

# パスイェ情報を用いた顧みられない熱帯病 創薬支援のための統合データベース iNTRODB の改良

蓮実 梢<sup>1,a)</sup> 石田 貴士<sup>2</sup> 秋山 泰<sup>2,3</sup>

**概要:** 標的となるタンパク質を定めて行う薬剤開発において、薬剤標的タンパク質の選定はとても重要となる。この標的タンパク質探索には、既に病原体のゲノム情報等を利用して探索を行うための統合的なデータベースシステムが提案されてきたが、その生化学経路情報については統合の対象とされていなかった。しかし、この情報を用いることであるタンパク質が病原体の生命維持に対して致命的であるかという議論が可能となるため、生化学経路情報の統合は標的タンパク質の探索のために有用であると考えられる。そこで本研究では、顧みられない熱帯病の新薬の標的タンパク質を探索するための統合データベースシステム iNTRODB に、トリパノソーマ科寄生原虫に関する生化学経路情報を追加し、またそこにゲノム等に関連する情報を表示するインタフェースを開発することで、標的タンパク質の探索の更なる効率化を目指した。

**キーワード:** 顧みられない熱帯病, iNTRODB, KEGG, トリパノソーマ

## Improvement of a database for supporting NTDs drug discovery based on biological pathway information

**Abstract:** In structure based-drug design, selecting a drug target protein is very important. For the target protein selection, several database systems integrating various related information, such as genomic information of pathogens and phenotypic information, have been proposed. However, biological pathway information, which may facilitate understanding the importance of proteins, has not been integrated. In this research, we integrated the biological pathway information about Trypanosomatidae family, protozoans parasites into a database system iNTRODB, which has been developed for selecting drug target protein of neglected tropical diseases. We also developed an interface to display pathway information with genome and protein information to improve the search process of drug target proteins.

**Keywords:** Neglected Tropical Diseases, iNTRODB, KEGG, Trypanosomatidae

### 1. はじめに

顧みられない熱帯病 (Neglected Tropical Diseases, 以下 NTDs) とは、世界保健機構によって定義された熱帯地域・貧困層を中心に蔓延している 17 の寄生虫・細菌感染症の総称である。多くの場合その感染には衛生状態や生活状態

が大きく関わるため、先進国の中ではあまり問題にならず [1], 所得水準の低い熱帯地域が主な感染地域であり、新薬の研究開発対象とされることは稀なことから、3 大感染症 (エイズ, マラリア, 結核) と比べ、十分な対策がとられてこなかった。しかし、1997 年の G8 北海道洞爺湖サミット首脳宣言を 1 つの契機として、近年ではこれらの疾患群にも焦点があてられ、新薬の開発が試みられるようになりつつある。現在では、複数の製薬企業や学術機関が非営利でこれらの疾患への新薬開発を開始しており、我々も国内製薬企業、学術機関が連携する NTDs 創薬コンソーシアムに参加し、新薬開発のための技術開発を行っている [2].

従来、新薬開発では、病原体を殺す化合物を多種の化合物からまず探し出し、見つかった化合物が病原体のどのタ

<sup>1</sup> 東京工業大学 工学部 情報工学科  
Dept. of computer science, Tokyo Institute of technology  
<sup>2</sup> 東京工業大学 大学院情報理工学研究科 計算工学専攻  
Graduate School of Information Science and Engineering,  
Tokyo Institute of technology  
<sup>3</sup> 東京工業大学 情報生命博士教育院  
Education Academy of Computational Life Science,  
Tokyo Institute of technology  
a) hasumi@bi.cs.titech.ac.jp

ンパク質に対して効果があるのか不明なまま、構造活性相関等で化合物の最適化を行うことで創薬が行われてきた。一方、近年ではまず薬剤の標的となるタンパク質を先に選定し、そのタンパク質に対して効果のある化合物を探索する標的タンパク質ベースの手法が主流となりつつある。特に X 線結晶解析等により立体構造の決定を行い、その情報を用いることでより効率的に新薬の探索を行う structure-based drug design (SBDD) の手法は様々な創薬で成功を収めている [3]。しかしそのような創薬手法においては、標的タンパク質の選定が必須となり、標的とするタンパク質の選択によって成果が大きく左右される。これは多くの遺伝子、タンパク質について基礎研究が進み、様々な標的タンパク質が既に多数提案されているヒトや主要な疾患の病原体についてはあまり大きな問題とならないが、寄生原虫のようにあまり研究の進んでいない生物が病原体である NTDs では既知の標的タンパク質の数が不十分であり、新薬開発の問題の一つとなっている [4]。

そのため NTDs 向け創薬において、この新薬標的タンパク質の発見を効率化するため、我々の研究グループでは integrated Neglected TROPical disease DataBase (iNTRODB) [5] と名付けたシステムの開発を行ってきた。iNTRODB は薬剤化合物情報を提供する ChEMBL[6] やタンパク質構造情報を提供する PDB[7] といった公的データベースの情報を統合し、それらの情報から標的として適したタンパク質を探索するための探索機能や予測機能を備えたシステムである。現在はトリパノソーマ科寄生原虫症やビルハルツ住血吸虫症等に対応している。特にトリパノソーマ科寄生原虫症については個別のタンパク質についての既存のフェノタイプ実験情報 [8] も統合されており、これは標的タンパク質候補の探索には有用な情報となっている。しかし多くの有用な情報が利用可能となる一方、現在のシステムでは寄生原虫のタンパク質間の関係性、より詳しく言えば生化学経路とそれに関連するタンパク質の知識が欠落してしまっている。もし生化学経路情報を利用することができれば、薬剤標的候補のタンパク質が寄生原虫の生命活動維持に致命的なタンパク質かどうかを生化学経路の観点からも議論が可能となり、有用であると考えられる。例えば最初に注目したタンパク質が何らかの理由で標的とすることができない場合でも、その経路の上流もしくは下流のタンパク質を探索することで、新たな標的候補の探索が可能になると考えられるからである。

現在、標的タンパク質候補の探索機能を持つ NTDs に関するデータベースは iNTRODB の他にもいくつか存在するが、残念ながら現時点ではそれらのシステムでも生化学経路情報は統合されていない。例えば、TDR Targets[9] は iNTRODB に近い機能を備えるデータベースであり、結核菌やトリパノソーマ科寄生原虫等に対して様々な探索項目から標的タンパク質を探索する機能が実装されており、

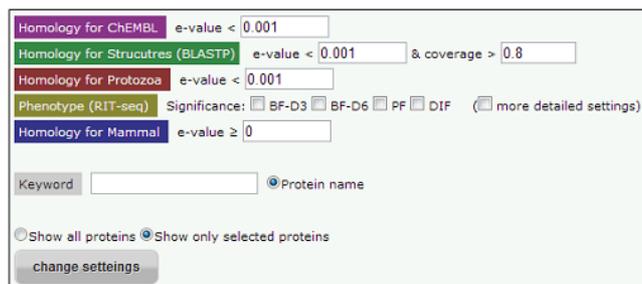


図 1 Genome-wide protein list の探索欄

Fig. 1 Search fields of genome-wide protein list.

GeneDB ID	Protein name	RIT-seq	ChEMBL hit	ChEMBL e-value	ChEMBL e-value (BLASTP)	PDB e-value (BLASTP)	Phospho (BLASTP)	Enzymes major	Developmental stage	Home organ	Mammal
TSD3_353_00	protein kinase, putative	no info.	ChEMBL2451	2e-20	1042	5e-23	0.88	2e-28	1e-118	2e-20	2e-21
TSD3_353_100	small GTP-binding protein, putative	no info.	ChEMBL2597	1e-28	1	1e-15	0.97	1e-112	1e-113	1e-20	2e-21
TSD3_160_0010	lysozyme, putative	1-1-1-8	ChEMBL2989	1e-28	17	2e-20	0.97	1e-149	0	2e-20	1e-21
TSD3_160_1680	cytidine deaminase, putative	1-1-1-1	ChEMBL1110	2e-22	37	2e-21	0.80	1e-51	2e-61	2e-20	2e-21
TSD3_160_1850	acetosuccinyl enolpyruvate, putative	1-1-1-1	ChEMBL203850	2e-18	11	1e-18	0.80	2e-58	1e-149	2e-17	2e-18

図 2 Genome-wide protein list の探索結果

Fig. 2 Search results of genome-wide protein list.

その探索項目の中に標的とする病原体にとって重要であるタンパク質かどうかという項目が備えられている。その病原体の生存に対して重要であるかどうかを調べられるという点で本研究の目的に近い機能ではあるが、その情報は既知の実験情報由来のものであり、生化学経路情報から推定されたものではなく、また関連する生化学経路情報を調べることも不可能である。そのため生化学経路情報の統合は創薬標的タンパク質候補の探索に、従来は利用されてこなかった新たな知見を導入するものである。そしてその情報を他のゲノム、立体構造と言った情報と組み合わせることで標的タンパク質候補探索の効率化、高精度化が期待できる。

そこで、本研究ではトリパノソーマ科寄生原虫の創薬標的タンパク質候補探索の更なる改善を目指し、iNTRODB にトリパノソーマ科寄生原虫の生化学経路情報を統合し、それらを表示するインタフェースの開発を行った。

## 2. 手法

### 2.1 iNTRODB

iNTRODB は NTDs について既存のデータベース情報を集約し、新薬標的として適したタンパク質の発見を目的とした探索機能を備えたデータベースであり、実験生物学者のようなユーザーにとっても簡潔でわかりやすいインタフェースによる探索機能を構築している。

本研究で開発した機能を実装するのは、寄生原虫のゲノム情報から薬剤候補として適したタンパク質を探索するための探索リスト genome-wide protein list である。現在 iNTRODB は複数の寄生原虫や病原体について対応しているが、本研究で機能開発を行うのはトリパノソーマ科に属するリーシュマニアメジャー (*Leishmania major*)、クルーズトリパノソーマ (*Trypanosoma cruzi*)、ブルーストリパノソーマ (*Trypanosoma brucei*) の 3 種類についてで

GeneDB ID	Protein name	ChEMBL Hit	ChEMBL E-value	ChEMBL # compounds	PDB e-value (BLASTP)	PDB coverage (BLASTP)	Leishmania major	Trypanosoma brucei	RIT-seq	Homo sapiens	Mammal	KEGG pathway
Tc00.1047033508947.80	arginine N-methyltransferase, type 1	ChEMBL1075069	2e-32	94	3e-37	0.84	1e-114	1e-151	others	5e-24	2e-23	No hit
Tc00.1047033510739.19	cystathione gamma lyase, putative	ChEMBL1075079	7e-13	23	2e-38	0.89	1e-104	1e-106	1-1-1-1	2e-31	2e-33	No hit
Tc00.1047033510741.10	cystathione gamma lyase, putative	ChEMBL1075079	2e-14	23	4e-57	0.98	1e-136	1e-145	1-1-1-1	2e-43	1e-51	No hit
Tc00.1047033508241.149	cystathionine beta-synthase, putative	ChEMBL1075088	9e-12	28	8e-28	0.84	3e-37	6e-43	1-1-1-1	2e-26	6e-27	No hit
Tc00.1047033507165.50	cysteine synthase, putative	ChEMBL1075088	9e-73	28	1e-171	0.98	1e-172	1e-45	1-1-1-1	4e-29	6e-28	ter00270R(C P H) ter00920R(C P H) ter01100R(C P H) ter01230R(C P H)
Tc00.1047033508241.140	cystathionine beta-synthase, putative, cysteine synthase, serine sulfhydrylase	ChEMBL1075088	2e-37	28	1e-110	0.86	0	0	1-1-1-1	1e-108	1e-113	ter00260R(C P H) ter00270R(C P H) ter01100R(C P H) ter01230R(C P H) ter00260R(C P H)

図 4 KEGG pathway を実装した genome-wide protein list  
Fig. 4 Genome-wide protein list that implement the KEGG pathway.

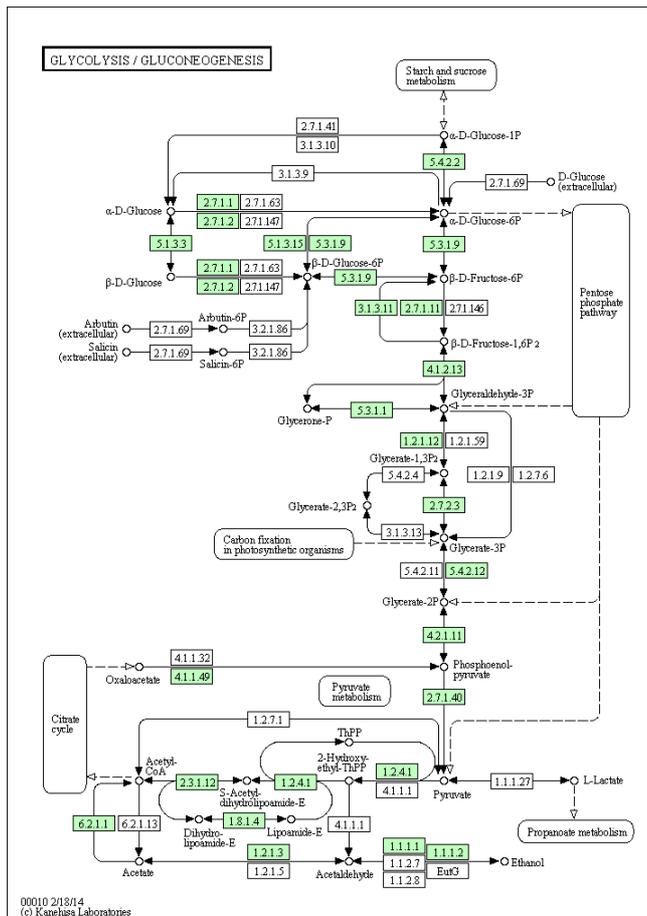


図 3 *T. cruzi* の解糖パスウェイ  
Fig. 3 Glycolytic pathway of *T. cruzi*.

ある。Genome-wide protein list の探索欄は図 1 のようになっており、ChEMBL 中の化合物配列との相溶性や立体構造との類似性等の探索条件を入力することで、図 2 のように探索条件に合致する寄生原虫のゲノムと各相溶性等について探索結果を表示する。本研究ではこの iNTRoDB に生化学経路情報を統合し、この genome-wide protein list に既知の生化学経路情報を表示するインタフェースを開発し、実装する。

## 2.2 KEGG パスウェイ

本研究では生化学経路情報のデータセットとして無償で扱う事ができ、多くの様々な生物種についての生化学

経路情報を備えている Kyoto Encyclopedia of Genes and Genomes (KEGG)[10] を用いた。京都大学化学研究所の金久實教授らによって開発された KEGG は生物の代謝やシグナル伝達等の生化学経路をデータベース化し、パスイとして公開している。KEGG パスイは KEGG データベース上にある様々な生物の各代謝経路を図として参照する事が可能であり、単に図を表示するだけでなく、ユーザーが指定した条件に合致するパスイを検索したり、パスイ中のタンパク質を着色したりすることができるツールも実装されている。

KEGG パスイの例として、クルーズトリパノソーマ (*T. cruzi*) の解糖経路のパスイを図 3 に示す。図 3 の各長方形が反応の各段階を受け持つ酵素タンパク質の働く場所を示している。各長方形の数字は EC 番号 [11] と呼ばれ、その酵素タンパク質がどのような特徴を持つものかを分類するための番号である。薄い緑色になっている四角はタンパク質の詳細が判明しているもので、選択することでそのタンパク質の詳細ページへ移動する。全てのパスイに固有の ID が付与されている。

## 2.3 KEGG パスイデータの統合

iNTRoDB の各タンパク質の情報は TriTrypDB[12] からの情報であり、内部的ではこの DB 由来の ID によってタンパク質は管理されている。これは KEGG パスイ上のデータとは異なる ID であるため、それらの関連付けを行う必要がある。ただし各タンパク質について KEGG 上に TriTrypDB 上で扱われている ID である GeneDB ID と KEGG データベースで扱われている KEGG ID との対応情報は存在するため、これを利用し、タンパク質の GeneDB ID とそのタンパク質が関連するパスイ ID とを関連付けたリストを作成した。この際タンパク質 1 つにつき関連するパスイが複数存在する場合もあることから、GeneDB ID とパスイ ID は 1 対 1 対応とは限らず、KEGG 上では各タンパク質のデータに KEGG ID とその KEGG ID と対応するパスイ ID が登録されている事を利用し、各 ID 対応リストを作成した。また KEGG パスイ上では表示されるタンパク質は KEGG ID ではなく先述の通り EC 番号で表示される。



図 5 KEGG パスウェイリンクの拡大図

Fig. 5 Enlarged view of the KEGG pathway column.

表 1 KEGG pathway カラムの説明

Table 1 Description of KEGG pathway column.

アイコン	説明
パスウェイ ID	カーソルを乗せるとパスウェイの名前が表示される.
R	選択すると RIT-seq で着色されたパスウェイを表示する.
C	選択すると検索条件 Homology for ChEMBL に基づいて着色されたパスウェイを表示する.
P	選択すると検索条件 Homology for Structures (BLASTP) に基づいて着色されたパスウェイを表示する.
H	選択すると検索条件 Homology for Mammal に基づいて着色されたパスウェイを表示する.

これらのパスウェイ情報は図 4 のように genome-wide protein list 上の右端に KEGG pathway のカラムを追加した。パスウェイが判明していないものは「No hit」と表示し、パスウェイが判明しているものはパスウェイ ID とアイコンを 4 つ並べ表示するよう実装を行った。図 5 は新たに追加したカラムを拡大したもので、パスウェイ ID とアイコンの意味は表 1 に示すものであり、アイコンを選択することで、KEGG パスウェイの図を表示する。KEGG パスウェイは選択された条件に応じて、その経路上での全てのタンパク質の背景色を着色しており、iNTRADB の情報を KEGG パスウェイ上でも確認できるようにしている。これは KEGG Mapper の機能を利用することで実現しており、また直接選択されたタンパク質は背景色を黒とし、一目でわかるように実装を行った。

## 2.4 KEGG パスウェイへのタンパク質アノテーション情報のマッピング

KEGG には先述の通りいくつかのツールが存在し、その 1 つにパスウェイ中のタンパク質を着色する KEGG Mapper-Search&Color Pathway[13] というツールが存在する。これはどのタンパク質を何色に着色するかを指示することで関係する全てのパスウェイ中の指定したタンパク質に任意の着色が可能となるツールである。

本研究ではトリパノソーマ科の各タンパク質の生化学経

表 2 RIT-seq の結果とカラーコード

Table 2 Color code the results of the RIT-seq.

数字	カラーコード	意味
0-0-0-0	blue	全ての場合で減少が確認された.
0-0-1-0	red	血流型のみ減少が確認された.
1-1-0-0	green	プロサイクリック型のみ減少が確認された.
1-1-1-0	purple	血流型からプロサイクリック型への分化・成長の場合で減少が確認された.
1-1-1-1	gray	全ての場合で減少が確認されなかった.
other	white	上記に当てはまらない場合.

路情報を得るために、KEGG パスウェイ情報を genome-wide protein list へ追加し、さらにこの機能を用いてユーザーが入力した探索条件に応じて KEGG パスウェイ上のタンパク質に着色を行うことで代謝経路から新薬標的発見の可能性を見出すことを目指してインタフェースの開発・実装を行った。

## 2.5 着色する探索条件

着色する探索条件として、化合物データベース (ChEMBL) に登録されたタンパク質との相同性 (Homology for ChEMBL), タンパク質の立体構造情報を蓄積したタンパク質構造データバンク (PDB) に登録されたタンパク質との相同性 (Homology for Structures), 哺乳類に存在するタンパク質との相同性 (Homology for Mammal) を示す 3 つを選択した。

さらにトリパノソーマ科のゲノムデータのアノテーション情報としてフェノタイプ情報がある。これは RNAi target sequencing (RIT-seq)[8] と呼ばれ、実験で得られたブルーストリパノソーマ (*T. brucei*) の寄生形態毎に行ったフェノタイプ実験の結果を示したものである。*T. brucei* の寄生形態には 4 種類あり、メタサイクリック型、血流型、プロサイクリック型、上鞭毛型がある。実験では成長して 3 日以内の血流型を“BFD3”, 成長して 3 日以上 6 日以内の血流型を“BFD6”, プロサイクリック型を“PF”, 血流型からプロサイクリック型への分化・成長の過程を“DIF”と分類し、それぞれについて RNAi target sequencing を行っている。iNTRADB では表 2 に示すように RNAi target sequencing で *T. brucei* が減少した場合を 0, 減少が確認されなかった場合を 1 とし, “BFD3”-“BFD6”-“PF”-“DIF”の順で 4 桁の数字とカラーコードで結果を示している。本研究ではこの RIT-seq 情報についても着色したパスウェイを表示することを可能としている。

## 3. 生化学経路表示機能の使用例

本研究で実装した機能を使用例を用いて説明する。探索結果から見たい条件で着色されたパスウェイを選択し、着色されたパスウェイを閲覧する。例として図 6 で強調され



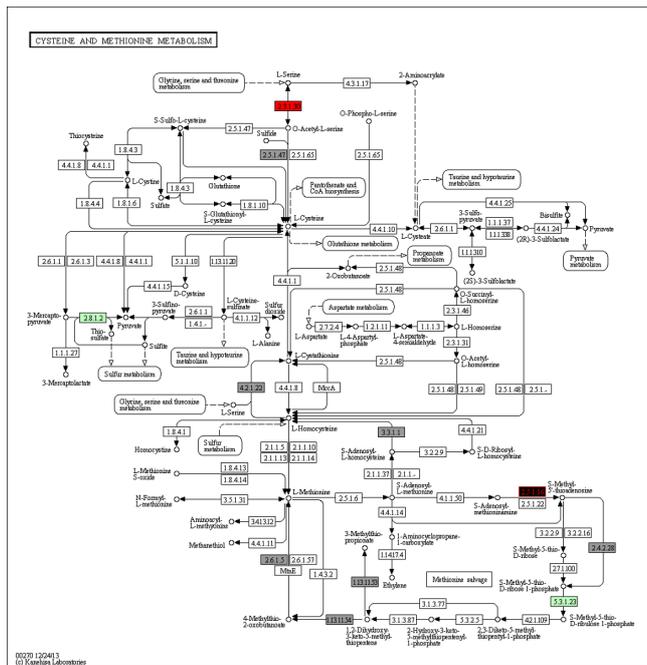


図 8 シスチン，メチオニン代謝 (RIT-seq によるカラーリング)  
Fig. 8 Cysteine, methionine metabolism (RIT-seq color).

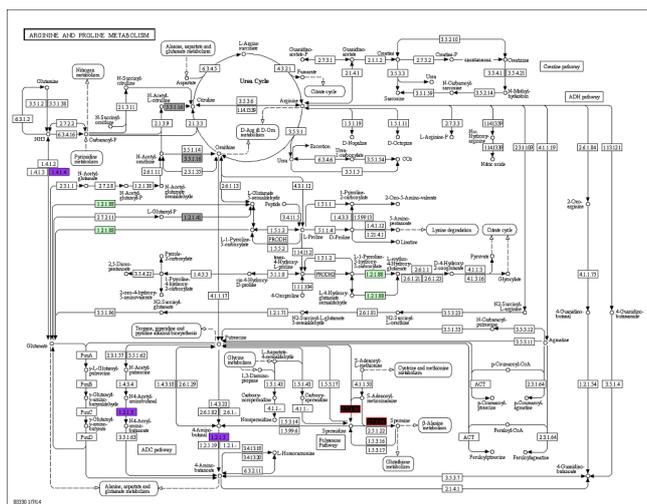


図 9 アルギニン，プロリン代謝 (RIT-seq によるカラーリング)  
Fig. 9 Arginine, proline metabolism (RIT-seq color).

質の発見の効率化を目指した。また本研究では iNTRODB に既知の生化学経路情報を単に統合しただけではなく、その情報をその他の創薬に係るの情報と同時に表示するインタフェースの開発を行った。さらに追加したインタフェースについて実際に既知の標的候補タンパク質を用いて有用性の検証を行った。その結果、残念ながら本研究で実装したインタフェースは直接標的候補タンパク質の発見にはつながらなかったが、従来のシステムでは得られなかった情報を視覚的に得ることは可能であり、候補となるタンパク質についての議論を行う際に有用なツールであると考えられる。

謝辞 iNTRODB の開発にあたり様々なご助言を賜りま

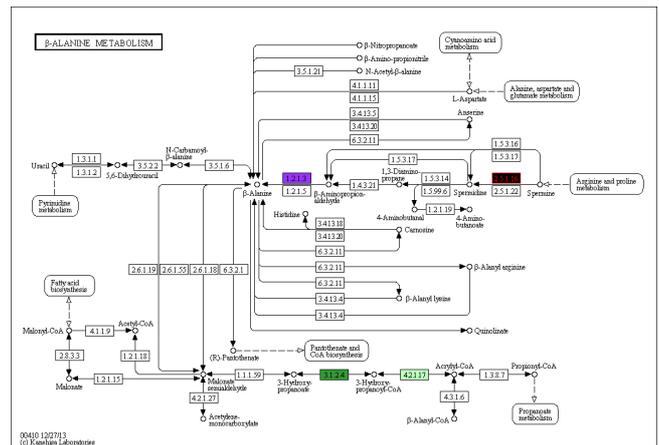


図 10  $\beta$ -アラニン代謝 (RIT-seq によるカラーリング)  
Fig. 10  $\beta$ -alanine metabolism (RIT-seq color).

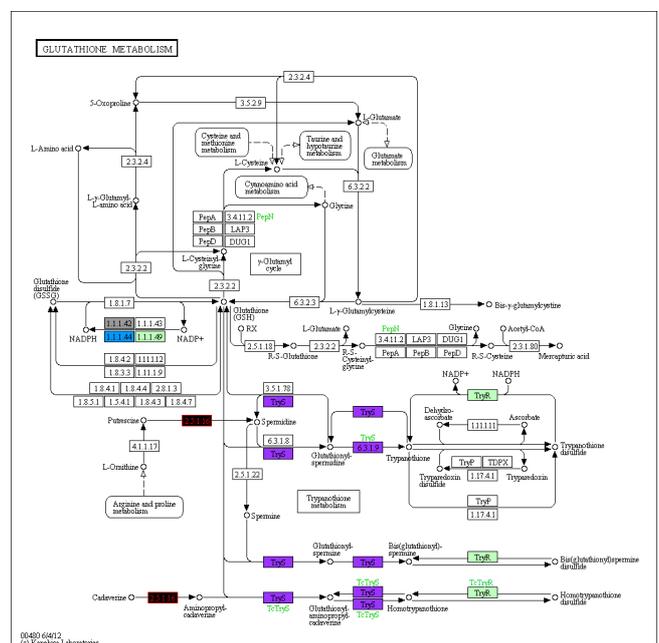


図 11 グルタチオン代謝 (RIT-seq によるカラーリング)  
Fig. 11 Glutathione metabolism (RIT-seq color).

した、東京大学 大学院医学系研究科 北潔教授、稲岡健ダニエル助教、アステラス製薬株式会社 折田正弥博士、大野一樹博士に心より御礼申し上げます。

#### 参考文献

- [1] 一盛和世: “顧みられない熱帯病対策-地球規模での現状と日本の貢献”, 2008.
- [2] “オールジャパンで NTDs 創薬に挑む”, *Astellas Square No.52*, No.5, pp.40-42, 2013.
- [3] Chand P., Babu Y. S., Bantia S., et al.: “Design and Synthesis of Benzoic Acid Derivatives as Influenza Neuraminidase Inhibitors Using Structure-Based Drug Design.”, *Journal of medicinal chemistry*, 40(25), pp.4030-4052, 1997.
- [4] 八木崇, 大久保昌美: “医薬品開発の機関と費用”, *JPM News Letter*, No.136, pp.33-35, 2010.
- [5] iNTRODB のホームページ,

<http://www.bi.cs.titech.ac.jp/introdb/>

- [6] Gaulton A., Bellis L. J., Bento A. P., *et al.*: “ChEMBL: a large-scale bioactivity database for drug discovery.”, *Nucleic acids research*, 40(D1), pp.D1100-D1107, 2012.
- [7] Sussman J. L., Lin D., Jiang J., *et al.*: “Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules.”, *Acta Crystallographica Section D: Biological Crystallography*, 54(6), pp.1078-1084, 1998.
- [8] Alsford S., Turner D.J., Obado S.O., *et al.*: “High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome.”, *Genome research*, 21(6), pp.915-924, 2011.
- [9] Agüero F., Al-Lazikani B., Aslett M., *et al.*: “Genomic-scale prioritization of drug targets: the TDR Targets database.”, *Nature Reviews Drug Discovery*, 7(11), pp.900-907, 2008.
- [10] Kanehisa M., Goto S.: “KEGG: kyoto encyclopedia of genes and genomes.”, *Nucleic acids research*, 28(1), pp.27-30, 2000.
- [11] Webb E. C.: “Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes (No. Ed. 6).”, *Academic Press*, 1992.
- [12] Aslett M., Aurrecoechea C., Berriman M., *et al.*: “TriTrypDB: a functional genomic resource for the Trypanosomatidae.”, *Nucleic acids research*, 38(suppl 1), pp.D457-D462, 2010.
- [13] Kanehisa M., Goto S., Sato Y., *et al.*: “KEGG for integration and interpretation of large-scale molecular data sets.”, *Nucleic acids research*, 40(D1), pp.D109-D114, 2012.
- [14] Alves-Ferreira M., Guimaraes A. C. R., Capriles P. V. D. S. Z., *et al.*: “A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information.”, *Memorias do Instituto Oswaldo Cruz*, 104(8), pp.1100-1110, 2009.
- [15] Heby O., Persson L., Rentala M.: “Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of African sleeping sickness, Chagas’ disease, and leishmaniasis.”, *Amino acids*, 33(2), pp.359-366, 2007.
- [16] Heby O., Roberts S. C., Ullman, B.: “Polyamine biosynthetic enzymes as drug targets in parasitic protozoa.” *Biochemical Society Transactions*, 31(2), pp.415-419, 2003.