

# GenomeDataAnalysis5

大阪大学大学院医学系研究科 遺伝統計学  
東京大学大学院医学系研究科 遺伝情報学  
理化学研究所生命医科学研究センター システム遺伝学チーム

<http://www.sg.med.osaka-u.ac.jp/index.html>

## GenomeDataAnalysis5

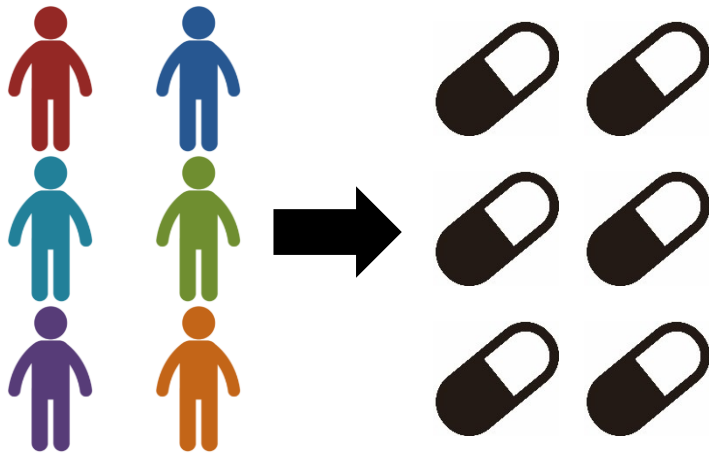
① Polygenic risk score(PRS)とゲノム個別化医療

② PRSice-2を使ったPRS解析

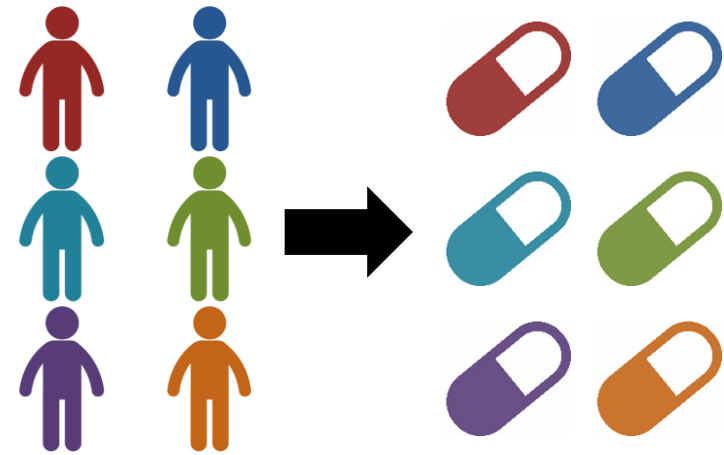
本講義資料は、Windows PC上で  
C:¥SummerSchoolにフォルダを配置すること  
を想定しています。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## これまでの医療

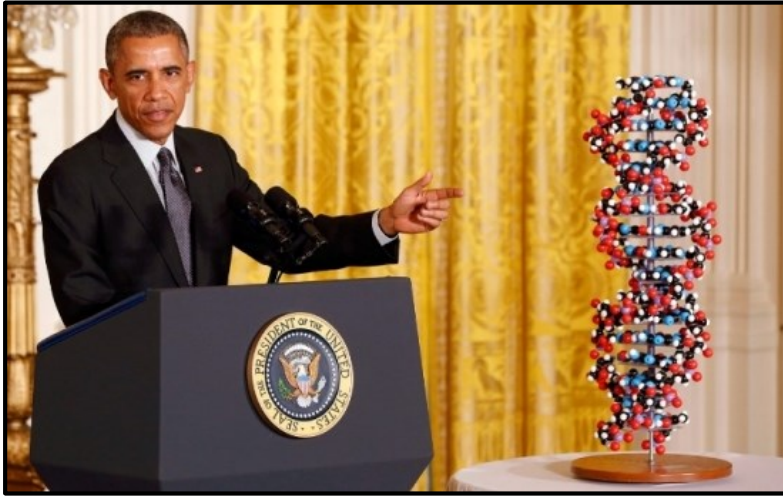


## 個別化医療



- 世の中には沢山の病気がありますが、同じ病気であっても、**個人の体質**や**ライフスタイルに応じて病態が異なる**ことが経験的に知られています。
- 例えば、病気のサブタイプ、薬の効果、副作用の確率などが異なります。
- 全員に画一的な標準治療を行うことも大事ですが、**個人の体質に応じて最適化された治療法の確立**が期待されています。
- それが、**個別化医療**です。

# ① Polygenic risk score (PRS)とゲノム個別化医療



## Precision Medicine Initiative



遺伝子、環境、ライフスタイルに  
関する個人ごとの違いを考慮した  
予防や治療法を確立すること

(<https://www.kantei.go.jp/jp/singi/kenkouiryou/genome/dai1/siryou04.pdf>)

- 2015年のオバマ大統領の一般教書演説で、“Precision Medicine Initiative”という言葉が注目を集めました。
- Precision Medicine Initiativeは、特に**個人のゲノム情報の違い**を活用した**個別化医療**の提言に重きを置いた点が画期的でした。
- **ゲノム個別化医療**の時代の幕開け、という解釈もできます。

# ① Polygenic risk score (PRS) とゲノム個別化医療

## ゲノム個別化医療の例：希少難病疾患



The screenshot shows the website of the Japanese Society of Human Genetics (JSHG). The main navigation bar includes links for 'Home', 'About Us', 'Education', 'Guidelines', 'Seminars', and 'Contact'. The current page is 'Chromosomal Abnormalities' (染色体異常をみつけたら) and displays a table of contents for a PDF document titled 'Chromosomal Abnormalities Table of Contents (8th Edition)'. The table lists 14 items, including 'Normal variants', 'Chromosomal number abnormalities', 'Chromosomal structural abnormalities', 'Microdeletion syndromes', 'Sex chromosome abnormalities', 'R-banding', 'Pre-natal diagnosis', 'Chromosomal instability', 'Leukemia and lymphoma', 'Cytogenetic classification', 'Karyotype analysis', 'Frequency of chromosomal abnormalities', and 'Chromosomal microarray'.

染色体異常をみつけたら 目次 (改訂八版)	
大項目を選んで小項目に移動します。その中から該当する項目を選べると pdf ファイルが起動します。	
大項目	
01 正常変異	08 染色体不安定性
02 染色体数の異常	09 白血病・固形腫瘍
03 染色体の構造異常	10 染色体検査の依頼
04 微細欠失症候群	11 核型の読み方
05 性染色体異常	12 染色体異常の頻度
06 リプロダクションの異常	13 染色体マイクロアレイ
07 出生前診断	14 その他

pdf ファイルを参照するには Adobe Reader が必要です。お持ちでない方は [Adobe サイト](#) から入手してください。

小項目

(<http://www.cytogen.jp/index/download.html>)

- ゲノム個別化医療への取り組みは、以前から行われています。
- 単一の遺伝子変異で発症する希少難病疾患においては、原因遺伝子変異を有する家系を対象とした、遺伝カウンセリングが行われています。
- 遺伝子変異の有無や将来的な発症リスク、他の家族や子供への影響など、心理面や社会面も含めた支援が行われています。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## ゲノム個別化医療の例：がん

### 阪大病院での がんゲノム遺伝子 パネル検査の流れ



- **遺伝子パネル検査**を用いてがんゲノム情報を取得し、どの体細胞遺伝子変異が原因でがんを発症したのか調べ、同定された原因遺伝子に基づき最適な治療薬を選択する、**がんゲノム医療**が始まっています。
- 特定のがん遺伝子(例:EGFR、ALK)の特定の変異が同定された症例に対する遺伝子阻害剤の投与は、個別化医療の先駆けと考えられます<sup>6</sup>。

# ① Polygenic risk score (PRS) とゲノム個別化医療

## ゲノム個別化医療の例: HLA 遺伝子型

### 薬剤副作用関連

### HLA 遺伝子型

・アロプリノール	B*58:01
・カルバマゼピン	A*31:01
・感冒薬	A*02:06
・クロザピン	B*59:01
・アバカビル	B*57:01
・サニルブジン	B*40:01
・チアマゾール	B*27:05

高尿酸血症治療剤  
日本薬局方

アロプリノール錠

**ザイロリック<sup>®</sup>錠50**

**ザイロリック<sup>®</sup>錠100**

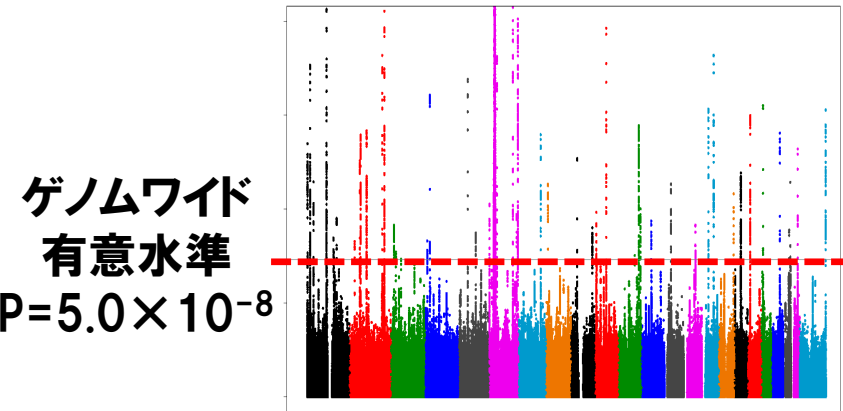
51例中全ての症例が HLA-B\*5801 保有者の報告がある<sup>3)</sup>。また、別の研究では、アロプリノールにより皮膚粘膜眼症候群及び中毒性発症した日本人及びヨーロッパ人において10例中4例(40%)、27例中15例(55%)が HLA-B\*5801 保有者であったとの報告もある<sup>4),5)</sup>。

- ・白血球の血液型を決定する **HLA 遺伝子型** は、広範なヒト形質に強いリスクを有しています。一部の HLA 遺伝子型は、薬剤重症副作用リスクを有するため、薬剤投与前の遺伝子型検査が推奨されています。
- ・薬剤添付文書に記載され、**個別化医療の社会実装例** として重要です。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## ゲノム個別化医療の例: GWAS関連SNP

### GWASによる疾患感受性SNPの同定



GWAS由来  
疾患感受性  
SNPの抽出



### GRSの推定

$$GRS = \sum_i^{risk\ SNP} \beta_i x_i$$

各SNPの  
リスク

各SNPの  
ジェノタイプ

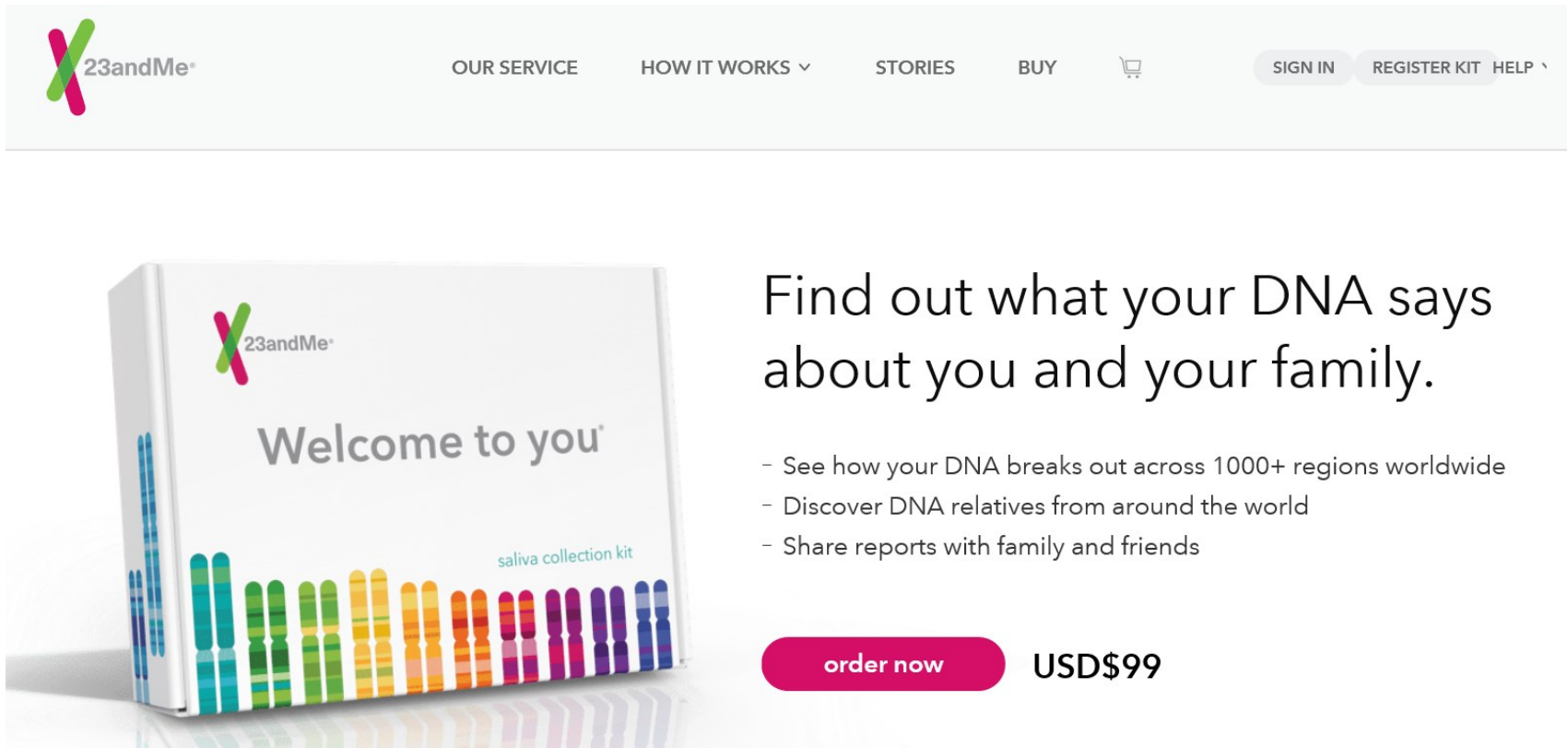


個人のゲノムワイド  
SNPジェノタイプの取得

- GWASが同定した疾患感受性SNPに基づく発症予測も始まりました。
- 各疾患感受性SNPにおいて、GWASで推定されたリスクと、推定対象の個人におけるジェノタイプを掛け合わせた上で、合計します。
- Genotype risk score (GRS)と呼ばれます。



# ① Polygenic risk score (PRS)とゲノム個別化医療

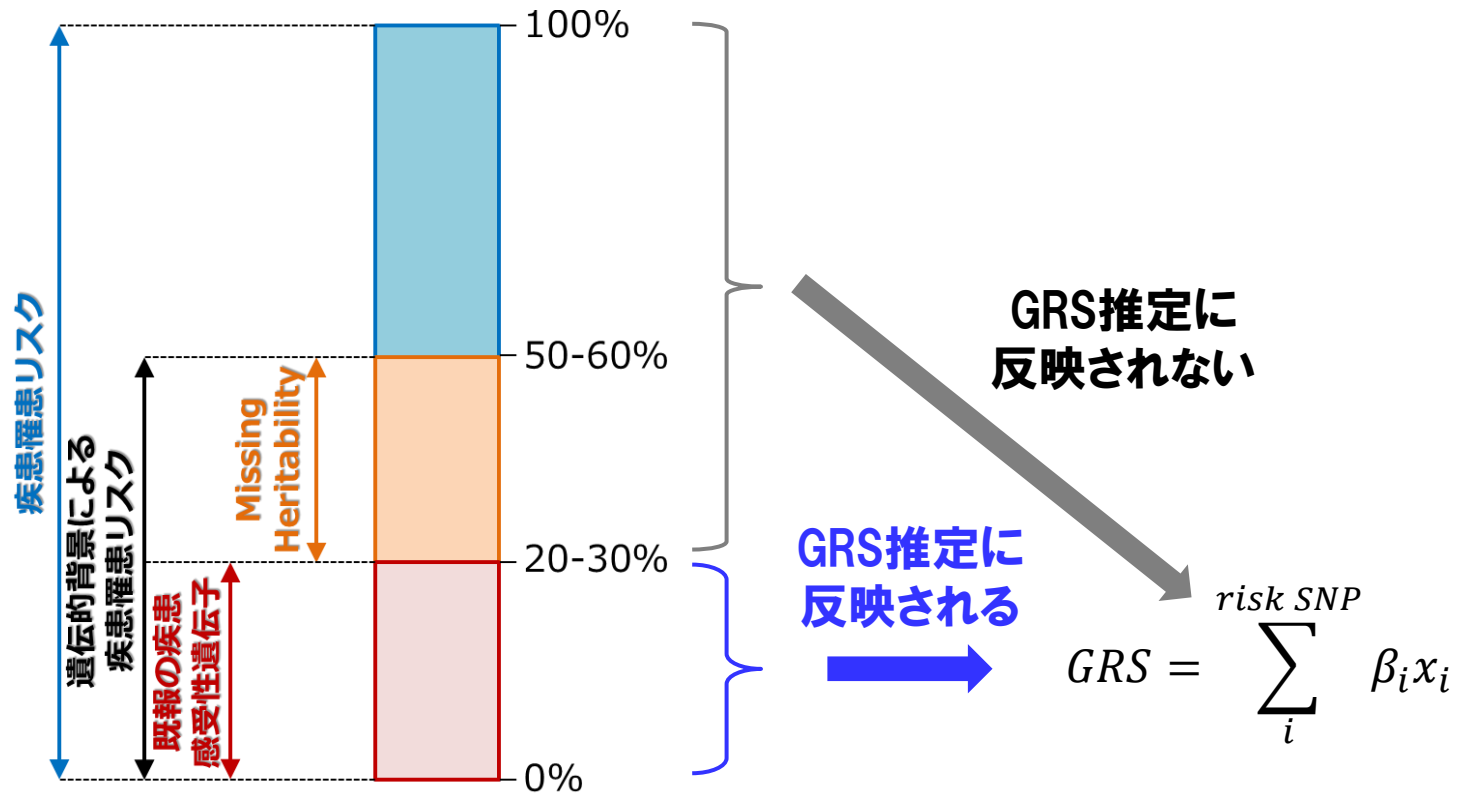


The image shows a screenshot of the 23andMe website. On the left is a product box for the 'Welcome to you' saliva collection kit, featuring the 23andMe logo and a row of colorful test tubes. To the right of the box, the text reads: 'Find out what your DNA says about you and your family.' Below this are three bullet points: '- See how your DNA breaks out across 1000+ regions worldwide', '- Discover DNA relatives from around the world', and '- Share reports with family and friends'. At the bottom right, there is a pink 'order now' button and the price 'USD\$99'. The website's navigation bar at the top includes links for 'OUR SERVICE', 'HOW IT WORKS', 'STORIES', 'BUY', 'SIGN IN', 'REGISTER KIT', and 'HELP'.

- 米国の23andMe社など複数の企業が、**個人が自主的に提供した試料**からゲノム情報を取得し、**GWAS疾患感受性SNPに基づくGRSを推定し**、**疾患発症予測結果を有償**で提供するサービスを始めています。
- 23andMe社は既に100万人以上のゲノム情報と自己申告に基づく臨床情報を収集し、多数のGWAS研究に参画しています。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## Missing heritabilityの謎

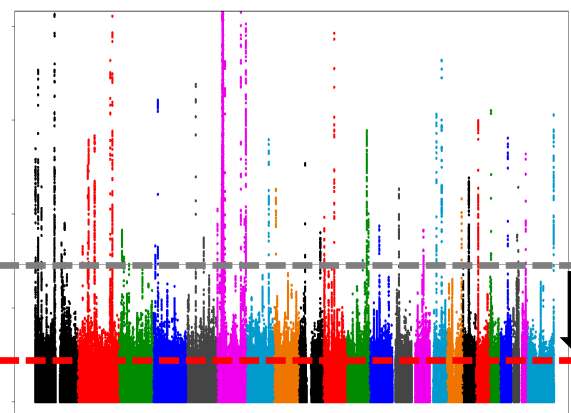


- しかし、ゲノムワイド有意水準( $P=5.0 \times 10^{-8}$ )を満たした少数のGWAS疾患感受性SNPだけでは、疾患の遺伝的背景の一部しか説明できないことが知られています(=“missing heritability”)。
- 現状では、臨床応用へのハードルはまだ高い状況と考えられています。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## ゲノム個別化医療の例: Polygenic Risk Score

GWASによる疾患感受性SNPの同定



緩い有意水準  
例:P=0.01

有意水準が緩くなった分、より多くのSNPがリスク予測に使われる

より多くのGWAS由来疾患感受性SNPの抽出

PRSの推定

$$PRS = \sum_i^{risk\ SNP} \beta_i x_i$$

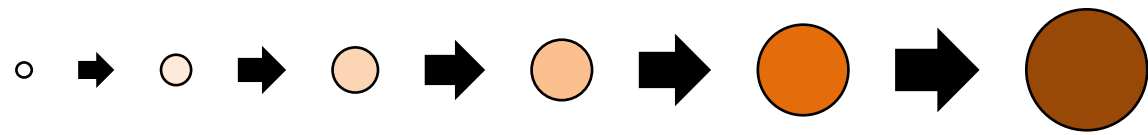
各SNPのリスク      各SNPのジェノタイプ

個人のゲノムワイドSNPジェノタイプの取得

- 感受性SNPによる発症予測精度を高めるにはどうすればいいでしょうか。
- リスク計算に採用するGWAS P値の閾値を、ゲノムワイド有意水準より緩くすると、より多くの感受性SNPが対象となり精度が改善します。
- Polygenic model (=ゲノム全体の無数のコモンバリエーションが弱い疾患リスクを有するという仮説)に基づくことから、polygenic risk score (PRS)と呼ばれます。

# ① Polygenic risk score (PRS)とゲノム個別化医療

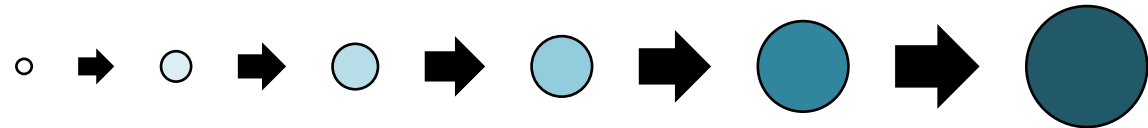
GWAS  
サンプル数



$$PRS = \sum_i^{risk\ SNP} \beta_i x_i$$

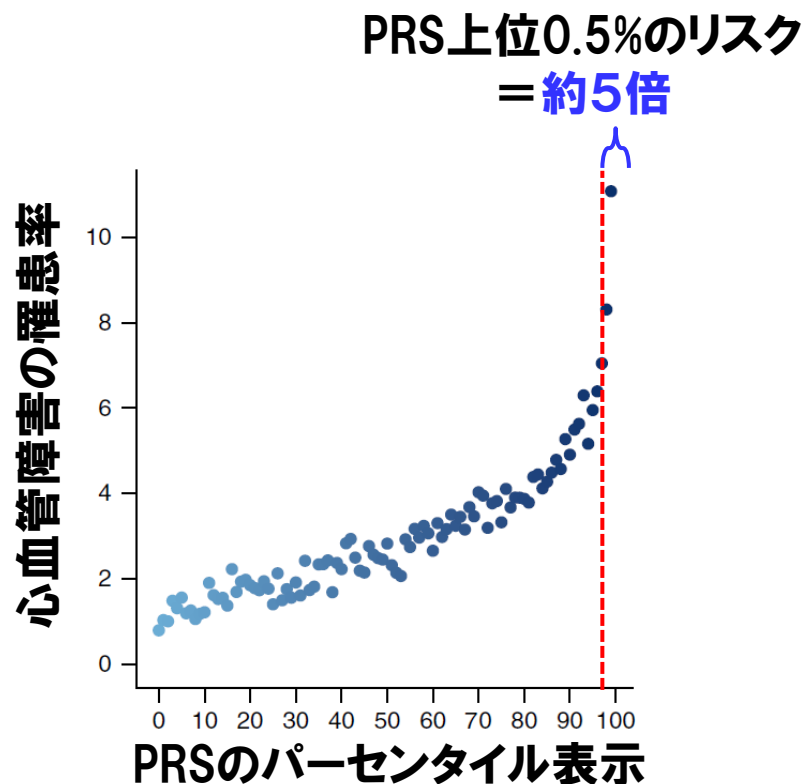
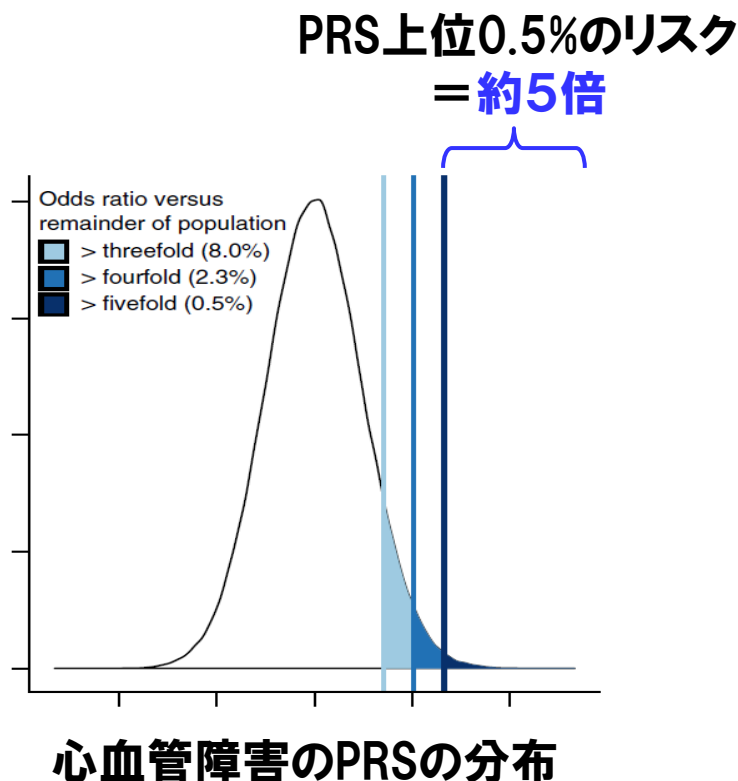


PRS  
予測精度



- 発症予測精度を高めるもう一つの方法として、**リスク計算モデルの構築**に使うGWASの**サンプルサイズを大きくすること**、が挙げられます。
- 数百人～数千人規模のGWASでは難しかった高精度のPRS予測が、**数万人～数十万人規模のGWAS**を用いることで可能になりつつあります。

# ① Polygenic risk score (PRS) とゲノム個別化医療



- ・欧米人集団20万人規模の心血管障害(CAD)GWAS解析の結果、**PRS上位0.5%の集団は5倍のリスク**を有することが明らかになりました。
- ・個々の感受性SNPが持つリスクは小さくても、大規模GWASに基づきゲノム領域全体のSNPのジェノタイプを組み合わせることで、効果的な予測が可能になることが示されました。

# ① Polygenic risk score (PRS) とゲノム個別化医療

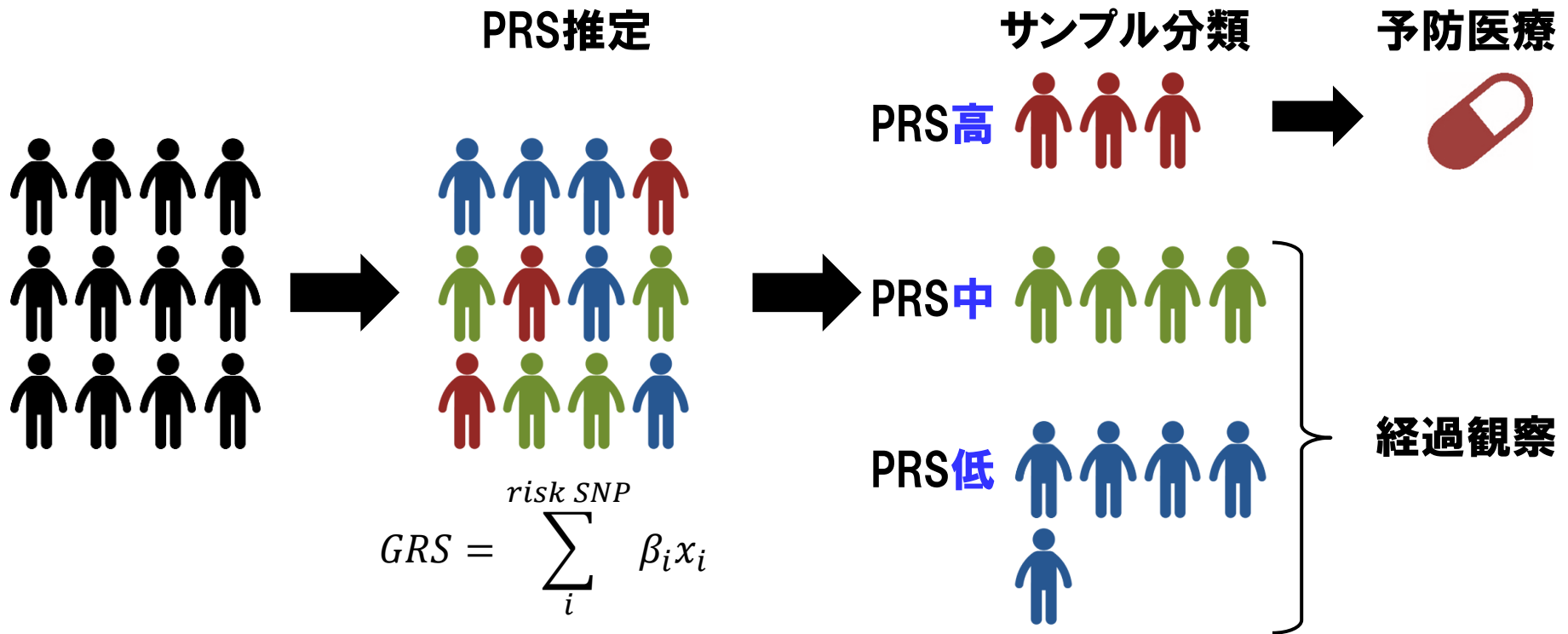


PRSを使うと、疾患発症をどこまで予測することが可能なのでしょうか？



誰が、いつ、どの疾患を発症するかを正確に予測することは、現状では難しいです。でも、集団内のサンプルを高リスク群と低リスク群に分類することは可能です。

# ① Polygenic risk score (PRS) とゲノム個別化医療



- 生活習慣病など、集団中の一定の割合のサンプルが、加齢に伴い発症する疾患においては、予め発症リスクの高いサンプル群を同定し、**予防医療**を施すことが有用です。
- 既知のリスク因子にPRSを加味することで、高リスク群サンプルの効率的な同定が可能になり、予防医療が推進されると期待されています。

# ① Polygenic risk score (PRS) とゲノム個別化医療



PRS計算の元となるGWAS解析結果には、どのスタディを採用するのがいいのでしょうか？一番サンプルサイズの大きいGWASが良いのでしょうか？

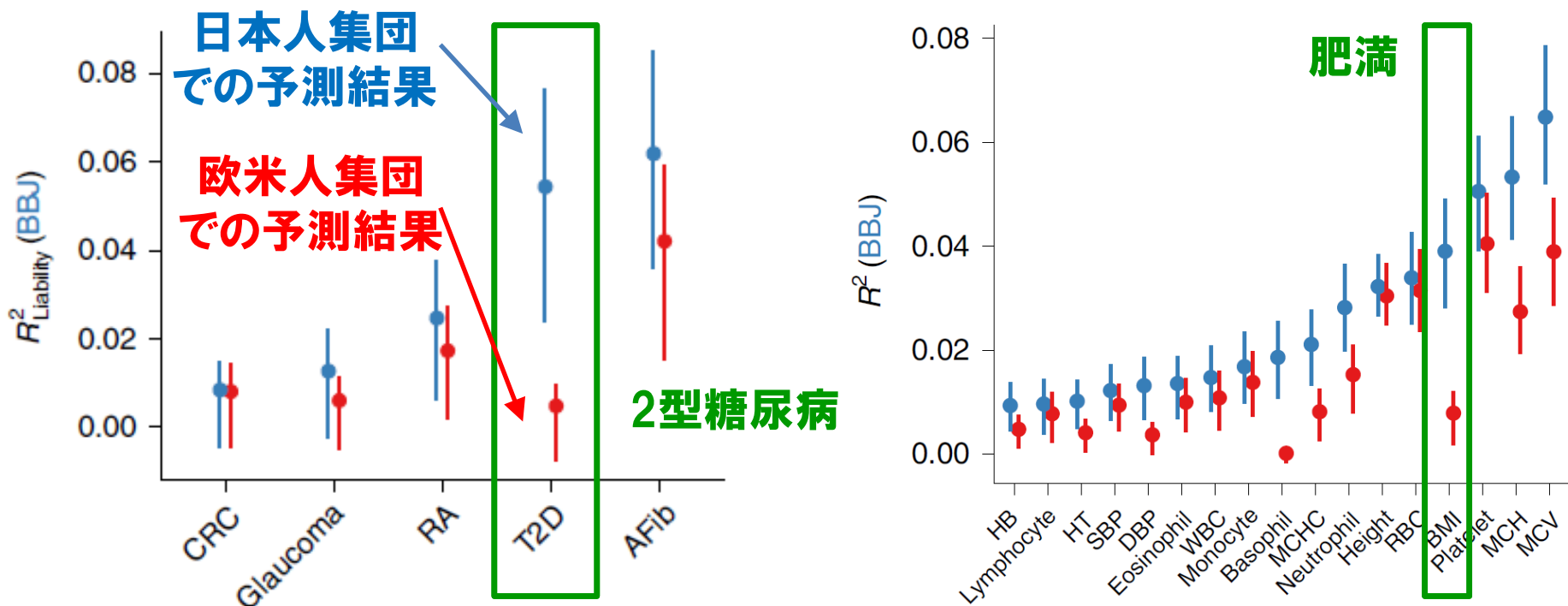


確かにGWASサンプルサイズが大きいことは重要です。しかし、**PRS推定の対象集団と同じ人種集団で実施されたGWASを採用することの必要性が指摘されています。**



# ① Polygenic risk score (PRS)とゲノム個別化医療

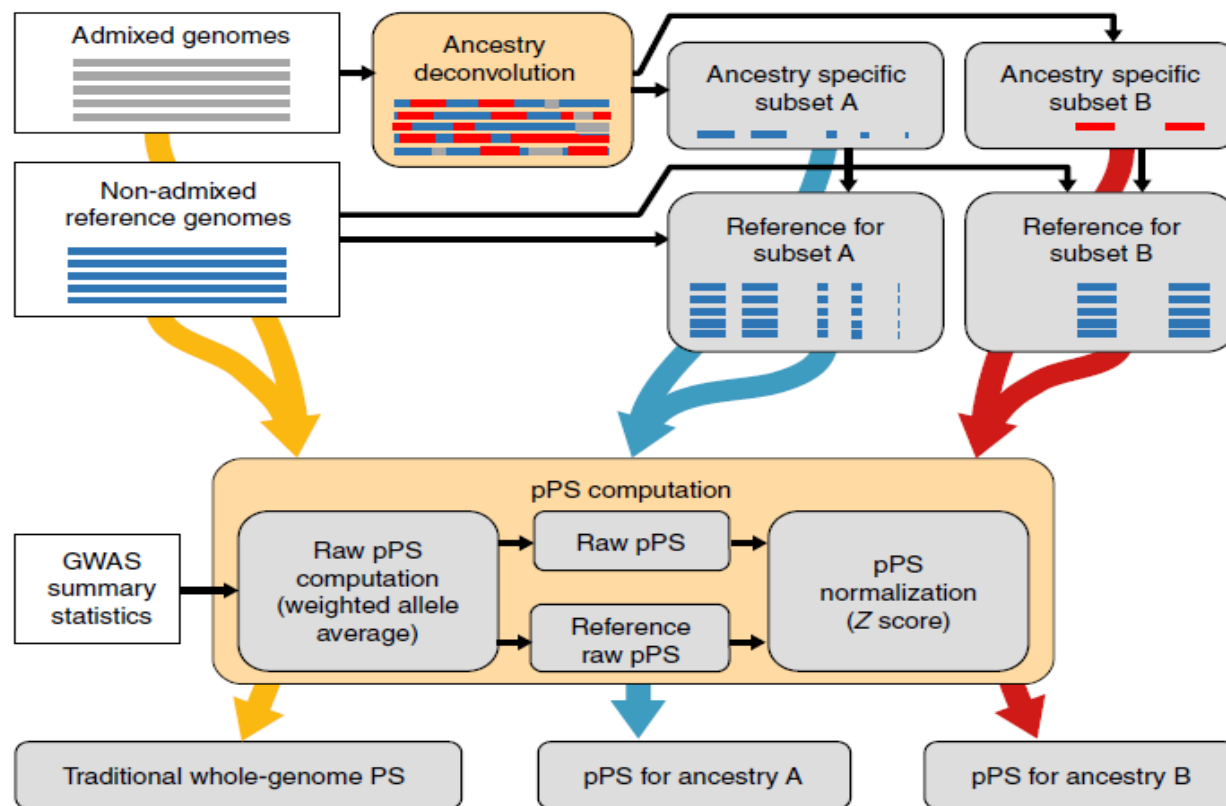
## 日本人集団ゲノム情報に基づくPRSを用いた疾患発症予測



- 日本人集団のGWAS結果に基づき、欧米人集団GWASデータにおけるPRSを推定した結果、特に**2型糖尿病および肥満において異人種間のPRS予測が困難なことが判明しました。**
- **PRSの異人種間互換性の低さが、特定の形質に依存的に存在しているようです。**

# ① Polygenic risk score (PRS)とゲノム個別化医療

## 各SNPの祖先集団を考慮したPRS計算

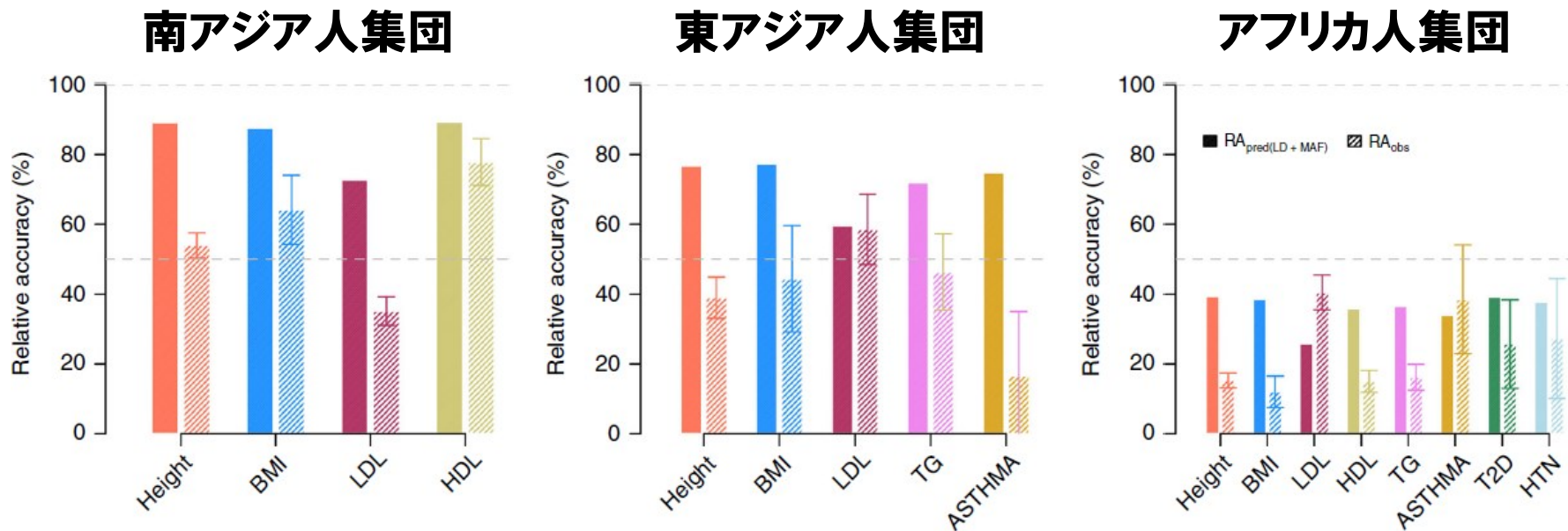


- 個人のヒトゲノム配列には、異なる祖先集団由来の配列が部分的に含まれていることが知られています(例:欧米人集団由来・アフリカ人集団由来)。
- PRS計算時に各SNPが含まれる祖先集団配列を考慮することで、人種間バイアスが(ある程度)補正可能になるという報告もあります。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## 欧米人集団GWAS結果に基づく他人種でのPRS予測能比較

PRS予測精度



SNPアレル頻度差やLD構造差を考慮したPRSモデル

■  $RA_{pred(LD + MAF)}$

▨  $RA_{obs}$

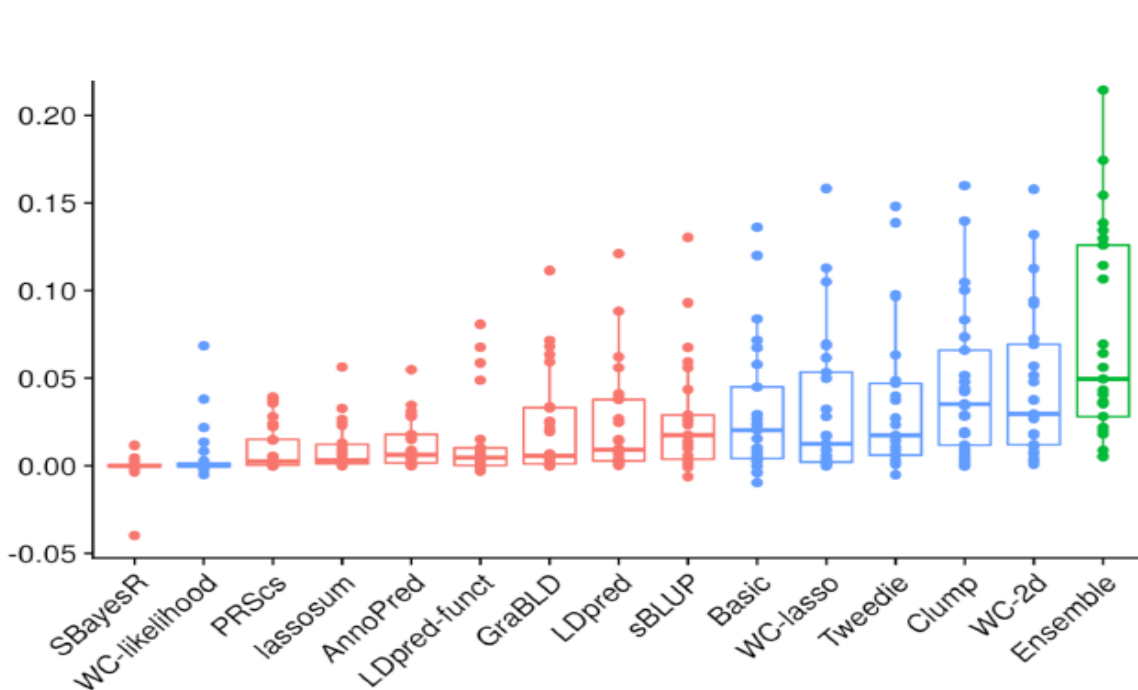
シンプルなPRSモデル

- PRSの異人種間互換性の低さの大半は、SNPアレル頻度や連鎖不平衡 (LD)構造の人種集団間の違いで説明可能である、という説もあります。
- 人種集団の遺伝的背景、形質特異的要因、コホート特異的要因、様々な要因が存在しているようです。

# ① Polygenic risk score (PRS) とゲノム個別化医療

## PRS 計算手法に依存した予測精度の違い

予測精度 (AUC) の相対的改善

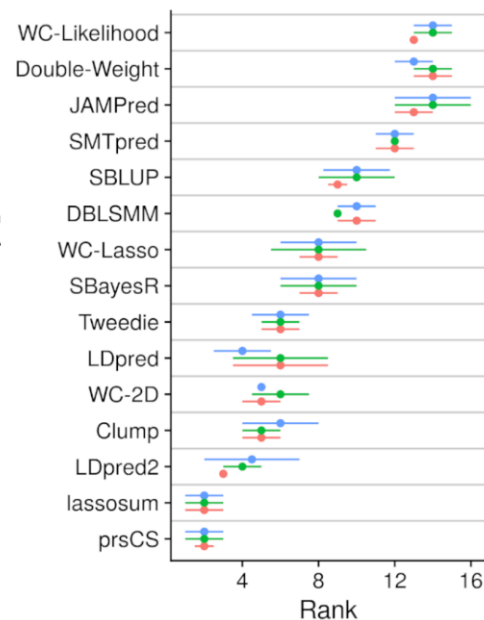


PRS 計算手法 (シンプル型・複雑型・組み合わせ型)

PRS 計算手法

Statistic

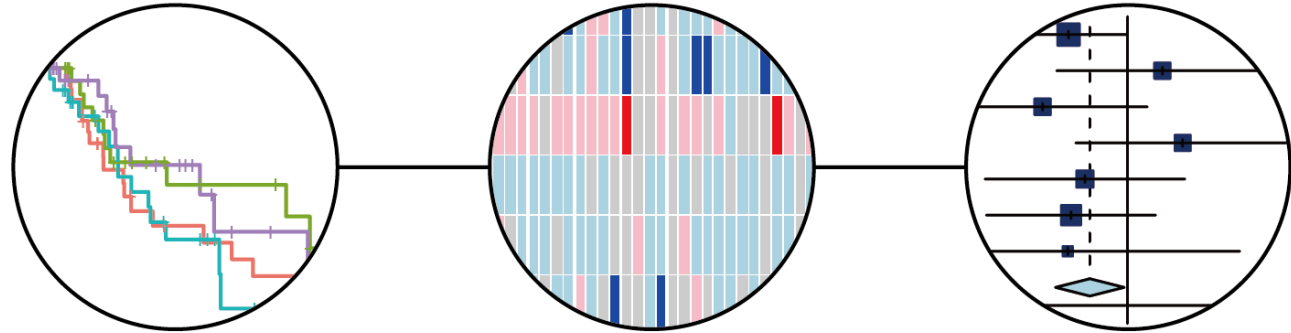
- Disease Incidence Difference (green)
- Concordance (red)
- Odds Ratio (blue)



予測精度順位

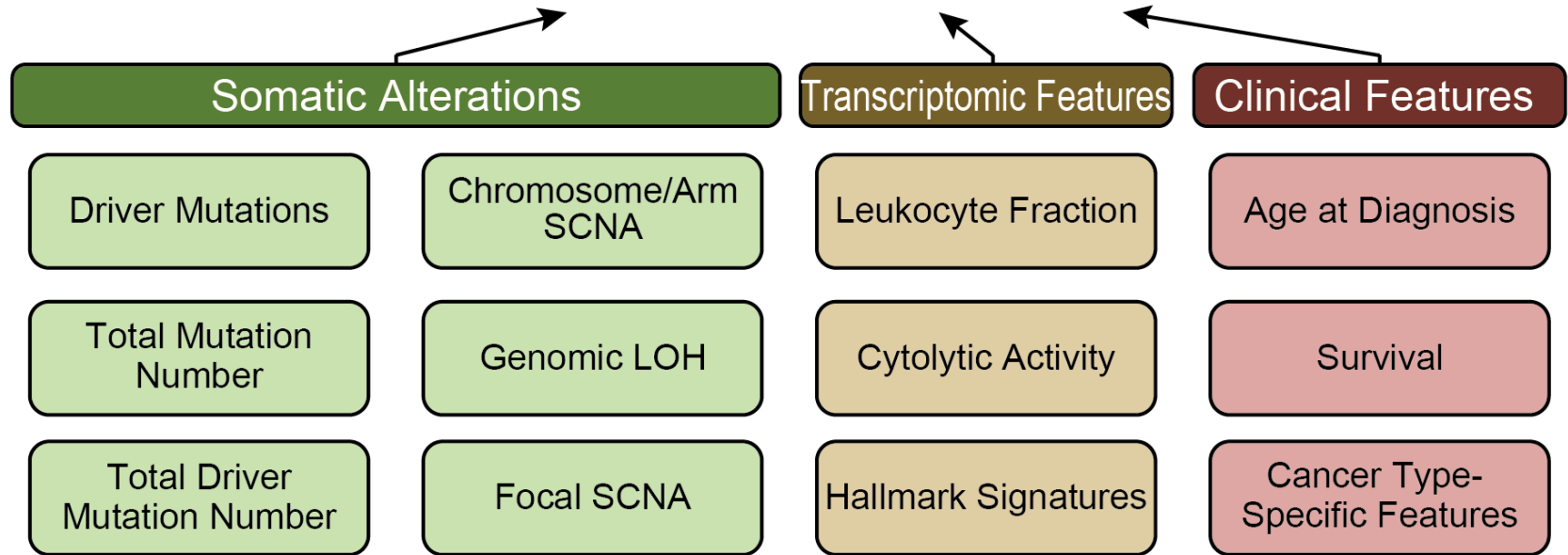
- 数多くのPRSの計算手法が開発されていますが、どのPRS計算手法が、どの形質に最適なのか、まだ結論が出ていません。
- 必ずしも複雑なPRS計算手法の予測精度が高いわけではない、という事実も判明しつつあります。

# ① Polygenic risk score (PRS)とゲノム個別化医療



Germline-Somatic Association Study

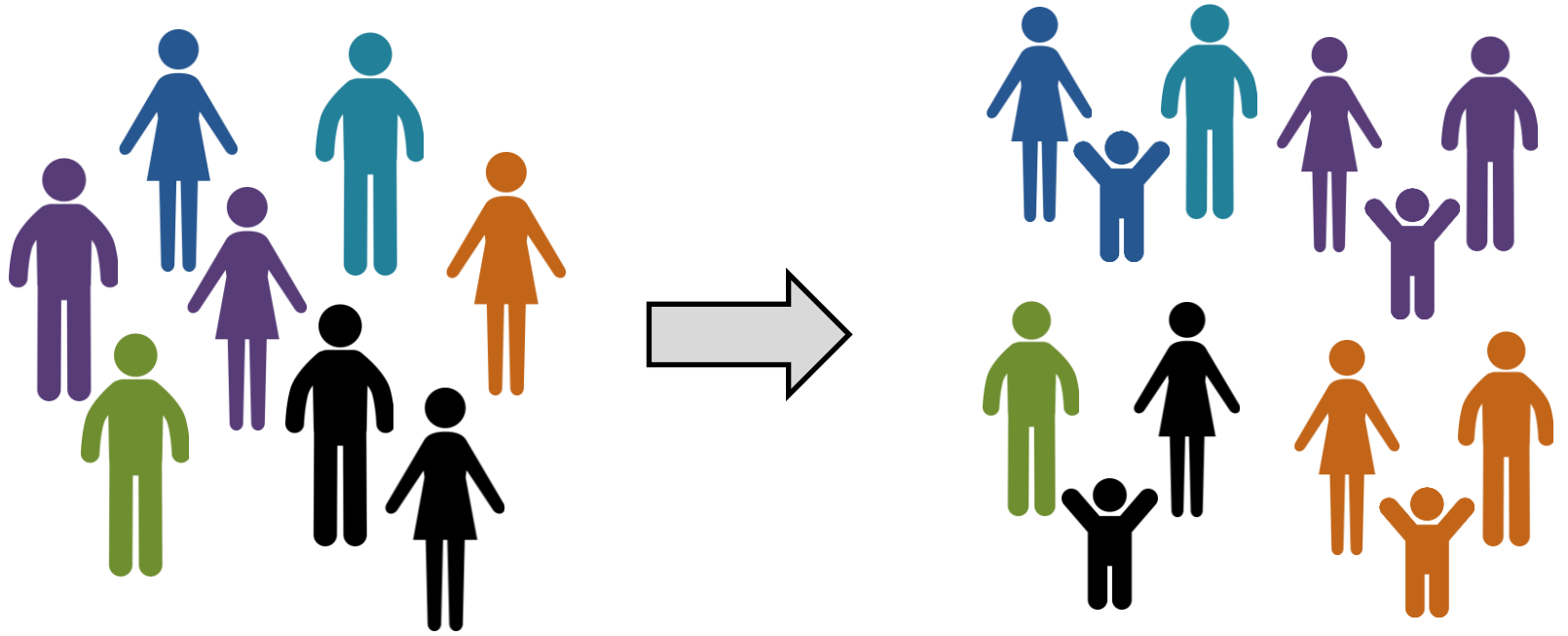
Somatic Alterations



・がんの生殖細胞・体細胞系列変異を比較し、がんPRSと相互作用する体細胞遺伝子変異の特徴が明らかに。

# ① Polygenic risk score (PRS)とゲノム個別化医療

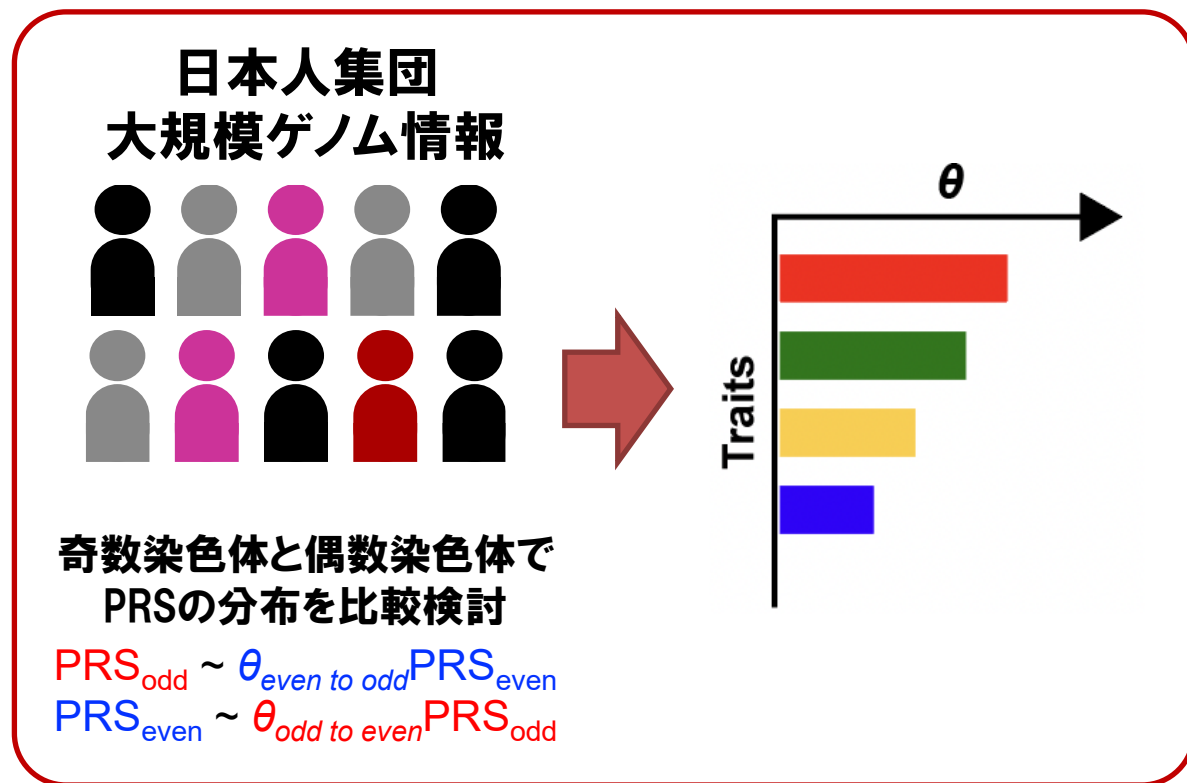
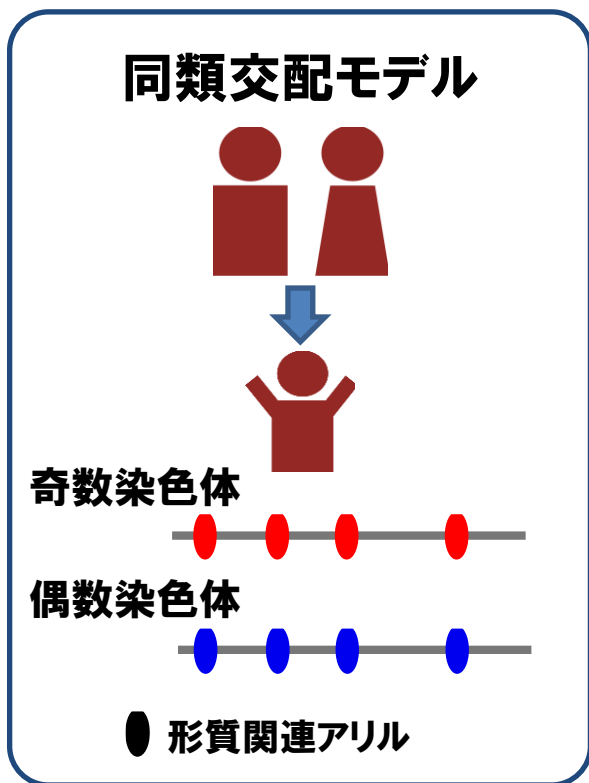
## ヒトにおける同類交配と集団ゲノム情報の関わり



- **同類交配**(assortative mating: AM)とは、類似した表現型を持つ個体同士が任意で予想されるよりも高い確率でペアを形成する現象です。
- 集団遺伝学においては、表現型の類似性の増加は**遺伝的類似性**につながり、表現型感受性遺伝子変異の集団中の不均衡分布をもたらす。
- 欧米人集団においては**身長**や**学歴**に対する同類交配が存在します。

# ① Polygenic risk score (PRS) とゲノム個別化医療

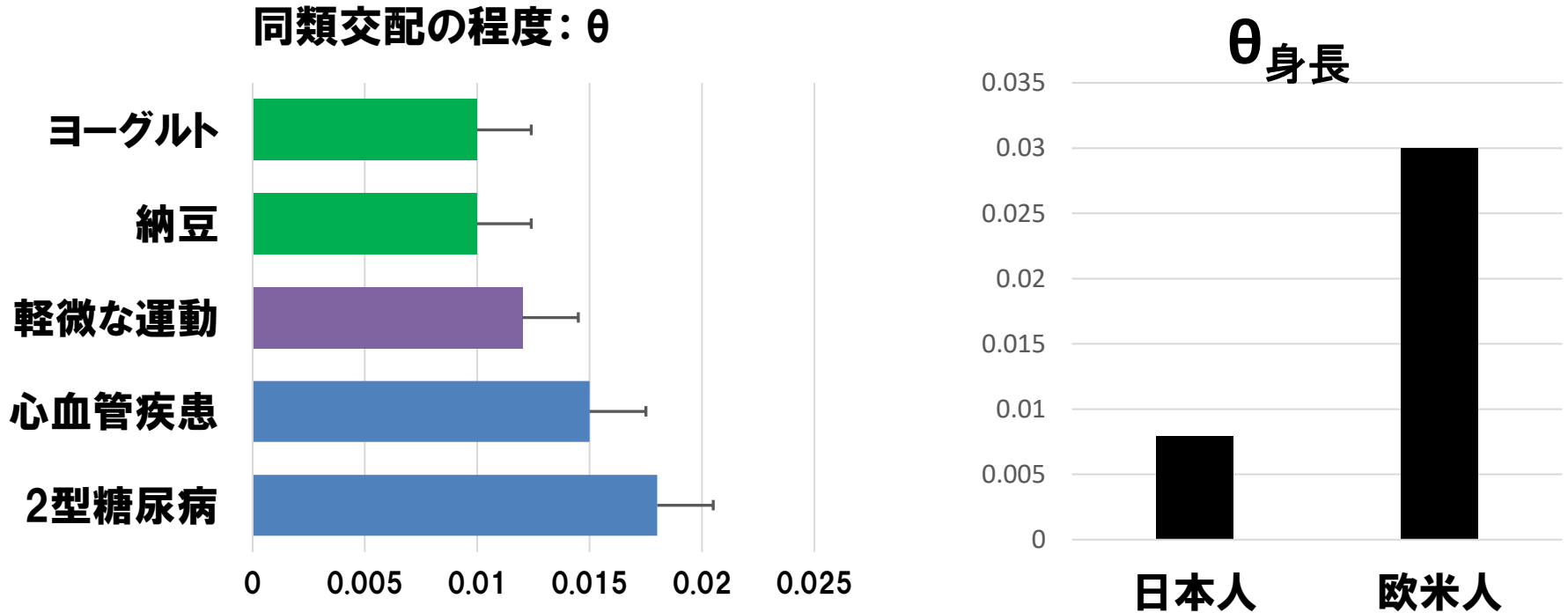
## Polygenic risk score を用いた同類交配の検証



- ゲノム全体の形質関連変異を統合するpolygenic risk score (PRS) を異なる染色体間で比較することで(例: 奇数vs偶数)、同類交配を検討可能。
- 日本人集団大規模ゲノム情報(バイオバンク・ジャパン)を対象に、多彩な形質のPRSを計算し、同類交配の影響を検討しました。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## Polygenic risk scoreを用いた同類交配の検証



- 日本人集団では、**2型糖尿病、心血管疾患、軽微な運動習慣、納豆消費量、ヨーグルト消費量**で、有意な同類交配の影響を観測。  
(食物摂取量には地域差があるので、集団構造化の影響も要検討)
- 身長に対する同類交配の影響は日本人集団では欧米人集団ほど顕著ではないことが明らかになりました。



# ① Polygenic risk score (PRS) とゲノム個別化医療

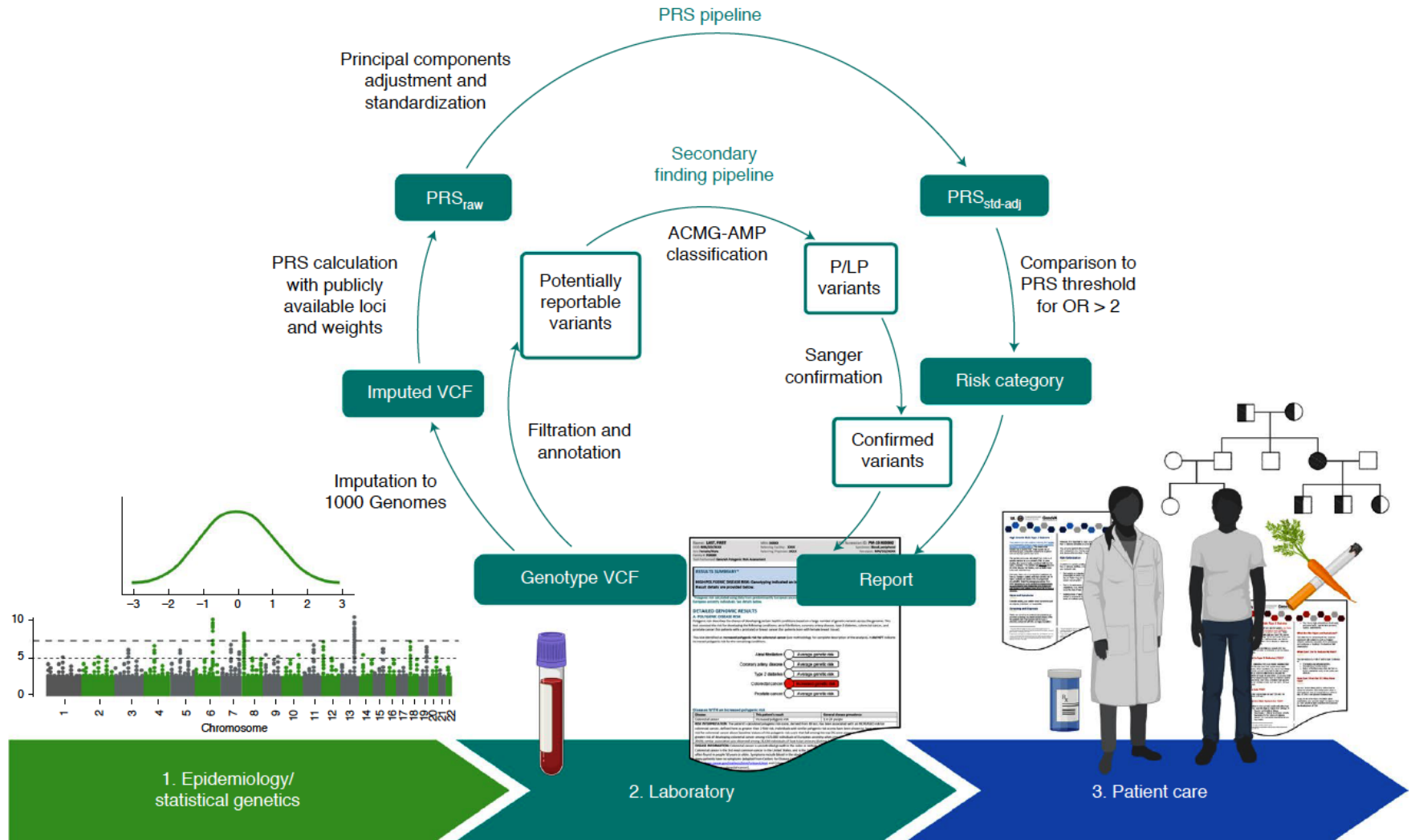
PRSを個人に還元するまでの「遠い」道のり



- PRSによる疾患発症予測精度向上のための課題整備が進められています。
- 一方、「**得られたPRSをどのように個人に還元するか**」については、未だ具体的な道筋が得られていない状況です。
- **医療機関・アカデミア・ベンチャーの密な連携**を通じてゲノム個別化医療社会実装に取り組んでいく時代が到来しています。

# ① Polygenic risk score (PRS)とゲノム個別化医療

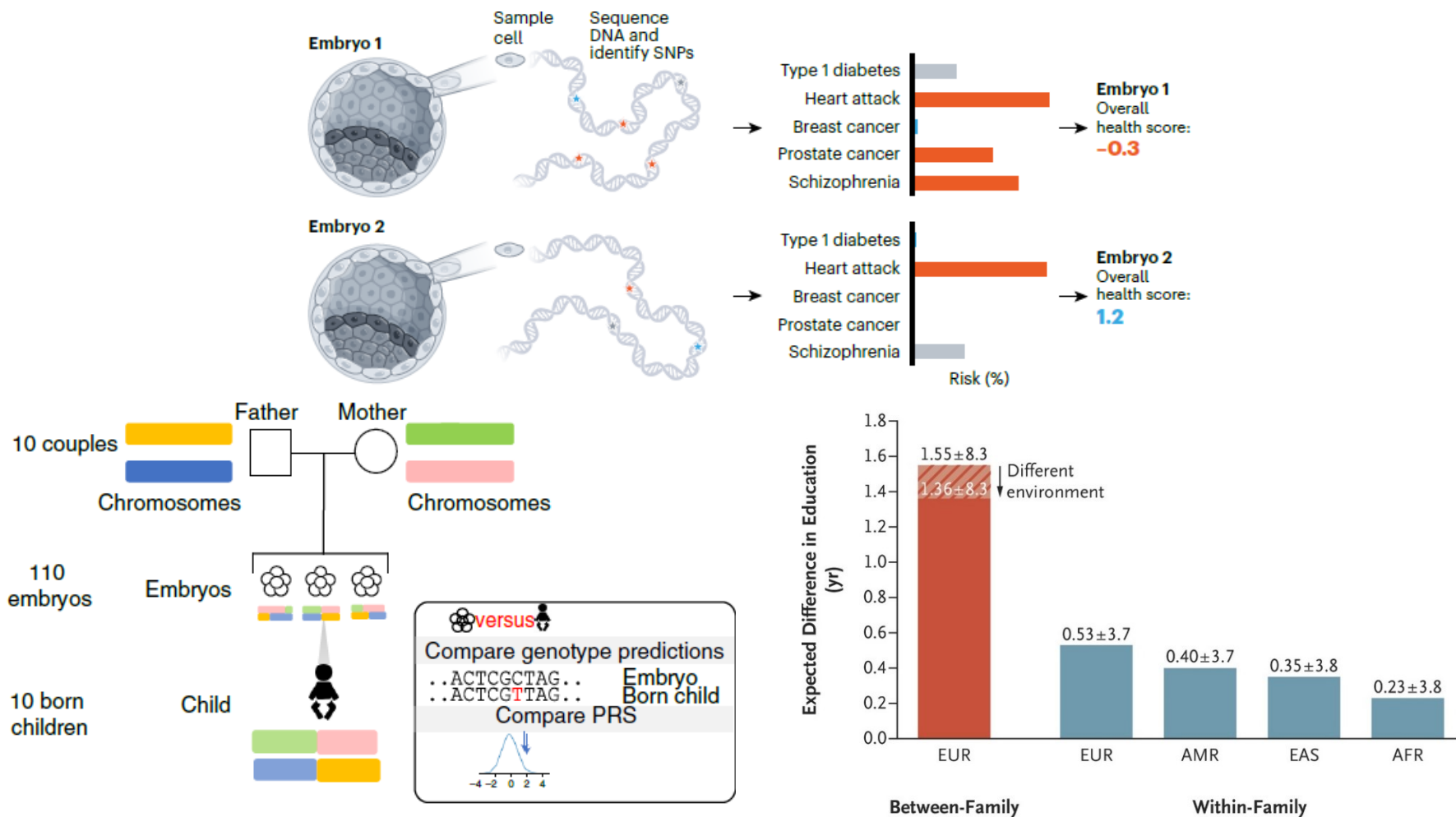
## Veterans Affairs Genomic MedicineにおけるPRS還元の取り組み



・実際にPRSを臨床現場に還元する枠組みも、少しずつ始まっているようです。

# ① Polygenic risk score (PRS)とゲノム個別化医療

## 出生前診断におけるPRS活用を巡る懸念



• 出生前診断で複数胚由来のゲノム情報からPRSを推定し、特定形質における推定値に基づき胚を選挙する医療行為について、懸念が提唱されています。

(Turley P et al. *NEJM* 2021, Forzano F et al. *Eur J Hum Genet* 2021, Kozlov *Nature* 2022, Kumar A et al. *Nat Med* 2022)

## GenomeDataAnalysis5

① Polygenic risk score(PRS)とゲノム個別化医療

② PRSice-2を使ったPRS解析

本講義資料は、Windows PC上で  
C:\SummerSchoolにフォルダを配置すること  
を想定しています。

## ② PRSice-2を使ったPRS解析

### 主なPRS解析ソフトウェア

ソフトウェア	URL	引用文献
LDpred	<a href="https://github.com/bvilhjal/ldpred">https://github.com/bvilhjal/ldpred</a>	Vilhjalmsson BJ et al. <i>Am J Hum Genet</i> 2015
lassosum	<a href="https://github.com/tshmak/lassosum">https://github.com/tshmak/lassosum</a>	Mak TSH et al. <i>Gebet Epidemiol</i> 2017
PRSice-2	<a href="https://www.prsice.info/">https://www.prsice.info/</a>	Choi SW et al. <i>GigaScience</i> 2019
PRS-CS	<a href="https://github.com/getian107/PRS-CS">https://github.com/getian107/PRS-CS</a>	Ge T et al. <i>Nat Commun</i> 2019

- PRS解析を実施するソフトウェアは、複数あります。
  - どのソフトウェアを使うと一番高精度にPRSを推定できるのかは、PRSモデルの精度評価の難しさもあり、**結論が出ていません**。
  - 本実習では、下記の理由から**PRSice-2**を使った演習を行います。
    - ①: **Linux**、Mac、Windows用の実行ファイルが提供されている。
    - ②: PRS解析結果の**プロット機能**が付いている。
    - ③: **親切なチュートリアル**がWebページに記載されている。
- (※PRSice-2がPRS計算に一番適切なソフトウェアというわけではありません!!)<sup>29</sup>

## ② PRSice-2を使ったPRS解析

### PRSice-2

<https://www.prsice.info/>

Home PRSice-2

Search docs

Home

QUICK START

- PRSice
- PRSet

DETAIL GUIDES

- PRSice
- PRSet

Available Commands

DEVELOPERS

- Compile from Source
- Development Decisions
- Useful Resources

MISC

- Additional Steps for MAC and Window users
- Frequently Asked Questions
- Archive
- Update Log

Docs » Home [Edit on GitHub](#)

**Announcement**

Our guide paper is now published in Nature Protocols:

Choi, S.W., Mak, T.S. & O'Reilly, P.F. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc (2020). <https://doi.org/10.1038/s41596-020-0353-1>

The online tutorial that accompanies the paper is [here](#)

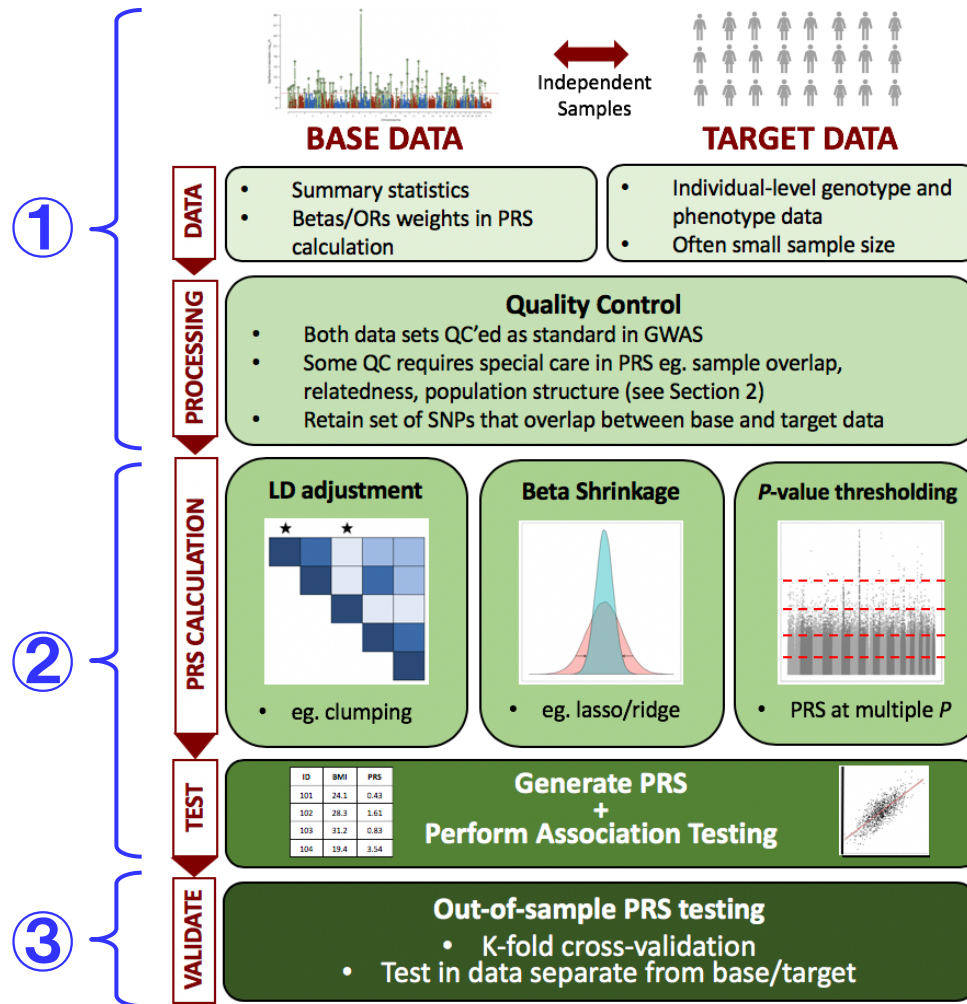
### PRSice-2: Polygenic Risk Score software

PRSice (pronounced 'precise') is a Polygenic Risk Score software for calculating, applying, evaluating and plotting the results of polygenic risk scores (PRS) analyses. Some of the features include:

1. High-resolution scoring (PRS calculated across a large number of P-value thresholds)
2. Identify Most predictive PRS
3. Empirical P-values output (not subject to over-fitting)
4. Genotyped (PLINK binary) and imputed (Oxford bgen v1.2) data input
5. Biobank-scale genotyped data can be analysed within hours
6. Incorporation of covariates
7. Application across multiple target traits simultaneously

• PRSice-2のソフトウェアは、テストデータと共に公開されています。 <sup>30</sup>

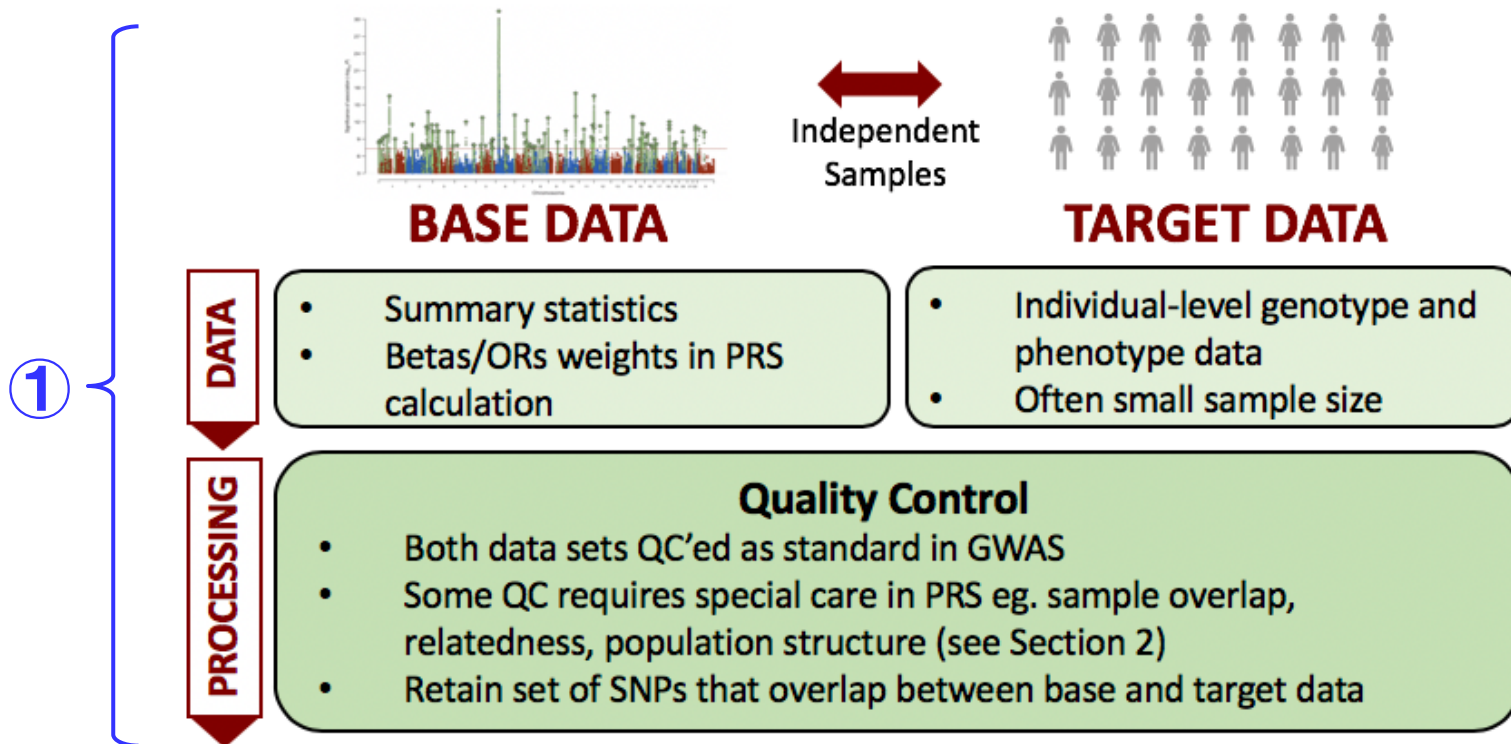
## ② PRSice-2を使ったPRS解析



• PRS解析のステップは、①: GWASおよび推定サンプルデータの準備、②: PRS計算モデルの構築、③: PRS計算結果の検証、に分かれます。

## ② PRSice-2を使ったPRS解析

### ①: GWASおよび推定サンプルデータの準備

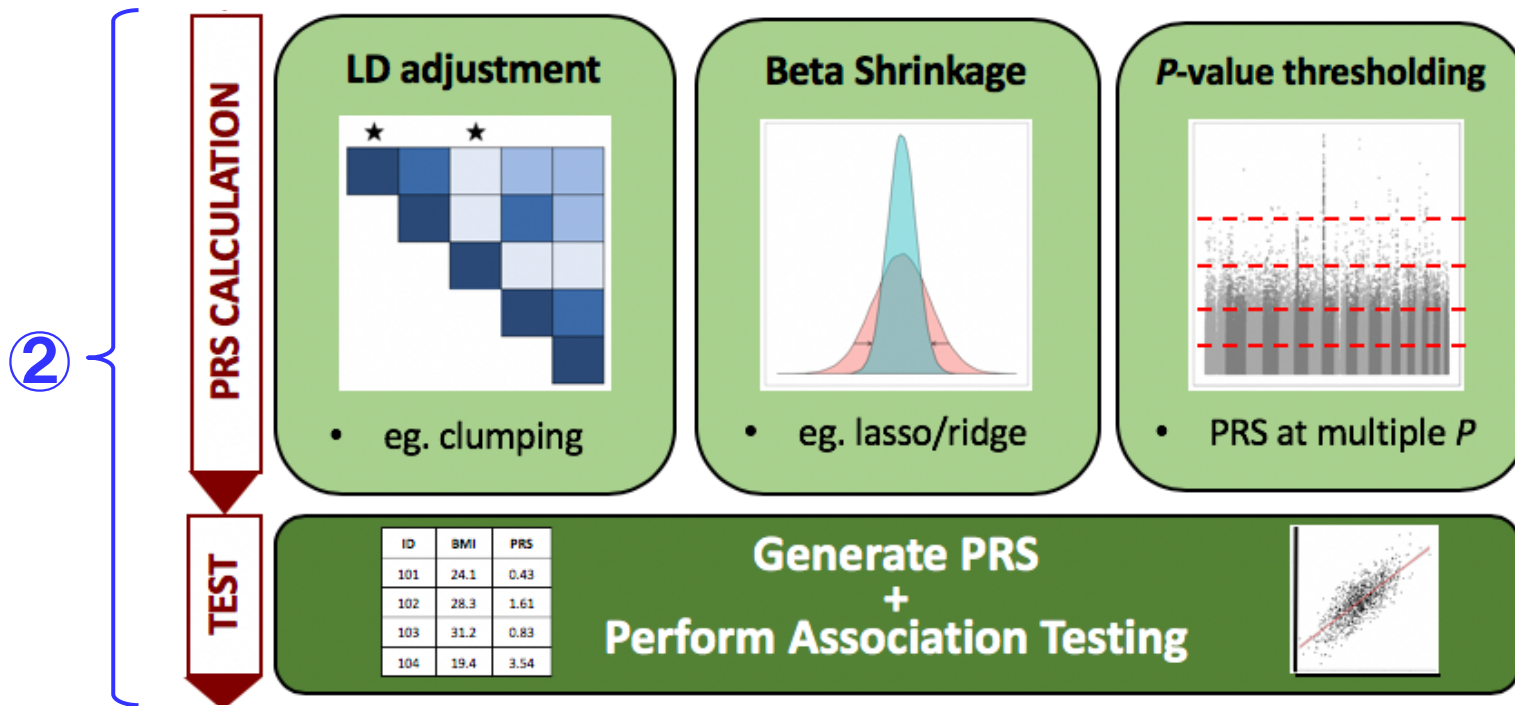


- PRS解析では、**予測モデル構築に使うGWAS結果(Base Data)**と、**PRS推定の対象となるGWASデータ(Target Data)**の2種類が必要です。
- **対象形質**を決定した上で、前者にはGWAS解析統計量(全ゲノムSNPにおけるオッズ比・P値)、後者には個人別ジェノタイプデータを使用します。
- 両者の間には、**重複サンプルが含まれていない**ことが推奨されます。<sup>32</sup>



# ② PRSice-2を使ったPRS解析

## ②: PRS計算モデルの構築



• 下記に留意して、PRS計算モデル構築に必要なリスクSNPを抽出します。

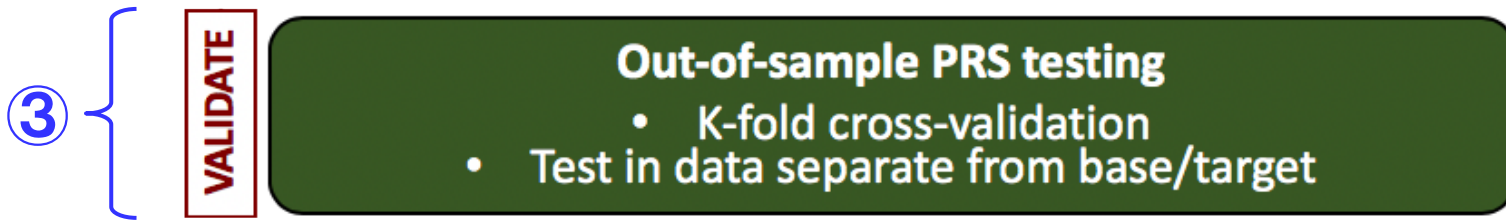
(i) 近傍でLD関係にある複数リスクSNP群から独立なSNPを選択。

(ii) GWAS実施時の各SNPのオッズ比推定バイアスの補正。

(iii) リスクSNP抽出時のP値の有意水準を複数候補の中から選択。

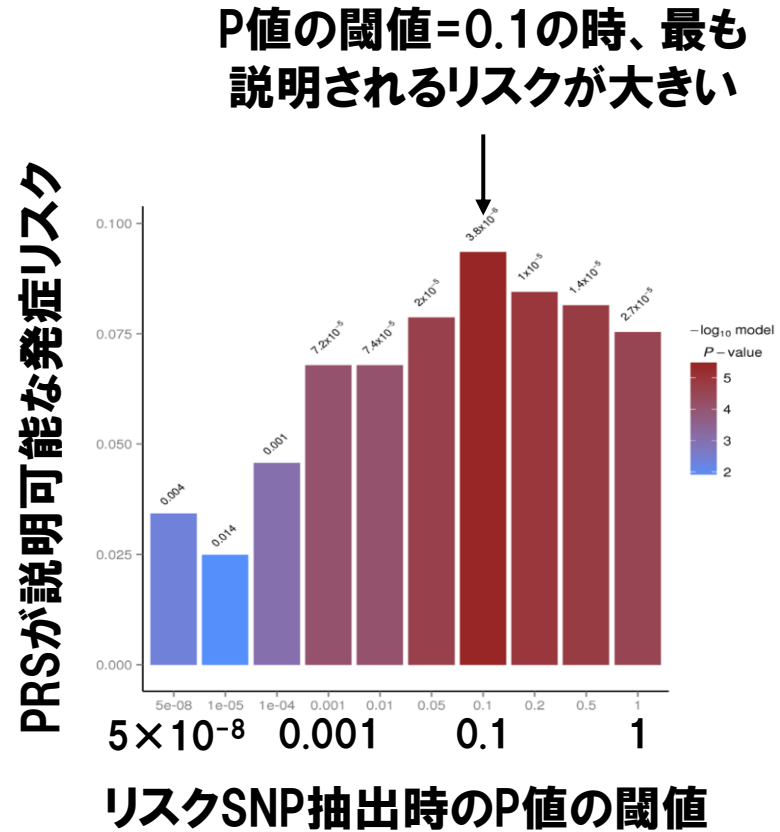
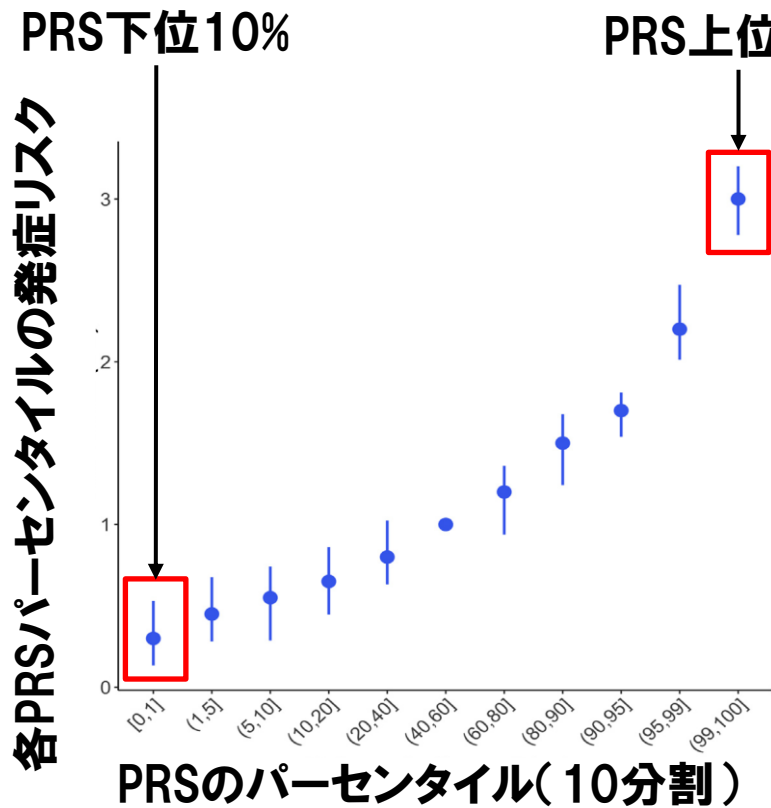
## ② PRSice-2を使ったPRS解析

### ③:PRS計算結果の検証



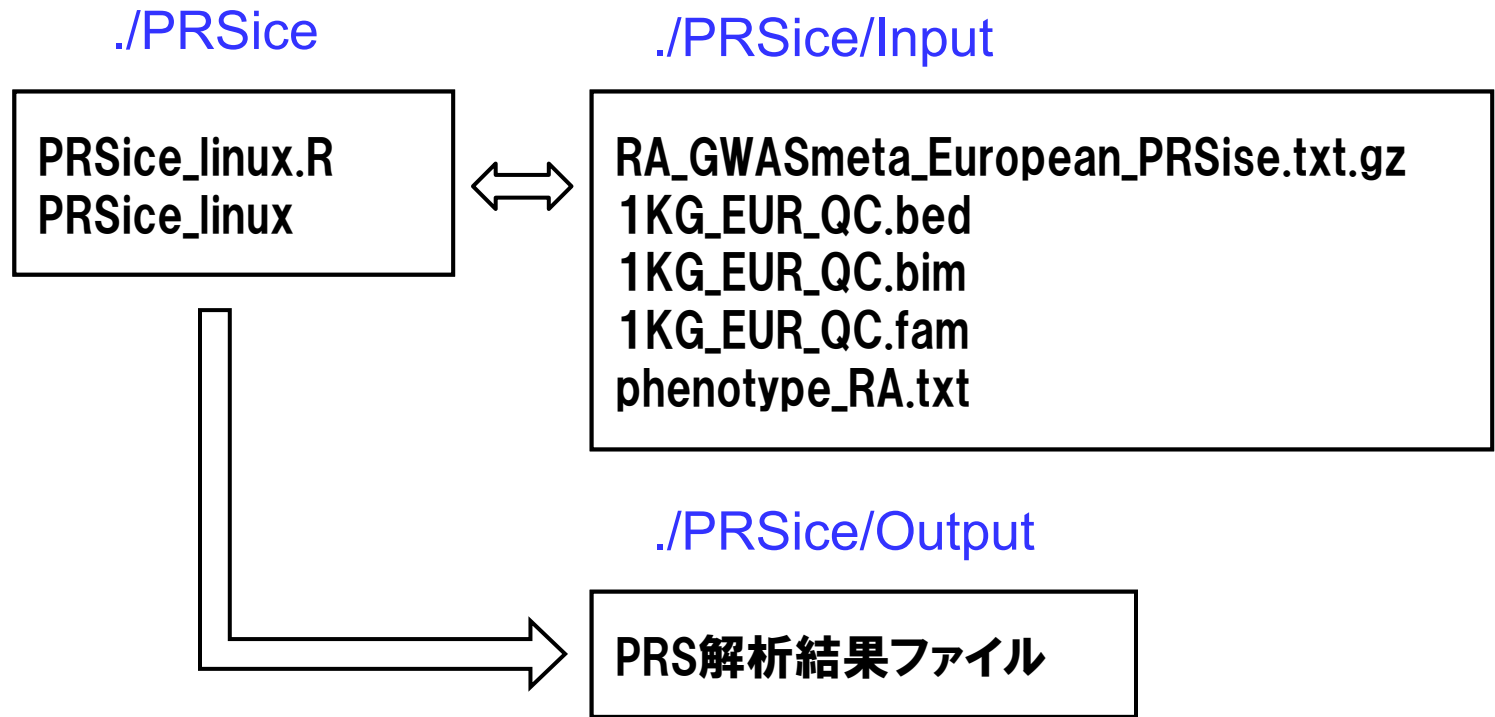
- 推定されたPRSと対象形質との関連を調べることで、**PRS推定精度を経験的に検証**します。
- Target Dataを分割して**クロスバリデーション法**での検討や、**追加の対象形質GWASデータ**をもう一つ用意しての検討が行われます。

## ② PRSice-2を使ったPRS解析



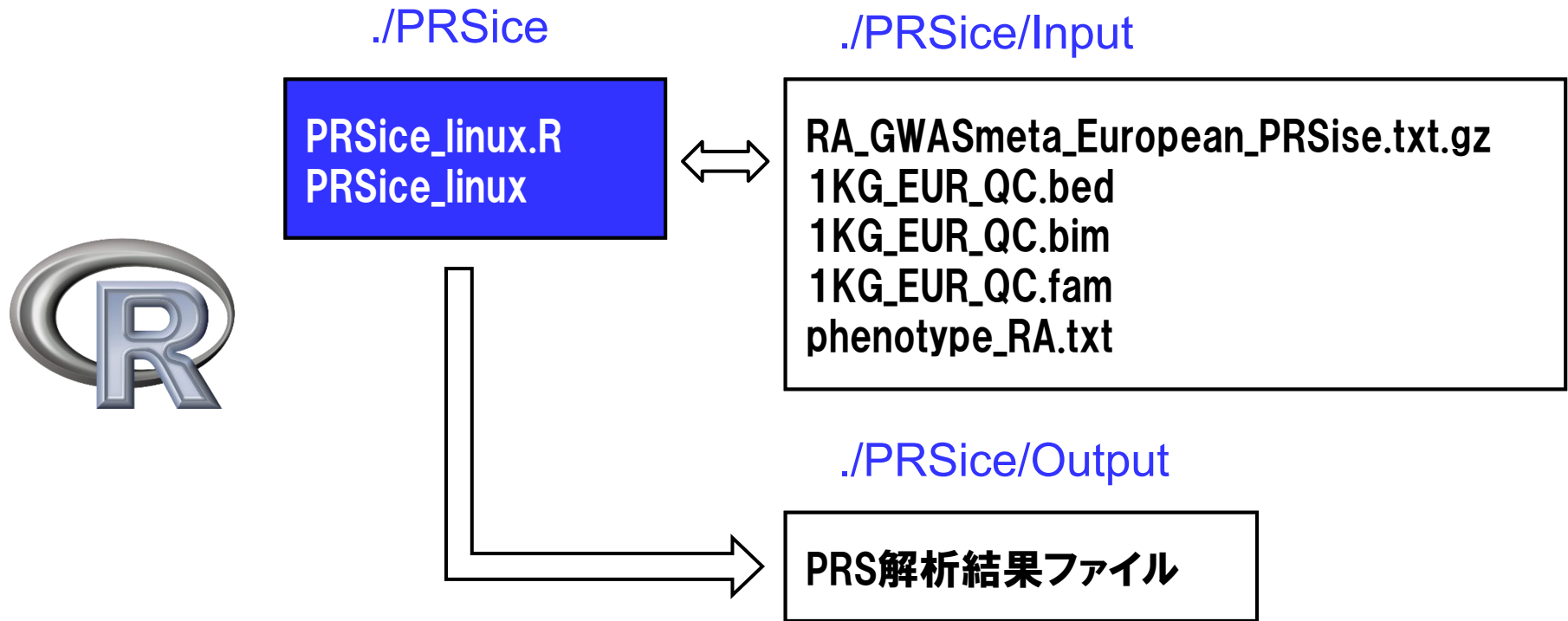
- PRSが集団中で説明可能なリスクは、**サンプルをPRS値の順に分割し** (例:10分割)、各グループのリスクをプロットすることでも可視化できます。
- **リスクSNP抽出時のP値の閾値を変化させた場合の、PRSが説明可能な疾患発症リスクの定量的な変化の検討も行われます。**

## ② PRSice-2を使ったPRS解析



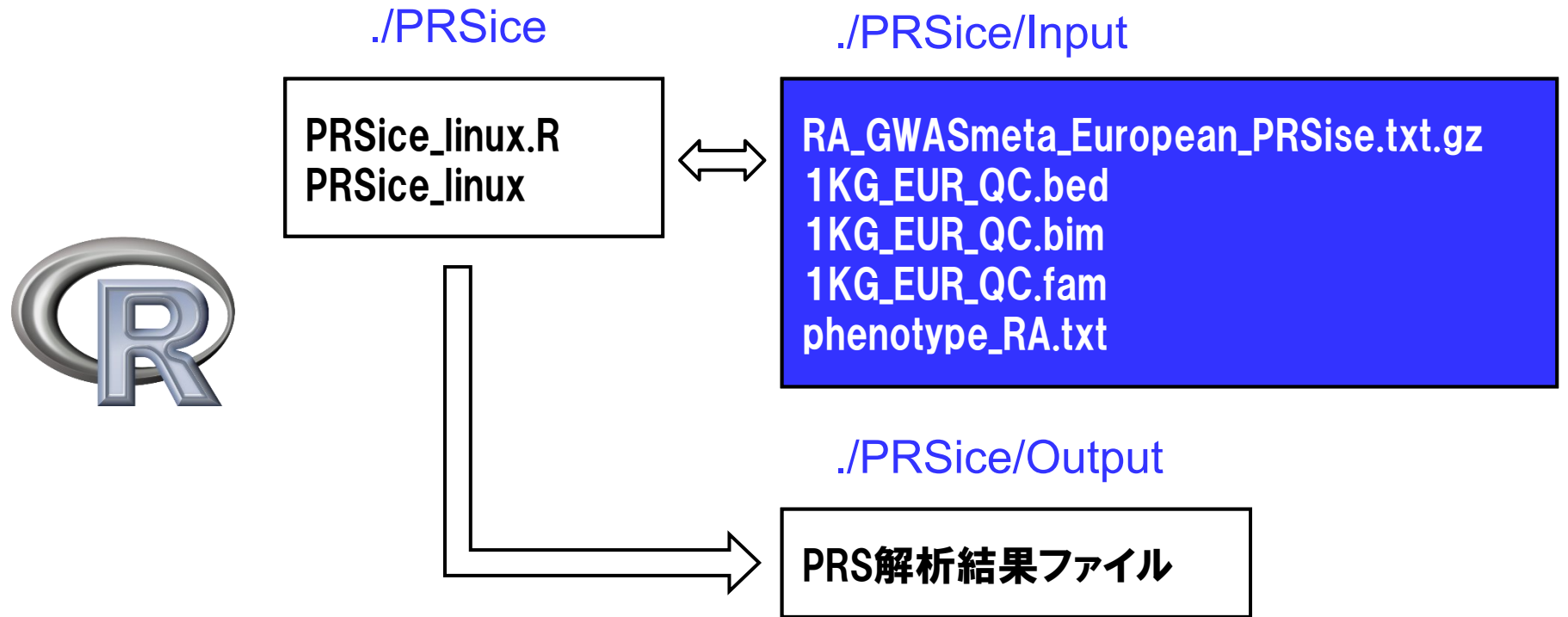
- PRSice-2の実施には、複数のファイル群を扱う必要があります。
- 各ファイル群を異なるフォルダに配置し、お互いを参照しながら解析を行うことで、ファイルの整理が容易になります。
- 解析や作図の都合上、Rを使用するので、Linux(WSL2, Cygwin)環境においてR(およびggplot2等の追加パッケージ)の事前インストールが必要です。

## ② PRSice-2を使ったPRS解析



- PRSice-2のメインの実行ファイルは、R言語で書かれたソースコードである、“`PRSice_linux.R`”になります。
- Linuxのコマンドライン上で`PRSice_linux.R`を実行すると、内部で解析ソフトウェア“`PRSice_linux`”が呼び出され、各種入力ファイルを読み込んでPRS解析を実施します。

## ② PRSice-2を使ったPRS解析



- 入力ファイルには、予測モデル構築に使うGWAS結果(Base Data)とPRS推定の対象となるGWASデータ(Target Data)の2種類があります。
- 前者に関節リウマチのGWAS結果(RA\_GWASmeta\_European\_PRSice.txt.gz)を、後者に1KG欧米人集団(1KG\_EUR\_QC.bed/bim/fam)を採用しました。
- TargetのGWASデータでも形質情報が必要です。今回は、シミュレーションで疑似的な関節リウマチ罹患情報を作成しました(phenotype\_RA.txt<sup>38</sup>)。

## ② PRSice-2を使ったPRS解析

### PRSice-2の入力ファイル

予測モデル構築に使うGWAS結果(Base Data)  
”RA\_GWASmeta\_European\_PRSice.txt.gz”

TargetのGWASデータの形質情報  
”phenotype\_RA.txt”



```
1 SNP>CHR>BP>A1>A2>P>OR↓
2 chr1:751343>1>751343>A>T>0.01>0.85↓
3 chr1:751756>1>751756>T>C>0.01>1.17↓
4 rs3094315>1>752566>A>G>0.0093>1.14↓
5 rs3131972>1>752721>A>G>0.009>0.88↓
6 rs3131971>1>752894>T>C>0.01>0.87↓
7 chr1:753405>1>753405>A>C>0.01>1.17↓
8 chr1:753425>1>753425>T>C>0.0097>0.87↓
9 rs2073814>1>753474>G>C>0.009>1.14↓
10 rs2073813>1>753541>A>G>0.0095>0.85↓
```



```
1 HG00096>HG00096>2↓
2 HG00097>HG00097>1↓
3 HG00099>HG00099>1↓
4 HG00100>HG00100>1↓
5 HG00101>HG00101>2↓
6 HG00102>HG00102>2↓
7 HG00103>HG00103>1↓
8 HG00104>HG00104>1↓
9 HG00106>HG00106>1↓
10 HG00108>HG00108>1↓
```

- ”SNP” : SNP ID
- ”CHR” : 染色体番号
- ”BP” : 染色体上の位置(base pair)
- ”A1” : アレル1
- ”A2” : アレル2
- ”P” : P値
- ”OR” : オッズ比(アレル1)

- 第1列目 : Family ID
- 第2列目 : Sample ID
- 第3列目 : Phenotype

※疑似形質であり、実際のサンプルの臨床情報とは無関係です。

## ② PRSice-2を使ったPRS解析

statgen@statgen-PC: ~

```
$ cd /mnt/c/SummerSchool/GenomeDataAnalysis5/PRSice/
```

※Cygwinの場合 /mnt/を/cygdrive/に変えてください

※Macの場合 資料を置いたフォルダを指定してください

statgen@statgen-PC: /mnt/c/SummerSchool/GenomeDataAnalysis5/PRSice

```
$ Rscript PRSice_linux.R --prsice PRSice_linux --dir . --out ./Output/1KG_EUR-
```

```
RAGWAS --base ./Input/RA_GWASmeta_European_PRSice.txt.gz --
```

```
target ./Input/1KG_EUR_QC --pheno ./Input/phenotype_RA.txt --thread 1 --stat OR --
```

```
binary-target T
```

※ファイル”PRSice\_Command.txt”を開いて、内容をLinuxコマンドにコピー&ペーストして下さい。

※Cygwinの場合、”PRSice\_Command\_cygwin.txt”を使用してください。

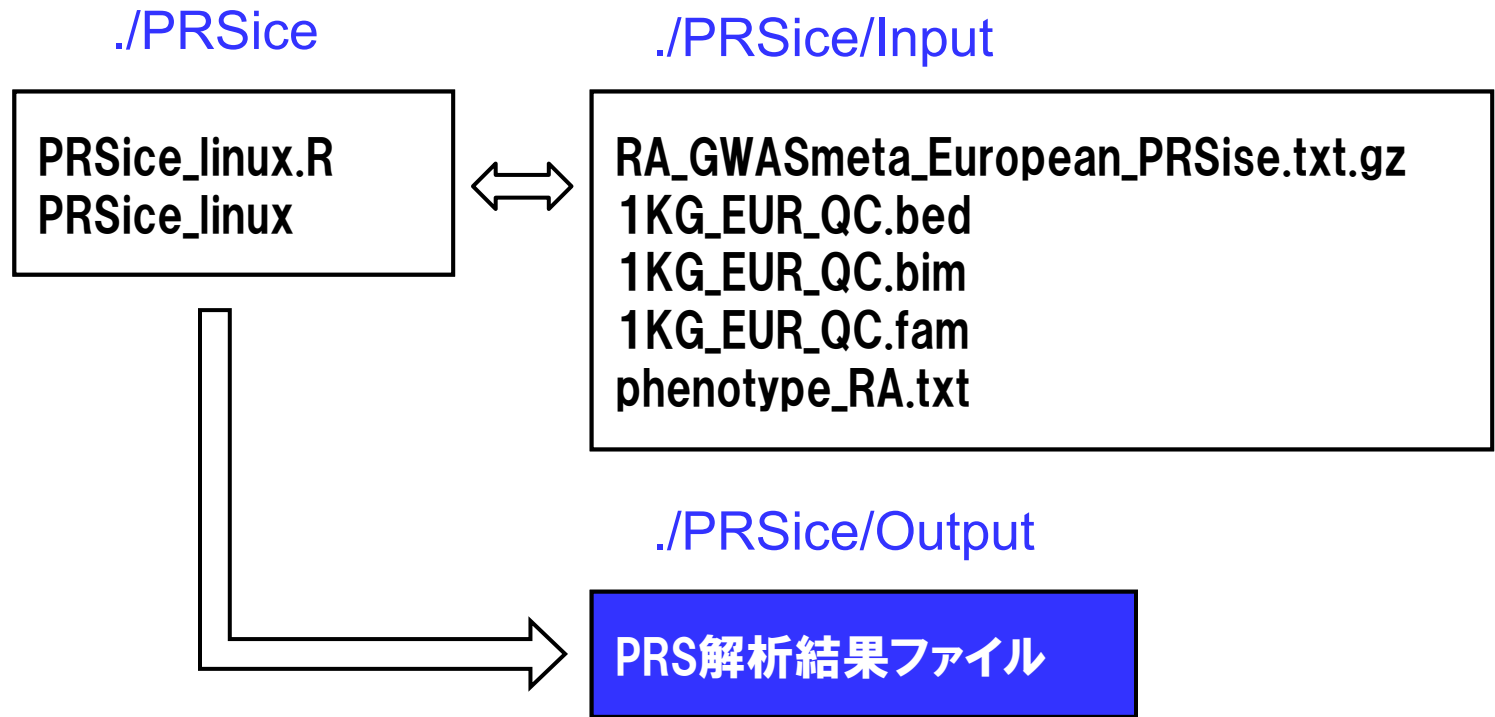
※Macの場合、”PRSice\_Command\_mac.txt”を使用してください。

• PRSice-2は、”Rscript PRSice\_linux.R --prsice (PRSice実行ファイル) --dir . --out (出力ファイル名) --base (Baseデータ) --target (Target Plinkファイル名) --pheno (Target 形質ファイル名) --thread (スレッド数) --binary-target (T or F)”という形で実行します。

• 各データの名前は、カレントディレクトリからの相対パスで表記可能です。



## ② PRSice-2を使ったPRS解析



- "XXX.prsice": 様々なP値の閾値におけるPRSモデルの結果
- "XXX.summary": 一番高精度だったP値の閾値におけるPRSモデルでの結果
- "XXX.best": 一番高精度だったP値の閾値における各サンプルのPRS値
- "XXX\_BARPLOT\_日付.png": P値の閾値を変化させた際の精度比較の図
- "XXX\_HIGH-RES\_PLOT\_日付.png": P値の閾値を変化させた際の精度比較の図
- "XXX.log": ログファイル

• PRSice-2は、P値の閾値を変化させながら最適なモデルを探索します。

## ② PRSice-2を使ったPRS解析

”XXX.prsice”：様々なP値の閾値におけるPRSモデルの結果

	A	B	C	D	E	F	G
1	Set	Threshold	R2	P	Coefficient	Standard.Error	Num_SNP
2	Base	5.00E-08	0.0122466	0.0630318	12.6059	6.78113	257
3	Base	5.01E-05	0.0126982	0.0583782	22.8194	12.0556	479
4	Base	0.00010005	0.01382	0.0484231	27.829	14.1004	561
5	Base	0.00015005	0.0139415	0.0474541	30.2877	15.2797	611
6	Base	0.00020005	0.0143518	0.0443572	33.0976	16.4607	658
7	Base	0.00025005	0.0146751	0.0420568	35.1525	17.2913	691
8	Base	0.00030005	0.015445	0.0370949	37.9226	18.1908	729

P値の閾値

PRSが説明  
可能な  
発症リスク

PRSと発症  
リスクの関連

”XXX.summary”：一番高精度だったP値の閾値におけるPRSモデルでの結果

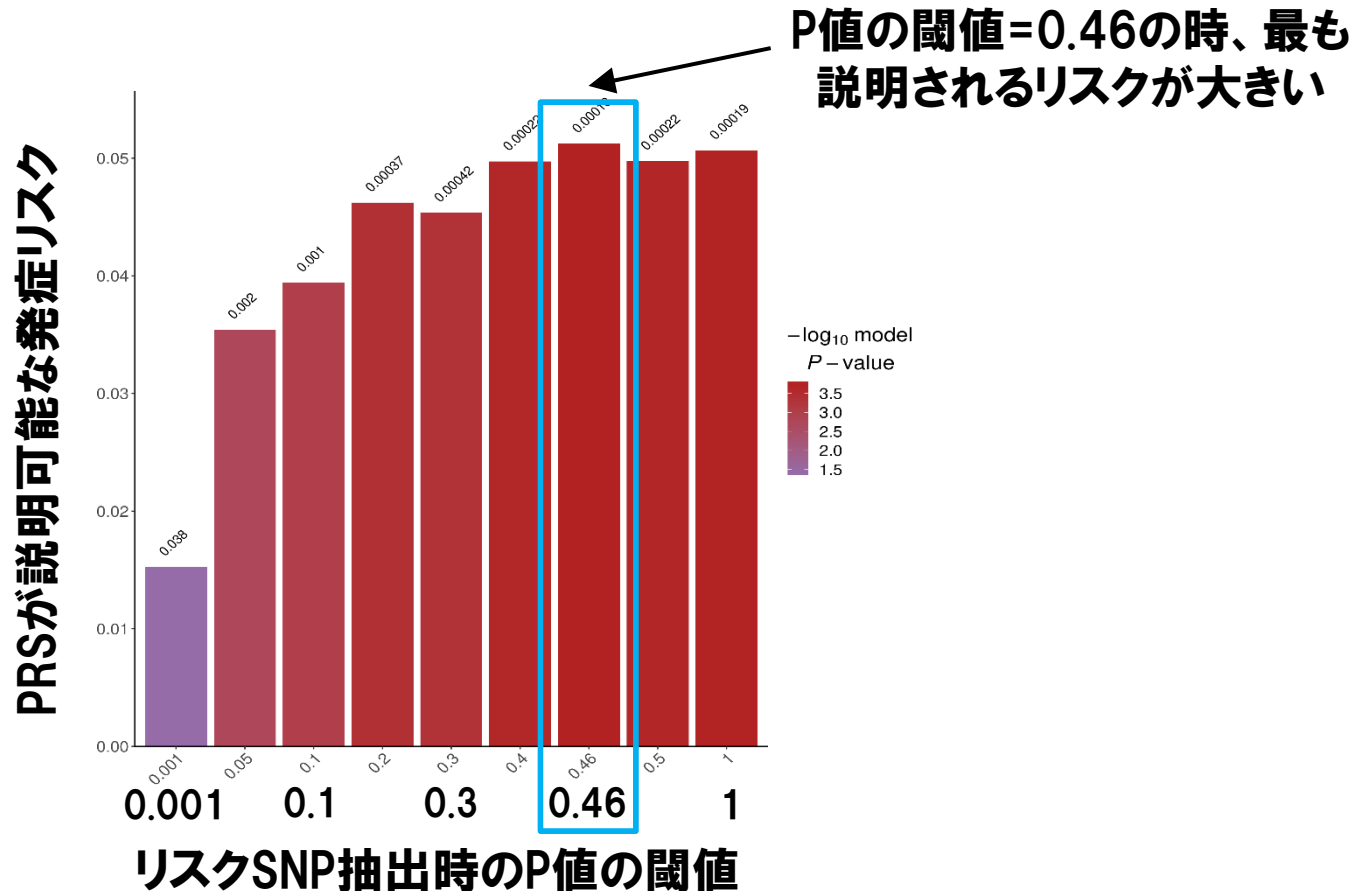
	A	B	C	D	E	F	G	H	I	J	K
1	Phenotype	Set	Threshold	PRS.R2	Full.R2	Null.R2	Prevalence	Coefficient	Standard.Error	P	Num_SNP
2	-	Base	0.46	0.0512432	0.0512432	0	-	5081.42	1356.18	0.000179059	101385

P値の閾値 PRSが説明  
可能な  
発症リスク

PRSと発症 リスク  
リスクの関連 SNP数

## ② PRSice-2を使ったPRS解析

”XXX\_BARPLOT\_日付.png”：P値の閾値を変化させた際の精度比較の図



- P値の閾値を0.46に設定した際のPRSモデルが、Target Dataにおいて対象形質との関連を最も有意に示すことがわかりました  
(シミュレーションによる疑似形質なので、実際には形質との関連は存在していません。)

## ② PRSice-2を使ったPRS解析

“XXX.best”：一番高精度だったP値の閾値における各サンプルのPRS値

```
1 FID IID In Regression PRS
2 HG00096 HG00096 Yes -0.000627027476
3 HG00097 HG00097 Yes -0.00058856461
4 HG00099 HG00099 Yes -0.000713382756
5 HG00100 HG00100 Yes -0.000571751231
6 HG00101 HG00101 Yes -0.000530233701
7 HG00102 HG00102 Yes -0.000583765514
```

\$1:Family ID    \$2:Individual ID    \$4:PRS値

```
statgen@statgen-PC: /mnt/c/SummerSchool/GenomeDataAnalysis5/PRSice
```

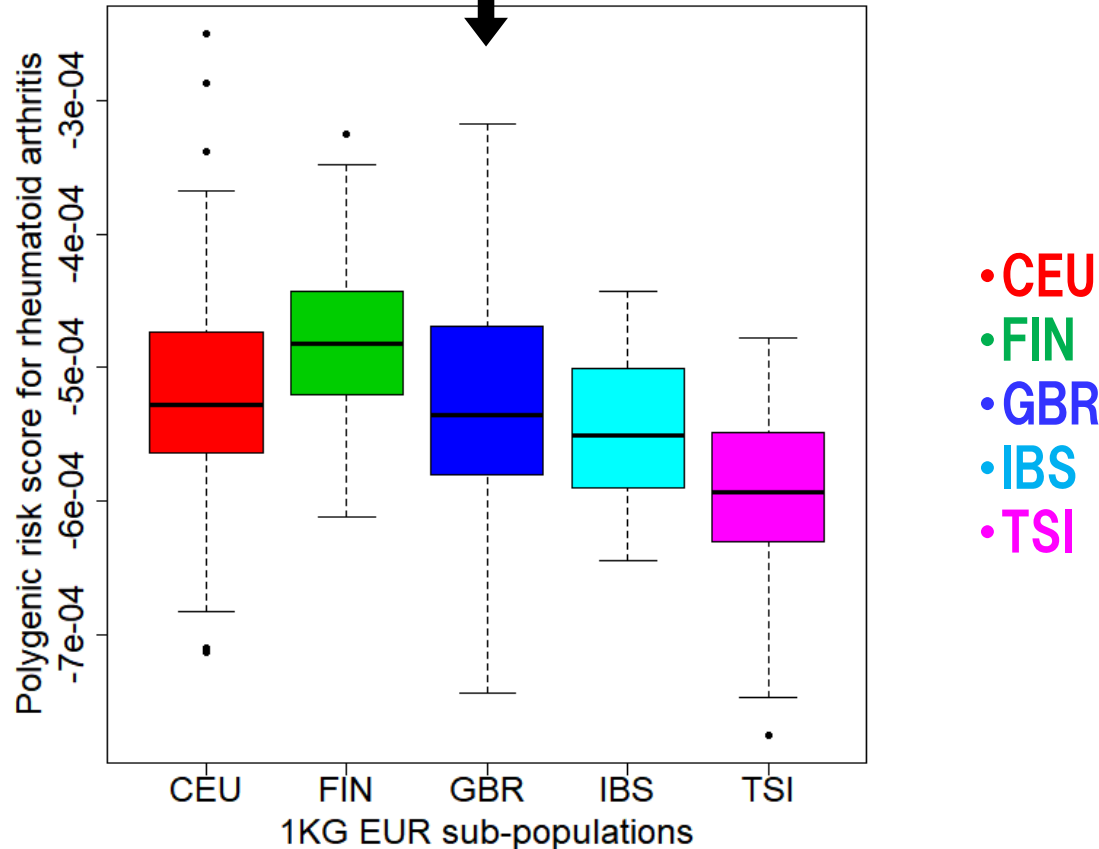
```
$ awk '$1!="FID" && $3="Yes" {print $1"¥t"$2"¥t"$4}' 1KG_EUR-RAGWAS.best > 1KG_EUR-RAGWAS.best.txt
```

- 一番高精度だったP値の閾値での、各サンプルのPRS値が出力されます。
- サンプル別PRS値が、出身地域によって異なるか調べてみましょう。
- 今回は疑似形質を使用したので、出身地域ごとの差はないはずです。  
(疑似形質を生成したシミュレーションでは、地域情報は使用していません。)

## ② PRSice-2を使ったPRS解析

分散分析による出身地域間の  
PRS値の差の検定:  $P < 2.2 \times 10^{-16}$

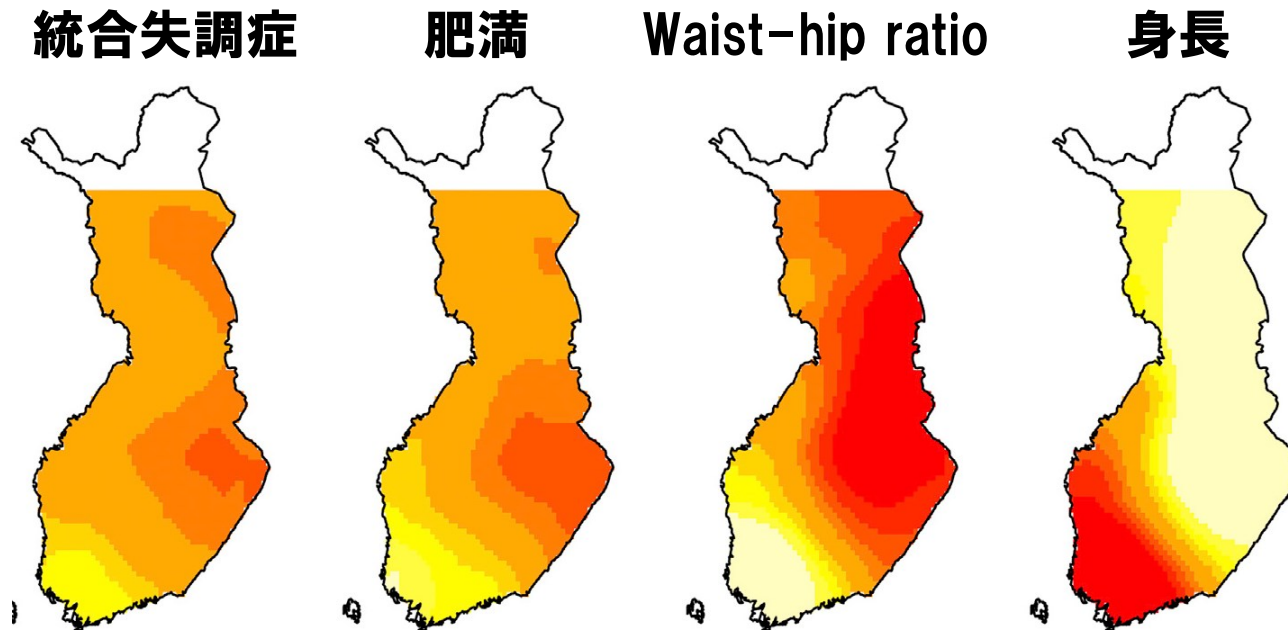
※ファイル”PlotPRSice.R”を開いて、内容をRに  
コピー&ペーストして下さい。



- (疑似形質を使用していることに加え) 今回の対象形質である関節リウマチでは、地域や人種集団による発症率の差は少ないことが知られています。
- しかし予想に反し、**出身地域間でPRS値の分布に差が生じていました**。<sup>45</sup>

## ② PRSice-2を使ったPRS解析

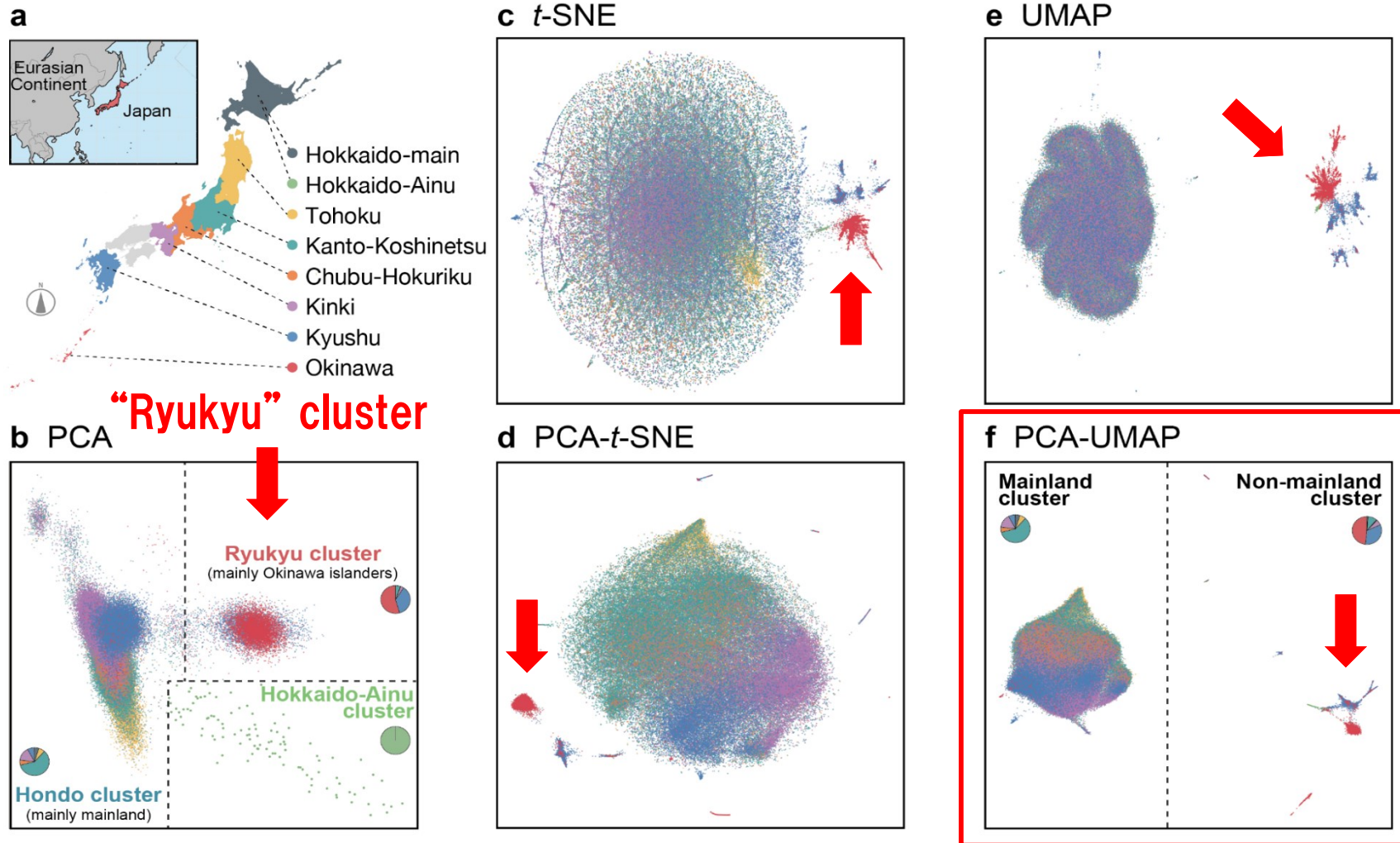
フィンランドバイオバンク(FinnGen)におけるPRS分布の地域差



- 最近の研究では、同じPRSモデルで計算をしても、**PRS値の分布が集団間(および同一集団内の異なる地域)で異なることが指摘されています。**
- 疾患発症率の地域差なのか、集団構造化による地域間の遺伝的背景の違い(=リスクSNPのアレル頻度分布の差)なのか、**結論が出ていません。**
- PRSを臨床現場に導入するには、**まだまだ検討すべきことが多そうです。**

### ③ 疾患病態解明・ゲノム創薬への展望

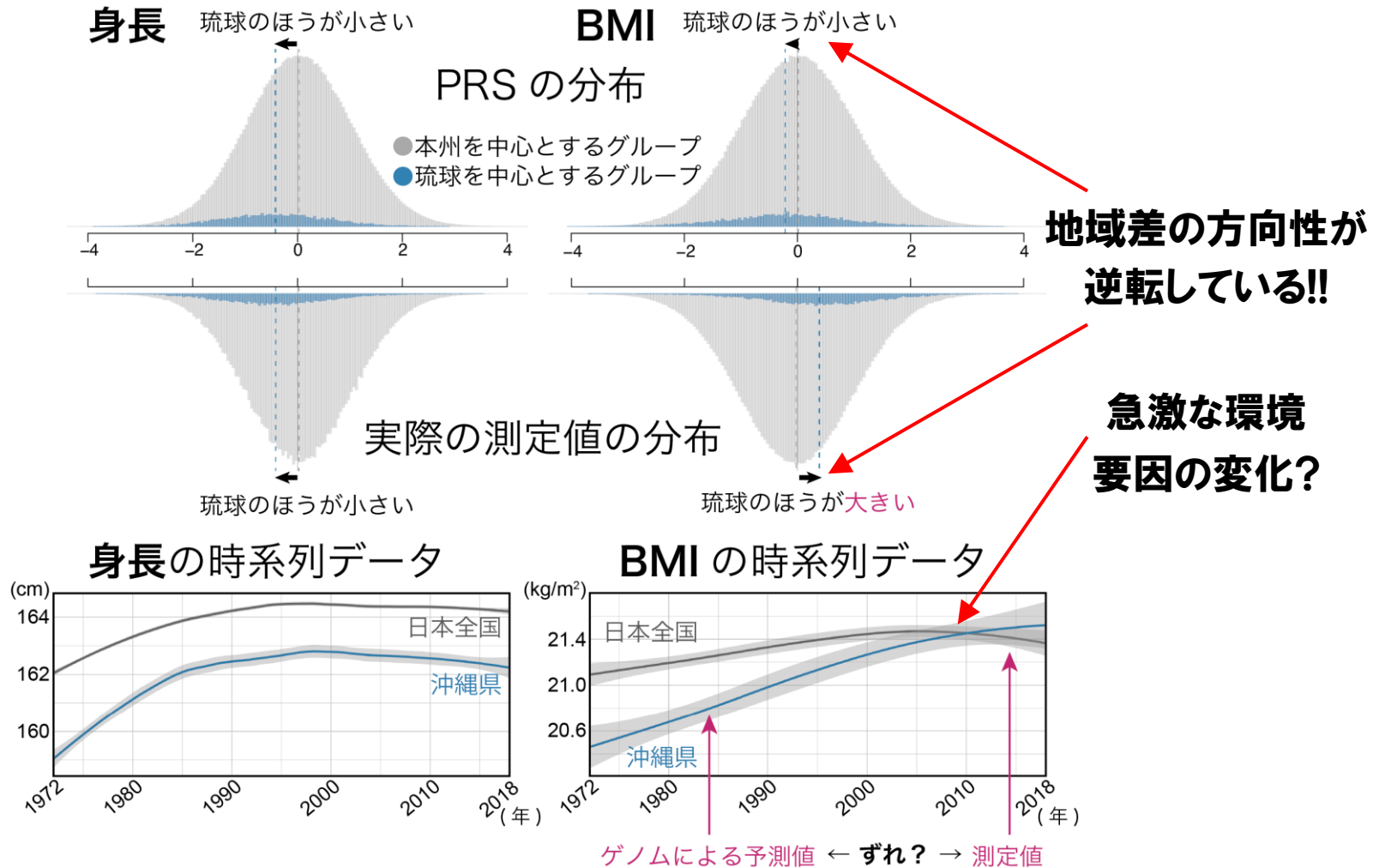
## 線形-非線形機械学習による日本人集団構造化の解明



• 線形-非線形のハイブリッド型機械学習により、日本人集団の詳細な構造化の解明(“Hondo”, “Ryukyu”, “Hokkaido-Ainu”)が可能になりました。

### ③ 疾患病態解明・ゲノム創薬への展望

## 日本人集団の構造化がゲノム個別化医療に与える影響



- 本土・琉球クラスター間において、**肥満の実測値とPRS予測値が逆の方向性を示す**ことが判明。沖縄における急激な環境要因の変化をゲノム情報が網羅できていない可能性が浮上しました。

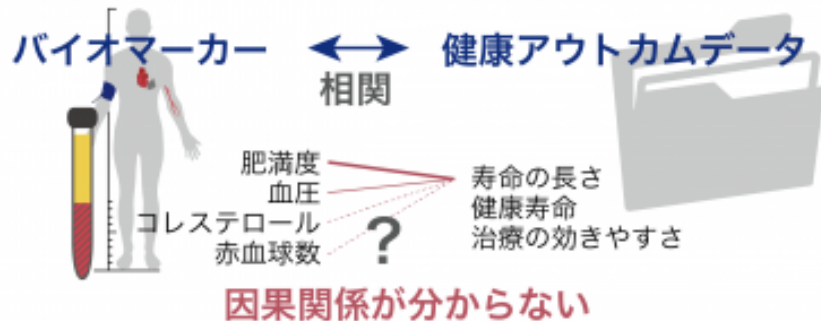
(Sakaue S et al. *Nat Commun* 2020)



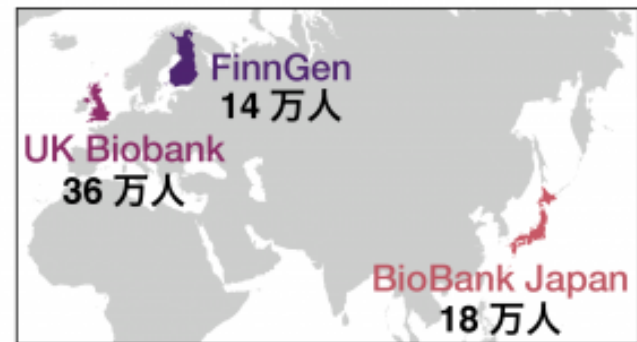
### ③ 疾患病態解明・ゲノム創薬への展望

## Polygenic risk scoreを活用した健康寿命バイオマーカー探索

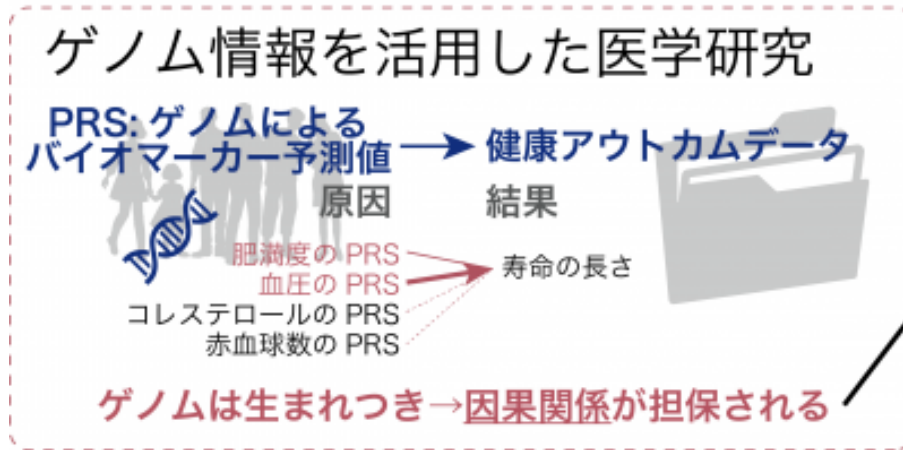
これまでの医学研究



世界3カ国 70万人のゲノムデータに応用



治療により健康アウトカムの改善が期待できるバイオマーカーを特定



- 国際バイオバンク67万人のpolygenic risk score (PRS)解析を実施。健康の指標バイオマーカーPRSと寿命との因果関係を横断的に調べる手法を開発。
- 高血圧・肥満が健康寿命を短縮し、その治療が健康寿命延伸に貢献する可能性を示すことができました。

(Sakaue S and Kanai M et al. *Nat Med* 2020)

## 終わりに

- ゲノム個別化医療に向けた取り組みについて、具体的な解析手法の紹介とともに、簡単になぞってみました。
- 大規模疾患ゲノムデータが構築され、ゲノム個別化医療の実現に社会からの期待が高まっています。
- これまでヒトゲノム研究に多額の公的研究費が投入されてきた点を考慮すると、社会還元という点からも、ゲノム個別化医療の実装は重要です。
- 大規模GWAS結果の活用や解析手法の改善により、PRSを用いたゲノム個別化医療への期待が高まっています。
- しかし、PRS分布には様々な理由でバイアスが生じる可能性も指摘されており、さらなる解析手法の改善が必要と考えられます。