

2017年7月18日

国立大学法人東北大学東北メディカル・メガバンク機構
学校法人岩手医科大学いわて東北メディカル・メガバンク機構
国立研究開発法人日本医療研究開発機構

日本人 3,554 人分の全ゲノムリファレンスパネルを作成

- 日本人を対象とするゲノム医療に大きく貢献 -

【発表のポイント】

- ◆ 東北大学東北メディカル・メガバンク機構（ToMMO）と岩手医科大学いわて東北メディカル・メガバンク機構（IMM）は、3,554 人の全ゲノムリファレンスパネル（3.5KJPN）を作成し、約 3,710 万個の一塩基変異（SNVs）^{*1} を収載することに成功しました。同 SNVs の 72%以上にあたる約 2,690 万個が、世界各地の SNVs を登録する国際データベースには存在しません。集団（民族集団など）が保有する SNVs の多くは集団ごとに特徴的なことが知られていることから、3.5KJPN には日本人に特徴的な SNVs が多数収載されていることが明らかになりました。今後、3.5KJPN は日本人を対象とするゲノム医療に大きく貢献することが期待されます。
- ◆ 3.5KJPN の約 32%は宮城県と岩手県以外の検体から構成されています。詳細に検討すると、日本列島内の地域集団の微細な違いは確認されるものの、他のアジア集団のゲノム情報とは、大きく、かつ明確に異なる日本列島出身者としてのまとまりが検出されました。これらのことから、これまで ToMMO が開発してきたリファレンスパネル（1KJPN、2KJPN）が日本人の特徴を幅広く反映するものであることが実証されました。合わせて、3.5KJPN は日本全国の地域集団の特徴を更に詳細に反映していることも明らかになりました。
- ◆ 3.5KJPN は、日本人の持つ 0.03%以上の SNVs をカバーするものと考えられます。ToMMO と IMM は本パネルに含まれるすべての SNVs の位置情報、アレル頻度^{*2} 情報及びアレル数情報を近日中に公開します。これらの情報により、現在日本医療研究開発機構（AMED）が実施中の、未診断疾患イニシアチブ（IRUD）事業における疾患候補遺伝子の絞り込み性能のさらなる向上が見込まれます。また、疾患への罹患リスク推定などの目的で利用されるマイクロアレイ解析での全ゲノム復元性能を向上させることも可能になります。このように 3.5KJPN 全ゲノムリファレンスパネルは本邦にお

けるゲノム医療の研究基盤として大いに活用されるものと期待されます。

【概要】

ToMMo と IMM は、東北メディカル・メガバンク計画が宮城県と岩手県で実施するコホート調査*³への協力者 3,344 人、および国立病院機構長崎医療センターにおける協力者 181 人、ながはま 0 次予防コホート事業*⁴への協力者 29 人（日本人一般住民合計 3,554 人分）の全ゲノム塩基配列を解析し、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（3.5KJPN）を作成することに成功しました。

3,554 人分の DNA 情報を次世代シーケンサーで読み取り、のべ約 329 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*⁵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約 3,710 万箇所 of SNVs を同パネルに収載しました。同 SNVs の 72% 以上にあたる約 2,690 万箇所は世界各地の SNVs を登録する国際データベースに報告されていない新規の SNVs でした。3.5KJPN は、日本人の持つ 0.03% 以上の SNVs をカバーするものと考えられます。

集団（民族集団など）が保有する SNVs の多くは集団ごとに特徴的なことが知られていることから、3.5KJPN は日本人に特徴的な SNVs が多数収載されていることが明らかになりました。今後、3.5KJPN は日本人を対象とするゲノム医療へ大きく貢献することが期待されます。

3.5KJPN の約 32% は宮城県と岩手県以外の検体から構成されており、詳細に検討すると日本列島内の地域集団の微細な違いは確認されるものの、他のアジア人の集団のゲノム情報とは、大きく、かつ明確に異なる日本列島出身者としてのまとまりが検出されました。これらのことから、これまで ToMMo が開発してきたリファレンスパネル（1KJPN、2KJPN）が日本人の特徴を反映していることも明らかになりました。合わせて、3.5KJPN は日本全国の地域集団の特徴を更に幅広く反映しています。

本成果について、平成 27 年度より、本計画の研究支援機関である AMED のデータシェアリング方針のもとに、当計画における解析結果は速やかに研究者コミュニティに共有することとしており、本リファレンスパネルによる SNVs の位置情報、アレル頻度情報及びアレル数情報は、ToMMo の専用ウェブサイト iJGVD（integrative Japanese Genome Variation Database; <http://ijgvd.megabank.tohoku.ac.jp>）で間もなく公開されます。

東北メディカル・メガバンク計画は、東日本大震災の被災地等で健康調査を行い、被災地の健康状態の改善と遺伝要因・環境要因を考慮した次世代型医療・予防の確立を目指したもので、事業の実施は、ToMMo と IMM とが連携して行っています。

【背景】

ToMMo と IMM は、2013 年から宮城県・岩手県の地域住民 15 万人規模のコホート調査を実施しています。ToMMo は、コホート調査に参加された宮城県住民の 1,070 人分の全ゲノム解読が完了したことを 2013 年 11 月に発表しました。2014 年 8 月には、1,070 人分の全ゲノム配列情報に基づく「全ゲノムリファレンスパネル」(1KJPN) のアレル頻度 5%以上の SNP 頻度情報をウェブサイト上で公開するとともに、翌年 12 月には、公開範囲を全ての SNVs 頻度・位置情報に拡充しました。さらに、2016 年 6 月には 2,049 人分の全ゲノム配列情報に基づく「全ゲノムリファレンスパネル」(2KJPN) を構築し、直ちに公開しました。

【詳細】

今回、次世代シーケンサーによる高精度なゲノム解析により、一般集団の日本人 3,554 人分からなる全ゲノムリファレンスパネル (3.5KJPN) を構築しました。

3.5KJPN 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) の広範囲にわたる約 3,710 万箇所の SNVs を含み (図 1)、そのうち 72%以上が新規の SNVs (表 1) でした (うち約 1,830 万箇所は 2KJPN 発表時に報告され、約 860 万箇所が今回新規となる) (備考 1)。

日本人集団の 1,000 人以上からなる公開されている全ゲノム情報は、2016 年 6 月に ToMMo が公開をおこなった 2KJPN 以外にはありません。

また、全ゲノム断片配列情報は、国際 1,000 人ゲノム計画 (1KGP) では、一人当たり 4~6 回分の解読量であるのに対し、3.5KJPN は、一人当たり最低 25 回分以上の解読量となっています。その結果、1KGP はアレル頻度 1%以上の限定された SNVs の同定が主な目標となっていました。3.5KJPN では、均質な日本人の集団に焦点をあて、3,554 人という大規模な日本人の全ゲノム情報を詳細に解析することにより、アレル頻度が極めて低い SNVs の同定も成功しました。

3.5KJPN は、検体が日本全国をカバーするように設計されています。そのため、東北メディカル・メガバンク計画がそもそも対象としてきた宮城・岩手両県からの協力者については、調査票の情報から特に母親の出身地 (備考 2) が両県以外の方々 941 人を含めています (2KJPN では 326 人でしたが、さらに東日本 (その他) 83 名、中部 242 名、西日本 290 名の検体を追加しています)。また、長崎医療センターにおける協力者 181 人と、ながはま 0 次予防コホート事業の協力者 29 人を合わせて 1,151 人の宮城・岩手両県以外の出身者の情報を含めることにより、3.5KJPN の約 32%は、宮城県と岩手県以外の検体から構成されています (図 2 参照)。これまで ToMMo が作製してきた全ゲノムリファレンスパネル (1KJPN と 2KJPN) よりも、日本人の集団を幅広く代表し

た全ゲノムリファレンスパネルを収集できているといえます。

3.5KJPNに含まれる個人毎の遺伝型を詳細に検討すると、既に報告されているとおり*6、東日本、中部、西日本の出身地による微細な分布の違いが確認されました（図4参照）。しかし、3.5KJPNを国際1,000人ゲノム計画で取得されたアジアの集団（中国3地域、ベトナム）と比較したところ、日本列島出身者の集団は大きく、かつ明確に異なる1つのクラスタを形成しました（図3参照）。

これまでToMMoが開発してきたリファレンスパネル（1KJPN、2KJPN）も、3.5KJPNと重複するクラスタを形成していることから、1KJPN、2KJPNについても日本人の特徴を幅広く反映するものであることが実証されました。合わせて、3.5KJPNは日本全国の地域集団の特徴を更に詳細に反映していることも明らかになりました。

【今後の展開】

本リファレンスパネルは、日本人を対象としたゲノムに関連するさまざまな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
- （例4）日本人集団およびアジアを中心とした他集団との多様性に関する研究による地域ごとの疾患の罹患率等との関連性の探索

近年、希少疾患、難病などの原因特定のため、患者さんのゲノム情報を次世代シーケンサーで解析する試験的な取り組みが始まっています。この解析では、一般集団のリファレンスパネルが、患者さんのゲノムに含まれる疾患と関わりのない塩基変異を除外する上で、重要な役割を果たします。具体的には、現在AMEDが実施中の、日本全国の診断がつかずに悩んでいる患者さん（未診断疾患患者）に対して遺伝学的解析結果を含めた総合的診断等を行う、未診断疾患イニシアチブ（IRUD）事業において、疾患候補遺伝子のさらなる絞り込みが期待されます。

また、疾患への罹患リスク推定などの目的で利用されるマイクロアレイ解析での全ゲノム復元性能を向上させることも可能になります。このように3.5KJPN全ゲノムリファレンスパネルは本邦におけるゲノム医療の研究基盤として大いに活用されるものと期待されます。

今後は、東北メディカル・メガバンク計画の第2段階の目標のひとつである8,000人規模の全ゲノムリファレンスパネルの構築に向け、全ゲノム解析を行

う人数をさらに増やすだけではなく、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た、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を進めていきます。また、東北メディカル・メガバンク計画における他地域出身のコホート参加者をさらに取り入れることや、日本の他地域のコホート事業などとも連携を進めることで、より日本人としての網羅性を高めていく予定です。

さらに、従来より使用してきた国際ヒトゲノム参照配列*7ではなく、2017年6月6日に ToMMo がリリースを行った最新版の日本人基準ゲノム配列*8 **Japanese Reference Genome version 2 (JRGv2)** などを活用して全ゲノムリファレンスパネルを構築することで、一塩基変異および構造多型の同定精度を向上させていきます。全国の研究者のゲノム研究を加速するとともに、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。

備考 1：世界各地の SNVs を登録する国際 SNP データベース version 138 (dbSNP138) を基準としている。

備考 2：東北メディカル・メガバンク計画の参加者は、コホート調査の調査票の回答情報から母親の出生地を「出身地」としている。

【参考図表】

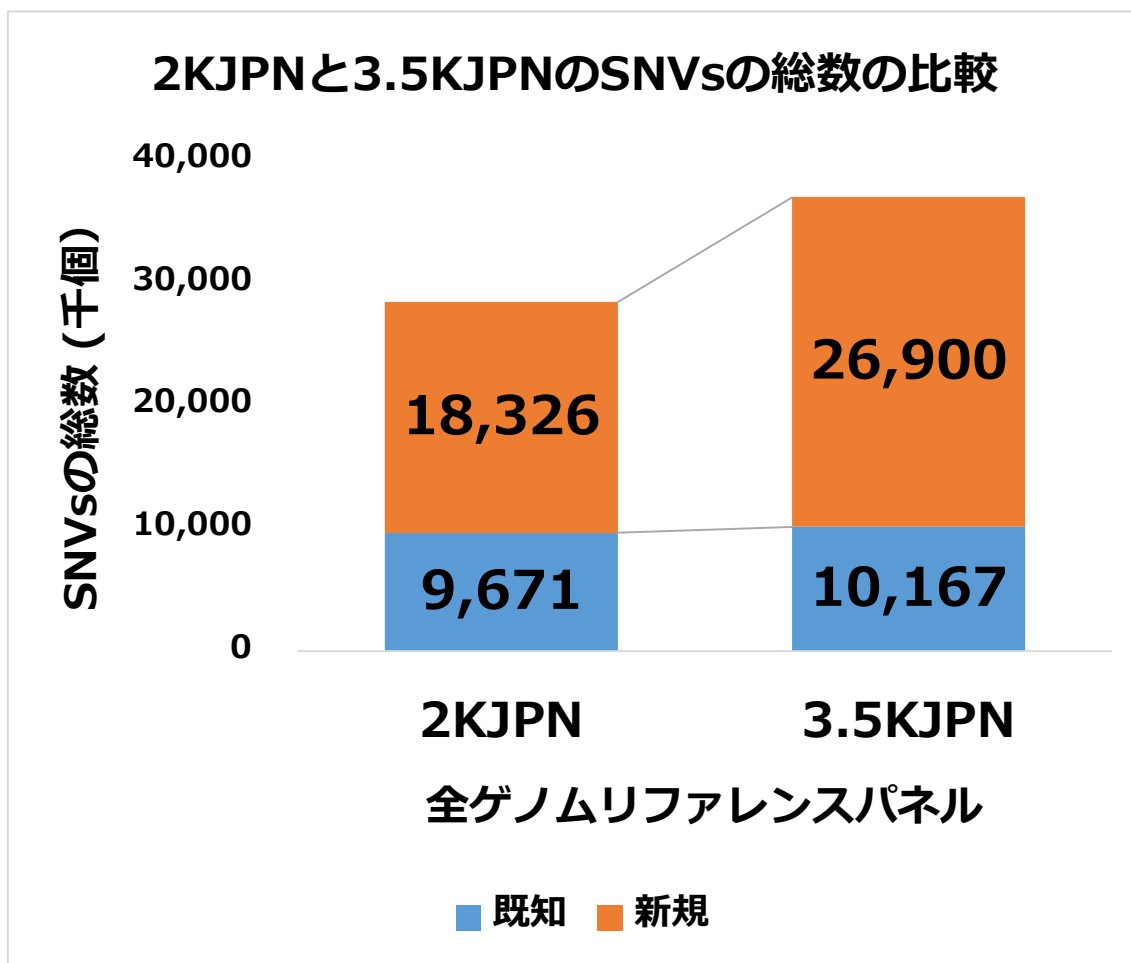


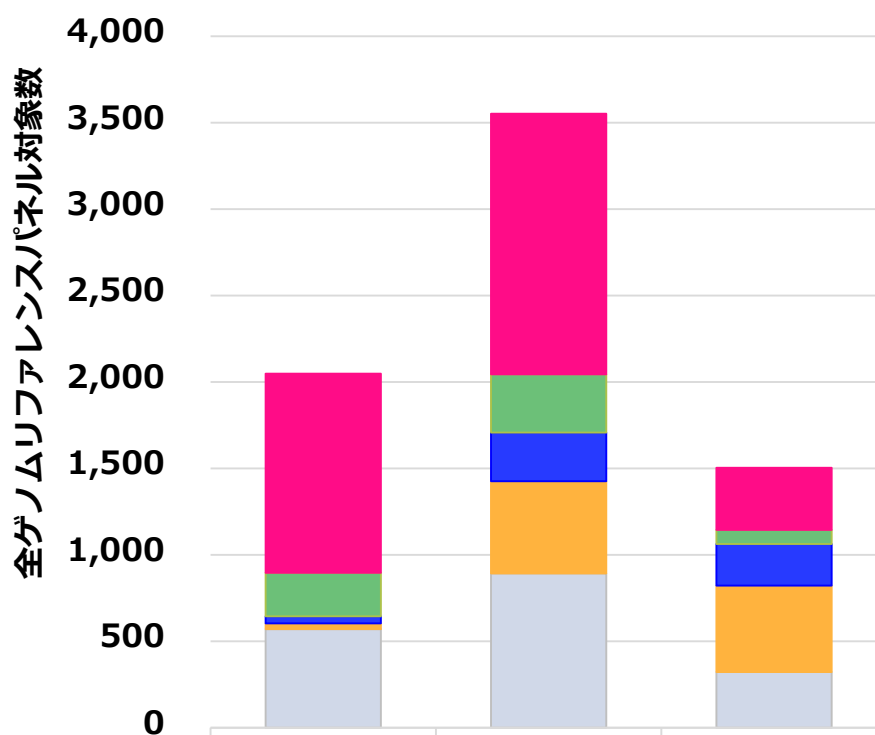
図1 2KJPNと3.5KJPNのSNVsの総数の比較

		2KJPN	3.5KJPN
対象人数		2,049 人	3,554 人
総塩基数		約 192 兆塩基	約 329 兆塩基
平均深度		32.3 ×	31.9 ×
SNVs 総数	新規	約 1,833 万個 (18,326,183 個)	約 2,690 万個 (26,900,894 個)
	既知 ※	約 967 万個 (9,671,410 個)	約 1,017 万個 (10,166,821 個)
	合計	約 2,800 万個 (27,997,593 個)	約 3,707 万個 (37,067,715 個)

表1 2KJPNと3.5KJPNのシーケンスデータおよびSNVsの総数の比較

※ここでの既知とは国際SNPデータベース version 138 (dbSNP138) に登録されているSNVsをさす。

3.5KJPNの他地域の拡充内訳（概要）



	2KJPN	3.5KJPN	増加人数
■ 東日本（宮城、岩手）	1,151	1,509	358
■ 東日本（その他）	253	336	83
■ 中部	41	283	242
■ 西日本	32	532	500
■ その他（未回答など）	572	894	322

図2 3.5KJPNにおける他地域（東日本（宮城、岩手以外）、西日本、中部）の検体の拡充の内訳（概要）

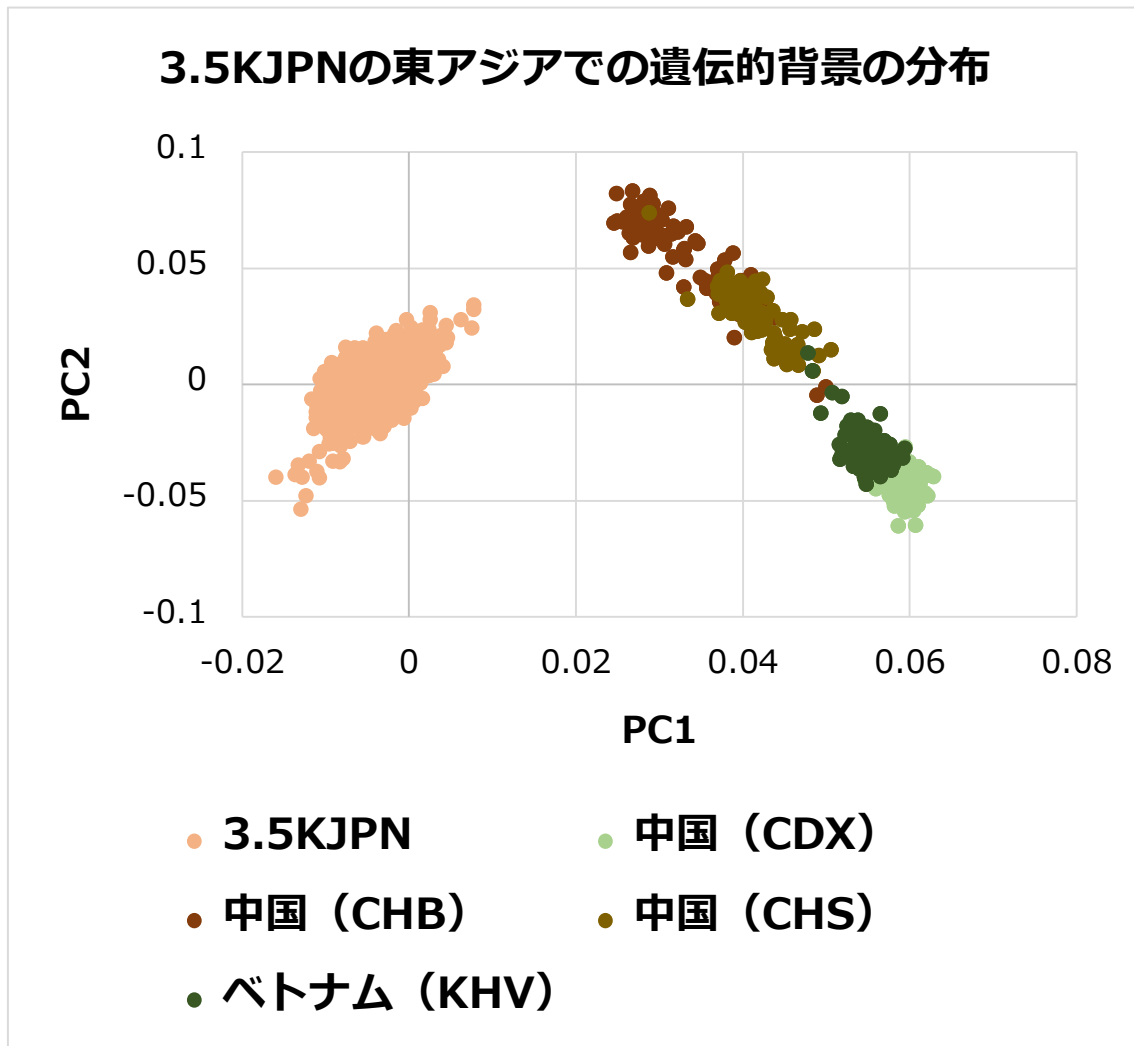


図3 3.5KJPNに含まれる検体のアジアにおける遺伝的背景の分布(主成分分析の結果による)

3.5KJPN : 3.5KJPN パネルに含まれるサンプル

中国 (CDX) : 国際 1000 人ゲノムに含まれる

中国 (Chinese Dai in Xishuangbanna) サンプル

中国 (CHB) : 国際 1000 人ゲノムに含まれる

中国 (Han Chinese in Beijing) サンプル

中国 (CHS) : 国際 1000 人ゲノムに含まれる

中国 (Southern Han Chinese) サンプル

ベトナム (KHV) : 国際 1000 人ゲノムに含まれるベトナム

(Kinh in Ho Chi Minh City, Vietnam) サンプル

※なお、地域情報が不明の検体については、プロットをおこなっていない。

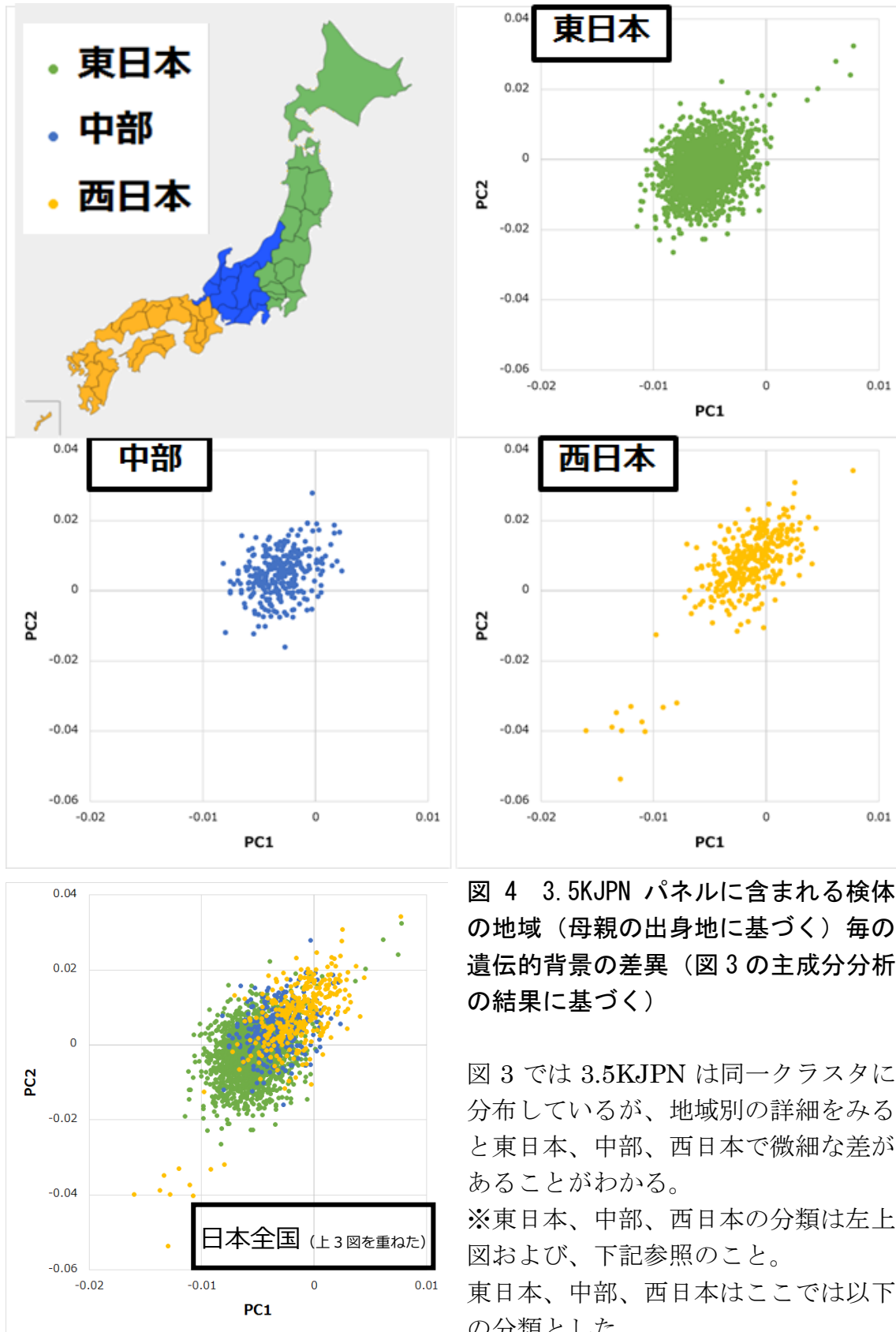


図 4 3.5KJPN パネルに含まれる検体の地域の（母親の出身地に基づく）毎の遺伝的背景の差異（図 3 の主成分分析の結果に基づく）

図 3 では 3.5KJPN は同一クラスターに分布しているが、地域別の詳細をみると東日本、中部、西日本で微細な差があることがわかる。

※東日本、中部、西日本の分類は左上図および、下記参照のこと。

東日本、中部、西日本はここでは以下の分類とした。

東日本（北海道、青森県、岩手県、秋田県、宮城県、山形県、福島県、茨城県、

栃木県、群馬県、埼玉県、千葉県、東京都、神奈川県)

中部 (山梨県、長野県、新潟県、富山県、石川県、福井県、静岡県、愛知県、岐阜県)

西日本 (三重県、滋賀県、京都府、大阪府、兵庫県、奈良県、和歌山県、鳥取県、島根県、岡山県、広島県、山口県、香川県、愛媛県、徳島県、高知県、福岡県、佐賀県、長崎県、熊本県、大分県、宮崎県、鹿児島県、沖縄県)

※集計において東北メディカル・メガバンク計画の参加者は、母親の出生地を地域として選択した。長崎医療センターの検体、ながはま0次予防コホート事業の検体は、西日本を地域とした。

【公開予定】

3,554人分の日本人の全ゲノムリファレンスパネル (3.5KJPN)

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



【用語解説】

*1 一塩基変異 (Single Nucleotide Variations; SNVs): ゲノムを比較すると1塩基だけ塩基配列が異なっている箇所が見つかることがあるが、このような箇所は一塩基変異と呼ばれる。なお、一定以上の頻度で確認され、世代を超えて伝わる (遺伝する) 一塩基変異のことを、特に一塩基多型 (Single Nucleotide Polymorphism; SNP) と呼ぶ場合がある。

*2 アレル頻度: ある集団における各々の対立遺伝子の存在の相対的頻度のこと。ある集団の中で特定のSNVが存在する割合を示すこともある。今回は抽出された日本人3,554人中がもつ7,108本の常染色体中での割合となる。

*3 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、生活習慣など環境要因と疾病の関係を解明するための調査。

*4 ながはま0次予防コホート事業: 京都大学が滋賀県長浜市と共に進めるコホート調査。長浜市民を対象に1万人規模で2007年から進められている。

*5 スーパーコンピュータ: 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、その機能を果た

すため、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を運用している。バイオバンクとは、生体試料を収集・保管し、研究利用のために提供を行う機関。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。

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

*6 米国人類遺伝学会機関紙 (American Journal of Human Genetics) 2008年8巻 (4号) p.445-456に掲載された研究。筆頭著者は、山口由美助教 (現在ToMMo所属)。

*7 国際ヒトゲノム参照配列：国際的な学術組織 Genome Reference Consortium が継続的に改訂を行っているヒトゲノムの全染色体の塩基配列。同配列は主に欧米の複数のヒトゲノムを読むことで構築されている。事実上、ヒトゲノムのデファクトスタンダードの塩基配列として全世界のヒトゲノム研究に利用されている。2017年7月現在、最新の国際ヒトゲノム参照配列はGRCh38である。

*8 日本人基準ゲノム配列：より日本人のゲノム情報を反映した参照配列。東北メディカル・メガバンク機構において、2016年に日本人基準ゲノムJRGv1 (Japanese Reference Genome version 1) を公開するとともに、2017年6月6日にJRGv2 (Japanese Reference Genome version 2) を公開した。

参考：プレスリリース「日本人の基準ゲノム配列(JRG)を決定 - 長鎖読みとり型次世代シーケンサーを用いて日本人のもつゲノム構造を解明 - 」

URL: <http://www.megabank.tohoku.ac.jp/news/15143>

参考：プレスリリース「日本人基準ゲノム配列、精度が向上した新版(JRGv2)を公開 一分子長鎖型シーケンサーを用いた複数名の高深度ゲノム情報を元に日本人に特徴的なゲノム情報を 10 倍に拡充」

URL: <http://www.megabank.tohoku.ac.jp/news/21425>

【お問い合わせ先】

(研究に関すること)

東北大学東北メディカル・メガバンク機構

ゲノム情報解析室 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：nagasaki@megabank.tohoku.ac.jp

東北大学東北メディカル・メガバンク機構

シーケンス解析室 室長

准教授 勝岡 史城 (かつおか ふみき)

電話番号：022-273-6214

Eメール：kfumiki@megabank.tohoku.ac.jp

(報道に関すること)

東北大学東北メディカル・メガバンク機構

広報戦略室 室長

長神 風二 (ながみ ふうじ)

電話番号：022-717-7908

ファクス：022-717-7923

Eメール：f-nagami@med.tohoku.ac.jp

(AMED 事業に関すること)

日本医療研究開発機構 (AMED)

基盤研究事業部 バイオバンク課

電話番号：03-6870-2228

Eメール：tohoku-mm@amed.go.jp