

遺伝統計学の世界へ ようこそ

岡田 随象

大阪大学大学院医学系研究科
遺伝統計学

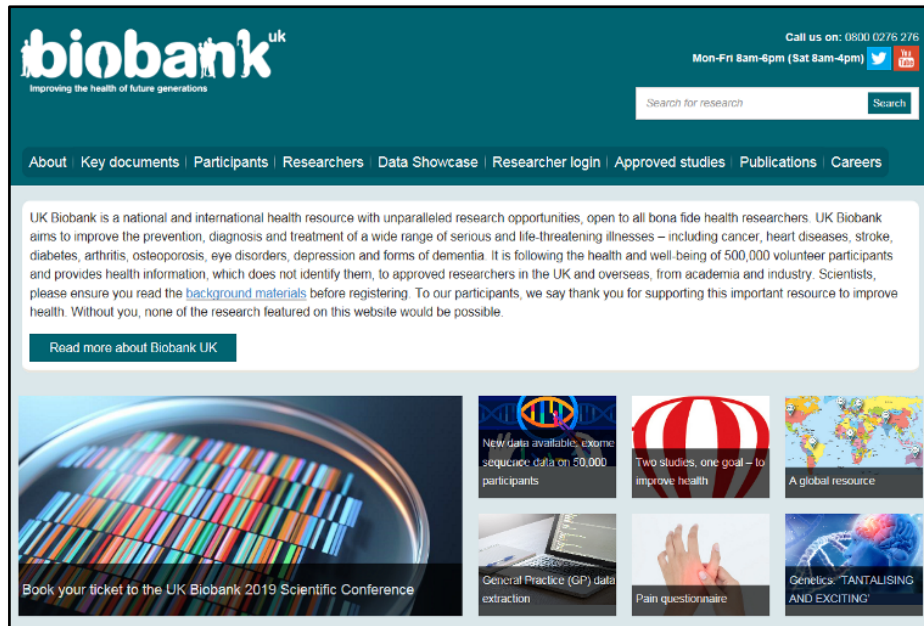
URL : <http://www.sg.med.osaka-u.ac.jp/index.html>



ヒトゲノム研究の大規模化

UKバイオバンク

(<https://www.ukbiobank.ac.uk/>)



The screenshot shows the UK Biobank website. At the top, the logo "biobank^{uk}" is displayed with the tagline "Improving the health of future generations". Below the logo is a search bar and a "Search" button. A navigation menu includes links for "About", "Key documents", "Participants", "Researchers", "Data Showcase", "Researcher login", "Approved studies", "Publications", and "Careers". The main content area features a paragraph about the bank's mission and a "Read more about Biobank UK" button. Below this are several featured images and text boxes: a DNA helix, a globe, a hand holding a DNA helix, and a DNA helix with the text "Genetics: 'TANTALISING AND EXCITING'".

TopMedプロジェクト

(<https://www.nhlbiwgs.org/>)

Trans-Omics for Precision Medicine (TOPMed) Program

What is the goal of the TOPMed program?

The goal of the TOPMed program is to generate scientific resources that will improve the understanding of heart, lung, blood, and sleep disorders and advance precision medicine. Precision medicine is an emerging approach to disease prevention and treatment that considers the unique genes and environment of each patient.

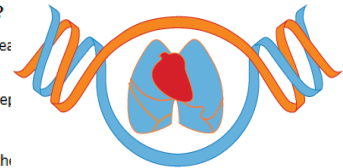
The TOPMed program collects whole-genome sequencing and other -omics data. In biology, -omics refers to measurable differences or changes in biological molecules, such as genes, metabolites, proteins, and RNA. The program will integrate -omics data with molecular, behavioral, imaging, environmental, and clinical data to improve the prevention and treatment of heart, lung, blood, and sleep disorders.

The TOPMed program will specifically support the National Heart, Lung, and Blood Institute (NHLBI) [Precision Medicine Activities](#). The program also complements the National Institutes of Health (NIH) Precision Medicine Initiative and [All of Us Research Program](#), which will collect data from one million or more people to help study a range of health issues and diseases.

How does the TOPMed program contribute to scientific discoveries?

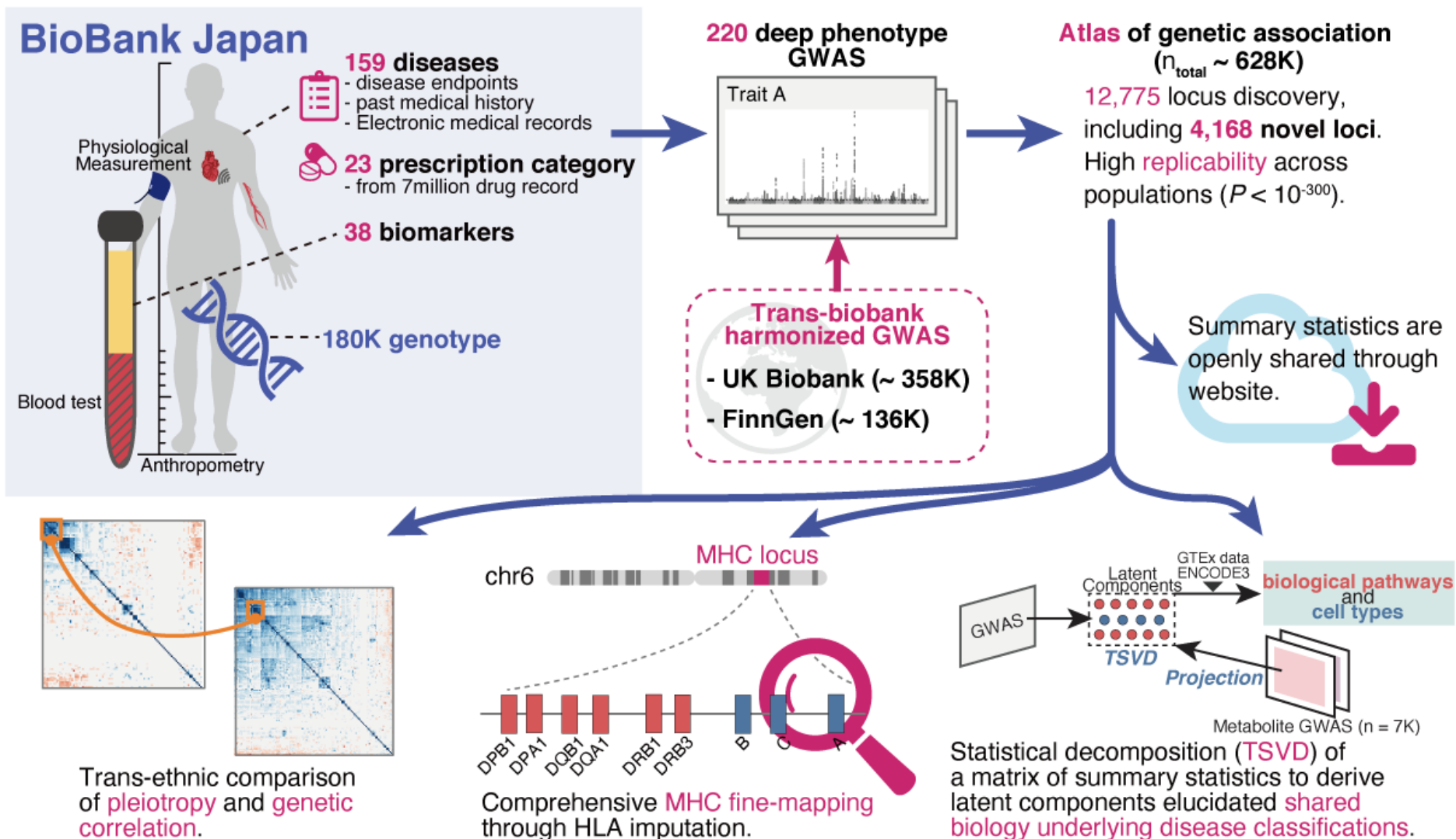
The TOPMed program will support research that furthers our understanding of how the program may lead to future scientific discoveries such as:

- Biomarkers that increase or decrease the risk of heart, lung, blood, and sleep
- Interactions between the environment and genes that affect health
- Potential targets for new treatments
- New ways to define heart, lung, blood, and sleep disorders or subtypes of the



- ヒトゲノム研究の大規模化を引き起こしています。
- UKバイオバンクは50万人のSNPデータと20万人の全エクソーム、TopMedプロジェクトは10万人の全ゲノムシーケンスを構築済みです。
- UKバイオバンクは、ゲノム・表現型情報を安価に分譲しています。

国際バイオバンク連携によるGWASメタ解析

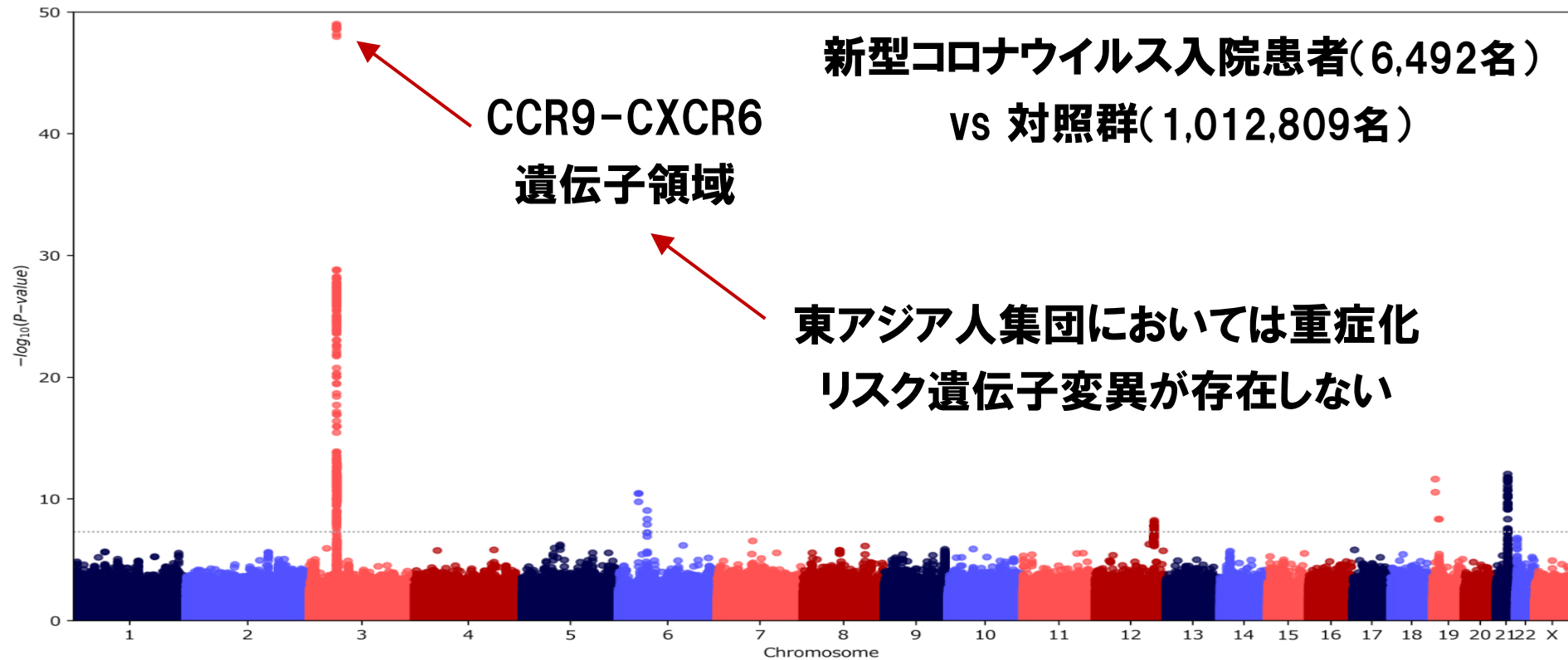


・国際バイオバンク連携63万人・220形質のGWASメタ解析を実施。

・全GWAS統計量を <https://pheweb.jp/> にて一般公開中。

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

国際コンソーシアムによる重症化ゲノム解析

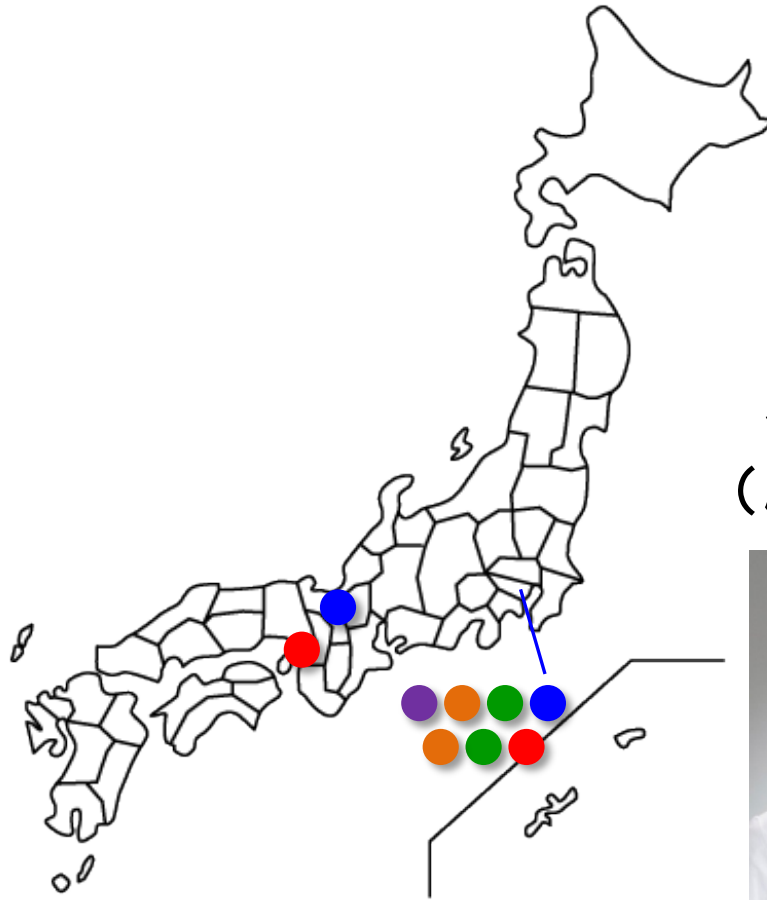


- COVID-19 Host Genetics Initiativeにより、**新型コロナウイルス感染者の重症化リスクに関するGWAS**が実施され、欧米人集団において、CCR9-CXCR6遺伝子領域が強いリスクを有することが明らかに。
- CCR9-CXCR6領域リスク遺伝子変異は東アジア人集団に存在せず、**日本人集団のCOVID-19ホストゲノム解析が不可欠**。

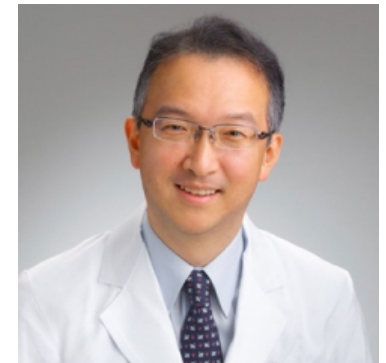
(The COVID-19 Host Genetics Initiative. *medRxiv* 2021)

コロナ制圧タスクフォースの結成

- ・慶應義塾大学
- ・東京医科歯科大学
- ・東京大学医科学研究所
- ・京都大学
- ・大阪大学
- ・北里大学
- ・120の関連協力病院



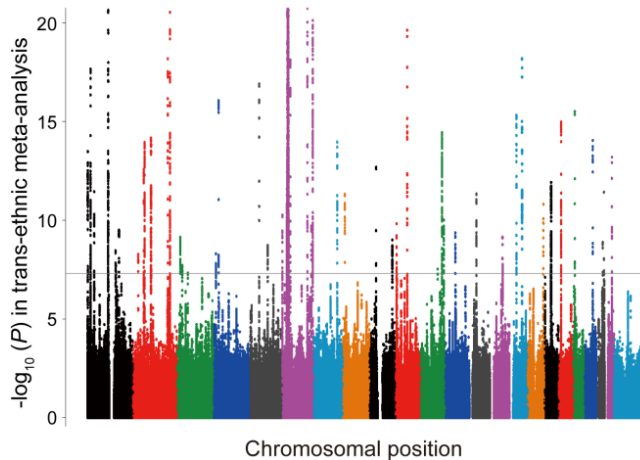
金井隆典先生
(慶應義塾大学)



- ・日本人集団における新型コロナウイルスの制圧を目指し、2020年5月に**有志連合:コロナ制圧タスクフォース**を結成。
- ・感染者から、ゲノム・RNA・血漿サンプルを収集を実施中。

ゲノム解析の成果をどのように活用すれば 病態解明・創薬・個別化医療に貢献できるの ほとんどわかっていない

Human Genetics

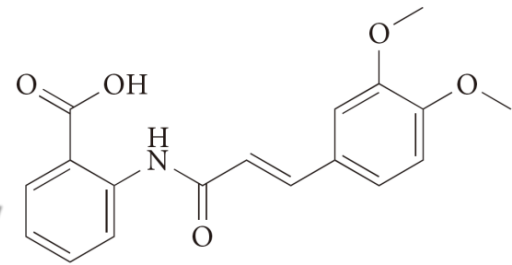


Trans-ethnic GWAS of rheumatoid arthritis identified 101 risk loci
(Okada Y. *Nature* 2014)



**Disease
biology**

**Drug
discovery**



**Personalized
Medicine**

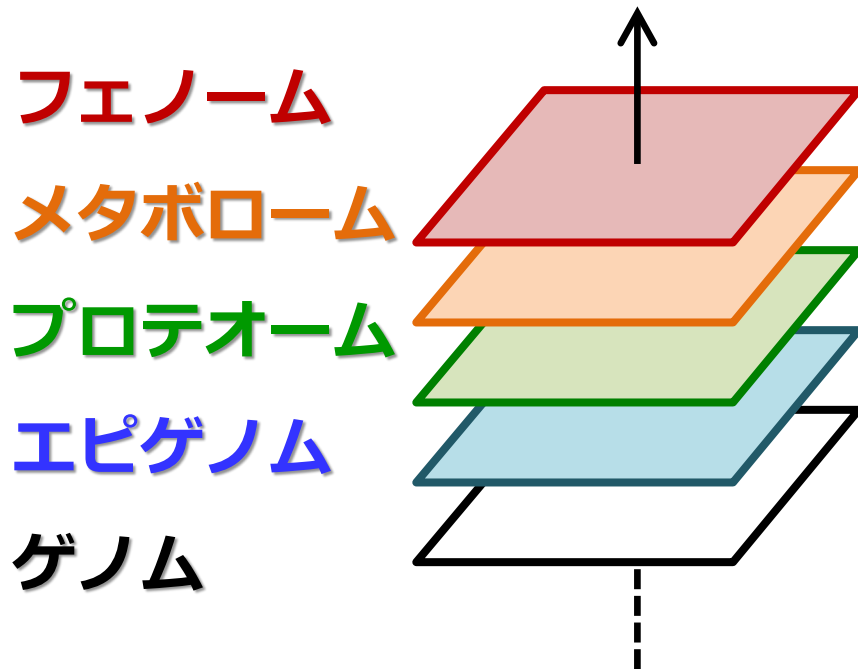
講演内容

- **横断的オミクス解析による病態解明**
- ヒトゲノム情報への人工知能の活用
- ゲノム個別化医療への挑戦

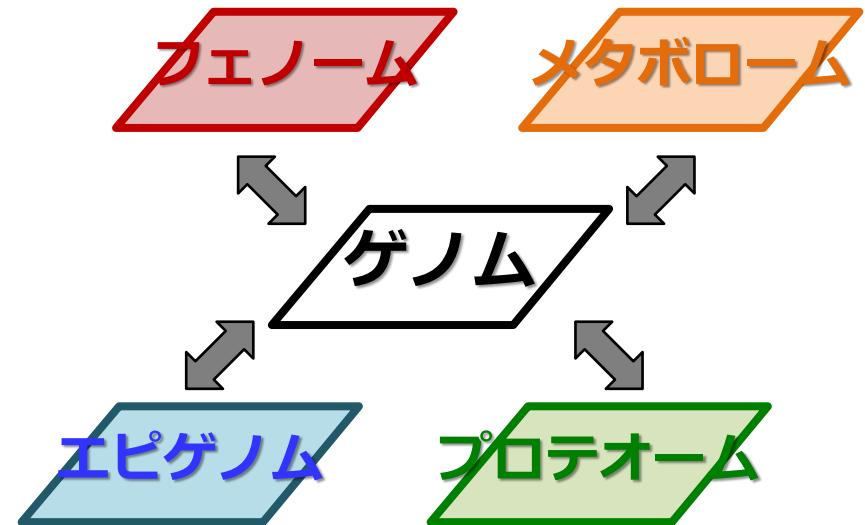


多層的 vs 横断的 オミクス解析

多層的オミクス解析



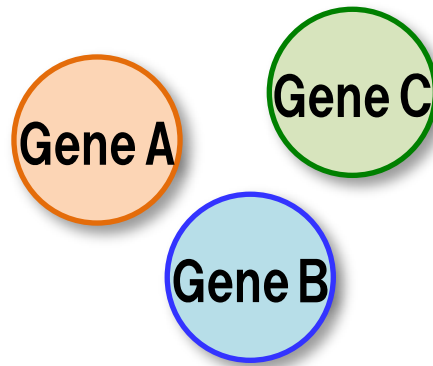
横断的オミクス解析



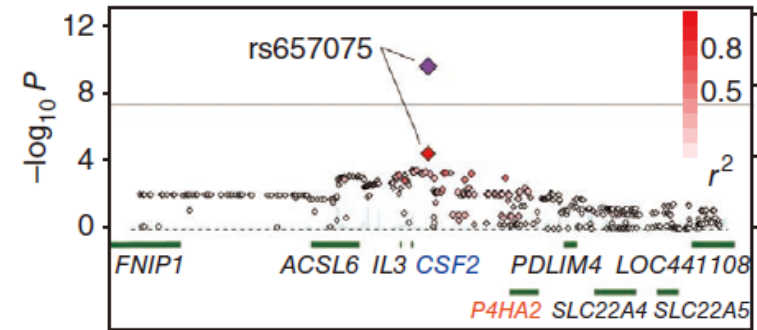
	多層的オミクス解析	横断的オミクス解析
網羅性	×	○
コスト	△	○
サンプル数	×	○
予測性	○	△

横断的オミクス解析の解析方法

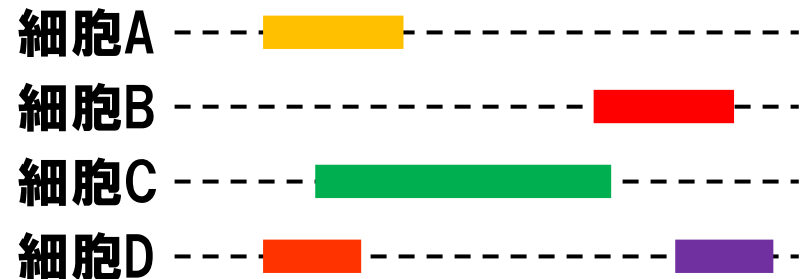
遺伝子発現
情報



バックアウトマウス
表現型スクリーニング

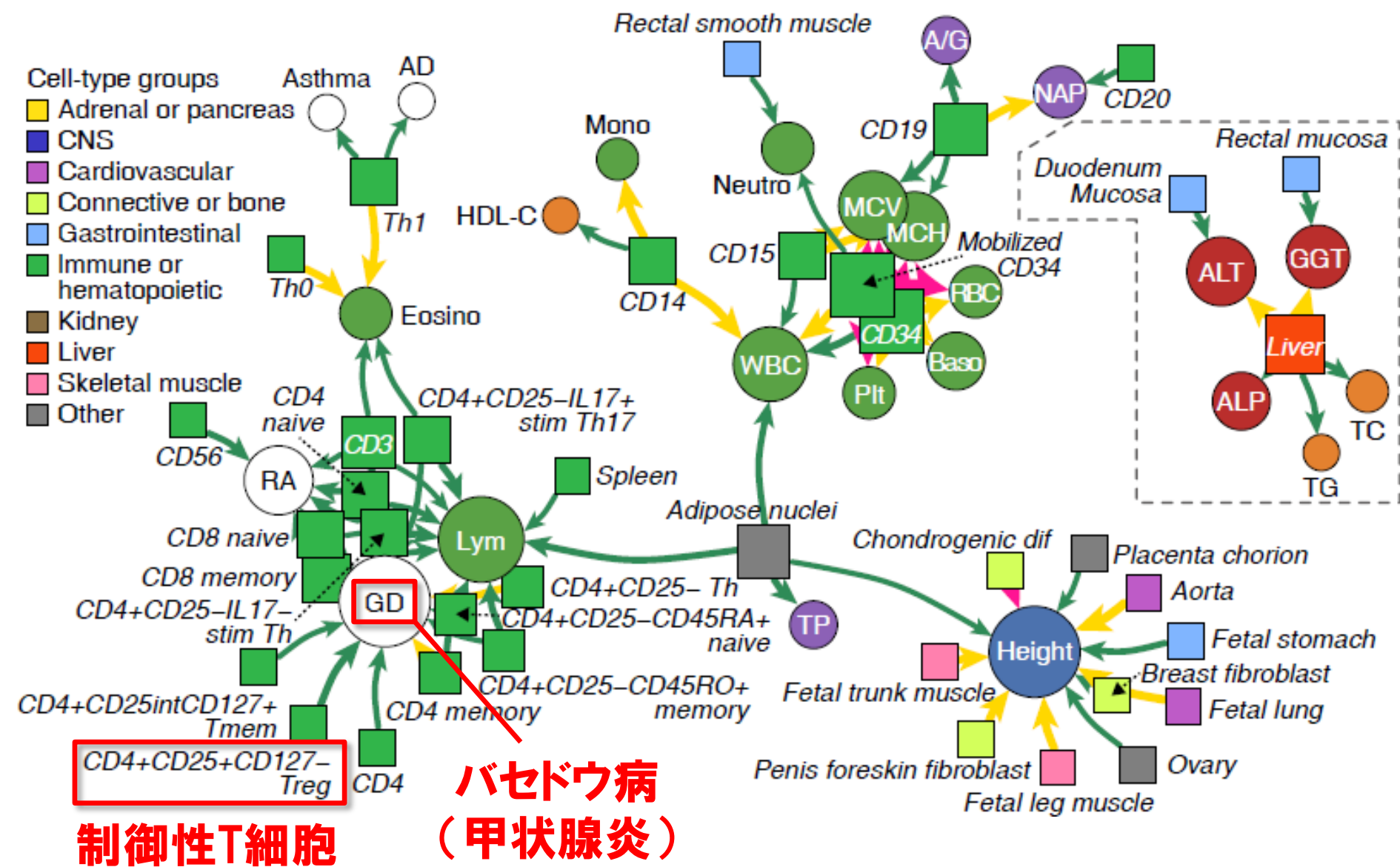


ヒトゲノム標準配列



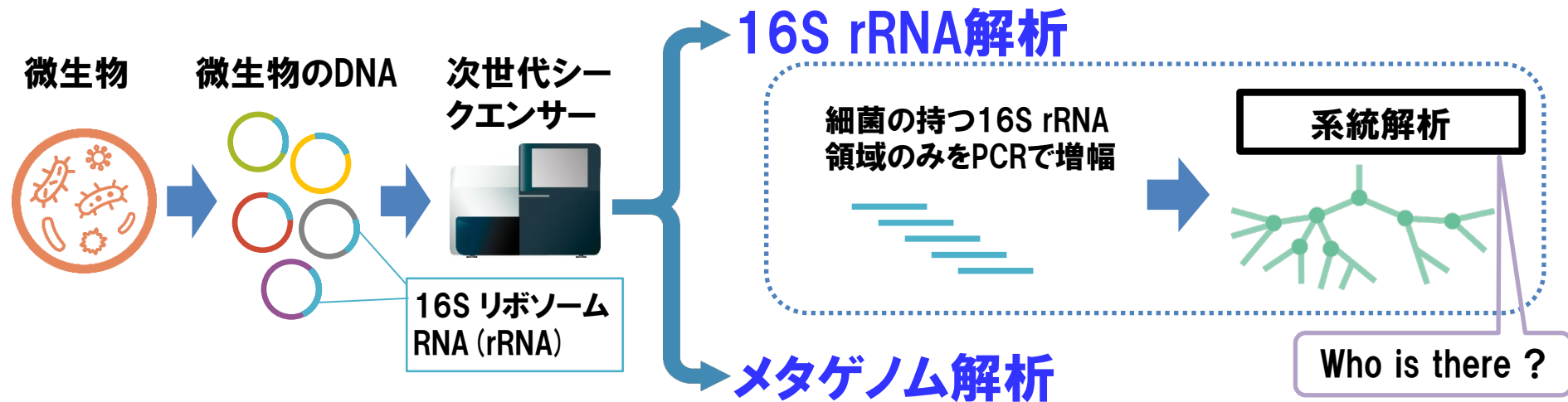
- 異なるサンプル集団由来のオミクス情報を**共通の情報単位**に変換することで、横断的な統合が可能になる。
- 近年、**ヒトゲノム標準配列に沿ったベクトル情報**に変換する横断的オミクス解析方法の開発がさかんである。

複数疾患の遺伝的背景の網羅的比較

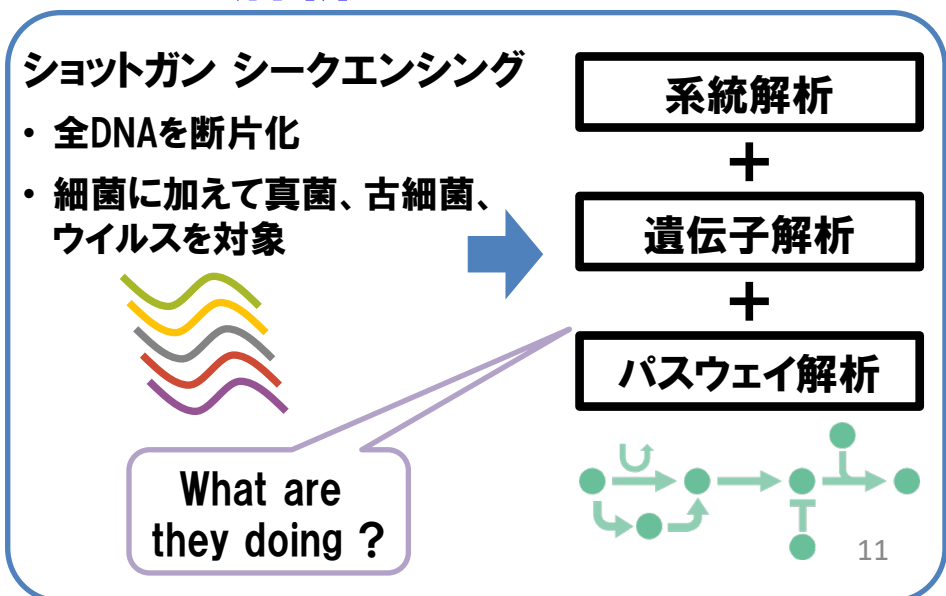
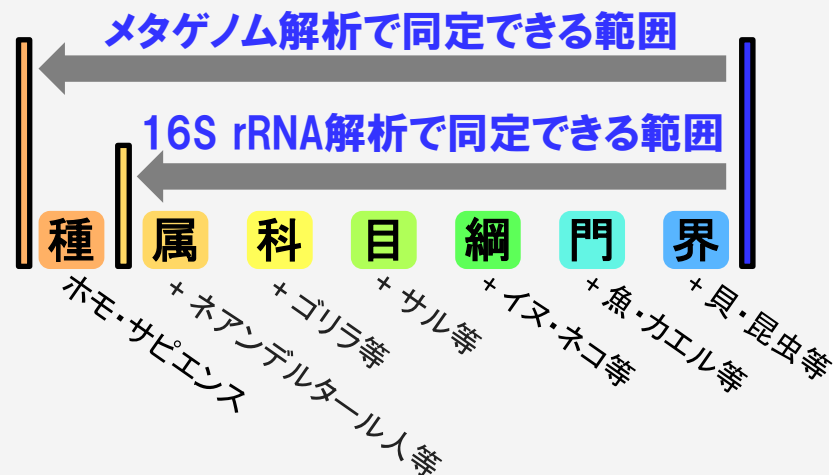


16S rRNA解析とメタゲノムショットガン解析

- ・環境中に含まれる**全微生物のゲノム情報の集合**に対する網羅的な解析。
- ・細菌のゲノムの一部を対象とする16S rRNA解析に代わる新しい手法。

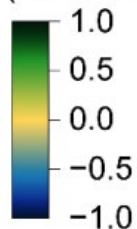


生物の階級・各解析の対象範囲

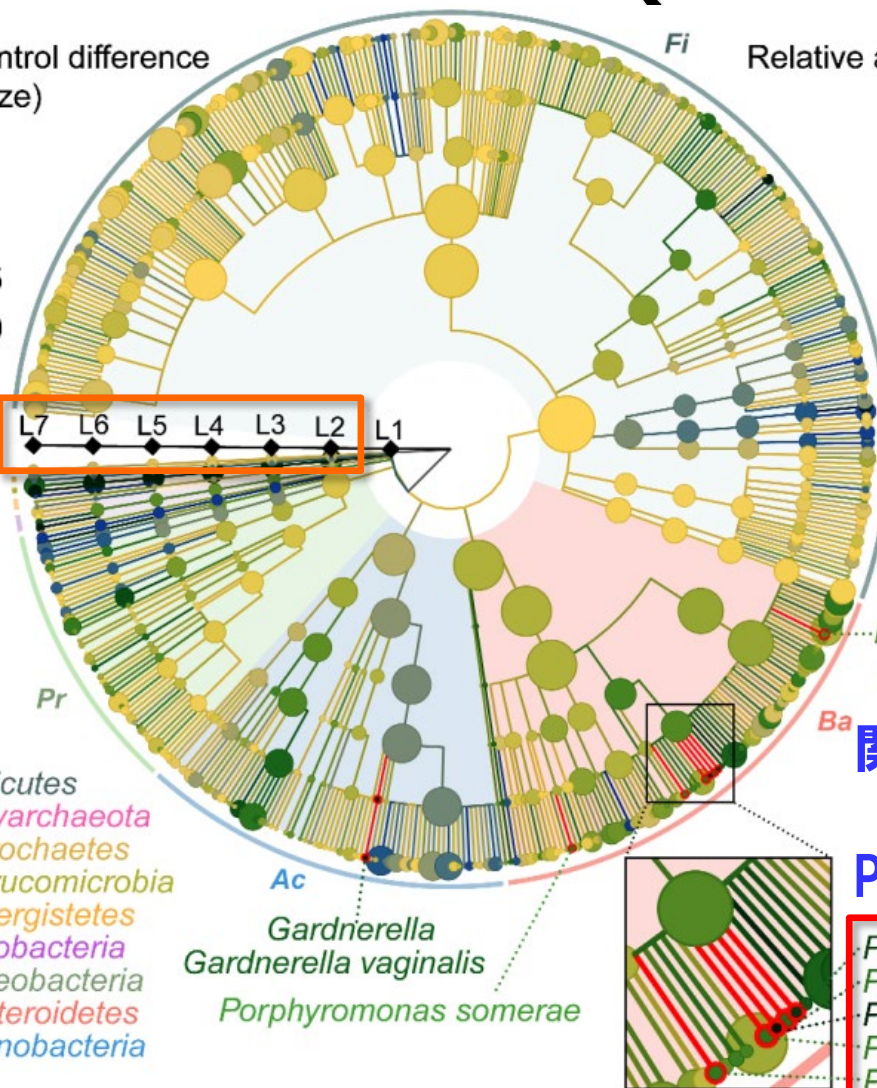


メタゲノムワイド関連解析(菌種系統解析)

Case-control difference
(effect size)



Relative abundance (%)



関節リウマチ患者の
腸内細菌叢で
Prevotella属が増加

Prevotella marshii
Prevotella disiens
Prevotella denticola
Prevotella corporis
Prevotella amnii

生物の階級・各解析の対象範囲

メタゲノム解析で同定できる範囲

16S rRNA解析で同定できる範囲

種 属 科 目 綱 門 界
+ ホモ・サピエンス
+ ネアンデルタール人等
+ ゴリラ等
+ サル等
+ イヌ・ネコ等
+ 魚・カエル等
+ 貝・昆虫等

・微生物叢ショットガンシーケンスに基づくメタゲノムワイド関連解析により、関節リウマチに特異的な菌種・代謝物・パスウェイを同定。

(特願2019-156566, Kishikawa T et al. *Ann Rheum Dis* 2020a)

ゲノム創薬の新しいモデル

十年～数十年の歳月

ヒット
化合物

24.3

第I相
臨床試験

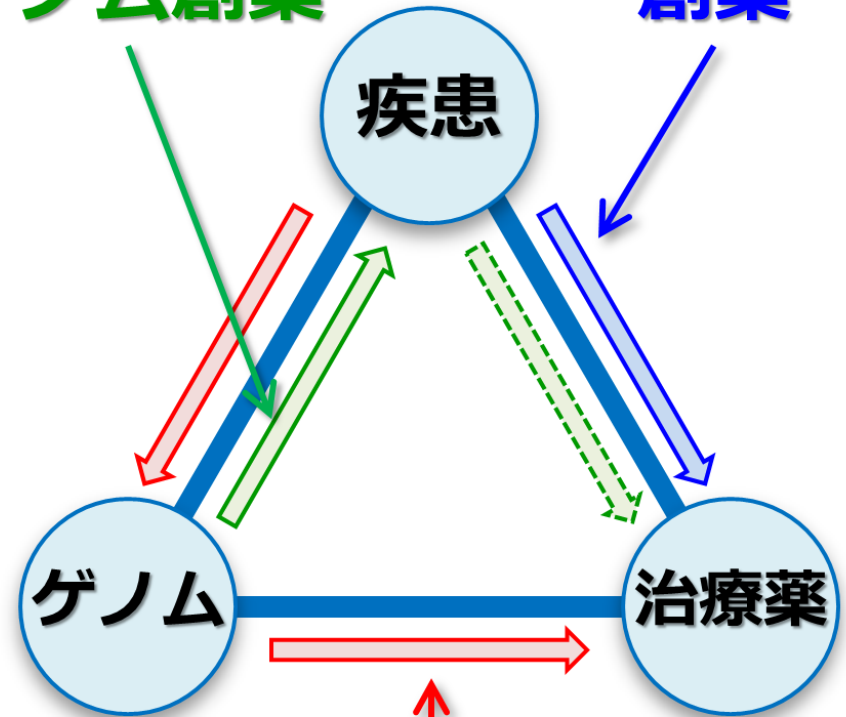
8.6

上市

1

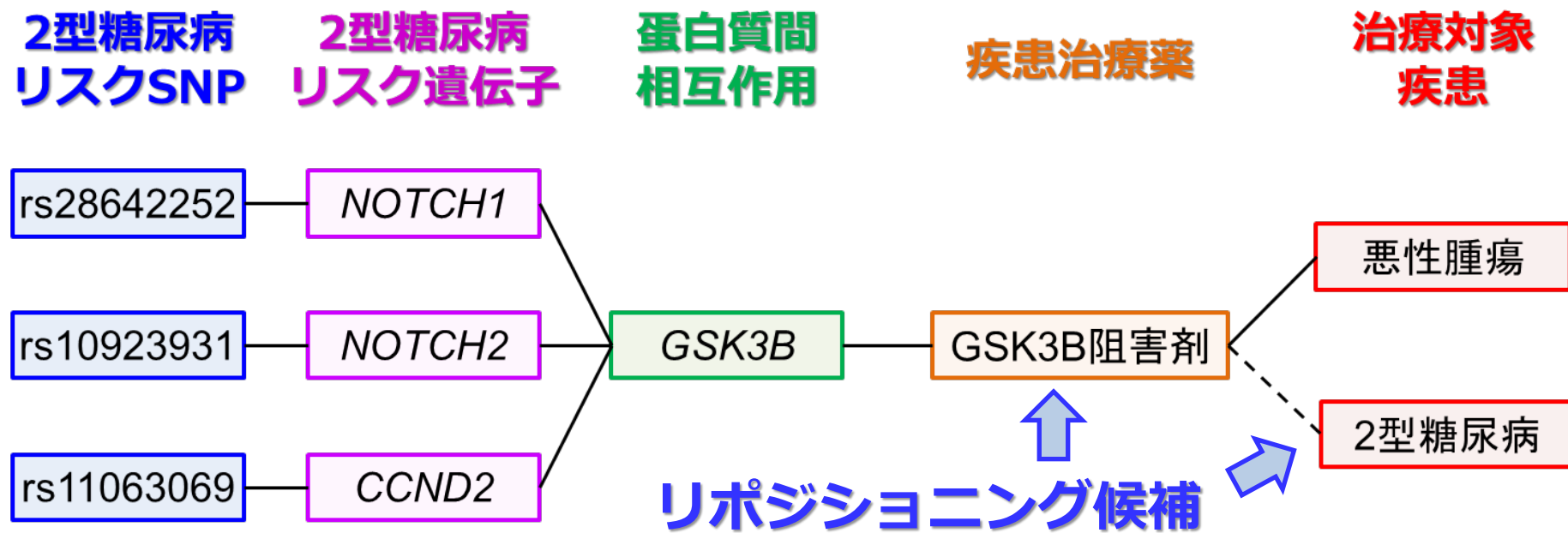
これまでの
ゲノム創薬

これまでの
創薬



これからの
ゲノム創薬

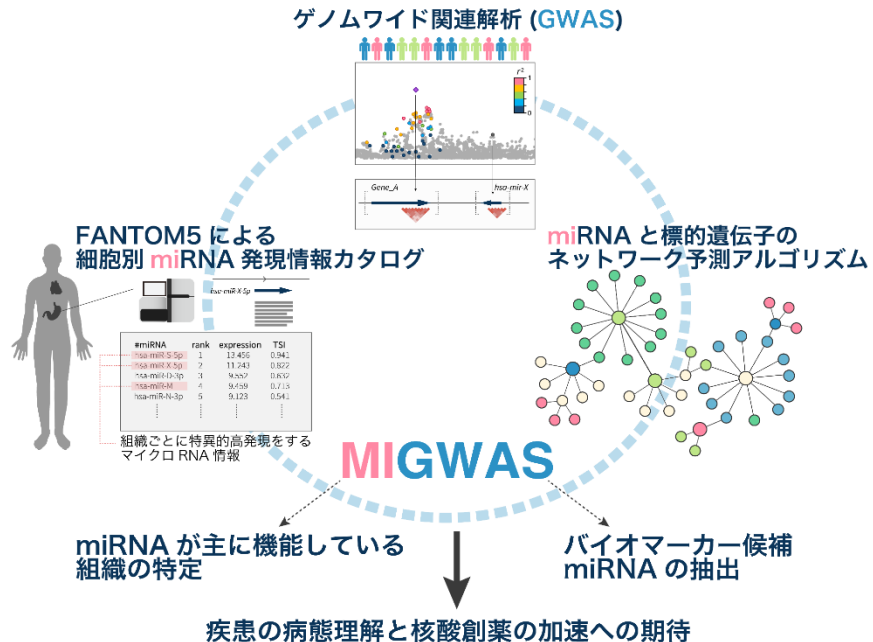
ゲノム解析に基づくドラッグ・リポジショニング



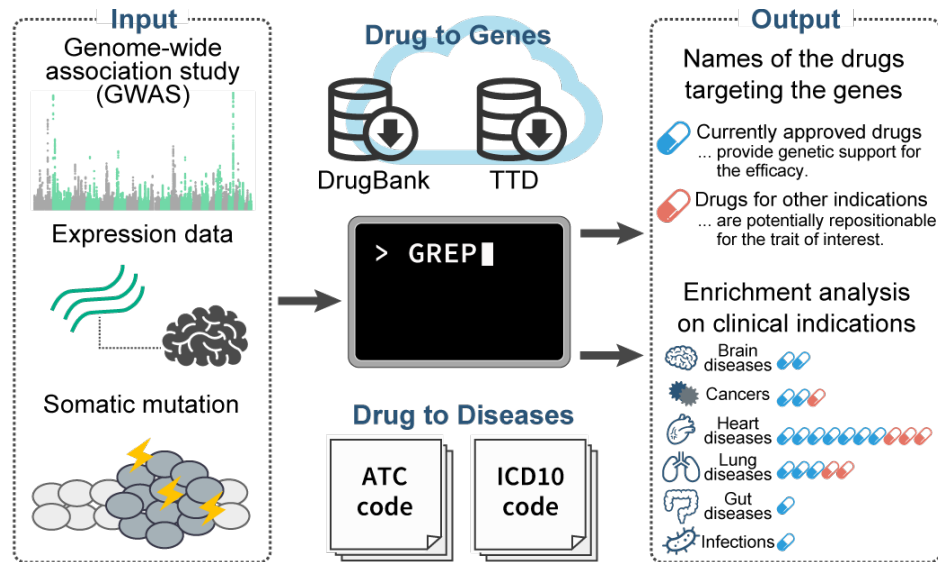
- ・ 既存の治療薬の他の疾患への適用拡大(**ドラッグ・リポジショニング**)が、新規創薬のコスト改善に貢献すると注目されている。
- ・ **疾患感受性遺伝子を直接標的とした治療薬**に注目することで、ドラッグ・リポジショニング候補探索が効率化できることを報告。
- ・ **GSK3B阻害剤**(2型糖尿病)や**CDK4/6阻害剤**(関節リウマチ)を同定。

疾患ゲノム情報に基づく新規創薬スクリーニング

MIGWAS: miRNA-target gene networks enrichment on GWAS



GREP: Genome for REpositioning drugs



- 疾患感受性遺伝子を直接標的とした治療薬に注目することで、**ドラッグ・リポジショニング**候補探索の効率化が可能。
- **MIGWAS**: 組織特異的miRNA発現情報に基づく核酸ゲノム創薬ソフト
- **GREP**: WHO ATC分類を活用した自動化リポジショニングソフト

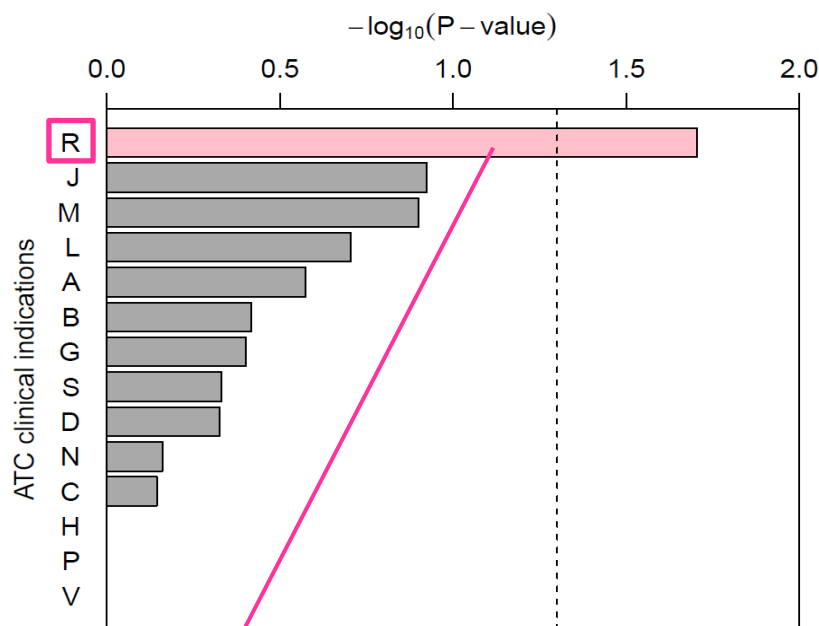
COVID-19 HGI GWAS A2 release 4 (very severe respiratory confirmed COVID)

ATC
code

A: alimentary tract and metabolism
 C: cardiovascular system
 B: blood and blood forming organs
 D: dermatologicals

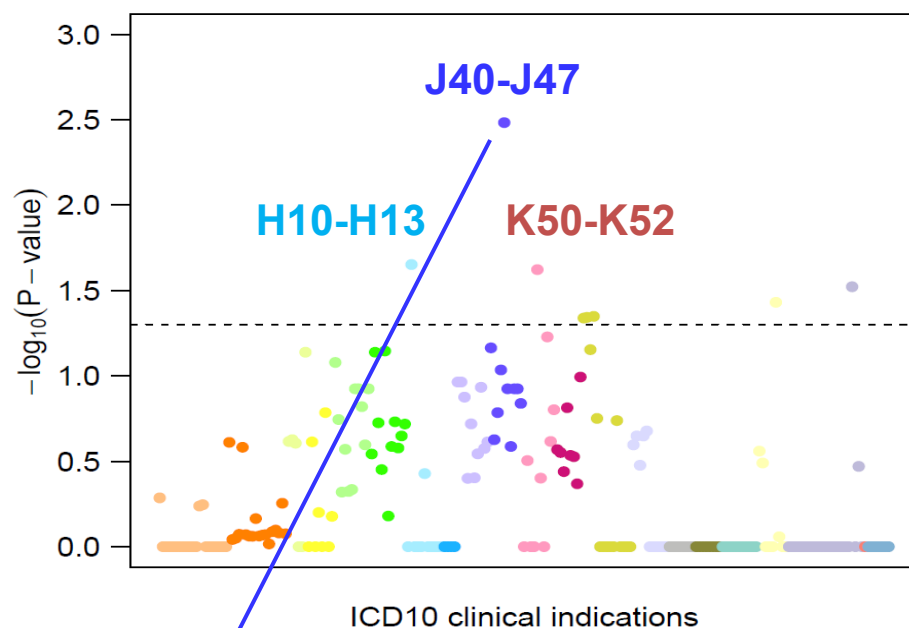
ICD10
code

● A-B: Certain infectious and parasitic diseases
 ● C00-D49: Neoplasms
 ● D50-D89: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
 ● E: Endocrine, nutritional and metabolic diseases



R: Respiratory system

DrugTarget	DrugName	Indication
PDE4A	Theophylline	COPD
PDE4A	Ibudilast	Psoriasis
PDE4A	Roflumilast	Asthma
ICAM1	Hyaluronic acid	Joint lubricant



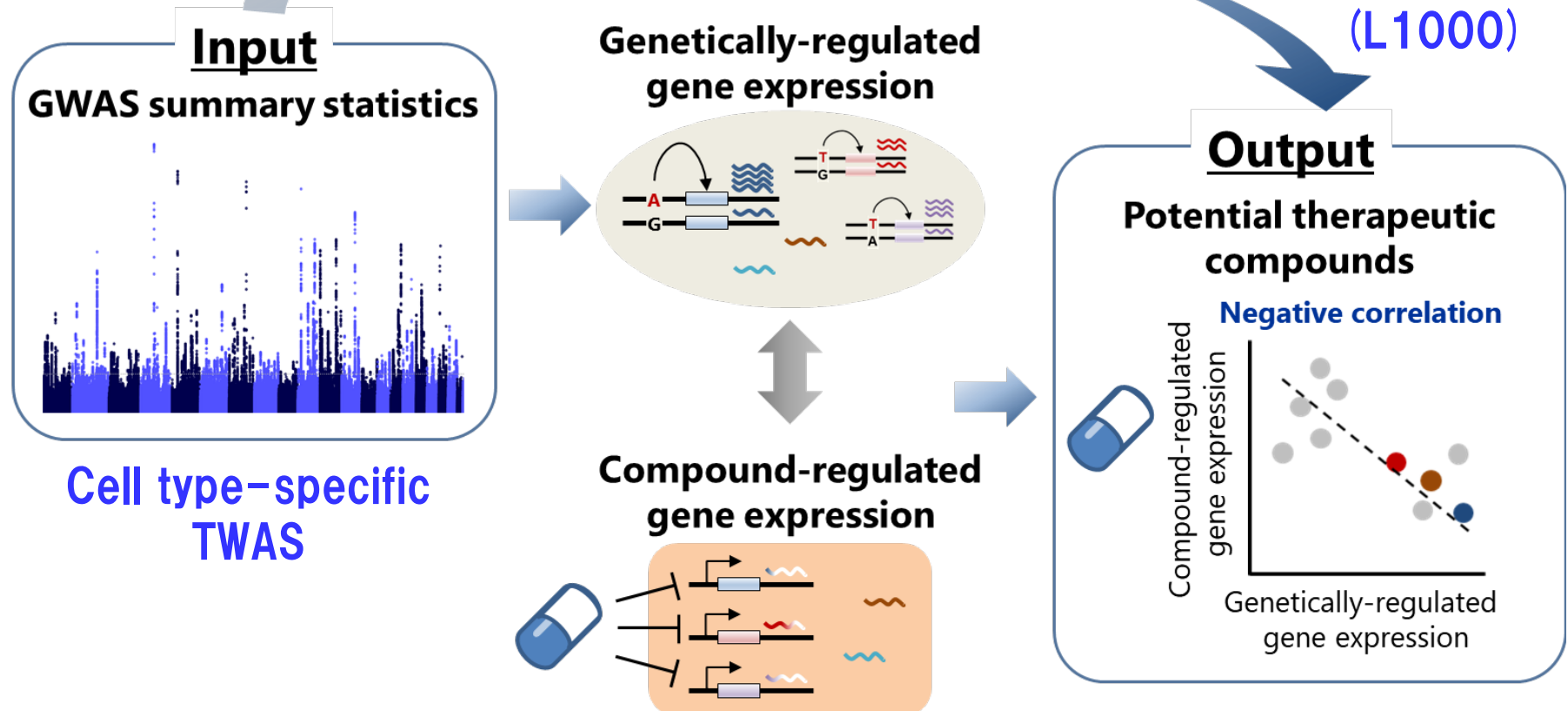
J40-J47: Chronic lower respiratory diseases

DrugTarget	DrugName	Indication
CCR1	AZD-4818	COPD
CCR3	ASM8	Asthma, rhinitis
CCR5	Maraviroc	Virus infection
PDE4A	Roflumilast, etc	Asthma

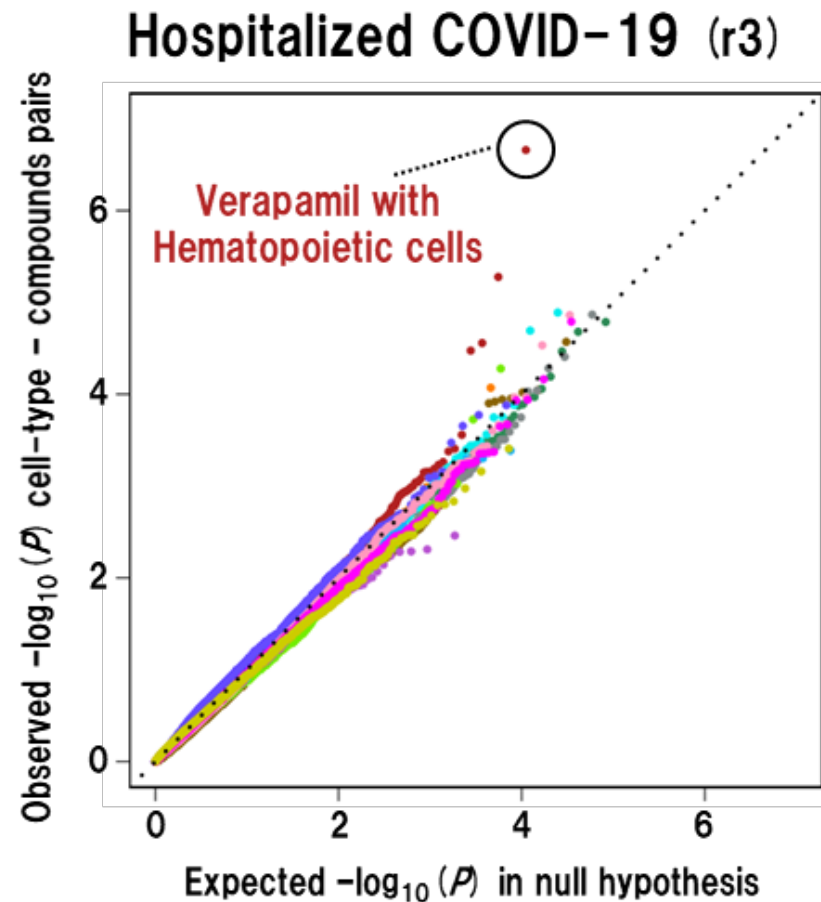
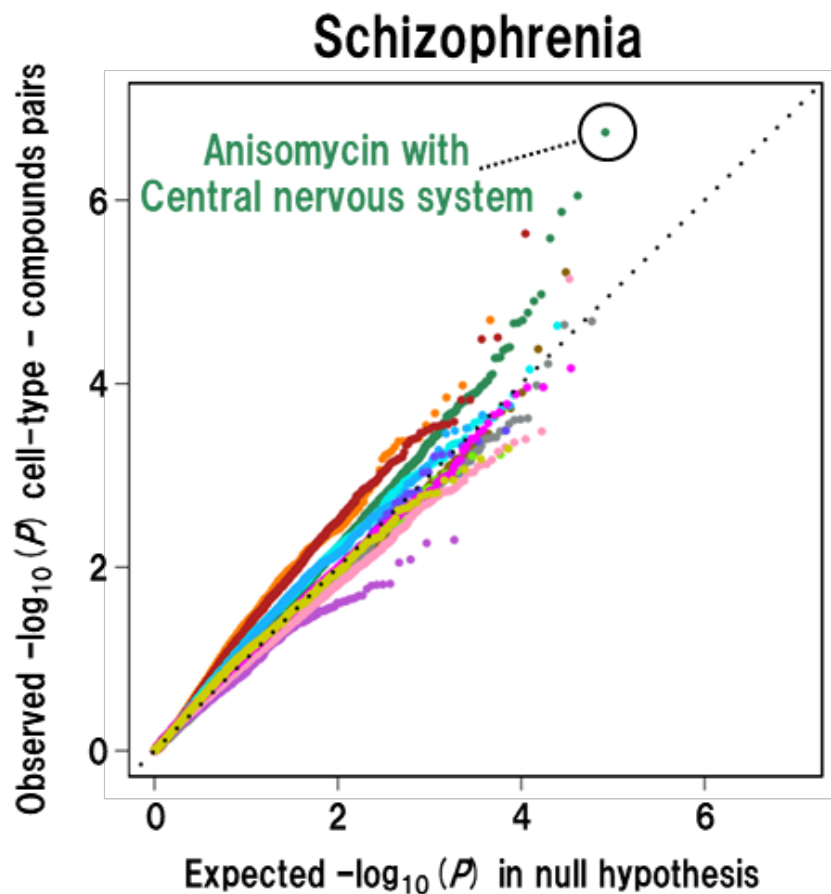
Trans-Phar

Integration of Transcriptome-wide association study
and Pharmacological database

LINCS
Connectivity map
(L1000)



Trans-Phar: TWASを用いたインシリコ・ドラッグスクリーニング



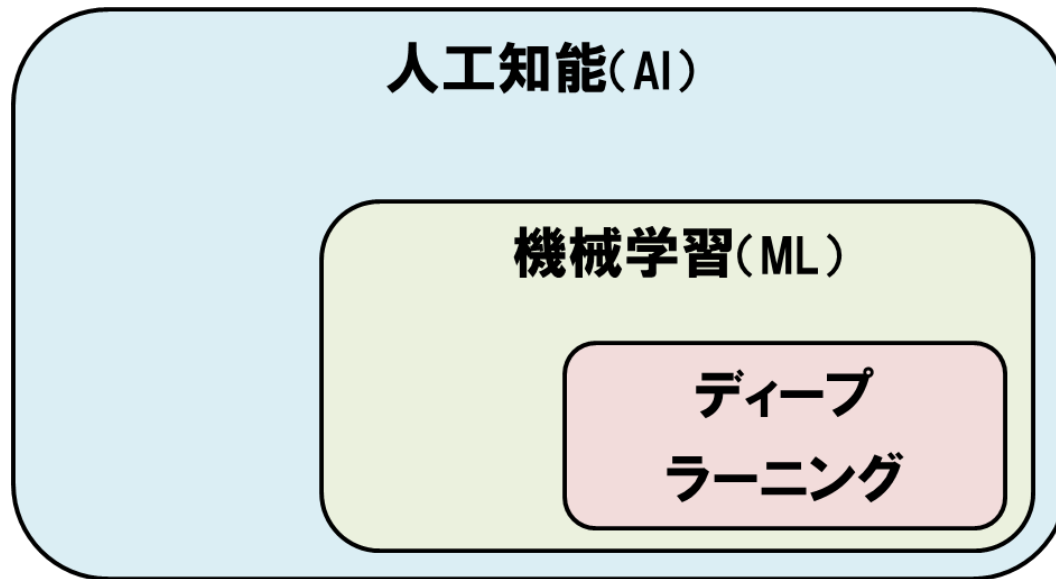
- Trans-Pharは、GWAS対象形質自身に適用済(治験中)や、リポジショニング候補として有用な化合物を、細胞組織別にスクリーニング可能(例: 統合失調症&中枢神経細胞、新型コロナウイルス感染症&血球細胞)。

講演内容

- ・ 横断的オミクス解析による病態解明
- ・ **ヒトゲノム情報への人工知能の活用**
- ・ ゲノム個別化医療への挑戦



人工知能・機械学習・深層学習の包含関係

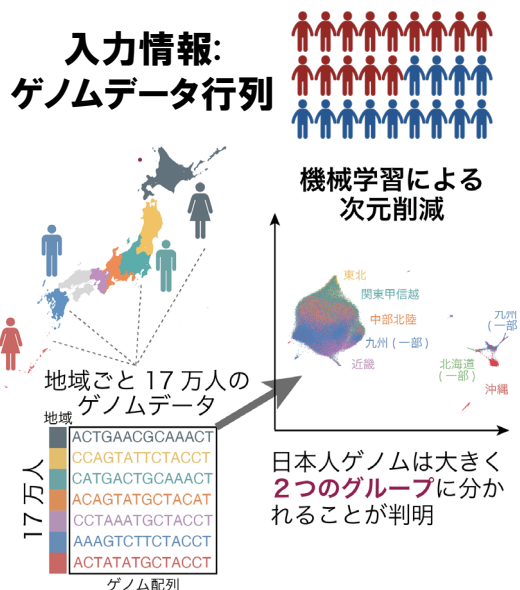


- 昨今注目されている**人工知能**([Artificial Intelligence; AI](#))や**機械学習**([Machine Learning; ML](#))もゲノム解析に有用です。
- 人工知能は、人間の知的活動をコンピューターで実現する技術です。
- 機械学習は、人工知能の実践的な解析技術の一つで、与えられたデータを学習して特徴を把握し、判断や予測を行うための手法になります。
- **ディープラーニング**は機械学習のアルゴリズムの一つで、脳の神経回路を模倣して、複数のデータ処理層を組み合わせていることが特徴です。

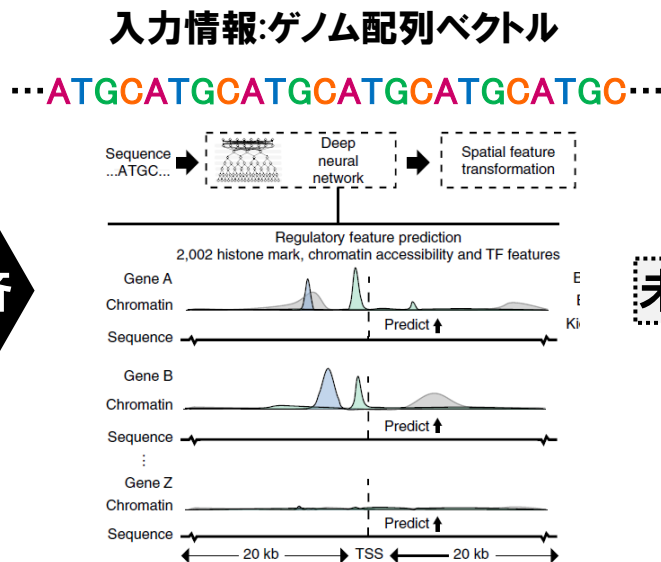
ヒトゲノムデータに対する機械学習の応用

- 人工知能技術(AI)における**深層学習の技術変革**が著しく、従来は困難であった問題の解決を生命科学の広い分野で達成しています。
- ヒトゲノム解析では、**集団ゲノムデータへの機械学習** → **参照ゲノム配列への深層学習**へと進化するも、**集団ゲノムデータへの深層学習**が課題。
- 集団ゲノムデータ行列への深層学習の適用は、高精度予測能を要する個別化医療実装に必須と期待されるが、フレームワークの構築は道半ばです。

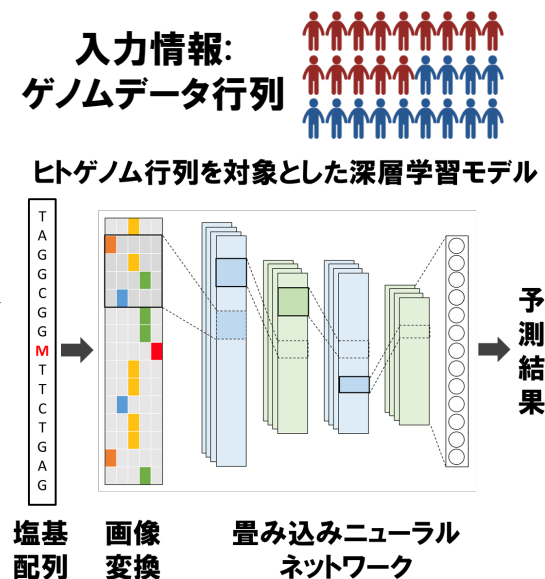
集団ゲノムデータ行列への教師なし機械学習



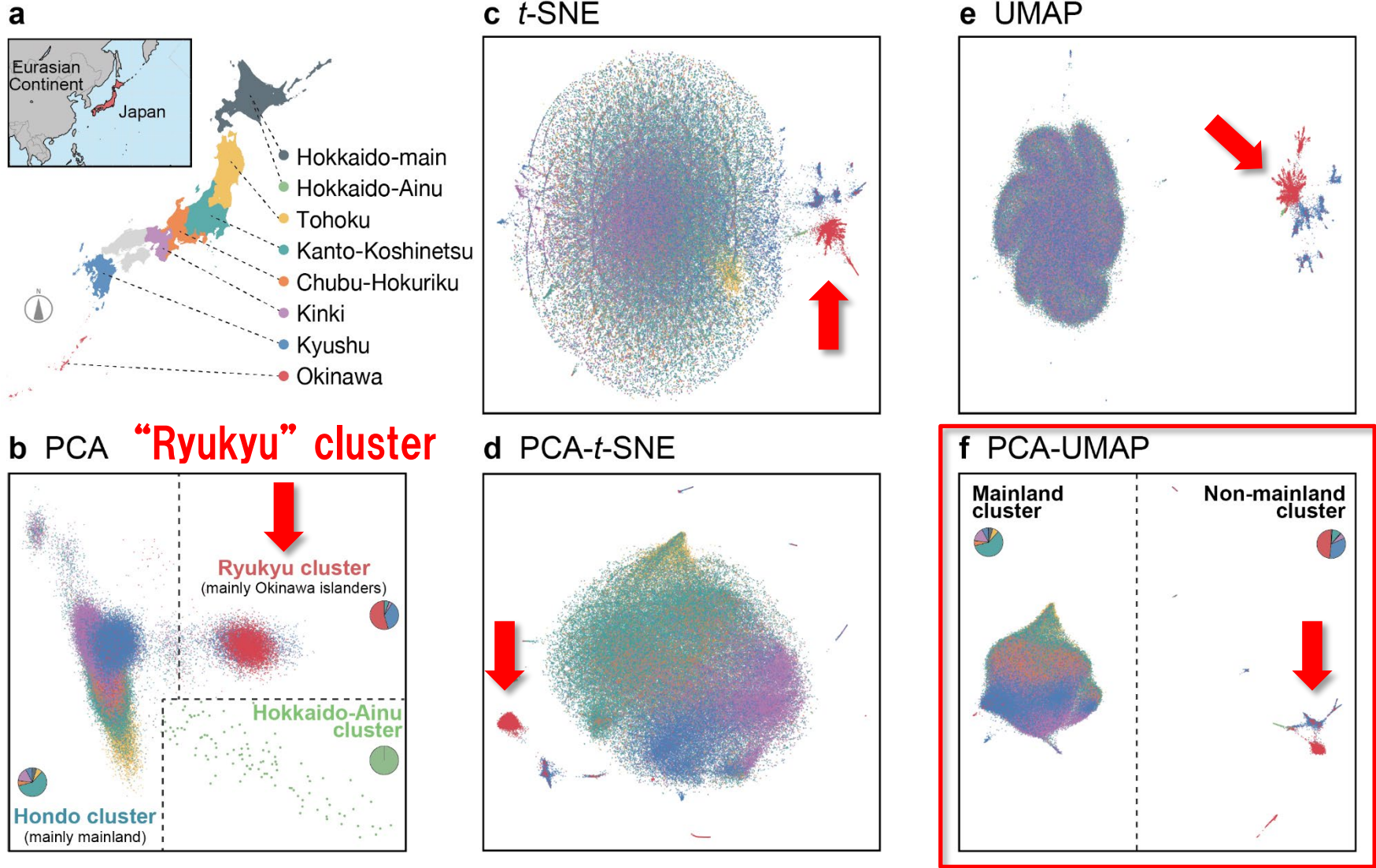
参照ゲノム配列ベクトルへの教師あり深層学習



集団ゲノムデータ行列への教師あり深層学習



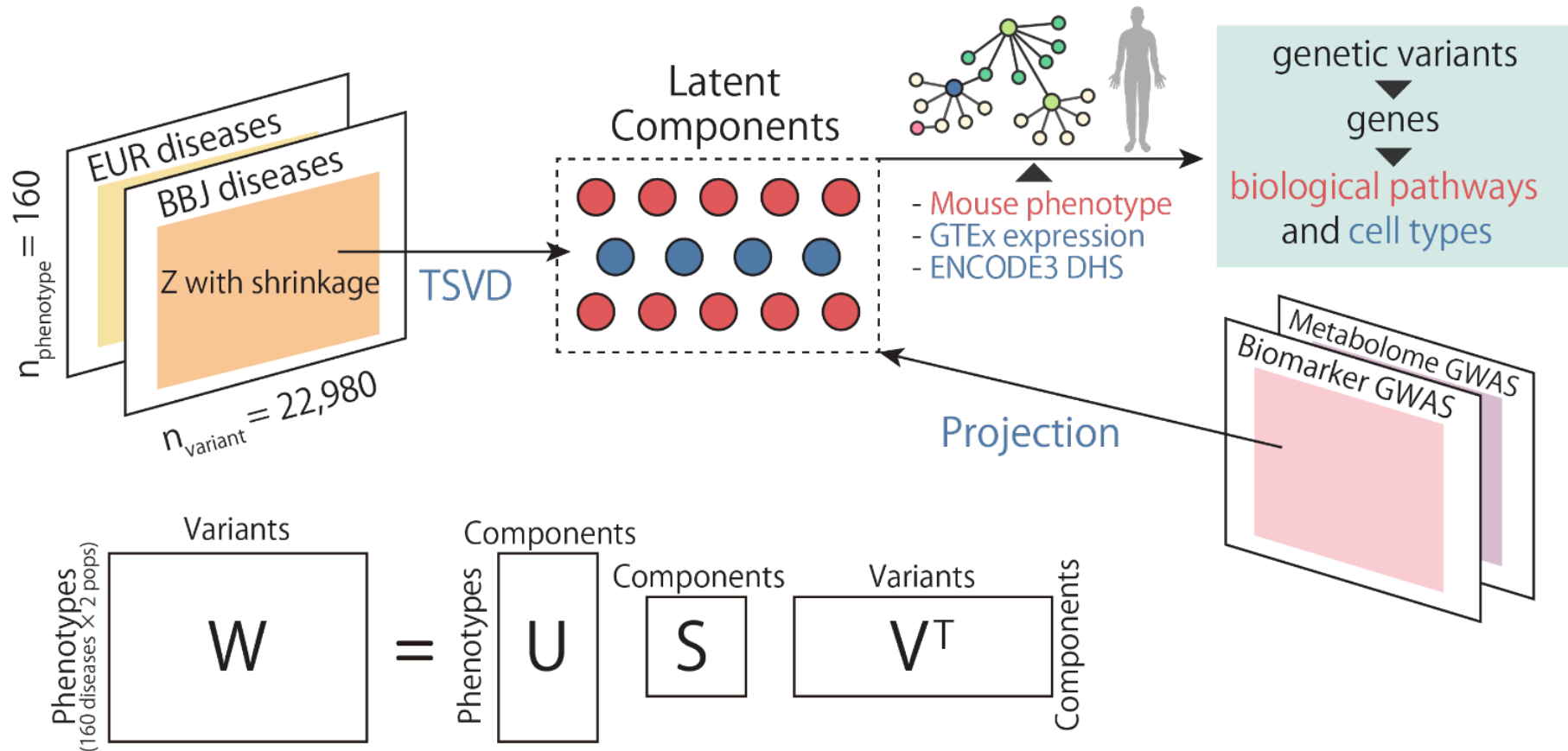
線形-非線形機械学習による日本人集団構造化の解明



・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”)が可能となった。

(Sakaue S et al. *Nat Commun* ²²2020)

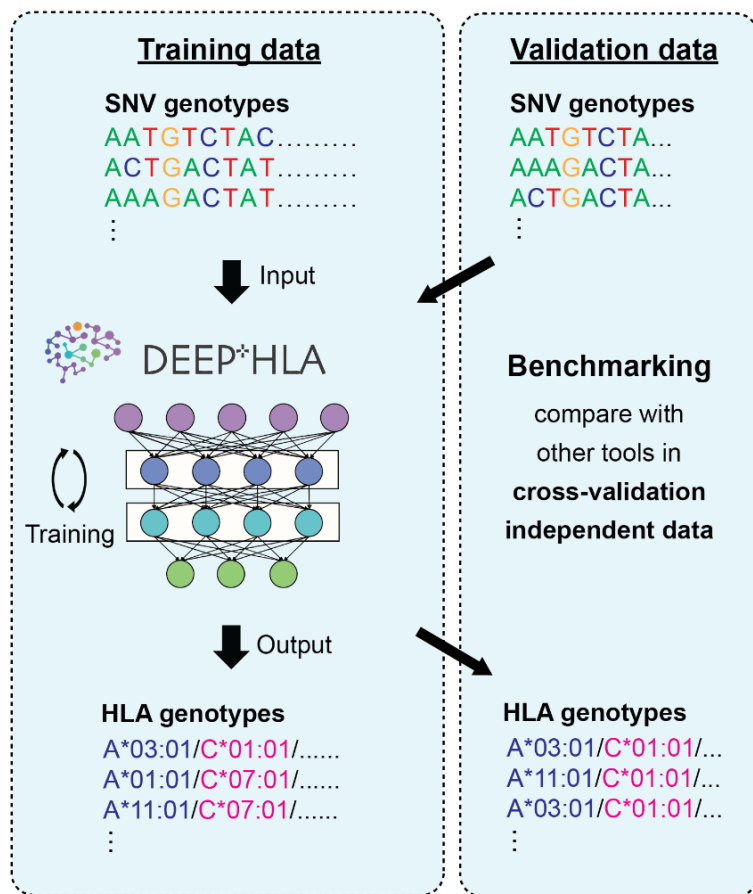
複数疾患の遺伝的背景の網羅的比較



- 多数のGWAS統計量への特異値分解(tSVD)の適用により、関節リウマチとSLE、喘息・アレルギー性鼻炎・アトピー性皮膚炎・など、複数の形質に共通した感受性遺伝子・パスウェイを抽出可能になった。

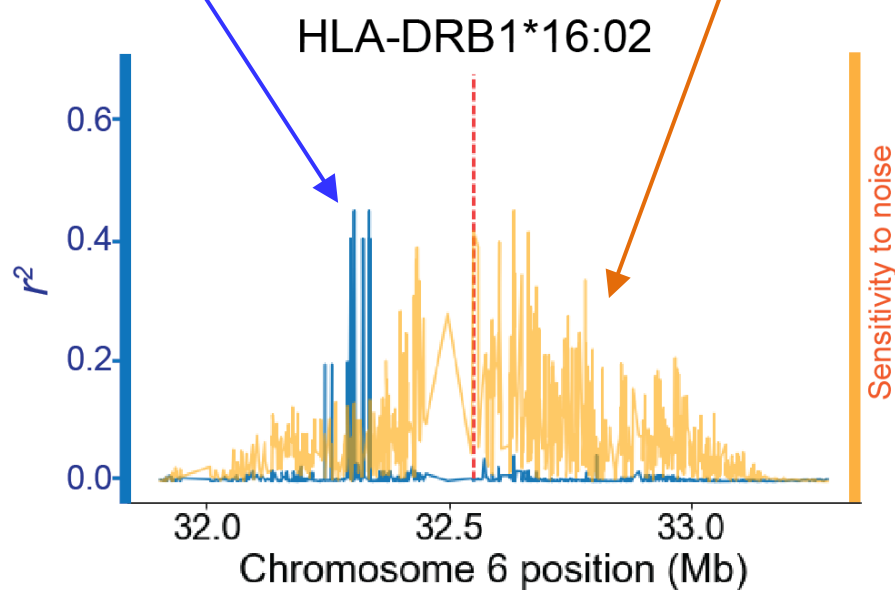
(Sakaue S and Kanai M et al. *medRxiv* 2020)

ヒト集団ゲノムデータ行列への深層学習の適用



既存の機械学習が
参考にする箇所

深層学習が
参考にする箇所



- 深層学習によるHLA imputation法(DEEP*HLA)を開発。MHC領域内のヒト集団ゲノム行列を画像変換することで深層学習の適用を可能にした。
- 近接する変異間の連鎖不平衡関係に依存しないため、従来の機械学習(例:マルコフ連鎖)と比較して、**稀なHLA遺伝子型**の推定精度が改善。計算時間・使用メモリも改善。

(Naito T et al. *Nat Commun* 2021)

講演内容

- 横断的オミクス解析による病態解明
- ヒトゲノム情報への人工知能の活用
- **ゲノム個別化医療への挑戦**



ゲノム個別化医療時代の幕開け



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>)

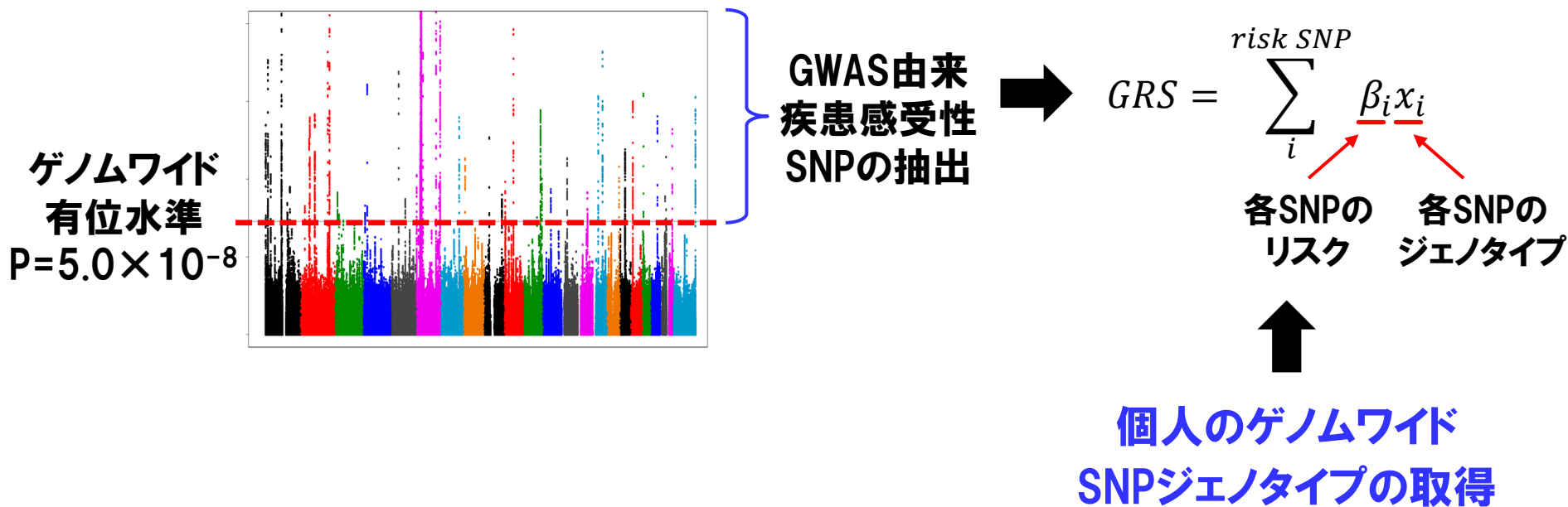
- 2015年のオバマ大統領の一般教書演説で、“**Precision Medicine Initiative**”という言葉が注目を集めました。
- Precision Medicine Initiativeは、特に**個人のゲノム情報の違いを活用した個別化医療**の提言に重きを置いた点が画期的でした。
- **ゲノム個別化医療**の時代の幕開け、という解釈もできます。

GWAS感受性遺伝子変異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

GRSの推定

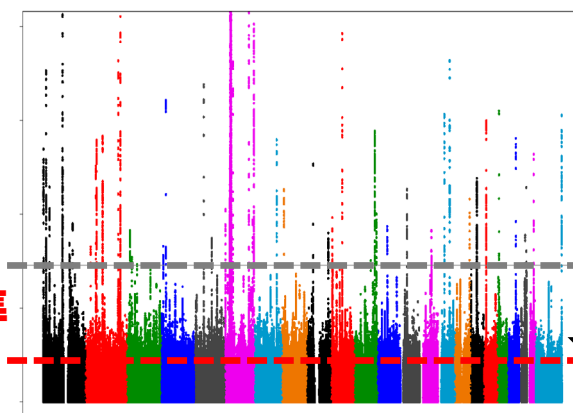


- GWASが同定した**疾患感受性SNP**に基づく発症予測も始まりました。
- 各疾患感受性SNPにおいて、GWASで推定されたリスクと、推定対象の個人におけるジェノタイプを掛け合わせた上で、合計します。
- **Genotype risk score (GRS)**と呼ばれます。

Polygenic risk score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



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

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

- 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、ゲノムワイド有位水準より
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。

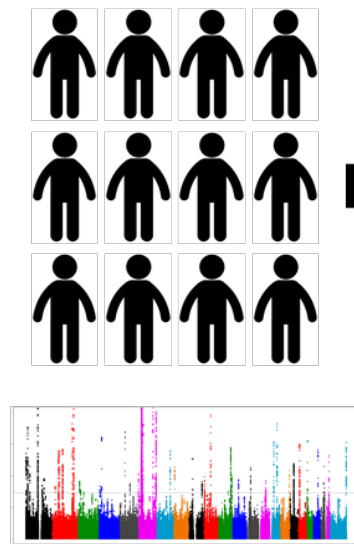
免疫オミクス・臨床情報に基づくRA患者層別化・ゲノム個別化医療

大規模RA GWAS
データ構築

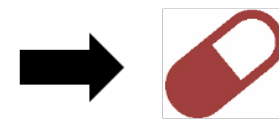
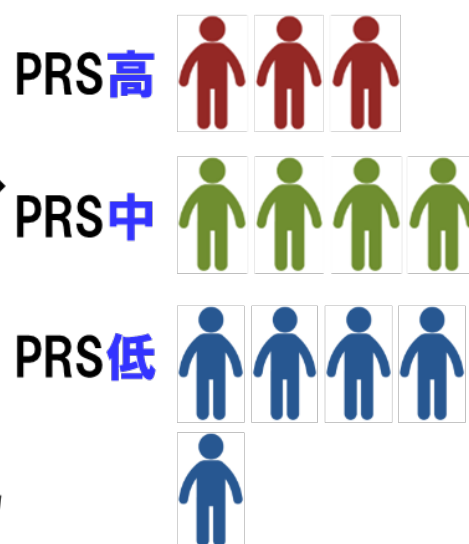
Polygenic Risk
Score推定

RA患者群
層別化モデル

RAゲノム
個別化医療



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



標準的
治療手段

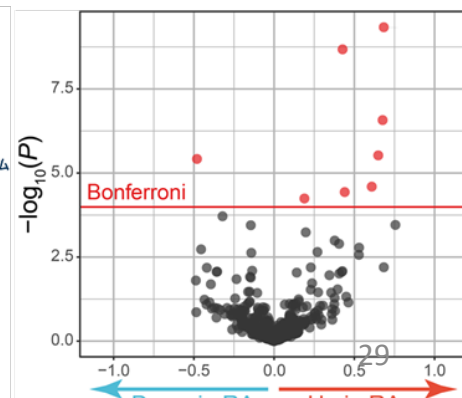
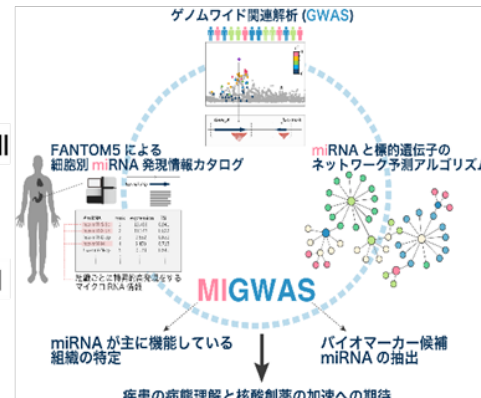
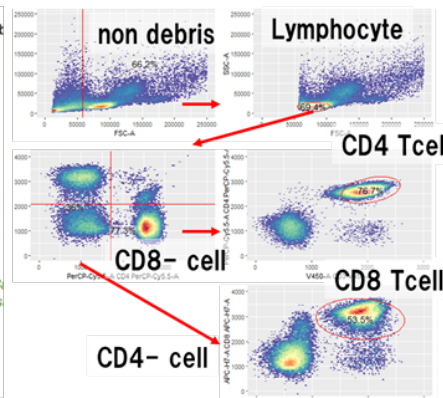
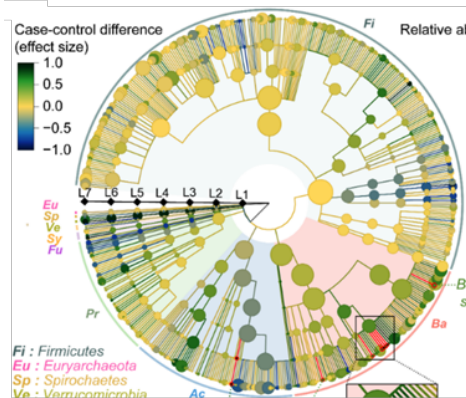
前向きコホート・バイオバンク由来のRA免疫オミクス情報・臨床情報

腸内細菌メタゲノム

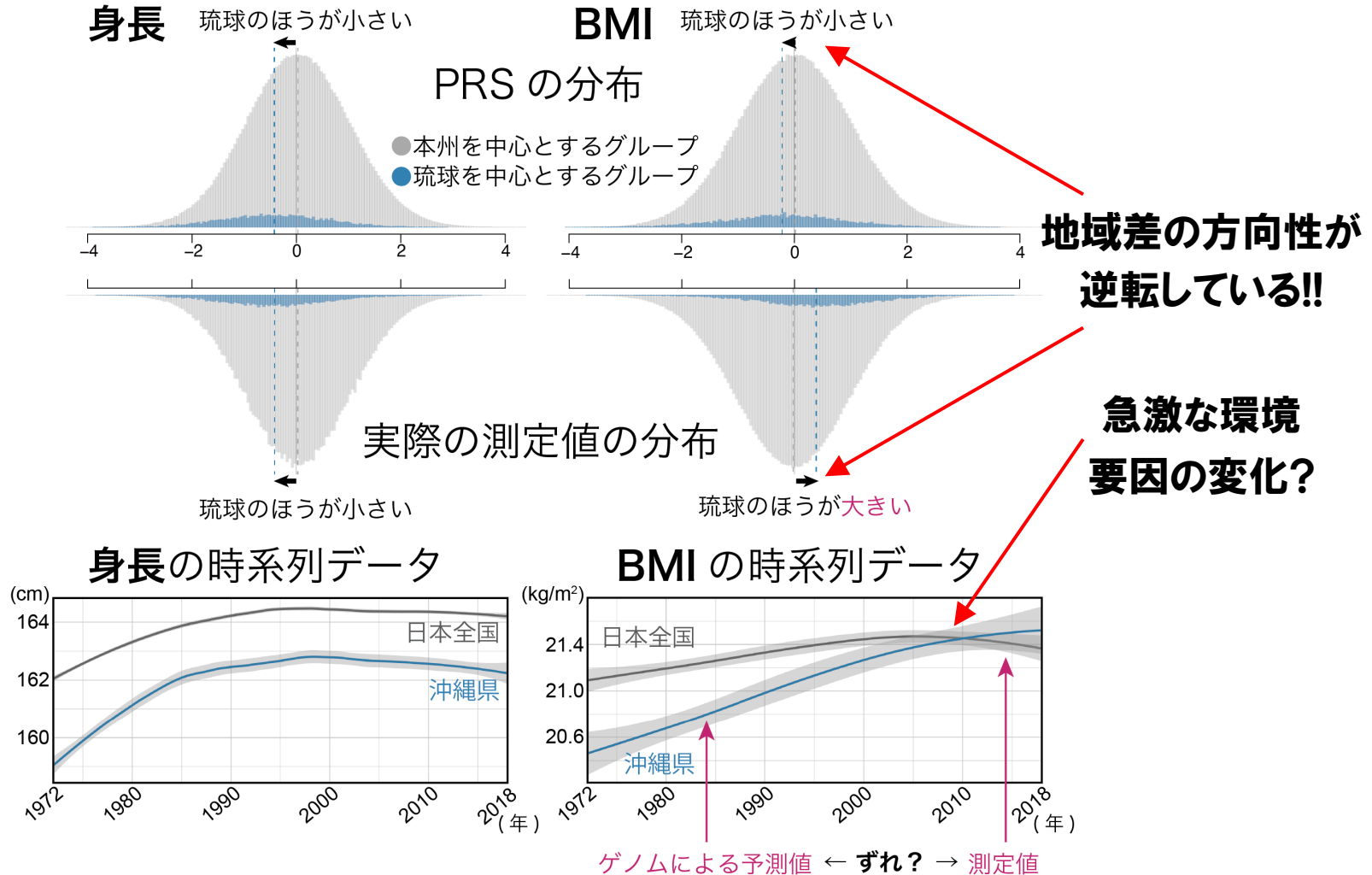
免疫フェノタイプ

マイクロRNA

メタボローム



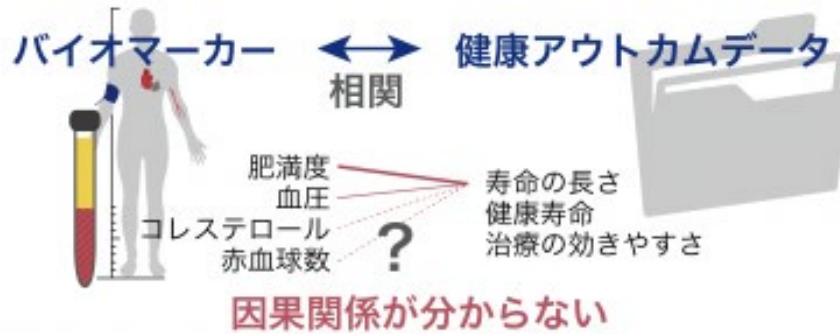
日本人集団構造化がゲノム個別化医療に与える影響



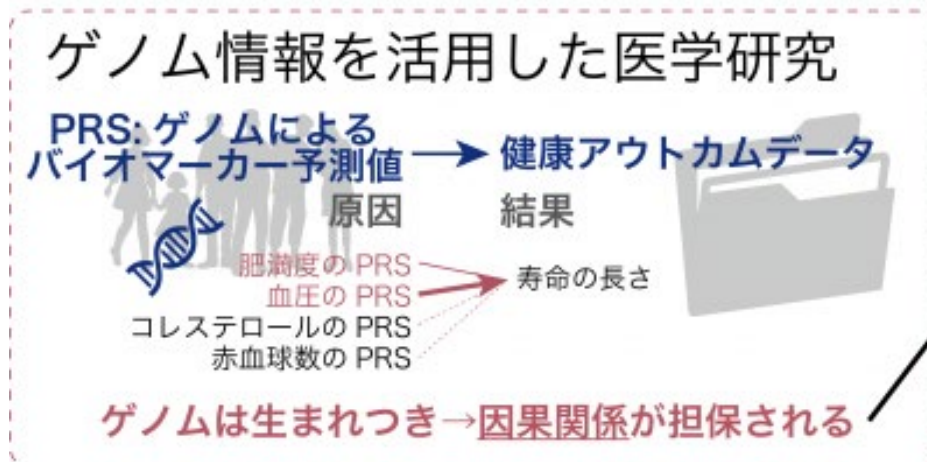
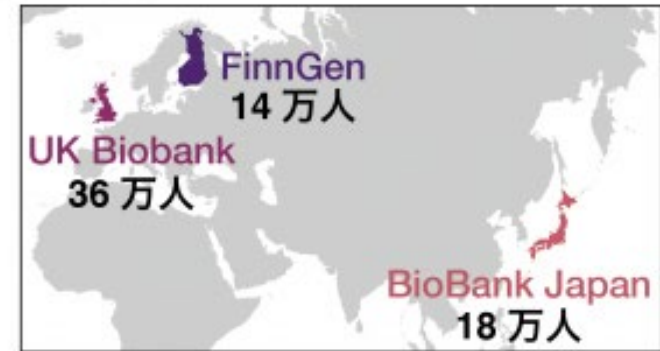
- ・本土・琉球クラスター間において、**肥満の実測値とPRS予測値が逆の方向性を示す**ことが判明。沖縄地域における急激な環境要因の変化をゲノム情報が網羅できていない可能性が浮上した。

Polygenic risk scoreとゲノム個別化医療

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PRS解析を実施。健康の指標かつ治療可能なバイオマーカーのPRSと寿命との因果関係を人種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した。

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

次代を担う若手研究者の人材育成

遺伝統計学・夏の学校@大阪大学の開催 (2020年度はオンライン開催:8/28-30)

医学部正規カリキュラムとして
ゲノムデータ解析演習の導入

遺伝統計学 夏の学校 @大阪大学 参加費：無料

夏の大阪で、ゲノム解析の世界に触れてみませんか？

日時：令和2年8月28日(金)～8月30日(日)

(原則として、三日間通じて参加をお願いします)

場所：ウェビナー形式でのオンライン参加

(新型コロナウイルス感染対策の観点より、本年度はオンライン形式での開催とします)

対象：大学学部生以上

(大学院生・社会人の方も歓迎です。保護者の了解が得られれば高校生も歓迎です。)

形式：講義・ゲノムデータ解析演習

(演習に使用するパソコンは、参加者各自に用意して頂きます)
(使用する解析用ゲノムデータは、夏の学校で用意し、設定資料と共に事前配布します)

定員：100名

(応募者多数の場合、所属や学年を考慮して調整させて頂く場合があります)

日程表

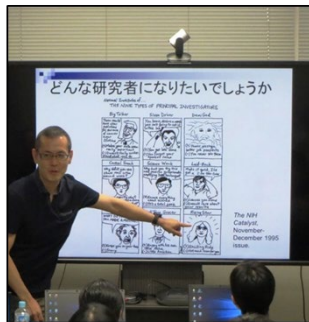
時間帯	一日目 (8/28)	二日目 (8/29)	三日目 (8/30)
午前 (9:00~12:00)	遺伝学入門 Linux入門	プログラミング 入門	ゲノムデータ 解析演習 ③・④・⑤
午後 (13:00~16:30)	統計学入門 統計ソフト入門	ゲノムデータ 解析演習①・②	

(講義演習内容は、変更となる場合があります)

URL : http://www.sg.med.osaka-u.ac.jp/school_2020.html

申込み : office@sg.med.osaka-u.ac.jp (担当：岡田 随象)

共催：AMEDゲノム医療実現推進プラットフォーム事業、バイオインフォマティクス・イニシアティブ



③ Cygwinを使ったLinuxコマンド実習

コマンド	機能
top	システム使用状況を表示
pwd	カレントディレクトリを表示
cd	カレントディレクトリを変更
ls	ディレクトリやファイル情報を表示
echo	文字列や変数の値を標準出力に表示
cp	ファイルやディレクトリをコピーする
mv	ファイルやディレクトリを移動する
rm	ファイルやディレクトリを消去する
wc	テキストファイルの大きさや行数を表示
gzip	ファイルとgz形式で圧縮
gunzip	gz形式のファイルを解凍
less	ファイル内容を一面ごとに表示
head	ファイル内容を行数単位で先頭から表示
tail	ファイル内容を行数単位で末尾から表示

*Linuxコマンドは多数存在します。頻回に使用するものから覚えましょう。

- ・公開データを用いたゲノム解析体験の提供を通じて、有望な国内若手人材を発掘中。
- ・ゲノムデータ解析実践演習講義「遺伝統計学・夏の学校@大阪大学」を毎年自主開催し、日本各地から若手バイオインフォマティシャン(の卵)が参加。羊土社より書籍化。
- ・ゲノムデータ解析演習を大阪大学医学部2回生の正規カリキュラムとして開始。

謝辞(敬称略)

大阪大学大学院医学系研究科

— 岸川敏博 吹田直政 平田潤 鈴木顕 小沼貴裕 岩崎航太郎 山本賢一 田中宏明
内藤龍彦 曾根原究人 白井雄也 曾根原究人 森井航 枝廣龍哉 佐々暢亜
友藤嘉彦 難波真一 茶谷亮輔 末成順子 東本弘子
熊ノ郷淳 竹田潔 奥崎大介 中村昇太 前田悠一 新居卓朗 荒瀬規子

バイオバンクジャパン・理化学研究所・東京大学医科学研究所

— 山本一彦 松田浩一 村上善則 桃沢幸秀 鎌谷洋一郎

京都大学 ゲノム医学センター/医学研究科

— 松田文彦 山田亮 大村浩一郎 高橋めい子

産業医科大学医学部

— 田中良哉

名古屋市立大学大学院医学研究科

— 森田明理

東京慈恵会医科大学

— 玉利真由美 広田朝光

Harvard Medical School, Broad Institute

— Mark Daly, Soumya Raychaudhuri
石垣和慶 坂上沙央里 金井仁弘

**スタッフ・大学院生
募集中!!**

**遺伝統計学・夏の学校
@大阪大学が
書籍化されました。**

amazon.com
販売サイト
QRコード

