Brief Genetics Report

Multi-SNP Analysis of MHC Region

Remarkable Conservation of HLA-A1-B8-DR3 Haplotype

Theresa A. Aly,[1] Elise Eller,[1] Akane Ide,[1] Katherine Gowan,[2] Sunanda R. Babu,[1] Henry A. Erlich,[3] Marian J. Rewers,[4] George S. Eisenbarth,[1] and Pamela R. Fain[1,2]

Technology has become available to cost-effectively analyze thousands of single nucleotide polymorphisms (SNPs). We recently confirmed by genotyping a small series of class I alleles and microsatellite markers that the extended haplotype HLA-A1-B8-DR3 (8.1 AH) at the major histocompatibility complex (MHC) is a common and conserved haplotype. To further evaluate the region of conservation of the DR3 haplotypes, we genotyped 31 8.1 AHs and 29 other DR3 haplotypes with a panel of 656 SNPs spanning 4.8 Mb in the MHC region. This multi-SNP evaluation revealed a 2.9-Mb region that was essentially invariable for all 31 8.1 AHs. The 31 8.1 AHs were >99.9% identical for 384 consecutive SNPs of the 656 SNPs analyzed. Future association studies of MHC-linked susceptibility to type 1 diabetes will need to account for the extensive conservation of the 8.1 AH, since individuals who carry this haplotype provide no information about the differential effects of the alleles that are present on this haplotype. *Diabetes* 55:1265–1269, 2006

More than 20 years ago, analysis of polymorphisms of complement genes, such as the 21-hydroxylase gene and alleles of class I and II major histocompatibility complex (MHC) genes, identified a number of MHC haplotypes that were termed "conserved extended haplotypes" or "ancestral haplotypes" (1–4). A Basque haplotype studied in U.S. French-Canadian populations with the HLA A30, Cw5, B18, BfF1, C4F, C4s°, DR3 haplotype was associated with diabetes susceptibility, and more recent studies have confirmed increased risk associated with this haplotype (5). The HLA-A1-B8-DR3 haplotype (8.1 AH) is one of the most common extended DR3 haplotypes, with a northern-European frequency of ~10%. The 8.1 AH consists of the HLA-A1, HLA-Cw7, HLA-B8, MICA-5.1, DR3, and DQ2 alleles (6) and has been sequenced by Stewart et al. (7). It has been associated with multiple immunological diseases, such as type 1 diabetes, celiac disease, systemic lupus erythematosus, common variable immunodeficiency, myasthenia gravis, and accelerated HIV disease (8–10). However, a recent study has reported that the 8.1 AH is not more diabetogenic compared with other DR3 haplotypes (6).

With the sequencing of the genome, the development of single nucleotide polymorphism (SNP) databases and haplotype maps, software to analyze haplotype blocks, and finally development of cost-effective large-scale SNP typing reagents, detailed multi-SNP analysis of the MHC region is now feasible. In this study, the analysis of multiple SNPs was necessary to describe the amount of conservation of the 8.1 AH or the proportion of alleles that were identical between 8.1 AHs. The basic scientific question we explored was how long and how conserved is the extended 8.1 AH? With our analysis of 656 SNPs of 31 8.1 AHs, we describe the 8.1 AH as having a remarkably long region (2.9 Mb) of >99.9% conservation.

RESEARCH DESIGN AND METHODS

In the ongoing prospective Diabetes Autoimmunity Study of the Young (DAISY), participants (most with Caucasian and Hispanic ancestry) were HLA typed and stratified into groups by family history of type 1A diabetes. Subgroups of DAISY children were enrolled for prospective follow-up of development of anti-islet autoantibodies and diabetes. DNA samples from DAISY families, including children and parents with and without type 1 diabetes, were genotyped at the HLA-A, -B, -DRB1, and -DQB1 loci with sequence-specific oligonucleotide genotyping as previously described (11). The MICA microsatellite marker was genotyped using fluorescence-based methods as previously described (12). DNA from 45 of these families (143 individuals) was analyzed with Illumina multiplex technology. Twenty of the individuals analyzed in these 45 families had type 1 diabetes, 10 additional nondiabetic individuals were persistently positive for anti-islet autoantibodies, and 117 were unaffected.

Selection of SNPs. First, all SNPs located in or within 6 kb of the coding regions of high-priority genes with a known or suspected autoimmune function were selected. For genes of lower priority, a single representative SNP was selected. In addition, all coding SNPs, regardless of gene priority, were included. Finally, SNPs were selected to break down any interval >30 kb. All chosen SNPs had minor allele frequencies of at least 0.10 and were validated with at least double-hit validation. A total of 656 SNPs spanning 4.8 Mb in the MHC region were included and successfully genotyped by Illumina (Fig. 1). The mean inter-SNP interval size was 6,309 bp (range 61–29,937 bp). Forty-nine percent of the intervals between adjacent SNPs were <2,000 bp.

Statistical analysis. SNP results for 13 homozygous DR3-DQ2 individuals and 11 homozygous DR4-DQ8 individuals were isolated, and all heterozygous loci were highlighted to illustrate regions of lower conservation. The Illumina genotype results were processed with the program PedCheck to assure that there was a Mendelian pattern of genotype inheritance for each family, and

Chromosome 6p

Centromere — Location of SNPs Genotyped — Telomere

Genes — DP — DQ DR — MICA B C — A

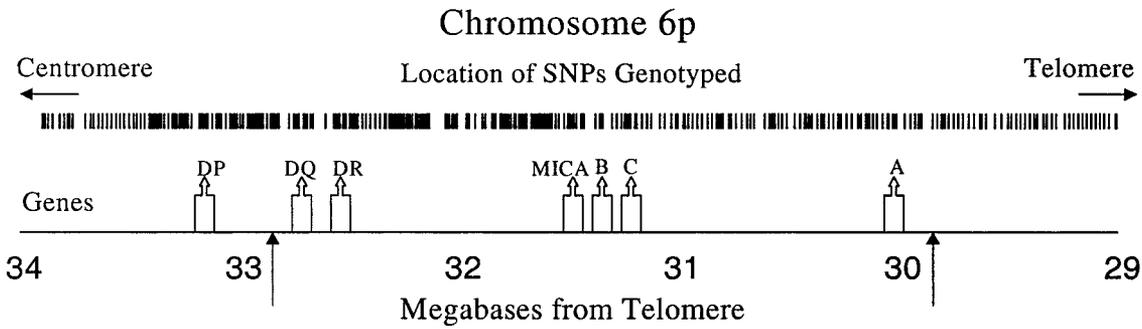34 — 33 — 32 — 31 — 30 — 29

Megabases from Telomere

FIG. 1. Each bar represents the location of each of the 656 SNPs genotyped in relation to representative genes in the MHC gene and distance (Mb) from the telomere. The arrows represent the centromeric and telomeric ends of the remarkable region of conservation.

Merlin was used to determine the phase of the haplotypes. DR3 haplotypes ($n = 60$) from the parents were stratified into the 8.1 AH group ($n = 31$); the HLA-B8-DR3, non-A1 group ($n = 16$); and the HLA-DR3, non-B8 group ($n = 13$). The 60 DR3 haplotypes were analyzed to determine the allelic frequencies at each of the 656 SNPs, and a consensus sequence of the more common or major alleles was established. Minor alleles along each individual DR3 haplotype were highlighted. Major alleles for this population and alleles that were not called due to ambiguities or unknown phase were not highlighted.

RESULTS

Five replicate DNA samples were genotyped for the 656 SNPs to assess the reproducibility of SNP allele calls, including one blind replicate that was not included in routine error screening analysis by Illumina. All called SNPs were identical for all five pairs of replicate samples with a reproducibility of 100%.

Of 13 DR3-DQ2/DR3-DQ2 individuals evaluated for homozygosity at the SNP loci, three were homozygous at HLA-B (B8/B8) and HLA-A (A1/A1), inheriting the 8.1 AH. These three individuals (left three columns in Fig. 2) were homozygous for 356 consecutive SNPs, without exception, spanning 2.9 Mb from rs362536 to rs3135391 (from nucleotide 29,634,918 to 32,518,964). A shorter but still dramatic region of conservation was present for the DR3-DQ2 homozygous individuals in whom one or more haplotype lacked HLA-A1. DR3-DQ2 homozygous individuals with a haplotype without HLA-B8 lacked the large region of
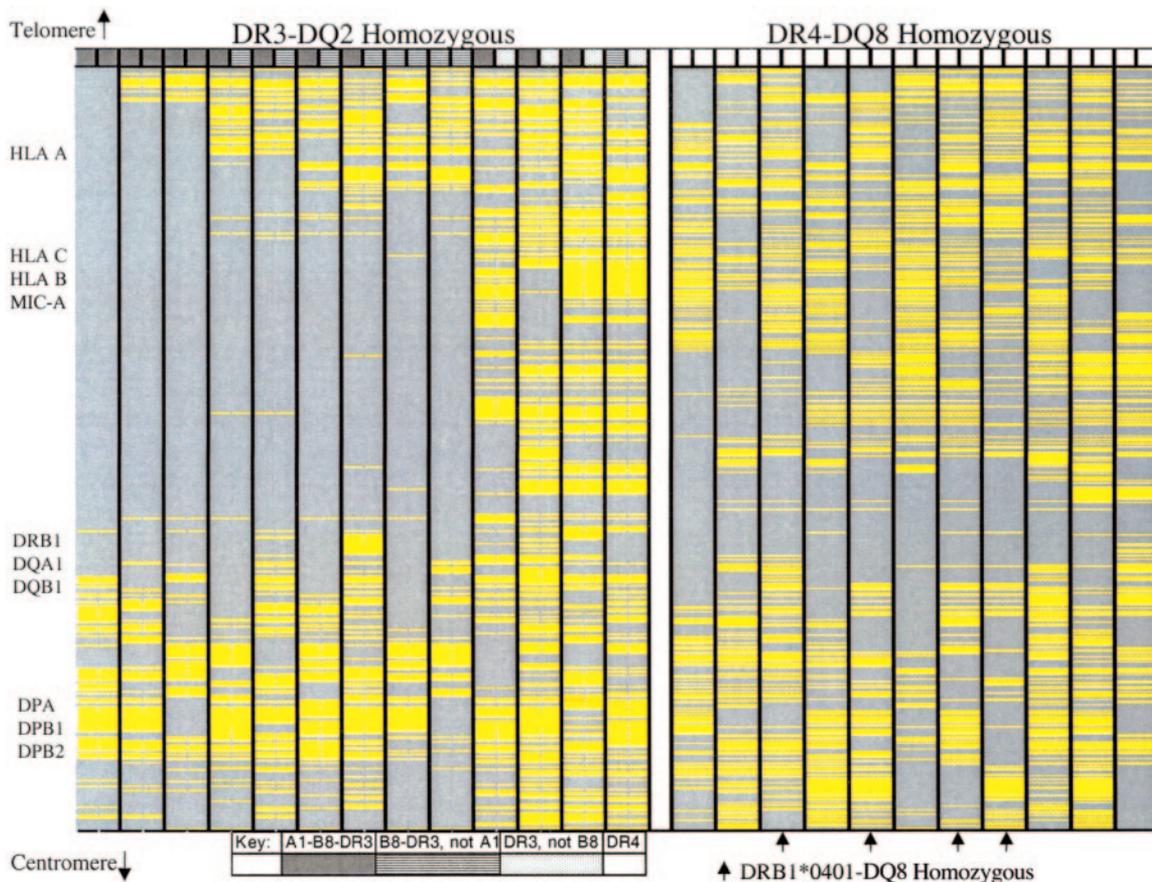

Telomere — DR3-DQ2 Homozygous — DR4-DQ8 Homozygous

HLA A

HLA C
HLA B
MIC-A

DRB1
DQA1
DQB1

DPA
DPB1
DPB2

Key: A1-B8-DR3 | B8-DR3, not A1 | DR3, not B8 | DR4

Centromere

▲ DRB1*0401-DQ8 Homozygous

FIG. 2. DR3/DR3 homozygotes ($n = 13$) are shown on the *left* and DR4/DR4 homozygotes ($n = 11$) on the *right*, with HLA-A and -B alleles identified at the *top of each column* (see key). The arrows below the columns indicate individuals that are homozygous for DRB1*0401-DQ8. The length of each column spans the 4.8-Mb MHC region evaluated with the 656 SNPs. Each of the 656 evenly spaced rows represents one SNP locus. Highlighted rows in each column represent heterozygous genotypes.
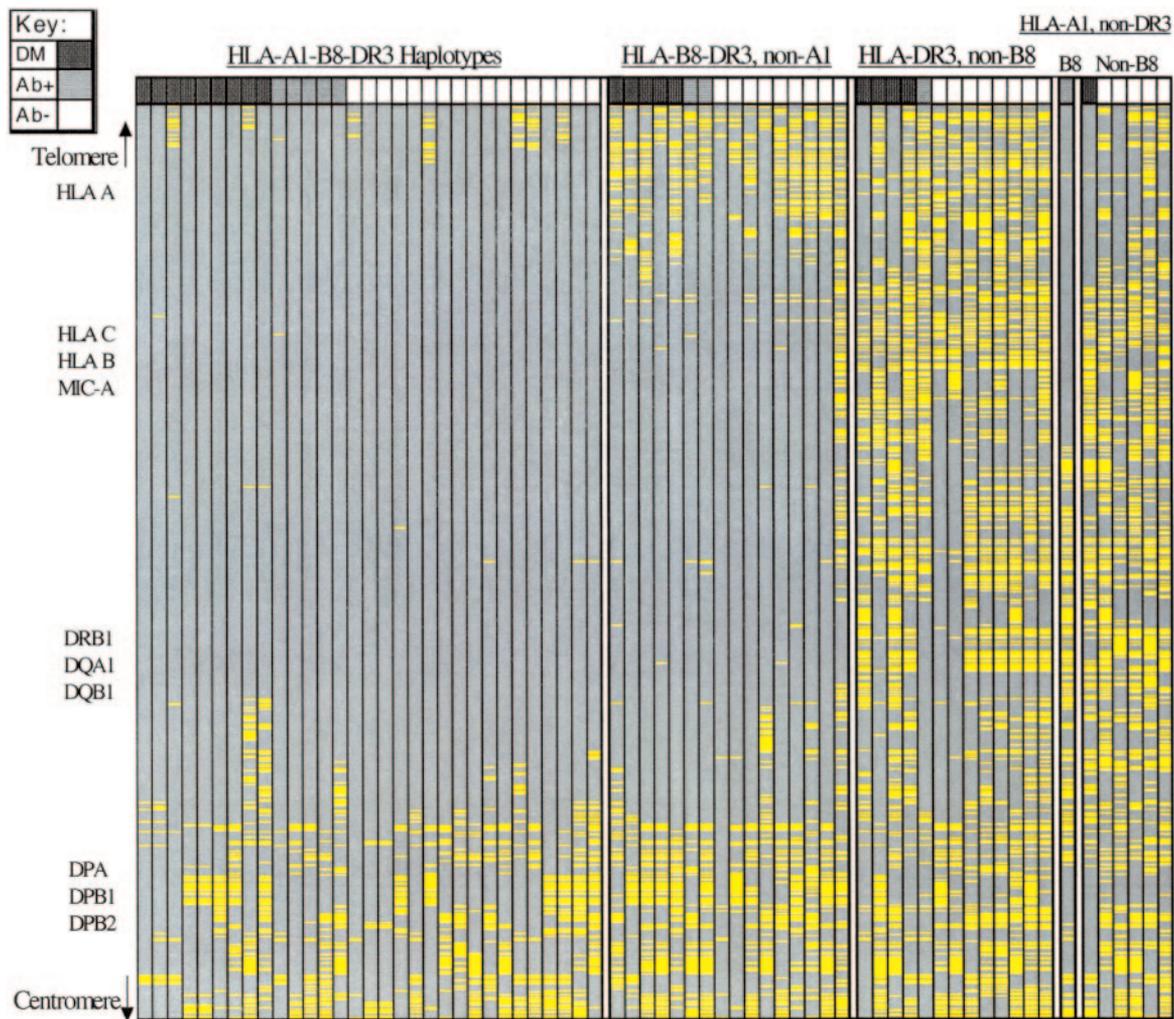
FIG. 3. The *left three groups* depict SNP results from all DR3 haplotypes (*n* = 60) stratified by 8.1 AH haplotypes (*n* = 31); HLA-B8-DR3, non-A1 haplotypes (*n* = 16); and HLA-DR3, non-B8 haplotypes (*n* = 13), substratified by diabetic and anti-islet autoantibody status. The *two panels to the right* depict SNP results from non-DR3 haplotypes (*n* = 7) stratified by an HLA-A1-B8-DR4 haplotype (*n* = 1) and by HLA-A1-non-DR3, non-B8 haplotypes (*n* = 6). The lower frequency allele (row) for each SNP along each haplotype column is highlighted.

conservation. Four DRB1*0401-DQ8 homozygous individuals had a short region of much less conservation surrounding the class II loci compared with the 8.1 AH homozygotes, similar to the remaining seven DR4-DQ8 individuals (Fig. 2).

Of the total 60 DR3 haplotypes from unrelated individuals, the HLA-A1 and -B8 alleles were present on 31 haplotypes. Similar to the analysis of the 8.1 AH homozygotes, the total group of 8.1 AHs had identical alleles for 98% of 384 consecutive SNPs (378 of 384), extending from rs1611165 to rs11759565 and defining a 2.9-Mb region of conservation from nucleotide 29,900,092 to 32,825,923, ~100 kb telomeric of HLA-A to 165 kb centromeric of DRB1 (arrows in Fig. 1). This region was >99.9% conserved, with only 9 variant alleles of the 10,768 alleles identified for the 384 SNPs in the 31 8.1 AHs [(10,768−10,769)/10,768 = 99.9%]. The conserved region stretched to the telomeric limit of the 4.8-Mb HLA region analyzed for 23 of the 31 haplotypes. For the entire MHC panel of 656 SNPs, the 31 8.1 AHs were significantly more conserved than the 29 other DR3 haplotypes analyzed. (Minor to major alleles in 8.1 AHs was 1,146 of 17,124 vs. 5,228 of 16,285 in the other DR3 haplotypes, χ^2 = 2,385, *P* < 0.0001.)

The group of HLA-B8-DR3, non-A1 haplotypes generally had a smaller region of conservation than the 8.1 AHs, with one of the haplotypes having much more variability (right column in the HLA-B8-DR3, non-A1 panel of Fig. 3). In contrast, no extended region of conservation was found with analysis of haplotypes with A1 but without DR3 alleles, although a non-DR3 (DR4) HLA-A1-B8 haplotype had a conserved region surrounding the HLA-A1-B8 loci (seven right columns in Fig. 3). DP alleles, located 33.1 Mb from the telomere, were ~325 kb outside of the region of extensive conservation of the 8.1 AH (Fig. 3), consistent with reports of one or more recombination hotspots centromeric to DQB1 (13,14).

After stratifying the 60 DR3 haplotypes by affected status (26 diabetic and anti–islet autoantibody–positive haplotypes vs. 34 unaffected haplotypes), none of the SNPs were associated with diabetic autoimmunity (Fig. 3). There was no difference in the conserved 8.1 AHs between the nine diabetic haplotypes, the five nondiabetic autoantibody-positive haplotypes, and the 17 autoantibody-negative haplotypes. The alleles for each haplotype of each of the 147 individuals analyzed for this study are available in the online appendix (available at http://diabetes.diabetes journals.org).

DISCUSSION

Our data show that 8.1 AH haplotypes have >99.9% conservation of alleles spanning a 2.9-Mb region of the MHC (equating to <0.1% allelic diversity). The conserved region reaches the telomeric limits of our SNP map for the majority of these haplotypes. The current results are based on three A1-B8-DR3 homozygotes together with 25 other A1-B8-DR3 haplotypes that were reconstructed from family data.

In contrast to SNPs, short tandem repeats (STRs) generally have a higher mutation rate. Thus, when Vorchevsky et al. (15) analyzed a 1.5-Mb region between the RING3 and HLA-B genes of the 8.1 AH with 23 STRs, they found an allelic diversity of 1.9%, suggesting low allelic variability at the STR loci, though not as low as the remarkable conservation of SNPs observed in the current study (1.9 vs. <0.1% allelic diversity). Malkki et al. (16) found extensive diversity for microsatellite alleles on the 8.1 AH in a study in which haplotype frequencies were determined for unrelated individuals rather than from family data (16). Their index of "haplotype-specific heterozygosity" was >0.20 for 4 of 12 microsatellites located between HLA-A and HLA-DQB1. Using the same index as a measure of SNP diversity for this region, we found that only 1 of 384 consecutive SNPs had a haplotype-specific heterozygosity >0.2. These differences could reflect imprecision in the estimation of haplotype frequencies from unrelated individuals, a higher mutation rate for microsatellites, or demographic differences in the populations surveyed.

The factors that lead to the extended conservation of the 8.1 AH are currently unknown, but this extended conservation could be due to natural selection, recombination suppression, or demographic factors such as population bottlenecks, genetic drift, or migration and admixture (17–20). The high frequency and extended length of the haplotype are key characteristics of recent positive selection, in which the frequency of a selectively favored allele increases rapidly over a period too short for the surrounding haplotype to become disrupted by recombination (21). The most credible example of recent positive selection is a common extended (~1 Mb) haplotype carrying alleles associated with lactase persistence, a phenotype that may have become advantageous with the relatively recent introduction of dairy farming (22). Interestingly, the 8.1 AH was not detected in a recent global analysis of long-range linkage disequilibrium (LD) across the MHC (14). In that study, the only indications of extended LD involved a 540-kb haplotype associated with DR2 (DRB1*1501). Their inability to detect the more extensive conservation of the 8.1 AH emphasizes the need to account for prior evidence for HLA-defined ancestral haplotypes in future population genetic analyses of the MHC.

Recombination suppression is another intriguing hypothesis for explaining the extensive conservation of the 8.1 AH, particularly in light of evidence that long-range LD may not be restricted to common MHC haplotypes (17). The MHC is characterized by remarkable sequence diversity, variable haplotype lengths, and differences in gene organization (23). Sequence diversity and structural differences between homologous chromosomes may disrupt the pairing and alignment that is essential for cross overs to occur (19). Sequence heterology that inhibits crossing over could also explain the observations of greater differences in recombination rates between siblings who share

one MHC haplotype compared with siblings who share two MHC haplotypes (24).

Our results demonstrating the extended length and remarkable conservation of the 8.1 AH have immediate implications for identifying specific genes and alleles that contribute to MHC-linked susceptibility to type 1 diabetes. In particular, our results imply that individuals carrying the 8.1 AH are essentially uninformative for assessing the association of variants that lie within the region of conservation with type 1 diabetes. Furthermore, the extended 8.1 AH may have a confounding effect in allelic association studies, resulting in misleading conclusions about diabetes susceptibility alleles. Data for association studies might be analyzed with and without the inclusion of the 8.1 AHs to assess for confounding effects. Nonetheless, our SNP data are important in providing a means of identifying individuals with recombinant fragments of the 8.1 AH for the identification of specific fragments that show the strongest association with disease (2). TRIMHAP (trimmed haplotype analysis for analyzing portions of extended haplotypes) is a free program written by R.B. Martin that might help with the identification of disease-associated fragments of the 8.1 AH (25). Further investigations in which the 8.1 AH and its recombinant fragments are characterized more definitively will be important for future association studies of MHC-linked susceptibility to type 1 diabetes.

REFERENCES

1. Alper CA, Awdeh Z, Yunis EJ: Conserved, extended MHC haplotypes. *Exp Clin Immunogenet* 9:58–71, 1992
2. Degli-Esposti MA, Abraham LJ, McCann V, Spies T, Christiansen FT, Dawkins RL: Ancestral haplotypes reveal the role of the central MHC in the immunogenetics of IDDM. *Immunogenet* 36:345–356, 1992
3. Dawkins RL, Christiansen FT, Kay PH, Garlepp M, McCluskey J, Hollingsworth PN, Zilko PJ: Disease associations with complotypes, supratypes and haplotypes. *Immunol Rev* 70:1–22, 1983
4. Alper CA, Awdeh ZL, Raum DD, Yunis EJ: Extended major histocompatibility complex haplotypes in man: role of alleles analogous to murine t mutants. *Clin Immunol Immunopathol* 24:276–285, 1982
5. Cambon-de Mouzon A, Ohayon E, Hauptmann G, Sevin A, Abbal M, Sommer E, Vergnes H, Ducos J: HLA-A, B, C, DR antigens, Bf, C4 and glyoxalase I (GLO) polymorphisms in French Basques with insulin-dependent diabetes mellitus (IDDM). *Tissue Antigens* 19:366–379, 1982
6. Ide A, Babu SR, Robles DT, Wang T, Erlich HA, Bugawan TL, Rewers M, Fain PR, Eisenbarth GS: "Extended" A1, B8, DR3 haplotype shows remarkable linkage disequilibrium but is similar to nonextended haplotypes in terms of diabetes risk. *Diabetes* 54:1879–1883, 2005
7. Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, Forbes S, Halls K, Howson JM, Humphray SJ, Hunt S, Mungall AJ, Osoegawa K, Palmer S, Roberts AN, Rogers J, Sims S, Wang Y, Wilming LG, Elliott JF, De Jong PJ, Sawcer S, Todd JA, Trowsdale J, Beck S: Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 14:1176–1187, 2004
8. Valdes AM, Wapelhorst B, Concannon P, Erlich HA, Thomson G, Noble JA: Extended DR3–D6S273-HLA-B haplotypes are associated with increased susceptibility to type 1 diabetes in US Caucasians. *Tissue Antigens* 65:115–119, 2005
9. Price P, Witt C, Allcock R, Sayer D, Garlepp M, Kok CC, French M, Mallal

S, Christiansen F: The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol Rev* 167:257–274, 1999

10. Bilbao JR, Martin-Pagola A, Perez de Nanclares G, Calvo B, Vitoria JC, Vazquez F, Castano L: HLA-DRB1 and MICA in autoimmunity: common associated alleles in autoimmune disorders. *Ann N Y Acad Sci* 1005:314–318, 2003

11. Bugawan TL, Erlich HA: Rapid typing of HLA-DQB1 DNA polymorphism using nonradioactive oligonucleotide probes and amplified DNA. *Immunogenetics* 33:163–170, 1991

12. Park YS, Sanjeevi CB, Robles D, Yu L, Rewers M, Gottlieb PA, Fain P, Eisenbarth GS: Additional association of intra-MHC genes, MICA and D6S273, with Addison's disease. *Tissue Antigens* 60:155–163, 2002

13. Jeffreys AJ, Ritchie A, Neumann R: High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum Mol Genet* 9:725–733, 2000

14. Miretti MM, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, Morrison J, Whittaker P, Lander ES, Cardon LR, Bentley DR, Rioux JD, Beck S, Deloukas P: A high-resolution linkage-disequilibrium map of the major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76:634–646, 2005

15. Vorechovsky I, Kralovicova J, Laycock MD, Webster AD, Marsh SG, Madrigal A, Hammarstrom L: Short tandem repeat (STR) haplotypes in HLA: an integrated 50-kb STR/linkage disequilibrium/gene map between the RING3 and HLA-B genes and identification of STR haplotype diversification in the class III region. *Eur J Hum Genet* 9:590–598, 2001

16. Malkki M, Single R, Carrington M, Thomson G, Petersdorf E: MHC microsatellite diversity and linkage disequilibrium among common HLA-A, HLA-B, DRB1 haplotypes: implications for unrelated donor hematopoietic transplantation and disease association studies. *Tissue Antigens* 66:114–124, 2005

17. Ahmad T, Neville M, Marshall SE, Armuzzi A, Mulcahy-Hawes K, Crawshaw J, Sato H, Ling K, Barnardo M, Goldthorpe S, Walton R, Bunce M, Jewell DP, Welsh KI: Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet* 12:647–656, 2003

18. Trowsdale J: HLA genomics in the third millenium. *Curr Opin Immunol* 17:1–7, 2005

19. Kauppi L, Jeffreys AJ, Keeney S: Where the crossovers are: recombination distributions in mammals. *Nat Rev Genet* 5:413–424, 2004

20. Yoshino M, Sagai T, Lindahl KF, Toyoda Y, Moriwaki K, Shiroishi T: Allele-dependent recombination frequency: homology requirement in meiotic recombination at the hot spot in the mouse major histocompatibility complex. *Genomics* 27:298–305, 1995

21. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837, 2002

22. Bersaglieri T, Sabeti PC, Patterson N, Vanderplog T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschorn JN: Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111–1120, 2004

23. Yunis EJ, Larsen CE, Fernandez-Vina M, Awdeh ZL, Romero T, Hansen JA, Alper CA: Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens* 62:1–20, 2003

24. Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M: High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am J Hum Genet* 71:759–776, 2002

25. MacLean CJ, Martin RB, Sham PC, Wang H, Straub RE, Kendler KS: The trimmed-haplotype test for linkage disequilibrium. *Am J Hum Genet* 66:1062–1075, 2000