

生化学反応ネットワーク 統合解析環境の拡張

西田 孝三

理化学研究所・生命システム研究センター(QBiC)

内容

1. 研究背景・目的
2. 研究開発の内容
3. ユースケース

(ゲノムスケールの)生化学反応ネットワーク解析の流れ

- 定量生物学
- 分析化学

- 統計、機械学習

- データベース、オントロジー
- ゲノムアノテーション

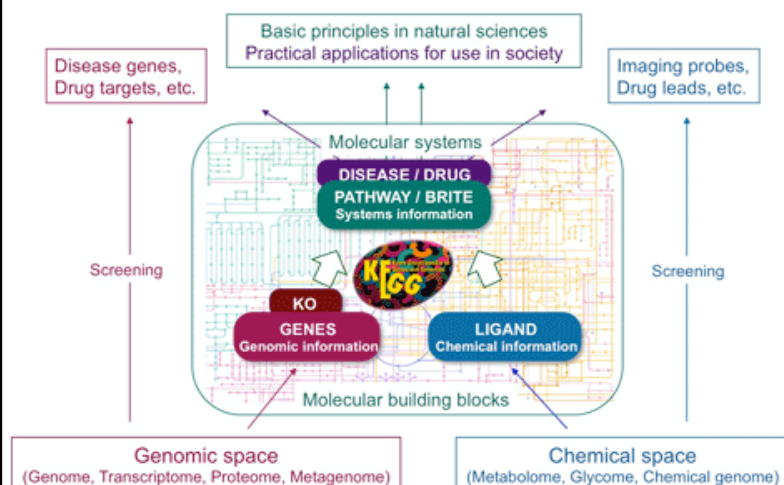
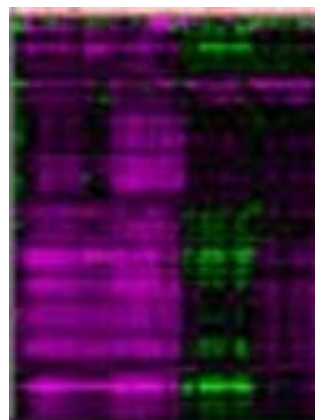
- ネットワーク
バイオロジー

- トランスクリプトーム (GeneChip, RNA-seq)
- メタボローム(LC, GC-MS)

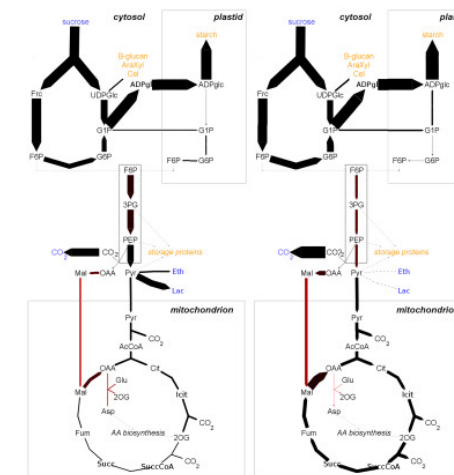
- 前処理(正規化)
- [統計]解析(有意差、相関、機械学習...)

反応ネットワーク(パスウェイ)への解析結果のマッピング(統合)・可視化

- パスウェイ比較・解析・可視化
- シミュレーションの制約条件への利用 (FBA...)



(www.genome.jp)



(Rohn, 2012)

(ゲノムスケールの)生化学反応ネットワーク解析の流れ

問題提起

- オミクス解析結果マッピング対象のパスウェイデータにおいて異なるデータベース間の情報を統合するソフトウェアが無い(前年度)
- 生化学反応ネットワーク解析の横の流れをつなぐ汎用性の高い統合環境が無い(本年度)

前年度研究開発における統合

大腸菌の前処理済み
大規模発現プロファイル
(E COLI EXPRESSION2,
<http://systemsbiology.ucsd.edu/InSilicoOrganisms/Ecoli/EcoliExpression2>)

詳細な大腸菌の代謝モデル
(iJO1366)

+

大腸菌のドラッグターゲット
(DRUGBANK)

+



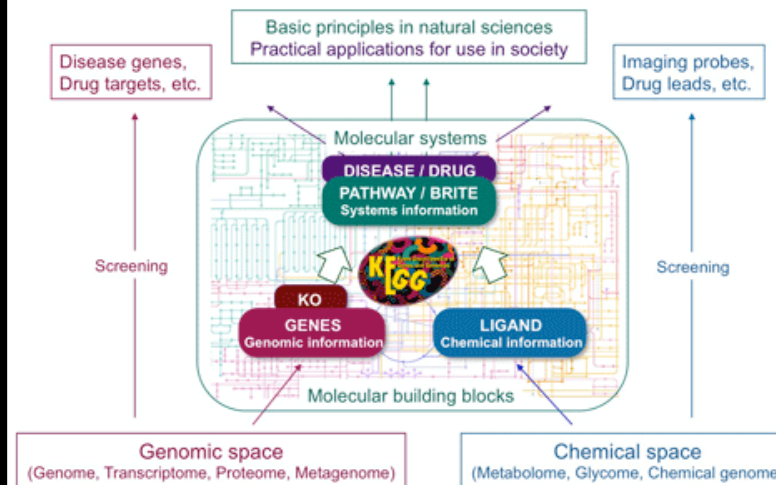
Cytoscape拡張機能
KEGGscapeの作成と
活用

残された課題

汎用性の向上

- 複数の生物種、オミクスデータ種に対応
- オミクス解析ワークフローとの統合

その上で一連のフローを通して
行える環境(Cyberinfrastructure)
への昇華



(www.genome.jp)

KEGG情報のみからは
得られない知見

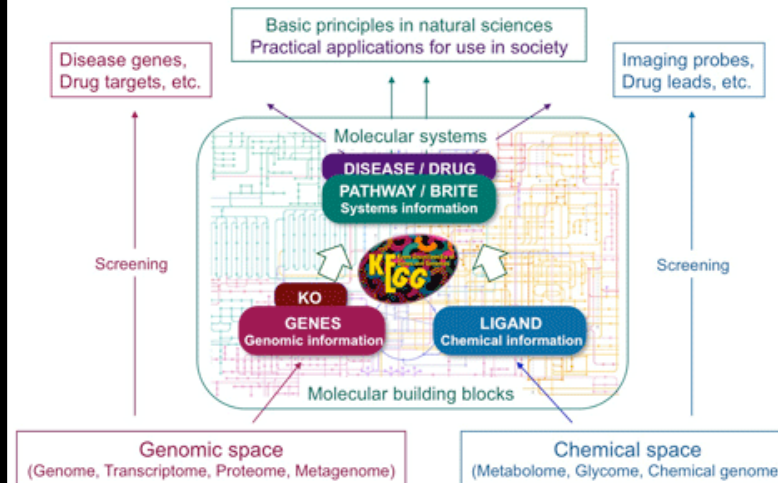
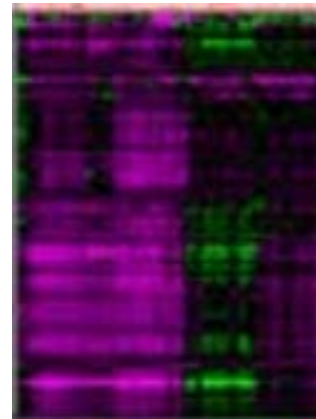
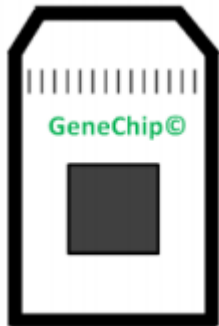
- ターゲット既知、
薬理作用未知の
ドラッグによる
代替反応経路の推定

今年度目指すもの

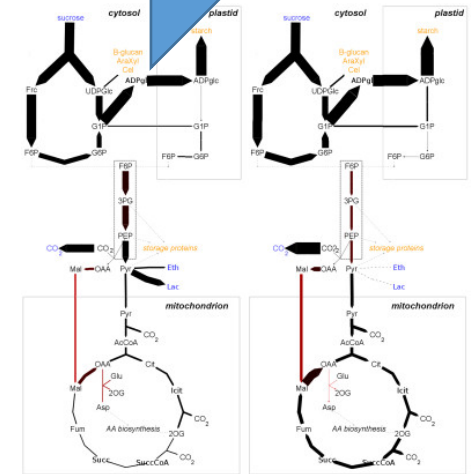
汎用的かつ各解析を最終結果まで一気通貫で行える
統合解析環境(Cyberinfrastructure)の実現

- メタボローム(LC, GC-MS)

遺伝子、タンパク質、機械学習...



(www.genome.jp)



(Rohn, 2012)

Cyberinfrastructureと開発の概要

IP[y]: Notebook UIとして採用、ウェブブラウザから利用可能

入力
生オミクス
データ

[再現、共有、拡張、インタラクティブ]性の高い解析環境

出力
データ統合
反応ネットワーク
の可視化



- 採用で得られる利点
- フレームワークとして成熟しており、組み込みによって汎用性が向上
 - 豊富なオミクス解析ワークフロー資源が得られる

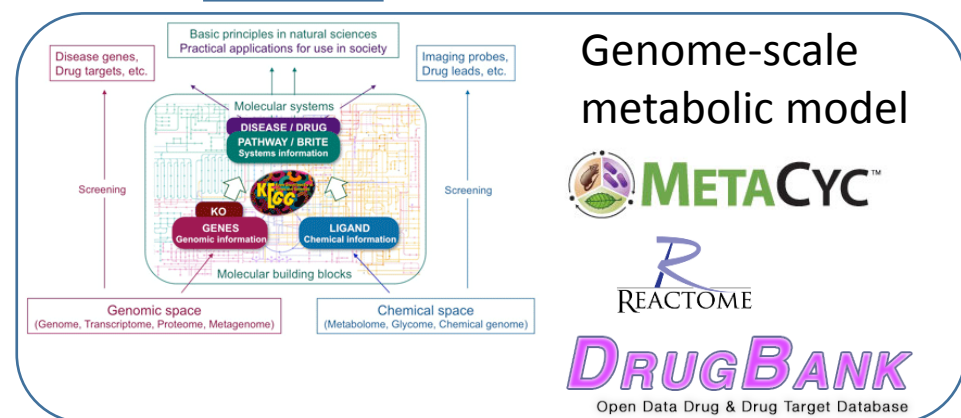


パスウェイ情報の統合

- KEGGのみからは得られない知見

RDFの利用

- 統合プロセスの汎用化



RESTインターフェース



既存研究とは異なり、Cytoscapeは可視化に徹し操作は外部プログラムからRESTで行う

- プログラム改変性・汎用化の向上

の統合や機能
モジュール間の橋渡し
をR, Python言語の
パッケージとして実装

データについて

- 今回はまず大腸菌ではなくシロイヌナズナを対象とした開発に注力
 - 大腸菌に関しては昨年度成果でパスウェイデータの統合の実績があり、シロイヌナズナにもほぼ同様の統合対象データが存在
- メタボロームプロファイルを収集したデータベースが存在する
(前年度採択者、福島敦史氏による成果
AtMetExpress[<http://atmetexpress.riken.jp/>])
 - メタボロームは機器によるレンジの違いがあり、また一度の実験で同定できる代謝物の数が少ない(約100個ほど) -> 通常データでは反応ネットワーク視点の知見獲得が困難
 - 「プロファイル」を収集し、再解析することを可能にした数少ない例(代謝物総数約1200)
- トランスクリプトーム、メタボロームの統合解析の実現をねらう

実施する研究項目とスケジュール

研究開発項目	平成26年				平成27年	
	9月	10月	11月	12月	1月	2月
1.. パスウェイ情報の RDF を利用した統合		←→				
2. Bioconductorを用いた解析とCytoscape上での可視化のパイプライン化			←→			
3. ウェブブラウザUIの実装					←→	
4. 広報用ウェブサイト、論文の作成						←→

ユースケース

メタボロームプロファイルを用いた Enrichment Analysis



上記の流れを一気通貫で

- 容易に
- 解析内容を変更可能に
- 結果を表や数値ではなくネットワーク上で見える形で出力を行う

メタボロームプロファイルを用いた Enrichment Analysis

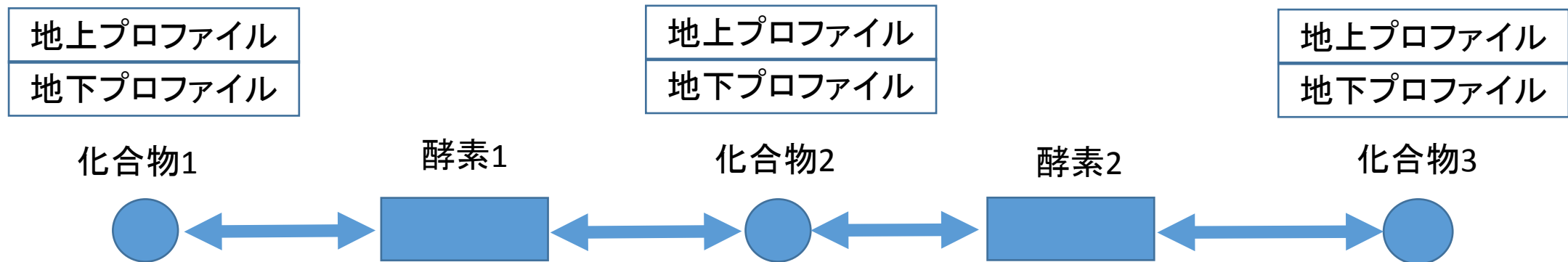


- ヒト以外の生物種で代謝物蓄積量のEnrichment Analysisが行われることは稀
- トランスクリプトーム解析でも共通
- AtMetExpressにはサンプル数の多い実験が含まれている

メタボロームプロファイルを用いた Enrichment Analysis



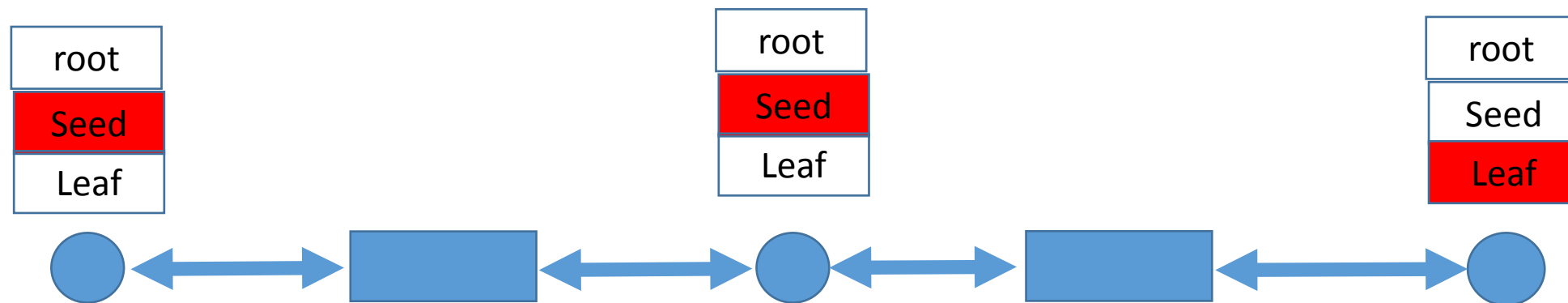
- KEGGパスウェイ上で組織毎の発現、model間の差異を視認できる



メタボロームプロファイルを用いた Enrichment Analysis



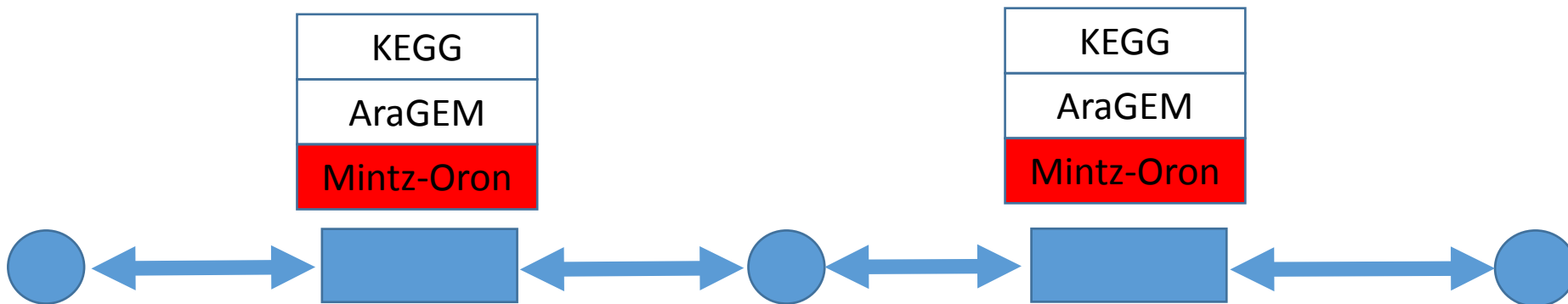
- KEGGパスウェイ上で組織毎の発現、model間の差異を視認できる



メタボロームプロファイルを用いた EnrichmentAnalysis



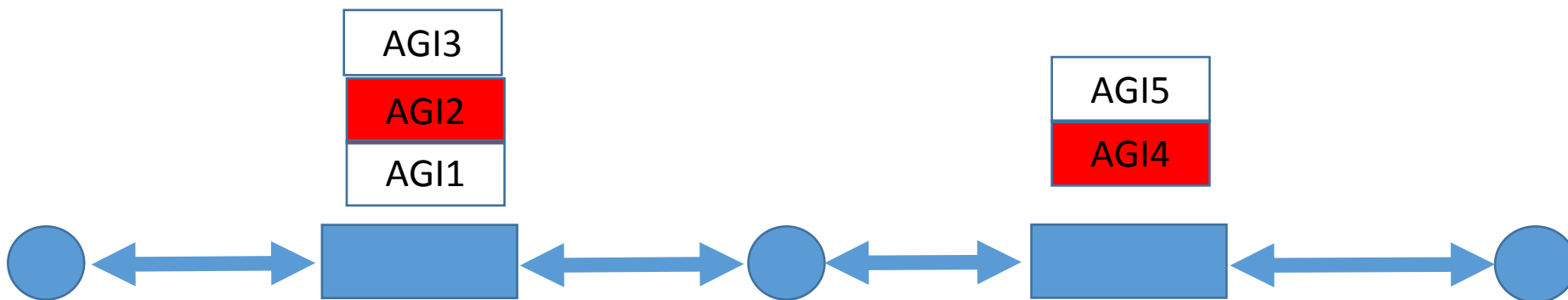
- モデル間アノテーション差異の可視化



メタボロームプロファイルを用いた Enrichment Analysis



- 同様にアイソフォーム遺伝子の発現の可視化へも適用可能 (トランスクリプトーム情報利用時)



メタボロームプロファイルを用いた Enrichment Analysis



- RDFの利用に基づいたEnrichment Analysisに用いるオントロジーの柔軟な切り替え
 - 通常はGO、KEGGパスウェイ
 - Linkdb RDFを利用した
 - 他DBのオントロジー (ChEBI)利用
 - 隣接2 KEGGパスウェイ

謝辞 (敬称略)

- 瀬々潤 (産総研CBRC)
- 福島敦史 (理研CSRS)
- 大野圭一郎 (UCSD)