

200831018A

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

テーラーメイド治療を目指した肝炎ウイルス
データベース構築に関する研究

(H19-肝炎一般-013)

平成 20 年度 総括・分担研究報告書

研究代表者 田中 靖人

平成 21 (2009) 年 3 月

目 次

I. 総括研究報告書	頁
テラーメイド治療を目指した肝炎ウイルスデータベース構築に関する研究 …………… 1 (名古屋市立大学 田中 靖人)	1
II. 分担研究報告書	
1. ゲノムワイド関連分析による肝炎の宿主遺伝要因の探索……………11 (東京大学 徳永 勝士)	11
2. 肝炎ウイルス統合データベースの構築……………15 (国立遺伝学研究所 五條掘 孝)	15
3. 肝炎ウイルスデータベースの更新及びその啓蒙……………33 (国立国際医療センター国府台病院 肝炎・免疫研究センター 溝上 雅史)	33
4. 肝炎の進展、治療反応性に寄与する宿主遺伝子発現データベースの構築……………36 (金沢大学 本多 政夫)	36
5. C型慢性肝炎に対するPEG-IFN・RBV併用療法における治療効果規定因子の データマイニング解析……………38 (武蔵野赤十字病院 黒崎 雅之)	38
6. 口腔扁平上皮癌患者におけるHCVと重複癌の検討……………41 (久留米大学 長尾 由実子)	41
7. B型肝炎ウイルス全ゲノムにおける遺伝子変異と肝発癌機序の解明……………45 (名古屋市立大学 田中 靖人)	45
III. 研究成果の刊行一覧……………51	51
IV. 研究成果の刊行物・別冊……………59	59

I. 総括研究報告書

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（平成20年度）

テラーメイド治療を目指した肝炎ウイルスデータベース構築に関する研究

研究代表者：田中靖人 名古屋市立大学大学院 医学研究科 臨床分子情報医学分野
准教授

研究要旨：

本研究は、肝炎ウイルス感染に対する応答性や薬剤応答性の個人差に関わるヒト及びウイルス両方の遺伝子要因を同一個体内で明らかにすることが大きな特徴である。さらにこれらのデータを網羅的に収集する組織作りを行い、統合的にデータベース化し解析することにより、ヒト側要因とウイルス側要因の両方を考慮した知見を得ることを目的とする。

1) 検体及び付帯情報の収集：ヒトゲノム・遺伝子解析研究に関する倫理指針に従い、全国15施設より検体及び付帯情報の収集を開始した。平成21年3月5日現在、慢性ウイルス性肝疾患患者から合計685検体を得た。**2) ヒトSNPs解析**：健常対照者200検体及び肝疾患群406検体について90万種のSNPタイピングを終了し、各SNPのアリル頻度、遺伝子型頻度等について健常者群、患者群、肝炎亜型患者群の間でゲノムワイド関連分析を行った結果、ゲノムワイド有意水準 (p 値 $\leq 5 \times 10^{-7}$) に達する3か所の遺伝子領域 (Gene X、Gene Y、Gene Z) を検出した。**3) 肝炎ウイルス統合データベースの構築**：患者SNPs情報を、その臨床的背景情報とともに管理、解析するデータベースを構築し、収集したデータの登録を行った。併せて、必要とされるデータ解析手法について検討し、同データベースを利用してそれを遂行するインターフェースのプロトタイプを構築した。**4) 肝炎テラーメイド治療を目指した各種要因の解明**：a) ペグインターフェロン・リバビリン (PEG-IFN/RBV) 併用療法の治療抵抗性因子の解明：治療前の肝組織と末梢血単核球 (PBMC) の遺伝子発現プロファイリングの解析及びデータマイニングを用いた網羅的解析により治療抵抗性要因を明らかにした。b) HBV関連肝癌進展に寄与する因子：HBVの全塩基配列を決定し、肝癌に寄与するウイルス変異を複数同定した。HCVの病態進展及び薬剤応答性に関連する領域の塩基配列も決定する。c) HBV・HCV感染に伴う肝外病変：HCVの肝外病変の一つである口腔扁平上皮癌におけるHCV感染と重複癌について検討した結果、重複癌発生に関わる因子は、Stage IV、HCV抗体陽性、70歳以上の年齢層であった。最終的には上記すべてのデータを統合し、肝炎テラーメイド医療（治療）を目指した統合型肝炎ウイルスデータベースを構築する。

研究分担者

五條堀 孝	国立遺伝学研究所 生命情報・DDBJ 研究センター 遺伝情報分析研究室 教授
徳永 勝士	東京大学大学院医学系研究科 人類遺伝学 教授
溝上 雅史	国立国際医療センター国府台病院 肝炎・免疫研究センター センター長
本多 政夫	金沢大学医学部 先端医療技術学 教授
黒崎 雅之	武蔵野赤十字病院 消化器科 部長
長尾 由実子	久留米大学医学部 消化器疾患情報講座 准教授

A. 研究目的

本邦における慢性肝疾患の原因ウイルスである B 型肝炎ウイルス (HBV) キャリアは約 150 万人、C 型肝炎ウイルス (HCV) キャリアは約 200 万人と推定され、その一部は慢性肝炎から肝硬変・肝細胞癌へ移行するがその機序は不明のままである。また、HCV にはペグインターフェロン・リバビリン併用療法が導入され、一定の効果は認められているが完治率は 50% 程度でその副作用の発現率は高い。今回の研究の目的は、(1) 肝炎ウイルス感染に対する応答性 (発症感受性及び病態進展) や薬剤応答性の個人差に関わるヒト及びウイルス両方の遺伝子要因を同一個体内で解析する。(2) 各種データを統括する組織作り、統合的にデータベース化・解析からヒト及びウイルス側要因の両方を考慮した知見を得る。(3) 得られたデータをもとに、病態進展の予測 (ハイリスク群の囲い込み) 及び適切な薬剤の選択を行い、テーラーメイド治療を目指す。(4) 統合型データベース (DB) をすべて公開し、一般臨床でも活用しやすいようなプラットフォームを作成する。

B. 研究方法

今年度より研究代表者となった田中靖人は、引き続き本研究の主研究施設である名古屋市立大学大学院医学研究科ヒト遺伝子解析倫理委員会の規定に基づいて、検体の採取を継続した。これまでに 16 施設においてヒト遺伝子解析倫理委員会の規定に基づいて検体採取が開始され、平成 21 年 3 月 5 日現在、685 検体が東大の SNP (一塩基多型) センターに届けられている。上記目的を達成するための研究概要を図 (P8) に示す。

1) ヒト SNPs を用いたゲノムワイド関連研究: 90 万種以上の SNP 解析用プローブが搭載された Affymetrix Genome-Wide Human SNP

Array 6.0 (以下、SNP Array 6.0) を用いるゲノムワイド関連分析は世界的にも始まったばかりであり、この戦略を用いた肝炎ウイルス感染症に関する研究は独創的である。慢性ウイルス性肝疾患患者 406 検体および健常対照群 200 検体の SNP タイピングし、患者群を病態および薬剤応答性に応じてサブグループに分類してゲノムワイド関連解析 (GWAS) を行うことにより、将来のテーラーメイド医療の基盤を構築する (徳永)。

2) 肝炎ウイルス統合データベースの構築: これまでに網羅的に収集されたヒト及びウイルスの情報を統合したデータベースを構築するが、これは未だ世界に存在せず、必要性が以前から唱えられていたものである。さらに、日本全国の肝臓病専門医から収集された付帯情報を活用して、ウイルス性肝炎の特徴づけ (プロファイリング) を詳細に行い、肝炎テーラーメイド治療の確立を目指す (五條堀、溝上、新井)。今年度は、徳永研究分担者により収集された SNP タイピングデータおよびゲノムワイド関連解析結果を、昨年度までに設計したデータベーススキーマに導入した。それを基に、必要とされる解析手法および利用者インターフェースを検討している。

3) 肝炎テーラーメイド治療を目指した各種要因の解明:

a) ペグインターフェロン・リバビリン (PEG-IFN/RBV) 併用療法の治療抵抗性因子の解明: IFN 療法の反応性の違いにはウイルス側因子に加え、宿主側因子の果たす役割が極めて重要と考えられる。GWAS に基づく SNP 解析により、IFN 療法に対する反応性の違いが明らかにされることが期待される一方で、本多研究分担者はゲノム上で認められる変化が実際に遺伝

子発現の変化として再現されるかを網羅的遺伝子発現により解析することが必要と考え、PEG-IFN/RBV 併用療法を施行したC型慢性肝炎57症例につき、治療前の肝組織と末梢血単核球(PBMC)の遺伝子発現プロファイリングをAffymetrix GeneChip (HG-U133 Plus2.0 Array)を用いて解析した。黒崎研究分担者は、PEG-IFN/RBV 併用療法を行った1b型高ウイルス量のC型慢性肝炎症例256例を対象とし、臨床背景、肝生検所見、一般検査、ウイルス学的検査、薬剤投与量を説明因子として、ウイルス学的著効と関連する因子について、SPSS Clementine 12.0のアルゴリズムC5.0を用いてデータマイニング解析を行った。

b) HBV・HCV 関連肝癌進展に寄与する因子：田中は溝上研究分担者と共同で、各種患者情報をmatchさせたcase-control studyとして、肝癌(HCC)患者76例及び非肝癌患者(non-HCC)77例から得られたHBVの全塩基配列を決定することで、HBV発癌に寄与するHBV遺伝子変異の同定を行った。今後は、ゲノムワイド解析により得られた病態進展に寄与する宿主因子とウイルス因子を統合し、肝癌ハイリスク群の囲い込みを目指す(田中、溝上)。HCVの病態進展及び薬剤応答性・耐性に関連する領域の塩基配列も決定する。

c) HBV・HCV 感染に伴う肝外病変：HCVの肝外病変の一つである口腔扁平上皮癌におけるHCV感染と重複癌について検討した。前癌病変である扁平苔癬を含めて、ゲノムワイドにもその要因を検討する(長尾)。

C. 研究結果

今年度より研究代表者となった田中靖人は、本研究の主研究施設である名古屋市立大学大

学院医学研究科ヒト遺伝子解析倫理委員会の規定に基づいて、検体の採取を継続した。これまでに16施設においてヒト遺伝子解析倫理委員会の規定に基づいて検体採取が開始され、平成21年3月5日現在、685検体が東大のSNPセンターに届けられた。

1) ヒト SNPs を用いたゲノムワイド関連研究：SNP Array 6.0による慢性ウイルス性肝疾患患者406検体および健常対照群200検体のSNPタイピングがすべて完了した。タイピング工程において品質水準に達した患者群396検体を病態および薬剤応答性に応じて8つのサブグループに分類し、健常対照群198検体と共にゲノムワイド関連分析を実施したところ、ゲノムワイド有意水準(p 値 $\leq 5 \times 10^{-7}$)に達する3か所の遺伝子領域(Gene X、Gene Y、Gene Z)を検出した。特に、サブグループ1とサブグループ2の関連解析からは、 p 値 $< 1.7 \times 10^{-12}$ (Odds ratio > 10)という非常に強い関連を示す遺伝子領域(Gene Z)が検出された。

2) 肝炎ウイルス統合データベースの構築：有効なデータの得られたHCV患者236件、HBV患者100件、対照群184件のSNPタイピング結果、および、6組のゲノムワイド関連解析結果についてデータベースへ登録した。患者の臨床情報は匿名化の上、同様に登録した。これらを検索、参照する利用者インターフェースを構築し、参考情報としてGWAS(東京大学)、dbSNP(NCBI)の対応する情報が参照できるよう、リンクを準備した。また、1997年以降、肝炎ウイルスデータベースは進化し続けており、現在では5種類の肝炎ウイルス、HAV 3,798件、HBV 21,068件、HCV 72,822件、HDV 1,008件、HEV 2,629件について、塩基配列を網羅的に収集、系統解析により整理しデータベース化して一般に公開している。

3) 肝炎テーラーメイド治療を目指した各種要因の解明:

a) PEG-IFN/RBV 併用療法の治療抵抗性因子の解明: 本多研究分担者は、併用療法を施行した症例の治療前の肝生検では興味深いことに Interferon stimulated gene (ISG) を高発現している群と低発現している 2 群に群別されることが明らかとなった。すなわち HCV 感染に伴う生体の反応が宿主によって異なっている可能性が示唆された。ISG を高発現している群では IFN 療法の反応が不良であり、ISG 低発現群では IFN 療法の反応が良好である傾向が認められた。黒崎研究分担者は、データマイニングを用いた網羅的解析により治療抵抗性要因を探索した。その結果、NS5A 遺伝子変異、年齢、AFP、Core アミノ酸置換の組み合わせにより、SVR の確率が 10% しか期待できない難治症例、および 72% 期待できる治療感受性症例を同定することが可能であった。一般検査に加えて HCV 遺伝子検査を説明変数に投入して解析することにより、予想精度は向上した。本解析により現時点で入手可能な臨床情報、ウイルス学的情報に基づいた治療効果予測アルゴリズムが構築できたが、更なる予測精度の向上のためには、治療抵抗性にかかわる HCV 遺伝子構造の詳細な検討が必要であり、それ以上に宿主因子の解析が必要である。

b) HBV・HCV 関連肝癌進展に寄与する因子: 本邦に広く分布し予後不良とされる HBV genotype C (HBV/C) 症例において、年齢、性別、HBe 抗原陽性率をマッチさせた患者 (HCC グループ) 一対照 (non-HCC グループ) 研究を行った。全塩基配列を決定した結果、PreS2 領域の欠失 (26% vs. 13%; $p=0.043$)、S 領域

の C373T 変異 (18% vs. 6%)、Polymerase 領域の C934A 変異 (70% vs. 6%; $p=0.031$)、X 領域の T1497V 変異 (28% vs. 8%; $p=0.0014$)、X 領域かつ box α の C1653T 変異 (54% vs. 32%; $p=0.0039$)、X 領域かつ BCP の T1753V、A1762T/G1764A 変異 (50% vs. 29%, 92% vs. 75%; $p=0.0081$, $p=0.0078$) が、HCC に寄与するウイルス変異であった。来年度以降これらの患者の SNPs を測定し、両者を合わせて検討することで、生体因子と HBV 因子の両方の関係が明らかになるとと思われる。

c) HBV・HCV 感染に伴う肝外病変: 口腔扁平上皮癌患者 60 例において、多重複癌の発生率は 35% (21/60)、HCV 抗体陽性率は 26.7% (16/60) であった。多変量解析により重複癌発生に関わる因子は、Stage IV、HCV 抗体陽性、70 歳以上の年齢層であった。また、前癌病変である扁平苔癬を含めて、SNP タイピングを実施した。

D. 考察

ウイルス性肝疾患は肝炎ウイルス感染症であるのでホストのヒト側要因と病原体であるウイルス側要因の両方の相互関係で病態は決定される。したがって、これら両方の要因を同時に検討しないことにはウイルス性肝疾患の本体を明らかにできない。しかしながら現在まで主にウイルス側要因に関する研究しか行なわれてこなかった。その理由として、ヒト側要因は多くの要因が複雑に絡み合っていることが予想されるがそれらを網羅的に解析する手法が存在しなかったからである。

近年、そのヒト側要因を一度に網羅的に約 90 万個の SNPs で測定するゲノムワイド関連解析 (GWAS) が可能となった。徳永研究分担者は、今回の検討で慢性ウイルス性肝疾患患者

396 検体と健常対照群 200 検体でゲノムワイド関連分析を実施した結果、ゲノムワイド有意水準 (p 値 $\leq 5 \times 10^{-7}$) に達する 3 か所の遺伝子領域を検出した。これらは非常に有望な SNPs と考えられるため、今後、新たなサンプルセットを準備し、検出された領域を中心に再現性の確認および絞り込みのため、DigiTag2法による高密度 SNP 関連解析を行う予定である。

また、我々が得意とするウイルス遺伝子解析技術を用いて、同一個体内での HBV 及び HCV 遺伝子配列を決定し、病態進展や薬剤耐性に寄与する変異を複数同定した。こうしたウイルス遺伝子情報は 1997 年以来、運用している肝炎ウイルスデータベースに登録し、すでに一部公開している。さらに、国立遺伝学研究所 DDBJ 研究センター長である五條堀研究分担者は、ヒト SNPs とウイルスゲノム解析に基づいた肝炎ウイルス統合データベースにおいて必要とされるデータ解析手法について検討し、同データベースを利用してそれを遂行するインターフェースのプロトタイプを構築した。本データベースにより該当情報および解析結果を管理、参照できることが示されたため、これをプラットフォームとして利用者が独自の解析を遂行することが可能と考えられる。

本多研究分担者らは、肝組織学的進行度やヒ肝組織と末梢血単核球 (PBMC) の遺伝子発現プロファイリングを検討し、IFN 療法の治療効果との関連性を明らかにした。これまでの問題点として、SNP データと遺伝子発現データを統合したデータベースがなかったことで、各々が独立して研究がなされていた。SNP 変異のある遺伝子の発現レベルを知りたい研究者、逆に発現変化のある遺伝子の SNP を知り

たい研究者も存在する実態を考えた場合、SNP データと遺伝子発現データを統合したデータベースもオプションとして組み込むことで、効率的な宿主因子の解明が期待される。

こうして当研究班の SNPs 解析により有意な宿主因子が同定されれば、治療効果予測の精度はさらに向上する。データマイニングによる decision tree model は、SNPs を含めた多彩な情報を臨床の場にわかりやすく提示する上で極めて有用な方法であり、患者個人の臨床情報に対応したオーダーメイド医療の実現において重要な役割を果たす可能性がある。

E. 結論

今年度までに明らかとされたウイルス遺伝子情報、SNP 情報、臨床情報はすでに統合され、肝炎ウイルス統合データベースのプロトタイプは完成した。将来的には、遺伝子発現データとの統合、さらにはデータマイニング解析によりウイルス、宿主、臨床情報 (病態、治療効果) の各因子を統合して解析することが可能である。臨床の分野においては、本データベースを参照することにより、患者 SNPs とウイルス変異の組み合わせから病態進展の予測及びハイリスク群の抽出が可能となり、テーラーメイド治療への展開が期待される。従って、適切な治療法の選択および新たな治療法の開発で患者の予後を改善するのみならず、肝硬変・肝癌という高度な医療が必要な患者数を減らすことにより、医療費の低減に繋がり、社会の福祉に寄与することができる。

F. 研究発表

1. 論文発表

1. Sugiyama M, **Tanaka Y**, Kurbanov F, Maruyama I, Shimada T, Takahashi S,

- Shirai T, Hino K, Sakaida I, Mizokami M. Direct Cytopathic Effects of Particular Hepatitis B Virus Genotypes in uPA/SCID Mouse with Human Hepatocytes. *Gastroenterology*. 136(2):652-662. 2009.
2. Yuen MF, **Tanaka Y**, Fong DYT, Fung J, Wong DKH, Yuen JCH, But DYK, Chan AOO, Wong BCY, Mizokami M, Lai CL. Independent Risk Factors and Predictive Score for the Development of Hepatocellular Carcinoma in Chronic Hepatitis B. *J Hepatol*. 50(1):80-88. 2009.
 3. Kurbanov F, **Tanaka Y**, Kramvis A, Simmonds P, Mizokami M. When should "I" consider a new hepatitis B virus genotype? *J Virol*. 82(16):8241-8242. 2008.
 4. Togashi H, Hashimoto C, Yokozawa J, Suzuki A, Sugahara K, Saito T, Yamaguchi I, Badawi H, Kainuma N, Aoyama M, Ohya H, Akatsuka T, **Tanaka Y**, Mizokami M, Kawata S. What can be revealed by extending the sensitivity of HBsAg detection to below the present limit? *J Hepatol*. 49(1):17-24. 2008.
 5. **Tanaka Y**, Sanchez LV, Sugiyama M, Sakamoto T, Kurbanov F, Tatematsu K, Roman S, Takahashi S, Shirai T, Panduro A, Mizokami M. Characteristics of hepatitis B virus genotype G coinfecting with genotype H in Chimeric Mice Carrying Human Hepatocytes. *Virology*. 376(2):408-415. 2008.
 6. Elkady A, **Tanaka Y**, Kurbanov F, Oynsuren T, Mizokami M. Virological and clinical implication of core promoter C1752/V1753 and T1764/G1766 mutations in hepatitis B virus genotype D infection in Mongolia. *J Gasrotol Hepatol*. 23(3):474-481. 2008.
 7. Yuen MF, **Tanaka Y**, Shinkai N, Poon RT, But DYK, Daniel Fong YT, Fung J, Wong DKH, Yuen JCH, Mizokami M, Lai CL. Risk for hepatocellular carcinoma with respect to hepatitis B virus genotypes B/ C, specific mutations of enhancer II/ core promoter/ precore regions and HBV DNA levels. *Gut*. 57(1):98-102. 2008.
2. 学会発表
 1. 2008 肝臓学会総会
B 型肝炎ウイルス T1497V, C1653T, T1753V 変異と肝臓との関連性～153 例の B 型肝炎ウイルス全塩基配列の解析～新海登, **田中靖人**, 伊藤清顕, 向出雅一, 長谷川泉, 大野智義, 朝比奈靖浩, 泉並木, 八橋弘, 城卓志, 溝上雅史.
 2. 18th APASL 2008 Conference (APASL 2008 Korea). Complete genomic analyses of 153 hepatitis B virus genotype C2 strains in Japan reveal association of T1497V, C1653T and T1753V mutations with hepatocellular carcinoma. Noboru Shinkai, **Yasuhito Tanaka**, Kiyooki Ito, Motokazu Mukaide, Izumi Hasegawa, Yasuhiro Asahina, Namiki Izumi, Hiroshi Yatsuhashi, Etsuro Orito, Takashi Joh, and Masashi Mizokami.
 - G. 知的所得権の所得状況
 1. 特許取得
なし

2. 実用新案登録
なし
3. その他
なし

研究の概要

全国15施設からの検体提供(各施設で匿名化)

* 平成21年3月現在

個人情報管理者

ヒトSNPs・ゲノムワイド解析

- ・ 健康人 200例 (徳永班員)
- ・ 慢性肝疾患 685例

ヒトSNPs情報

GWASデータベース
(東大人類学)

ヒトSNPデータベース
(jSNP, dbSNP, ...)

患者付帯情報の収集・解析

(田中、研究協力者)

肝炎ウイルス統合データベースの構築

- ・ ウイルス配列DBの更新
- ・ 患者SNP DB及び患者情報DBの設計

(五條堀、溝上、新井)

肝炎ウイルスデータベース

公共DNAデータベース
(DDBJ/EMBL/GenBank)

肝炎ウイルス塩基配列決定

- ・ 肝癌症例 150例 (田中、溝上)
- ・ 慢性肝炎HBV 140例
- ・ 慢性肝炎HCV 351例

肝炎ウイルスゲノム情報

肝炎テラーメイド治療を目指した各種要因の解明(黒崎、本多、長尾)

テラーメイド医療で利用

肝炎・免疫情報センター
(国立国際医療センター)

ヒトゲノム倫理委員会承認状況(21年度～19施設)

(敬称略)

施設名	代表者	ヒトゲノム倫理委員会	ご担当者
東京大学	徳永勝士	○	西田奈央、上原靖加
北海道大学	髭修平	○	髭修平
岩手医科大学	鈴木一幸	○	阿部弘一
埼玉医科大学	持田智	○	持田智
山梨大学	榎本信幸	○	前川伸哉
武蔵野赤十字病院	黒崎雅之	○	黒崎雅之
金沢大学	本多政夫	○	酒井明人
信州大学	田中榮司	○	松本晶博
京都府立医科大	伊藤義人	○	伊藤義人
国立大阪医療センター	三田英治	○	三田英治
鳥取大学	村脇義和	○	大山賢治
山口大学	坂井田功	○	是永匡紹
愛媛大学	恩地森一	○	日浅陽一
久留米大学	長尾由実子	○	長尾由実子
国立国際医療センター 国府台病院	溝上雅史	申請中	正木尚彦
東京医科歯科大学	坂本直哉	○	中川美奈
兵庫医科大学	西口修平	申請中	榎本平之
川崎医科大学	日野啓輔	申請中	日野啓輔
名古屋市立大学	田中靖人	○ (2007.3.30)	田中靖人

II. 分担研究、研究協力報告書

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）
分担研究報告書（平成20年度）

テーラーメイド治療を目指した肝炎ウイルスデータベース構築に関する研究

研究分担者：徳永勝士 東京大学大学院医学系研究科 人類遺伝学 教授
研究協力者：西田奈央 東京大学大学院医学系研究科 人類遺伝学 特任助教

分担研究課題：ゲノムワイド関連分析による肝炎の宿主遺伝要因の探索

研究要旨：肝炎ウイルスに感染した宿主側の病態促進因子及び薬剤応答性を規定する遺伝要因の解明を目的として、90万種類以上のSNPを対象としたゲノムワイド関連分析を行った。現在までに、慢性ウイルス性肝疾患患者396検体および健常対照群200検体のSNPタイピングがすべて完了し、Overall call rateは平均でそれぞれ99.46%、99.71%となった。患者群396検体を病態および薬剤応答性に応じて8つのサブグループに分類し、健常対照群200検体とともにゲノムワイド関連分析を実施したところ、ゲノムワイド有意水準（ p 値 $\leq 5 \times 10^{-7}$ ）に達する3か所の遺伝子領域を検出した。このうち、特にサブグループ1とサブグループ2の関連解析からは、 p 値 $< 10^{-12}$ (Odds ratio > 10) という非常に強い関連を示す遺伝子領域（Gene Z）が検出された。今後、新たなサンプルセットを準備し、検出された遺伝子領域において再現性確認および絞り込みのため、DigiTag2法による高密度SNP関連解析を行う予定である。

A. 研究目的

B型肝炎ウイルスおよびC型肝炎ウイルスに感染した宿主を対象としたゲノムワイド関連分析を行うことにより、宿主側の肝病態進展に寄与する遺伝因子、治療効果に寄与する遺伝因子、ウイルス感染感受性に寄与する遺伝因子を探索することを目的とする。

B. 研究方法

90万種類以上のSNP解析用プローブが搭載されたAffymetrix Genome-Wide Human SNP Array 6.0（以下、SNP Array 6.0）を用いて、慢性ウイルス性肝疾患患者400検体および健常対照群200検体のSNPタイピングを行い、患者群を病態および薬剤応答性に応じてサブグループに分類してゲノムワイド関連解析を行う。

SNP Array 6.0は、制限酵素によるゲノムDNAの断片化とマイクロアレイによるタイピングの

手法に改良を加えることにより、大規模なタイピングを行える手法として確立された。解析対象となるSNPは、公共のSNPデータベースおよびPerlegen社に登録されている約220万種のSNPから遺伝学的情報量が最大化されるように、また連鎖不平衡やHapMapプロジェクトからの情報も考慮して選択された約44万種のSNPに、Tag SNP、X染色体およびY染色体に存在するSNPなどを加えた90万種類以上のSNPである。

SNP Array 6.0によるSNPタイピングは、ゲノムの複雑さを低減しマイクロアレイへのハイブリダイゼーション効率を上げるための酵素反応ステップと、洗浄・染色装置（Fluidics Station 450）およびマイクロアレイ用スキャナー（GeneChip Scanner 3000 7G）を用いた検出ステップで構成される（図1）。SNPタイピングは1検体につき合計500ngのゲノムDNAを使用し、2種類の制限酵素（Sty I、Nsp I）を用いて実施さ

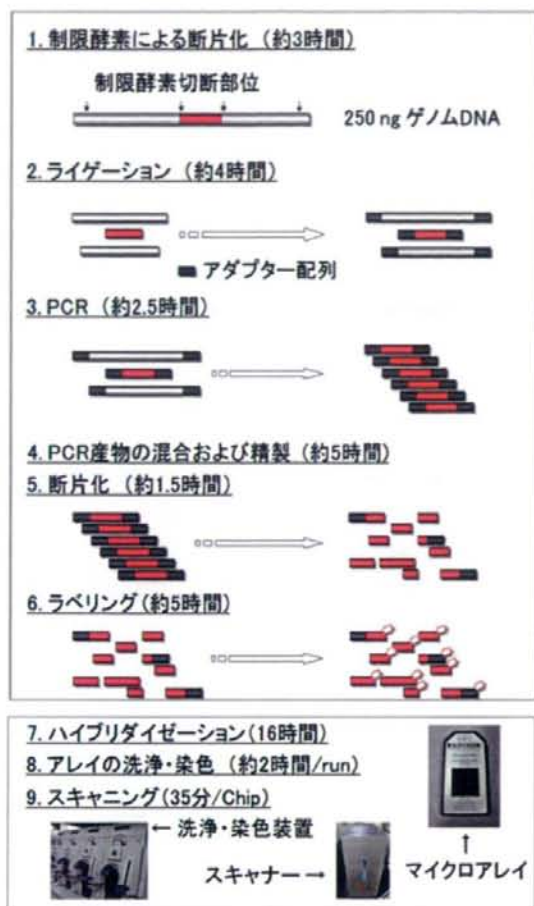


図 1 SNP Array 6.0 による SNP タイピング

れる。制限酵素によるゲノム DNA の断片化を行った後、断片化されたゲノム DNA の両末端にアダプター配列をライゲーション反応により付加する。アダプター配列は、続く PCR で使用されるプライマーと相同な配列を持ち、また制限酵素認識配列を突出端として持つ 2 本鎖 DNA である。2 種類の制限酵素 (Sty I、Nsp I) のそれぞれに対して用意されるアダプター配列は、制限酵素認識配列を除いて共通の配列を持っているので共通のプライマーを使用して PCR を行うことができる。PCR では、目的の長さを持った DNA 断片 (250-1100 bp) だけが選択的に増幅される。続いて、Sty I および Nsp I それぞれの PCR 産物を混合した後、混合産物を精製し、DNase I 制限酵

素による断片化を行う。ここで断片化された PCR 産物は平均長で 180 bp 以下となる。最後に terminal deoxynucleotidyl transferase 酵素反応により、断片化された PCR 産物の末端にビオチンを導入する。

続いて、専用のマイクロアレイ (GeneChip アレイ) を用いてハイブリダイゼーションを行う。マイクロアレイへのハイブリダイゼーションが終了した後、洗浄・染色装置を用いてマイクロアレイの洗浄および蛍光染色を行う。蛍光染色は、蛍光分子で標識されたストレプトアビジンが、上述のビオチン導入された PCR 断片に結合することにより行われる。また、洗浄・染色装置内ではビオチン修飾された抗ストレプトアビジン抗体を用いてシグナルの増強が行われる。最後に蛍光染色されたマイクロアレイを専用のスキャナーで画像データとして読み取り、続いて専用のソフトウェア (Affymetrix Genotyping Console 2.0) を用いて各 SNP の遺伝子型を決定する。

複数の施設で行われた SNP Array 6.0 による SNP 解析の結果から、Overall call rate (全 909,622 種類の SNP のうちで遺伝子型が決定された SNP の割合) は平均 99% 以上となり、また、HapMap データベースに登録された SNP との遺伝子型一致率は 99.7% を超えることが Affymetrix 社から報告されている。また、タイピング結果が悪いことが明らかとなっている 3,022 種類の SNP をクオリティーコントロール (QC) として用いており、QC call rate (全 3,022 種類の SNP のうちで遺伝子型が決定された SNP の割合) が 86% を下回る検体は解析から除外される。

C. 研究結果

SNP Array 6.0 による慢性ウイルス性肝疾患患者 406 検体および健常対照群 200 検体の SNP タイピングがすべて完了した。なお、患者群 406 検体のうちの 10 検体は、タイピング工程において品質水準に達しなかったため、タイピングデータの

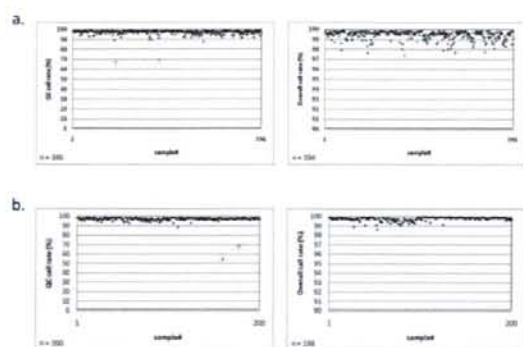


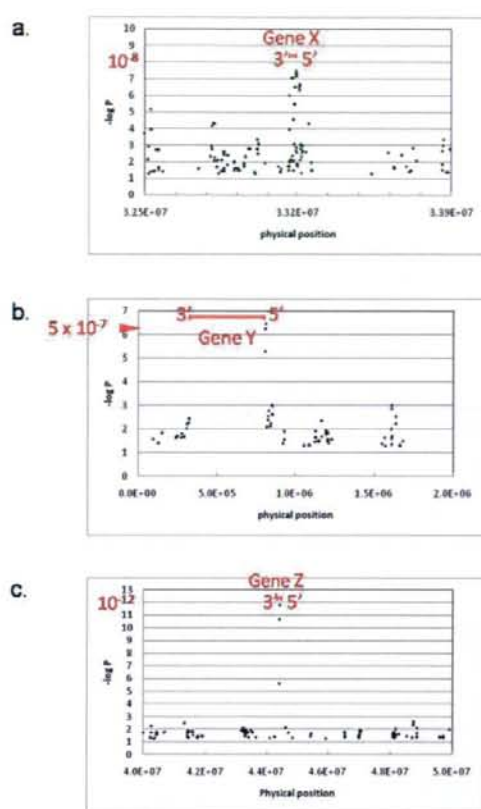
図 2 QC call rate および Overall call rate

a. 慢性ウイルス性肝疾患患者 396 検体のタイピング結果 b. 健常対照群 200 検体のタイピング結果
取得を断念した。QC call rate の平均は、慢性ウイルス性肝疾患患者で 97.29%、健常対照者群で 97.37% となり、また、QC call rate が 86% を下回る検体は、患者群 396 検体のうち 2 検体、健常対照者群 200 検体のうち 2 検体であった。Overall call rate は、QC call rate が 86% を上回った検体で決定され、それぞれ 99.46% (n=394)、99.71% (n=198) となった (図 2)。

ゲノムワイド関連解析を行う上で、解析の大きな障害となる偽陽性関連の発生を抑えることが、真の疾患関連遺伝子領域を効率よく探索することにつながる。偽陽性関連発生の大きな要因として、「集団の構造化」と「タイピングエラー」の二つが挙げられる。そこで我々は、今回解析した患者群 396 検体および健常群 200 検体の両群において、集団の構造化が起きていないことを主成分分析法で確認した。また、QC call rate の閾値を 86% から 95% まで引き上げて不良データの除去を行い、また患者群・健常群をまとめて遺伝子型の決定を行うことで、タイピングエラーを効果的に排除できることが我々の解析から明らかとなった。

患者群 396 検体を病態および薬剤応答性に応じて 8 つのサブグループに分類し、健常対照群

図 3 QC call rate および Overall call rate



a. HBV 感染患者群と健常対照群の関連解析で検出された遺伝子領域 (Gene X) b. HBV 感染患者群と HCV 感染患者群の関連解析で検出された遺伝子領域 (Gene Y) c. サブグループ 1 とサブグループ 2 の関連解析で検出された遺伝子領域 (Gene Z)

198 検体と共にゲノムワイド関連分析を実施したところ、ゲノムワイド有意水準 (p 値 $\leq 5 \times 10^{-7}$) に達する 3 か所の遺伝子領域 (Gene X、Gene Y、Gene Z) を検出した (図 3)。HBV 感染患者群と健常対照群および HBV 感染患者群と HCV 感染患者群をそれぞれ比較した関連解析からは、それぞれ遺伝子領域 Gene X [p 値 $< 3.5 \times 10^{-8}$ (Odds ratio > 3)]、および遺伝子領域 Gene Y [p 値 $< 3.6 \times 10^{-7}$ (Odds ratio < 0.5)] が検出された。特に、サブグループ 1 とサブグループ 2 の関連解析からは、

p 値 $< 1.7 \times 10^{-12}$ (Odds ratio > 10) という非常に強い関連を示す遺伝子領域 (Gene Z) が検出された。

D. 考察

慢性ウイルス性肝疾患患者 396 検体と健常対照群 200 検体でゲノムワイド関連分析を実施した結果、ゲノムワイド有意水準 (p 値 $\leq 5 \times 10^{-7}$) に達する 3 か所の遺伝子領域を検出した。サブグループごとの関連分析において、集団の構造化やタイピングエラーが原因で引き起こされる顕著なインフレーションは見られず、また P 値 $< 10^{-4}$ の SNP について散布図を確認した結果、偽陽性関連は効果的に排除できたと考えられる。今後、新たなサンプルセットを準備し、検出された領域において再現性の確認および絞り込みのため、DigiTag2 法による高密度 SNP 関連解析を行う予定である。

E. 結論

肝炎ウイルスに感染した宿主側の病態促進因子及び薬剤応答性を規定する遺伝要因はまだ明らかとなっていない。今回のゲノムワイド関連分析から、ゲノムワイド有意水準 (p 値 $\leq 5 \times 10^{-7}$) に達する 3 か所の遺伝子領域を含めて、宿主側の遺伝要因の候補 (p 値 $\leq 10^{-4}$) が多数検出されており、宿主側の病態促進因子および薬剤応答性を規定する遺伝要因が特定されることが期待される。

F. 研究発表

1. 論文発表

- 1) 西田奈央、徳永勝士：ゲノムワイド SNP タイピング技術の現状と将来、医学のあゆみ、225 : 721-725 (2008)
- 2) 西田奈央、徳永勝士：ゲノムワイド関連分析による多因子疾患遺伝子の探索、肝疾患 Review 2008~2009: 85-91 (2008)

- 3) 小池麻子、西田奈央、徳永勝士：ゲノムワイド関連解析データベースの開発、蛋白核酸酵素、53(7) : 882-887 (2008)
- 4) Nishida N., et al.: Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals, BMC Genomics, 9: 431 (2008)
- 5) Miyagawa T., et al.: Appropriate data cleaning methods for genome-wide association study, J. Hum. Genet., 53: 886-893 (2008)

2. 学会発表

該当なし

G. 知的所得権の所得状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

該当なし

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）
分担研究報告書（平成20年度）

テーラーメイド治療を目指した肝炎ウイルスデータベース構築に関する研究

研究分担者：五條 堀 孝 国立遺伝学研究所 生命情報・DDBJ 研究センター
遺伝情報分析研究室 教授

分担研究課題：肝炎ウイルス統合データベースの構築

研究要旨：

宿主（ヒト）側と肝炎ウイルス側双方の要因に注目し、広範なサンプルより両データを収集し、患者より得られた臨床情報を加えて多次的にデータベース化し解析する。本年度は患者SNPs情報を、その臨床的背景情報とともに管理、解析するデータベースを構築し、収集したデータの登録を行った。併せて、必要とされるデータ解析手法について検討し、同データベースを利用してそれを遂行するインターフェースのプロトタイプを構築した。

する情報が参照できるよう、リンクを準備した。

A. 研究目的

本研究は、ヒト側・ウイルス側双方の要因を統合的に解析することが特徴であり、それに必要な基盤として、両データを網羅的に収集し相互参照の解析を可能とするデータベースを構築する。

B. 研究方法

分担研究者・徳永教授により収集された SNP タイピングデータおよびゲノムワイド関連解析結果を、昨年度までに設計したデータベーススキーマに導入する。それを基に、必要とされる解析手法および利用者インターフェースを検討する。

C. 研究結果（参考資料1, 2）

有効なデータの得られた HCV 患者 236 件、HBV 患者 100 件、対照群 184 件の SNP タイピング結果、および、6 組のゲノムワイド関連解析結果についてデータベースへ登録した。患者の臨床情報は匿名化の上、同様に登録した。これらを検索、参照する利用者インターフェースを構築し、参考情報として GWAS(東京大学)、dbSNP(NCBI)の対応

D. 考察

本データベースにより該当情報および解析結果を管理、参照できることが示されたため、これをプラットフォームとして利用者が独自の解析を遂行することが可能と考えられる。

E. 結論

昨年度までの成果を基にデータベースを構築し、実際に該当情報を登録した。データ管理のフレームワークは完成したと考えられるため、次年度以降、利用者インターフェースと解析機能の向上に注力できる。

F. 研究発表

1. 論文発表

- 1 Jin, L., Kryukov, K., Suzuki, Y., Imanishi, T., Ikeo, K. and Gojobori, T. (2009). The evolutionary study of small RNA-directed gene silencing pathways by investigating RNase III enzymes *Gene* (in press).
- 2 Fukuchi, S., Homma, K., Sakamoto, S., Sugawara, H., Tateno, Y., Gojobori, T. and Nishikawa, K. (2009). The GTOP database in 2009: updated content and novel features to expand and deepen

- insights into protein structures and functions. *Nucleic Acids Res.* 37(Database issue):D333-7.
- 3 Sugawara, H., Ikeo, K., Fukuchi, S., Gojobori, T. and Tateno, Y. (2009). DDBJ Dealing with Mass Data Produced by the Second Generation Sequencer. *Nucleic Acids Res.* 37(Database issue):D16-8.
 - 4 Shimada, MK., Matsumoto, R., Hayakawa, Y., Sanbonmatsu, R., Gough, C., Yamaguchi-Kabata, Y., Yamasaki, C., Imanishi, T. and Gojobori, T. (2009) VarySysDB: a human genetic polymorphism database based on all H-InvDB transcripts. *Nucleic Acids Res.* 37(Database issue):D810-5.
 - 5 Jin, L., Kryukov, K., Clemente, J., Komiyama, T., Suzuki, Y., Imanishi, T., Ikeo, K. and Gojobori, T. (2008). The evolutionary relationship between gene duplication and alternative splicing. *Gene* 427(1-2): 19-31.
 - 6 Sakai, H., Itoh, T. and Gojobori, T. (2008). Processed Pseudogenes and Their Functional Resurrection in the Human and Mouse Genomes. *Encyclopedia of Life Science (ELS)*. Pp. 1-6. John Wiley & Sons, Ltd.
 - 7 Yamaguchi-Kabata, Y., Shimada, MK., Hayakawa, Y., Minoshima, S., Chakraborty, R., Gojobori, T. and Imanishi, T. (2008). Distribution and effects of nonsense polymorphisms in human genes. *PLoS One* 3(10): e3393.
 - 8 Takeda, J., Suzuki, Y., Sakate, R., Sato, Y., Seki, M., Irie, T., Takeuchi, N., Ueda, T., Nakao, M., Sugano, S., Gojobori, T. and Imanishi, T. (2008). Low conservation and species-specific evolution of alternative splicing in humans and mice analysed with comparative genomics using well-annotated full-length cDNAs. *Nucleic Acids Res.* 36(20): 6386-6395.
 - 9 Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, DP., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O. and Yon Rhee, S. (2008). Big data: The future of biocuration. *Nature* 455(7209): 47-50.
 - 10 Genome Information Integration Project and H-Invitational 2 Consortium: Yamasaki, C., Imanishi, T., Gojobori, T. et al. (2008). The H-Invitational Database (H-InvDB), a comprehensive annotation resource for human genes and transcripts. *Nucleic Acids Res.* 36(Database issue), D793-799.
 - 11 Matsuya, A., Sakate, R., Kawahara, Y., Koyanagi, KO., Sato, Y., Fujii, Y., Yamasaki, C., Habara, T., Nakaoka, H., Todokoro, F., Yamaguchi, K., Endo, T., Oota, S., Makalowski, W., Ikeo, K., Suzuki, Y., Hanada, K., Hashimoto, K., Hirai, M., Iwama, H., Saitou, N., Hiraki, AT., Jin, H., Kaneko, Y., Kanno, M., Murakami, K., Noda, AO., Saichi, N., Sanbonmatsu, R., Suzuki, M., Takeda, J., Tanaka, M., Gojobori, T., Imanishi, T. and Itoh, T. (2008). Evola: Ortholog database of all human genes in H-InvDB with manual curation of phylogenetic trees *Nucleic Acids Res.* 36(Database issue), D787-792.
 - 12 Sugawara, H., Ogasawara, O., Okubo, K., Gojobori, T. and Tateno, Y. (2008). DDBJ with New System and Face. *Nucleic Acids Res.* 36(Database issue), D22-24.
- ## 2. 学会発表
- 1 五條堀 孝 (2008) 「オミックス医療のためのデータベース戦略-パーソナルゲノム時代を迎えて」、第1回オミックス医療研究会シンポジウム・定期講演会、学術総合センター (東京) 7月3日
 - 2 五條堀 孝 (2008) 「ゲノム科学の最前線-肝炎ウイルスと肝臓病のゲノム的研究アプローチ」、第6回肝臓病研究会シンポジウム、六本木アカデミーヒルズ (東京) 7月5日 (特別講演)
 - 3 五條堀 孝 (2008) 「大進化を論じるためのオミックス的基盤を考える」、「大進化・論」(S10)、第10回日本進化学会大会、東京大学駒場キャンパス (東京) 8月24日
 - 4 T. Gojobori (2008) "Personal Genome Sequencing and its Strategic Construction of Database", The Human Variome Project "Collection of Human GENE VARIATION", HGM 2008, Hyderabad International Convention Centre (HICC) (Hyderabad, India) 9月27日
 - 5 T. Gojobori (2008) "Sequencing Revolution and Human Genome Network", Session 3 "Genome Informatics to Genome Biology", HGM 2008, Hyderabad International Convention Centre (HICC) (Hyderabad, India) 9月29日
- ## G. 知的所得権の所得状況
1. 特許取得
なし
 2. 実用新案登録
なし
 3. その他
なし

参考資料1：統合型肝炎ウイルスデータベース設計構築

図1. ウイルス・ホスト因子の統合（シェーマ）

図2. 統合型肝炎ウイルスデータベースの構築。Affymetrix SNP Array によりコールされた SNP を GWAS 解析（ここでは HCV 陽性 .vs. 健常者）し、有意な関連が得られそうな候補を抽出する。データベースには、Affymetrix SNP Array により検出可能な全 SNP の情報、SNP コールの結果、GWAS により関連が認められた SNP とその統計値、各患者の臨床情報が格納される。

図3. 統合型データベースの構造。データベースの概念的な構造を示す。各患者に関連付けて、臨床情報（診断情報）、感染ウイルスの種類とサブタイプ、各 SNP のジェノタイプがまとめられている。ウイルス情報からは HVDB へ、SNP の種類からは GWAS-DB および dbSNP へリンクし、その情報を探索することが出来る。

図4. SNP の検索。GWAS で得られた統計値や、SNP そのものの情報（関連する遺伝子、ゲノム上の位置、種類 cSNP/sSNP/iSNP/gSNP、など）、その他キーワードから SNP を絞り込み、リストアップできる。ここで、GWAS-DB、dbSNP の対応するエン트리へリンクできる。

図5. 候補 SNP の再検索。素データに戻って2群を再定義し、それを基に GWAS を行うことで、改めて候補 SNP 探索を行うことも出来る（機能準備中）。

まとめ

1) フレームワークとしての DB は完成

- データの導入は随時可能、既導入データは、
 - ・ 健常者 184 件、HBV 感染者 100 件、HCV 感染者 236 件のタイピング結果
 - ・ 同、臨床情報（継続中）
 - ・ 6 件の関連解析結果
- 検索、結果参照、外部情報（GWAS、dbSNP）参照は可能

2) To do

- 統計解析ツールの準備
 - ・ 特に、サンプルグループの新規定義と、関連解析の再実行
- セキュリティを考慮した形での公開

参考資料2：統合型データベース設計