

(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  
2014年1月9日(09.01.2014)



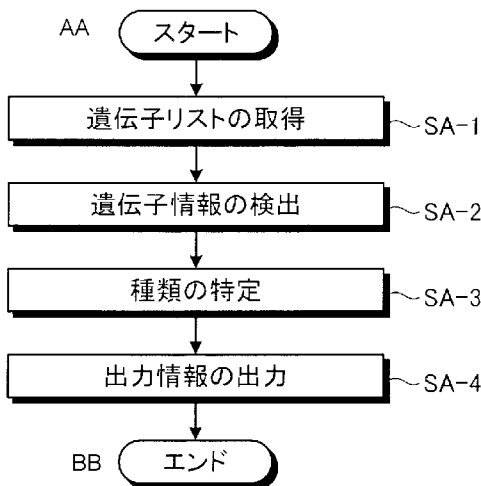
(10) 国際公開番号  
WO 2014/007363 A1

- (51) 国際特許分類:  
G06F 19/20 (2011.01)
- (21) 国際出願番号: PCT/JP2013/068489
- (22) 国際出願日: 2013年7月5日(05.07.2013)
- (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 2012-151695 2012年7月5日(05.07.2012) JP
- (71) 出願人: 独立行政法人科学技術振興機構 (NATIONAL INSTITUTE OF JAPAN SCIENCE AND TECHNOLOGY AGENCY) [JP/JP]; 〒3320012 埼玉県川口市本町4丁目1番8号 Saitama (JP).
- (72) 発明者: シューメーカー, ジェイソン (SHOE-MAKER, Jason); 〒1088639 東京都港区白金台4-6-1 独立行政法人科学技術振興機構内 Tokyo (JP). 北野 宏明 (KITANO, Hiroaki); 〒1088639 東京都港区白金台4-6-1 独立行政法人科学技術振興機構内 Tokyo (JP). 河岡 義裕 (KAWAOKA, Yoshihiro); 〒1088639 東京都港区白金台4-6-1 独立行政法人科学技術振興機構内 Tokyo (JP).
- (74) 代理人: 酒井 宏明 (SAKAI, Hiroaki); 〒1006020 東京都千代田区霞が関三丁目2番5号 霞が関ビルディング 酒井国際特許事務所 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, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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, 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).

[続葉有]

(54) Title: CELL TYPING DEVICE, CELL TYPING METHOD, AND PROGRAM

(54) 発明の名称: 細胞種解析装置、細胞種解析方法、および、プログラム



- SA-1 Acquire gene list
- SA-2 Detect genetic information
- SA-3 Specify type
- SA-4 Output information to be output
- AA Start
- BB End

(57) Abstract: The present invention acquires a gene list containing genetic information, detects stored genetic information that matches the genetic information contained in the acquired gene list, specifies the type of cell or tissue that corresponds to the gene list on the basis of the detected genetic information and stored cell or tissue information, and outputs the specified information to be output that contains cell information comprising the cell type and/or tissue information comprising the tissue type.

(57) 要約: 本発明は、遺伝子情報を含む遺伝子リストを取得し、取得された遺伝子リストに含まれる遺伝子情報と一致する、記憶された遺伝子情報を検出し、検出された遺伝子情報、および、記憶された細胞情報または組織情報に基づき、遺伝子リストに対応する細胞または組織の種類を特定し、特定された、細胞の種類を含む細胞情報、および/または、組織の種類を含む組織情報を含む出力情報を出力させる。

WO 2014/007363 A1

添付公開書類:

- 国際調査報告 (条約第 21 条(3))

## 明 細 書

発明の名称：

細胞種解析装置、細胞種解析方法、および、プログラム

### 技術分野

[0001] 本発明は、細胞種解析装置、細胞種解析方法、および、プログラムに関する。

### 背景技術

[0002] 従来から、遺伝子発現解析を行う技術が開示されている。

[0003] 非特許文献1に記載のDAVIDにおいては、遺伝子またはタンパク質から生物学的意義を系統的に抽出することを目的として統合された生物学的知識ベースおよび分析ツールから構成される技術が開示されている。

[0004] また、非特許文献2に記載のGATHERにおいては、データを統合して遺伝子グループの新規機能を発見し、遺伝子の特徴をアノテーションする技術が開示されている。

[0005] また、非特許文献3に記載の遺伝子発現バーコードにおいては、単一のマイクロアレイのデータを用いて全ての遺伝子の発現状態について精度よく分類するためのデータベースが開示されている。

[0006] また、非特許文献4に記載のBioGPSにおいては、ユーザに様々なオンライン遺伝子アノテーション手段を効果的に集め、アクセスするためのユーザインターフェースを提供する技術が開示されている。

### 先行技術文献

#### 非特許文献

[0007] 非特許文献1：Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

非特許文献2: JT Chang and JR Nevins. GATHER: A Systems Approach to Interpreting Genomic Signatures. *Bioinformatics* 22 (23), 2006.

非特許文献3: McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA. The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes *Nucleic Acids Research*. 2011 Jan; 39 (suppl 1): D1011-D1015

非特許文献4: Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodges CL, Haase J, Janes J, Huss JW 3rd, Su AI. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources, *Genome Biology*, 2009, 10: R130

## 発明の概要

### 発明が解決しようとする課題

[0008] しかしながら、非特許文献1乃至4に記載の従来 of 遺伝子発現解析技術においては、生物学的機能に基づき解析を行っているため、遺伝子発現の細胞種毎の違いを区別、決定することができないという問題点を有していた。

[0009] 本発明は、上記問題点に鑑みてなされたもので、細胞種に依存して発現量が高くなる遺伝子を細胞種のシグネチャとして利用してデータ解析を行うことができる細胞種解析装置、細胞種解析方法、および、プログラムを提供す

ることを目的とする。

### 課題を解決するための手段

[0010] このような目的を達成するため、本発明の細胞種解析装置は、出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置であって、上記記憶部は、細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶する特異的高発現記憶手段、を備え、上記制御部は、上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得手段と、上記遺伝子リスト取得手段により取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記憶された上記遺伝子情報を検出する遺伝子情報検出手段と、上記遺伝子情報検出手段により検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定手段と、上記種類特定手段により特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力手段と、を備えたことを特徴とする。

[0011] また、本発明の細胞種解析装置は、上記記載の細胞種解析装置において、上記記憶部は、上記細胞または上記組織における遺伝子発現情報を記憶する遺伝子発現情報記憶手段、を更に備え、上記制御部は、上記遺伝子発現情報記憶手段に記憶された上記細胞または上記組織における上記遺伝子発現情報に基づき、当該細胞または当該組織において特異的に高発現となる上記遺伝子を特定し、当該遺伝子に関する上記遺伝子情報と、当該細胞に関する上記細胞情報または当該組織に関する組織情報と、を対応付けて上記特異的高発現記憶手段に格納する遺伝子情報格納手段、を更に備えたことを特徴とする。

[0012] また、本発明の細胞種解析装置は、上記記載の細胞種解析装置において、上記記憶部は、複数種類の上記細胞を含む上記組織における遺伝子発現情報

を記憶する組織発現情報記憶手段、を更に備え、上記制御部は、上記組織発現情報記憶手段に記憶された上記組織における上記遺伝子発現情報に基づく遺伝子発現のユニークなパターンを識別することにより、当該組織における遺伝子発現情報を複数のクラスタにクラスタリングするクラスタリング手段、を更に備え、上記遺伝子リスト取得手段は、上記クラスタリング手段によりクラスタリングされた上記クラスタに含まれる上記遺伝子発現情報に対応する上記遺伝子情報に基づき、遺伝子セット解析を用いて上記遺伝子リストを取得することを特徴とする。

[0013] また、本発明の細胞種解析方法は、出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置において実行される細胞種解析方法であって、上記記憶部は、細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶する特異的高発現記憶手段、を備え、上記制御部において実行される、上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得ステップと、上記遺伝子リスト取得ステップにて取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記憶された上記遺伝子情報を検出する遺伝子情報検出ステップと、上記遺伝子情報検出ステップにて検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定ステップと、上記種類特定ステップにて特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力ステップと、を含むことを特徴とする。

[0014] また、本発明のプログラムは、出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置に実行させるためのプログラムであって、上記記憶部は、細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と

、を対応付けて記憶する特異的高発現記憶手段、を備え、上記制御部において、上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得ステップと、上記遺伝子リスト取得ステップにて取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記憶された上記遺伝子情報を検出する遺伝子情報検出ステップと、上記遺伝子情報検出ステップにて検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定ステップと、上記種類特定ステップにて特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力ステップと、を実行させることを特徴とする。

### 発明の効果

[0015] この発明によれば、遺伝子情報を含む遺伝子リストを取得し、取得された遺伝子リストに含まれる遺伝子情報と一致する、記憶された遺伝子情報を検出し、検出された遺伝子情報、および、記憶された細胞情報に基づき、遺伝子リストに対応する細胞または組織の種類を特定し、特定された、細胞の種類を含む細胞情報、および／または、組織の種類を含む組織情報を含む出力情報を出力させるので、生体サンプルの分類に使用することができるという効果を奏する。また、この発明によれば、ユーザ定義の遺伝子リストから、細胞種毎の発現の差異を特定することができるという効果を奏する。また、この発明によれば、*in vivo*でのマイクロアレイデータが真の遺伝子発現を反映するための継続的な分析を保証する新しい分析ワークフローを提供することができるという効果を奏する。

[0016] また、この発明によれば、記憶された細胞または組織における遺伝子発現情報に基づき、当該細胞または当該組織において特異的に高発現となる遺伝子を特定し、当該遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて格納するので、細胞の種

類のいずれかの分類に広く適用することが可能な汎用的なデータベースを提供することができるという効果を奏する。

[0017] また、この発明によれば、記憶された組織における遺伝子発現情報に基づく遺伝子発現のユニークなパターンを識別することにより、当該組織における遺伝子発現情報を複数のクラスタにクラスタリングし、クラスタリングされたクラスタに含まれる遺伝子発現情報に対応する遺伝子情報に基づき、遺伝子セット解析を用いて遺伝子リストを取得するので、いくつか特有の遺伝子発現特性を持つRNAサンプルから、個々の発現パターンを検出するために、遺伝子のクラスタリングと遺伝子セット解析とを組み合わせた生体サンプルの分類をすることができるという効果を奏する。また、この発明によれば、サンプル内の細胞の種類を決定したり、腫瘍領域に特異的な遺伝子発現を同定したりする場合などに広く適用できるという効果を奏する。

### 図面の簡単な説明

- [0018] [図1]図1は、本実施の形態の基本原理を示すフローチャートである。
- [図2]図2は、本実施の形態における細胞種解析装置の構成の一例を示すブロック図である。
- [図3]図3は、本実施の形態の細胞種解析装置の処理の一例を示すフローチャートである。
- [図4]図4は、本実施の形態におけるH E C Sデータベース構築処理の一例を示す概念図である。
- [図5]図5は、本実施の形態におけるマウスH E C Sデータベースにおけるデータ構造の一例を示す図である。
- [図6]図6は、本実施の形態におけるヒトH E C Sデータベースにおけるデータ構造の一例を示す図である。
- [図7]図7は、本実施形態における細胞型毎のH E C S遺伝子数の分布図である。
- [図8]図8は、本実施形態におけるサンプルが分離された部位の違いによる分類の一例を示す図である。



[図9]図9は、本実施の形態における細胞種解析装置の処理の一例を示すフローチャートである。

[図10]図10は、本実施形態における入力画面の一例を示す図である。

[図11]図11は、本実施形態における入力画面の一例を示す図である。

[図12]図12は、本実施形態における遺伝子情報検出の一例を示す図である。

[図13]図13は、本実施形態における遺伝子情報検出の一例を示す図である。

[図14]図14は、本実施形態における出力情報の一例を示す図である。

[図15]図15は、本実施形態における出力情報の一例を示す図である。

[図16]図16は、本実施形態における出力情報の一例を示す図である。

[図17]図17は、本実施形態におけるH E C Sデータベースを用いたデータ処理を行うワークフローの一例を示す図である。

[図18]図18は、本実施形態における組織の染色の一例を示す図である。

[図19]図19は、本実施形態における遺伝子発現変化のクラスタ解析の一例を示す図である。

[図20]図20は、本実施形態におけるクラスタ解析のヒートマップの一例を示す図である。

### 発明を実施するための形態

[0019] 以下に、本発明にかかる細胞種解析装置、細胞種解析方法、および、プログラムの実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

[0020] [本発明の実施の形態の概要]

以下、本発明の実施の形態の概要について図1を参照して説明し、その後、本実施の形態の構成および処理等について詳細に説明する。図1は、本実施の形態の基本原理を示すフローチャートである。本実施の形態は、概略的に、以下の基本的特徴を有する。

[0021] すなわち、本実施の形態の細胞種解析装置の制御部は、図1に示すように

、遺伝子情報を含む遺伝子リストを取得する（ステップS A-1）。

[0022] そして、細胞種解析装置の制御部は、ステップS A-1にて取得した遺伝子リストに含まれる遺伝子情報と一致する、記憶部に記憶された遺伝子情報を検出する（ステップS A-2）。

[0023] そして、細胞種解析装置の制御部は、ステップS A-1にて検出した遺伝子情報、および、記憶部に記憶された細胞情報または組織情報に基づき、ステップS A-1にて入力された遺伝子リストに対応する細胞または組織の種類を特定する（ステップS A-3）。

[0024] そして、細胞種解析装置の制御部は、ステップS A-2にて特定した、細胞の種類を含む細胞情報、および／または、組織の種類を含む組織情報を含む出力情報を出力部を介して出力させ（ステップS A-4）、処理を終了する。

[0025] 以上で、本実施の形態の概要の説明を終える。

[0026] [細胞種解析装置100の構成]

次に、本実施の形態における細胞種解析装置100の構成の詳細について、図2を参照して以下に説明する。図2は、本実施の形態における細胞種解析装置100の構成の一例を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。ここで、本実施の形態における細胞種解析装置100においては、各構成が一筐体内に全て備えられ、単独で処理を行うもの（スタンドアローン型）を、細胞種解析装置100として説明するが、当該実施例に限らず、各構成が分離した筐体内に備えられ、ネットワーク300等を介して接続されて一つの概念としての装置を構成するもの（例えば、クラウドコンピューティング等）であってもよい。

[0027] 図2において、外部システム200は、ネットワーク300を介して、細胞種解析装置100と相互に接続され、遺伝子発現情報、遺伝子情報、細胞情報、および／もしくは、組織情報等に関する外部データベース、ならびに／または、ユーザインターフェース等を実行するウェブサイトを提供する機能等を有していてもよい。

[0028] ここで、外部システム200は、WEBサーバやASPサーバ等として構成していてもよい。また、外部システム200のハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成していてもよい。また、外部システム200の各機能は、外部システム200のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現されてもよい。

[0029] また、ネットワーク300は、細胞種解析装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネット等である。

[0030] また、細胞種解析装置100は、概略的に、制御部102と通信制御インターフェース部104と記憶部106と入出力制御インターフェース部108とを備える。ここで、細胞種解析装置100は、更に、表示部112を少なくとも含む出力部、および、入力部114を備えていてもよい。また、出力部は、更に、音声出力部、および、印刷出力部等を含んでいてもよい。ここで、制御部102は、細胞種解析装置100の全体を統括的に制御するCPU等である。また、通信制御インターフェース部104は、通信回線等に接続されるルータ等の通信装置（図示せず）に接続されるインターフェースであり、入出力制御インターフェース部108は、出力部、および、入力部114に接続されるインターフェースである。また、記憶部106は、各種のデータベースやテーブルなどを格納する装置である。これら細胞種解析装置100の各部は任意の通信路を介して通信可能に接続されている。更に、この細胞種解析装置100は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク300に通信可能に接続されている。

[0031] 記憶部106に格納される各種のデータベースやテーブル（遺伝子発現情報データベース106a、組織発現情報データベース106b、および、特異的高発現データベース106c）は、固定ディスク装置等のストレージ手段である。例えば、記憶部106は、各種処理に用いる各種のプログラム、

テーブル、ファイル、データベース、および、ウェブページ等を格納する。

[0032] これら記憶部106の各構成要素のうち、遺伝子発現情報データベース106aは、細胞または組織における遺伝子発現情報を記憶する遺伝子発現情報記憶手段である。ここで、遺伝子発現情報は、マイクロアレイの測定結果であってもよい。ここで、遺伝子発現情報データベース106aは、細胞または組織におけるタンパク質発現情報を記憶していてもよい。ここで、タンパク質発現情報は、質量分析法(Mass Spectrometry、例えば、LC/MS、または、GC/MS等)による測定結果であってもよい。これら遺伝子発現情報、および、タンパク質発現情報は、遺伝子発現情報データベース106aに予め記憶されており、細胞種解析装置100の制御部102は、定期的に、および/または、制御部102による処理に応じてネットワーク300を介して最新のデータを外部システム200等からダウンロードして遺伝子発現情報データベース106aに記憶された遺伝子発現情報、および、タンパク質発現情報をアップデートしてもよい。

[0033] また、組織発現情報データベース106bは、複数種類の細胞を含む組織における遺伝子発現情報を記憶する組織発現情報記憶手段である。ここで、組織発現情報データベース106bは、複数種類の細胞を含む組織におけるタンパク質発現情報を記憶していてもよい。これら遺伝子発現情報、および、タンパク質発現情報は、組織発現情報データベース106bに予め記憶されており、細胞種解析装置100の制御部102は、定期的に、および/または、制御部102による処理に応じてネットワーク300を介して最新のデータを外部システム200等からダウンロードして組織発現情報データベース106bに記憶された遺伝子発現情報、および、タンパク質発現情報をアップデートしてもよい。

[0034] また、特異的高発現データベース106cは、細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶する特異的高発現記憶手段である。ここで、特異的高発現データベース106cは、

細胞または組織の種類に依存して特異的に高発現となるタンパク質に関するタンパク質情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶していてもよい。また、特異的高発現データベース106cは、外部システム200（例えば、パブリックデータベース等）から（ネットワーク300を介して）集めた遺伝子発現情報に基づき、細胞種に依存して発現量の高くなる遺伝子（HECS: High Expressed Cell-Specific genes）を特定したデータベースであってもよい。また、特異的高発現データベース106cは、バックグラウンドリファレンスとして機能する遺伝子のシグネチャとなるデータを記憶するために、比較的限定された条件下で非常に高く発現される遺伝子セットを収集し、単一の条件下で必ずしも排他的とならないようになっていてもよい。

[0035] また、通信制御インターフェース部104は、細胞種解析装置100とネットワーク300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部104は、外部システム200、および、他の端末等と通信回線を介してデータを通信する機能を有する。

[0036] また、入出力制御インターフェース部108は、出力部（表示部112）、および、入力部114の制御を行う。

[0037] ここで、表示部112としては、アプリケーション等の表示画面を表示する表示手段（例えば、液晶または有機EL等から構成されるディスプレイ、モニタ、または、タッチパネル等）であってもよい。また、入力部114は、例えば、キー入力部、タッチパネル、コントロールパッド（例えば、タッチパッド、および、ゲームパッド等）、マウス、キーボード、または、マイク等であってもよい。

[0038] また、図2において、制御部102は、OS（Operating System）等の制御プログラムや、各種の処理手順等を規定したプログラム、および、所要データを格納するための内部メモリを有する。そして、制御

部102は、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部102は、機能概念的に、遺伝子情報格納部102a、クラスタリング部102b、遺伝子リスト取得部102c、遺伝子情報検出部102d、種類特定部102e、および、情報出力部102fを備える。

[0039] このうち、遺伝子情報格納部102aは、遺伝子発現情報データベース106aに記憶された細胞または組織における遺伝子発現情報に基づき、当該細胞または当該組織において特異的に高発現となる遺伝子を特定し、当該遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて特異的高発現データベース106cに格納する遺伝子情報格納手段である。ここで、遺伝子発現情報データベース106aに記憶された細胞または組織におけるタンパク質発現情報に基づき、当該細胞または当該組織において特異的に高発現となるタンパク質を特定し、当該タンパク質に関するタンパク質情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて特異的高発現データベース106cに格納してもよい。

[0040] また、クラスタリング部102bは、組織発現情報データベース106bに記憶された組織における遺伝子発現情報に基づく遺伝子発現のユニークなパターンを識別することにより、当該組織における遺伝子発現情報を複数のクラスタにクラスタリングするクラスタリング手段である。ここで、クラスタリング部102bは、組織発現情報データベース106bに記憶された組織におけるタンパク質発現情報に基づくタンパク質発現のユニークなパターンを識別することにより、当該組織におけるタンパク質発現情報を複数のクラスタにクラスタリングしてもよい。

[0041] また、遺伝子リスト取得部102cは、遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得手段である。ここで、遺伝子リスト取得部102cは、クラスタリング部102bによりクラスタリングされたクラスタに含まれる遺伝子発現情報に対応する遺伝子情報に基づき、遺伝子セット解析を

用いて遺伝子リストを取得してもよい。また、遺伝子リスト取得部102cは、クラスタリング部102bによりクラスタリングされたクラスタに含まれるタンパク質発現情報に対応するタンパク質情報に基づき、タンパク質リストを取得してもよい。また、遺伝子リスト取得部102cは、入力部114を介して入力された遺伝子リスト、または、タンパク質リストを取得してもよい。

[0042] また、遺伝子情報検出部102dは、遺伝子リスト取得部102cにより取得された遺伝子リストに含まれる遺伝子情報と一致する、特異的高発現データベース106cに記憶された遺伝子情報を検出する遺伝子情報検出手段である。ここで、遺伝子情報検出部102dは、遺伝子リスト取得部102cにより取得されたタンパク質リストに含まれるタンパク質情報と一致する、特異的高発現データベース106cに記憶されたタンパク質情報を検出してもよい。

[0043] また、種類特定部102eは、遺伝子情報検出部102dにより検出された遺伝子情報、および、特異的高発現データベース106cに記憶された細胞情報または組織情報に基づき、遺伝子リストに対応する細胞または組織の種類を特定する種類特定手段である。ここで、種類特定部102eは、遺伝子情報検出部102dにより検出されたタンパク質情報、および、特異的高発現データベース106cに記憶された細胞情報または組織情報に基づき、タンパク質リストに対応する細胞または組織の種類を特定してもよい。また、種類特定部102eは、更に、種類を特定した細胞または組織毎の *Enrichment* (特徴抽出) スコアを計算してもよい。

[0044] また、情報出力部102fは、情報を出力部を介して出力させる情報出力手段である。ここで、情報出力部102fは、種類特定部102eにより特定された、細胞の種類を含む細胞情報、および/または、組織の種類を含む組織情報を含む出力情報を出力部を介して出力させてもよい。ここで、出力情報は、更に、種類特定部102eにより計算された特徴抽出スコアを含んでいてもよい。また、情報出力部102fは、情報を表示部112に表示さ

せてもよい。また、情報出力部102fは、情報を印刷出力部を介して出力させてもよい。また、情報出力部102fは、遺伝子リスト、または、タンパク質リストを入力するための表示情報（入力画面）を表示部112に表示させてもよい。

[0045] 以上で、本実施の形態における細胞種解析装置100の構成の一例の説明を終える。

[0046] [細胞種解析装置100の処理]

次に、このように構成された本実施の形態における細胞種解析装置100の処理の詳細について、以下に図3乃至図20を参照して詳細に説明する。

[0047] [HECSデータベース構築処理]

まず、本実施形態におけるHECSデータベース構築処理の一例について、図3乃至図8を参照して説明する。図3は、本実施の形態における細胞種解析装置100の処理の一例を示すフローチャートである。

[0048] 図3に示すように、遺伝子情報格納部102aは、遺伝子発現情報データベース106aに記憶された細胞または組織における遺伝子発現情報（マイクロアレイの測定結果）を取得する（ステップSB-1）。ここで、遺伝子情報格納部102aは、ネットワーク300を介して外部システム200（例えば、遺伝子発現情報を蓄積したデータベース等）から遺伝子発現情報を取得してもよい。

[0049] そして、遺伝子情報格納部102aは、ステップSB-1にて取得したマイクロアレイの測定結果に基づき、当該細胞または当該組織において特異的に高発現となる遺伝子（HECS遺伝子）を特定する（ステップSB-2）。

[0050] そして、遺伝子情報格納部102aは、ステップSB-2にて特定した遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて特異的高発現データベース106c（HECSデータベース）に格納し（ステップSB-3）、処理を終了する。

[0051] ここで、図4を参照して、本実施の形態におけるHECSデータベース構



築処理の一例について説明する。図4は、本実施の形態におけるHECSデータベース構築処理の一例を示す概念図である。

[0052] 図4に示すように、遺伝子情報格納部102aは、外部システム200（例えば、BioGPS等）からネットワーク300を介してマウスのマイクロアレイデータ（例えば、MOE430遺伝子アトラス等）、および、ヒトのマイクロアレイデータ（例えば、U133A、および、GNF1HのGeneChipデータ等）を取得する。そして、遺伝子情報格納部102aは、マウスおよびヒトそれぞれのマイクロアレイデータに対して測定サンプル数で平均をとり、プローブの中央値が全ての細胞種および組織間で15倍（マウス）、または、10倍（ヒト）以上の遺伝子発現を示すプローブ（遺伝子断片）を1つの細胞種または組織に割り当てる、すなわち、当該プローブを1つの細胞種または組織にアサインする。そして、遺伝子情報格納部102aは、1つの細胞種または組織に割り当てたプローブにアノテーションファイル（例えば、Bioconductorで利用可能なAffymetrixマウスゲノム4302.0アレイ（mouse4302version2.5.0）、および、AffymetrixヒトゲノムU133セット（hgu133aversion2.5.0）のアノテーションファイル等）を用いてアノテーションを追加することで、1つの細胞種または組織において特異的に高発現となる遺伝子（HECS遺伝子）を特定する。そして、遺伝子情報格納部102aは、同じ細胞種または組織に複数回アサインされた遺伝子を削除し（すなわち、一つの細胞種または組織に同一遺伝子に含まれる複数のプローブがアサインされた場合、データの重複を避けるために一つの遺伝子として扱い）、特定した遺伝子に関する遺伝子情報と、細胞種に関する細胞情報または組織に関する組織情報と、を対応付けて特異的高発現データベース106c（マウスHECSデータベース、および、ヒトHECSデータベース）に格納する。

[0053] すなわち、マウスHECSデータベース、および、ヒトHECSデータベースは、比較的少数の細胞型（細胞種）において、非常に高く発現される遺

伝子に関する遺伝子情報で構成されていてもよい。また、遺伝子発現情報は、BioGPS等からダウンロードしたものを利用してよい。また、BioConductorから入手可能なアノテーションファイルを利用してアノテーションを追加し、同じ組織または細胞型に複数回割り当てられている遺伝子を除去することによりデータベースを最小化してもよい。

[0054] また、図5および図6を参照して、本実施の形態におけるHECSデータベースにおけるデータ構造の一例について説明する。図5は、本実施の形態におけるマウスHECSデータベースにおけるデータ構造の一例を示す図である。図6は、本実施の形態におけるヒトHECSデータベースにおけるデータ構造の一例を示す図である。

[0055] 図5に示すように、本実施形態におけるマウスHECSデータベースには、3つのdata.frames(行列)を含んでいる。また、図6に示すように、本実施形態におけるヒトHECSデータベースには、3つのdata.frames(行列)を含んでいる。ここで、図5および図6に示すように、本実施形態におけるHECSデータベースに含まれる行列としては、Rスクリプトで互換性のある組織名を示す「Tissue(組織)」、同じ行に細胞または組織のHECS遺伝子として同定された1遺伝子のEntrezGeneID(識別子)を示す「EntrezGene」、および、EntrezGene識別子に一致する公式遺伝子シンボルを示す「Symbol(シンボル)」から構成される、HECS遺伝子と細胞または組織との組み合わせごとに1行を構成する「CT」を含んでいてもよい。また、図5および図6に示すように、本実施形態におけるHECSデータベースに含まれる行列としては、データベース上で利用可能な細胞または組織の識別文字列を示す「Tissues」を含んでいてもよい。また、図5および図6に示すように、本実施形態におけるHECSデータベースに含まれる行列としては、マイクロアレイデータに対応するEntrezgene識別子、および、マイクロアレイデータに対応する公式遺伝子シンボルを含むユニークな遺伝子識別子のリストであり、特徴抽出解析中に利用するGene

Universeに相当する「Universe」を含んでいてもよい。

[0056] また、プローブを割り当てる場合のカットオフは、遺伝子の一意性、各条件下の遺伝子のシグネチャとして機能することができた遺伝子数を均衡化することで、慎重に選択されてもよい。ここで、図7を参照して、本実施形態における閾値選択処理の一例について説明する。図7は、本実施形態における細胞型毎のHECS遺伝子数の分布図である。

[0057] 図7左に示すように、HECS遺伝子を定義するために使われる閾値の基準である、細胞型あたりのHECS遺伝子数の分布は、マウス（図7左上）およびヒト（図7左下）の遺伝子発現データに対して、全ての細胞型で、遺伝子発現データの中央値の2.5倍から25倍の順に上昇している。また、図7右に示すように、閾値の値を一意的に定めるために、所定の整数n以下の細胞型でマップされたHECS遺伝子の割合（すなわち、累積度数）を、マウス（図7右上）およびヒト（図7右下）の遺伝子発現データに対して決定することで、それぞれの異なる閾値を決定している。

[0058] また、分類方法は、あらかじめ決定しておく必要があり、本実施形態におけるHECSデータベースの場合、細胞および組織の種類（タイプ）で分類されているが、ガンを対象にするデータベースの場合は、サンプルが分離された部位の違いによる分類、ガンの成長を抑制する複数の薬剤の薬効の違いによる分類、同一の腫瘍において異なる部位または分離時間によるサンプルに対する単一薬剤の有効性の違いによる分類等による分類であってもよい。

[0059] ここで、図8を参照して、本実施形態におけるサンプルが分離された部位の違いによる分類の一例について説明する。図8は、本実施形態におけるサンプルが分離された部位の違いによる分類の一例を示す図である。

[0060] 図8に示すように、本実施形態においては、腫瘍内のサンプルが分離された部位（Zone1、Zone2、および、Zone3）の違いにより、HECS遺伝子を特定し、HECS遺伝子に関する遺伝子情報と、Zone1、Zone2、または、Zone3に関する組織情報と、を対応付けたHECSデータベースを構築してもよい。このように適切な分類データベースを

構築することで、近年、腫瘍内の細胞集団が大きく異なることで効果的な治療が妨げられているガン治療において、医師は、患者の腫瘍（ガン）の複数部位のサンプリングから、本発明を利用してデータをクラスタ化し、腫瘍の異なる部位に対して選択的に異なる薬剤化合物を投与するべきかどうかを判断できるようになる。

[0061] 以上で、本実施形態におけるH E C Sデータベース構築処理の一例の説明を終える。

[0062] [細胞種解析処理]

次に、本実施形態における細胞種解析処理の一例について、図9乃至図20を参照して説明する。図9は、本実施の形態における細胞種解析装置100の処理の一例を示すフローチャートである。

[0063] 図9に示すように、クラスタリング部102bは、組織発現情報データベース106bに記憶された組織における遺伝子発現情報を取得する（ステップSC-1）。

[0064] そして、クラスタリング部102bは、ステップSC-1にて取得した遺伝子発現情報に基づく遺伝子発現のユニークなパターンを識別することにより、当該組織における遺伝子発現情報を複数のクラスタにクラスタリングする（ステップSC-2）。

[0065] そして、遺伝子リスト取得部102cは、クラスタリング部102bによりクラスタリングされたクラスタに含まれる遺伝子発現情報に対応する遺伝子情報に基づき、遺伝子セット解析を用いて遺伝子リストを取得する（ステップSC-3）。なお、本実施形態においては、自動的に遺伝子リストを取得する一例について説明したが、遺伝子リスト取得部102cは、情報出力部102fに遺伝子リストを入力するための表示情報（入力画面）を表示部112に表示させ、ユーザにより入力部114を介して入力された遺伝子リストを取得してもよい。ここで、ユーザにより入力部114を介して入力された遺伝子リストは、誘導多能性幹細胞、または、腫瘍等において高発現となる遺伝子に関する遺伝子情報を含んでいてもよい。

[0066] ここで、図10および図11を参照して、本実施形態における入力画面の一例を説明する。図10および図11は、本実施形態における入力画面の一例を示す図である。

[0067] 図10に示すように、本実施形態における入力画面においては、対象となる生物種を選択するドロップダウンリスト、遺伝子ID型を選択するドロップダウンリスト、データの区切り記号（スペース、コンマ、または、タブ）を選択するラジオボタン、遺伝子リストを入力するテキストボックス、および、遺伝子リストの入力（アップロード）を完了させるボタン（Submit Formボタン）を含んでいてもよい。また、図11に示すように、本実施形態における入力画面においては、対象となる生物種を選択するドロップダウンリスト、遺伝子ID型を選択するドロップダウンリスト、データの区切り記号を選択するラジオボタン、複数（グループ）の遺伝子リストを入力するテキストボックス、および、遺伝子リストの入力（アップロード）を完了させるボタン（Submit Formボタン）を含んでいてもよい。

[0068] 図9に戻り、遺伝子情報検出部102dは、遺伝子リスト取得部102cにより取得された遺伝子リストに含まれる遺伝子情報と一致する、特異的高発現データベース106cに記憶された遺伝子情報を検出する（ステップS C-4）。

[0069] ここで、図12および図13を参照して、本実施形態における遺伝子情報検出の一例を説明する。図12および図13は、本実施形態における遺伝子情報検出の一例を示す図である。

[0070] 図12に示すように、本実施形態における遺伝子情報検出結果を示す画面においては、ユーザが1つの遺伝子リストを入力画面において入力した場合の結果を示しており、リスト名を表示するテキストボックス、ユーザが入力した遺伝子リストの遺伝子数を表示するテキストボックス、当該遺伝子リストの遺伝子と一致する特異的高発現データベース106cに記憶された、入力画面にて指定した生物種の遺伝子数を表示するテキストボックス、および、各細胞型の特徴抽出スコアを計算させ、特徴抽出スコアを表示させるボタ

ン (Continue to Enrichment ボタン) を含んでもよい。また、図 13 に示すように、本実施形態における遺伝子情報検出結果を示す画面においては、ユーザが複数の遺伝子リストを入力画面において入力した場合の結果を示しており、リスト名 (グループラベル) を表示するテキストボックス、ユーザが入力した遺伝子リストの遺伝子数を表示するテキストボックス、当該遺伝子リストの遺伝子と一致する特異的高発現データベース 106c に記憶された、入力画面にて指定した生物種の遺伝子数を表示するテキストボックス、および、グループ毎に各細胞型の特徴抽出スコアを計算させ、特徴抽出スコアを表示させるボタン (Continue to Enrichment ボタン) を含んでもよい。

[0071] 図 9 に戻り、種類特定部 102e は、遺伝子情報検出部 102d により検出された遺伝子情報、および、特異的高発現データベース 106c に記憶された細胞情報または組織情報に基づき、遺伝子リストに対応する細胞または組織の種類を特定し、種類を特定した細胞または組織毎の特徴抽出スコアを計算する (ステップ SC-5)。例えば、種類特定部 102e は、R スクリプトにて、ユーザがアップロードした情報を収集し、適切なデータを HEC S データベースから選択して、細胞または組織毎の特徴抽出スコアを計算してもよい。

[0072] そして、情報出力部 102f は、種類特定部 102e により特定された、細胞の種類を含む細胞情報、および/もしくは、組織の種類を含む組織情報、ならびに、種類特定部 102e により計算された特徴抽出スコアを含む出力情報を表示部 112 に表示させ (ステップ SC-6)、処理を終了する。例えば、情報出力部 102f は、種類特定部 102e により計算された特徴抽出スコアに基づくレーダーチャート、または、ヒートマップを含む出力情報を生成して、当該出力情報を表示部 112 に表示させてもよい。

[0073] ここで、図 14 乃至図 16 を参照して、本実施形態における出力情報の一例について説明する。図 14 乃至図 16 は、本実施形態における出力情報の一例を示す図である。

[0074] 図14は、本実施形態において、健常マウスの肺（非感染組織）と、インフルエンザに感染したマウスの肺（感染組織）と、において異なる発現をする遺伝子に関する遺伝子発現データを用いて、HECSデータベースを使った解析（細胞種の特定）を行った結果（出力情報）を示している。図14に示すように、本発明を利用することによって、感染組織においては、マクロファージが深く肺に浸透しており、インフルエンザの感染により、体が損傷を修復しようとして、リンパ球の移動を促していることを裏付ける解析結果を示すことができた。

[0075] また、図15に示すように、情報出力部102fは、ユーザにより1つの遺伝子リストがアップロードされた場合、HECSデータベースで識別された特徴抽出スコアが高い上位20の細胞または組織を示すレーダーチャート（出力情報）を表示部112に表示させてもよい。ここで、図15においては、図14に示す解析において特徴抽出スコアが高い上位20の細胞または組織をレーダーチャートで示しており、マクロファージを多く含むサンプルであることがわかる。また、図15に示すように、ユーザが表示画面をマウス（入力部114）で右クリックした場合、HECSデータベースの全細胞型に対する特徴抽出スコアをダウンロードできるようにしてもよい。

[0076] また、図16に示すように、情報出力部102fは、ユーザにより複数（Group1-Group4）の遺伝子リストがアップロードされた場合、HECSデータベースで識別された特徴抽出スコアが高い細胞または組織について、遺伝子リストのグループ毎のヒートマップ（出力情報）を表示部112に表示させてもよい。ここで、図16に示すように、ユーザが表示画面をマウス（入力部114）で右クリックした場合、HECSデータベースの全細胞型に対する特徴抽出スコアをダウンロードできるようにしてもよい。また、図16に示すように、本実施形態においては、加重ランキング戦略を用いて、各クラスタの中で最もEnrich強化された細胞の種類を示すヒートマップを表示することができる。

[0077] また、図17を参照して、本実施形態におけるHECSデータベースを用

いたデータ処理を行うワークフローの一例について説明する。図17は、本実施形態におけるHECSデータベースを用いたデータ処理を行うワークフローの一例を示す図である。

[0078] 図17に示すように、まず、ユーザは、インターフェースを介してデータのアップロードを行う（ユーザ定義の遺伝子リストの入力、遺伝子識別子の選択、および、生物種の選択）。例えば、ユーザは、表計算ソフトから直接適切なパラメータ（例えば、種、および、区切り記号等）を選択し、任意の遺伝子リストを入力フィールドにコピーアンドペーストする。このように、区切り記号等のパラメータをユーザに選択可能とすることで、同時に複数の遺伝子リストをアップロードできるユーザフレンドリーなインターフェースを提供し、迅速な分析を可能としている。そして、遺伝子情報検出部102dは、アップロードされた遺伝子リストをパース（構文解析）するユーザデータの検証を行い、HECSデータベースと遺伝子リストとの照合を行うデータ予備処理（HECSデータベースで検出されることを確認する予備処理）を実行する。そして、種類特定部102eは、予備処理後、HECSデータベース内の全ての細胞型について、フィッシャーの正確確率検定（Fisher Exact Test）による計算処理を利用した特徴抽出解析（Enrichment Analysis）を行い、特徴抽出スコア（log10（Benjamini調整P値））を計算する。そして、情報出力部102fは、細胞種分類のグラフを表示部112に表示させ、結果をテキストファイル形式で印刷出力部を介して出力させる。

[0079] また、図18乃至図20を参照して、本実施形態におけるクラスタリング解析と組み合わせて細胞遊走の結果である特定の遺伝子発現パターンを決定した処理結果の一例について説明する。図18は、本実施形態における組織の染色の一例を示す図である。図19は、本実施形態における遺伝子発現変化のクラスタ解析の一例を示す図である。図20は、本実施形態におけるクラスタ解析のヒートマップの一例を示す図である。

[0080] 本実施形態における遺伝子発現情報のクラスタリングにおいては、遺伝的



に混在した環境に由来するRNAサンプル内の異なるパターンの識別手段、すなわち、K平均法、Kクラスタリング、または、加重遺伝子相関関係等のクラスタリング手法が包括的にサポートされており、時間経過とともに、さまざまな種類の細胞が移入した単一の組織において、遺伝子発現データを日単位（例えば、1日から5日等）で収集し、検討している。ここで、図18および図19には、制御組織に対して有為に変異して発現している遺伝子を特定し、遺伝子発現のユニークなパターンを識別するために、遺伝子発現情報をクラスタリングした結果を示している。ここで、本実施形態においては、WGCNA (<http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA/>) を使用したクラスタリング研究に用いた潜在的な結果を示し、固有のプロファイルをもつ4つのクラスタに分けて分析している。その結果、3つのクラスタ（クラスタA、クラスタB、および、クラスタC）は、図18および図19に示すように、リンパ球数の変化（それぞれマクロファージ、B細胞、および、T細胞）に高い相関性が示された。

[0081] また、図20に示すように、本実施形態においては、各クラスタ内に含まれる遺伝子を分析し、どのクラスタで遺伝子が発現しているか表示することができ、また、細胞の移動を表すクラスタを区別して表示することができ、更に、観測された遺伝子調整の役割を担う細胞の種類を決定することも可能となる。このように、図20に示すように、クラスタDのみが、従来のバイオインフォマティクス技術を用いて更なる分析を行うにふさわしいクラスタであることがわかる。一方、図20に示すように、クラスタA、クラスタB、および、クラスタCの各クラスタは、感染時のリンパ球数（それぞれマクロファージ、B細胞、および、T細胞）の相対的变化を反映していることがわかる。すなわち、感染した組織では、B細胞、T細胞、および、マクロファージがサンプル領域に移行し、その結果、細胞型の増加（減少）に比例して、保存された発現パターンの結果として、各細胞におけるRNAの量が特

異的な値を示している。また、本実施形態におけるクラスタリングは、同定されていない遺伝子セットに対して、有為差のある発現パターンを検出するために用いられてもよい。ここで、本実施形態の手法をガン組織に適用する場合には、指定した分類法に従って、例えば、日毎に、腫瘍や腫瘍全体にわたって異なる部位からサンプルを収集し、クラスタリングを行ってもよい。また、本実施形態におけるクラスタリングの最後のステップとしては、各分類にクラスタ化された遺伝子の各セットをリンクする作業であり、遺伝子セット解析を用いて行われてもよく、各クラスタ内の各分類と遺伝子リストとの間の照合は、フィッシャーの直接確率検定を用いてもよい。

[0082] また、本実施形態における手法を、誘導多能性幹細胞（IPS細胞）、または、胚性幹細胞（ES細胞）等に適用した場合、培養後の細胞の遺伝子発現情報を用いることで、当該細胞が分化誘導（例えば、神経細胞等に分化誘導）されたのか、または、未分化の細胞が増加等しただけなのかを判別することができる。

[0083] また、一般的に、生体組織検査（生検）で摘出されたガン細胞は、健康な組織または他のガン細胞と比較され、異常な成長を促進する遺伝子ターゲットがあるか否かを識別し、それが適切な薬物標的であるかを確認することが行われている。しかしながら、これらの検査では、生検試料に背景組織が多く含まれているといった低品質のサンプルに対して行われることが多く、正しい結果が得られない場合があった。そこで、本実施形態における手法を腫瘍に適用することで、HECSデータベースを用いて、発現する遺伝子が生検に含まれている背景組織の量が増大した結果であるのか確認することができる。したがって、本実施形態における手法においては、背景組織の混入による遺伝子発現量の差異が、更なる分析で影響がでないように処理することによって、生検サンプルの品質に対処することができる。すなわち、本実施形態における手法を腫瘍の病理検査等の際に適用することで、当該腫瘍にどのようなタイプのガン細胞が含まれるかを判別でき、医療現場において投薬する薬剤の選択、および、薬剤の投薬量の変更等を効率的に行

うことができる。

[0084] 以上で、本実施の形態における細胞種解析装置100の処理の一例の説明を終える。

[0085] [他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

[0086] 例えば、細胞種解析装置100がスタンドアローンの形態で処理を行う場合を一例に説明したが、細胞種解析装置100は、クライアント端末（細胞種解析装置100とは別筐体である）からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するようにしてもよい。

[0087] また、実施の形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

[0088] このほか、上記文献中や図面中で示した処理手順、制御手順、具体的名称、各処理の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

[0089] また、細胞種解析装置100に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

[0090] 例えば、細胞種解析装置100の各装置が備える処理機能、特に制御部102にて行われる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該CPUにて解釈実行されるプログラムにて実現してもよく、また、ワイヤードロジックによるハードウェアとして実現してもよい。尚、プログラムは、後述する、コンピュータに本発明に係る方法を実行させるためのプログラム化された命令を含む、一時的でないコンピュータ読み取り可能な記録媒体に記録

されており、必要に応じて細胞種解析装置100に機械的に読み取られる。すなわち、ROMまたはHDD (Hard Disk Drive) などの記憶部106などには、OS (Operating System) と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAMにロードされることによって実行され、CPUと協働して制御部を構成する。

[0091] また、このコンピュータプログラムは、細胞種解析装置100に対して任意のネットワーク300を介して接続されたアプリケーションプログラムサーバに記憶されていてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

[0092] また、本発明に係るプログラムを、コンピュータ読み取り可能な記録媒体に格納してもよく、また、プログラム製品として構成することもできる。ここで、この「記録媒体」とは、メモリーカード、USBメモリ、SDカード、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD、および、Blu-ray (登録商標) Disc等の任意の「可搬用の物理媒体」を含むものとする。

[0093] また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

[0094] 記憶部106に格納される各種のデータベース等(遺伝子発現情報データベース106a、組織発現情報データベース106b、および、特異的高発現データベース106c)は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、および、光ディスク等

のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラム、テーブル、データベース、および、ウェブページ用ファイル等を格納する。

[0095] また、細胞種解析装置100は、既知のデスクトップ型またはノート型のパーソナルコンピュータ、携帯電話、スマートフォン、PHS、およびPDA等の携帯端末装置、ならびに、ワークステーション等の情報処理装置として構成してもよく、また、該情報処理装置に任意の周辺装置を接続して構成してもよい。また、細胞種解析装置100は、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

[0096] 更に、装置の分散・統合の具体的形態は図示するものに限られず、その全部または一部を、各種の付加等に応じて、または、機能負荷に応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。すなわち、上述した実施の形態を任意に組み合わせて実施してもよく、実施の形態を選択的に実施してもよい。

### 産業上の利用可能性

[0097] 以上詳述に説明したように、本発明によれば、細胞種に依存して発現量が高くなる遺伝子を細胞種のシグネチャとして利用してデータ解析を行うことができる細胞種解析装置、細胞種解析方法、および、プログラムを提供することができるので、特に病態診断等の医療、製薬、創薬、ならびに、*in vivo*の遺伝子発現プロファイリングおよび細胞計数研究等の生物学研究などの様々な分野において極めて有用である。

### 符号の説明

- [0098] 100 細胞種解析装置  
102 制御部  
102a 遺伝子情報格納部  
102b クラスタリング部  
102c 遺伝子リスト取得部

- 1 0 2 d 遺伝子情報検出部
- 1 0 2 e 種類特定部
- 1 0 2 f 情報出力部
- 1 0 4 通信制御インターフェース部
- 1 0 6 記憶部
- 1 0 6 a 遺伝子発現情報データベース
- 1 0 6 b 組織発現情報データベース
- 1 0 6 c 特異的高発現データベース
- 1 0 8 入出力制御インターフェース部
- 1 1 2 表示部
- 1 1 4 入力部
- 2 0 0 外部システム
- 3 0 0 ネットワーク

## 請求の範囲

[請求項1]

出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置であって、

上記記憶部は、

細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶する特異的高発現記憶手段、

を備え、

上記制御部は、

上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得手段と、

上記遺伝子リスト取得手段により取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記憶された上記遺伝子情報を検出する遺伝子情報検出手段と、

上記遺伝子情報検出手段により検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定手段と、

上記種類特定手段により特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力手段と、

を備えたことを特徴とする細胞種解析装置。

[請求項2]

請求項1に記載の細胞種解析装置において、

上記記憶部は、

上記細胞または上記組織における遺伝子発現情報を記憶する遺伝子発現情報記憶手段、

を更に備え、

上記制御部は、

上記遺伝子発現情報記憶手段に記憶された上記細胞または上記組織における上記遺伝子発現情報に基づき、当該細胞または当該組織において特異的に高発現となる上記遺伝子を特定し、当該遺伝子に関する上記遺伝子情報と、当該細胞に関する上記細胞情報または当該組織に関する組織情報と、を対応付けて上記特異的高発現記憶手段に格納する遺伝子情報格納手段、

を更に備えたことを特徴とする細胞種解析装置。

[請求項3]

請求項1または2に記載の細胞種解析装置において、

上記記憶部は、

複数種類の上記細胞を含む上記組織における遺伝子発現情報を記憶する組織発現情報記憶手段、

を更に備え、

上記制御部は、

上記組織発現情報記憶手段に記憶された上記組織における上記遺伝子発現情報に基づく遺伝子発現のユニークなパターンを識別することにより、当該組織における遺伝子発現情報を複数のクラスタにクラスタリングするクラスタリング手段、

を更に備え、

上記遺伝子リスト取得手段は、

上記クラスタリング手段によりクラスタリングされた上記クラスタに含まれる上記遺伝子発現情報に対応する上記遺伝子情報に基づき、遺伝子セット解析を用いて上記遺伝子リストを取得することを特徴とする細胞種解析装置。

[請求項4]

出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置において実行される細胞種解析方法であって、

上記記憶部は、

細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する



る組織情報と、を対応付けて記憶する特異的高発現記憶手段、

を備え、

上記制御部において実行される、

上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得ステップと、

上記遺伝子リスト取得ステップにて取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記憶された上記遺伝子情報を検出する遺伝子情報検出ステップと、

上記遺伝子情報検出ステップにて検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定ステップと、

上記種類特定ステップにて特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力ステップと、

を含むことを特徴とする細胞種解析方法。

[請求項5]

出力部と記憶部と制御部とを少なくとも備えた細胞種解析装置に実行させるためのプログラムであって、

上記記憶部は、

細胞または組織の種類に依存して特異的に高発現となる遺伝子に関する遺伝子情報と、当該細胞に関する細胞情報または当該組織に関する組織情報と、を対応付けて記憶する特異的高発現記憶手段、

を備え、

上記制御部において、

上記遺伝子情報を含む遺伝子リストを取得する遺伝子リスト取得ステップと、

上記遺伝子リスト取得ステップにて取得された上記遺伝子リストに含まれる上記遺伝子情報と一致する、上記特異的高発現記憶手段に記

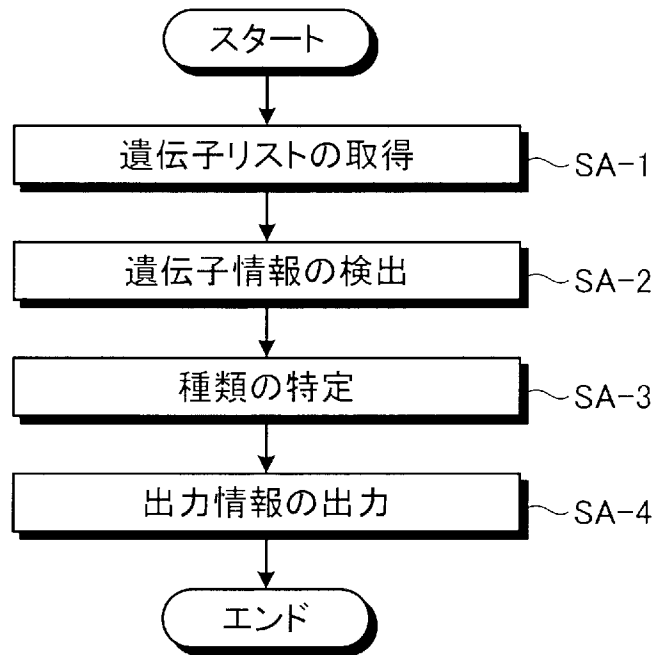
憶された上記遺伝子情報を検出する遺伝子情報検出ステップと、

上記遺伝子情報検出ステップにて検出された上記遺伝子情報、および、上記特異的高発現記憶手段に記憶された上記細胞情報または上記組織情報に基づき、上記遺伝子リストに対応する上記細胞または上記組織の種類を特定する種類特定ステップと、

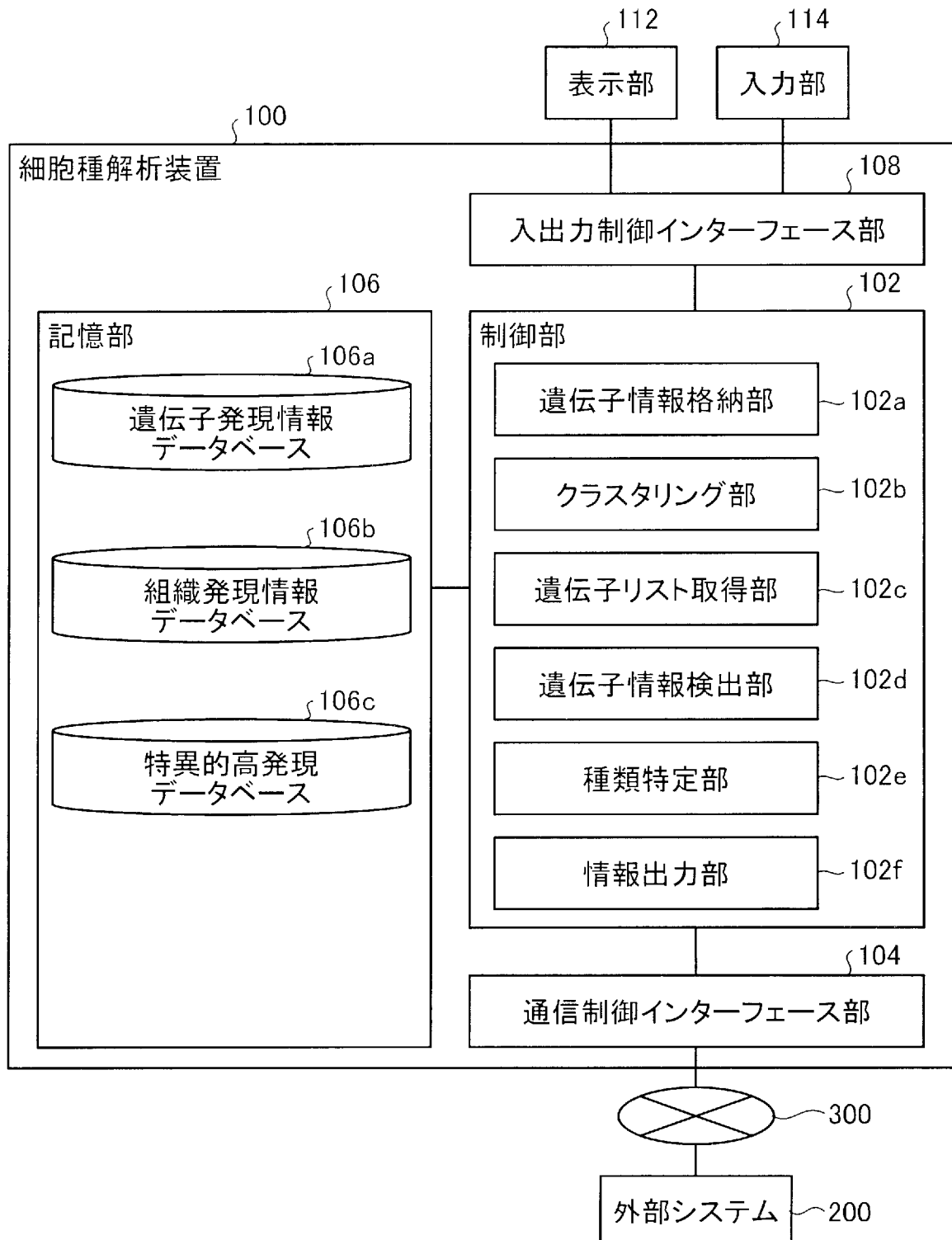
上記種類特定ステップにて特定された、上記細胞の種類を含む上記細胞情報、および／または、上記組織の種類を含む上記組織情報を含む出力情報を上記出力部を介して出力させる情報出力ステップと、

を実行させるためのプログラム。

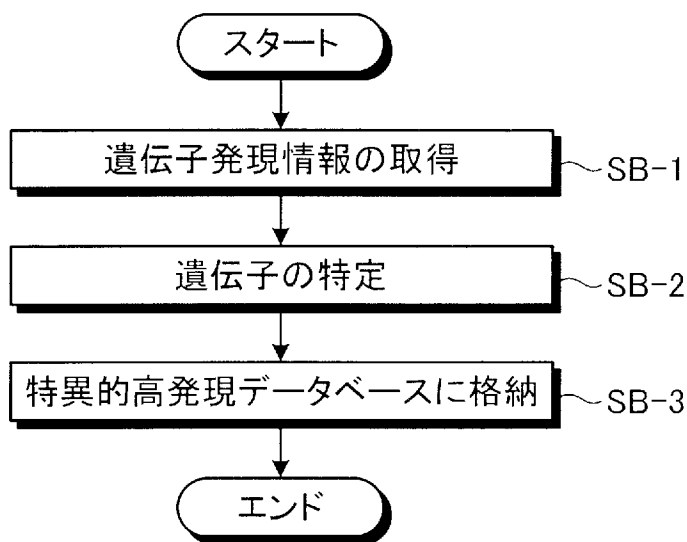
[図1]



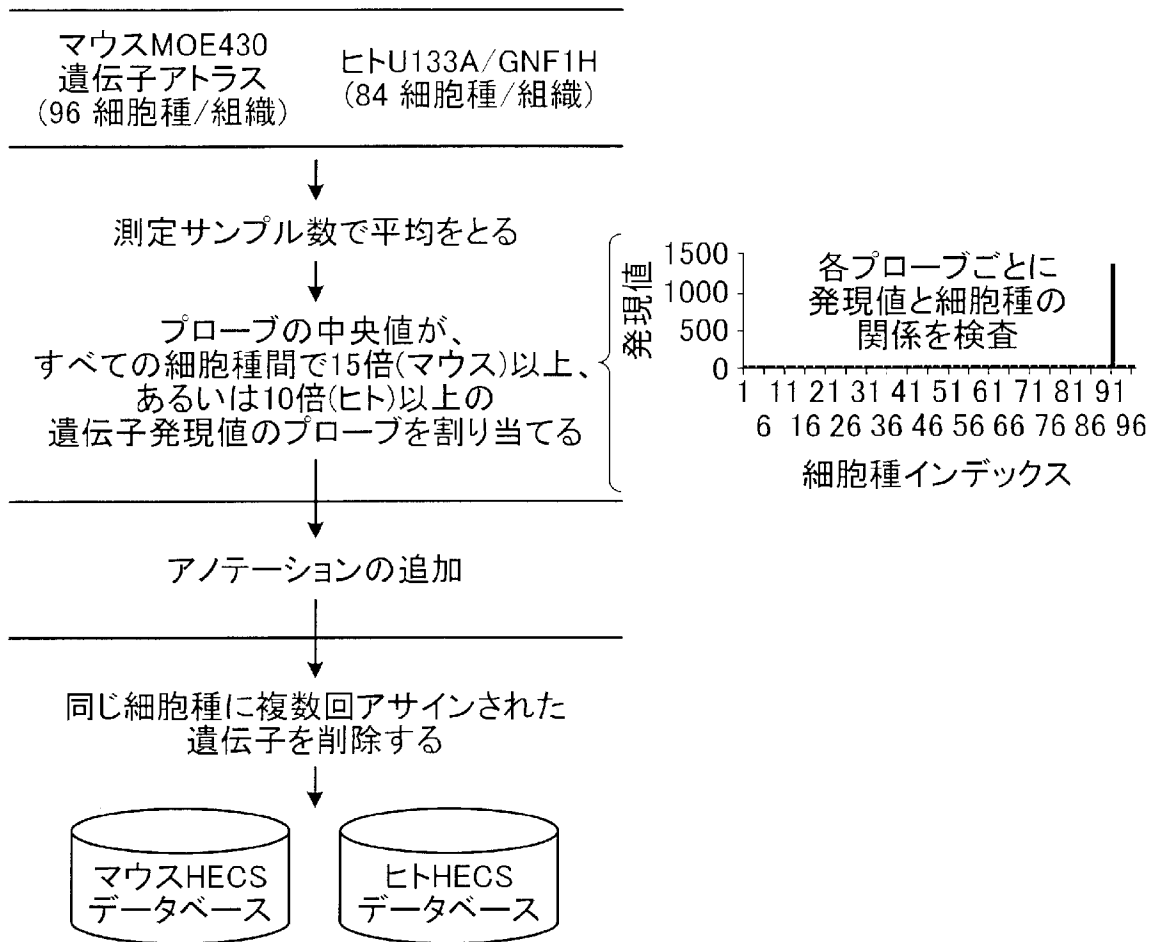
[図2]



[図3]



[図4]



[図5]

CT

Tissue	EntrezGene	Symbol
adipose_brown	11302	Aatk
adipose_brown	11303	Abca1
adipose_brown	11303	Abca1
adipose_brown	11363	Acadl
adipose_brown	11364	Acadm
adipose_brown	11370	Acadvl
adipose_brown	11409	Acads
adipose_brown	11429	Aco2
adipose_brown	11429	Aco2
adipose_brown	11433	Acp5

Universe

EntrezGene	Symbol
11972	Atp6v0d1
57437	Golga7
100678	Psph
60409	Trappc4
13481	Dpm2
19173	Psmb5
52585	Dhrs1
19042	Ppm1a
66340	Psenen
17222	Anapc1

Tissues

"X3T3.L1"
"B.cells_GL7_negative_KLH"
"B.cells_GL7_positive_Alum"
"B.cells_GL7_positive_KLH"
"B.cells_GL7negative_Alum"
"B.cells_marginal_zone"
"Baf3"
"C2C12"
"C3H_10T1_2"
"MEF"

[図6]

CT

Tissue	EntrezGene	Symbol
Adipocyte	284119	PTRF
Adipocyte	284119	PTRF
Adipocyte	169611	OLFML2A
Adipocyte	158471	PRUNE2
Adipocyte	125058	TBC1D16
Adipocyte	116039	OSR2
Adipocyte	114885	OSBPL11
Adipocyte	113146	AHNAK2
Adipocyte	112464	PRKCDBP
Adipocyte	91851	CHRD1

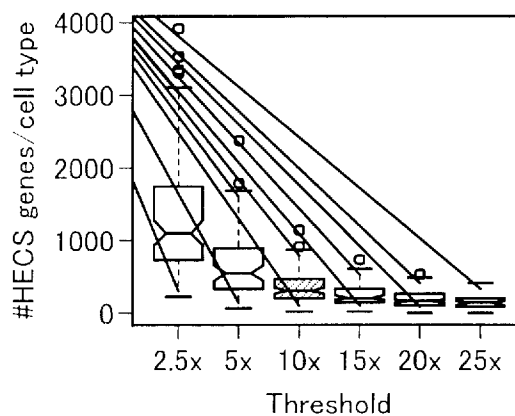
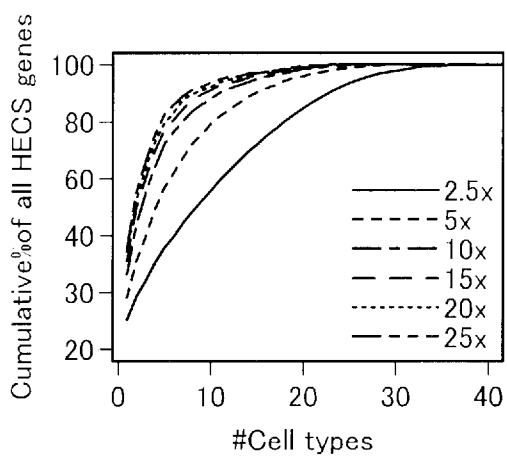
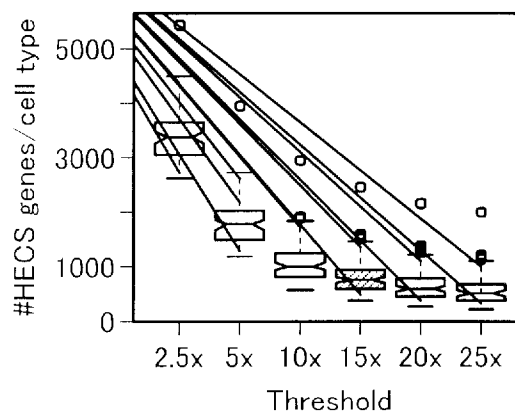
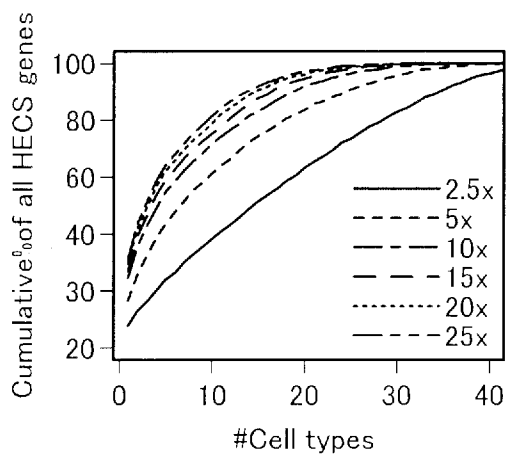
Universe

EntrezGene	Symbol
5982	RFC2
3310	HSPA6
7849	PAX8
2978	GUCA1A
7318	UBA7
7067	THRA
11099	PTPN21
6352	CCL5
1571	CYP2E1
2049	EPHB3

Tissues

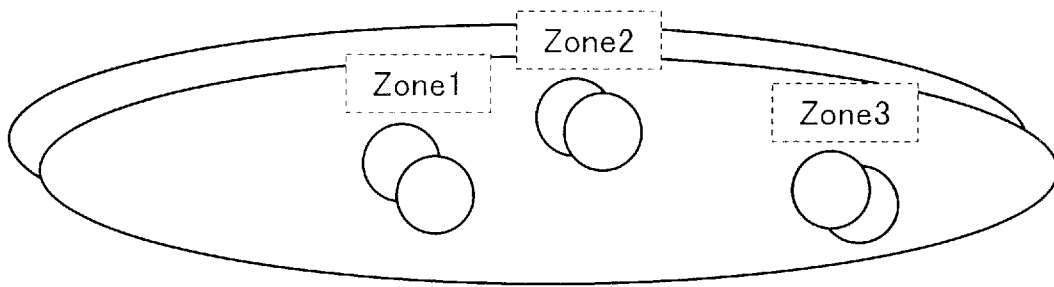
"X721_B_lymphoblasts"
"Adipocyte"
"AdrenalCortex"
"Adrenalgland"
"Amygdala"
"Appendix"
"AtrioventricularNode"
"BDCA4_DentriticCells"
"Bonemarrow"
BronchialEpithelialCells

[7]

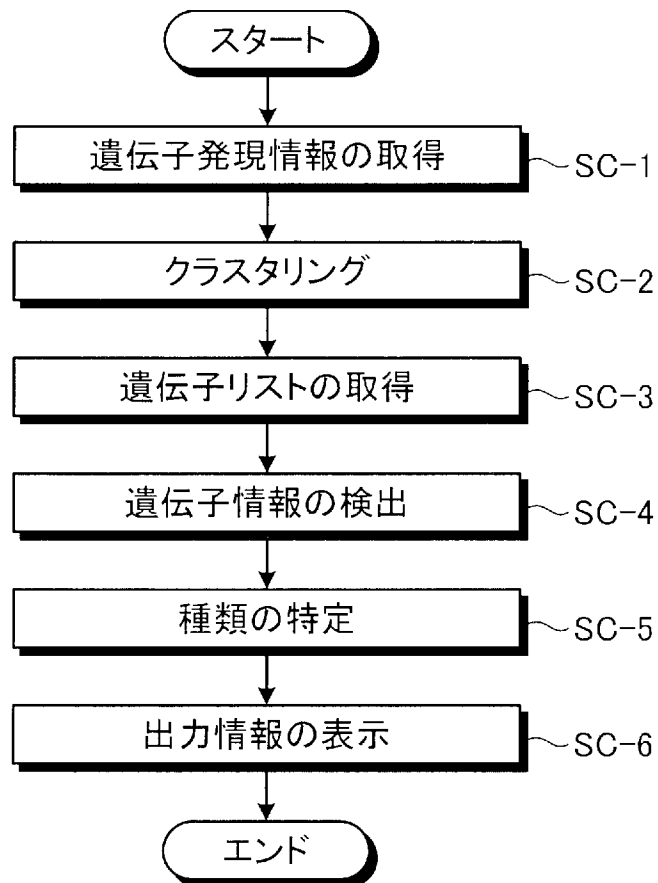




[図8]



[図9]



[図10]

Select Species:

Upload gene list      Gene ID type:

· Simple Example  
· Advanced Example

Separator (if multiple lists; limit 20): None  Comma   
Tab

Input gene list:

71701  
16391  
80287  
68591  
15039  
217684  
19698  
226695  
NULL  
76933  
75841

[図11]

Select Species:

Upload gene list      Gene ID type:

· Simple Example  
· Advanced Example

Separator (if multiple lists; limit 20): None  Comma   
Tab

Input gene list:

20292, Group1  
20293, Group1  
24047, Group1  
20296, Group1  
20297, Group1  
20302, Group1  
20303, Group1  
20306, Group1  
54199, Group1  
14825, Group1  
15945, Group1

[図12]

Upload gene list

- Simple Example
- Advanced Example

Preliminary processing.

The table below describes the number of genes in each list submitted which are in the database.

List Name	# User Genes	# User Genes in DB
User List	541	487

[Continue to Enrichment](#)

[図13]

Upload gene list

- Simple Example
- Advanced Example

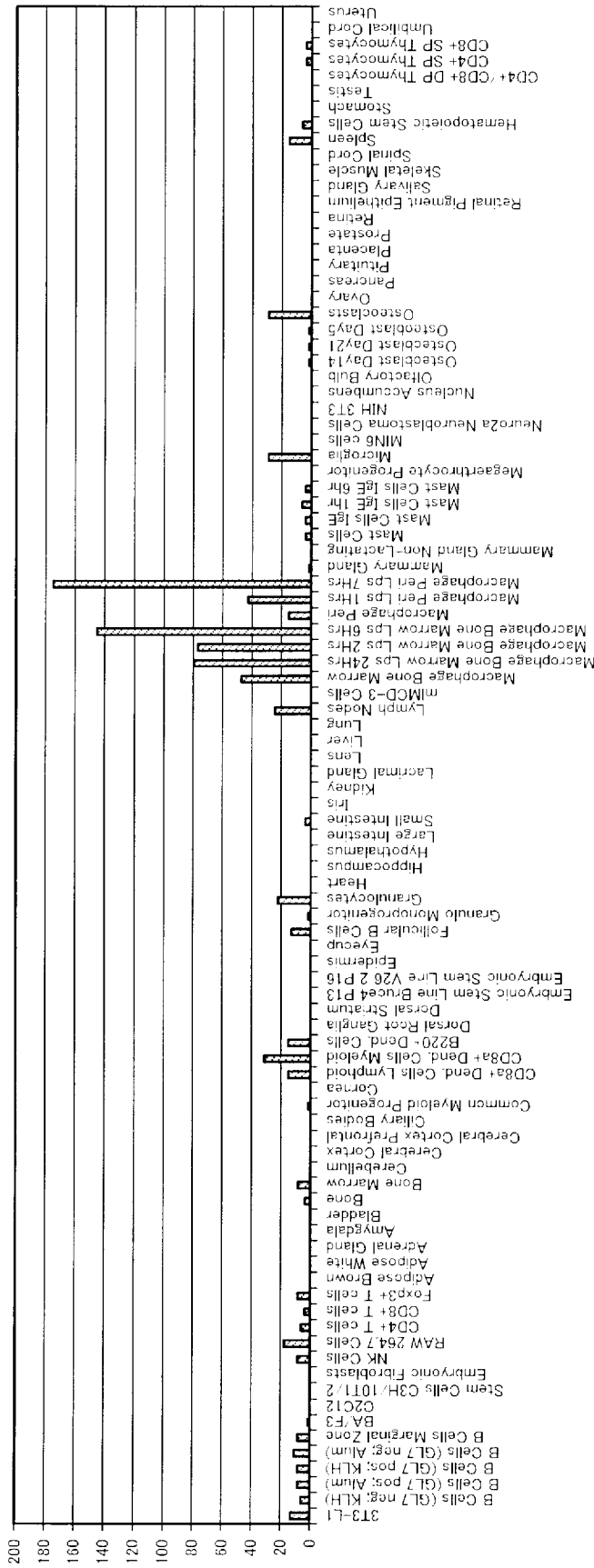
Preliminary processing.

The table below describes the number of genes in each list submitted which are in the database.

List Name	# User Genes	# User Genes in DB
Group1	106	94
Group2	87	83
Group3	106	96
Group4	57	46

[Continue to Enrichment](#)

[ 14 ]

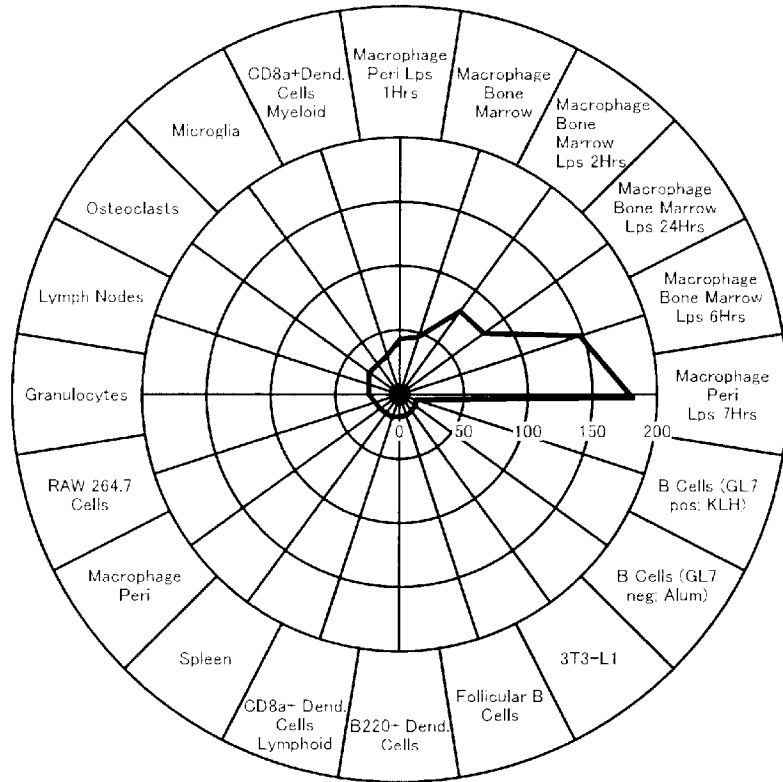


[圖15]

The 20 most enriched cell types. We recommend an enrichment score of 2 or greater to be considered significant.

Upload gene list

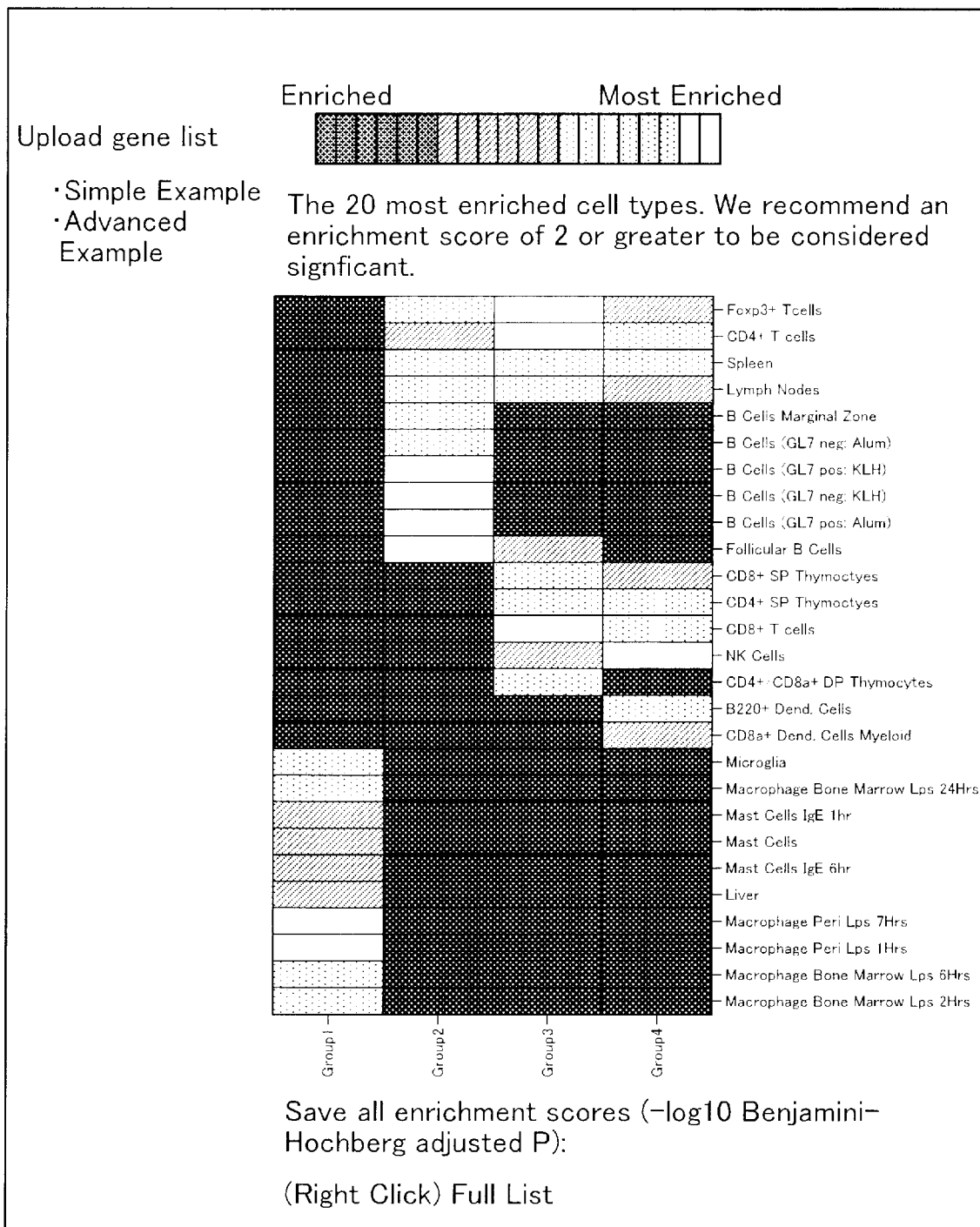
- Simple Example
- Advanced Example



Save all enrichment scores ( $-\log_{10}$  Benjamini-Hochberg adjusted P):

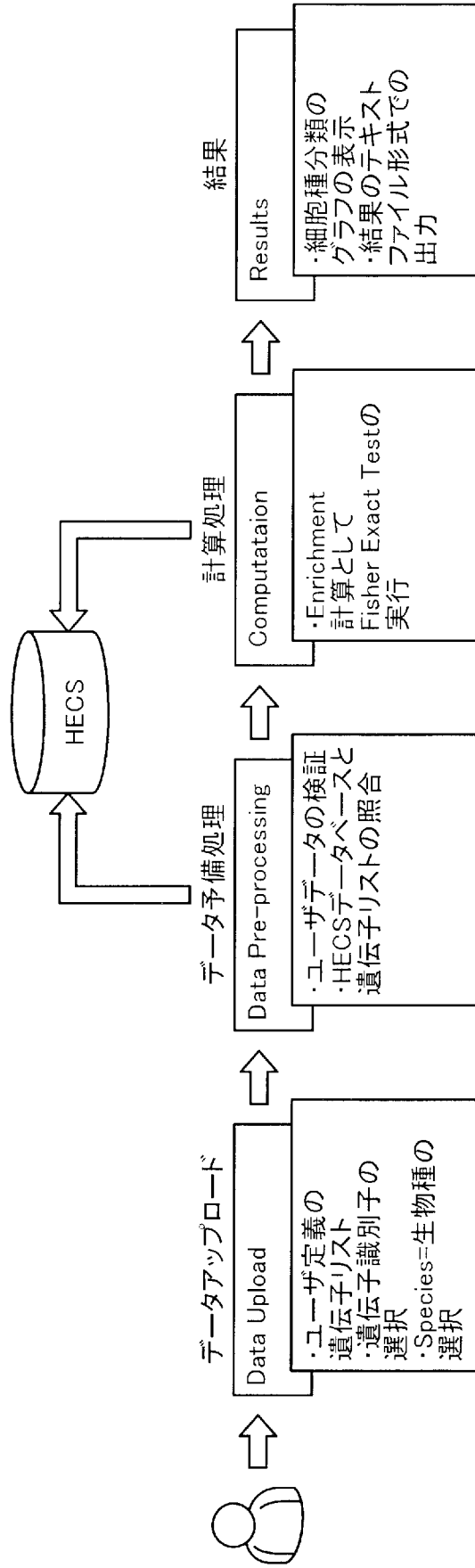
(Right Click) Full List

[16]

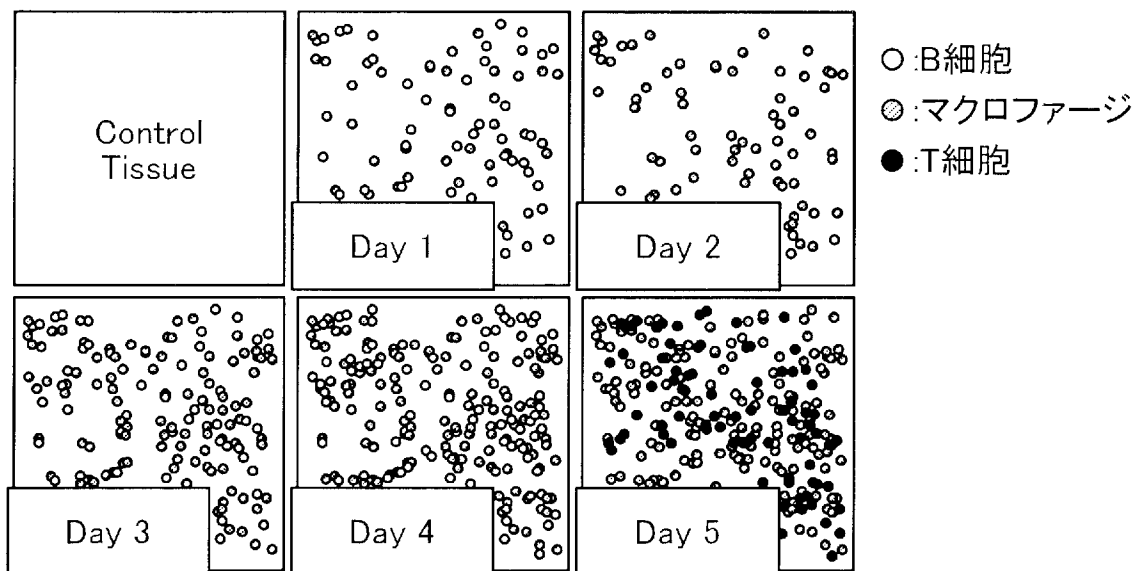


[図17]

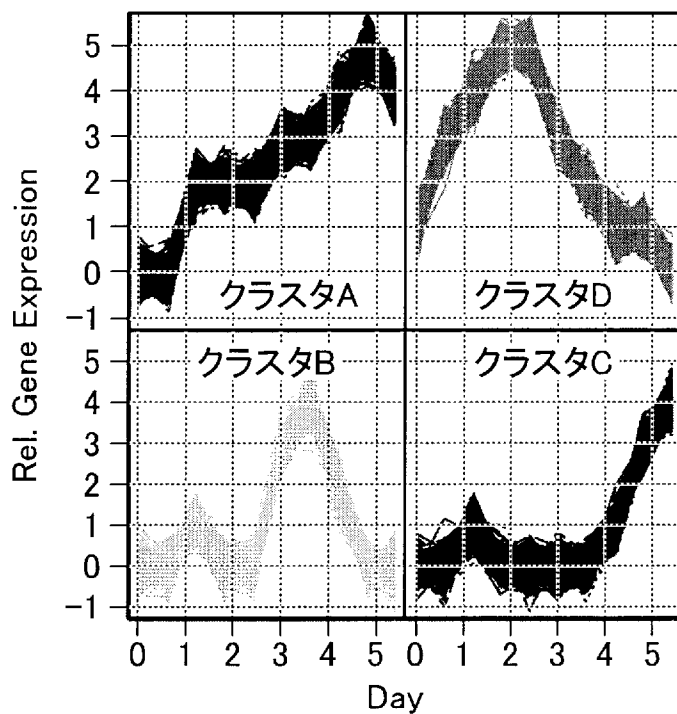
ワークフロー



[図18]

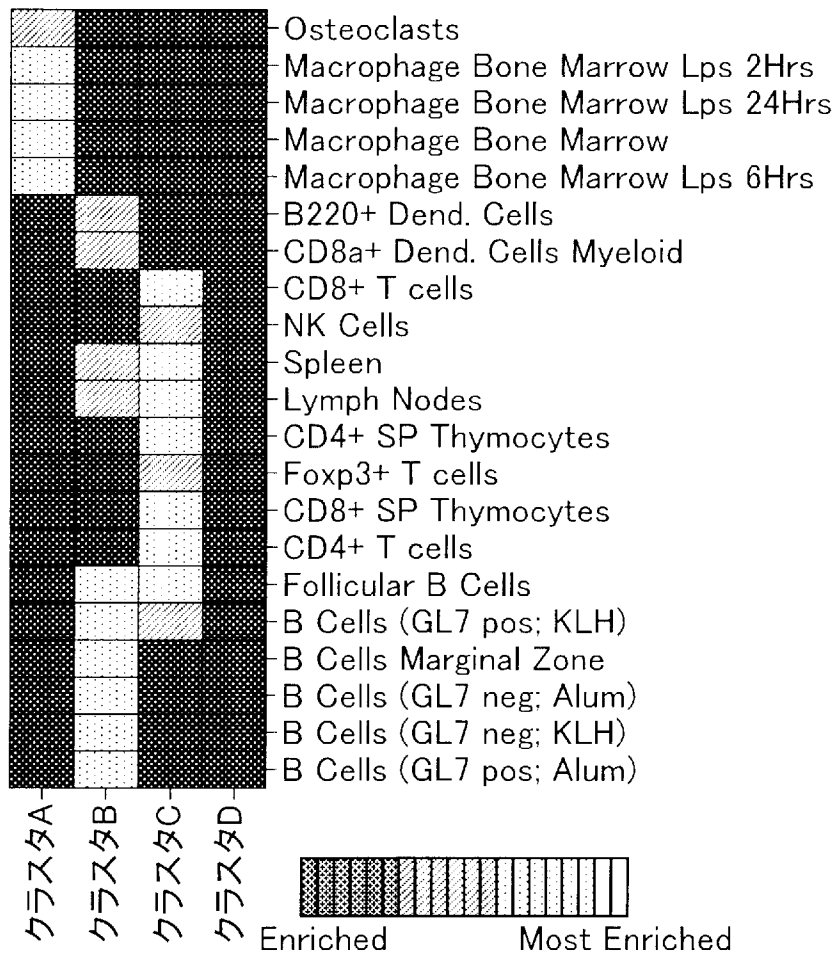


[図19]





[図20]



**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/JP2013/068489

**A. CLASSIFICATION OF SUBJECT MATTER**  
G06F19/20(2011.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
G06F19/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2006-092478 A (National Institute of Advanced Industrial Science and Technology), 06 April 2006 (06.04.2006), claim 1; paragraphs [0034] to [0035] & US 2007/0172833 A1	1-2, 4-5
X	HORTON, P.B., RaPiDS: an algorithm for rapid expression profile database search, Genome informatics. International Conference on Genome Informatics, 2006, Vol.17, No.2, p.67-76	1-2, 4-5
X	Wataru FUJIBUCHI, "Development of Cell Knowledge Base and Prediction of Cell Types and Characteristics by Gene Expression Profiles", IPSJ SIG Notes, 07 October 2005 (07.10.2005), vol.2005, no.99, pages 33 to 37	1-2, 4-5

Further documents are listed in the continuation of Box C.       See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 19 September, 2013 (19.09.13)	Date of mailing of the international search report 01 October, 2013 (01.10.13)
--	---

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2013/068489

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2006-271385 A (DNA Chip Research Inc.), 12 October 2006 (12.10.2006), claims & US 2005/0260572 A1 & WO 2002/072828 A1	1-2, 4-5
A	WO 2006/088208 A1 (Dainippon Sumitomo Pharma Co., Ltd.), 24 August 2006 (24.08.2006), entire text; all drawings (Family: none)	1-5
A	Kiyoshi YANAGISAWA, "Proteomic signature and outcome for patients with resected non-small cell lung cancer", Japanese Journal of Clinical Medicine (special extra issue) Haigan, 28 August 2008 (28.08.2008), vol.66, no.941, pages 143 to 150	1-5

A. 発明の属する分野の分類（国際特許分類（IPC）） Int.Cl. G06F19/20(2011.01)i		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） Int.Cl. G06F19/20		
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 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2013年 日本国実用新案登録公報 1996-2013年 日本国登録実用新案公報 1994-2013年		
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）		
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
X	JP 2006-092478 A（独立行政法人産業技術総合研究所）2006.04.06, 請求項1, 段落【0034】-【0035】 & US 2007/0172833 A1	1-2, 4-5
X	HORTON, P.B., RaPiDS: an algorithm for rapid expression profile database search, Genome informatics. International Conference on Genome Informatics, 2006, Vol.17, No.2, p.67-76	1-2, 4-5
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <span style="margin-left: 200px;"><input type="checkbox"/> パテントファミリーに関する別紙を参照。</span>		
* 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矛盾するものではなく、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 「&」同一パテントファミリー文献		
国際調査を完了した日 19.09.2013	国際調査報告の発送日 01.10.2013	
国際調査機関の名称及びあて先 日本国特許庁（ISA/J P） 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 宮久保 博幸 電話番号 03-3581-1101 内線 3562	5 L   3 1 3 6

C (続き) . 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
X	藤渕航, 細胞の知識ベース開発と遺伝子発現プロファイルによる細胞種と特徴予測, 情報処理学会研究報告, 2005.10.07, 第2005巻, 第99号, p. 33-37	1-2, 4-5
X	JP 2006-271385 A (株式会社DNAチップ研究所) 2006.10.12, 特許請求の範囲 & US 2005/0260572 A1 & WO 2002/072828 A1	1-2, 4-5
A	WO 2006/088208 A1 (大日本住友製薬株式会社) 2006.08.24, 全文, 全図 (ファミリーなし)	1-5
A	柳潭聖, タンパク発現パターンに基づく再発予測, 日本臨床 (増刊) 肺癌, 2008.08.28, 第66巻, 第941号, p. 143-150	1-5