

# 情報理論で金持ちになる



緒方法親 博士(農学)<sup>1,2,3</sup>

<sup>1</sup> 株式会社日本バイオデータ 代表取締役

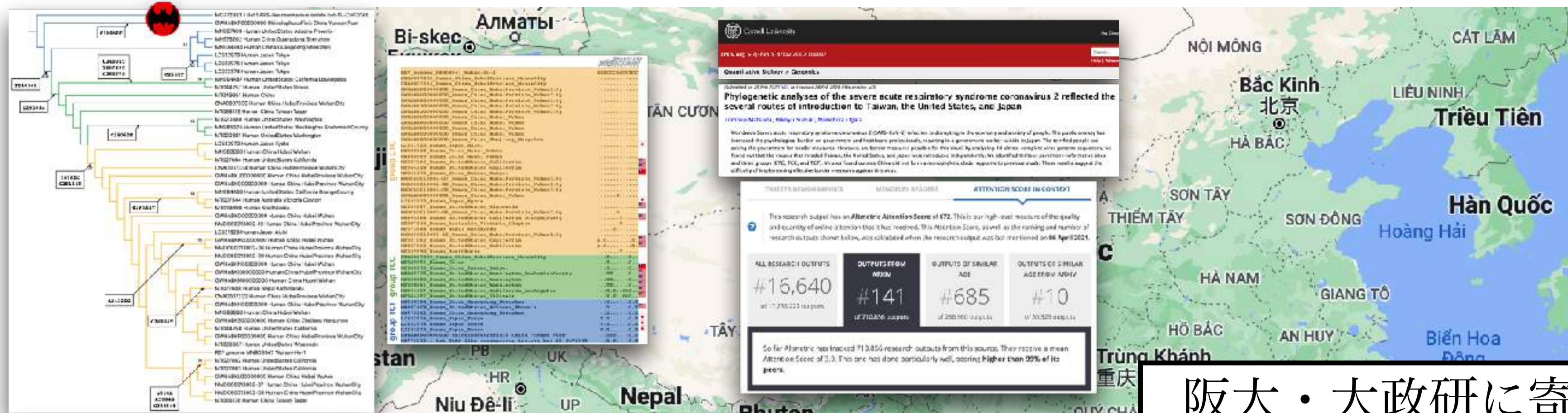
<sup>2</sup> 大阪大学大学院工学研究科生物工学専攻 招へい教授

<sup>3</sup> 次世代バイオ医薬品製造技術研究組合 顧問 (ゲノム技術)

# 緒方と主な会社について

新型コロナウイルスの変異株を世界で最初に発見&報告

設立	2013年
資本金	200千円
従業員	4人
論文数	32



阪大・大政研に寄生

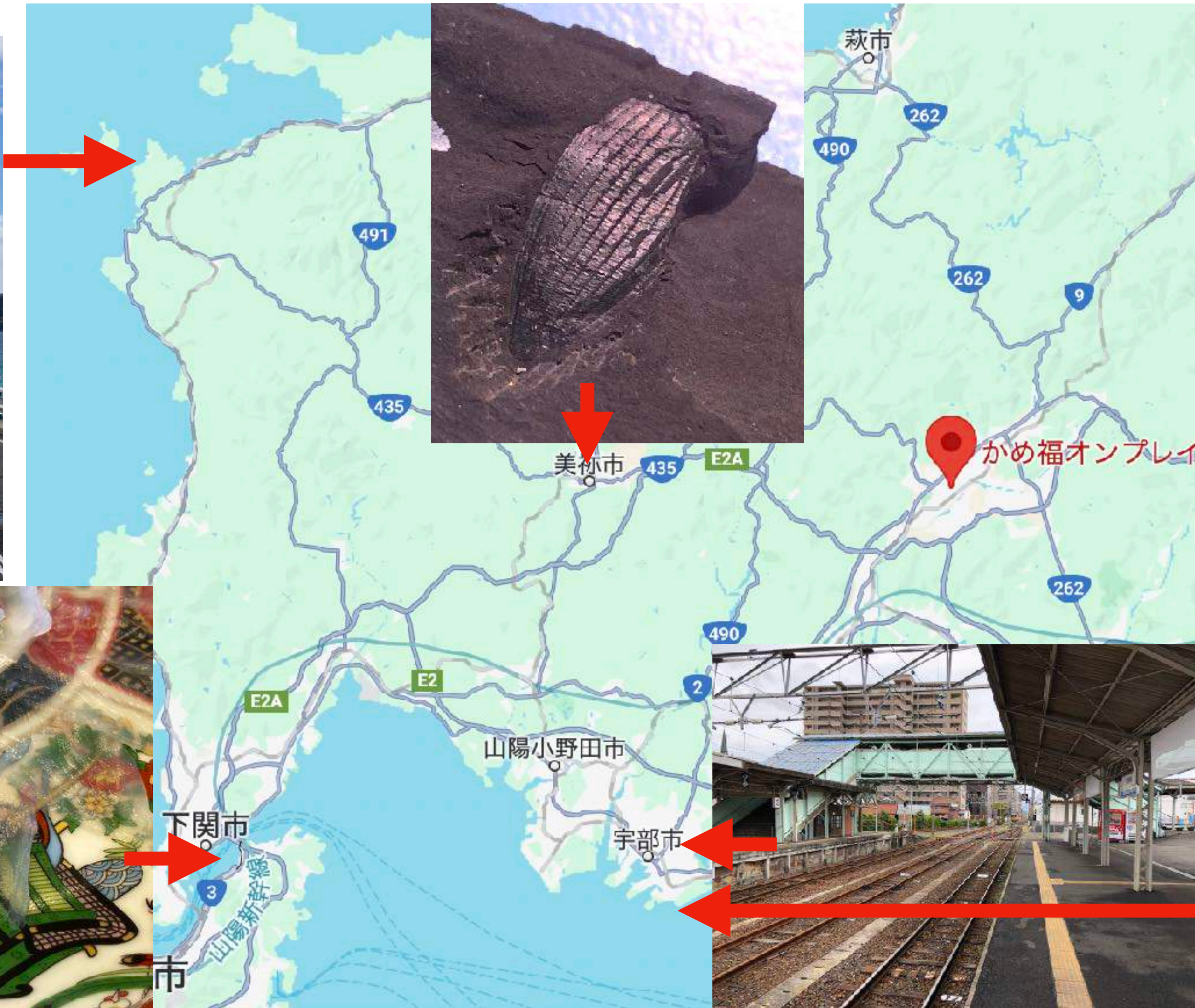


Work Space (川崎)



Nishikigoi Farm

# 山口県とゲノム科学と情報理論



# ポジティブリスト式免疫で防衛できるのか

## ポジティブリスト式

実施してよいことが規定されている。  
警察や自衛隊で採用されている。  
昆虫の免疫（攻撃してよい標的が定義されている）

## ネガティブリスト式

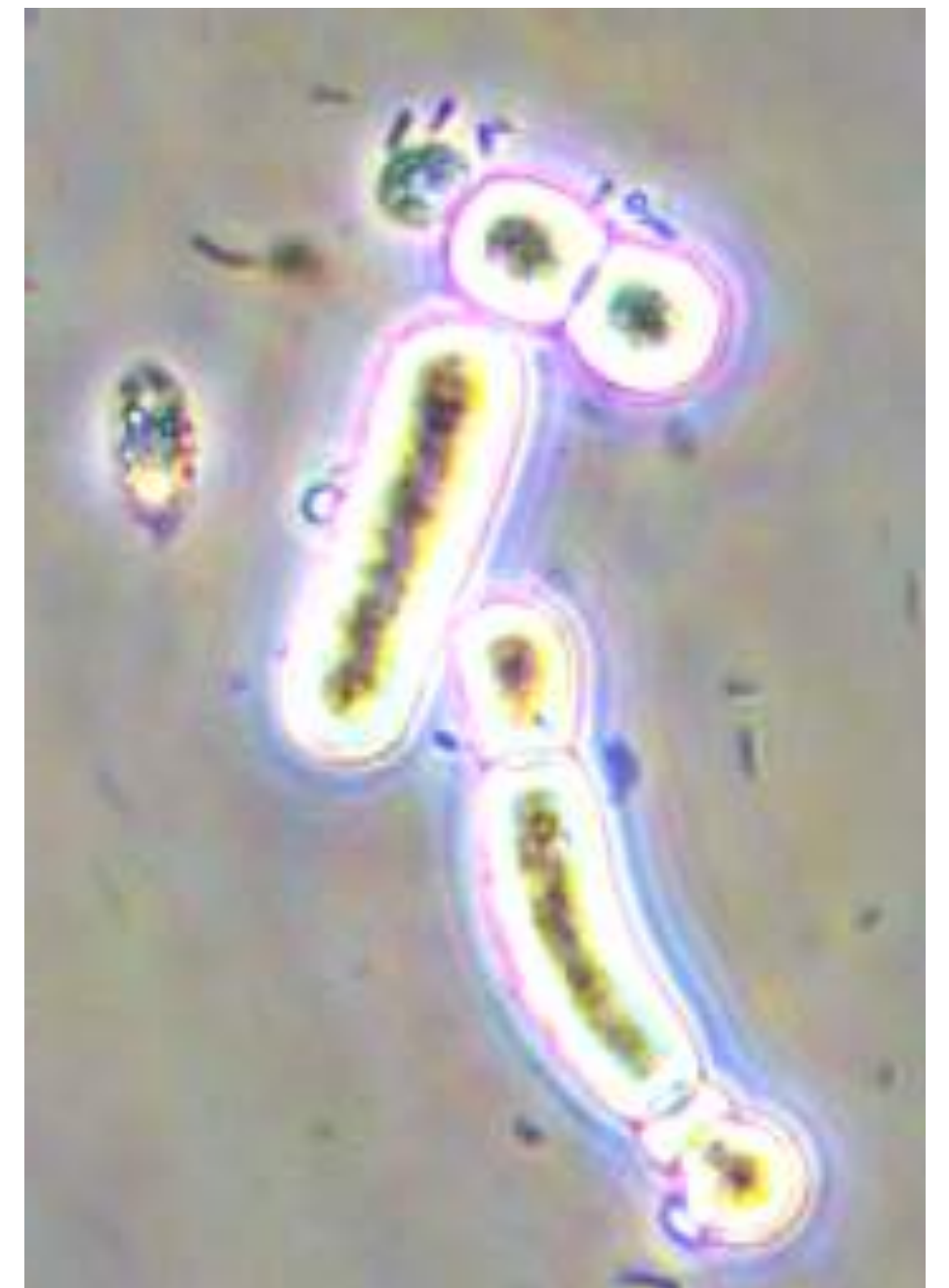
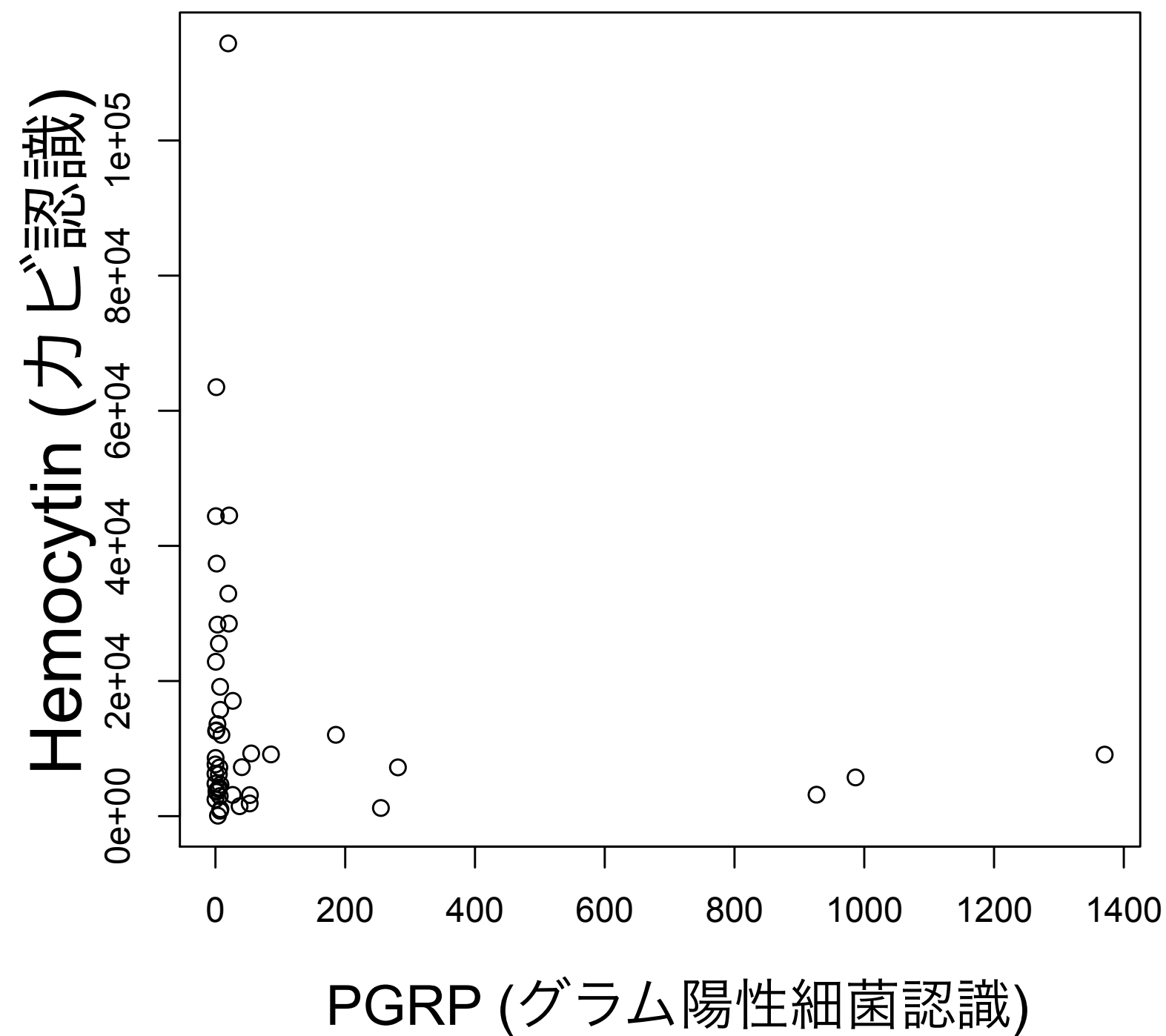
禁止事項が規定されている。  
軍隊で採用されている。  
ヒトの免疫（自分以外は全部攻撃する）

ポジティブリスト式の方が低コスト運用できるらしい。

“軍隊や自衛隊に対する法的権限の付与要領についての一考察 -公共財の限界費用価格形成原理に基づく権限付与モデル- 陸上自衛隊教育訓練研究本部 奥田昌弘

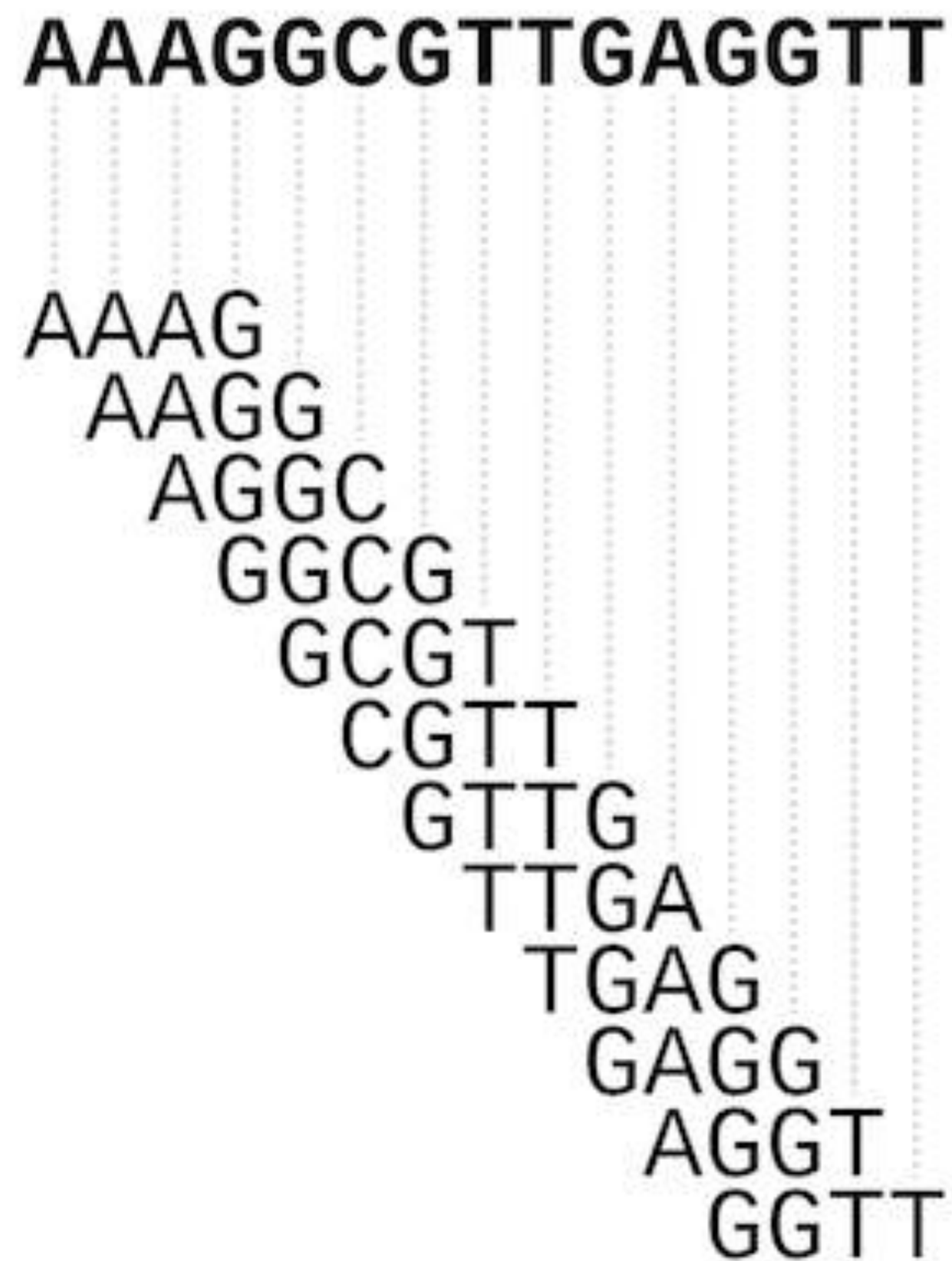
カブトムシ幼虫の血球を個々で見ると  
ヘモシチン遺伝子とPGRP遺伝子は  
同時に高発現にならない。

細菌を食べる血球と  
カビを食べる血球がいるが、  
両方食べる血球は少ない。  
培養して出てくるのはカビ好き。

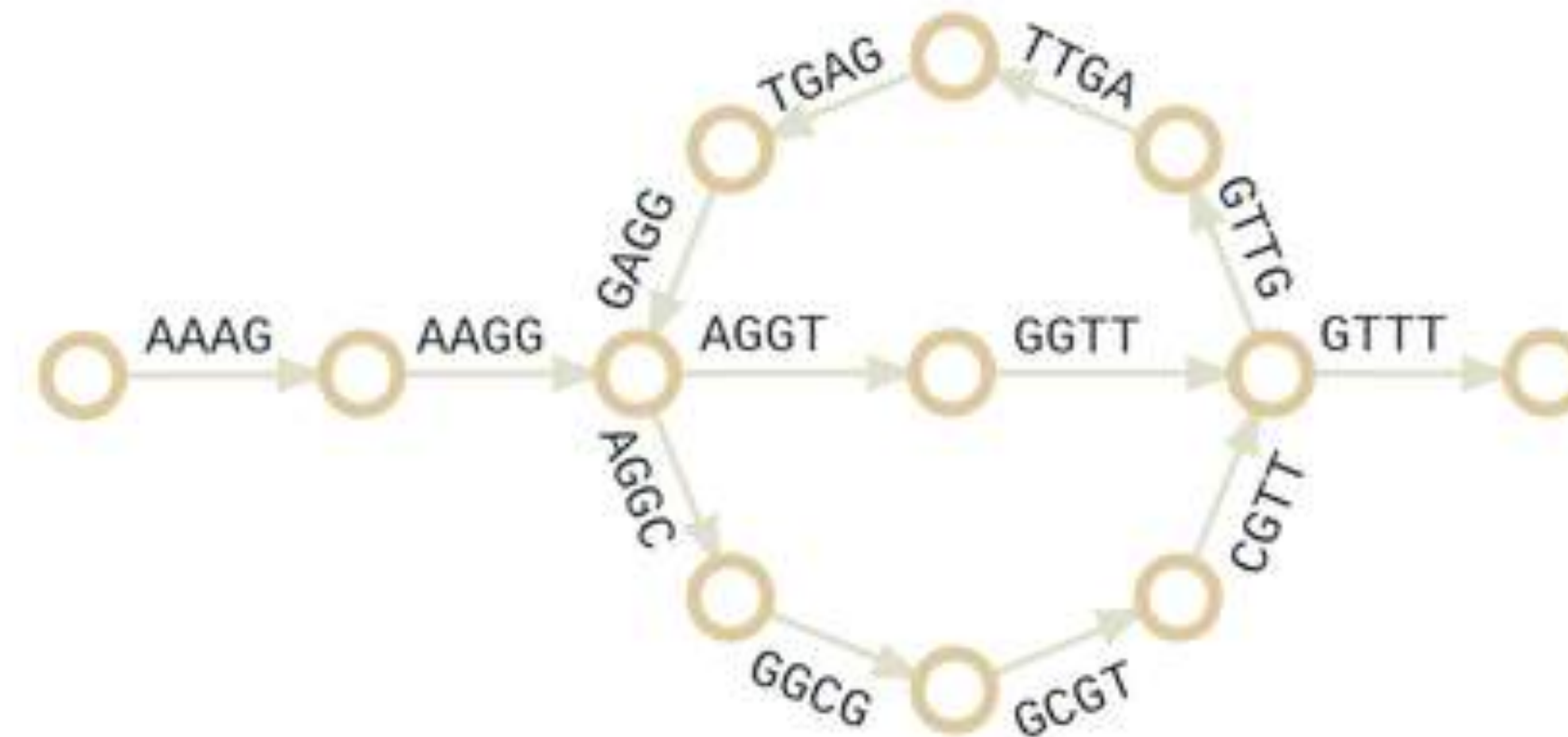


# de Bruin graph を使ってゲノム配列を復元

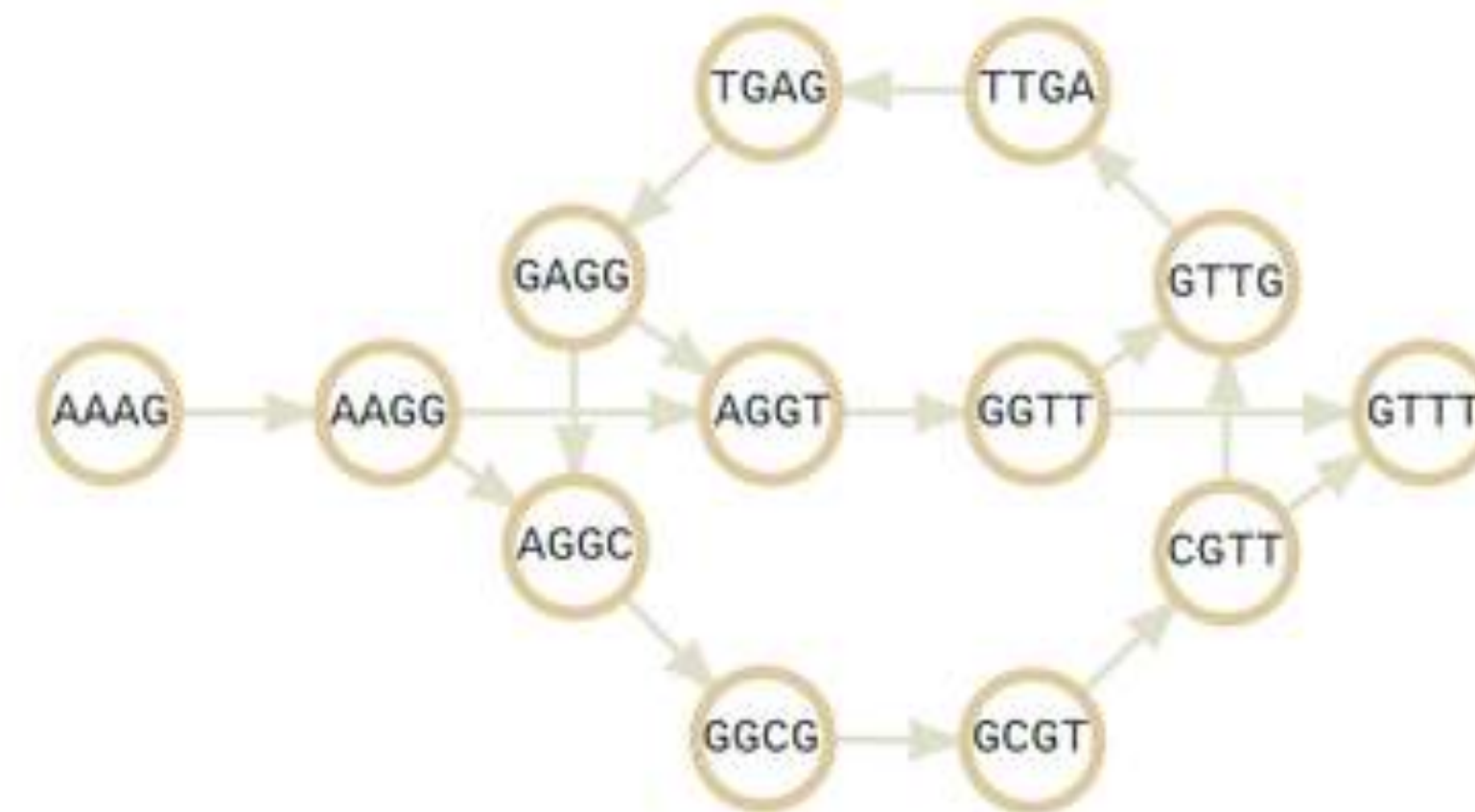
A. Short read to  $k$ -mers ( $k=4$ )



B. Eulerian de Bruijn graph



C. Hamiltonian de Bruijn graph



# ゲノムを何に使うのか

## Mendelian Randomization

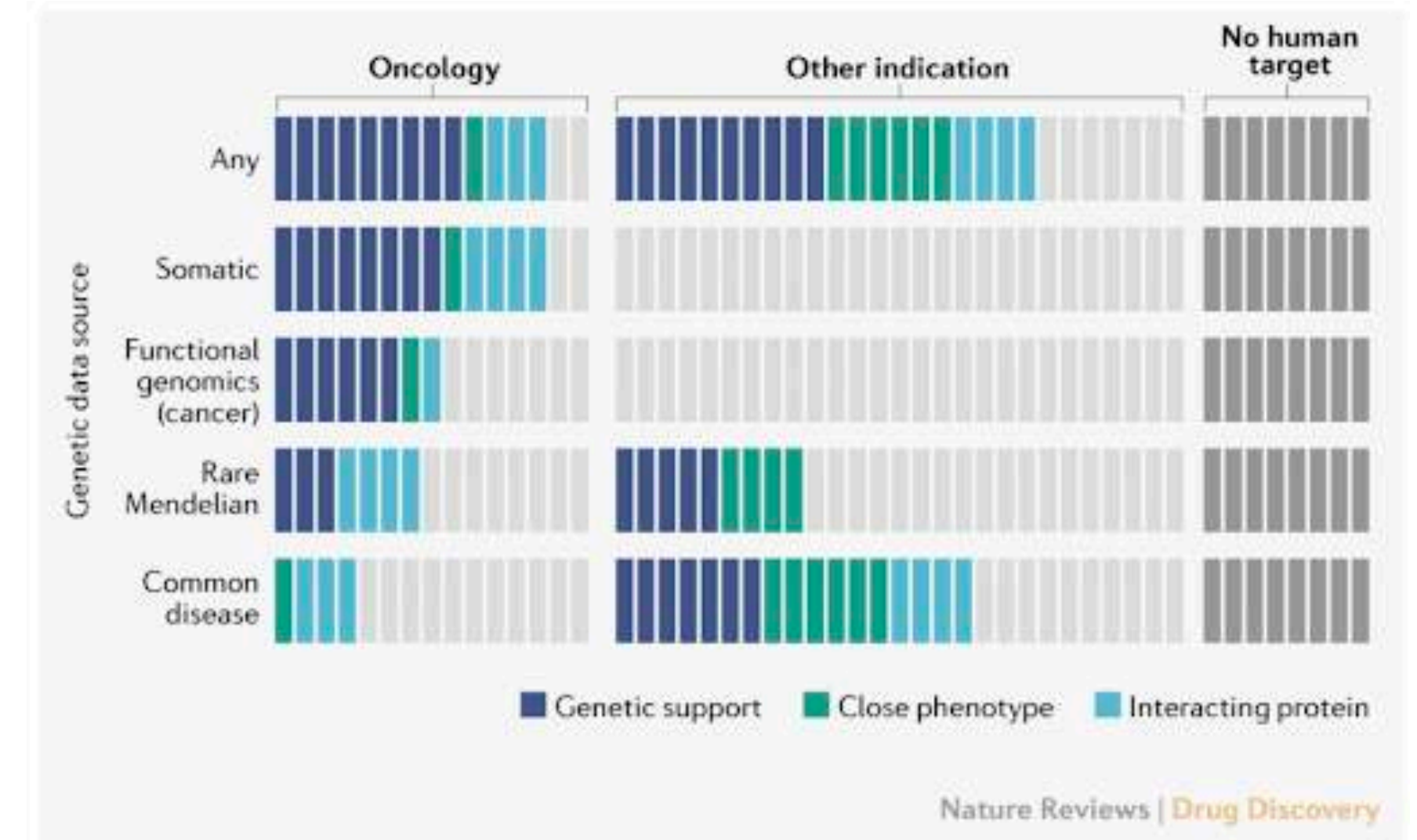
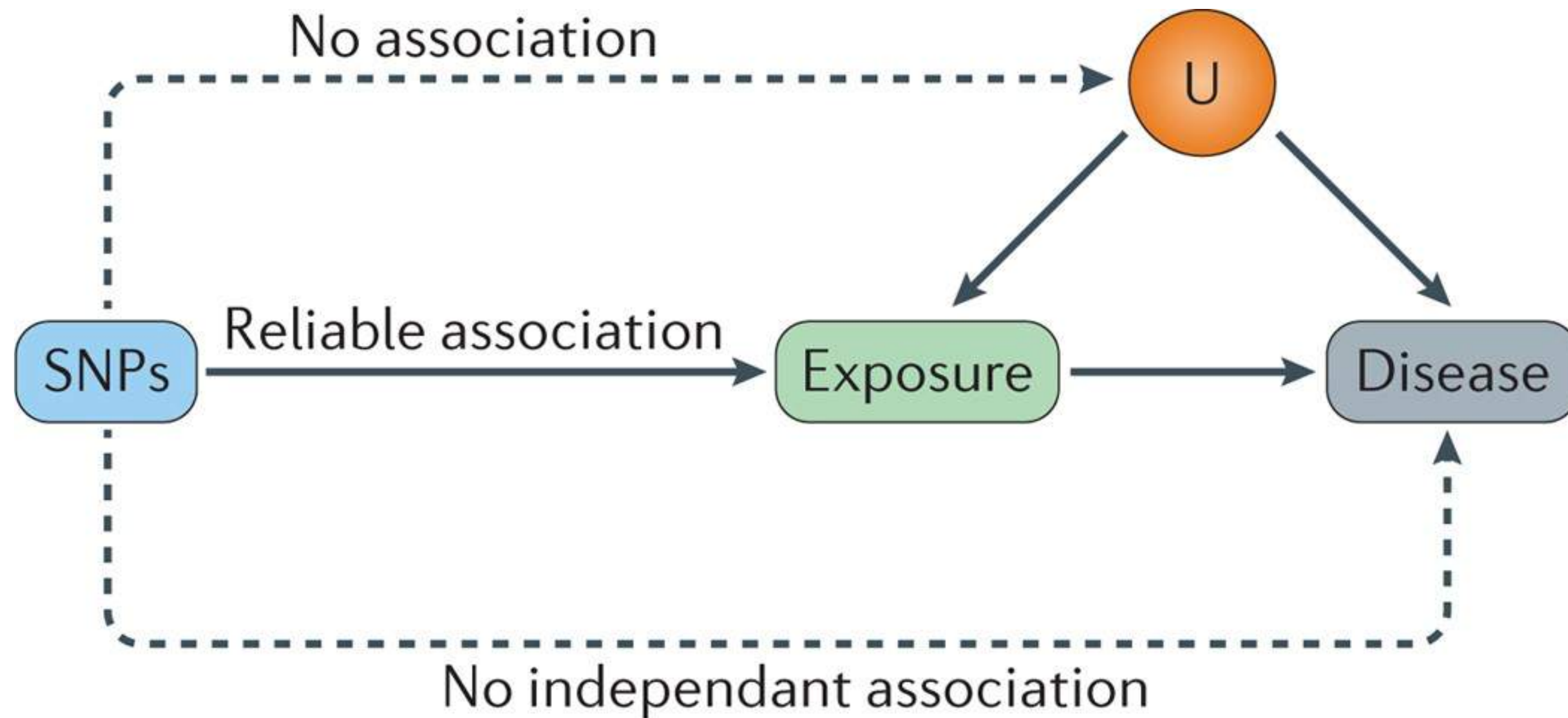
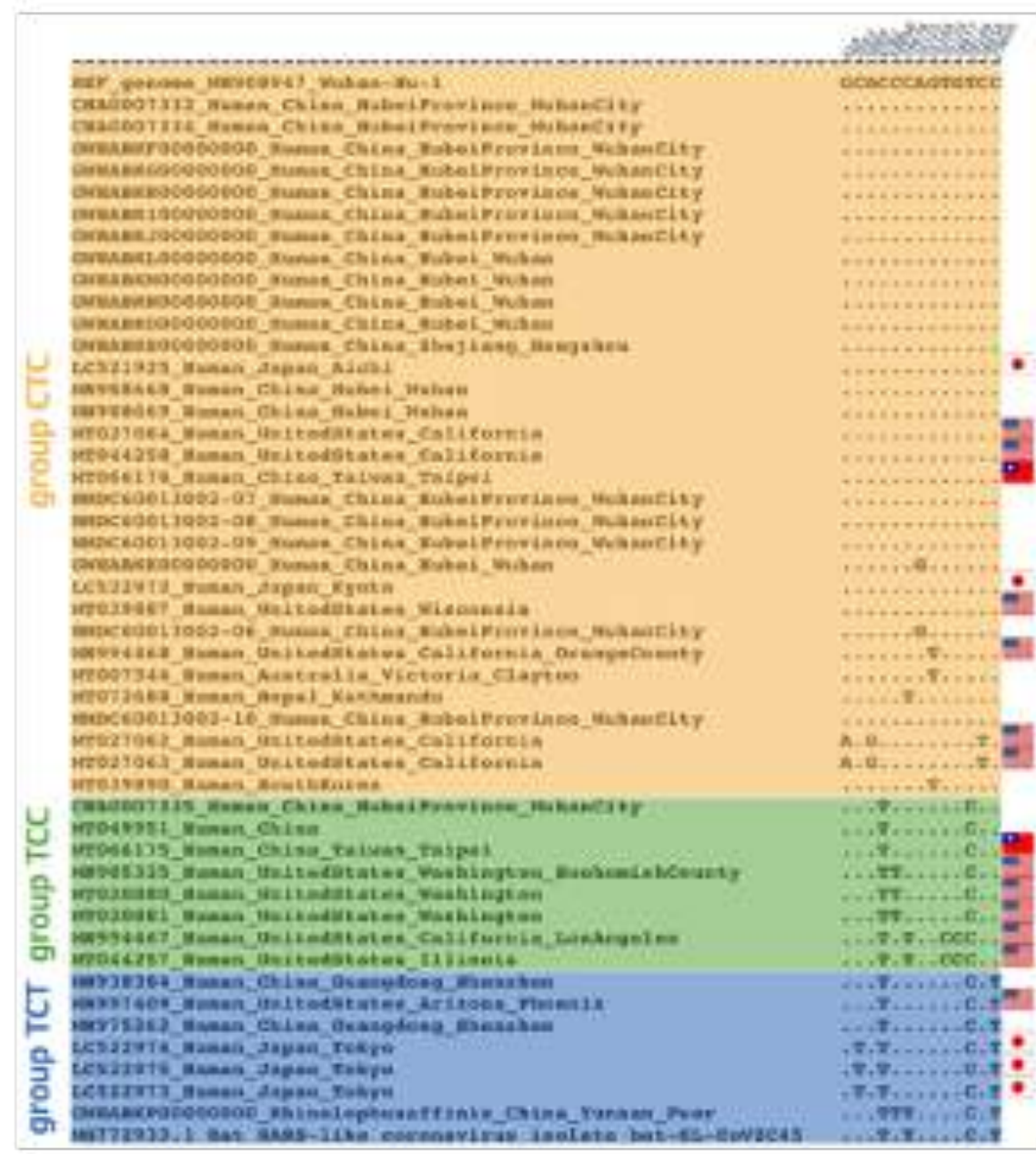
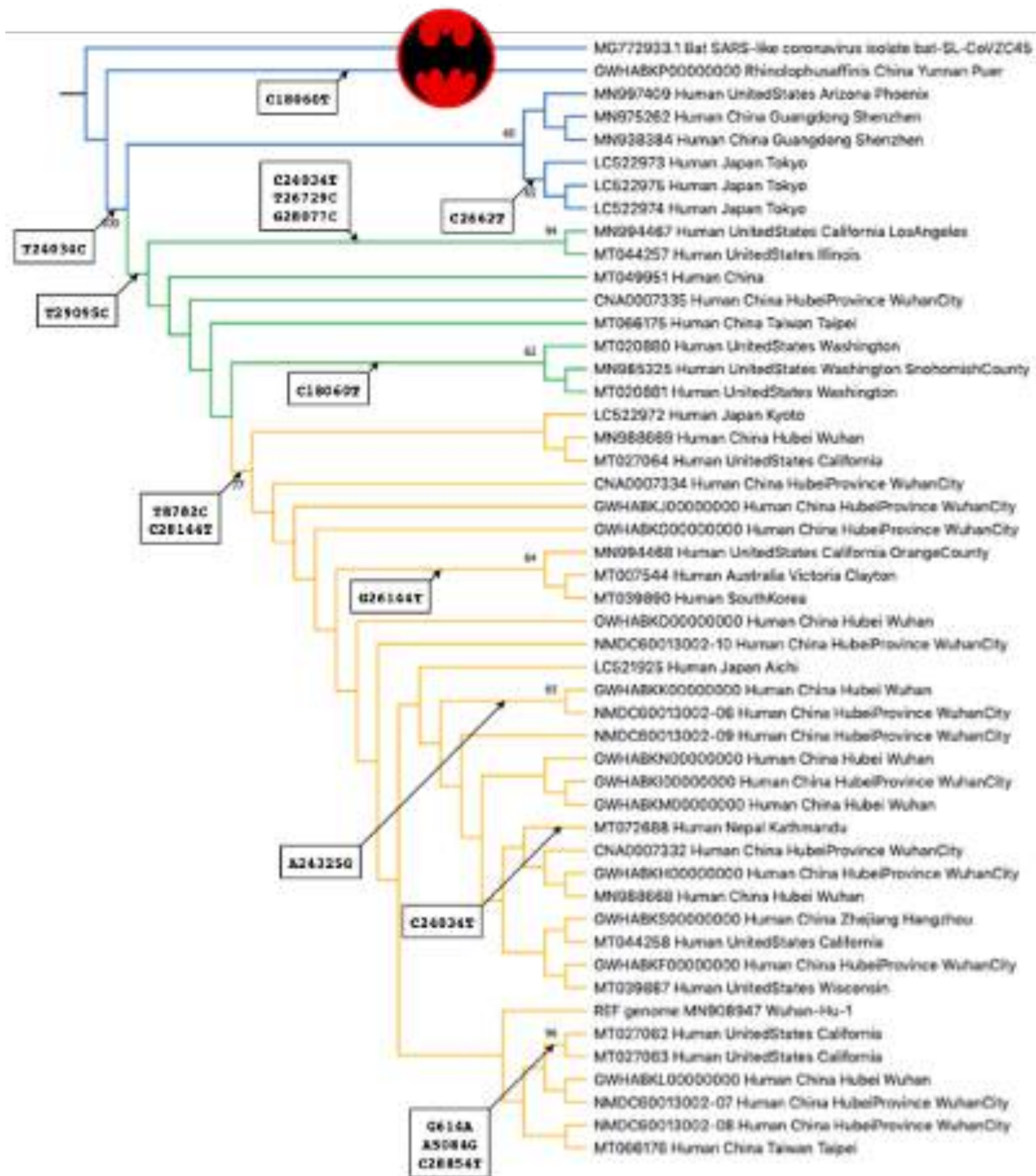


Fig. 1 | Supporting human genetic evidence for new drugs approved by the FDA in 2021. Availability of genetic evidence implicating each of the 50 drug targets (columns) as likely causal genes for the disease for which the drug is indicated. Genetic data sources are stratified (rows) based on the predominant nature of the genetic information. Evidence is classified based on whether it directly supports the gene–indication pair or any of the closely related phenotypes or proteins. See Supplementary information for details and an expanded figure. Data source: [Open Targets Platform](#) (November 2021).

<https://www.nature.com/articles/d41573-022-00120-3>

Nature Reviews | **Cardiology**

# 感染症の追跡にも使えた



Cornell University  
arXiv.org > q-bio > arXiv:2002.08802

Quantitative Biology > Genomics

(Submitted on 29 Feb 2020 (v1); last revised 28 Feb 2020 (this version, v2))

### Phylogenetic analyses of the severe acute respiratory syndrome coronavirus 2 reflected the several routes of introduction to Taiwan, the United States, and Japan

Tomoko Matsuda, Hikoyu Suzuki, Norichika Ogata

Worldwide Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is disrupting in the economy and anxiety of people. The public anxiety has increased the psychological burden on government and healthcare professionals, resulting in a government worker suicide in Japan. The terrified people are asking the government for border measures. However, are border measures possible for this virus? By analyzing 48 almost complete virus genome sequences, we found out that the viruses that invaded Taiwan, the United States, and Japan were introduced independently. We identified thirteen parsimony-informative sites and three groups (CTC, TCC, and TCT). Viruses found outside China did not form a monophyletic clade, opposite to previous study. These results suggest the difficulty of implementing effective border measures against this virus.

TWITTER DEMOGRAPHICS MENDELEY READERS ATTENTION SCORE IN CONTEXT

This research output has an **Altmetric Attention Score of 672**. This is our high-level measure of the quality and quantity of online attention that it has received. This Attention Score, as well as the ranking and number of research outputs shown below, was calculated when the research output was last mentioned on **06 April 2021**.

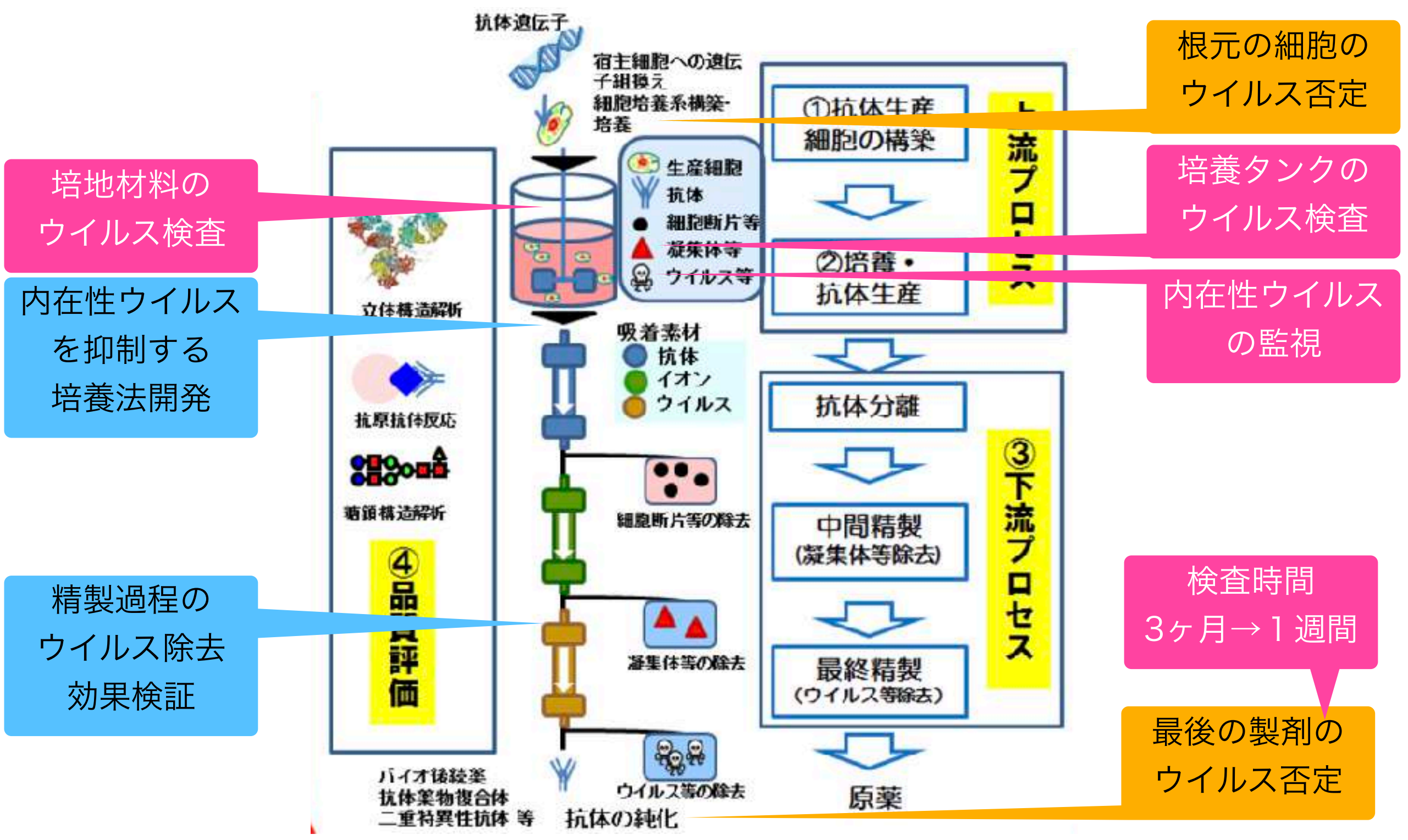
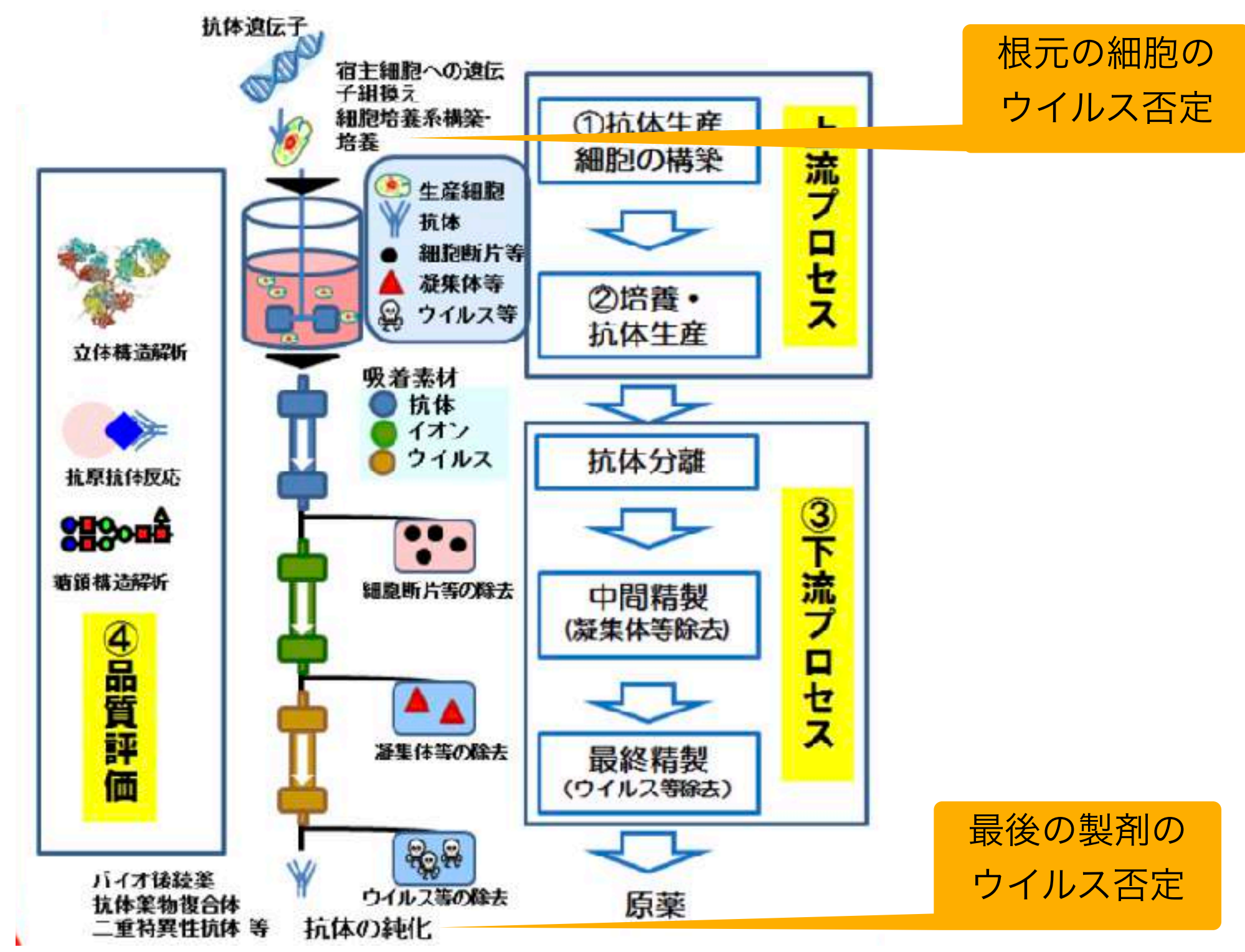
ALL RESEARCH OUTPUTS	OUTPUTS FROM ARXIV	OUTPUTS OF SIMILAR AGE	OUTPUTS OF SIMILAR AGE FROM ARXIV
#16,640 of 17,736,227 outputs	#141 of 710,856 outputs	#685 of 266,960 outputs	#10 of 30,325 outputs

So far Altmetric has tracked 710,856 research outputs from this source. They receive a mean Attention Score of 3.9. This one has done particularly well, scoring **higher than 99% of its peers**.

# バイオ医薬品製造工程のウイルス管理

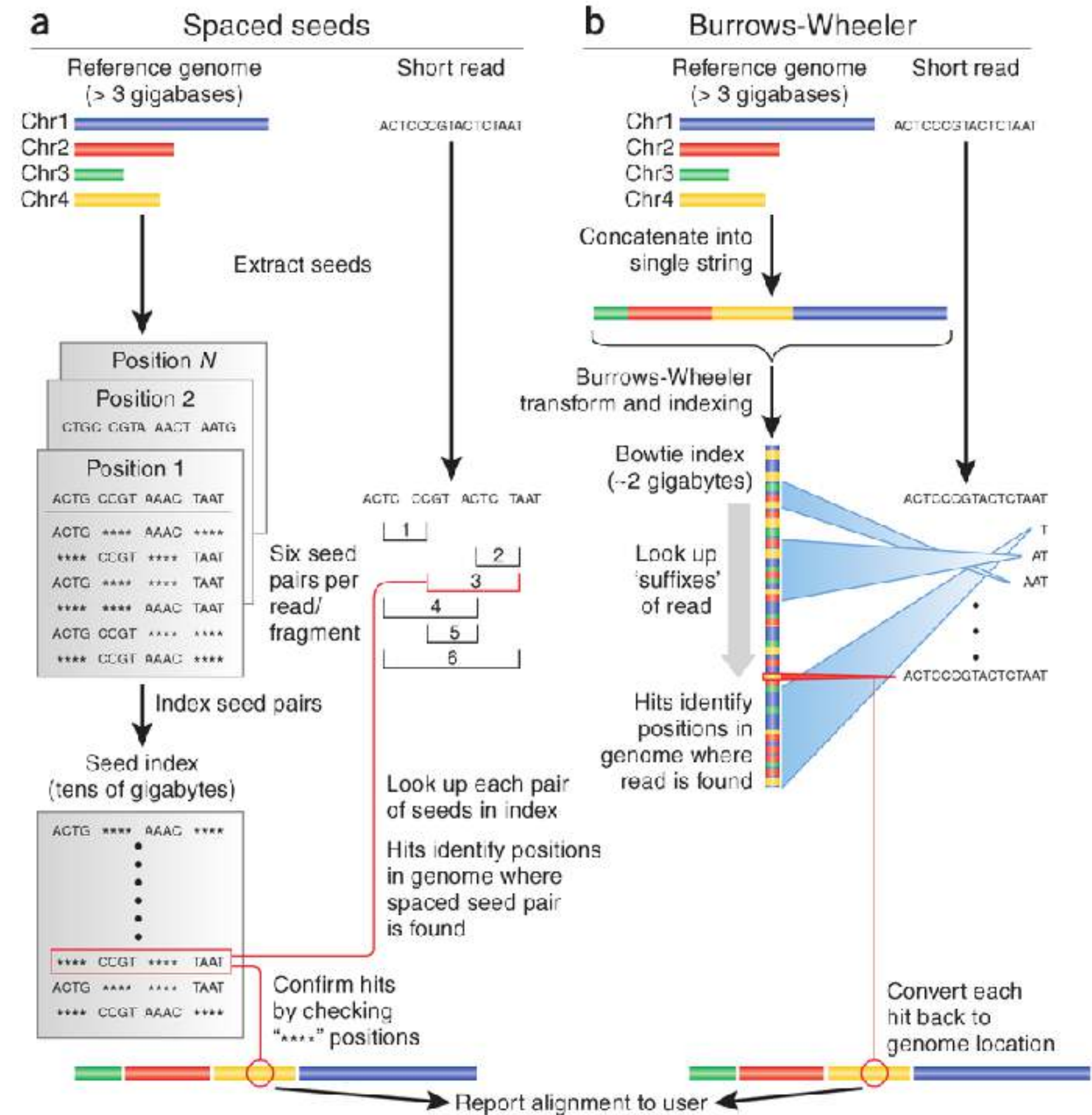
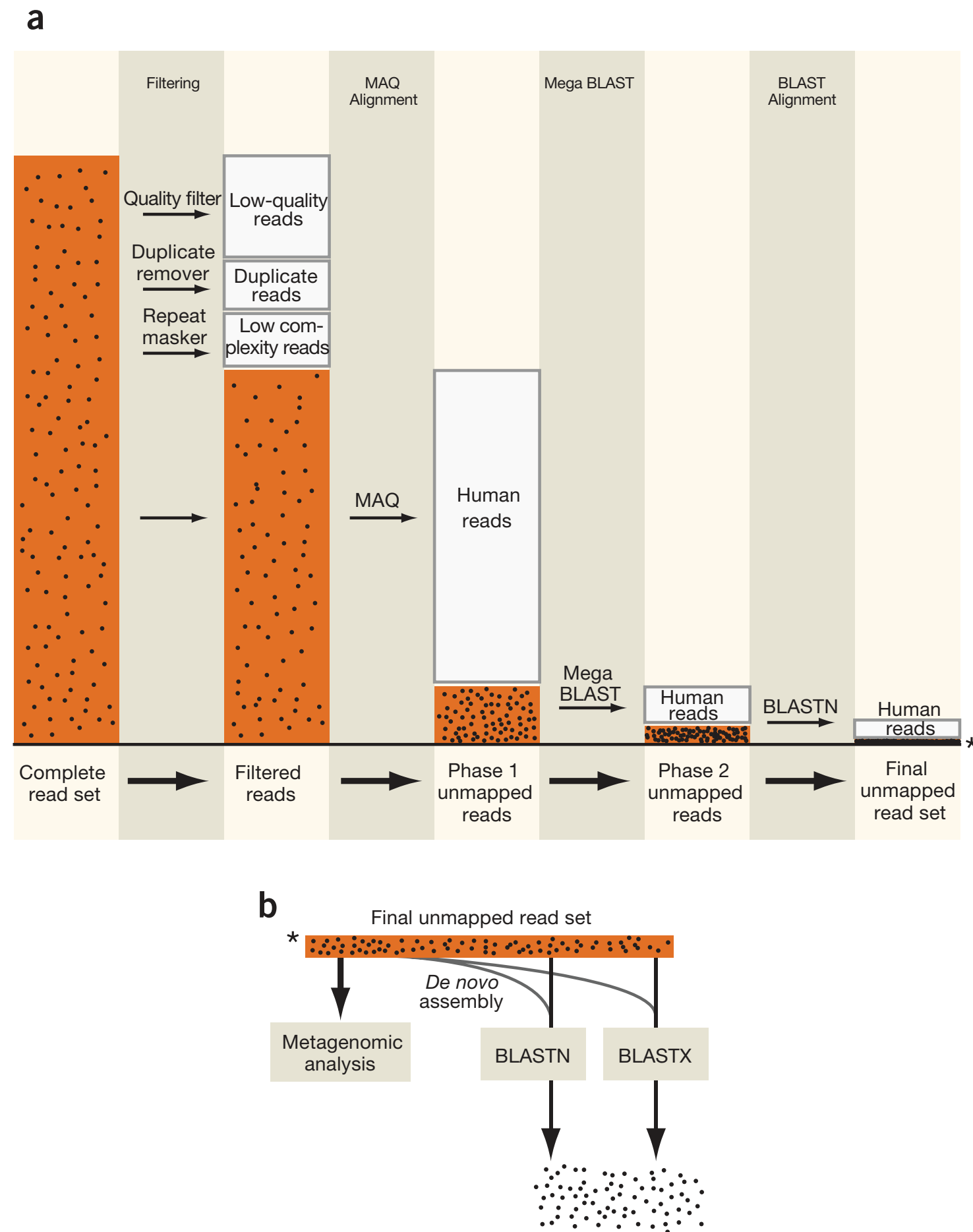
## Before ICH Q5A (R2)

## After ICH Q5A (R2)





# Burrows-Wheeler 変換の活躍



**Figure 1** The PathSeq workflow. (a) Conceptual workflow of the subtractive phase of PathSeq. The size of the read set (orange bars) is proportional to the number of reads at the indicated step in a typical run of the method. The black dots in the bars represent pathogen-derived sequences, which become progressively concentrated. The steps in this conceptual workflow have been reordered for concision (see **Supplementary Methods** for actual ordering). (b) Conceptual workflow of the analytic phase of PathSeq. The asterisk indicates the unmapped read-set that is carried over from the subtractive phase.

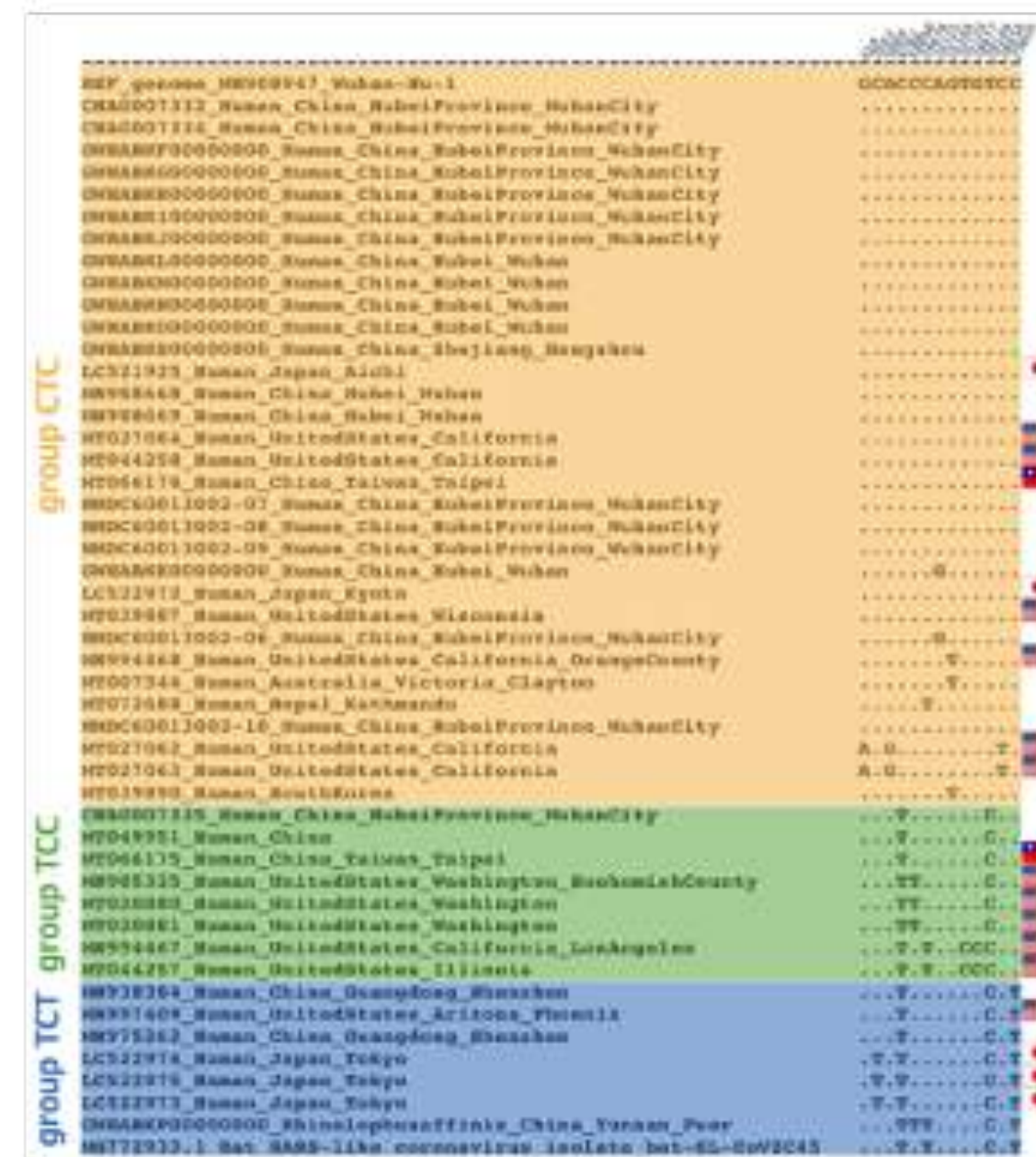
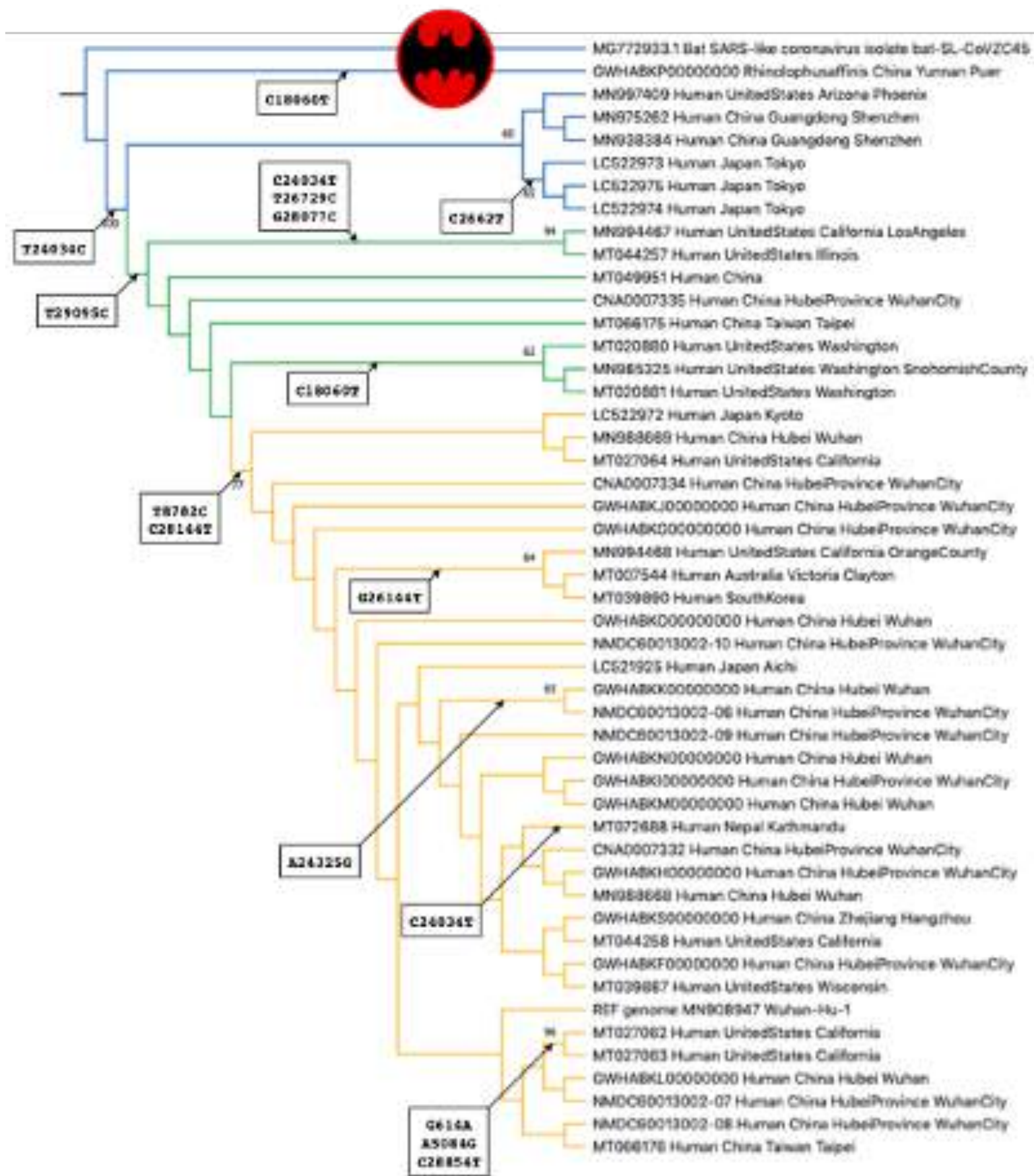
How to map billions of short reads onto genomes  
 Nature Biotechnology 27(5):455-7 (2009)  
 DOI:10.1038/nbt0509-455

# ここまでのまとめ

- ゲノム科学に情報理論が使われている
  - ゲノム配列を手に入れるためにオイラーパスが
  - ゲノム配列を利用するためにbw変換が使われている
  - どちらも（たぶん）儲かってない
- 
- そもそも儲ける話ではなかったか
  - 儲けて何になるのか

# 太陽は無料

コロナの変異株を見つけても儲からないし、それをつかって感染追跡しても儲からないし、病院からはPCRのプライマーくれと言われる



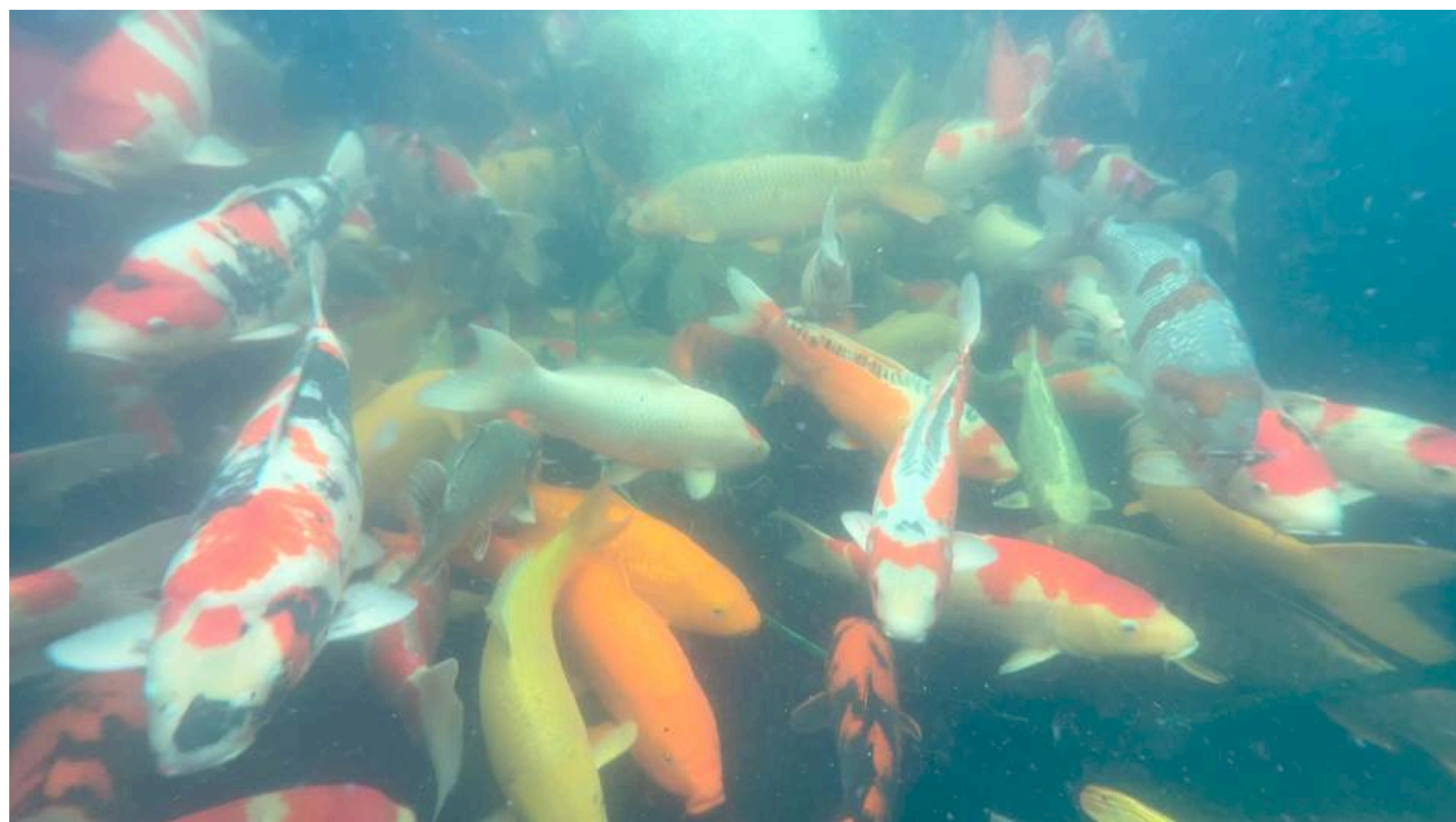
Cornell University  
arXiv.org > q-bio > arXiv:2002.08802  
Quantitative Biology > Genomics  
Submitted on 28 Feb 2020 (v1); last revised 28 Feb 2020 (this version, v2)  
**Phylogenetic analyses of the severe acute respiratory syndrome coronavirus 2 reflected the several routes of introduction to Taiwan, the United States, and Japan**  
Tomoko Matsuda, Hikoju Suzuki, Norichika Ogata  
Worldwide Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is disrupting in the economy and anxiety of people. The public anxiety has increased the psychological burden on government and healthcare professionals, resulting in a government worker suicide in Japan. The terrified people are asking the government for border measures. However, are border measures possible for this virus? By analyzing 48 almost complete virus genome sequences, we found out that the viruses that invaded Taiwan, the United States, and Japan were introduced independently. We identified thirteen parsimony-informative sites and three groups (CTC, TCC, and TCT). Viruses found outside China did not form a monophyletic clade, opposite to previous study. These results suggest the difficulty of implementing effective border measures against this virus.

TWITTER DEMOGRAPHICS MENDELEY READERS ATTENTION SCORE IN CONTEXT  
This research output has an **Altmetric Attention Score of 672**. This is our high-level measure of the quality and quantity of online attention that it has received. This Attention Score, as well as the ranking and number of research outputs shown below, was calculated when the research output was last mentioned on **06 April 2021**.

ALL RESEARCH OUTPUTS	OUTPUTS FROM ARXIV	OUTPUTS OF SIMILAR AGE	OUTPUTS OF SIMILAR AGE FROM ARXIV
#16,640 of 17,736,227 outputs	#141 of 710,856 outputs	#685 of 266,960 outputs	#10 of 30,325 outputs

So far Altmetric has tracked 710,856 research outputs from this source. They receive a mean Attention Score of 3.9. This one has done particularly well, scoring **higher than 99% of its peers**.

# ワクチンマナーを生贄に養鯉場を召喚



# 貧しさを恐れる心を破壊する

## ◎ 本学同窓生の緒方法親氏に紺綬褒章が授与されました

本学に多額のご寄附をされた、緒方法親氏（株式会社日本バイオデータ 代表取締役社長、本学卒業・修了生）に、2019年9月28日(土)に日本国政府から紺綬褒章が授与され、2019年12月12日(木)に本学の犬野弘幸学長から褒章および褒状が伝達されました。

伝達式ののち、懇談会を開催し、緒方氏のご家族、元担任の千葉一裕農学研究院長および元指導教員の岩淵喜久男名誉教授も参加され、大変和やかな雰囲気の中に盛会に終わりました。

緒方氏は、本学農学部をご卒業後、農学府・連合農学研究科を修了され、2019年7月には、学長主催の企業トップセミナーで、「農工大で年2億円ラボの20代PI博士をつくろう」と題してご講演をいただきました。

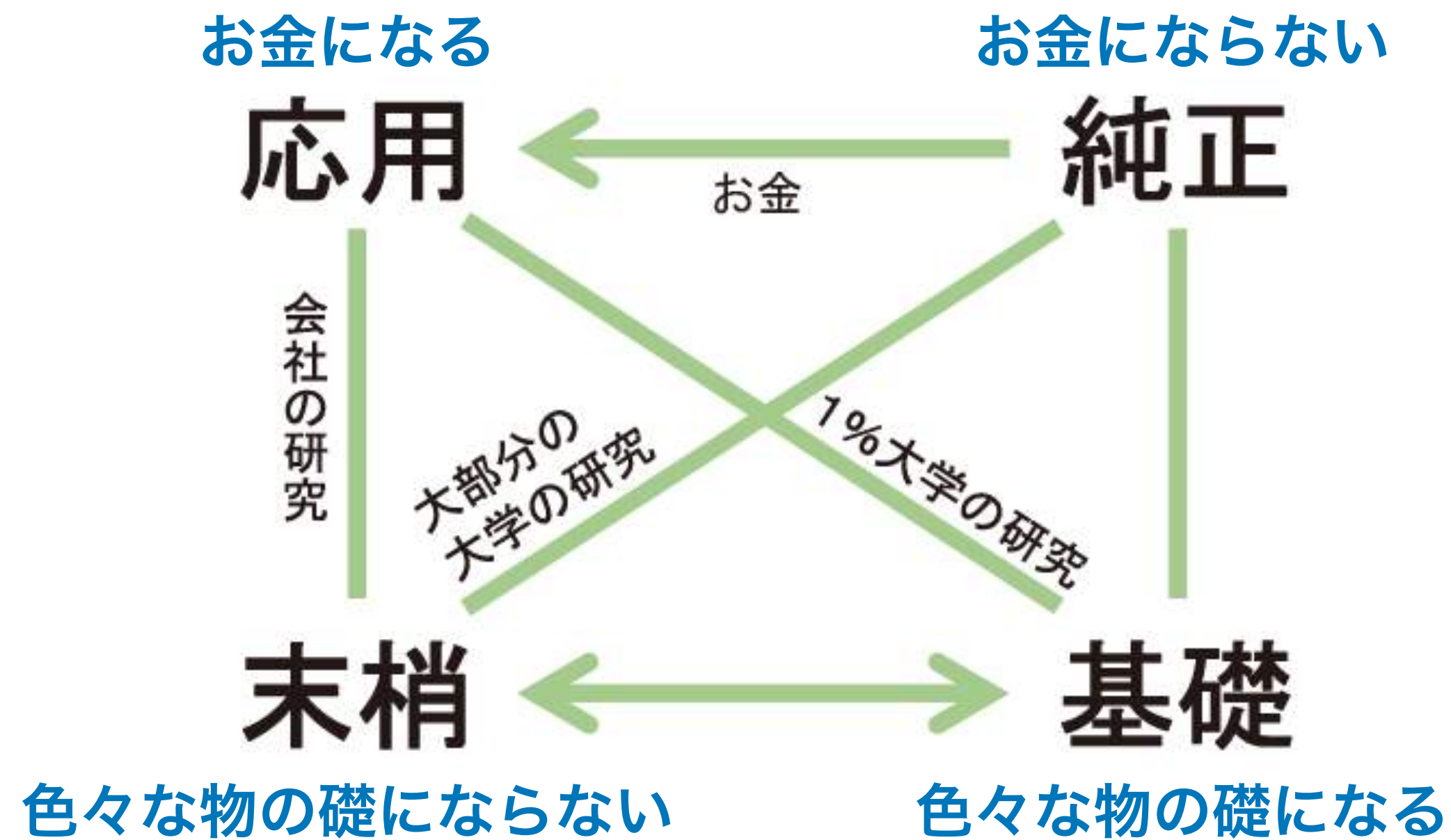
本学での紺綬褒章の授与者は、遠藤章特別栄誉教授に続いて、2人目となります。



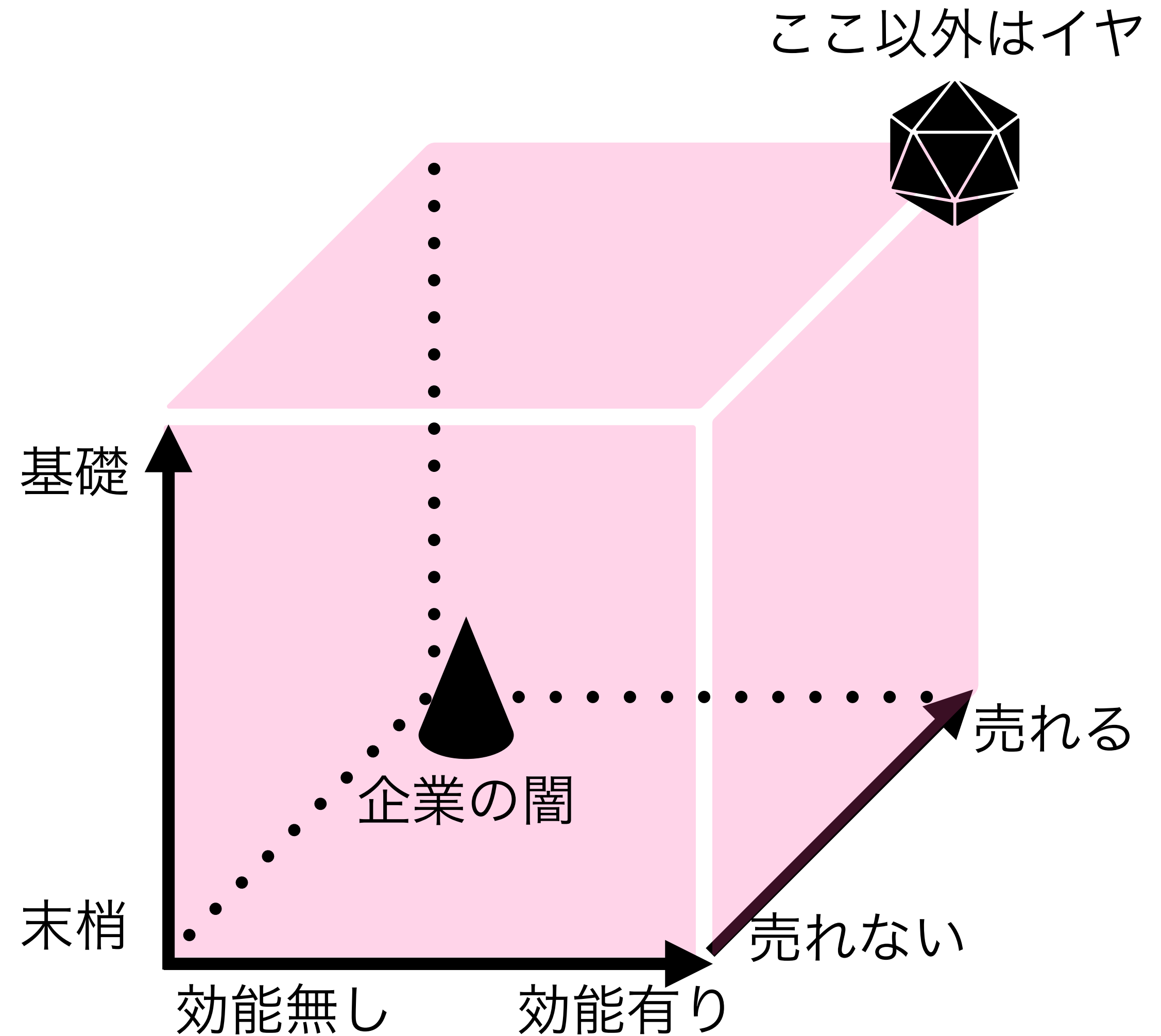
貧しさとは貧しさを怖れる心

# 応用基礎研究のすすめ

基礎研究⇔応用研究という比較は誤っている



JST/産学官ジャーナル/2018年10月/応用研究と純正研究  
元ネタ・上田良二「応用基礎研究のすすめ」  
原題「研究開発にみる日本の後進性」。「蟻塔」9月号、1980年









# 特許庁以外は大抵許してくれる

News & Trend + 連載をフォロー

## 2012年の特許庁システム開発中止、開発費全額返納のなぜ

浅川 直輝 日経コンピュータ 2014.07.31

CPU/GPUの排熱が最新サーバーの差別化ポイントとなる、その理由は？ PR  
【サステナビリティに向けた新イノベーション】サーバーの「液浸冷却」とは？ PR  
守りながら攻める“製造DX”とは 適材適所で使い分けるハイブリッドクラウド PR

東芝ソリューションとアクセンチュアが、2012年に開発を中止した特許庁システムの開発費に利子を加えた約56億円を、同庁に返納していたことが分かった。特許庁によれば、2013年8月に合意が成立、同年9月に返納金として両社から約56億円が支払われたという。

同システムの開発では、特許庁は東芝ソリューションに対し、2009年度までの4年間で約24億8700万円を、アクセンチュアには2011年度までに約29億6400万円、計54億5100万円を支払っている。つまり、両社は受け取った開発費全額を、利子付きで特許庁に返納したことになる。

# 細胞の話させてください

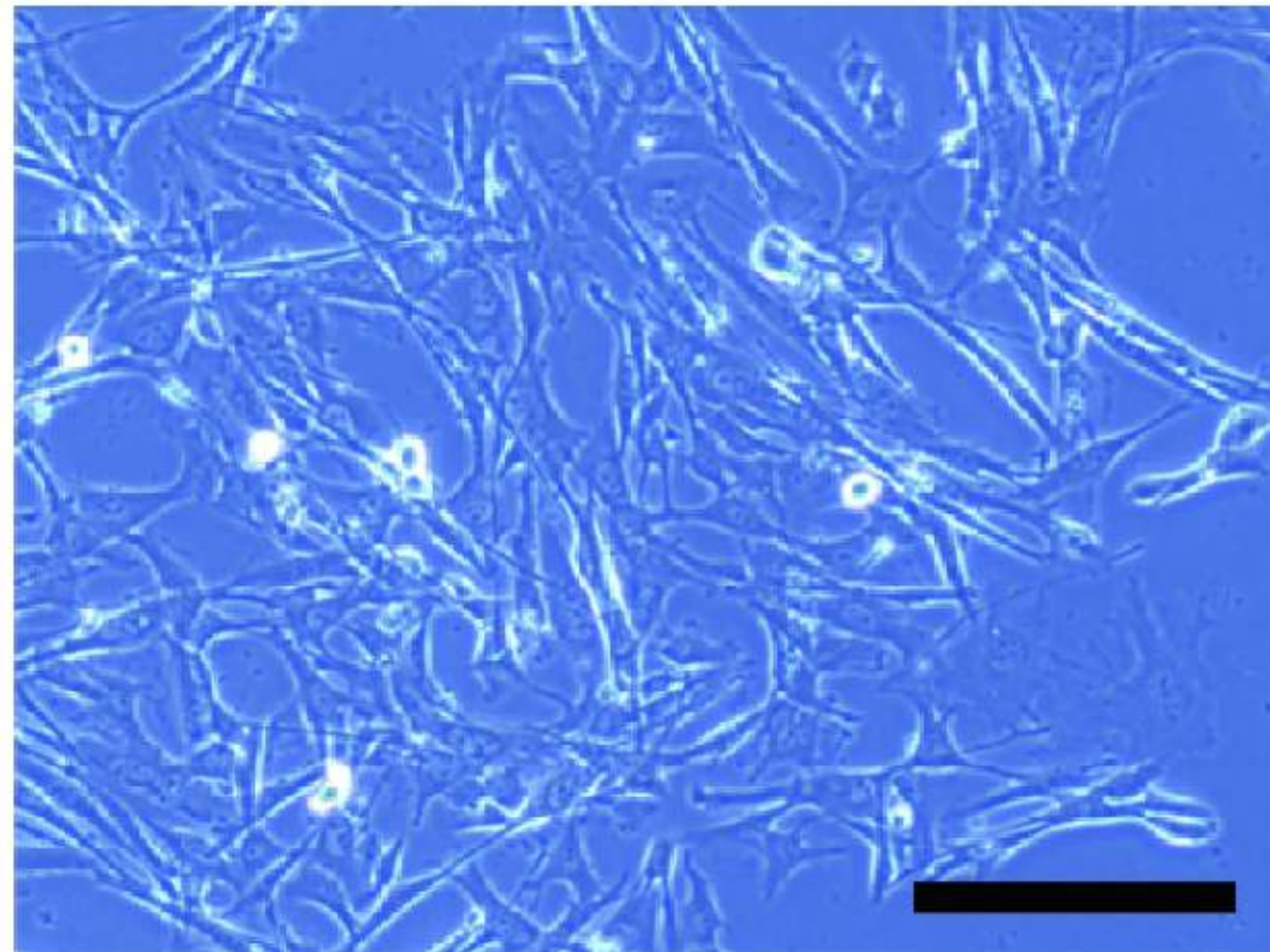
多細胞生物の細胞を培養していると不死化することがある。

” "Immortality"という言葉の意味は、自明のことと考えるかもしれないが、そうではない。最近のある論文で、「"partial immortality"を獲得した細胞」と述べているものがあつた。これは矛盾した言葉であるが、かなりの的を得た表現法のひとつであろう。”

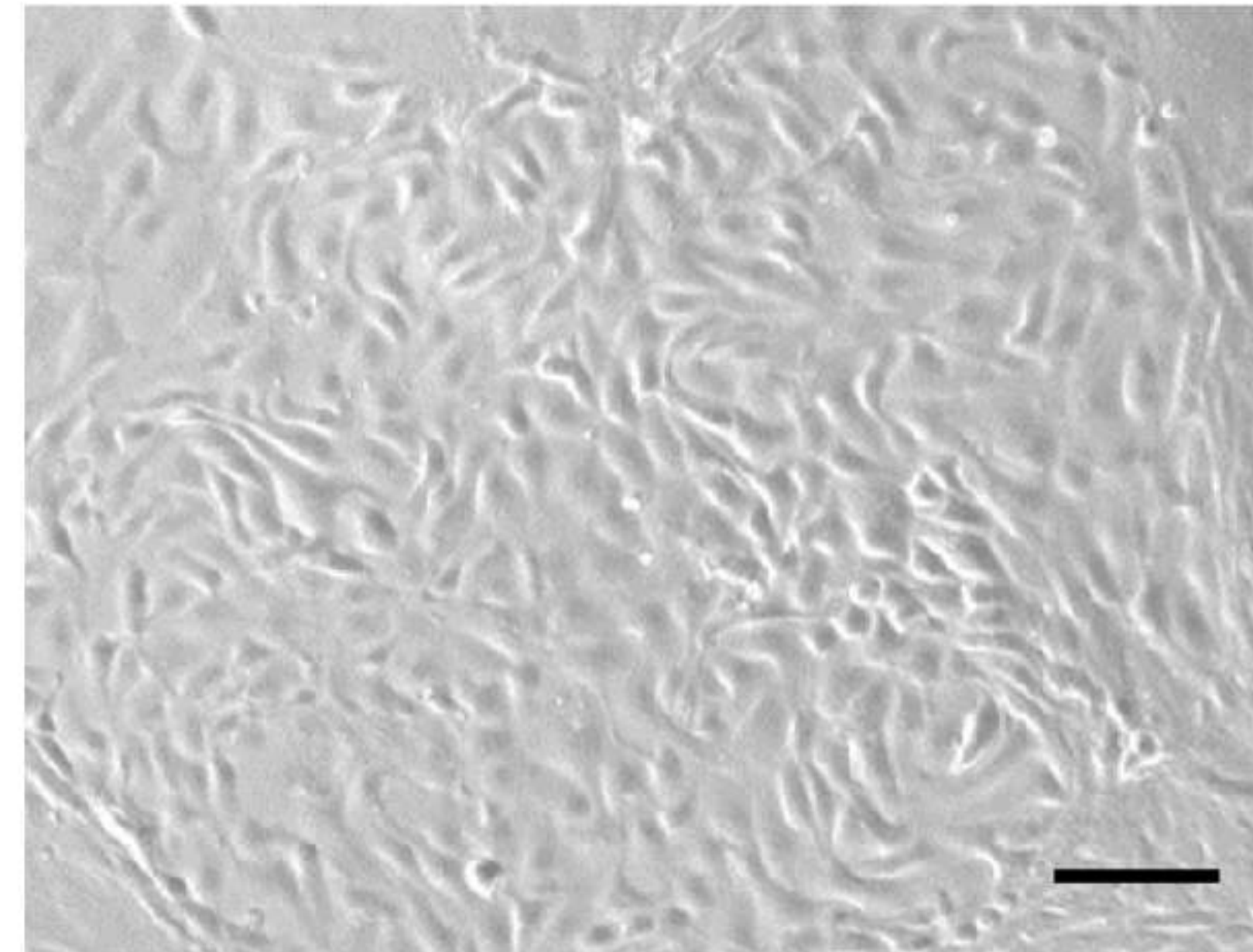
ヒト細胞の老化・不死化・癌化/難波正義、渡邊正己、L.Hayflick 編著/共立出版/p.10-18

The Role of Telomeres in the Mortality and Alleged Immortality of Cultured Human Cells

ヒト培養細胞の老化と不死化におけるテロメアの役割/Leonard Hayflick/杉原進介 訳



PMID: 33077772



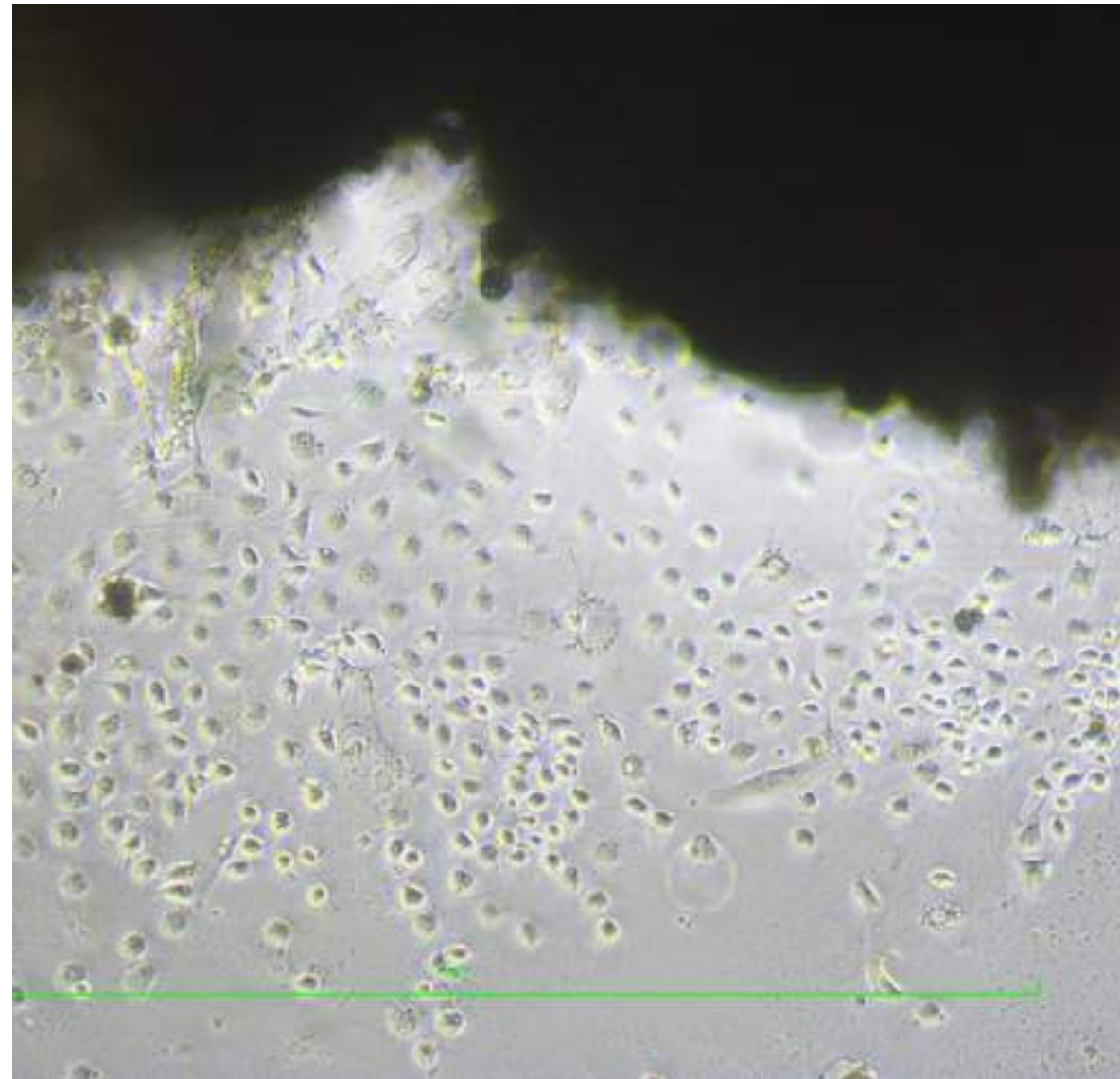
PMID: 35358245

死から不死の生まれる過程を知りたい

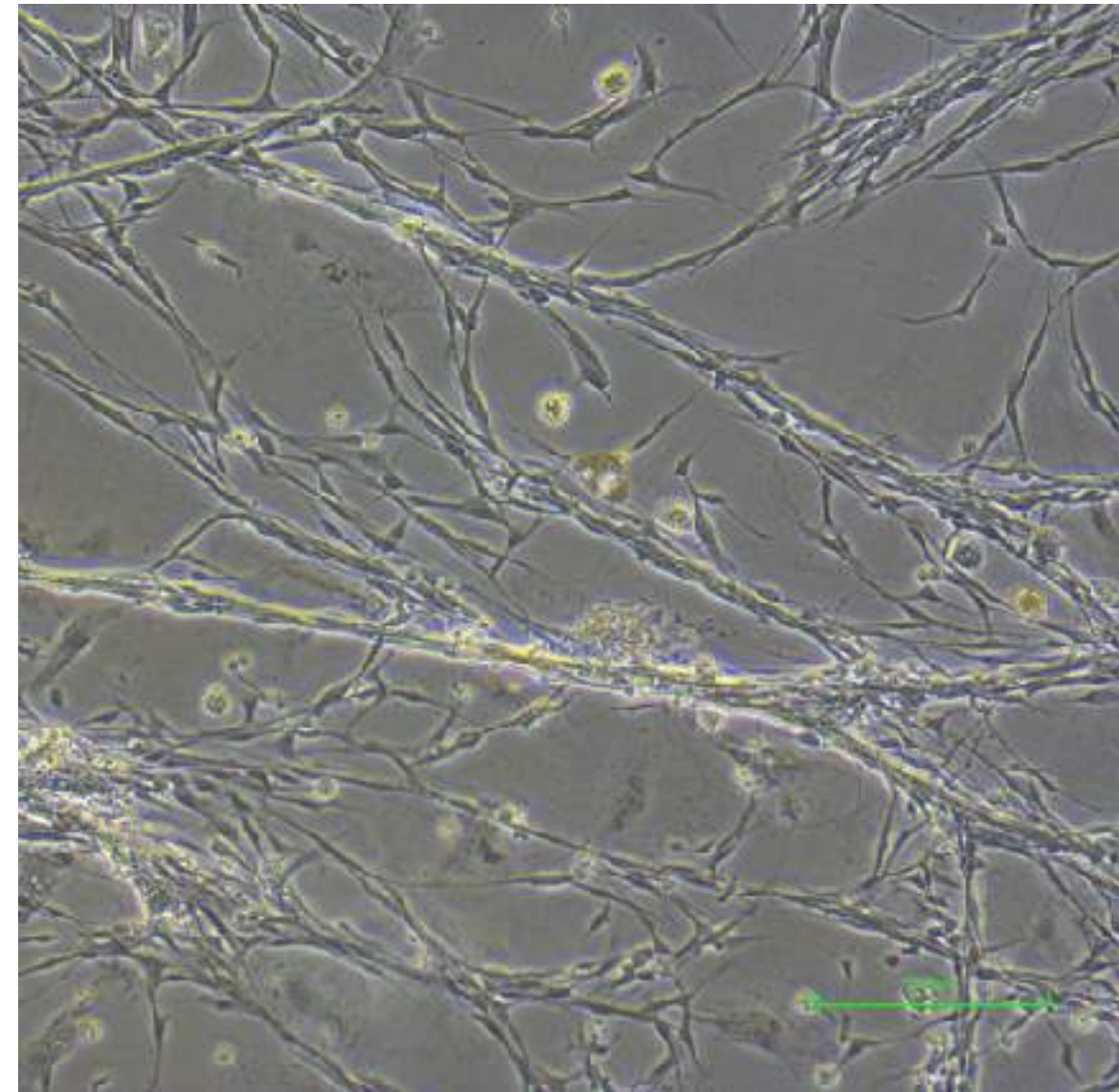


# 細胞の話させてください

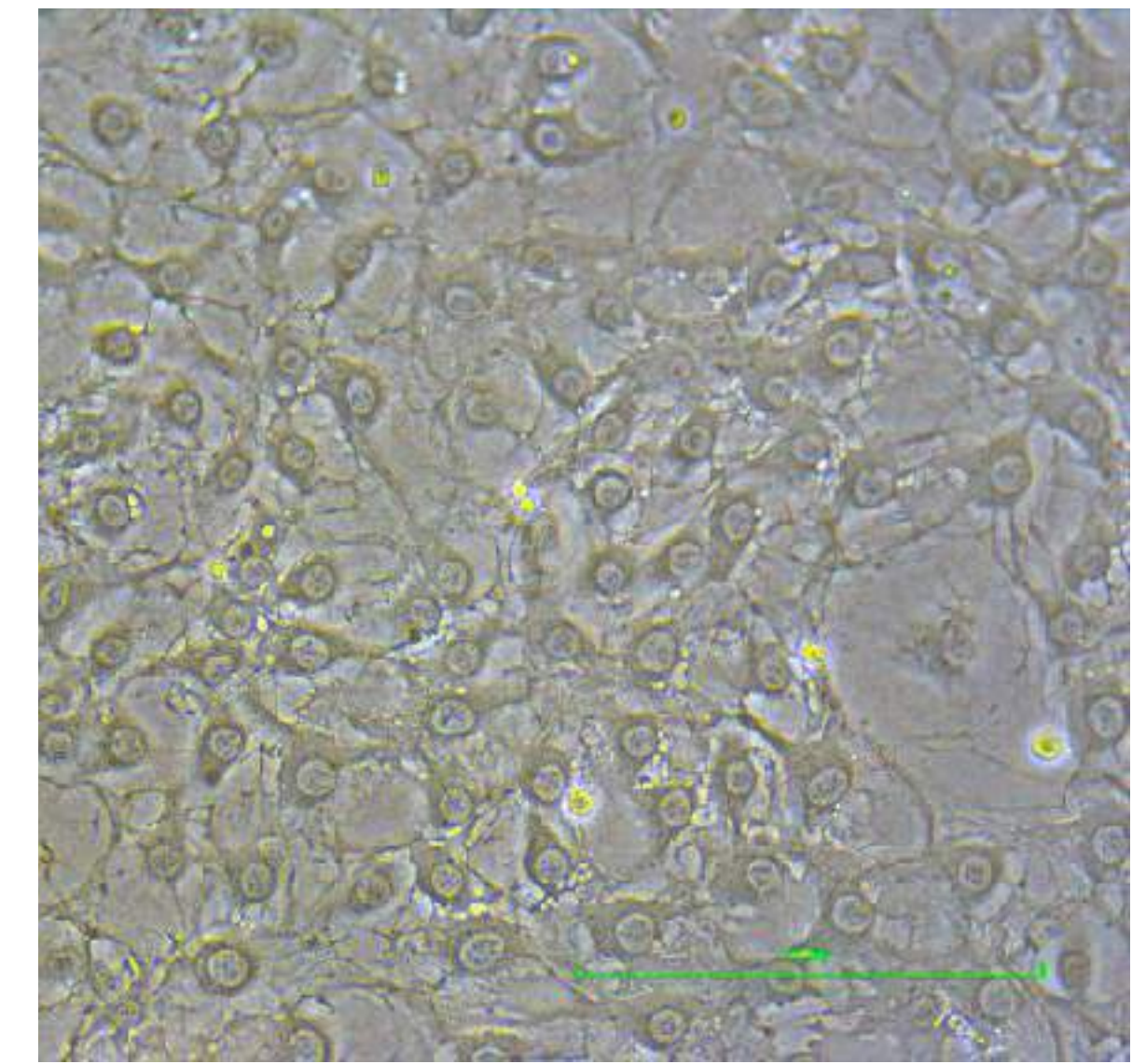
個体の中の細胞



初代培養の細胞



樹立細胞系の細胞

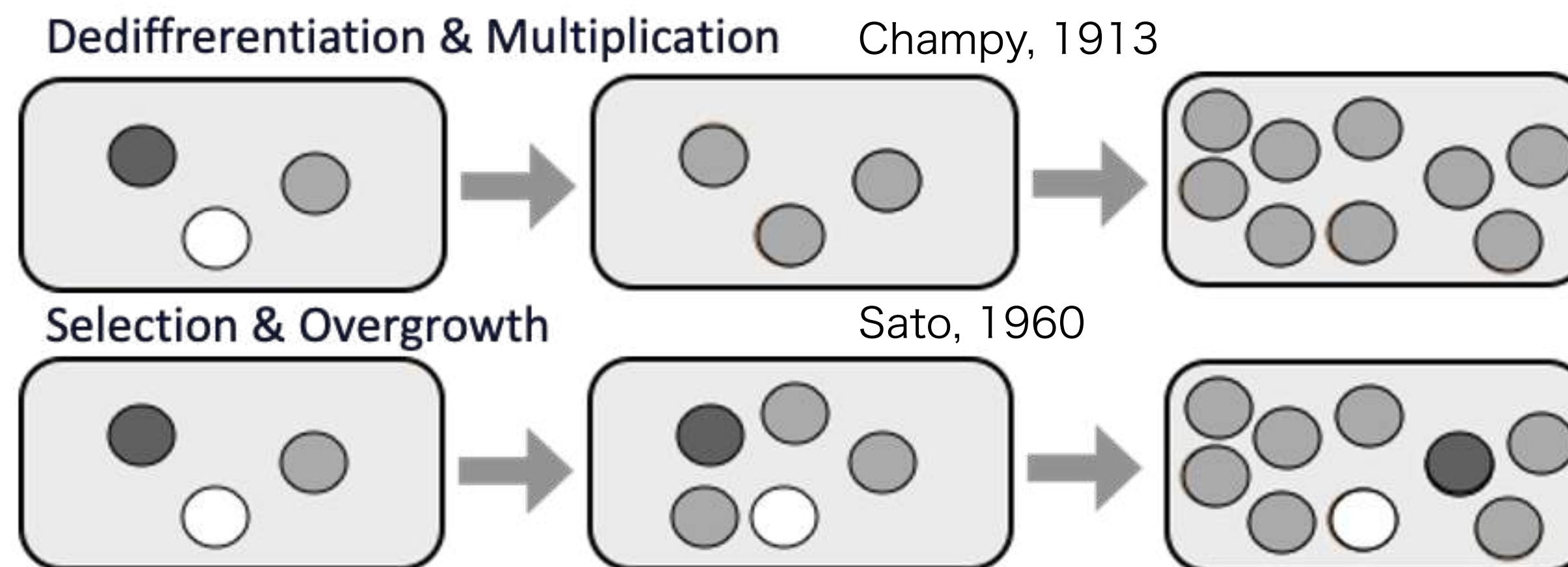


いずれ死ぬ

何かが起こる

(集団としては)不死

# 細胞は培養によって脱分化するのだろうか



J'ai montré dans les publications citées plus haut qu'on ne pouvait pas cultiver un tissu ou un organe en lui conservant ses attributs caractéristiques, c'est-à-dire multiplication des cellules celles-ci se dédiffénciaient.

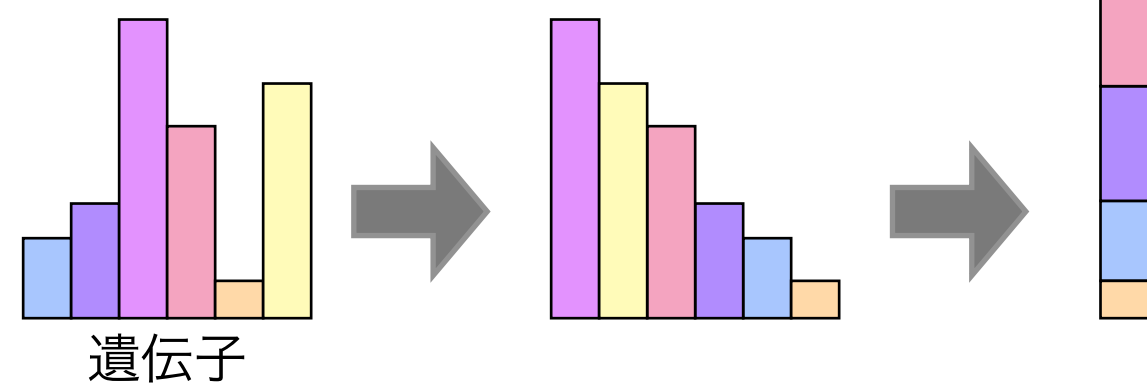
上記の研究に示した通り、培養する前の組織の特徴を維持したままで培養し続けることはできなかった。脱分化することで細胞は増殖する。 Champy, 1913

⇒細胞の増殖抜きにトランスクリプトームから見ると、培養によってシャノンエントロピーが増大する。

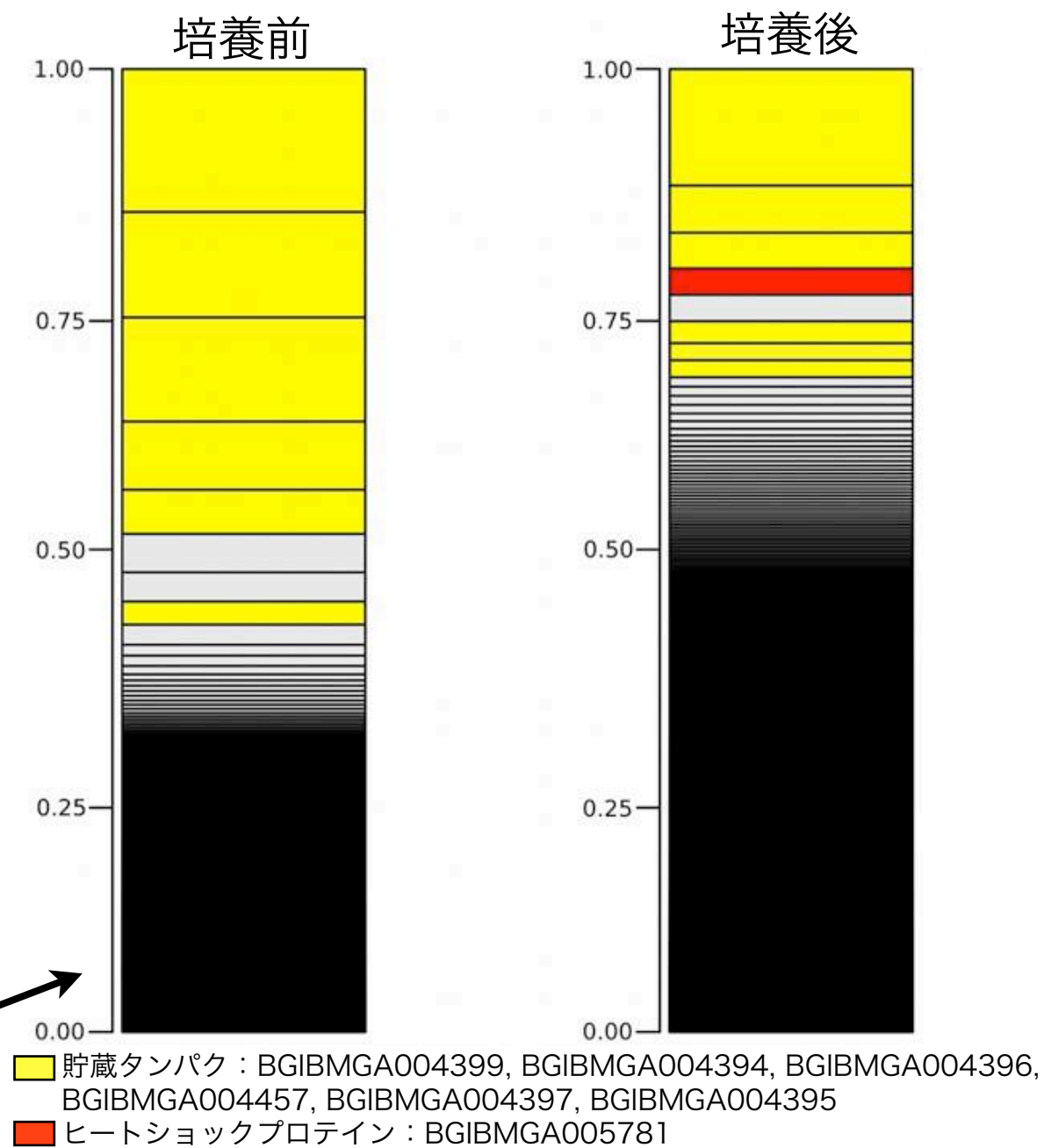
⇒分化/脱分化の程度はトランスクリプトームの情報エントロピーとして定量評価可能そう。

# 細胞は培養によって脱分化するのだろうか

●全遺伝子の発現量を1本のグラフで示した。



黒く塗りつぶされたように見える部分には10,000個程度の遺伝子が含まれている。



ぼく「左の図をひとつの値にしてくれ」  
西村「シャノンエントロピー使え」

$$H_j = - \sum_{i=1}^g p_{ij} \log_2(p_{ij})$$

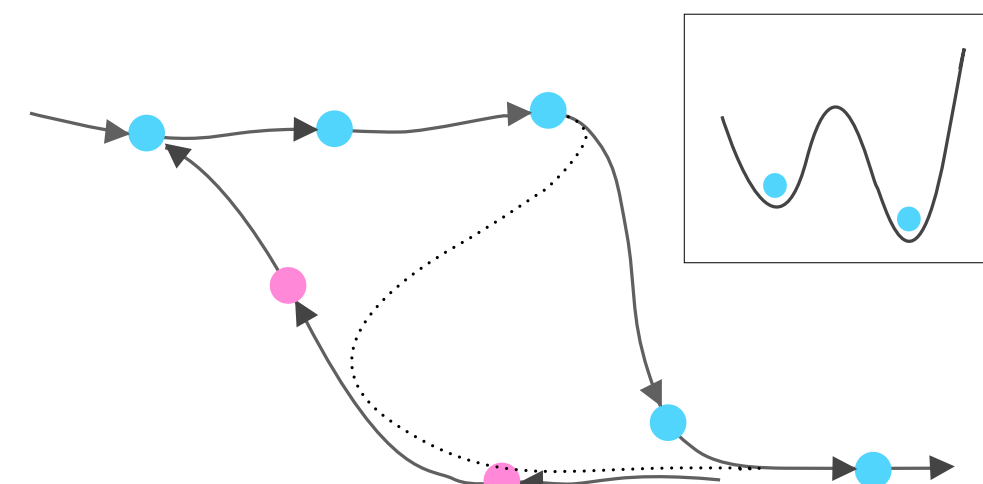
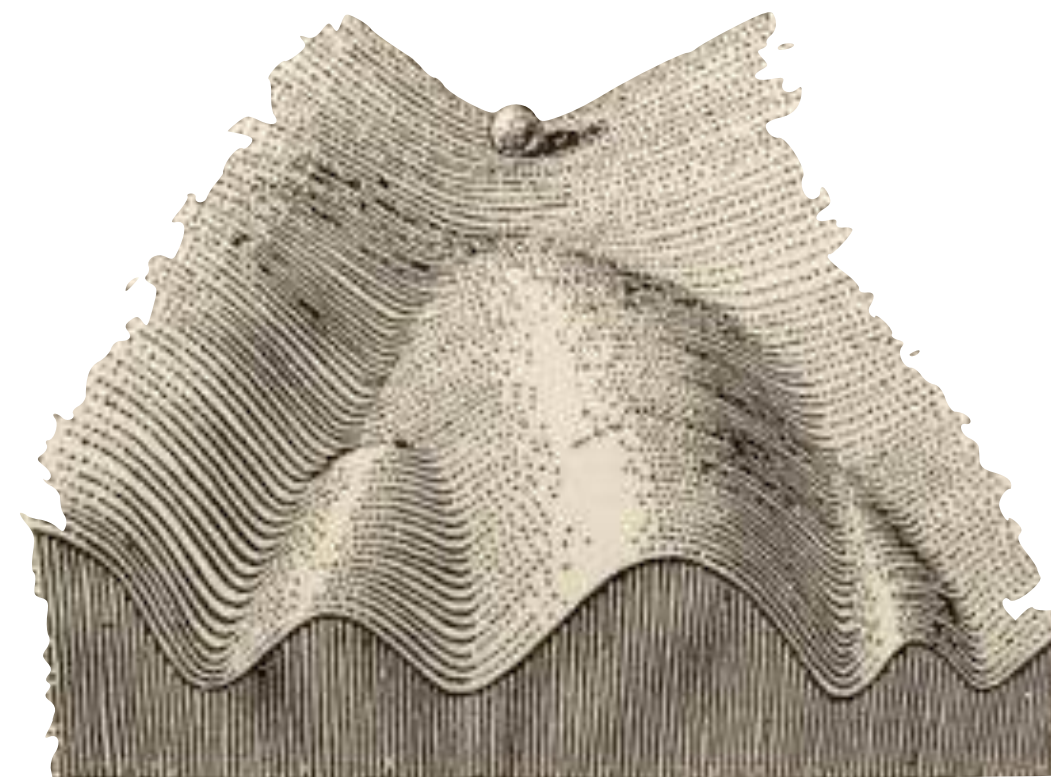
任意の組織*j*の任意の遺伝子*i*の発現量を*P<sub>ij</sub>*とした時の、その組織のトランスクリプトームの不均一度(*H<sub>j</sub>*)

実験区	<i>H<sub>j</sub></i>
培養前	6.488197
培養後	8.518913

Shannon, 1948; Martinez and Reyes-Valdes, 2008

# この値の変化は信じても良いだろうか

トランスクリプトームのシャノンエントロピーの性質を調べる  
ふたつの世界があるなら過去にどこにいたか (=ヒストリー) に依存する



“双極安定性という言葉は、スイッチがヒステリシスを示すときに使う”

ワトソン 遺伝子の分子生物学 6版 p.722

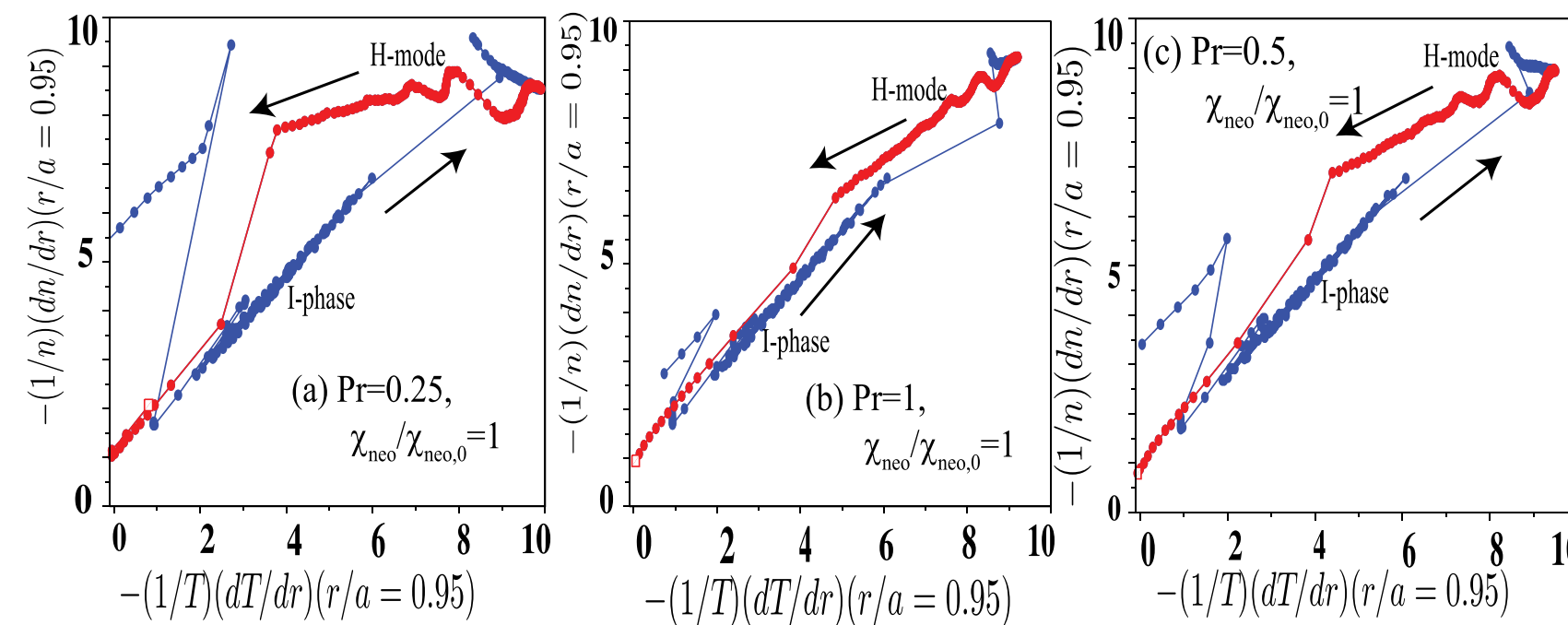
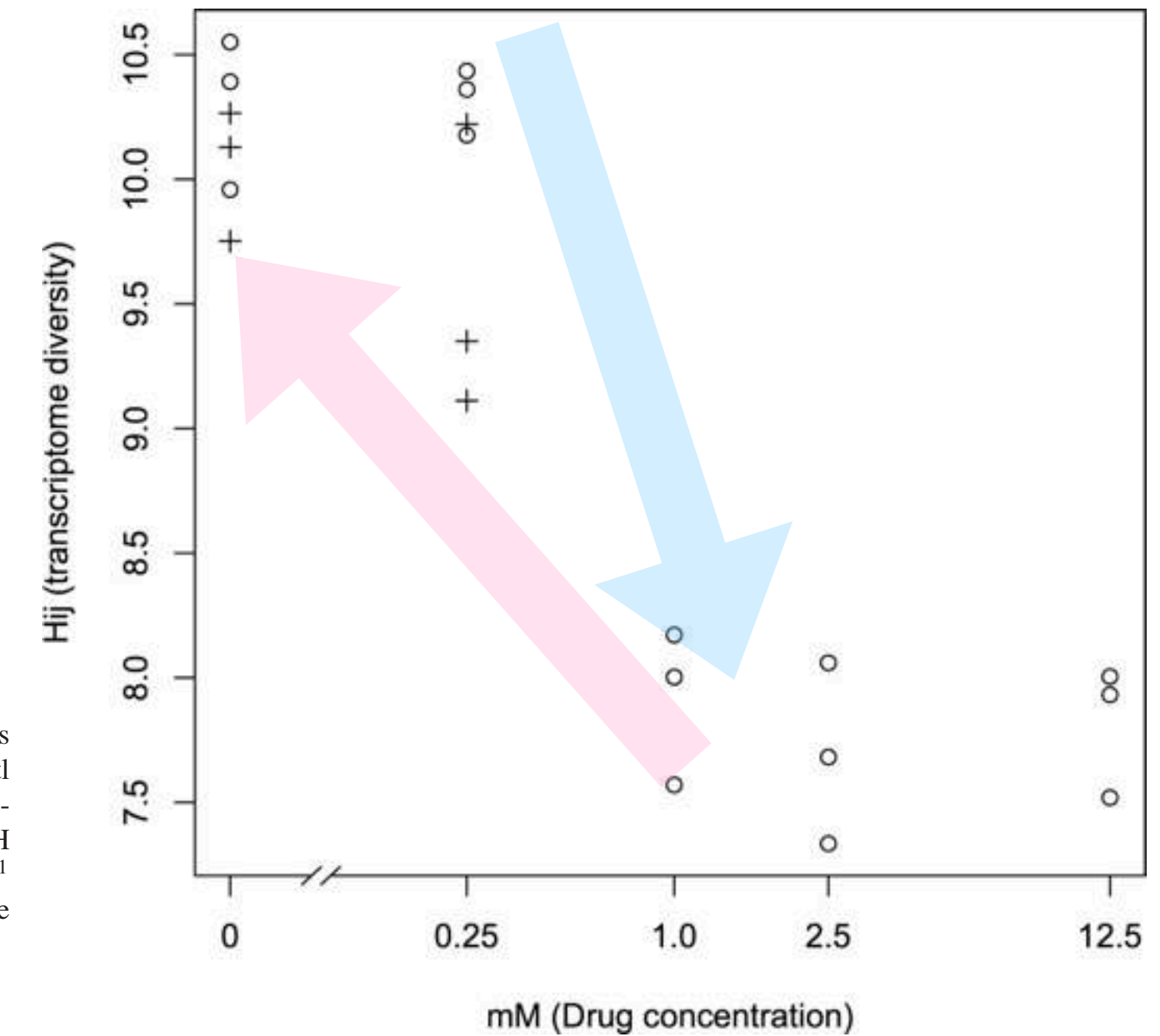
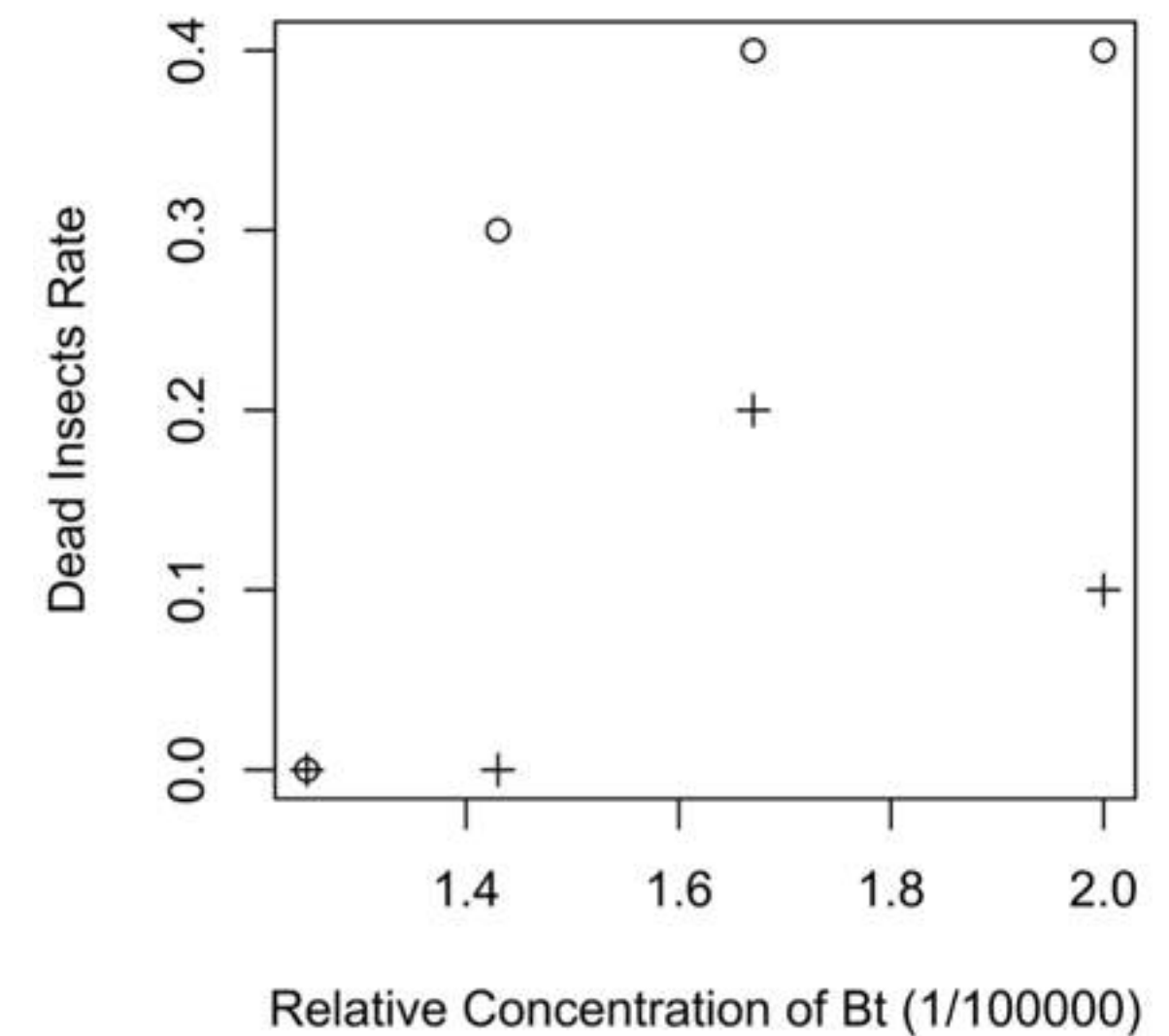
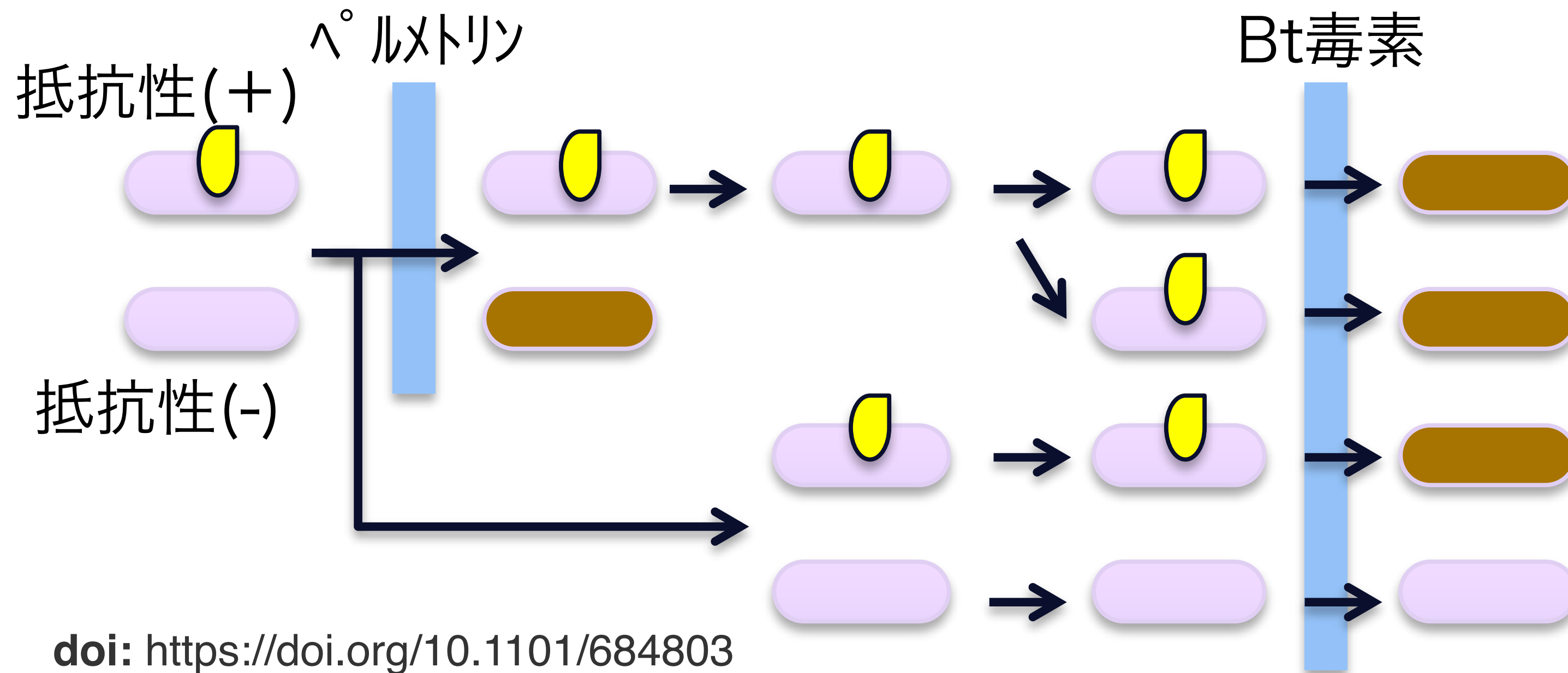
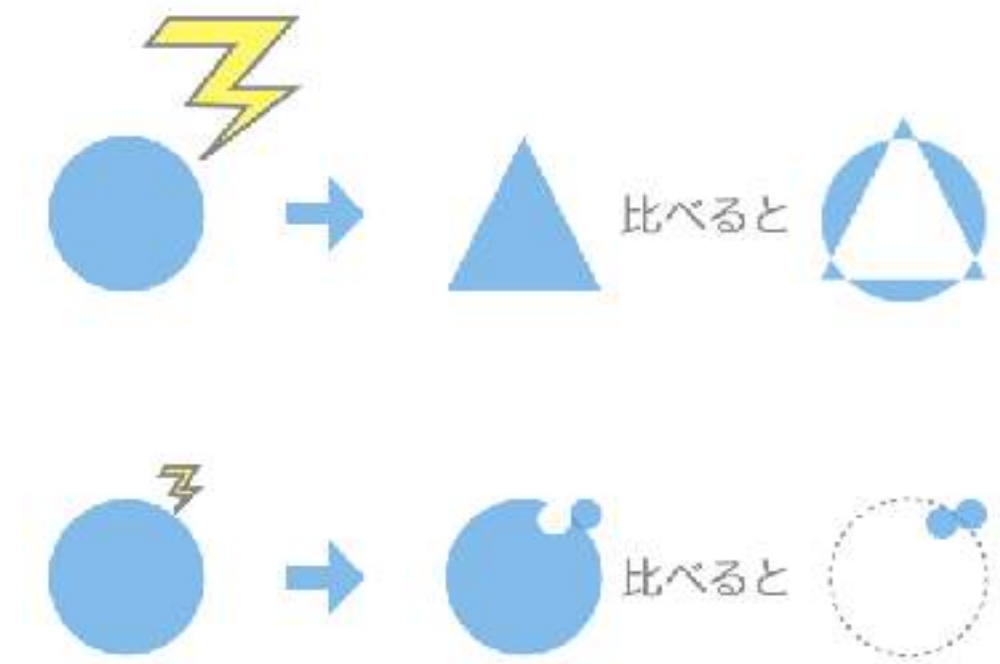
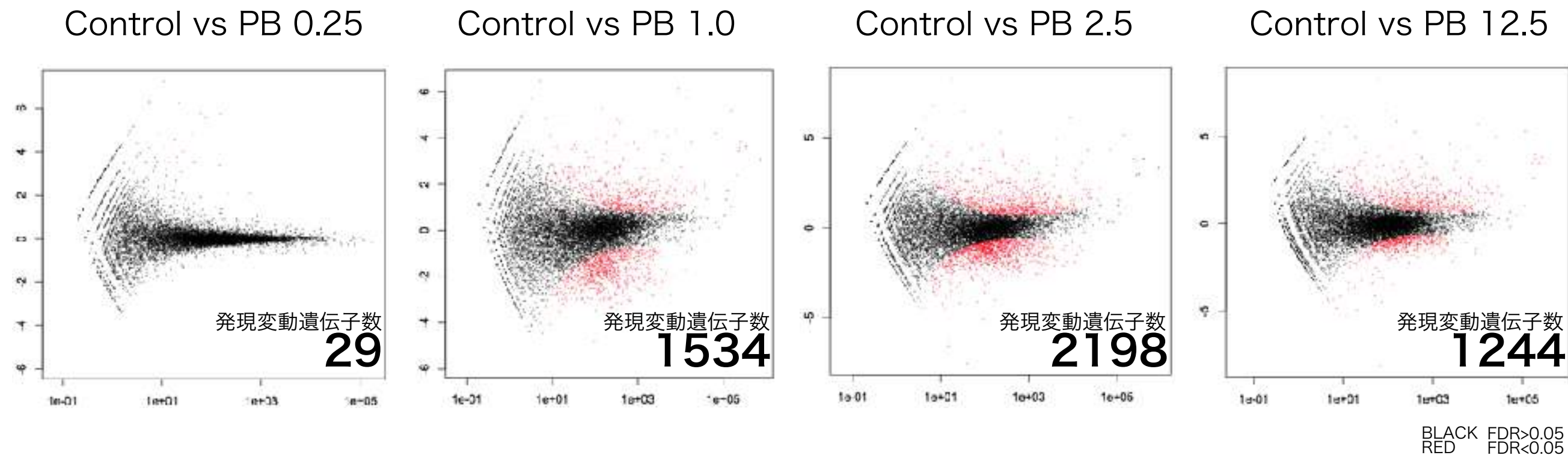
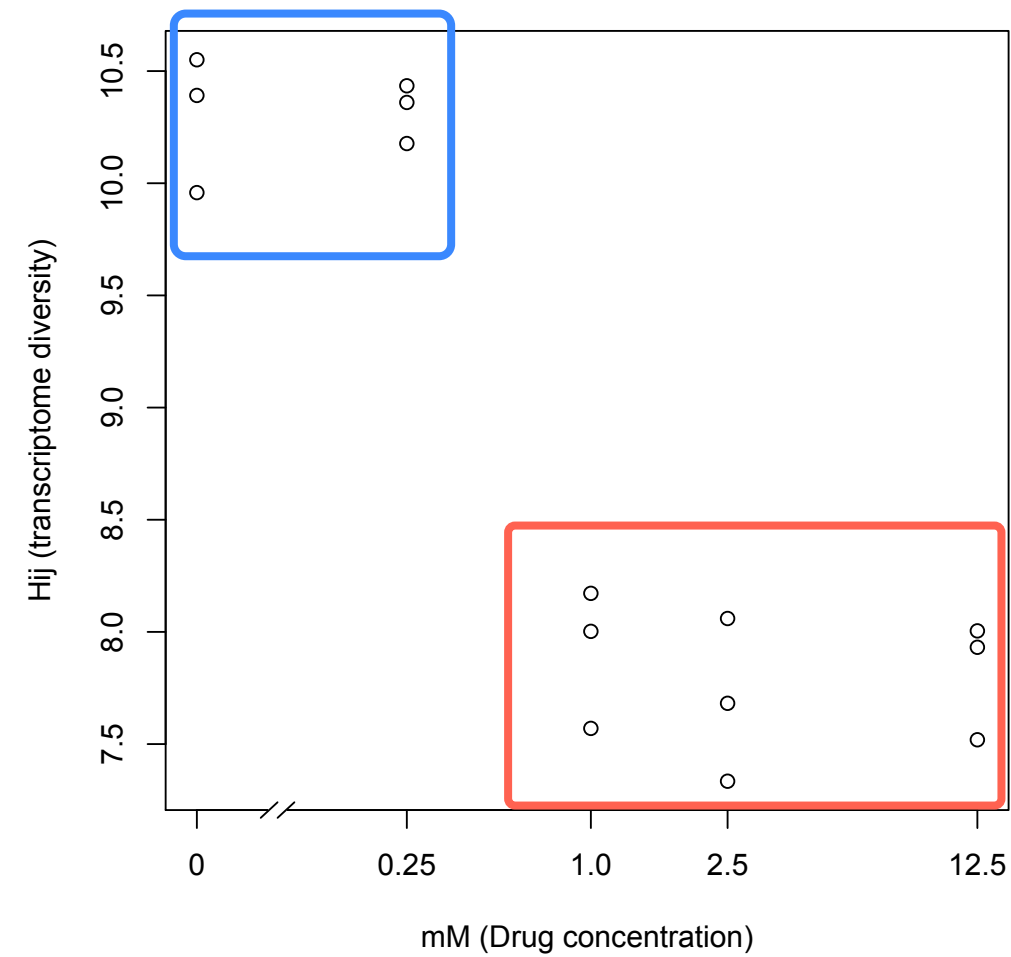


FIG. 7. Relative hysteresis in  $L_T^{-1}$  vs  $L_n^{-1}$  at  $r/a = 0.95$ , with different Prandtl numbers. Blue plots in  $L_T^{-1} \sim L_n^{-1}$  indicate the evolution through the L  $\rightarrow$  H transition, while red plots in  $L_T^{-1} \leq L_n^{-1}$  indicate the evolution through the H  $\rightarrow$  L back transition.



Spatio-temporal evolution of the H  $\rightarrow$  L back transition  
K. Miki, P. H. Diamond, L. Schmitz, D. C. McDonald, T. Estrada et al.  
Phys. Plasmas 20, 062304 (2013); doi: 10.1063/1.4812555

# 実験の予備検討（濃度決定）につかえた

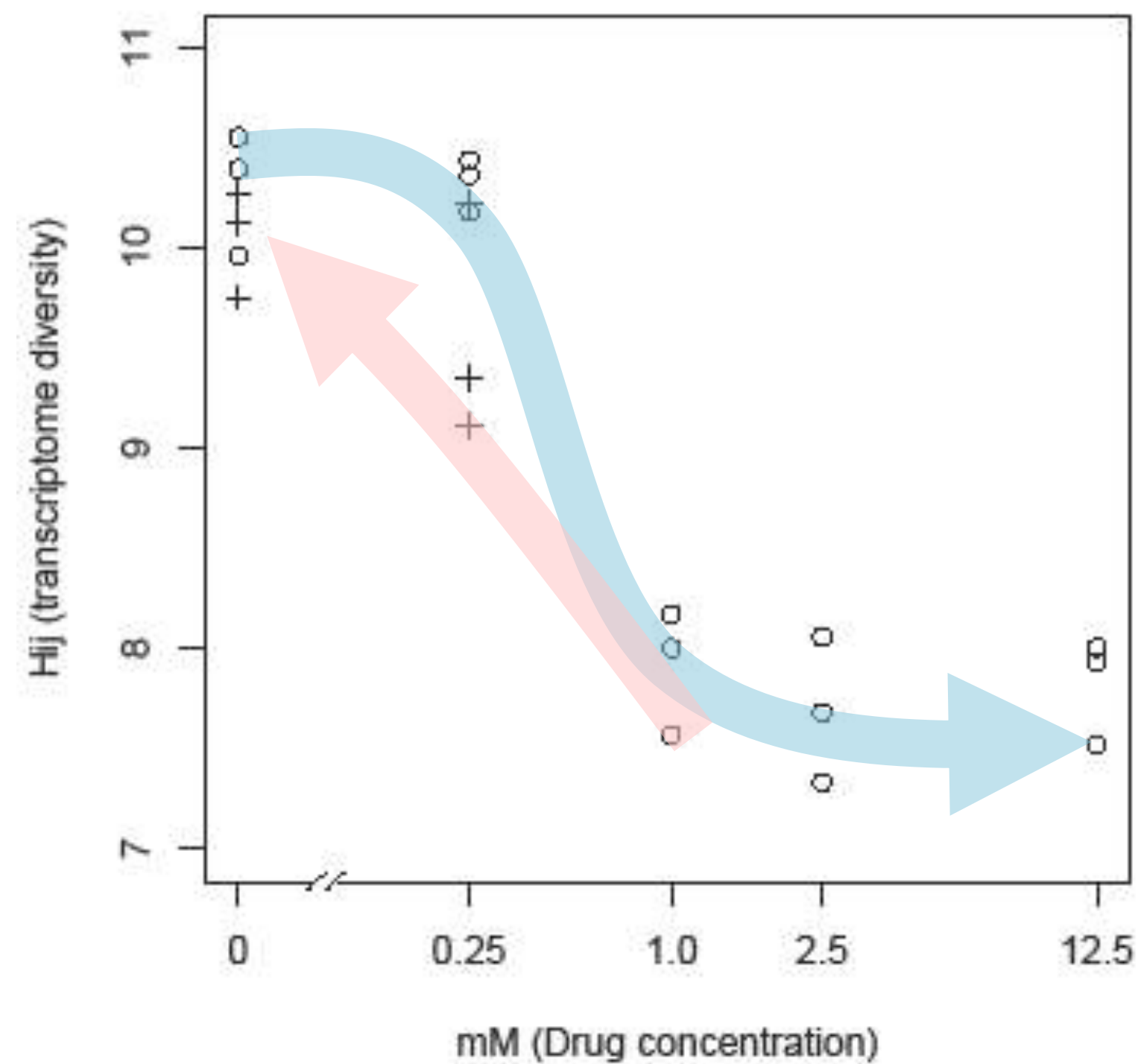


doi: <https://doi.org/10.1101/684803>

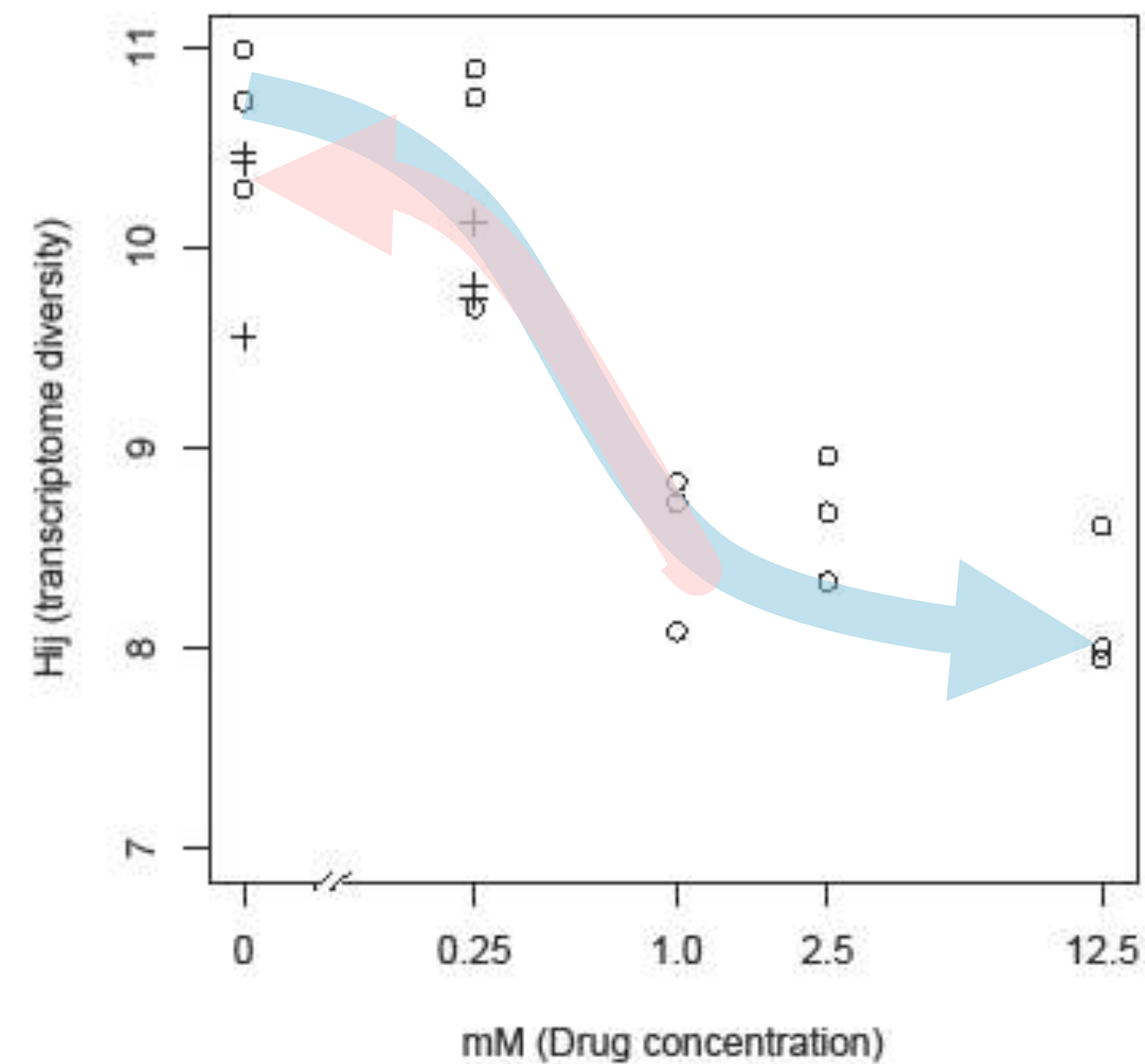
# 実験の予備検討（培地選び）につかえた

ヒステリシスの面積を安定的なmilieuの指標としてつかう

条件① 細胞に合う培地（確立された培地）

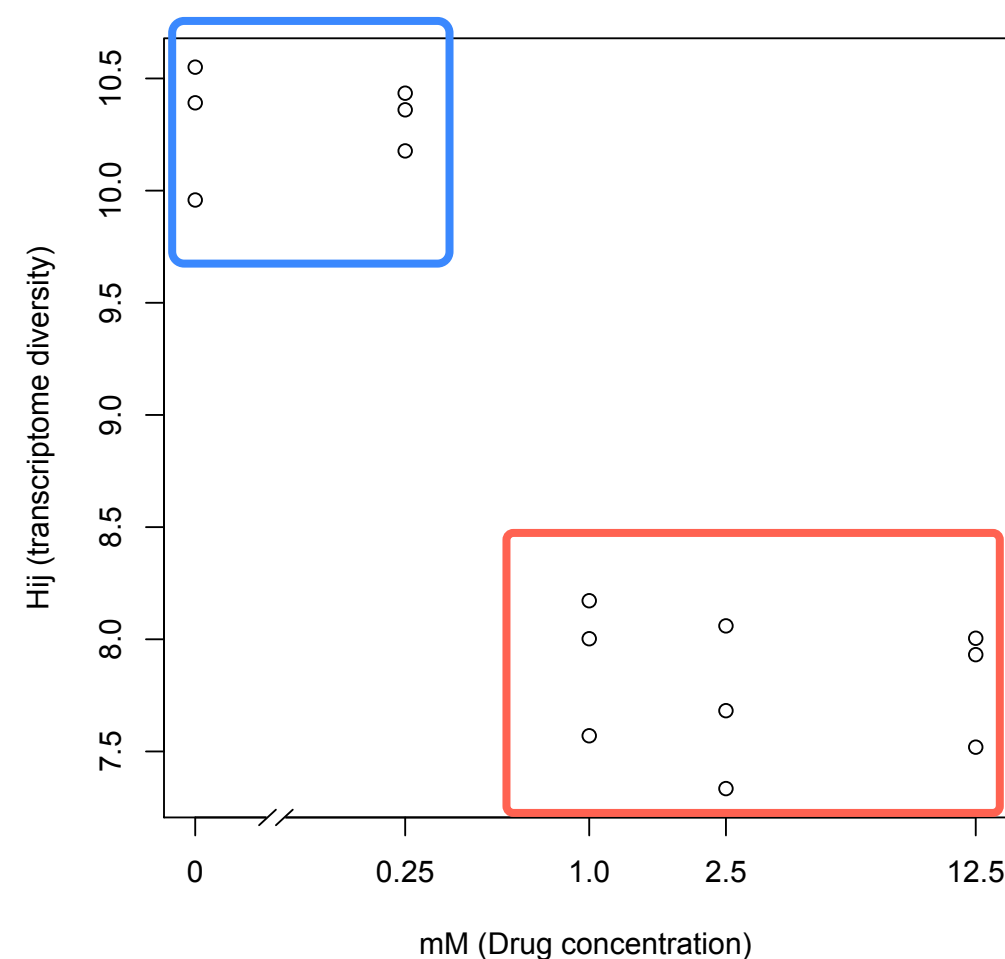


条件② 細胞に合わない培地

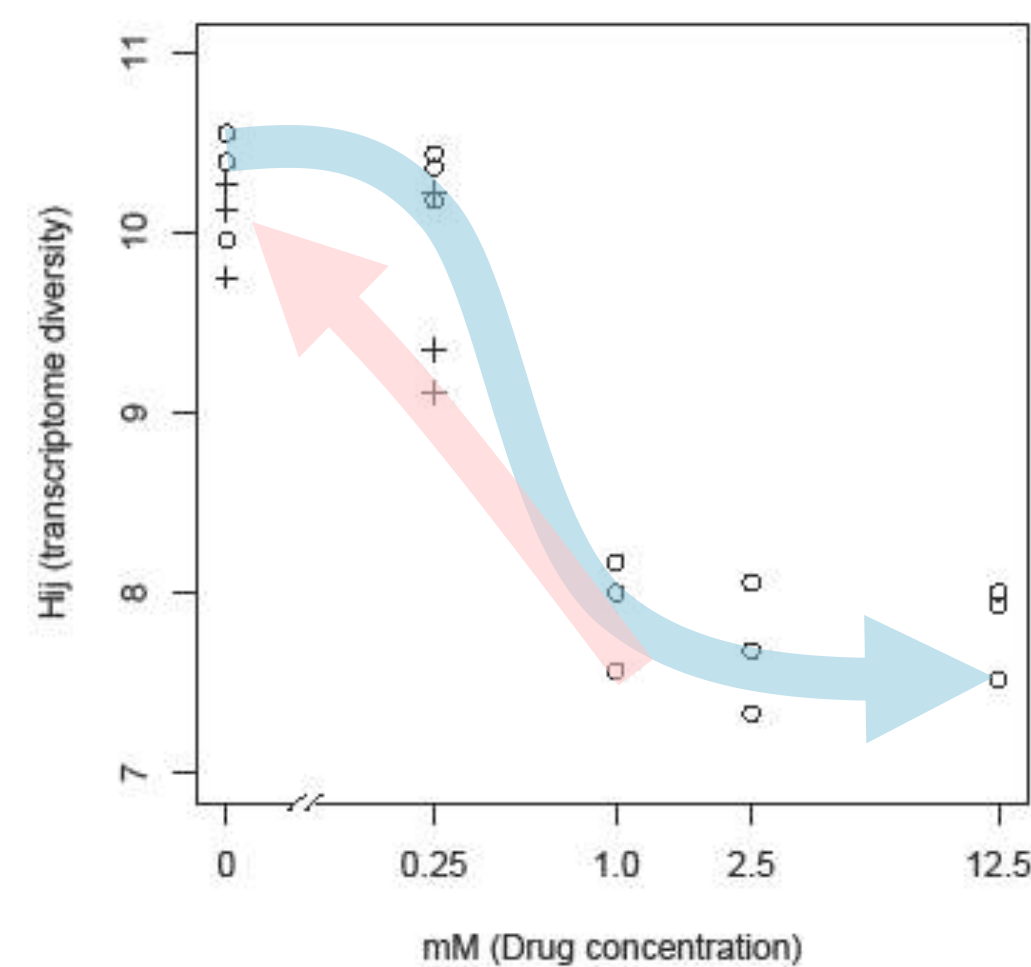


○ 濃度を上げて行った際の値  
× 濃度を下げて行った際の値

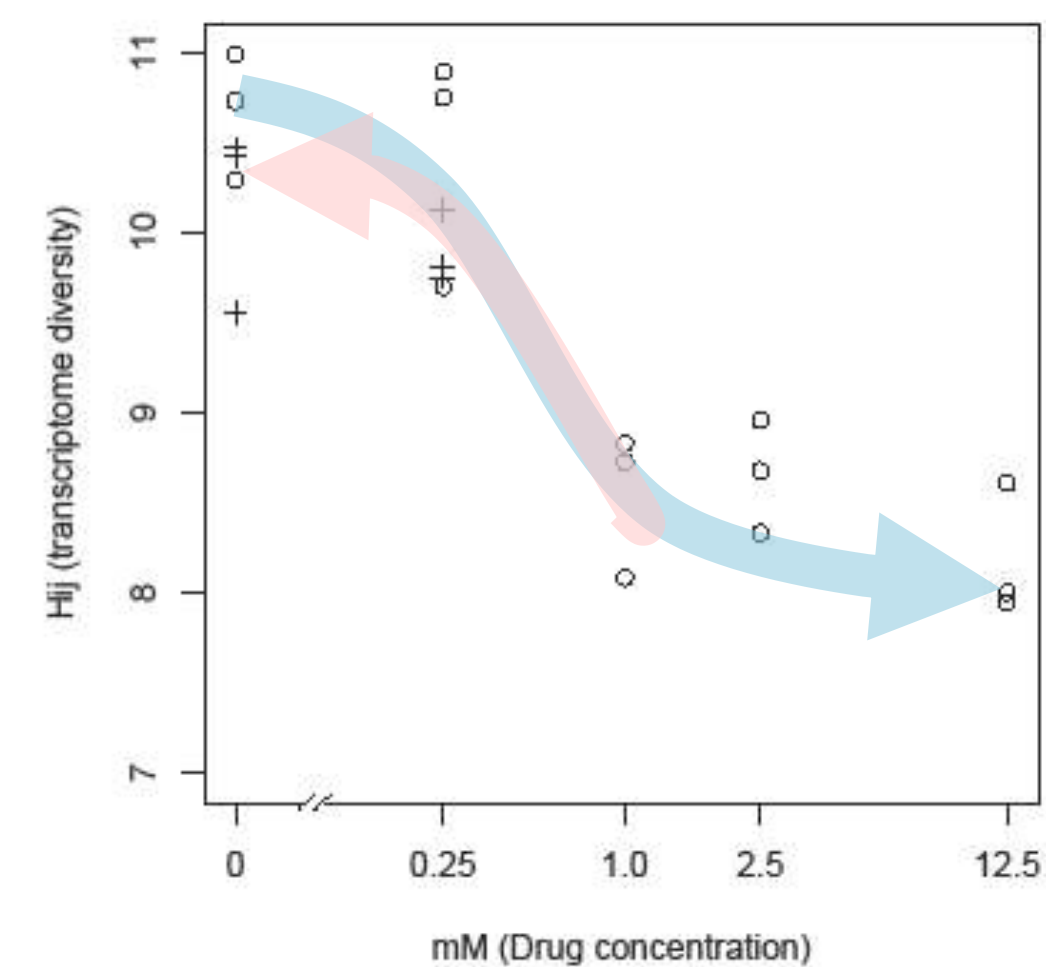
# どちらもあまり儲からなかった



条件① 細胞に合う培地 (確立された培地)



条件② 細胞に合わない培地

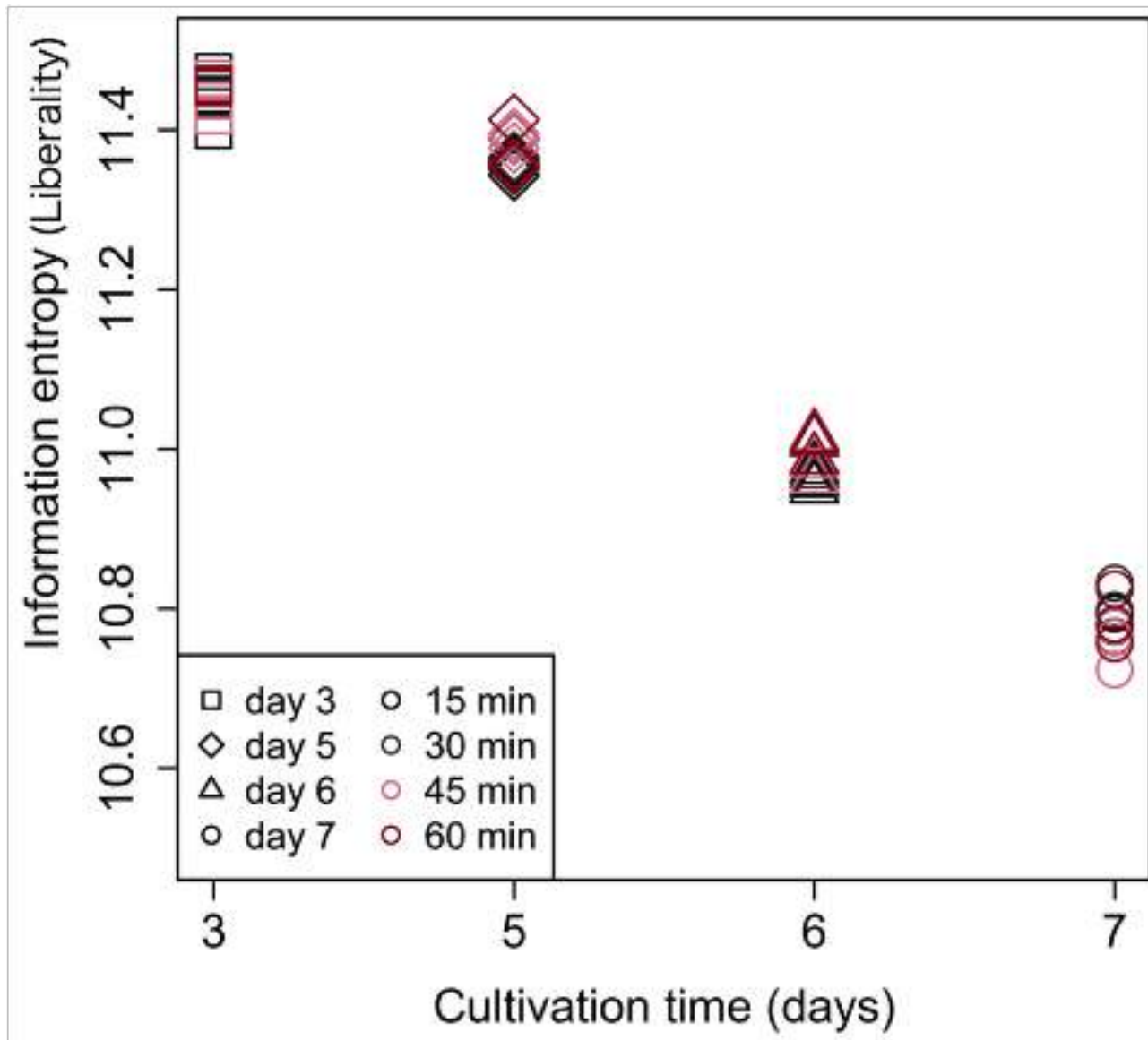


○ 濃度を上げて行った際の値  
× 濃度を下げて行った際の値

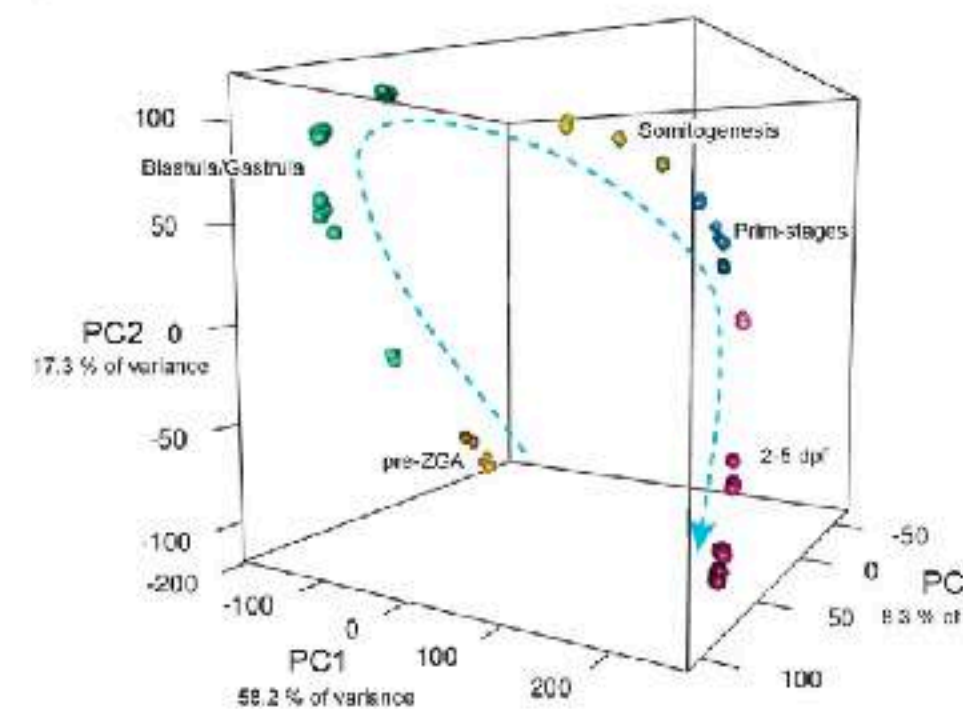
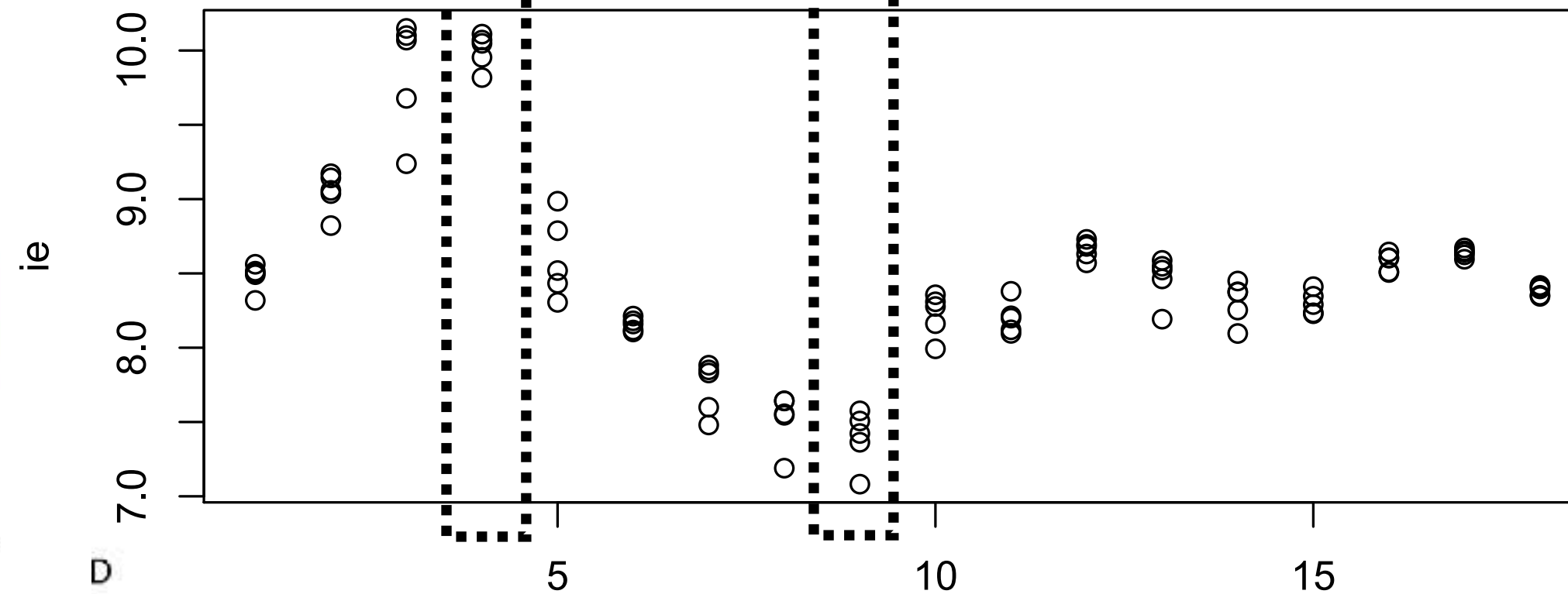
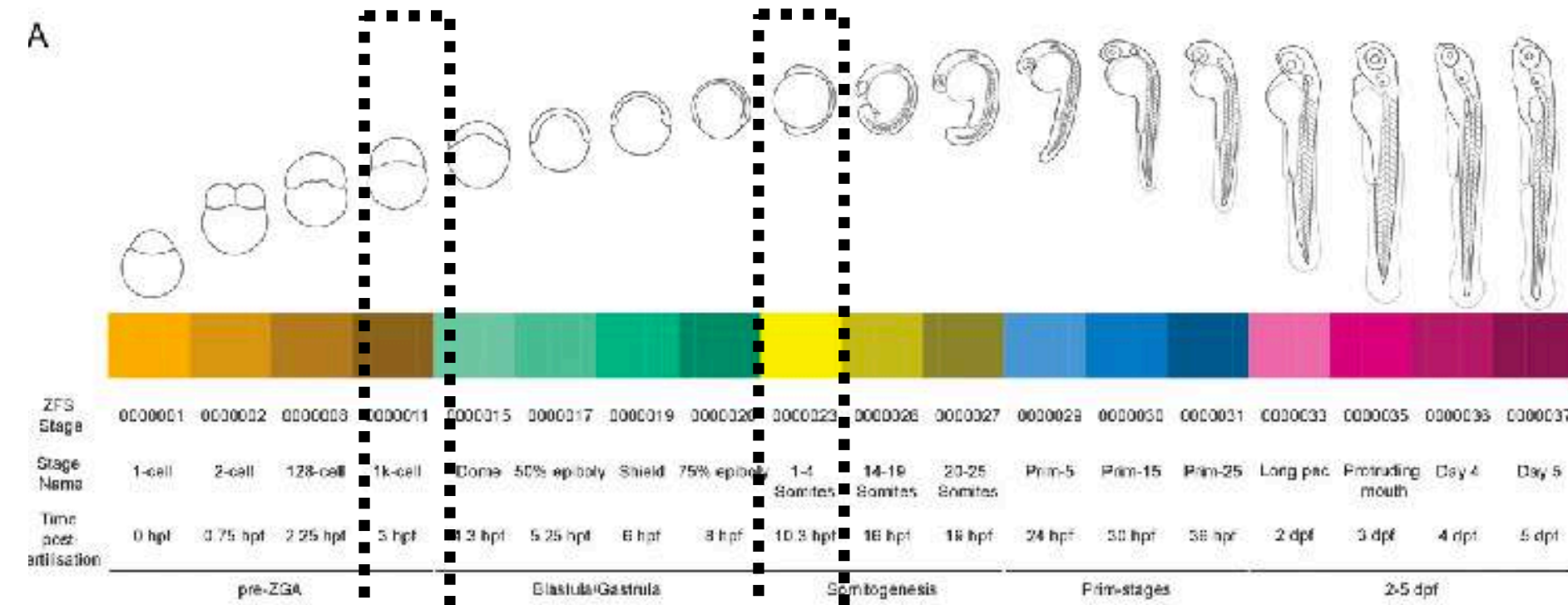
ピボットするしかない

# ただの経時変化追っかけは需要あった

## バイオ医薬品製造

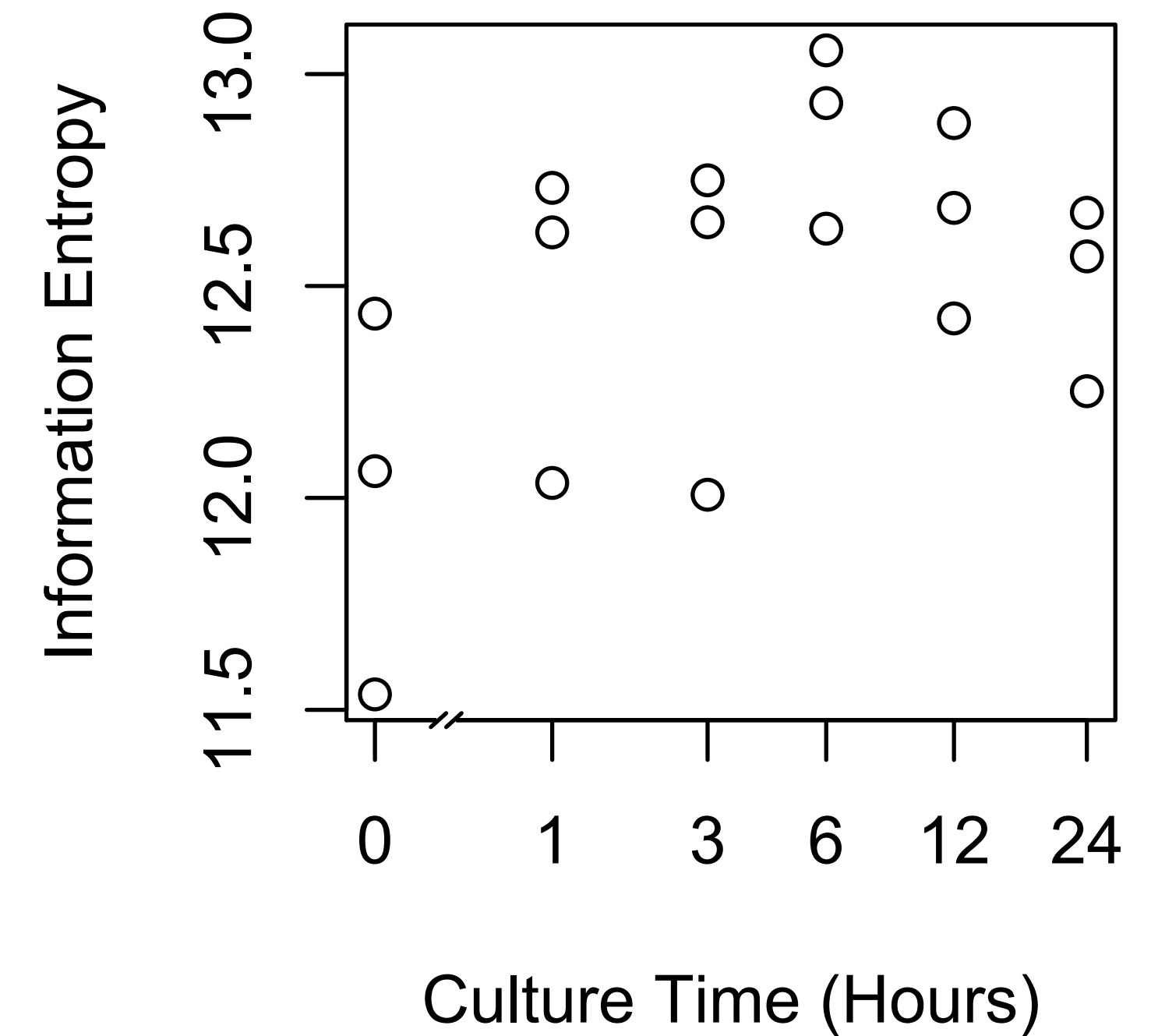


PMID: 34527435



## 胚発生

## 植物の再生

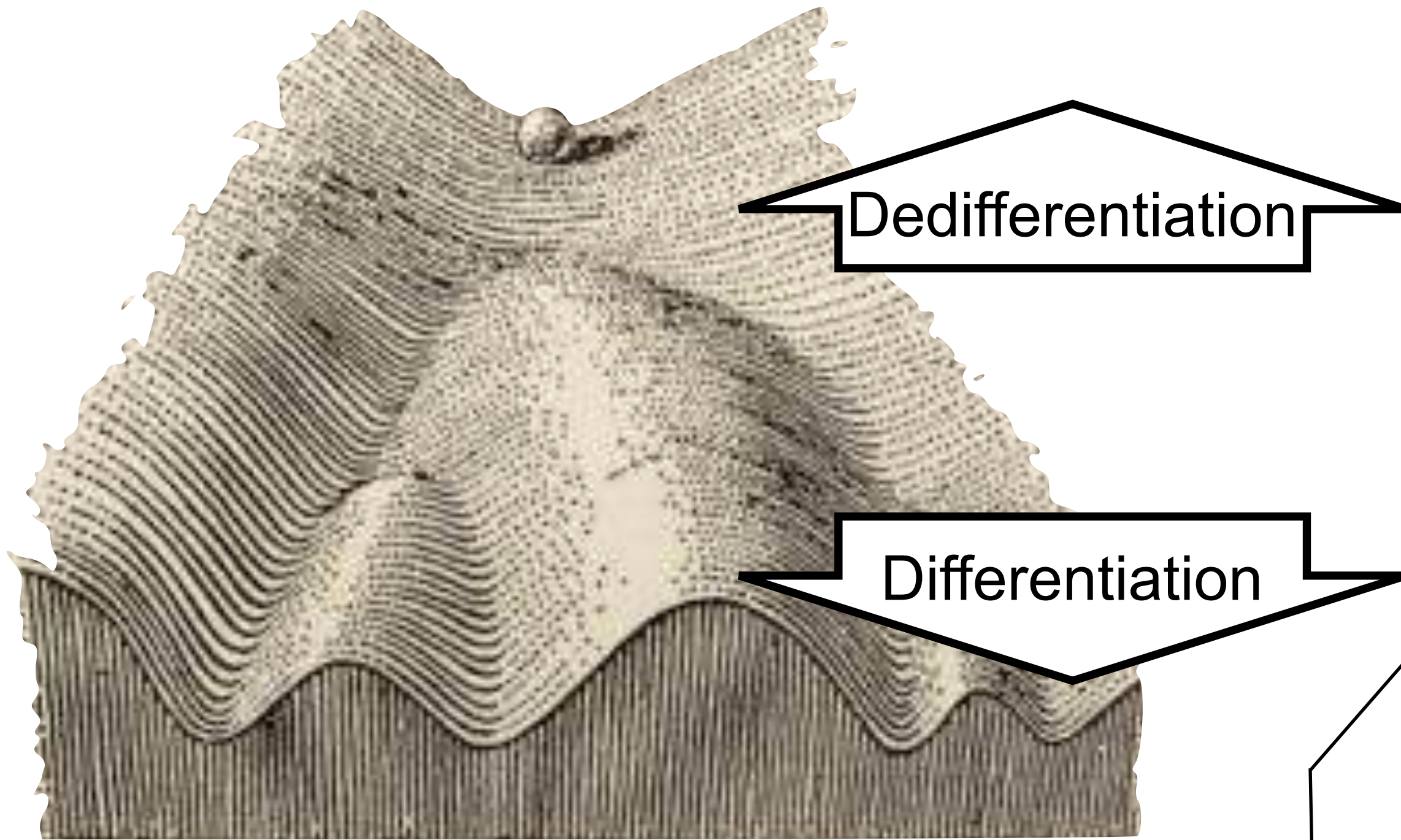


細菌でも酵母でもカビでも同じように使える



# 分化の度合いをLiberality と呼ぶことにした

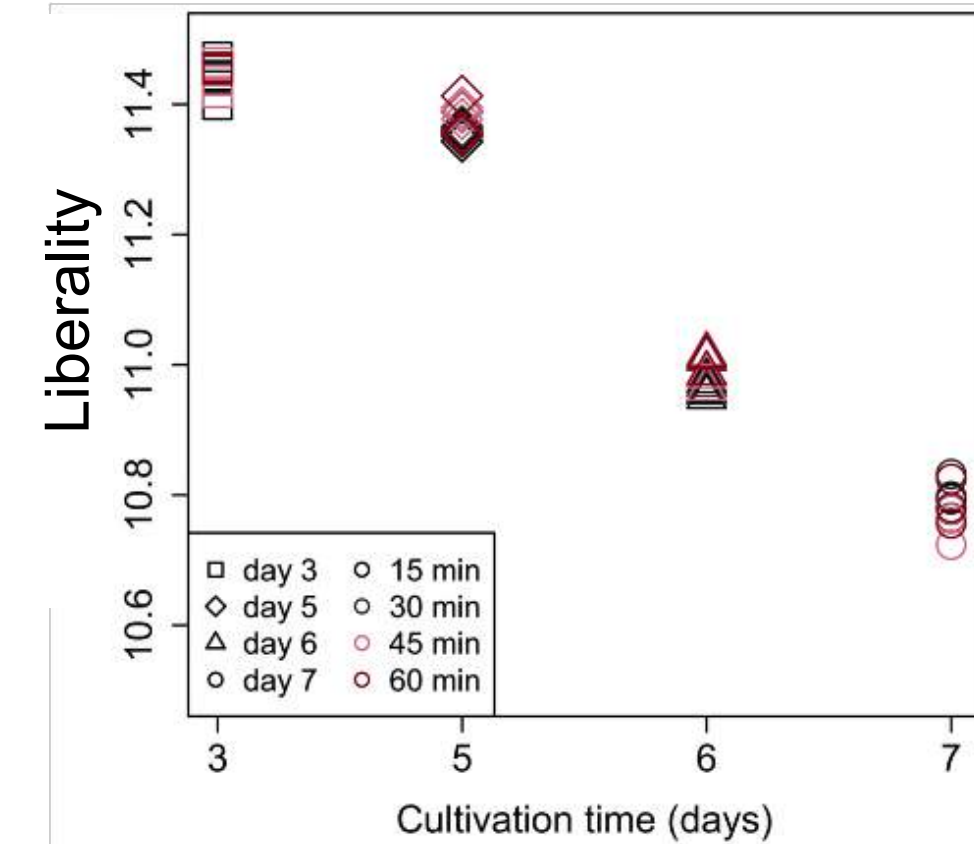
Liberality is quantified cellular dedifferentiation/differentiation states.



"En plaçant un tissu ou un organe dans un milieu ... permettant la survie (culture de tissus ou d'organes), on libère le tissu ou l'organe de toutes les stimulations ou inhibitions qu'exercent sur lui par la voie du milieu intérieur normal,..."

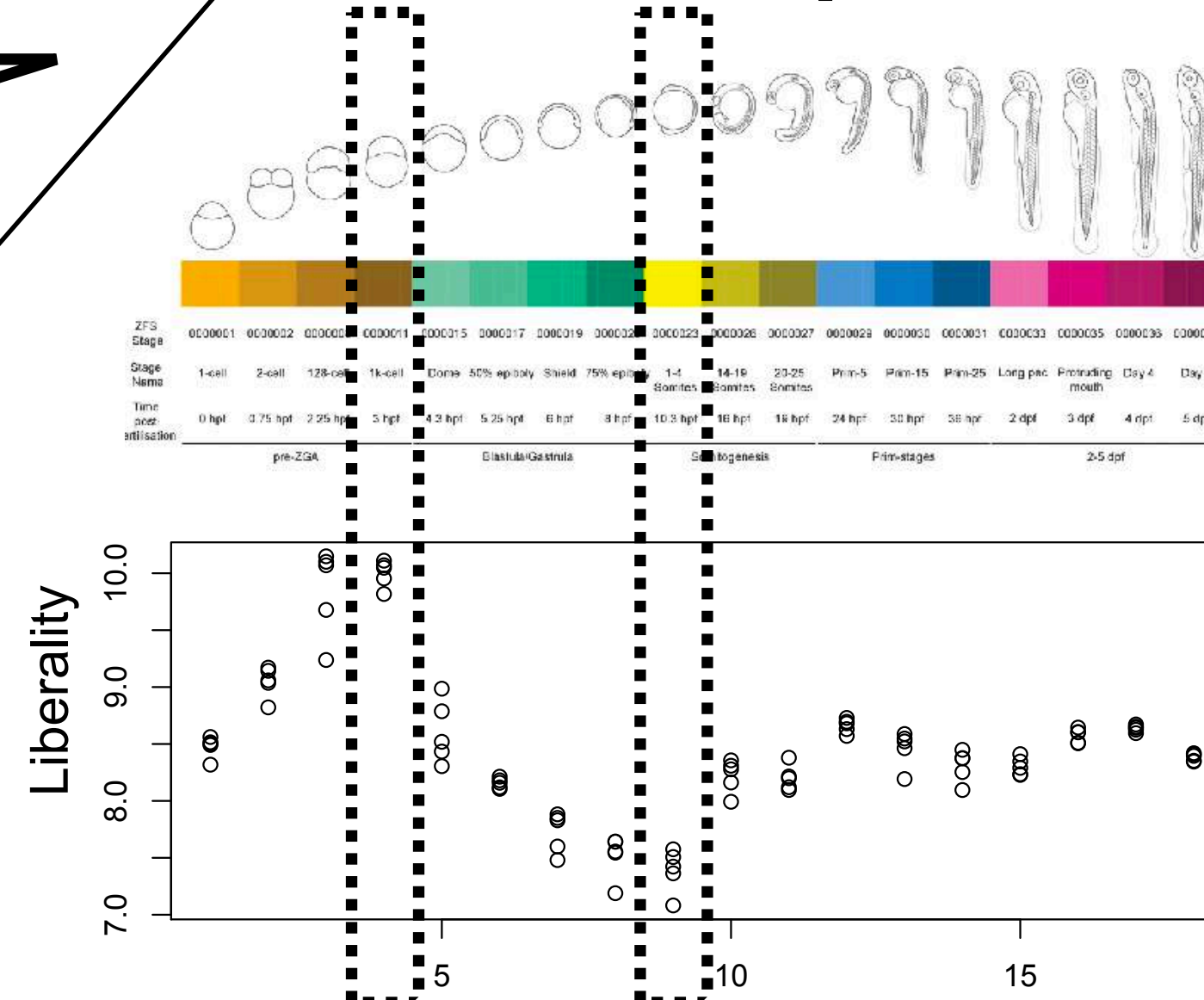
L'EXPÉRIMENTATION EN BIOLOGIE ANIMALE  
Georges CANGUILHEM

## in Animal Cell Culture

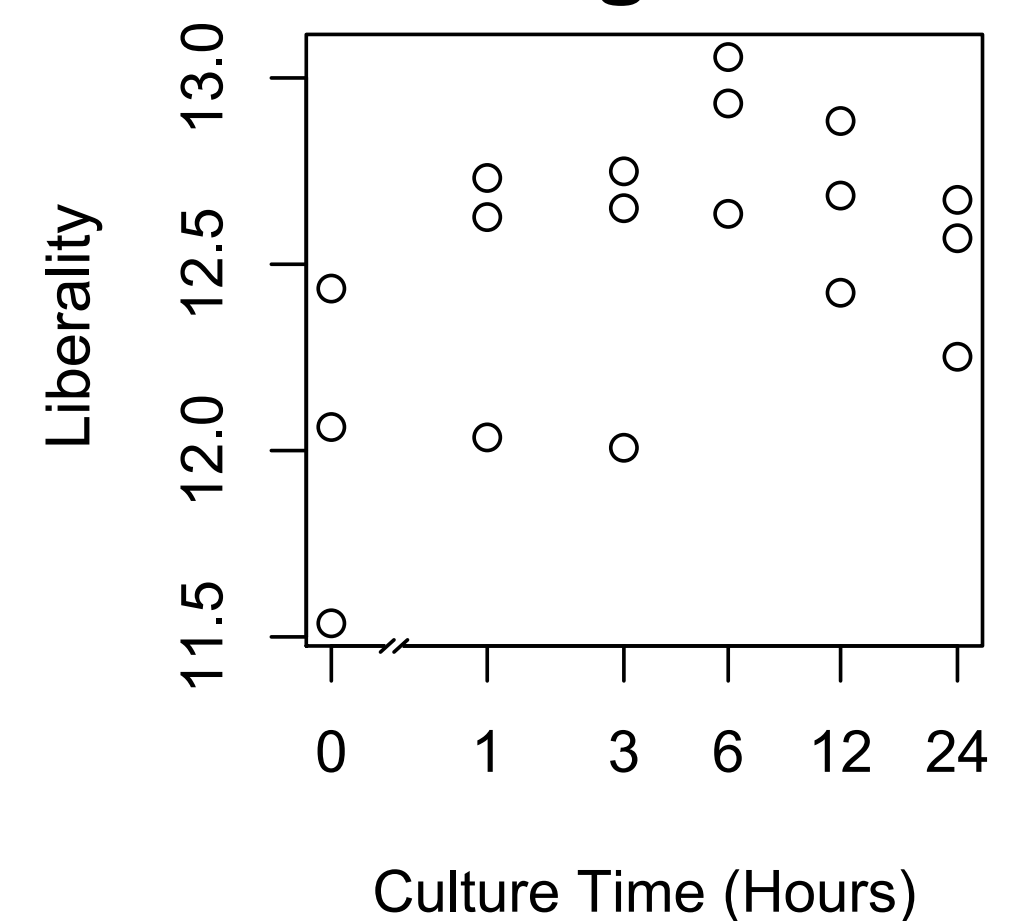


PMID: 34527435

## in Fish Development



## in Plant Regeneration



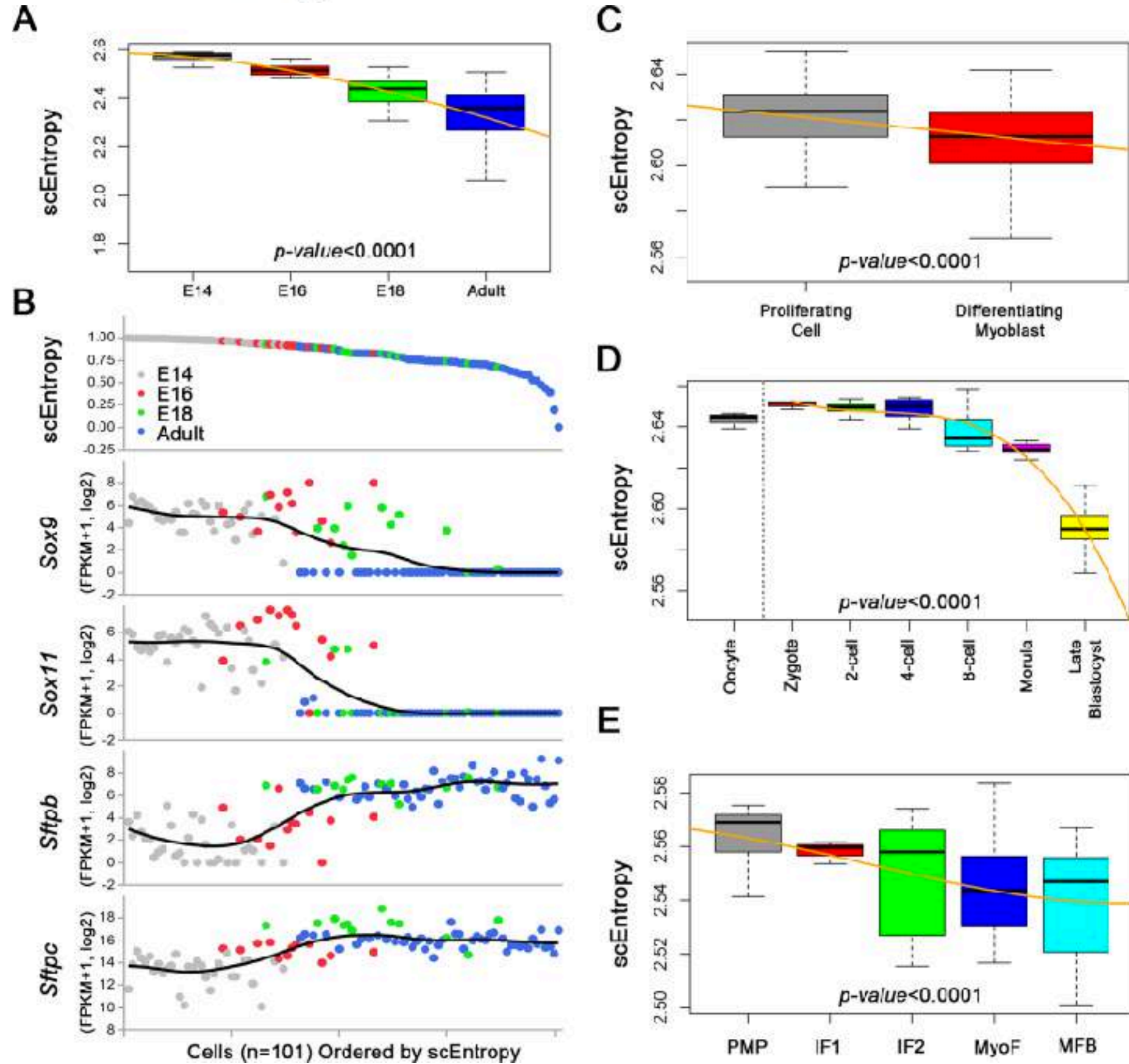
# 他グループによる追試

## Nucleic Acids Research

Nucleic Acids Res. 2017 Apr 20; 45(7): e64.  
Published online 2016 Dec 20. doi: 10.1093/nar/gkw1278

PMCID: PMC5397210  
PMID: 27998929

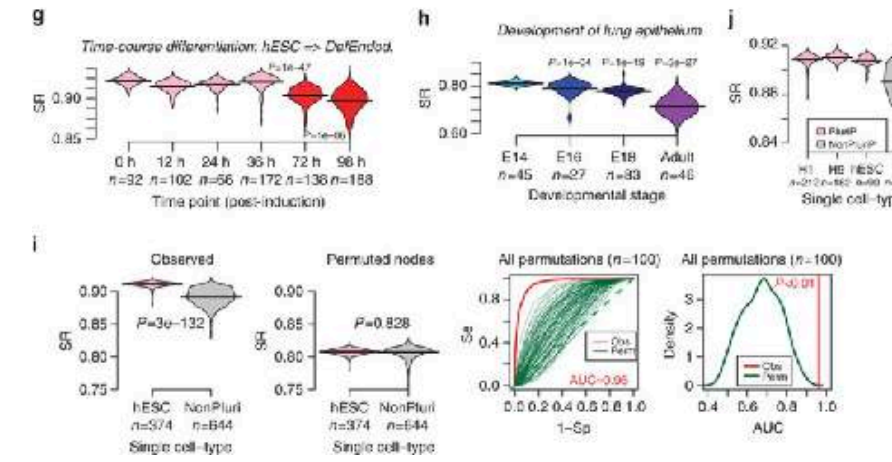
### SLICE: determining cell differentiation and lineage based on single cell entropy



Nat Commun. 2017; 8: 15599.  
Published online 2017 Jun 1. doi: 10.1038/ncomms15599

PMCID: PMC5461595  
PMID: 28569036

### Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome

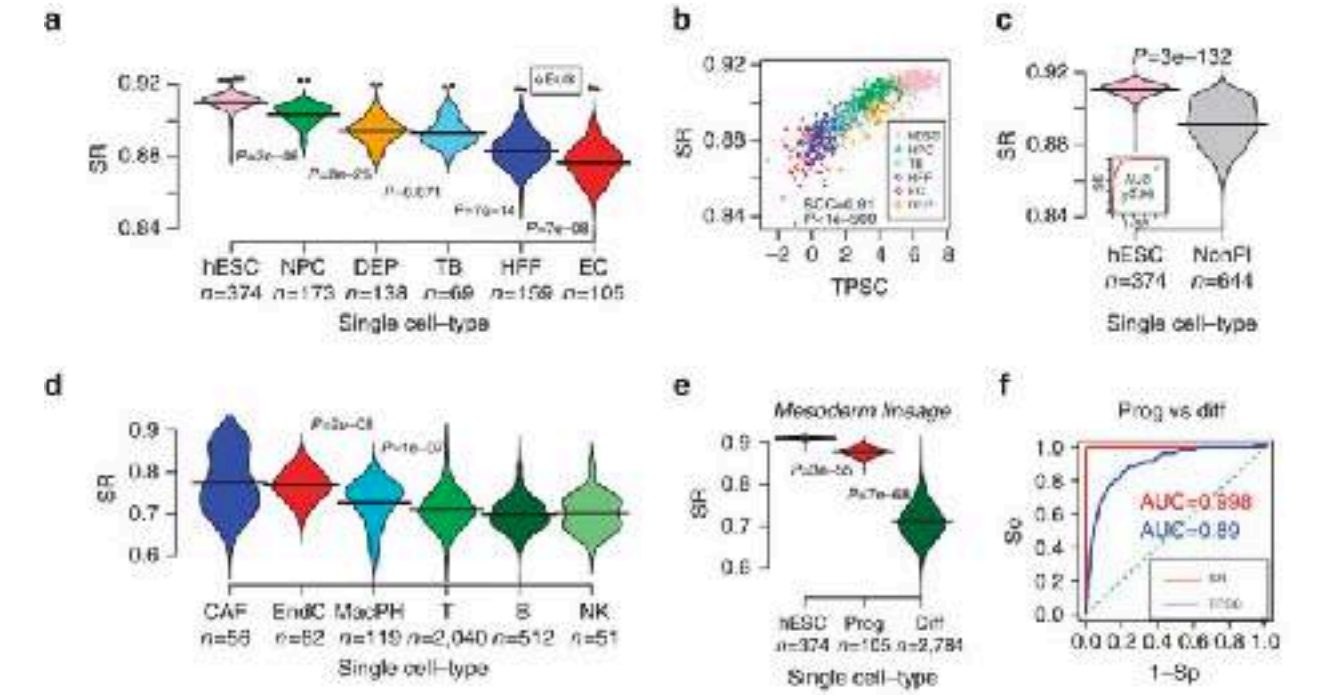
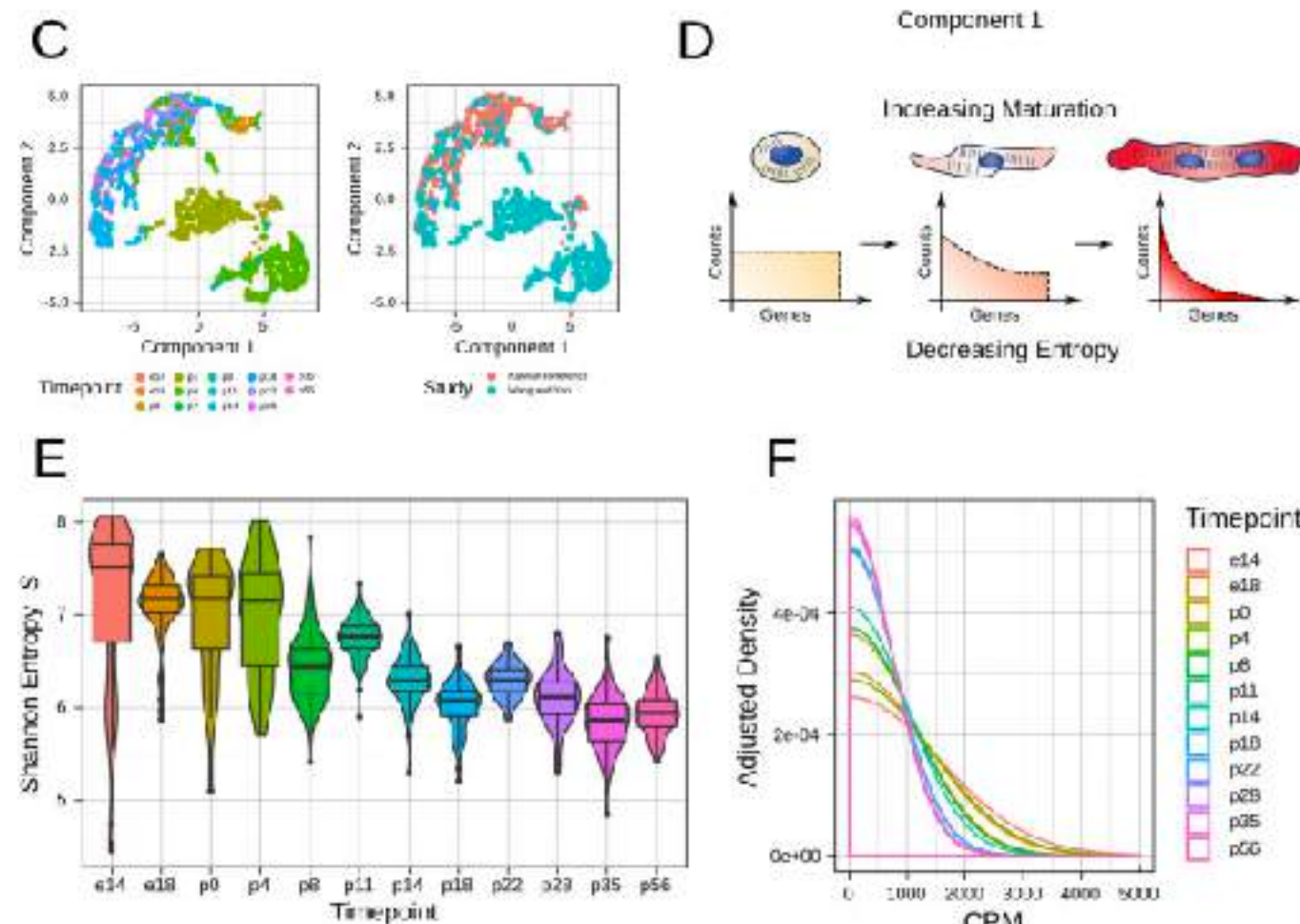


## PLOS COMPUTATIONAL BIOLOGY

PLoS Comput Biol. 2021 Sep; 17(9): e1009305.  
Published online 2021 Sep 17. doi: 10.1371/journal.pcbi.1009305

PMCID: PMC8448341  
PMID: 34534204

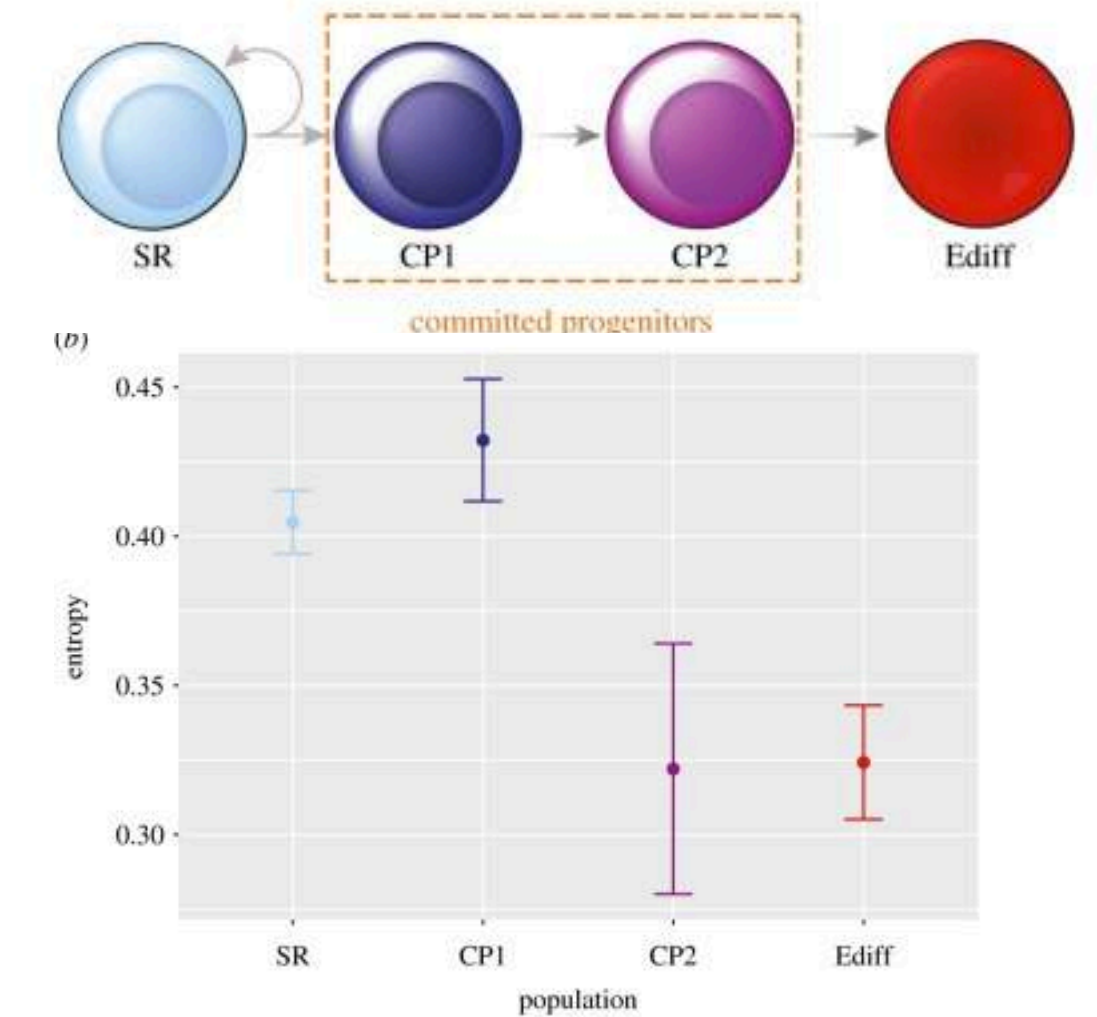
### Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level



Interface Focus. 2018 Dec 6; 8(6): 20180040.  
Published online 2018 Oct 19. doi: 10.1098/rsfs.2018.0040

PMCID: PMC6227807  
PMID: 30443337

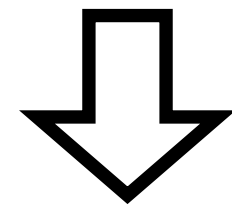
### Haematopoietic stem cells: entropic landscapes of differentiation



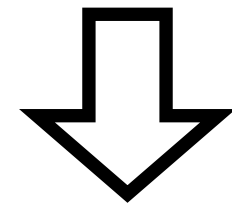
# より一般化を目指して

**OLD**

Cells

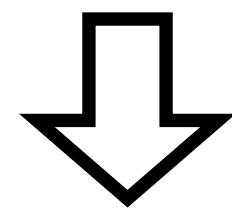


RNA-seq **Character data**

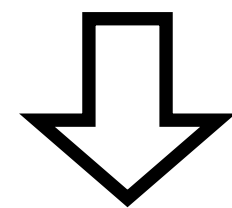


Mapping to Genomes **Character data**

Reference data



Tag Counting **Numerical data**

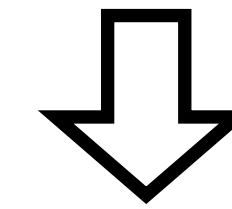


Measuring Transcriptome Entropy **Numerical data**

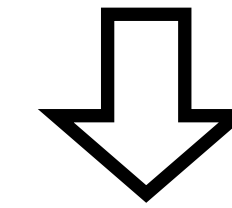
Shannon Entropy and LZ Complexity has similarity.

**NEW**

Cells



RNA-seq **Character data**



Measuring LZ Complexity **Numerical data**

Very easy. Very fast. For example...

DRR099453.fastq	1.73 GB	DRR099491.fastq	1.82 GB
DRR099453.fastq.bz2	244.5 MB	DRR099491.fastq.bz2	253 MB

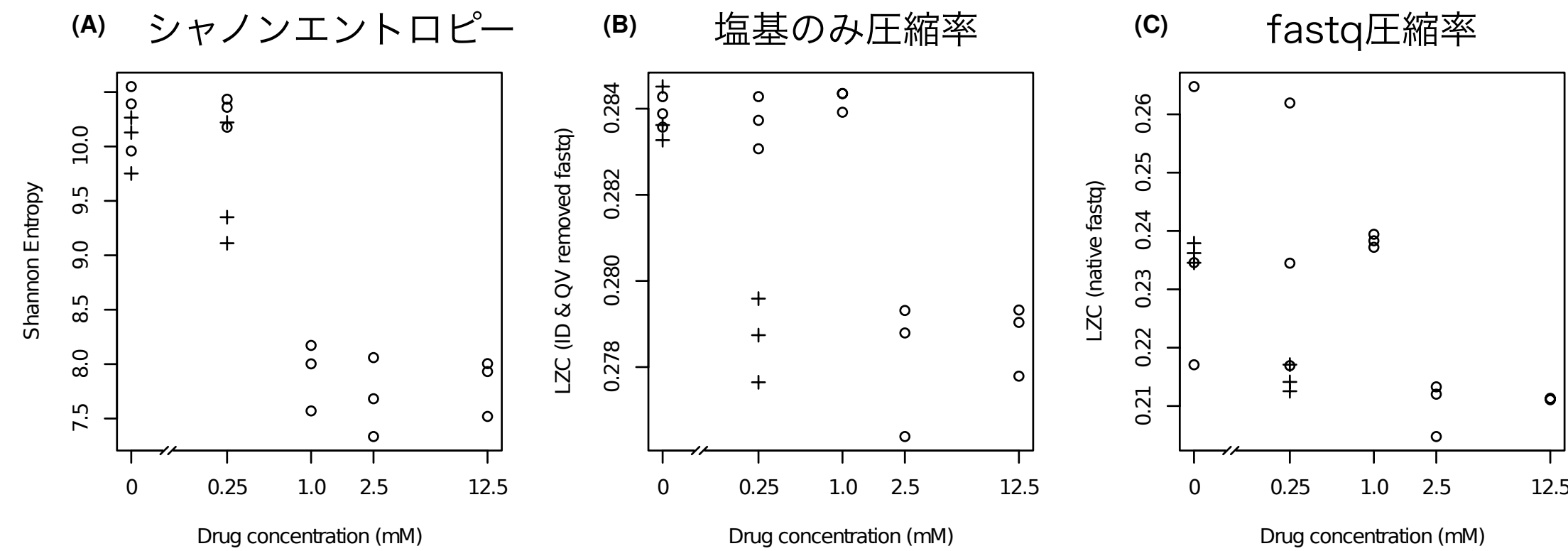
$$244.5\text{MB}/1.73\text{GB} = 0.1351$$

$$253\text{MB}/1.82\text{GB} = 0.1388$$

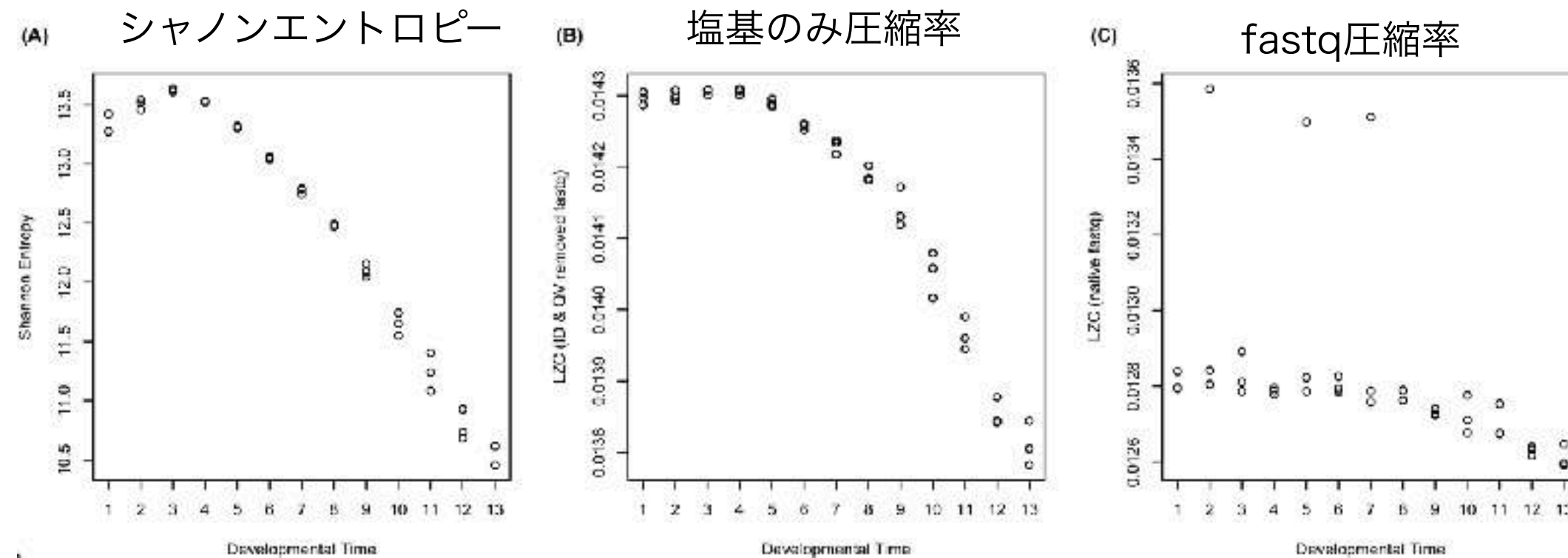
- Independent from Genome Data
- Independent from Mapping Tools
- Genome & Mapping Tools are always in Refinement = VARIABLE

# より一般化を目指して

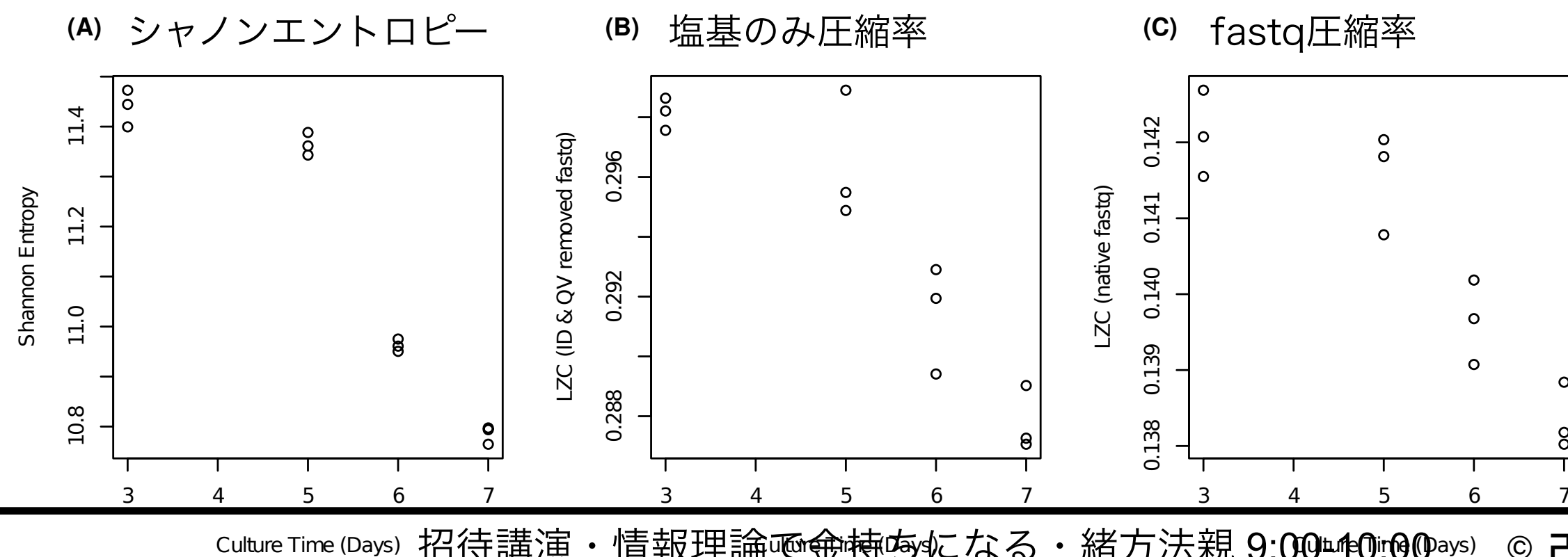
2010年  
カイコ



コムギ



2020年  
CHO

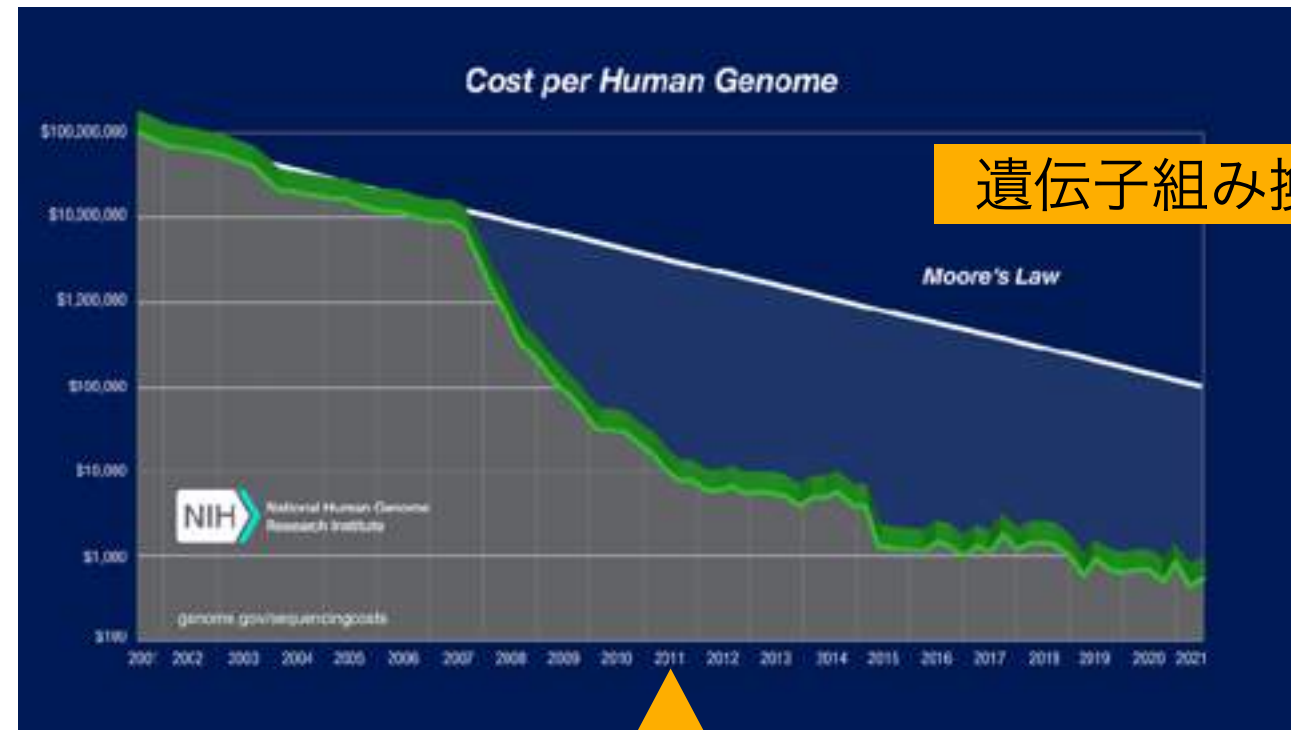


- シannonエントロピーとコルモゴロフ複雑性が似ているという話を読んだことがあった

- KCは測定できないが、LZ複雑性であれば圧縮で測定可能

- 測定装置の性能向上に伴っていつの間にかデータの加工なしにLZ複雑性からLiberalityが測れるようになっていた！

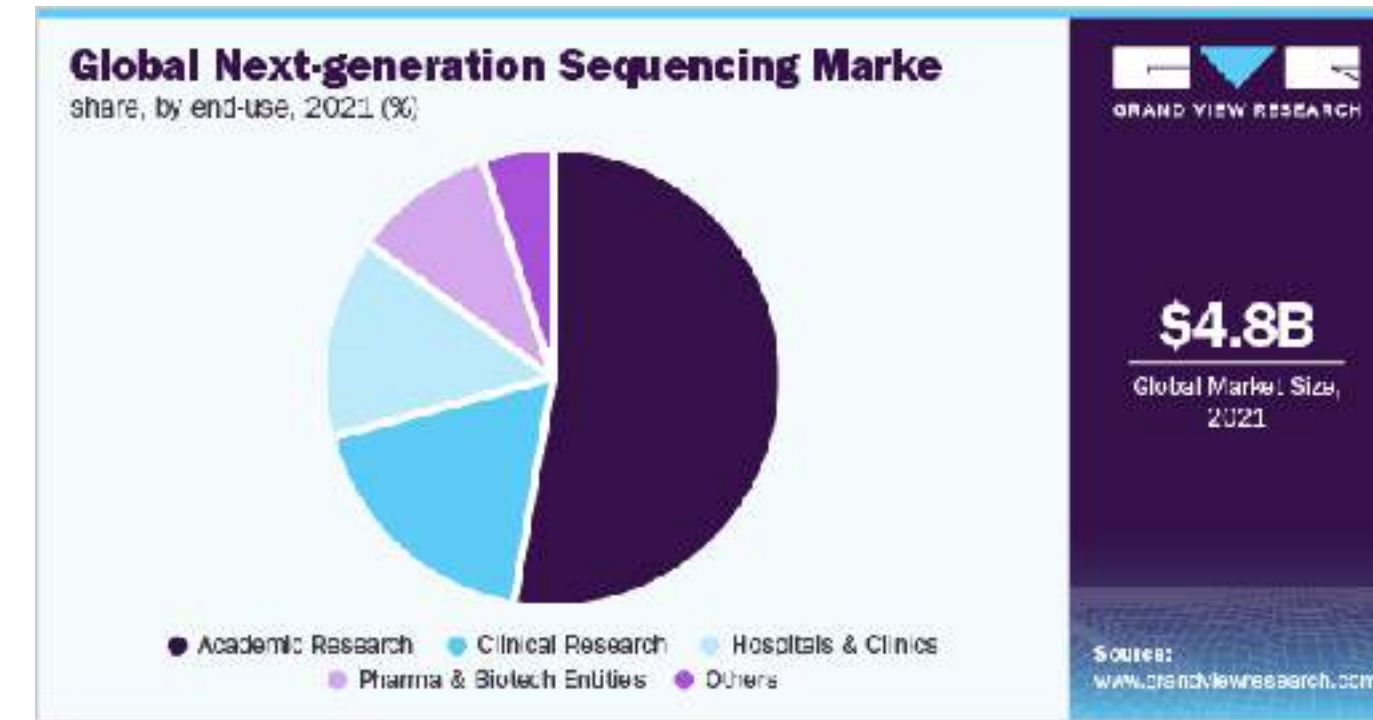
# どこから金を取るか (キヨキヨ)



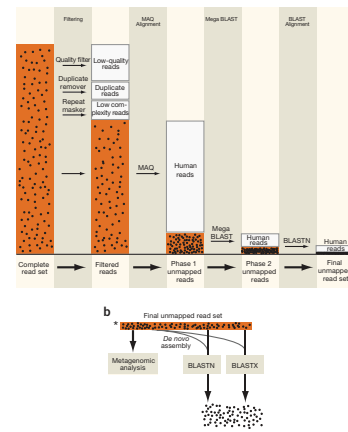
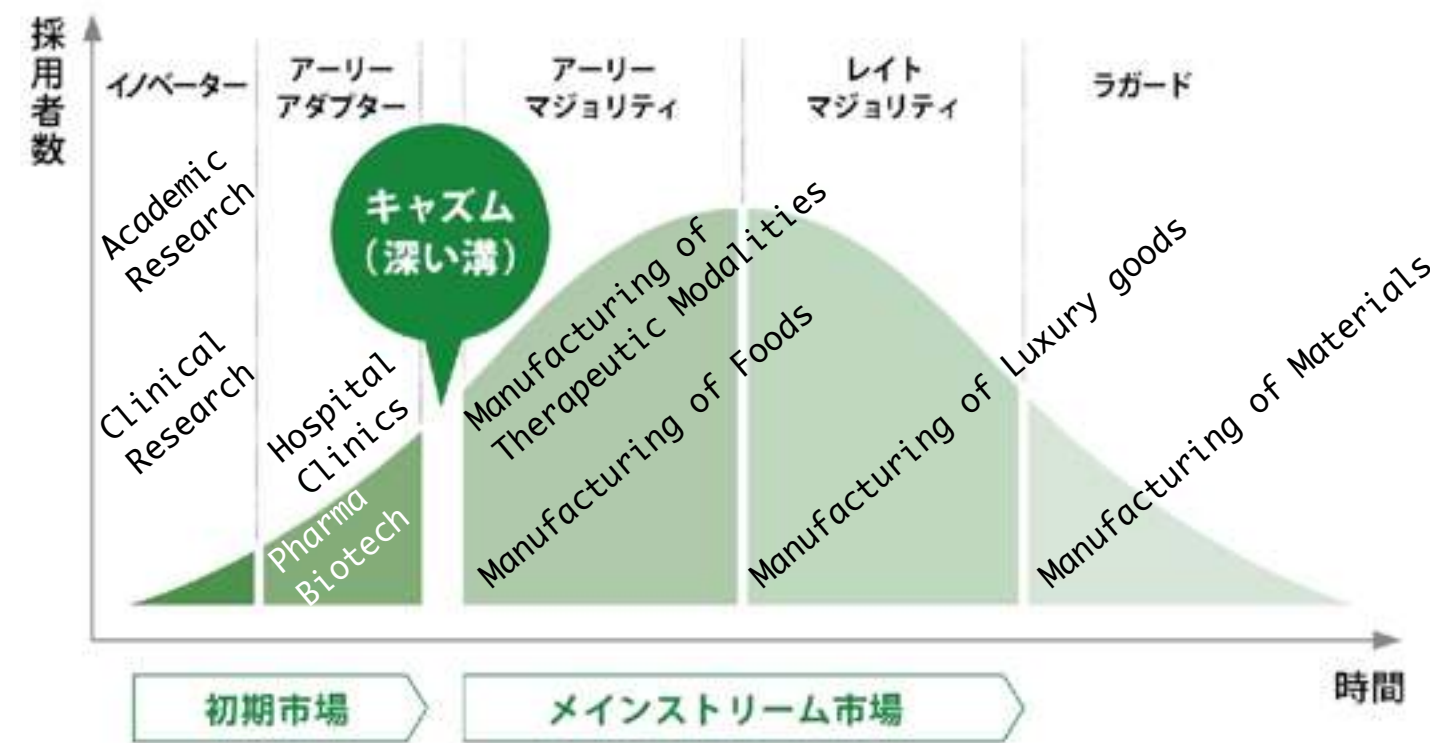
遺伝子組み換え技術の浸透と同じ

← 一方、データ解析コストはこの間あまり変わらなかった。

Nature Biotech. の患部シーケンス論文



遺伝子組み換え技術の浸透と同じ

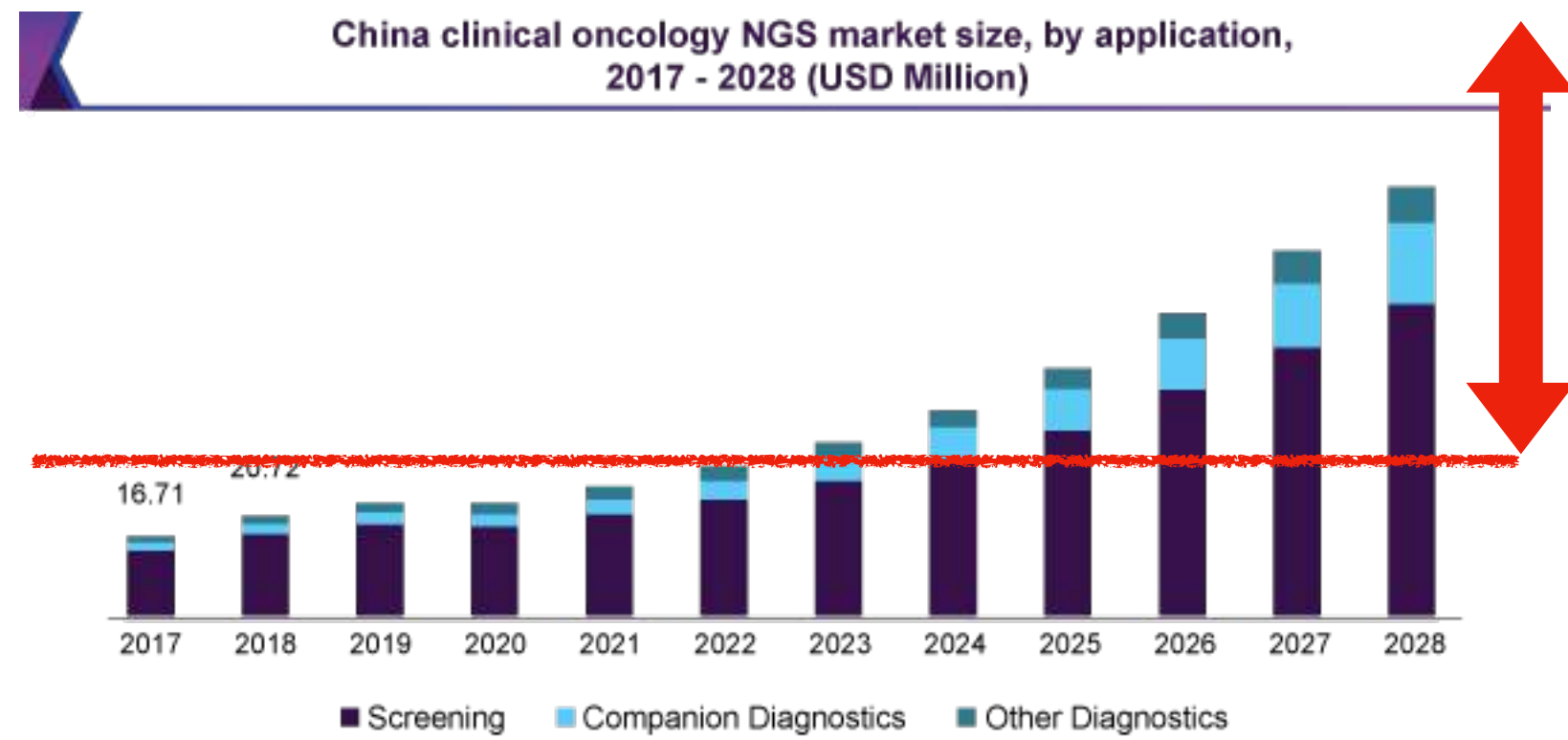


Technology Creeping ↓

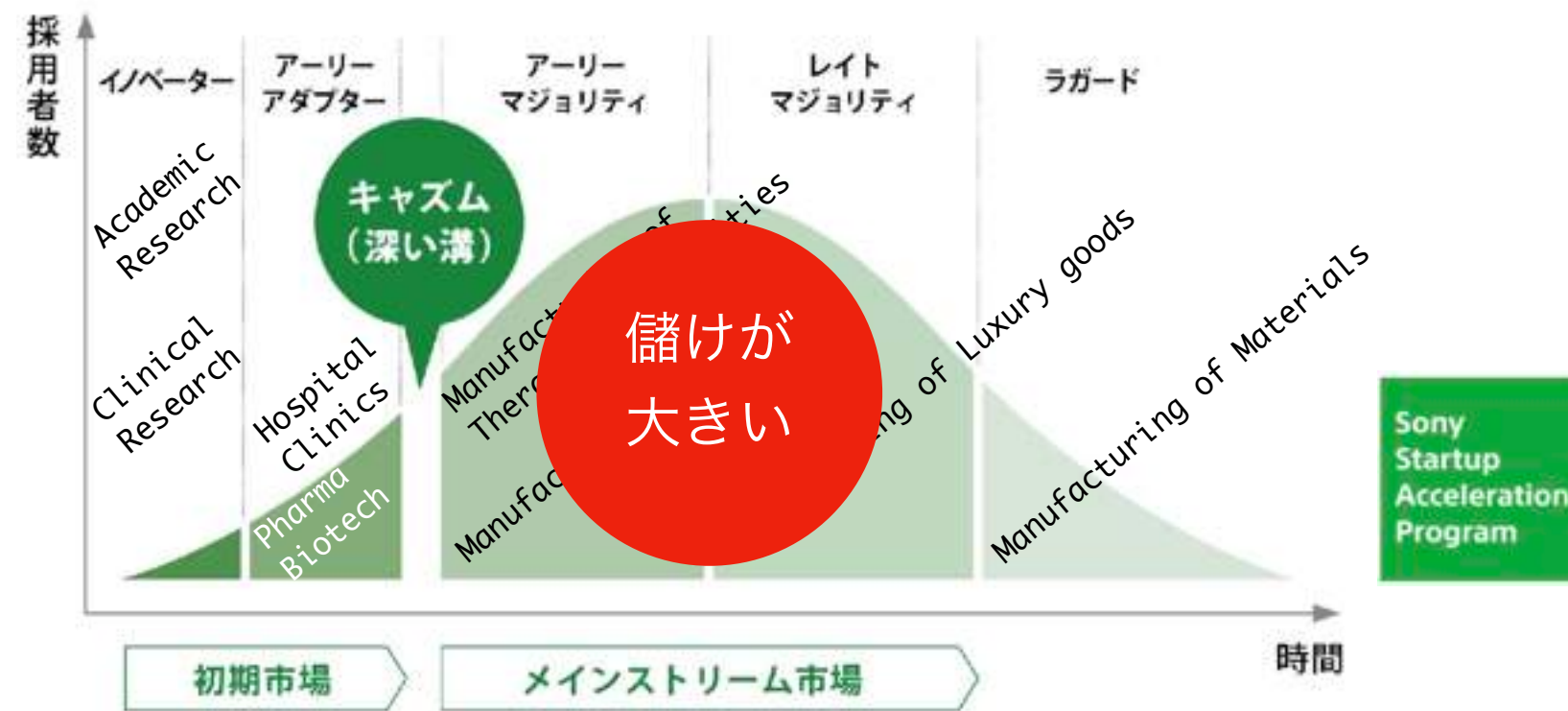
- Academic Research
- Clinical Research
- Hospital and Clinics
- Pharma and Biotech
- Manufacturing of Therapeutic Modalities (MABs, Cells etc.)
- Manufacturing of Luxury goods (Perfume, Supplements)
- Manufacturing of Foods, Ingredients
- Manufacturing of Materials (Polymer)

- Viral Safety
- HCP Safety
- Process Reproducibility
- Product Heterogeneity
- Cellular Heterogeneity

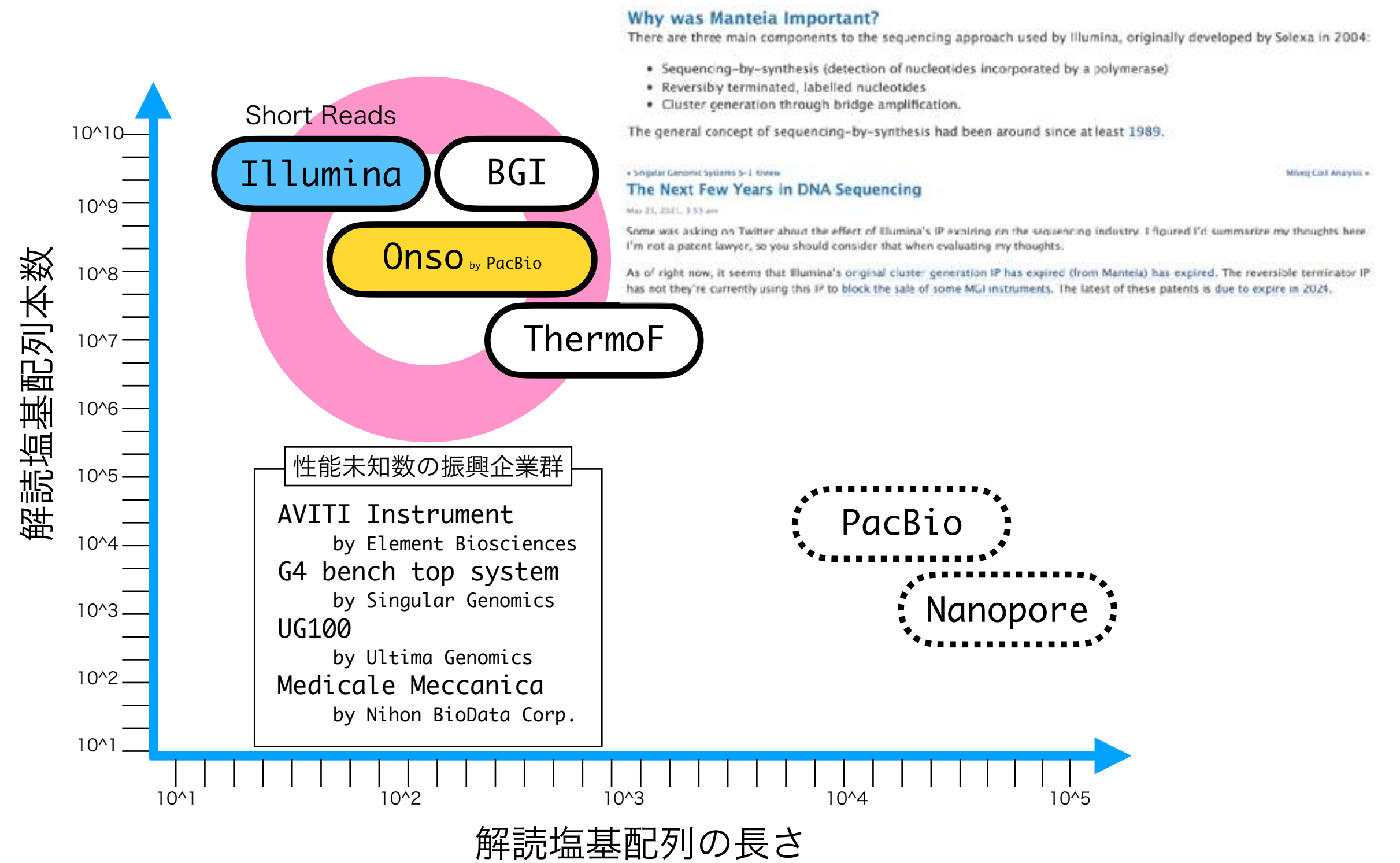
# ショートリード戦国時代



Source: www.grandviewresearch.com

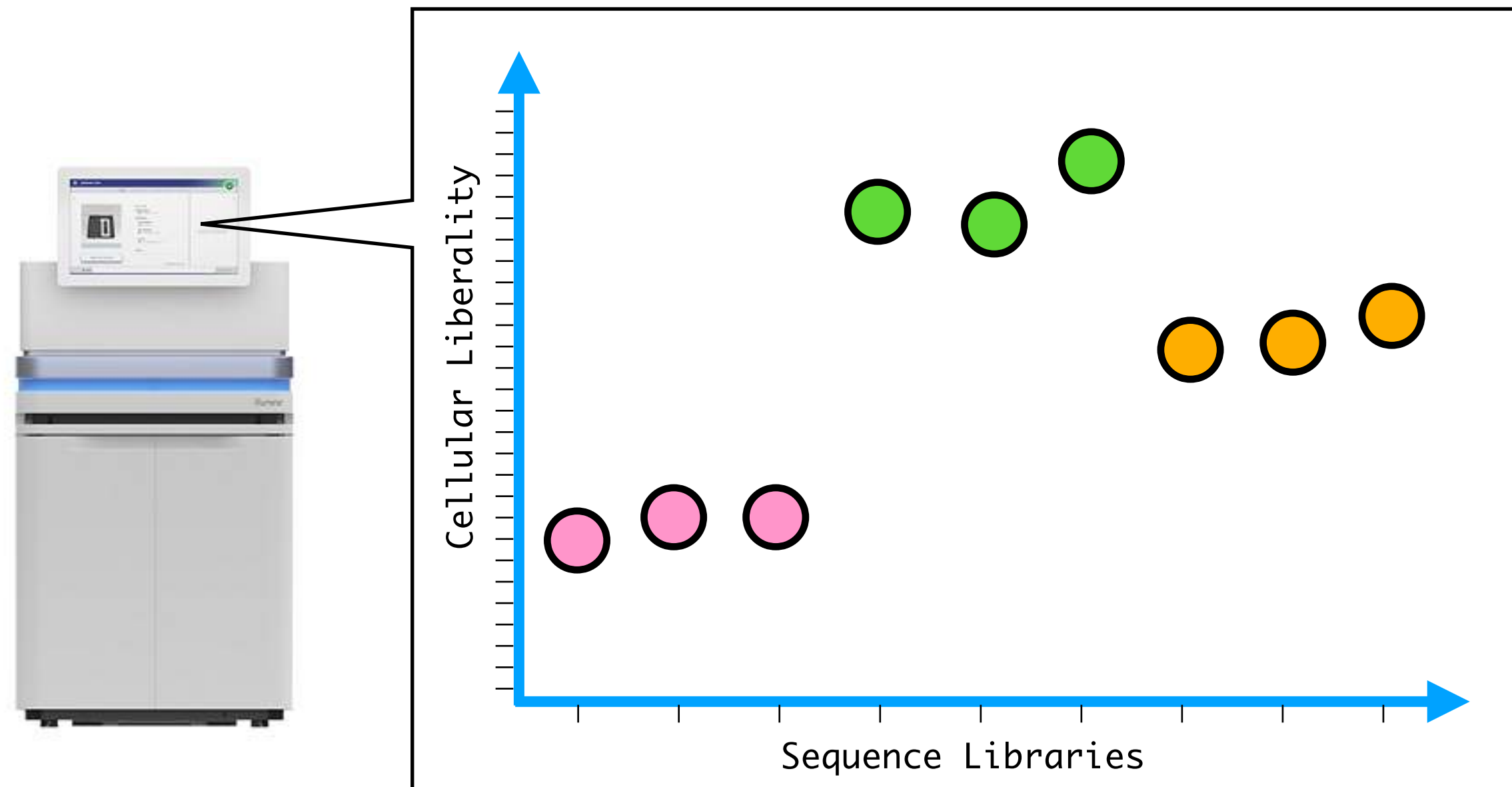


イルミナが持つMantia特許が2024年に時効を迎える。



# LZ複雑性を使った方法を装置に載せよう

細胞状態グラフをシーケンサー本体に表示できる



# 文脈散歩

"問題解決型リーダーたちはつねに使えるアイディアはないか、  
とほかの文脈を探し歩いているものなのだ。"

p.29



座長推薦💪

