

# 「ライフサイエンス分野の統合データベース整備事業」18年度成果の概要

### I. DB整備戦略のための研究俯瞰提供サイト (分野を知りデータを知る)

- 1. 学会要旨統合サイト
 
- 2. データバンク統合目次
  - DNAバンク目次
 
  - 遺伝子発現バンク目次
 
- 3. 戦略立案資料(報告)
 

### II. 統合技術開発・提供 (統合に足りない技術をつくる)

- 1. 辞書シソーラス
  - 遺伝子名称シソーラス
  - 生物学名日本語一般名対応辞書
  - 施設名称辞書
- 2. オントロジー、分類機
  - 動植物解剖学自動分類タガー
  - 都市名国名自動検出タガー
  - 解剖学3Dポリゴンマン辞書
  - 3D-アナトモグラフィーAPI
  - ベンチメソッドオントロジー
- 3. キュレーター支援技術開発
 
- 4. 癌研究知識表現技術開発
 
- 5. 多型知識表現技術開発
 

### III. ポータルサービス (使える情報サイトに案内する)

- 1. DBポータル
 
- 2. WEBリソースポータル
 
- 3. ポータルサイト連携のための調査
 

### IV. 分子データベース整理統合 (1次データの統合整理)

- 1. ヒト遺伝子発現統合
 
- 2. 植物ESTボディーマップ
 

### V. 人材の育成 (長期的な取り組みにむけて)

教育資料編纂・閲覧システム (調査,教科書作成)

- 1. キュレーター教育
- 2. アノテーター教育
- 3. DBマネジャー教育

DBマネージャー教育用の教科書




# 「ライフサイエンス分野の統合データベース整備事業」18年度成果項目

## 1. データベース統合戦略立案および評価

項目
ゲノム注釈とデータベース間の連携における課題 国内外DBの俯瞰と質的量的比較
ライフサイエンス分野の研究の俯瞰調査
検索アルゴリズムを含めた知識情報技術の動向調査
臨床情報や医療統計の現状調査

## 2. データベース統合化基盤技術開発

基盤知識表現技術開発	1) 辞書シソーラス 1. 遺伝子名称シソーラス 2. 生物学名日本語一般名対応辞書 3. 施設名称辞書
	2) オントロジー、分類機 1. 動植物解剖学自動分類タガー 2. 都市名国名自動検出タガー 3. 解剖学3Dポリゴンマン辞書 4. 3DアナトモグラフィAPI 5. メソッドオントロジーとの連携システムの開発
	3) 分子データベース整理統合 1. ヒト遺伝子発現統合
癌研究知識表現技術開発	
多型知識表現技術開発	
キュレーター支援技術開発	

## 3. 人材の育成

キュレーターの育成
アナテーターの育成
DBマネージャーの育成

## 4. データベース(DB)等ポータル構築

データベースポータル構築
WEBリソースポータル構築
ポータルサイト連携
文献情報との連携調査

# 「ゲノム注釈とデータベース間の連携における課題」

- 目的

- モデル植物のゲノム統合型データベース群の調査を行い、データベース間の連携状況の現状を知るとともに、あるべき将来像を示す。

- 実施事項

1. 「イネ」ならびに「シロイヌナズナ」のデータベースについてコンテンツや相互の連携状況をサイトから調査するとともに、聴き取りとアンケート調査から問題点を抽出。
2. 調査結果を踏まえ、将来的にライフサイエンスの統合型データベースのあるべき姿を列挙し、必要な機能を実証的に検討するためのプロトタイピングを行う。

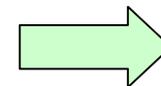
- 成果

- 全ゲノム塩基配列が決定されている2植物に関するデータベース群に関するコンテンツ調査と聴き取りならびにアンケート調査を行い、問題点を抽出するとともに、将来の統合型DBへの提言を含む調査報告をとりまとめた

*Oriza sativa* (イネ) 統合型 DB 群



調査



統合DB間の連携状況に関する調査報告書

*Arabidopsis thaliana* (シロイヌナズナ) 統合型 DB 群





# 「ライフサイエンス分野の研究の俯瞰調査」

## ライフサイエンス分野の俯瞰のための連携システム開発

### 目的

1. 現在、プラナリアの脳の遺伝子情報を発信するOh!脳サイト(以下URL)を、脳の進化-発生-働きに関する内容の充実を図る。

#### 【具体的内容】

- 脳のみならずと遺伝子に関する情報の充実
  - 脳の働き、なりたち、進化の過程に関する情報の充実
  - プラナリアを事例に、細胞と遺伝子情報のデータベース化を図り、関係性を視覚的に把握できるシステムを構築
2. 脳の働きについて、プラナリアの細胞と遺伝子の中核に、1,000個の細胞と約40個の遺伝子情報のマトリクスのデータベースを構築するとともに、細胞単位で発現を視覚的に把握することができる(俯瞰できる)ビューワー機能を構築する。
3. 当システム開発に伴い、既存サイトと今回拡張する機能も含め、サイトマップを作成する。

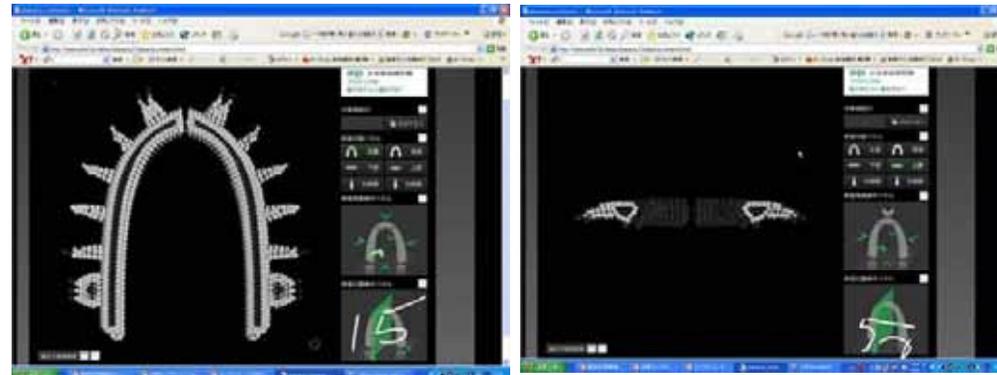
### 成果

- プラナリアの脳細胞に関するデータベースの構築
  - プラナリアの脳細胞の情報と遺伝子情報の相関性を表現。
- 3次元表現による遺伝子発現情報の作成
  - WEBブラウザを通じて3次元的表現により、脳細胞と遺伝子情報の相関性を表現。
  - また、インタラクティブなユーザーインターフェイスにより、細部にわたる情報を把握できる仕組みを実現。

#### ▼細胞情報と遺伝子情報のBDシステム



#### ▼3次元表現による遺伝子発現情報を表現

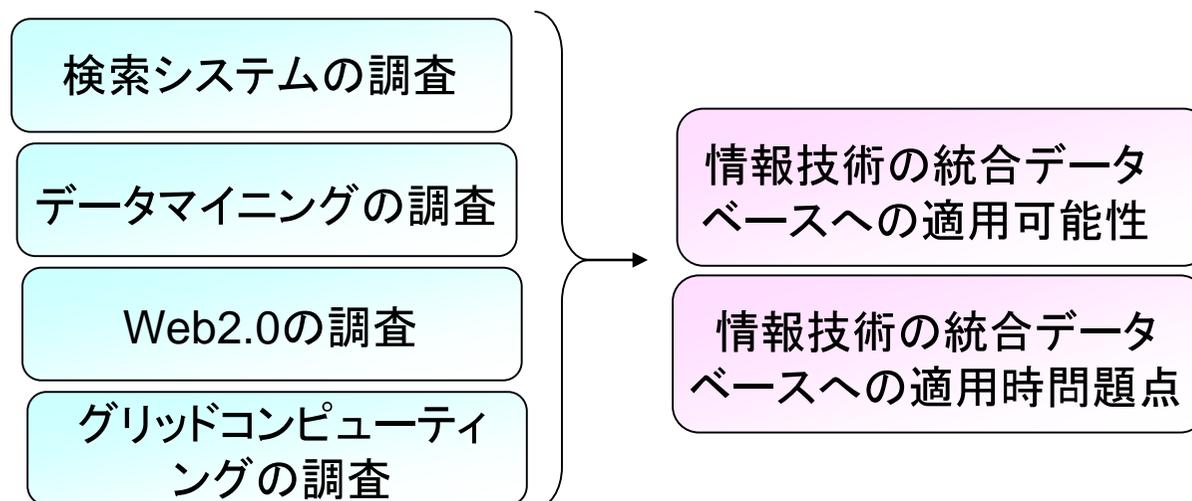


↑背面焦点(レイヤー15)

↑上部焦点(レイヤー52)

# 「検索アルゴリズムを含めた知識情報技術の動向調査」 統合処理技術の動向調査

- 目的
  - 急速に展開する情報処理技術の動向、中でも知識情報処理の動向を文献資料、ネットワーク上での情報検索、専門家からの聞き取り調査、実地見聞などを駆使し分析を加え、その成果を報告書としてまとめる。
- 実施事項
  1. 検索システム、データマイニング、Web2.0、グリッドコンピューティングの4項目について調査を行った。
  2. 上記の4項目の調査結果を基に、上記情報技術の統合データベースへの適用可能性について議論し、また、その問題点を考察した。
- 成果
  1. 検索システム、データマイニング、Web2.0、グリッドコンピューティングの4項目についての調査報告
  2. 上記情報技術の統合データベースへの適用可能性とその問題点についての報告



# 「検索アルゴリズムを含めた知識情報技術の動向調査」

---

## 統合データベースに関する計算機資源の調査

- 目的: 統合データベースに必要とされる計算機資源(CPU数、disk容量など)を予測し、効率的な計算機環境を整える。
- 実施事項: 文献、インターネット、及び、公開されているプログラムの計算速度、及びメモリー使用量、disk使用量を計算し、また、今後、新たに利用されられると思われる技術については、他分野での同技術を用いているプログラム速度を参考として見積もりを行った。
- 成果: テキストマイニング、Webクローラなどの8項目において重点的に調査した。

# 「臨床情報や医療統計の現状調査」

---

## 遺伝統計学分野における解析技術の基礎調査

- **目的**

- 個体の遺伝子多型と表現型との関連を解析するための有効な手法である遺伝統計学的解析手法の動向を調査する。

- **実施事項**

- 文献・インターネットによる調査を行ない、遺伝統計学分野の解析手法・解析アルゴリズムの特徴を整理して報告書を作成した。

- **成果**

1. 解析手法・解析アルゴリズムの調査

- 遺伝統計学分野で用いられる連鎖解析, 連鎖不平衡解析(ハプロタイプ解析), QTL解析等の解析手法と, それぞれの解析手法における代表的なアルゴリズムの調査を行ない, 特徴および長所・短所を評価した。

2. ソフトウェアの調査

- 連鎖解析, 連鎖不平衡解析(ハプロタイプ解析), QTL解析を行なうためのソフトウェアに関して, 実装されているアルゴリズム, 動作環境, 入出力などを調査し, 特徴および長所・短所を評価した。

# 「臨床情報や医療統計の現状調査」

## 医療情報、疾患・健康情報の調査

### 目的

- 臨床データの活用を視野に、本プロジェクトにおける医療情報への取組みに関する方向性を探る。
- 医療統計に関わるデータ産生の現状と解析手法の調査を行い、本プロジェクトにおける取組み指針を得る

### 実施事項

1. 電子カルテ等における臨床データの規格化とデータ共有化の現状と動向の調査
2. 医学統制用語等の現状調査
3. 臨床データ活用上の課題の論点整理
4. 我が国における代表的なコホート研究の統計分析手法の調査
5. 統計分析手法の課題と方向性の整理

### 成果

1. 地域臨床データ共有において、現在できていることと課題意識を整理
  - HL7等の標準規格の役割、データ抽出のためのカルテデータのモデル化の意味、インセンティブ等
2. 生活習慣病を中心としたコホート研究の幅広い事例調査、統計分析手法、成果等の整理
  - 生活習慣病を中心としたコホート研究の事例を分類・整理
  - 統計分析手法を可能な範囲で論文等から抽出
  - データオリジンの課題の抽出
  - 地域、全国、世界でのデータ連携の動向の整理 等

# 「基盤知識表現技術開発」

## メソッドオントロジーとの連携システムの開発

- **目的**
  - データを産生し、データを解析するメソッド名についてウェット系、ドライ系の両方から整理を進め、データ統合技術内での意味づけ関連づけを正確にするとともに、利用者にわかりやすいメソッド検索システムや辞書の提供を図る。
- **実施事項**
  1. ウェット及びドライ両方の実験手法名を収集し辞書構築を行った。
  2. バイオインフォマティクス解析メソッドオントロジーを利用した検索プログラムを作成した。
  3. ウェット系の実験手法を整理しデータや論文との連携を図るためのオントロジーについて検討した。
- **成果**
  1. ウェット系実験手法名の収集と辞書及びオントロジー作成。
    - ウェット系の実験手法名を英文、和文双方の書籍、文献から幅広く語彙を収集し辞書の素材構築を行った。
  2. ドライ系メソッドオントロジーを利用した検索プログラムを作成しWebリソースポータルに実装した。
    - 様々な名称がつけられているバイオインフォのメソッドを利用者の観点から分類し検索可能にした。

### ウェット系の実験手法名辞書の構築

見出し	同義/類義語	英訳	(英) 同義/類義語	解説
トリチウムチミン取り込み実験	[3H]チミン取り込み実験	[3H]thymidine incorporation experiment		DNAの合成速度や合成量を調べたり、3Hで放射線標識したDNAを得る目的で、生物を[3H]チミンで標識する実験。
エンハンサートラップ法	エンハンサートラップ、エンハンサートラップ実験	enhancer trap	enhancer-trap experiment	DNA組換え技術を利用し、活性の指標となる遺伝子とプロモーターの組み合わせがエンハンサーの近傍に位置したときに遺伝子活性が上昇することを利用し、エンハンサー領域を同定、単離する実験法。
紫外線照射	UV照射、UV照射実験、紫外線照射実験	ultraviolet irradiation	Ultraviolet radiation	主として紫外線による損傷や影響を調べる目的で、物体や生物を紫外線にあてること。
トレーサー	トレーサー実験	tracer, orbital	tracer experiment	一般的には、ある事象の変化を追跡観察するための鍵となるもの。生物実験では、分子中の特定の原子を放射線や蛍光で標識した特定の部位。
同位体トレーサー法	トレーサー実験	isotope tracer technique	tracer experiment	同位体を利用した実験技術で、特定の部位を同位体で標識した化合物を生物に投与し、同位体の存在を目印に、特定物質の代謝や合成経路、存在部位を調べる技術。

### Webリソースポータルサイトの検索システム

# 「基盤知識表現技術開発」辞書シソーラス

- 目的
  - 辞書シソーラスとして、1. 遺伝子名称シソーラス、2. 生物学名日本語一般名対応辞書、及び3. 施設名称辞書の作成を行う。
- 実施事項
  - 1. 9種類の生物の遺伝子名シソーラスの構築を行った。
  - 2. 生物学名日本語一般名対応辞書(登録データ数合計14028種)の構築を行った。
  - 3. 施設名称研究室名称の表記ゆれに対応するための、施設名称辞書を開発した。
- 成果
  - 1. 遺伝子名称シソーラス
    - さまざまなデータベースや論文で利用されている表記を専門的キュレータが編集して、遺伝子が持つ多様な表記のあいだの関係を明かにした遺伝子名シソーラス Ver1.0を開発した。Ver1.0 はヒトをはじめ9種類の生物をカバーしている。
  - 2. 生物学名日本語一般名対応辞書
    - 塩基配列登録の多いものから順に学名:日本語一般名(標準和名が存在しない場合、その生物を説明する一般的な名称)を対応させた日本語生物種辞書(登録データ数合計14028種)を開発した。主要な73種類についてはさらに認識を容易にするアイコンを付与した。
  - 3. 施設名称辞書
    - 学会要旨DB化における施設名称研究室名称の表記ゆれ(大阪大、阪大、大阪大学大学院など)に対応するために、施設名称辞書を開発した。これにより、同一の研究室の同一テーマを一塊として把握し国内の研究動向の把握を容易にすることが可能になった。

# 「基盤知識表現技術開発」オントロジー、分類機

## 1. 動植物解剖学自動分類タガー 2. 都市名国名自動検出タガー

### 目的

- 解剖学用語の名称を自動的にカテゴリー分類するプログラム(動植物解剖学自動分類タガー)、及び論文やDBレコードの国別分類のための辞書(都市名国名自動検出タガー)の開発を行う。

### 実施事項

#### 1. 動植物解剖学自動分類タガー

解剖学用語、すなわち臓器・器官・部位の名称を、専門家が作成したルール化(振興調整費DB統合のための調査研究にて作成)を用いて自動的にカテゴリー分類するプログラム(動植物解剖学自動分類タガー)を開発した

#### 2. 都市名国名自動検出タガー

論文やDBレコードの国別分類のための辞書(都市名国名自動検出タガー)を開発した。

### 成果

#### 1. 動植物解剖学自動分類タガー

解剖学用語、すなわち臓器・器官・部位の名称を、専門家が作成したルール化を用いて自動的にカテゴリー分類するプログラム(動植物解剖学自動分類タガー)を開発した。それぞれ1万を越すESTライブラリ/material表現を手作業で同定分類し作成したルールに基づいて構築した。

#### 2. 都市名国名自動検出タガー

論文やDBレコードの国別分類のための辞書(都市名国名自動検出タガー)を開発した。これにより、国名称の未記載や国名称にみられる表記ゆれを吸収可能になった。

# 「基盤知識表現技術開発」オントロジー、分類機

## 3.解剖学3Dポリゴンマン辞書 4.3DアナトモグラフィーAPI

### 目的

- 解剖学用語、すなわち臓器・器官・部位の名称やそれらに属する概念をモデル人間中の3次元座標(3Dポリゴンマン)で定義した辞書(解剖学3Dポリゴンマン辞書)を開発し、解剖学用語が付与された手持ちのデータを解剖学3Dポリゴンマン辞書にマッピングして俯瞰可能なアナトモグラフィー(新造語)を開発する。

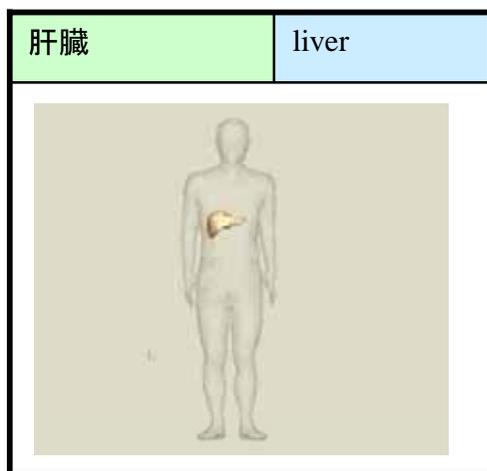
### 実施事項

1. 解剖学3Dポリゴンマン辞書 3D編集ソフト(FreeForm Modelling)を用いて、数値人体モデルデータベース(独立行政法人情報通信研究機構が開発)を基盤に、人体解剖模型・図譜等を参考に詳細化を行った。
2. 3DアナトモグラフィーAPI 解剖学3Dポリゴンマン辞書へのユーザデータマッピング機能を実装した。

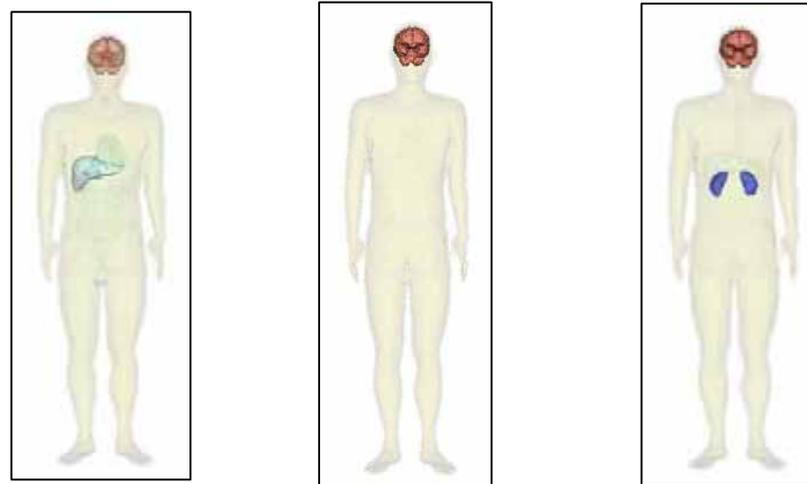
### 成果

1. 解剖学3Dポリゴンマン辞書の作成
  - 解剖学用語を3次元座標で定義することにより、概念を曖昧性なく表現。従来のツリー型表現(いわゆる解剖オントロジー)と違い、多角的に破綻しない表現が可能になり、ボクセルデータに比べエディットが飛躍的に容易になった。
2. 3DアナトモグラフィーAPIの作成
  - ユーザデータ(例: 器官別の発現解析データ)の解剖学3Dポリゴンマン辞書へのマッピングによって、解剖学の観点からのデータ俯瞰を実現。
  - 由来の異なるデータの同時マッピングによって、それらのデータの統合的な比較観察を実現。
  - マッピング結果は、静止画像もしくは動画で得ることができ、webサービスとして提供予定。

### 解剖学3Dポリゴンマン辞書



### 3Dアナトモグラフィー



# 「基盤知識表現技術開発」 オントロジー、分類機

## 5. メソッドオントロジーとの連携システムの開発

- **目的**
  - データを産生し、データを解析するメソッド名についてウェット系、ドライ系の両方から整理を進め、データ統合技術内での意味づけ関連づけを正確にするとともに、利用者にわかりやすいメソッド検索システムや辞書の提供を図る。
- **実施事項**
  1. ウェット及びドライ両方の実験手法名を収集し辞書構築を行った。
  2. バイオインフォマティクス解析メソッドオントロジーを利用した検索プログラムを作成した。
  3. ウェット系の実験手法を整理しデータや論文との連携を図るためのオントロジーについて検討した。
- **成果**
  1. ウェット系実験手法名の収集と辞書及びオントロジー作成。
    - ウェット系の実験手法名を英文、和文双方の書籍、文献から幅広く語彙を収集し辞書の素材構築を行った。
  2. ドライ系メソッドオントロジーを利用した検索プログラムを作成しWebリソースポータルに実装した。
    - 様々な名称がつけられているバイオインフォのメソッドを利用者の観点から分類し検索可能にした。

### ウェット系の実験手法名辞書の構築

見出し	同義/類義語	英訳	(英) 同義/類義語	解説
トリチウムチミン取り込み実験	[3H]チミン取り込み実験	[3H]thymidine incorporation experiment		DNAの合成速度や合成量を調べたり、3Hで放射線標識したDNAを得る目的で、生物を[3H]チミンで標識する実験。
エンハンサートラップ法	エンハンサートラップ、エンハンサートラップ実験	enhancer trap	enhancer-trap experiment	DNA組換え技術を利用し、活性の指標となる遺伝子とプロモーターの組み合わせがエンハンサーの近所に位置したときに遺伝子活性が上昇することを利用し、エンハンサー領域を同定、単離する実験法。
紫外線照射	UV照射、UV照射実験、紫外線照射実験	ultraviolet irradiation	Ultraviolet radiation	主として紫外線による損傷や影響を調べる目的で、物体や生物を紫外線にあてること。
トレーサー	トレーサー実験	tracer, orbital	tracer experiment	一般的には、ある事象の変化を追跡観察するための鍵となるもの。生物実験では、分子中の特定の原子や基を放射線や蛍光で標識した特定の部位。
同位体トレーサー法	トレーサー実験	isotope tracer technique	tracer experiment	同位体を利用した実験技術で、特定の部位を同位体で標識した化合物を生物に投与し、同位体の存在を目印に、特定物質の代謝や合成経路、存在部位を調べる技術。

### Webリソースポータルサイトの検索システム

# 「基盤知識表現技術開発」 分子データベース整理統合

## 1. ヒト遺伝子発現統合

- 目的**
  - ヒト遺伝子の解剖学的な発現パターンデータの統合サイトを構築する。
- 実施事項**
  - 発現パターンは、測定法毎に異なる場合があることが知られており、ここではできるだけ客観的な発現パターンの解釈を可能にするために、できるだけ多くの測定方法(5種類)に基づく発現データを表示可能にした。
- 成果**
  - 遺伝子の発現パターンを、3種類の生物学的な分類(似た発現パターン、染色体上での隣接、同じ遺伝子ファミリー)に応じて表示することを可能にした。組織情報は、開発した動物解剖学自動分類タグで処理し整理分類しており、これにより異なるプロジェクトから得られた発現データ間の比較が可能になった。また、開発した遺伝子名称シソーラスと3Dアナトモグラフィーを、検索部分と表示部分にそれぞれ用いた。



# 「癌研究知識表現技術開発」

- 目的
  - 実験的なデータベース作成等を通じて癌の分子データと臨床情報の統合、表現を想定ユーザーにわかりやすい形で実現する。
- 実施事項
  1. 癌遺伝子発現臨床情報データベースの機能拡張。
  2. 臨床情報の整理(成人病センター乳腺内分泌外科症例)。
- 成果
  - CGED (Cancer Gene Expression Database)の機能拡張。
  - 従来の機能に加えて臨床情報から遺伝子を検索する機能を追加。

追加機能画面→

(例)

転移のある癌とない癌  
で発現の異なる遺伝  
子を検索する

Search Result

Search Condition.

Cancer Type	Clinical Info.	p value	q value
1	--	<0.05	--

q value list

Cancer Type	Clinical Info.	q value
1	Breast (BC) ER +.	0.157
2	Breast (BC) p53 0/1/null	0.211
3	Thyroid (TC) tumor_type FA/CP	0.429
4	Gastric (GC) gender F/M	0.687
5	Gastric (GC) lymph node 0/1/null	0.778

1 Breast (BC) ER +. q value=0.157

5 genes are found. (select all genes) (Remove selection)

	ID	Accession	p value	Gene name	cancer data availability
1	GS1042	NM_001904	0.00125	Catenin (cadherin-associated protein), beta 1, 85kDa	CC,BC,GC,TC,BC_Do,JHCC,EC
2	GS6707	NM_001331	0.01509	catenin (cadherin-associated protein), delta 1	BC,BC_Do
3	GS6707	NM_001331	0.01509	catenin (cadherin-associated protein), delta 1	BC,BC_Do
4	GS6740	NM_001331	0.02345	Catenin (cadherin-associated protein), delta 1	BC,BC_Do
5	GS7006	NM_001904	0.03312	Catenin (cadherin-associated protein), beta 1, 85kDa	BC,BC_Do

Display mosaic plot  
Cancer: Breast

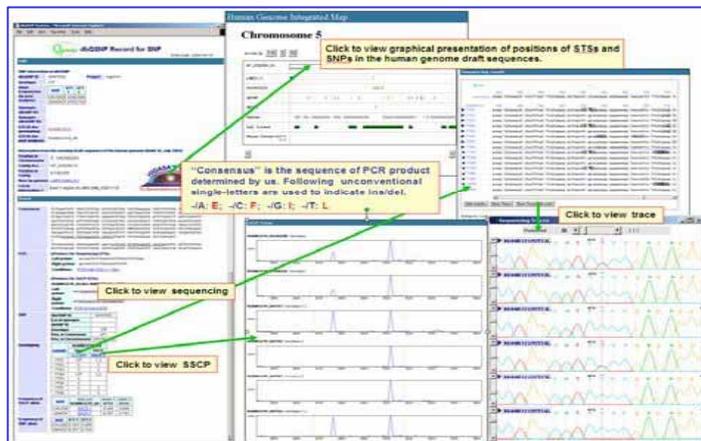
Selected genes  
Display

Similarity search  
The number of genes: 20  
Display

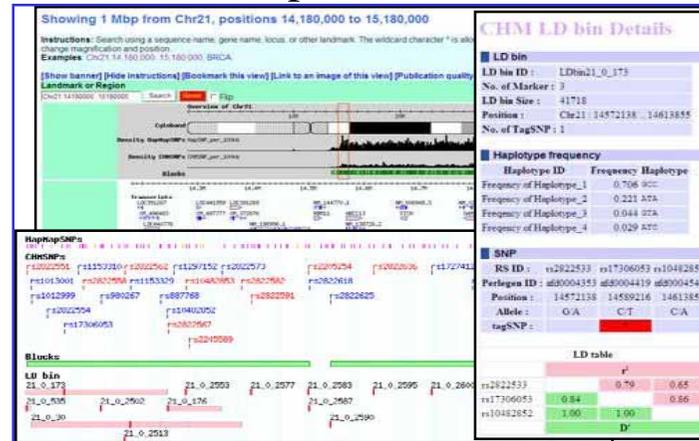
# 「多型知識表現技術開発」

- 目的
  - 日本人ゲノム多型情報を高度化し、医療情報との統合のためのデータポータビリティを図る。
- 実施事項
  1. 疾患の主要な遺伝的要因である遺伝子発現調節領域多型の公開データベース「dbQSNP」を拡充する。
  2. 要因遺伝子探索に必須な全ゲノム確定ハプロタイプ構造の公開データベース「D-HaploDB」を拡充する。
  3. 他のデータベースとの多型データ相互利用促進のため上記2データベースの標準言語(XML)化を行う。
- 成果
  1. 「dbQSNP」の拡充
    - 種々の疾患要因候補遺伝子領域にあるSNP情報を収集し、同データベースに記述されたSNPを  $1.0 \times 10^4$  個超とした。
  2. 「D-HaploDB」の拡充
    - 既存の280k個のSNPに加えて500K個のSNPによる確定ハプロタイプ情報を収集・追加し、同データベースを高度に精細化した。
  3. 「dbQSNP」及び「D-HaploDB」のXML化
    - 上記2データベースのPML (XMLを多型情報記述に特化させた標準言語)版を構築した。

「dbQSNP」の拡充



「D-HaploDB」の拡充



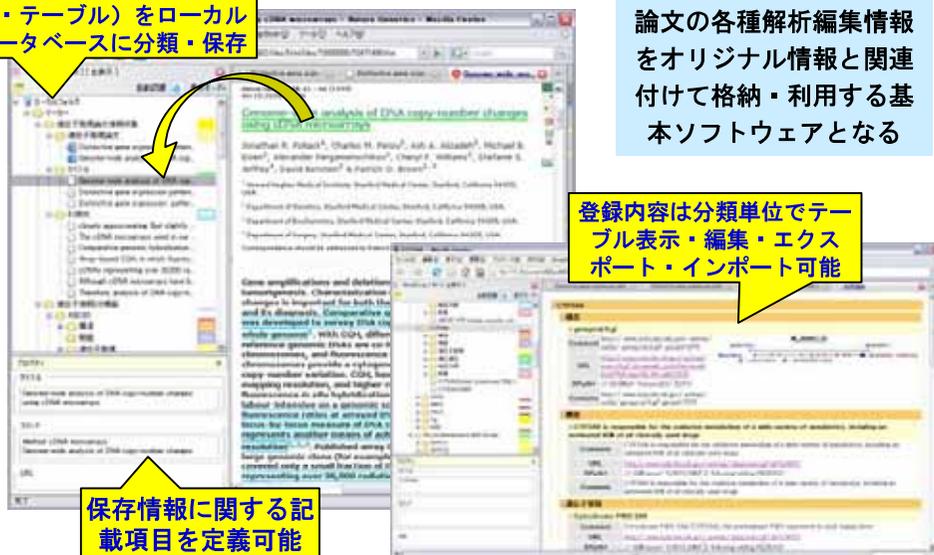
➡ 両DBのPML化

# 「キュレーション支援技術開発」

- 目的
  - 各分野専門家の手による文献やDBからの事実の切り取りと再配置による知識の整理を支援するウェブブラウザのプラグインアプリケーション“ScrapParty”を開発を行う。
- 実施事項
  1. 論文情報解析・編集ソフトウェアのベースシステムの開発を行った。
  2. 論文情報解析・編集用各種解析モジュールの開発を行った。
- 成果
  1. 論文要素情報スクラップソフトウェアの作成
    - 電子化ドキュメント上のオブジェクトをクライアントサイドで分類・蓄積するシステムを完成。
  2. 論文情報解析・編集用各種モジュールを作成
    - 辞書マッピング、ナビゲーション、論文構成認識、引用論文情報収集を行う基本モジュールを作成。
    - 各種モジュールは、ウェブブラウザのプラグインアプリケーション“ScrapParty”の追加モジュールとして開発。
  3. ScrapPartyにより、論文やWebを渡り歩いて重要箇所だけドラッグドロップすることにより、自動的にURLやページ座標情報が記録され、さらに記事を並べた後からのメモ書きが可能になる。収集記事はxml形式で書き出すことができ他人とも共有可能である。

## 論文要素スクラップソフトウェア

論文上の要素（文・イメージ・テーブル）をローカルデータベースに分類・保存



論文の各種解析編集情報をオリジナル情報と関連付けて格納・利用する基本ソフトウェアとなる

登録内容は分類単位でテーブル表示・編集・エクスポート・インポート可能

保存情報に関する記載項目を定義可能

## 論文情報解析・編集用モジュール

論文構造解析・引用論文情報収集モジュール



辞書マッピングモジュール

# 「キュレーターの育成」 キュレーションの実際の調査

- 目的
  - 現在実際にキュレーションが行われている機関での作業内容や流れを調査する。
- 実施事項(図1参照)
  1. 資料(文献等)および、当該施設がWebに公開している情報をもとに、キュレーション作業の実際を調査した。
  2. 調査をもとに、教育教材作成のための内部レポートを作成した。
- 成果
  1. 言葉の定義の明確化(人材育成の教材作成、職業としての確立にむけて)
    - 著名な生物学データベースの領域でのキュレーションとキュレーターの作業の実際が明らかになった。
  2. 著名キュレーション型DBの比較(表1参照)
    - 各々のDBの人員的な規模や、キュレーション、キュレーターがDB全体のなかでどのように関与しているのが明らかになった。



表1 代表的なキュレーション型DBのスタッフ規模

DB	組織	キュレーター	コア・キュレーター	ソフトウェア技術者、プログラマー	参考
RefSeq	NCBI	30	3	38	<a href="http://www.ncbi.nlm.nih.gov/RefSeq/staffcredits.html">http://www.ncbi.nlm.nih.gov/RefSeq/staffcredits.html</a>
	EBI	26	5	36	<a href="http://www.ebi.ac.uk/Information/Staff/viewgallery_seqdb.php?cid=1">http://www.ebi.ac.uk/Information/Staff/viewgallery_seqdb.php?cid=1</a>
UniProtKB	PIR	12	2	3	<a href="http://pir.georgetown.edu/pirwww/about/staff.shtml">http://pir.georgetown.edu/pirwww/about/staff.shtml</a>
	SIB	52 <sup>1)</sup>	5	13	<a href="http://au.expasy.org/people/swissprot.html">http://au.expasy.org/people/swissprot.html</a>
PATHWAY MAP BRITE KO	KEGG	21 <sup>2)</sup>	2 <sup>3)</sup>	4	<a href="http://kanehisa.luiickyo-u.ac.jp/people.html">http://kanehisa.luiickyo-u.ac.jp/people.html</a>

<sup>1)</sup>1.キュレーターという肩書きがなくアナレーターとなっていたが、その方々がいわゆるキュレーション作業もすると判断した  
<sup>2)</sup>2.KEGGは基本的な生物情報だけでなく、化学物質から疾患情報、薬剤情報など非常に広範囲にわたるDBを作成している。21名という数字はそれら多岐にわたるDBの作成にあたるスタッフの総数であり、実際に生体分野を担当しているスタッフは数名であると考えられる  
<sup>3)</sup>3.いわゆるコア・キュレーターという肩書きはない。前出の21名スタッフが自分の担当分野において状況に応じてコア・キュレーターの役割もはたしてると考えられる

# 「アノテーターの育成」 教育用システム作成と教材作成

- **目的**

- ライフサイエンスデータ統合の基盤となる「ゲノムアノテーション」の実務に携わるアノテーターを支援する教育用システムを作成し、同時に実践的なアノテーター教育テキストを作成する。

- **実施事項**

1. ゲノムアノテーションのノウハウを抽出。
2. アノテーターによる手動アノテーションの手順を模倣・再現するプログラムを作成。
3. 「アノテーション教育テキスト」をWiki上に作成した。

- **成果**

1. 大量ゲノムアノテーションの現場で高度な技術を持つ実務者からのノウハウ抽出を行った。
2. DBやツールの利用方法やその結果の解釈を含み、実行したアノテーションの根拠を明示することでアノテーション初学者の理解を助け、学習を促進するシステムを作成
3. アノテーションにかかわる知識や注意点を文書化しゲノム解析型統合DB構築に役立つ教科書とした。



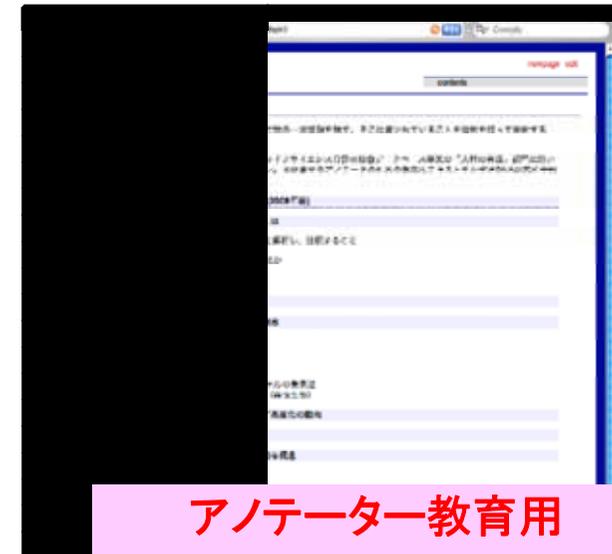
推定結果の提示  
推定根拠の明示



アノテーション学習者



アノテーションの  
ノウハウ抽出と  
プログラム化



アノテーター教育用  
コンテンツの作成

# 「DBマネージャーの育成」 教育用システム作成と教材作成

- 目的
  - ライフサイエンス関係のデータベースを構築・維持管理の実際を行う、DBマネージャーのための教育資料編纂・閲覧システムと実際のコンテンツ構築を行う
- 実施事項
  1. qwikwebを用いて、wikiと電子メールによって知識の蓄積が可能な教育資料編纂・閲覧システムを作成した。
  2. 作成したシステムを用いてDBマネージャー養成のための実習書の「維持管理編」を作成した
- 成果
  1. wikiと電子メールを介した知識蓄積ができる教育資料編纂・閲覧システムの構築
    - Wikiだけではなく、電子メールからの入力も可能とすることでより多くの有識者から情報を集めることが可能に
  2. DBマネージャー養成のための実習書(維持管理編)作成
    - 構築した教育資料編纂・閲覧システムを使って、DBの日々の管理、アップデート、バックアップを中心に作成
    - 個別の関連する事項に関して具体例も併せて解説

## 教育資料編纂・閲覧システム



## DBマネージャー養成のための実習書



# 「データベースポータル構築」

## ・目的

ー利用者を目的にかなったDBやリソースに誘導するポータルサイトを構築する。

## ・実施事項

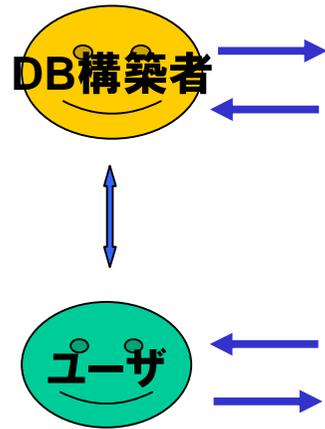
ーDBの俯瞰や戦略立案の目的で作成したDBディレクトリと分類による一覧表示。

ー利用者向け日本語解説を付与し、今後、利用者が書き込み可能なシステムを構築。

## ・成果

ーDB構築者や利用者が書き込み可能なDBポータルを構築。

## 情報の閲覧と書き込みシステム



# 「WEBリソースポータル構築」



**Webリソースポータルサイト**

- ①リソースのキーワード検索
- ②リソースの分野、目的別提示

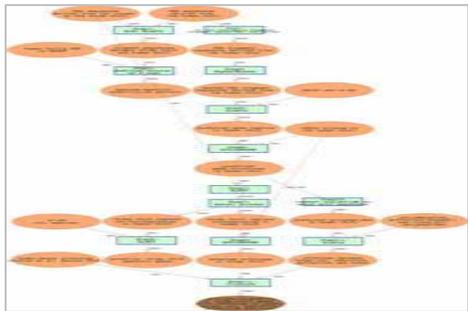


## メソッド

- ・カテゴリ
- ・キーワード
- ・解説(目的、特徴など)
- ・提供サイト
- ・参考文献
- ・引用文献

## ワークフロー

- ・メソッド構成
- ・カテゴリ
- ・キーワード
- ・解説(目的、構成など)
- ・提供サイト
- ・引用論文



リソース詳細表示

リソース名	tRNAscan-SE
別称	
解説	ゲノム配列からtRNAを予測するツール。tRNAの候補を探し、tRNAの共分散モデルを用いて判定する。
キーワード	
カテゴリ	塩基配列からtRNAを予測するゲノム配列構造を特定/予測する>非コードRNA領域を特定/予測する 分野別ゲノム解析/アノテーション/非コードRNA領域予測>[ArabidopsisChr1_Flow] tRNA領域を特定/予測する>Step14 分野別ゲノム解析/アノテーション/非コードRNA領域予測>[Aspergillus_Flow1] tRNA領域を特定/予測する>Step1
提供サイト	UCSC Pasteur Institute
参考文献	tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.
引用文献	
入出力データ	

**研究事例論文**

- ・メソッドのハイライト,リンク
- ・ワークフローへのリンク



# 「ポータルサイト連携のための調査」

---

1. 「Jabion」日本語バイオポータルサイト(<http://www.bioportal.jp>)の調査
  2. 「ライフサイエンスの広場」文部科学省ライフサイエンスポータルサイト(<http://www.lifescience-mext.jp/index.html>)の調査
  3. 「UMIN」大学病院医療情報ネットワーク(<http://www.umin.ac.jp>)の調査
    - ・一般公開用ページの掲載内容の概要
  4. 上記1, 2の調査結果を基に有用な連携の仕組みアイデア提示
    - ・掲載内容から相互利用に有効な事項を選び、組み合わせを提示。
    - ・利用者の目的からサイトマップを提示。
- ・有効なアイデアの実装(今後の課題)
- ・ポータル連携と個別ポータルの支援枠組みの提示(今後の課題)

## 「文献情報との連携調査」

---

JSTが運営している文献情報提供事業で提供されている文献情報を活用し、遺伝子情報に遺伝子機能を付加する調査を実施した。100程度の抄録から遺伝子関連の表現型や疾患名を抽出した。

GenLibi (<http://gene.jst.go.jp/>) としてテスト中。(統合DB予算はないが、提案書に記載)

遺伝子と関連していると判断する基準や実際の抽出作業(今後の課題)