



東北大学



東北メディカル・メガバンク機構  
TOHOKU MEDICAL MEGABANK ORGANIZATION

2015年8月20日

報道機関各位

東北大学 東北メディカル・メガバンク機構

## 日本人 1,070 人の高精度全ゲノムデータの統合的な解析に成功 ～お米の消化の遺伝子の個人差や HLA の詳細などが統合解析からみえてくる～

### 【発表のポイント】

- ・ 1,000 人以上の全ゲノム情報に基づく日本人がもつアレル頻度 0.1%以上の SNVs をほぼ (90%以上) 網羅した解析を世界で初めて行いました。本研究を通じて明らかにされたアレル頻度 5.0%以上の SNVs の頻度情報については、リファレンスパネルの概要版として公開 (2014年8月) してきましたが、今回の論文発表をもって完成版としました。
- ・ 日本人の全ゲノムの中の 2 万個以上の遺伝子のほぼ全ての領域におけるコピー数の違いについての詳細プロフィールを世界で初めてデータベース化しました。
- ・ 日本人を対象としたあらゆるゲノムに関連した研究に役に立つことが期待されます。
  - (例 1) 希少疾患の原因変異かどうかの推定
  - (例 2) 遺伝子の機能に関わる個人差の原因となる変異の探索
  - (例 3) 日本人特有のゲノムに基づくリスク診断・医療・創薬のための基盤情報

### ◆将来の展望

当事業および地域のコホートとの連携によるリファレンスパネルの人数規模の拡大、また、解析手法の改良によるさらなるリファレンスパネルの高精度化を行う予定です。さらに、全国の研究者への分譲や共同研究による利活用の展開も行います。

### 【概要】

東北大学東北メディカル・メガバンク機構ゲノム解析部門を中心とした研究チームは、東北メディカル・メガバンク計画のコホート調査\*1に参加した健常な日本人1,070人分の全ゲノムを解析し、精度検証を進めることで、日本人の全ゲノムリファレンスパネル (以下、1KJPN) を構築しました。

今回、1,070人分のDNA情報を次世代シーケンサーで読み取り、のべ約100兆塩基もの高品質な全ゲノム断片配列情報を解読し、スーパーコンピュータ\*2による情報解析技術と他の手法による実験結果による検証とを組み合わせることで、最終的に信頼度の高い2,120万箇所の一塩基バリエーション\*3 (single-nucleotide variants : 以下 SNVs) を発見しました。これら

SNVsのうち1,200万箇所はこれまで国際データベースに報告されていない新規のSNVsでした。

今回の1KJPNでは、SNVs同定対象の常染色体領域上の日本人がもつアレル頻度0.1%以上のSNVsをほぼ(90%以上)網羅できていることがスーパーコンピュータによる集団遺伝学のモデルを用いたシミュレーションにより確かめられています。

研究チームは、SNVsに加え、日本人がもつ340万箇所の数十塩基以内の挿入及び欠失(うち新規約200万箇所)や、全ゲノム中の2万個以上の遺伝子のほぼすべての領域におけるコピー数変化の詳細プロファイルの作成についても世界で初めて成功しました。

また、お米に含まれるデンプンの消化に関係するアミラーゼ遺伝子(AMY1)が偶数個の単位で個人差があることや、免疫疾患に関連するヒト白血球型抗原(HLA)遺伝子<sup>\*4</sup>の詳細プロファイルなどの作成などにも成功しました。

この一連の研究で発見された情報は、日本人に固有な体質・疾患の関連遺伝子を大規模に探索研究するための基盤情報であり、日本人の個別化予防・医療研究を加速する重要な成果です。本研究成果は、2015年8月21日以降に英国科学誌「Nature Communications (ネイチャー・コミュニケーションズ)」オンライン版で公開されます。

## 【背景】

東北大学東北メディカル・メガバンク機構(以下、ToMMo)は、岩手医科大学と共に東日本大震災からの復興事業として、東北メディカル・メガバンク計画に取り組んでおり、両大学は、宮城県・岩手県の地域住民15万人規模のコホート調査を2013年から実施しています。

ToMMoでは、このコホート調査に参加された宮城県住民の1,070人分の全ゲノム配列に基づく「全ゲノムリファレンスパネル」(以下、1KJPNパネル<sup>\*5</sup>)の構築を行っており、2013年11月には、1,000人分の全ゲノム解読が完了<sup>\*6</sup>したことを発表し、続けて、2014年8月には1KJPNパネルのアレル頻度<sup>\*7</sup>5%以上のSNP頻度情報につき、ウェブサイト上で公開<sup>\*8</sup>しています。また、2015年7月には日本人に最適化されたSNPアレイ「ジャポニカアレイ<sup>®</sup>」<sup>\*9</sup>の設計に成功した報告がJournal of Human Genetics誌に掲載されました。

## 【詳細】

次世代シーケンサーによる高精度なゲノム解析により、東北メディカル・メガバンク計画のコホート調査に参加した健常な日本人1,070人分からなる1KJPNパネルの構築に成功しました。

1KJPNパネルは、アレル頻度の極めて低いSNVs(0.1%以下)から高いSNVs(5%以上)の広範囲にわたる2,120万箇所のSNVsを含み、そのうち56.6%が新規のSNVs(表1)でした。

これまでは日本人集団の全ゲノム情報に基づく情報は国際1,000人ゲノム計画(1KGP)<sup>\*10</sup>に含まれた日本人89人分の情報しかありませんでした。また、全ゲノム断片配列情報は、当機構では一人当たり平均32.4回分の解読量であるのに対して4~6回分の解読量に限られており1KGPではアレル頻度1%以上の限定されたSNVの同定が主な目標となっていました。し

かし、今回、均質な日本人の集団に焦点をあてたことで1,070人という大規模な日本人の全ゲノム詳細情報を解析することで、アレル頻度が極めて低いSNVsの同定にも成功しました（図1）。

1KJPNパネルでは、SNVs同定対象の常染色体領域上の日本人がもつアレル頻度0.1%以上のSNVsをほぼ（90%以上）網羅できていることをスーパーコンピュータによる集団遺伝学のモデルを用いたシミュレーションにより検証しています。

さらに、情報解析を行うことで、340万箇所の百塩基以内の挿入及び欠失や、全ゲノム中の25,923個の遺伝子領域におけるコピー数の違いについてのプロファイル作成についても成功しました。

また、1,070人の全ゲノム断片配列情報を情報解析することで、お米を中心とした食生活を行っているアジア人で重要と考えられているアミラーゼ遺伝子について、日本人集団において偶数個を単位とした重複がおきていることを推定し、検証を行いました（図2）。さらに、臓器移植適合や、免疫疾患などにかかわると考えられているHLA遺伝子のうち、MHCクラス1（HLA-A）の日本人集団におけるHLA型頻度情報の詳細プロファイル情報を研究チームが開発を行ったソフトウェアHLA-VBSeqを使って作成することにも成功しました（図3）。

ToMMoでは一連の事業を通じて、東日本大震災後の東北地方において、早期に社会に還元できる成果を創出し復興への貢献を目指しています。

### 【今後の展開】

今後は全ゲノム解析を行う人数を数千人規模に増やしていただくだけではなく、東北メディカル・メガバンク計画のコホート調査の一つである三世代コホートの家系情報なども活用して、より高精度の日本人の全ゲノムリファレンスパネルの作成の構築を進めていきます。また、東北地方だけではなく、他の地域のコホート事業などとも連携を進めることで、日本における異なる地域の全ゲノム情報についても全ゲノムリファレンスパネルに加えてより網羅的なパネルとなるよう拡張をしていく予定です。

また、1KJPNパネルの頻度情報などについて分譲を進め、全国の研究者のゲノム研究を加速するとともに、共同研究の公募（第3回 2015年8月末締切など）を通じてさらなる高精度化やより高度な情報解析を進めていきます。

### 【研究について】

日本人1,070人の全ゲノム配列解読は、国の東日本大震災からの復興事業として、「東北メディカル・メガバンク計画」の一環として行われました。平成27年度より、国立研究開発法人 日本医療研究開発機構（AMED）が本計画の研究支援担当機関の役割を果たしています。

【参考図表】

	総サンプル数	1,070	
	総シーケンサー読取塩基数	100.4兆塩基	
	1人あたりのカバー率	32.4x	
	合計	21,221,195	
情報	既知変異	9,219,783	
	新規変異	12,001,412	
	新規の割合	56.55%	
	1人あたりの平均一塩基変異数	2,716,853	
	1人あたりの平均ヘテロ変異数	1,532,773	
			1,532,773
	欠失長	1塩基から99塩基まで	100塩基以上
欠失	総欠失数	1,969,302	47,343
	新規の数	1,429,636	-
	新規の割合	72.60%	-
	1人あたりの平均数	190,857	2,654
	挿入長	1塩基から99塩基まで	100塩基以上
挿入	総挿入数	1,384,230	9,354
	新規の数	1,037,839	9,354
	新規の割合	74.98%	-
	1人あたりの平均数	159,359	45

表1 : 日本人 1,070 人の解析結果

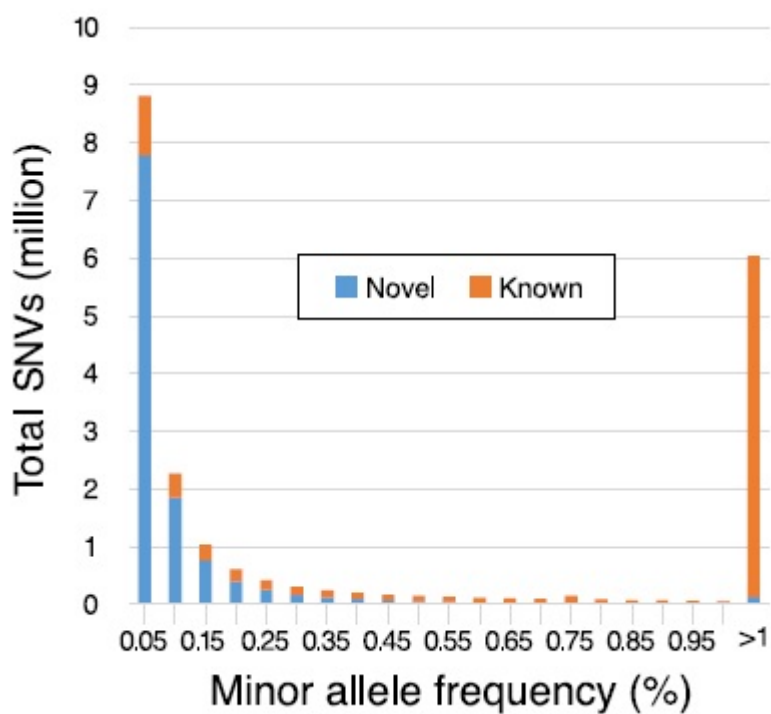


図1 : 新規と既知のSNVのアレル頻度毎の総数、青：新規の総数、オレンジ：国際データベースに報告されている総数

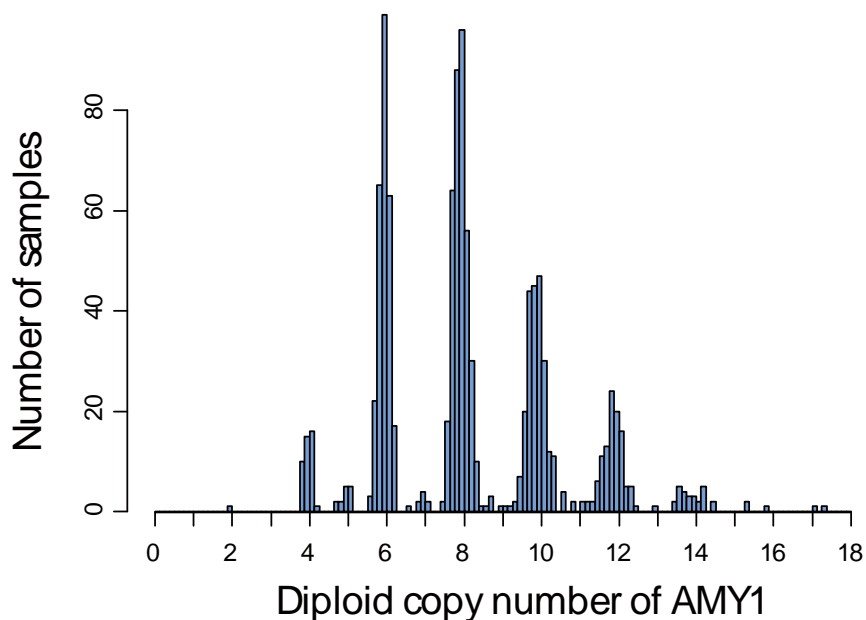


図2 : アミラーゼの個人毎に推定された遺伝子のコピー数、横軸：染色体2本でのアミラーゼ遺伝子の総個数、縦軸：1070人での割合

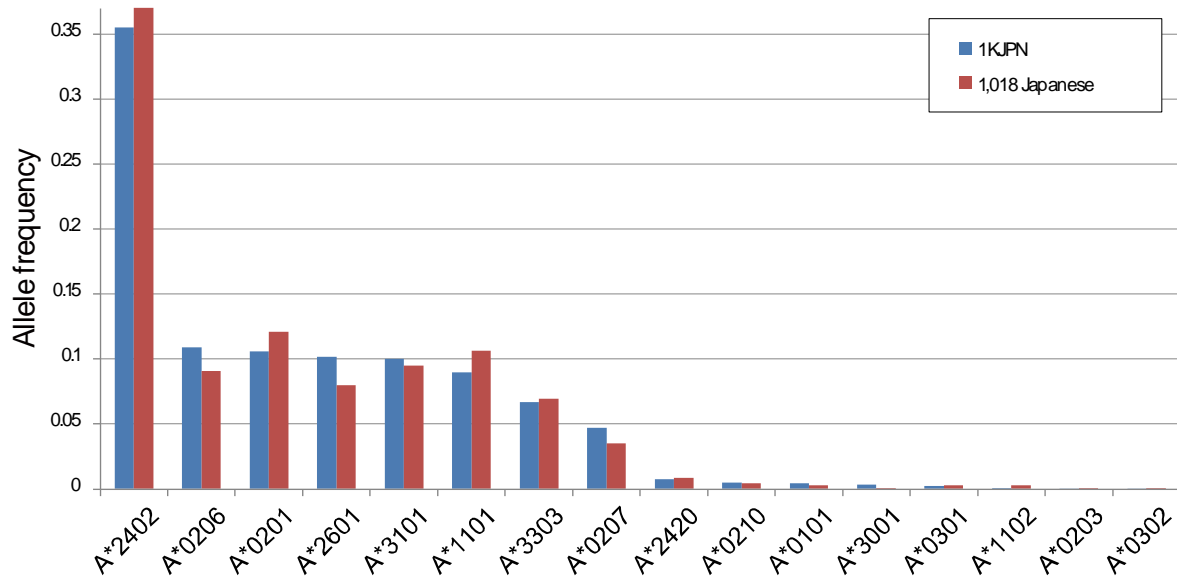


図 3： 1KJPN で推定された HLA-A の各型の頻度情報（青）、他手法で 1,018 人について過去に観測された頻度情報（赤）

### 【参考】

<東北メディカル・メガバンク計画について>

本計画は、東日本大震災を受け、被災地住民の健康不安の解消に貢献するとともに、個別化予防等の東北発の次世代医療を実現するため、ゲノム情報を含むゲノムコホート研究等を実施し、被災地域の復興を推進するものです。

被災地に医療関係人材を派遣して地域医療の復興に貢献するとともに、15 万人規模の地域住民コホートと三世帯コホートを形成し、そこで得られる生体試料、健康情報、診療情報等を収集してバイオバンク<sup>\*1</sup>を構築します。さらに、ゲノム情報、診療情報等を解析することで、個別化予防等の次世代医療に結びつく成果を創出することを目指しています。また、得られた生体試料や解析成果を同意の内容等に十分留意し、個人情報保護のための匿名化等の適切な措置を施した上で、外部に提供することや、コホート調査や解析研究を行うための多様な人材の育成も行っています。

本計画の事業の実施は、東北大学東北メディカル・メガバンク機構と岩手医科大学いわて東北メディカル・メガバンク機構とが連携して行っています。

### 【用語解説】

\*1 コホート調査：ある特定の人々の集団を一定期間にわたって追跡し、生活習慣などの環境要因・遺伝的要因などと疾病の関係を解明するための調査のこと。

\*2 スーパーコンピュータシステム：東北メディカル・メガバンク機構は複合バイオバンクとしてデータバンクおよび解析の機能も併せ持っており、ライフサイエンス分野では日本最大級のスーパーコンピュータシステムの本格運用を開始している。

URL: <http://sc.megabank.tohoku.ac.jp/>

\*3 一塩基バリエーション(SNV)：個人間でゲノムの一塩基が異なる状態。なお、通常は一定以上の頻度(通常1%)で確認されるSNVを特に一塩基多型(Single Nucleotide Polymorphism) SNPと呼ぶ場合がある。

\*4 ヒト白血球型抗原(HLA)遺伝子：ヒトの主要組織適合遺伝子複合体(MHC)。クラスI分子は内因性抗原を抗原提示し、クラスII分子は外来性抗原を抗原提示する働きを持つ。MHC分子が抗原提示を行うことにより、免疫系が細菌、ウイルスなどの非自己抗原を認識し、感染を防御するように働く。

\*5 1KJPNパネル：大規模な人数の全ゲノム解読を行った結果を総合し、DNA配列の多型の頻度などの情報をまとめることで、今後のゲノム研究の参照情報となるよう、東北大学東北メディカル・メガバンク機構が構築を進めている全ゲノムリファレンスパネルのこと。現在、1,070人のコホート参加者の全ゲノム解読により、アレル頻度0.5%程度までの稀少SNPの同定を進めている。今後は、診療情報や生活習慣情報、血液・尿などのデータとを統合することで、我が国における次世代医療を目指す研究に幅広く活用可能なデータベースとして構築することを目指している。

\*6 1,000人分の全ゲノム解読が完了：2013年11月29日、ToMMoは東北メディカル・メガバンク計画のコホート調査事業に参加した宮城県在住の健常な日本人1,000人分の全ゲノムを解読した。

参考：プレスリリース「疾病原因探索の基盤となる1,000人分の全ゲノム配列の高精度解読を完了 ~1,500万個におよぶ新たな遺伝子多型を収集~」

URL: <http://www.megabank.tohoku.ac.jp/news/682>

\*7 アレル頻度：ここでは、ある集団のSNVの中で観察数の少ない塩基(A,T,G,C)の頻度をいう。今回は抽出された日本人1,070人中がもつ2,140本の常染色体での割合となる。集団間でアレル頻度が大きく異なるSNVがあることが知られている。

\*8 アレル頻度5%以上のSNP頻度情報につき、ToMMoのポータルサイト Integrative Japanese Genome Variation Databaseウェブサイト (<http://ijgvd.megabank.tohoku.ac.jp/>)で公開

\*9 ジャポニカアレイ® : 66万箇所のSNPを搭載した日本人に最適化された初のSNPアレイ、1KJPNパネルを組み合わせることで頻度1%以上のSNVsのうち90% (約824万箇所) を復元することができる。

\*10 国際1,000人ゲノム計画 (1KGP) : 人類集団の詳細な遺伝的多様性を行うことを目指し、世界各地の約1,000人の全ゲノムシーケンスを行った国際研究計画。当計画は、現在解析人数を2,535人にまで拡張したphase3が進行中。

\*11 バイオバンク : 生体試料を収集・保管し、研究利用のために提供を行う。東北メディカル・メガバンク計画のバイオバンクは、コホート調査の参加者から血液・尿などの生体試料を集める。

### 【論文名】

Rare variant discovery by deep whole-genome sequencing of 1070 Japanese Individuals

『1,070人の日本人全ゲノム高深度解析によるレアバリエントの発見』

掲載予定誌 : *Nature Communications*

Masao Nagasaki\*, Jun Yasuda\*, Fumiki Katsuoka\*, Naoki Nariai, Kaname Kojima, Yosuke Kawai, Yumi Yamaguchi-Kabata, Junji Yokozawa, Inaho Danjoh, Sakae Saito, Yukuto Sato, Takahiro Mimori, Kaoru Tsuda, Rumiko Saito, Xiaoqing Pan, Satoshi Nishikawa, Shin Ito, Yoko Kuroki, Osamu Tanabe, Nobuo Fuse, Shinichi Kuriyama, Hideyasu Kiyomoto, Atsushi Hozawa, Naoko Minegishi, James Douglas Engel, Kengo Kinoshita, Shigeo Kure, Nobuo Yaegashi, ToMMo Japanese Reference Panel Project & Masayuki Yamamoto

\* These authors contributed equally to this work

ToMMo Japanese Reference Panel Project is the following members including main authors in this manuscript: Abe Michiaki, Akito Tsuboi, Fuji Nagami, Hiroshi Kawame, Hiroaki Tomita, Ichiro Tsuji, Jun Nakaya, Junichi Sugawara, Kichiya Suzuki, Masahiro Kikuya, Naoki Nakaya, Noriko Osumi, Riu Yamashita, Soichi Ogishima, Takako Takai, Teiji Tominaga, Yasuyuki Taki, Yoichi Suzuki

#### 【お問い合わせ先】

<研究内容に関すること>

東北大学東北メディカル・メガバンク機構

インシリコ解析 室長

教授 長崎 正朗 (ながさき まさお)

電話番号 : 022-273-6051

Eメール : 1kjpn@sc.megabank.tohoku.ac.jp

<報道担当>

東北大学東北メディカル・メガバンク機構

長神 風二 (ながみ ふうじ)

電話番号 : 022-717-7908

ファックス : 022-717-7923

Eメール : f-nagami@med.tohoku.ac.jp