiency and DQB1*02:01 are associated with type 1 diabetes and SLE (Supplementary Fig. 2). Although DQB1*06:02 is strongly associated with HERV-K(C4), it lacks any association with C4A or C4B and is only weakly associated with total C4 and C4S (Supplementary Fig. 2). HERV-K(C4) CN captures the association of type 1 diabetes with both DQB1*02:01 and *06:02, suggesting that the retroviral insertion rather than total C4 or C4A is most strongly linked to type 1 diabetes.
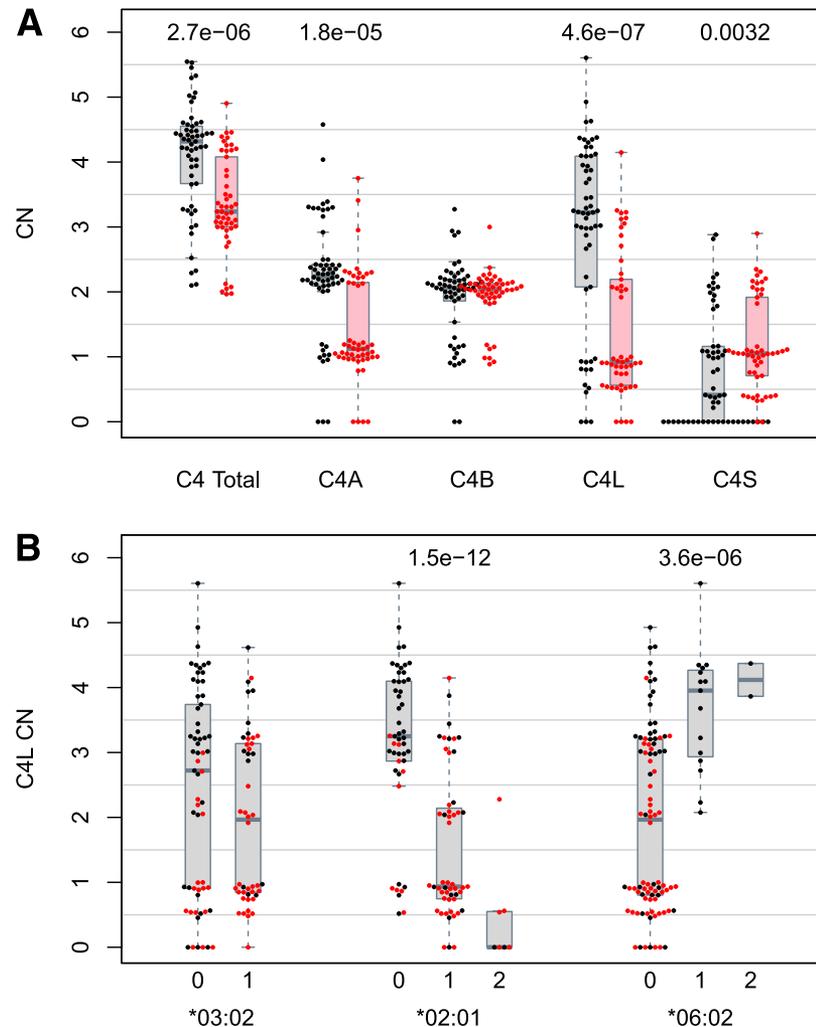
Figure 2—C4 variant CN related to type 1 diabetes disease status and associated MHC class II alleles in 57 control subjects and 50 type 1 diabetic patients. *A*: CN estimates of total C4, C4A, C4B, C4L [HERV-K(C4)], and C4S for control subjects (black) and type 1 diabetic patients (red). Significance was determined by Wilcoxon rank sum test; only significant *P* values are shown. *B*: Estimates of C4L CN related to number of HLA-DBQ1 alleles for control subjects (black) and type 1 diabetic patients (red). No individuals were homozygous for DQB1*03:02. Simple linear regression with C4L CN estimate as the response variable was used to determine significance; only significant *P* values are shown.

The HERV-K(C4) CN–Type 1 Diabetes Association Is Not Simply Due to Linkage With MHC II Genes

Only 5 of 50 originally studied type 1 diabetes samples and 15 of 57 control samples lacked disease-associated alleles. Therefore, we identified an additional 17 type 1 diabetes and 33 control samples lacking these alleles and assessed their C4 variation (Table 1). Even in the absence of these alleles, patients still had significantly fewer copies of HERV-K(C4) (Fig. 3A). Of note, no significant association with total C4, C4A, or C4B CNs was found, supporting the notion that the length variant is most strongly associated with disease. Furthermore, the significant difference in HERV-K(C4) CN in samples lacking these alleles suggests that DQB1*02:01 and *06:02 are associated with type 1 diabetes through linkage with the RCCX region.

HERV-K(C4) CN Differs in DQB1*06:02-Positive Type 1 Diabetic Patients and Control Subjects

DQB1*06:02 is exceedingly rare in type 1 diabetic patients but relatively common in healthy individuals (frequencies of 0.4% and 12%, respectively) (5). Of the 50 type 1 diabetic patients first studied, none were DQB1*06:02 positive. We identified nine additional patient samples positive for the DQB1*06:02 allele and compared their HERV-K(C4) CNs to 18 *06:02-positive control subjects (Table 1 and Fig. 3B). These patients had significantly fewer copies of the retroviral insertion, even though *06:02 is associated with an increase in HERV-K(C4) CN (Fig. 2B). This finding suggests that increased HERV-K(C4) CN and not DQB1*06:02 is protective against type 1 diabetes.
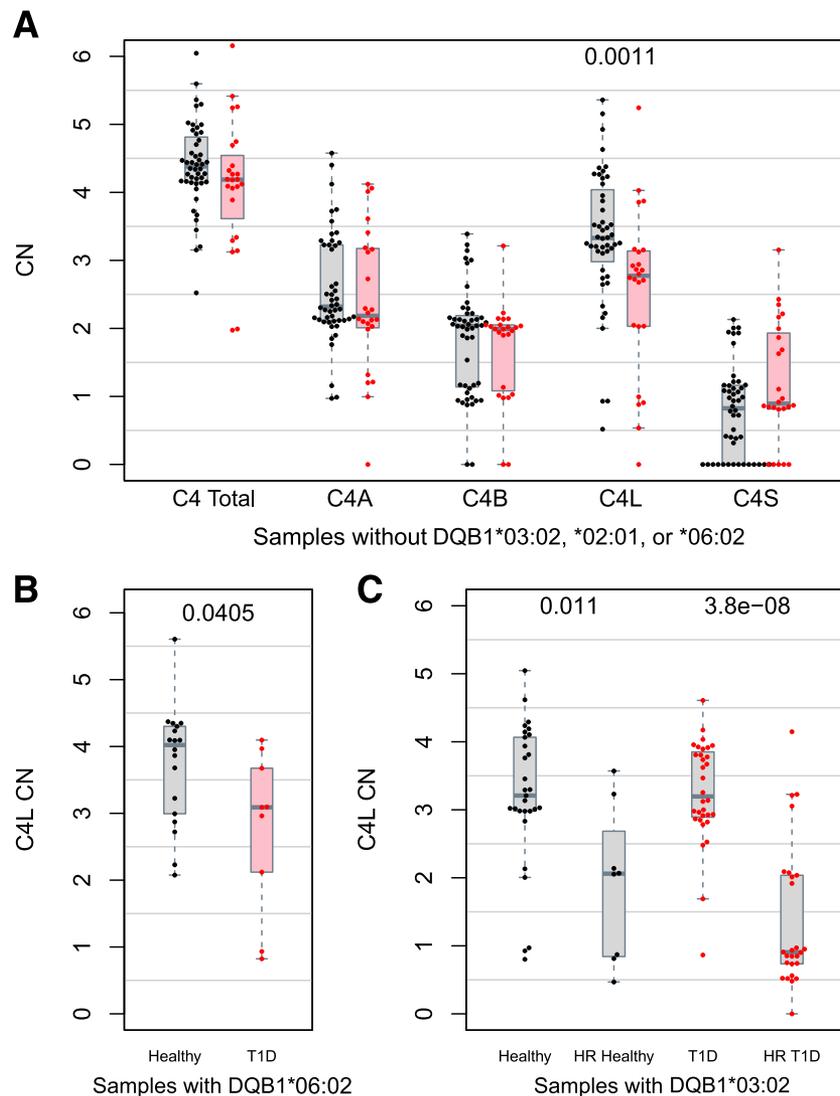
Figure 3—C4 variant CN related to type 1 diabetes disease status in patients lacking or having type 1 diabetes–associated MHC class II alleles. *A*: CN estimates of total C4, C4A, C4B, C4L [HERV-K(C4)], and C4S for control subjects (black) and type 1 diabetic patients (red) lacking DQB1*03:02, *02:01, and *06:02. *B*: Estimates of C4L CN in DQB1*06:02-positive control subjects and type 1 diabetic patients. *C*: Estimates of C4L CN in DQB1*03:02-positive control subjects and type 1 diabetic patients. Patients who are heterozygous with DQB1*02:01 are high risk. Wilcoxon rank sum tests were used to determine significance in all three plots; only significant *P* values are shown. HR, high risk; T1D, type 1 diabetes.

The High-Risk DQB1*03:02/DQB1*02:01 Genotype Is Associated With Fewer Copies of HERV-K(C4)

Figure 2*B* shows that HERV-K(C4) CN is not associated with DQB1*03:02, the HLA allele most strongly linked to type 1 diabetes risk. Furthermore, DQB1*03:02 has significantly different epitope binding avidity in type 1 diabetes (23). No such difference in epitope binding is seen for DQB1*02:01 or *06:02; thus, these haplotypes may be associated with separate mechanisms of action. To study whether high-risk individuals have fewer copies of HERV-K(C4), we identified samples from cohort 1 with DQB1*03:02 but lacking the protective *06:02, resulting in 16 control and 31 type 1 diabetes samples. We identified an additional 24 control and 27 type 1 diabetes samples from cohort 2 (Table 1), giving a total of 40

DQB1*03:02 control and 58 *03:02 patient samples. As expected, high-risk DQB1*03:02/DQB1*02:01 individuals have significantly fewer copies of HERV-K(C4), whereas no association was seen when DQB1*02:01 is absent (Fig. 3*C*). DQB1*03:02 is not linked to HERV-K(C4), suggesting two mechanisms of action in type 1 diabetes: one related to MHC binding avidity and associated with DQB1*03:02 and one unknown mechanism associated with HERV-K(C4) CN. The high risk associated with heterozygosity could be the result of these two mechanisms working in tandem.

DISCUSSION

This study provides evidence for the first time to our knowledge that low HERV-K(C4) CN is associated with

type 1 diabetes. HERV-K(C4) CN alone may account for the relationship between the DRB1*03:01/DQB1*02:01 and HLA-DRB1*15:01/DQB1*06:02 haplotypes and type 1 diabetes risk. Furthermore, this difference in CN is observed both in individuals lacking any disease-associated HLA type and in individuals with the protective DQB1*06:02 allele, suggesting that the association of HERV-K(C4) with type 1 diabetes is not simply due to linkage between HERV-K(C4) and the MHC class II region. This difference in CN is consistently seen with HERV-K(C4) but not for the acidic and basic isoforms or total CN.

The relationship of HERV-K(C4) with type 1 diabetes suggests that disease-associated MHC II DRB1/DQB1 alleles can be split into two groups: those not in linkage with HERV-K(C4) CN like DQB1*03:02, which has been shown to differ significantly in HLA epitope binding avidity (23), and those in linkage with HERV-K(C4) CN like DQB1*02:01 and *06:02. For this latter group, the mechanism of action may not involve MHC II binding but, instead, may act through an unknown function of HERV-K(C4) or some neighboring gene. If HERV-K(C4) plays a functional role in type 1 diabetes, one would expect its CN to correlate with transcription level, which we verified through quantitative PCR (Supplementary Fig. 3).

We profiled >200 samples; however, the sample size is small when investigating specific HLA types, especially for DQB1*06:02. Additional studies are needed to replicate these preliminary findings and to determine any functional role of HERV-K(C4) in type 1 diabetes. As an intronic insertion, HERV-K(C4) may influence C4 RNA processing, ultimately affecting the complement pathway. Alternatively, it may have an impact on other endogenous or exogenous viruses by affecting pathways that sense viral RNA/DNA or through antisense inhibition of a homologous virus (19). These findings have potential implications for other HLA-DQB1*02:01– and DQB1*06:02–linked autoimmune diseases, such as celiac disease, Graves disease, and Hashimoto thyroiditis (24,25), and offer a novel avenue of research in such diseases.

References

1. Nepom GT. The major histocompatibility complex. In *Harrison's Principles of Internal Medicine*. 18th ed. Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J, Eds. New York, McGraw-Hill Medical Publishing Division, 2012, p. 2685–2694

2. Pociot F, Akolkar B, Concannon P, et al. Genetics of type 1 diabetes: what's next? Diabetes 2010;59:1561–1571

3. Sheehy MJ, Scharf SJ, Rowe JR, et al. A diabetes-susceptible HLA haplotype is best defined by a combination of HLA-DR and -DQ alleles. J Clin Invest 1989;83:830–835

4. Nepom GT, Erlich H. MHC class-II molecules and autoimmunity. Annu Rev Immunol 1991;9:493–525

5. Erlich H, Valdes AM, Noble J, et al.; Type 1 Diabetes Genetics Consortium. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes 2008;57:1084–1092

6. Raum D, Awdeh Z, Yunis EJ, Alper CA, Gabbay KH. Extended major histocompatibility complex haplotypes in type I diabetes mellitus. J Clin Invest 1984;74:449–454

7. Price P, Witt C, Allcock R, et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol Rev 1999;167:257–274

8. Bánlaki Z, Doleschall M, Rajczy K, Fust G, Szilágyi A. Fine-tuned characterization of RCCX copy number variants and their relationship with extended MHC haplotypes. Genes Immun 2012;13:530–535

9. Barba G, Rittner C, Schneider PM. Genetic basis of human complement C4A deficiency. Detection of a point mutation leading to nonexpression. J Clin Invest 1993;91:1681–1686

10. Marcelli-Barge A, Poirier JC, Chantome R, Deschamps I, Hors J, Colombani J. Marked shortage of C4B DNA polymorphism among insulin-dependent diabetic patients. Res Immunol 1990;141:117–128

11. Jenhani F, Bardi R, Gorgi Y, Ayed K, Jeddi M. C4 polymorphism in multiplex families with insulin dependent diabetes in the Tunisian population: standard C4 typing methods and RFLP analysis. J Autoimmun 1992;5:149–160

12. Bertrams J, Hintzen U, Schlicht V, et al. Gene and haplotype frequencies of the fourth component of complement (C4) in type 1 diabetics and normal controls. Immunobiology 1984;166:335–344

13. McCluskey J, McCann VJ, Kay PH, et al. HLA and complement allotypes in Type 1 (insulin-dependent) diabetes. Diabetologia 1983;24:162–165

14. Tassabehji M, Strachan T, Anderson M, Campbell RD, Collier S, Lako M. Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. Nucleic Acids Res 1994;22:5211–5217

15. Wouters D, van Schouwenburg P, van der Horst A, et al. High-throughput analysis of the C4 polymorphism by a combination of MLPA and isotype-specific ELISA's. Mol Immunol 2009;46:592–600

16. Hohn O, Hanke K, Bannert N. HERV-K(HML-2), the best preserved family of HERVs: endogenization, expression, and implications in health and disease. Front Oncol 2013;3:246

17. Shin W, Lee J, Son S-Y, Ahn K, Kim H-S, Han K. Human-specific HERV-K insertion causes genomic variations in the human genome. PLoS ONE 2013;8:e60605

18. Mack M, Bender K, Schneider PM. Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region. Immunogenetics 2004;56:321–332

19. Schneider PM, Witzel-Schlömp K, Rittner C, Zhang L. The endogenous retroviral insertion in the human complement C4 gene modulates the expression of homologous genes by antisense inhibition. Immunogenetics 2001;53:1–9

20. Wu YL, Savelli SL, Yang Y, et al. Sensitive and specific real-time polymerase chain reaction assays to accurately determine copy number variations (CNVs) of human complement C4A, C4B, C4-long, C4-short, and RCCX modules: elucidation of C4 CNVs in 50 consanguineous subjects with defined HLA genotypes. J Immunol 2007;179:3012–3025

21. Gersuk VH, Nepom GT. A real-time polymerase chain reaction assay for the rapid identification of the autoimmune disease-associated allele HLA-DQB1*0602. Tissue Antigens 2009;73:335–340

22. Gersuk VH, Nepom GT. A real-time PCR approach for rapid high resolution subtyping of HLA-DRB1*04. J Immunol Methods 2006;317:64–70

23. Kwok WW, Schwarz D, Nepom BS, Hock RA, Thurtle PS, Nepom GT. HLA-DQ molecules form alpha-beta heterodimers of mixed allotype. J Immunol 1988;141:3123–3127

24. Alper CA, Awdeh ZL, Yunis EJ. Complotypes, extended haplotypes, male segregation distortion, and disease markers. Hum Immunol 1986;15:366–373

25. Skanes VM, Barnard J, Farid N, et al. Class III alleles and high-risk MHC haplotypes in type I diabetes mellitus, Graves' disease and Hashimoto's thyroiditis. Mol Biol Med 1986;3:143–157