

多因子疾患研究における ゲノムデータ共有の意義

国立国際医療研究センター（NCGM） 研究所

ゲノム医科学プロジェクト

ナショナルセンターバイオバンクネットワーク（NCBN）

中央バイオバンク

徳永勝士

ゲノム全域解析の主な方法

• DNAマイクロアレイ (SNPアレイ)

- あらかじめ設計された50万～数100万ヶ所のSNP（一塩基多型）のタイピングが可能
- **さまざまな多因子疾患、薬剤応答性などの研究**

• 次世代シーケンサー (NGS)

- 大量のゲノムシーケンスを解読可能
- SNP/SNVに加え、挿入や欠失など複雑なゲノム変異も解析可能
- SNPアレイに比べて高価
- **がん、単一遺伝子疾患、希少疾患、未診断疾患の研究・診断**

多因子疾患のゲノム研究

- **ゲノムワイド関連解析 (GWAS)** : 疾患群とコントロール群の間で頻度差のあるSNPをゲノム全域から探索
- **効果 (オッズ比) が小さなSNP**を検出するために**多数検体 (数千～数万／数十万)**を解析する必要、**多施設共同研究**が一般的
- 近年は**国際共同研究**によって、ひとつの集団の解析では検出困難な疾患関連SNPを探索する研究が多い

(1)

国内外のデータ共有による 大規模解析の意義

より多くの遺伝要因の特定
より正確なリスク予測
発症機序の理解
創薬標的の候補

原発性胆汁性胆管炎のGWAS：データ共有の意義

アジア初のGWAS：3領域 (487 患者/476 健常者)

(Nakamura *et al.* Am J Hum Genet 2012)



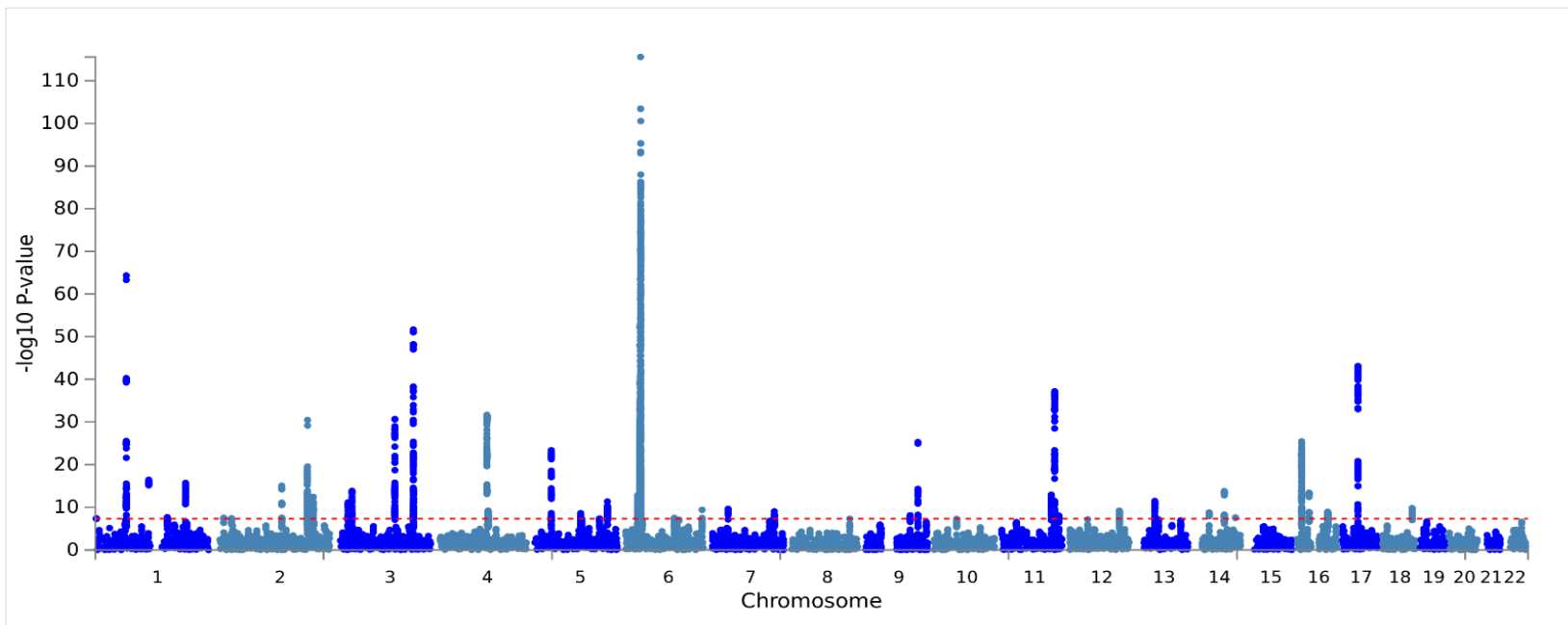
日本人拡大GWAS：8領域 (894 患者/1,029 健常者)

(Kawashima *et al.* Hum Mol Genet 2017)



国際メタGWAS：56領域 (10,516 患者/20,772 健常者)

日本全国多施設共同
→国際共同により
多数の遺伝要因特定



英国、日本、中国、米国、イタリア、カナダの共同研究

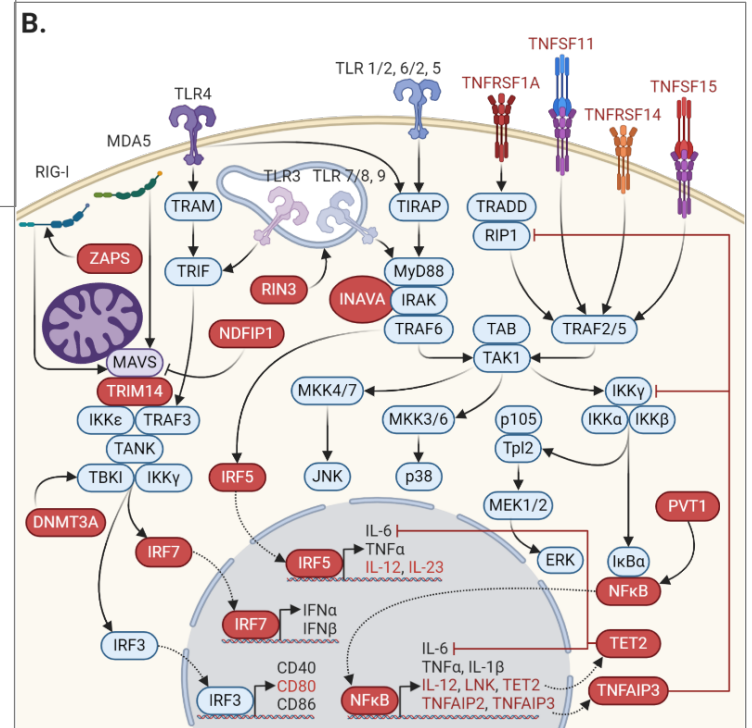
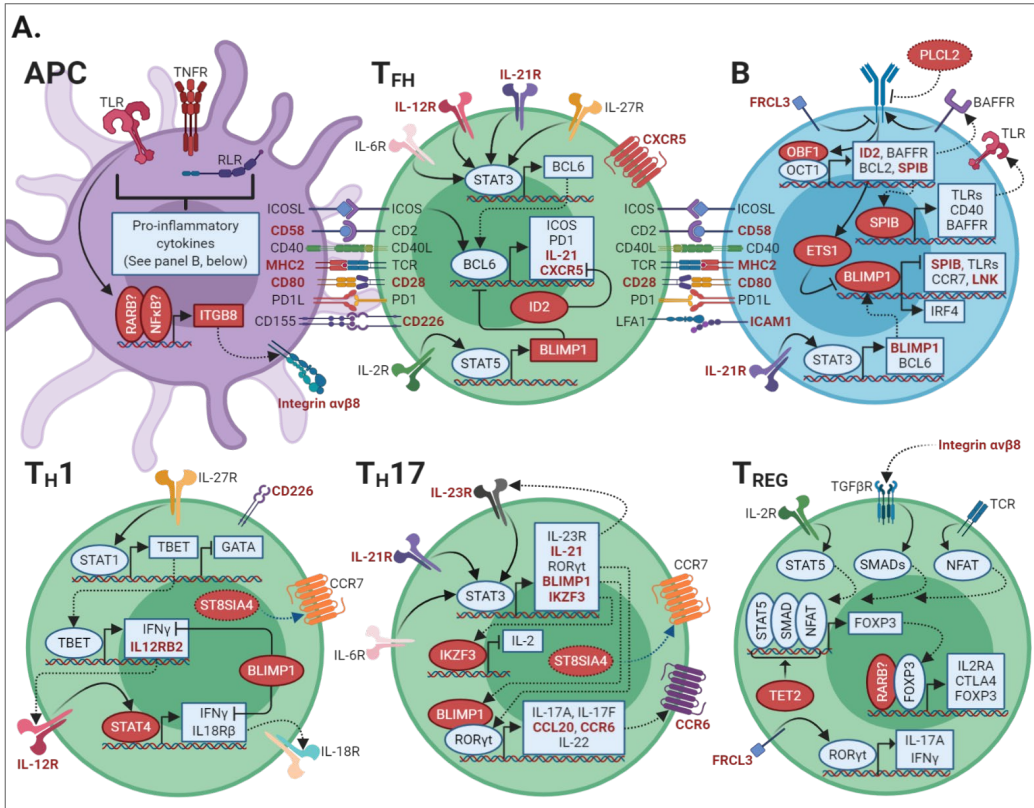
患者：10,516名、対照者：20,772名

ゲノムワイド有意水準を満たす56領域の検出

原発性胆汁性胆管炎 発症機序の理解へ

T細胞活性化の重要性

パターン認識受容体や
TNFシグナルの重要性



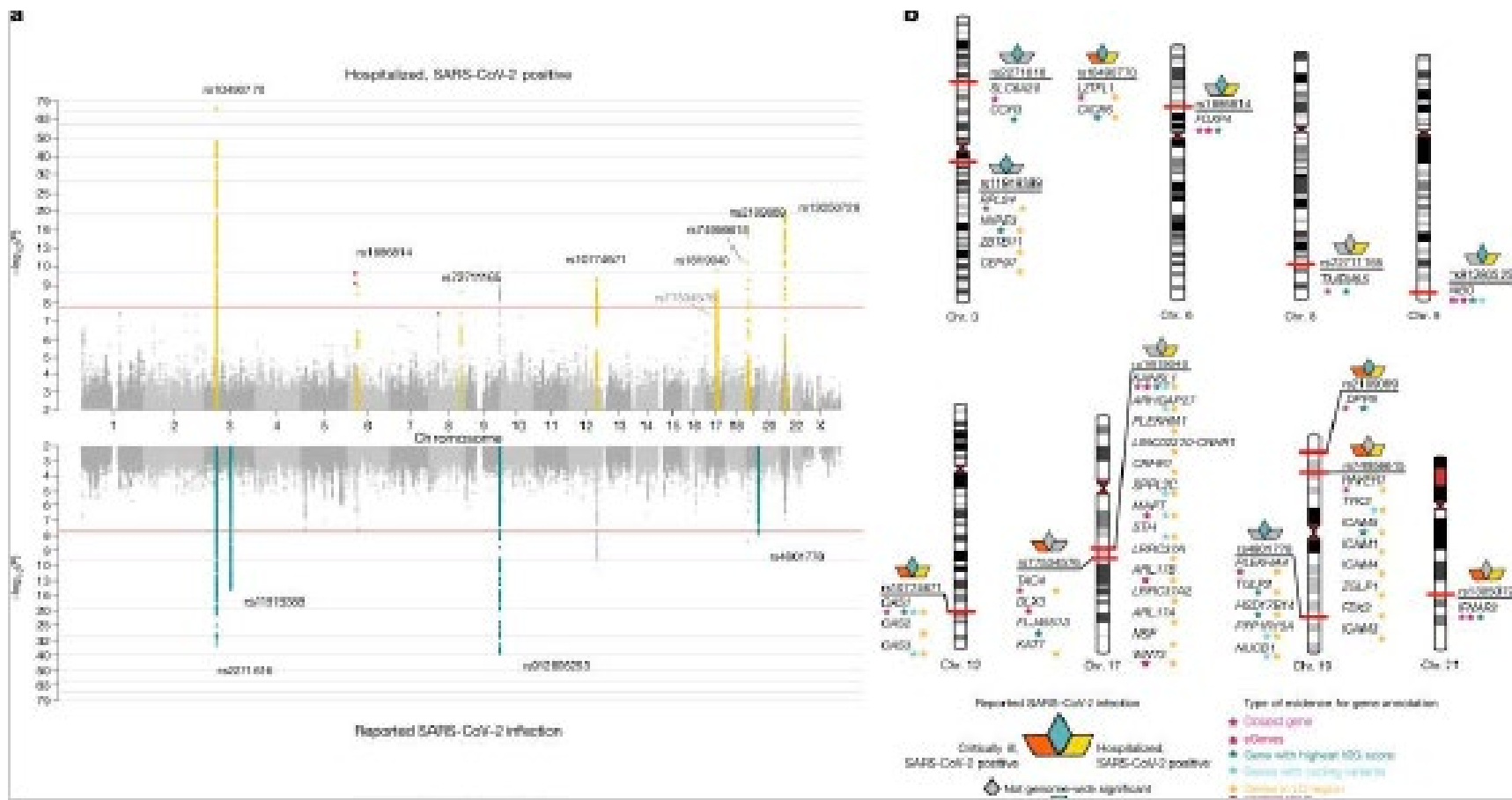
(赤色タンパクの遺伝子がGWASで検出)

ドラッグリポジショニング (既存薬
再開発) の可能性も提案した

Cordell et al. J Hepatol 2021

最大規模の国際共同研究による COVID-19感染症のGWAS

13遺伝子領域が検出されたが、予想されたほど遺伝要因は強くなかった



患者49,562人のゲノム解析、19カ国・46プロジェクトの共同

(2)

全ゲノム解析データの価値

遺伝病／希少疾患のみならず
多因子疾患研究にも役立つ
より多くの成果が得られる

多因子疾患のゲノム研究の課題

• コントロールデータの共有

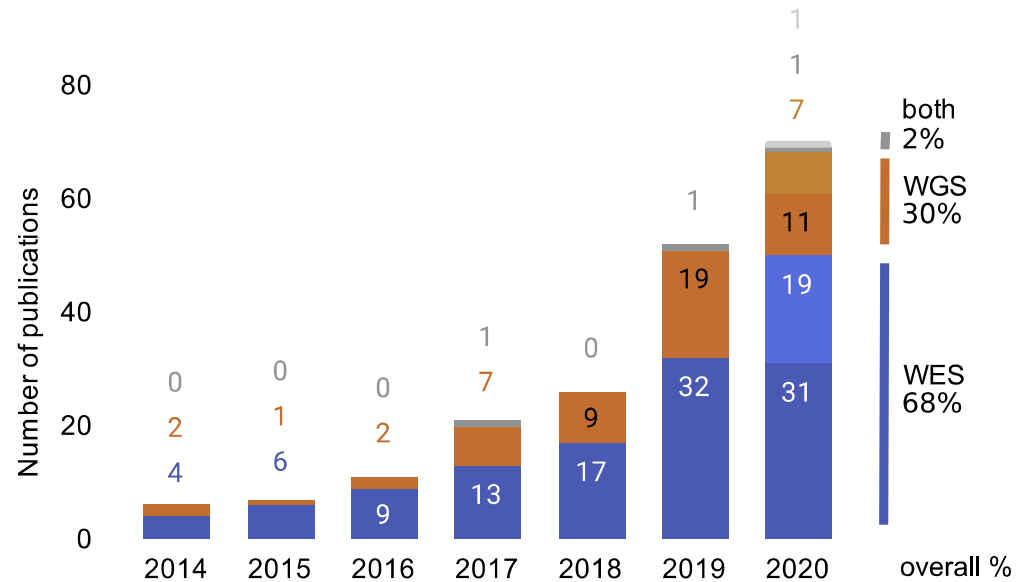
- プロジェクト毎に症例(ケース)群とともに対照(コントロール)群を収集することは困難、特にコントロールデータを共有することが多い
- ただし、データベース/バイオバンクのコントロールデータを利用する場合は、症例群も同じSNPアレイで解析するのが望ましい

• レアバリエント・構造変異の検出が困難

- SNPアレイ→インピュテーションによりほぼ全てのSNPを推定できるが、構造変異やレアバリエントなど解析困難

全ゲノム解析(WGS)を用いたGWAS

- 解析費用は下がり続けている（ただし、SNPアレイ解析より高価）
- バイオバンクがWGS解析を実施しデータ提供を行う例が増えている
 - UK Biobank など
 - NCBN, BBJ, ToMMo
- SNPアレイに代わりWGSをGWASに用いる研究も増えている
- 横断的データ共有の必要性と意義はさらに大きい

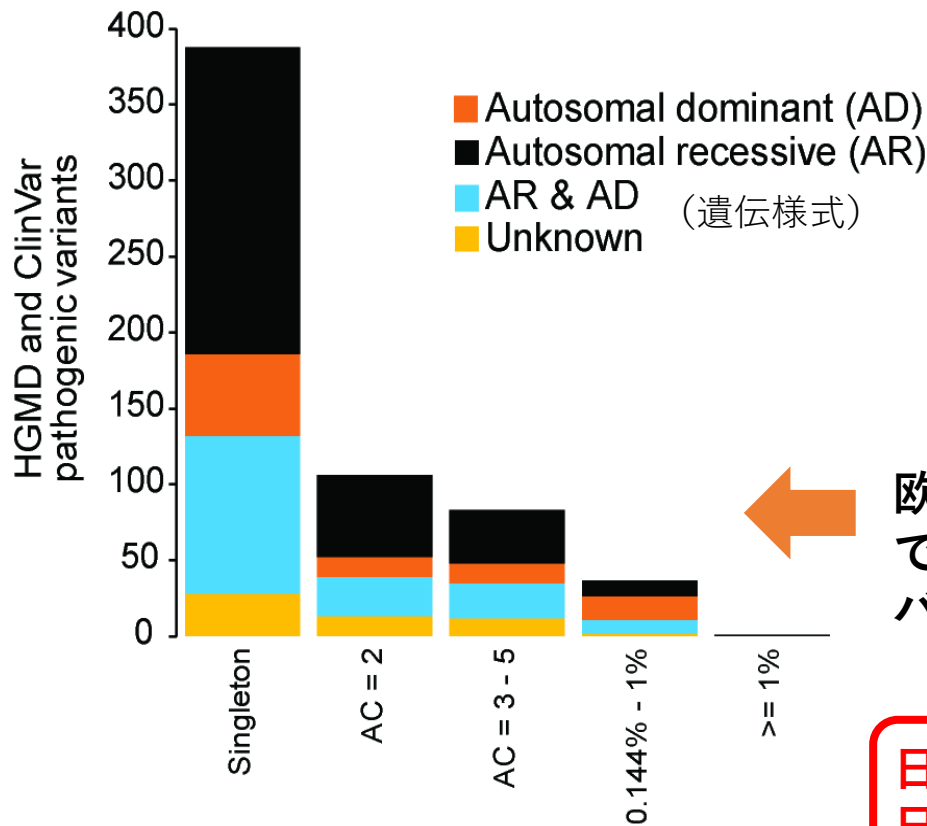


McMahon et al. Cell Genomics 2021

WESまたはWGSを使ったGWAS論文数の増加
(2020年でWGS-GWASは11報)

日本人の大規模全ゲノム解析の必要性

アジア諸集団の大規模全ゲノム解析 (64カ国、200以上の集団の試料)



欧米のデータベース (HGMD, ClinVar) で誤まって病因変異とされているバリエントが多数検出された

日本人のゲノム医療を実現するため、日本人の標準バリエントデータが必要

縦軸：HGMD, ClinVarで病因とされている変異のうち、本研究で観察された種類数
横軸：観察回数、頻度

GenomeAsia100K Consortium, Nature (2019)

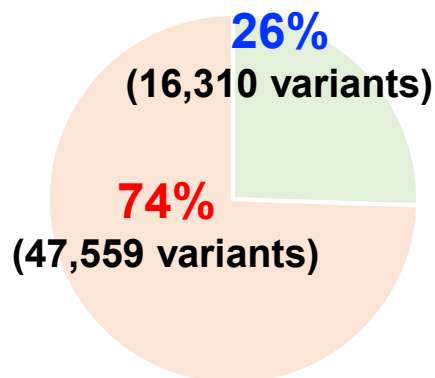
MGeND (疾患関連バリエーションDB) 登録データ数

Data type	Variants	GWAS	HLA allele
がん	185,828 (147,086)	---	---
希少疾患 / 難病	15,399 (2,989)	---	---
感染症	1,276 (1,276)	155,100 (155,100)	1,979 (1,803)
認知症 / 難聴	11,408 (7,668) APOE: 17,876 (5,451)	1,171 (410)	---
その他	1,100 (770)	14,321,737 (584)	---

2021年9月時点

括弧内は公開データ数

ClinVar登録 バリエーション との比較

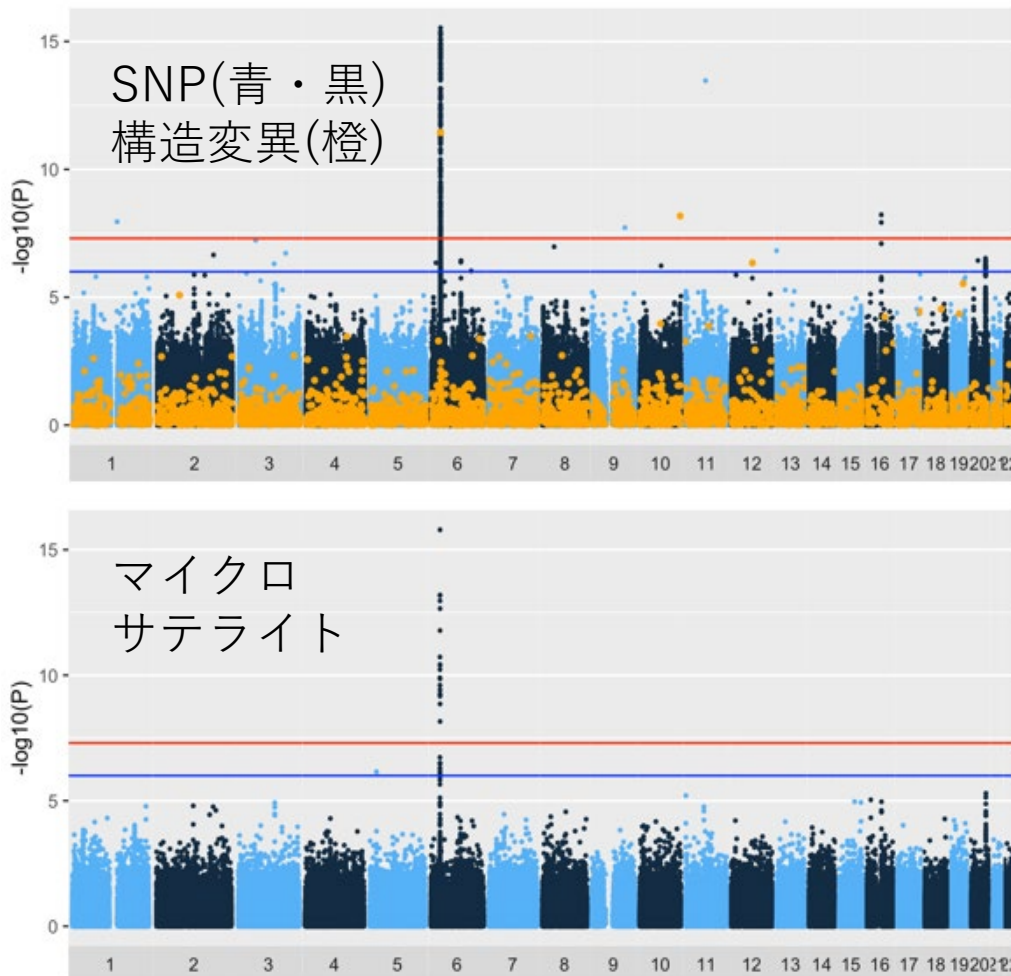


- ClinVarにも存在
- MGeNDのみに存在

MGeND (2021.09, vcf)
unique variants: 63,869

ClinVar (2021.09, vcf)
unique variants (hg38): 1,025,226

重症薬疹スティーブンス・ジョンソン症候群のWGS解析を使ったGWAS



日本人患者133名と
対照者418名の**WGS**
解析 → GWAS

SNP-GWASによる
既知の関連SNPに
加え、**構造変異**や
マイクロサテライト
の関連や**レアバリア
ントの集積遺伝子**も
検出

(3)

大規模データサーバ・
計算環境の提供

より広範かつ深い利活用が可能になる

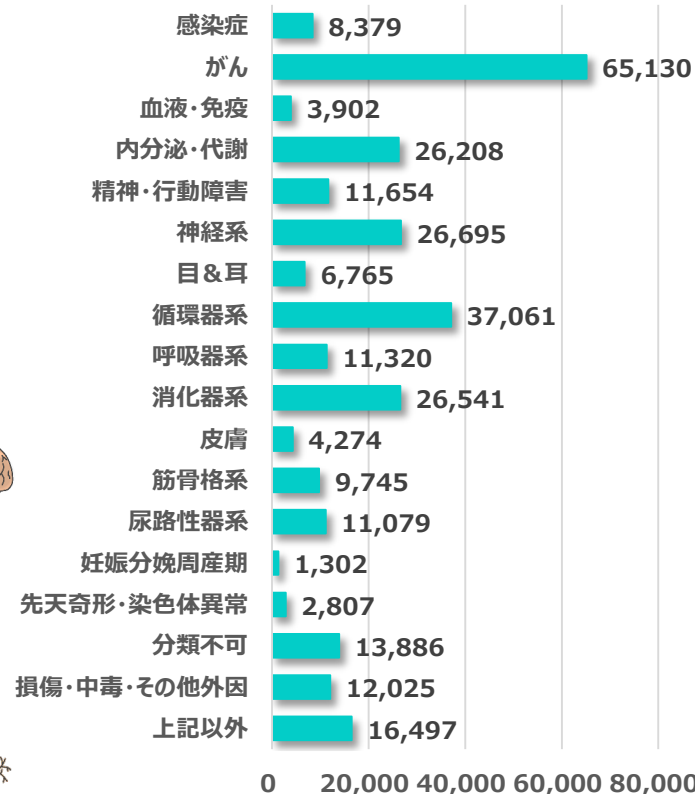
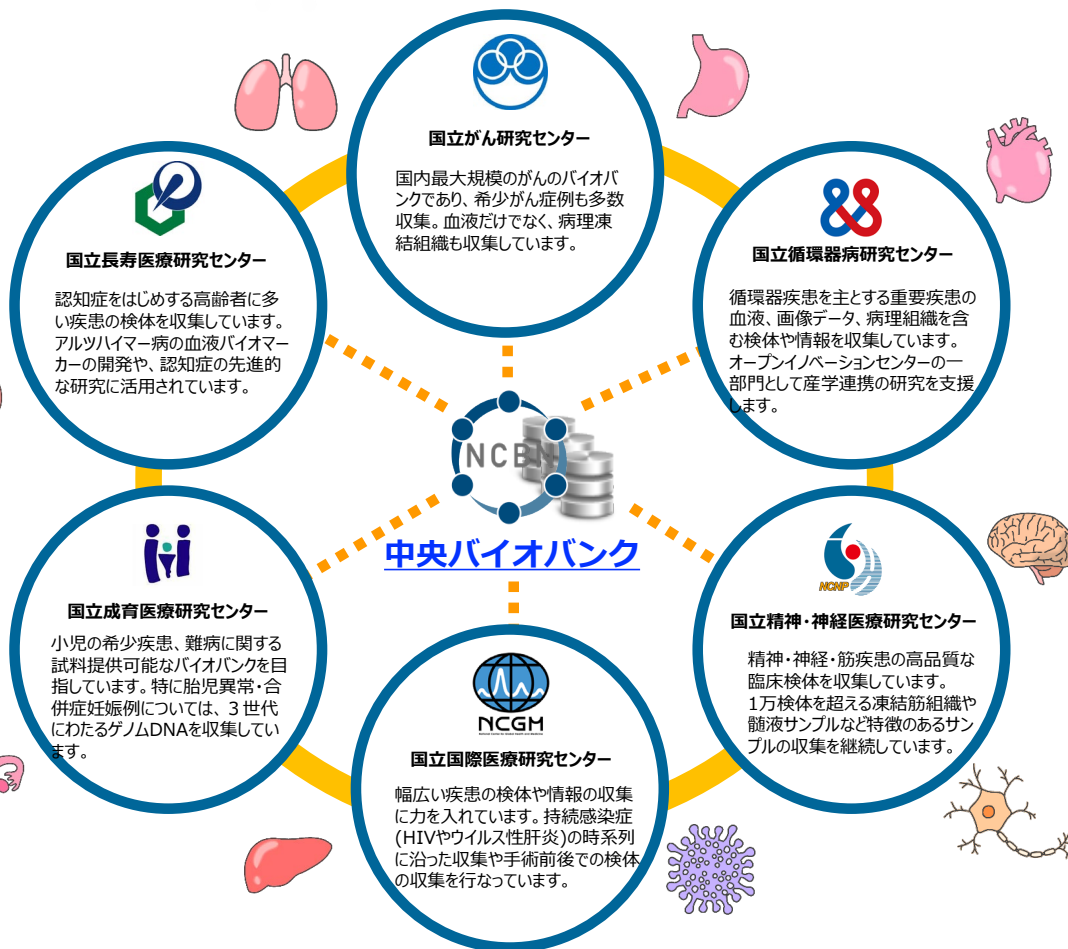
多因子疾患のWGS解析に必要なデータ解析環境

- コントロール群もケース群も全ゲノムデータの量が膨大になり、もはや輸送／転送が困難
- WGS-GWASでは必要な計算リソースが膨大
- 大規模データベース (UK Biobankなど)がvisiting環境へ移行
- 国際共同研究に必要な環境の変化
 - 国外の大量ゲノムデータの受け入れ
 - コンソーシアム研究で用いるパイプラインの導入
 - 海外からのvisiting環境へのアクセスを実現

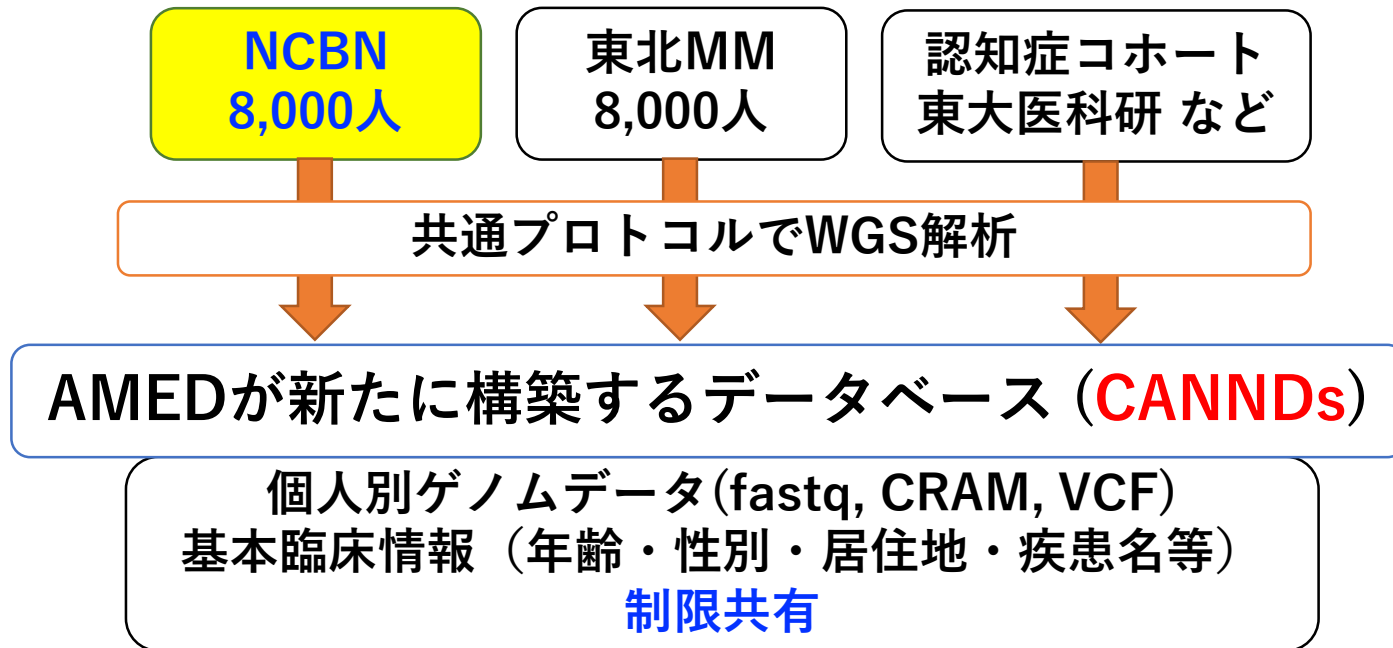
ナショナルセンターバイオバンクネットワーク(NCBN)

疾患患者検体、高度な専門性・詳細な臨床情報、追跡可能性、一括検索データベース、提供実績：>1,000回、10万検体

登録者数>12,200
試料数>416,000



コントロール群全ゲノム解析プロジェクト



5つのNCバイオバンク(NCGM, NCVC, NCGG, NCNP, NCCHD)が共同して、総計9,850検体の全ゲノムシーケンス解析を完了

がん、難病などのゲノム医学研究のため、研究者のゲノムデータ共有（検体毎の全ゲノムデータ）を促進

すでに11の共同研究を開始、まずは基本臨床情報で十分、必要に応じ詳細な臨床情報を各NCより提供可能