

細胞分化解析を目指した 網羅的ヒト細胞データベース「CELLPEDIA」

幡野 晶子^{1,2}, Harry Amri Moesa^{1,3}, 千葉啓和¹,
谷口丈晃⁴, 永家聖^{1,2}, 山根木康嗣⁵, 藤渕航¹

本研究では、細胞を人体の物理的な位置に基づいて 2000 種類以上に細分化し、細胞画像、形態値、文献、遺伝子発現データを収集し統合した、細胞の網羅的データベースを作成した。さらに成体の組織幹細胞、前駆細胞を約 60 種類に分類し、情報を収集した。本データベースを用いることにより、細胞の形態値、遺伝子発現データをもとに、細胞の個性を定量的に解析することが可能である。

CELLPEDIA: Comprehensive human cell database toward cell differentiation analysis

Akiko Hatano^{1,2}, Harry Amri Moesa^{1,3}, Hirokazu Chiba¹,
Takeaki Taniguchi⁴, Satoshi Nagaie^{1,2}, Koji Yamanegi⁵,
Wataru Fujibuchi¹

In this research, we construct a comprehensive database for human cells, in which various types of information including cell images, morphological measurements, related journal articles and gene expression data are collected and integrated. In this database, the human cells are subdivided into more than 2,000 cells based on their physical locations. Similarly, adult stem cells and progenitor cells are classified into about 60 kinds of cells, and their related information is collected. By using our database, cell characteristics can be quantitatively analyzed.

1. はじめに

ヒトには 200 種類以上、約 60 兆個の細胞が存在していると考えられ、その構造と機能を理解することは、人体の正常な機能、疾患メカニズムの解明において重要である。さらに同じ名称がついている細胞であっても、細胞の年齢、周囲の環境、人体における物理的存在位置等の要因により、性質が異なる可能性がある。このような細胞の個性を認識することは生体を理解するために必要不可欠である。そのため我々は成体に含まれる正常細胞、幹細胞を物理的位置に基づいて細分化するとともに、培養細胞を加えた 3 種類の細胞カテゴリーを設け、ヒト細胞に関する網羅的データベースを作成した。これにより、今後得られるであろう 1 細胞レベルのデータにも対応できる、細胞の電子辞書的役割を備えた細胞のデータベースの提供を可能とした。

2. CELLPEDIA の構成

CELLPEDIA は細胞分類表、細胞画像、文献、遺伝子発現データの 4 つの基本情報と、細胞画像に対してそれらを集約し、細胞画像の形態値情報、オントロジーを加えた統合ページから構成されている。さらに遺伝子発現データを用いて作成した遺伝子モジュール辞書、キーワードによる検索機能を付随している。トップページと統合ページをそれぞれ図 1 に示す。また構成内容に関して以下に順を追って説明する。

¹産業技術総合研究所 生命情報工学研究センター

²東京医科歯科大学大学院 歯学総合研究科

³東京大学医科学研究所 ヒトゲノム解析センター

⁴(株)三菱総合研究所

⁵兵庫医科大学 病理学講座 機能病理部門

¹Computational Biology Research Center, Advanced Industrial Science and Technology

²Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University

³Human Genome Center, Institute of Medical Science, University of Tokyo

⁴Mitsubishi Research Institute Inc

⁵Department of Clinical Pathology, Hyogo College Of Medicine

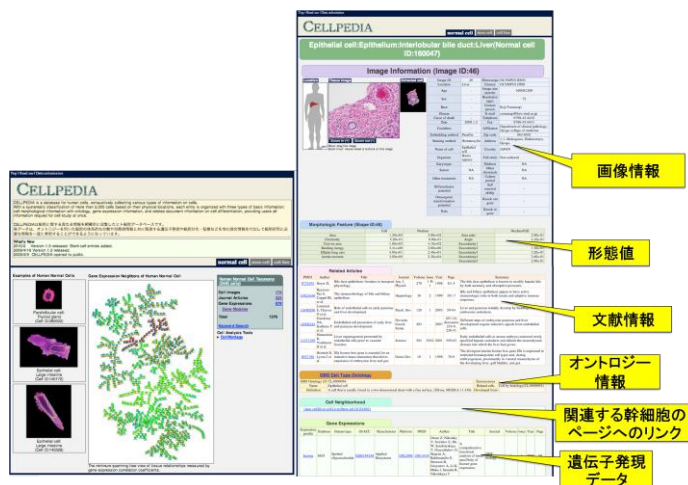


図 1 トップページ (左) と細胞情報統合ページ (右)

2.1 細胞分類

現在分類されている細胞は、十二指腸や大腸に存在する杯細胞のように、存在する環境、すなわち物理的存在位置が異なっている、同じ名称がつけられている細胞も少なくない。そのような細胞と物理的位置数の例を表 1 に示す。これらの細胞は遺伝子発現レベルでは、異なる性質を持つ可能性がある。そのため我々は、今後 1 細胞レベルでの細胞機能解析が可能となる時代に対応できる、詳細な細胞分類表を作成した。具体的には、人体の物理的位置に基づき、各器官または組織から細胞までを最大 8 階層まで階層化し、細胞を 2446 種類に細分化した。さらに幹細胞についても 62 種類に分類し、同様に情報を収集した (図 2)。

表 1 細胞の物理的存在位置数

細胞名	物理的存在位置数
fibroblast	297
smooth muscle cell	264
epithelial cell	122
endothelial cell	118
nerve cell	44
goblet cell	31

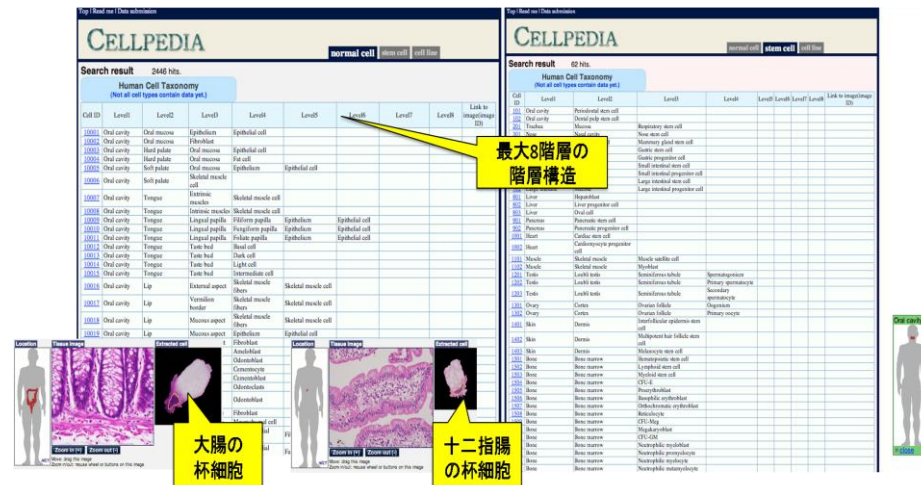


図 2 細胞の分類表、物理的存在位置の異なる杯細胞の例 (左) と幹細胞の分類表 (右)

2.2 細胞画像

本研究では、物理的位置に基づいて分類した、ヒト細胞カタログに対応する各細胞画像を収集し、細胞の電子辞書の役割を持つデータベース構築を目標としている。現在の画像登録数は、組織画像を含め正常細胞が 174 件、幹細胞が 30 件である。各画像に関して、画像登録者が提供した情報を記載した (図 3)。顕微鏡の倍率に関しては今後追加する予定である。

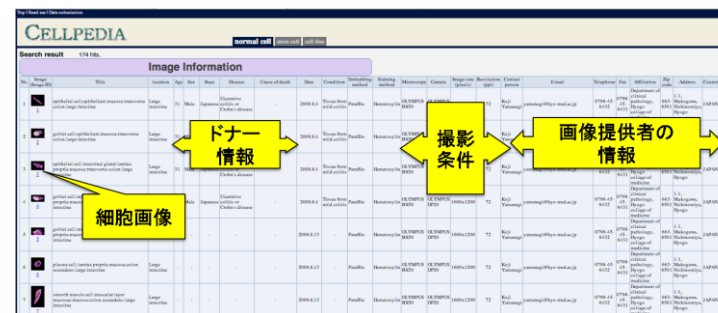


図 3 撮影装置や画像登録者情報を記載した画像情報のページ

2.3 文献データ

各細胞の生理学的機能、特異的に発現している遺伝子等の特徴を記載している論文を検索し、正常細胞に関して 324 件、幹細胞に関して 76 件の論文を収集した。雑誌名、著者名等の情報を表示するとともに、Pubmed や本データベース中の関連する細胞の統合ページにリンクを張ってある。さらに医学をバックグラウンドとする研究者が論文を読み、内容を要約して表示することにより、細胞の特徴を理解しやすくなった (図 4)。



図 4 文献情報

2.4 遺伝子発現データ

公共のデータベース GEO[1]、ArrayExpress[2]から正常細胞に関して 878 件、幹細胞に関して 92 件の遺伝子発現データを収集し、組織特徴的遺伝子、発現分布、ヒートマップと遺伝子リスト、論文や実験情報を表示した。現在 1 細胞レベルのデータが少ないため、組織あるいは複数の細胞から得られたデータがほとんどとなっている。1 枚のアレイ上で発現量の多い上位 10 遺伝子、発現量の少ない下位 10 遺伝子を表示するとともに、発現量の分布をヒストグラムで表した。さらに全発現データを用いて、自己組織化マップ (SOM) を作成することにより、発現量のビジュアル化を可能とした (図 5)。ヒストグラム、SOM の図をクリックするとその値を示す遺伝子リストが表示される。

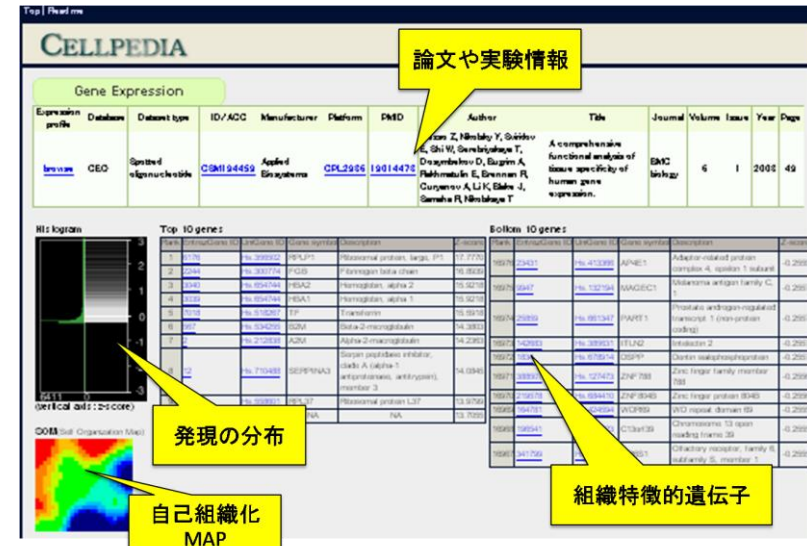


図 5 遺伝子発現データのページ

2.5 形態値

各組織画像から、画像中に含まれる 1 細胞を対象に、独自に開発したツールである Cytometorica[3]を用いて、細胞の形態値を測定した。多くが HE 染色によるノイズの大きい画像であり、NIH-Image、CellProfiler のような他の測定技術では、細胞の輪郭を認識することが困難であった。しかし Cytometorica を用いることにより、ほとんどの HE 染色画像に対して形態値を得ることができた。測定した形態値は、細胞と細胞核の面積、凸包面積、円形度、慣性モーメント、bending energy、楕円長軸長、細胞と核の面積比率、細胞の主軸核の比、細胞内における核の偏在性である (図 6)。

Morphologic Feature (Shape ID:127)				
	Cell	Nucleus		Nucleus/Cell
Area	8.88e+02	3.65e+02	Area ratio	4.11e-01
Circularity	6.23e-01	8.58e-01	Angle	1.99e-02
Convex area	1.05e+03	4.08e+02	Excentricity1	4.47e-02
Bending energy	2.00e+00	1.31e+00	Excentricity2	3.87e-02
Elliptic long axis	3.66e+01	2.38e+01	Excentricity3	7.19e-02
Inertia moment	1.29e+05	2.18e+04	Excentricity4	5.94e-02
			Excentricity5	8.33e-02

図 6 Cytometorica により測定された細胞の形態値

2.6 OBO cell type ontology

細胞のオントロジー情報を提供している The open biomedical ontology [4]に掲載された細胞名と、本データベース中の細胞名との関連づけを行った。正常細胞に関して 87 件、幹細胞 42 件に対し、オントロジー情報を表示した。

2.7 遺伝子のモジュール辞書

我々が開発した遺伝子モジュール高速マイニングシステム、SAMURAI [5]を用い、収集した遺伝子発現データからバイクラスタリング法により、ヒト正常細胞間の遺伝子モジュールデータを作成した。遺伝子モジュールを抽出することにより、組織特異的、あるいは異なる組織で共通する遺伝子発現パターンを両方を探索することが可能である。各モジュールのページには KEGG パスウェイ解析、遺伝子オントロジーに関する情報を記載しており、細胞接着性、代謝パスウェイ等、モジュールが関連している生物学的機能を確認することができる。

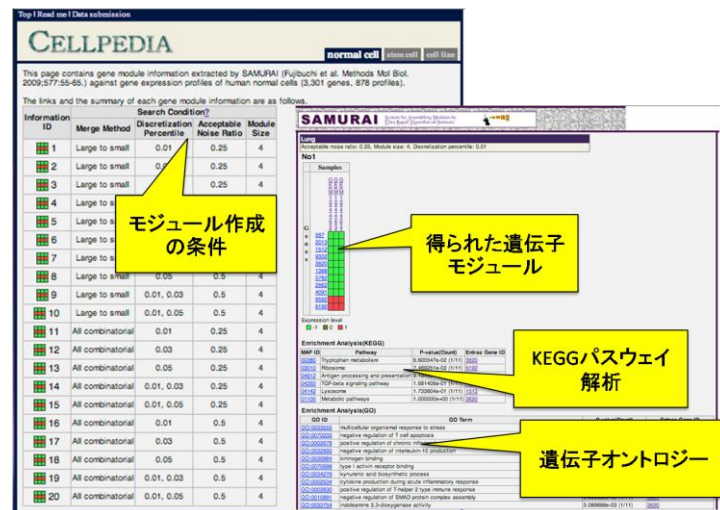


図 7 遺伝子モジュールのリスト (左) と各モジュールの詳細ページ (右)

3. CELLPEDIA の応用 (細胞の形態値、遺伝子発現データを組み合わせた細胞の分化解析)

細胞は同じ組織中に存在する、同種の細胞であっても、細胞周囲の環境や、細胞の年齢等、様々な要因により、遺伝子発現や細胞の形態に違いが生じていると考えられる。従って本研究で測定した形態値は、個々の細胞の状態を反映した値と捉えること

ができる。形態値を用いた解析例の 1 つとして、形態値と遺伝子発現データを組み合わせ、細胞の形態に関連する遺伝子の同定が挙げられる。さらに応用として、細胞が成熟するまでの各段階の形態値を抽出し、整理することにより、形態値情報に基づいた細胞の分化状態の予測や、遺伝子モジュールとの組み合わせによる細胞の分化転換経路の探索等、細胞の分化解析を目指している。

4. おわりに

本研究では、画像、論文、遺伝子発現等、細胞に関する情報を網羅的に集め、これら膨大なデータを収容することができる、ヒト細胞統合データベース「CELLPEDIA」を構築した。今後は細胞画像の充実に加えて、細胞分化や分化転換情報、細胞レベルの遺伝子発現データを増やし、多くの研究者に利用される「電子細胞辞書」的役割を果たすデータベースを構築する予定である。さらに独自に抽出した形態値情報と収集した遺伝子発現データを組み合わせることにより、細胞分化解析を可能とするデータベースを目指している。

参考文献

- 1) Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>
- 2) ArrayExpress, <http://www.ebi.ac.uk/microarray-as/ae/>
- 3) 細胞輪郭抽出装置、細胞輪郭抽出方法およびプログラム 特願 2006-331416、2006
- 4) The Open Biomedical Ontology, <http://www.obofoundry.org/>
- 5) Fujibuchi et al. *Methods Mol Biol.* 2009;577:55-65.