

遺伝統計学チームの研究ミッション

生命科学（医学や農学）のビッグデータを機械学習・人工知能ベースの遺伝統計学で解析を行い、ヒトの疾患や農業形質などの複雑な現象に寄与する要因を探り出す。今回、我々研究チームの取り組む課題のうち、疾患リスク予測を代表して紹介する。

I 大規模ゲノムコホートデータへの疾患リスク予測手法の適用

背景

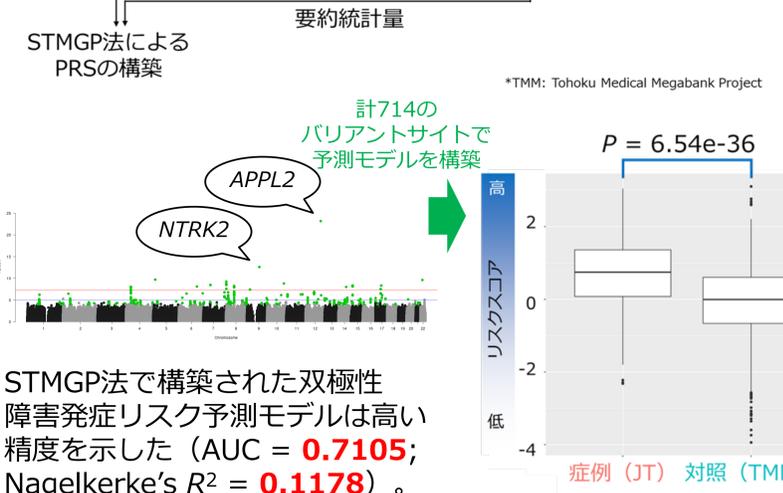
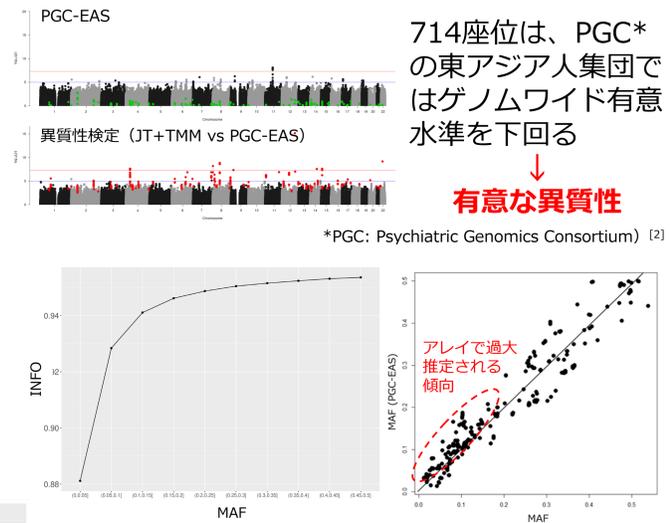
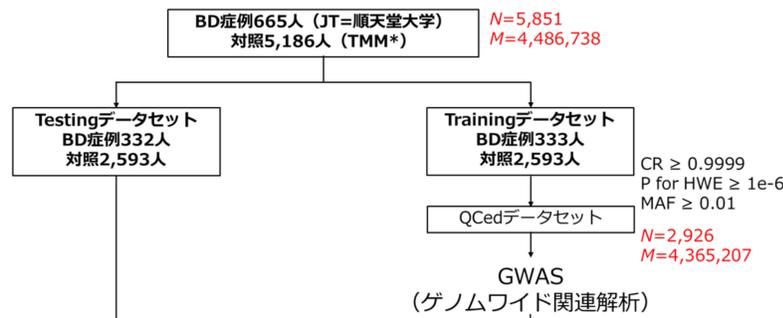
近年、ゲノムデータを活用した疾患リスク予測の有用性が注目を集めている。当チームが開発したSTMGP法（smooth-threshold multivariate genetic prediction）^[1]は、他の手法と比較して高い予測精度を示しているが、疾患リスク予測システムの社会実装の実現に向けて、より多くの疾患、より大規模なデータでの検証を進めている。

これまでの進捗

[2024年度までの成果]
 UKバイオバンクおよび東北メディカル・メガバンク計画 地域住民コホートのLDL-CデータにSTMGP法を適用し、既存の手法よりも高い予測精度を示すことを確認した。また、分割計算により、予測精度を維持しながら計算負荷の軽減を実現し、また外部で公開されているGWAS（ゲノムワイド関連解析）の要約統計量を取り込むことで、予測精度が向上することも併せて示した。

[2025年度の成果]
 順天堂大学医学部 精神医学講座との共同研究として、双極性障害665人、および対照群として東北メディカル・メガバンク計画5,186人のWGS（全ゲノムシーケンス）データにSTMGP法を適用し、既存の手法を上回る予測性能を実現することができた。

データおよび解析



この性能をもたらした要因について検証

MAF（マイナーアレル頻度）が低いサイトはimputation精度（INFO）も低く、アレイで頻度が過大推定される傾向

考察

本研究で用いたデータとPGCの結果の差異
 ↓
プラットフォーム（WGS vs アレイ）の違い
 （PGCはアレイデータが中心であり、主に低MAFのサイトにおけるimputationバイアスが両者の結果の異質性を招いた可能性）

今後の予定

- さらなる予測精度向上に向けて、引き続きWGSデータを上積みし、サンプルサイズの拡大を進める
- 関連多型の数や効果など、遺伝的アーキテクチャの異なる他の疾患/形質でも予測精度を評価する

II 臨床サブタイプに対するポリジェニックリスク予測手法の開発

背景・目的

■ PRSの予測精度を高めるため、複数のGWAS要約統計量を祖先性横断的にメタ統合してデータを大規模化する流れが主流となっているが、その精度は限定的（例えば2型糖尿病（T2D）PRS AUC ~0.66程度に留まる）

■ PRSの予測精度を妨げる要因

- 臨床的に異質性のある疾患をあたかも単一で均一な疾患とみなしたモデル構築
- T2Dはその代表的疾患（臨床病態の異なる4つのサブタイプが存在）

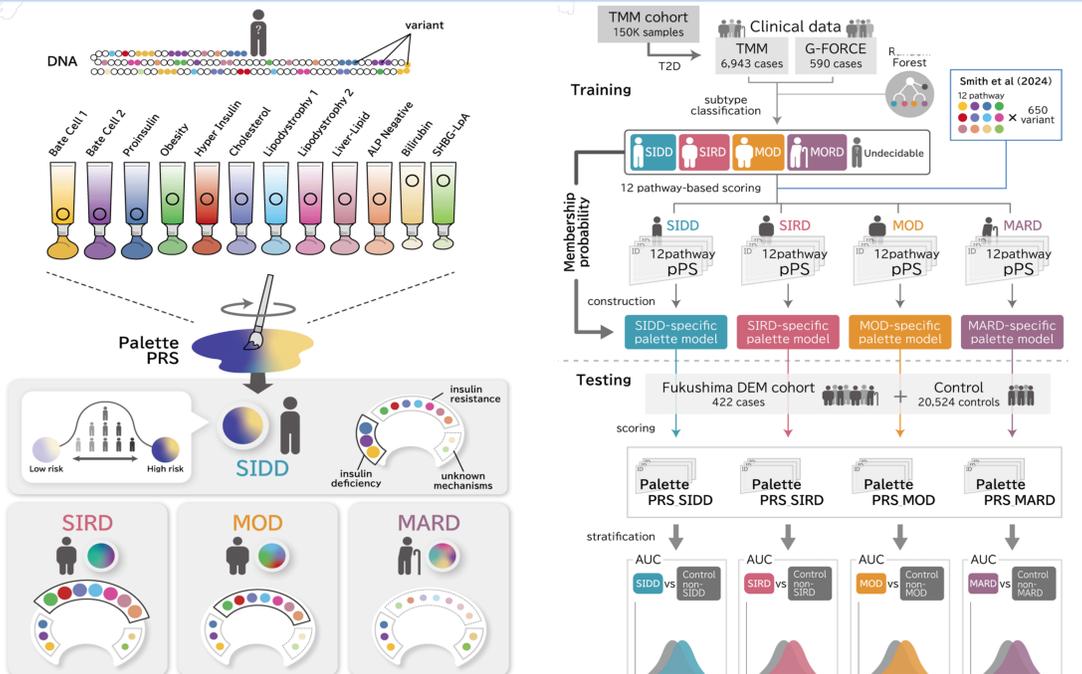
K-meansクラスタリングによる糖尿病サブタイプ分類 ^{[1][2]}

サブタイプ	特徴	日本 (%)	スウェーデン (%)
1型糖尿病	Cluster1: 重症・自己免疫型糖尿病(SAID) GAD抗体陽性、インスリン枯渇、高血糖、若年発症	5.4%	6.4%
2型糖尿病	Cluster2: 重症・インスリン欠乏型糖尿病(SIDD) GAD抗体陰性、高度のインスリン欠乏、高血糖	19.0%	17.5%
	Cluster3: 重症・インスリン抵抗性型糖尿病(SIRD) 高度肥満、重度のインスリン抵抗性、脂肪肝	7.2%	15.3%
	Cluster4: 軽症・肥満関連型糖尿病(MOD) 軽度肥満、軽度のインスリン抵抗性	28.9%	21.6%
	Cluster5: 軽症・加齢関連型糖尿病(MARD) 高齢発症、軽度インスリン分泌不全	39.5%	39.1%

■ サブタイプに最適化したポリジェニックリスク予測モデルを構築し予測精度向上を図る

- 臨床的異質性をもたらす複数の病態生理学的経路（パスウェイ）の存在および重要度の違いに着目

Palette PRSによる臨床サブタイプリスク予測の枠組み

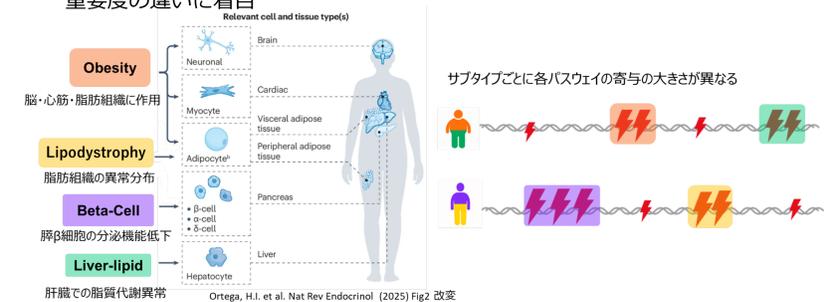
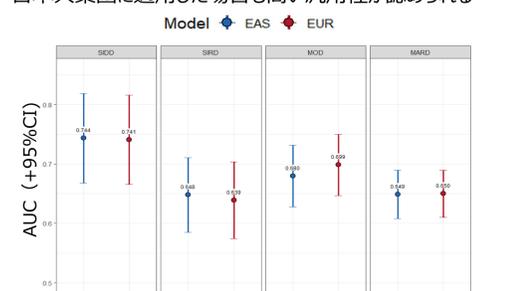


実データへの適用結果

■ 最大規模のサブタイプ層別化GWASから構築したモデルと比べて大幅に予測精度を上回る

Population with T2D subtypes	AUC (+95%CI) in testing dataset	
	Palette PRSs	Subtype-stratified GWAS PRSs
SIDD	0.744 (0.668-0.819)	0.547 (0.467-0.627)
SIRD	0.648 (0.585-0.711)	0.509 (0.481-0.580)
MOD	0.680 (0.628-0.732)	0.549 (0.493-0.605)
MARD	0.649 (0.608-0.690)	0.554 (0.510-0.598)

Ukbiobank British集団（EUR）から構築したモデルを日本人集団に適用した場合も高い汎用性が認められる



▶ ベイジアン非負値行列因子分解（bNMF）により発見された12種類の糖尿病関連パスウェイ^[3]を基に、各臨床サブタイプに最適化したPalette PRS^[4]を開発

[1] Tanabe H, et al. J Clin Med. 9:7, 2020.
 [2] Ahlqvist E, et al. Lancet Diabetes Endocrinol. 6:361-369, 2018.
 [3] Smith K, et al. Nat Med. 30:1065-1074, 2024.
 [4] Miyake A, et al. medRxiv [preprint]. 2026.