Classification of the etiologies of acute liver failure in Japan: A report by the Intractable Hepato-Biliary Diseases Study Group of Japan.	**Mochida S**, Takikawa Y, Nakayama N, Oketani M, Naiki T, Yamagishi Y, Fujiwara K, Ichida T, Tsubouchi H.	Hepatol Res.	2014 in press	国内
A proposal for management of rheumatic disease patients with hepatitis B virus infection receiving immunosuppressive therapy.	Harigai M, **Mochida S**, Mimura T, Koike T, Miyasaka N	Mod Rheumatol.	2014	国内
Cellular senescence checkpoint function determines differential Notch1-dependent oncogenic and tumor suppressor activities.	Kagawa S, **Natsuizaka M**, Whelan KA, Facompre N, Naganuma S, Ohashi S, Kinugasa H, Egloff AM, Basu D, Gimotty PA, Klein-Szanto AJ, Bass A, Wong K, Diehl JA, Rustgi AK, Nakagawa H	Oncogene.	2014 in press	国外
CTNNB1 mutational analysis of solid-pseudopapillary neoplasms of the pancreas using EUS-guided fine-needle aspiration and next-generation deep sequencing.	Kubota Y, Kawakami H, **Natsuizaka M**, Kawakubo K, Marukawa K, Kudo T, Abe Y, Kubo K, Kuwatani M, Hatanaka Y, Mitsuhashi T, Matsuno Y, Sakamoto N	J Gastroenterol	2014 in press	国内
Inhibition of Notch signaling enhances transdifferentiation of the esophageal squamous epithelium towards a Barrett's-like metaplasia via KLF4.	Vega ME, Giroux V, **Natsuizaka M**, Liu M, Klein-Szanto AJ, Stairs DB, Nakagawa H, Wang KK, Wang TC, Lynch JP, K Rustgi AK	Cell Cycle	2014	国外
Mitochondrial SOD2 regulates epithelial-mesenchymal transition and cell populations defined by differential CD44 expression.	Kinugasa H, Whelan KA, Tanaka K, **Natsuizaka M**, Long A, Guo A, Chang S, Kagawa S, Srinivasan S, Guha M, Yamamoto K, St. Clair DK, Avadhani NG, Diehl JA, Nakagawa H	Oncogene.	2014 in press	国外
Serum granulysin levels as a predictor of serious telaprevir-induced dermatological reaction.	Suda G, Yamamoto Y, Nagasaka A, Furuya K, Kudou M, Chuganji Y, Tsukuda Y, Tsunematsu S, Sato G, Terashita K, Nakai M, Horimoto H, Sho T, **Natsuizaka M**, Ogawa K, Ohnishi S, Chuma M, Fujita Y, Abe R, Taniguchi M, Nakagawa M, Asahina Y, Sakamoto N for the NORTE study group	Hepatol Res.	2014 in press	国内
Hepatosplenic Gamma-delta T-cell Lymphoma Associated with Epstein-Barr Virus.	Tsunematsu S, **Natsuizaka M**, Fujita H, Otsuka N, Terashita K, Sato F, Kobayashi T, Nakai M, Tsukuda Y, Horimoto H, Sho T, Suda G, Nakanishi M, Hashino S, Chuma M, Sakamoto N	Intern Med	2014	国外

Intratumoral artery on contrast-enhanced computed tomography imaging: differentiating intrahepatic cholangiocarcinoma from poorly differentiated hepatocellular carcinoma	Tsunematsu S, Chuma M, Kamiyama T, Miyamoto N, Yabusaki S, Hatanaka K, Mitsuhashi T, Kamachi H, Yokoo H, Kakisaka T, Tsuruga Y, Orimo T, Wakayama K, Ito J, Sato F, Terashita K, Nakai M, Tsukuda Y, Sho T, Suda G, Morikawa K, **Natsuizaka M**, Nakanishi M, Ogawa K, Taketomi A, Matsuno Y, Sakamoto N	Abdom Imaging	2015 in press	国外
Comparison of improved prognosis between hepatitis B- and hepatitis C- related hepatocellular carcinoma.	Minami T, Tateishi R, Shiina S, Nakagomi R, Kondo M, Fujiwara N, Mikami S, Sato M, Uchino K, Enooku K, Nakagawa H, Asaoka Y, Kondo Y, Yoshida H, **Koike K.**	Hepatol Res.	2015 in press	国内
Specific delivery of microRNA93 into HBV-replicating hepatocytes downregulates protein expression of liver cancer susceptible gene MICA.	Ohno M, Otsuka M, Kishikawa T, Shibata C, Yoshikawa T, Takata A, Muroyama R, Kowatari N, Sato M, Kato N, Kuroda S, **Koike K.**	Oncotarget.	2014	国外
Potential associations between perihepatic lymph node enlargement and liver fibrosis, hepatocellular injury or hepatocarcinogenesis in chronic hepatitis B virus infection.	Sato M, Hikita H, Hagiwara S, Sato M, Soroida Y, Suzuki A, Gotoh H, Iwai T, Kojima S, Matsuura T, Yotsuyanagi H, **Koike K**, Yatomi Y, Ikeda H.	Hepatol Res.	2014 in press	国内
Multiplication of alpha-fetoprotein and protein induced by vitamin K absence-II is a powerful predictor of prognosis and recurrence in hepatocellular carcinoma patients after a hepatectomy.	Kamiyama T, Yokoo H, Kakisaka T, Orimo T, Wakayama K, Kamachi H, Tsuruga Y, Yamashita K, Shimamura T, Todo S, **Taketomi A.**	Hepatol Res.	2014	国内
Suprabasin as a novel tumor endothelial cell marker.	Alam MT, Nagao-Kitamoto H, Ohga N, Akiyama K, Maishi N, Kawamoto T, Shinohara N, **Taketomi A**, Shindoh M, Hida Y, Hida K.	Cancer Sci.	2014	国外
A rare point mutation in the Ras oncogene in hepatocellular carcinoma.	**Taketomi A**, Shirabe K, Muto J, Yoshiya S, Motomura T, Mano Y, Ikegami T, Yoshizumi T, Sugio K, Maehara Y.	Surg Today.	2013	国外
Heat shock factor 1 accelerates hepatocellular carcinoma development by activating nuclear factor-κB/mitogen-activated protein kinase.	Chuma M, Sakamoto N, Nakai A, Hige S, Nakanishi M, Natsuizaka M, Suda G, Sho T, Hatanaka K, Matsuno Y, Yokoo H, Kamiyama T, **Taketomi A**, Fujii G, Tashiro K, Hikiba Y, Fujimoto M, Asaka M, Maeda S.	Carcinogenesis.	2014	国外

Diacylglycerol kinase δ phosphorylates phosphatidylcholine-specific phospholipase C-dependent, palmitic acid-containing diacylglycerol species in response to high glucose levels.	Sakai H, Kado S, **Taketomi A**, Sakane F.	J Biol Chem.	2014	国外
最新の消化器癌術前術後化学療法】 肝細胞癌	若山 顕治, 神山 俊哉, **武冨 紹信**.	消化器外科	2014	国内
Dysregulation of Retinoic Acid Receptor Diminishes Hepatocyte Permissiveness to Hepatitis B Virus Infection through Modulation of NTCP Expression.	Tsukuda S, Watashi K, Iwamoto M, Suzuki R, Aizaki H, Okada M, Sugiyama M, Kojima S, Tanaka Y, Mizokami M, Li J, Tong S, **Wakita T.**	J Biol Chem.	2014 in press	国外
Formation of covalently closed circular DNA in Hep38.7-Tet cells, a tetracycline inducible hepatitis B virus expression cell line.	Ogura N, Watashi K, Noguchi T, **Wakita T.**	Biochem Biophys Res Commun.	2014	国外
Cyclosporin A and its analogs inhibit hepatitis B virus entry into cultured hepatocytes through targeting a membrane transporter NTCP.	Watashi K, Sluder A, Daito T, Matsunaga S, Ryo A, Nagamori S, Iwamoto M, Nakajima S, Tsukuda S, Borroto-Esoda K, Sugiyama M, Tanaka Y, Kanai Y, Kusuhara H, Mizokami M, **Wakita T.**	Hepatology.	2014	国外

Ⅳ．研究成果の刊行物・別刷

Genes **2014**, *5*, 84-96; doi:10.3390/genes5010084

OPEN ACCESS

genes

ISSN 2073-4425

www.mdpi.com/journal/genes

Review

Lessons from Genome-Wide Search for Disease-Related Genes with Special Reference to HLA-Disease Associations

Katsushi Tokunaga

Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo 113-0013, Japan; E-Mail: tokunaga@m.u-tokyo.ac.jp

Received: 8 January 2014; in revised form: 11 February 2014 / Accepted: 12 February 2014 / Published: 26 February 2014

Abstract: The relationships between diseases and genetic factors are by no means uniform. Single-gene diseases are caused primarily by rare mutations of specific genes. Although each single-gene disease has a low prevalence, there are an estimated 5000 or more such diseases in the world. In contrast, multifactorial diseases are diseases in which both genetic and environmental factors are involved in onset. These include a variety of diseases, such as diabetes and autoimmune diseases, and onset is caused by a range of various environmental factors together with a number of genetic factors. With the astonishing advances in genome analysis technology in recent years and the accumulation of data on human genome variation, there has been a rapid progress in research involving genome-wide searches for genes related to diseases. Many of these studies have led to the recognition of the importance of the human leucocyte antigen (HLA) gene complex. Here, the current state and future challenges of genome-wide exploratory research into variations that are associated with disease susceptibilities and drug/therapy responses are described, mainly with reference to our own experience in this field.

Keywords: genome-wide association study (GWAS); multifactorial disease; disease susceptibility gene; drug response gene; HLA genes

1. Development of Genome-Wide Searches

The greatest attraction of the strategy of genome-wide searches for genes related to diseases is the potential for the discovery of the involvement of completely new genes that could not have been predicted using existing knowledge or data. The previous method for genome-wide search of

multifactorial disease-susceptibility genes was non-parametric linkage analysis, which does not presuppose any specific inheritance mode. One such method is the affected sib-pair method. However, it is not easy to collect a large number of samples with affected sib-pairs, so the detection power of this method is inevitably low [1]. Consequently, only limited results have been obtained so far.

The genome-wide association study (GWAS), however, makes use of the high statistical power of association analysis traditionally used for investigating the possible involvement of specific candidate genes, and applies it genome-wide [1]. Two pioneering GWAS studies were carried out in Japan. One was the first single nucleotide polymorphism (SNP)-based GWAS for myocardial infarction, which utilized an approximately 90,000 SNPs [2]. The other was the first microsatellite-based GWAS for rheumatoid arthritis, which used approximately 30,000 microsatellite polymorphisms [3]. However, only a few research groups adopted either of these platforms, due to the labor and cost they involved.

GWAS advanced to a new stage from 2006 onward, mainly as a result of two developments in infrastructure. The first was information infrastructure, typified by the Database of Single Nucleotide Polymorphisms (dbSNP) [4], the International HapMap Project [5] and the 1000 Genomes Project [6], which gathered together a vast range of information of genome variation that spanned the entire human genome. The other development was in technology infrastructure; this was the commercial release of platforms that allowed the analysis of several thousands of samples performed on several hundreds of thousands of SNPs and could be carried out relatively easily. The application of these developments meant that SNP-based GWAS became a broad-based, practical strategy, and in 2007, several studies were published from large-scale collaborations between multiple institutions. The subsequent rush to discover gene polymorphisms associated with different diseases or traits using GWAS was dramatic, and over 1600 types of significant associations with 250 diseases or traits have been reported [7]. Nevertheless, attention should be paid for GWAS in ethnically diverse populations, since the genome-wide SNP typing chips have been designed based on mainly European data, these chips may have limited utility in certain populations.

2. Identified Susceptibility Genes to Multifactorial Diseases

2.1. Population Differences in Disease Susceptibility Genes

A disease for which GWAS have shown striking results is type II diabetes. In 2007, several groups from Europe and North America reported results from different GWAS on several thousand patients and controls [8–11]. Over 11 susceptibility loci were identified, and over half of these were newly discovered. The following year, two independent groups from Japan reported a new susceptibility gene, *KCNQ1* [12,13]. Table 1 shows a comparison between European and Japanese populations of the allele frequency, odds ratio and *p*-value of *TCF7L2*, the most important susceptibility gene found in European populations, and *KCNQ1*, which was discovered in Japanese. *TCF7L2* showed a *p*-value of 10^{-48} in European populations, indicating a definite association with type II diabetes [8]. Among Japanese, however, the *p*-value is at a level of no more than 10^{-4} [14]. The main reason for this is the difference in minor allele frequency, which is lower in Japanese by an order of magnitude. Consequently, although the odds ratio is similar to European populations, no clear association was observed in an analysis of 2000 patients and 2000 healthy controls. A contrasting relationship can be seen with *KCNQ1* [12]. The

p-value for Japanese samples was 10^{-29}, indicating a definite association with type II diabetes, and the same clear association was found for Korean and Chinese samples. However, although European samples showed the same tendency of the odds ratio, the *p*-value was at a level of no more than 10^{-4}.

Table 1. Population differences of susceptibility genes to type II diabetes.

Gen (SNP)	Population	Odds Ratio	*p*	Minor Allele Frequency
TCF7L2 (rs7903146)	European [8–11]	1.37	1.0×10^{-48}	0.31/0.25
TCF7L2 (rs7903146)	Japanese [14]	1.70	7.0×10^{-4}	0.05/0.02
KCNQ1 (rs2237892)	European [12]	1.29	7.8×10^{-4}	0.03/0.05
KCNQ1 (rs2237892)	Japanese [12]	1.43	3.0×10^{-29}	0.31/0.40

In other words, the main type II diabetes-susceptibility genes for European and East Asian populations, respectively, are, in fact, shared susceptibility genes by both populations, but because they differ greatly in frequency, their contribution in each respective population is different.

Several genetic factors, in addition to environmental factors, such as stress, are involved in the onset of narcolepsy, one of the hypersomnia. In the past, the only gene well established as a genetic factor for narcolepsy was *HLA-DR/DQ* [15]; then, we carried out a GWAS to search for new genetic factors [16]. As a result, an SNP located between *CPT1B* and *CHKB* on Chromosome 22 was found to be associated with narcolepsy. Japanese and Koreans were found to have similar allele frequency and both showed a significant association. However, although the odds ratio showed similar trends in European Americans and African Americans, we could not find a significant difference association, because of the low frequency of the susceptibility allele. We have also experienced significant population differences in other diseases, including tuberculosis [17], rheumatoid arthritis [18], glaucoma [19] and primary biliary cirrhosis [20].

The above diseases serve as examples of different contributions of multiple genetic factors in each population. Consequently, the study of each individual population would be essential to build a complete picture of the important genetic factors to complex diseases in the various human populations.

2.2. Susceptibility Genes Common to Different Diseases

There has been an increase in the number of reports of genetic factors that are common to different diseases. *GPC5* (glypican-5) has been found to be a susceptibility gene common to nephrotic syndrome diseases, such as membranous nephropathy, immunoglobulin A nephropathy and diabetic nephropathy (Table 2) [21]. We further confirmed the expression of the GPC5 protein in the glomerular podocytes and showed that the risk allele is associated with a high level of GPC5 expression.

Table 2. Common susceptibility gene *GPC5* (glypican 5) for acquired nephrotic syndrome [21].

Panel	Case: Minor Allele Frequency	Control: Minor Allele Frequency	*p* *	Odds Ratio
1	0.237	0.167	5.8×10^{-3}	2.33 (1.25–4.35)
2	0.195	0.159	2.0×10^{-5}	3.44 (1.89–6.25)
3	0.224	0.174	8.7×10^{-6}	2.39 (1.61–3.55)
Combined	0.219	0.168	6.0×10^{-11}	2.54 (1.91–3.40)

* Based on the recessive model of the minor allele (GG + GA *vs.* AA).

Meta-analysis of the largest-scale GWAS in Japan on rheumatoid arthritis (RA) led to the discovery of susceptibility genes that are common to various different autoimmune disorders [18]. The GWAS was performed on approximately 4000 patients and 17,000 controls, and a replication study was carried out with 5000 patients and 22,000 controls. In addition to previously reported susceptibility genes, nine new susceptibility genes were discovered. Among these are several susceptibility genes that have been also reported for systemic lupus erythematosus (SLE) and Graves' disease.

Another example in our recent experience was primary biliary cirrhosis [20]. We performed a GWAS by a nation-wide collaboration; as a result, we discovered two new susceptibility genes. Interestingly, one of these, *TNFSF15*, has also been reported as a susceptibility gene for inflammatory bowel disease, including Crohn's disease and ulcerative colitis. There are numerous other reports of genetic factors that are found to be common to various autoimmune and inflammatory diseases [22,23].

The presence of common susceptibility genes for different diseases suggests that at least part of the pathogenic mechanism of these diseases is shared. These results may contribute to the elucidation of the pathogenic mechanism of these diseases and to the development of new therapies.

2.3. Towards the Understanding of Pathogenic Mechanisms

As mentioned earlier, the new narcolepsy-susceptibility region, *CPT1B/CHKB*, was discovered through a GWAS performed to search for genetic factors other than the established factor, *HLA* [16]. Subjects possessing the risk allele of the susceptibility SNP showed significantly lower levels of mRNA expression of both *CPT1B* and *CHKB*. We also observed that narcolepsy patients show abnormally low levels of carnitine [24], on which *CPT1B* (carnitine palmitoyltransferase 1B) is relevant, and that carnitine improves the sleep of the patients [25]. Carnitine is known as the transporter of long-chain fatty acids into mitochondria, thus playing a crucial role in energy production.

Moreover, the new susceptibility gene, *TRA* (T cell receptor α), was discovered through a GWAS performed by a joint international research group [26]. SNPs located in the J region of *TRA* showed significant associations with narcolepsy in European and Asian populations. *TRA* and *HLA* are key molecules in the regulation of immune response in the acquired immunity. The same joint international research group also found that a polymorphism of *P2RY11*, which is also involved in the regulation of the immune system, is associated with narcolepsy [27]. From these results, it may be assumed that narcolepsy onset has at least two mechanisms: both autoimmunity to orexin (hypocretin)-producing cells and a disorder of fatty acid β-oxidation.

If we appreciate that multiple susceptibility genes that have been discovered belong to specific pathways or networks, they will provide useful hints toward clarifying the mechanism of disease onset or disease formation and also developing new drugs.

3. Identified Response Genes to Drugs/Therapies

3.1. Development of New Gene Tests

GWAS studies are extremely useful in the search for drug-response genes. We performed a GWAS as part of a multi-institutional joint research group investigating hepatitis C virus related diseases. As a result of this GWAS, we discovered that *IL28B* on Chromosome 19 was strongly associated with

non-responder patients to the combined therapy of PEGylated interferon-alpha and ribavirin for chronic hepatitis C [28]. This was a completely unexpected result. The GWAS was performed on only 78 non-responders and 64 responders to this therapy; nevertheless a p-value at the level of 10^{-12} was obtained, reaching the genome-wide significance level (Figure 1). About 70%–80% of the non-responding patients possessed the minor alleles of several SNPs in the *IL28B* region, and combining the replication study data, the p-value was 10^{-27}–10^{-32} and the odds ratio was 17–30 (Figure 2).

Figure 1. A genome-wide association study (GWAS) on the response to the combined therapy of PEGylated interferon-alpha and ribavirin for chronic hepatitis C identified two SNPs on Chromosome 19 [28].

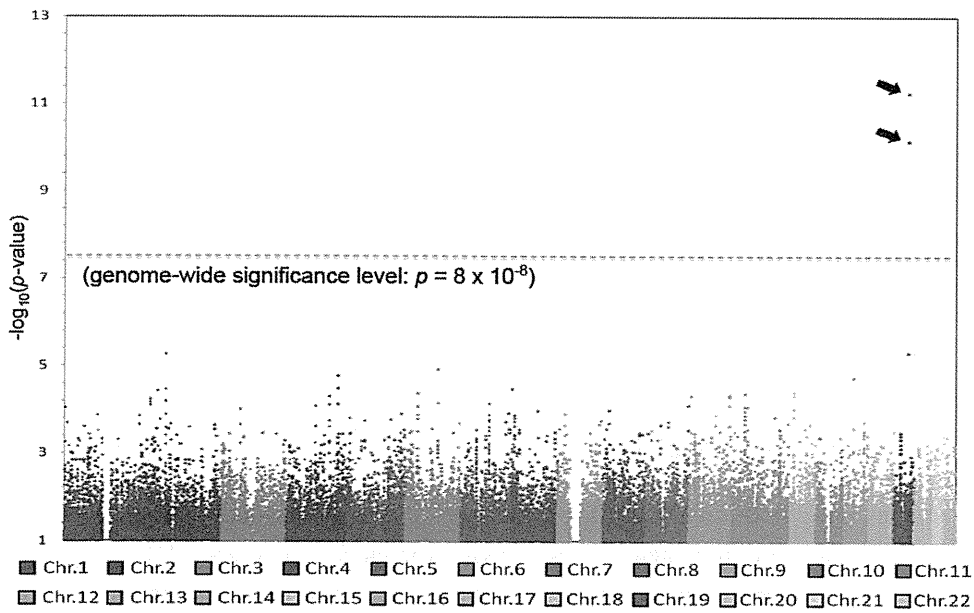
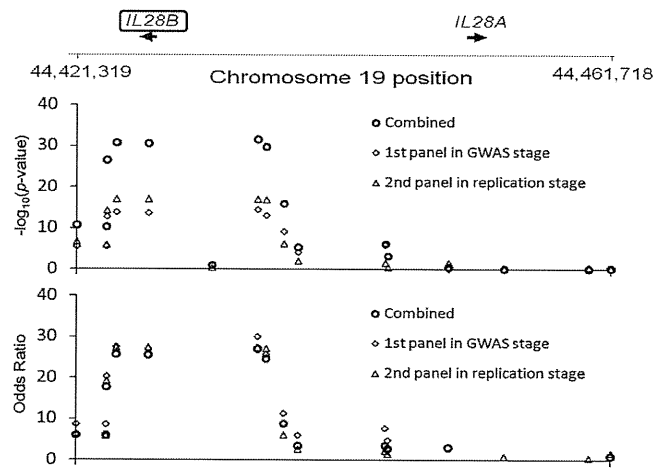


Figure 2. The strong association of *IL28B* with therapy response for chronic hepatitis C: 80% of non-responders possess the minor allele [28].



Response to the interferon-alpha therapy had been considered to be determined mainly by the virus genotype and concentration. However, the discovery that response is, in fact, mostly determined by a human genetic factor had a major impact. *IL28B* SNP typing has already been introduced into the routine clinical testing in Japan and is used as important reference data in the determination of therapeutic strategies.

3.2. Identification of New Therapeutic Targets

The discovery of *IL28B*, which is strongly associated with response to treatment for hepatitis C, indicated another highly interesting possibility. *IL28B* is a member of the interferon λ family and is assumed to exhibit its defensive activity against viral infection mediated by similar receptors and intracellular signal transduction pathway as interferon α, which was used in the treatment of hepatitis C. *IL28B* itself is therefore expected to be a powerful contender for the development of new hepatitis C drugs. In fact, *IL-29*, a member of the same family, has already been subjected to clinical trial for a new drug.

In addition to the above, genes involved in response to many drugs have been reported, and an increasing number of genetic factors are being identified for the first time as a result of GWAS. Drug-response genes generally tend to show greater odds ratios than disease-susceptibility genes, so that even with a relatively small sample size, there is a high likelihood of being able to identify the relevant gene. Ever greater results may therefore be expected in the future.

4. Particular Importance of HLA

4.1. Immune-Mediated Diseases and HLA

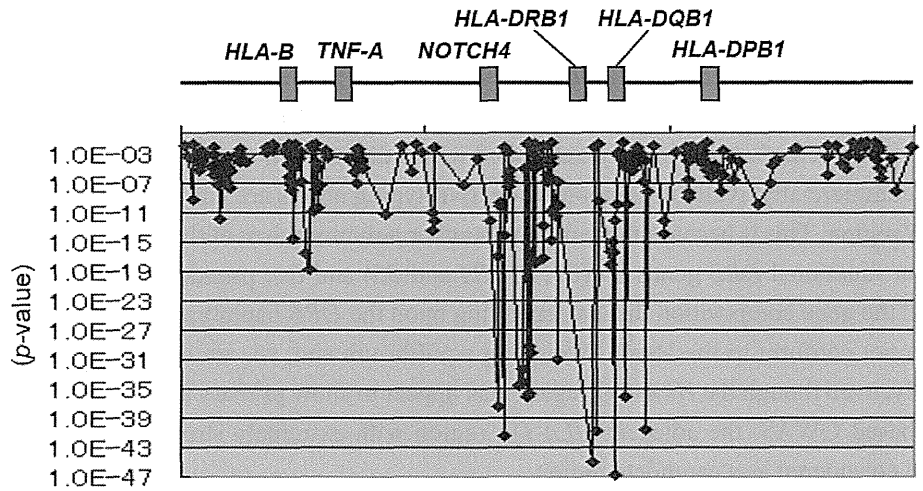
GWAS studies have been conducted for a number of diseases to date, and many of these have reported *HLA* as a susceptibility gene. In our own experience, narcolepsy [16], hepatitis B [29], rheumatoid arthritis [18], primary biliary cirrhosis [20], Stevens–Johnson syndrome, insulin autoimmune syndrome and type I diabetes have all shown strong association with certain *HLA* gene(s). Of these, narcolepsy, rheumatoid arthritis, primary biliary cirrhosis, type I diabetes and insulin autoimmune syndrome were associated most strongly with the *HLA-DR* and *HLA-DQ* regions, while hepatitis B and Stevens-Johnson syndrome were associated most strongly with the *HLA-DP* and *HLA-A* genes, respectively.

With regard to narcolepsy, Juji *et al.* [30] first reported in 1984 an extremely strong association with *HLA-DR2* (*HLA-DRB1*1501-DQB1*0602* haplotype according to the recent sequence-level nomenclature). We also found an extremely strong association between narcolepsy and the *HLA-DR/DQ* region with an SNP-based GWAS (Figure 3) [16]. If the results of HLA analysis in European and African populations are considered together, the primary susceptibility allele is assumed to be *HLA-DQB1*0602*.

Numerous GWAS have also been carried out for rheumatoid arthritis in Japan and elsewhere, and the *HLA-DR/DQ* region has been shown to have stronger association than any other region of the genome [18]. *HLA-DR4* has been known to be strongly associated with rheumatoid arthritis since the latter half of the 1970s; recent analysis at the sequence level has shown that *DRB1*0401* is most strongly associated in

European populations and *DRB1*0405* among Japanese. However, there are several other *DRB1* alleles that also exhibit susceptibility or resistance, and a hierarchy may be seen in their odds ratios.

Figure 3. GWAS confirmed the most strong association of the *HLA-DR/DQ* region with narcolepsy [16].



With primary biliary cirrhosis, also, the *HLA-DR/DQ* region showed the strongest association in the GWAS of European populations [31] and in the first GWAS of an Asian population [20]. From the analysis of *HLA* itself, *HLA-DRB1*0803-DQB1*0602* and *HLA-DRB1*0405-DQB1*0401* have been reported as susceptible haplotypes in the Japanese population [32], while *HLA-DRB1*0801-DQB1*04* was reported in European descendants [33].

4.2. Drug Hypersensitivity and HLA

There has also been great interest in *HLA* in its association with drug hypersensitivity. In 2002, it was reported that nearly 80% of patients who showed a hypersensitivity against the HIV drug, abacavir, possessed *HLA-B*5701*, with an odds ratio of 117 [34]. In 2004, a group from Taiwan found that of 44 patients with Stevens–Johnson syndrome induced by carbamazepine used for epilepsy seizures or as a psychotropic drug, all had *HLA-B*1502* [35]. However, less than 0.1% of Japanese possess *HLA-B*5701*, while *HLA-B*1502* is extremely rare. Consequently, it was predicted that the associations observed in the previous reports are hardly seen at all among Japanese.

In fact, Ozeki *et al.* [36] reported that adverse reactions in the skin as a result of carbamazepine are associated with *HLA-A*3101*. We reported independently that Stevens–Johnson syndrome/toxic epidermal necrolysis accompanied by eye manifestations caused by certain types of cold remedies is associated with *HLA-A*0206* [37]. Now, GWAS for this type of Stevens-Johnson syndrome has identified new susceptibility gene(s). Accordingly, GWAS can be powerful tool to investigate hypersensitivity to different kinds of drugs, and there is particular interest in associations with the *HLA* gene complex.

4.3. Characteristics of HLA and the Importance of HLA Typing

There are a number of unique characteristics of *HLA* genes and their polymorphisms, which indicates the limitation of SNP-based analysis and the importance of typing *HLA* genes themselves. First, the *HLA* genes are broadly classified into the Class I and Class II genes. Genes that exhibit high degrees of polymorphisms include *HLA-A*, *-B* and *-C* in Class I and *HLA-DRB1*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1* in Class II. Including *HLA* and non-HLA genes, a total of some 130 genes encoding proteins are densely located within a physical distance of about 4 Mbp on the short arm of Chromosome 6. They also show stronger linkage disequilibria than any other region of human genome. For these reasons, specifying a gene locus that is primarily associated with a disease is no easy task.

Second, commercially available genome-wide SNP typing arrays are unable to analyze the SNPs of the *HLA-DR* region. This is because there is copy number polymorphism of the *DRB* genes in the region: there are four functional *DRB* genes (*DRB1*, *B3*, *B4* and *B5*) and five pseudogenes *(DRB2*, *B6*, *B7*, *B8* and *B9*), and the gene composition differs depending upon the *DRB* haplotype. The SNPs of this region therefore do not conform to the Hardy–Weinberg equilibrium and, so, are not included on the arrays. Consequently, even though the *HLA-DQ* region may appear to show primary association from the results of an SNP-based GWAS, the adjacent *HLA-DR* region with extremely strong linkage disequilibrium must also be considered as a candidate region.

Third, genes in the Class II region are each adjacent on the genome as a pair, comprising an A gene and a B gene, and are linked to each other with a strong linkage disequilibrium. It is therefore very difficult to specify which gene of the pair is the primary one.

Fourth, as mentioned above, the *HLA* gene exhibits a high degree of polymorphism, and there are a huge number of alleles. There are almost no SNPs or SNP haplotypes that correspond one-on-one to individual *HLA* alleles. For example, more than 1300 alleles of *HLA-DRB1* have been admitted worldwide to date; for example, around 20 alleles with relatively high frequency and a great number of rare alleles have been found in the Japanese population; however, this sort of subclassification is not possible from SNP haplotypes.

Furthermore, a major feature is that a striking diversity between different populations can be observed. In other words, many *HLA* alleles are distributed only in certain regional populations.

Imputation of *HLA* alleles using *HLA* region SNP data is reported to have an accuracy of over 94% in European populations [38–40]. However, it is not perfect, especially for infrequent alleles, and the imputation is not yet fully available in Japanese or other Asian populations. The typing of the *HLA* genes is preferable for specifying *HLA* alleles directly involved in susceptibility, because there are multiple susceptibility alleles and resistance alleles, as well as 'neutral' alleles, and for many of these, the odds ratios are not consistent.

With regard to the *HLA*-associated diseases, therefore, detailed analysis, including the typing of the *HLA* genes themselves, are necessary to identify the primary *HLA* genes and alleles for each individual disease. These data will prove invaluable in clarifying the molecular mechanism through which HLA is associated with disease.

5. Conclusions and Issues for the Future

There are two hypotheses regarding the involvement of genome variation in common diseases: the common disease (common variants hypothesis and the common disease) and the rare variants hypothesis. In this regard, there is the argument that the common variants identified by GWAS as causing susceptibility to multifactorial diseases can only account for a small proportion of the genetic factors of disease, so that rare variants must also be important. This was symbolized by the term "missing heritability" [41], when only around 20 susceptibility loci for type II diabetes had been identified. Even in total, these could only explain about 5% of heritability. To date, over 60 common susceptibility loci have been identified, and this number is increasing all the time as a result of GWAS and meta-analyses carried out on greater scales. Further, it has been shown by the latest statistical analysis using all the GWAS data that around 40%–60% of all genetic factors can be explained. Therefore, it is assumed that there are still a great many relatively weak common susceptibility variants that have yet to be discovered.

To put it differently, we have not yet utilized the data obtained from GWAS to the fullest extent. For example, susceptibility genes that are not discovered by gathering samples from patients with the same disease name may be discovered by collecting detailed clinical data for each patient and then carrying out an analysis focused on clinical subsets. Considering a common disease from the viewpoint of its genetic architecture, the disease could be a collection of the many diseases that resemble each other, but also exhibit heterogeneity. Furthermore, it is likely that many susceptibility gene polymorphisms do not reach the so-called genome-wide significance level and, instead, exhibit moderate *p*-values. Establishing a method to identify the real susceptibility loci from this gray area is an issue that will need to be resolved in the future. It will be necessary to develop new methods that synthesize data from genetic ontology, pathway/network informatics and other fields and to establish statistical methods that can detect both intra-gene and inter-gene interactions. Our collaborators developed one such method that greatly improves the detection power of susceptibility loci [42].

Other than investigation by means of SNPs, there is also a need to clarify the degree to which variation, such as copy number variation (CNV) and short insertion/deletion variation, account for genetic factors in disease. Massive sequencing using next-generation sequencers is leading to astounding developments; to date, it has been very useful in identifying single genes responsible for hereditary diseases, and it has recently started to be applied to the search for susceptibility genes of multifactorial diseases. Until now, exome analysis has not turned up major results with respect to multifactorial disease. Considering that the majority of susceptibility SNPs identified by GWAS have been discovered in regions that regulate gene expression rather than in regions that code proteins, large-scale whole genome sequencing with a large number of patient and control samples may be needed. Then, the major challenge for the future is to establish a system to extract valuable data from the huge data produced by this new technology and to detect variants associated with certain multifactorial diseases.

HLA is already essential in clinical testing, such as organ and bone marrow transplantation and platelet transfusion. In addition, its association with over 100 types of diseases, including various autoimmune and inflammatory disorders, as well as infectious diseases, has been reported since the 1970s. Research aimed at understanding the mechanism of *HLA*-disease association commenced in the 1980s, but even now, the mechanism is not clearly known. In the 1990s, also, researchers carried out

many analyses of antigenic peptides eluted from *HLA* molecules prepared from mass cultured cells and analyses of T-cell clones created from patient samples, but were unable to gain a complete understanding of pathogenic peptides or the mechanisms of disease onset. It is hoped that there will be breakthroughs in the search for solutions to the huge riddle of disease mechanisms through advances, such as the diversity analysis of each *HLA* haplotype using next-generation sequencers, expression analysis of each *HLA* molecule using the latest protein chemistry and high-order structure analysis of the *HLA*-antigenic peptide-T-cell receptor complex.

Finally, the sharing of a huge amount of data produced by genome-wide variation analyses on various diseases through public databases, such as the Database of Genotypes and Phenotypes (dbGaP) [43], European Genome-Phenome Archive (EGA) [44] and GWAS Central [45], is crucial for the promotion of the complete identification of disease susceptibility genes and the understanding of the molecular mechanism of disease onset. We have also developed a public database for studies on the Japanese population [46–48].

Acknowledgments

This work was supported by grants-in-aid for scientific research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, the Japanese Ministry of Health, Labor and Welfare and the Japan Science and Technology Agency.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Risch, N. Searching for genetic determinants in the millennium. *Nature* **2000**, *405*, 847–856.
2. Ozaki, K.; Ohnishi, Y.; Iida, A.; Sekine, A.; Yamada, R.; Tsunoda, T.; Sato, H.; Hori, M.; Nakamura, Y.; Tanaka, T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **2002**, *32*, 650–654.
3. Tamiya, G.; Shinya, M.; Imanishi, T.; Ikuta, T.; Makino, S.; Okamoto, K.; Furugaki, K.; Matsumoto, T.; Mano, S.; Ando, S.; *et al.* Whole genome association study of rheumatoid arthritis using 27039 microsatellites. *Hum. Mol. Genet.* **2005**, *14*, 2305–2321.
4. Database of Single Nucleotide Polymorphisms (dbSNP). Available online: http://www.ncbi.nlm.nih.gov/snp/ (accessed on 11 February 2014).
5. Database of HapMap Project. Available online: http://hapmap.ncbi.nlm.nih.gov/ (accessed on 11 February 2014).
6. Database of 1000 Genomes Project. Available online: http://www.1000genomes.org/ (accessed on 11 February 2014).
7. A Catalog of Published Genome-Wide Association Studies. Available online: http://www.genome.gov/gwastudies/ (accessed on 11 February 2014).

8. Sladek, R.; Rocheleau, G.; Rung, J.; Dina, C.; Shen, L.; Serre, D.; Boutin, P.; Vincent, D.; Belisle, A.; Hadjadj, S.; *et al.* Genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **2007**, *445*, 881–885.
9. Saxena, R.; Voight, B.F.; Lyssenko, V.; Burtt, N.P.; de Bakker, P.I.; Chen, H.; Roix, J.J.; Kathiresan, S.; Hirschhorn, J.N.; Daly, M.J.; *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **2007**, *316*, 1331–1336.
10. Scott, L.J.; Mohlke, K.L.; Bonnycastle, L.L.; Willer, C.J.; Li, Y.; Duren, W.L.; Erdos, M.R.; Stringham, H.M; Chines, P.S.; Jackson, A.U.; *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **2007**, *316*, 1341–1345.
11. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661–678.
12. Yasuda, K.; Miyake, K.; Horikawa, Y.; Hara, K.; Osawa, H.; Furuta, H.; Hirota, Y.; Mori, H.; Jonsson, A.; Sato, Y.; *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **2008**, *40*, 1092–1097.
13. Unoki, H.; Takahashi, A.; Kawaguchi, T.; Hara, K.; Horikoshi, M.; Andersen, G.; Ng, D.P.; Holmkvist, J.; Borch-Johnsen, K.; Jørgensen, T.; *et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **2008**, *40*, 1098–1102.
14. Miyake, K.; Horikawa, Y.; Hara, K.; Yasuda, K.; Osawa, H.; Furuta, H.; Hirota, Y.; Yamagata, K.; Hinokio, Y.; Oka, Y.; *et al.* Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4,087 Japanese subjects. *J. Hum. Genet.* **2008**, *53*, 174–180.
15. Matsuki, K.; Juji, T.; Tokunaga, K.; Naohara, T.; Satake, M.; Honda, Y. Human histocompatibility leukocyte antigen (HLA) haplotype frequencies estimated from the data on HLA class I, II, and III antigens in 111 Japanese narcoleptics. *J. Clin. Invest.* **1985**, *76*, 2078–2083.
16. Miyagawa, T.; Kawashima, M.; Nishida, N.; Ohashi, J.; Kimura, R.; Fujimoto, A.; Shimada, M.; Morishita, S.; Shigeta, T.; Lin, L.; *et al.* Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. *Nat. Genet.* **2008**, *40*, 1324–1328.
17. Mahasirimongkol, S.; Yanai, H.; Mushiroda, T.; Promphittayarat, W.; Wattanapokayakit, S.; Phromjai, J.; Yuliwulandari, R.; Wichukchinda, N.; Yowang, A.; Yamada, N.; *et al.* Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus foe young tuberculosis. *J. Hum. Genet.* **2012**, *57*, 363–367.
18. Okada, Y.; Terao, C.; Ikari, K.; Kochi, Y.; Ohmura, K.; Suzuki, A.; Kawaguchi, T.; Stahl, E.A.; Kurreeman, F.A.; Nishida, N.; *et al.* Meta-analysis of genome-wide association studies identifies multiple novel loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet.* **2012**, *44*, 511–516.
19. Takamoto, M.; Kaburaki, T.; Mabuchi, A.; Araie, M.; Amano, S.; Aihara, M.; Tomidokoro, A.; Iwase, A.; Mabuchi, F.; Kashiwagi, K.; *et al.* Common variants on chromosome 9q21 are associated with normal tension glaucoma. *PLoS One* **2012**, *7*, e40107.
20. Nakamura, M.; Nishida, N.; Kawashima, M.; Aiba, Y.; Tanaka, A.; Yasunami, M.; Nakamura, H.; Komori, A.; Nakamuta, M.; Zeniya, M.; *et al.* Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am. J. Hum. Genet.* **2012**, *91*, 721–728.

21. Okamoto, K.; Tokunaga, K.; Doi, K.; Fujita, T.; Suzuki, H.; Katoh, T.; Watanabe, T.; Nishida, N.; Mabuchi, A.; Takahashi, A.; *et al.* Common variation in GPC5 is associated with acquired nephrotic syndrome. *Nat. Genet.* **2011**, *43*, 459–463.
22. Cotsapas, C.; Voight, B.F.; Rossin, E.; Lage, K.; Neale, B.M.; Wallace, C.; Abecasis, G.R.; Barrett, J.C.; Behrens, T.; Cho, J.; *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **2011**, *7*, e1002254.
23. Parkes, M.; Cortes, A.; van Heel, D.A.; Brown, M.A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **2013**, *14*, 661–673.
24. Miyagawa, T.; Miyadera, H.; Tanaka, S.; Kawashima, M.; Shimada, M.; Honda, Y.; Tokunaga, K.; Honda, M. Abnormally low serum acylcarnitine level in narcolepsy. *Sleep* **2011**, *34*, 349–353.
25. Miyagawa, T.; Kawamura, H.; Obuchi, M.; Ikesaki, A.; Ozaki, A.; Tokunaga, K.; Inoue, Y.; Honda, M. Effects of oral L-carnitine administration in narcolepsy patients: A randomized, double-blind, cross-over and placebo-controlled trial. *PLoS One* **2013**, *8*, e53707.
26. Hallmayer, J.; Faraco, J.; Lin, L.; Hesselson, S.; Winkelmann, J.; Kawashima, M.; Mayer, G.; Plazzi, G.; Nevsimalova, S.; Bourgin, P.; *et al.* Narcolepsy is strongly associated with the TCR alpha locus. *Nat. Genet.* **2009**, *41*, 708–711.
27. Kornum, B.R.; Kawashima, M.; Faraco, J.; Lin, L.; Rico, T.J.; Hesselson, S.; Axtell, R.C.; Kuipers, H.; Weiner, K.; Hamacher, A.; *et al.* Common variants in P2RY11 are associated with narcolepsy. *Nat. Genet.* **2011**, *43*, 66–71.
28. Tanaka, Y.; Nishida, N.; Sugiyama, M.; Kurosaki, M.; Matsuura, K.; Sakamoto, N.; Nakagawa, M.; Korenaga, M.; Hino, K.; Hige, S.; *et al.* Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* **2009**, *41*, 1105–1109.
29. Nishida, N.; Sawai, H.; Matsuura, K.; Sugiyama, M.; Ahn, S.H.; Park, J.Y.; Hige, S.; Kang, J.H.; Suzuki, K.; Kurosaki, M.; *et al.* Genome-wide association study confirming association of *HLA-DP* with protection against chronic hepatitis B and viral clearance in Japanese and Korean. *PLoS One* **2012**, *7*, e39175.
30. Juji, T.; Satake, M.; Honda, Y.; Doi, Y. *HLA* antigens in Japanese patients with narcolepsy. All patients were DR2 positive. *Tissue Antigens* **1984**, *24*, 316–319.
31. Hirschfield, G.M.; Liu, X.; Xu, C.; Lu, Y.; Xie, G.; Lu, Y.; Gu, X.; Walker, E.J.; Jing, K.; Juran, B.D.; *et al.* Primary biliary cirrhosis associated with *HLA*, *IL12A*, and *IL12RB2* variants. *N. Engl. J. Med.* **2009**, *360*, 2544–2555.
32. Nakamura, M.; Yasunami, M.; Kondo, H.; Horie, H.; Aiba, Y.; Komori, A.; Migita, K.; Yatsuhashi, H.; Ito, M.; Shimoda, S.; *et al.* Analysis of *HLA-DRB1* polymorphisms in Japanese patients with primary biliary cirrhosis (PBC): The *HLA-DRB1* polymorphism determines the relative risk of antinuclear antibodies for disease progression in PBC. *Hepatol. Res.* **2010**, *40*, 494–504.
33. Begovich, A.B.; Klitz, W.; Moonsamy, P.V.; van de Water, J.; Peltz, G.; Gershwin, M.E. Genes within the *HLA* class II region confer both predisposition and resistance to primary biliary cirrhosis. *Tissue Antigens* **1994**, *43*, 71–77.

34. Mallal, S.; Nolan, D.; Witt, C.; Masel, G.; Martin, A.M.; Moore, C.; Sayer, D.; Castley, A.; Mamotte, C.; Maxwell, D.; *et al.* Association between presence of *HLA-B*5701*, *HLA-DR7* and *HLA-DQ3* and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **2002**, *359*, 727–732.
35. Chung, W.H.; Hung, S.I.; Hong, H.S.; Hsih, M.S.; Yang, L.C.; Ho, H.C.; Wu, J.Y.; Chen, Y.T. Medical genetics: A marker for Stevens-Johnson syndrome. *Nature* **2004**, *428*, 486.
36. Ozeki, T.; Mushiroda, T.; Yowang, A.; Takahashi, A.; Kubo, M.; Shirakata, Y.; Ikezawa, Z.; Iijima, M.; Shiohara, T.; Hashimoto, K.; *et al.* Genome-wide association study identifies *HLA-A*3101* allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. *Hum. Mol. Genet.* **2011**, *20*, 1034–1041.
37. Ueta, M.; Tokunaga, K.; Sotozono, C.; Inatomi, T.; Yabe, T.; Matsushita, M.; Mitsuishi, Y.; Kinoshita, S. *HLA* class Ⅰ and Ⅱ gene polymorphisms in Stevens-Johnson syndrome with ocular complications in Japanese. *Mol. Vis.* **2008**, *14*, 550–555.
38. Dilthey, A.; Leslie, S.; Moutsianas, L.; Shen, J.; Cox, C.; Nelson, M.R.; McVean, G. Multi-population classical *HLA* type imputation. *PLoS Comput. Biol.* **2013**, *9*, e1002877.
39. Zheng, X.; Shen, J.; Cox, C.; Wakefield, J.C.; Ehm, M.G.; Nelson, M.R.; Weir, B.S. *HIBAG-HLA* genotype imputation with attribute bagging. *Pharmacogenomics J.* **2013**, doi:10.1038/tpj.2013.18.
40. Jia, X.; Han, B.; Onengut-Gumuscu, S.; Chen, W.M.; Concannon, P.J.; Rich, S.S.; Raychaudhuri, S.; de Bakker, P.I. Imputing amino acid po*lymorphisms in hum*an leucocyte antigens. *PLoS One* **2013**, *8*, e64683.
41. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **2008**, *456*, 18–21.
42. Dinu, I.; Mahasirimongkol, S.; Liu, Q.; Yanai, H.; El-Din, N.S.; Kreiter, E.; Wu, X.; Jabbari, S.; Tokunaga, K.; Yasui, Y. SNP-SNP interactions discovered by logic regression explain Crohn's disease genetics. *PLoS One* **2012**, *7*, e43035.
43. Database of Genotypes and Phenotypes (dbGaP). Available online: http://www.ncbi.nlm.nih.gov/gap/ (accessed on 11 February 2014).
44. European Genome-Phenome Archive (EGA). Available online: https://www.ebi.ac.uk/ega/ (accessed on 11 February 2014).
45. GWAS Central. Available online: https://www.gwascentral.org/ (accessed on 11 February 2014).
46. Koike, A.; Nishida, N.; Inoue, I.; Tsuji, S.; Tokunaga, K. Genome-wide association database developed in the Japanese Integrated Database Project. *J. Hum. Genet.* **2009**, *54*, 543–546.
47. Koike, A.; Nishida, N.; Yamashita, D.; Tokunaga, K. Comparative analysis of copy number variation detection methods and database construction. *BMC Genet.* **2011**, *12*, e29.
48. Human Genome Variation Database. Available online: https://gwas.biosciencedbc.jp/index.html/ (accessed on 11 February 2014).

The Journal of Clinical Investigation RESEARCH ARTICLE

Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA

Hiroko Miyadera,[1,2] Jun Ohashi,[3] Åke Lernmark,[4] Toshio Kitamura,[5] and Katsushi Tokunaga[1]

[1]Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [2]Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Chiba, Japan. [3]Faculty of Medicine, University of Tsukuba, Ibaraki, Japan. [4]Department of Clinical Sciences, Lund University/CRC, Skåne University Hospital SUS, Malmö, Sweden. [5]Division of Cellular Therapy and Division of Stem Cell Signaling, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Sciences, The University of Tokyo, Tokyo, Japan.

Polymorphisms within *HLA* gene loci are strongly associated with susceptibility to autoimmune disorders; however, it is not clear how genetic variations in these loci confer a disease risk. Here, we devised a cell-surface MHC expression assay to detect allelic differences in the intrinsic stability of HLA-DQ proteins. We found extreme variation in cell-surface MHC density among *HLA-DQ* alleles, indicating a dynamic allelic hierarchy in the intrinsic stability of HLA-DQ proteins. Using the case-control data for type 1 diabetes (T1D) for the Swedish and Japanese populations, we determined that T1D risk–associated *HLA-DQ* haplotypes, which also increase risk for autoimmune endocrinopathies and other autoimmune disorders, encode unstable proteins, whereas the T1D-protective haplotypes encode the most stable HLA-DQ proteins. Among the amino acid variants of *HLA-DQ*, alterations in 47α, the residue that is located on the outside of the peptide-binding groove and acts as a key stability regulator, showed strong association with T1D. Evolutionary analysis suggested that 47α variants have been the target of positive diversifying selection. Our study demonstrates a steep allelic hierarchy in the intrinsic stability of HLA-DQ that is associated with T1D risk and protection, suggesting that HLA instability mediates the development of autoimmune disorders.

Introduction

HLA (also known as the MHC in other vertebrates) proteins present self- and non-self peptides to the T cell receptor (TCR) to maintain self-tolerance and adapted immunity (ref. 1 and Figure 1A). Certain *HLA-DR-DQ* haplotypes, such as *DR3-DQA1*05-DQB1*02:01* (*DR3-DQ2.5*) and *DR4-DQA1*03-DQB1*03:02* (*DR4-DQ8.3*) in Europeans, confer a risk for autoimmune diseases, including type 1 diabetes (T1D), celiac disease, and autoimmune endocrinopathy (2–7). In East Asian populations (Japanese) *DR9-DQA1*03-DQB1*03:03* (*DR9-DQ9.3*) and *DR4-DQA1*03-DQB1*04:01* (*DR4-DQ4.3*) are the major risk factors for T1D (8–10) and other autoimmune endocrinopathies (11) (Table 1 and see Supplemental Figure 1 for abbreviations for the *DQA1-DQB1* haplotypes [*DQ* haplotype];; supplemental material available online with this article; doi:10.1172/JCI74961DS1). Despite accumulating genetic evidence, the mechanism through which particular *HLA* alleles confer risk for autoimmune diseases has not been fully uncovered.

The *HLA*-autoimmunity association has been generally explained by the allelic differences in self-epitope presentation (see, for instance, refs. 12, 13). However, the binding affinity and specificity in the MHC-self-epitope interaction are highly variable. Studies of T1D, multiple sclerosis, and other autoimmune disorders have found that the disease-relevant self-epitopes interact with the MHC with high or low affinity or in kinetically unfavorable registers (14–29). Some of these self-epitopes bind to both the disease risk and neutral/protective allele products (i.e., promiscuous binders) (15, 30–32). Although self-epitope presentation is critical in autoimmune pathogenesis, the above findings suggest that additional factors may also contribute to the allele-specific disease risk.

In the 1990s, Kwok's and Unanue's groups and other researchers reported that the T1D risk alleles of *HLA-DQ* and murine *I-A* encode SDS unstable proteins (33–36). The SDS stability measures the migration of non-boiled MHC class II (MHC II) on SDS-PAGE (37) and was initially regarded as an indicator of peptide occupancy. It was later found that SDS stability reflects the stabilization of the peptide-MHC (pMHC) at the P1 and P9 pockets and at the extended peptide residues (38–45). In some of these and other studies, however, SDS stability was not affected by the peptide-binding affinity (41, 42, 46) and was maintained through the peptide-independent stabilization (46). The mechanism of SDS stability, and hence its relevance to the MHC protein function, has remained controversial.

The stability of the pMHC is maintained through the heterodimerization of the α and β subunits and peptide presentation (Supplemental Figure 2). The interaction of the peptide side chain atoms with MHC stabilizes the pMHC in a peptide-specific manner and has been extensively analyzed (1). In this study, we focused on the possibilities that the MHC stability might differ intrinsically among the alleles and that this stability may be associated with autoimmunity. The intrinsic stability of the MHC protein in this study refers to the MHC stability that is formed through the α/β assembly and peptide main chain interactions. The contribution of both the polymorphic and nonpolymorphic residues in the heterodimerization and peptide main chain interactions suggests that MHC stability might differ intrinsically among alleles. However, it has not been possible to measure the intrinsic stability of MHC protein

Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: December 27, 2013; **Accepted:** November 6, 2014.
Reference information: *J Clin Invest.* 2015;125(1):275–291. doi:10.1172/JCI74961.

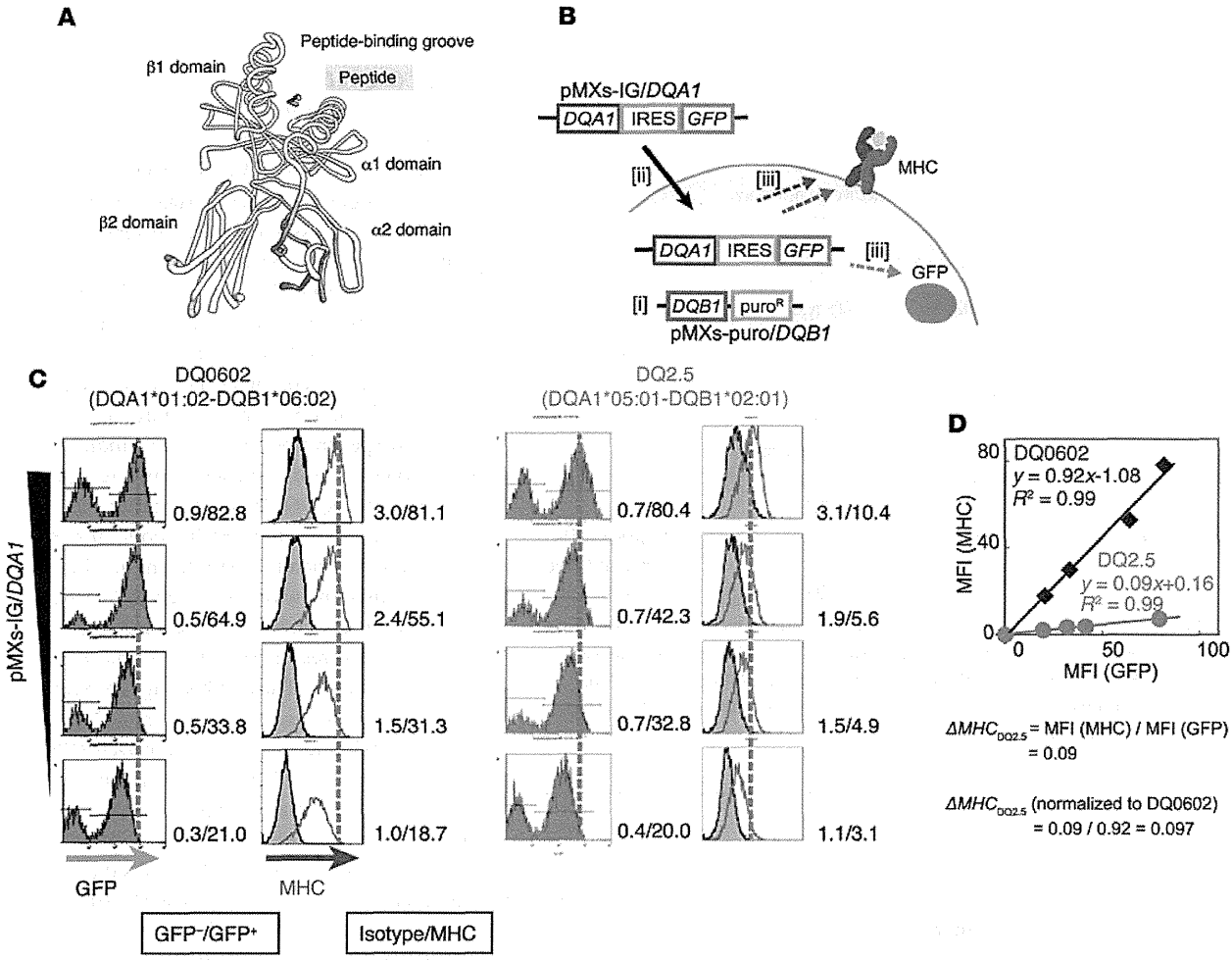


Figure 1. Measurement of Δ*MHC*. (**A**) Structure of MHC II (DQ0602 [pdb: 1uvq]) (70). MHC II is a heterodimeric transmembrane glycoprotein that is composed of one α and one β subunit. The α1 and β1 domains constitute the peptide-binding groove, and the α2 and β2 domains form the constant domain. In the case of HLA-DQ, *HLA-DQA1* and *-DQB1* encode the α and β subunits, respectively. (**B**) Outline of the Δ*MHC* assay. The Δ*MHC* assay measures cell-surface MHC expression levels normalized to the internal control GFP. The *HLA-DQB1*–stable cells (i) were transduced with the retroviral vector pMXs-IG/DQA1 (52) (ii). HLA-DQ was expressed on the cell surface in the presence of both the *HLA-DQA1* and *-DQB1*. Cell-surface HLA-DQ expression and cytosolic GFP expression were measured by flow cytometry (iii). (**C**) Representative data from the Δ*MHC* assay for DQ0602 (left panels) and DQ2.5 (right panels). To quantify cell-surface MHC expression, the *HLA-DQB1*–stable cell line was transduced with a graded concentration of retrovirus containing pMXs-IG/*DQA1*. Expression levels of both HLA-DQ and GFP increased with the concentration of the retrovirus. Numbers indicate the MFI for GFP-negative and -positive cells and for the isotype control and anti–HLA II β (WR18). Dashed lines indicate the highest MFI in each sample set. GFP (green), anti–HLA II β (WR18) (magenta), and isotype control (black). (**D**) The increase in cell-surface MHC expression relative to GFP (Δ*MHC*) was calculated by plotting the MFI (GFP) and MFI (MHC) (Supplemental Figure 3A). The Δ*MHC* for each *HLA-DQ* allelic pair was normalized to the Δ*MHC* of DQ0602, which was measured on the same day. The Δ*MHC* for DQ2.5 and its normalized value are shown. See also Supplemental Figures 3–5 and Methods.

or to demonstrate its allelic differences, because the pMHC is usually stabilized through both the peptide main chain and side chain interactions.

To detect the potential allelic differences in the intrinsic stability of the MHC protein, we used an alternative approach to the conventional stability assays. Specifically, instead of analyzing protein stability itself, we measured the biological outcome, the cell-surface expression of MHC protein. We quantified the amount of cell-surface MHC in engineered conditions and confirmed, through the use of mutagenesis and the model peptides, that the level of cell-surface MHC protein density (referred to herein as the Δ*MHC*) reflects the intrinsic stability of the MHC protein. Δ*MHC* was then used to analyze the relationship between the intrinsic stability of MHC protein and autoimmune disease risk. Δ*MHC* measures the combined outcomes of the heterodimer assembly, cell-surface transport, and turnover, but not the chemical or physical stability of the MHC protein. However, for simplicity, Δ*MHC* is used as an equivalent to the protein stability in this article.

In this study, we identified an allelic diversity in the intrinsic stability of HLA-DQ that has been maintained through evolution and is associated with genetic risk for T1D. Our study provides a new framework through which to interpret the *HLA*-autoimmunity association profiles and uncover the mechanism of autoimmunity.

Table 1. Associations of *DR-DQ* haplotypes with autoimmune and other immune disorders

DR-DQA1-DQB1 haplotype	Susceptible	Protective
*DR3-DQ2.5 (DRB1*03-DQA1*05-DQB1*02)*[A]	APS type II	
	Celiac disease	
	Selective IgA deficiency	
	MS	
	SLE	
	T1D	
*DR4-DQ8.3 (DRB1*04-DQA1*03-DQB1*03:02)*[A]	APS type II	
	Celiac disease	
	T1D	
*DR9-DQ9.3 (DRB1*09-DQA1*03-DQB1*03:03)*[B]	APS type III	
	T1D	
	Microscopic polyangiitis	
*DR4-DQ4.3 (DRB1*04-DQA1*03-DQB1*04:01)*[B]	APS type III	
	T1D	
*DR15-DQ0602 (DRB1*15-DQA1*01:02-DQB1*06:02)*	Narcolepsy	APS types II and III
	MS	Selective IgA deficiency
	SLE	T1D

[A]Haplotypes commonly observed in Europeans. [B]Haplotypes commonly observed in the Japanese population. APS type II (6); celiac disease (56, 137); selective IgA deficiency (5, 138); MS (4, 5, 139); SLE (4, 5); T1D (Europeans) (59, 79, 80); APS type III (11); T1D (Japanese) (10); microscopic polyangiitis (140); narcolepsy (141). MS, multiple sclerosis; SLE, systemic lupus erythematosus.

Results

Measurement of ΔMHC. We generated an MHC expression system using fibroblasts (NIH3T3, murine embryonic fibroblasts) as expression hosts and GFP as an internal control. MHC II is expressed in a functionally intact form in fibroblasts (47, 48) but in inappropriately paired or unassembled forms, the α and β subunits are retained during intracellular transport and are degraded (49, 50). The observation that cell-surface MHC expression on fibroblasts is altered through the gain and loss of hydrogen bond(s) (H-bond) between the MHC and the peptide (51) indicates that subtle changes in the net stability of the pMHC can be detected using cell-surface MHC protein expression levels. H2-DM, which is expressed in antigen-presenting cells (APCs) and stabilizes pMHC, may be absent in this expression system.

We established *HLA-DQB1*–stable cells using the retrovirus vector pMXs-puro and the packaging cell line PLAT-E (52, 53). We then transduced the *HLA-DQB1*–stable cells with a retrovirus containing pMXs-IG/*DQA1* (Figure 1B). Using a graded concentration of retrovirus particles, it was possible to express both the HLA-DQ and GFP at several different levels (Figure 1C). Cell-surface HLA-DQ and cytosolic GFP were measured by flow cytometry using the pan–HLA II β mAb (WR18). The mean fluorescence intensity (MFI) for both the MHC [MFI (MHC)] and the GFP [MFI (GFP)] showed good linear correlation (R^2 >0.9). The increase in MFI (MHC) relative to MFI (GFP) (slope in Figure 1D), which indicates the amount of cell-surface MHC normalized to GFP, was calculated for each *HLA-DQ* allelic pair and was designated as *ΔMHC* (Figure 1, C and D, and Supplemental Figure 3, A and B). To minimize interassay variation, *ΔMHC* was normalized to *ΔMHC* for the *DQA1*01:02-DQB1*06:02* haplotype product (DQ0602), which is highly SDS stable (36) and showed one of the highest *ΔMHC* values among the tested alleles. Hereafter, the *ΔMHC* values that were normalized to the *ΔMHC* of DQ0602 are indicated in the figures unless otherwise specified. Representative *ΔMHC* assay data are presented in Supplemental Figure 4, A and B. *ΔMHC* was measured for the major *HLA-DQ* alleles in worldwide populations and in their possible *trans* combinations, given that the *trans DQA1-DQB1* pair forms heterodimers (54), and certain *trans* combinations are associated with autoimmunity (55–59). In this study, the *HLA* allele and haplotype protein products are indicated using the nonitalic version of the gene name (e.g., DQ0602 represents the *DQ0602* haplotype product).

Figure 2A shows the *ΔMHC* profile for HLA-DQ. *ΔMHC* varied by nearly 100-fold among the *HLA-DQ* alleles. Consistent with earlier work (60–62), *HLA-DQA1* and *DQB1* alleles of the same evolutional sublineage (63) expressed HLA-DQ on the cell surface (Figure 2, A and B). These sublineages are referred to herein as the subgroups DQ2/3/4 and DQ5/6. *HLA-DQA1*02*, **03*, and **05* and certain *DQB1*06* alleles also expressed HLA-DQ on the cell surface (Figure 2A). HLA-DQ cell-surface expression was not detectable in the absence of *HLA-DQA1* or in the presence of the incompatible *HLA-DQA1* alleles (Supplemental Figure 5, A–C, and Supplemental Table 1). The HLA-DQ cell-surface expression pattern and the assembly of the DQα and DQβ subunits were confirmed using stable insect cells (*Drosophila melanogaster* S2) (Supplemental Figure 6, A–D).

Among the major *DQ* haplotypes, *DQ0602* and *DQ9.2* showed the highest *ΔMHC*, whereas the *ΔMHC* of the *DQA1*01:04-DQB1*05:01* (*DQ0501*) product was below the threshold (Figure 2, A and C, and Supplemental Table 1). The conserved hierarchy in *ΔMHC* among the DQ5/6 (DQA1*01:02 > *01:01, *01:03 > *01:04 and DQB1*06:02 > *06:01, *06:03 > *06:04 > *05:03 > *05:02 > *05:01) (Figure 2A and Supplemental Figure 7, A–D) indicates that polymorphic variants in each subunit act independently of the variants in the other subunit in the regulation of *ΔMHC*. In DQ2/3/4, *ΔMHC* decreased in the order of DQA1*02 > *03, *05 > *04 > *06 and DQB1*03:01, *03:03, *04 > *02, *03:02, except for the high *ΔMHC* value of *DQA1*03-DQB1*02* (*DQ2.3*) (Figure 2A and Supplemental Figure 7, E–H). The mechanism that stabilizes DQ2.3 was not identified in this study. The hierarchy in *ΔMHC* for the major *DQ* haplotype products was similar to the hierarchy in the SDS stability (DQ0602 > DQ0603 > DQ0604, DQ7.3, DQ0501, DQ8.3, DQ2.5) that was measured using the cell lysates of *DQ*-homologous B lymphoblastoid cell lines (B-LCLs) (36). SDS stability may be sensitive to the variants at 57β, given that DQ9.3 (*ΔMHC* = 0.12, carries Asp57β) is SDS stable (43), whereas DQ2.2 (*ΔMHC* = 0.4–0.6, carries non-Asp57β) is reported as SDS unstable (43).

The mAb WR18 stains diverse *HLA-DQ* (64, 65), *-DR*, and *-DP* allele products (H. Miyadera, unpublished observations), indicating that the mAb WR18 recognizes a common epitope on HLA II β.