

生命情報ビッグデータからの 発見技術

工学院大学 工学部 電気電子工学科
教授 福岡 豊

従来技術とその問題点(1/5)

DNAマイクロアレイなどの技術が発展し、生命情報のデータが大規模化している。近年では、異種データの統合的な解析法も提案されている。

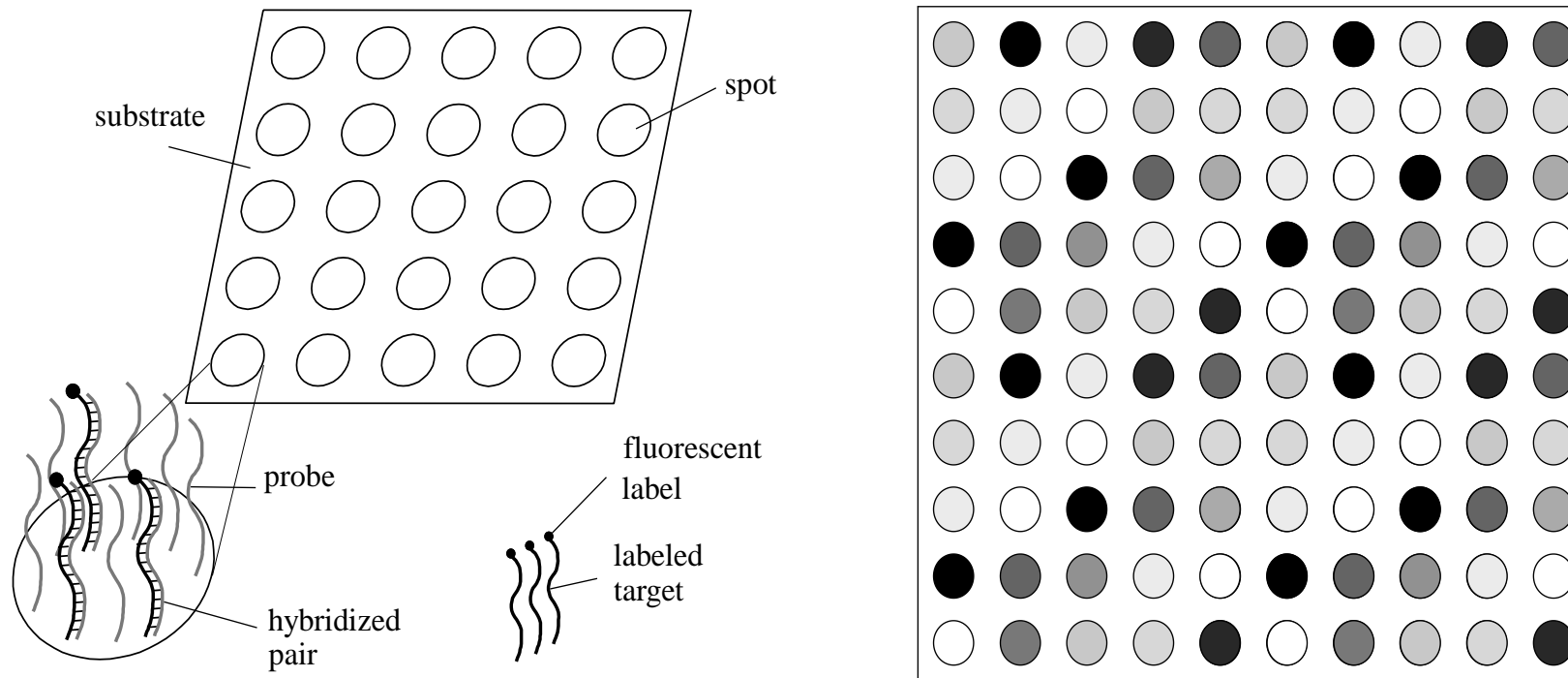
生命情報ビッグデータ

パラメータ(遺伝子など)数に比べて、測定条件数が少ないというデータ構造のため、数理的な処理法によって有用な知見を発見することが困難である。

従来技術とその問題点(2/5)

DNAマイクロアレイ

数千～万遺伝子の発現を調べる技術



従来技術とその問題点 (3/5)

マイクロアレイのデータ形式

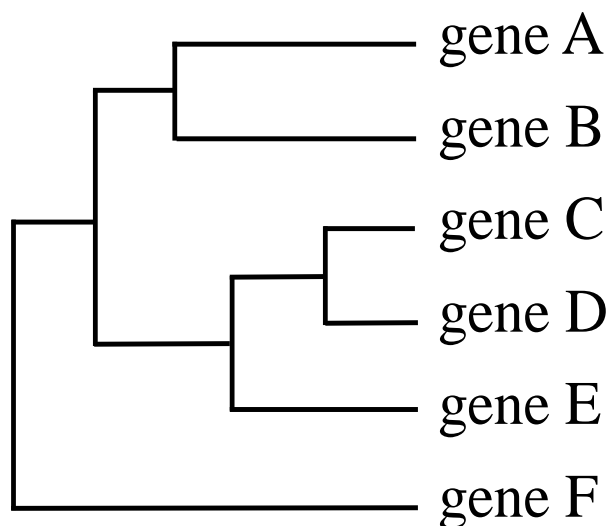
	測定条件1	測定条件2	測定条件3	⋯⋯	測定条件L
遺伝子1	$X(1, 1)$	$X(1, 2)$	$X(1, 3)$	⋯	$X(1, L)$
遺伝子2	$X(2, 1)$	$X(2, 2)$	$X(2, 3)$	⋯	$X(2, L)$
遺伝子3	$X(3, 1)$	$X(3, 2)$	$X(3, 3)$	⋯	$X(3, L)$
⋯	⋯	⋯	⋯	⋯	⋯
遺伝子 <i>i</i>	$X(i, 1)$	$X(i, 2)$	$X(i, 3)$	⋯	$X(i, L)$
⋯	⋯	⋯	⋯	⋯	⋯
遺伝子 <i>N</i>	$X(N, 1)$	$X(N, 2)$	$X(N, 3)$	⋯	$X(N, L)$

N : 総遺伝子数、 L : 測定条件数 $N \gg L$

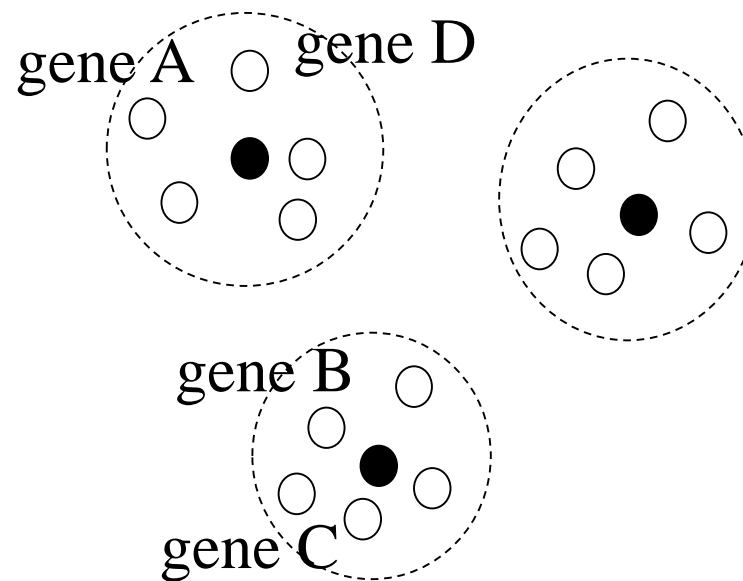
従来技術とその問題点(4/5)

初期のデータ処理法

階層的クラスタリング法



Self Organizing Map法 (非階層的クラスタリング)

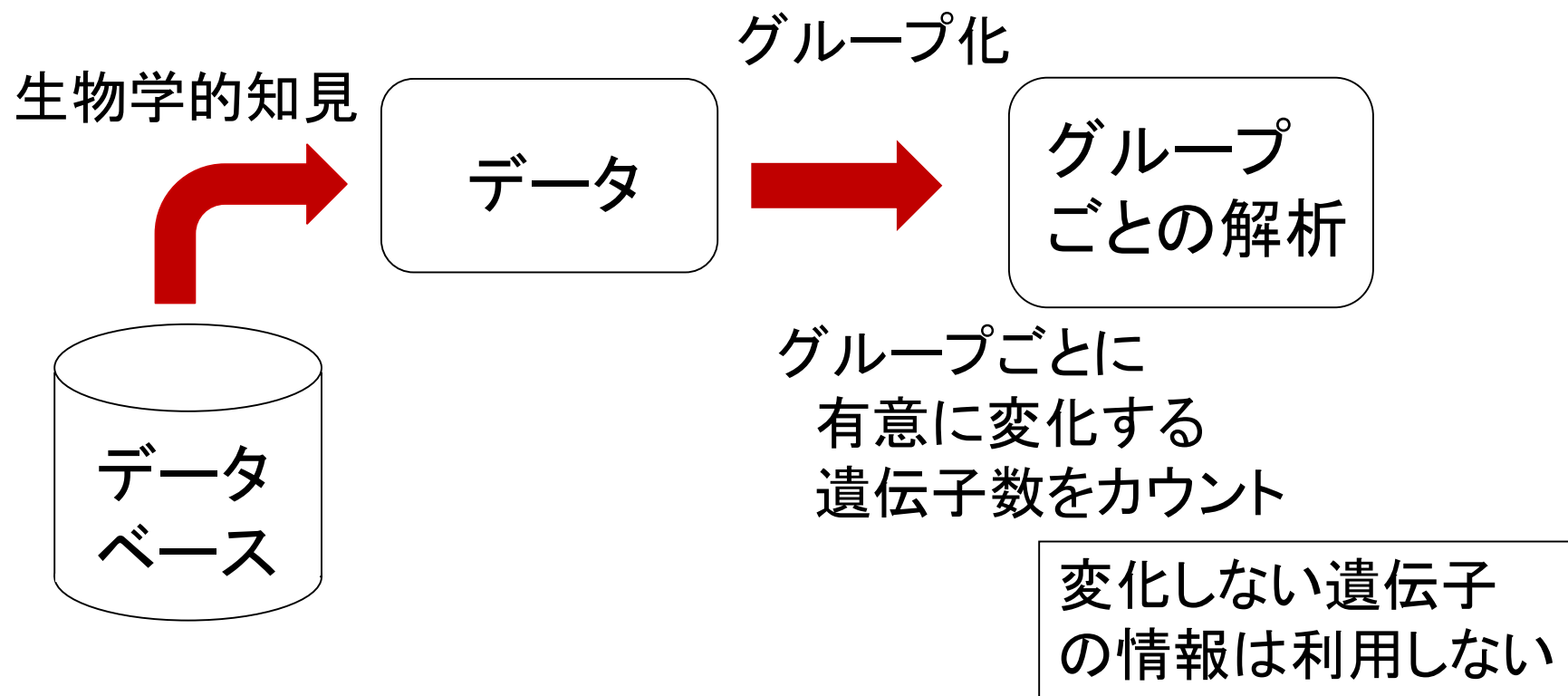


類似度で分類 → 遺伝子数 ≧ 患者数のため、
結果が**ノイズの影響を受けやすい**

従来技術とその問題点 (5/5)

生物学的知見を採り入れたグループ化

Gene Set Enrichment Analysis法 (GSEA法)

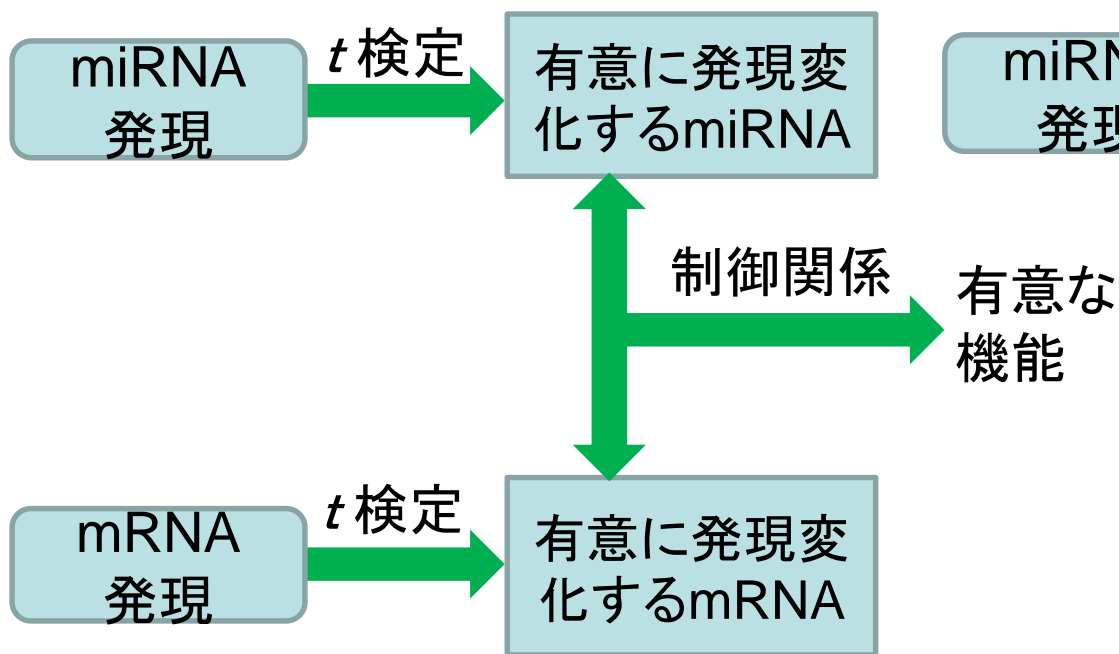


新技術の特徴・従来技術との比較(1/2)

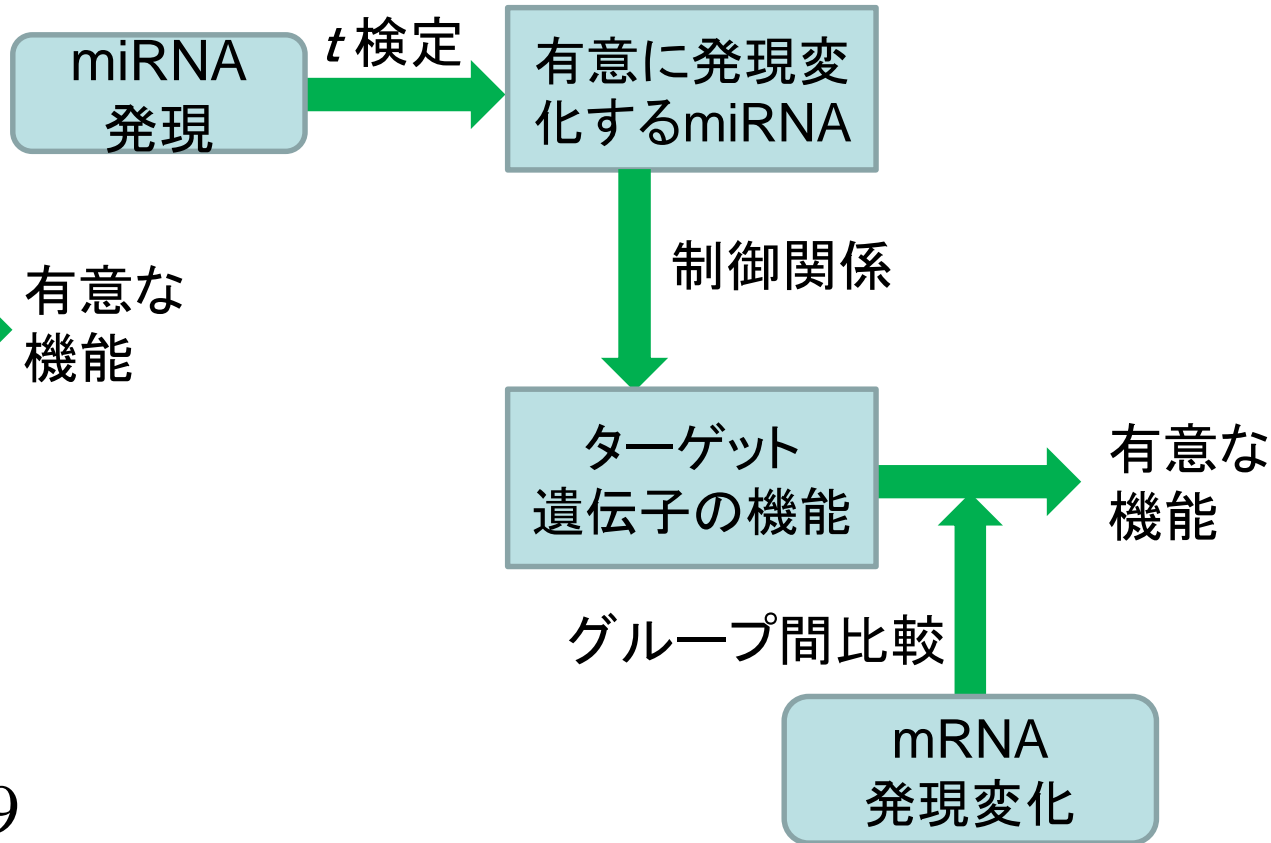
- 本技術は、このようなビッグデータを
生物学的な知見に基づきグループ化して、
統計的に比較するものである。
- この方法によって、
測定条件(患者など)数が少ない場合でも、
有用な知見を発見できる。

新技術の特徴・従来技術との比較 (2/2)

従来の方法



提案法



Seungyoonら、NAR、2009

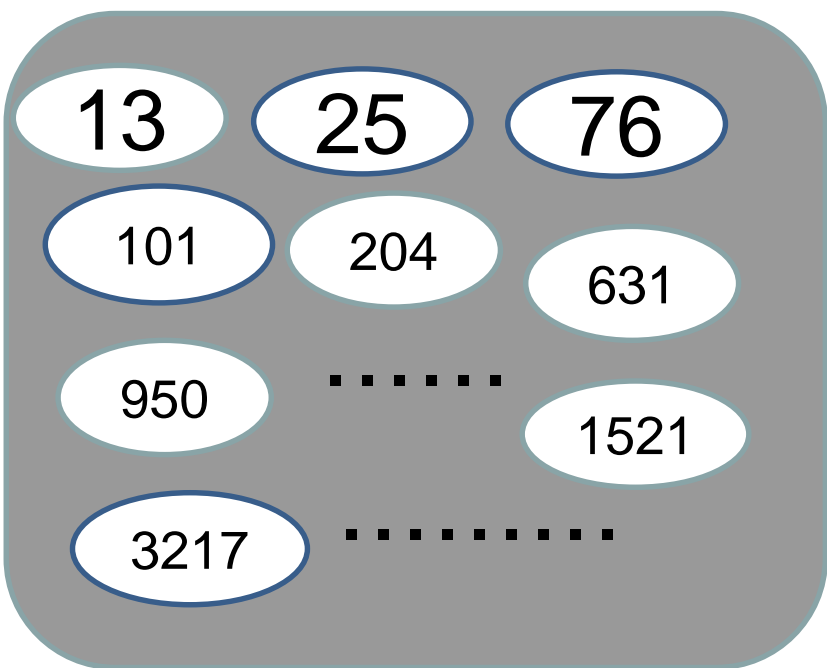
新技術の特徴(1/4)

- 本技術では、データ構造を考慮して、
これまで捨てていたデータもグループに取り
込むことで、精度を高め、適用範囲を広げた。

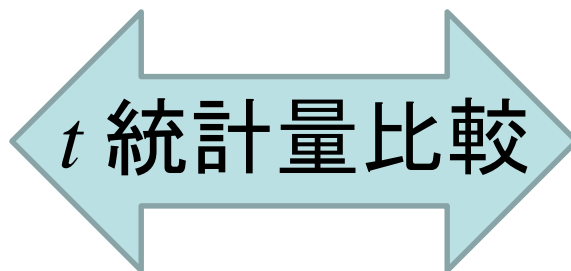
新技術の特徴(2/4)

- 機能やパスウェイごとのグループで比較
ウイルコクソンの順位和検定

対象群



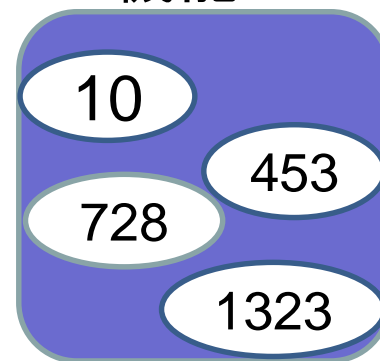
ランダムに1000遺伝子を選択



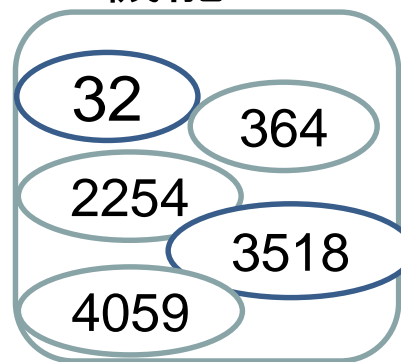
大きく変化する
機能を検出

ターゲット遺伝子群

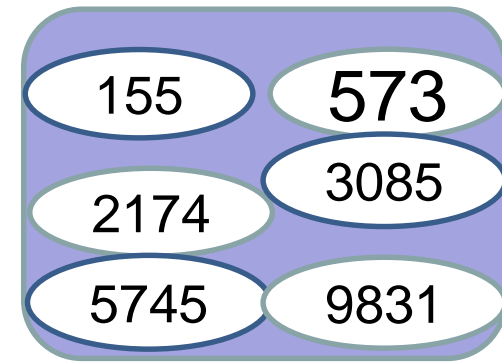
機能A



機能B



機能C



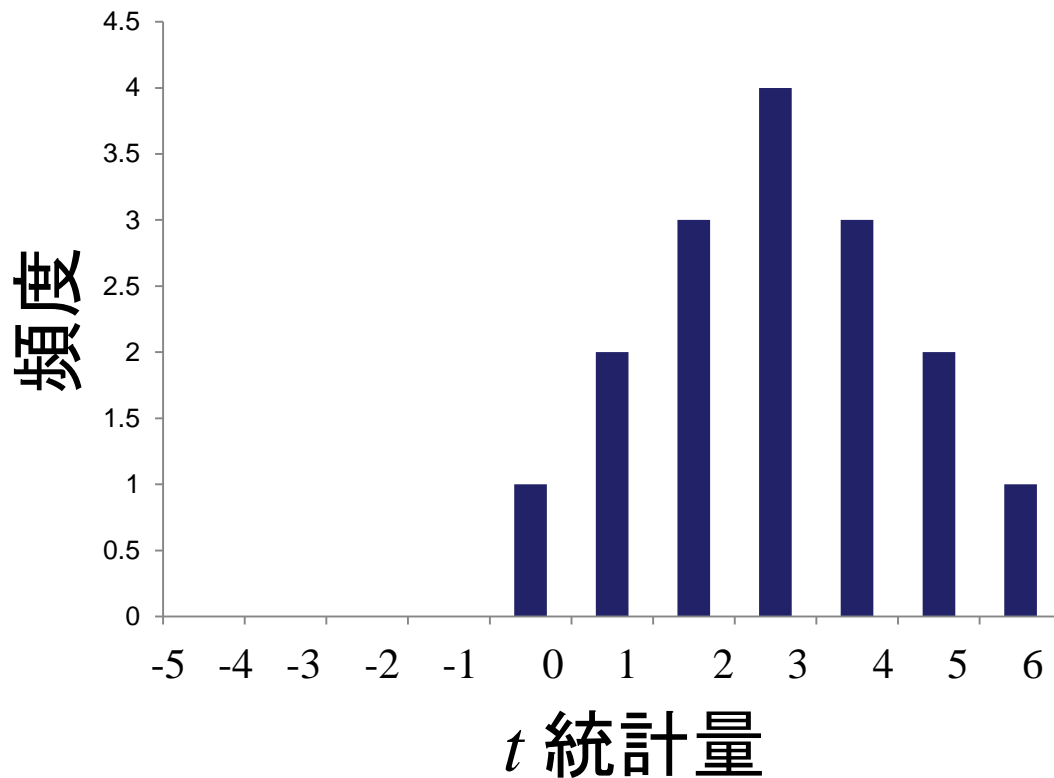
新技術の特徴(3/4)

- 先見的知識を活用したビッグデータ解析
- 異種データの統合的解析
- 測定条件数が少ない場合(培養細胞を用いた研究等)でも適用可能

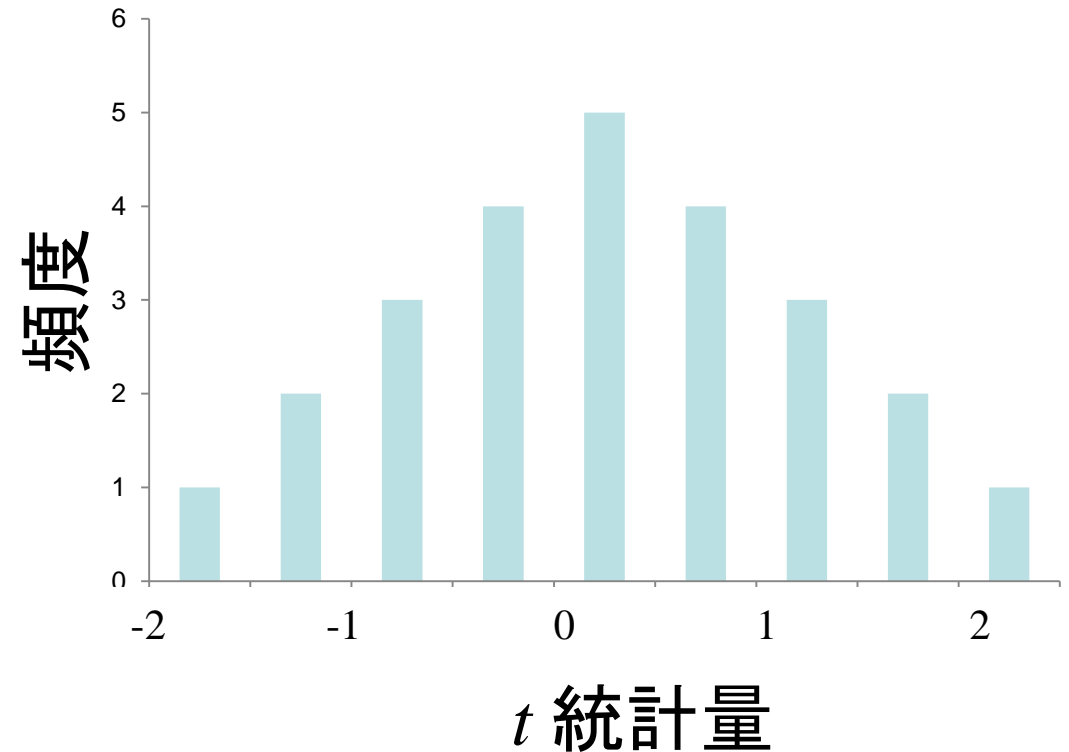
新技術の特徴(4/4)

- t 統計量の分布を比較

ターゲット遺伝子



対象群



新技術の適用例(1/3)

miRNAに対する t 検定結果

有意差がある3個のmiRNA

miRNA	p 値
miR-A	0.021
miR-B	0.033
miR-C	0.034

3つの内2つでは有意な遺伝子機能は得られなかった。

新技術の適用例(2/3)

有意に多くのmiRNA-Aのターゲット遺伝子を含む機能

遺伝子機能群1(転写に関する機能)				
カテゴリ	遺伝子機能	個数	p値	Gene ID
GO_BP	transcription, DNA-templated	240	4.63E-04	10346, 147687, 219749, 6774, 7753, ...
GO_MF	DNA binding	208	5.65E-04	3725, 7494, 147687, 219749, 6774, ...
GO_BP	regulation of transcription, DNA-templated	170	5.97E-02	2146, 4302, 10346, 219749, 147687, ...
遺伝子機能群2(リン酸に関する機能)				
カテゴリ	遺伝子機能	個数	p値	Gene ID
GO_MF	ATP binding	192	1.74E-04	84930, 79705, 54617, 80216, 29028, ...
GO_MF	protein serine/threonine kinase activity	60	3.23E-04	2580, 84930, 79705, 57538, 80216, ...
GO_BP	protein phosphorylation	69	5.63E-04	2580, 79705, 57538, 332, 8573, ...
GO_MF	protein kinase activity	52	6.54E-03	2580, 79705, 8573, 6352, 50488, ...

新技術の適用例(3/3)

miR-Aによって有意に発現変化した遺伝子機能

カテゴリ	機能	wilcoxon_p	遺伝子数
GOTERM_MF_DIRECT	GO_0005524_ATP binding	2.95E-06	192
GOTERM_BP_DIRECT	GO_0075733_intracellular transport of virus	2.13E-03	12
GOTERM_MF_DIRECT	GO_0004672_protein kinase activity	4.50E-03	52
GOTERM_BP_DIRECT	GO_0030522_intracellular receptor signaling pathway	5.07E-03	10
GOTERM_MF_DIRECT	GO_0098641_cadherin binding involved in cell_cell adhesion	1.23E-02	49
KEGG_PATHWAY	hsa04012_ErbB signaling pathway	2.37E-02	15
KEGG_PATHWAY	hsa04917_Prolactin signaling pathway	2.55E-02	16
GOTERM_BP_DIRECT	GO_0006409_tRNA export from nucleus	3.06E-02	8
GOTERM_BP_DIRECT	GO_0010827_regulation of glucose transport	3.06E-02	8
GOTERM_BP_DIRECT	GO_0098609_cell_cell adhesion	3.27E-02	48

想定される用途

- がんをはじめとする疾患のバイオマーカー探索
- 医療・健康分野のビッグデータに基づく要因分析
- ビッグデータのマイニング

実用化に向けた課題

- 現在、培養細胞から得られたデータに適用した。しかし、有意な結果が得られるmiRNAの数が少ない。
- 今後、データベースを更新して、最新の知見を用いた処理を行う。
- 実用化に向けて、他のデータセットに適用し、問題点の洗い出しを行う。

企業への期待

- 有意な結果が得られるmiRNAが少ない点については、データベースの更新などによって、有用性を示せると考えている。
- 未解析のビッグデータ(医療分野に限らず)を持つ、企業との共同研究を希望。
- また、発見技術を開発中の企業、データサイエンス分野への展開を考えている企業には、本技術の導入が有効と思われる。

他機関との連携の経歴

産学連携

- 2016年- (株)フィットラボ社と共同研究実施
電位治療器が自律神経系に及ぼす影響の解析

大学間連携

- 2013年- 東京電機大学
容量結合性電極による生体センシング技術の開発
- 2015年- 東京医科大学
生命情報ビッグデータの統合解析技術の開発
- 2016年- 東京医科歯科大学
ビデオ眼振計を用いた眼位平面の解析技術の開発

本研究は工学院大学と東京医科大学の医工連携プロジェクトの一環として行われています。

共同研究者：東京医科大学 医学総合研究所
分子腫瘍研究部門 大屋敷純子 教授、梅津知宏 講師
医学部 分子病理学分野 黒田雅彦 教授



お問い合わせ先

工学院大学 研究戦略部 研究推進課

担当者：西原・山岸

Tel : 042-628-4940

kenkyu@sc.kogakuin.ac.jp