

COXPRES  
provides co-regulatory information

# 公共の遺伝子発現データの利用 ～データを俯瞰して落とし穴を避ける～

All words

Search

» Publication

» Help

## Search

- GeneTable
- EdgeAnnotation
- CoExSearch
- etc.

## Browse

## Target species

Human (Hsa)

e.a. [aene] [aade] [aene]

実験家のためのデータ駆動科学セミナー

「データドライなライフサイエンスへの誘い」

2020年7月17日(金)

Mouse (Mmu)

e.a. [aene] [aade] [aene]

Rat (Rno)

e.a. [aene] [aade] [aene]

## What's new

2020.06.27

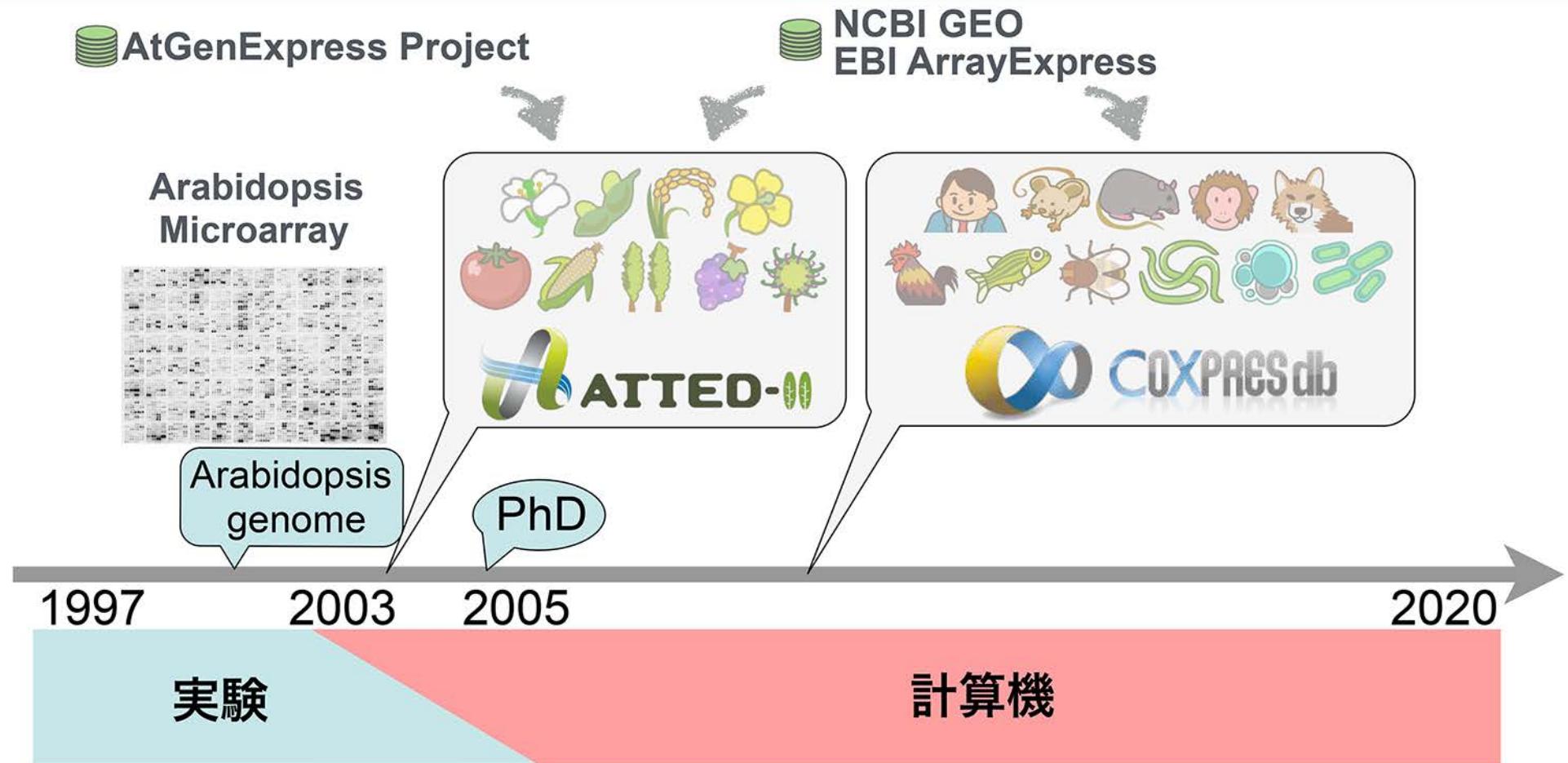
The API is upgraded, allowing a multiple gene query. Uniprot ID query is acceptable as well as Entrez Gene ID [API].

2019.07.01

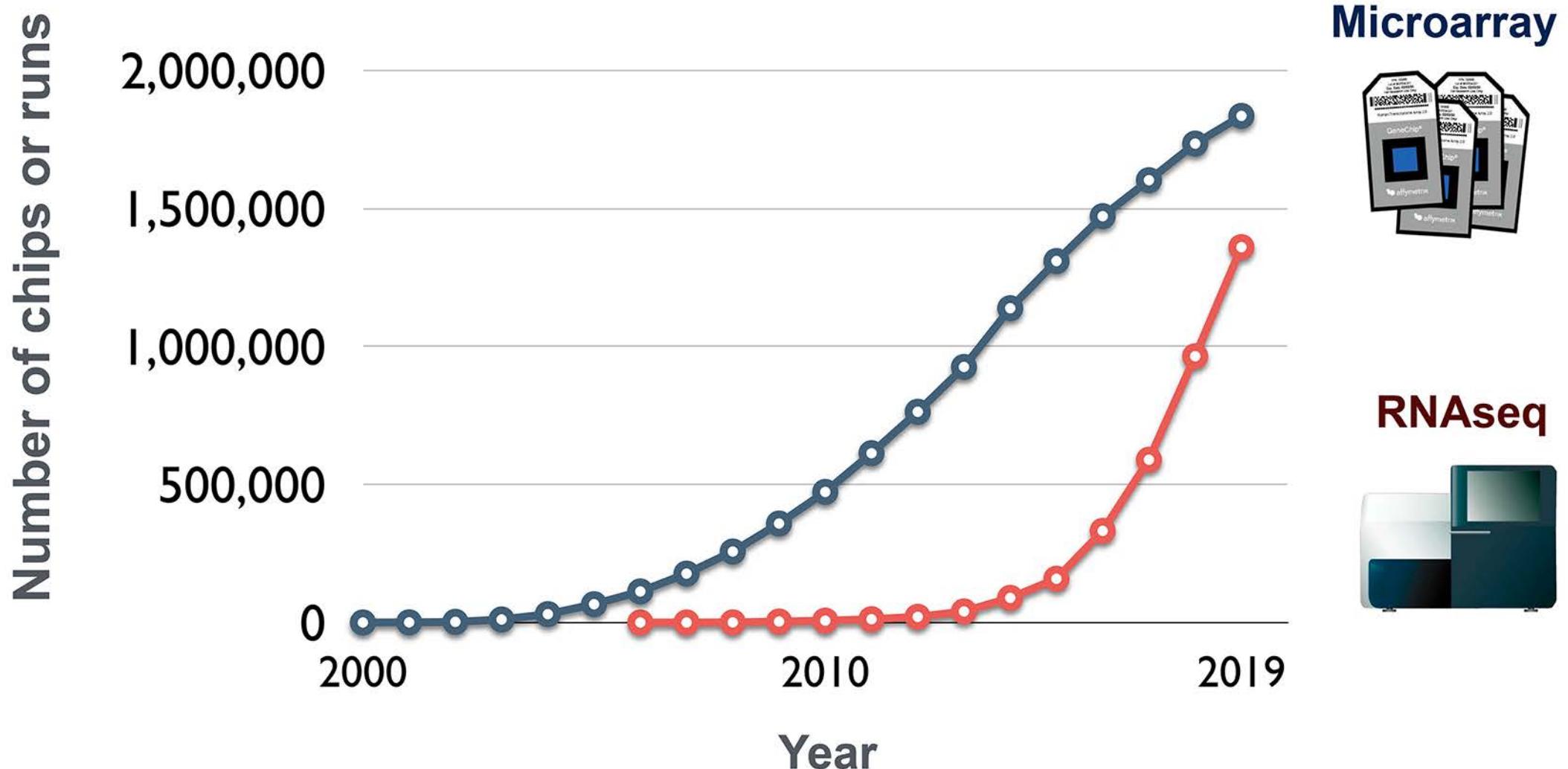
COXPRESdb ver. 7.3 is released. This version up is to improve response time for user queries.

more

# 自己紹介



# 遺伝子発現量データの蓄積



Microarray



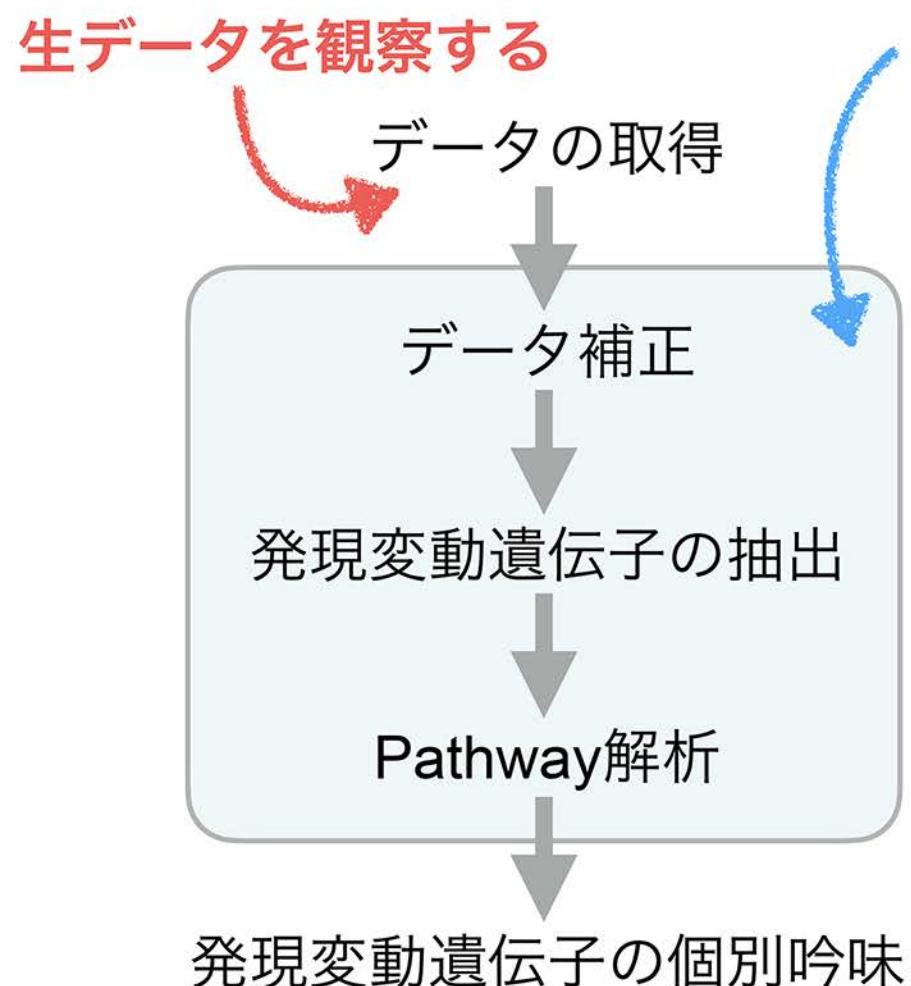
RNAseq



NCBI GEO Repository browserに基づく集計値

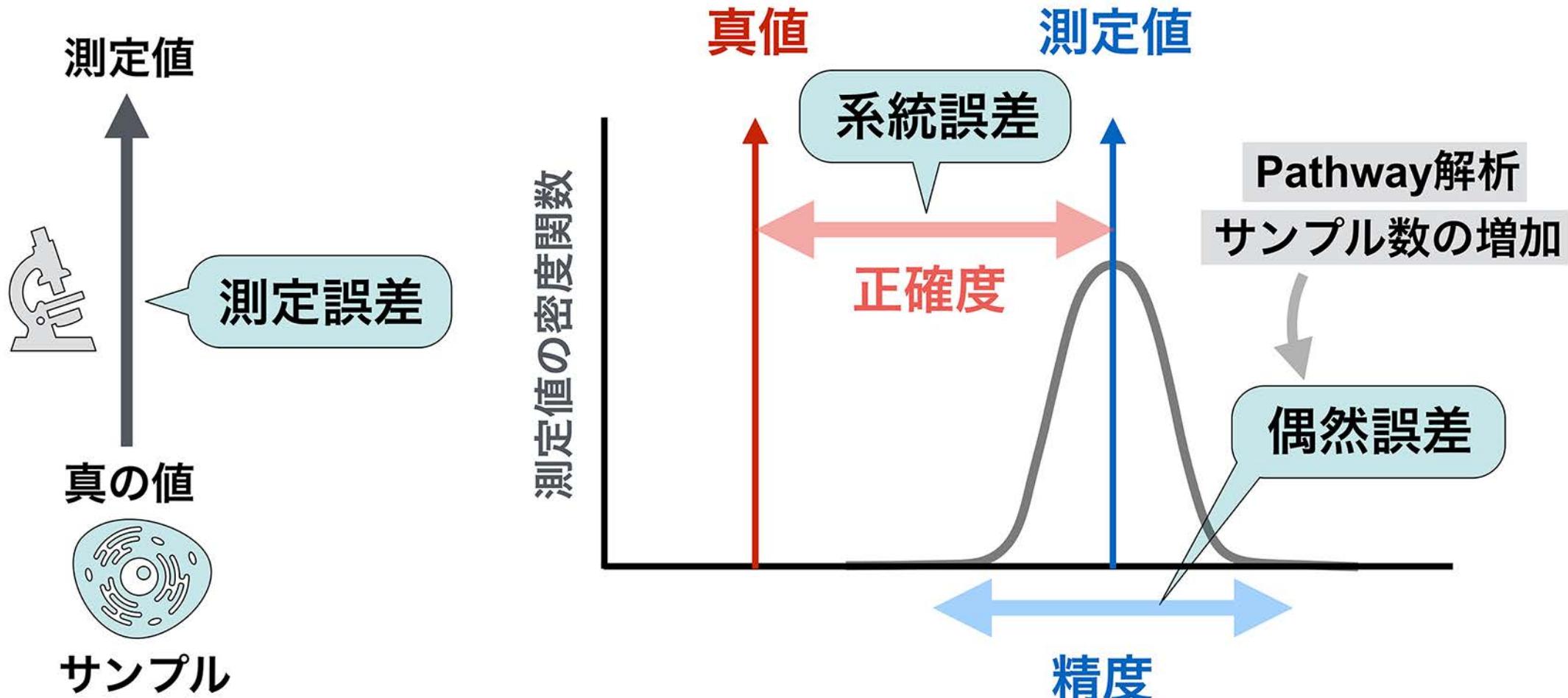
画像 © 2016 DBCLS TogoTV

# 遺伝子発現量解析の標準的フロー



<https://togotv.dbcls.jp/20190922.html>

# 測定値は測定誤差を含んでいる



系統誤差を補正するには、真値についての仮定が必要

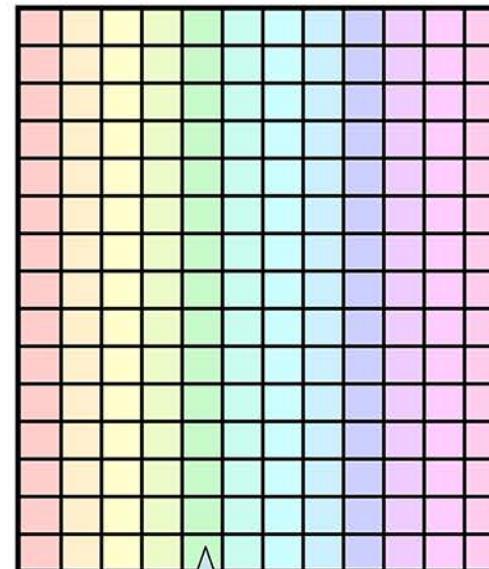
# デモ解析：Microarray

## TRANSBIG コンソーシアム



(11837 遺伝子)

22283 プローブ



MAS5 normalization

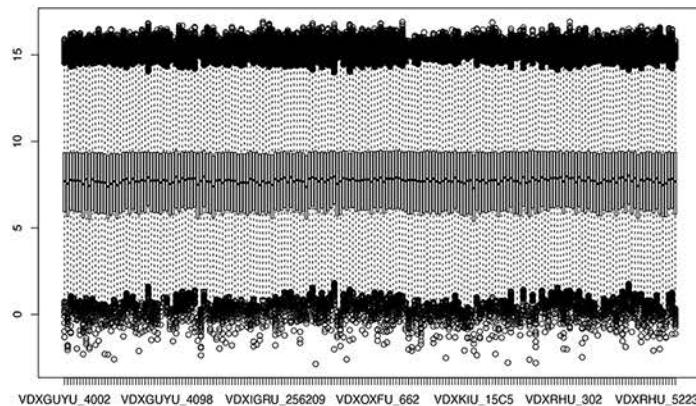
(GPL96) Affymetrix Human Genome U133A Array

- ・病院名 (5)
- ・年齢 [year]
- ・がん組織の大きさ [mm]
- ・手術方式 (2)
- ・組織型 (7)
- ・悪性度 (3)
- etc.

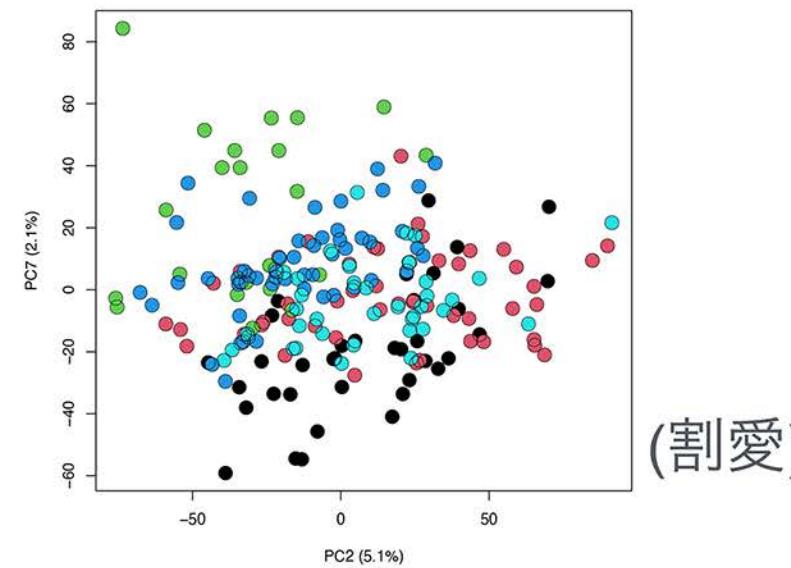
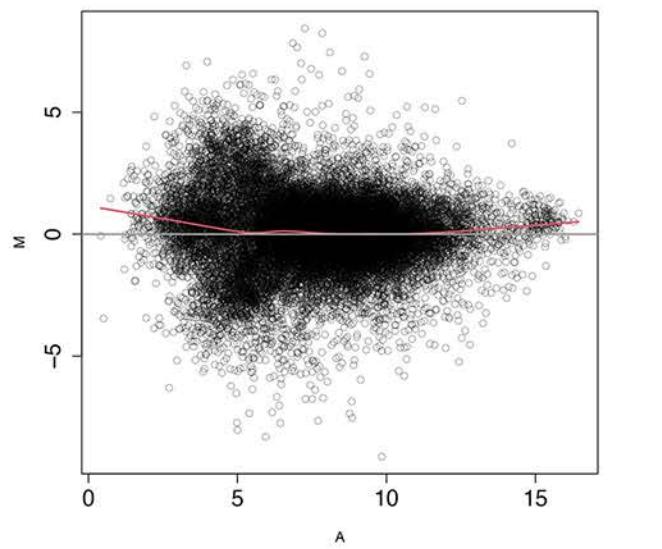
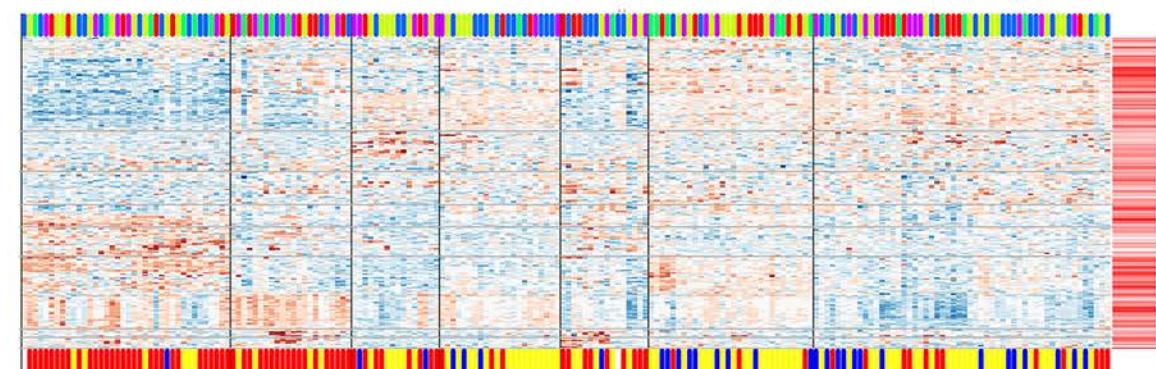
Desmedt+ (Clin Cancer Res 07)  
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7390>

# (生)データの俯瞰

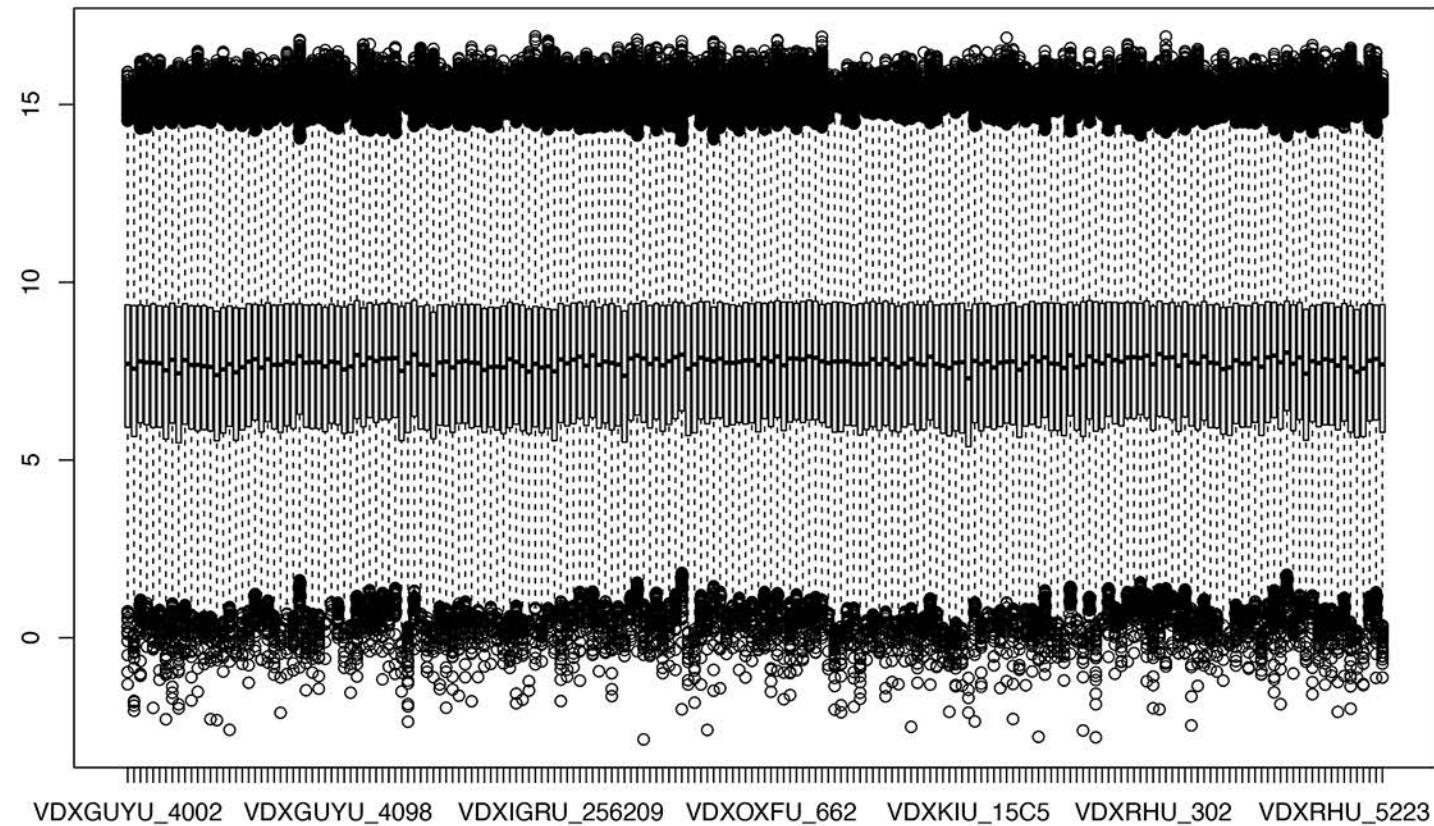
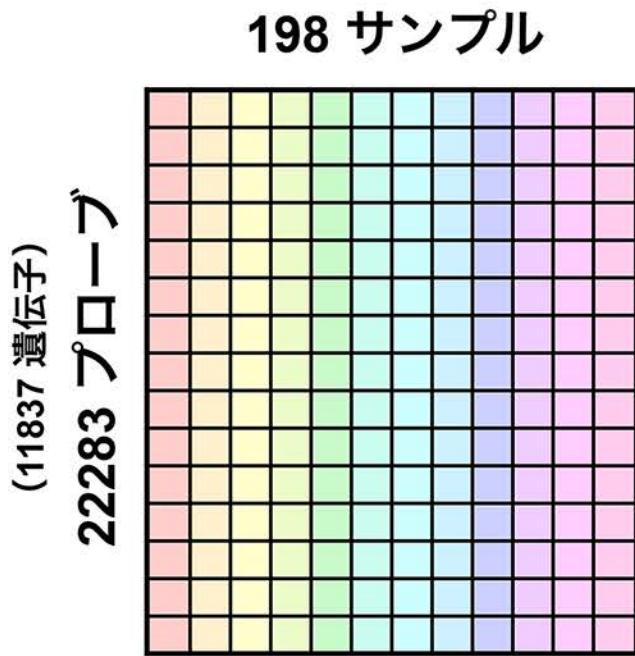
測定値の分布



サンプル全体の様子



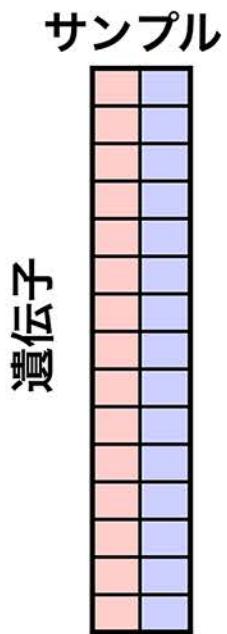
# (1) サンプルの偏り by 測定値の分布



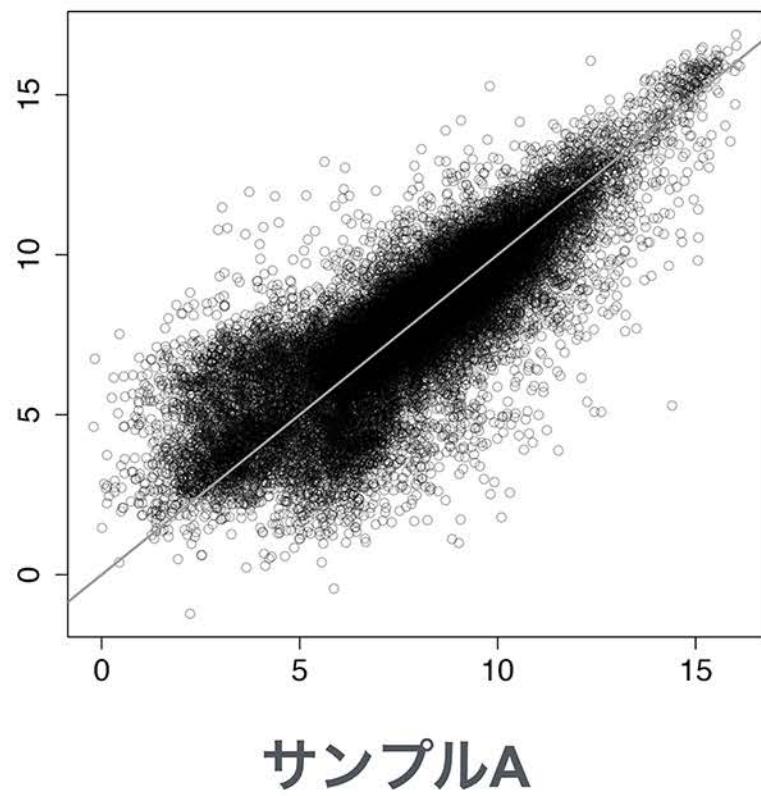
シグナル値の分布は、サンプルが異なってもほぼ同一

## (2) サンプルの偏り by ペアで比較

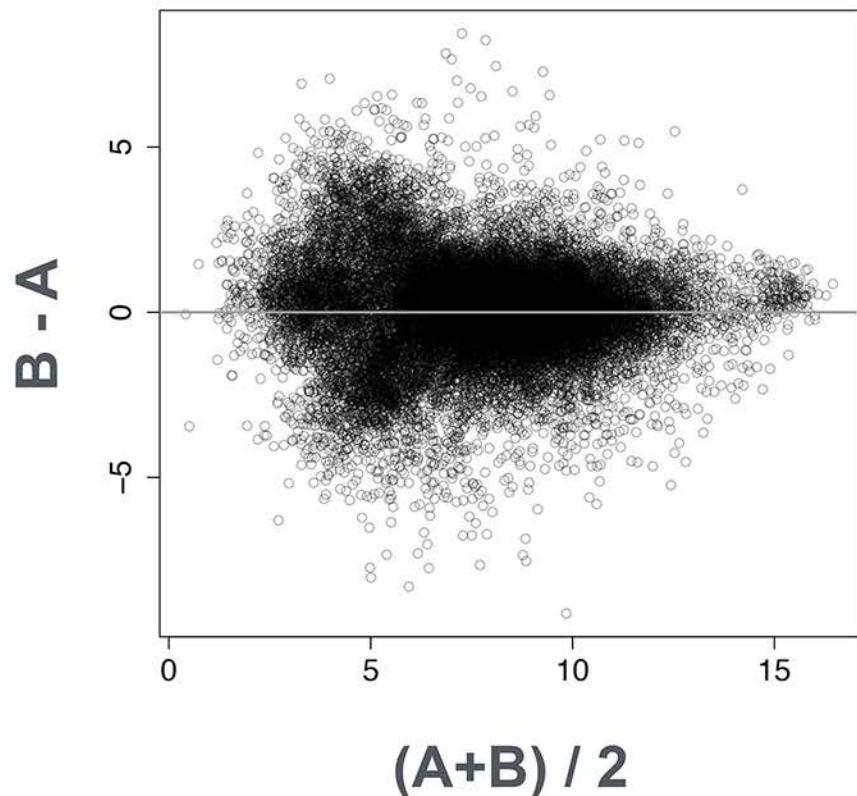
2サンプル



発現量の散布図



MAプロット



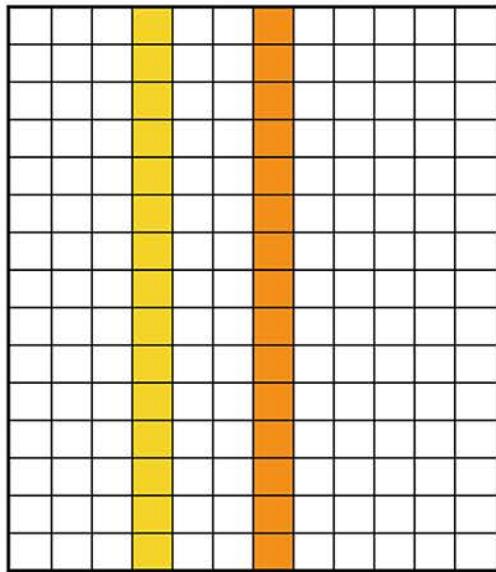
MAプロットは元の数値データの全てを図示したもの

## (2) サンプルの偏り by ペアで比較

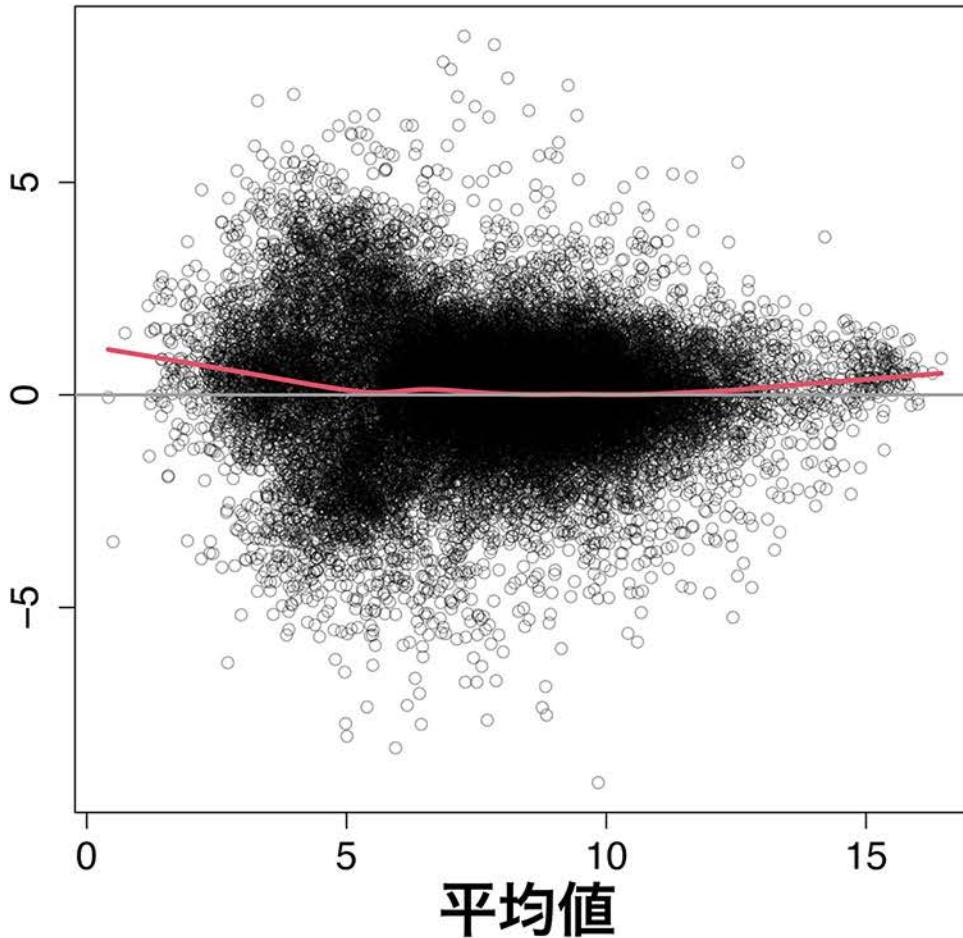
GSM178023, GSM177998

198 サンプル

(11837 遺伝子)  
22283 プローブ



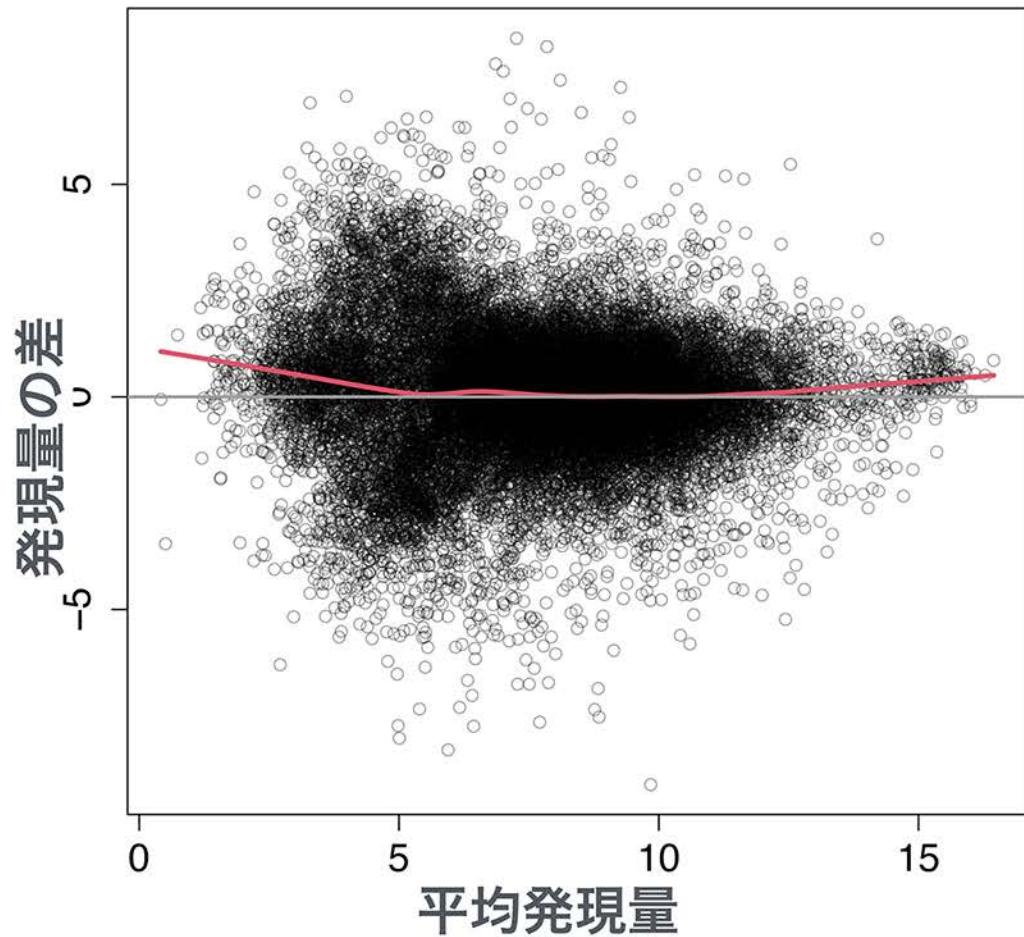
発現量の差



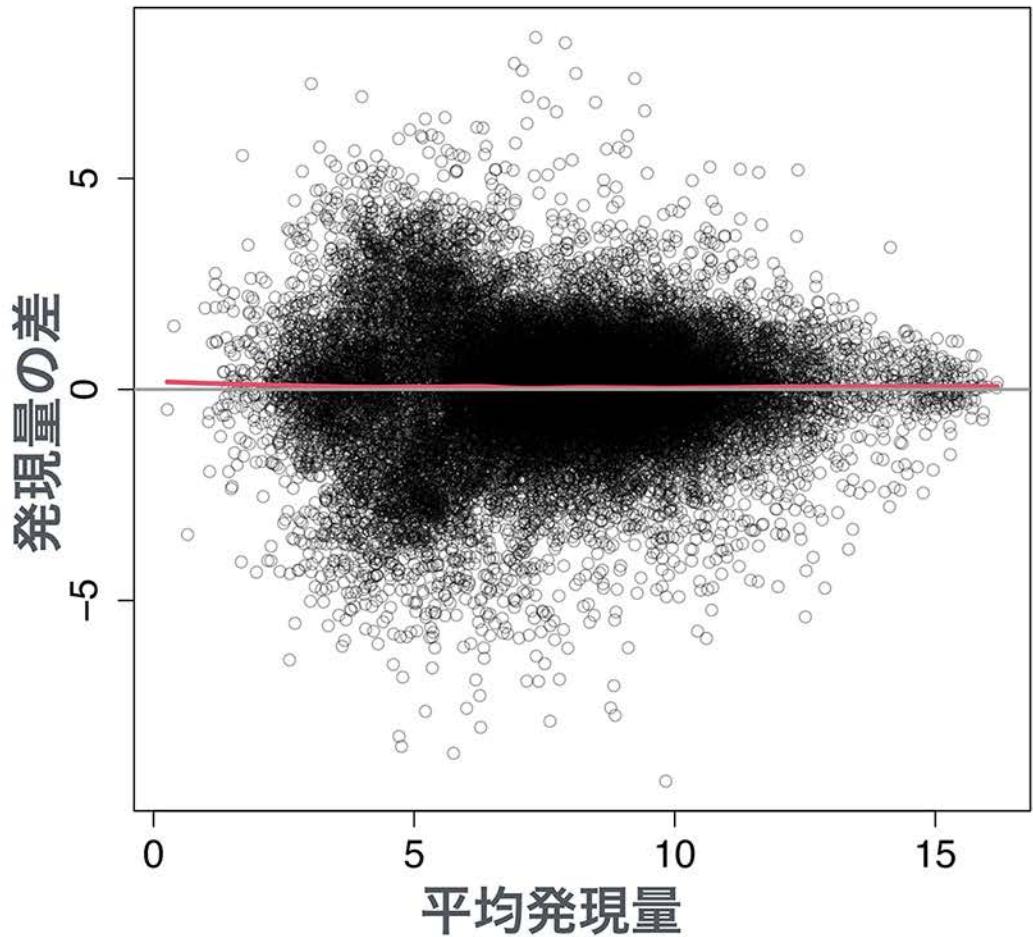
発現量依存の系統誤差が少しある

# 発現量依存の偏りを補正する

補正前

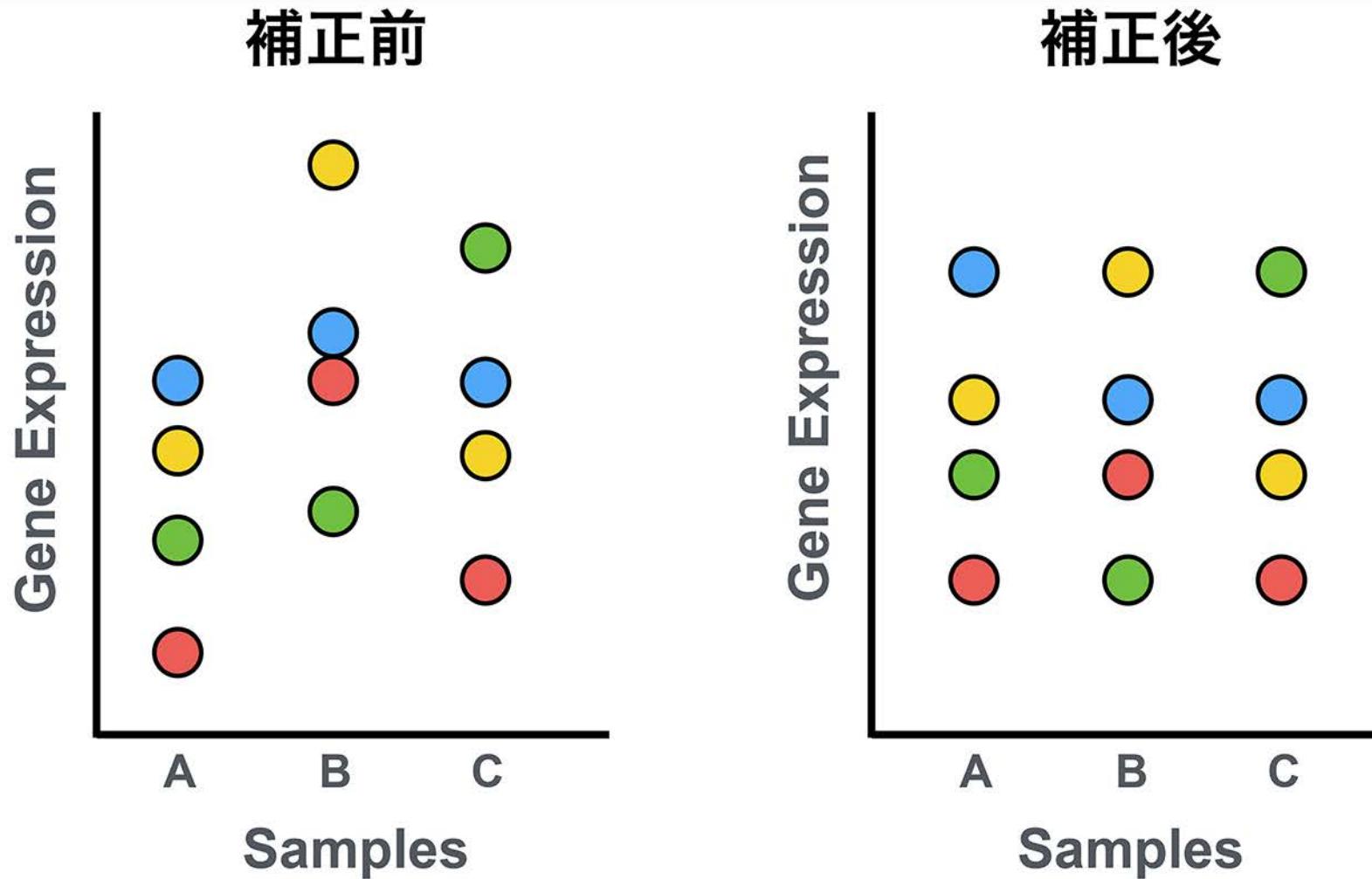


補正後



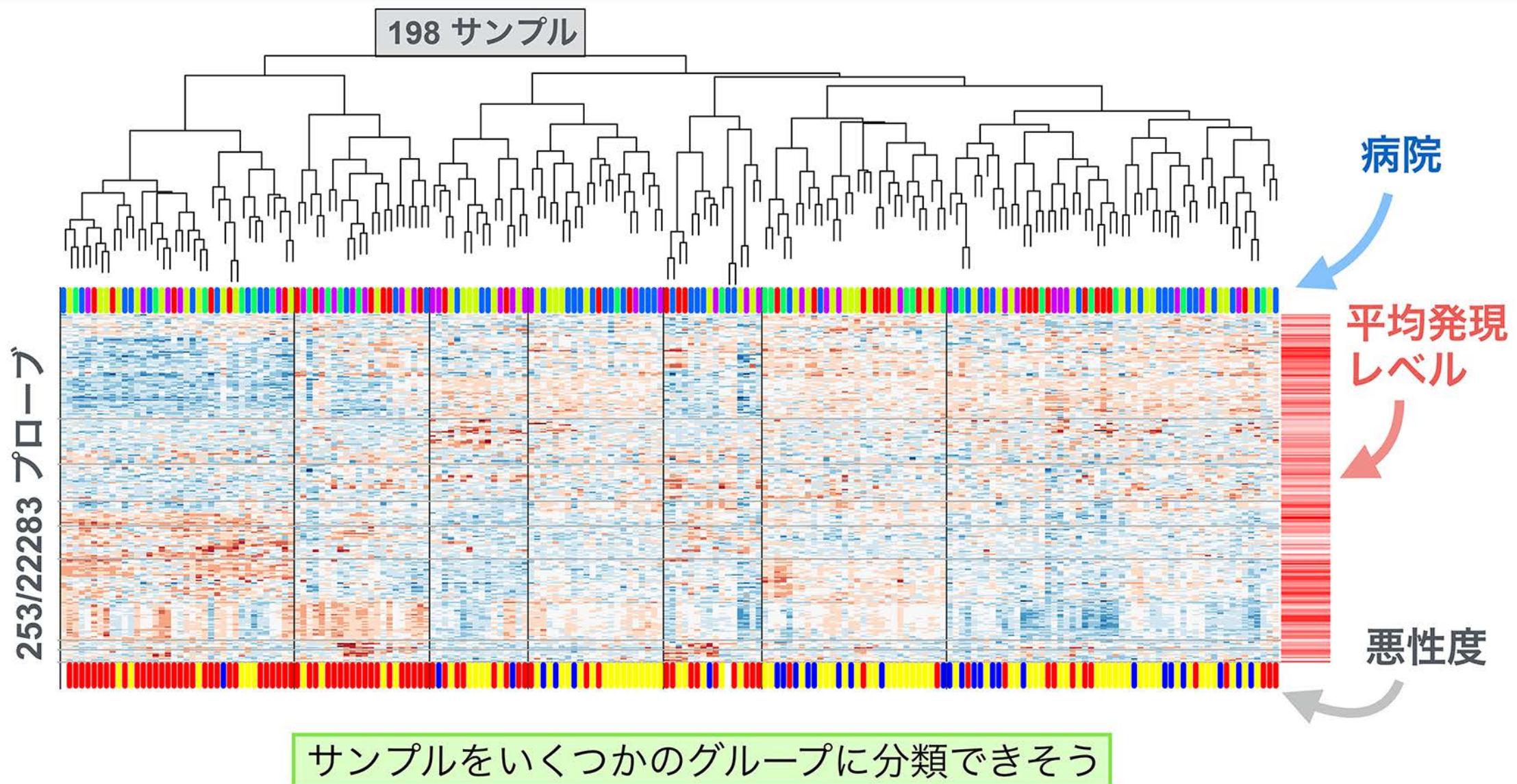
発現量依存の偏りを補正することはできる

# Quantile 補正

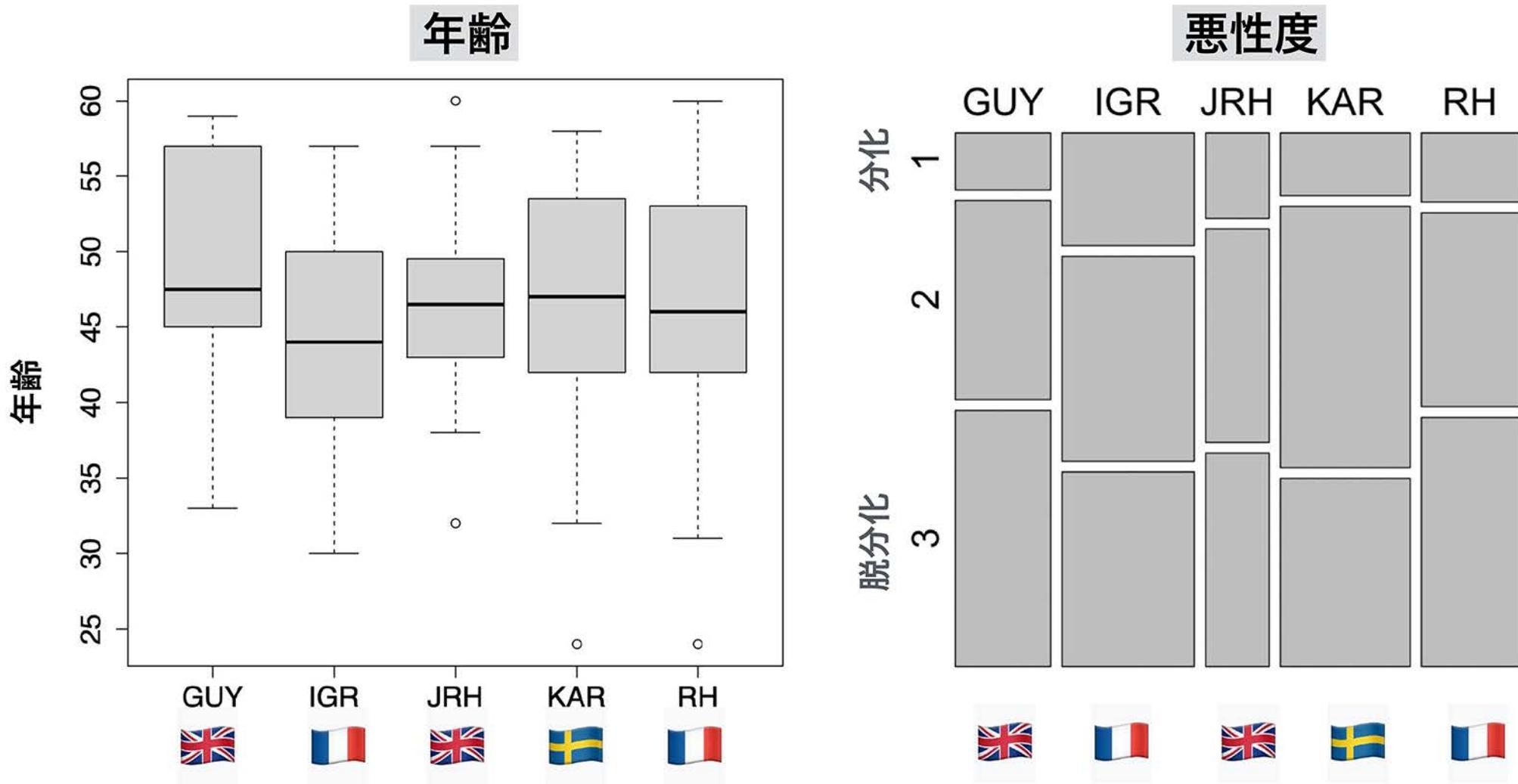


全てのサンプルが、本来は同一の分布をしていると仮定

### (3) サンプル全体の様子 (ヒートマップ)



# 病院の違いを補正して良いか？



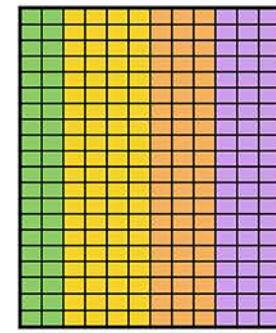
判断には専門知識が必要

# 補正

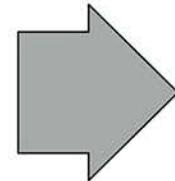
Batchの偏りを除去  
(ComBat)

分布を揃える  
(Quantile補正)

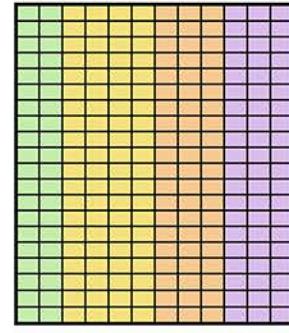
22283 Probes



198 Patients

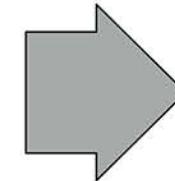


198 Patients

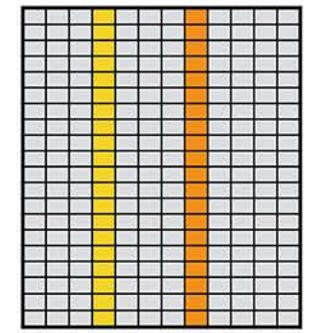


22283 Probes

バッチ固有の傾向は  
系統誤差、偶然誤差



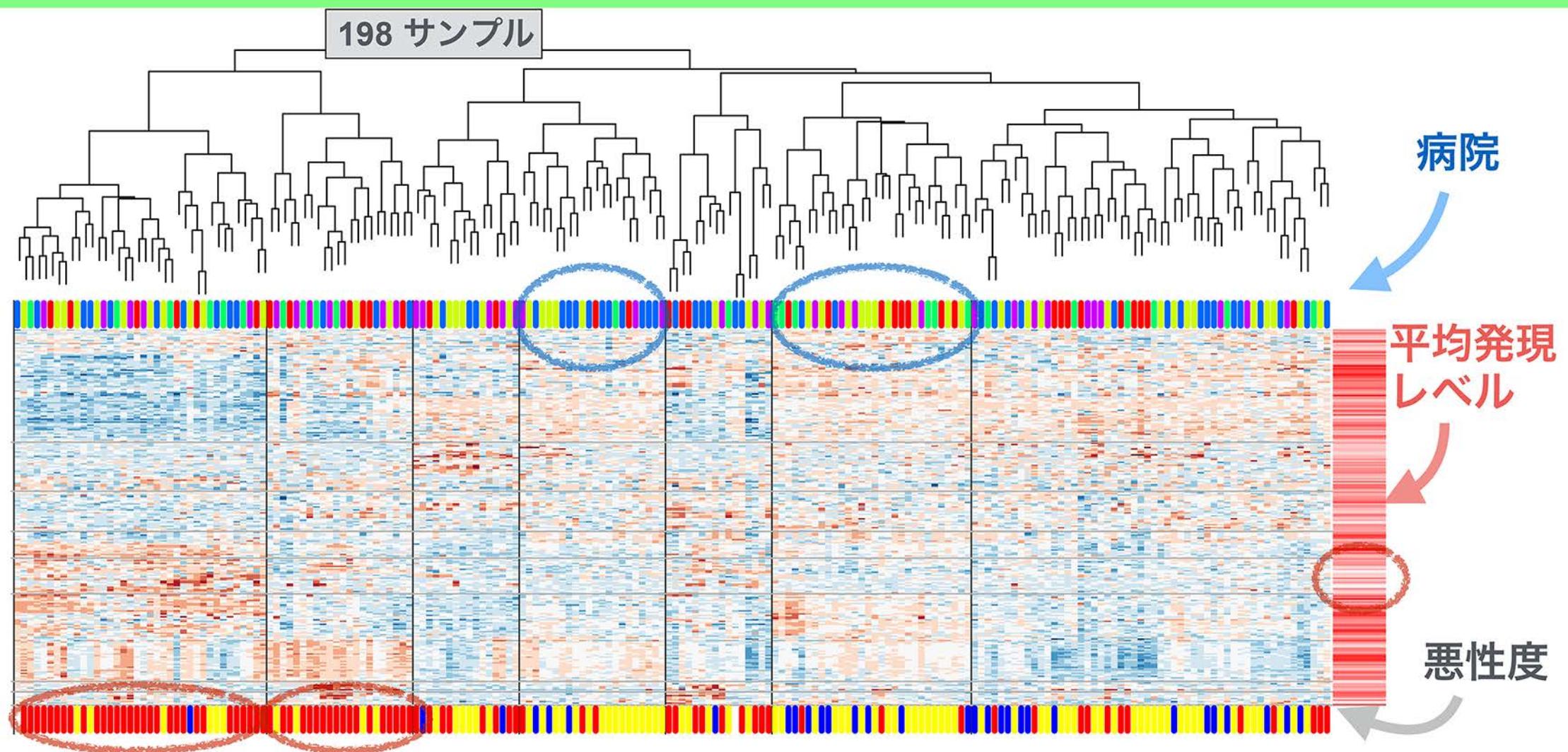
198 Patients



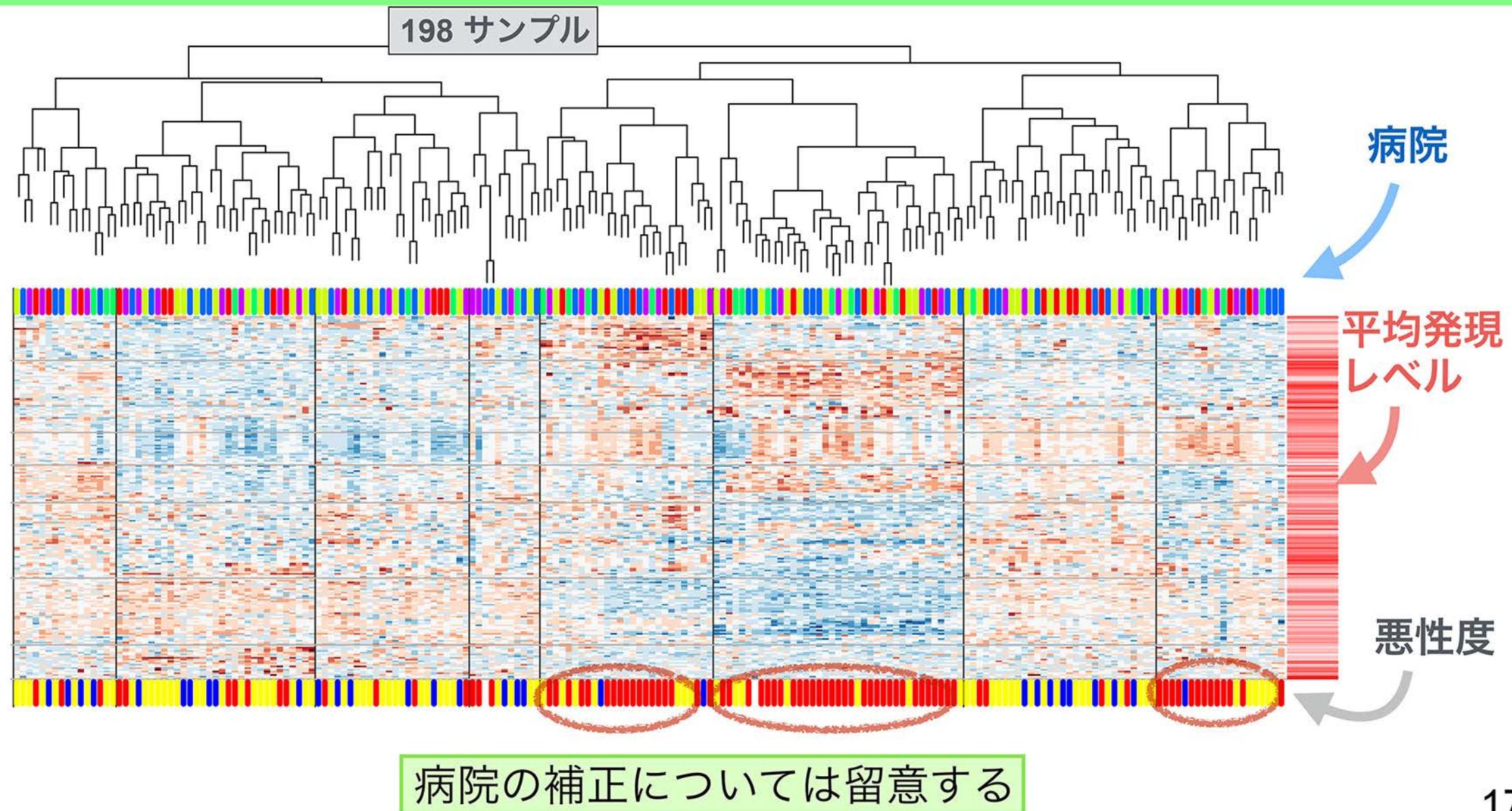
22283 Probes

全サンプルの  
分布が同一

# 生データ



# 補正後（おそらく妥当）



# COXPRESdb



last update: 2020.06.28

## COXPRESdb

provides co-regulated gene relationships to estimate gene functions.

All words



Search

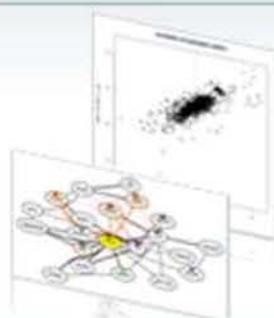
### Search

- GeneTable
- EdgeAnnotation
- CoExSearch
- etc.



### Draw

- NetworkDrawer
- HCluster
- CoexViewer
- etc.



### Browse

### Bulk download

## Information

- » Overview
- » Tutorial Movie
- » API
- » Publication
- » Help

## What's new

2020.06.27

→ The API is upgraded, allowing a multiple gene query. Uniprot ID query is acceptable as well as Entrez Gene ID [API].

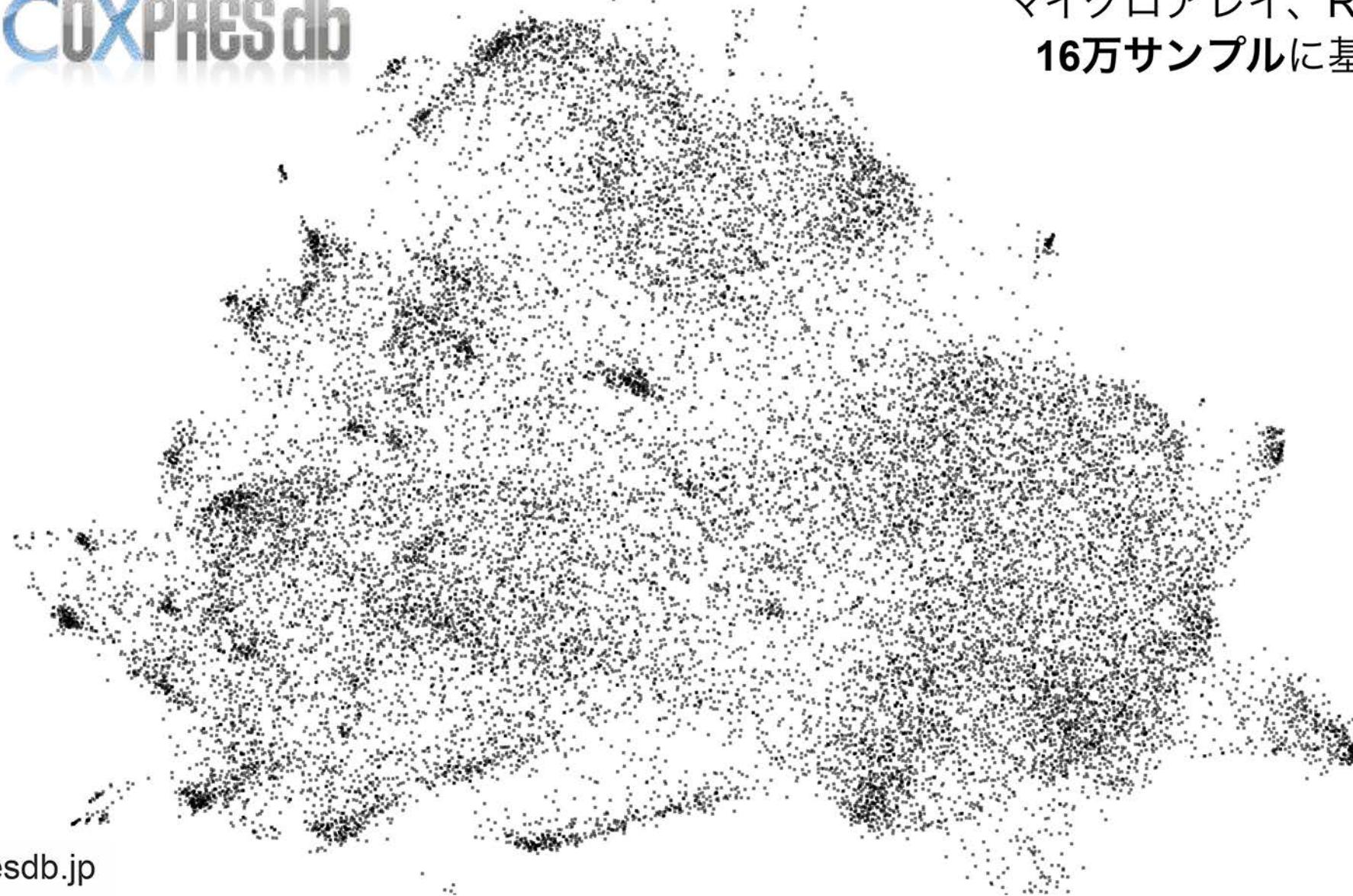
2019.07.01

→ COXPRESdb ver. 7.3 is released. This

# ヒトの遺伝子共発現ネットワーク



マイクロアレイ、RNAseq  
16万サンプルに基づく

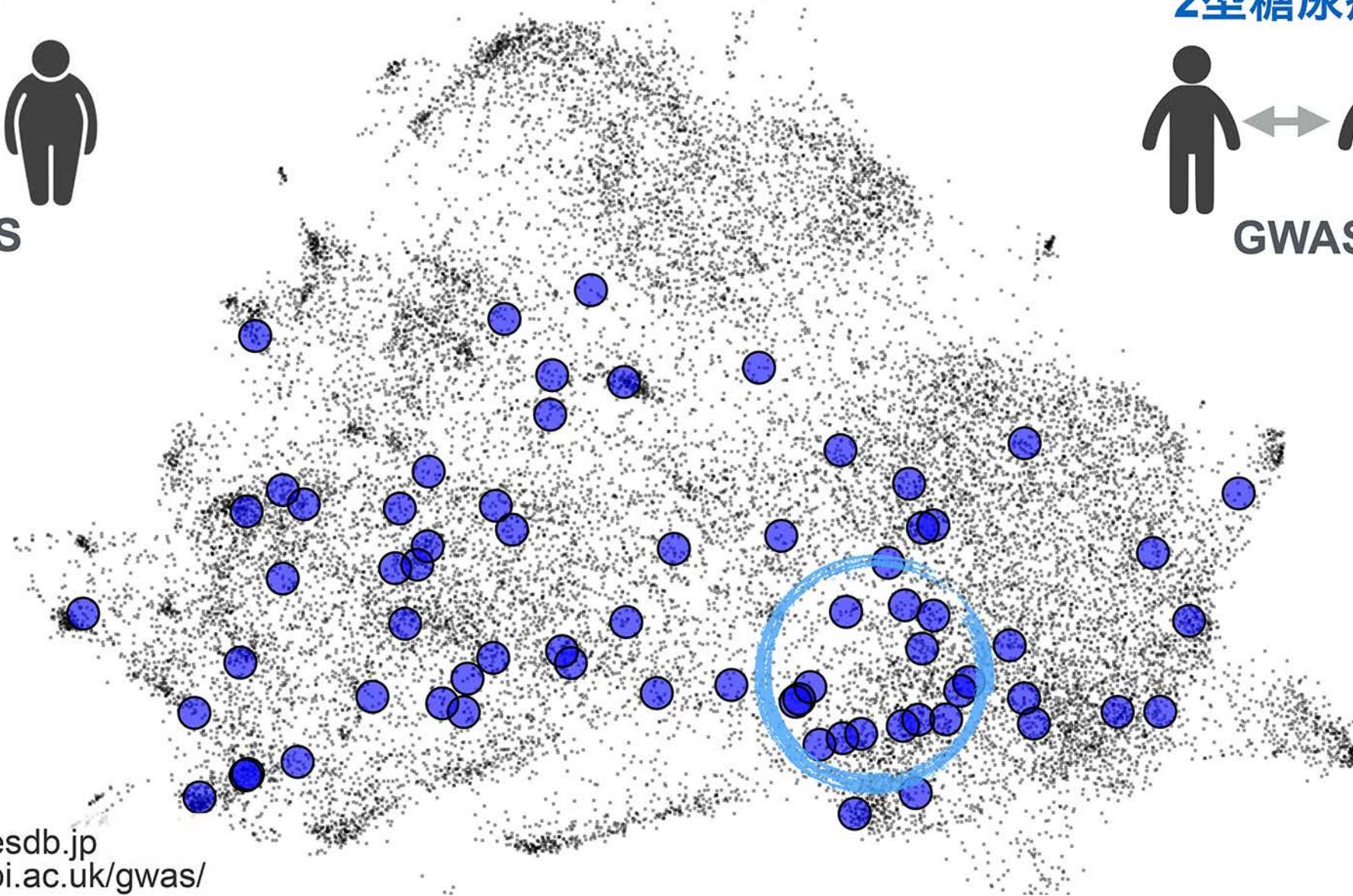
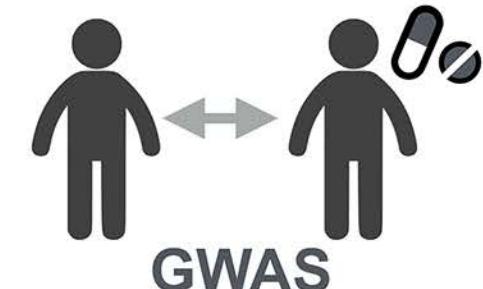


# (例) 糖尿病の遺伝子

BMI

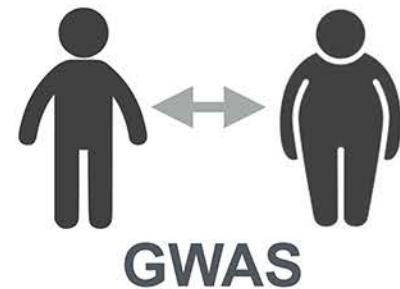


2型糖尿病

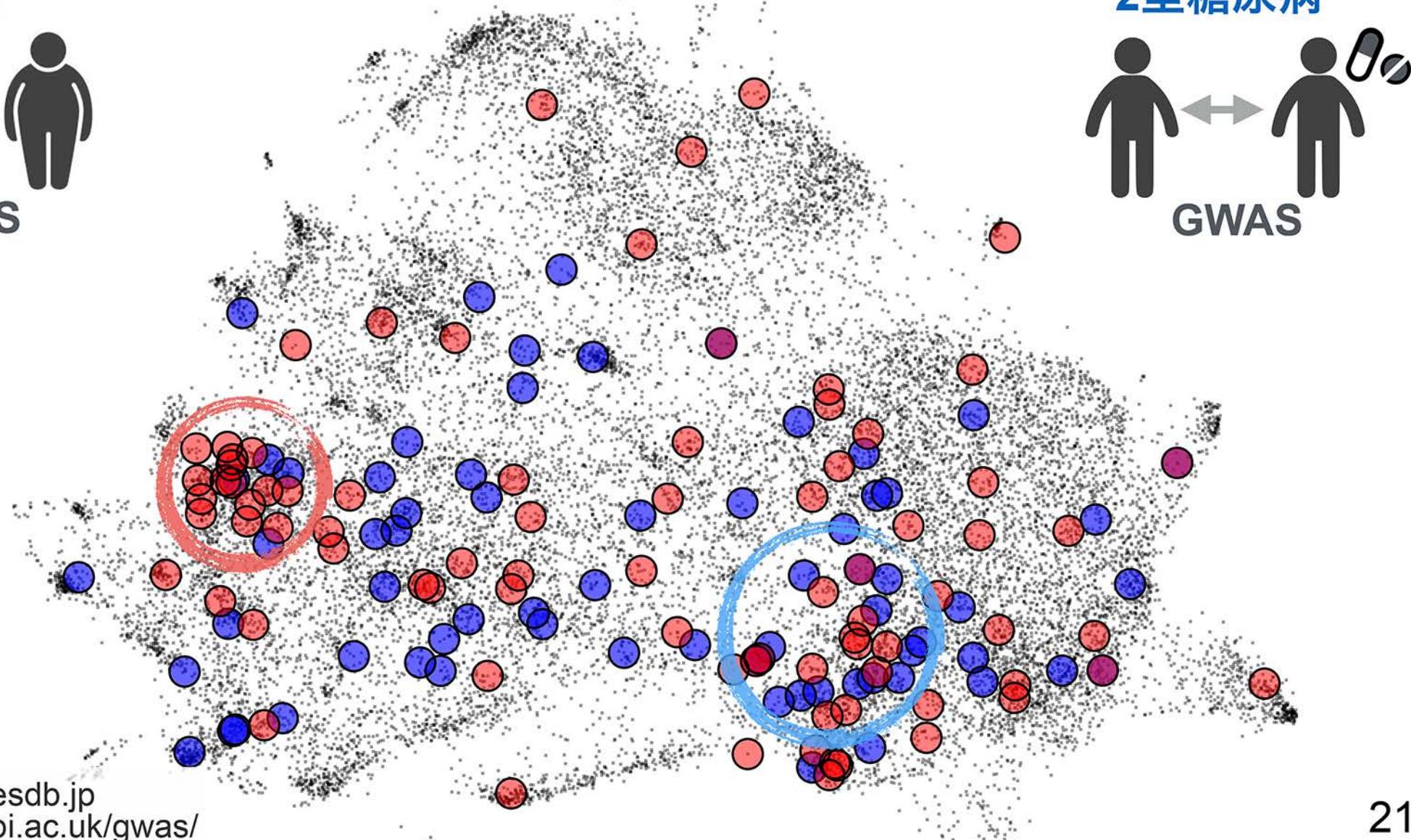
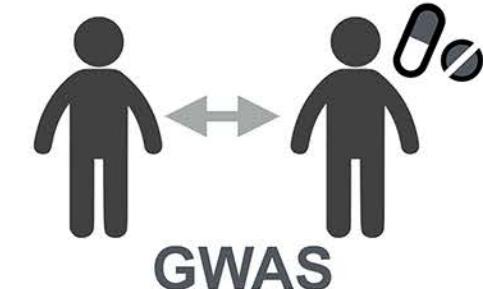


# (例) 糖尿病の遺伝子

BMI



2型糖尿病



# まとめ

## 補正の考え方

- ・補正前のデータを理解し、そのデータの有効性を想像する
- ・補正により真の情報も失われる（偽の情報が生まれる）

## 補正の具体的なポイント

- ・等価情報を考える（平均を引いたら、平均値も参照する）
- ・強力な補正（仮定の強い補正）はできれば後回し
- ・測定に付随する情報（今回は病院）は補正に有用

## その他

- ・生物学的知識はできるだけ後に使う
- ・下流解析の一つとして、共発現データベースもご活用ください

# デモ解析の補正で用いたR command

## Quantile補正

- preprocessCoreライブラリ normalize.quantiles

## バッチ補正 (病院の補正)

- svaライブラリ (BioConductor) ComBat