

# データ産生・利活用における 省庁枠を超えたオールジャパン体制の必要性について

東京大学 大学院新領域創成科学研究科  
メディカル情報生命専攻/  
生命データサイエンスセンター  
東京大学 ゲノム医科学連携研究機構

鈴木 穰

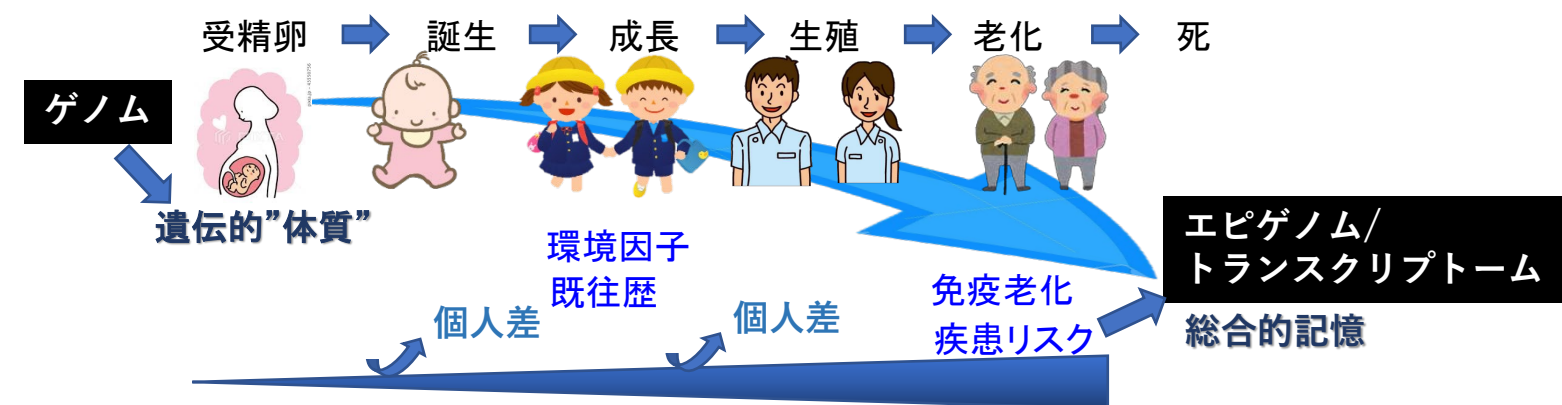
大学共同利用機関法人 情報・システム研究機構  
国立遺伝学研究所  
生命情報・DDBJセンター長

有田 正規

# 背景

# 新しいタイプの計測データ(特に動的データ)が続々生まれている

- 疾患リスク・未病状態は、個々人の エピゲノム/トランスクリプトーム/プロテオーム動態で決まる。
- つまり生来のゲノム配列だけでなく、既往歴・環境要因の総和としての情報が「治療への鍵」。



取得できるデータはゲノムだけでない。画像データ、数値データなど大きく変化。



さらなる超大量データ時代へ

新しい測定データは既に大量に産出されている。  
しかも行き場がない。

ムーンショット目標2 (超早期疾患予防)  
AMED革新的がん医療  
AMED先端ゲノム (GRIFIN)  
今後、AMEDワクチン開発 (SCARDA)



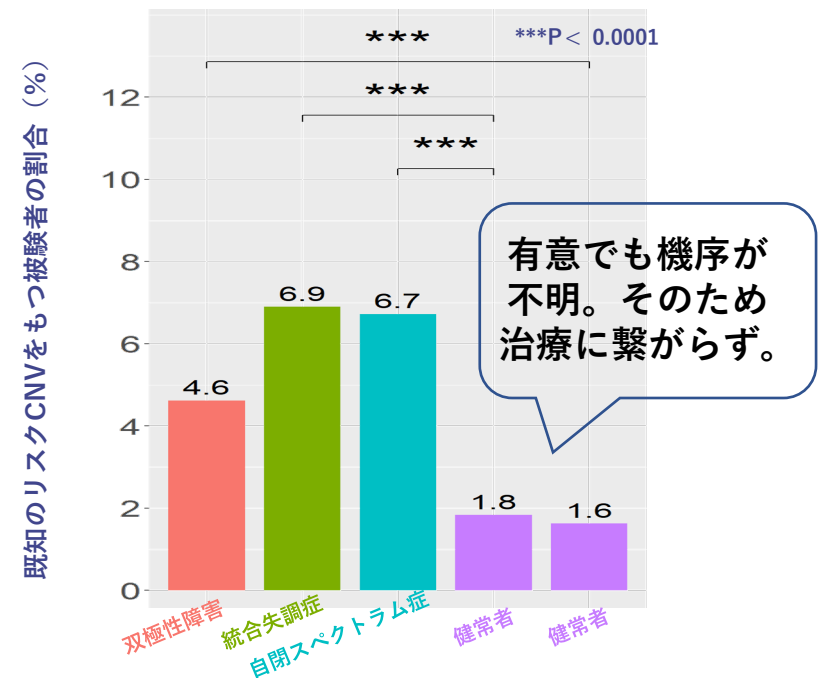
出てくるデータの標準化や受け皿が必要。現状では対応できる組織(データセンター)がない。

# 背景

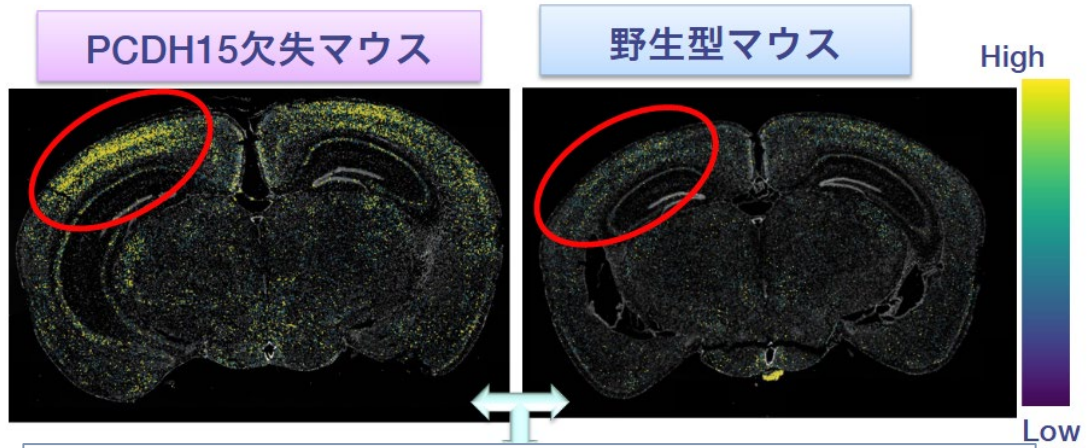
# (ゲノム以外の)空間トランスクリプトームが精神疾患解明に役立つ

PCDH15 遺伝子と双極性障害は有意に相関 (既知情報)

- 8,708例のゲノム解析
- 双極性障害 1,818 例
  - 統合失調症 3,014例
  - 自閉スペクトラム症 1,205例
  - 健常者 2,671例



PCDH15 遺伝子の空間トランスクリプトーム情報が作用解明に役立つ



脳の特定部位で発現向上することがマウスで判明

作用機序には遺伝子発現の空間情報が必須

### Xenium in situ プロファイリング

マウス脳の空間カタログ化->ヒト脳へも

Neuron	Oligodendrocyte	Layer2/3	Layer4
Glutamatergic	GABAergic	Layer5	Layer6

ヒト死後検体脳で解析済み。しかし結果を発信 (共有) する手段が無い。(学術誌に渡しても死蔵されるだけ)

こうしたデータを、有効に活用できるインフラ構築が必須

新規モダリティデータ  
(非配列データ)  
の登場を契機に

新規の開発計画が必要  
-多様なデータを標準化  
-データ利活用の基盤技術を開発

**機能ゲノムデータベース**  
=基礎研究から各階層データに  
付加価値を与える

個人情報に当てはまりにくいデータを扱う場  
(空間トランスクリプトーム、画像情報、行動データなど)  
・画像情報の効果的利活用  
・ヘルスケア・行動データの利活用

データ利活用の階層

基礎研究

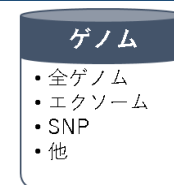
階層1 非制限公開しうるデータ



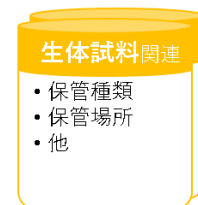
集約/  
自由な発想

階層2 制限共有あるいは制限公開レベル

ゲノム = 各階層をまたいだ利活用

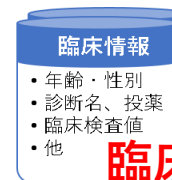


集約



データ  
カタログ

階層3 臨床情報  
公開が一般に困難

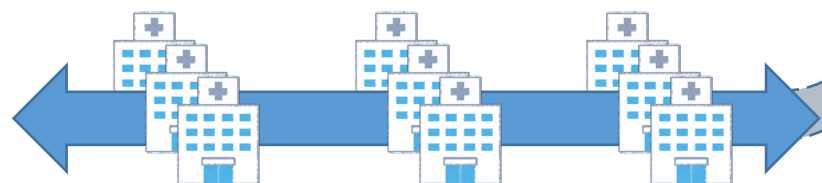
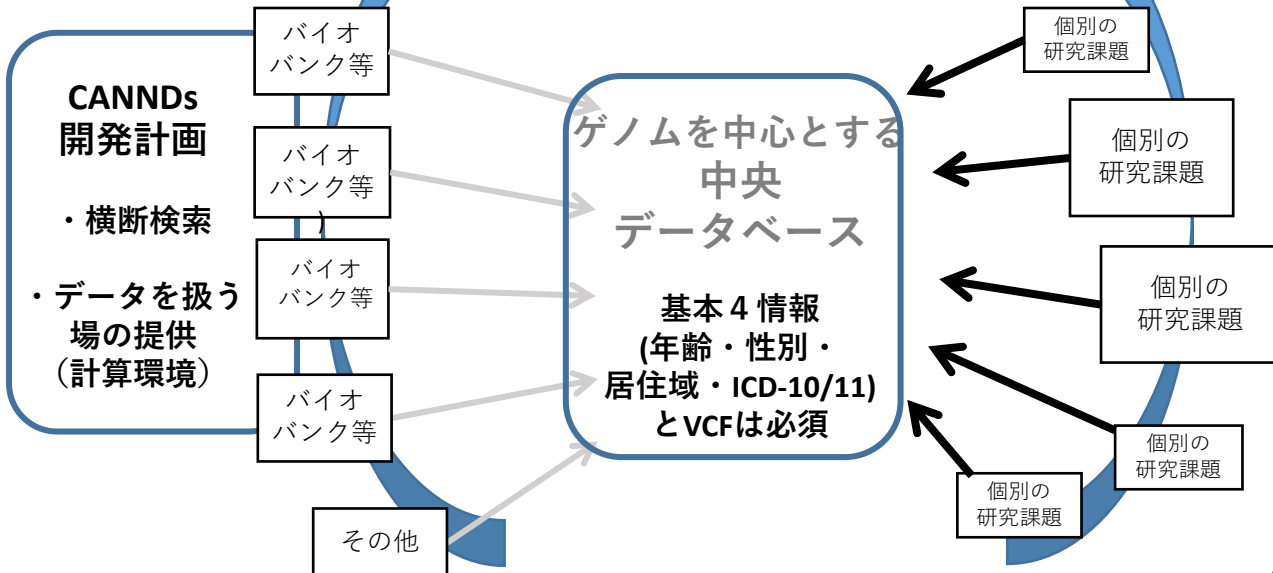


分散

臨床研究と患者還元

既存バンク等はナショナル・データベースとデータを共有、かつ自前でも解析、成果発信。

情報サービスを提供しない個別研究課題は中央データベースへの登録が必須。



# まとめ

**ゲノムだけではない、空間等オミクスデータに対応する国レベルの受け皿が必要**

**(それを各階層につなぐ中央(セントラル)データベースの構築は喫緊の課題)**

- 空間等オミクスデータは、ゲノム配列情報の価値を飛躍的に向上させる
- データ利活用の枠組みは弾力的に改変しなければ各階層が混乱：データシェアに禍根

**ヒトゲノムや生物多様性データには、中央(セントラル)データベースが必要**

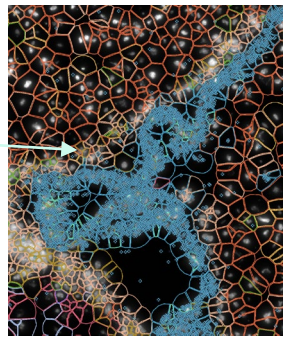
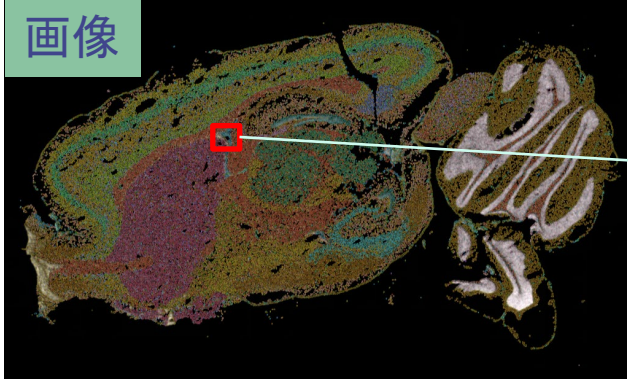
- G7声明や生物多様性条約の決定に対し、対応できるデータ管理体制がない
- データが散逸・消失するばかりでは、DXしても意味がない

**データ基盤には国の強いリーダーシップが不可欠  
産官学・省庁の枠を超えたALL JAPANでの連携を！**

### 空間解析

### 式

Mouse brain (Xenium explore)

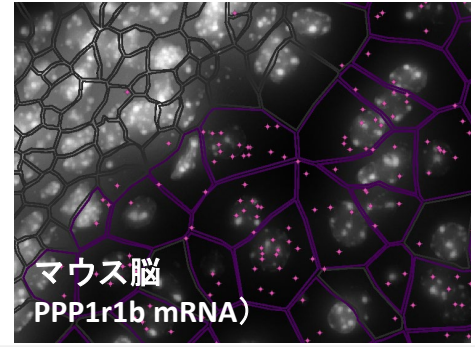


**画像**

**mRNAの座標**

```

id", "overlaps_nucleu
Lyz2", 633.43896, 629.
281483566645249, -1, 0, "Lyz2", 639.2887, 609.1
281483566645250, -1, 0, "Lyz2", 641.2621, 400.4
281483566645251, -1, 0, "Nxp3", 644.0111, 776.
281483566645252, -1, 0, "Lyz2", 645.5328, 55.87
281483566645253, -1, 0, "Bhlhe40", 648.7455, 54
281483566645254, -1, 0, "Lyz2", 651.8875, 453.2
281483566645255, -1, 0, "Lyz2", 652.56384, 697.
281483566645256, -1, 0, "Lyz2", 653.0201, 240.3
  
```



マウス脳  
PPP1r1b mRNA

**細胞/核の座標**

```

centroid", "transcript_counts", "control_probe_counts", "control_co
6171875, 272, 1, 0, 273, 476.35328125, 36.80234375
2, 629.6219362948827, 7375.214257812499, 18, 0, 0, 18, 9.4828125, 4.0640625
3, 836.3118560791015, 7371.93876953125, 370, 0, 0, 370, 205.37062500000002, 64.93468750000001
4, 820.2961730957031, 7369.998046875, 242, 0, 0, 242, 157.95656250000002, 49.581562500000004
5, 842.2104949951172, 7338.1296875, 372, 0, 0, 372, 404.058125, 12.23734375
6, 831.3291046142577, 7396.638574218749, 123, 0, 0, 123, 62.812343750000004, 12.869531250000001
  
```

### 細胞ごとに集計した遺伝子発現量

```

metadata_json: { "software_
541 245060 15666824
1 1 4
2 1 5
9 1 1
11 1 3
12 1 3
  
```

### シングルセル解析 (固定サンプル)

```

%%MatrixMarket matrix coordinate integer general
%metadata_json: ["software_version": "cellranger-7.1.0", "format_version": 2]
18082 10301 27984338
2 1 1
7 1 1
19 1 2
21 1 1
27 1 3
32 1 1
35 1 1
  
```

1列目の数字 : 遺伝子名  
2列目の数字 : 細胞ID  
3列目の数字 : 発現情報

### 血漿プロテオーム解析

```

SampleID;Index;OlinkID;UniProt;Assay;MissingFreq;Panel;Panel_Lot_Nr;PlateID;QC_Warning;LOD;NPX;Normaliz
sample_001;1;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.2
sample_009;2;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.3
sample_017;3;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_025;4;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_035;5;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_043;6;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_051;7;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_054_1;8;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_269_1;9;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_125;10;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_056;11;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_269_1;12;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
sample_002;13;O1D20419;P60568;IL2;0.9659;Inflammation;B04411;Olink_Plate_Layout_0831_mod;PASS;0.5811;0.
  
```

タンパク質毎の発現量(NPX値)が記載されたファイル

全て数字 : 個人識別符号を含まない

これらの情報だけでもな本質的に機能情報を表現する