

生命科学辞書とオントロジーの自動構築法の開発

●大久保 公策

国立遺伝学研究所 生命情報・DDBJ 研究センター 遺伝子発現解析研究室

<研究の目的と進め方>

目的： 本課題は生命科学の専門用語間の意味的な関係を自動的に算出することで用語が作る専門分野の知識を形式的に表現する方法の開発です。算出された用語の意味関係が全体として表現する知識は

- (1) 生命科学の文書やデータベースレコード等の並べ替え
- (2) 記述的データを用いて行う探索的研究での専門知識を動員などで利用することを想定しています。

分子生物学や遺伝学の手法で進める基礎医学領域では疾患サンプルの遺伝子発現解析や遺伝性疾患の原因遺伝子単離などがよく行われていますが、これらの知識の生物学的な利用やまた基礎生物学からの解釈の助けを期待する際にまさしく両者間での知識のトランスレーションが問題となっています。

専門用語の意味的な関係を計算機で扱える形に表現する方法には二つのアプローチがあります。一つはオントロジーに代表されるように人間が宣言的に行う方法です。宣言的方法では意味関係は概念の上下関係をつないだ人間が理解しやすいグラフ構造（階層分類（木）や DAG）や表現します。一番近い上下関係にある用語同士をリンクさせることを繰り返して得られます。いまひとつは文書をデータとして用語の分布から統計的に計算する方法です。統計的方法では用語の文書間などでの出現パターンの距離で用語の意味の関係を表現しようとする方法でパターンを距離に直す手法が様々な検討されています。

基礎から臨床までのすべての生命科学領域をカバーした用語集には NLM の MeSH が知られています。MeSH は NLM が扱う文献に対するインデックス名称として集められた用語であり、実際に文書で使用される用語との重複は必ずしも多くありません。

また基礎生命科学領域では特に遺伝子の機能を表現し生物種を超えた知識の整理を可能にする道具として GeneOntology が広く使われています。すなわち両者とも宣言的な手法であり、広汎な医学知識を有効に表現する統計的手法は見られません。

これら宣言的な手法はいわばマンパワーの投下によって一定の水準での目的を達成しています。しかし索引用語集と構造の絶え間ない更新と新規文書や新規遺伝子への絶え間ない索引付けを必要とし、文書の爆発と分野の知識の更新や再編成に常に対応できる機械化可能な方法が必ず必要であると考えました。

一般に統計的手法も宣言的手法もそれぞれに長所と短所が知られています。統計的に行くとデータとして利用した文書に構造が依存することや出来上がった関係が人間に理解しにくい、また手作業で後から修正しにくいことが問題です。宣言的に行くと当然宣言者に依存し、同じ宣言者でも再現性の低い問題があります。また枝の距離は対象分野の認識の解像度を表すために出来上がった構造は局所的な意味関係しか表せない（大きな距離には意味が

ない）という問題があります。

一般的には大規模に多様な内容を扱う場合は統計的に、限定分野の精緻な用語に対しては宣言的に行うことが適当です。生命系全体では解析的に研究を進めたために理解が時代とともに詳細化していったために構造や概念の名称が階層的になっている特徴から宣言的な用語の構造化になじむと考えられていますが、医学系では分子名称から疾患や症状に関わる名称の数は膨大であり同じ宣言ルールですべてをカバーすることは容易ではないと考えられています。本計画では用語の意味関係を構造化する方法の開発ですが、作られた方法は利用によってしか評価ができないためにタスクを二つもうけ宣言的分類と統計的手法の両者を行いました。

タスク 1. ESTライブラリ情報からの効率的解剖用語の収集と半自動分類

タスク 2. 教科書を利用した専門用語の収集と構造化

<研究開始時の研究計画>

タスク 1.

臓器名称が自動的に検出可能でその臓器の系統分類にしたがって離散会合操作することが可能になれば蓄積している多種類の生物の遺伝子発現データを使って異なる生物の遺伝子発現パターンを比較することが可能です。遺伝子間の関係はオーソログ情報を利用すれば遺伝子の臓器別の発現様式のファミリーメンバー間での発現パターンの変化などを調べることが可能になり、遺伝子発現の進化についての洞察を得ることが出来ます。

たとえば現在広く知られている大野進氏の重複遺伝子生成仮説ではひとつの遺伝子が重複するときに重複と同時に発現パターンも分割されるので重複したコピー遺伝子が生き残るという説明がある。この説明はいくつかの臓器特異的に進化したパラログファミリーをよく説明するが多くの遺伝子の例で確認されてはいません。

<研究期間の成果>

- 1) ライブラリ情報を集めるために、EST-DDBJ エントリをライブラリ単位にまとめ、それぞれのライブラリの中からの材料に関する記載のあるフィールドを集める。
- 2) 分類効果の大きなルールにするために EST 数の大きなライブラリからトップ 100 のライブラリの該当フィールドを見て、臓器単位に分類する。
- 3) 分類の根拠を与えたパターンを集めて正規表現のパターン辞書を作る。
- 4) パターン辞書と同じ 100 のライブラリ情報を与えて、誤回答の例につきアドホックな例外集を作る。
- 5) 確認後のこりのすべてのライブラリ情報を処理する。
- 6) 分類不能であったライブラリから大きなものを 100 選んで手作業で分類しパターン辞書に加える。

- 7) 分類不能が1割以下になるまで3) から6) を繰り返す
分類機を他の哺乳類にあてはめすべてのESTデータを種 x 臓器分類で整理しなおす。

のステップを繰り返す

http://bodymap.jp/organ_tissue_rule

に記載されているような解剖学名称ボタン辞書を得ました。

以降下のような経緯でボタン辞書の公開とデータの更新を行ってきました。

現在は50種類の生物の600種類のライブラリ由来の3500万件のESTレコードを40臓器に分類して公開しています。設計後5年を経過した分類機ですがヒトのライブラリ分類力はいまだに80%を超えておりクリーン数では90%以上が分類できています。開発後の維持修正作業のリストは以下のようです。

DDBJ リリース 63 に登録されている50種類の生物の合計

2008-05-12 DDBJ release73 processed

12 new animals 1918583 new ESTs

2007-04-19 * Gene name search: Gene symbol support

DDBJ release68 processed

2 new animals 1196172 new ESTs

2006-09-06 * DDBJ release66 processed

1 new animal 600597 new ESTs

2006-06-04 * DDBJ release65 processed

4 new animals 3720010 new ESTs

Data filter added to every expression profile representation

* anatomical tagger service opened to the public

2006-01-26 * RefSeq, Entrez Gene ID supported in ID search

* Minor bugs in bar chart representation corrected

* Bar chart modified to be linkable

2005-12-22 * DDBJ release63 processed

6 new animal 2067185 new ESTs

* Organ dictionary modified: Ver.2005-11-25.1

* Status dictionary modified: Ver.2005-11-25.1

* Process dictionary modified: Ver.2005-11-25.

* Gene name search: UniProt/Swiss-Prot protein name support

* Homology search(blastp) supported

さらに同じ分類機をSAGE・EST・マイクロアレイ・iAFLPの測定法に適用し、異なるプラットフォームでの発現パターンを比較するビューアサイトを作成した。

http://okubolab.genes.nig.ac.jp/bodymap_i/

それぞれの方法が与えるパターンを比較すると相互の不一致に一定の傾向を見出した。すなわち遺伝子それぞれにライブラリ別の濃度を与えその最大値を使って遺伝子を発現検出のしやすさで容易・中等度・困難3分類すると特に困難な部分で著しく3つの方法由来の発現パターンに一致が見られないことが明らかになった。困難部分ではマイクロアレイはハウスキーピングパターンが大半を占めるが、タグ同定法では逆にどこかの組織に偏る傾向が強く、iAFLP法ではその中間であった。

<国内外での成果の位置づけ>

開発した自動分類機はBodyMap-Xs、植物bodymap、統合データベースプロジェクトにおけるGEO目次などに利用されています。分類機自体ではなくその使用例としての種間の遺伝子発現の比較を公判に行っている例は見られません。種を超えて遺伝子の発現パターンを比較するところみはマイクロアレイベースのヒト霊長類ネズミ程度の同一臓器での発現量の比較が多く見られる。いずれも特定臓器での遺伝子の発現量が“中立”に進化している

という予定調和を支持するものである。BodyMap-Xsの本体は実は解剖名称の自動分類プログラムである。これはむしろ主要な生物データセンターに存在するそれぞれの生物の解剖オントロジーであろう。オントロジーと比べると概念数もすくなく階層的に構造化されているものではないがライブラリ材料の分類というタスクに特化して自動分類に使えるボタン辞書は他に存在しない。

<達成できなかったこと、予想外の困難、その理由>

分類材料を対象として宣言的に分類する方法はたとえ対象が数千におよんでも再帰的に行うことで比較的容易に行うことができた。しかしながら頻繁に繰り返されるライブラリデータの増加にともない生じる分類不能例を用いて適宜分類辞書にボタンを加えることは自身が時間を避けなかったために不可能であった。広く医学知識を利用しなければならないタスクを以外系の知識の少ないグループ内で行うことの困難さを痛感しました。

一方で解剖学名称の医学系で広く利用される用語集にはNominaAnatomicaがありさらに構造化されていないNAをis-aとpart-ofを中心に関係付けたオントロジーにFoundationModelofontologyが既に公開された。FMAはTAを構造化するために人為的なクラス概念を多数付加しており細胞名称までを加えたとはいえそのノード数は7万を超えている。われわれの目的にFMAのマッピングだけではマッチ率は数十パーセントと低いがこれらの活動をライブラリ名称の自由なグループ化に使用するためにもこれらの外部の既存オントロジーを実際にライブラリの記述に用いられている用語にマップしてゆく方法を作ることが広く利用されることを保障すると考えられます。

<今後の課題、展望>

OBOで公開されているFMAオントロジーを長年つくってきたDr.Mejinoと会い、現状と方向性を確認しました。

最大の解剖分類の電子版であるアメリカ病理学会のオントロジーは北米以外では自由に利用できない事情もあり今後の詳細化は彼らのものを利用することに決定しました。

一方で彼らの辞書は架空の分類名称が多またDAGも多いために文書にそのままぶつけるには好ましくありません。彼らと連絡を取りながら目的を果たすオントロジーに縮小し、さらにこれまで作成したボタン辞書分類機を用いて実際に使用されているボタンが同じように分類されるかをテストし、両者を利用したボタンに富んだ分類機を作成しGEOなどそのほかの発現データや文書の分類に進む計画です。

<研究開始時の研究計画>

タスク2 統計的用語の構造化

臨床医学から基礎医学まで医科学系では解剖学および疾患名称や病理学名称が特に多種類の用語を利用しています。近年これらの医学概念が遺伝子やたんぱく質のレベルで詳細に研究されています。ここでの遺伝子や蛋白質と臓器や疾患との関係は上位下位の概念関係や部分全体の関係ではなく、トピカルリレーションと呼ばれる場面と役者の関係です。すなわちテニスボールとラケットや野球のボールとバットの関係はオントロジーでよく使われる上位下位(is a)や部分全体(part of)では記述できません。これらの関係がtopicalな関係でこれは競技について記述した文書などで簡単に検出可能です。

このように医科学の文書や遺伝子の機能説明の意味的な操作や

分類のためには上位下位や部分全体だけでなくトピカルな関係を測れる用語の分布を用いた統計的な分類が適当です。

文書を材料に用語の関係を作る方法は用語の意味を登場する文書の組み合わせで表現するベクトル空間法が代表的です。文書の内容を全用語数の次元のベクトル(文書ベクトル)と考えれば、同じ文書に共起の多い用語ベクトルが大きな内積(余弦)を与え全く共起のない用語同士は直行する(余弦ゼロ)という直感的な関係を与えます。本計画でもこのベクトル空間法を用います。

しかしながらこの方法を生命科学のように専門用語が多く同義や類義などの関連にみちた語彙を持つ分野にそのままあてはめると以下のような問題点があります。

- 1) 集める文書には同義語や表記ゆれ類義語がみちておりそのまま機械的にインデックスしたのでは類似用語の関係も検出できない。
- 2) 文書の集合によって用語の共起関係などがことなり用語集合の分類に再現性がない。
- 3) 得られた類似用語を特徴付ける方法がタグクラウドなどの用語による方法以外に友好なものがない。

以上の問題を克服するために

- 1) データとする文書集合として教科書を利用する。
 - 2) 教科書内教科書間での類義語関係を検出するためにLatent Semantic Analysis (LSA)を用いる。
- の二つのアイデアを提案しました。

アイデア1) a. 文書集合として教科書を用いると教科書の内容は非常にトピックの粒度が調整されており、内容の重複や不足など任意の文書集合から出発するよりも有利であること。

b. 教科書では既に著者が重要であると考えた用語の選択による索引付けが終わっていること。これは複数の単語でつくられる用語のどこまでを意味単位であると認めるか、いわゆるチャンキングの問題を回避できる。

c. 教科書は解剖学や生理学など分野によってわかれており、それぞれの小分野の専門用語を分離採取可能でしかもその分野の体系に従った意味関係を検出可能である。

d. LSA(後述)では文書(ここではページ)も用語と同じ空間にマップされるので単語の内容が目次を使えば項目見出しでも表現できる。

などの利点が考えられます。

アイデア2) LSAは90年代にベル研究所で開発されたベクトル空間法による高感度な文書の検索方法です。その中心となっているアイデアは用語x文書のスパースなデータに用語間の共起による意味的な類似性を反映しようというものです。データ行列ではすべて無関係として扱っている用語AとBでも文書中でよく共起すれば意味的に近いと考えAだけが使われている文書にも弱くBでも索引付けしてあげるという方法です。データ内の用語パターンをさらにデータに当てはめるといふ少し難しい計算をはじめのスパースなデータ行列を次元を下げて近似した行列に誤差としてあらわれるもともとは0であった多数のセル中の値が数学的にも用語間の共起度数を反映した索引値に等しいことが証明されています。LSAでは用語の意味も文書の内容も同じ空間のベクトルとして扱うために教科書の索引データによる用語の構造化に利用できるはずである。

<研究期間の成果>

以上のアイデアに基づいて以下のような方法を考案しました。

- 1) 分野別に教科書データを用いて十分な数の用語とそれぞれ独

立な内容であるはずの教科書の項目によるベクトル空間を作成する。

教科書150冊合計5万ページ強の目次および索引データを収集し管理データベース化しました。教科書リストは http://222.151.240.4/project/vbob/nonsense_dic/bob_textbook.txt にあります。

- 2) 教科書を分野別に分類しひとつの分野が数冊の教科書の和のデータで構成できるようにします。教科書索引には同義語を括弧つき表記で表現したものや A See B で表記したものが、これらから同義語辞書を作成します。
 - 3) A gene, A protein, A などほぼ同義語として使われる用語パラエティーを作る接尾単語のリストを作成し、教科書間の索引のマージを進めます。
 - 4) パッケージプログラムを用いて用語xページ行列を次元下げします。この際後の計算の負担を減らし実際によく使われている50次元での近似を行う。いわばあらゆるページ内容は独立の50項目の線形和で近似できるという過程である。
 - 5) 適当なプログラムライブラリを用いてSingular Value Decompositionを行い、近似行列を得ます。得られた近似行列はそのまま用語とページの関係を与えます。
 - 6) SVDで作った用語ベクトルは任意の文書の内容の表現のためにDB化して保持します。
 - 7) 近似行列が与える用語とページの関係オリジナルの索引と比較して望ましい潜在的関係が得られているかを確認します。
 - 8) 類似パターンを与える用語を目視でチェックして同義語の検出を行い、十分量検出したのちにステップ3の同義語辞書に加え5以下を繰り返します。
 - 9) 50次元の用語意味ベクトルとページ意味ベクトルが出来上がります。
- ここまでで**用語の意味の形式表現**が完了します。

用語の意味的表現を利用した文書内容の表現

任意の医学文書は教科書用語で索引付けします。

- 1) 文書内容は索引付けされた用語をあらわす用語ベクトルの線形和として表現します。用語の重み付けには標準的なTF/IDF(ありふれた用語は小さく、その文書だけで繰り返す用語は重く)
- 2) 文書同士の関係は意味ベクトルの余弦で与えられます。
- 3) 文書内容のわかりやすい表現には文書ベクトルと教科書ページベクトルの余弦によって得られる教科書の目次項目との類似度パターンで得られます。
- 4) 文書に与えられたベクトルを利用して文書間のすべての余弦を計算してクラスタリングを行えば文書群の内容によるクラスタリングを行えます。
- 5) 出来上がった文書クラスターの意味づけは教科書項目との類似性パターンを用いて行います。
- 6) 特定の文書に類似する内容の文書は余弦の小さい順に序列化し類似文書をクラスタリングすることで類似する観点別に分解することも可能です。

遺伝子機能の形式表現

遺伝子の機能表現を文書として扱います。ここでは Entrez Gene で遺伝子に付与されている文献アブストラクト、複数あればそれをつないだ文書を作り、使用しました。文書と全く同じ方法で、遺伝子機能をベクトル表現できます。

表現の妥当性の検討（方法全体の評価）

遺伝子の機能表現が妥当であるか否かの、また GO などに比べ優れたところがあるかの検討は以下のように行いました。

- ① 遺伝子ファミリーの機能表現の共通性およびファミリー内の差異の表現力
- ② 遺伝子発現データの特徴解釈の論文との比較

①は Pfam などの DB で与えられている進化的な遺伝子ファミリーのグループ別に形式表現を受けた遺伝子（教科書のページとの類似パターン）を並べパタンの共通な部分および異なった部分について知られている知識を再現しているかを目視する。GOA の場合にはファミリーはたいてい同じ GO が与えられていたが時に本法では組織特異性が表現できていることがありました。

②は論文などからマイクロアレイデータを取得し遺伝子発現のパターン順に遺伝子列を作り、その順に教科書ページパターンを与えられた遺伝子を並べ替えました。または遺伝子間の類似性行列を色調の濃淡表示であらわした遺伝子 x 遺伝子の機能距離行列を作成して、同じように並べ替えると、機能的に類似した遺伝子が発現データによってクラスターされている場合には濃意部分を作る四角形が対角線上に見えました。

著作権上問題のある目次や索引データがそのまま DL できないように画像で表現したテストサイトを以下の URL に構築しました。

計算が膨大であり更新やサーバー維持のためのコストも相当額であるために長期維持は不可能です。

BOB (publicly open: <http://www.ebob.jp/>)

アカウントパスワードは作業の保存の為ですので自由に設定できます。

研究機関内に行った公開作業およびチューニングは以下のようです。

2009/08/28	Active	Corei7	証拠公開サーバー http://www.ebob.jp
2008/01/29	Active	BasicCore1	I E 7 , Firefox2に対応
2007/11/06	Active	Basic2	IE7, Firefox2に対応
2007/06/12	Active	Basic2	スペース更新、オブジェクト表示高速化
2007/01/23	Active	Basic2	複数教科書詳細表示機
2006/12/26	Active	Basic2	Test for Patent Data(Term他のマッピング情報は入っていません)
2006/11/21	Active	Basic2	Short/日本語目次対応
2006/08/29	Active	Basic2	Post対応、Bookshelf他性能向上、PubMed最新データ(2004-2006/7)登録
2006/07/21	Active	Basic3	indexデータから教科書不要セクションを除去(注:教科書ビューで、表示域が残っています、タームDBへのリンクが古いまです)
2006/07/18	Active	Basic3	Bookshelf表示の改善(高速化、バグフィックス)、Object(Gene)最新版: Samplesのchrも(注:UCSC base..NCBIと少し違う)、Term検索改善、その他
2006/07/12	Active	Basic2	オリジナルページデータ確認・アノテーション機能、Bookshelf表示の高速化第1弾
2006/07/05	Active	Basic3	オリジナルページデータ確認・アノテーション機能の追加、辞書の修正(ph1, ii)
2006/06/28	Active	Basic2	全教科書+1冊教科書、小さな修正多数、まだ少し遅い、辞書の修正(diabetes)
2006/06/20	Active	Basic2	ストップワードをSVDの前に実行

2006/06/14	Active	Basic3	Book Shelf 機能その2 一括個別教科書ソートが遅い
2006/05/31	Active	Basic2	Book Shelf 機能開発開始版(まだまだですが)、ヘアカール機能実装
2006/05/30	Active	Basic2	開発者専用: BOB スペース調査のためのテストバージョン
2006/05/18	Active	Basic3	オブジェクトベクトルの単位ベクトル化対応(Ajax対応第2版)、PubMed => GeneID更新
2006/05/09	Active	Basic2	Ajax対応初版
2006/04/11	Active	Basic3	新bio_stop_list.txt
2006/04/6	Active	Basic2	bio_stop_list.txtから、不良ターム削除
2006/03/28	Active	Basic2	SVD: page x termマトリックス重み付け有り版を用意→以後は有りを採用 当面、開発フェーズにおけるPubMedサポート年範囲を狭める(4年分→1年分)
2006/03/22	Deleted	Basic2	BaseForm0_OriginalTerm.seeのシノニムへの混入により辞書が崩壊
2006/03/15	Active	JST1	
2006/03/08	Active	Basic1	

<国内外での成果の位置づけ>

比較すべきものは 文献情報を利用して遺伝子に関係化する様々のサービスであろう。 PubGene に代表される遺伝子ネットワークの抽出が代表である。一見すると遺伝子を構造にする点で類似しているが動作原理はまったく違う。教科書をフレームワークにして文書を構造化し構造の部分を教科書の表題で表現するような試みは幸いまだない。一方で検索結果をアドホックにクラスター化してそれぞれを名づけるやりかたは単純な検索結果の序列化や固定的なサイテーション情報などでの関係化の次の方向となりつつある。この点では教科書を選んで好きな視点から無構造の文書をグループ化する本法は類似しています。

宣言的オントロジー: 専門家が一人もしくは小集団で自らの理解に基づいてドメイン知識を概念関係などで表現する手法は過去10年の間に生命科学系では広く受け入れられてきた。特にテキストマイニングやセマンティックウェブなどの大量の文書を対象に、検索したり操作したりする分野との連携の容易さから他の分野よりも盛んに言及、利用される道具として定着しています。

その記述形式は標準化され、統一され、それを踏まえたフリーの編集ソフトは多数作成されて公共の利用に供されている。また多くのチームが作成したオントロジーは遺伝子オントロジーの創始者である Ashburner らが主催する OpenBioMedicalOntologie などのサイトに集積され検索利用が容易になっている。

動的クラスタリング: あらかじめ外部から基準を与えることなく検索で得られた文書群を相互の類似性で分類する方法では得られた文書クラスターの名づけが課題となっている。

<達成できなかったこと、予想外の困難、その理由>

開始時に比べるとマイクロアレイの凋落に加えてGOの流行がはなはだしく、GO以外で遺伝子機能の表現を行う研究はほとんど見られない。しかしながらGOの表現にもさまざまな課題があり、遺伝子機能の自動表現は完成すれば十分に役立つものと考えられる。また遺伝子に限らずあらゆる文書をベクトル化して教科書ページとの関連性パターンで表現する方法は文書検索や、情報取得へと一般化できる。

検索後の数百のヒットの中身の表現はV I V I S I M Oに代表されるクラスタリングとグループの名づけの巧みさの問題とするのが一般的である。

BOBではクラスタリング後のグループの名づけは絶対知識座標である教科書の項目見出しで行うわけであるので科学分野に限定すればV I V I S I M O的手法をしのぐと期待される。

国内の情報学系の一般解説誌にはそのユニークさから取り上げられることが多く、JST 研究開発戦略センターにおいても分野俯瞰ツール作成法として採用され、国内の教科書出版社の試験的な利用が計画されています。情報整理の方法として教科書をフレームとして利用する方法は国際特許が成立しました。しかし実データにはすべて著作権があるために利用サイトは公開できず論文投稿にも至っていない

文書のグループ化やその内容を好みのアスペクトから眺められる俯瞰を与える方法として、そしてマイクロアレイその他で作られる遺伝子のグループの共通性を探すなどの目的で利用可能な方法であるが以下の点が問題である。

1. 教科書の目次および索引の2次的利用にあたりすべての出版社とライセンス契約が必要であることが明らかとなった。
サイトの公共利用によって第三者に部分的にも当該内容の配布をすることになるためにライセンス契約を行ったとしても公開することが出来ないことが最大の問題である。
2. 教科書で収集した用語自身や多くの見出し項目にはその選択や序列以外に著作権は認められないので一般に再利用可能な形で公開することができる。
3. 遺伝子記述やPubMedの内容が更新される度に数万の用語による索引付けと更新された用語 x 文書ベクトル群の用語空間へのプロジェクションのための計算を行う必要があり計算資源の維持が難しい

<今後の課題、展望>

知的構造体を任意の情報整理に利用するアイデアの独創性は特許成立 (P2005-259088A) によって認められた。適当な条件でのライセンス下の利用で実用化公開は可能である。

またグレイ解剖学などの著作権切れの古典的教科書は自由に使用することも可能である。

<研究期間の全成果公表リスト>

1) 論文

1. 0708101552
Hoshino H, Uchida T, Otsuki T, Kawamoto S, Okubo K, Takeichi M, Chisaka O.: Cornichon-like Protein Facilitates Secretion of HB-EGF and Regulates Proper Development of Cranial Nerves Molecular Biology of the Cell Apr Vol.18 D1143-1152(2007).
2. 0603021854
Ogasawara O, Otsuji M, Watanabe K, Iizuka T, Tamura T, Hishiki T, Kawamoto S, Okubo K.: BodyMap-Xs: anatomical breakdown of 17 million animal ESTs for cross-species comparison of gene expression., Nucleic Acids Res.,34(Database issue):D628-31 (2006)
3. 0603021855
Okubo K, Sugawara H, Gojobori T, Tateno Y., DDBJ in preparation for overview of research activities behind data submissions.,Nucleic Acids Res., 34(Database issue):D6-9. (2006)
4. 0603021856
Itoh K, Kawasaki S, Kawamoto S, Seishima M, Chiba H,

- Michibata H, Wakimoto K, Imai Y, Minesaki Y, Otsuji M, Okubo K., Identification of differentially expressed genes in psoriasis using expression profiling approaches. Exp Dermatol., 14(9):667-74, (2005)
5. 0603021857
Kaimori JY, Takenaka M, Okubo K.
Quantification of gene expression in mouse and human renal proximal tubules. Methods Mol Biol., 293,209-19.(2005)
 6. 0603021859
Tanino M, -----, Auffray C, Hide W, Okubo K.
The Human Anatomic Gene Expression Library (H-ANGEL) Nucleic Acids Res. 2005 Jan 1;33(Database issue):D567-72.
 7. 0603031716
Tateno Y, Saitou N, Okubo K, Sugawara H, Gojobori T. DDBJ in collaboration with mass-sequencing teams on annotation.Nucleic Acids Res. 2005 Jan 1;33(Database issue):D25-8.
 8. 0603031655
Michibata H, Chiba H, Wakimoto K, Seishima M, Kawasaki S, Okubo K, Mitsui H, Torii H, Imai Y.
Identification and characterization of a novel component of the cornified envelope, cornifelin. BBRC. 2004 Jun 11;318(4):803-13.
 9. 0603031700
Imanishi T,---- Okubo K, Wagner L, Wiemann S, Strausberg RL, Isogai T, Auffray C, Nomura N, Gojobori T, Sugano S. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol. 2004 Jun;2(6):e162. Epub 2004 Apr 20.
 10. 0603031708
Chiba H, Michibata H, Wakimoto K, Seishima M, Kawasaki S, Okubo K, Mitsui H, Torii H, Imai Y.
Cloning of a gene for a novel epithelium-specific cytosolic phospholipase A2, cPLA2delta, induced in psoriatic skin. J Biol Chem. 2004 Mar 26;279(13):12890-7. Epub 2004 Jan 2)
- ##### 2) 図書
1. 0606201847
Okubo K,
「文献知識の計算機による利用がなぜ必要か」
「蛋白質核酸酵素」編集部 vol.50 共立出版株式会社
 - 3) データベース
1. 0802251729
Okubo K,
BOB(試験公開サーバー)
<http://222.151.240.6/project/bob/080208-/simple.cgi>
 2. 0606231539
Tamura, T., Iizuka, T, Watanabe, K., and Okubo, K., BodyMap-Xs -- CGED Source Classification Database <http://cged.genes.nig.ac.jp/scd/>
 3. 0603031800
Tanino, M., Debily, MA., Tamura, T., Hishiki, T., Ogasawara, O., Murakawa, K., Kawamoto, S., Itoh, K., Watanabe, H-Angel <http://www.jbirc.aist.go.jp/hinv/h-angel/wge.top>
 4. 0603031752
Ogasawara, O., Otsuji, M., Watanabe, K, Iizuka, T., Tamura, T, Hishiki, T., Kawamoto, S., Okubo, K. BodyMap-Xs <http://bodymap.jp>