

## 1.高度化のための技術調査【概要】

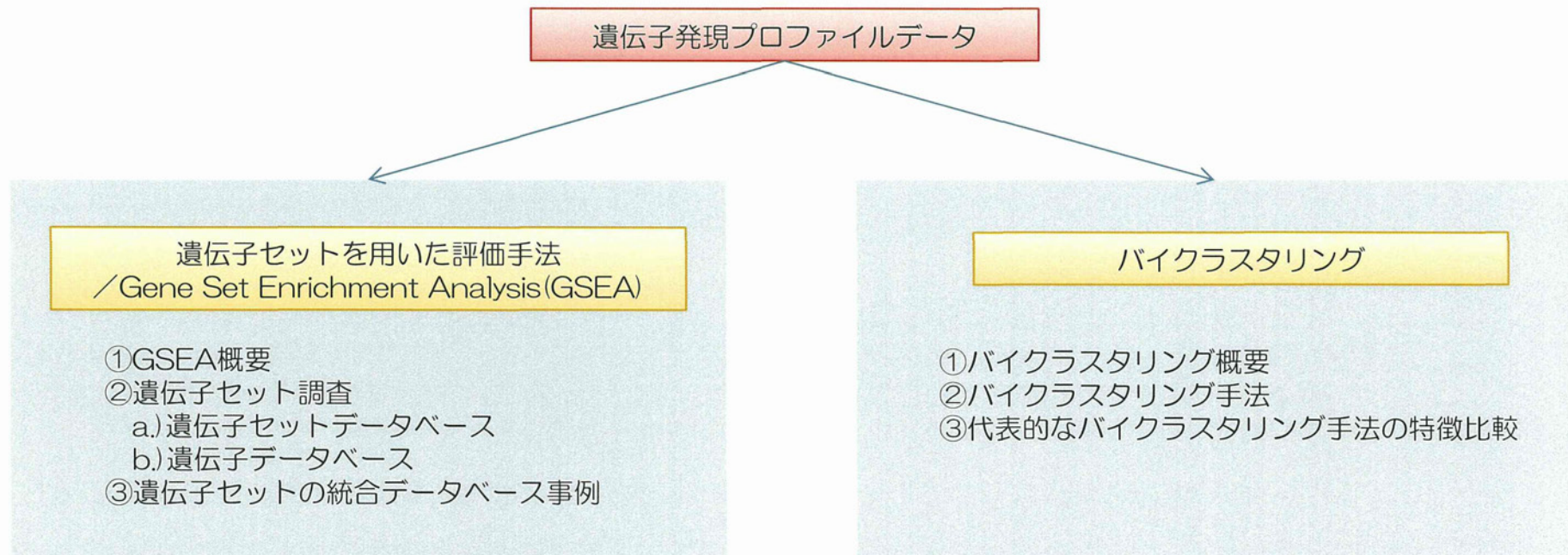
本調査では、本システムの高度化のために有効な技術について検討を行う。

本調査中「解析結果データの比較」における比較解析機能では、ユーザーの入力するクエリとデータベースの比較を行っている。

この比較機能をより高度化する方法として、ある共通した性質を持つ遺伝子セットを用いて、ユーザーが目的とする遺伝子群を評価することが考えられる。また、未知の遺伝子群について、その遺伝子群をクラスタリングすることである共通の性質をもつ遺伝子セットを特定することも有効であると思料される。

本調査では、まず遺伝子セットを用いた評価手法としてGene Set Enrichment Analysis(GSEA)を挙げ、その概要を述べる。次にこの手法で用いることができる遺伝子セットについて述べ、最後に遺伝子セットの概念をデータベースの高度化へ用いた事例を紹介する。

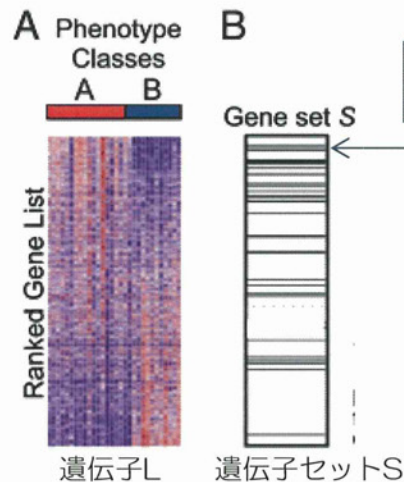
続いて、未知の遺伝子群をクラスタリングする手法としてバイクラスタリングを挙げ、その概要を述べる。バイクラスタリングには多くの提案手法が存在するが、本調査ではそのうち代表的な手法についてそれぞれの概要を述べた後、それらの手法を比較するいくつかの論文を引用して評価を行う。



## 2-1. Gene Set Enrichment Analysis

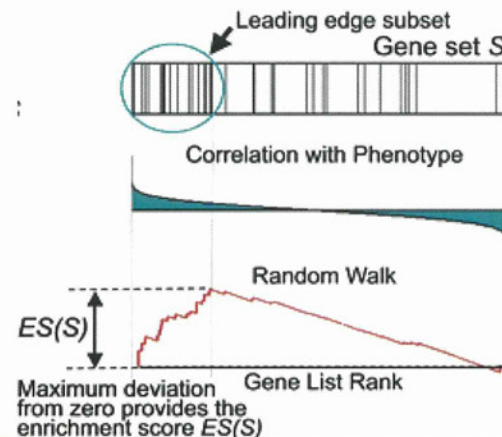
あらかじめ定義されている遺伝子セット中の遺伝子と、実験で得られた遺伝子の表現型との相関を調べる。遺伝子セットとして評価することで、新たに重要な遺伝子を発見することができる。

## 概要



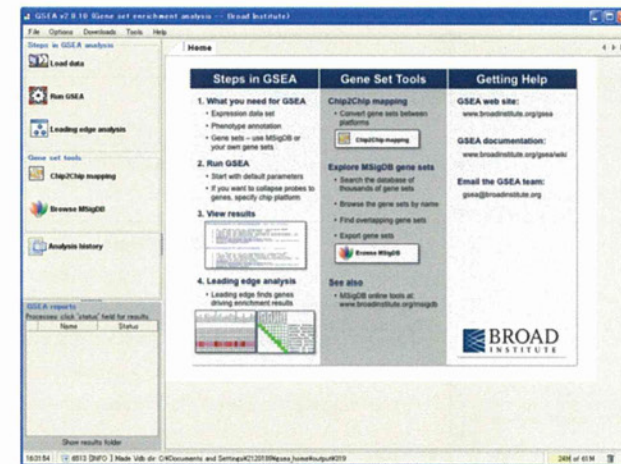
2クラスにラベル付されるサンプルの遺伝子発現プロファイルにおいて、発現値とクラス差異の間の相関に基づいてランク付けされた遺伝子リストL（左図A）と、そのうちある機能をもつ遺伝子セットSに含まれる遺伝子の出現位置（左図B）。

遺伝子リストLのランキング上位から順に、遺伝子セットSに含まれる遺伝子があればスコアを加算、無ければ減算していく。ゼロからの最大偏差を Enrichment Score (ES) とし、ESに到達するまでに含まれる遺伝子をLeading edge サブセットと定義。これは遺伝子セットSの機能の核となる遺伝子群と考えることができる。



## 事例

GSEAは以下のソフトウェアで提供されている。



ダウンロード：

<http://www.broadinstitute.org/gsea/downloads.jsp>

遺伝子セット：

GSEAに利用可能な遺伝子セットは数多く存在する（\*ページ参照）が、代表的なものとしてMSigDBがある。MSigDBは上記GSEAソフトウェアでの利用を目的として作成されたデータベースであり、6つの主要なカテゴリに分類されている（<http://www.broadinstitute.org/gsea/msigdb/index.jsp>）

## 2-2.遺伝子セット調査①

### ①遺伝子セットデータベース

データベース名	対象生物	対象遺伝子	データセット詳細	元となるデータベース/ データベース作成方法	概要	リンク	最新版データ 数
MSigDB	ヒト モデル生物	全遺伝子	c1 染色体 c2 化学的・遺伝的摂動とパスウェイ（化学的・遺伝的摂動における遺伝子セット/カノニカルパスウェイにおける（複数データベースからの）遺伝子セット/ BioCarta遺伝子セット/KEGG遺伝子セット/ Reactome遺伝子セット） c3 cis-regulatoryモチーフ（microRNAターゲット/ 転写因子ターゲット） c4 癌由来マイクロアレイデータ c5 GO c6 発がん性をもつ遺伝子セット	c1 HUGO, Unigeneのデータを解析して取得（両データで不一致の際にはUnigeneを採用）。 c2 BioCarta, CORUM MIPS, KEGG, Pathway Interaction Database, Reactome, SigmaAldrich Signaling Gateway, Signal Transduction KE, SuperArray, L2L, MYC Target Gene Database からキュレート、又は 340以上のPubMed論文から抽出。 c3 Xieらの論文（2005）。 c4 Brentani, Caballeroらの論文による380の癌関連遺伝子のリストをはじめ、Human tissue compendium, Global Cancer Map, NCI-60 cell lines, Novartis carcinoma compendiumの4つの関連する遺伝子セットのデータセットを利用（CGN）。KEGG,GOなど（GM）。 c5 Biological Process Ontology、Cellular Component Ontology、Molecular Function Ontology c6 NCBI GEOのマイクロアレイデータ、または既知の癌遺伝子の摂動に関わる、未発表のプロファイリング実験から作成。少数の発がん性の特徴遺伝子は論文からキュレート。	GSEAソフトウェアにおける利用を目的として作成。オリジナルの研究論文、GOやKEGG、TRANSFAC、L2Lといった特化されたリソースに由来する遺伝子セットを含む。手動のキュレーションと自動化されたコンピュータによる方法によって収集されており、最も多い遺伝子セットを含むデータベースとなっている。MSigDBに含まれる遺伝子セットは、下記6種類のコレクションに分けられる。 c1：染色体の座標でグルーピングした遺伝子セット。 c2：オンラインのパスウェイデータ、PubMedの出版物、専門家の知識からキュレートした化学的・遺伝的摂動、カノニカルパスウェイの遺伝子セット。 c3：ヒト、マウス、ラット、犬のゲノムにおける、保存されたcis-regulatoryモチーフを共有する遺伝子を含むセット。 c4：がん由来のマイクロアレイデータの大規模なコレクションを用いて計算機で作成したリスト。以下2種類を含む。 CGN---380の癌関連遺伝子において発現パターンが類似する遺伝子発現プロファイル群で定義される遺伝子セット。 CM---Segalらによって2004年に定義された遺伝子セット。KEGG,GOなどのような様々なリソースから遺伝子セット（モジュール）を収集。癌に関連するマイクロアレイデータの大規模な一貫を分析することで、様々な癌の状態において重要な変化がみられる456のモジュールを特定した。 c5：Gene Ontology (GO) から作成したリスト。 c6：癌で制御不能となる細胞のパスウェイの特徴を示す遺伝子セット。	■データベースリンク <a href="http://www.broadinstitute.org/gsea/msigdb/index.jsp">http://www.broadinstitute.org/gsea/msigdb/index.jsp</a> ■原文 Liberzon A et al.Molecular signatures database (MSigDB) 3.0. <i>Bioinformatics</i> . 2011 Jun 15;27(12):1739-40	8,513 遺伝子 セット
GeneSigDB	ヒト マウス ラット	癌、幹細胞、免疫細胞、発生、肺の病気における遺伝子発現	論文ごとにグルーピングされた遺伝子セット	PubMedでインデックスを付けられた、1,604の出版論文から自動・手動で遺伝子セットを収集。	出版論文から手動でキュレートした、癌と癌に関連する特徴遺伝子のデータベース。全てのデータはgene set enrichment analysisに用いることができるよう、gmtフォーマットやR/Bioconductorデータファイルを含む多数のフォーマットでダウンロード可能。	■データベースリンク <a href="http://compbio.fci.harvard.edu/genesigdb/">http://compbio.fci.harvard.edu/genesigdb/</a> ■原文 Culhane AC et al.GeneSigDB: a manually curated database and resource for analysis of gene expression signatures. <i>Nucleic Acids Research</i> , 2012, Vol. 40	3,515 特徴 遺伝子
CCancer	ヒト ハツカネズミ ドブネズミ等	遺伝子リストの半分 以上が癌に関連する 研究から抽出された ものと推定。	論文ごとにグルーピングされた遺伝子セット	~100のピアレビューされたジャーナルで出版されている3000の論文から自動的に抽出された、3500以上の遺伝子リストをカバーしている。	CCancerは自動的に収集された、様々な生物学・臨床の実証実験で報告されている遺伝子リストのデータベースである。	■データベースリンク <a href="http://www.bioprofiling.de/CCancer.html">http://www.bioprofiling.de/CCancer.html</a> ■原文 Sabine Dietmann et al.CCancer: a bird's eye view on gene lists reported in cancer-related studies. <i>Nucleic Acids Research</i> , 2010, Vol. 38	3500以上の 遺伝子リスト
HPD	ヒト	ヒトのタンパク質、 遺伝子、RNA転写 産物、酵素、シグナル 伝達、代謝反応、 遺伝子制御	パスウェイリスト	NCI Pathway Interaction Database (PID), Reactome, BioCarta, KEGGからキュレート、もしくはProtein Lounge Web sitesでインデックスされた異なるヒト のパスウェイデータを統合	複数のパスウェイデータのデータベースからの、異なるヒトのデータを統合することで構築。 ヒトのタンパク質、遺伝子、RNA転写産物、酵素、シグナル伝達、代謝反応、遺伝子制御をつなぐ 包括的なオンライン閲覧を提供。	■データベースリンク <a href="http://discern.uits.jiu.edu/8340/HPD/">http://discern.uits.jiu.edu/8340/HPD/</a> ■原文 SR Chowbina et al.JHPD: an online integrated human pathway database enabling systems biology studies. <i>BMC Bioinformatics</i> 2009, 10(Suppl 11)	999のヒトの パスウェイと 59,341の分子

## 2-2.遺伝子セット調査②

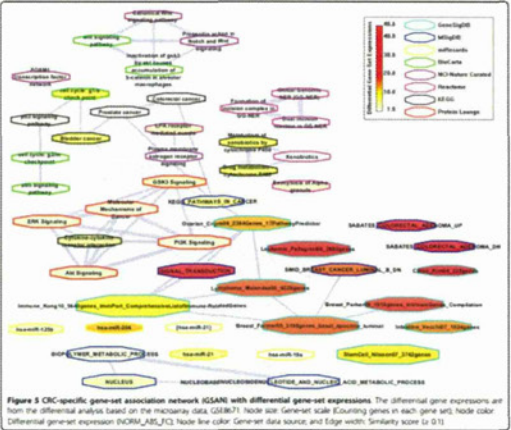
## ②遺伝子データベース

※各データベースは対象遺伝子が特定の機能をもつため、データベース全体を遺伝子セットとして利用することが可能。

データベース名	対象生物	対象遺伝子	データセット詳細	元となるデータベース/ データベース作成方法	概要	リンク	最新版データ 数
Hembase	ヒト	赤芽細胞、赤血球生成に関連する遺伝子	染色体位置、それぞれのエントリーに付帯する独自の名称とアノテーション	発達上の段階である初期のヒト赤芽細胞からのmRNA遺伝子ライブラリに由来する、15,752遺伝子が含まれる。これらのESTは3つの異なる赤芽細胞ライブラリのハイブループリントシーケンシングから得られる。	ヒト赤血球トランスクリプトーム研究のための、統合ブラウザかつ遺伝子ポータルサイト。検索クエリは名前、キーワード、染色体座様で実行され、検索結果は元となる配列データと、その検索の現在の情報にアクセスするためヒトゲノムの主要な3ブラウザにリンクしている。	<ul style="list-style-type: none"> <li>■データベースリンク <a href="http://hembase.niddk.nih.gov/">http://hembase.niddk.nih.gov/</a></li> <li>■原文 Sung-Ho Goh et al.Hembase browser and genome portal for hematology and erythroid biology.Nucleic Acids Research, 2004, Vol. 32</li> </ul>	赤芽細胞のESTsから得られた15,752の遺伝子と、赤血球生成に関連する380遺伝子(2003)
EpoDB	脊椎動物	赤血球	DNA配列、構造的特徴、タンパク質情報、遺伝子発現情報と転写因子結合部位	GenBank, SWISS-PROT, Transfac, TRRD, GERDから取得	脊椎動物の赤血球細胞に関連する遺伝子のデータベース。データ統合とエントリー更新のため、新たなプロトコルを開発。BLASTをベースとしたアルゴリズムを用いて、同じ遺伝子を表示するGenBankのエントリーをグループ化している。この配列類似性プロトコルはEpoDBに入れるべき新たなエントリーの特定にも用いられる。	<ul style="list-style-type: none"> <li>■データベースリンク <a href="http://www.cbil.upenn.edu/EpoDB/">http://www.cbil.upenn.edu/EpoDB/</a></li> </ul>	
LymphTF-DB	マウス	T細胞、B細胞の転写因子とそのターゲット遺伝子	転写因子、ターゲット遺伝子、転写因子結合部位、転写因子相互作用 (※補足: 転写因子を検索すると、関連する遺伝子のリストが出てくるが、それぞれの情報については遺伝子名をクリックするとNCBIのサイトへ飛びようになっている)	マウスの転写因子と様々な発達段階における特定のターゲット遺伝子の活動について手動でキュレート。	B細胞、T細胞の発達段階では、多数の転写因子によって制御された異なる遺伝子発現を必要とする。このデータベースは、個々の転写因子とある発達段階におけるそれらに特有のターゲットの間の相互作用を把握している。このような相互作用を発達上の進行の機能として蓄積することによって、リンパ球の発達を促す制御ネットワークの解明の前進につなげることを目指す。発達段階における転写因子のターゲット遺伝子の相互作用のためのクエリに加え、ターゲット遺伝子の制御配列を転写開始位置と転写結合位置とともにグラフィカルに表示する。	<ul style="list-style-type: none"> <li>■データベースリンク <a href="http://www.jupui.edu/~tfinter/">http://www.jupui.edu/~tfinter/</a></li> <li>■原文 Childress PJ et al.LymphTF-DB: a database of transcription factors involved in lymphocyte development.Genes Immun. 2007 Jun;8(4):360-5</li> </ul>	111転写因子 110ターゲット遺伝子 78マップされた転写因子結合部位 402転写因子相互作用
BloodExpress	マウス	血液細胞(前駆体、成熟細胞を含む)37の異なるマウスの血液細胞型を含む	遺伝子セット/遺伝子	15の異なる研究による271のマイクロアレイから得られた発現データを統合。これらの発現データはGene Expression OmnibusとStemBaseから取得、もしくは著者から提供を受けている。	前駆細胞と最終的に分化した細胞をともに含むネズミ科の血液細胞型の大部分をカバーしているデータベース。 遺伝子で検索するインターフェースはその遺伝子の造血における発現パターンを、機能的アノテーションと類似した発現パターンの遺伝子リストとともに返す。細胞種類で検索するインターフェースは種の遺伝子機能カテゴリによるフィルタリングによって、血液分化における特定の時点で発現する遺伝子の特定を可能にする。このようにして、BloodExpressは造血過程における新たな遺伝子機能の発見のためのプラットフォームを構成している。	<ul style="list-style-type: none"> <li>■データベースリンク <a href="http://hscl.cimr.cam.ac.uk/bloodexpress/">http://hscl.cimr.cam.ac.uk/bloodexpress/</a></li> <li>■原文 D Miranda-Saavedra et al.BloodExpress: a database of gene expression in mouse haematopoiesis.Nucleic Acids Research, 2009, Vol. 37</li> </ul>	15の異なる研究による271のマイクロアレイから得られた発現データを統合。
StemBase	マウス ヒト ラット	幹細胞の遺伝子	DNAマイクロアレイデータ	マウスとヒト(大部分がAffymetrixのマイクロアレイデータ)から得られた幹細胞と関連細胞のサンプルにおける遺伝子発現測定結果を収集。	ヒトとマウス幹細胞遺伝子発現データのオンラインの最も大きなレポジトリの一つとして、幹細胞のコントロールと分化に関わる遺伝子機能の発見を促進するためCanadian Stem Cell Networkによって最初に作られたDNAマイクロアレイデータのシンプルなwebインターフェース。	<ul style="list-style-type: none"> <li>■データベースリンク <a href="http://www.stembase.ca/?path/">http://www.stembase.ca/?path/</a></li> <li>■原文 R Sandie et al.Recent developments in StemBase: a tool to study gene expression in human and murine stem cells.BMC Research Notes 2009, 2:39</li> </ul>	60の異なる実験による210の幹細胞サンプルから得られた遺伝子発現データ

## 2-3.遺伝子セットの統合データベース事例

PAGED: a pathway and gene-set enrichment database to enable molecular phenotype discoveries

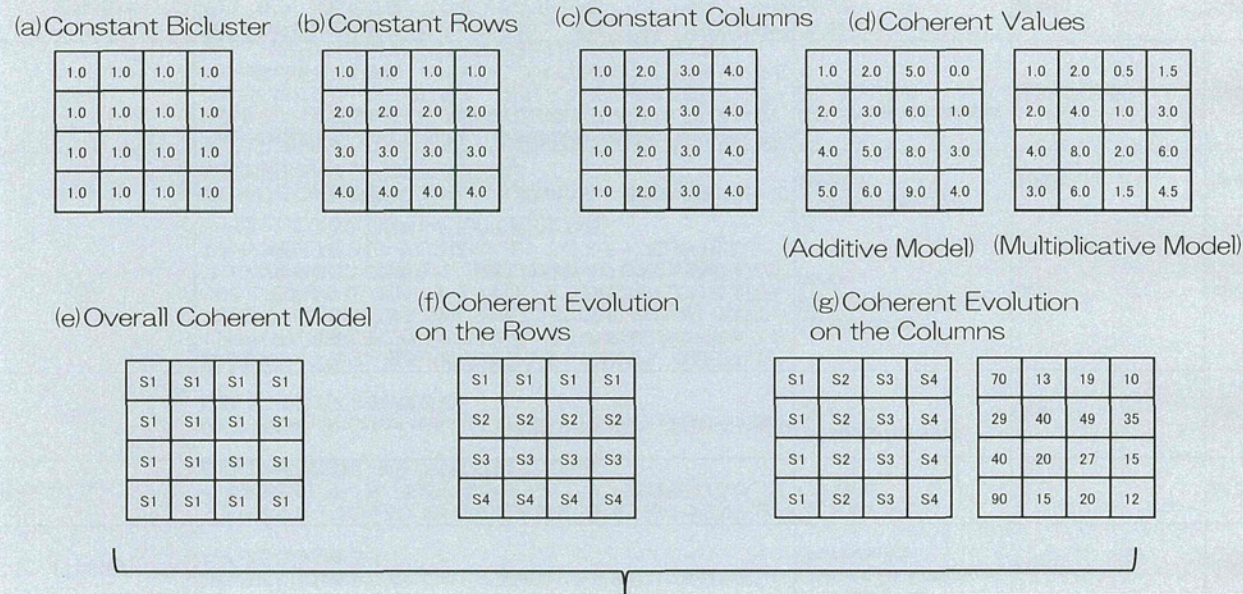
概要	<p>疾病特有のパスウェイ、特徴遺伝子、microRNAターゲット、ネットワークモジュールについての包括的な研究を可能とするための統合されたオンラインデータベース。</p> <p>25,242の遺伝子セット、61,413の遺伝子、20の生物、1,275,560の5つの主要なカテゴリからのレコード（HPDからのパスウェイデータ、OMIM・GADからの遺伝子レベルの疾病データ、MSigDBとGeneSigDBからのトランスクリプトムレベルの特徴遺伝子、miRecordsからのポストトランスクリプトムmicroRNAデータ、HAPPIからのプロテオームレベルのデータ）を含む。</p> <p>さらに、遺伝子セット関連性ネットワーク：GSANSを定義することにより、遺伝子セット間の関係性の検証を可能とした。パスウェイ、特徴遺伝子、microRNAターゲット、そして（または）ネットワークモジュールの間の関係を表示できることである。これらの遺伝子セットをベースとした関係はGSANSとして可視化することができる。</p>
入力データ	<p>①疾病名    ②マイクロアレイデータを解析して得られたFCの高い遺伝子のセット    ③検索結果（①、②）から得られた遺伝子セット</p>
出力データ	<p>①疾病関連遺伝子セット（遺伝子セット名に基づいて検索、または疾病関連遺伝子のプロファイルから検索）②類似遺伝子セット ③疾病特有遺伝子セットの関連ネットワーク（GSANS）。発現レベルの可視化も可能。</p>
提案手法	<p>Gene-set association network(GSANS)は以下のパラメータによって定義される。</p> <ul style="list-style-type: none"> <li>・ノード：遺伝子セット</li> <li>・エッジ：2遺伝子間の関連</li> <li>・ノードサイズ：遺伝子セット中の遺伝子数</li> <li>・ノードの色：遺伝子セットにおける、全ての識別的遺伝子発現のコントロールとの発現量の比（NORM_ABS_FC）</li> <li>・ノードラインの色：遺伝子セットのデータソース</li> <li>・エッジの幅：類似度スコア</li> </ul>  <p>Figure 3 CRC-specific gene-set association network (GSANS) with differential gene-set expressions. The differential gene expressions are from the differential analysis based on the microarray data (GSE8371). Node size: Gene-set size. Node color: Differential gene-set expression (NORM_ABS_FC). Node line color: Gene-set data source. Edge width: Similarity score (&lt;math&gt;\ge 0.1&lt;/math&gt;).</p>
原文	<p>Hui Huang et al., PAGED: a pathway and gene-set enrichment database to enable molecular phenotype discoveries, <i>BMC Bioinformatics</i>, 2012 PAGEDリンク：<a href="http://bio.informatics.iupui.edu/PAGED">http://bio.informatics.iupui.edu/PAGED</a></p>

## 3-1. バイクラスタリング

遺伝子と実験群に対して同時にクラスタリングを行う手法。  
 特定の実験群でのみ類似の発現パターンを示す遺伝子群を発見することができる。

## 概要

## ■ バイクラスタのタイプ例



実際の発現値ではなく、実験群において  
 上方制御または下方制御される遺伝子の  
 サブセットでクラスタリング

## ■ バイクラスタリング手法

大きく以下の5つに分類される。

## ① Iterative Row and Column Clustering Combination

既存のクラスタリング手法をデータ行列の行と列それぞれに適用し、その結果を結合してバイクラスタを得る手法。

## ② Divide-and-Conquer

分割統治法。非常に速いが、バイクラスタを特定する前に分割してしまうことがあり、結果的に良いバイクラスタを見つけれない可能性があるという欠点をもつ。

## ③ Greedy Iterative Search

局所的なゲインを最大化するように行または列を追加または削除することでバイクラスタを生成する方法。高精度ではないが、高速処理が可能。

## ④ Exhaustive Bicluster Enumeration

全探索法。確実に最も良いバイクラスタを見つけることが可能だが、その高い複雑性ゆえに、バイクラスタのサイズ制限を仮定しないと実行されないという重大な欠点をもつ。

## ⑤ Distribution Parameter Identification

統計モデルを与えられていると仮定し、反復法を用いてある基準を最小化することでデータを生成するための分布パラメータを特定することを目指す。

出所) Sara C. Maderia and Arlindo L. Oliveira, Biclustering Algorithm for Biological Data Analysis: A Survey, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 1 Issue 1, 2004, pp. 24 - 45

## 3-2. バイクラスタリング手法例

## ■ バイクラスタリング手法例

手法名	概要	手法タイプ	前処理	バイクラスタタイプ	利用方法
CTWC	遺伝子群と実験群それぞれに対して階層的クラスタリングを繰り返し、バイクラスタを得る。	①Iterative Row and Column Clustering Combination	なし	(c) Constant Columns	-
Block Clustering	マトリックス（ブロック）の分散の合計が最も削減するようにマトリックスを2分割していき、予め定めた数のバイクラスタが得られるか、分散の合計値が閾値を下回った時点で終了とする。	②Divide-and-Conquer	なし	(a) Constant Values	-
CC	MRS (Mean Squared residue) が閾値よりも低い、すなわち分散が小さいバイクラスタを探索する。	③Greedy Iterative Search	なし	(d) Coherent Values	BicATをダウンロードすることで使用可能 <a href="http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php">http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php</a>
ISA	最初のステップでは、遺伝子群G0がランダムに選ばれ、遺伝子スコア1がそれぞれの遺伝子に割り当てられる。全ての実験における実験スコアがこれらの遺伝子で計算され、絶対スコアが閾値よりも高い実験群がC0として選ばれる。次のステップでは、全ての遺伝子における遺伝子スコアが実験群C0で計算され、定義した閾値よりも大きい遺伝子スコアをもつ遺伝子群がG1として選ばれる。このステップを繰り返し、Gn=Gn-1となるまで収束した時点で終了となる。	⑤Distribution Parameter Identification	なし	(a) Constant Bicluster	BicATをダウンロードすることで使用可能 <a href="http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php">http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php</a>
OPSM	遺伝子ごとに発現値をランク付けし（離散化）し、複数の実験において順位が保存されるバイクラスタを探索する。	③Greedy Iterative Search	離散化	(g) Coherent Evolution	BicATをダウンロードすることで使用可能 <a href="http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php">http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php</a>
BiMax	遺伝子の発現値を0/1で2値化し、探索領域の一番最初の行の遺伝子の発現値によって重複を許して実験条件によるサブマトリックスを作成。それぞれのサブマトリックス毎に、発現している実験数で遺伝子をソートし探索領域を分割する。それぞれについて再帰的にクラスタリングを行い、1のみから成るバイクラスタを得られた時点で終了となる。	④Exhaustive Bicluster Enumeration	離散化（2値化）	(a) Constant Bicluster	BicATをダウンロードすることで使用可能 <a href="http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php">http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php</a>
xMotif	遺伝子の発現状態を離散化によって定義し、実験にわたって発現状態が同じ、かつ、実験数が最大になるような遺伝子群（バイクラスタ）を探索する。	③Greedy Iterative Search	離散化	(f) Coherent Evolution	BicATをダウンロードすることで使用可能 <a href="http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php">http://www.tik.ee.ethz.ch/sop/bicat/?page=bicat.php</a>
QUBIC	正負の記号をもつ整数と0の値で表現された遺伝子発現マトリックスにおいて、シードとなる類似度の高い遺伝子のペアを選び、0の値をとる実験を除外して初期のバイクラスタを作成したのち、あらかじめ定めた閾値を超えない範囲でそのグループを拡大し、バイクラスタを決定。	③Greedy Iterative Search	なし	(a) (d) Constant and Coherent values	Bioconductorのパッケージ ( <a href="http://www.bioconductor.org/packages/2.11/bioc/html/rqubic.html">http://www.bioconductor.org/packages/2.11/bioc/html/rqubic.html</a> )
SAMBA	サンプルと遺伝子を頂点とし、それらの間が辺で結ばれる2分グラフを考える、あるサンプル（実験）において、ある遺伝子の発現量が上方/下方制御されている場合、そのサンプルと遺伝子を辺で結び、そのグラフ中にあるクリークをバイクラスタのシードとして抽出する。次に、バイクラスタ（サブグラフ）に対して頂点を追加/削除してみ、バイクラスタが改善される（サブグラフ中の辺の数が増える）場合にその追加/削除を受理する。	④Exhaustive Bicluster Enumeration	離散化	(e) Coherent Evolution	Expanderをダウンロードすることで使用可能 (登録が必要) <a href="http://acgt.cs.tau.ac.il/expander/">http://acgt.cs.tau.ac.il/expander/</a>
SAMURAI	飽和アイテム集合列挙法 (LCM) に基づく高速遺伝子モジュールマイニング法で、データの離散化及び、ユーザーのクエリファイルと類似したモジュールを探索することで不要データを圧縮し高速性を実現。最小サポート値以上の実験で同じ発現パターンを持つ遺伝子の最大集合をクラスタリングしている。	④Exhaustive Bicluster Enumeration	離散化	(e) (f) (g) Coherent Evolution	CELLPEDIA( <a href="http://cellpedia.cbrc.jp/cgi-bin/index.cgi">http://cellpedia.cbrc.jp/cgi-bin/index.cgi</a> )内のツールとして以下で提供 <a href="http://samurai.cbrc.jp/cgi-bin/index.cgi?page=2">http://samurai.cbrc.jp/cgi-bin/index.cgi?page=2</a>

出所) Sara C. Maderia and Arlindo L. Oliveira, Biclustering Algorithm for Biological Data Analysis: A Survey, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 1 Issue 1, 2004, pp. 24 - 45

藤津涉他、マイクロアレイデータ統計解析プロトコール、羊土社、2008年

## 3-3.代表的なバイクラスタリング手法の特徴比較

■以下に前頁に記載したバイクラスタリング手法のうち代表的なものについて、複数の比較論文(①~③)を引用し評価を行った結果を示す。

手法名	ノイズ耐性			バイクラスタの オーバーラップへの耐性			処理時間		バイクラスタ 数	バイクラスタ サイズ	実データ (GOによるfunctional enrichmentで 評価)
	①	②	③	①	②	③	①	④	①	②	③
CC	×	△	×	×	×	-	×(※2)	-	-	◎	×
ISA	○	/	○	◎	/	○	×	-	◎	/	○
OPSM	×	×	△	-	○	○	×	×	×	○	◎
BiMax	◎	/	○(※1)	×	/	◎	○	-	-	/	○
xMotif	×	/	×	×	/	-	×	-	◎	/	/
QUBIC	◎	/	/	×	/	/	○	-	○	/	/
SAMBA	/	-	△	/	○	◎	/	○	/	-	○
SAMURAI	/	/	/	/	/	/	/	○	/	/	/

※1：但しバイクラスタのタイプに影響を受ける  
 ※2：遺伝子数が増えるにつれ急激に増加

- ・バイクラスタリング手法はターゲットとするバイクラスタタイプによって選択する必要があるため、一概に手法を比較評価することは難しい。
- ・ISA,BiMax,SAMBAは比較的欠点が少なく、ソフトウェアも提供されているため使いやすいと考えられる。
- ・SAMURAIは比較的新規手法であり、高速処理が特徴だが、比較論文が少なく評価が困難。

## 【参照論文】

- ①Kemal Eren et al., A comparative analysis of biclustering algorithm for gene expression data,Briefing in Bioinformatics,2012  
 ②D Bozdag et al., Comparative analysis of biclustering algorithms,Proceeding BCB '10 Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology,2010,pp. 265-274  
 ③Amela Prelic,A systematic comparison and evaluation of biclustering methods for gene expression data,Bioinformatics,2006,Vol. 22 no. 9,pp.1122-1129  
 ④[http://hinv.jp/pdf/20100914/20100914\\_H22\\_IDB-lecture1\\_fujibuchi.pdf](http://hinv.jp/pdf/20100914/20100914_H22_IDB-lecture1_fujibuchi.pdf)



# 実験データ管理システム 管理者向けマニュアル

平成25年3月

日本電気株式会社

## 使用上のご注意

本システムのご使用にあたり、以下の点にご注意ください。

### 動作環境

- 本システムは、ファイアウォールなど、お客様側にてセキュリティ対策済みの LAN 内で運用されることを前提としています。
- 本システムのクライアント動作環境は下記を前提としております。  
オペレーティングシステム: Microsoft Windows 7  
ブラウザ: Google Chrome 25 以降

### 禁止事項および免責事項

- 本システムに関わる禁止事項および免責事項は、東京大学様と日本電気株式会社間で締結した実験データ管理システムの開発委託契約に準ずるものとします。

Copyright (c) NEC Corporation 2013. All rights reserved.

## 目次

1. はじめに .....	1
1-1. 実験データ管理システムについて .....	1
1-2. 利用者権限について .....	1
1-3. 研究グループについて .....	2
2. 利用者と研究グループの管理 .....	3
2-1. 利用者の登録・変更 .....	3
2-2. 研究グループの登録 .....	6
2-3. その他のシステム管理機能 .....	10
2-3-1. 使用履歴を表示する .....	10
2-3-2. バージョン情報を表示する .....	11
付録1. [システム情報を更新する]ボタンについて .....	13

## 1. はじめに

### 1-1. 実験データ管理システムについて

本書では、東京大学医科学研究所様に導入した「実験データ管理システム」(以下、「本システム」と記します。)にてユーザの登録や研究グループの登録を行う操作について説明します。

なお、本システムは、日本電気株式会社製の研究支援システム「BIOPRISM」をベースとしています。

### 1-2. 利用者権限について

本システムでは、利用者ごとに権限を設定することができます。権限を設定することによって、操作可能な機能および閲覧可能な情報の範囲を制限することができます。本システムにおける利用者権限の種類は次の通りです。

#### 【システム管理者】

利用者の登録および利用者権限の変更と、研究グループの登録および変更を行うことができます。

#### 【研究管理者】および【研究責任者】

匿名化済みの患者情報の閲覧、匿名化の差し戻しが行えます。

本システムでは、他に次のような権限が定義できますが、バイオバンク匿名化システムの患者情報管理に関しては利用しません。

- ・ オリジナルサンプル責任者（サンプル管理機能を使用する場合に管理責任者となります）
- ・ 受入担当者（サンプル管理機能を使用する場合に受入処理を行います）
- ・ 研究担当者（実験データを管理する場合に設定します）
- ・ 実験補助（実験データの一部を入力するための権限です）
- ・ 匿名化点検者（サンプルに関して患者同意確認と匿名化が済んでいることをチェックします）
- ・ 臨床情報入力者（臨床情報を登録する場合、項目の定義を行うことができます）
- ・ 個人情報管理者（患者情報の登録を行うことができます）

本システムにおける上述の各権限は、登録されている利用者に自由に割り当てることができます。例えば、役割の分担範囲が重ならないような研究室であれば、一人に一つの権限を設定し、一人が担う役割が多い研究室であれば、一人に複数の権限を設定する等が可能です。

### 1-3. 研究グループについて

本システムでは、登録されたデータの閲覧を適切に制限できるよう、「研究グループ」を設定できるようになっています。部門や研究プロジェクトごとに研究グループを定義し、所属する利用者を設定することによって、同じ部門内あるいはプロジェクト内でのみデータを共有することが可能です。ひとりの利用者を複数の研究グループに所属させることも可能です。

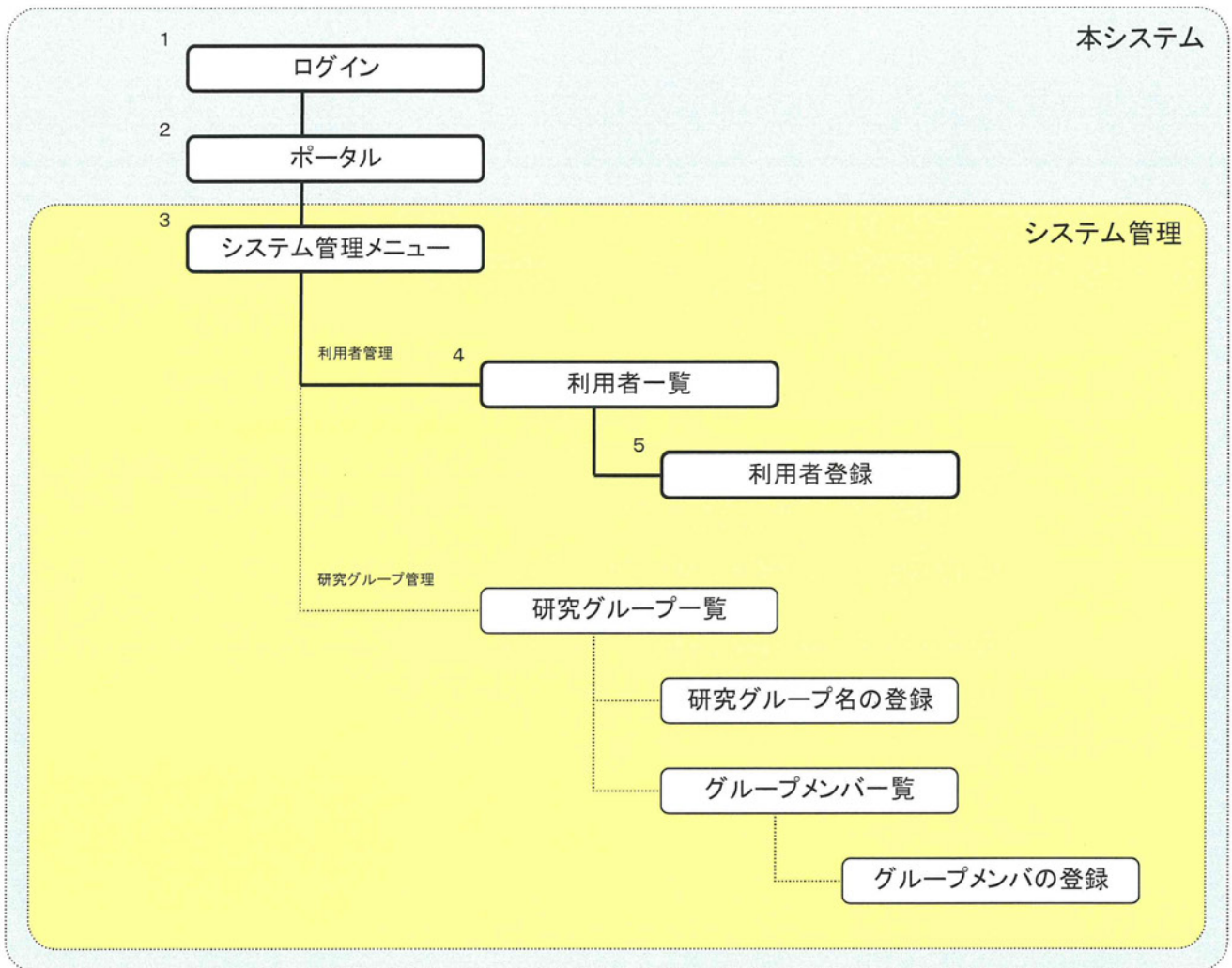
## 2. 利用者と研究グループの管理

本システムをご利用頂くには、本システムにおける「システム管理者」が、利用者の登録、研究グループの登録などを行う必要があります。(注:本システムにおける「システム管理者」はサーバや OS の「管理者」とは異なります。)

### 2-1. 利用者の登録・変更

本システムの利用者を登録する操作手順について以下に説明します。利用者の登録を行うには、次の権限を持ったユーザーでログインしてください。

システム管理者



図中の数字は、本節における手順番号です。

図1 利用者の登録・変更に関わる機能

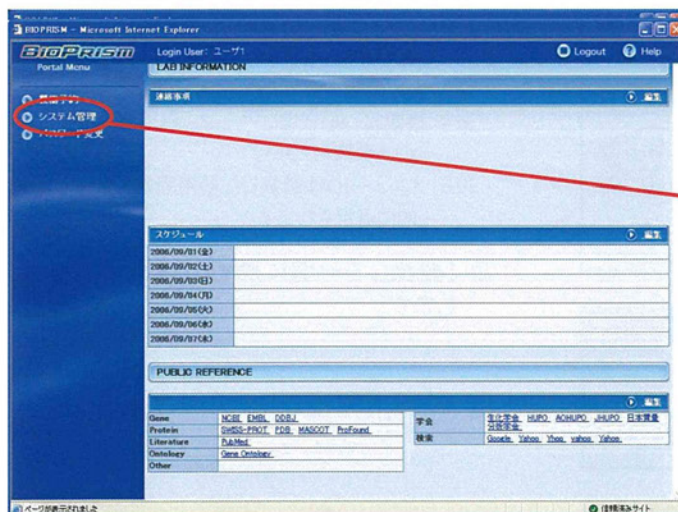
出荷状態の本システムには、“システム管理者”権限を持つ利用者1名のみが登録されています。まずこの利用者でログインし、利用利用者の登録を行います。利用者登録や、自分以外の利用者パスワードの変更、利用者の無効化などは“システム管理者”権限を持った利用者のみが行えます。パスワードの管理は十分にご注意ください。



### 手順1. 本システムへのログイン

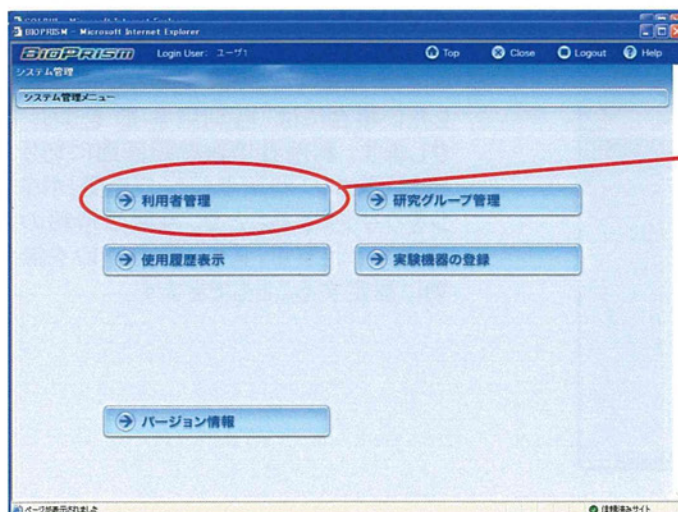
- ① ブラウザでサーバに接続し、利用者ID、パスワードを入力します。  
サーバの URL は、  
http://[サーバの IP アドレス]:8080/portal/  
のように指定します。
- ② [Login]ボタンをクリックします。  
◆ [Exit]ボタンをクリックすると、ウィンドウが閉じて終了します。

注意: クライアント PC を起動してから初めてログイン画面を表示する場合、利用者 ID が入力できるようになるまで時間がかかる場合があります。PC を再起動するまでは、2 回目よりすぐにログインできます。



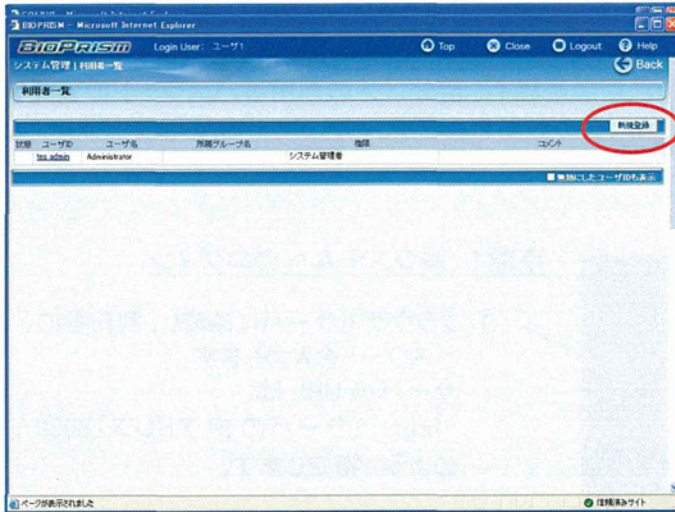
### 手順2. ポータル画面

- ① メニューの[システム管理]をクリックします。



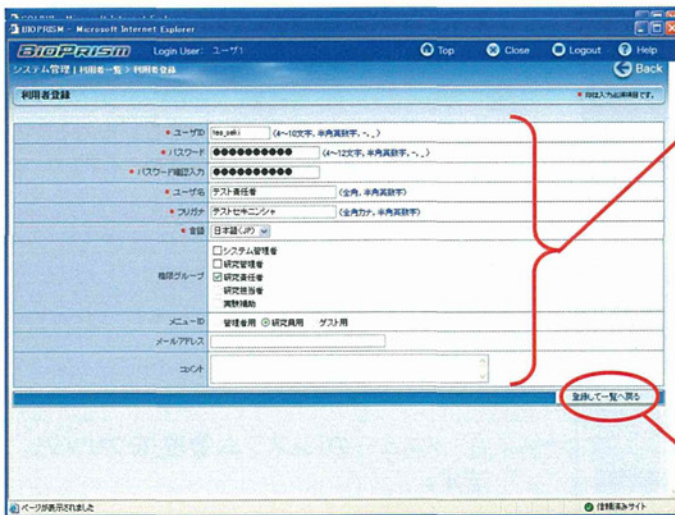
### 手順3. システム管理の Top 画面

- ① [利用者管理]ボタンをクリックします。



#### 手順4. 利用者一覧画面

- ① [新規登録]ボタンをクリックします。

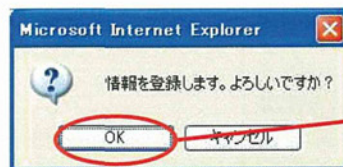


#### 手順5. 利用者の登録

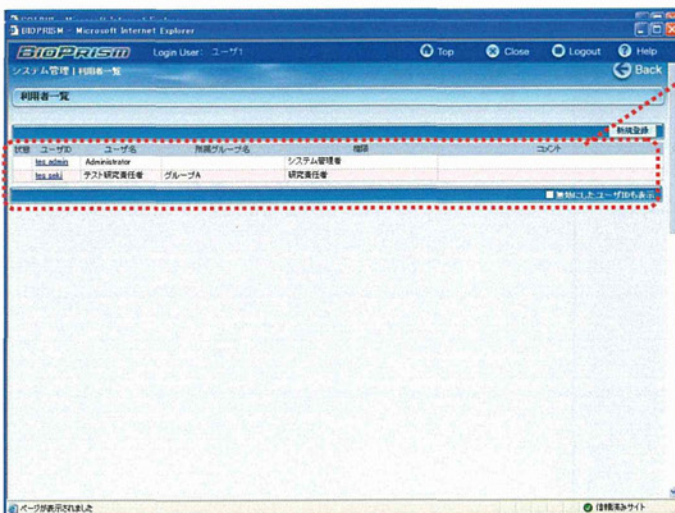
- ① 「利用者ID」※1「パスワード」「パスワード確認入力」「利用者名」※2「フリガナ」「言語」※3「権限グループ」※4を入力または選択します。必要に応じて「メールアドレス」「コメント」を入力します。※5

- ※1 利用者IDはログイン時に使用します。
- ※2 利用者はログイン後、各画面の Login User 欄に表示されます。
- ※3 言語は現在日本語のみ選択可能です。
- ※4 権限については、「1-2. 利用者権限について」を参照ください。
- ※5 メニューIDは選択した利用者権限により自動的に選択されます。

- ② [登録して一覧に戻る]ボタンをクリックします。



- ③ [OK]ボタンをクリックします。



- ④ 利用者一覧に、登録した利用者が表示されます。
- ⑤ 利用者権限や利用者情報などを変更したい場合には、各利用者 ID をクリックします。利用者情報参照画面に切り替わります。[利用者情報の変更]ボタンをクリックすることで、権限や情報の変更が行えます。また、利用者 ID を無効に設定することもできます。



## 2-2. 研究グループの登録

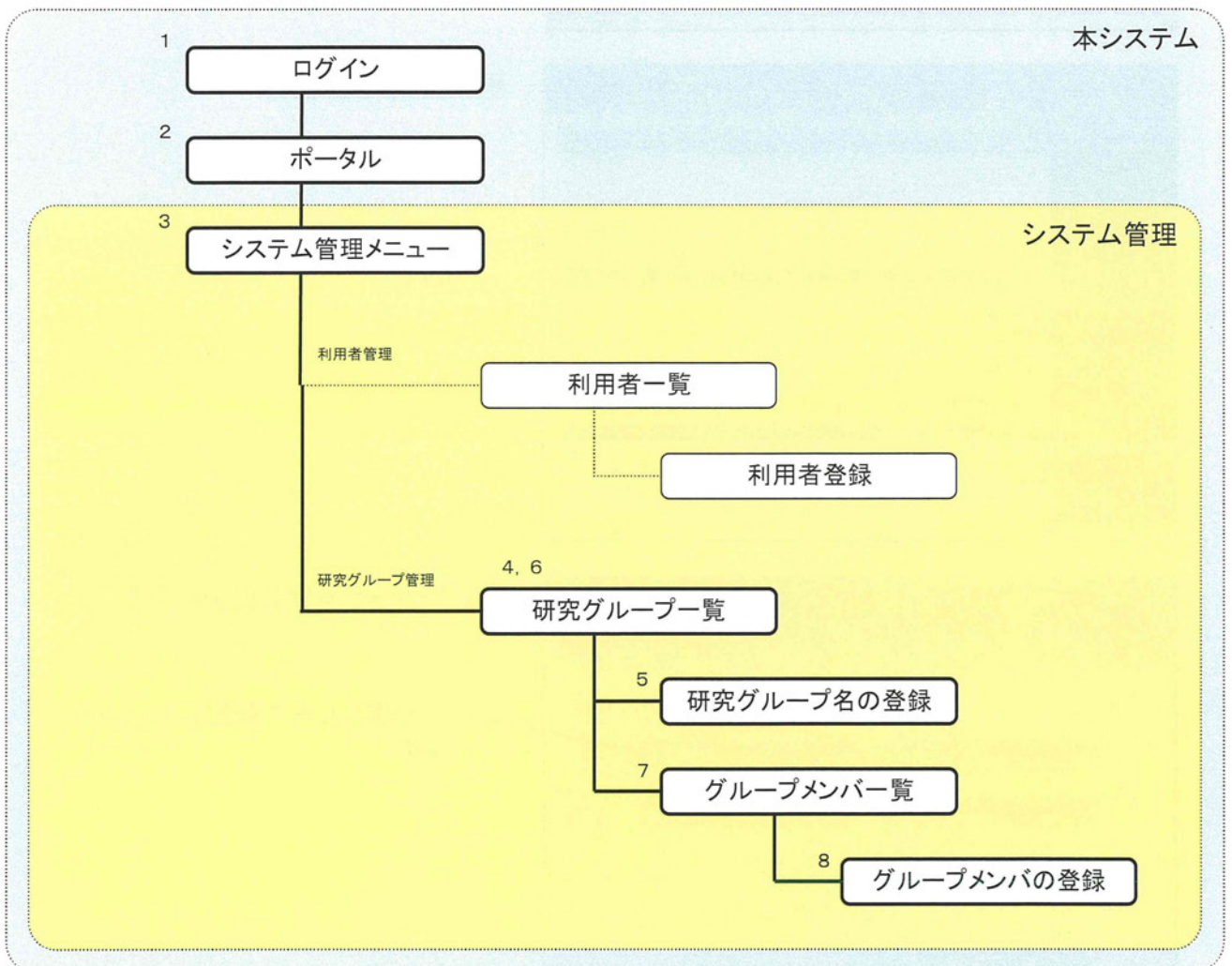
研究グループの登録を行うには、次のいずれかの権限を持ったユーザーでログインしてください。

システム管理者

研究管理者

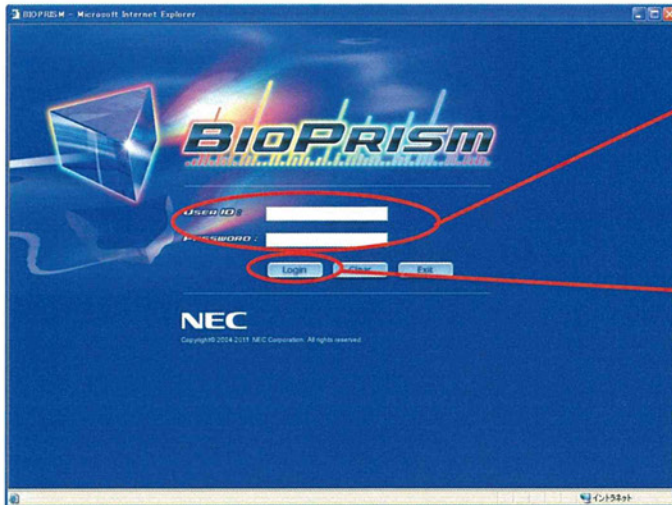
本システムは複数の研究室や研究プロジェクトによる利用を想定しています。このため、同じ研究室や研究チームに所属する利用者だけがデータを共有できるよう、研究グループによるデータ管理を行います。データは、研究グループ単位で登録され、利用者は所属する研究グループ以外のデータを参照することはできません。利用者は複数の研究グループに所属することができます。

研究グループを登録する操作手順について、以下に説明します。



図中の数字は、本節における手順番号です。

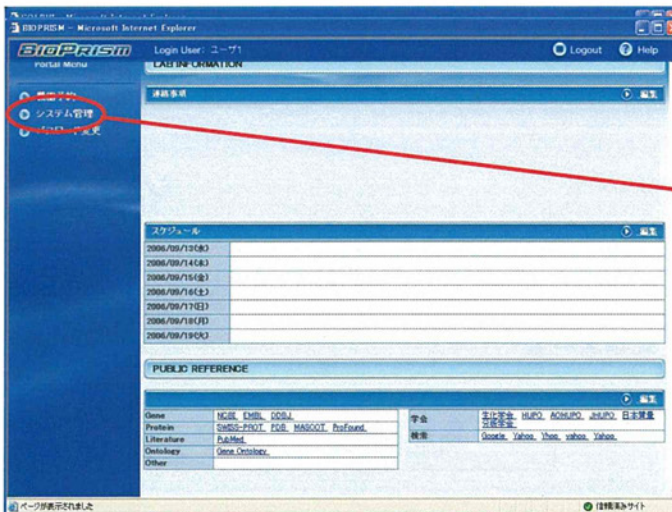
図2 研究グループの登録に関わる機能



### 手順1. 本システムへのログイン

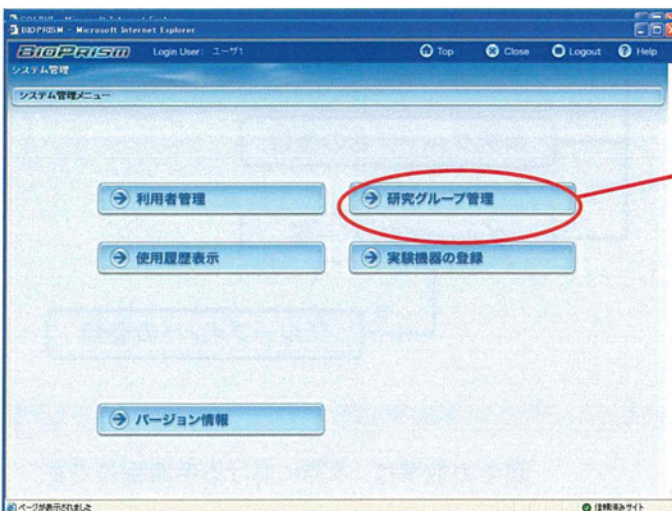
- ② ブラウザでサーバに接続し、利用者ID、パスワードを入力します。  
サーバの URL は、  
http://[サーバの IP アドレス]:8080/portal/  
のように指定します。
- ② [Login]ボタンをクリックします。

注意: クライアント PC を起動してから初めてログイン画面を表示する場合、利用者 ID が入力できるようになるまで時間がかかる場合があります。PC を再起動するまでは、2回目よりすぐにログインできます。



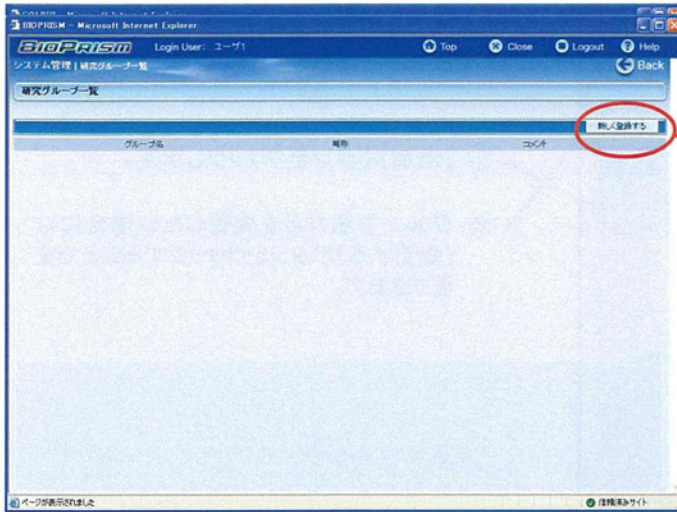
### 手順2. ポータル画面

- ① メニューの[システム管理]をクリックします。



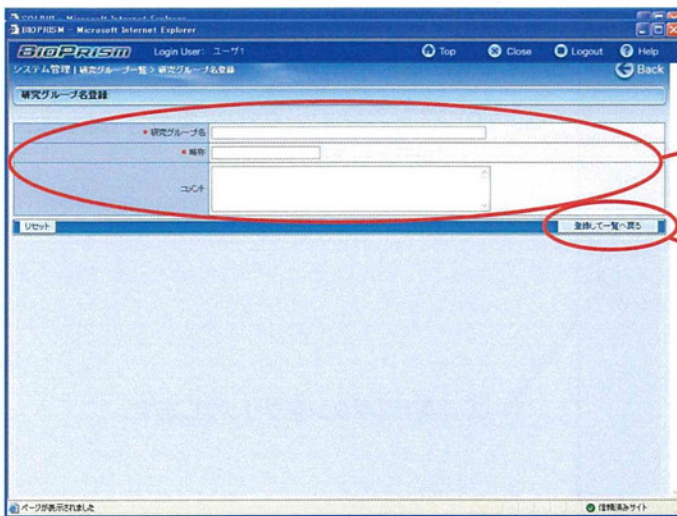
### 手順3. システム管理の Top 画面

- ① [研究グループ管理]ボタンをクリックします。



#### 手順4. 研究グループ一覧

① [新しく登録する]ボタンをクリックします。

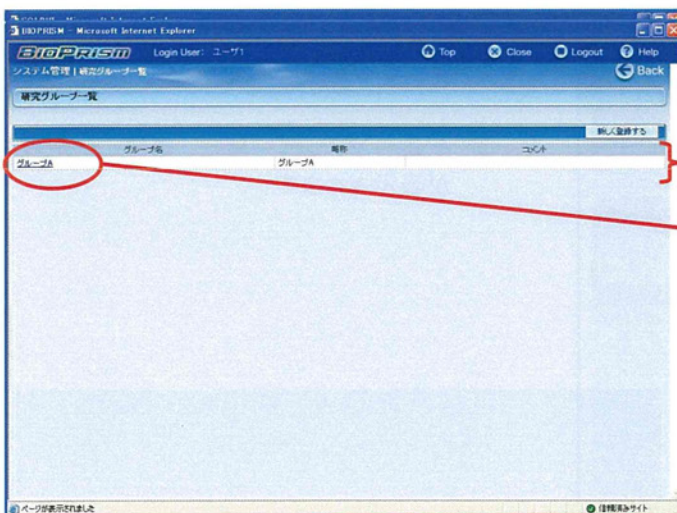


#### 手順5. 研究グループ名の登録

- ① 必須項目「研究グループ名」「略称」と任意項目「コメント」を入力します。
- ② [登録して一覧に戻る]ボタンをクリックします。

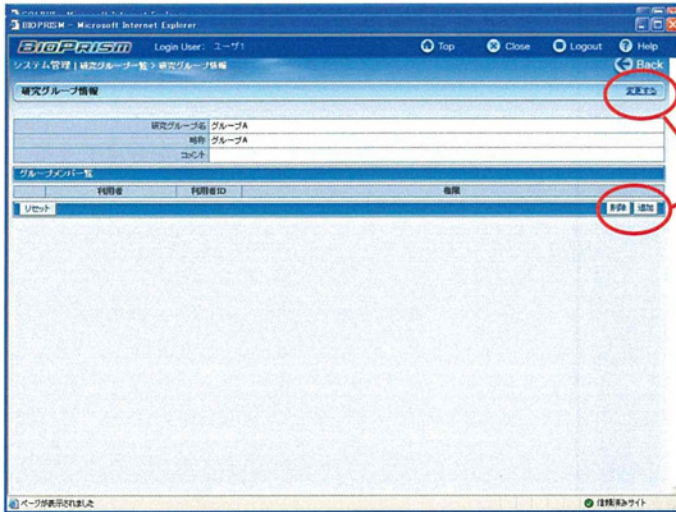


③ [OK]ボタンをクリックします。



#### 手順6. 研究グループ一覧

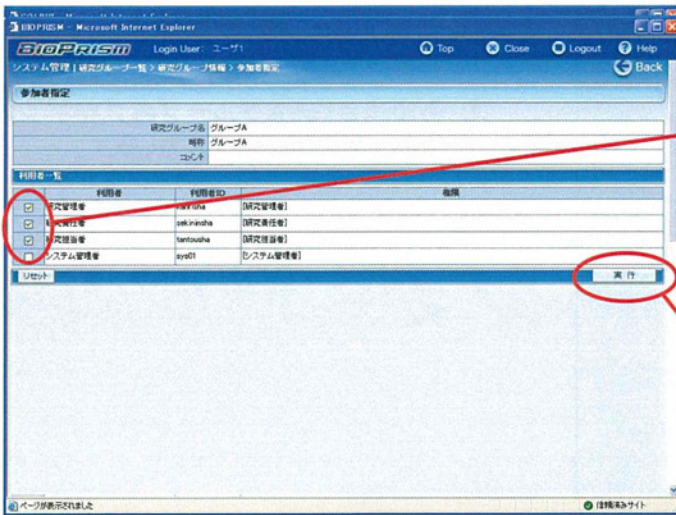
- ① 研究グループ一覧に、登録した研究グループが表示されます。
- ② 登録した研究グループ名をクリックします。



### 手順7. グループメンバー一覧画面

① [追加]ボタンをクリックします。

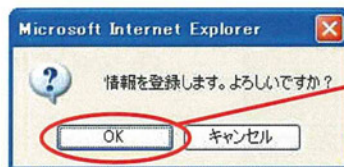
※ グループ名などを変更したい場合には [変更する]ボタンをクリックすることで変更できます。



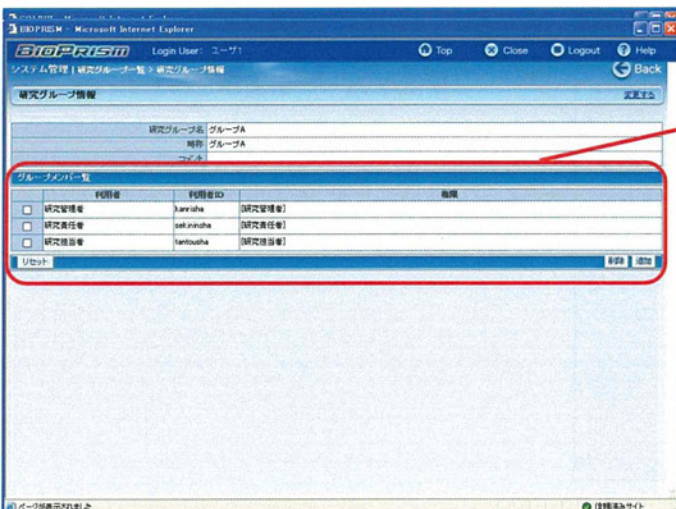
### 手順8. グループメンバーの登録

① グループメンバーとして登録する利用者にチェックを付けます。

② [実行]ボタンをクリックします。



③ [OK]ボタンをクリックします。



④ グループメンバー一覧に、登録したメンバーが表示されます。