



東北大学

2015年6月24日  
東北大学 東北メディカル・メガバンク機構

## 日本人に最適化された SNP アレイ「ジャポニカアレイ®」を設計 ～約 66 万個の SNP 情報を搭載した個別化予防・医療研究を加速する解析ツール～

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の長崎正朗教授、河合洋介講師らは、日本人集団のもつSNP<sup>\*1</sup>を高精度で取得できる全ゲノム領域を網羅する、日本人に最適化されたSNPアレイ「ジャポニカアレイ®」の設計に世界で初めて成功しました。

ジャポニカアレイ®の設計は、東北メディカル・メガバンク計画のコホート調査<sup>\*2</sup>に協力した1,070人分の全ゲノム情報を活用し、独自のSNP選択アルゴリズムを開発・実装してスーパーコンピュータ上で解析することで実現されました。

遺伝子型インピュテーション<sup>\*3</sup>（以下、インピュテーション）技術を用いることで約66万個のSNPを搭載したジャポニカアレイ®から最大2,000万SNPを取得可能です。

ジャポニカアレイ®は既存の同等数のSNPが搭載されているアレイと比べてインピュテーションの精度が10%以上向上し、また、3倍以上の数のSNPが搭載されている既存のSNPアレイとほぼ同等またはそれ以上の性能を発揮します。

この研究は、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為の基盤解析ツールであり、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。本研究成果の詳細は、2015年6月25日（英国時間）Journal of Human Genetics誌の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>\*4</sup>）の構築を行ってお

り、2013年11月には、1000人分の全ゲノム解読が完了\*5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5%以上のSNP頻度情報につき、ウェブサイト上で公開\*6しています。

次世代シーケンサーを用いて30億塩基対すべてを解読する全ゲノムシーケンス解析が提案されていますが、ゲノムワイド関連解析 (Genome-Wide Association Study: GWAS) 等に広く活用できるほどのスループットやコストには達していません。

そこで、遺伝子型を解析する研究では、SNPアレイを用いて数万～数百万程度のDNA配列を簡便に読むことにより、30億塩基対を統計学的手法により推定する方法 (インピュテーション\*3) が脚光を浴びるようになりました。このインピュテーションには、全ゲノムの情報を有するリファレンスパネルとSNPアレイが必要となります。これまで広く用いられてきたのは、国際1,000人ゲノム計画\*7で作成された少数の日本人集団を含むリファレンスパネルでした。また、SNPアレイは、国際HapMap計画\*8や国際1,000人ゲノム計画をもとに設計されたアレイでした。そのため、これらを用いた日本人ゲノムのインピュテーションによる解析は十分とはいえず、日本人向けのSNPアレイの開発が待たれていました。

ジャポニカアレイ®が国内の多くの研究機関、病院などで疾患の研究に活用されることで個別化予防・医療につながる原因遺伝子の発見やリスクSNPを同定するための強力なツールになることが期待されます。

## 【詳細】

約64万のタグSNP\*9と約2万の機能SNPを搭載し、約66万SNP (659,253 SNP) から構成されたジャポニカアレイ®を開発しました。(表1)

今回、その性能評価を行ったところ、既存のカタログアレイに対し、ジャポニカアレイ®でのインピュテーションの精度が同程度のSNPが搭載されているアレイに比べて10%以上、3倍以上のSNPが搭載されているアレイと同等またはそれ以上の性能が達成できることを示しました。(図1)

また、今回の発表にあわせて、研究者の興味のあるSNPが搭載されているかを評価できるようにジャポニカアレイ®に搭載されているSNPのリストについて一般公開\*10を開始しました。

さらに、ジャポニカアレイ®を用いて未搭載のSNPがどの程度の精度でインピュテーション可能かについて問い合わせを行うことができるポータルサイトを公開\*11しました。

ジャポニカアレイ®は、株式会社 東芝へのライセンスのもと、一般にジェノタイプングサービスの提供が行われています。

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】

ゲノムワイド関連解析に必要な全ゲノム規模のインピュテーションを行うには膨大な計算量を必要とします。本研究に合わせてジャポニカアレイ®と1KJPNを使った大規模インピュテーションのためのスーパーコンピュータで稼働する計算パイプラインの構築を行いました。現在、ToMMoで構築中の1KJPNパネルについて、サンプル数を増やすことで、アレル頻度の低いSNPについてもさらにインピュテーション精度を上げ、さらなる改善を行っていく予定です。成果を通じて、他地域でのゲノムコホート研究での活用を含め、我が国のゲノム研究全体に裨益していきたいと考えています。

## 【研究について】

日本人1,070人の全ゲノム配列解読及びこれに基づく全ゲノム領域を対象としたSNPアレイ「ジャポニカアレイ®」の設計は、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】

カテゴリ	SNP 数
タグ SNP	638,269
薬理ゲノミクス	2,028
Y 染色体	275
ミトコンドリア	70
NHGRI GWAS catalog	10,798
HLA 領域	3,906
その他	3,990
合計（上記は重複を含む）	659,253

表1 : ジャポニカアレイ®の構成

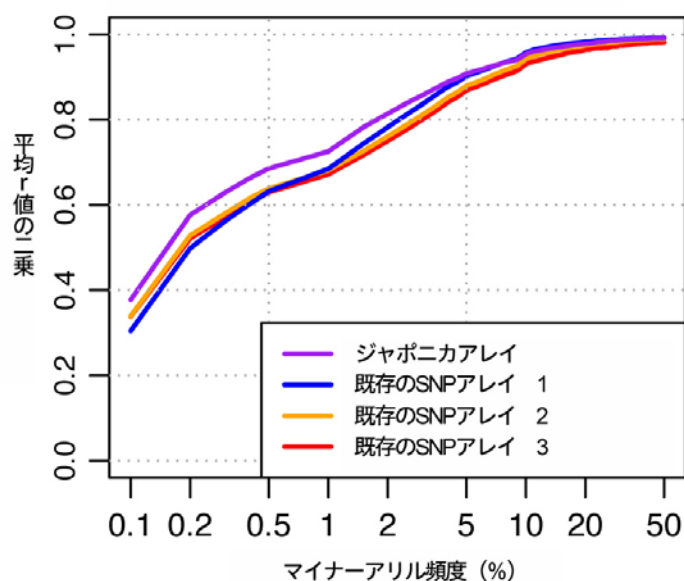


図1 : 1KJPNパネルとそれぞれ異なるSNPアレイを用いたインピュテーション結果

#### 【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代コホートを形成し、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>\*12</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 一塩基多型(SNP)：個人間でゲノムの1塩基が異なる状態。通常は一定以上の頻度で存在するものをSNPと呼ぶ。

\*2 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。

\*3 (遺伝子型) インピュテーション: リファレンスパネルと組み合わせることでSNPアレイによる計測をしていない数百万のSNPの遺伝子型を推定する情報科学的手法。

\*4 1KJPN: 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。現在、1070人のコホート参加者の全ゲノム解読により、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・尿の解析のデータなど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。

\*5 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>

\*6 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

\*7 国際1,000人ゲノム計画: 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。

\*8 国際HapMap計画: 世界各地の様々な人類集団のSNPタイピングを行いヒトの遺伝的多様性の全貌を明らかにすることを目標とした国際研究計画。

\*9 タグSNP: SNPリストの中から日本人集団がもつ2,000万以上のSNPのうち、インピュテーションの性能が良いSNPアレイ上に設計されたSNPの一覧

\*10 SNPのリストの一般公開: 本論文発表と同時に、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のウェブサイトで公開を開始している。なお、SNPリストと選択アルゴリズムについては特許出願中である。

URL: <http://nagasakilab.csml.org/en/japonica>

\*11 ジャポニカアレイ®未搭載のSNPのインピュテーション精度を問合わせるポータルサイトを公開

URL: <http://nagasakilab.csml.org/en/japonica>

\*12 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】

Yosuke Kawai, Takahiro Mimori, Kaname Kojima, Naoki Nariai, Inaho Danjoh, Rumiko Saito, Jun Yasuda, Masayuki Yamamoto, Masao Nagasaki

Japonica array: Improved genotype imputation by designing a population-specific SNP array with 1,070 Japanese individuals

「ジャポニカアレイ®： 1070人の日本人の情報に基づく日本人集団に適したSNPアレイの設計によるインピュテーションの改良」

掲載予定誌： *Journal of Human Genetics*

### 【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： [snparray@sc.megabank.tohoku.ac.jp](mailto:snparray@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](mailto:f-nagami@med.tohoku.ac.jp)



ジャポニカレイ®のロゴ