



分野:生命科学・医学系 キーワード:ゲノムワイド関連解析、健康・医療データ、ゲノム個別化医療、特異値分解

## 国際バイオバンク連携によるヒト疾患リスク遺伝子アトラスを構築 日本主導型の国際メタアナリシスでゲノム情報に基づく疾患の精密分類を提案

### 【研究成果のポイント】

- ◆ バイオバンク・ジャパン 18万人のゲノムデータを基に過去最大 220 の健康・医療データ(多因子疾患・希少疾患・バイオマーカー<sup>※1</sup>・服薬データ)に対する網羅的ゲノムワイド関連解析(GWAS)<sup>※2</sup>を実施。
- ◆ 国際バイオバンク連携によりイギリス・フィンランドのバイオバンク<sup>※3</sup>と計 63 万人のメタアナリシスを実施し、5,000 以上の新規遺伝的リスク関連領域を発見。GWAS 結果を公開するデータベース PheWeb.jp を構築し、データシェアリングを推進。
- ◆ GWAS 要約統計量<sup>※4</sup>の特異値分解<sup>※5</sup>と、東北大学東北メディカル・メガバンク機構のメタボローム<sup>※6</sup>データ等、他のオミクスデータへのプロジェクション解析<sup>※7</sup>を実施することで、ゲノム情報を基にした従来の疾患分類の精密・層別化方法を提唱。

### ❖ 概要

全世界のヒトゲノムデータの蓄積に伴ってこの 20 年ほどで実施されてきた GWAS により、ヒトの遺伝的変異と将来の疾患発症リスクとの関係が網羅的に判明しつつあります。得られたゲノムデータを基に、長い医学の歴史から構築されてきた多彩なヒトの病気の分類について、客観的に振り返って見たらどうなるでしょうか？ この疑問に適切に答えるためには、従来のゲノム研究の課題である、①:ゲノム研究データが欧米人集団に偏重して構築されていること、②:研究対象となる疾患の網羅性が低かったこと、③:得られた大量の GWAS 結果を医学的・生物学的に解釈する方法論が確立されていないこと、の 3 点を解決する必要がありました。

大阪大学大学院医学系研究科の坂上沙央里助教(研究当時、現ハーバード大学医学部博士研究員)、金井仁弘特別研究学生(ハーバード大学医学部 博士課程)、岡田随象教授(遺伝統計学 / 理化学研究所生命医科学研究センター自己免疫疾患研究チーム 客員主管研究員)、東京大学大学院新領域創成科学研究科 松田浩一教授らの研究グループは、国際バイオバンク連携を通じて、バイオバンク・ジャパン(日本)・UK バイオバンク(英国)・FinnGen(フィンランド)の計 63 万人のゲノムデータと健康・医療データの網羅的な解析を実施しました。多因子疾患・希少疾患・バイオマーカー・服薬データまでを網羅する過去最大 220 のヒト形質のゲノムワイド関連解析(GWAS)により、ヒト疾患に関わる 5,000 以上の新規遺伝的リスク関連領域が発見されました。研究グループは、各疾患の遺伝的リスク構造が遺伝的集団<sup>※8</sup>を超え共有されていることを示すとともに、得られたゲノム解析アトラスを医療に役立てる方法として、GWAS 要約統計量の特異値分解し遺伝学を基にした疾患群の再分類を試みました(図1)。東北大学東北メディカル・メガバンク機構が構築したメタボロームデータ等、他のオミクスデータへのプロジェクション解析、エピゲノム情報やパスウェイ等との多分野融合的な解析を実施し、GWAS 要約統計量行列を数理的に分解する工程を経て、より複雑な疾患(例:心筋梗塞)のリスクを単純な形質の足し合わせとして解釈すること(例:コレステロールと血圧)や、ゲノム変異を基にした疾患の精密分類・層別化方法の提唱(例:I型アレルギーとIV型アレルギー疾患の分類)に成功しました。さらに、本研究結果を広く利用促進することを目的にデータシェアリングサイト PheWeb.jp (<https://pheweb.jp/>)を構築しデータを無償・非制限公開し、Polygenic risk score(PRS)<sup>※9</sup>をはじめとするゲノム個別化医療の社会実装基盤として広く利用されることが期待されます。

本研究成果は、米国科学誌「Nature Genetics」に、10月1日(金)午前0時(日本時間)に公開されました。

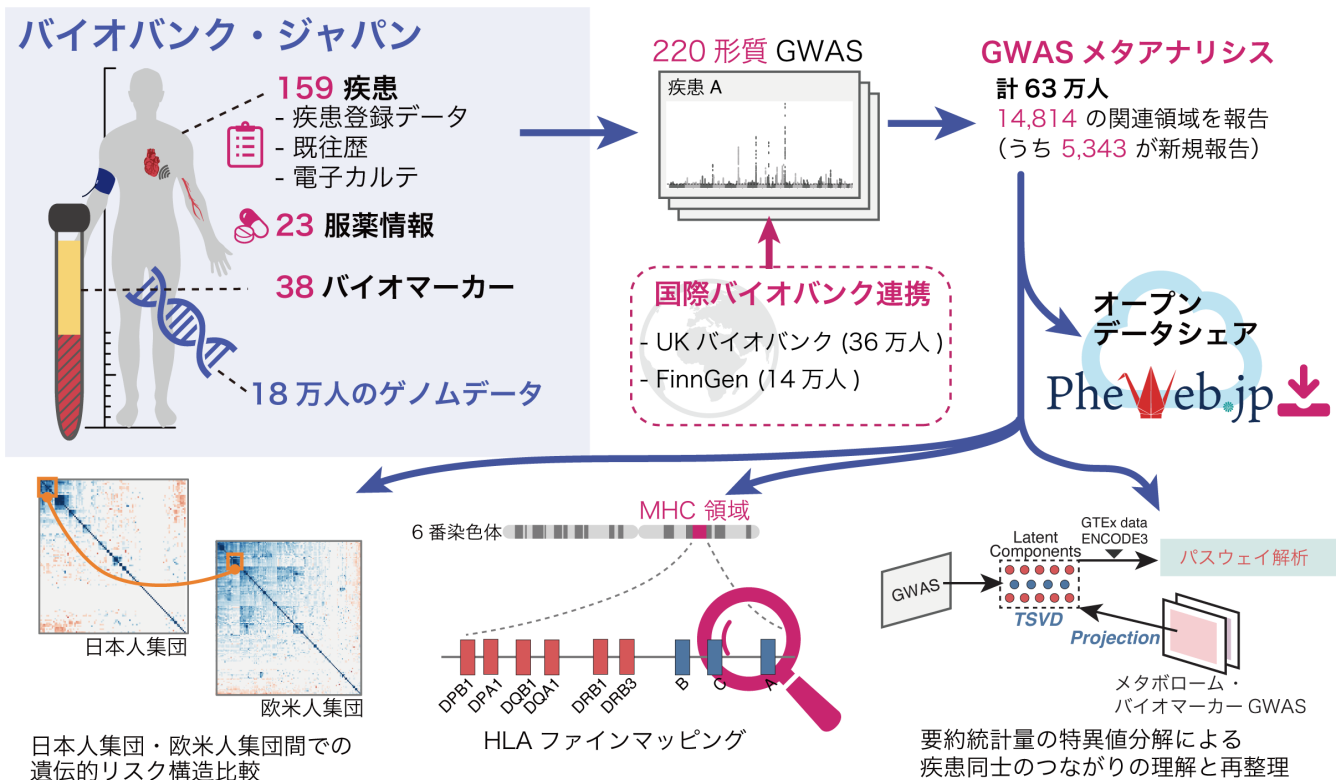


図 1. 日本主導の国際バイオバンク連携により、220 のヒト形質の GWAS と横断的メタアナリシスを実施して結果を公開するとともに、医療に役立つ方法論を提唱

### ❖ 研究の背景

古来より、ヒト疾患の診断分類は、患者の多彩な症候を観察し、似た症状を分類整理し、それに基づき最適な治療を選ぶために経験的に構築されてきました。先端的な検査方法や画像診断が発達した現在においても、従来の診断分類は徐々に形を変えながら踏襲されていますが、ゲノムデータを基にこの分類を客観的に振り返って見たらどうなるでしょうか？ゲノムに基づく疾患分類によって従来の分類の正しさを確かめることができるかもしれませんし、あるいは似通った疾患群によりよい精密な分類を与えることが可能になるかもしれません。数十万人規模のゲノムデータと多彩な臨床データを蓄えたバイオバンクが各国に存在する現在、国際共同研究を通じてこの疑問に初めて答えを出すことができるようになってきました。

### ❖ 本研究の成果

研究グループは、国際バイオバンク連携を通じてバイオバンク・ジャパン(日本)・UK バイオバンク(英国)・FinnGen(フィンランド)の計 63 万人のゲノムデータと、多因子疾患・希少疾患・バイオマーカー・服薬データまでを網羅するヘルスデータをテキストマイニングを用いて網羅的に解析し、過去最大 220 のヒト形質のゲノムワイド関連解析(GWAS)を実施しました。本研究では、従来のゲノム研究の課題であった欧米人集団へのゲノムデータ偏重に対して、アジア人集団として最大規模を誇るバイオバンク・ジャパンのデータを基底に、UK バイオバンク・FinnGen と協調した幅広い形質を対象としたことに意義があります。その結果、集団特異的変異を含む 14,814 の疾患感受性領域の発見に至り、このうち 5,343 の領域が新規報告となりました。今回初めて網羅的に多数の疾患の遺伝的構造を日本人集団と欧米人集団とで比較したところ、各疾患の遺伝的リスク構造は遺伝的集団を超えて共通しており、ゲノムによって定義された類似疾患グループ(例: 心筋梗塞・不安定狭心症などの心血管疾患)も遺伝的集団間で共有されていることがわかりました。これは、地域や環境要因を超えて歴史的に形成されてきた診断分類の妥当性を、大規模ゲノム解析によって再確認できた知見と言えるかもしれません。

## Press Release

現在のゲノム研究の最大の課題は、GWAS 解析から得られた大量の要約統計量が複雑すぎて、それらを生物学的に解釈し、直接的に医療に役立てるのが困難であることでした。そこで研究グループは、全ゲノムに渡る遺伝的変異ごとの疾患リスク統計量行列を、特異値分解(SVD)という数理統計学手法を用いて行列分解しました。さらに分解の結果得られたゲノムに基づく疾患リスク要素を、バイオマーカーGWAS や東北大学東北メディカル・メガバンク機構で実施したメタボーム GWAS 結果へ投影(プロジェクション)しました(図 2)。これにより、心筋梗塞のリスク変異がコレステロールと血圧を説明するリスク変異の組み合わせで説明されるなど、複雑な疾患のリスクの背景をより単純な形質のリスク要素の和で説明したり、それらの要素を説明する生物学的パスウェイと重要な細胞組織を特定したりすることが可能になりました。この分解要素を基にして、似通った複数のアレルギー疾患を、喘息やアレルギー性鼻炎など古典的に I 型アレルギーと分類されてきた疾患群と、金属アレルギーや接触性皮膚炎など古典的に IV 型アレルギーと分類されてきた疾患群との 2 群に分類することに成功しました。この結果は、遺伝学を基にした疾患診断の再分類の成功例として意義深いものと考えています。

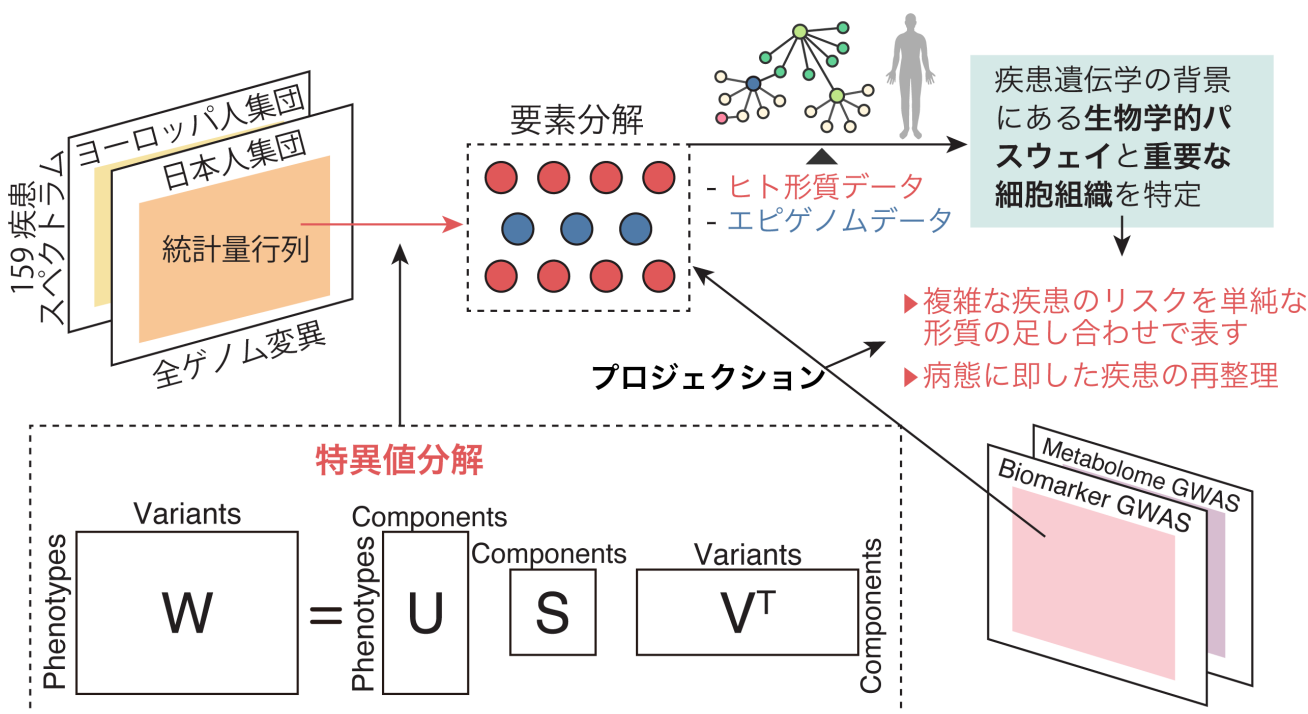


図 2. 要約統計量行列に特異値分解を行い、多彩な疾患群のオミクス情報を基に再整理し、異なる疾患に共通する要素を抽出して、その背景にある生物学的パスウェイを特定することに成功。

### ❖ 本研究成果が社会に与える影響（本研究成果の意義）

本研究の成果はわが国発のゲノム解析データ基盤として今後のさらなる国際連携のために重要であり、GWAS 解析結果をデータシェアリングサイト PheWeb.jp (<https://pheweb.jp/>)にて無償・非制限公開を実施しています。このデータを基にした新規ゲノム解析手法の開発や、Polygenic risk score (PRS)をはじめとするゲノム個別化医療の社会実装基盤として広く利用されることが期待されます。

### ❖ 特記事項

本研究成果は、2021年10月1日(金)午前0時(日本時間)[9月30日(木)午後4時(英国時間)]に米国科学誌「Nature Genetics」(オンライン)に掲載されました。

【タイトル】“A cross-population atlas of genetic associations for 220 human phenotypes”

【著者名】Saori Sakaue<sup>1-5, 46, \*</sup>, Masahiro Kanai<sup>1, 5-9, 46</sup>, Yosuke Tanigawa<sup>10</sup>, Juha Karjalainen<sup>5-7, 9</sup>, Mitja Kurki<sup>5-7, 9</sup>, Seizo Koshiba<sup>11, 12</sup>, Akira Narita<sup>11</sup>, Takahiro Konuma<sup>1</sup>, Kenichi Yamamoto<sup>1, 13, 14</sup>, Masato Akiyama<sup>2, 15</sup>,



国立大学法人 大阪大学

〒565-0871 大阪府吹田市山田丘 1-1

TEL: 06-6877-5111 (代)

www.osaka-u.ac.jp

## Press Release

Kazuyoshi Ishigaki<sup>2-5</sup>, Akari Suzuki<sup>16</sup>, Ken Suzuki<sup>1</sup>, Wataru Obara<sup>17</sup>, Ken Yamaji<sup>18</sup>, Kazuhisa Takahashi<sup>19</sup>, Satoshi Asai<sup>20,21</sup>, Yasuo Takahashi<sup>21</sup>, Takao Suzuki<sup>22</sup>, Nobuaki Shinozaki<sup>22</sup>, Hiroki Yamaguchi<sup>23</sup>, Shiro Minami<sup>24</sup>, Shigeo Murayama<sup>25</sup>, Kozo Yoshimori<sup>26</sup>, Satoshi Nagayama<sup>27</sup>, Daisuke Obata<sup>28</sup>, Masahiko Higashiyama<sup>29</sup>, Akihide Masumoto<sup>30</sup>, Yukihiro Koretsune<sup>31</sup>, FinnGen, Kaoru Ito<sup>32</sup>, Chikashi Terao<sup>2</sup>, Toshimasa Yamauchi<sup>33</sup>, Issei Komuro<sup>34</sup>, Takashi Kadowaki<sup>33,35</sup>, Gen Tamiya<sup>11,12,36,37</sup>, Masayuki Yamamoto<sup>11,12,36</sup>, Yusuke Nakamura<sup>38,39</sup>, Michiaki Kubo<sup>40</sup>, Yoshinori Murakami<sup>41</sup>, Kazuhiko Yamamoto<sup>16</sup>, Yoichiro Kamatani<sup>2,42</sup>, Aarno Palotie<sup>5,9,43</sup>, Manuel A. Rivas<sup>10</sup>, Mark J. Daly<sup>5-7,9</sup>, Koichi Matsuda<sup>44, \*</sup>, Yukinori Okada<sup>1,2,14,42,45, \*</sup>

(\* 責任著者)

### 【所属】

1. 大阪大学大学院医学系研究科 遺伝統計学
2. 理化学研究所生命医科学研究センター ゲノム解析応用研究チーム
3. ハーバード大学医学部 Center for Data Sciences
4. プリガム&ウィメンズ病院 Divisions of Genetics and Rheumatology
5. ブロード研究所 Program in Medical and Population Genetics
6. マサチューセッツ総合病院 Analytic and Translational Genetics Unit
7. ブロード研究所 Stanley Center for Psychiatric Research
8. ハーバード大学医学部 Department of Biomedical Informatics
9. ヘルシンキ 大学 Institute for Molecular Medicine Finland (FIMM)
10. スタンフォード大学 Department of Biomedical Data Science
11. 東北大学 東北メディカル・メガバンク機構
12. 東北大学 未来型医療創成センター(INGEM)
13. 大阪大学大学院医学系研究科 小児科学
14. 大阪大学免疫学フロンティア研究センター 免疫統計学
15. 九州大学医学部 眼科
16. 理化学研究所生命医科学研究センター 自己免疫疾患研究チーム
17. 岩手医科大学 泌尿器科学講座
18. 順天堂大学大学院 膠原病・リウマチ内科学
19. 順天堂大学大学院 呼吸器内科学
20. 日本大学医学部 生体機能医学系薬理学分野
21. 日本大学医学部 臨床試験研究センター
22. 徳洲会グループ
23. 日本医科大学 血液内科
24. 日本医科大学 生体機能制御学分野
25. 東京都健康長寿医療センター
26. 結核予防会 複十字病院
27. がん研究会 がん研究所
28. 滋賀医科大学 医学部附属病院臨床研究開発センター
29. 大阪国際がんセンター 呼吸器外科
30. 麻生飯塚病院
31. 国立病院機構大阪医療センター
32. 理化学研究所生命医科学研究センター 循環器ゲノミクス・インフォマティクス研究チーム



## Press Release

33. 東京大学医学部附属病院 糖尿病・代謝内科
34. 東京大学医学部附属病院 循環器内科
35. 虎の門病院
36. 東北大学大学院医学系研究科
37. 理化学研究所 革新知能統合研究センター
38. 東京大学医科学研究所 ヒトゲノム解析センター
39. がんプレジジョン医療研究センター
40. 理化学研究所生命医科学研究センター
41. 東京大学医科学研究所 癌・細胞増殖部門 人癌病因遺伝子分野
42. 東京大学大学院新領域創成科学研究科 メディカル情報生命専攻 複雑形質ゲノム解析分野
43. マサチューセッツ総合病院 Psychiatric & Neurodevelopmental Genetics Unit
44. 東京大学大学院新領域創成科学研究科 メディカル情報生命専攻 クリニカルシーケンス分野
45. 大阪大学先導的学際研究機構 生命医科学融合フロンティア研究部門
46. 共同筆頭著者

【DOI 番号】10.1038/s41588-021-00931-x

なお、本研究は、日本医療研究開発機構(AMED) ゲノム医療実現バイオバンク利活用プログラム:B-cure のうち、ゲノム研究バイオバンク(旧:疾患克服に向けたゲノム医療実現プロジェクト(オーダーメイド医療の実現プログラム))、ならびにゲノム医療実現推進プラットフォーム・先端ゲノム研究開発:GRIFIN「遺伝統計学に基づく日本人集団のゲノム個別化医療の実装」の一環として行われ、大阪大学大学院医学系研究科バイオインフォマティクス・イニシアティブの協力を得て行われました。本研究で使用したサンプルは、「オーダーメイド医療の実現プログラム」において収集されたものです。また、本研究で使用した一部のデータは AMED による「東北メディカル・メガバンク計画」において得られました。

### ❖ 用語説明

#### ※1 バイオマーカー

ヒトの疾患病態や生物的特徴の説明に役立つ、測定指標のこと。体重、血圧などの身体測定値や、コレステロールや尿酸などの血液検査値が含まれる。

#### ※2 ゲノムワイド関連解析研究(GWAS: Genome Wide Association Study)

ヒトゲノム配列上に存在する数百～数千万カ所の遺伝的変異(とくに一塩基多型・SNP)とヒトの疾患発症リスクや個人差(形質)との関連を網羅的に検討する、遺伝統計解析手法。

#### ※3 バイオバンク

疾患疫学や病態研究などを目的に、多数のヒトの DNA、血清、尿、組織などの検体を収集、蓄積、管理する施設のこと。近年では国家レベルで数十万人を対象とするバイオバンクが構築され、個人の検体とともに電子カルテ上の臨床情報やその後の予後などの追跡情報も蓄積される例が多い。

#### ※4 要約統計量 (summary statistics)

GWAS の関連解析結果の主となる統計量をまとめた表のことで、一般的には全ゲノムに渡る遺伝的変異の名称や位置情報と、線形回帰やロジスティック回帰の結果の P 値(=統計学的な有意性)・beta 値(=リスクの方向性・大き



## Press Release

さ)・標準誤差、解析サンプル数などの情報がまとめられたもののことを指す。元のゲノムデータや形質データと異なりプライバシーの問題がないため、GWAS 結果を他機関に共有する際の基本的な情報として利用される。

### ※5 特異値分解 (SVD: singular value decomposition)

線形代数において、正方行列に限らない任意の行列を複数の行列の積として分解する手法の一つ。

任意の実行列  $A$  に対して得られる特異値分解  $A=U\Sigma V$  では、 $\Sigma$  は  $A$  の特異値を対角項にもつ矩形行列、 $U$  と  $V$  は直交行列となる。本論文では truncated SVD という応用手法により、指定した階数( $< \text{rank}(A)$ )までの分解を実施し、複雑な遺伝子変異—疾患間の遺伝的相関関係を、得られた成分の組み合わせとして表現した。

### ※6 メタボローム (metabolome)

生体内の低分子化合物質(代謝物: metabolite)の総体(-ome)のこと。質量分析法等により網羅的に代謝物を測定する技術が近年発達し、数千種類におよぶ代謝物の量を測定することができる。

### ※7 プロジェクション解析

疾患形質に対する特異値分解によって得られた遺伝的変異と各要素の関係の大きさと符号を表す重み情報を用いて、新たな GWAS 要約統計量を特異値分解の要素に投影し、それぞれの要素に対してどのくらいの貢献度があるかを定量する解析のこと。

### ※8 遺伝的集団 (population)

英語における population、すなわちある遺伝学的背景を共有した任意の集団(データから科学的に定義された任意の集団)のことで、本研究においては日本語訳の一例として「遺伝的集団」を使用している。より狭義の ancestry(遺伝学的背景を共有した集団かつ、歴史的にも遺伝的祖先を共有していると考えられる集団、日本語訳の例: 遺伝的祖先)や、社会的に構築されたニュアンスを含む race(日本語訳の例: 人種)や ethnicity(日本語訳の例: 民族)とは異なる意味で使用している。

### ※9 Polygenic risk score (PRS)

大規模なゲノムワイド関連解析研究(GWAS)により疾患や形質との関連が示唆された数十～数千の遺伝的変異の重み付きの和を個人ごとに計算したスコア。このスコアは実際の疾患発症リスクと相関することが示されており、集団内でスコアの分布を調べることで、特にその疾患のリスクが高い個人を特定することができる。

#### 【研究者のコメント】<坂上沙央里>

ゲノム解析技術の著しい進歩により世界中で実施されてきた GWAS から、すでに 25 万以上の疾患感受性領域が報告されていますが、得られた結果が複雑すぎて解釈が困難であることがしばしば批判として挙げられてきました。遺伝的背景・診断基準や環境要因が異なる集団においてもゲノムに基づく病気同士の関係が概ね共通していることや、ヒトの病気の複雑な遺伝的リスクを行列分解を用いて単純な形質の遺伝的リスクの和で表すことができたことは、当たり前のように見えて本研究で初めて示されたことであり、結果を見たときの喜びは忘れられません。研究施設も国も超えて、異分野横断的な解析を実施できる環境を作ってくださった共同研究者の方々、また何よりバイオバンク参加者の方々に深くお礼を申し上げます。本プロジェクトでは積極的にデータ公開を実施して世界中の研究者に研究成果を使用していただくことを目標としており、すでに私達の解析結果を利用した論文が多数発表されていることも嬉しく感じています。



## Press Release

### ❖ 本件に関する問い合わせ先

<研究に関すること>

岡田 随象(おかだ ゆきのり)

大阪大学 大学院医学系研究科 遺伝統計学 教授

TEL: 06-6879-3971 FAX: 06-6879-3975

E-mail: yokada[AT]sg.med.osaka-u.ac.jp

坂上 沙央里(さかうえ さおり)

ハーバード大学医学部 博士研究員

E-mail: ssakaue[AT]bwh.harvard.edu

<報道に関すること>

大阪大学大学院医学系研究科 広報室

TEL: 06-6879-3388 FAX: 06-6879-3399

Email: medpr[AT]office.med.osaka-u.ac.jp

東京大学大学院新領域創成科学研究科 広報室

TEL: 04-7136-5450

E-mail: press[AT]k.u-tokyo.ac.jp

理化学研究所 広報室 報道担当

E-mail: ex-press[AT]riken.jp

東北大学東北メディカル・メガバンク機構 広報戦略室

TEL: 022-717-7908 FAX: 022-717-7923

E-mail: pr[AT]megabank.tohoku.ac.jp

東京都健康長寿医療センター 総務係広報担当

E-mail: kouhou[AT]tmghig.jp

<AMED 事業に関すること>

日本医療研究開発機構

ゲノム・データ基盤事業部 ゲノム医療基盤研究開発課

ゲノム医療実現バイオバンク利活用プログラム: B-cure

(ゲノム医療実現推進プラットフォーム・先端ゲノム研究開発: GRIFIN)

TEL: 03-6870-2228 FAX: 03-6870-2243

E-mail: genome-platform[AT]amed.go.jp

※Email は上記アドレス[AT]の部分を@に変えてください。