

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局

(43) 国際公開日  
2024年3月7日(07.03.2024)



(10) 国際公開番号  
**WO 2024/048079 A1**

(51) 国際特許分類:  
C12N 5/10 (2006.01) C12P 21/08 (2006.01)  
C12N 5/071 (2010.01) G16B 40/00 (2019.01)  
C12P 1/00 (2006.01)

(21) 国際出願番号: PCT/JP2023/025263

(22) 国際出願日: 2023年7月7日(07.07.2023)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

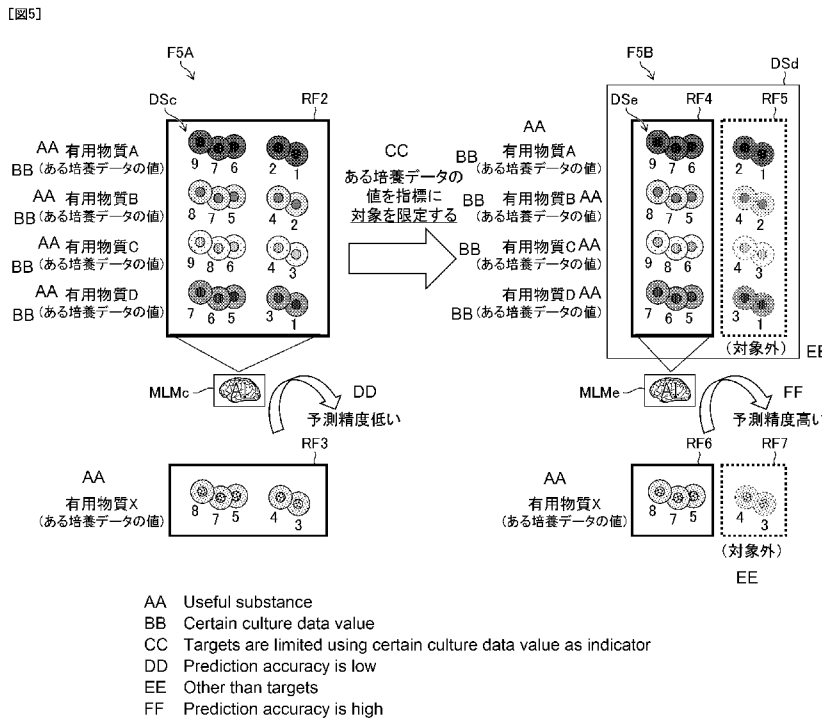
(30) 優先権データ:  
特願 2022-138789 2022年8月31日(31.08.2022) JP

(71) 出願人: 富士フイルム株式会社 (FUJIFILM CORPORATION) [JP/JP]; 〒1068620 東京都港区西麻布2丁目2番30号 Tokyo (JP).

(72) 発明者: 梅川 正夫 (UMEKAWA Masao); 〒2588577 神奈川県足柄上郡開成町牛島577番地 富士フイルム株式会社内 Kanagawa (JP). 鈴木 貴文(SUZUKI Takafumi); 〒2588577 神奈川県足柄上郡開成町牛島577番地 富士フイルム株式会社内 Kanagawa (JP). 佐藤 政寛(SATO Masahiro); 〒2588577 神奈川県足柄上郡開成町牛島577番地 富士フイルム株式会社内 Kanagawa (JP). 長瀬 雅也(NAGASE Masaya); 〒2588577 神奈川県足柄上郡開成町牛島577番地 富士フイルム株式会社内 Kanagawa (JP). 松浦 達也(MATSUURA Tatsuya); 〒2588577 神

(54) Title: METHOD FOR PREDICTING PRODUCTION STABILITY OF CLONE THAT PRODUCES USEFUL SUBSTANCE, INFORMATION PROCESSING DEVICE, PROGRAM, AND PREDICTION MODEL GENERATION METHOD

(54) 発明の名称: 有用物質を産生するクローンの産生安定性を予測する方法、情報処理装置、プログラムおよび予測モデル生成方法



(57) Abstract: Provided are: a method for predicting, at high accuracy and low cost, the production stability of a clone that produces a useful substance; an information processing device; a program; and a prediction model generation method. According to the present invention, at least one processor executes: acquiring at least one piece of clone culture data for

奈川県足柄上郡開成町牛島577番地 富士フィルム株式会社内 Kanagawa (JP). 村上裕太(MURAKAMI Yuta); 〒2588577 神奈川県足柄上郡開成町牛島577番地 富士フィルム株式会社内 Kanagawa (JP).

(74) 代理人: 松浦 憲三 (MATSUURA Kenzo); 〒1600023 東京都新宿区西新宿一丁目8番1号 新宿ビルディング5階 新都心国際特許事務所 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

a clone that produces a useful substance; analyzing the culture data to limit a clone to be predicted; and predicting the production stability of the useful substance by means of the clone to be predicted, by using data measured for the clone to be predicted. The production stability may be defined by the presence or absence of a change in the production amount of a useful substance at the start of culture and after culture for a predetermined period.

(57) 要約: 有用物質を産生するクローンの産生安定性を高精度かつ低コストに予測できる方法、情報処理装置、プログラムおよび予測モデル生成方法を提供する。1つ以上のプロセッサが、有用物質を産生するクローンについて1種類以上のクローンの培養データを取得することと、培養データを解析して予測対象のクローンを限定することと、予測対象のクローンについて測定されたデータを用いて、予測対象のクローンによる有用物質の産生安定性を予測することと、を実行する。産生安定性は、培養開始時と所定期間培養後における有用物質の産生量の変化の有無により定義されてよい。

## 明 細 書

発明の名称：

有用物質を産生するクローンの産生安定性を予測する方法、情報処理装置、プログラムおよび予測モデル生成方法

### 技術分野

[0001] 本開示は、有用物質を産生するクローンの産生安定性を予測する情報処理技術および機械学習技術に関する。

### 背景技術

[0002] 近年、従来の化学合成では作製が困難であった複雑な有用物質を細胞に作らせる製造法の産業利用が進んでいる。その一例がバイオ医薬品であり、世界の医薬品売り上げランキングTOP10では半数以上の品目数で、約3分の2の売上額を占めている。バイオ医薬品は、従来の低分子医薬品に比べて、複雑なタンパク質等を活用したものであり、人工的に化学合成するのは非常に難しい。そのため、バイオ医薬品の一例である抗体医薬品は、例えばCHO細胞 (Chinese Hamster Ovary cells) 等に所望のヒトタンパク質に対応する遺伝子を挿入し、細胞機能によって所望タンパク質を産生させ、これを抽出および精製して抗体医薬品を製造する生産方法が広く普及している。

[0003] 上述の様な細胞への遺伝子の挿入は、細かい制御が不可能なため、大量の細胞に一斉に遺伝子を挿入するのが一般的である。その際、生成される個々の細胞は遺伝子の挿入位置がランダムであることを踏まえ、医薬品としての抗体を安定化し、品質保証するため、多くの規制当局から、遺伝子挿入後に抗体産生を担う細胞が単一細胞由来であり、継代培養によってその性質が変化しないこと、所謂モノクロナリティが求められている。

[0004] そこで、遺伝子の挿入位置がランダムな個々の細胞から単一の細胞を抽出し、その単一細胞を増殖させて細胞クローン（以下、クローンという）を作成し、このクローンに抗体を産生させることによってモノクロナリティを担保している。本発明のクローンとは、遺伝子的に同一な細胞の集団、または

、その集団を構成する細胞を意味する。

[0005] 一方で、産業化においては、良質な抗体産生能を持つクローンが求められている。ここで、良質な抗体産生能とは、現時点において高い抗体産生能力があること、および、長期の培養期間においても抗体産生能力が安定していることである。前述の様に、遺伝子の挿入位置がランダムな個々の細胞から作成されるクローンは抗体産生能力にばらつきがあり、クローン毎に良質な抗体産生能かを判別する必要がある。現時点で抗体産生能力が高い高産生クローンであるか否かは2週間の規格試験によって判別可能であるが、長期の培養期間において抗体産生能力が安定しているか否かの産生安定性の判別については、実際に数か月間程度の長期培養による実験的な検証（安定性試験）が必須となっている。

[0006] このような背景の下、特許文献1では、現時点で得られるクローンの遺伝子発現データから数か月先のクローンの組換えタンパク質の産生安定性を予測する手法が提案されている。また、非特許文献1では、クローン開発の早期の段階において組換えタンパク質の安定発現を予測できるマーカー遺伝子を同定し、クローン開発の早期段階において組換えタンパク質の産生安定性を予測する方法が提案されている。

## 先行技術文献

### 特許文献

[0007] 特許文献1：国際公開第2016/075216号

### 非特許文献

[0008] 非特許文献1：Uros Jamnikar, Petra Nikolic, Ales Belic, Marjanca Blas, Dominik Gaser, Andrej Francky, Holger Laux, Andrej Blejec, Spela Baebler and Kristina Gruden, "Transcriptome study and identification of potential marker genes related to the stable expression of recombinant proteins in CHO clones" BMC Biotechnology volume 15, Article number 98 (2015).

## 発明の概要

### 発明が解決しようとする課題

[0009] しかし、特許文献1に記載の方法は、予測精度の点で十分とは言えない。また、多数のクローンに対する遺伝子解析などは一般に高額な費用を要するため、組換えタンパク質の産生安定性を予測することにより得られるコストダウン効果を、予測のための遺伝子解析などによるコストアップが減退させるといった問題もあった。コスト抑制のために、産生安定性の予測対象のクローン数を絞り込むことが考えられるが、そうすると予測対象中の産生安定性の高いクローン数も減ることになり、結果的に得られる産生安定性の高いクローン数が少なくなってしまうことになり、単純に予測対象のクローン数を絞り込むことも難しかった。

[0010] 本開示が解決しようとする第1の課題は、高い精度でクローンにおける有用物質の産生安定性を予測する手段を提供することである。第2の課題は、クローンにおける有用物質の産生安定性の予測コストを低減する手段を提供することである。

[0011] 本開示はこのような事情を鑑みてなされたものであり、有用物質を産生するクローンの産生安定性を高精度かつ低コストに予測することができる方法、情報処理装置、プログラムおよび予測モデル生成方法を提供することを目的とする。

### 課題を解決するための手段

[0012] 本開示の第1態様に係る方法は、有用物質を産生するクローンの産生安定性を予測する方法であって、1つ以上のプロセッサが、1種類以上のクローンの培養データを取得することと、培養データを解析して予測対象のクローンを限定することと、予測対象のクローンについて測定されたデータを用いて、予測対象のクローンによる有用物質の産生安定性を予測することと、を実行する。

[0013] 第1態様によれば、培養データから得られる情報を基に予測対象を限定して産生安定性の予測を行うため、対象を限定しない場合と比較して、高い精

度で産生安定性を予測することが可能になる。また、予測対象であるクローンに限定して予測に必要なデータの取得を行えばよいため、コスト抑制が可能である。

[0014] 予測する産生安定性は、実際には数か月間の長期培養によって実験的に検証されている産生安定性と同様に、数か月先の将来のクローンの状態を表すものであってよい。例えば、長期培養後も初期の産生量が維持されているか否かという観点から産生安定性が評価されてよい。第1態様によれば、長期培養が必要な安定性試験の結果を高精度かつ低コストで予測できる。

[0015] 本開示の第2態様に係る方法は、第1態様に係る方法において、産生安定性は、培養開始時と所定期間培養後における有用物質の産生量の変化の有無により定義される構成であってもよい。

[0016] 本開示の第3態様に係る方法は、第1態様または第2態様に係る方法において、1つ以上のプロセッサが、培養データから得られる指標と、指標に関する閾値とを設定することを含み、指標の値と閾値とに基づき予測対象を限定する構成であってもよい。

[0017] 本開示の第4態様に係る方法は、第3態様に係る方法において、閾値は、産生安定性の予測精度が予測対象を限定しない場合よりも高くなるように調整される構成であってもよい。

[0018] 本開示の第5態様に係る方法は、第3態様または第4態様に係る方法において、閾値は、指標の値についての順位を用いて定義される構成であってもよい。なお、「順位」は、複数のクローンについての指標の値を降順に並べた場合の順位と、昇順に並べた場合の順位とがあり得る。例えば、閾値は、複数のクローンを含む集団における相対順位の上位40%などのように定義されてよい。

[0019] 本開示の第6態様に係る方法は、第3態様から第5態様のいずれか一態様に係る方法において、予測対象は、指標の値の上位集団であってもよい。

[0020] 本開示の第7態様に係る方法は、第3態様から第6態様のいずれか一態様に係る方法において、指標は、有用物質の産生量であってもよい。

- [0021] 本開示の第8態様に係る方法は、第3態様から第6態様のいずれか一態様に係る方法において、指標は、積分生存細胞密度であってもよい。
- [0022] 本開示の第9態様に係る方法は、第3態様から第6態様のいずれか一態様に係る方法において、指標は、乳酸濃度であってもよい。
- [0023] 本開示の第10態様に係る方法は、第1態様から第9態様のいずれか一態様に係る方法において、産生安定性の予測に用いるデータは、1つ以上の遺伝子発現レベルを含む構成であってもよい。
- [0024] 本開示の第11態様に係る方法は、第1態様から第10態様のいずれか一態様に係る方法において、1つ以上のプロセッサが、予測対象のデータの入力を受けて、安定または不安定の2クラス分類を行うモデルを用いて産生安定性を予測する構成であってもよい。
- [0025] 本開示の第12態様に係る方法は、第11態様に係る方法において、モデルは、予測対象のクローンと同様の限定をした訓練用のクローンについてのデータと正解の安定性ラベルとが関連付けされた複数の訓練データを用いた機械学習によって訓練されたモデルであってもよい。
- [0026] 本開示の第13態様に係る方法は、第12態様に係る方法において、複数の訓練データは、産生する有用物質が異なる複数種類のクローンについての訓練データを含み、1つ以上のプロセッサが、モデルの訓練に使用された有用物質とは別の有用物質を産生するクローンについての産生安定性を予測する構成であってもよい。
- [0027] 本開示の第14態様に係る方法は、第1態様から第13態様のいずれか一態様に係る方法において、有用物質は、医薬品原料であるタンパク質、ペプチド、およびウイルスのうちいずれかであってもよい。
- [0028] 本開示の第15態様に係る方法は、第1態様から第14態様のいずれか一態様に係る方法において、有用物質は、抗体、または抗体様タンパク質であってもよい。
- [0029] 本開示の第16態様に係る方法は、第1態様から第15態様のいずれか一態様に係る方法において、クローンは、脊椎動物由来細胞であってもよい。

- [0030] 本開示の第17態様に係る方法は、第1態様から第15態様のいずれか一態様に係る方法において、クローンは、哺乳類由来細胞であってもよい。
- [0031] 本開示の第18態様に係る方法は、第1態様から第15態様のいずれか一態様に係る方法において、クローンは、CHO細胞またはHEK細胞 (Human Embryonic Kidney cells) であってもよい。
- [0032] 本開示の第19態様に係る情報処理装置は、1つ以上のプロセッサと、1つ以上のプロセッサに実行させる命令が記憶される1つ以上の記憶装置と、を備え、1つ以上のプロセッサは、有用物質を産生するクローンについて1種類以上のクローンの培養データを取得し、培養データを解析して予測対象のクローンを限定し、予測対象のクローンについて測定されたデータを用いて、予測対象のクローンによる有用物質の産生安定性を予測する。
- [0033] 第19態様に係る情報処理装置について、第2態様から第18態様のいずれか一態様の方法と同様の態様を含む構成とすることができる。
- [0034] 本開示の第20態様に係るプログラムは、コンピュータに、有用物質を産生するクローンについて1種類以上のクローンの培養データを取得する機能と、培養データを解析して予測対象のクローンを限定する機能と、予測対象のクローンについて測定されたデータを用いて、予測対象のクローンによる有用物質の産生安定性を予測する機能と、を実現させる。
- [0035] 第20態様に係るプログラムについて、第2態様から第18態様のいずれか一態様の方法と同様の態様を含む構成とすることができる。
- [0036] 本開示の第21態様に係る予測モデル生成方法は、有用物質を産生するクローンの産生安定性を予測する機能をコンピュータに実現させる予測モデルを生成する予測モデル生成方法であって、1つ以上のプロセッサを含むシステムが、1種類以上のクローンの培養データを取得することと、培養データを解析して予測対象のクローンを限定することと、予測対象に該当するクローンについて測定されたデータと正解の安定性ラベルとが関連付けされた複数の訓練データを用いて機械学習を行い、データの入力に対する予測モデルの出力が正解の安定性ラベルに近づくように予測モデルを訓練することと、

を含む。

[0037] 第21態様に係る予測モデル生成方法について、第2態様から第18態様のいずれか一態様の方法と同様の態様を含む構成とすることができる。

### 発明の効果

[0038] 本開示によれば、培養データを解析して得られる情報を基に予測対象が適切に限定され、有用物質を産生するクローンの産生安定性を高精度に予測することが可能になる。また、本開示によれば、予測対象が限定されることにより、産生安定性の予測コストを抑制でき、低コストで予測が可能である。

### 図面の簡単な説明

[0039] [図1]図1は、抗体医薬品の生産工程の概要を示す説明図である。

[図2]図2は、クローンによる抗体産生量の変化の例を示すグラフである。

[図3]図3は、本実施形態によって実現される安定性予測AI (Artificial Intelligence) の役割を概説する説明図である。

[図4]図4は、遺伝子発現データを基に産生安定性を予測する機械学習モデルの概念図である。

[図5]図5は、本実施形態に係るクローンの産生安定性予測方法の概要を示す説明図である。

[図6]図6は、モデルの訓練および評価に用いるデータセットの例を示す図表である。

[図7]図7は、培養データのある指標による対象の絞り込みの例を示すグラフである。

[図8]図8は、評価サンプルとして用意した5種類の抗体産生CHO細胞のクローン数と安定性ラベルの付与例を示す図表である。

[図9]図9は、各抗体種において抗体生産量の値が相対順位の上位40%に該当するクローン数と安定性ラベルの付与例を示す図表である。

[図10]図10は、各抗体種において積分生存細胞密度の値が相対順位の上位60%に該当するクローン数と安定性ラベルの付与例を示す図表である。

[図11]図11は、各抗体種において乳酸濃度の値が相対順位の上位40%に

該当するクローン数と安定性ラベルの付与例を示す図表である。

[図12]図12は、実施形態に係る情報処理装置の機能的構成を示すブロック図である。

[図13]図13は、情報処理装置のハードウェア構成の例を示すブロック図である。

[図14]図14は、産生安定性予測モデルを生成するための機械学習の処理を実行する機械学習装置のハードウェア構成の例を示すブロック図である。

[図15]図15は、機械学習装置が実行する機械学習方法の例を示すフローチャートである。

[図16]図16は、実施形態に係る情報処理装置が実行する情報処理方法の例を示すフローチャートである。

### 発明を実施するための形態

[0040] 以下、添付図面に従って本発明の好ましい実施形態について詳細に説明する。

[0041] 《抗体医薬品の生産工程の概要》

バイオ医薬品の中でも薬効面と安全面の両立性の高さから市場が拡大している抗体医薬品は、複雑な構造を持つタンパク質である抗体を安定的に産生できる動物細胞のクローンを用いて生産されている。以下では、有用物質として抗体を例にとり説明する。図1は、抗体医薬品の生産工程の概要を示す説明図である。抗体医薬品を生産するまでのプロセスは、[1]クローン作製フェーズと、[2]プロセス開発フェーズと、[3]GMP (Good Manufacturing Practice) 製造フェーズと、を含む。

[0042] クローン作製フェーズは、抗体医薬品の生産に適した動物細胞に対して、ベクターを加えて遺伝子組み換えを行い、複数のクローンの候補を作製する工程と、これら複数の候補の中から、抗体の産生量、細胞増殖性、繰り返し増殖しても細胞特性が変化しない品質安定性などの点で優れたクローンをスクリーニングする工程と、を含む。

[0043] プロセス開発フェーズは、スクリーニングしたクローンを用いて、GMP

製造に必要な生産プロセス（培養条件、精製条件など）を開発するフェーズである。

[0044] GMP製造フェーズでは、確立した生産プロセスのもと、クローンを培養して増殖させ、クローンに抗体を産生させる。さらに、その抗体を精製して製剤化することにより、抗体医薬品が出来上がる。

[0045] 抗体をクローンに産生させる場合、長期間にわたってその産生性が変化しないこと（安定であること）が求められている。そのため、なるべく多種類のクローンを作製しておき、そこから産生性が安定なクローンを選抜することが行われるが、従来は、数か月の連続培養を要する実験的な検証が必要なため、負荷が高くなっている。

[0046] 図2は、クローンによる抗体産生量の変化の例を示すグラフである。縦軸は抗体の産生性を表し、横軸は経過時間（タイムポイント）を表す。「抗体の産生性」は、クローンが産生する抗体の単位時間当たりの抗体産生量で表される。

[0047] 図2には、クローンが産生する抗体の量が長期間（2～3か月）にわたってどのくらい変化するかをプロットしたグラフが示されている。グラフG1は、産生性が安定しているクローンについての抗体産生量の変化を示すグラフである。グラフG2は産生性が不安定なクローンについての抗体産生量の変化を示すグラフである。グラフG1に示すように、産生性が安定しているクローンは、現時点から2～3か月経過しても産生性が概ね変わらず、現時点と概ね変わらない産生性を維持することができる。これに対し、グラフG2に示すように、産生性が不安定なクローンは2～3か月の間に次第に産生性が低下する。

[0048] 本発明では、「現時点」とは、2週間の規格試験時点、または、規格試験の終了した時点、すなわち、産生安定性を判別するための培養が開始される時点である。また、「現時点の抗体の産生性」とは、2週間の規格試験におけるクローンが産生する単位時間あたりの抗体産生量である。

[0049] 遺伝子導入により抗体を産生する細胞を作った場合に、図2に示すように

、安定なクローンと不安定なクローンとの両方のものが作られてしまう。したがって、クローン作製フェーズにおいては、多種類のクローンを作製し、その中からグラフG1のような振る舞いを示す産生性が安定なクローンを選抜することが行われる。

[0050] 図2のような産生性の振る舞いは、クローンの種類によって様々であり、従来は、クローンに作らせる抗体の種類が変わる度に、図2と同じような実験を行ってそれぞれのクローンの産生安定性を評価しなければならなかった。

[0051] これに対し、本開示の実施形態では、現時点におけるクローンから得られる情報を基に、数か月先の抗体の産生安定性を精度よく予測する仕組みを提案する。ここで、「現時点におけるクローンから得られる情報」とは、2週間の規格試験においてクローンから得られる情報である。予測の目的変数である抗体の産生安定性とは、現時点と数か月の期間培養後との抗体の産生量の変化の有無で定義することができる。ここでの「数か月」とは例えば2か月以上の期間であり、例示的には2～3か月であってよい。また、継代を所定の回数行うまでの期間としてもよい。期間の設定は、クローンの増殖能力に基づいて決めてもよいし、実際に抗体の製造を行う際のクローンの培養期間に基づいて決めてもよい。「現時点」は図2のグラフの左端に示す培養初期の時点、つまり、2週間の規格試験が終了した時点であり、抗体の産生安定性を判別するための培養が開始される時点である。産生性が「安定」であるとは、現時点と数か月先とで抗体の産生量の変化がないことである。「変化がない」とは、変化の量が許容範囲内であり、実質的に変化がないものと見なしうる場合を含む。産生性が「不安定」であるとは、現時点と数か月先とで抗体の産生量の変化があること、多くの場合は産生量が低下することである。産生性の変化があるとみなす閾値は、任意に設定できるが、例えば現時点の産生量に対して±30%や±20%であってよい。

[0052] 《未知クローンへの汎化性能について》

図3は、本実施形態によって実現される安定性予測AI (Artificial Inte

lligence) の役割を説明する説明図である。図3に示すように、クローン作製フェーズでは、宿主細胞に対して、作りたい有用物質の遺伝子の設計図を導入する遺伝子導入が行われる。例えば、宿主細胞に対して有用物質Aを作る設計図を遺伝子導入した場合は、有用物質Aを産生する細胞が得られる。このような産生細胞は確率的にできるため、有用物質Aを産生しない細胞や産生量が不十分な細胞も作られてしまう。このため、まずはこの段階で簡便な試験を行い、有用物質Aを十分に産生し得る高産生なクローンを選抜することが行われる。

[0053] その後、従来であれば、図2で説明したように、2～3か月間の安定性試験を行い、数か月にわたって有用物質Aを作り続けられるかどうかを確認し、産生安定性のあるクローンを選抜する。

[0054] 本実施形態では、従来の安定性試験に代替する手段として、安定性予測A1を構築し、現時点でのクローンの状態、つまり2週間の規格試験におけるクローンの状態を測定して得られるプロフィールを基に、安定性予測A1によって2～3か月後の状態（産生性の変化）を予測する。

[0055] 細胞に産生させる有用物質（例えば抗体）の種類は、目的によって多種多様であることから、細胞が産生する有用物質の種類によらず、産生安定性を予測できるモデルを構築することが望まれる。すなわち、未知の抗体種に対してロバストに抗体産生安定性を予測するモデルが好ましい。

[0056] 安定性予測A1に適用するモデルを学習する際には、対象とする有用物質を事前に知ることはできず、モデルの学習時に使用した有用物質の種類と、学習後にモデルに予測させる対象のクローンが産生する有用物質とは別の種類となり得る。つまり、未知の有用物質種に対してロバストに産生安定性を精度よく予測するモデルが好ましく、有用物質種をドメインとしたドメイン汎化性のある予測モデルを構築することが好ましい。

[0057] 《産生安定性を予測する機械学習モデルの概要》

本実施形態では、クローン作製フェーズにおいて、現時点のクローンの情報から、2～3か月先の産生性の変化の有無を推定（予測）すること、すな

わち、有用物質の産生安定性を予測することを可能とする安定性予測AIを構築する。より具体的には、クローンの現時点（規格試験時）の遺伝子発現データの入力を受けて有用物質の産生安定性を示す安定性ラベルを出力するモデルを構築する。より詳しくは、クローンの一部を規格試験に用いるクローンとし、別の一部を遺伝子発現データの取得のための遺伝子解析にかけるクローンとすることで、規格試験に用いるクローンの遺伝子発現データを取得する。安定性ラベルは、「安定」であることを示す値の「1」または「不安定」であることを示す値の「0」の2値で表すことができる。産生安定性を予測する予測モデルは、「安定」または「不安定」のクラス分類を行う2クラス分類モデルであってよい。

[0058] 遺伝子発現データは、1つ以上の遺伝子レベルを含む。本実施形態に用いる遺伝子発現データは、複数の遺伝子のそれぞれの遺伝子発現レベルを数値化したデータを含む。遺伝子発現データは、例えば、RNA (ribonucleic acid) シーケンス解析によって得ることができる。遺伝子発現量を示す値は、例えば、正の整数をとるカウント値であり、対数変換して特徴量として用いることができる。

[0059] 図4は、遺伝子発現データを基に産生安定性を予測する機械学習モデルMLMの概念図である。図4の矩形枠RF1の内側には、訓練データのデータセットの例が示されている。図4では、複数のクローンA～Nのそれぞれの現時点（規格試験時）の遺伝子発現データをヒートマップによって可視化した遺伝子発現パターンGEPとして表している。遺伝子発現パターンGEPの横軸は遺伝子の種類を表しており、複数の遺伝子のそれぞれの遺伝子発現レベルが2色のグラデーション（ヒートマップ）によって表現されている。遺伝子発現データに含まれる遺伝子a, b, c, d・・・の種類数は、例えば安定なクローンと不安定なクローンについて全遺伝子発現データを取得し、安定なクローンと不安定なクローン間の2群の統計学的な有意確率を用いて選択した300～400種類が好ましい。さらに遺伝子の種類数を絞り込む場合は、選択した遺伝子を用いて遺伝子の種類数を増減させながら機械学

習モデルMLMを実際に訓練し、予測性能が高くなる種類数を探索して、例えば50～100種類の遺伝子に絞り込むことが好ましい。なお、ここでは全遺伝子発現データを取得したが、必ずしも全遺伝子発現データを取得する必要はなく、無作為に一部の遺伝子を選択し、その遺伝子発現データを取得してもよい。図示の制約上ヒートマップの色を表現できないため、代わりに、赤を「R」、青を「B」、白を「W」と表示している。赤（R）は、遺伝子発現レベルが相対的に高いことを表し、青（B）は遺伝子発現レベルが相対的に低いことを表す。白（W）は、遺伝子発現レベルが中間的な値であることを表す。

[0060] 複数のクローンA～Nのそれぞれは、規格試験後の数か月間の培養による実験的検証に基づき、「安定」または「不安定」であることが確認されており、各クローンA～Nに対して「安定」または「不安定」を示す安定性ラベル（正解ラベル）が付与されている。こうして、複数のクローンA～Nのそれぞれの現時点の遺伝子発現データと正解の安定性ラベルとが関連付け（紐付け）された複数の訓練データを含むデータセットが用意される。そして、複数の訓練データを用いて機械学習モデルMLMを訓練し、機械学習モデルMLMに安定または不安定の遺伝子パターンを学習させる。こうして訓練された学習済み（訓練済み）の機械学習モデルMLMに対して、未知のクローンXの現時点（規格試験時）の遺伝子発現データを入力すると、機械学習モデルMLMは入力された遺伝子発現データから産生安定性を予測し、「安定」または「不安定」のラベルを予測結果として出力する。なお、図4では、未知のクローンXに対して、機械学習モデルMLMが「安定」であると予測した例が示されている。

[0061] 《実施形態の概要：予測対象を限定して有用物質の産生安定性を予測するモデルを構築する》

有用物質を産生するクローンには様々な特性があるため、種類を問わず全てのクローンの産生安定性を高精度に予測することは難しい。本実施形態では、現時点（規格試験時）の各クローンの培養データから得られる指標に基

づき予測対象を限定することにより、高精度な予測を実現する。ここで培養データとは、クローンについて培養装置あるいは細胞を含む培養液を一部サンプリングして専用装置を用いて測定できる一般的なデータである。

[0062] 図5は、本実施形態に係るクローンの有用物質の産生安定性予測方法の概要を示す説明図である。図5の左図F5Aには予測対象を限定しない場合の比較例を示し、図5の右図F5Bに本実施形態による方法の概要を示す。

[0063] 左図F5Aの予測対象を限定しない場合について説明する。左図F5Aの矩形枠RF2内には、有用物質A～Dを産生する複数種類のクローンの訓練データを含むデータセットDS<sub>c</sub>が模式的に示されている。このデータセットDS<sub>c</sub>は、各有用物質A～Dについて5クローンずつ、計20クローンの訓練データを含む。ここで訓練データは、20クローンについて、それぞれの規格試験時の遺伝子発現データと正解の安定性ラベルとが関連付け（紐付け）されたデータである。図5の各クローンの下部に表示している「9」、「7」、「6」などの値は、規格試験時の各クローンのある培養データの測定値を表している。なお、測定値ではなく、測定値から取得できる各クローンにおける相対的なレベルを表してもよい。ここでは、左図F5Aについて説明したが、右図F5Bについても同様である。

[0064] 左図F5Aでは、訓練データを限定せずに、データセットDS<sub>c</sub>の全ての訓練データを用いて機械学習モデルMLM<sub>c</sub>を訓練し、学習済み（訓練済み）のモデルを用いて未知の有用物質Xを産生する複数種類のクローンの産生安定性を予測することを示している。この場合、予測対象である未知の有用物質Xを産生する複数種類のクローンについても特に限定せず、矩形枠RF3内に示すように、未知の有用物質Xを産生する5種類のクローンのすべてを対象に産生安定性の予測を行う。産生安定性の予測は、未知の有用物質Xを産生する5種類のクローンのすべてに対して、現時点（規格試験時）の遺伝子発現データを取得し、学習済み（訓練済み）のモデルに入力することで行うが、その予測精度は低い。

[0065] 次に右図F5Bの本実施形態による方法について説明する。左図F5Aの

予測対象を限定しない方法に対し、右図 F 5 B に示す方法では、規格試験時のある培養データの値を指標にして予測対象を限定する。まず、ある培養データの値に注目して閾値を決め、データセット D S d に含まれるクローンの集団をグループ分けする。例えば、指標とする培養データの値が閾値に対して相対的に大きいものと、小さいものとの 2 つのグループに分ける。ここでは閾値を「5」とし、指標とする培養データの値が「5」以上である集団を訓練の対象とし、培養データの値が「5」より小さい集団は対象外とする例を示している。この閾値処理により、矩形枠 R F 4 内に示すように各有用物質 A ~ D について 3 クローンずつ、計 12 クローンの訓練データを対象として残し、これら限定された集団の訓練データを含むデータセット D S e を機械学習モデル M L M e の訓練に用いる。その一方、破線の矩形枠 R F 5 内に示す 8 クローンの訓練データ、すなわち、閾値の条件を満たしていないクローンの訓練データは処理の対象外とする。

[0066] こうして、対象が限定されたデータセット D S e を用いて機械学習モデル M L M e を訓練する。そして、学習済みのモデルを用いて未知の有用物質 X を産生するクローンの産生安定性を予測する際にも、その予測対象のクローンは、モデルの訓練に用いたデータセット D S e のクローンの集団と同じように、指標とする培養データの値について閾値を適用し、閾値による限定条件を満たすもの（指標の値が閾値より上位の集団）に限定して予測を行う。矩形枠 R F 6 内に示す 3 種類のクローンは、予測対象に該当するクローンを表している。また、破線の矩形枠 R F 7 内に示す 2 種類のクローンは、予測対象外のクローンを表している。このように予測対象を限定して予測を行うことにより、高い予測精度を実現できる。さらに、破線の矩形枠 R F 7 内に示す予測対象外のクローンは、遺伝子発現データ取得が不要であることから、遺伝子解析のコストを抑制できる。

[0067] 《訓練および評価に用いるデータセットの例》

図 6 に、モデルの訓練および評価に用いるデータセットの例を示す。図 6 の上段には、有用物質としての抗体 A を産生するクローンについてのデータ

セットD S Aの例を示し、下段には有用物質としての抗体Bを産生するクローンについてのデータセットD S Bの例を示す。図示は省略するが、有用物質としての他種類の抗体を産生するクローンについてのデータセットも同様である。

[0068] データセットD S Aは、複数のクローンA C L jのそれぞれについて測定された規格試験時の培養データと、規格試験時の遺伝子発現データと、安定性試験によって得られた正解の安定性ラベルとを含む。添字のjは、クローンを識別するインデックス番号を表す。培養データは、例えば、抗体産生量、積分生存細胞密度 (integral viable cell density: I V C D)、乳酸濃度、p Hなど1つ以上の項目を含んでよい。培養データは、培養装置あるいは細胞を含む培養液を一部サンプリングして専用装置を用いて測定できる一般的なデータであってよく、例えば、細胞総数、細胞分泌物質の量、細胞産生物質の量、細胞代謝物質の量および培地成分量のうち1つ以上を含んでいてもよい。図6に示す表の各セル内の文字記号 (添字jを付した記号) は、対応するデータ項目の値を表している。

[0069] データセットD S Bについても同様である。データセットD S Aに含まれるクローンA C L jの個数n aとデータセットD S Bに含まれるクローンB C L jの個数n bは異なっていてよい。

[0070] このように用意された複数のドメイン (有用物質種) のデータセットからある培養データの指標に注目して対象の絞り込み (限定) を行う。

[0071] 《予測対象の絞り込みの例》

図7は、培養データのある指標による予測対象の絞り込みの例を示すグラフである。横軸に、複数の有用物質A~Eのそれぞれを産生する複数種類のクローンが並んでいる。縦軸は、規格試験時の培養データより得られたある指標の値である。なお、図7に示すクローンは、モデルの訓練 (学習) に用いるクローンである。

[0072] 図7に示すように、産生する有用物質種の異なるクローンによって、培養データより得られたある指標の分布範囲が異なることがある。この場合、図

5で説明したように、指標の値に対して閾値を決め、その閾値との相対的な大小関係に基づいてクローンを2つの集団に分け、訓練に用いるクローンの集団と訓練の対象外とするクローンの集団とを決めたのでは、産生する有用物質種によって訓練の対象となるクローンの数にバラツキがでてしまう。例えば、指標の閾値を2.5とし、閾値以上の値のクローンの集団を訓練に用いるとした場合、有用物質Bを産生するクローンは訓練に使用しないことになってしまう。

[0073] そこで、例えば、図7に示すように、各有用物質A～Dを産生するクローンについて、培養データより得られるある指標の相対的上位X%（Top-X%）に訓練対象を限定するとしてもよい。ここで相対的上位X%とは、各有用物質A～Dのそれぞれを産生するクローンの集団において、培養データより得られるある指標について降順に並べたときの上位X%（Top-X%）を意味する。限定条件となる閾値に相当する「X%」という基準は、各有用物質A～Eからのサンプリングの数が概ね同じ位の数になるように調整されることが好ましい。相対的上位X%は本開示における「指標の値の順位を用いて定義される閾値」の一例である。

[0074] このように訓練対象を限定しても、その中に産生性が安定なクローンと不安定なクローンとが存在し得る。そして、学習（訓練）済みモデルを用いて未知の有用物質Yを産生するクローンの産生安定性を予測する際にも、その予測対象のクローンは、モデルの訓練に用いたクローンと同じように、培養データより得られるある指標に関して上位X%のクローンに限定して予測を行う。

[0075] ここでは、複数の有用物質A～Eのそれぞれを産生する複数種類のクローンをモデルの訓練に用いたが、必ずしも異なる有用物質を産生するクローンを複数種類用いる必要はなく、例えば、有用物質Aを産生するクローンのみを訓練に用いてもよい。この場合、訓練に用いるクローンの集団の限定方法は、規格試験時のある培養データの値に注目して閾値を設定し、閾値との相対的な大小関係に基づいて行ってもよいし、規格試験時の培養データより得

られたある指標の値に関して上位X%としてもよい。また、相対的上位X%としたが、培養データより得られたある指標によっては、相対的下位X%としてもよい。

[0076] 予測対象を限定する際の培養データの指標と閾値は、用意されたデータセットから、試行錯誤的に仮説と検証とを繰り返す作業によって決定されてもよい。あるいはまた、予測対象を限定する際の培養データの指標と閾値は、用意されたデータセットから探索的な解析を行うことにより決定することができる。

[0077] 例えば、図7に示すような5つの有用物質A～E（ドメイン）のデータセットが存在する場合、プロセッサを含む情報処理装置により、フィルタ法（Filter Method）などの特徴選択の手法を用いて、5つのドメインのそれぞれにおいて各特徴量と目的変数（安定性ラベル）との関連度をそれぞれ評価し、5ドメイン中例えば4ドメイン以上で関連度が高い特徴量をドメイン普遍性が高い特徴量とする。情報処理装置は、全データの中からある指標に着目して特定の条件を満たすデータをサブセットとして抽出し、抽出したサブセットについて、ドメイン普遍性が高い特徴量の個数を基にドメイン汎化性評価を行う。ドメイン普遍性が高い特徴量の個数が多い場合、ドメイン汎化性の高いサブセットと評価される。ドメイン汎化性の高いサブセットのデータを訓練データとして用いて予測モデルの学習（訓練）を行うことにより、学習済みのモデルは、学習時と同様の条件で対象を限定した集団（サブセット）に対しては、他のドメイン（有用物質種）に対してもロバストに産生安定性を予測可能である。

[0078] 抗体産生クローンの場合、予測対象の限定に有効な培養データの指標は、例えば、抗体産生量、積分生存細胞密度、乳酸濃度などであり、これらのいずれかの指標の値の上位集団を対象とすることで、高精度の産生安定性予測が可能であることが確認された。

[0079] 《有用物質の例》

有用物質は、抗体に限らず、抗体様タンパク質であってもよい。有用物質

は、医薬品原料であるタンパク質、ペプチド、およびウイルスのうちいずれかであってよい。

[0080] 《クローンの例》

有用物質を産生するクローンは、脊椎動物由来細胞であってよい。クローンは、例えば、哺乳類由来細胞であってよい。クローンは、CHO細胞またはHEK細胞であってもよい。

[0081] 《実施例》

以下、本開示の技術を適用した実施例1～3を説明する。各実施例1～3に共通する構成は次の通りである。すなわち、有用物質を抗体とし、産生細胞をCHO細胞とする。評価サンプルとして5種類の抗体産生CHO細胞のクローンについて、それぞれ複数種のクローンを用意し、RNAシーケンス（RNA-Seq）解析にて2週間の規格試験にて測定した全遺伝子発現レベルから100種類の遺伝子発現レベルを選択して説明変数とし、安定または不安定の2クラスに分類するロジステック回帰モデルを学習器とした、5分割クロスバリデーションを実施して予測モデルの訓練（学習）を行い、性能評価はPRAUC（Area Under the Precision-Recall Curve）を用いた例を示す。説明変数に用いる遺伝子発現レベルの種類数は、実施例1～3において、統計学的な有意確率を用いて選択した300～400種の遺伝子を用いて、種類数を増減させながら予測モデルの訓練（学習）を実際に行い、予測性能が高くなる種類数を探索することで100種類とした。なお、規格試験は、クローン（CHO細胞）の播種数は $5 \times 10^5$  cells/mL、40mLのフラスコで浮遊培養で行った。

[0082] 5分割クロスバリデーションは、5種類の抗体種ごとに分割して、未学習の抗体種による性能を評価した。すなわち、4種類の抗体種のデータセットを訓練（学習）用のデータとして用い、残りの1種類の抗体種のデータセットを性能評価用のテストデータとして用いた。

[0083] 図8は、評価サンプルとして用意した5種類の抗体産生CHO細胞のクローン数と安定性ラベルの付与例を示す図表である。評価サンプルとして5種

類の抗体産生細胞を182クローン用意し、規格試験と同様の条件で2か月細胞培養することで、各クローンに対して安定性ラベル（「安定」または「不安定」）を付与した（図8参照）。例えば、抗体Aを産生するクローンは計24クローンあり、そのうち「安定」のラベルが付与されたものが7クローン、「不安定」のラベルが付与されたものが17クローンである。また、182クローンのそれぞれについて、規格試験時に培養データと遺伝子発現データを取得し、クローン毎に遺伝子発現データと安定性ラベルとか紐付けられ、訓練データを構成している。

[0084] [実施例1]

実施例1では、予測対象を「相対的高産生なクローン」に限定した安定性予測を行う例を説明する。ここで、「相対的高産生なクローン」とは、有用物質の産生量が相対的に高いクローンを意味する。

[0085] 予測対象を「相対的高産生なクローン」に限定した安定性予測を行う際の、予測モデルの訓練に用いる訓練対象のクローンの限定方法について説明する。なお、訓練対象のクローンを限定することは、訓練において予測モデルに予測させる対象のクローンを限定すること、つまり予測モデルによる予測対象のクローンを限定することに相当する。

[0086] 訓練対象のクローンの限定方法は、全182クローンの規格試験時の培養データから「抗体産生量」に注目し、予測モデルでの予測性能が高くなるように閾値を探索して、各抗体種で相対順位の上位40%のクローンに限定する方法とした。ここで、「抗体産生量」は、例えば規格試験における2週間（14日間）の抗体産生量の積算量とすることができる。または、規格試験中のある期間、例えば期間、例えば10日間の抗体産生量の積算量としてもよく、計測期間で除算して単位時間あたりの抗体産生量としてもよい。「上位40%」は閾値の一例である。図9は、各抗体種において抗体生産量の値が相対順位の上位40%に該当するクローン数と安定性ラベルの付与例を示す図表である。

[0087] 図9には、各抗体種において相対順位の上位40%に該当する計73クロ

ーンの例が示されている。図9に示す73クローンの規格試験時の遺伝子発現データと安定性ラベルとが紐付けされた訓練データを含む抗体種ごとのデータセットを用いて5分割クロスバリデーションを実施した。このように訓練対象を限定した結果、学習済みの予測モデルの予測性能はP R A U Cの値が0.743となった。なお、学習済みの予測モデルを用いて未知の有用物質の産生安定性を予測する際の予測対象のクローンについても、訓練対象の限定と同様に規格試験時（現時点）の培養データを解析し、「抗体産生量」の上位40%に限定して予測を行うこととする。

[0088] [比較例]

これに対し、訓練対象を限定せずに、図8に示す182クローンの全データを含むデータセットを用いて、同様の学習を行い、5分割クロスバリデーションを実施した場合に得られる比較例に係る予測モデルの予測性能はP R A U Cの値が0.503であった。なお、予測対象は、訓練対象と同様に対象を限定せずに行うこととする。実施例1によって予測対象を限定した予測モデルの性能は、比較例に係る予測モデルよりも高精度であることが確認された。

[0089] この結果は、実施例1の方法により生成される予測モデルを用いて、未知の有用物質に対して高精度に予測可能であることを示すと同時に、相対的に高産生なクローンに限定することは有用物質の産生クローンの選抜工程において全く障害にならず、高精度で予測できる対象に限定することにより低コストで実施可能なため、本開示による安定性予測は実用可能であると考えられる。

[0090] [実施例2]

実施例2では、予測対象を「相対的に細胞密度の高いクローン」に限定した安定性予測を行う例を説明する。まず、予測対象を「相対的に細胞密度の高いクローン」に限定した安定性予測を行う際の、予測モデルを訓練するための訓練対象のクローンの限定方法について説明する。実施例1と同様に、図8に示す全182クローンの規格試験時の培養データから「積分生存細胞

密度（IVCD）」に着目し、予測モデルでの予測性能が高くなるように閾値を探索して、各抗体種で相対順位の上位60%のクローンに限定する方法とした。ここで、「相対的に細胞密度の高いクローン」は、例えば規格試験における2週間（14日間）の「積分生存細胞密度（IVCD）」に基づいて取得することができる。または、規格試験中のある期間、例えば10日間の「積分生存細胞密度（IVCD）」に基づいて取得してもよい。「上位60%」は閾値の一例である。図10は、各抗体種において積分生存細胞密度の値が相対順位の上位60%に該当するクローン数と安定性ラベルの付与例を示す図表である。

[0091] 図10には、各抗体種において相対順位の上位60%に該当する計109クローンの例が示されている。図10に示す109クローンの規格試験時の遺伝子発現データと安定性ラベルとが紐付けされた訓練データを含む抗体種ごとのデータセットを用いて5分割クロスバリデーションを実施した。このように訓練対象を限定した結果、学習済みの予測モデルの予測性能はPRAUCの値が0.647となった。すなわち、実施例2によって予測対象を限定した予測モデルの性能は、対象を限定しない比較例に係る予測モデルのPRAUC（0.503）よりも高精度であることが確認された。なお、学習済みの予測モデルを用いて未知の有用物質の産生安定性を予測する際の予測対象のクローンについても、訓練対象の限定と同様に規格試験時（現時点）の培養データを解析し、「積分生存細胞密度（IVCD）」の上位60%に限定して予測を行うこととする。

[0092] この結果は、実施例2の方法により生成される予測モデルを用いて未知の有用物質に対して高精度に予測可能であることを示すと同時に、相対的に生存細胞密度が高いクローンに限定することは有用物質の産生クローンの選抜工程において障害にならず、高精度で予測できる対象に限定することにより低コストで実施可能なため、本開示による安定性予測は実用可能であると考えられる。

[0093] [実施例3]

実施例3では、「相対的に乳酸濃度の高いクローン」に限定した安定性予測を行う例を説明する。まず、予測対象を「相対的に乳酸濃度の高いクローン」に限定した安定性予測を行う際の、予測モデルを訓練するための訓練対象のクローンの限定方法について説明する。実施例1と同様に、図8に示す全182クローンの2週間の規格試験の培養データからクローンを培養している培養液の「乳酸濃度」に着目し、2週間（14日間）の内の各時点、例えば一日毎に測定された培養液の「乳酸濃度」の中央値を代表値として、各クローンの「乳酸濃度」を取得する。そして、予測モデルでの予測性能が高くなるように閾値を探索して、各抗体種で相対順位の上位40%のクローンに限定する方法とした。「上位40%」は閾値の一例である。図11は、各抗体種において乳酸濃度の値が相対順位の上位40%に該当するクローン数と安定性ラベルの付与例を示す図表である。

[0094] 図11は、各抗体種において相対順位の上位40%に該当する計72クローンの例が示されている。なお、図9と比較してクローンの数が1クローン少ない理由は、乳酸濃度の測定において、1クローンのデータ欠損があったためである。

[0095] 図11に示す72クローンの規格試験時の遺伝子発現データと安定性ラベルとが紐付けされた訓練データを含む抗体種ごとのデータセットを用いて5分割クロスバリデーションを実施した。このように対象を限定した結果、学習済みの予測モデルの予測性能はPRAUCの値が0.613となった。すなわち、実施例3によって予測対象を限定した予測モデルの性能は、対象を限定しない比較例に係る予測モデルのPRAUC（0.503）よりも高精度であることが確認された。なお、学習済みの予測モデルを用いて未知の有用物質の産生安定性を予測する際の予測対象のクローンについても、訓練対象の限定と同様に規格試験時（現時点）の培養データを解析し、「乳酸濃度」の上位40%に限定して予測を行うこととする。

[0096] この結果は、未知の有用物質に対して高精度に予測可能であることを示すと同時に、相対的に乳酸濃度が高いクローンに限定することは有用物質の産

生クローンの選抜工程において障害にならず、高精度で予測できる対象に限定することにより低コストで実施可能なため、本開示による安定性予測は実用可能であると考えられる。

[0097] 《情報処理装置の構成例》

図12は、実施形態に係る情報処理装置10の機能的構成を示すブロック図である。情報処理装置10は、データ取得部12と、予測対象限定部14と、産生安定性予測モデル16と、処理結果出力部18と、を備える。情報処理装置10の各種機能は、コンピュータのハードウェアとソフトウェアとの組み合わせによって実現し得る。情報処理装置10の物理的形態は特に限定されず、サーバコンピュータであってもよいし、ワークステーションであってもよく、パーソナルコンピュータあるいはタブレット端末などであってもよい。

[0098] データ取得部12は、有用物質を産生するクローンについての1種類以上のクローンの培養データおよび遺伝子発現データを含む各種データを取得する。

[0099] 予測対象限定部14は、培養データ解析部20と、限定条件判定部22とを含み、入力された1種類以上のクローンの培養データを解析して予測対象のクローンを限定する。培養データ解析部20は、培養データの解析を行う。限定条件判定部22は、培養データの解析結果を基に、閾値により対象を限定する。なお、説明の便宜上、培養データ解析部20と限定条件判定部22とを分けて記載しているが、限定条件判定部22は培養データ解析部20に含まれていてもよい。また、培養データ解析部20が予測対象限定部14として機能すると理解してもよい。

[0100] 培養データ解析部20は、入力されたデータセットから予測対象を限定するための指標と閾値とを決定する処理を実行し得る。なお、予測対象の限定条件となる指標と閾値については、培養データ解析部20による解析結果に基づいて設定されてもよいし、不図示の別の情報処理装置等を用いた探索処理の結果などによって事前に把握されている既知の情報として予測対象限定

部 1 4 に設定されてもよい。

- [0101] 産生安定性予測モデル 1 6 には、機械学習モデルが適用される。産生安定性予測モデル 1 6 は、予測対象であるクローンの現時点の遺伝子発現データの入力を受け付けて、入力された遺伝子発現データを基にクローンの産生安定性を予測して安定性ラベルを出力する 2 クラス分類モデルであってよい。産生安定性予測モデル 1 6 は、図 5 の右図 F 5 B にて説明した方法により対象を限定した訓練データを用いて訓練される。産生安定性予測モデル 1 6 に入力される遺伝子発現データは、1 つ以上の遺伝子発現レベルを含む。産生安定性予測モデル 1 6 に入力される遺伝子発現データには、複数の遺伝子の発現レベルのデータが含まれていてもよい。説明変数として用いる特徴量は、公知の特徴量選択の手法により選択されてもよい。
- [0102] 処理結果出力部 1 8 は、産生安定性予測モデル 1 6 の予測結果を含む処理結果を出力する。処理結果出力部 1 8 は、例えば、処理結果を表示させる処理、処理結果をデータベース等に記録する処理、および処理結果を印刷させる処理のうち少なくとも 1 つの処理を行う構成であってよい。
- [0103] 図 1 3 は、情報処理装置 1 0 のハードウェア構成の例を示すブロック図である。ここでは、1 台のコンピュータを用いて情報処理装置 1 0 の処理機能を実現する例を述べるが、情報処理装置 1 0 の処理機能は、複数台のコンピュータを用いて構成されるコンピュータシステムによって実現してもよい。
- [0104] 情報処理装置 1 0 は、プロセッサ 1 0 2 と、非一時的な有体物であるコンピュータ可読媒体 1 0 4 と、通信インターフェース 1 0 6 と、入出力インターフェース 1 0 8 と、バス 1 1 0 と、を備える。プロセッサ 1 0 2 は、バス 1 1 0 を介してコンピュータ可読媒体 1 0 4、通信インターフェース 1 0 6 および入出力インターフェース 1 0 8 と接続される。
- [0105] プロセッサ 1 0 2 は CPU (Central Processing Unit) を含む。プロセッサ 1 0 2 は GPU (Graphics Processing Unit) を含んでもよい。コンピュータ可読媒体 1 0 4 は、主記憶装置であるメモリ 1 1 2 および補助記憶装置であるストレージ 1 1 4 を含む。コンピュータ可読媒体 1 0 4 は、例えば、

半導体メモリ、ハードディスク (Hard Disk Drive : HDD) 装置、もしくはソリッドステートドライブ (Solid State Drive : SSD) 装置またはこれらの複数の組み合わせであってよい。コンピュータ可読媒体 104 は本開示における「記憶装置」の一例である。

- [0106] コンピュータ可読媒体 104 は、1種類以上のクローンの培養データおよび遺伝子発現データなどの各種のデータを記憶するデータ記憶領域 120 を含む。また、コンピュータ可読媒体 104 には、予測対象限定プログラム 140、産生安定性予測モデル 16、処理結果出力プログラム 180 および表示制御プログラム 190 を含む複数のプログラム、並びにデータ等が記憶される。「プログラム」という用語はプログラムモジュールの概念を含み、プログラムに準じる命令を含む。プロセッサ 102 は、コンピュータ可読媒体 104 に記憶されたプログラムの命令を実行することにより、各種の処理部として機能する。
- [0107] 予測対象限定プログラム 140 は、培養データを解析して予測対象を限定する処理を実行させる命令を含む。予測対象限定プログラム 140 は、培養データ解析プログラム 142 と限定条件判定プログラム 144 とを含んで構成されてもよい。培養データ解析プログラム 142 は、1種類以上のクローンの培養データを解析する処理を実行させる命令を含む。培養データ解析プログラム 142 は、データセットから予測対象を絞り込むための指標と閾値を探索する処理を実行させる命令を含んでもよい。
- [0108] 限定条件判定プログラム 144 は、培養データ解析プログラム 142 の解析結果を利用し、限定条件として定められた指標と閾値に基づいて予測対象を限定する処理を実行させる命令とを含む。
- [0109] 産生安定性予測モデル 16 は、限定条件を満たす予測対象に係るクローンの遺伝子発現データの入力を受け付けて、産生安定性を予測する処理を実行させる命令を含む。
- [0110] 処理結果出力プログラム 180 は、産生安定性予測モデル 16 によって予測された産生安定性を含む処理結果を出力する処理を実行させる命令を含む

。表示制御プログラム190は、表示装置154への表示出力に必要な表示用信号を生成し、表示装置154の表示制御を実行させる命令を含む。

[0111] 通信インターフェース106は、有線または無線により外部装置との通信処理を行い、外部装置との間で情報のやり取りを行う。情報処理装置10は、通信インターフェース106を介して不図示の通信回線に接続される。通信回線は、ローカルエリアネットワークであってもよいし、ワイドエリアネットワークであってもよく、これらの組み合わせであってもよい。通信インターフェース106は、データの入力を受け付けるデータ取得部の役割を担うことができる。

[0112] 情報処理装置10は、入力装置152と、表示装置154とを備えていてもよい。入力装置152は、例えば、キーボード、マウス、マルチタッチパネル、もしくはその他のポインティングデバイス、もしくは、音声入力装置、またはこれらの適宜の組み合わせによって構成される。表示装置154は、例えば、液晶ディスプレイ、有機EL (organic electro-luminescence: OEL) ディスプレイ、もしくは、プロジェクタ、またはこれらの適宜の組み合わせによって構成される。入力装置152と表示装置154とは、入出力インターフェース108を介してプロセッサ102と接続される。なお、タッチパネルのように入力装置152と表示装置154とが一体的に構成されてもよく、タッチパネル式のタブレット端末のように、情報処理装置10と入力装置152と表示装置154とが一体的に構成されてもよい。

[0113] 《機械学習装置の構成例》

図14は、産生安定性予測モデル16を生成するための機械学習の処理を実行する機械学習装置300のハードウェア構成の例を示すブロック図である。ここでは、1台のコンピュータを用いて機械学習装置300の処理機能を実現する例を述べるが、機械学習装置300の処理機能は、複数台のコンピュータを用いて構成されるコンピュータシステムによって実現してもよい。

[0114] 機械学習装置300は、プロセッサ302と、非一時的な有体物であるコ

ンピュータ可読媒体304と、通信インターフェース306と、入出力インターフェース308と、バス310と、を備える。コンピュータ可読媒体304は、メモリ312およびストレージ314を含む。プロセッサ302は、バス310を介してコンピュータ可読媒体304、通信インターフェース306および入出力インターフェース308と接続される。入力装置352および表示装置354は入出力インターフェース308を介してバス310に接続される。

[0115] 機械学習装置300のハードウェア構成は、図6で説明した情報処理装置10の対応する要素と同様であってよい。機械学習装置300の形態は、サーバコンピュータであってもよいし、パーソナルコンピュータであってもよく、ワークステーションであってもよい。機械学習装置300は本開示における「1つ以上のプロセッサを含むシステム」の一例である。

[0116] 機械学習装置300は、通信インターフェース306を介して不図示の通信回線に接続され、データ保存部550などの外部装置と通信可能に接続される。データ保存部550は、複数の訓練データを含むデータセットが保存されているストレージを含む。データ保存部550には、図6に例示したような複数のドメインの全データを含むデータセットが保存されていてもよいし、予測対象として限定された対象のサンプルのみのデータを含むデータセットが保存されていてもよい。なお、データ保存部550は、機械学習装置300内のストレージ314に構築されてもよい。

[0117] コンピュータ可読媒体304には、予測対象限定プログラム320、学習処理プログラム330および表示制御プログラム340を含む複数のプログラム並びにデータ等が記憶される。予測対象限定プログラム320は、図12で説明した予測対象限定プログラム140と同様であってよい。表示制御プログラム340は、図12で説明した表示制御プログラム190と同様であってよい。

[0118] コンピュータ可読媒体304は、予測対象データ記憶領域322を含む。予測対象データ記憶領域322には、限定された予測対象に該当する訓練デ

ータが記憶される。データ保存部550に保存されているデータセットから予測対象限定プログラム320によって該当する訓練データが適時にサンプリングされてもよいし、予め予測対象のみのデータセットがサブセットとして抽出されていてもよい。

[0119] 学習処理プログラム330は、データ取得プログラム400と、機械学習モデルである予測モデル410と、損失算出プログラム430と、オプティマイザ440と、を含む。データ取得プログラム400は、予測対象データ記憶領域322から訓練データを取得する処理を実行させる命令を含む。データ取得プログラム400を介して取得された訓練データは予測モデル410に入力される。

[0120] 損失算出プログラム430は、予測モデル410から出力される安定性ラベルの予測値と、正解の安定性ラベルとの誤差を示す損失を算出する処理を実行させる命令を含む。オプティマイザ440は、算出された損失から予測モデル410のパラメータの更新量を算出し、予測モデル410のパラメータを更新する処理を実行させる命令を含む。オプティマイザ440は、例えば確率的勾配降下法 (Stochastic Gradient Descent : SGD) などの手法により、パラメータの最適化を行ってもよい。

[0121] 《機械学習方法のフローチャート》

図15は、機械学習装置300が実行する機械学習方法の例を示すフローチャートである。ここでは、図6に例示したような、機械学習に用いるデータセットが用意されているものとして説明する。ステップS102において、プロセッサ302は、用意されたデータセットから培養データを取得する。

[0122] ステップS104において、プロセッサ302は、培養データを解析し、訓練対象を限定する。プロセッサ302は、予め指定された培養データの指標と閾値に従い、限定条件を満たす対象サンプルのデータであるか、限定条件を満たさない対象外サンプルのデータであるかを選別してもよいし、培養データから限定条件とする指標と閾値を探索し、対象サンプルのデータと対

象外サンプルのデータとを選別してもよい。

- [0123] ステップS106において、プロセッサ302は、限定条件を満たすクロンのデータのみを用いて機械学習を行い、予測モデル410を訓練する。すなわち、プロセッサ302は、限定条件を満たすサンプルの遺伝子発現データを予測モデル410に入力し、予測モデル410から出力される安定性ラベルの予測値と、正解の安定性ラベルとの誤差を示す損失を算出する。プロセッサ302は、算出された損失に基づき予測モデル410のパラメータの更新量を算出し、パラメータを更新する。こうして、プロセッサ302は、予測モデル410に入力したデータに対する予測モデル410からの出力（予測値）が正解の安定性ラベルに近づくように予測モデル410を訓練する。なお、予測モデル410のパラメータの更新はミニバッチの単位で実施されてもよい。
- [0124] ステップS108において、プロセッサ302は、学習を終了するか否かを判定する。学習の終了条件は、損失の値に基づいて定められていてもよいし、パラメータの更新回数に基づいて定められていてもよい。損失の値に基づく方法としては、例えば、損失が規定の範囲内に収束していることを学習終了条件としてよい。また、更新回数に基づく方法としては、例えば、更新回数が規定回数に到達したことを学習終了条件としてよい。あるいは、訓練データとは別にモデルの性能評価用のデータセットを用意しておき、評価用のデータを用いた評価値に基づいて学習終了の可否を判定してもよい。
- [0125] ステップS108の判定結果がNo判定である場合、プロセッサ302はステップS106に戻り、学習処理を継続する。一方、ステップS108の判定結果がYes判定である場合、プロセッサ302は図12のフローチャートを終了する。
- [0126] 学習済みの予測モデル410は、産生安定性予測モデル16として情報処理装置10に組み込まれる。機械学習装置300が実行する機械学習方法は、産生安定性予測モデル16を生成する方法と理解することができ、本開示における予測モデル生成方法の一例である。

[0127] 《産生安定性の予測を行う情報処理方法のフローチャート》

図16は、情報処理装置10が実行する情報処理方法の例を示すフローチャートである。ステップS202において、プロセッサ102は、有用物質を産生するクローンについて測定された培養データを取得する。プロセッサ102は、不図示のデータ保存サーバなどからデータを自動的に取得してもよいし、ユーザインターフェースを介してデータの指定の入力を受け付け、指定されたクローンについてのデータを取得してもよい。

[0128] ステップS204において、プロセッサ102は、培養データを解析し、予測対象を限定する。プロセッサ102は、産生安定性予測モデル16を訓練した際に訓練対象を限定した条件と同じ限定条件を適用して予測対象を限定する。なお、このステップS204により、予測対象が限定された後に、予測対象に該当するクローンについて遺伝子発現データの計測を実施することにより、全クローンの遺伝子解析を実施する場合と比較して、作業負荷およびコストの低減が可能である。

[0129] ステップS206において、プロセッサ102は、予測対象に該当するクローンの遺伝子発現データを産生安定性予測モデル16に入力し、産生安定性予測モデル16によって安定性を予測する。

[0130] ステップS208において、プロセッサ102は、産生安定性予測モデル16から出力された予測結果を出力する。この産生安定性の予測結果を基に、産生クローンの選抜を行うことができる。

[0131] ステップS208の後、プロセッサ102は、図16のフローチャートを終了する。

[0132] 《コンピュータを動作させるプログラムについて》

実施形態に係る情報処理装置10および機械学習装置300の各装置における処理機能の一部または全部をコンピュータに実現させるプログラムを、光ディスク、磁気ディスク、もしくは、半導体メモリその他の有体物たる非一時的な情報記憶媒体であるコンピュータ可読媒体に記録し、この情報記憶媒体を通じてプログラムを提供することが可能である。

- [0133] またこのような有体物たる非一時的なコンピュータ可読媒体にプログラムを記憶させて提供する態様に代えて、インターネットなどの電気通信回線を利用してプログラム信号をダウンロードサービスとして提供することも可能である。
- [0134] さらに、上述の各装置における処理機能の一部または全部をクラウドコンピューティングによって実現してもよく、また、SaaS (Software as a Service) として提供することも可能である。
- [0135] 《各処理部のハードウェア構成について》
- 情報処理装置10におけるデータ取得部12、予測対象限定部14、産生安定性予測モデル16を含む安定性予測部、処理結果出力部18、培養データ解析部20、限定条件判定部22、機械学習装置300における予測モデル410を含む学習部、損失算出部、パラメータ更新量算出部、パラメータ更新部などの各種の処理を実行する処理部 (processing unit) のハードウェア的な構造は、例えば、次に示すような各種のプロセッサ (processor) である。
- [0136] 各種のプロセッサには、プログラムを実行して各種の処理部として機能する汎用的なプロセッサであるCPU、GPU、FPGA (Field Programmable Gate Array) などの製造後に回路構成を変更可能なプロセッサであるプログラマブルロジックデバイス (Programmable Logic Device: PLD)、ASIC (Application Specific Integrated Circuit) などの特定の処理を実行させるために専用に設計された回路構成を有するプロセッサである専用電気回路などが含まれる。
- [0137] 1つの処理部は、これら各種のプロセッサのうちの1つで構成されていてもよいし、同種または異種の2つ以上のプロセッサで構成されてもよい。例えば、1つの処理部は、複数のFPGA、あるいは、CPUとFPGAの組み合わせ、またはCPUとGPUの組み合わせによって構成されてもよい。また、複数の処理部を1つのプロセッサで構成してもよい。複数の処理部を1つのプロセッサで構成する例としては、第一に、クライアントやサーバな

どのコンピュータに代表されるように、1つ以上のCPUとソフトウェアの組み合わせで1つのプロセッサを構成し、このプロセッサが複数の処理部として機能する形態がある。第二に、システムオンチップ (System On Chip : SoC) などに代表されるように、複数の処理部を含むシステム全体の機能を1つのIC (Integrated Circuit) チップで実現するプロセッサを使用する形態がある。このように、各種の処理部は、ハードウェア的な構造として、上記各種のプロセッサを1つ以上用いて構成される。

[0138] さらに、これらの各種のプロセッサのハードウェア的な構造は、より具体的には、半導体素子などの回路素子を組み合わせた電気回路 (circuitry) である。

[0139] 《実施形態の利点》

上述した実施形態に係る産生クローンの産生安定性を予測する方法およびその方法を実行する情報処理装置10によれば、次のような効果が得られる。

[0140] [1] 現時点 (規格試験時) の培養データの指標に基づき適切に予測対象のクローンが限定されるため、予測対象のクローンについて高い精度で産生安定性を予測することができる。

[0141] [2] 予測対象のクローンに限定して遺伝子解析 (RNA-Seq解析) を行えばよいため、全クローンについて遺伝子解析を行う場合と比較して、コスト抑制が可能である。

[0142] [3] 従来の安定性試験の代わりに、本実施形態に係る方法を適用することにより、産生細胞の開発工程の期間短縮および低コスト化を実現できる。

[0143] 《その他》

本開示は上述した実施形態に限定されるものではなく、本開示の技術的思想の趣旨を逸脱しない範囲で種々の変形が可能である。

## 符号の説明

[0144] 10 情報処理装置  
12 データ取得部

- 1 4 予測対象限定部
- 1 6 産生安定性予測モデル
- 1 8 処理結果出力部
- 2 0 培養データ解析部
- 2 2 限定条件判定部
- 1 0 2 プロセッサ
- 1 0 4 コンピュータ可読媒体
- 1 0 6 通信インターフェース
- 1 0 8 入出力インターフェース
- 1 1 0 バス
- 1 1 2 メモリ
- 1 1 4 ストレージ
- 1 2 0 データ記憶領域
- 1 4 0 予測対象限定プログラム
- 1 4 2 培養データ解析プログラム
- 1 4 4 限定条件判定プログラム
- 1 5 2 入力装置
- 1 5 4 表示装置
- 1 8 0 処理結果出力プログラム
- 1 9 0 表示制御プログラム
- 3 0 0 機械学習装置
- 3 0 2 プロセッサ
- 3 0 4 コンピュータ可読媒体
- 3 0 6 通信インターフェース
- 3 0 8 入出力インターフェース
- 3 1 0 バス
- 3 1 2 メモリ
- 3 1 4 ストレージ

- 3 2 0 予測対象限定プログラム
- 3 2 2 予測対象データ記憶領域
- 3 3 0 学習処理プログラム
- 3 4 0 表示制御プログラム
- 3 5 2 入力装置
- 3 5 4 表示装置
- 4 0 0 データ取得プログラム
- 4 1 0 予測モデル
- 4 3 0 損失算出プログラム
- 4 4 0 オプティマイザ
- 5 5 0 データ保存部
- DSA、DSB データセット
- DS c、DS d、DS e データセット
- F 5 A 左図
- F 5 B 右図
- G 1 グラフ
- G 2 グラフ
- G E P 遺伝子発現パターン
- MLM 機械学習モデル
- MLM c、MLM e 機械学習モデル
- RF 1～RF 7 矩形枠
- S 1 0 2～S 1 0 8 機械学習方法のステップ
- S 2 0 2～S 2 0 8 産生安定性を予測する情報処理方法のステップ

## 請求の範囲

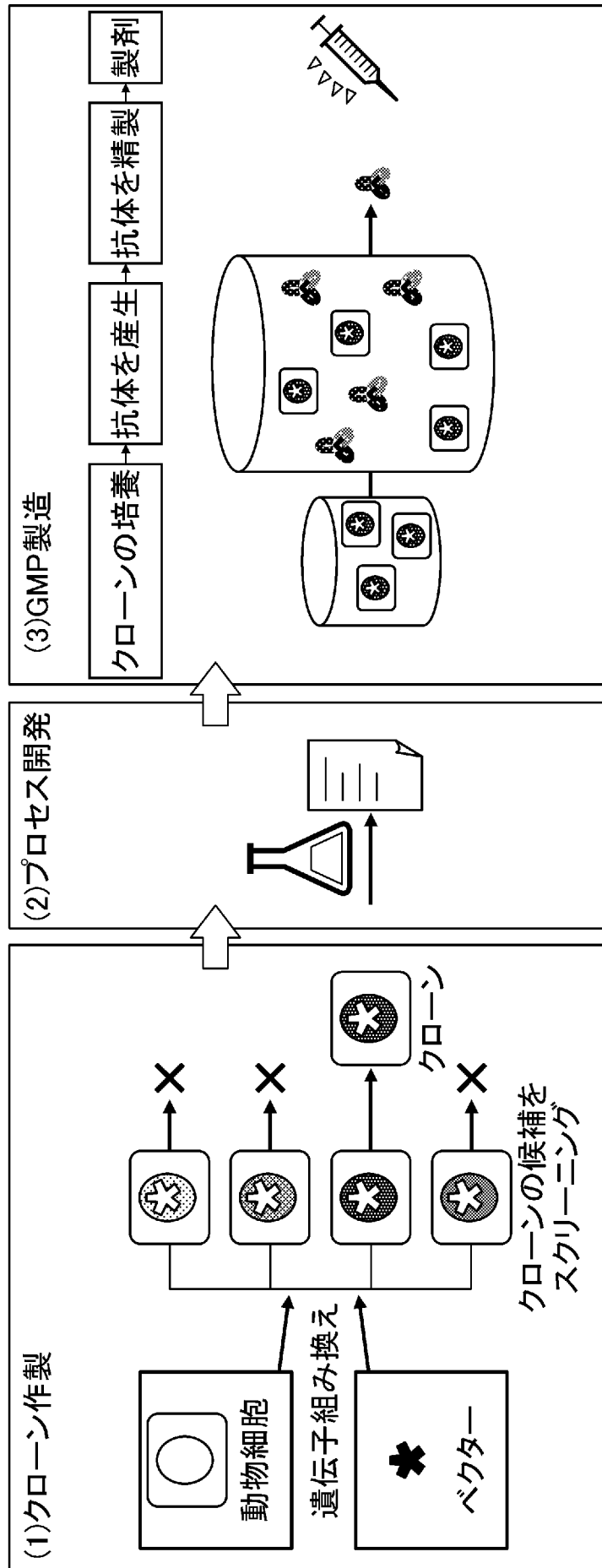
- [請求項1] 有用物質を産生するクローンの産生安定性を予測する方法であって、
- 、
- 1つ以上のプロセッサが、
- 1種類以上の前記クローンの培養データを取得することと、
- 前記培養データを解析して予測対象のクローンを限定することと、
- 前記予測対象のクローンについて測定されたデータを用いて、前記予測対象のクローンによる前記有用物質の産生安定性を予測することと、
- を実行する、方法。
- [請求項2] 前記産生安定性は、培養開始時と所定期間培養後における前記有用物質の産生量の変化の有無により定義される、
- 請求項1に記載の方法。
- [請求項3] 前記1つ以上のプロセッサが、
- 前記培養データから得られる指標と、前記指標に関する閾値とを設定し、
- 前記指標の値と前記閾値とに基づき前記予測対象を限定する、
- 請求項1に記載の方法。
- [請求項4] 前記閾値は、前記産生安定性の予測精度が前記予測対象を限定しない場合よりも高くなるように調整される、
- 請求項3に記載の方法。
- [請求項5] 前記閾値は、前記指標の値についての順位を用いて定義される、
- 請求項3に記載の方法。
- [請求項6] 前記予測対象は、前記指標の値の上位集団である、
- 請求項3に記載の方法。
- [請求項7] 前記指標は、前記有用物質の産生量である、
- 請求項3から6のいずれか一項に記載の方法。
- [請求項8] 前記指標は、積分生存細胞密度である、

- 請求項 3 から 6 のいずれか一項に記載の方法。
- [請求項9] 前記指標は、乳酸濃度である、  
請求項 3 から 6 のいずれか一項に記載の方法。
- [請求項10] 前記産生安定性の予測に用いる前記データは、1 つ以上の遺伝子発現レベルを含む、  
請求項 1 から 6 のいずれか一項に記載の方法。
- [請求項11] 前記 1 つ以上のプロセッサが、  
前記予測対象の前記データの入力を受けて、安定または不安定の 2 クラス分類を行うモデルを用いて前記産生安定性を予測する、  
請求項 1 から 6 のいずれか一項に記載の方法。
- [請求項12] 前記モデルは、前記予測対象のクローンと同様の限定をした訓練用のクローンについての前記データと正解の安定性ラベルとが関連付けられた複数の訓練データを用いた機械学習によって訓練されたモデルである、  
請求項 1 1 に記載の方法。
- [請求項13] 前記複数の訓練データは、産生する有用物質が異なる複数種類のクローンについての前記訓練データを含み、  
前記 1 つ以上のプロセッサが、前記モデルの訓練に使用された有用物質とは別の有用物質を産生するクローンについての産生安定性を予測する、  
請求項 1 2 に記載の方法。
- [請求項14] 前記有用物質は、医薬品原料であるタンパク質、ペプチド、およびウイルスのうちいずれかである、  
請求項 1 から 6 のいずれか一項に記載の方法。
- [請求項15] 前記有用物質は、抗体、または抗体様タンパク質である、  
請求項 1 から 6 のいずれか一項に記載の方法。
- [請求項16] 前記クローンは、脊椎動物由来細胞である、  
請求項 1 から 6 のいずれか一項に記載の方法。

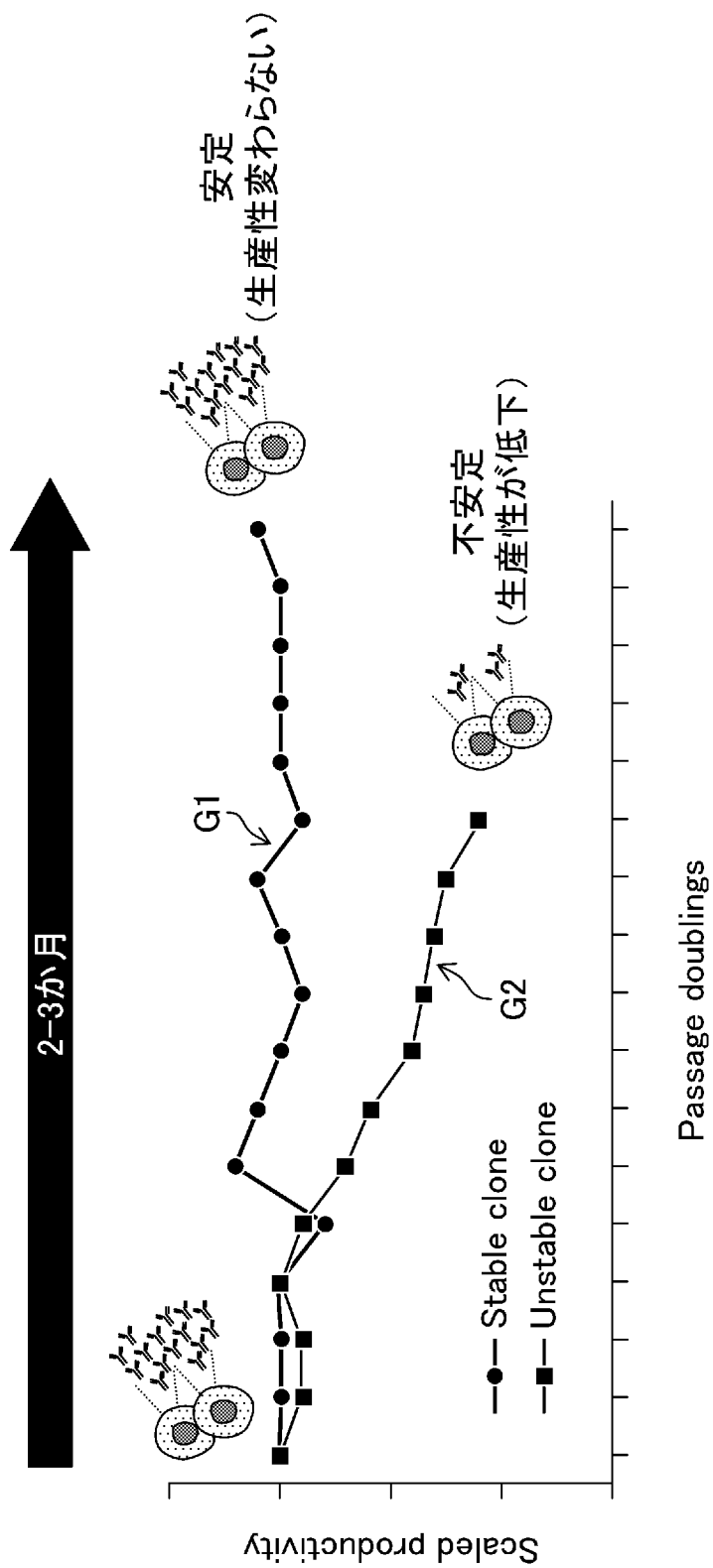
- [請求項17] 前記クローンは、哺乳類由来細胞である、  
請求項1から6のいずれか一項に記載の方法。
- [請求項18] 前記クローンは、CHO細胞またはHEK細胞である、  
請求項1から6のいずれか一項に記載の方法。
- [請求項19] 1つ以上のプロセッサと、  
前記1つ以上のプロセッサに実行させる命令が記憶される1つ以上の記憶装置と、を備え、  
前記1つ以上のプロセッサは、  
有用物質を産生するクローンについて1種類以上のクローンの培養データを取得し、  
前記培養データを解析して予測対象のクローンを限定し、  
前記予測対象のクローンについて測定されたデータを用いて、前記予測対象のクローンによる前記有用物質の産生安定性を予測する、  
情報処理装置。
- [請求項20] コンピュータに、  
有用物質を産生するクローンについて1種類以上のクローンの培養データを取得する機能と、  
前記培養データを解析して予測対象のクローンを限定する機能と、  
前記予測対象のクローンについて測定されたデータを用いて、前記予測対象のクローンによる前記有用物質の産生安定性を予測する機能と、  
を実現させるプログラム。
- [請求項21] 非一時的かつコンピュータ読取可能な記録媒体であって、請求項20に記載のプログラムが記録された記録媒体。
- [請求項22] 有用物質を産生するクローンの産生安定性を予測する機能をコンピュータに実現させる予測モデルを生成する予測モデル生成方法であって、  
1つ以上のプロセッサを含むシステムが、

1 種類以上の前記クローンの培養データを取得することと、  
前記培養データを解析して予測対象のクローンを限定することと、  
前記予測対象に該当するクローンについて測定されたデータと正解の安定性ラベルとが関連付けされた複数の訓練データを用いて機械学習を行い、前記データの入力に対する前記予測モデルの出力が前記正解の安定性ラベルに近づくように前記予測モデルを訓練することと、  
を含む予測モデル生成方法。

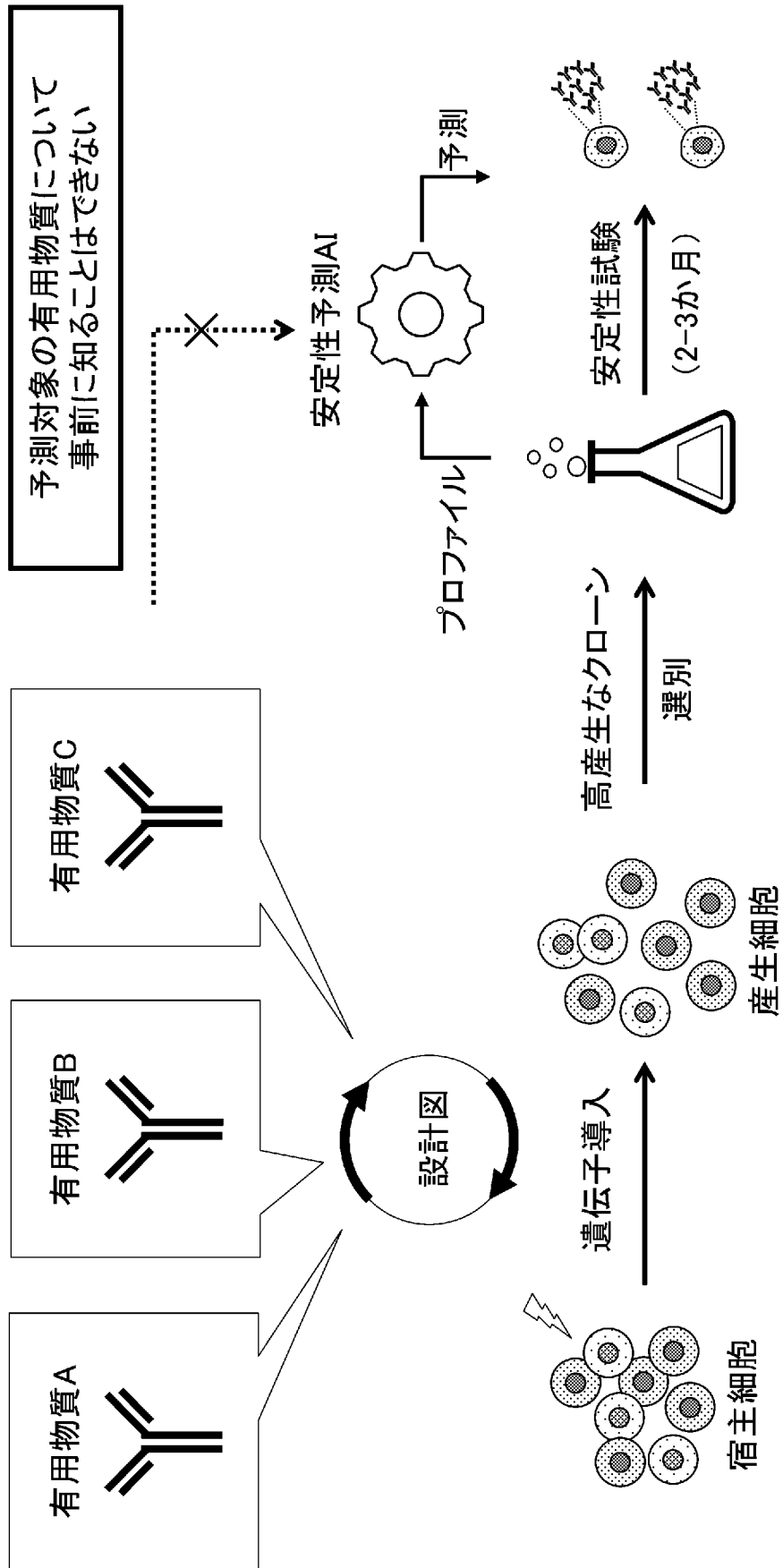
[図1]



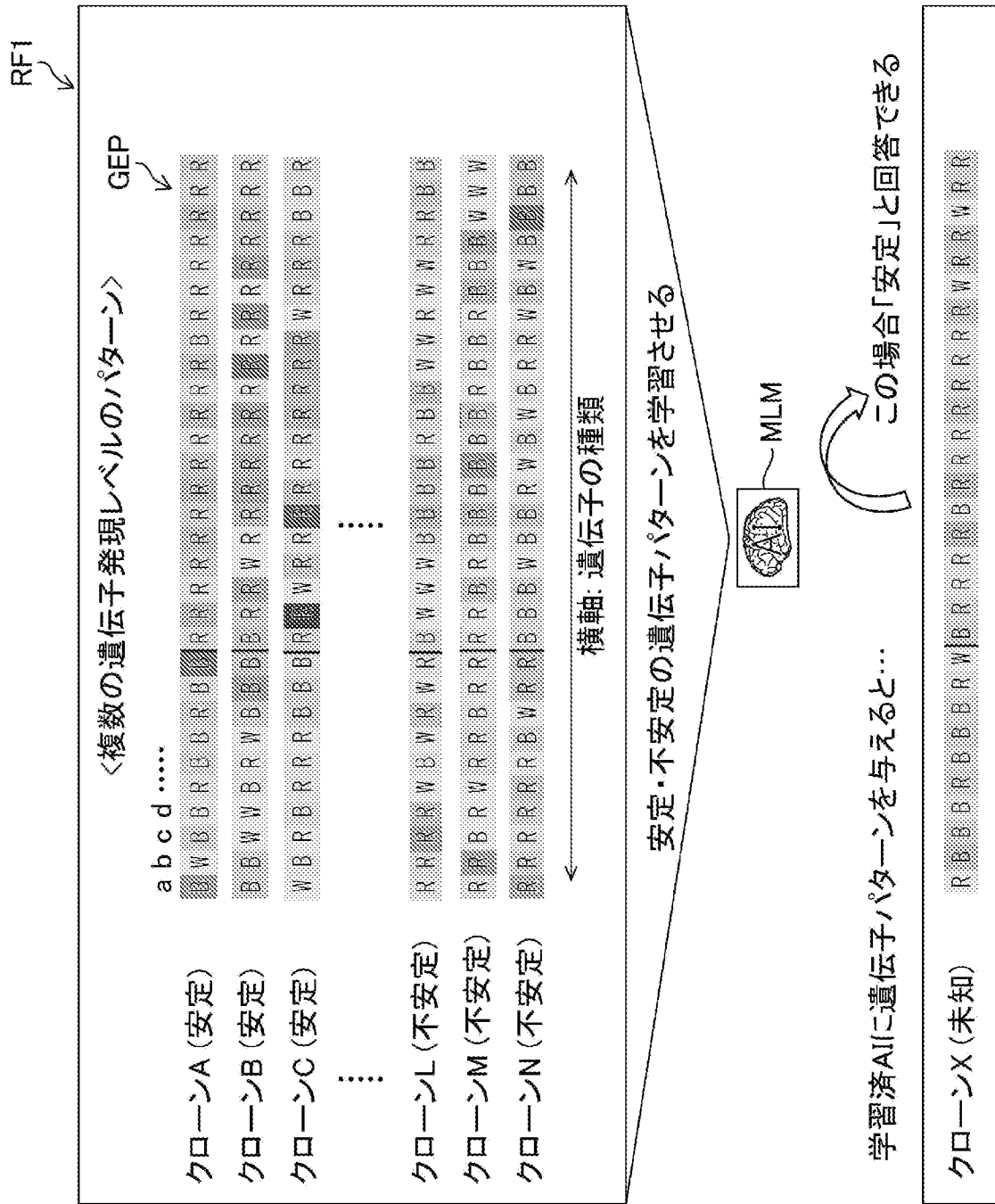
[図2]



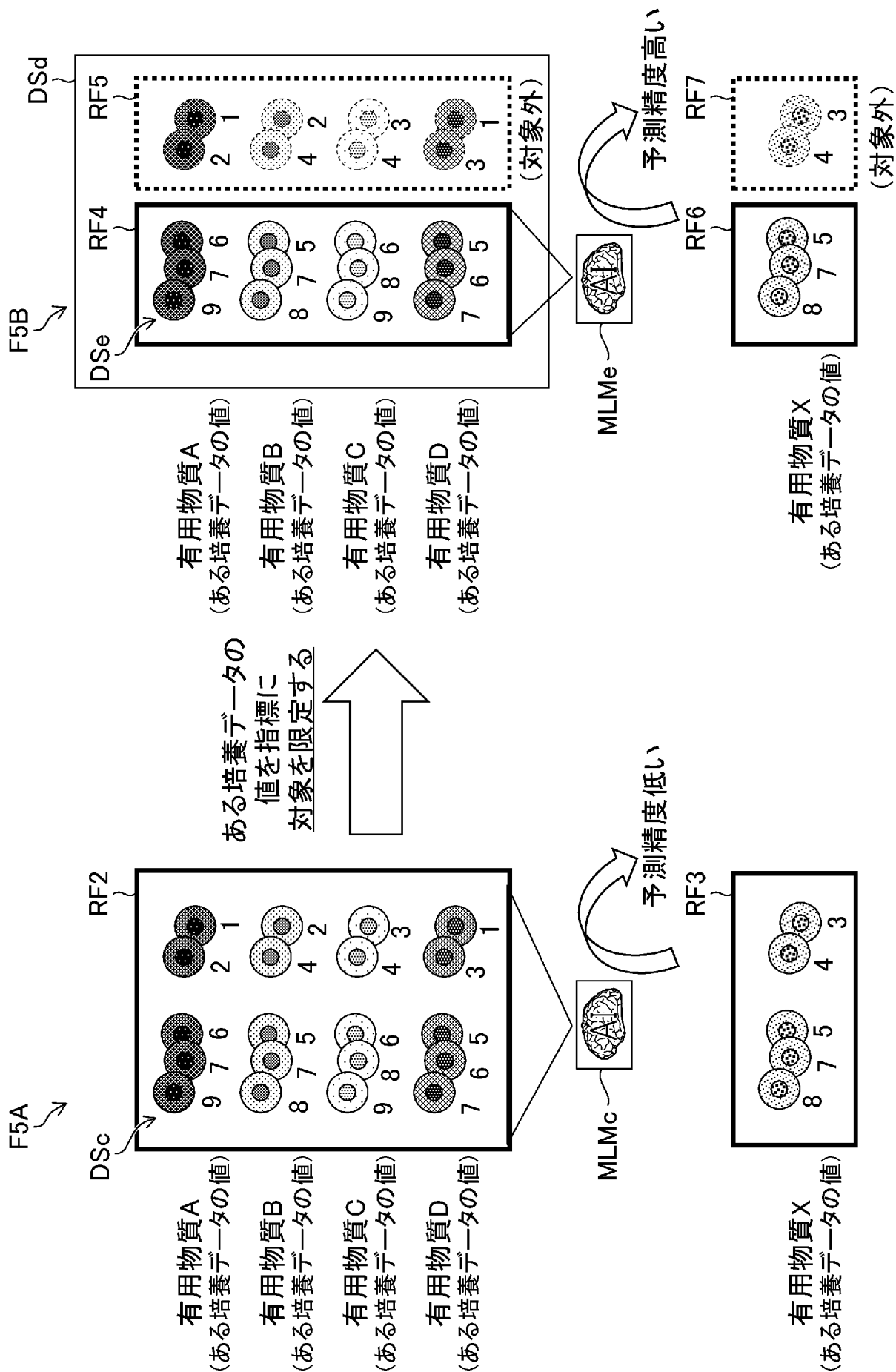
[図3]



[図4]



[図5]



[図6]

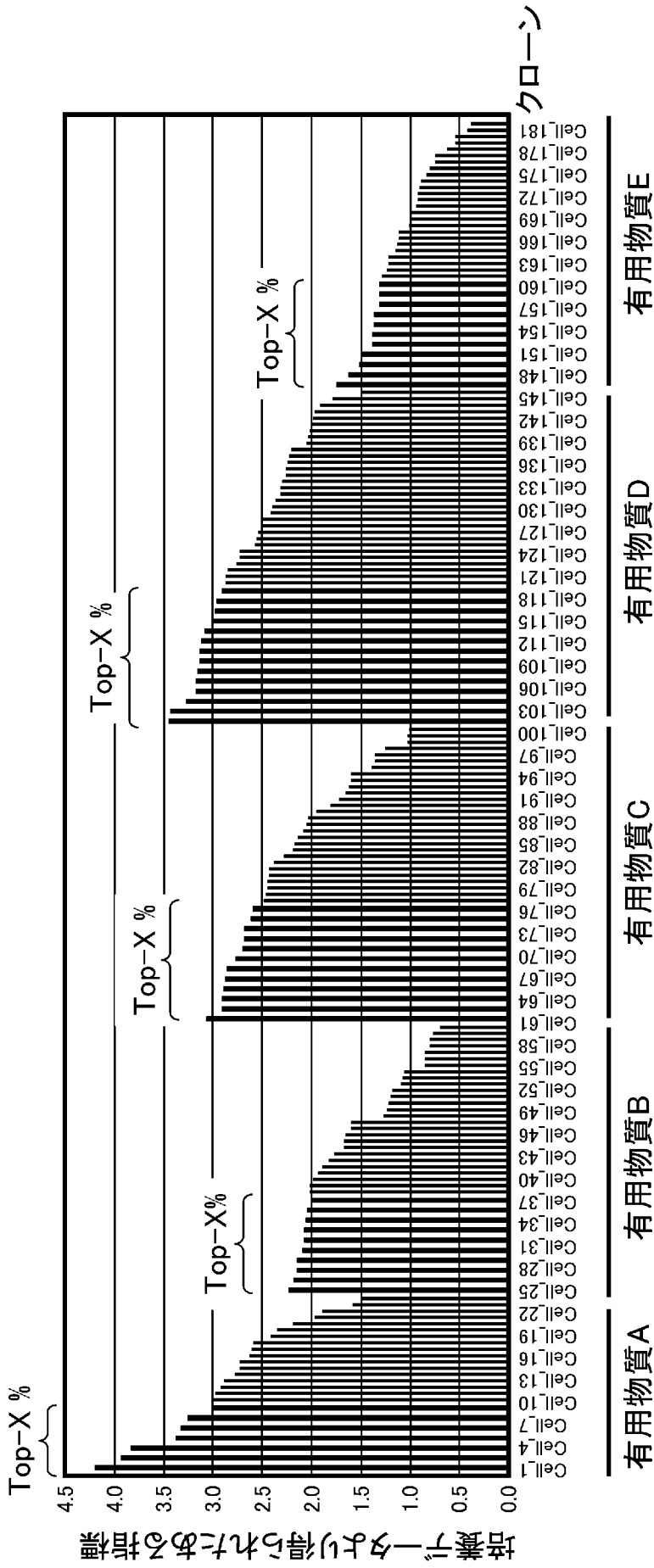
DSA 

有用物質:抗体A クローン	培養データ				遺伝子発現レベル				安定性ラベル
	抗体産生量	積分生存細胞密度	乳酸濃度	...	遺伝子1	遺伝子2	...	遺伝子N	
ACL1	APA1	CDA1	LCA1	:	G1A1	G2A1	:	GNA1	TLA1
ACL2	APA2	CDA2	LCA2	:	G1A2	G2A2	:	GNA2	TLA2
:	:	:	:	:	:	:	:	:	:
ACLj	APAj	CDAj	LCAj	:	G1Aj	G2Aj	:	GNAj	TLAj
:	:	:	:	:	:	:	:	:	:
ACLna	APAna	CDAAna	LCAAna	:	G1Ana	G2Ana	:	GNAAna	TLAna

DSB 

有用物質:抗体B クローン	培養データ				遺伝子発現レベル				安定性ラベル
	抗体産生量	積分生存細胞密度	乳酸濃度	...	遺伝子1	遺伝子2	...	遺伝子N	
BCL1	APB1	CDB1	LCB1	:	G1B1	G2B1	:	GNB1	TLB1
BCL2	APB2	CDB2	LCB2	:	G1B2	G2B2	:	GNB2	TLB2
:	:	:	:	:	:	:	:	:	:
BCLj	APBj	CDBj	LCBj	:	G1Bj	G2Bj	:	GNBj	TLBj
:	:	:	:	:	:	:	:	:	:
BCLnb	APBnb	CDBnb	LCBnb	:	G1Bnb	G2Bnb	:	GNBnb	TLBnb

[図7]



[図8]

## 抗体産生CHO細胞のクローン数と安定性ラベル

	抗体A	抗体B	抗体C	抗体D	抗体E	計
安定	7	5	20	21	23	76
不安定	17	32	20	24	13	106
計	24	37	40	45	36	182

[図9]

## 相対的高産生に限定した際のクローン数と安定性ラベル

	抗体A	抗体B	抗体C	抗体D	抗体E	計
安定	5	0	11	3	13	32
不安定	5	14	5	15	2	41
計	10	14	16	18	15	73

[図10]

## 相対的高細胞密度に限定した際のクローン数と安定性ラベル

	抗体A	抗体B	抗体C	抗体D	抗体E	計
安定	7	5	15	15	15	57
不安定	9	17	9	11	6	52
計	16	22	24	36	21	109

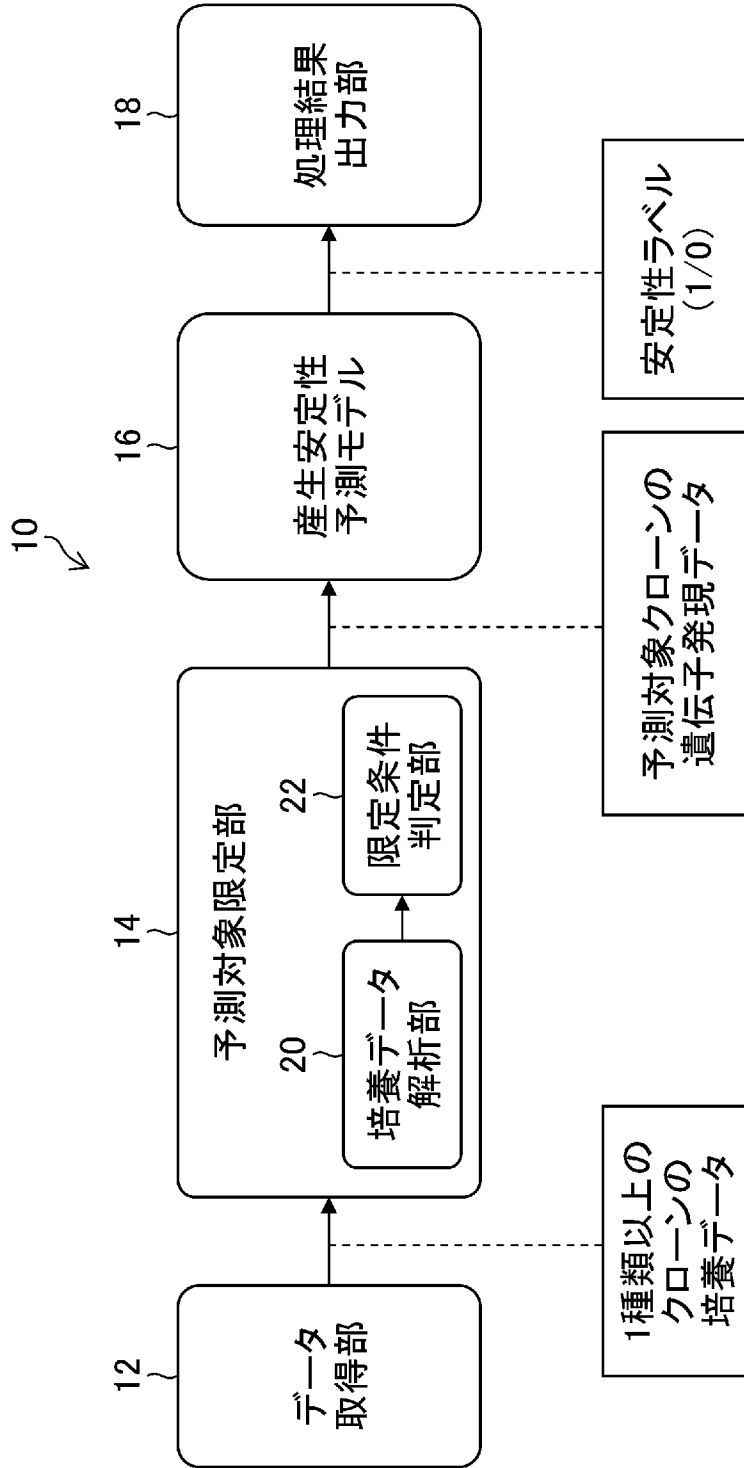
[図11]

## 相対的高乳酸濃度に限定した際のクローン数と安定性ラベル

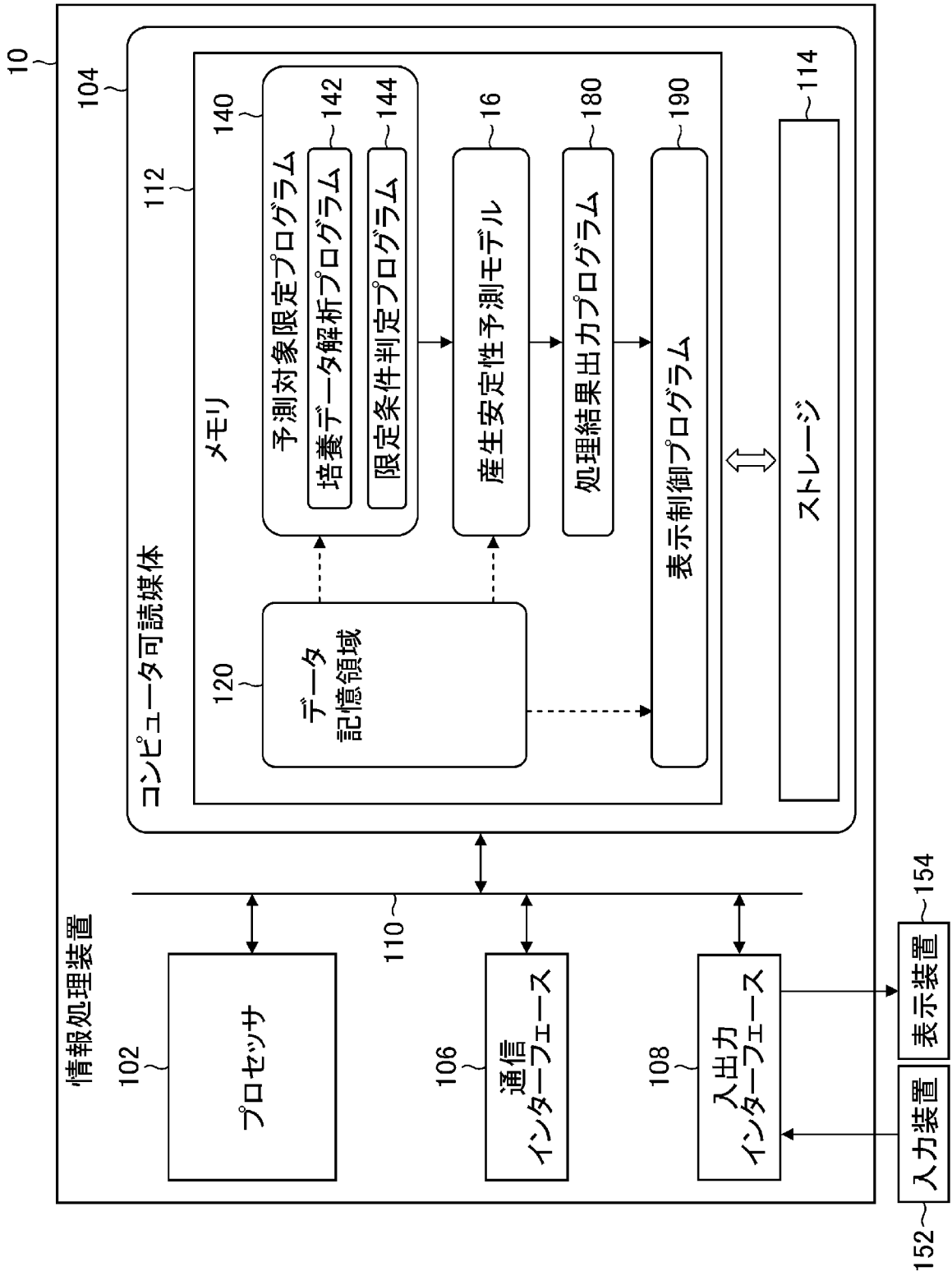
	抗体A	抗体B	抗体C	抗体D	抗体E	計
安定	5	1	7	15	8	36
不安定	5	12	9	4	6	36
計	10	13*	16	19	14	72

(\*データ欠損1クローン)

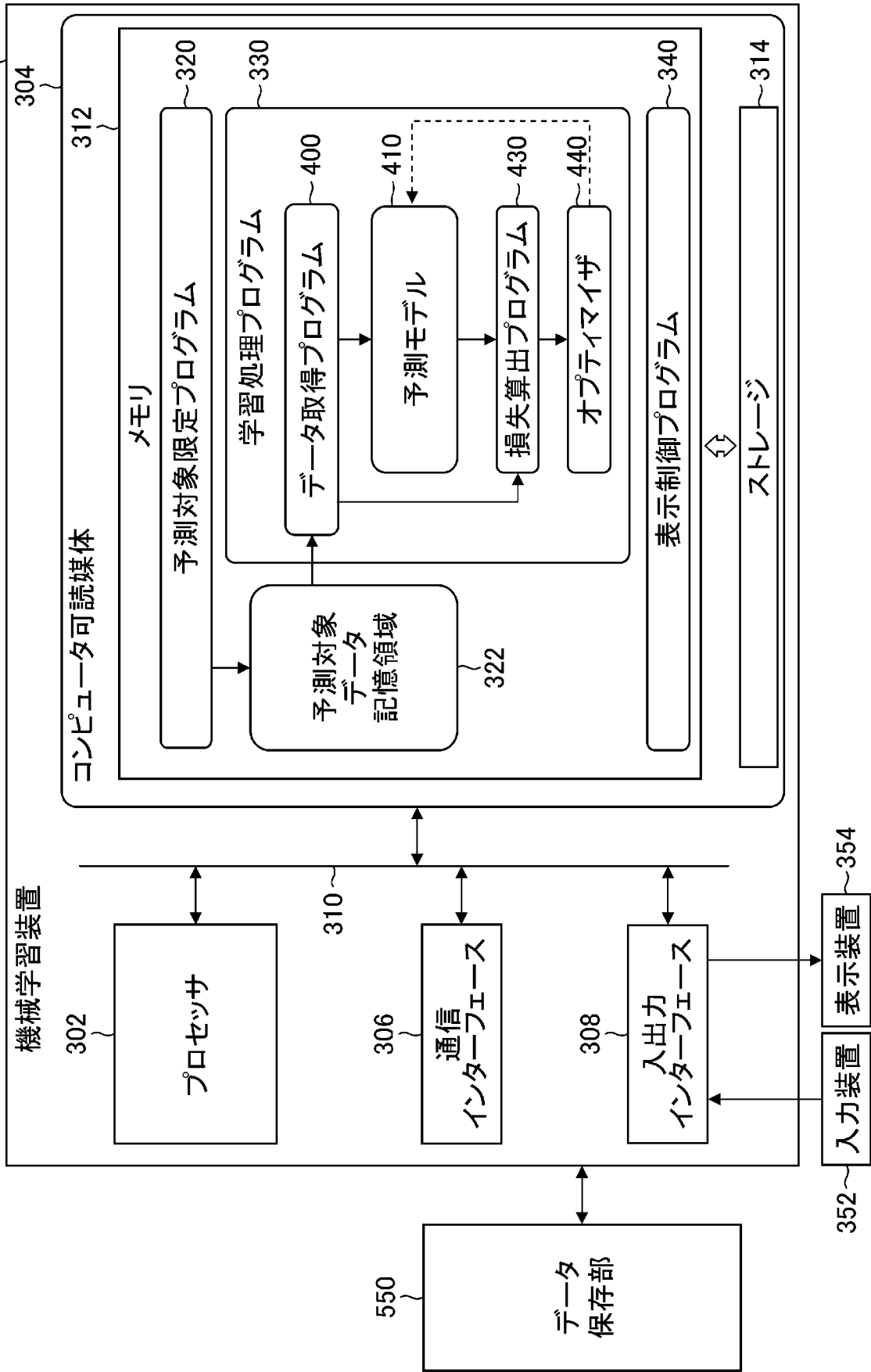
[図12]



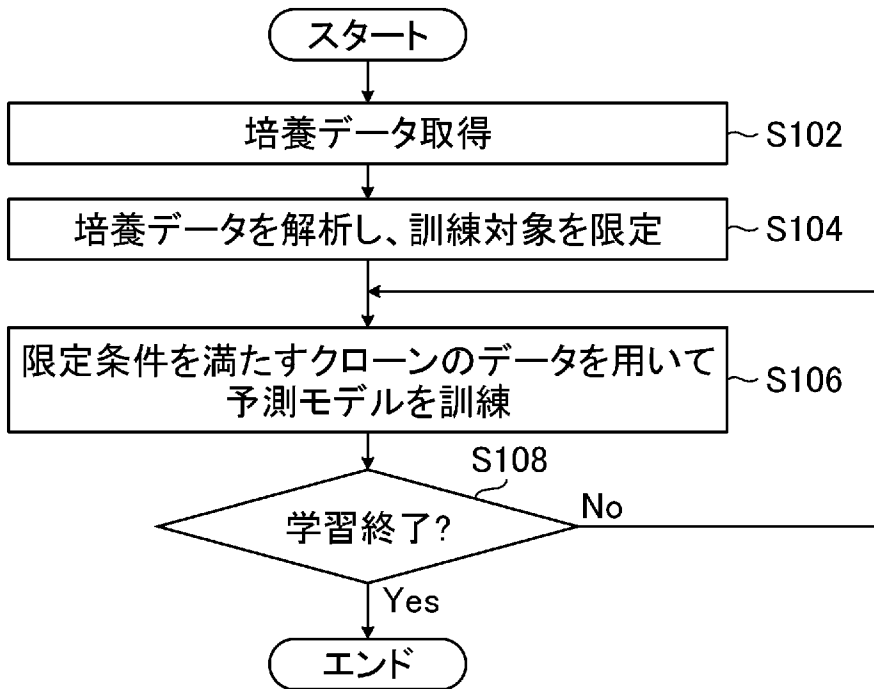
[図13]



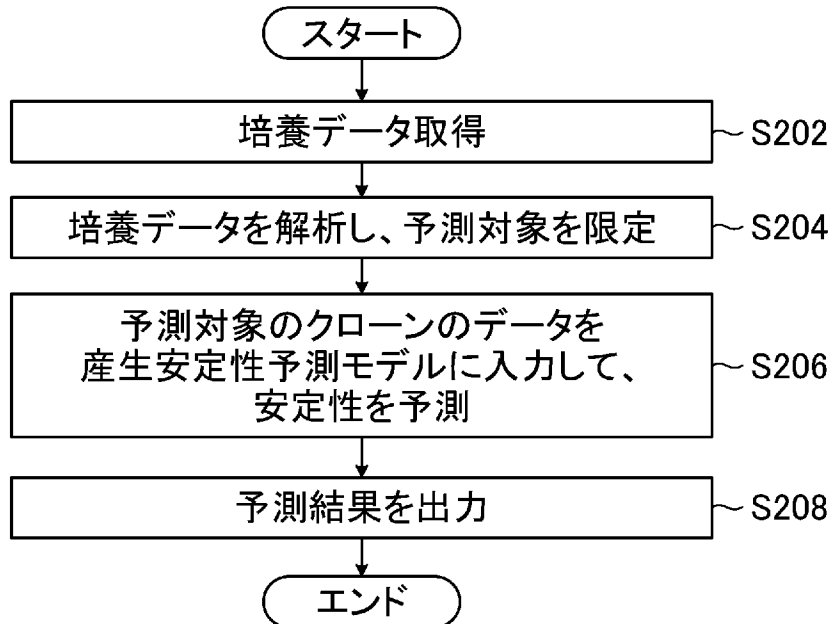
[図14]



[図15]



[図16]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/025263

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
<i>C12N 5/10</i> (2006.01)i; <i>C12N 5/071</i> (2010.01)i; <i>C12P 1/00</i> (2006.01)i; <i>C12P 21/08</i> (2006.01)i; <i>G16B 40/00</i> (2019.01)i FI: C12N5/10; G16B40/00; C12P21/08; C12P1/00; C12N5/071		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) C12N5/10; C12N5/071; C12P1/00; C12P21/08; G16B40/00		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2023 Registered utility model specifications of Japan 1996-2023 Published registered utility model applications of Japan 1994-2023		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JSTPlus/JMEDPlus/JST7580 (JDreamIII); CAlus/MEDLINE/EMBASE/BIOSIS (STN)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2022-533003 A (AMGEN INC) 21 July 2022 (2022-07-21) examples, paragraph [0083]	1-22
Y	JP 2008-502365 A (THE REGENTS OF THE UNIVERSITY OF MICHIGAN) 31 January 2008 (2008-01-31) example 1	1-22
Y	JP 2010-517530 A (CENTOCOR INC) 27 May 2010 (2010-05-27) claims	1-22
Y	JP 2021-534782 A (F. HOFFMANN-LA ROCHE AG) 16 December 2021 (2021-12-16) example 4	1-22
Y	JP 2021-503291 A (LONZA LTD) 12 February 2021 (2021-02-12) paragraph [0004]	1-22
Y	WO 2016/075216 A1 (LEK PHARMACEUTICALS D.D.) 19 May 2016 (2016-05-19) claims	1-22
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search <b>25 July 2023</b>		Date of mailing of the international search report <b>08 August 2023</b>
Name and mailing address of the ISA/JP <b>Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan</b>		Authorized officer  Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

**PCT/JP2023/025263**

<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2021/166824 A1 (FUJIFILM CORP) 26 August 2021 (2021-08-26) claims	1-22
A	JP 2017-511688 A (VALITACELL LIMITED) 27 April 2017 (2017-04-27) claims	1-22

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/JP2023/025263**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2022-533003	A	21 July 2022	US 2022/0228102 paragraph [0092], claims	A1
				WO 2020/223422	A1
				EP 3963589	A1
JP	2008-502365	A	31 January 2008	US 2006/0014138 example 1	A1
				WO 2005/123993	A2
				EP 1774020	A4
JP	2010-517530	A	27 May 2010	US 2009/0191545 claims	A1
				WO 2008/094871	A2
				EP 2129790	A4
JP	2021-534782	A	16 December 2021	US 2021/0257045 example 4	A1
				WO 2020/043601	A1
				EP 3844505	A1
				CN 112639478	A
				KR 10-2021-0035875	A
JP	2021-503291	A	12 February 2021	US 2020/0354666 paragraph [0004]	A1
				WO 2019/100040	A1
				EP 3714036	A1
				KR 10-2020-0086348	A
				CN 111630148	A
WO	2016/075216	A1	19 May 2016	US 2017/0314082 claims	A1
				EP 3218514	A1
WO	2021/166824	A1	26 August 2021	US 2022/0380717 claims	A1
				EP 4108758	A1
				CN 115135752	A
JP	2017-511688	A	27 April 2017	US 2017/0017891 claims	A1
				WO 2015/113704	A1
				EP 2902493	A1
				CN 106164285	A

<p>A. 発明の属する分野の分類（国際特許分類（IPC））                  C12N 5/10(2006.01)i; C12N 5/071(2010.01)i; C12P 1/00(2006.01)i; C12P 21/08(2006.01)i;                  G16B 40/00(2019.01)i                  FI: C12N5/10; G16B40/00; C12P21/08; C12P1/00; C12N5/071</p>										
<p>B. 調査を行った分野</p>										
<p>調査を行った最小限資料（国際特許分類（IPC））                  C12N5/10; C12N5/071; C12P1/00; C12P21/08; G16B40/00</p>										
<p>最小限資料以外の資料で調査を行った分野に含まれるもの</p> <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2023年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2023年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2023年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2023年	日本国実用新案登録公報	1996 - 2023年	日本国登録実用新案公報	1994 - 2023年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2023年									
日本国実用新案登録公報	1996 - 2023年									
日本国登録実用新案公報	1994 - 2023年									
<p>国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）                  JSTPlus/JMEDPlus/JST7580 (JDreamIII); CPlus/MEDLINE/EMBASE/BIOSIS (STN)</p>										
<p>C. 関連すると認められる文献</p>										
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号								
Y	JP 2022-533003 A (アムジエン・インコーポレーテッド) 21.07.2022 (2022 - 07 - 21) 特許請求の範囲、 [0083]	1-22								
Y	JP 2008-502365 A (ザ リージェンツ オブ ザ ユニバーシティ オブ ミシガン) 31.01.2008 (2008 - 01 - 31) 実施例 1	1-22								
Y	JP 2010-517530 A (セントカー・インコーポレーテッド) 27.05.2010 (2010 - 05 - 27) 特許請求の範囲	1-22								
Y	JP 2021-534782 A (エフ. ホフマン-ラ ロシュ アーゲー) 16.12.2021 (2021 - 12 - 16) 実施例 4	1-22								
Y	JP 2021-503291 A (ロンザ リミテッド) 12.02.2021 (2021 - 02 - 12) [0004]	1-22								
<p><input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。</p>										
<p>* 引用文献のカテゴリー                  “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの                  “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの                  “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）                  “O” 口頭による開示、使用、展示等に言及する文献                  “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献                  “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの                  “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの                  “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの                  “&amp;” 同一パテントファミリー文献</p>										
国際調査を完了した日	25.07.2023	国際調査報告の発送日 08.08.2023								
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官）  鳥居 敬司 4N 4045  電話番号 03-3581-1101 内線 3448									

C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	WO 2016/075216 A1 (LEK PHARMACEUTICALS D.D.) 19.05.2016 (2016 - 05 - 19) claims	1-22
A	WO 2021/166824 A1 (富士フイルム株式会社) 26.08.2021 (2021 - 08 - 26) 特許請求の範囲	1-22
A	JP 2017-511688 A (バリタセル リミテッド) 27.04.2017 (2017 - 04 - 27) 特許請求の範囲	1-22

国際調査報告  
 パテントファミリーに関する情報

国際出願番号

PCT/JP2023/025263

引用文献			公表日	パテントファミリー文献			公表日
JP	2022-533003	A	21.07.2022	US	2022/0228102	A1	
					[0092], claims		
				WO	2020/223422	A1	
				EP	3963589	A1	
JP	2008-502365	A	31.01.2008	US	2006/0014138	A1	
					Example 1		
				WO	2005/123993	A2	
				EP	1774020	A4	
JP	2010-517530	A	27.05.2010	US	2009/0191545	A1	
					claims		
				WO	2008/094871	A2	
				EP	2129790	A4	
JP	2021-534782	A	16.12.2021	US	2021/0257045	A1	
					Example 4		
				WO	2020/043601	A1	
				EP	3844505	A1	
				CN	112639478	A	
				KR	10-2021-0035875	A	
JP	2021-503291	A	12.02.2021	US	2020/0354666	A1	
					[0004]		
				WO	2019/100040	A1	
				EP	3714036	A1	
				KR	10-2020-0086348	A	
				CN	111630148	A	
WO	2016/075216	A1	19.05.2016	US	2017/0314082	A1	
					claims		
				EP	3218514	A1	
WO	2021/166824	A1	26.08.2021	US	2022/0380717	A1	
					claims		
				EP	4108758	A1	
				CN	115135752	A	
JP	2017-511688	A	27.04.2017	US	2017/0017891	A1	
					claims		
				WO	2015/113704	A1	
				EP	2902493	A1	
				CN	106164285	A	