

令和 5 年 6 月 1 日現在

機関番号：15401

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K08288

研究課題名（和文）宿主と腸内共生菌のゲノム情報を用いた機械学習による非ウイルス性肝発癌リスクの予測

研究課題名（英文）Prediction of the risk of non-viral hepatocarcinogenesis by machine learning using genomic information of host and intestinal commensal bacteria

研究代表者

三木 大樹（Miki, Daiki）

広島大学・医系科学研究科（医）・講師

研究者番号：10584592

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：非ウイルス性の肝発癌が増加傾向である。その原因の一つが非アルコール性脂肪肝炎（NAFLD）の増加である。一方で、腸内細菌叢が糖尿病や肥満、さらには肝疾患を含めた種々の疾患に関与することが明らかとなってきた。本研究では、ヒトゲノムとその転写産物であるトランスクリプト、および共生微生物である腸内細菌のゲノム、これらの情報を機械学習の手法により統合的に解析し、非ウイルス性肝発癌リスク因子の抽出を目指した。その結果、複数の肝発癌予測モデルを構築することができた。各モデルの予測能については、今後、前向きにその精度を確認していく予定である。

研究成果の学術的意義や社会的意義

これまで経験を積み上げてきたヒトゲノムあるいはがんゲノムやトランスクリプト研究といった特徴的な基盤の上に、さらにメタゲノムという共生する非宿主生物由来の新たな情報を加えることで、独自のアプローチで非ウイルス性肝発癌のリスク因子に迫ることを目指す。食生活の変化を含めた環境要因などの外的要因がリスクを高めるとも推測されるが、それに影響をうけて可逆的に変化をするメタゲノム情報と、一生変わることなく安定なヒトゲノム情報とを網羅的に統合解析するのは、新たな手法である。これまでの単階層解析では見えてこなかった未知の特徴量の抽出を行い、より有用な疾患リスク評価を行うことで、これら疾患群の診療への貢献を目指す。

研究成果の概要（英文）：Non-viral liver cancers are on the rise. One of the causes is the increase in non-alcoholic steatohepatitis (NAFLD). On the other hand, it has become clear that the gut microbiota is involved in various diseases including diabetes, obesity, and even liver diseases. In this study, we aimed to extract non-viral hepatocarcinogenic risk factors by integrating the human genome, its transcripts, and the genomes of commensal microorganisms, i.e., intestinal bacteria, using machine learning methods. As a result, we were able to construct several models for predicting hepatocarcinogenesis. We plan to prospectively confirm the prediction accuracy of each model.

研究分野：ゲノム医科学、肝臓学

キーワード：ヒトゲノム 腸内細菌 非ウイルス性肝発癌 機械学習

## 1. 研究開始当初の背景

わが国における肝癌の原因として、最多の etiology が C 型肝炎ウイルス (HCV)、次いで B 型肝炎ウイルス (HBV) によるものであったが、近い将来、それらに取って代わる勢いで非ウイルス性の肝癌が増えてきた。その主な原因が非アルコール性脂肪肝炎 (NASH) と考えられている。わが国における糖尿病罹患人口および肥満人口の増加とともに、NASH が増加しており、国内に 100 万人以上存在すると見積もられている。肝癌と診断された時点で、肝硬変となった背景肝には肝細胞の脂肪化などが見られず、NASH に起因した肝癌とは確定しづらい症例が多々あるが、その中でアルコール性や自己免疫性肝炎や原発性胆汁性胆管炎などが否定的な場合には、肝発癌の原因つまり何がリスク因子となったのかの判断が困難である。また、脂肪肝が無く、肝硬変にも至っていない線維化の比較的軽度な肝臓にも、発癌をきたす症例がある。今まさに増加を続けている非ウイルス性の肝癌について、そのリスク因子を明らかにすることが、喫緊の課題である。

これまでに腸内細菌叢 (腸内フローラ) の乱れと様々な疾患との関係が多数報告されている。炎症性腸疾患をはじめとした消化管疾患の他、肥満や糖尿病などの生活習慣病 (Qin J, *et al. Nature* 2012, Karlsson FH, *et al. Nature* 2013)、およびそれらと関連する非アルコール性脂肪性肝疾患 (NAFLD) や NASH (Mouzaki M, *et al. Hepatology* 2013, Boursier J, *et al. Hepatology* 2016) にも関係することが報告されているものの、まだ一定の見解が得られているとは言えない。一方、最近ではヒトの遺伝的要因と腸内細菌叢との関わりを示唆する報告が出てきており (Goodrich JK, *et al. Cell* 2014, Goodrich JK, *et al. Science* 2016)、この点についてもさらなる検証が待たれている。さらに、疾患の発症・進展における両者の関わりや相互作用については、ほとんど分かっていない。

これまでの疫学研究、ヒトゲノムやがんゲノム研究、あるいはメタゲノム研究、それぞれでは、非ウイルス性肝癌がなぜ増加しているのか、そのリスク因子は何であるのか、明確な答えは出ていない。このリスク因子を一つでも多く明らかにして、高リスク群を囲い込み、より効率的なスクリーニングを行うことが必要である。

しかし、非ウイルス性肝癌がなぜ増加しているのか、そのリスク因子は何であるのか、現時点では明確な答えが出ておらず、高リスク群の囲い込みには決定打が無い状況である。申請者はこれまで、肝疾患に関するヒトゲノム研究を中心に行ってきた。その易罹患性に関連する遺伝的要因を明らかにしてきた。しかし、ヒトゲノムで規定されるリスクの効果量は小さいため、病態メカニズム解明のヒントになったり、一定の疫学的な意義はあるものの、個人でのリスク評価にはあまり有効ではなかった。そこでゲノミクスからトランスクリプトミクスにも研究対象を広げた。血清中 miRNA 量が肝炎の etiology の違いや病期・病勢の違いを反映し得ることを報告した経験もあるが (*PLoS One* 2012, *J Med Virol* 2013, *J Viral Hepat* 2014, *J Infect* 2015)、これらはウイルス性肝炎を対象としていた。ウイルス性肝炎では、宿主因子に加えて、肝細胞内に慢性感染する HBV や HCV を併せて解析する必要があることから、ウイルスゲノムの次世代シーケンシング解析も多数実施し、その変異と宿主遺伝子多型との間の相関関係を明らかにした (*J Hepatol* 2015, *J Infect Dis* 2016)。さらに、癌に対しては、癌特異的な変異に着目して血中循環腫瘍 DNA を経時的に検討して、術後再発モニタリングなどに有用であることを示した (*Cell Mol Gastroenterol Hepatol* 2015, *Int J Cancer* 2018)。このような経緯の中で、何らかの外的要因 (食習慣を含めた環境変化など) がリスクとなって非ウイルス性の肝癌が増加しているという可能性を考えたときに、腸管内に共生する重さにして 1kg 超、数にして数十兆個の腸内細菌叢の関与が大きいのではないかと想起された。実際、糖尿病や肥満、さらには NASH/NAFLD については、腸内細菌叢の関与がすでに報告されている。しかし、検討症例数が不十分、人種の違い、他の交絡因子、機能的意義が不明、などの問題もあり、コンセンサスに至っているとは言い難い。肝癌については、肥満に伴って増加する腸内細菌の代謝産物が肝癌発症を促進する (Yoshimoto S, *et al. Nature* 2013)、腸内細菌叢のうち 1 次胆汁酸を 2 次胆汁酸に変換する過程に関わる細菌が、間接的に肝癌に対する免疫反応を制御する (Ma C, *et al. Science* 2018) といったことが、マウスモデルを用いた実験で示されている。今回の研究では、1 万人規模の GWAS データを活用して (ゲノミクス)、トランスクリプトミクスから、共生微生物のメタゲノミクスまでを含めた解析を行った上で、すべてを統合解析するという着想を得た。これを実現するためには機械学習が最適であるとの考えに至った。

また、腸内細菌叢は、食生活を主とした生活習慣に大きな影響を受けて形成され、加齢によって変動することが知られている。また、抗生物質の他、プロトンポンプ阻害薬や NSAIDs などによっても影響を受けるとされる。宿主の遺伝的要因がどの程度、腸内細菌叢の構成に影響を与えるのかについては、まだ十分に分かっていない。双生児研究などから、環境因子に比べるとその影響は弱いと考えられる一方で、ビタミン D 受容体 (Wang J, *et al. Nat Genet* 2016) や乳糖分解酵素 (Bonder MJ, *et al. Nat Genet* 2016, Kato K, *et al. PLoS One* 2018) の遺伝子多型が特定の腸内細菌と関連することが報告されている。また、メタボリック症候群 (Lim MY, *et al. Gut* 2017) や炎症性腸疾患 (Imhann F, *et al. Gut* 2018) では、疾患に関連する遺伝的要因が特定の腸内細菌と関連して病態を修飾している可能性が示唆されている。一方、肝疾患にお

いてはこのような報告は見当たらない。申請者らは、ここ2年間で500例超の糞便サンプルを収集しており、現在も収集中である。ウイルス性および非ウイルス性肝疾患、健常人を含む計431例の腸内細菌叢解析を実施した結果、etiology別に特徴的なパターンが確認された(2019年11月の米国肝臓病学会で発表)。今後、年齢や背景肝や服薬情報などの詳細な情報も含めた上で、より多数例での解析が必要である。

## 2. 研究の目的

申請者は、ヒトゲノムに見られる個人差・多様性の代表である一塩基多型(SNP)に着目して、全ゲノム領域を網羅的に調べるゲノムワイド関連解析(GWAS)という手法を主に用いて、C型肝炎(Miki, *et al.* *Nat Genet* 2011)、C型肝炎(Miki, *et al.* *PLoS One* 2013)、C型肝炎硬変(Urabe, *et al.* *J Hepatol* 2013)、HCV治療応答(Ochi, *et al.* *J Gen Virol* 2014)と関連するSNPを報告してきた。また、NASH発症に関連するSNPが、NASH自体よりも発症により強い効果を示すことを報告した(日米癌合同会議、2019年2月、USA)。しかし、個々のSNPが有する効果量は決して大きくはなく、ヒトゲノム情報だけでは高リスク症例を効率的に囲い込むことが困難であった。そんな中、次世代シーケンサーの登場で、全ゲノム/全エクソーム解析やその転写産物を網羅的にみるトランスクリプトーム解析が可能となった。肝臓の全ゲノム解析で、癌ゲノムに生じた変異にはetiology間で大きな違いが認められなかったのに対し、トランスクリプトーム解析ではetiology間の違いが認められた。

一方、次世代シーケンサーの発展は、細菌のゲノム解析にも恩恵をもたらした。腸内細菌叢の構成は、主に食習慣などの生活歴に影響されながら成長過程で変化し、成人となつてからは比較的安定であると考えられているが、宿主の遺伝的背景がどの程度影響するのかという点についてはほとんど分かっていない。

これまで経験を積み上げてきたヒトゲノムあるいは癌ゲノムやトランスクリプト研究といった特徴的な基盤の上に、さらにメタゲノムという共生する非宿主生物由来の新たな情報を加えることで、独自のアプローチで非ウイルス性肝臓のリスク因子に迫ることを目指す。食生活の変化を含めた環境要因などの外的要因がリスクを高めるとも推測されるが、それに影響を受けて可逆的に変化をするメタゲノム情報と、一生変わることなく安定なヒトゲノム情報とを網羅的に統合解析することで、新たな知見を得られる可能性がある。疾患リスクの探索において、宿主と非宿主である共生微生物の情報を機械学習に供するという新たな解析手法を確立したい。

## 3. 研究の方法

### (1) 腸内細菌叢(メタゲノム)

糞便サンプル収集:非ウイルス性肝発症例および非発症例、比較する対象としてのウイルス性肝疾患症例および健常人について、臨床情報とともに収集する。腸内細菌叢には個人差が大きく、症例ごとに臨床像が多様であることから、機械学習に供するにはより大きなサンプル数、いわゆるビッグデータが必要である。合計1,000例を目標とする。

16S rRNA解析:DNAを抽出し、菌種間で保存されたゲノム領域に設定した共通プライマーを用いてPCR反応を行う。ライブラリー調整後、次世代シーケンサー(Ion PGM)により配列を決定する。可変領域の相同性を細菌データベースの参照配列に照らし、分類・同定を行い、組成比を評価する。

### (2) ヒトゲノムおよびトランスクリプト

ヒトゲノム解析:これまでに蓄積された1万人規模のGWASデータを活用する。1万人のうち、2000人が健常人、8000人が肝疾患患者であり、肝疾患患者のうち、2500人は肝発症した症例である。Etiology別では、HCV関連が4000人、HBV関連が3000人、その他(非B非C)が1000人となっている。

トランスクリプト:上記のGWAS症例のうち、のべ500例超でトランスクリプト解析が実施済みである(血清、血球、癌部・非癌部肝組織)。

### (3) 機械学習

メタゲノム+ヒトゲノム+臨床情報を投入して、非ウイルス性肝発症の予測、リスク因子の抽出を試みる。

## 4. 研究成果

初年度は、今後の解析に供するための糞便サンプル収集と、それらに付随する臨床情報の収集とデータクリーニングを行った。具体的には、非ウイルス性肝発症例および非発症例、比較する対象としてのウイルス性肝疾患症例および健常人についてである。二年度目には収集したサンプル数が1000に到達し、順次、DNA抽出、16S rRNA解析に供した。得られたデータセットを用いて、機械学習プラットフォームの構築を行い、PCAやクラスタリング、多変量解析等の試験運用を開始した。三年度目には、当該症例についてヒトゲノム情報(GWASデータ)が有るものに関しては、統合解析を進めた。その結果、複数の肝発症予測モデルを構築できた。各モデルの予測能については、今後、前向きにその精度を確認していく予定である。

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 8件 / うち国際共著 1件 / うちオープンアクセス 2件）

1. 著者名 Nagaoki Y, Hyogo H, Ando Y, Kosaka Y, Uchikawa S, Nishida Y, Teraoka Y, Morio K, Fujino H, Ono A, Nakahara T, Murakami E, Yamauchi M, Okamoto W, Kawaoka T, Tsuge M, Hiramatsu A, Miki D, Imamura M, Takahashi S, Chayama K, Aikata H.	4. 巻 21(1)
2. 論文標題 Increasing incidence of non-HBV- and non-HCV-related hepatocellular carcinoma: single-institution 20-year study.	5. 発行年 2021年
3. 雑誌名 BMC Gastroenterol.	6. 最初と最後の頁 306
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12876-021-01884-5	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 三木大樹、茶山一彰	4. 巻 2(5)
2. 論文標題 C型肝炎による発癌のリスク因子	5. 発行年 2020年
3. 雑誌名 消化器内科	6. 最初と最後の頁 76-82
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 三木大樹、相方浩、茶山一彰	4. 巻 63(4)
2. 論文標題 肝炎ウイルス排除後の肝発がん	5. 発行年 2021年
3. 雑誌名 月刊薬事	6. 最初と最後の頁 77-82
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Daijo K, Nakahara T, Inagaki Y, Nanba M, Nishida Y, Uchikawa S, Kodama K, Oya K, Morio K, Fujino H, Ono A, Murakami E, Yamauchi M, Kawaoka T, Miki D, Tsuge M, Hiramatsu A, Hayes CN, Imamura M, Aikata H, Ochi H, Chayama K.	4. 巻 35(8)
2. 論文標題 Risk factors for histological progression of non-alcoholic steatohepatitis analyzed from repeated biopsy cases.	5. 発行年 2020年
3. 雑誌名 J Gastroenterol Hepatol.	6. 最初と最後の頁 1412-1419
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/jgh.14968	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakae S, Matoba N, Low SK, Okada Y, Terao C, Amariuta T, Gazal S, Kochi Y, Horikoshi M, Suzuki K, Ito K, Koyama S, Ozaki K, Niida S, Sakata Y, Sakata Y, Kohno T, Shiraishi K, Momozawa Y, Hirata M, Matsuda K, Ikeda M, Iwata N, Miki D, et al.	4. 巻 52(7)
2. 論文標題 Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases.	5. 発行年 2020年
3. 雑誌名 Nat Genet.	6. 最初と最後の頁 669-679
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41588-020-0640-3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Yamamoto M, Kobayashi T, Mashima H, Miki D, Kuroda S, Hamaoka M, Aikata H, Chayama K, Ohdan H.	4. 巻 80
2. 論文標題 PD1 gene polymorphism is associated with a poor prognosis in hepatocellular carcinoma following liver resection, cohort study.	5. 発行年 2020年
3. 雑誌名 Int J Surg.	6. 最初と最後の頁 84-90
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ijssu.2020.05.007	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nagaoki Y, Imamura M, Teraoka Y, Morio K, Fujino H, Ono A, Nakahara T, Murakami E, Yamauchi M, Kawaoka T, Miki D, Tsuge M, Hiramatsu A, Hayes CN, Chayama K, Aikata H.	4. 巻 50(11)
2. 論文標題 Impact of viral eradication by direct-acting antivirals on the risk of hepatocellular carcinoma development, prognosis, and portal hypertension in hepatitis C virus-related compensated cirrhosis patients.	5. 発行年 2020年
3. 雑誌名 Hepatol Res.	6. 最初と最後の頁 1222-1233
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/hepr.13554	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Miki D, Akita T, Kurisu A, Kawaoka T, Nakajima T, Hige S, Karino Y, Toyoda H, Kumada T, Tsuge M, Hiramatsu A, Imamura M, Aikata H, Hayes CN, Honda K, Seike M, Akuta N, Kobayashi M, Kumada H, Tanaka J, Chayama K.	4. 巻 55(12)
2. 論文標題 PNPLA3 and HLA-DQB1 polymorphisms are associated with hepatocellular carcinoma after hepatitis C virus eradication.	5. 発行年 2020年
3. 雑誌名 J Gastroenterol.	6. 最初と最後の頁 1162-1170
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00535-020-01731-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Liu S, Murakami E, Nakahara T, Ohya K, Teraoka Y, Makokha GN, Uchida T, Morio K, Fujino H, Ono A, Yamauchi M, Kawaoka T, Miki D, Tsuge M, Hiramatsu A, Abe-Chayama H, Hayes NC, Imamura M, Aikata H, Chayama K.	4. 巻 23(16)
2. 論文標題 In vitro analysis of hepatic stellate cell activation influenced by transmembrane 6 superfamily 2 polymorphism.	5. 発行年 2021年
3. 雑誌名 Mol Med Rep.	6. 最初と最後の頁 1-7
掲載論文のDOI (デジタルオブジェクト識別子) 10.3892/mmr.2020.11654	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tada T, Toyoda H, Kumada T, Kurisu A, Sugiyama A, Akita T, Ohisa M, Aikata H, Miki D, Chayama K, Tanaka J.	4. 巻 28(3)
2. 論文標題 Comparison of liver disease state progression in patients with eradication of versus persistent infection with hepatitis C virus: Markov chain analysis.	5. 発行年 2021年
3. 雑誌名 J Viral Hepat.	6. 最初と最後の頁 538-547
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/jvh.13444	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計6件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 三木大樹, 田中純子, 茶山一彰
2. 発表標題 SVR後新規肝発癌リスク層別化における経時的AFP変化の意義
3. 学会等名 第29回日本消化器関連学会週間
4. 発表年 2021年

1. 発表者名 三木大樹, 相方浩, 藤野初江, 大野敦司, 山内理海, 柘植雅貴, 茶山一彰
2. 発表標題 SVR後の肝発癌リスク予測に有用な血清AFP値に関する検討
3. 学会等名 第80回日本癌学会学術総会
4. 発表年 2021年

1. 発表者名 三木大樹、田中純子、茶山一彰
2. 発表標題 C型肝炎ウイルス排除後の新規肝発癌とPNPLA3遺伝子多型の検討
3. 学会等名 第57回日本肝臓学会総会
4. 発表年 2021年

1. 発表者名 三木大樹、田中純子、茶山一彰
2. 発表標題 C型肝炎ウイルス排除後の新規肝発癌に関わるリスク因子の検討
3. 学会等名 第28回日本消化器関連学会週間
4. 発表年 2020年

1. 発表者名 三木大樹、相方浩、藤野初江、大野敦司、山内理海、柘植雅貴、茶山一彰
2. 発表標題 PNPLA3遺伝子多型はHCV排除後の肝発癌に関連する
3. 学会等名 第79回日本癌学会学術総会
4. 発表年 2020年

1. 発表者名 三木大樹、河岡友和、茶山一彰
2. 発表標題 肝疾患における病因特異的な腸内細菌叢パターンの探索
3. 学会等名 第106回日本消化器病学会総会
4. 発表年 2020年

〔図書〕 計1件

1. 著者名 三木大樹、茶山一彰	4. 発行年 2021年
2. 出版社 科学評論社	5. 総ページ数 6
3. 書名 消化器・肝臓内科 第10巻第6号「SVR後肝発癌と遺伝子多型」	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------