厚生労働科学研究費補助金

肝炎等克服緊急対策研究事業

ウイルス性慢性肝疾患の病態に影響を与える

miRNA多型の網羅的探索に関する研究

平成２５年度　総括研究報告書

研究代表者　　三木　大樹

平成２６（２０１４）年３月

目　　次

厚生労働科学研究費補助金（肝炎等克服緊急対策研究事業）

総括研究報告書

ウイルス性慢性肝疾患の病態に影響を与えるmiRNA多型の網羅的探索に関する研究

研究代表者　三木　大樹　独立行政法人理化学研究所　研究員

研究要旨：肝疾患の診療現場では、宿主のゲノム情報を活用した個別化医療が現実のものとなりつつあるが、依然として宿主の遺伝的要因が全て解明された訳ではないし、医療の個別化も満足のいくものではない。最近、機能を持ったノン・コーディングRNAの存在が明らかになってきたが、中でもmiRNAは肝疾患を含む多くの疾患で新規バイオマーカーとして期待され、これを核酸医薬の標的としたC型慢性肝炎治療薬の臨床試験が進行中である。これまでのGWASやマイクロアレイの情報、複数の統合データベースを利用して包括的に解析することで、GWASでは明らかにできなかったウイルス性慢性肝疾患の病態へ影響を与えるmiRNA多型の同定を目指して研究を行っている。肝炎患者の表現型（量的・質的形質）に関するGWASデータを利用し、各解析で得られたシグナルの近傍に位置するmiRNA多型を網羅的に囲い込むと同時に、マイクロアレイや統合データベースから得られる発現量や組織特異性などの多面的な情報を利用して、解析候補となるmiRNA多型を抽出した。また、免疫関連遺伝子に注目し、それらを標的とするmiRNAを抽出し、多型解析候補とした。こうして抽出された100以上の候補SNPにつき、5,000人規模でのgenotypingを行い、表現型との関連について統計学的検証を行ったところ、既報のものも含め、いくつかの関連が確認できており、追認試験の準備を進めている。今後、追認試験を行った上で、多型と表現型との関連について再現性が確認できたものについては、順次、機能的な解析へと進める予定である。同時に、バイオマーカーとしての役割が期待できる多型に関しては、臨床的な有用性を評価して、臨床応用の可否を検討する。新規バイオマーカー開発や創薬に結びつく、あるいは個別化医療への貢献など、何らかの形で臨床応用に資する結果を目指して、本研究計画を進めて行く。

＜研究協力者＞
豊田成司
JA北海道厚生連　札幌厚生病院　院長

狩野吉康
JA北海道厚生連　札幌厚生病院　副院長

芥田憲夫
国家公務員共済組合連合会　虎の門病院　医長

越智秀典
広島大学大学院　消化器・代謝内科学　講師

ヘイズ・ネルソン
広島大学大学院　消化器・代謝内科学　特任准教授

A．研究目的
　ゲノムワイド関連解析（GWAS）によって、C 型慢性肝炎患者に対するインターフェロン治療の効果を強く規定する *IL28B* 遺伝子多型が発見され、ガイドラインにも盛り込まれるなど、すでに臨床応用がなされている。他にも、GWAS は B 型慢性肝炎・C 型慢性肝炎の易罹患性、肝硬変・肝癌への進展などに関連する遺伝子多型あるいは一塩基多型（SNP）を明らかにしてきた。しかし、GWAS ではカバーしきれないゲノム領域の存在や、統計学的検出力の不足などの問題が依然として残っている。一方で、機能を有するノン・コーディング RNA、中でも21〜25塩基ほどのmiRNAが肝疾患を含む多くの疾患で新規バイオマーカーとして期待され、これを核酸医薬の標的とした C 型慢性肝炎治療薬の臨床試験が進行中であるなど、注目が集まっている。わが国における肝炎対策が急務であるとい

う点では、臨床応用に比較的近い位置にあると考えられる miRNA を研究対象にすることは期待度が高いと思われる。また、医療の個別化を意識する点では、SNP 解析という実績のある手法を用いることは妥当であると思われる。発展目覚しい種々の大規模データベースを活用することで、効率的に、これまでの GWAS では明らかにできなかったウイルス性慢性肝疾患の病態・種々の表現型に関連する miRNA の多型の同定を目指す。

B．研究方法
　これまでのGWASの成果を活かしながら、急速に増加・充実している種々のデータベースの情報（miRNA配列、人種別SNP頻度、遺伝子発現量、組織特異性、Gene Ontology、パスウェイ、等）を積極的に統合し、活用するとこで、GWASでは明らかにし得なかったmiRNA多型を効率的に拾い上げる。
　1．網羅的アプローチ：GWASデータあるいは公開されたデータベースを利用し、解析対象SNPを網羅的に絞り込む。
　2．候補アプローチ：ウイルス性肝疾患の病態に重要であると考えられている免疫システムや炎症、線維化、癌関連の遺伝子を標的とし得るmiRNAの近傍のSNPについては、優先的に解析候補として扱う。
　上記、1．および2．によってSNPを抽出した。順次、スクリーニング用のサンプルを用いてPCR-based SNP genotyping（Invader assay）を行い、統計学的検定を行っている。
　サンプルと付随する臨床情報は虎の門病院、札幌厚生病院、広島大学病院とその関連病院といった研究協力施設で収集されており、解析対象とする種々の表現型については、各研究協力者からのアドバイスを受けながら選別を行った。
（倫理面への配慮）

「ヒトゲノム・遺伝子解析研究に関する倫理指針」に基づいて、インフォームド・コンセントが得られた検体を用いて実施した。所属研究機関の倫理委員会の承認を得て実施した。

C．研究結果

まず、我々が行った日本人ウイルス性慢性肝疾患の GWAS データから、種々の病態との関連が示唆される遺伝子座位の近傍に位置する miRNA のうち、pri-miRNA あるいは pre-miRNA あるいは mature miRNA 上に SNP を有するものをデータベースから検索した。このうち、アジア人におけるマイナーアレル頻度が 1%以上の SNP を対象とした。また、miRNA 発現の組織特異性など、公開された複数の統合データベースから得られる情報も含めて、多面的な評価を行った上で優先順位を付け、解析候補となる miRNA 多型を抽出した。

網羅的アプローチと並行して、すでに他国から報告のあった肝疾患関連 miRNA 多型や、免疫システムなどの有望な候補遺伝子あるいはパスウェイを制御し得ることが seed 配列との相補性等から予測される miRNA 多型を、解析候補として抽出した。肝疾患の病態形成に重要と考えられるその他の遺伝子を標的とし得る miRNA についても現在、検討中である。

このようにして、これまでに約 200 個の SNP を解析候補として抽出した。順次、それぞれの解析対象となる表現型に対応したスクリーニング用サンプル（最大で HBV および HCV 関連肝疾患患者各 2,000 人、健常人 1,000 人の規模）を用いて、SNP genotyping を行っている。すでに統計学的検定を行ったものの中で、発癌やウイルス量等との関連が示唆される miRNA 多型を複数認めた。これらについては、次年度に追認試験を行い、慎重に再現性を確認していく。再現性が確認ができたものについては、表現型に影響を与えるメカニズムを明らかにするための機能解析を予定する。

一方で、網羅的アプローチとしての候補 miRNA 多型の絞込みを行う際に利用するさまざまの表現型に関する GWAS データの解析において、本年度中にいくつかの新たな成果があった。C 型慢性肝炎から肝硬変への進展、C 型慢性肝炎の易罹患性について、HLA の多様性が重要であることを遺伝学的に示すことができた。

また、今後予定され得る miRNA の機能解析を行う上での、*in vitro* あるいは *in vivo* の実験系も段階的に構築を進めている。

D．考察

本研究計画初年度である本年度の研究は概ね当初の予定通り進捗した。現在は候補 miRNA 多型の選択とスクリーニングセットでの genotyping をほぼ終えた段階に過ぎず、次年度以降の追認試験の結果を踏まえて慎重に評価する必要があると考えられるものの、既報のものも含めて、表現型との関連が示唆されるものが複数認められている。

また、GWAS データの解析において得られた「HCV 感染の慢性化には HLA-DQB1*03 を規定する SNP が強く関連している」という知見と、「その HLA と、世界共通で HCV 持続感染に重要と考えられている *IFNL4* 多型とが相互作用している可能性がある」という知見を得たため、これついても、現在、解析中である。この経路における miRNA の持つ影響力についても、多型解析の視点から迫ってみたい。

ウイルス性肝疾患の多様な表現型について、網羅的に miRNA の多型との関連を検討するような、包括的・体系的研究はこれまでに行われておらず、本研究から新しい、臨床に有用な成果を見出せる

可能性は十分にあると考えている。

E．結論

　GWASデータや公共のデータベースを活用して解析候補となるmiRNA関連のSNPを抽出し、実際に相当数のサンプルを用いてgenotypingを行った結果、ウイルス性慢性肝疾患の病態との関連が示唆されるSNPを複数同定することができたが、今後の追認試験の結果がさらに重要である。

　GWAS解析の結果、C型慢性感染においてHLA-DQB1およびIFNL4の遺伝子多型が重要であること、およびそれらの相互作用を示唆する知見を得た。

　ウイルス性肝疾患の様々な病態を明らかにするため、主にmiRNAや免疫系について、遺伝子多型解析を進めて行くこととした。

F．健康危険情報
　なし

G．研究発表
1．論文発表

1) Miki D, Ochi H, Takahashi A, Hayes CN, Urabe Y, Abe H, Kawaoka T, Tsuge M, Hiraga N, Imamura M, Kawakami Y, Aikata H, Takahashi S, Akuta N, Suzuki F, Ikeda K, Kumada H, Karino Y, Toyota J, Tsunoda T, Kubo M, Kamatani N, Nakamura Y, Chayama K. HLA-DQB1*03 confers susceptibility to chronic hepatitis C in Japanese: a genome-wide association study. *PLoS One*. 8(12):e84226, Dec, 2013.

2) Kosaka K, Hiraga N, Imamura M, Yoshimi S, Murakami E, Nakahara T, Honda Y, Ono A, Kawaoka T, Tsuge M, Abe H, Hayes CN, Miki D, Aikata H, Ochi H, Ishida Y, Tateno C,

Yoshizato K, Sasaki T, Chayama K. A novel TK-NOG based humanized mouse model for the study of HBV and HCV infections. *Biochem Biophys Res Commun*. 441(1):230-5, Nov, 2013.

3) Tsuge M, Murakami E, Imamura M, Abe H, Miki D, Hiraga N, Takahashi S, Ochi H, Nelson Hayes C, Ginba H, Matsuyama K, Kawakami H, Chayama K. Serum HBV RNA and HBeAg are useful markers for the safe discontinuation of nucleotide analogue treatments in chronic hepatitis B patients. *J Gastroenterol*. 48(10):1188-204, Oct, 2013.

4) 茶山 一彰, 越智 秀典, 三木 大樹. 肝疾患の疾患関連遺伝子 GWAS を含めた最近の知見から. 日本消化器病学会雑誌 110 巻 9 号 Page1577-1590 (2013.09)

5) Abe H, Hayes CN, Hiraga N, Imamura M, Tsuge M, Miki D, Takahashi S, Ochi H, Chayama K. A translational study of resistance emergence using sequential direct-acting antiviral agents for hepatitis C using ultra-deep sequencing. *Am J Gastroenterol*. 108(9):1464-72, Sep, 2013.

6) Lange CM, Miki D, Ochi H, Nischalke HD, Bojunga J, Bibert S, Morikawa K, Gouttenoire J, Cerny A, Dufour JF, Gorgievski-Hrisoho M, Heim MH, Malinverni R, Müllhaupt B, Negro F, Semela D, Kutalik Z, Müller T, Spengler U, Berg T, Chayama K, Moradpour D, Bochud PY; Hiroshima Liver Study Group; Swiss Hepatitis C Cohort Study Group. Genetic analyses reveal a role for vitamin D insufficiency in HCV-associated hepatocellular carcinoma development. *PLoS One*. 8(5):e64053, May,

2013.

7) Arataki K, Hayes CN, Akamatsu S, Akiyama R, Abe H, Tsuge M, Miki D, Ochi H, Hiraga N, Imamura M, Takahashi S, Aikata H, Kawaoka T, Kawakami H, Ohishi W, Chayama K. Circulating microRNA-22 correlates with microRNA-122 and represents viral replication and liver injury in patients with chronic hepatitis B. *J Med Virol.* 85(5):789-98, May, 2013.

8) Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, Hosono N, Otsuka M, Tateishi R, Lo PH, Tanikawa C, Omata M, Koike K, Miki D, Abe H, Kamatani N, Toyota J, Kumada H, Kubo M, Chayama K, Nakamura Y, Matsuda K. A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. *J Hepatol.* 58(5):875-82, May, 2013.

9) Lo PH, Urabe Y, Kumar V, Tanikawa C, Koike K, Kato N, Miki D, Chayama K, Kubo M, Nakamura Y, Matsuda K. Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk. *PLoS One.* 8(4):e61279, Apr, 2013.

10) Nakano Y, Chayama K, Ochi H, Toshishige M, Hayashida Y, Miki D, Hayes CN, Suzuki H, Tokuyama T, Oda N, Suenari K, Uchimura-Makita Y, Kajihara K, Sairaku A, Motoda C, Fujiwara M, Watanabe Y, Yoshida Y, Ohkubo K, Watanabe I, Nogami A, Hasegawa K, Watanabe H, Endo N, Aiba T, Shimizu W, Ohno S, Horie M, Arihiro K, Tashiro S, Makita N, Kihara Y. A nonsynonymous polymorphism in semaphorin 3A as a risk factor for human unexplained cardiac arrest with documented ventricular fibrillation. *PLoS Genet.* 9(4):e1003364, Apr, 2013.

2. 学会発表

1) Miki D, Ochi H, Hayes CN, Abe H, Kawaoka T, Tsuge M, Hiraga N, Imamura M, Kawakami Y, Aikata H, Takahashi S, Akuta N, Suzuki F, Ikeda K, Kumada H, Karino Y, Toyota J, Chayama K. HLA-DQ confers susceptibility to chronic hepatitis C in Japanese: a genome-wide association study. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/5

2) Hayes CN, Akamatsu S, Tsuge M, Miki D, Hiraga N, Abe H, Imamura M, Takahashi S, Ochi H, Chayama K. Integrative analysis of early and late changes in microRNA and gene expression profiles in human hepatocyte chimeric mice following infection with hepatitis B virus or hepatitis C virus. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/5

3) Karino Y, Nakajima T, Ozeki I, Hige S, Kimura M, Arakawa T, Kuwata Y, Sato T, Ohmura T, Toyota J, Ochi H, Miki D, Chayama K. Analysis of the usefulness of HCV core amino acid substitution, serum IP-10 level and IFNL4 genotype as the predictive factor of anti-HCV effect in triple therapy including protease inhibitor. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/5

4) Kawaoka T, Imamura M, Aikata H, Miki D,

Ochi H, Tashiro H, Ohdan H, Chayama K. Relationship between survival rate at 10 years and virological response to IFN therapy in patients with recurrent hepatitis C after living donor liver transplantation. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/4

5) Hayes CN, Akamatsu S, Tsuge M, <u>Miki D</u>, Hiraga N, Abe H, Imamura M, Takahashi S, Ochi H, Chayama K. Early and late changes in gene expression profiles following infection with hepatitis B or C virus in human hepatocyte chimeric mice. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/4

6) Abe H, Hayes CN, Hiraga N, Imamura M, <u>Miki D</u>, Tsuge M, Ochi H, Chayama K. Expression levels of IFNL4 and other IFN-lambdas are determined by dinucleotide polymorphism of IFNL4 and correlate with basal expression levels of interferon stimulated genes and effect of interferon. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/4

7) Tsuge M, Murakami E, Imamura M, Abe H, <u>Miki D</u>, Hiraga N, Ochi H, Hayes CN, Ginba H, Matsuyama K, Kawakami H, Chayama K. Monitoring serum HBV RNA is useful for predicting rebound of hepatitis after the discontinuation of nucleotide analogue therapy in chronic hepatitis B patients. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/3

8) Murakami E, Tsuge M, Hiraga N, Kawaoka T, Ono A, <u>Miki D</u>, Abe H, Imamura M, Aikata H, Ochi H, Hayes CN, Chayama K. Antiviral effect of tenofovir disoproxil fumarate on drug-resistant HBV clones and different susceptibility between HBV genotype A and C. *The Liver Meeting 2013*. Washington DC, USA. 2013/11/3

9) <u>三木大樹</u>、越智秀典、茶山一彰. 肝細胞癌における宿主および腫瘍のゲノムワイド解析. *JDDW2013*. 東京. 2013/10/9

10) <u>三木大樹</u>、相方浩、卜部祐司、柘植雅貴、川上由育、中村祐輔、茶山一彰. DEPDC5遺伝子座位は日本人のHBV関連肝細胞癌およびC型慢性肝炎とは関連しない. *第72回日本癌学会学術総会*. 東京. 2013/10/5

11) 阿部弘美、ヘイズ・ネルソン、平賀伸彦、今村道雄、柘植雅貴、<u>三木大樹</u>、高橋祥一、越智秀典、茶山一彰. ヒト肝細胞キメラマウスと次世代シーケンサーを用いた新規抗ウイルス薬に対するHCVゲノムの解析. *第49回日本肝臓学会総会*. 東京. 2013/6/7

12) <u>三木大樹</u>、大石和佳、越智秀典、ヘイズ・ネルソン、阿部弘美、柘植雅貴、今村道雄、鎌谷直之、中村祐輔、茶山一彰. HCV Genotype 1・高ウイルス量症例に対するペグ IFNα2b・リバビリン併用療法の治療効果と血中アディポサイトカイン濃度についての検討. *第50回日本臨床分子医学会学術集会*. 東京. 2013/4/12

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

研究成果の刊行に関する一覧表

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Miki D, Ochi H, Takahashi A, Hayes CN, Urabe Y, Abe H, Kawaoka T, Tsuge M, Hiraga N, Imamura M, Kawakami Y, Aikata H, Takahashi S, Akuta N, Suzuki F, Ikeda K, Kumada H, Karino Y, Toyota J, Tsunoda T, Kubo M, Kamatani N, Nakamura Y, Chayama K.	HLA-DQB1*03 confers susceptibility to chronic hepatitis C in Japanese: a genome-wide association study.	PLoS One	8(12)	e84226	2013
Kosaka K, Hiraga N, Imamura M, Yoshimi S, Murakami E, Nakahara T, Honda Y, Ono A, Kawaoka T, Tsuge M, Abe H, Hayes CN, Miki D, Aikata H, Ochi H, Ishida Y, Tateno C, Yoshizato K, Sasaki T, Chayama K.	A novel TK-NOG based humanized mouse model for the study of HBV and HCV infections.	Biochem Biophys Res Commun	441(1)	230-5	2013
Tsuge M, Murakami E, Imamura M, Abe H, Miki D, Hiraga N, Takahashi S, Ochi H, Nelson Hayes C, Ginba H, Matsuyama K, Kawakami H, Chayama K.	Serum HBV RNA and HBeAg are useful markers for the safe discontinuation of nucleotide analogue treatments in chronic hepatitis B patients.	J Gastroenterol	48(10)	1188-204	2013

茶山 一彰, 越智 秀典, 三木 大樹.	肝疾患の疾患関連遺伝子 GWASを含めた最近の知見から.	日本消化器病学会雑誌	110巻9号	1577-90	2013
Abe H, Hayes CN, Hiraga N, Imamura M, Tsuge M, Miki D, Takahashi S, Ochi H, Chayama K.	A translational study of resistance emergence using sequential direct-acting antiviral agents for hepatitis C using ultra-deep sequencing.	Am J Gastroenterol	108(9)	1464-72	2013
Lange CM, Miki D, Ochi H, Nischalke HD, Bojunga J, Bibert S, Morikawa K, Gouttenoire J, Cerny A, Dufour JF, Gorgievski-Hrisoho M, Heim MH, Malinverni R, Müllhaupt B, Negro F, Semela D, Kutalik Z, Müller T, Spengler U, Berg T, Chayama K, Moradpour D, Bochud PY; Hiroshima Liver Study Group; Swiss Hepatitis C Cohort Study Group.	Genetic analyses reveal a role for vitamin D insufficiency in HCV-associated hepatocellular carcinoma development.	PLoS One	8(5)	e64053	2013
Arataki K, Hayes CN, Akamatsu S, Akiyama R, Abe H, Tsuge M, Miki D, Ochi H, Hiraga N, Imamura M, Takahashi S, Aikata H, Kawaoka T, Kawakami H, Ohishi W, Chayama K.	Circulating microRNA-22 correlates with microRNA-122 and represents viral replication and liver injury in patients with chronic hepatitis B.	J Med Virol	85(5)	789-98	2013

Urabe Y, Ochi H, Kato N, Kumar V, Takahashi A, Muroyama R, Hosono N, Otsuka M, Tateishi R, Lo PH, Tanikawa C, Omata M, Koike K, Miki D, Abe H, Kamatani N, Toyota J, Kumada H, Kubo M, Chayama K, Nakamura Y, Matsuda K.	A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region.	J Hepatol	58(5)	875-82	2013	
Lo PH, Urabe Y, Kumar V, Tanikawa C, Koike K, Kato N, Miki D, Chayama K, Kubo M, Nakamura Y, Matsuda K.	Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk.	PLoS One	8(4)	e61279	2013	
Nakano Y, Chayama K, Ochi H, Toshishige M, Hayashida Y, Miki D, Hayes CN, Suzuki H, Tokuyama T, Oda N, Suenari K, Uchimura-Makita Y, Kajihara K, Sairaku A, Motoda C, Fujiwara M, Watanabe Y, Yoshida Y, Ohkubo K, Watanabe I, Nogami A, Hasegawa K, Watanabe H, Endo N, Aiba T, Shimizu W, Ohno S, Horie M, Arihiro K, Tashiro S, Makita N, Kihara Y.	A nonsynonymous polymorphism in semaphorin 3A as a risk factor for human unexplained cardiac arrest with documented ventricular fibrillation.	PLoS Genet	9(4)	e1003364	2013	

HLA-DQB1*03 Confers Susceptibility to Chronic Hepatitis C in Japanese: A Genome-Wide Association Study

Daiki Miki[1,2,3©], Hidenori Ochi[1,2,3©], Atsushi Takahashi[4], C. Nelson Hayes[1,2,3], Yuji Urabe[2], Hiromi Abe[1,2], Tomokazu Kawaoka[1,2,3], Masataka Tsuge[3,5], Nobuhiko Hiraga[1,2,3], Michio Imamura[1,2,3], Yoshiiku Kawakami[1,2,3], Hiroshi Aikata[1,2,3], Shoichi Takahashi[2,3], Norio Akuta[6], Fumitaka Suzuki[6], Kenji Ikeda[6], Hiromitsu Kumada[6], Yoshiyasu Karino[7], Joji Toyota[7], Tatsuhiko Tsunoda[8], Michiaki Kubo[9], Naoyuki Kamatani[4], Yusuke Nakamura[10], Kazuaki Chayama[1,2,3*]

1 Laboratory for Digestive Diseases, Center for Integrative Medical Sciences, RIKEN, Hiroshima, Japan, 2 Department of Gastroenterology and Metabolism, Applied Life Sciences, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan, 3 Liver Research Project Center, Hiroshima University, Hiroshima, Japan, 4 Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan, 5 Natural Science Center for Basic Research and Development, Hiroshima University, Hiroshima, Japan, 6 Department of Hepatology, Toranomon Hospital, Tokyo, Japan, 7 Department of Gastroenterology, Sapporo Kosei General Hospital, Hokkaido, Japan, 8 Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan, 9 Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan, 10 Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, University of Tokyo, Tokyo, Japan

Abstract

Hepatitis C virus (HCV) establishes a chronic infection in 70-80% of infected individuals. Many researchers have examined the effect of human leukocyte antigen (HLA) on viral persistence because of its critical role in the immune response against exposure to HCV, but almost all studies have proven to be inconclusive. To identify genetic risk factors for chronic HCV infection, we analyzed 458,207 single nucleotide polymorphisms (SNPs) in 481 chronic HCV patients and 2,963 controls in a Japanese cohort. Next, we performed a replication study with an independent panel of 4,358 cases and 1,114 controls. We further confirmed the association in 1,379 cases and 25,817 controls. In the GWAS phase, we found 17 SNPs that showed suggestive association ($P < 1 \times 10^{-5}$). After the first replication study, we found one intronic SNP in the HLA-DQ locus associated with chronic HCV infection, and when we combined the two studies, the association reached the level of genome-wide significance. In the second replication study, we again confirmed the association ($P_{combined} = 3.59 \times 10^{-16}$, odds ratio [OR] = 0.79). Subsequent analysis revealed another SNP, rs1130380, with a stronger association (OR=0.72). This nucleotide substitution causes an amino acid substitution (R55P) in the HLA-DQB1 protein specific to the DQB1*03 allele, which is common worldwide. In addition, we confirmed an association with the previously reported IFNL3-IFNL4 locus and propose that the effect of DQB1*03 on HCV persistence might be affected by the IFNL4 polymorphism. Our findings suggest that a common amino acid substitution in HLA-DQB1 affects susceptibility to chronic infection with HCV in the Japanese population and may not be independent of the IFNL4 genotype.

Introduction

At least 1.5 million people in Japan and more than 200 million people worldwide are chronically infected with the hepatitis C virus (HCV) [1,2]. In many western countries and Japan, HCV infection is the most common risk factor for hepatocellular carcinoma (HCC) [3,4]. HCV establishes a chronic infection in approximately 70-80% of infected individuals [5,6], and host genetic factors in addition to viral factors are assumed to partially explain the heterogeneity in HCV persistence or clearance. Many researchers have examined the effect of human leukocyte antigen (HLA) on viral

Table 1. Summary of GWAS and replication studies.

SNP		Case			Control			MAF					
locus	Study	TT	TC	CC	TT	TC	CC	Case	Control	OR	(95%CI)[a]	P[b]	P_{het}[c]
rs9275572	GWAS	96	217	168	366	1317	1280	0.425	0.346	0.71	(0.62-0.82)	2.62E-06	
HLA-DQB1	1st replication	634	2049	1673	119	474	519	0.381	0.320	0.77	(0.69-0.85)	1.40E-07	
	2nd replication	206	695	477	3279	11284	10260	0.402	0.359	0.84	(0.77-0.90)	7.63E-06	
	Combined GWAS and 1st replication study[d]									0.75	(0.69-0.81)	2.04E-12	4.25E-01
	Combined all studies[e]									0.79	(0.75-0.84)	3.59E-16	1.13E-01

MAF; minor allele frequency, OR; odds ratio, CI; confidence interval.

Odds ratios and P values for independence test were calculated by the Mantel-Haenszel method.

[a]Odds ratios of protective allele (C) from two-by-two allele frequency table. [b]P value of Cochran-Armitage trend test.

[c]Result of Breslow-Day test. [d]Meta-analysis of GWAS and 1st replication study. [e]Meta-analysis of all three studies.

doi: 10.1371/journal.pone.0084226.t001

persistence because of its critical role in the immune response to HCV [7-10] and because numerous HLA class II alleles have been reported to affect HCV clearance; however, results of almost all such studies have been inconclusive. On the other hand, a strong association between HCV clearance and variants close to interferon-λ3 (IFNL3), previously known as IL28B, has been consistently replicated in multiple studies, including two genome-wide association studies (GWAS). Those studies have been performed mainly using individuals of European and African, but not Asian, ancestry [11-13]. Recently it was reported that a dinucleotide polymorphism (ss469415590) creates or disrupts an open reading frame belonging to a newly discovered gene, IFNL4 [14]. It is well known that there are great differences among ethnic groups in allele frequencies in both HLA class II and IFNL4 variants or variants from the IFNL3-IFNL4 region. To identify genetic markers associated with chronic HCV infection in the Japanese population, we conducted a case control study consisting of a GWAS and two replication studies that included 6,218 cases with chronic HCV infection and 29,894 controls for a total of 36,112 Japanese individuals. In Japan, the prevalence of chronic hepatitis C is approximately 1%, and it was reported that the frequency of individuals testing positive for antibody to HCV (anti-HCV) was 0.49% among 3,485,648 Japanese first-time blood donors [15]. As the risk of exposure to HCV is low in Japan and the potential occurrence of HCV infection in controls is unlikely to affect the results, we tried to detect loci involved in susceptibility to HCV persistence using a large number of individuals in the general population as controls.

Results

GWAS and Replication Studies

In the GWAS phase, single nucleotide polymorphism (SNP) genotyping was performed using the Illumina HumanHap610-Quad BeadChip for cases and the Illumina HumanHap550v3 BeadChip for controls. Genotype concordance between these two BeadChips was 99.99% among 182 duplicated samples, indicating the low possibility of genotyping error. 458,207 SNPs passed quality control filters and were analyzed using an additive model for genotype-phenotype association in 481

patients with chronic HCV infection and 2,963 controls. Principal component analysis revealed no population substructure in our population. In addition, a quantile-quantile plot using the results of the Cochran-Armitage trend test showed that the inflation factor, λ, was 1.007, indicating a low probability of false-positive associations resulting from population stratification (Figure S1A). Using the additive model, one SNP reached the genome-wide significance level for association after Bonferroni correction (calculated as $P < 0.05/458,207 = P < 1.09 \times 10^{-7}$), and another 24 SNPs showed suggestive association ($P < 1 \times 10^{-5}$) in our GWAS (Figure S1B and Table S1). Next, we conducted the first replication study to validate the results of the GWAS phase using 4,358 cases and 1,114 controls. We performed genotyping of 18 SNPs with P values $< 1 \times 10^{-5}$ in the GWAS phase, after excluding 7 SNPs with $r^2 = 1$ in the same locus based on the HapMap-JPT database, and successfully genotyped all SNPs using multiplex PCR-based Invader assay (Table S2), except for rs9275563, which is moderately linked with rs9275572 ($r^2 = 0.56$). Only rs9275572 in the HLA-DQ locus on chromosome 6 showed significant replication of the association ($P < 0.05/17$), and the association reached a genome-wide significance level when we combined the two studies using the Mantel-Haenszel method ($P = 2.04 \times 10^{-12}$, odds ratio [OR] = 0.75; Table 1). We further confirmed the association in 1,379 cases and 25,817 controls and found a highly significant association with chronic HCV after meta-analysis of all three studies ($P_{combined} = 3.59 \times 10^{-16}$, OR = 0.79, 95% confidence interval [CI] 0.75-0.84). As shown in Table 1, we observed no heterogeneity among the three studies (heterogeneity test $P = 0.113$).

Multiple Logistic Regression Analysis and Stratified Analysis

We next evaluated the association in more detail using the first replication set. Both HCV case and healthy control samples were collected at Hiroshima University. After adjusting for gender and age using multiple logistic regression analysis, the rs9275572 C allele remained highly significant with an OR = 0.79 (95% CI 0.70-0.89) (Table S3). Our study also confirmed a stronger protective effect of female gender compared to male gender [12,16]. Subsequently, we analyzed

the effect of rs9275572 according to gender and age. As shown in Table S4A, we observed no significant difference between subgroups. In addition to host genetic factors, such as HLA alleles, the distribution of viral genotypes is also highly variable around the world, and it has been suggested that different HCV genotypes result in different clinical outcomes [17,18]. For example, HCV genotype 1, the most common genotype in Japan, shows stronger resistance against IFN-based anti-viral therapy than HCV genotype 2, the second most frequently encountered genotype in Japan [19,20]. Hence, we stratified HCV cases by HCV genotype and compared each with healthy controls. We observed that the effect tended to be slightly greater in genotype 1 (OR = 0.74 [0.66-0.82]) than genotype 2 (OR = 0.83 [0.74-0.94]), but there was no significant difference (heterogeneity test P = 0.127) (Table S4B).

Detailed Analysis of HLA-DQ alleles

Other than rs9275572, which is located within the HLA-DQ locus (Figure S2), several SNPs within the major histocompatibility complex (MHC) region showed suggestive associations but failed to be validated in a replication study. The linkage disequilibrium (LD) structure around the HLA-DQ region indicates that the landmark SNP is located in an LD block containing HLA-DQB1 and HLA-DQA1 (Figure 1 and Figure S3). Therefore, we next genotyped HLA-DQA1 and HLA-DQB1 alleles by direct sequencing of exon 2 using all 1,114 healthy controls of the first replication set and the same number of cases with HCV genotype 1 obtained from the same set and collected from the same geographical area. Over 95% of samples were successfully genotyped for each HLA-DQ, and common alleles with frequencies of more than 1 % were determined (7 and 11 alleles for HLA-DQA1 and HLA-DQB1, respectively) (Table 2, Table S5A and S5B). Although no allele showed significant association (P < 0.05/7) among HLA-DQA1 alleles, the HLA-DQB1*0303 allele showed significant association among HLA-DQB1 alleles (P < 0.05/12, OR = 0.69). We also performed haplotype analysis using 1,025 cases and 1,038 controls that had been successfully genotyped for HLA-DQA1, HLA-DQB1 and rs9275572 to investigate the effects of haplotype combinations; however, no haplotype showed stronger association than the single association of HLA-DQB1*0303 (Table S6). In our Japanese cohort, DQB1*0303 was the most frequent of the DQB1*03 alleles and showed the strongest effect against HCV persistence, but we also found that all DQB1*03 alleles tended to have a protective effect. Hence, we speculate that not only DQB1*0303, but all DQB1*03 alleles share protective residues. We searched for a variant that is common to all DQB1*03 alleles but that differs from non-DQB1*03 alleles and found that SNP rs1130380 is a variant specific to DQB1*03 (Table S5A and S5B). Then, we evaluated the single marker effect of several nucleotide variants, which are useful to distinguish among common DQB1 alleles using the same individuals with the haplotype analysis, and finally determined that rs1130380 (P = 6.08 × 10⁻⁷, OR = 0.72) had a stronger effect than the landmark SNP, rs9275572 (P = 2.07 × 10⁻⁵, OR = 0.76) (Table S7). After mutual adjustment between rs1130380 and

rs9275572 using multiple logistic regression analysis, the association for both variants was remarkably attenuated (P = 2.38 × 10⁻³ and P = 0.0534, respectively), suggesting their mutual linkage (D' = 0.99, r^2 = 0.28) (Figure S4). Therefore, these variants probably represent the same genetic signal and rs1130380 could be more compatible as a representative variant of the signal. This nucleotide polymorphism causes an amino acid substitution at position 55 on the HLA-DQB1 molecule. We examined the position of this amino acid substitution using on a three-dimensional structure of HLA-DQ molecule that was determined by X-ray diffraction method [21]. As shown in Figure S5, the amino acid at β55 forms part of a peptide-binding pocket in the HLA-DQ molecule, suggesting a critical role in antigen presentation. The change from Arg to Pro implies a physicochemcial change from basic to hydrophobic (Figure S6), but we have no relevant data with which to show a change of affinity to certain peptides at this time. Further functional analysis is therefore needed.

Evaluation of IFNL3-IFNL4 locus in GWAS

In spite of our failure to detect it in our GWAS screen, variants close to IFNL3 have been reported to be associated with spontaneous clearance of HCV, so we tried to confirm the association using our GWAS data. Because genotype data for rs12979860 was not available in our platform, we checked the genotype data using another frequently reported SNP, rs8099917, which is highly linked with rs12979860 in the Japanese population (r^2 = 0.94) [22,23]. As shown in Table S8, the association was confirmed in our GWAS set (P = 3.22 × 10⁻³, OR = 0.73) with the same allele conferring a protective effect.

Analysis of IFNL3-IFNL4 locus and HLA-DQ*03

We genotyped a IFNL4 variant, ss469415590, and two representative variants close to IFNL3, rs12979860 and rs8099917, by Invader assay using the same individuals used in the detailed analysis of the HLA-DQ alleles above (1,114 cases and 1,114 controls). Each of these three polymorphisms were successfully genotyped in 1,074 cases and 1,073 controls. rs12979860 and ss469415590 were in almost complete LD, and ss469415590 showed the strongest association with HCV persistence under the additive model (P = 2.96 × 10⁻³, OR = 0.75) (Table 3). After adjusting for the effect of the HLA-DQ*03 allele using multiple logistic regression analysis, the significance level of ss469415590 increased slightly (P = 1.67 × 10⁻³, OR=0.74). We additionally analyzed the association under different genetic models other than additive. Among several genetic models, the IFNL4 variant, ss469415590, showed the strongest association under the recessive model for the protective allele (the dominant model for the risk allele), whereas rs1130380 showed the strongest association under the allelic model (Table S9). We further evaluated the effect of the DQB1*03 allele stratified by ss469415590 genotype (Table 4). Interestingly, DQB1*03 showed a stronger effect in individuals with the ss469415590-TT/TT genotype, and this tendency became more pronounced for rs9275572, the landmark SNP in our GWAS (heterogeneity test P = 2.45 × 10⁻³).
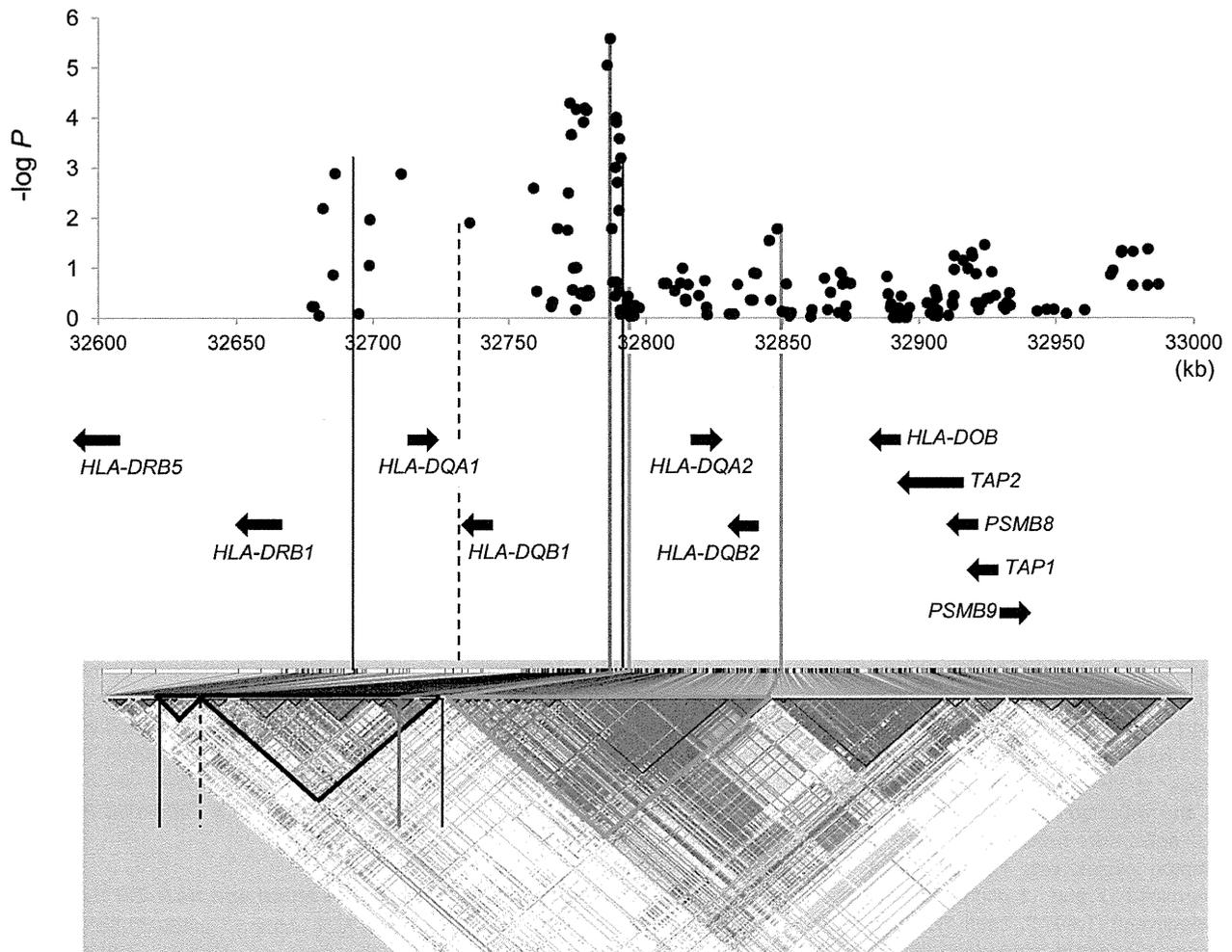
Figure 1. Linkage disequilibrium structure around the HLA-DQ region based on _D'_ using HapMap-JPT Data. P-value plot and genomic structure of the GWAS stage. Blue lines represent the LD block containing the _HLA-DQA2_ and _HLA-DQB2_ loci. Black lines represent two LD blocks containing the _HLA-DQA1_ and _HLA-DQB1_ loci, and the black dotted line represents the boundary between them. The red line represents the most strongly associated SNP, rs9275572, which is located within the _HLA-DQB1_ locus. The LD maps were created using HaploView software.

doi: 10.1371/journal.pone.0084226.g001

Discussion

The MHC region containing the _HLA_ locus is widely believed to play an important role in viral infection, and a recent GWAS revealed a strong association between the HLA-DP and _DQ_ locus and hepatitis B virus (HBV) persistence [24,25] as well as HBV-induced HCC [26] in East Asia. In our GWAS with two replication studies, we aimed to clarify the genetic factors involved in chronic HCV infection, and we successfully identified an association with the _HLA-DQB1_ locus based on the landmark SNP rs9275572. An association between the MICA locus and the risk of developing HCV-induced HCC was previously reported in a Japanese GWAS [27], and this study identified rs9275572 as a second associated SNP. HCV-positive cases in the previous study, which are independent of

ours, showed a similar minor allele frequency (MAF) of rs9275572 to our cases (0.407 in GWAS stage and 0.429 in replication stage), which strongly supports our results. Furthermore, according to the 1000 Genomes database, the MAF of this SNP is 0.348 in the Japanese population, which is consistent with the MAF of our controls.

On the other hand, while we observed no significant heterogeneity across the three studies, there were some differences in MAF among studies in both case and control populations. For both cases and controls, MAFs were lower in the first replication study than in the other two studies. Among the three studies, only the first replication set was collected from the same geographical area, therefore, geographical differences might affect the differences in MAF among studies, but this is only speculation. There are also differences in the

Table 3. Case-control association analysis of 5 variants using 1,074 chronic hepatitis C cases and 1,073 healthy controls.

SNP	Allele [1/2]	Case 11	Case 12	Case 22	Control 11	Control 12	Control 22	HWE-p[a] Case	HWE-p[a] Control	Protective AF Case	Protective AF Control	OR[b]	(95%CI)	P[c]
rs8099917	G/T	15	235	824	10	188	875	0.704	0.978	0.877	0.903	0.76	(0.63-0.92)	5.53E-03
rs12979860	T/C	16	244	814	11	194	868	0.637	0.965	0.872	0.899	0.76	(0.63-0.92)	4.08E-03
ss469415590	ΔG/TT	16	244	814	11	192	870	0.637	0.911	0.872	0.900	0.75	(0.62-0.91)	2.96E-03
rs9275572	T/C	166	492	416	112	459	502	0.305	0.641	0.616	0.682	0.75	(0.66-0.85)	9.64E-06
rs1130380	C/G	120	433	521	172	478	423	0.039	0.060	0.313	0.383	0.73	(0.65-0.83)	3.38E-06

AF; allele frequency, OR; odds ratio, CI; confidence interval.
[a]P value of Hardy-Weinberg equilibrium test. [b]Odds ratio of protective allele. [c]P value of Cochran-Armitage trend test.
doi: 10.1371/journal.pone.0084226.t003

Table 2. Effect of common *HLA-DQA1* and *HLA-DQB1* haplotypes on chronic hepatitis C.

		Case n	Case freq.(%)	Control n	Control freq.(%)	OR[a]	(95%CI)	P[b]
DQA1	*0101	304	14.3	320	14.4	0.99	(0.84-1.18)	9.53E-01
	*0102	302	14.2	263	11.8	1.24	(1.04-1.48)	1.84E-02
	*0103	445	21.0	423	19.0	1.13	(0.97-1.31)	1.08E-01
	*0301	830	39.1	924	41.5	0.90	(0.80-1.02)	1.02E-01
	*0401	65	3.1	58	2.6	1.18	(0.82-1.69)	3.66E-01
	*0501	138	6.5	177	8.0	0.80	(0.64-1.01)	6.44E-02
	*0601	30	1.4	49	2.2	0.64	(0.40-1.01)	5.15E-02
DQB1	*0301	181	8.4	207	9.6	0.86	(0.70-1.06)	1.69E-01
	*0302	220	10.2	245	11.4	0.89	(0.73-1.08)	2.24E-01
	*0303	272	12.6	372	17.3	0.69	(0.59-0.82)	2.05E-05
	*0401	305	14.2	282	13.1	1.10	(0.92-1.30)	3.01E-01
	*0402	114	5.3	84	3.9	1.38	(1.03-1.84)	2.86E-02
	*0501	147	6.8	168	7.8	0.87	(0.69-1.09)	2.22E-01
	*0502	59	2.7	61	2.8	0.97	(0.67-1.39)	8.57E-01
	*0503	116	5.4	78	3.6	1.52	(1.13-2.03)	5.14E-03
	*0601	444	20.6	409	19.0	1.11	(0.95-1.29)	1.76E-01
	*0602	155	7.2	128	5.9	1.23	(0.96-1.57)	9.52E-02
	*0604	119	5.5	98	4.5	1.23	(0.93-1.62)	1.42E-01

[a]Odds ratios of haplotype from two-by-two frequency table. [b]P value of chi-squared test.
Only alleles with frequencies of more than 1% were shown.
doi: 10.1371/journal.pone.0084226.t002

gender and age distribution among studies (Table S10). Nonetheless, it is difficult to find a suitable explanation for the differences in MAF at this time.

In the current study, we failed to detect the previously most replicated *IFNL3-IFNL4* locus during GWAS screening even though the association indeed existed in our GWAS set ($P = 3.22 \times 10^{-3}$, OR = 0.73; Table S8). The odds ratio for rs8099917 is nearly the same as for rs9275572, but the P values were quite different, with rs9275572 ranked in the top 9 candidate SNPS, whereas the other is far below at number 1,640. This crucial difference may be explained by differences in allele frequencies. The minor allele frequency of rs9275572 is over three times higher than that of rs8099917, and the statistical power to detect each SNP in our GWAS was calculated to be only 65% and 8%, respectively, compared to the 80% recommended to detect an association of the expected effect size. We are concerned that there are several remaining undetected SNPs that might affect HCV persistence or clearance, in addition to the *IFNL3-IFNL4* locus. To overcome this problem, it will be necessary to use a larger sample size, as well as more appropriate control samples. In fact, the major limitation of our study involves the selection of controls. Undoubtedly, the most appropriate control set would include patients who have been exposed to HCV (anti-HCV positive) but who spontaneously cleared the infection (HCV-RNA negative). Although the risk of exposure to HCV in the Japanese general population is very low, making it difficult to locate suitable control subjects, we admit concede that our choice of study design has reduced the statistical power of the study and might have undermined the conclusions.

HLA-DQs belong to the HLA class II molecules that form α-β heterodimers on the cell surface of antigen-presenting cells, such as macrophages, dendritic cells and B lymphocytes. HLA-DQ genes play a central role in immune-mediated diseases by presenting peptides derived from extracellular proteins to CD4-positive T lymphocytes. Both *HLA-DQB1* and *HLA-DQA1* are highly polymorphic, especially in exon 2, which encodes antigen-binding sites that confer binding specificity. We speculated that the association of the landmark SNP with chronic hepatitis C might reflect variation in HLA-DQ antigen-binding sites and that somewhat different effects of rs9275572 between HCV genotypes might reflect differences in binding affinities between the HLA-DQ pocket and HCV genotype-specific epitopes (Table S4B). Finally, we found that a single amino acid polymorphism (R55P) corresponding to DQB1*03,

Table 4. Effect of HLA-DQB1*03 and rs9275572 on the susceptibility of chronic hepatitis C stratified by genotype of ss469415590.

ss469415590	Allele	Case			Control			Protective AF					
	[1/2]	11	12	22	11	12	22	Case	Control	OR	(95%CI)[a]	P[b]	P_het[c]
TT/TT	HLA-DQB1	80	330	404	137	386	347	0.301	0.379	0.70	(0.61-0.81)	3.09E-06	3.36E-01
∆G/TT or ∆G/∆G	[*03 / non-*03]	40	103	117	35	92	76	0.352	0.399	0.82	(0.63-1.07)	1.61E-01	
TT/TT	rs9275572	134	389	291	89	367	414	0.596	0.687	0.67	(0.59-0.78)	6.31E-08	2.45E-03
∆G/TT or ∆G/∆G	[T / C]	32	103	125	23	92	88	0.679	0.660	1.09	(0.83-1.43)	5.56E-01	

AF; allele frequency, OR; odds ratio, CI; confidence interval.

[a]Odds ratios of protective allele from two-by-two allele frequency table. [b]P value of Cochran-Armitage trend test. [c]Result of Breslow-Day test.

a very common allele worldwide [28], has the strongest protective effect against HCV persistence. Among DQB1*03 alleles, DQB1*0301 has the highest frequency various populations worldwide and has been reported in association with viral clearance in several European and American studies [7,13]. A suggestive paper by Cangussu et al. found an association between protection against HCV persistence and not only DQB1*0301 but also DQB1*0302 and DQB1*0303 in a white Brazilian cohort [29]. They suggested that HLA-DQ beta molecules responsible for protection are present not only in molecules encoded by DQB1*0301 but also by DQB1*0302 and DQB1*0303, and the common polymorphic residues shared by DQB1*03 alleles are responsible for the selection of particular HCV epitopes. In the current study, we identified that not only the allele with the strongest effect, DQB1*0303, but all DQB1*03 alleles share protective residues at position 55 in at least the Japanese and white Brazilian populations. Although the human MHC region encompasses a complex and extended LD structure, and LD among SNPs in the HLA-class II locus is known to vary among ethnic populations, various immunological functions of MHC molecules have gradually become clear. Several issues had to be overcome in order improve our understanding of whether or not the HLA-DQB1 molecule has common functions across multiple populations.

In this study, we also made an intriguing observation that HLA-DQB1*03 showed a stronger effect in individuals with the ss469415590-TT/TT genotype (Table 4), suggesting that the effect of HLA-DQB1*03 on HCV persistence might be affected by the genotype of the IFNL4 polymorphism. The fact that in the Japanese population the frequency of the ss469415590-TT/TT genotype is nearly equal to that of the rs12979860-CC genotype (JPT 78.7%, CEU 57.6%, AFR 15.4% based on the 1000 Genomes database), with a higher frequency than most other populations, might partially explain the differences among the top loci in European versus Japanese GWAS results. Finally, we determined the relationship between HLA-DQB1*03 (rs1130380) and IFNL4 variants (ss469415590) in cases and controls (Figure S7A and S7B). In cases, the proportion of individuals with the protective ss469415590-TT/TT genotype significantly increased while the proportion of protective DQB1*03 alleles decreased, but this tendency was not observed in controls. During any phase of HCV infection, these

two polymorphisms might affect each other in complementary or exclusive fashion.

A recent GWAS reported that variants close to IFNL3 and DQB1*0301 are independently associated with spontaneous resolution of HCV infection [13]. On the other hand, our study newly suggests that a single amino acid polymorphism, common throughout the world, at position 55 on the HLA-DQB1 molecule has the strongest association with chronic HCV infection and may not be independent of the IFNL4 polymorphism.

In summary, we conducted a GWAS followed by two independent replication studies and detected genetic variants in the HLA-DQ genes strongly associated with chronic hepatitis C in the Japanese population. Our findings suggest that variation in antigen-binding sites involved in antigen presentation on HLA-DQ molecules might play an important role in HCV persistence or clearance. Further research is required to determine the distinct roles of HLA-DQB1 and IFNL4 in chronic HCV infection.

Materials and Methods

Study Population

Characteristics of each case-control group are shown in Table S10. Case samples used in this study were obtained between 2002 and 2012 from Toranomon Hospital Department of Hepatology, Sapporo Kosei General Hospital and Hiroshima University-affiliated hospitals. We used 481 samples from Toranomon Hospital as cases for the GWAS and the remaining 1,004 samples, along with 375 samples from Sapporo Kosei General Hospital, as cases for the second replication study. We used all 4,358 samples from Hiroshima University-affiliated hospitals as cases for the first replication study. All cases had abnormal levels of serum alanine aminotransferase for more than 6 months and were positive for both anti-HCV and serum HCV RNA. All patients were negative for hepatitis B surface antigen (HBsAg), had no evidence of other liver diseases, and had not received immunosuppressive therapy before enrollment. Control samples used for the GWAS and second replication studies were obtained from BioBank Japan at the Institute of Medical Science, University of Tokyo [30]. The control groups included 2,057 individuals for the GWAS and

25,817 independent samples for the second replication study and consisted of GWAS samples for unrelated diseases in the BioBank Japan project. We excluded samples that were positive for HBsAg or anti-HCV or that were registered as positive for any liver disease. We also obtained DNA from 906 Japanese control subjects from volunteers without liver disease in the Osaka-Midosuji Rotary Club, Osaka, Japan. In addition, we obtained 1,114 samples as healthy controls for the second replication study from volunteers without any liver diseases at Hiroshima University, Hiroshima, Japan.

Ethics statement

All subjects received a detailed explanation, and all signed a written informed consent form. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki and was approved a priori by the ethical committees at University of Tokyo, Hiroshima University, Toranomon Hospital and Sapporo Kosei General Hospital and by the Ethical Committee at the SNP Research Center, the Institute of Physical and Chemical Research (RIKEN), Yokohama.

SNP Genotyping and Quality Control

Genomic DNA was extracted from peripheral blood leukocytes using a standard method. For the GWAS stage, we genotyped 496 cases using the Illumina HumanHap610-Quad BeadChip (San Diego, CA, USA). We excluded 2 samples with call rates < 0.98. 13 other samples with apparent kinship or sample duplication were excluded from the analysis based on PI_HAT values (> 0.4). We assessed population stratification using the smartpca program in the EIGENSOFT package using SNPs informative for the Japanese population according to a previously described method [31]. Analysis was performed based on the GWAS data and the Japanese (JPT), Han-Chinese (CHB), European (CEU), and African (YRI) individuals from the International HapMap Project. Principal component analysis identified no outliers from the JPT/CHB clusters. 458,207 autosomal SNPs passed the quality control filters (call rate ≥ 0.99 in both cases and controls, a MAF > 0 and a Hardy-Weinberg equilibrium P value ≥ 1.0×10^{-6} in controls). We used multiplex-PCR-based Invader assays (Third Wave Technologies, Madison, WI, USA) for the replication studies [32]. Samples for both cases and controls were distributed randomly on genotyping plates in both phases of the study, and all persons performing genotyping and interpretation of results were blind to case / control status.

HLA-DQA1 and HLA-DQB1 Genotyping

We analyzed HLA-DQ genotypes using 1,114 cases and 1,114 controls in the first replication set. Exon 2 of the HLA-DQA1 and HLA-DQB1 genes were amplified and directly sequenced following the protocol of the International Histocompatibility Workshop Group [33]. HLA-DQA1 and DQB1 alleles were determined based on the alignment database of dbMHC (http://www.ncbi.nlm.nih.gov/gv/mhc/align.cgi?cmd=aligndisplay&locus_name=HLA-DQB1&banner=1#).

Protein Analysis

To represent the position of amino acid substitution caused by the SNP, we edited the image of a three-dimensional structure of HLA-DQ molecule that was previously determined by X-ray diffraction method [21] using Protein Data Bank Japan (PDBj) Viewer (http://www.pdbj.org/index_j.html). The amino acid numbering excludes the 32 amino acid signal peptide.

Statistical Analysis

Genotype-based associations were tested using the Cochran-Armitage trend test. ORs and CIs were calculated from a two-by-two allele frequency table. Combined analysis was performed following the Mantel-Haenszel method. Heterogeneity among studies was examined using the Breslow-Day test. We used HaploView software to analyze the association of haplotypes and LD values between HLA-DQs and SNPs [34]. Chi-squared or Fisher exact tests were used to analyze categorical data, as appropriate. For general statistical analysis, we used the R statistical environment version 2.12.0 (http://www.cran.r-project.org/) or PLINK 1.06 (http://pngu.mgh.harvard.edu/~purcell/plink/) [35]. Power analysis was performed using Power for Genetic Association Analyses software [36]. SNP allele frequencies were collected from the 1000 Genomes database (http://browser.1000genomes.org/).

Supporting Information

Figure S1. Results of the GWAS. (A) Quantile-quantile plot for the Cochran-Armitage trend tests for the GWAS phase. The horizontal axis represents P values expected under a null distribution, and the vertical axis shows the observed P values. Under the null hypothesis of no association at any locus, the points would be expected to fall along the line (y=x). (B) Manhattan plot showing the $-\log_{10} P$ value of each SNP calculated using the 1-d.f. Cochran-Armitage trend test. The red line shows the Bonferroni cutoff for genome-wide significance ($P = 1.09 \times 10^{-7}$) given the number of SNPs analyzed in this study (0.05/458207).
(TIF)

Figure S2. Case-control association results of the MHC region. P-value plot and genomic structure of the GWAS stage within the extended MHC region of chromosome 6. The black dotted line represents the SNP with the strongest association, rs9275572, which is located within the HLA-DQ locus.
(TIF)

Figure S3. Linkage disequilibrium structure around the HLA-DQ region based on D' and r^2 using GWAS data. The red line represents the most strongly associated SNP, rs9275572, which is located within the HLA-DQB1 locus. The LD maps were created using HaploView software.
(TIF)

Figure S4. Linkage disequilibrium structure between rs1130380 and rs9275572 based on D' and r^2 using data

from 2,063 individuals (1,025 cases and 1,038 controls). The LD maps were created using HaploView software. (TIF)

Figure S5. The position of the amino acid at β55 on the HLA-DQ molecule. An edited representation of the three-dimensional structure of HLA-DQ molecule that was previously determined by X-ray diffraction method [21] is shown. Protein Data Bank Japan (PDBj) Viewer (http://www.pdbj.org/index_j.html) was used for editing. The alpha and beta chains are represented by green and yellow, respectively. The amino acid at β55 is indicated as red. Amino acid numbering excludes the 32 amino acid signal peptide. (TIF)

Figure S6. Chromatograms of different alleles of rs1130380 and their predicted effects on protein translation. Two different PCR amplifications were performed for genotyping of *HLA-DQB1* alleles: amplicon 1 for the *02/03/04* alleles and amplicon 2 for the *05/06* alleles [33]. The G to C nucleotide change of codon 55 leads to amino acid substitution from Arg to Pro. The properties of Arg and Pro are basic and hydrophobic, respectively. We also observed the G to T nucleotide change, but this variant was not considered for further analysis because of its low frequency ($< 1\%$). (TIF)

Figure S7. Relationship between *HLA-DQB1*03* (rs1130380) and *IFNL4* variant (ss469415590) in chronic HCV patients (A) and healthy controls (B). Proportion of individuals with ss469415590 TT/TT in cases and controls according to HLA-DQB1*03 status. *P* values were calculated using the chi-squared test. (TIF)

Table S1. Results of GWAS. (PDF)

Table S2. Results of 1st replication study. (PDF)

Table S3. Multiple logistic regression analysis for the risk of chronic hepatitis C using 1st replication samples (4,347 cases and 1,097 controls). (PDF)

Table S4. (A) Effect of rs9275572 on the susceptibility of chronic hepatitis C stratified by gender and age using 1st replication samples. (B) Effect of rs9275572 on the susceptibility of chronic hepatitis C stratified by HCV genotype using 1st replication samples. (PDF)

Table S5. (A) Useful variants for genotyping common *HLA-DQA1* haplotypes in the Japanese population. (B) Useful variants for genotyping common *HLA-DQB1* haplotypes in the Japanese population. (PDF)

Table S6. Haplotype analysis was performed using the landmark SNP (rs9275572) and 12 variants shown in Table S5A and S5B by HaploView software. (PDF)

Table S7. Single marker effect on chronic hepatitis C. (PDF)

Table S8. Results of GWAS of two SNPs. (PDF)

Table S9. Case-control association analysis of 5 variants under different genetic models. (PDF)

Table S10. Basic characteristics of study population. (PDF)

Author Contributions

Conceived and designed the experiments: DM HO MK NK YN KC. Performed the experiments: DM HO YU H. Abe MK. Analyzed the data: DM HO AT TT NK KC. Contributed reagents/materials/analysis tools: TK MT NH MI Y. Karino H. Aikata ST NA FS KI HK Y. Kawakami JT MK NK YN KC. Wrote the manuscript: DM HO CNH KC.

References

1. Chayama K, Hayes CN, Ohishi W, Kawakami Y (2013) Treatment of chronic hepatitis C virus infection in Japan: update on therapy and guidelines. J Gastroenterol 48: 1-12. doi:10.1007/s00535-012-0714-9. PubMed: 23188091.
2. Lavanchy D (2009) The global burden of hepatitis C. Liver Int 29Suppl 1 : 74-81. doi:10.1111/j.1478-3231.2008.01702.x. PubMed: 19207969.
3. Barrera JM, Bruguera M, Ercilla MG, Gil C, Celis R et al. (1995) Persistent hepatitis C viremia after acute self-limiting posttransfusion hepatitis C. Hepatology 21: 639-644. doi:10.1002/hep.1840210306. PubMed: 7533121.
4. Welzel TM, Morgan TR, Bonkovsky HL, Naishadham D, Pfeiffer RM et al. (2009) Variants in interferon-alpha pathway genes and response to pegylated interferon-Alpha2a plus ribavirin for treatment of chronic hepatitis C virus infection in the hepatitis C antiviral long-term treatment against cirrhosis trial. Hepatology 49: 1847-1858. doi:10.1002/hep.22877. PubMed: 19434718.

5. Shepard CW, Finelli L, Alter MJ (2005) Global epidemiology of hepatitis C virus infection. Lancet Infect Dis 5: 558-567. doi:10.1016/S1473-3099(05)70216-4. PubMed: 16122679.

6. Prieto M, Olaso V, Verdú C, Córdoba J, Gisbert C et al. (1995) Does the healthy hepatitis C virus carrier state really exist? An analysis using polymerase chain reaction. Hepatology 22: 413-417. doi:10.1016/0270-9139(95)95374-4. PubMed: 7635408.

7. Singh R, Kaul R, Kaul A, Khan K (2007) A comparative review of HLA associations with hepatitis B and C viral infections across global populations. World J Gastroenterol 13: 1770-1787. PubMed: 17465466.

8. Kuzushita N, Hayashi N, Moribe T, Katayama K, Kanto T et al. (1998) Influence of HLA haplotypes on the clinical courses of individuals infected with hepatitis C virus. Hepatology 27: 240-244. doi:10.1002/hep.510270136. PubMed: 9425943.

9. Yoshizawa K, Ota M, Saito S, Maruyama A, Yamaura T et al. (2003) Long-term follow-up of hepatitis C virus infection: HLA class II loci influences the natural history of the disease. Tissue Antigens 61: 159-165. doi:10.1034/j.1399-0039.2003.00015.x. PubMed: 12694584.

10. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X et al. (2004) HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. Science 305: 872-874. doi:10.1126/science.1097670. PubMed: 15297676.

11. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D et al. (2009) Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 461: 798-801. doi:10.1038/nature08463. PubMed: 19759533.

12. Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, et al. (2010) Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology 138: 1338-1345.e7

13. Duggal P, Thio CL, Wojcik GL, Goedert JJ, Mangia A et al. (2013) Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann Intern Med 158: 235-245. doi:10.7326/0003-4819-158-4-201302190-00003. PubMed: 23420232.

14. Prokunina-Olsson L, Muchmore B, Tang W, Pfeiffer RM, Park H et al. (2013) A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat Genet 45: 164-171. doi:10.1038/ng.2521. PubMed: 23291588.

15. Tanaka J, Kumagai J, Katayama K, Komiya Y, Mizui M et al. (2004) Sex- and age-specific carriers of hepatitis B and C viruses in Japan estimated by the prevalence in the Mart,485,648 first-time blood donors during 1995-2000. Intervirology 47: 32-40. doi:10.1159/000076640. PubMed: 15044834.

16. Bakr I, Rekacewicz C, El Hosseiny M, Ismail S, El Daly M et al. (2006) Higher clearance of hepatitis C virus infection in females compared with males. Gut 55: 1183-1187. PubMed: 16434426.

17. Kasahara A, Hayashi N, Mochizuki K, Takayanagi M, Yoshioka K et al. (1998) Risk factors for hepatocellular carcinoma and its incidence after interferon treatment in patients with chronic hepatitis C. Osaka Liver Disease Study Group. Hepatology 27: 1394-1402. doi:10.1002/hep.510270529. PubMed: 9581697.

18. Haydon GH, Jarvis LM, Simmonds P, Harrison DJ, Garden OJ et al. (1997) Association between chronic hepatitis C infection and hepatocellular carcinoma in a Scottish population. Gut 40: 128-132. PubMed: 9155590.

19. Kawaoka T, Hayes CN, Ohishi W, Ochi H, Maekawa T et al. (2011) Predictive value of the IL28B polymorphism on the effect of interferon therapy in chronic hepatitis C patients with genotypes 2a and 2b. J Hepatol 54: 408-414. doi:10.1016/j.jhep.2010.07.032. PubMed: 21112660.

20. Hiraga N, Imamura M, Hatakeyama T, Kitamura S, Mitsui F et al. (2009) Absence of viral interference and different susceptibility to interferon between hepatitis B virus and hepatitis C virus in human

hepatocyte chimeric mice. J Hepatol 51: 1046-1054. doi:10.1016/j.jhep.2009.09.002. PubMed: 19853955.

21. Siebold C, Hansen BE, Wyer JR, Harlos K, Esnouf RE et al. (2004) Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proc Natl Acad Sci U S A 101: 1999-2004. doi:10.1073/pnas.0308458100. PubMed: 14769912.

22. Ochi H, Maekawa T, Abe H, Hayashida Y, Nakano R et al. (2011) IL-28B predicts response to chronic hepatitis C therapy--fine-mapping and replication study in Asian populations. J Gen Virol 92: 1071-1081. doi:10.1099/vir.0.029124-0. PubMed: 21228123.

23. Chayama K, Hayes CN, Imamura M (2012) Impact of interleukin-28B genotype on in vitro and in vivo systems of hepatitis C virus replication. Hepatol Res 42: 841-853. doi:10.1111/j.1872-034X.2012.01002.x. PubMed: 22524382.

24. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M et al. (2011) A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Hum Mol Genet 20: 3884-3892. doi:10.1093/hmg/ddr301. PubMed: 21750111.

25. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A et al. (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet 41: 591-595. doi:10.1038/ng.348. PubMed: 19349983.

26. Jiang DK, Sun J, Cao G, Liu Y, Lin D et al. (2013) Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. Nat Genet 45: 72-75. doi:10.1038/ng.2483. PubMed: 23242368.

27. Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R et al. (2011) Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat Genet 43: 455-458. doi:10.1038/ng.809. PubMed: 21499248.

28. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y et al. (2008) Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol 69: 443-464. doi:10.1016/j.humimm.2008.05.001. PubMed: 18638659.

29. Cangussu LO, Teixeira R, Campos EF, Rampim GF, Mingoti SA et al. (2011) HLA class II alleles and chronic hepatitis C virus infection. Scand J Immunol 74: 282-287. doi:10.1111/j.1365-3083.2011.02568.x. PubMed: 21535077.

30. Nakamura Y (2007) The BioBank Japan Project. Clin Adv Hematol Oncol 5: 696-697. PubMed: 17982410.

31. Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N et al. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83: 445-456. doi:10.1016/j.ajhg.2008.08.019. PubMed: 18817904.

32. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H et al. (2001) A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 46: 471-477. doi:10.1007/s100380170047. PubMed: 11501945.

33. van der Zwan A, Griffith B, Rozemuller E, Williams T, Tilanus MGJ, Tilanus MGJ, ed. (2002) IHWG Technical Manual Genomic Analysis of the Human MHC: DNA-Based Typing for HLA Alleles and Linked Polymorphisms. Seattle: International Histocompatibility Working Group.

34. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265. doi:10.1093/bioinformatics/bth457. PubMed: 15297300.

35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575. doi:10.1086/519795. PubMed: 17701901.

36. Menashe I, Rosenberg PS, Chen BE (2008) PGA: power calculator for case-control genetic association analyses. BMC Genet 9: 36. doi:10.1186/1471-2350-9-36. PubMed: 18477402.