

---

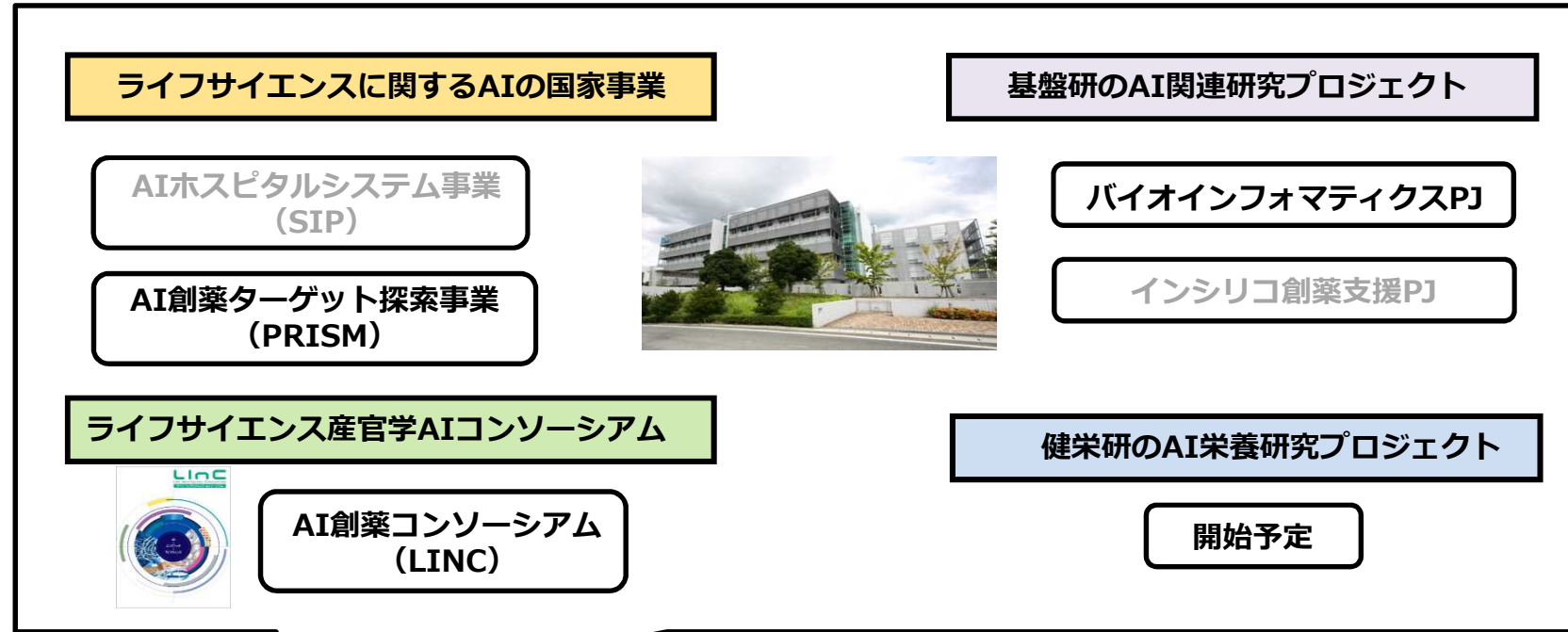
# AI健康・医薬研究センターについて

---

国立研究開発法人 医薬基盤・健康・栄養研究所  
AI健康・医薬研究センター

水口 賢司

# 医薬基盤・健康・栄養研究所のAI機能の統合化



- ① 基盤研のAI関連プロジェクト
- ② AI関連の国家事業 (SIP, PRISM)
- ③ 産官学AIコンソーシアム (LINC)
- ④ 健栄研のAI栄養 (開始予定)
- ⑤ 知財・契約等のインキュベーション機能

**AI健康・医薬研究センター**

2019年4月設立

# AI健康・医薬研究センターの組織

日本語名称 : AI健康・医薬研究センター

英語名称 : Artificial Intelligence Center for Health and Biomedical Research

略称 : ArCHER (アーチャー)

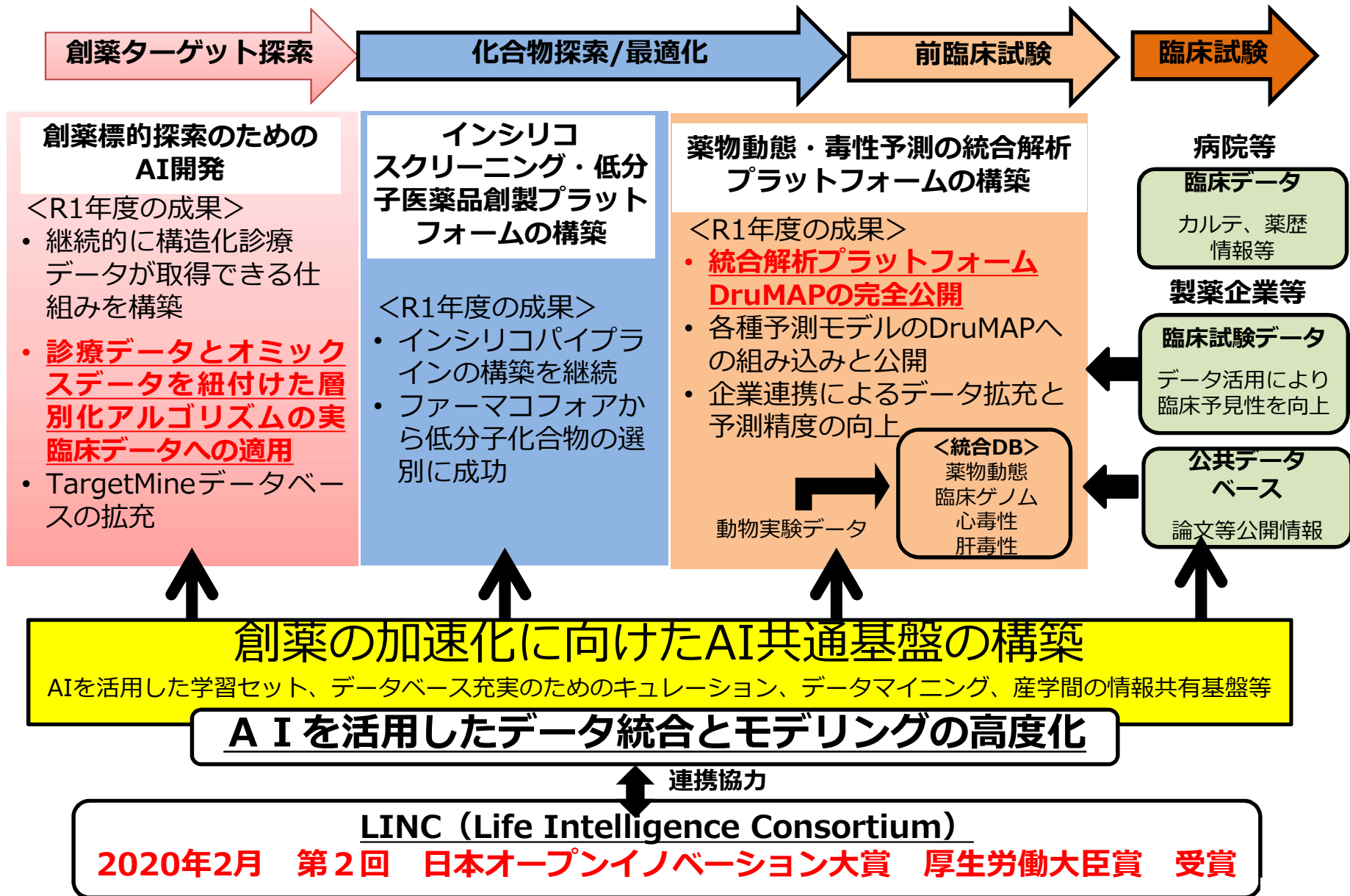
医薬基盤・健康・栄養研究所

医薬基盤研究所

AI健康・医薬研究センター

- バイオインフォマティクスプロジェクト  
(2020年11月時点 : 21名)
  - プロジェクト業務
  - PRISM業務
  - LINC業務
- AI研究企画推進室  
(2020年11月時点 : 5名)
  - 知財・契約・渉外・企画
  - ITインフラ関連の支援
  - PRISM・LINC業務のマネジメント&サポート

# AI/インフォマティクス研究の成果



# AIによる新規創薬ターゲット 探索 (PRISMプロジェクト)

# ヒトで有効な創薬標的の探索

## 臨床情報

複数の医療機関と協力

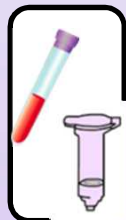
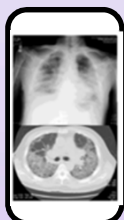
・特発性肺線維症 (IPF) + 肺がん

### 構造化した診療情報

電子カルテ

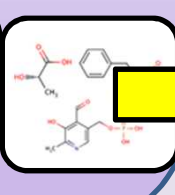
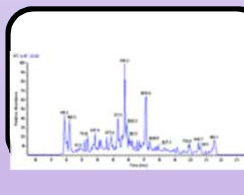
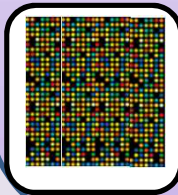
画像

血液・組織



### 多層の分子情報

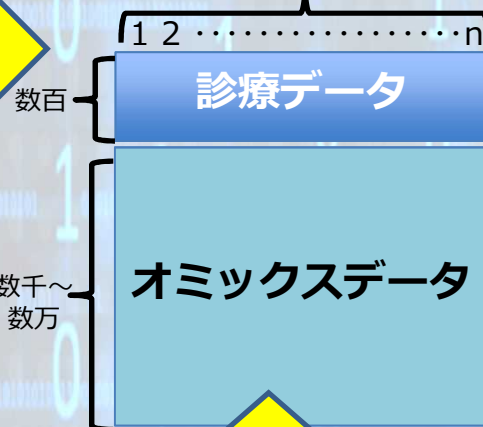
トランスクリプトーム プロテオーム メタボローム



## 人工知能 (AI)

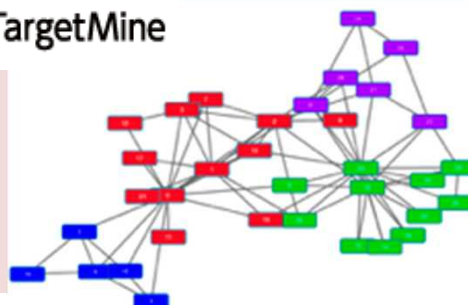
診療情報と分子情報との紐付

症例数



分子ネットワークへのマッピングと解釈

TargetMine



自然言語処理

## 新規治療法の探索

疾患の原因となる生体分子

同定された創薬ターゲット候補を細胞および動物により検証

細胞



動物



# ヒトで有効な創薬標的の探索

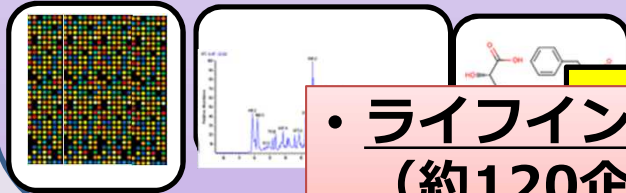
## 臨床情報

複数の医療機関と協力

- 約300件の臨床情報を構造化 (阪大)
- 神奈川循環器呼吸器病センターにおけるデータ収集を開始
- 患者層別化AIを完成

## 多層の分子情報

トランスクリプトーム プロテオーム メタボローム



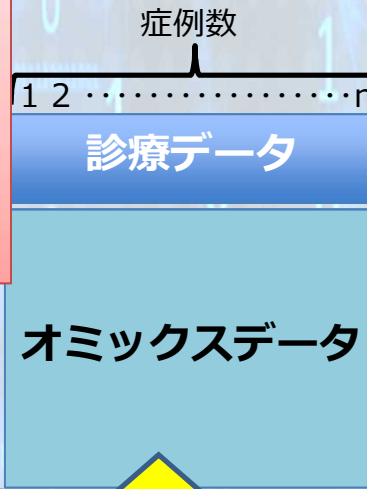
## 既知情報

臨床研究文献、  
分子生物学文献

自然言語処理

## 人工知能 (AI)

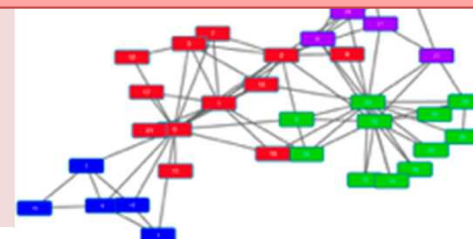
診療情報と分子情報との



## 新規治療法の探索

疾患の原因となる  
生体分子

同定された創薬  
ターゲット候補  
を細胞および動物  
により検証



# 診療情報とオミックスデータを紐づけるAI

タンパク質 A	HIGH
タンパク質 B	HIGH
タンパク質 C	HIGH
タンパク質 D	HIGH

これらのタンパク質発現量が高い傾向の患者は

牽引性気管支拡張:p:下葉|両側肺  
すりガラス影:p:下葉|両側肺  
間質性肺炎:p:  
網状影:p:下葉|両側肺

CT画像の読影所見にこれらのキーワードが含まれる傾向がある

← **IPFの特徴**

**実データを用いて、AIによるデータ駆動的な患者層別化と標的探索に成功した**



# 薬物動態予測システム

# 創薬支援インフォマティクスシステム構築

～企業連携による大規模データベース及び高精度化予測モデル構築～



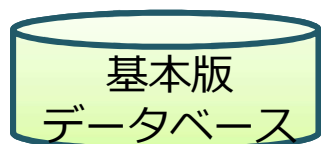
研究統括：医薬基盤健康栄養研究所 水口賢司



研究代表：理化学研究所 本間光貴

## 基本版： データベース・予測システム

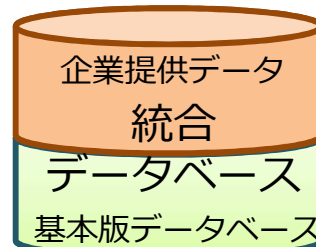
- ◆ 基本版データベース
- ★化学構造式★各種公開データベース (ChEMBL, KEGG Drug, hERG centralなど)からのキュレーションデータ
- ★市販化合物を用いた新規取得データ



インシリコ予測  
モデル  
・薬物動態  
・心毒性  
(hERG阻害)

## 企業連携版： データベース・予測システム

- ◆ 企業提供データ (約24,300化合物)
- ★構造記述子★実創薬データ (溶解性・代謝安定性・血漿タンパク結合・hERG阻害など)
- ★企業提供化合物を用いた新規取得データ



高精度化インシリコ予測モデル  
・薬物動態  
・心毒性  
(hERG阻害)



連動



データ及び化合物を提供

統合データベース・予測モデルを提供

## 基本版： 公開用Webアプリ

- ◆ 拡張機能
- ★データベース検索
- ★血漿中濃度推移のグラフ表示機能
- ★ケミカルスペース表示機能
- ★解析ツール
- ★構築した予測モデルの統合

## 参画 企業

アステラス製薬  
大塚製薬  
小野薬品工業  
第一三共  
武田薬品工業  
田辺三菱製薬  
日本たばこ産業

## 創薬 ブースター

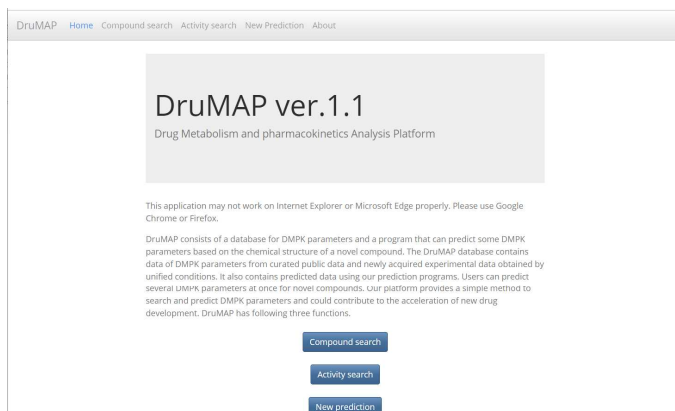
創薬支援ネットワークでの活用

産学官連携によるアカデミア・製薬企業での創薬基盤強化

# 統合DB及び薬物動態予測モデルの完全公開 (DruMAP)

## 統合データベースの完全公開

2020/07/21 DruMAP Ver. 1.1	化合物数
基盤研での試験データ	480
公共データ(curated)	12,078
肝毒性 DBへのリンク	620
hERG データ	9,114
Cav, Kv, Nav データ	2,211
薬物動態予測データ	27,239
合計	30,633



\*; 2019年度に構築したモデル (12)  
 \*; 2019年度にDruMAPにて公開したモデル (13)  
 \*; 2019年度に発表した論文 (3報)  
 \*; 2020年度にDruMAPにて公開したモデル (1)  
 ○; 公開準備中のモデル

## 予測モデルの構築・DruMAPへの組み込みと公開

No.	薬物動態パラメータ	細胞・動物種・ 酵素種・手法など	公開	論文
1	膜透過性	Caco-2	●*	Esaki et al. 2019*
2	溶解度	Dried-DMSO	●*	
3	吸収率*	-	●*	
4	膜透過性*	LLC-PK1	●*	
5	P-glycoprotein 輸送能*	ヒト	●*	
6	脳ホモジネート結合	ラット	●*	Esaki et al. 2019*
7	血漿タンパク結合1	ヒト	●	Watanabe et al, 2018
8	血漿タンパク結合2	ヒト	●	
9	血漿タンパク結合1*	ラット	●*	
10	血漿タンパク結合2*	ラット	●*	
11	血液血漿濃度比*	ラット	○	
12	組織分布*	ラット	○	On going
13	中枢移行性*	ヒト・ラット	○	On going
14	肝固有クリアランス1	ヒト	●	Esaki et al. 2019
15	肝固有クリアランス2	肝ミクロソーム	●*	
16	CYP代謝確率*	ヒト	●*	
17	CYP代謝部位*	ヒトCYP1A2, CYP3A4	●*	
18	尿中未変化体排泄率	ヒト	●*	Watanabe et al, 2019*
19	尿中排泄型*	ヒト	●*	
20	腎クリアランス*	ヒト	●*	

# エコシステム（商用版）

本プロジェクト成果物であるデータベース、システムの維持と発展を目的に、企業連携で作成した予測モデル含めてシステムを商用化する

## 予測モデルの作成

## 予測

①  
インプット



②  
機械学習



③  
検証



④  
予測



⑤  
予測結果表示



⑥  
リデザイン



学習用データ整理

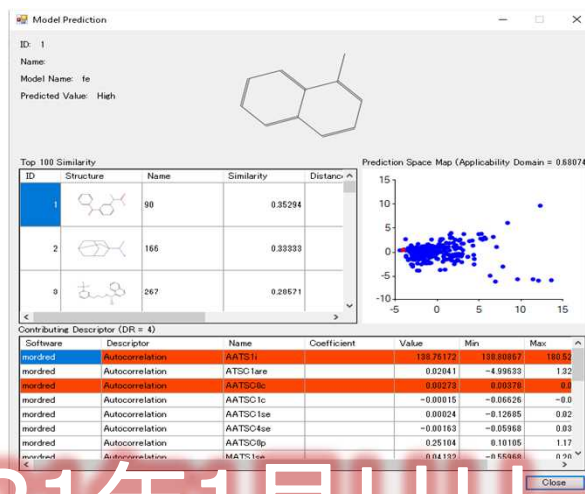
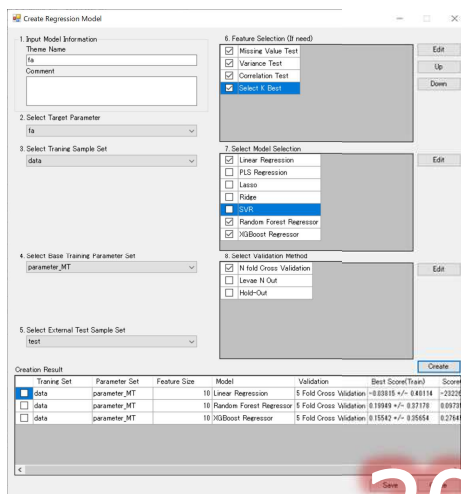
独自モデル作成  
既存モデル改良

モデルの  
評価/検証

化学構造式のみでの  
薬物動態・毒性予測

信頼性・妥当性の  
評価

開発初期段階での  
薬物動態・毒性予測



2021年1月リリース予定

# ヒトを対象とした生活習慣と腸内環境、健康に関する研究 @ 医薬基盤・健康・栄養研究所 (NIBIOHN)



NIBIOHNの  
コホート研究

匿名化番号	<input type="checkbox"/>
性別	<input type="checkbox"/>
年齢	<input type="checkbox"/>
M値	<input type="checkbox"/>
インスリン第1相	<input type="checkbox"/>
インスリン第2相	<input type="checkbox"/>
中性脂肪含量(骨格筋)	<input type="checkbox"/>
中性脂肪含量(肝臓)	<input type="checkbox"/>
アディポネクチン	<input type="checkbox"/>
レプチン	<input type="checkbox"/>



約200項目

約100項目

日本各地にお住まいの方の生活習慣や健康診断情報

睡眠



遺伝子多型

ゲノムワイド  
関連解析など

生活習慣

食事 (+BDHQによる  
各栄養素摂取情報)、  
睡眠、運動、服薬、病歴

生理指標

健康診断データ (血液デー  
タ、血圧、BMIなど)  
身体活動 (運動強度など)

約1,000項目

約100項目

測定項目

- (1)健康診断データ
- (2)疾患情報
- (3)血液 (血清、血漿)
- (4)糞便 (保存液、凍結)
- (5)唾液 (凍結)
- (6)食生活 (BDHQなど)
- (7)身体活動量 (活動量計)

マイクロバイーム



メタボローム

統合データベースの作成



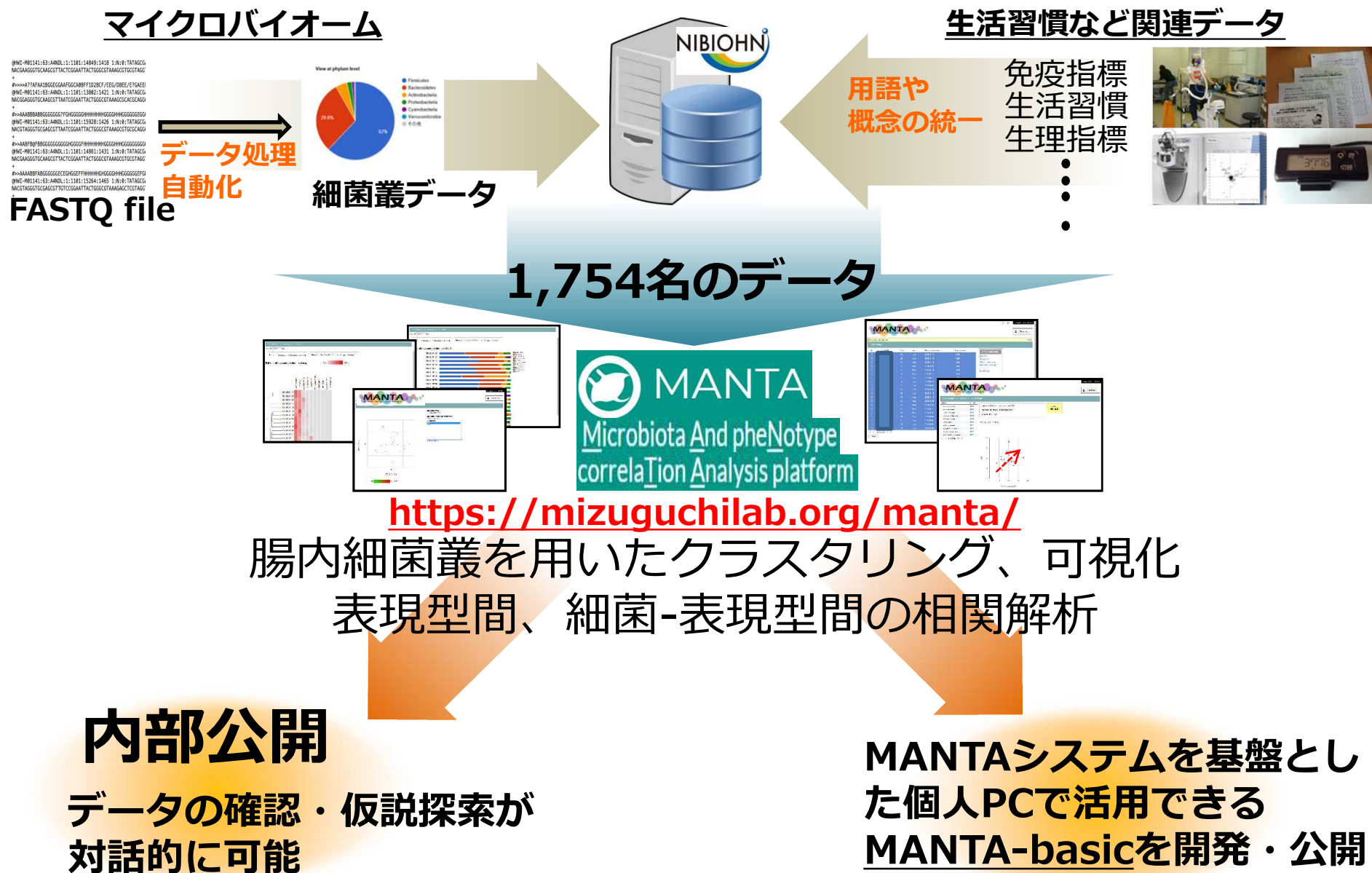
約1,000項目

免疫指標

抗体 (ワクチン)、サイ  
トカイン、代謝物など

情報科学を駆使  
した最先端解析

# MANTAプラットフォームを用いたデータベースの活用



# 今後に向けて

## ➤ データ統合と整備

- 公知データの構造化、知識化
- データ共有、再利用を可能にする仕組み作り

## ➤ 人材育成

- アカデミア、企業、海外からの人材受入
- マッチングの場の提供



## ➤ オープンプラットフォームの構築

- 既存のプラットフォーム（LINC等）との連携
- データばかりでなく、AIアルゴリズムを含むシステムの公開を通じ、研究開発の促進と人材の育成を図る