

ゲノム医療協議会 講演資料

NTT DATA
Trusted Global Innovator

NTTデータのゲノム解析ビジネスとCANNDsへの期待

2022年12月23日(金)

(株) NTTデータ 技術革新統括本部 技術開発本部

田端 佑介

業界の近況

- ゲノムデータが蓄積されることで研究が進み、徐々にゲノムの活用事例が出始め、ゲノムへの理解が浸透。
- 大規模なデータを解析できるアプリケーションと基盤の開発がセットだが、IT化の遅れが業界課題。

社会背景

ゲノムデータの普及

個別化医療への期待

ゲノム活用基盤の需要

AI活用の促進

ゲノム解析市場の技術構成

アプリケーション

IT基盤

医学・生命科学・創薬

- シーケンシング
- 検査技術
- 素材開発
- etc.

データサイエンス

ビッグデータ（計算・蓄積）

流通・セキュリティ

IT企業の強みを活かせる領域

業界課題

IT化の遅れ

データ蓄積の遅れ

AI人材の不足

国内IT企業の
参入余地

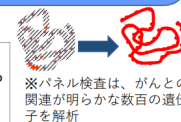
国家的なゲノム収集・解析・実用化の推進

ゲノム医療・AI活用の推進

- 全ゲノム情報等を活用したがんや難病の原因究明、新たな診断・治療法等の開発、個人に最適化された患者本位の医療の提供
- AIを用いた保健医療サービスの高度化・現場の負担軽減

【取組の加速化】

- 全ゲノム解析等によるがん・難病の原因究明や診断・治療法開発に向けた実行計画の策定
- AI活用の先行事例の着実な開発・実装



※パネル検査は、がんとの関連が明らかな数百の遺伝子を解析

引用元：今後のデータヘルス改革の進め方について（令和元年9月）資料より抜粋
<https://www.mhlw.go.jp/content/12601000/000545973.pdf>

国家的なゲノム・データサイエンティスト育成



文科省 データ関連人材育成プログラム（2017）

【カリキュラム】

- データサイエンス
- ビッグデータ
- AI創薬

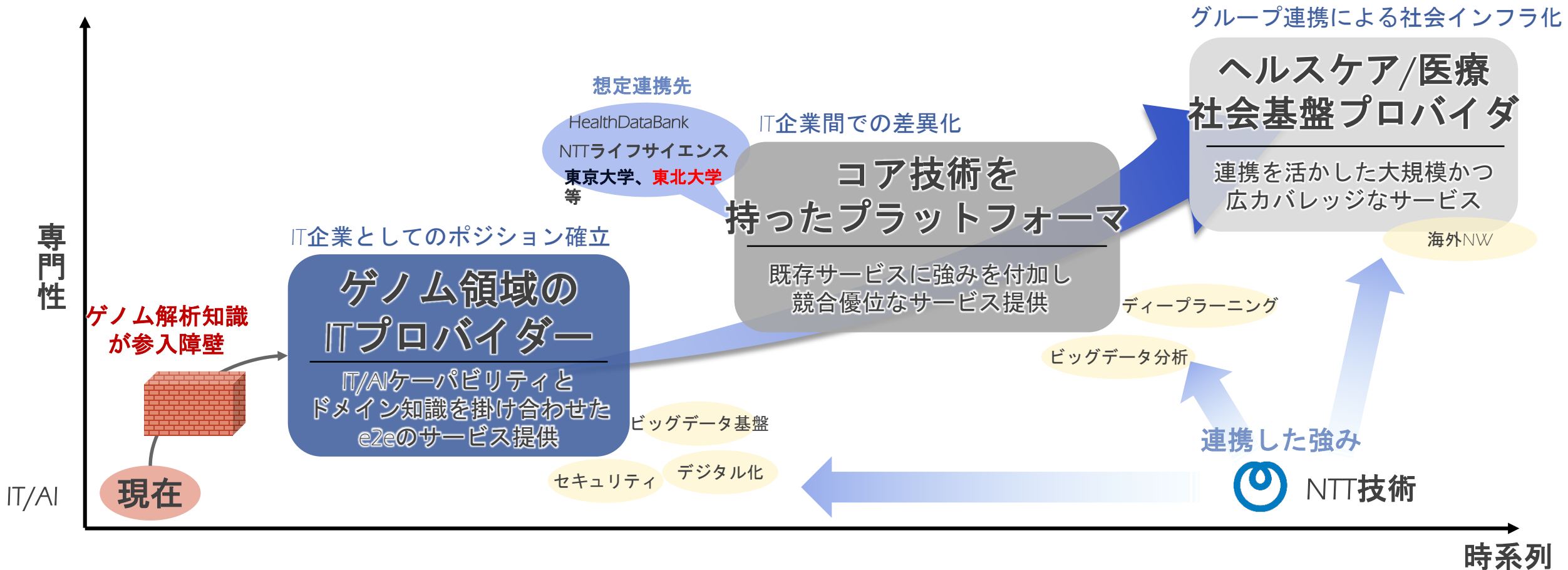
IT企業が続々と共同研究を始めている



テーマは三者三様

NTTデータの方針

- 参入障壁であるゲノム解析知識の習得から立ち上げ、サービスを提供しながらコア技術を獲得し、プラットフォームの中核企業として社会インフラ構築を目指す。
- 大学との共同研究を中心とした技術提供と検証で、ビジネスフェージビリティ調査を行いコア技術を育てる。



(連携例・東北大学) ゲノム疾患リスク予測(PRS)

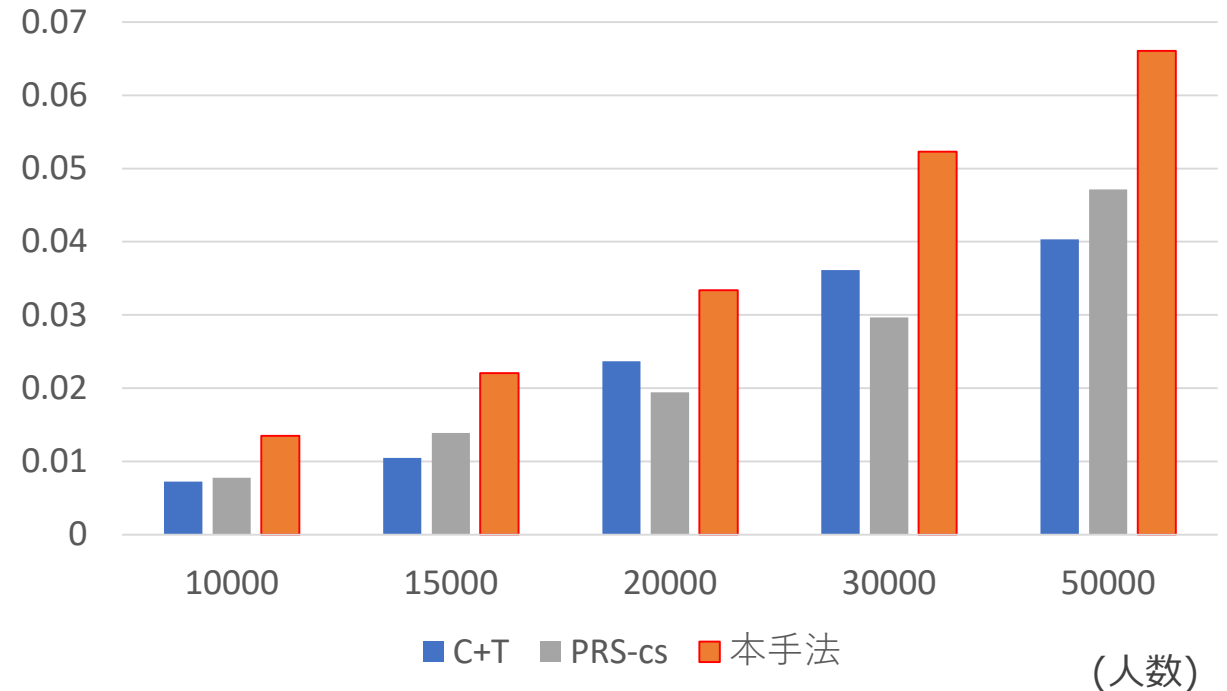
- 東北大学の機械学習をコア技術とし、共同開発及び大規模データでの検証を実施。現状の各手法の精度と基盤に求められる計算量を確認。
- ビジネス化に向けて研究を実装し、ゲノムデータ解析サービスの開発・技術提供を行う。(2023年本格化)

PRSの過去文献

事例	推定対象	人数	バリエーション数
Ueki & Tamiya (2016)	アルツハイマー病	約7百	約870万
Takahashi et. al. (2020)	抑うつ症状	約5千	約600万
共同研究 (2022)	糖尿病、痛風、高脂血症等 計41形質	約5万	約1,200万

HbA1cの予測精度

(HbA1cとリスクスコアのピアソン相関係数(R2))



第四回日本メディカルAI学会にて発表

提供サービスイメージ(医療)

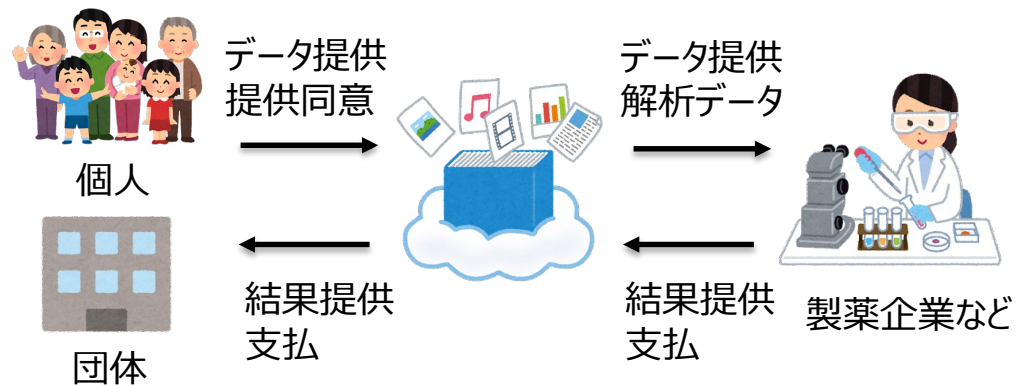
- ゲノムデータの解析結果と各種医療データを組み合わせて、幅広いサービス提供が可能。
- ゲノムは環境の影響を受け難く、既存の解析に疾患の遺伝要因が加わるのがメリット。

	ユースケース例	必要なデータ	概要
予防	ゲノム疾患予測 ライフスタイル改善	医療, 処方, 検査, 生活, ゲノム	疾患発症リスクの低減, 最適な健康維持・増進策の選択
診断	診断補助	医療, 処方, 検査, 生活, ゲノム +センサー経由のデータ	オンライン診療における診断補助AI, 適切な診断や個人に最適化された治療
	患者と医師のマッチング	医療, ゲノム +医療機関・医師の情報等	オンライン診療における医療リソース効率的活用, 質の高い医療
	がん早期発見	医療, 検査, 生活, ゲノム	生体バイオマーカーや他情報も含めたがんの早期発見
治療	緊急搬送	医療 + 患者の位置情報, ゲノム 病院情報 (手術室の空き状況等)	最短・適切な施設への搬送による救命率向上, 医療リソースの効率的活用
	治療方針の決定	医療, 処方, 検査, 生活, ゲノム	適切な診断や個人に最適化された治療
	服薬管理、残薬量のモニタリング	処方, 検査, 生活, ゲノム +センサー経由のデータ	服薬コンプライアンスへの寄与, 適切な残薬補充, 及び自動的な配送手配等
治験	治験データとしての利活用 治験リクルートメント	医療, 処方, 検査, 生活, ゲノム , 社会, 研究	データ収集の効率化・精度向上、適切な被験者選 定 (コスト削減)
	治療効果予測モデルや 薬物動態予測などのシミュレーション	医療, 処方, 検査, 生活, ゲノム , 社会, 研究	精度の高い予測による被験者の安全性確保, 治験の成功確率向上
	医薬品の安全性モニタリング・ 安全監視対策への利用	医療, 処方, 検査, 生活, ゲノム	適正な使用・管理・添付文書改訂

提供サービスイメージ(医療以外例)

医療以外への活用は業界が模索中。新たなサービス開拓となる可能性あり。

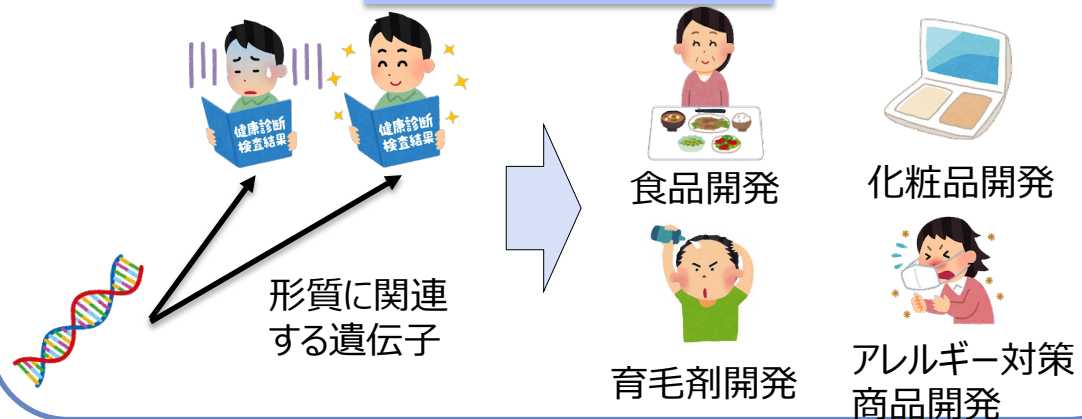
情報銀行



アグリゲノミクス



新商品開発

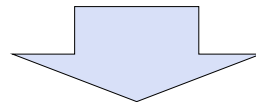


予測の高度化



ビジネス化を阻害する主な課題

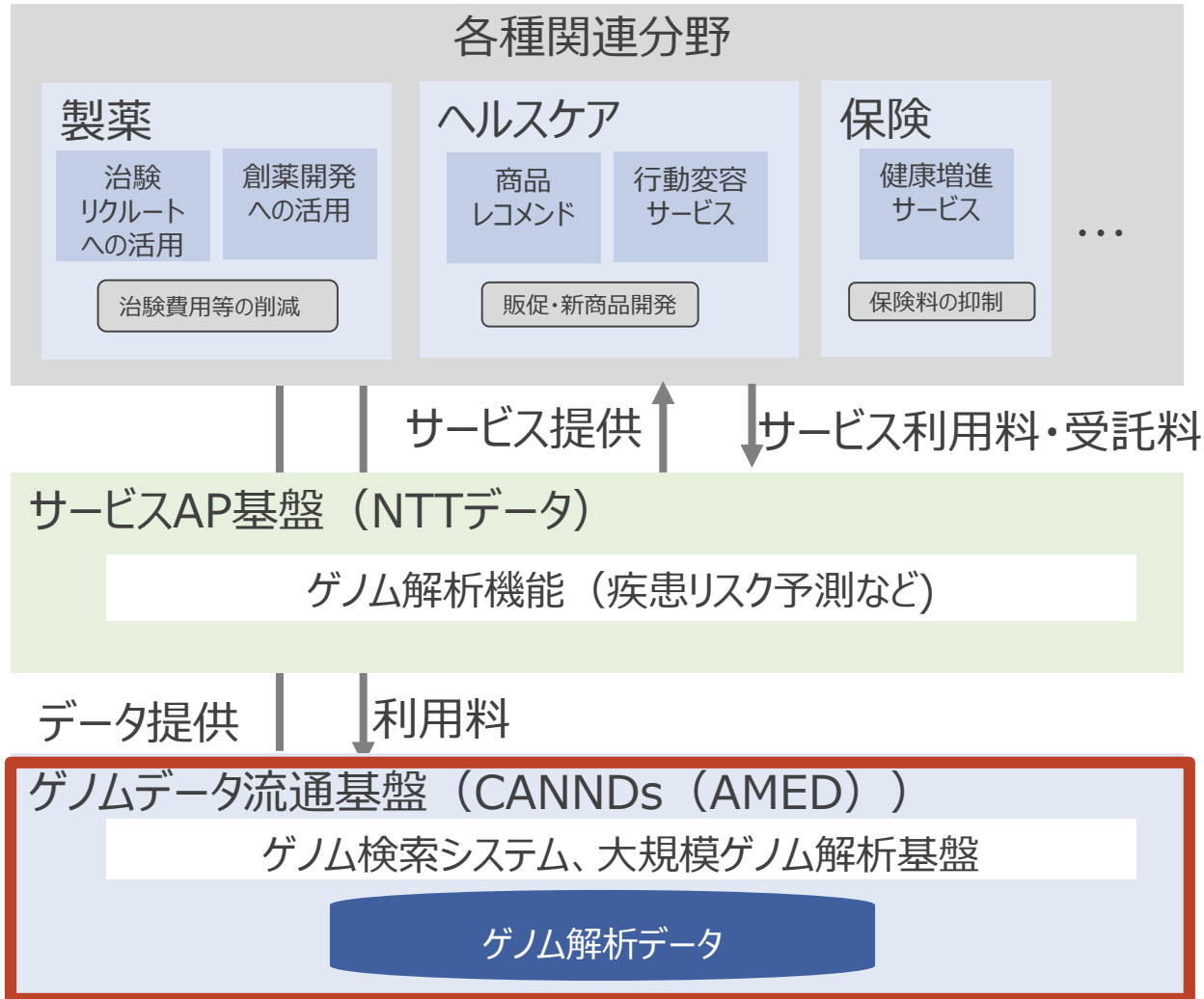
- ビジネスの前提にある、ゲノムデータが入手しづらい。
 - 究極の個人情報。各コホートの審査は厳重で、他社からの融通もきかない。
 - 自社でデータ収集すると数万円/検体。人数を集めると大変高額で、1社では厳しい。
 - 各データは収集目的・商用利用(二次利用)が限られており、取得地域、年齢層、性別などに偏りがある。
- 投資判断がつかない。
 - 市中の情報が少なく、かつ、専門家向けで、ビジネスに寄与するレベルの手法・精度、基盤環境が見えない。
 - ビジネスに必要なデータ規模がわからず、コストが見通せない。
 - 検証で確かめたいが、検証自体が容易ではない。
- ゲノム解析技術やAI技術などを保有する人財不足。



国が主体で進める、CANNDsが実現すれば、解決できる可能性が高い。

CANNDs連携イメージと期待

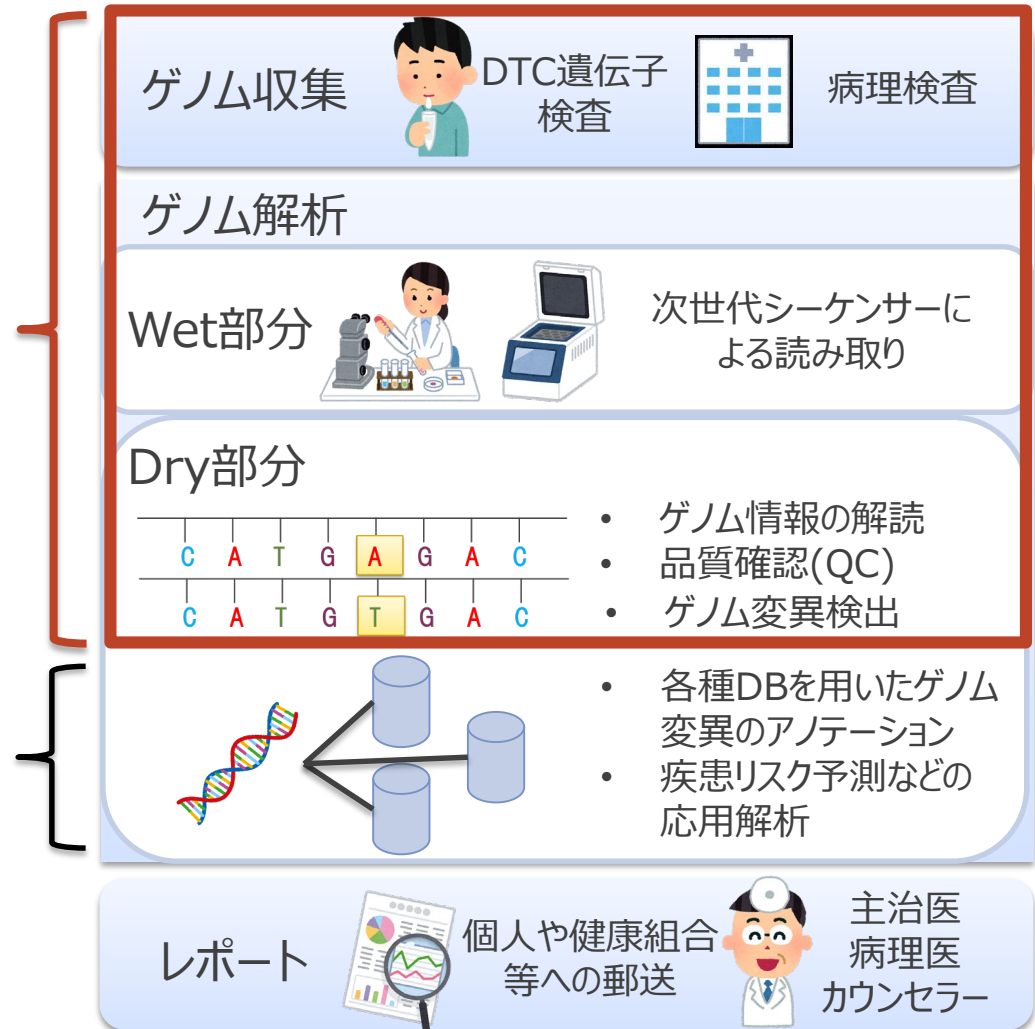
- ビジネス化で課題となる、大規模なゲノムデータはCANNDsを活用し、弊社はゲノム解析サービスに注力したい。
- 投資判断となる検証や、人材育成にもCANNDsを利用したい。



できればCANNDsへ連携したい

弊社が主に注力

ゲノム解析フロー (一例)



CANNDsへの要望(1/2)

- 大規模な個人のゲノムデータと分析基盤の提供。
 - 小規模なゲノムデータは自社収集としても、大部分はCANNDsから入手。
 - 大容量のメモリを擁する検証基盤が必要。
 - ある程度のデータクレンジングや計算はCANNDs側で実施し、ゲノムデータに反映。
- 各種調査で集めた調査票の公開。
 - 最低限は参加時年齢、性別、形質値(健康診断の値、疾患有無、バイオマーカーの値等)。
 - データ品質に関する情報や解析手段(シーケンサー、プロトコール、ソフト、アプリケーション等)。

【ゲノム疾患リスク予測における、東北大の機械学習手法の場合】

各被験者個人のバリエーションが入っているゲノムデータ

被験者数：数十万人規模（希望は数百万人）

バリエーション数：数千万バリエーション規模

メモリ：数百GBのメモリ

データ形式：バリエーションの情報が入っている、PLINK(*)で扱える形式のデータ(bgen/sample、vcf、bed/bim/fam等)

形質：健康診断の値

データに付与されている値：Infoスコア(ゲノムデータの品質に関する情報)

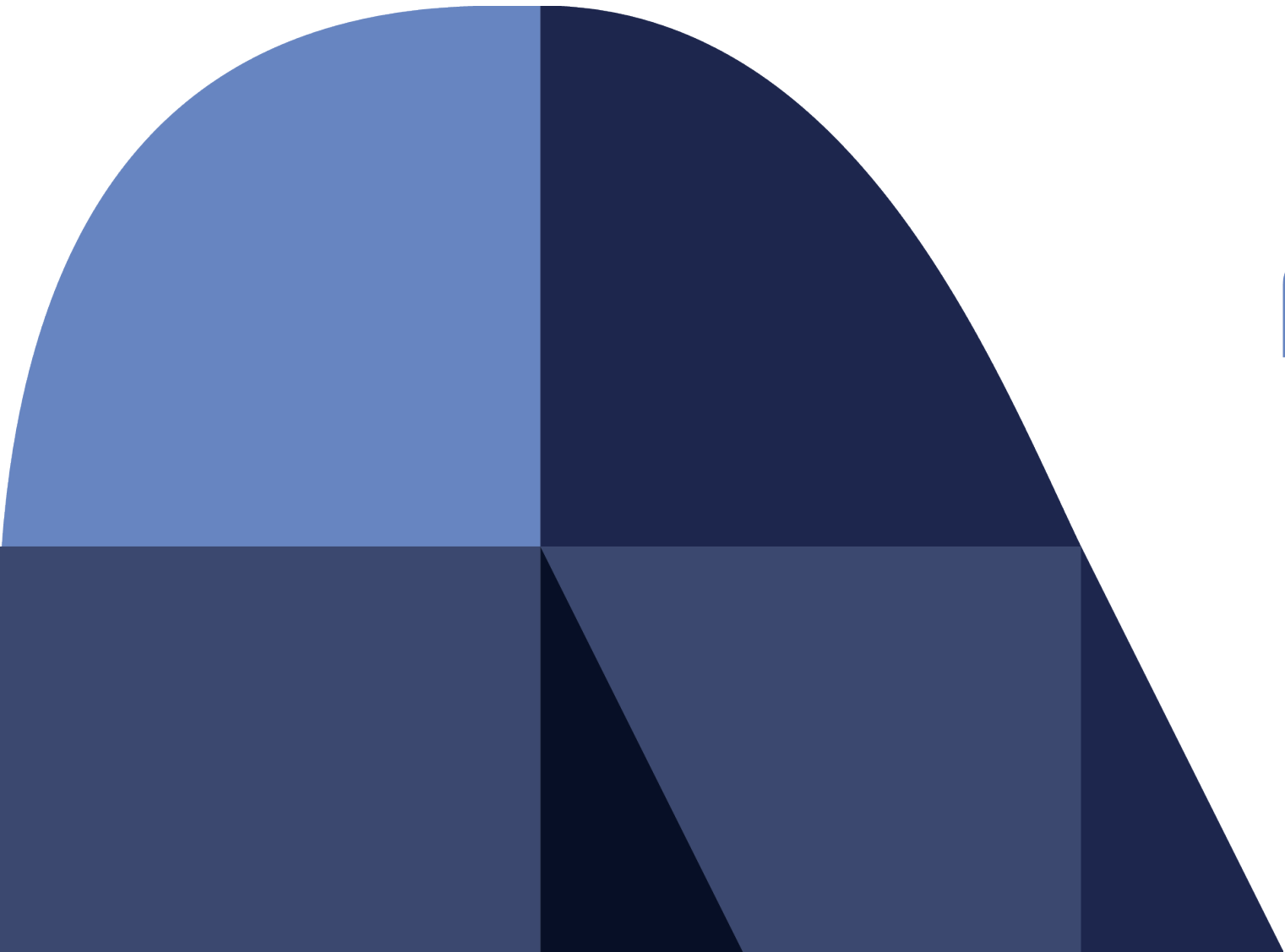
(*)PLINK：オープンソースの全ゲノム関連解析ツールセット

Author: Shaun Purcell

URL: <http://pngu.mgh.harvard.edu/purcell/plink/>

CANNDsへの要望(2/2)

- セキュリティ(個人情報)を意識しないレベルの、ゲノムデータのハンドリング。
→ セキュリティの確保はできる限りCANNDs側が実施。
- 解析結果は幅広く商用利用(二次利用)可能。
→ 商用利用に関する同意書のようなものがあれば良い。
- CANNDs側や他研究で実施した解析結果の公開。
→ 商用利用(二次利用)や利用者側が解析の正しさを確認できるため、解析結果(GWASの要約統計量等)を公開してほしい。
→ 審査はあると思うが、個人レベルの解析結果が利用できると良い。
→ 特に論文や特許の権利が切れたものは出してほしい。
→ 公開場所の希望はない。



NTT DATA
Trusted Global Innovator