

# トーゴーの日に想うこと

九州大学医学研究院  
伊藤 隆司



# そもそも私はなぜここにいるのか — JST BI事業と私の関わり

## BIRD発足 2001

バイオインフォマティクス推進事業 一期生 (2001~2006)

## NBDC準備室発足 2010

トーゴーの日@本郷キャンパス浅野地区 (2010.10.5)

「ライフサイエンスの未来へ 10年先のデータベースを考える」

“オーミクス研究者から見たデータベース：データ生産者の目線”

## NBDC発足 2011

基盤技術開発プログラムアドバイザー (長洲毅志 総括 2011~2013)

NBDC運営委員会委員 (堀田凱樹 委員長 2014~2019)

統合化推進プログラムアドバイザー (高木利久 総括 2014~2017)

同 上 アドバイザー (長洲毅志 総括 2017~2019)

同 上 副統括 (長洲毅志 総括 2019~2020)

同 上 統括 (2020~)

トーゴーの日@JST & Web (2021.10.5)

「データベース統合 10年のあゆみと未来への期待」



# BIRDの頃

## BIRD一期生 (2001~2006)

生命情報データベースの高度化・標準化

金久 實、菅原秀明、高木利久、中村春木

創造的な生物・情報知識融合型の研究開発

情報科学と生物科学との融合型アプローチによる研究開発

森下真一、伊藤隆司、大浪修一、伊藤 啓

情報生物学に関する創造的な研究開発

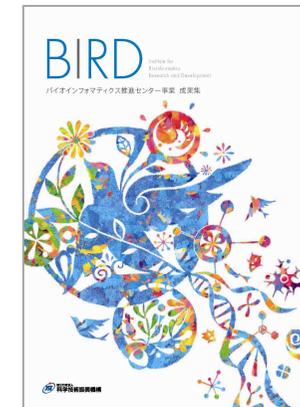
矢田哲史、肥後順一、岸野洋久

当時のキーワード

ポストゲノム (ゲノム界限ではポストシーケンス)

ドライとウェット

仮説発見型・データ駆動型



### ゲノム情報生物学

ポストシーケンスのゲノム科学⑥

<http://rr2.nakayamashoten.co.jp/products/4-521-61031-5>

### プロテオミクス

タンパク質の  
系統的・網羅的解析

ポストシーケンスのゲノム科学③

[https://honto.jp/netstore/pd-book\\_01959559.html](https://honto.jp/netstore/pd-book_01959559.html)

# データ統合とその変容

## 生命科学研究の常態

“論文Aと論文Bを読んで思いついたことを  
自分の実験系で確かめて論文Cを書きました”  
= データ統合による仮説生成と実験的検証

## 変わったのはデータの量と質

ゲノムワイド → 計算機が必要になる  
再利用性が高い

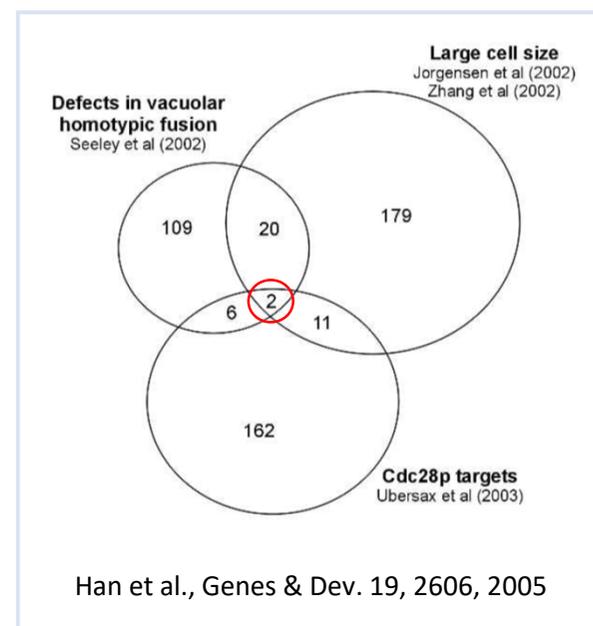
共通プラットフォーム → 比較しやすい

当時印象に残った論文

公開データ phenome x 3 + proteome 

細胞周期 → (Cln3 → Bem1) → 液胞機能

「オミクスデータはこう使わないと…」

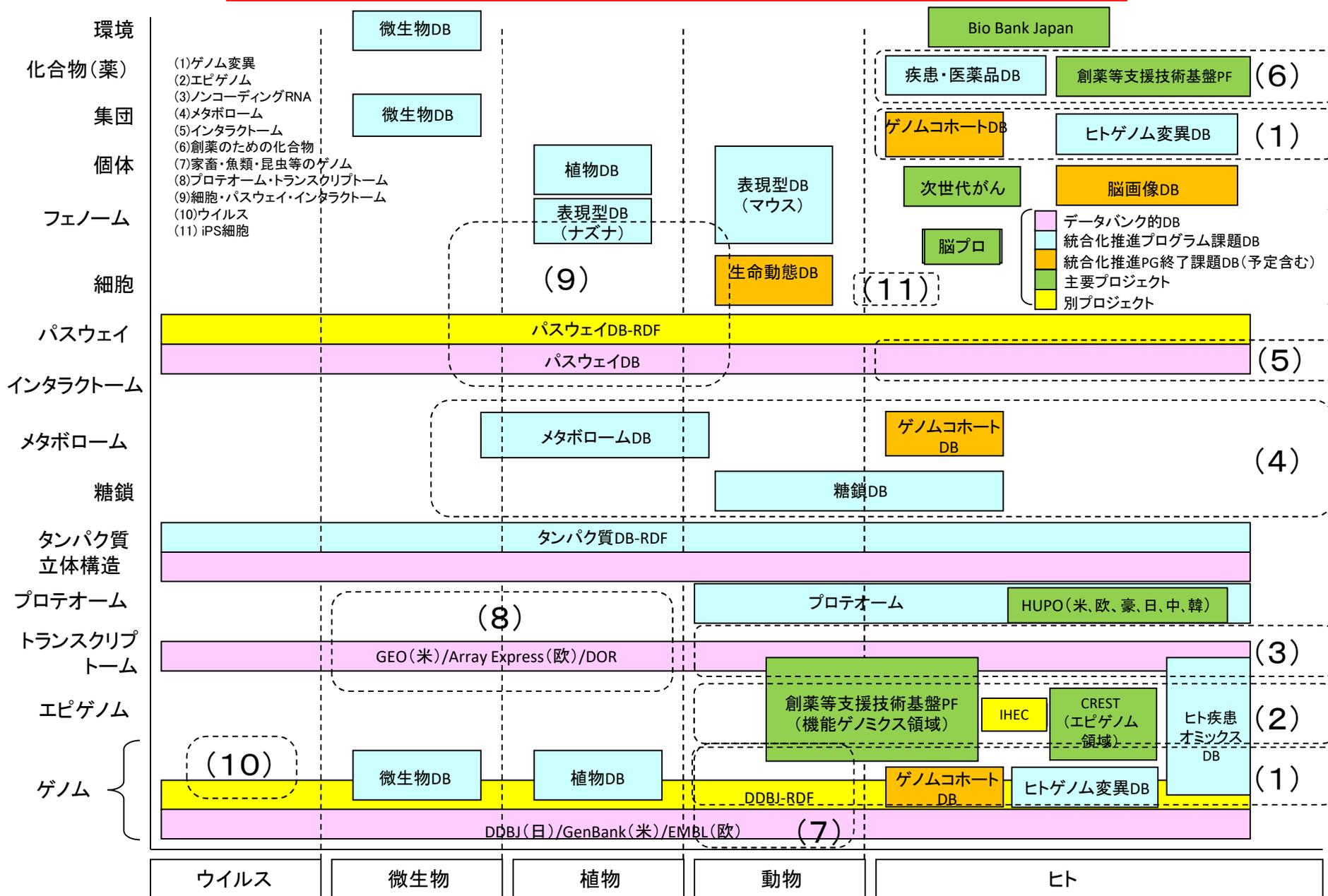


# 階層を跨いだオミクスデータの統合によって 新しい知識を得る仕組み

- そもそもデータはあるのか？
- あるとしたら、どこかに集められているのか？
  - ☞ まだ集められていないデータは何か？



# 全体の俯瞰に基づく戦略的なDB育成



# 統合化推進プログラムで 現在支援中のデータベース

## 研究アドバイザー

岩崎 渉	東京大学 大学院新領域創成科学研究科 教授
鎌田 真由美	京都大学 大学院医学研究科 准教授 (2020年9月~)
清水 佳奈	早稲田大学 理工学術院 教授 (2020年7月~)
白水 美香子	理化学研究所 生命機能科学研究センター タンパク質機能・構造研究チーム チームリーダー (2020年6月~)
瀬々 潤	株式会社ヒューマノーム研究所 代表取締役社長
馬場 健史	九州大学 生体防御医学研究所 附属トランスオミクス医学研究センター 教授 (2020年6月~)
山本 一夫	東京大学 大学院新領域創成科学研究科 教授 (2020年6月~)
吉田 哲郎	ナノキャリア株式会社 研究部 部長 (2020年6月~)

(2021年4月現在；五十音順に掲載)

- 国際的にも認知された老舗のDB
- BIRD時代からの流れを引き継ぐDB
- 戦略的に育成したDB
- ユーザー目線の新興DB

### MetaboBank

物質循環を考慮したメタボロミクス情報基盤

有田 正規

進行中

統合化推進プログラム

2018年度採択



プロテオームデータベースの機能深化と連携基盤強化

石濱 泰

進行中

統合化推進プログラム

2018年度採択

### ChIP-Atlas

エピゲノミクス統合データベースの開発と機能拡充

沖 真弥

進行中

統合化推進プログラム

2017年度採択



ゲノム・疾患・医薬品のネットワークデータベース

金久 寛

進行中

統合化推進プログラム

2017年度採択



糖鎖科学ポータル構築

木下 聖子

進行中

統合化推進プログラム

2017年度採択



蛋白質構造データバンクのデータ検証高度化と統合化

栗栖 源嗣

進行中

統合化推進プログラム

2017年度採択



データサイエンスを加速させる微生物統合データベースの高度実用化開発

黒川 顕

進行中

統合化推進プログラム

2017年度採択



疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オームデータの統合

菅野 純夫

進行中

統合化推進プログラム

2017年度採択



個体ゲノム時代に向けた植物ゲノム情報解析基盤の構築

田畑 哲之

進行中

統合化推進プログラム

2017年度採択

# 階層を跨いだオミクスデータの統合によって 新しい知識を得る仕組み

- そもそもデータはあるのか？
- あるとしたら、どこかに集められているのか？
  - ☞ まだ集められていないデータは何か？
- 各所に集められたデータは、つなげやすいのか？
  - ☞ つながる形のデータベース構築を支援

## **NBDC & DBCLS** = データ統合の基盤構築

バイオサイエンスのエッセンシャル・ワーカー  
地味だからこそ、支えなくてはならない部分

# 課題：サステナビリティ

## DBの中核部

- 競争的資金による維持継承（自転車操業）の限界  
= 成果発表費用までで支援が終わる通常のバイオ研究との違い
- 維持継承を評価する文化（評価軸の複線化）

## 作るヒトと活かすヒト

- 必要とされる2種類の人材  
データベース構築者 エッセンシャル・ワーカー（ボックス）  
バイオデータ科学者 ストーリー・テラー（フォワード）  
参考：NBDCワークショップ「データ駆動型研究の推進と課題」
- 双方の人材を評価する文化（評価軸の複線化）

結局は科学者とそれを取り巻く社会の理解

# 課題：新しいタイプのデータへの対応

## 生物学(者)は元来 greedy

物理・化学・数学・情報科学…  
使えるモノは何でも採り入れる性質  
(中身を完全に理解していなくても)

### ノーベル賞を受賞した実験手法

- 2002 MS, NMR
- 2006 RNAi
- 2007 KOマウス
- 2008 GFP
- 2014 超解像顕微鏡
- 2017 クライオEM
- 2020 ゲノム編集
- 2021 ?

## 実験手法の進歩

↓ *cf.* NGS

## 研究スタイルの変容

↓ *cf.* 個人ゲノム、一細胞解析

## データ(ベース)の変容

↓

## データ統合の需要 ↑

(含む非バイオデータ)

## 知識発見加速への要望 ↑

未来のバイオサイエンスを支える基盤としての成長発展

# 20年前のキーワード

ポストゲノム（ゲノム界隈ではポストシーケンス）

- シーケンス以外のデータも充実（データの多様化）
- シーケンスデータ自体もNGS登場以降に大きく変容
- 統合すべきデータは多様化・大規模化の一途

ドライとウェット

- 若手は多かれ少なかれバイリンガル
- 解析ツールと学習環境の充実や実験キットの進歩が双方向性のクロスボーダーを後押し
- 求められるものが変化？

仮説発見型・データ駆動型

- AI駆動の基盤もデータ(ベース)統合
- 既存の枠組み内での発見の加速
- 改めて目指すべきは新しいレベルでの生命法則の発見？

