

2021年度第1回

ヒトゲノム研究倫理を考える会

AI・ビッグデータを用いた ヒトゲノム研究の倫理を考える

開催レポート



開催日時 2021.7.5(月) 15:00-17:00

開催形式 オンラインセミナー

開催レポート

2021年度第1回「ヒトゲノム研究倫理を考える会」
—パンデミック・ゲノム研究の倫理を考える—

日時：2021年7月5日(月)

ウェビナー（オンラインセミナー動画配信）

<https://www.genomics-society.jp/news/event/post-20210705.php/>

2021年7月5日に、第1回「ヒトゲノム研究倫理を考える会」がウェビナー（オンラインセミナー動画配信）にて開催された。近年、ヒトゲノム・臨床情報を含めたデータの蓄積が進み、ビッグデータと呼ばれるまでになっており、人工知能（AI）が研究や医療で用いられるケースも増大していることから、今回は「AI・ビッグデータを用いたヒトゲノム研究の倫理」をテーマとして、この分野に取り組んでおられる3名の方にご登壇いただいた。主催者（大阪大学大学院医学系研究科、加藤和人）の開会の挨拶の後、AI・ビッグデータを用いる研究の現状と課題について山口大学大学院医学系研究科の中津井雅彦氏、人工知能・機械学習によるゲノム解析研究の最新の動向について東北大学大学院医学系研究科の田宮元氏が講演を行い、神戸薬科大学社会科学研究所の小門穂氏が指定発言を行った。その後の質疑応答・総合討論では、ウェブ上の参加者も含め、活発な議論が繰り広げられた。

以下、各氏の講演、総合討論中の主な質疑の内容を簡単に紹介し、当日の発表資料を後ろに掲載する。また、講演の動画も併せて公開しており、ぜひご参照いただきたい。

中津井氏の講演タイトルは「AI・ビッグデータを用いたライフサイエンス・ゲノム研究の現状と課題」であった。最初に、AIに関する基本的な事項やこれまでの歴史や現状、ゲノム研究においてどのようにAIが活用されつつあるのかなどについて説明された。その後、「臨床ゲノム情報統合データベース（MGeND）」、「AI創薬コンソーシアム（LINC）」で現在行われている取り組みについて紹介された。最後に、画像判別のAIが摂動によって誤った判断をしてしまう例を挙げ、こうした課題があることも認識しておく必要があると話された。

田宮氏の講演タイトルは、「人工知能・機械学習による大規模ゲノムコホート・バイオバンク解析」であった。最初に、ゲノムワイド関連解析（GWAS）研究のこれまでの流れについて紹介された。ポリジェニック・リスク・スコア（PRS）の概要について説明があった後、これまでに開発されてきた手法、機械学習の活用に関して紹介された。最後に、PRSは、民族背景の違いにより形質によっては正確な結果が予測できない場合があり、これまでゲノム研究の主要な対象となつてこなかった民族集団では精密医療にアクセスできない可能性が示唆されていること、これに対して、日本人のゲノムデータの蓄積など、今後の検討や取り組みの必要性について言及された。

小門氏の指定発言では、社会の信頼を得た上で、社会の共有財であるデータを活用し、発展させていくことの重要性が指摘された。医療倫理の観点から、技術の利用に

おける格差や差別への懸念、またそれらがきちんと対処されるかという懸念をいかに払拭するかが、社会の信頼を得るために必要ではないかと話された。また、こうしたことについて患者や当事者を含めた社会全体で考えていくことが重要なのではないかと発言された。

続いて、参加者を交えて質疑応答・総合討論が行われた。事前およびウェビナー中に寄せられた質問をもとに活発な議論が交わされた。その際取り上げられた主な質問や議論は以下の通りである。

- ビッグデータの取り扱いと AI で得られる結果の理解
- クラウドなど施設外でのデータの取り扱いやセキュリティ、産学連携での二次利用を含めたデータの取り扱い
- 教師データとして利用する場合の同意内容
- データ利用の際のセキュリティ条件の現状
- 被験者の募集の際にはメリットだけでなくデメリットも公正に記載するべきではないか
- 未来の医療に必須であるこの分野のリテラシーを向上させていく仕組み・設計のあり方
- ポリジェニック・リスク・スコアの臨床応用の時期に関する現在の予測
- 医用画像に関する AI について、画像データのみでなく、その周りのメタデータの構造化が貢献しているのか、画像自体の処理が AI に向くのか
- AI を用いた顔認証で差別につながらないのか
- 遺伝子バリエーションの機能解析がされていないものが多く、当事者に情報を伝えられないのはとても残念だ
- 日本人基準ゲノム JG1 の解析対象者の出身地
- リファレンス作成等は、ヨーロッパ系やアフリカ系との違いを見るという観点から、中国と協力しての研究は行われていないのか
- ゲノム解析研究におけるジェンダーバランスの扱い

最後に加藤教授より、2つの講演で、AIに関する将来の明るい展望だけではなく現状での課題についても言及があったことで、倫理について考えている人々への良いメッセージにつながったこと、技術の臨床応用が活発になる前の段階で今回のテーマについて取り上げ、検討することができたことの意義に関して発言があり、多くの方々にご参加いただいたことへの感謝の言葉の後、閉会となった。

AI・ビッグデータを用いた ライフサイエンス・ゲノム研究の 現状と課題

中津井 雅彦

山口大学大学院医学系研究科・医学部附属病院

AIシステム医学・医療研究教育センター



YAMAGUCHI
UNIVERSITY

2021年度第1回ヒトゲノム研究倫理を考える会

「AI・ビッグデータを用いたヒトゲノム研究の倫理を考える会」

人工知能が人間を超える？



2011年 人工知能ワトソンが
史上最強クイズ王に勝利



2016年 囲碁の人工知能が
世界トップクラス棋士に勝利

「AI」 「人工知能」

人間の知能を人工的に実現する

1956年 米国東部 ダートマス会議 人工知能に関する世界初の国際会議
「人工知能 (Artificial Intelligence)」の命名
ジョン・マッカーシー

明確な定義はない

人工知能専門家による定義

東京大学・松尾豊先生

「人工的に作られた人間のような知能、ないしはそれを作る技術」

公立はこだて未来大学・松原仁先生

「究極には人間と区別がつかない人工的な知能のこと」

大阪大学・浅田稔先生

「知能の定義が明確でないので、人工知能を明確に定義できない」

3

「AI」の整理



強いAI

心を持つ

強いAIはまだできていない！
実現しているのは弱い人工知能



弱いAI

限定された知能によって一見知的な問題解決が行われれば良い

4

人工知能が人間を超える？

(再掲)



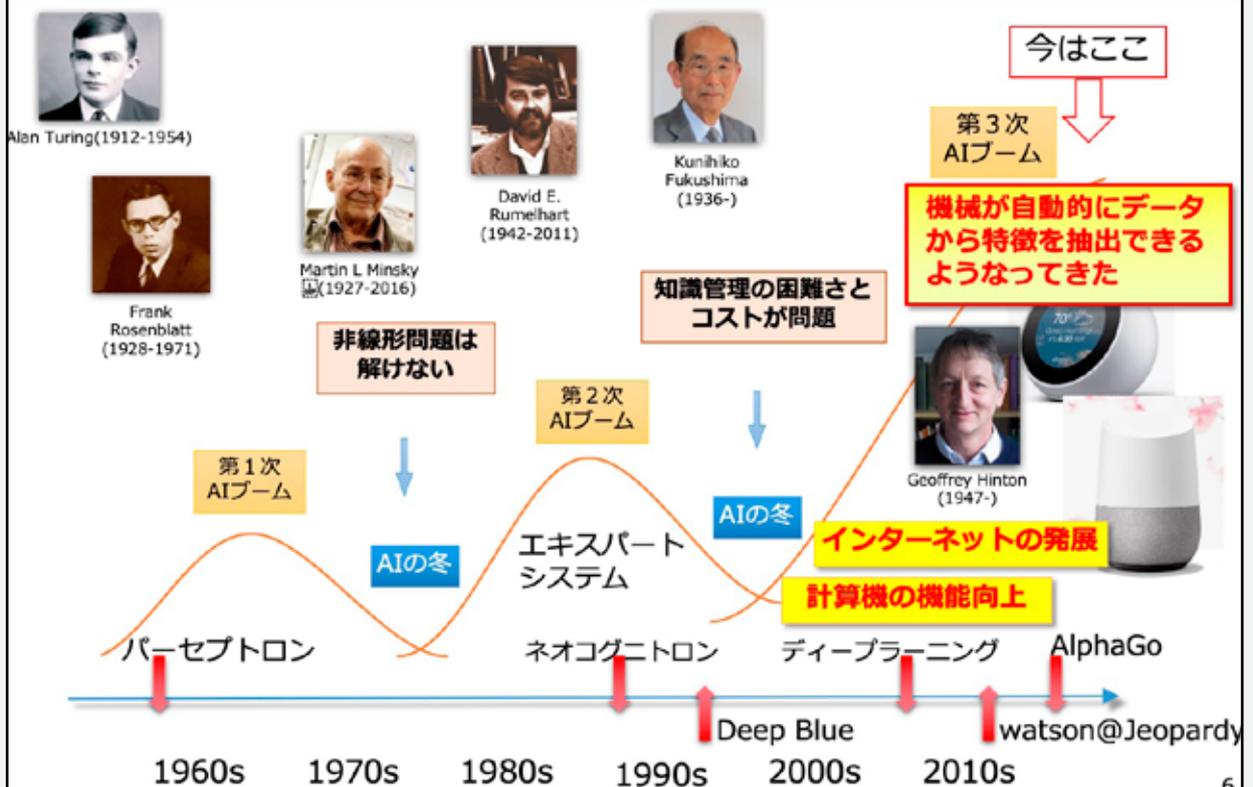
2011年 人工知能ワトソンが
史上最強クイズ王に勝利



2016年 囲碁の人工知能が
世界トップクラス棋士に勝利

5

AIブーム



6

第一次・第二次AIブームでの課題

第一次AIブーム：推論、探索

経路探索やチェスなど厳密に定義されたルールの中では通用するが、現実の問題はもっと複雑で解けない

第二次AIブーム：「知識」の活用

知識をどう表現するのか、知識をどのように計算機に入力していくのが課題
知識を書ききれない
扱う知識をコンピュータが決められない

特徴量設計の問題

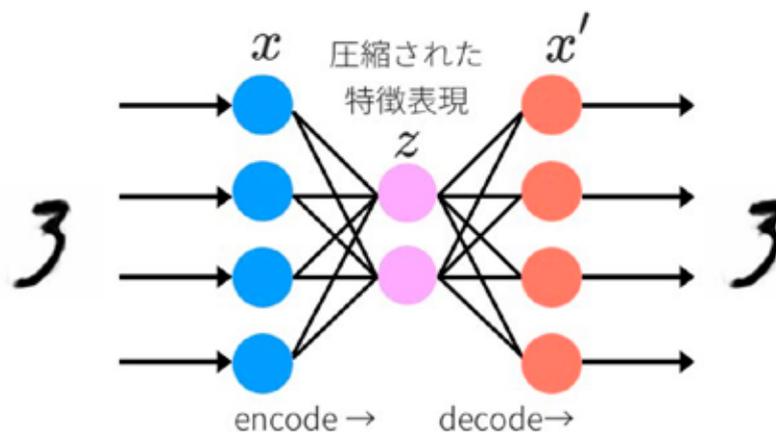
「世界からどの特徴に抽出して情報を取り出すか」を人がやらなければならない

コンピュータが与えられたデータから注目すべき特徴を見つけ、その特徴の程度を表す特徴量を得ることができればクリアされる

7

自己符号化器 (AutoEncoder, 2006)

入力と同じものを出力するニューラルネットワーク



一見無意味に見えるが、

- 次元圧縮に活用できる
- **データを表現する特徴を自動的に抽出できる**

8

第三次AIブームを支えるもの



カンブリア爆発
5億3000万年前

「目」の獲得

- インターネットの発明、情報共有
- 電子カルテの導入
- 検査機器、センサーの進化
- IoTの普及

「脳」の獲得

ハードウェア

- 「ムーアの法則」、半導体集積度の向上
- 並列化技術の進歩
- ゲーム業界でのGPUの進歩、GPGPU

ソフトウェア

- 数学・統計・推論・知識の基盤技術
- ニューラルネットワークに関する理論・技術
 - 特徴の自動抽出
 - 特徴のカテゴリライズ

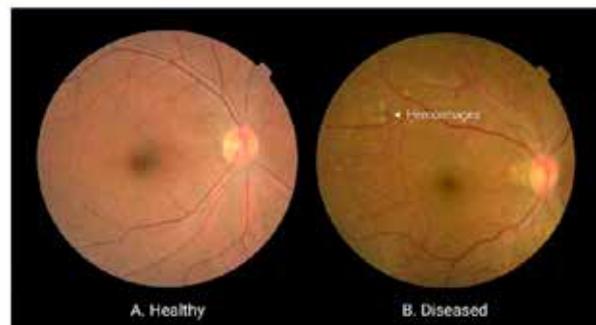
9

糖尿病性網膜症 (2018. 4)



IDX-DR Deepmind

2018, FDAが診断可能な初の検査機器として認可

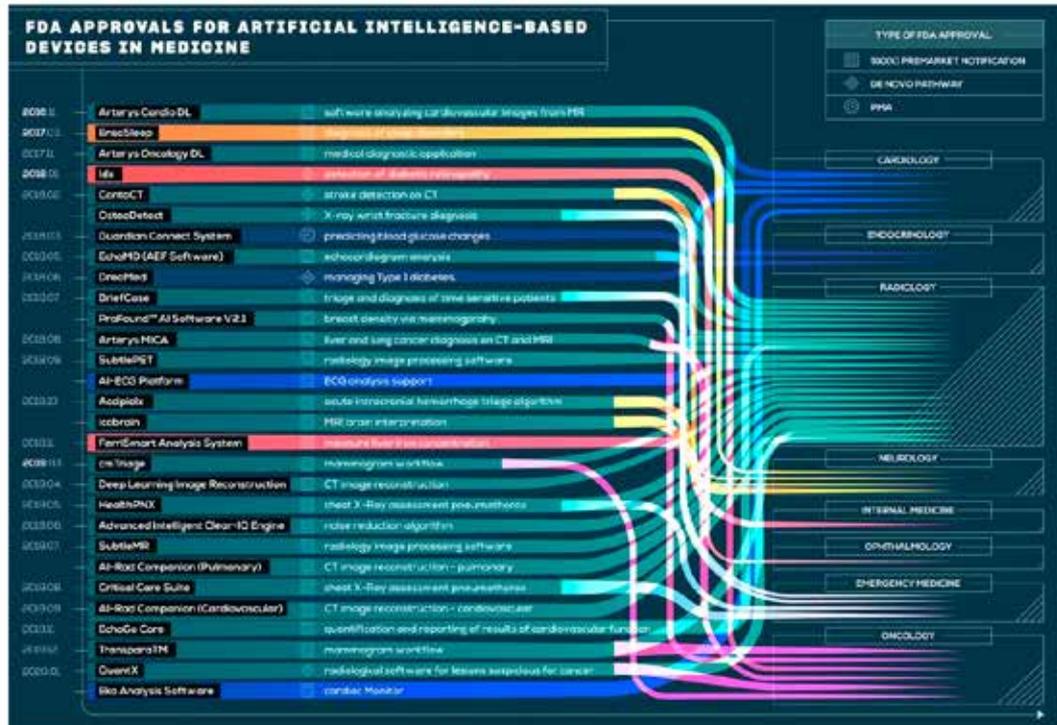


糖尿病性網膜症をスクリーニングするために用いられた網膜画像の例。
左が健康の網膜の眼底、右が増殖前網膜症の眼底。硝子体出血がみられる。

感度 97%, 特異度 87%

10

FDAにより認可されたAI/ML医療機器



Benjamins, S., Dhunoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* 3, 118 (2020). <https://doi.org/10.1038/s41746-020-00324-0>

AIシステム医学・医療研究教育センター (AISMEC)

AISMEC
AI Systems Medicine
Research and Training Center

多層的な生命機能
基礎医学・臨床医学
AIシステム医学・医療研究教育センター

システムバイオロジー
人工知能機械学習
医学・医療
創薬・治療
創薬・治療

山口大学



- システムバイオインフォマティクス講座 (浅井)
- 公衆衛生・予防医学講座 (田邊)
- 医療情報判断学講座 (石田)
- システム再生病態医学講座 (清木)

顧問：北野宏明先生 山口大学客員教授
(SONY コンピュータサイエンス研究所所長)



ゲノム研究における人工知能

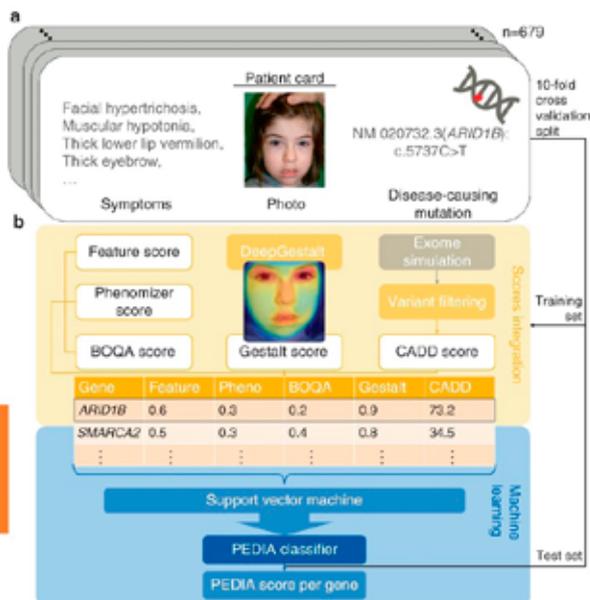
Phenotype-to-genotype mapping

Image to genetic diagnosis

DeepGestalt (CNNベースの顔画像解析アルゴリズム),
 PEDIA

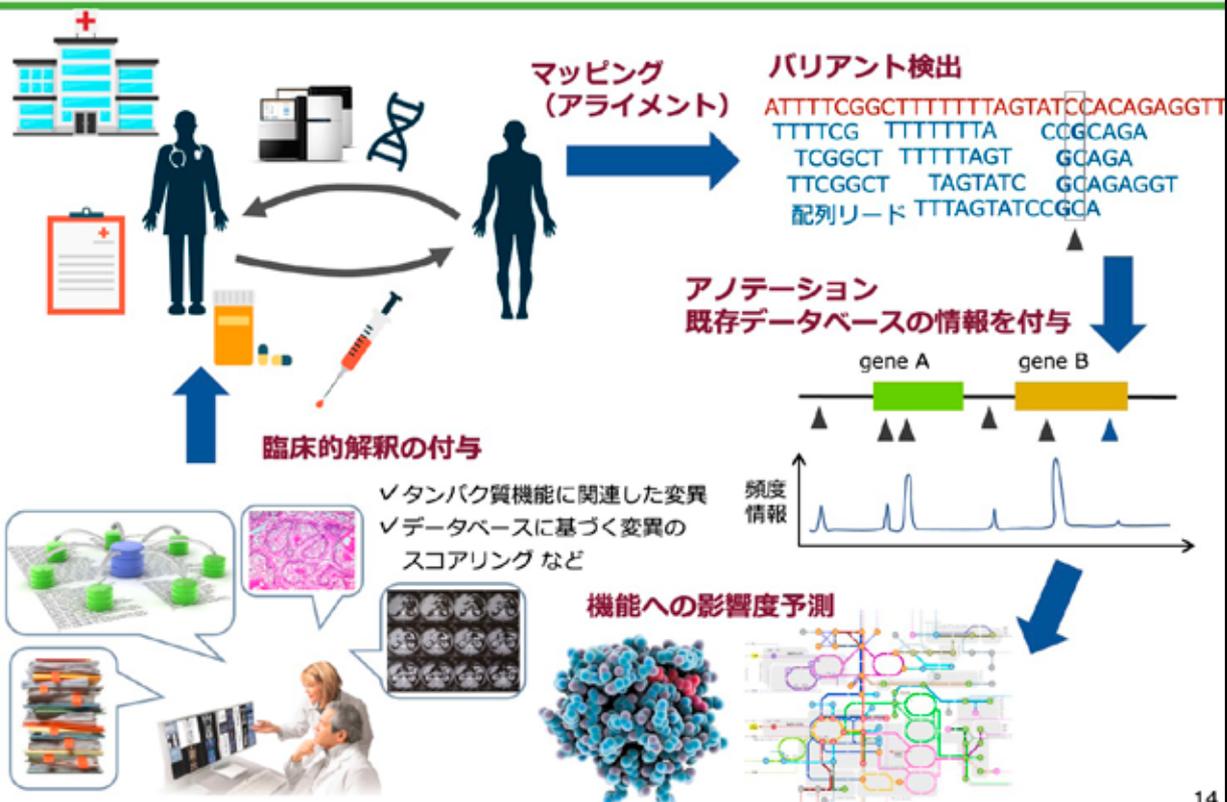
顔写真から抽出された表現型の特徴を使用して、単一遺伝子障害の候補バリエントを正確に優先順位付け
 (679人、105疾患)

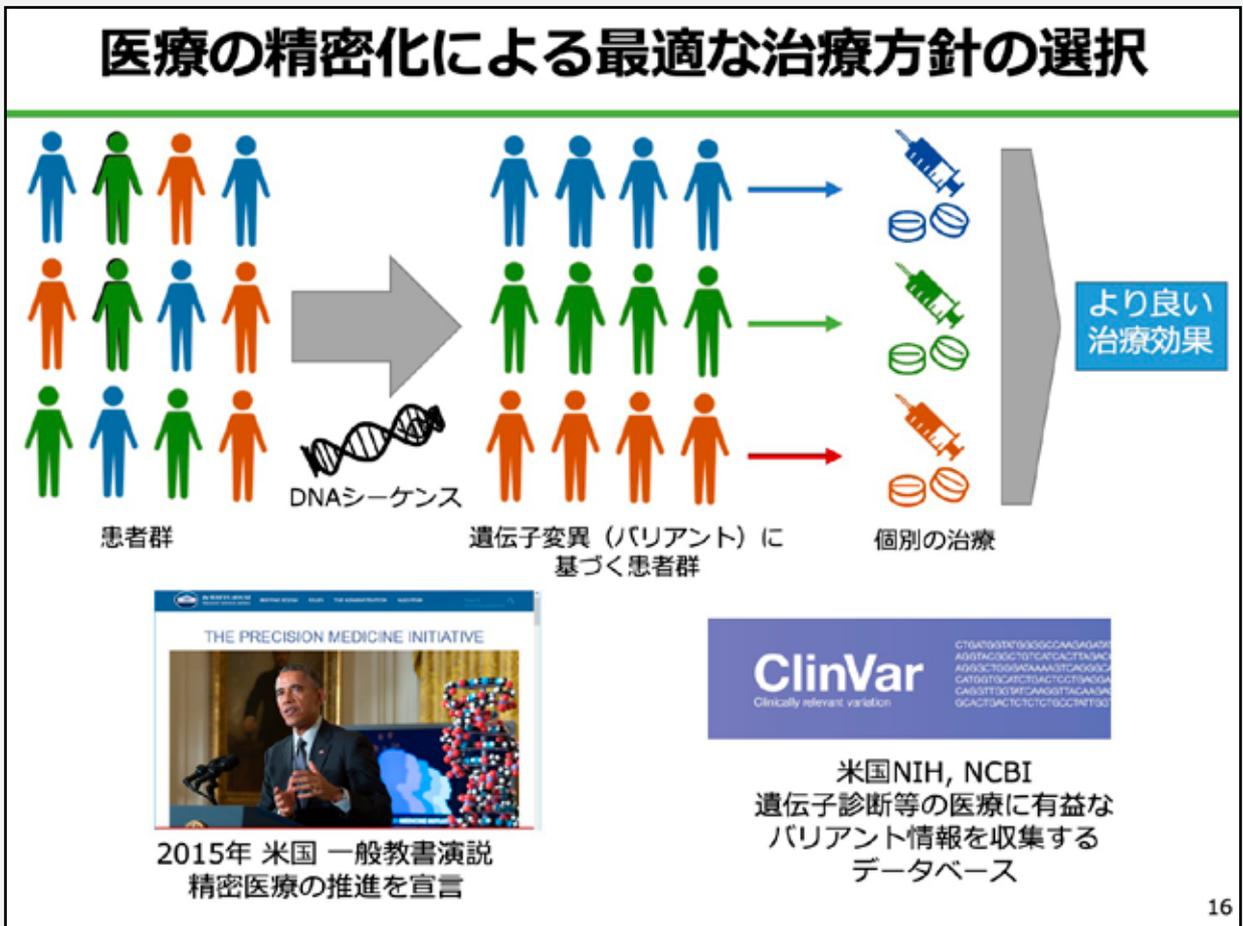
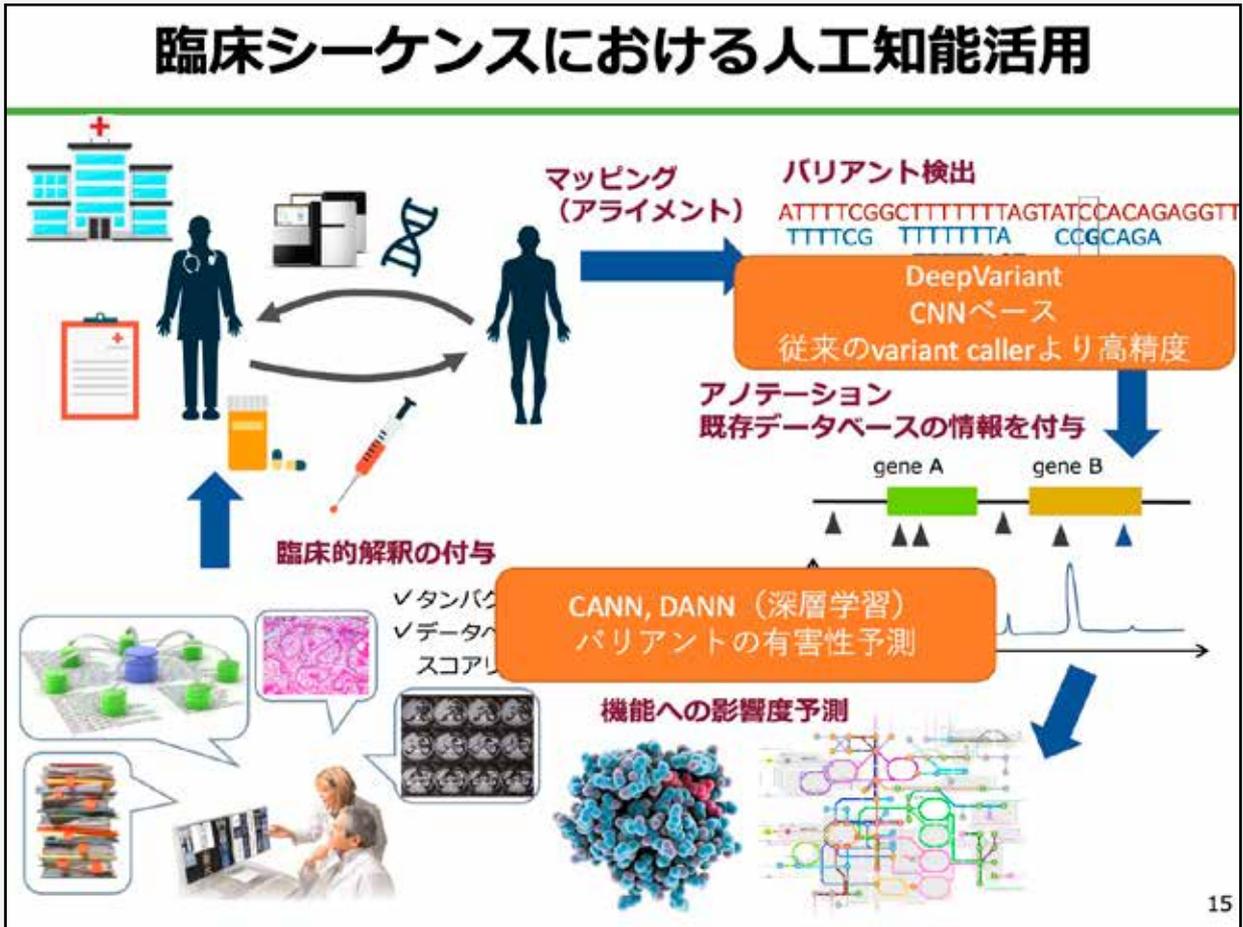
AIベースのコンピュータービジョンシステムは、画像に埋め込まれた複雑な表現型に基づいて、個人のゲノムに存在するゲノム異常を予測できる可能性がある



Hsieh T-C, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, et al. PEDIA: prioritization of exome data by image analysis. Genet Med. 2019.

ゲノム医療における臨床シーケンスの流れ





バリアントの疾患関連性の評価

ACMGガイドライン

米国臨床遺伝・ゲノム学会 (American College of Medical Genetics and Genomics; ACMG)

根拠の強さ

	Design		Pathogenic				Pathogenicity	Rational	Category
	Strong	Supporting	Supporting	Moderate	Strong	Very strong			
Population data	MAF is too high for disorder BA/BSI OR observation in controls inconsistent with disease penetrance BSI			Absent in population databases PM2		Prevalence in affecteds statistically increased over controls PS4		Very Strong	PVS1
Computational		Multiple lines of	Multiple lines of	None observed change		Same amino acid		Strong	PS1 PS2 PS3
		Observed in cis with a pathogenic variant BP2		Reputable source in trans with a pathogenic variant P03					BS3 BS4 BP1 BP2 BP3 BP4 BP5 BP6 BP7
Other database		Reputable source without shared data = design BP6	Reputable source = pathogenic PPS				Benign		
Other data		Found in case with an alternate cause BPS	Pathology phenotype or FIM highly specific for gene P04				Supporting		

臨床的意義のカテゴリ

※メンデル遺伝病に関わるバリアントに対してACMG/AMPで推奨される用語

Benign … 病原性なし

Likely benign … 病原性なしの可能性が高い

Likely pathogenic … 病原性ありの可能性が高い

Pathogenic … 病原性あり

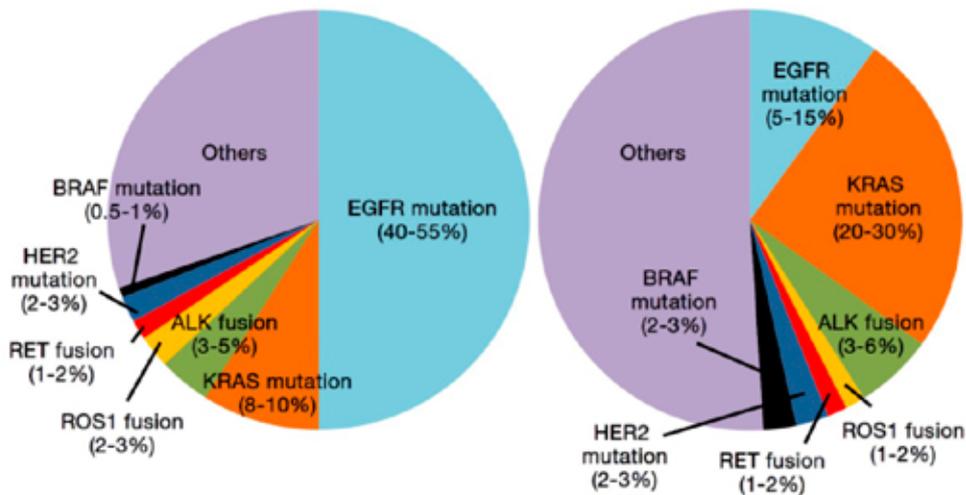
Uncertain significance … 解釈不明 (上記のいずれにも分類できなかったもの)

Richards, S. et al., 2015. Genet. Med. 17 (5): 405-424.

地域によって、遺伝的特徴が異なる

東アジア (日本・韓国・中国)

米国・ヨーロッパ



肺腺癌 (LADC) におけるドライバー遺伝子の変異頻度の比較

Kohno, T. et al., Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer, Transl. Lung Cancer Res. 2015;4(2):156-164

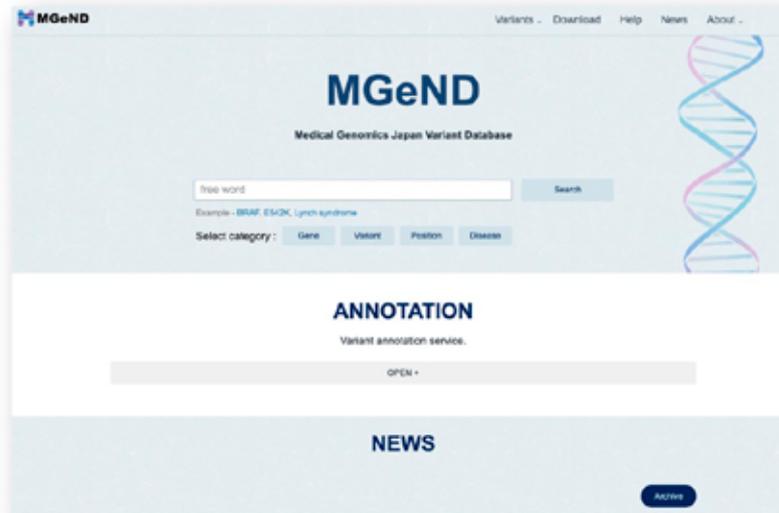
日本人の遺伝的特徴を反映したデータベースが必要

臨床ゲノム情報統合データベース

MGeND (Medical Genomics Japan Variant Database)

<https://mgend.med.kyoto-u.ac.jp/>

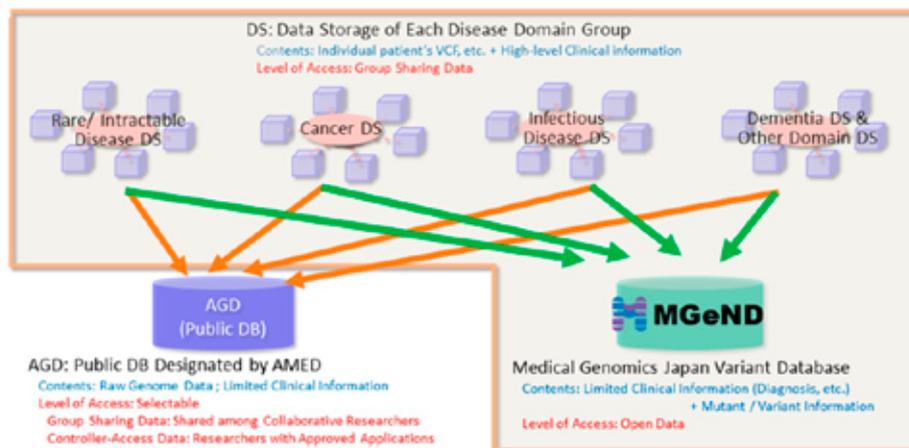
- ◆ 日本人疾患ゲノムの集積と世界への発信
- ◆ キュレーションに必要な情報の提供と臨床現場利用
- ◆ 多種多様な疾患やゲノム情報の統合的なデータベース化
- ◆ 疾患横断的な検索



19

AMED 臨床ゲノム情報統合データベース整備事業

AMED…国立研究開発法人 日本医療研究開発機構



AMEDデータシェアリングポリシー

- 制限共有データ
 - ✓ DMPに記載された研究者およびデータアクセス申請を承認された研究者と共有するデータ
 - ✓ 対象となる研究者は、データの生産や蓄積・活用に貢献できる研究者等
- 制限公開データ
 - ✓ 当該データベースの規約等に従って、利用目的、利用方法を明らかにしたうえで、データアクセス申請を承認された研究者が利用することが可能なデータ
- 非制限公開データ
 - ✓ アクセスに制限なく誰でも利用することが可能なデータ

MGeND

20

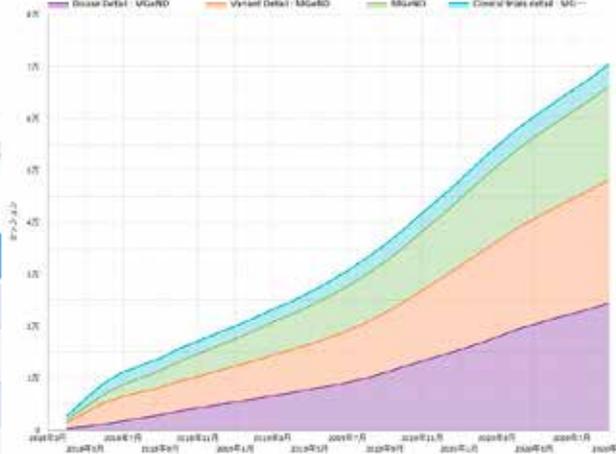
アクセス・利用ユーザーについて

(2018/03/18 -- 2020/10/10)

- トータルセッション数 : 80,868
- 新規ユーザー (ユニーク) : 50,407 (49,982)

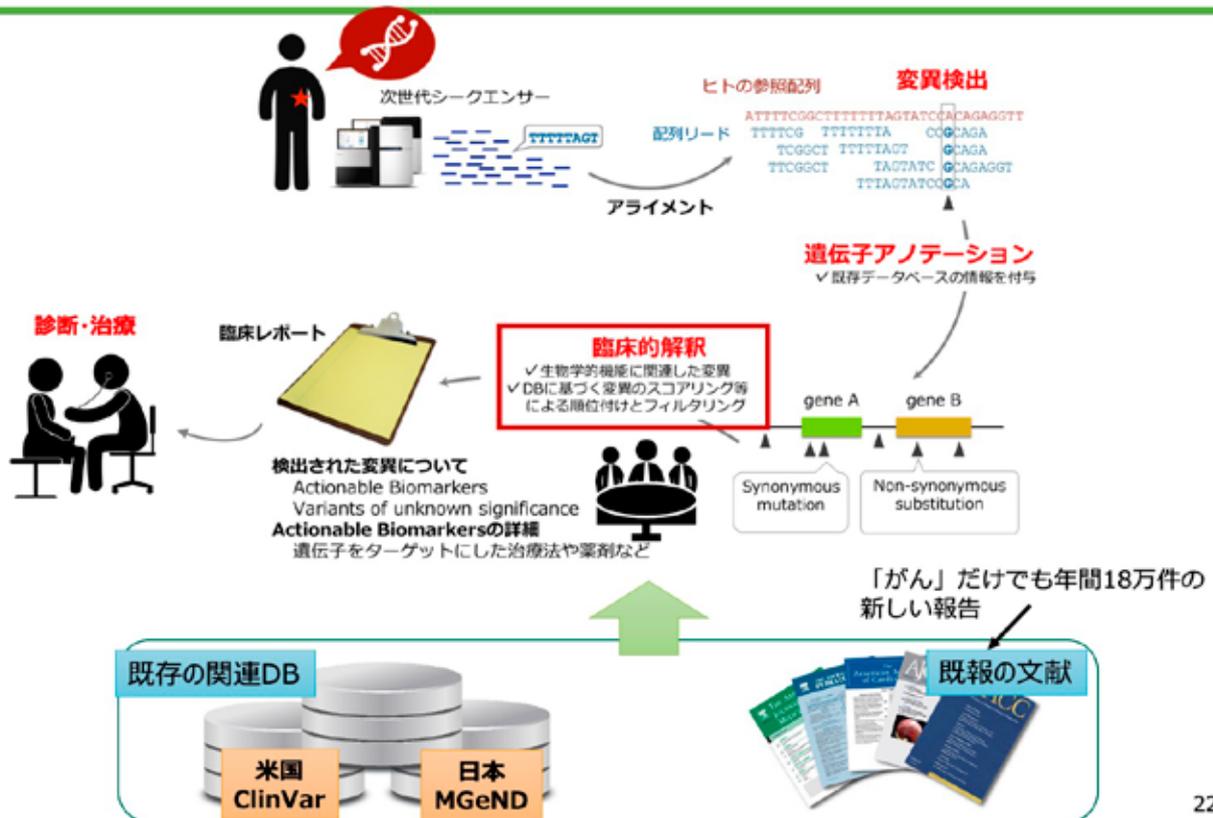


	Country
1	United states
2	Japan
3	India
4	United Kingdom
5	Germany



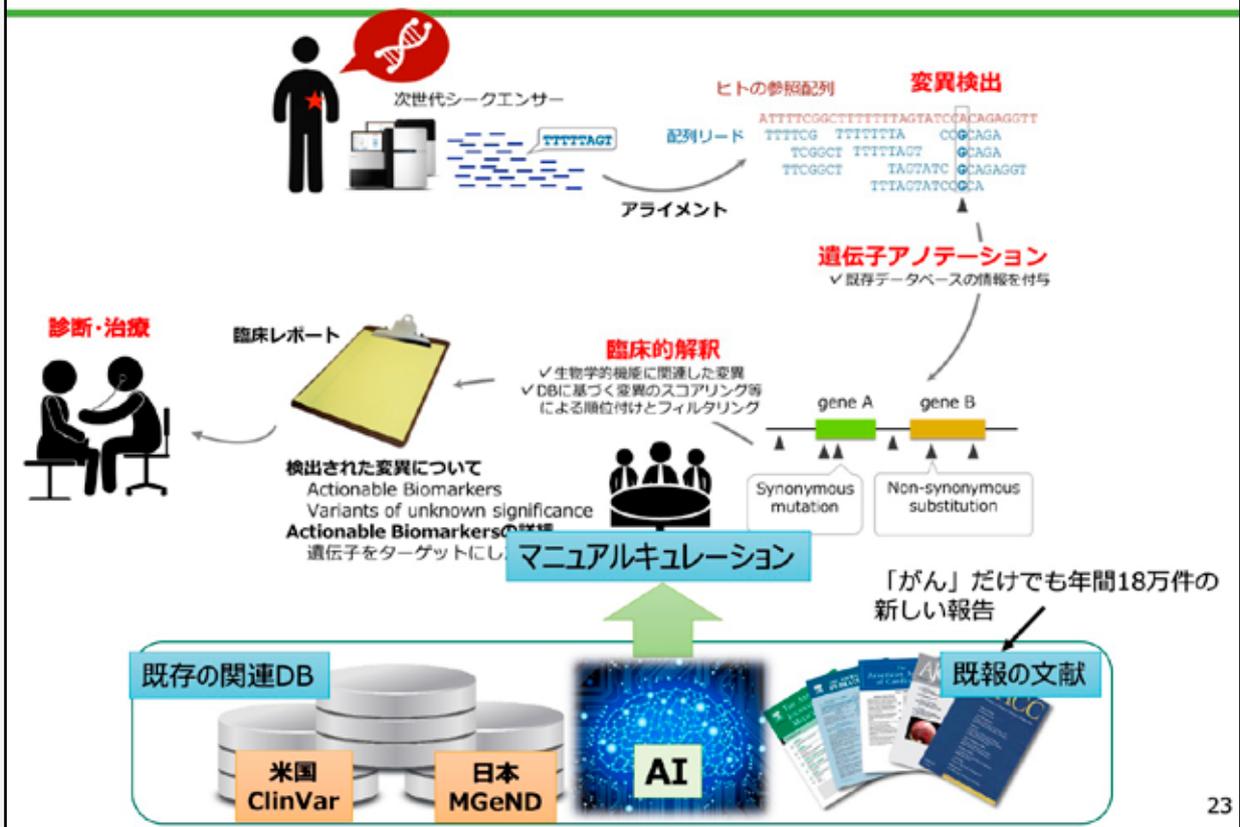
21

ゲノム医療実践に向けての課題



22

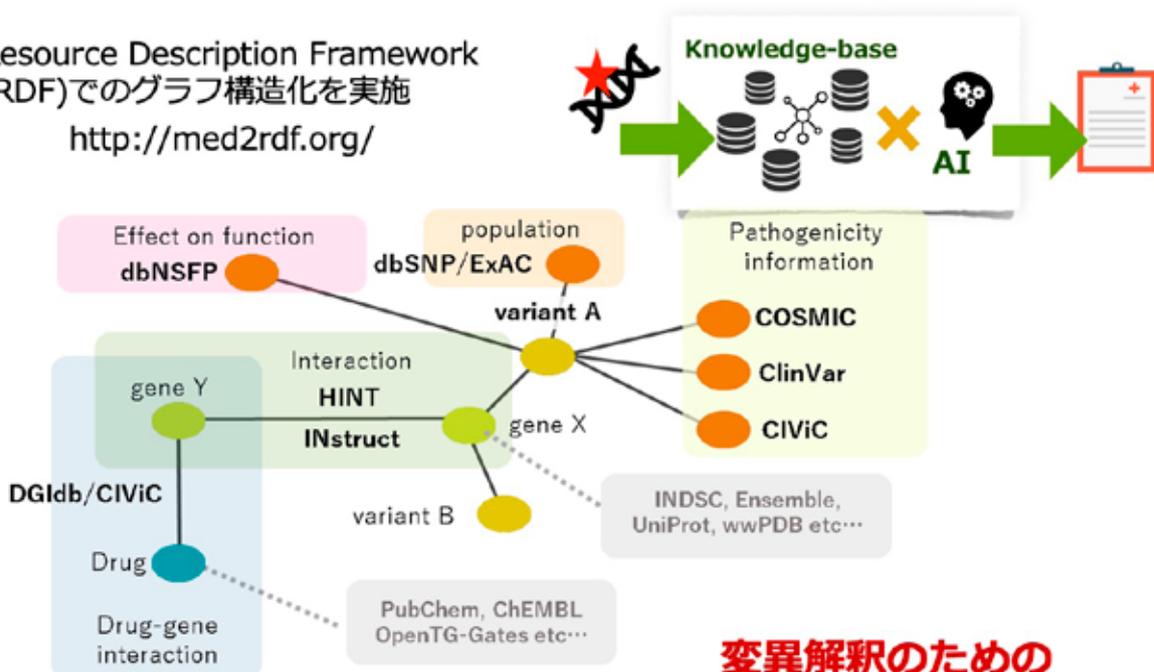
AIによるバリアントの臨床的解釈の支援



23

多岐にわたるデータベースを結合し、AIの知識基盤を構築する

Resource Description Framework (RDF)でのグラフ構造化を実施
<http://med2rdf.org/>



**変異解釈のための
 知識ベースとして整備
 (これまでに18データベース)²⁴**

MGeND Intelligence: AIによるバリアントの疾患関連性予測

AIによるPathogenicity予測スコア

Variant name	AA change	Origin	Sample number	MGeND annotation	AI Pathogenicity Score	DB	SeqEff impact	dbNSFP predictor
chr7:g.140481402C>T (rs121913355)	BRAF p.G469E		1		High (5 red bars)	dbSNP	missense_variant, sequence_feature	Multiple predictors
chr7:g.140501336C>G (rs180177034)	BRAF p.A246P		1		High (5 red bars)	dbSNP	missense_variant	Multiple predictors
chr7:g.140477853C>G (rs180177036)	BRAF p.L480F		1		Medium (3 orange bars)	dbSNP	missense_variant	Multiple predictors
chr7:g.1404	BRAF		1		High (5 red bars)	dbSNP		Multiple predictors

25

MGeND Intelligence: 関連文献の提示

Related Papers

自然言語処理による
関連論文の提示

Title	Citation	R
Epithelial-Mesenchymal Transition and Somatic Alteration in Colorectal Cancer with and without Peritoneal Carcinomatosis	BioMed Research International Hindawi Publishing Corporation 2014-2014-8-3	★★★★★
BRAF-activated long non-coding RNA contributes to cell proliferation and activates autophagy in papillary thyroid carcinoma	Oncology Letters D.A Spandidos 8-5-2014-8-28	★★★★★
Current Understanding of BRAF Alterations in Diagnosis, Prognosis, and Therapeutic Targeting in Pediatric Low-Grade Gliomas	Frontiers in Oncology Frontiers Media S.A. 5-2015-3-03	★★★★★
Targeting the Checkpoint to Kill Cancer Cells	Biomolecules MDPI 5-3-2015-8-18	★★★★★
Similar but different: distinct roles for KRAS and BRAF oncogenes in colorectal cancer development and therapy resistance	Oncotarget Impact Journals LLC 6-25-2015-7-30	★★★★★

Related Drug Papers

Drug	Disease	Variation	Sensitivity	Evidence Level	References
vemurafenib	ランゲルハンス細胞腫瘍	活性化変異	感受性	2A	一頁ページ
	悪性黒色腫	活性化変異	感受性	1A	皮膚悪性腫瘍診療ガイドライン
	悪性黒色腫 (BRAF V600E変異)	融合	無反応性	3	一頁ページ
	非小細胞肺癌	活性化変異	感受性	2A	一頁ページ
	皮膚癌	活性化変異	感受性	1A	皮膚悪性腫瘍診療ガイドライン
	悪性黒色腫	活性化変異	感受性	1B	一頁ページ
	悪性黒色腫	融合	感受性	3A	一頁ページ

薬剤情報とともに
関連論文・ガイドライン提示

26

推定根拠を説明できるAI



京都大学
KYOTO UNIVERSITY



2021年3月11日
富士通株式会社
国立大学法人京都大学

富士通と京都大学、AIの推定根拠を説明できる遺伝子変異の病原性推定の検証システムを開発

ゲノム医療における未知の遺伝子変異に基づく治療検討を支援

富士通株式会社(注1)(以下、富士通)と国立大学法人京都大学(注2)(以下、京都大学)大学院医学研究科の奥野恭史教授らの研究グループは、国立研究開発法人日本医療研究開発機構(以下、日本医療研究開発機構)の「臨床ゲノム情報統合データベース整備事業」(注3)において、特定の遺伝子変異(注4)が「何らかの疾患の原因になる可能性(以下、病原性)の有無を推定し、その推定根拠を添えて治療方針を検討する医師やゲノム医療の研究者に示すことができるAIを活用した検証システム「MGeND Intelligence」(エムジェンド インテリジェンス)を開発しました。京都大学は、2021年4月より、共同研究者や協力機関などに対して「MGeND Intelligence」の利用提供を予定しています。

「MGeND Intelligence」は、遺伝子変異の病原性の有無を高精度に推定できる病原性推定AI、推定の根拠を示す説明可能AI、関連する論文記載の検索を支援する文献探索支援AIから構成されます。京都大学が2018年に公開した臨床ゲノム情報統合データベース「MGeND」と連携し、病原性の有無が未知の変異も含めて医療従事者や研究者による遺伝子変異の調査・臨床的解釈の作業を支援します。

「MGeND Intelligence」を利用することで、がんをはじめとしたゲノム医療での治療方針検討など、診療支援につながるのと同時に、患者にとって最適な医療提供が加速されることが期待できます。

27

AI創薬コンソーシアム LINC

2016年11月～2020年9月、産官学 110団体が参画するAIコンソーシアム

予防・先制医療

WG1. 未病・先制医療

- 健康診断データによる発症予測
- マイクロバイオーム・オミクスデータ解析
- デジタルヘルス

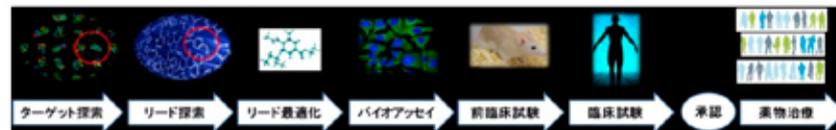
メディシナルケミストリー

WG4. 分子シミュレーション

- タンパク質立体構造・機能予測
- AIによるドッキング計算高度化
- 分子動力学計算によるAI活用
- AIを用いた高精度分子力場

WG5. メドケム・分子設計・ADMET

- 合成経路予測
- 分子設計AI
- 化合物記述子表現
- QSAR/QSPR/ in vitro ADMET予測



WG3. 創薬テーマ創出

- 有望提案先や研究テーマの自動探索
- 標的分子探索
- ドラッグリポジショニング

WG6. トランスレーショナルリサーチ

- 非臨床データからのヒトADMET予測
- 疾患メカニズム解明・ブリッジング予測

WG2. 臨床・診断

- がんゲノム医療におけるAI活用
- シミュレーションによる細胞分離
- AIによる病理画像処理
- AIによる電子カルテ処理

WG7. バイオリジクス・製剤・ロボティクス

- バイオリジクス関連AI
- 結晶形・製剤関連AI
- 調剤ロボティクス

WG8. 治験・市販後・メディカルアフェアーズ

- AIによる治験の効率化
- 有害事象の情報基盤
- 製品Q&Aシステム
- アウトカムリサーチ・医療技術評価

バイオメディカル・基礎から臨床への開発プロセス

診断・治療

WG9. 知識ベース・NLP

- 知識ベース

知識ベース / 計算機サーバー

WG10. AI基盤

- ライフサイエンスのためのAI基盤

28

AI創薬コンソーシアム LINC

H28年 12月開始 テーマ提案、調査、AI設計

H29年7月 AI開発開始

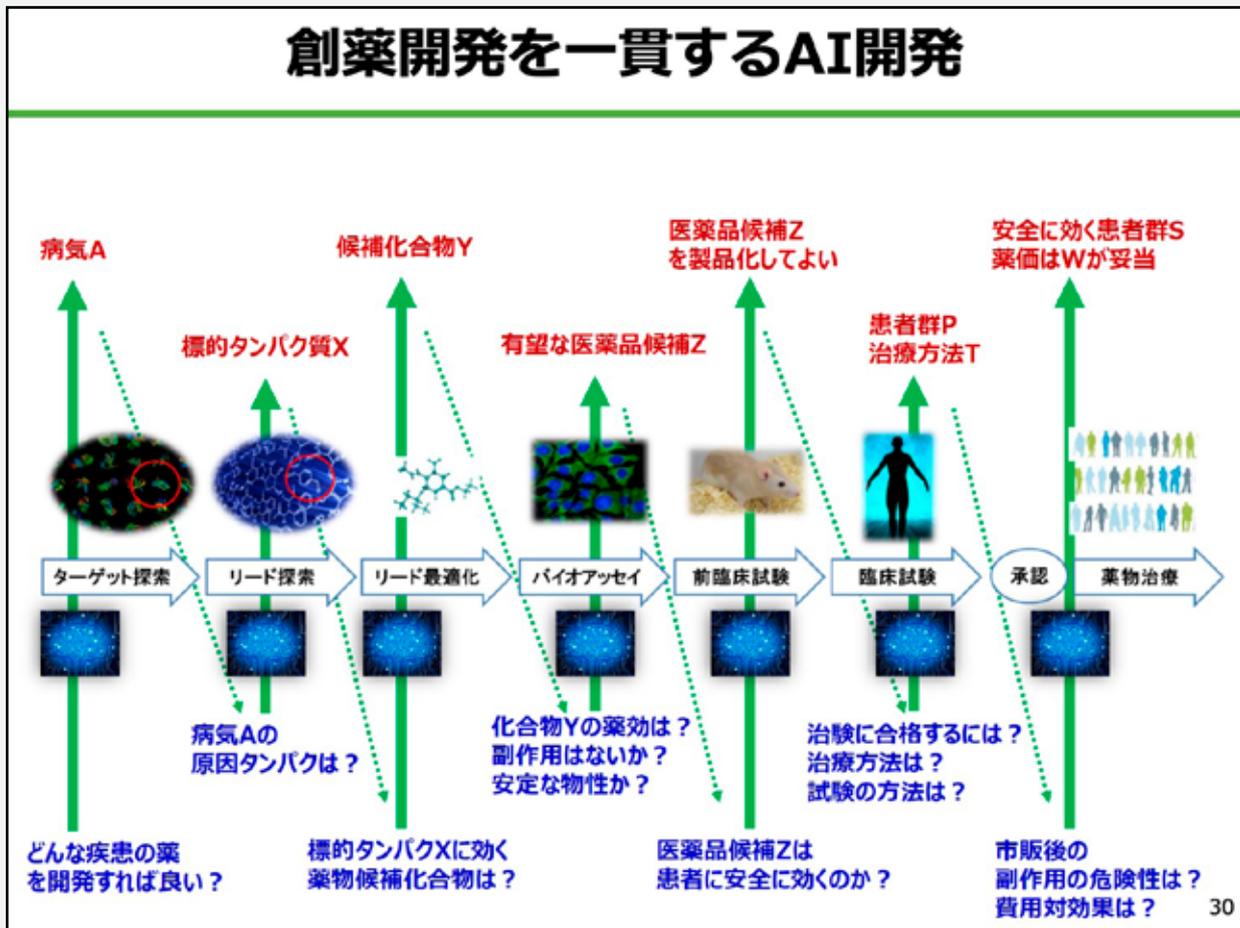
創薬AIで50社連
武田やNEC 新薬探し

創薬AI開発へ
期間・費用を半減
京大と企業70社

ライフインテリジェンスコンソーシアム (LINC)
京大・理研・医薬健栄研等、ライフ系企業、IT系企業等
約130業・団体

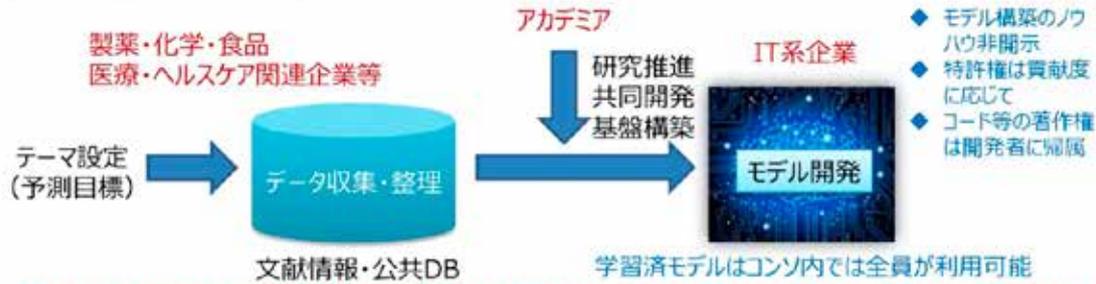
R3年4月 社団法人化

29

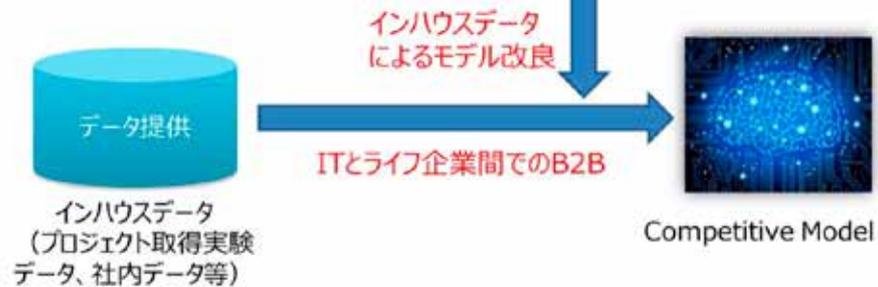


ライフ・インテリジェンスコンソーシアム (LINC)

Pre-Competitive Area 参加費は無料だが、参画機関はマンパワーを持出し、協調してAIを開発すること



Competitive Area



約30PJのほとんどでプロトタイプ構築
製品化5件、特許化1件
論文4報、学会発表37件

LINC
成果事例集
- 3年間のアウトプット -

2021.02



旧LINCの成果と反省点を踏まえ、
2021年4月に新LINCとして社団法人を立ち上げ

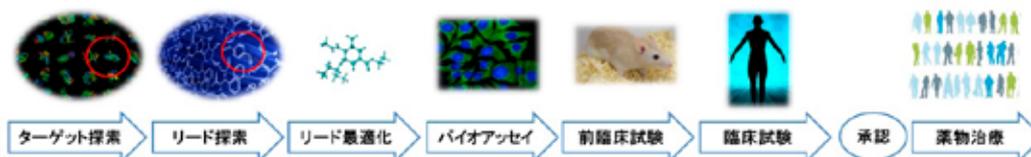
LINCのWG

WG03. バイオロジクス・ニューモダリティ

WG01. デジタルヘルス&メディスン・RWD解析

WG04. 分子シミュレーション

WG05. メディシナルケミストリー



WG02. 標的探索・バイオインフォマティクス

WG07. ADMET・トランスレーショナルリサーチ

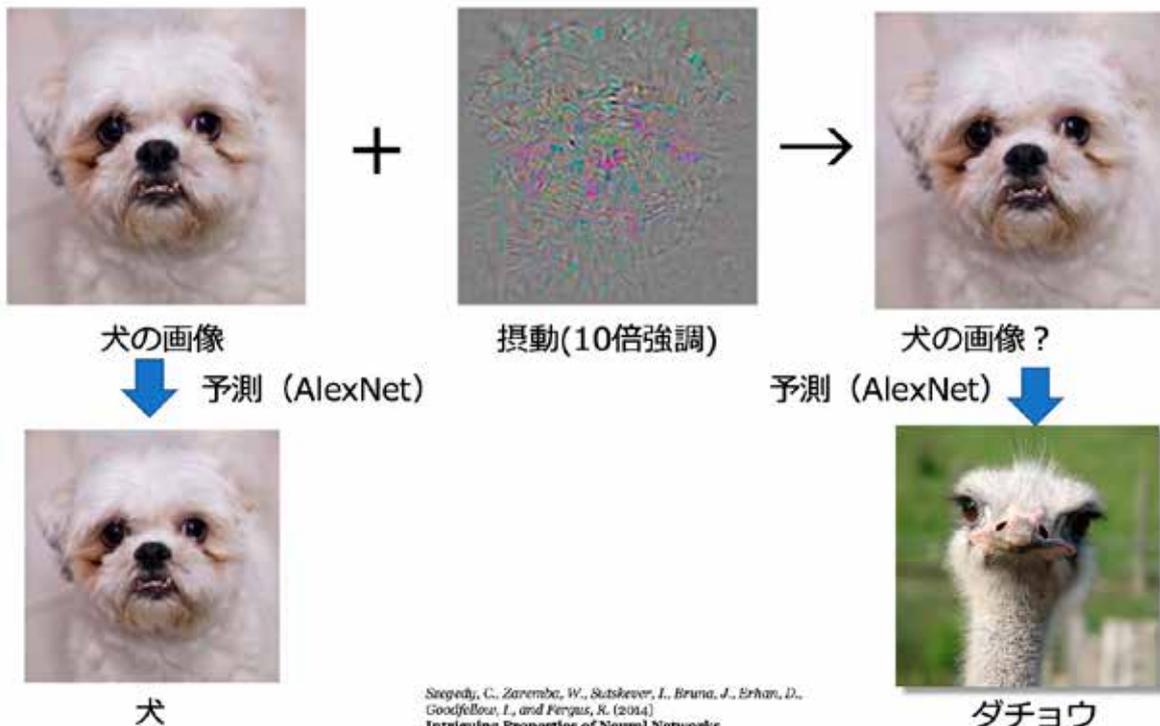
WG06. 製剤・マテリアルインフォマティクス

WG8. 治験・市販後・メディカルアフェアーズ

WG00. AI/DB基盤

AIの不安定性

AIは非直感的なふるまいをすることがある



Szegedy, C., Zarembka, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing Properties of Neural Networks

AIの不安定性



予測 (AlexNet)



Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing Properties of Neural Networks

35

謝辞

山口大学大学院医学系研究科

システムバイオインフォマティクス講座

浅井義之 教授
早野崇英 講師

AIシステム医学・医療研究教育センター
(AISMEC)

安部武志 助教

講座・AISMECメンバーの皆様

AMED 臨床ゲノム情報統合データベース
整備事業

ご参画の先生方

三菱スペース・ソフトウェア株式会社
富士通株式会社

京都大学大学院医学系研究科

奥野恭史 教授
鎌田真由美 准教授
岡本里香 特定准教授
小島諒介 特定講師
内野詠一郎 特定助教

理化学研究所

安倍理加 連携促進
コーディネーター

国立健康・栄養研究所

荒木通啓 上級研究員

ライフインテリジェンス・コンソーシアム

参加企業、参加者の皆様

36

2021年度第1回ヒトゲノム研究倫理を考える会(2021/7/5)

人工知能・機械学習による大規模ゲノムコホート・バイオバンク解析

田宮 元

東北大学・医学系研究科
東北メディカルメガバンク機構・GWASセンター
理化学研究所AIPセンター



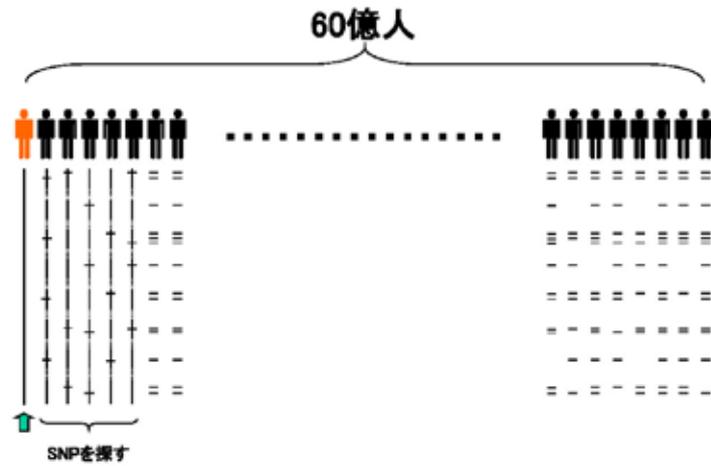
田宮 元, 共立出版2015

ヒトゲノムドラフト配列の解読完了

ヒトゲノム30億塩基対のドラフト版解読完了(2000年)



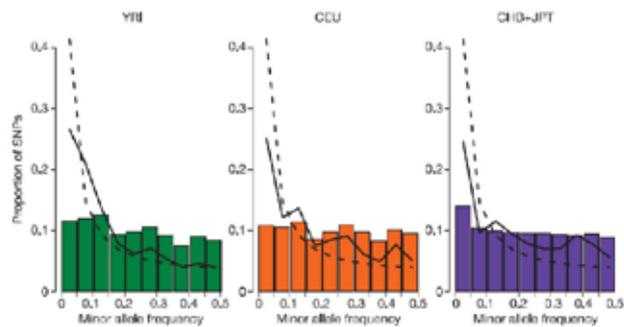
SNPの作業的概念



SNP~HapMapプロジェクト

3

ゲノムワイドなSNPカタログが発表されている



Vol 449 | 18 October 2007 | 4630-4638 | nature04258

nature

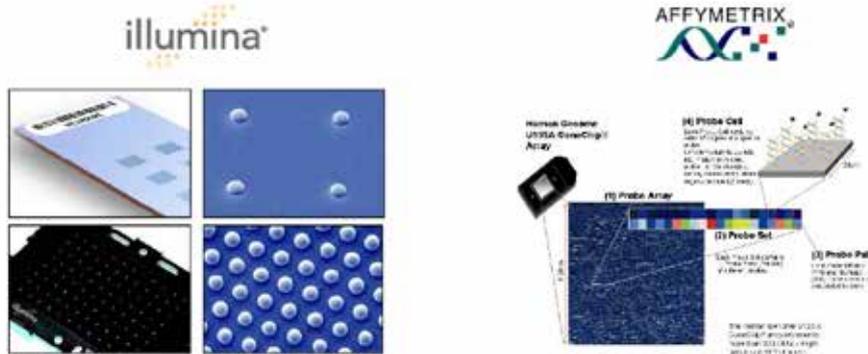
ARTICLES

A second generation human haplotype map of over 3.1 million SNPs

The International HapMap Consortium*

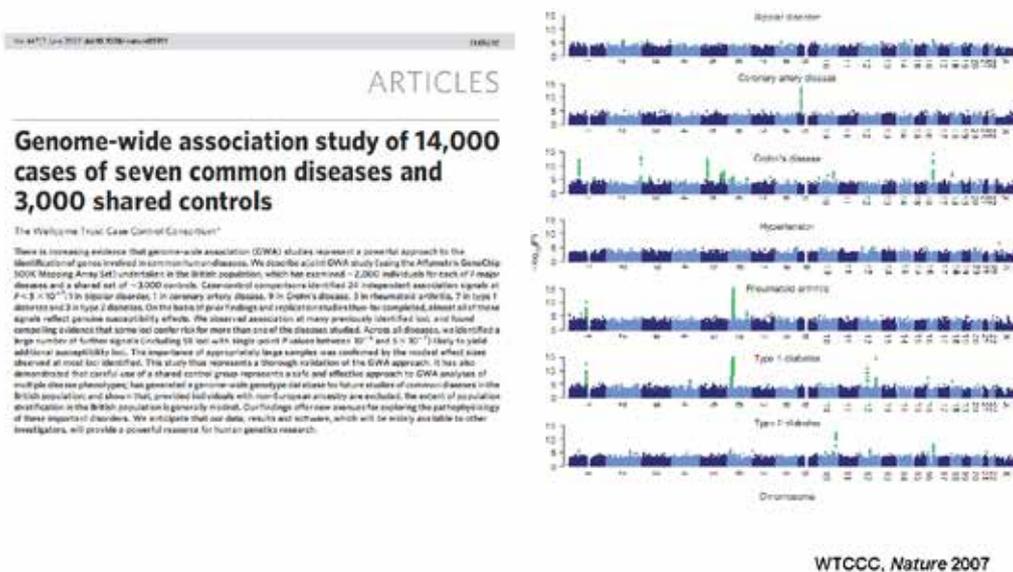
4

全ゲノムSNPジェノタイピングプラットフォーム



すでに商用プラットフォームで高密度SNPが利用可能

欧米での大規模なSNP-GWAS研究の例



WTCCC, Nature 2007

GWASの10年

2007年～

最初の大規模GWAS: サンプルサイズ=1000～5000人罹患者
冠動脈疾患、T2D、前立腺がん、結腸がんなど
1～3の座位(疾患遺伝率の5%未満しか説明しない; 失われた遺伝率)

～2012年

サンプルサイズ増加～1万人罹患者
10のオーダーの座位(疾患遺伝率の10%しか説明しない)
→GWASへの批判や拒否反応

2013年～

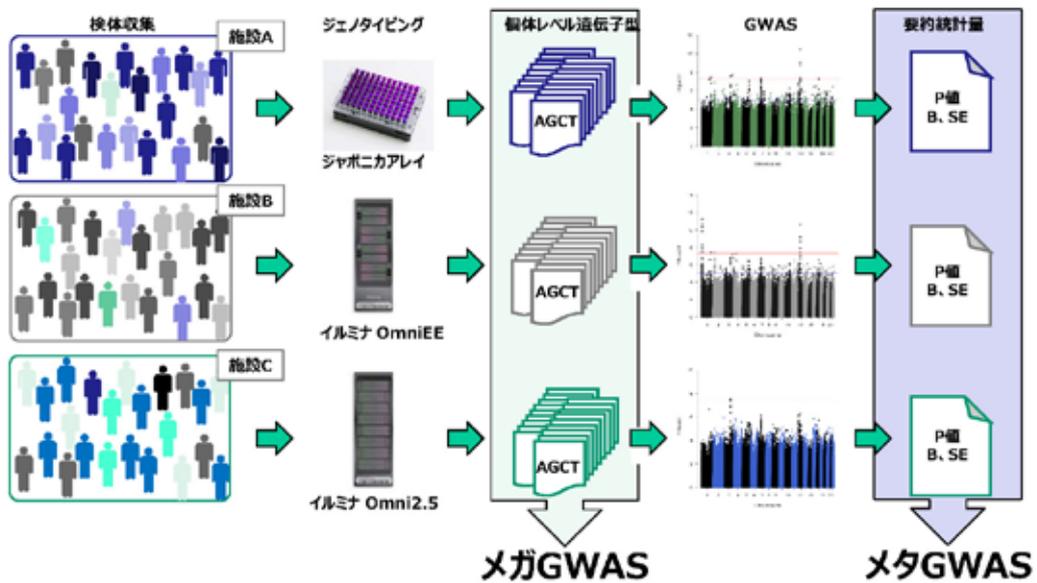
DTC遺伝検査サービス(～1200万人)
23andMeなどがGWAS発見を商用化して個々人に遺伝リスクを回付し始めた
検査の解析的・臨床的妥当性を立証を欠く→ゲノム占い
↓
合衆国FDAが23andMeの個人向けゲノムサービスの販売を一時的に禁止

～2018年

大規模メタGWAS解析: サンプルサイズ10万人以上
冠動脈疾患やT2D、乳がんやアルツハイマー病
100程度の座位で疾患遺伝率の20～30%が説明されている

7

メガGWASとメタGWAS



8

PRSの実用化(5疾患)



Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

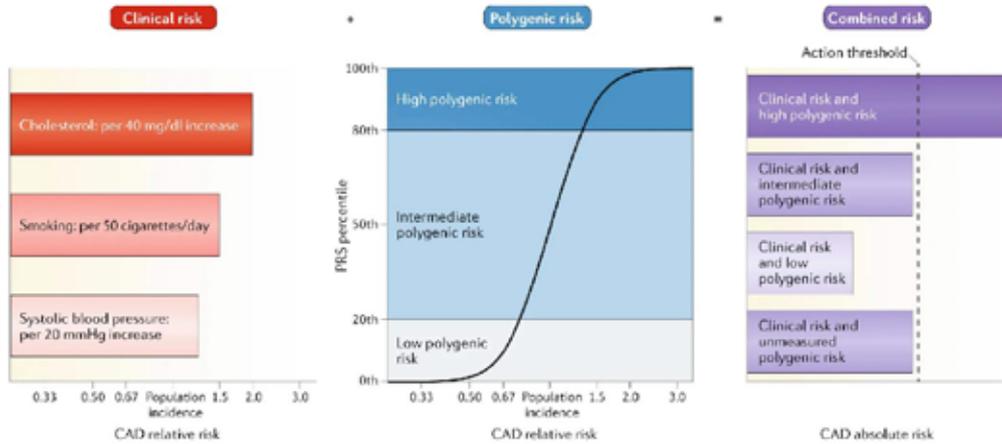
Amit V. Khera^{1,2,3,4,5}, Mark Chaffin^{4,5}, Krishna G. Aragam^{1,2,3,4}, Mary E. Haas⁴, Carolina Roselli⁴, Seung Hoan Choi⁴, Pradeep Natarajan^{2,3,4}, Eric S. Lander⁴, Steven A. Lubitz^{2,3,4}, Patrick T. Ellinor^{2,3,4} and Sekar Kathiresan^{1,2,3,4*}

PRSによる高リスク集団の同定

Table 2 | Proportion of the population at three-, four- and fivefold increased risk for each of the five common diseases

High GPS definition	Individuals in testing dataset (n)	% of individuals
Odds ratio ≥ 3.0		
CAD	23,119/288,978	8.0
Atrial fibrillation	17,627/288,978	6.1
Type 2 diabetes	10,099/288,978	3.5
Inflammatory bowel disease	9,209/288,978	3.2
Breast cancer	2,369/157,895	1.5
Any of the five diseases	57,115/288,978	19.8
Odds ratio ≥ 4.0		
CAD	6,631/288,978	2.3
Atrial fibrillation	4,335/288,978	1.5
Type 2 diabetes	578/288,978	0.2
Inflammatory bowel disease	2,297/288,978	0.8
Breast cancer	474/157,895	0.3
Any of the five diseases	14,019/288,978	4.9
Odds ratio ≥ 5.0		
CAD	1,443/288,978	0.5
Atrial fibrillation	2,020/288,978	0.7
Type 2 diabetes	144/288,978	0.05
Inflammatory bowel disease	571/288,978	0.2
Breast cancer	158/157,895	0.1
Any of the five diseases	4,305/288,978	1.5

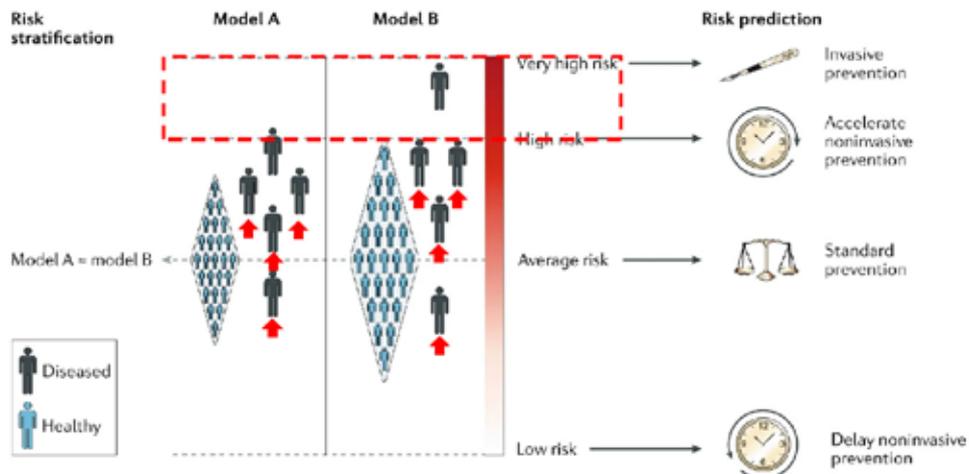
他の臨床的リスクとの組み合わせ



Torkamani *NRG* 2018

11

高リスク群の同定



Torkamani *NRG* 2018

12

現状のPRSの欠点

- 20~30%の狭義の遺伝率しか利用できていない(高リスク群の同定のみ)
- 高リスク対低リスクのみで、スクリーニング目的に向かない可能性
- 広義の遺伝率のために整備されていない
- 環境成分の取り込みのために整備されていない
- 民族特異性があり、欧米のPRSは日本人では利用できない(日本人でのデータが足りない)

13

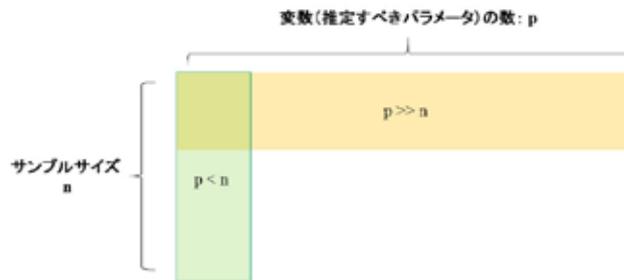
重要なこと

- 疾患に真に寄与するSNP(感受性SNP)を用いる
- 実際には以下の問題が付きまとう
 - 1)間違ったSNPを含めてしまう(包含問題)
 - 2)正しいSNPを外してしまう(除外問題)
- 発見段階での問題...

14

感受性SNPの同定 (GWAS)

- 理想
多因子疾患では、多因子と効果の取り出しを、理想的には重回帰式で行うべき
- 現実
数十万～100万ものSNPを全て含めた重回帰式を解くことは出来ない ($p \gg n$ 問題)
- 折衷
各SNPで単回帰を行って、最後に多重検定補正をほどこす (現行の単回帰GWAS)



15

感受性SNPの同定 (GWAS)

$$y = \beta_0 + \beta_1 snp_1 + \beta_2 snp_2 + \dots + \beta_{50万} snp_{50万}$$

↓ ばらばらに分解する

$$y = \beta_0 + \beta_1 snp_1 + (\beta_2 age + \beta_3 sex)$$

$$y = \beta_0 + \beta_1 snp_2 + (\beta_2 age + \beta_3 sex)$$

⋮

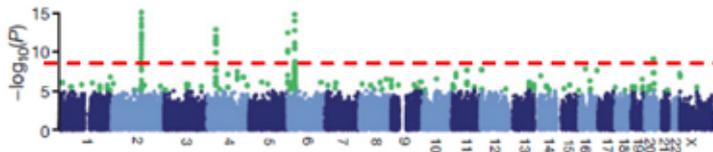
$$y = \beta_0 + \beta_1 snp_{50万} + (\beta_2 age + \beta_3 sex)$$

50万回の単回帰分析(一般線形)の後にBonferroniの補正を行う
(ゲノムワイド有意水準は $\alpha = 5e-8$)

16

単回帰GWASの要約

マンハッタンプロット



ゲノムワイド有意水準
 5×10^{-8} を超えるものは
どの程度あるの？
ギリギリのものは？

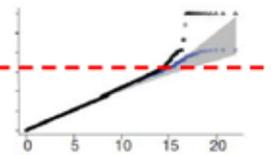
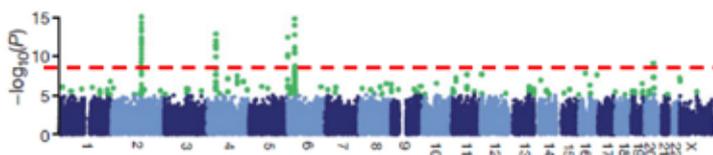
WTCCC Nature 2007

17

単回帰GWASの要約

マンハッタンプロット

QQプロット



染色体座標順

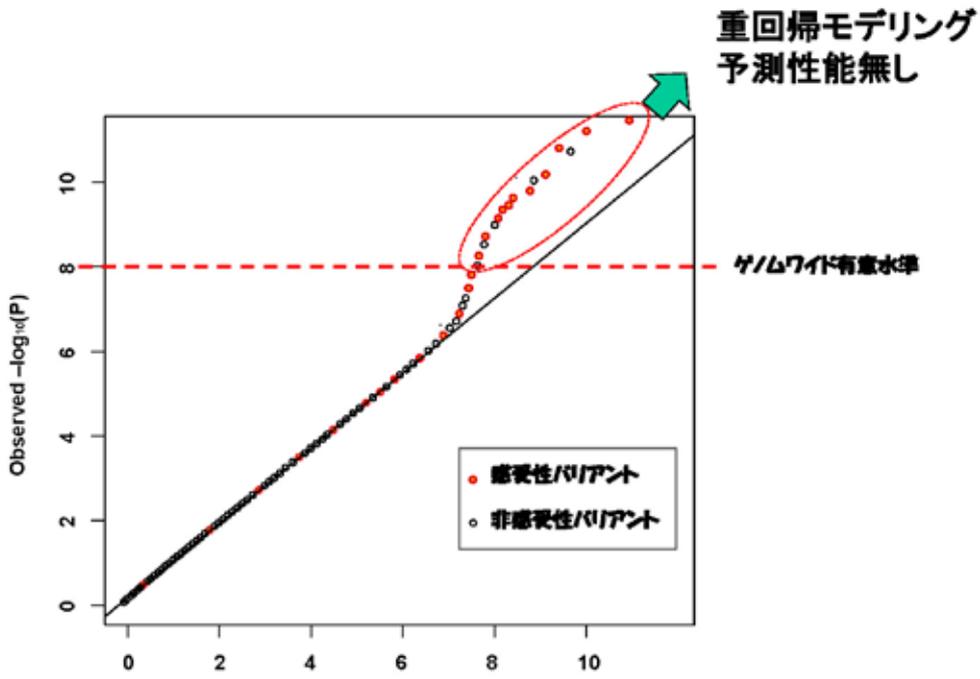
P値の大きさ順

ならべ変える

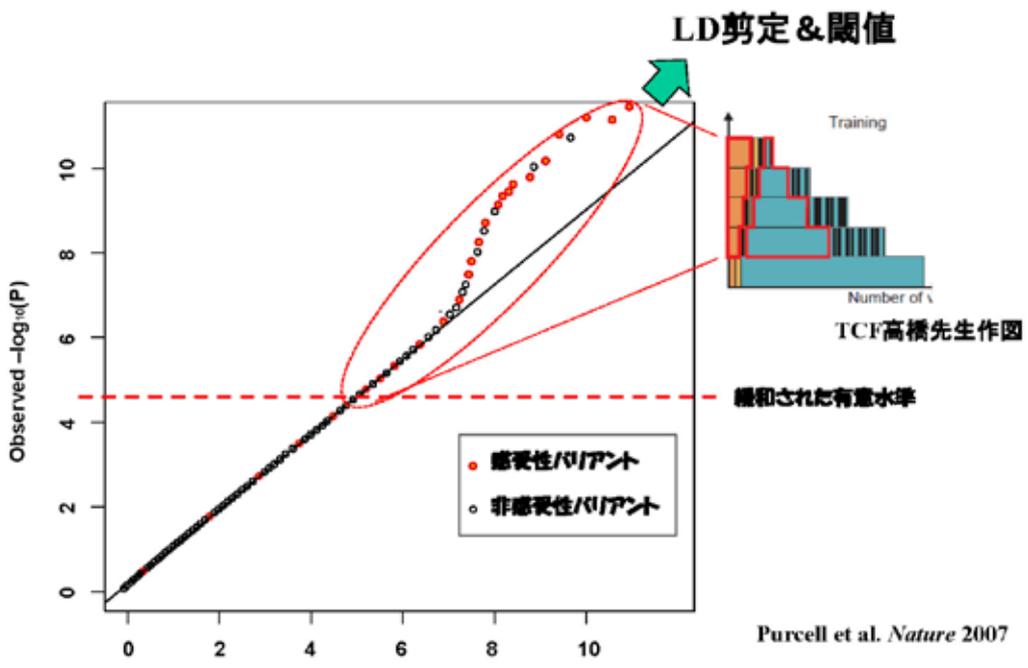
WTCCC Nature 2007

18

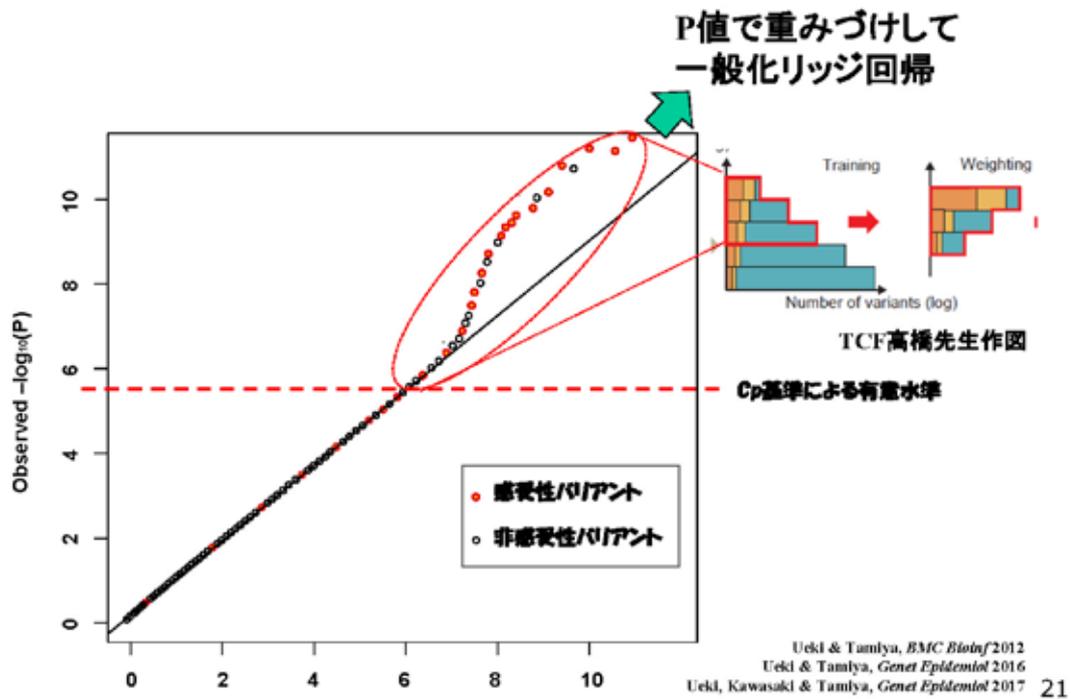
リスクバリアントの抽出：従来法（多重検定補正）



緩和水準に基づく方法（Purcell法；PRS）



統計的機械学習に基づく方法 (stmgp法)



統計的機械学習による リスク予測 (一例: ADNI)

- アルツハイマー病コホート800人で800万SNPs

Covariates	(ii)			OptP	NS
	STMGP	GCTA	GS		
(a)	0.61 (0.05)	0.54 (0.09)	0.51 (0.09)	4.8 (0.4)	143 (104)
(b)	0.63 (0.08)	0.44 (0.08)	0.51 (0.09)	4.8 (0.4)	113 (82)

Covariates	(iii)			OptP	NS
	STMGP	GCTA	GS		
(a)	0.60 (0.15)	0.51 (0.17)	0.52 (0.18)	3.5 (0)	3523 (309)
(b)	0.69 (0.15)	0.47 (0.15)	0.55 (0.21)	3.5 (0)	3601 (296)

統計的機械学習手法

GBLUP

Purcell法

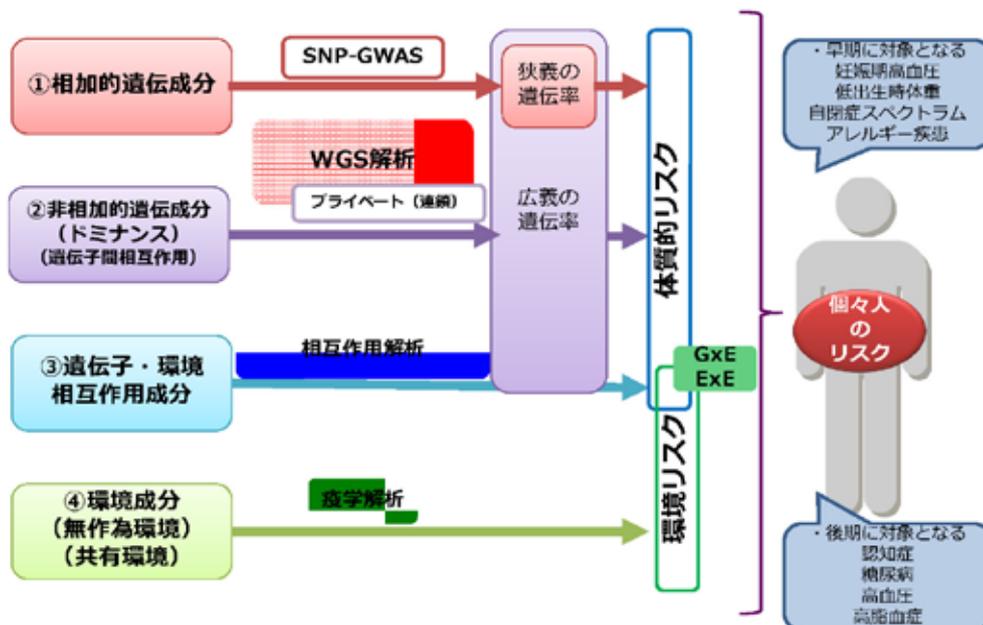
Ueki & Tamiya, *Genet Epidemiol* 2016

統計的機械学習による リスク予測の利点

- 高効率 & 高性能
- 重回帰式中に、環境効果や遺伝子 × 環境相互作用を同時に取り込める
- リスク式で、リスク遺伝子と共にリスク遺伝子 × 環境相互作用を扱える。
- (他のモデルでは統計学的保証を持って扱えない)

23

概論：疾患・健康状態に関する個々人の包括的リスク予測



従来のゲノム医学でのリスク予測は①のみを不正確に扱うしかなかった。
東北MMでは①、②、③、④を適切に取り込んだリスク予測を行う。

24

真の予防医学



ゲノムリスク予測の今後

- 日本人での大規模データの蓄積
- 欧米人のバイアスを含まないゲノムデータ整備
- 遺伝子 × 環境相互作用や遺伝子環境相互作用への対応 (超高次元化)
- ToMMoはこれらの方向性にフィットしている

民族集団ごとのGWAS解析

Genomics is failing on diversity

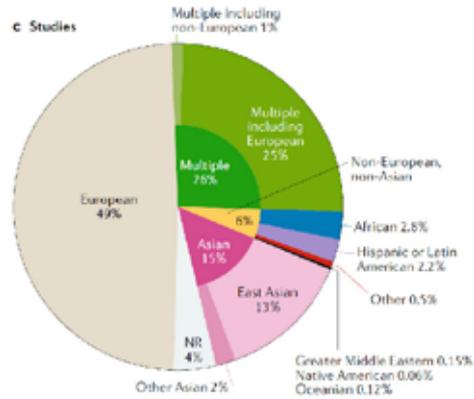
An analysis by Alice B. Popejoy and Stephanie M. Fullerton indicates that some populations are still being left behind on the road to precision medicine.

A 2019 analysis revealed that 95% of participants in genome-wide association studies (GWAS) were of European descent. As the stakes rise for the generalization of genetic findings to diverse groups of people, it is a stark reminder of how far we have to go to ensure that the benefits of genomic medicine belong to everybody.

Our findings indicate that the proportion of individuals included in GWAS who are not of European descent has increased to nearly 10%. Much of this increase is a result of more studies being done in diverse populations of lower ancestry. The degree of linkage disequilibrium in these populations is lower than in European populations, and this can lead to false discoveries in GWAS.

Over the past decade, GWAS have been the main tool for the discovery of genetic factors underlying common diseases. Tens of thousands of genetic variants have been associated with a wide range of traits.

915 studies (n= 110,291,046)



Gurdasani *NRG* 2019

27

trans-ethnic portability

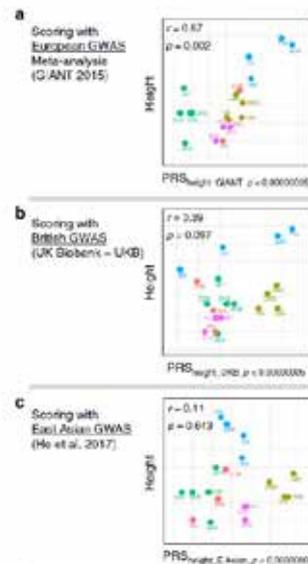
ARTICLE

[DOI:10.1093/bioinformatics/btz014](#) OPEN

Analysis of polygenic risk score usage and performance in diverse human populations

L. Duncan¹, H. Shen¹, S. Colaye², J. Meijer³, K. Resler¹, M. Fejerman⁴, R. Peterson⁵ & B. Domingue⁶

A historical tendency to use European ancestry samples hinders medical genetics research, including the use of polygenic scores, which are individual-level metrics of genetic risk. We analyze the first decade of polygenic scoring studies (2008–2017, *bioRxiv*), and find that 67% of studies included exclusively European ancestry participants and another 19% included only East Asian ancestry participants. Only 3.8% of studies were among cohorts of African, Hispanic, or Indigenous peoples. We find that predictive performance of European ancestry-derived polygenic scores is lower in non-European ancestry samples (e.g. African ancestry samples: $r = -0.07$, $d = 24$, $p = 3.7 \times 10^{-10}$), and we describe the effects of recombination on polygenic score distributions for admixed populations. These findings highlight the need for improved treatment of linkage disequilibrium and variant frequencies when applying polygenic scoring to cohorts of non-European ancestry, and outline the rationale for large-scale GWAS in diverse human populations.

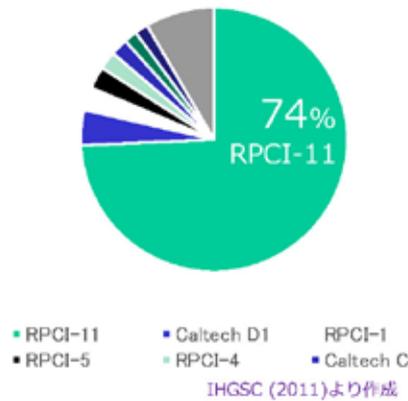


28

ザ・リファレンスの民族構成

- 13数名程度のハプロイドモザイク？

ザ・リファレンスのBACライブラリーの構成



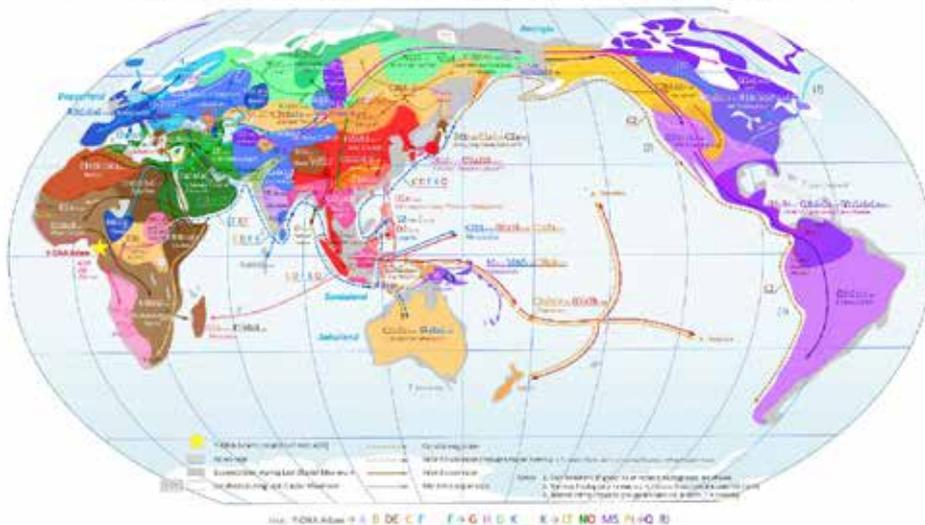
RP-11は
何人？



29

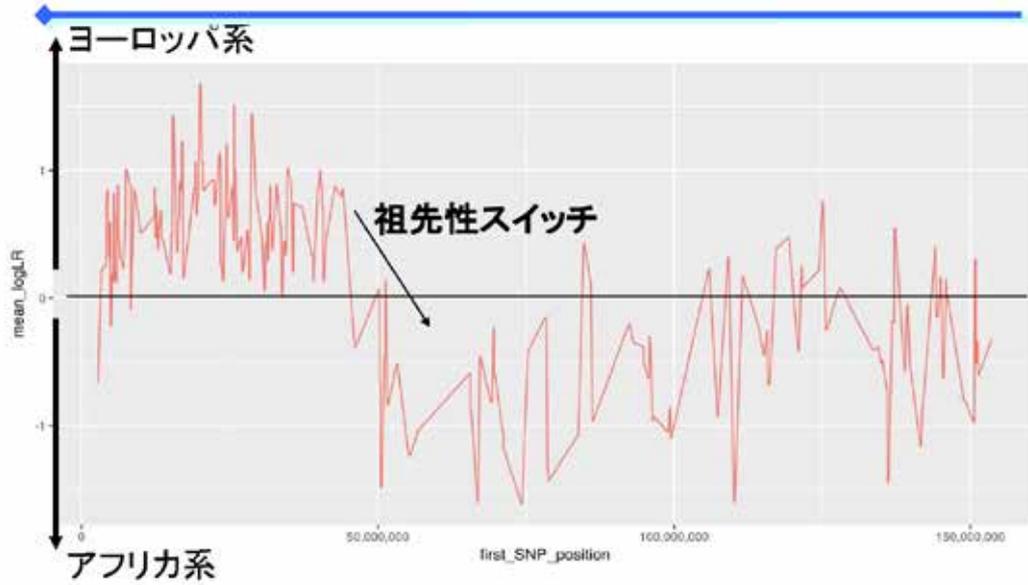
RP-11のYハプログループ

- R1b1a1a2a1a(西ヨーロッパ系の白人？)



30

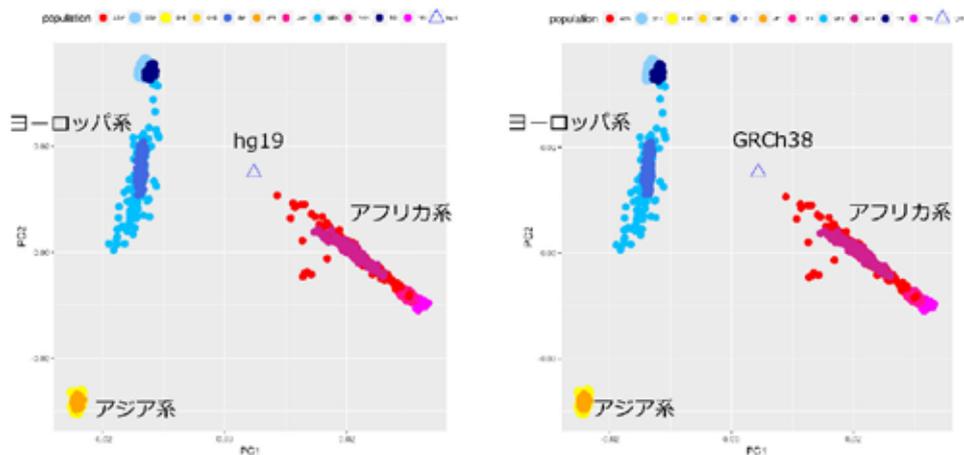
RP-11のX染色体祖先性



31

ザ・リファレンスの位置

- hapmapサンプルとの位置関係 (PCA)



32

日本人基準ゲノムJG1の作成



ARTICLE

Check for updates

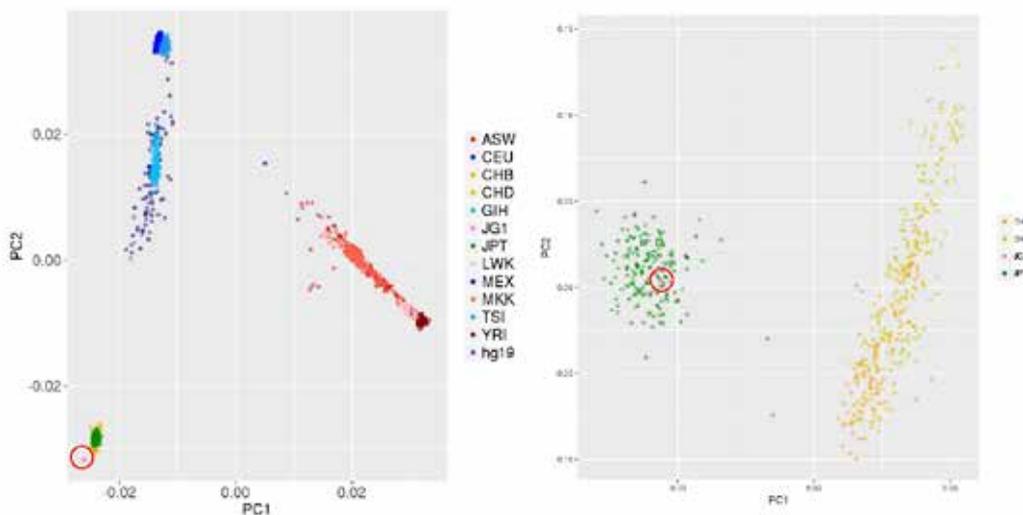
<https://doi.org/10.1038/s41467-020-20146-8> OPEN

Construction and integration of three *de novo* Japanese human genome assemblies toward a population-specific reference

Jun Takayama^{1,2,3}, Shu Tadaka², Kenji Yano^{2,3}, Fumiki Katsuoka^{1,2}, Chinatsu Gocho², Takamitsu Funayama², Satoshi Makino², Yasunobu Okamura^{1,2}, Atsuo Kikuchi⁴, Sachiyo Sugimoto², Junko Kawashima², Akihito Otsuki², Mika Sakurai-Yageta², Jun Yasuda², Shigeo Kure^{2,4}, Kengo Kinoshita^{1,2,5}, Masayuki Yamamoto^{1,2} & Gen Tamiya^{1,2,3}

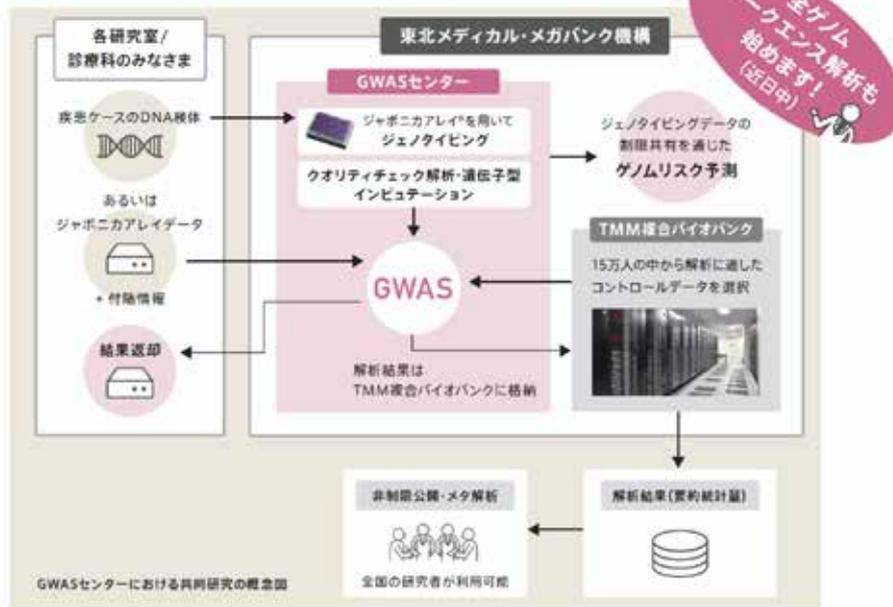
33

JG1の祖先性



34

GWASセンター (共同研究サービス)



2021年度第1回ヒトゲノム研究倫理を考える会

AI・ビッグデータを用いた ヒトゲノム研究の倫理を考える



WEB開催

開催
日時

2021年7月5日(月) 15:00-17:00

近年急速にヒトゲノム・臨床情報を含めたデータの蓄積が進み、ビッグデータと呼ばれるまでになっています。また、人工知能(AI)が研究や医療で用いられるケースも増大しています。一方、このような技術・手法の発展とその研究・医療への応用において新たな倫理的課題も指摘されています。そこで今回、「AI・ビッグデータを用いたヒトゲノム研究の倫理」をテーマに、ヒトゲノム研究倫理を考える会を開催いたします。全国どこからでも参加できるウェビナー形式で開催しますので、是非ご参加下さい。



開催形式 ウェビナー(オンラインセミナー)

- 事前に参加登録を頂いた方に当日参加用URLをお知らせします。
- ブラウザから誰でも参加できるシステム(ユーザ登録不要)を使用します。
- パソコン・スマホで全国どこからでもご参加頂けます。

15:00~15:05 「開会の挨拶」

加藤 和人(大阪大学大学院医学系研究科)

15:05~15:35 「AI・ビッグデータを用いたライフサイエンス・ゲノム研究の現状と課題」

中津井 雅彦(山口大学大学院医学系研究科・医学部附属病院 AI システム医学・医療研究教育センター)

15:35~16:05 「人工知能・機械学習による大規模ゲノムコホート・バイオバンク解析」

田宮 元(東北大学大学院医学系研究科、理化学研究所革新知能統合研究センター)

16:05~17:00 質疑応答・総合討論

指定発言: 小門 穂(神戸薬科大学社会科学研究室)

対象 大学・研究機関の倫理審査関係者、研究者など

参加費 無料

セミナーの詳細
参加登録はこちら

参加登録 下記ページのフォームから参加登録をお願いします。

www.genomics-society.jp/news/event/post-20210705.php/ ※ 事前参加登録: 7月2日(金) 12:00まで



主催 文部科学省科学研究費新学術領域「先進ゲノム支援」 ゲノム科学と社会ユニット (GS ユニット)
問合せ 大阪大学大学院医学系研究科医の倫理と公共政策学 email: workshop@eth.med.osaka-u.ac.jp

告知チラシデザイン (A4)

発行日: 2022年3月20日

発行: 文部科学省科学研究費新学術領域「先進ゲノム支援」

ゲノム科学と社会ユニット

(大阪大学大学院医学系研究科 社会医学講座 医の倫理と公共政策学 加藤和人研究室)

編集: ゲノム科学と社会ユニット (GS ユニット)

デザイン: YUMA DESIGN

